Volume 17 Number 23 1989 Nucleic Acids Research

An enlarged largest subunit of *Plasmodium falciparum* RNA polymerase II defines conserved and variable RNA polymerase domains

Wu-Bo Li, David J.Bzik, Haoming Gu, Manami Tanaka, Barbara A.Fox and Joseph Inselburg

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

Received September 11, 1989; Revised and Accepted November 1, 1989 EMBL accession no. X16561

ABSTRACT

We have isolated the gene encoding the largest subunit of RNA polymerase II from *Plasmodium falciparum*. The RPII gene is expressed in the asexual erythrocytic stages of the parasite as a 9 kb mRNA, and is present as a single copy gene located on chromosome 3. The *P. falciparum* RPII subunit is the largest (2452 amino acids) eukaryotic RPII subunit, and it contains enlarged variable regions that clearly separate and define five conserved regions of the eukaryotic RPII largest subunits. A distinctive carboxyl-terminal domain contains a short highly conserved heptapeptide repeat domain which is bounded on its 5′ side by a highly diverged heptapeptide repeat domain, and is bounded on its 3′ side by a long carboxyl-terminal extension.

INTRODUCTION

Transcription in eukaryotes is directed by three classes of nuclear RNA polymerases (1). The genes encoding the largest subunits of eukaryotic RNA polymerases I (RPI) (2), II (RPII) (3−8), and III (RPIII) (3) have been isolated and are single copy genes with the exception of *Trypanasoma brucei* RPII, which contains two alleles (7,8). RPII is a complex enzyme consisting of multiple subunits, and is responsible for the transcription of protein coding genes. DNA sequence analysis of the cloned largest RPII subunit from different species (3−8) revealed nucleotide and amino acid homology with the largest subunit of *E. coli* RNA polymerase (β′) (9), yeast RPI (2), and yeast RPIII (3). The largest subunit of eukaryotic RPII (to be subsequently referred to as RPII subunit) contains a repeated heptapeptide sequence in its carboxyl-terminal domain (CTD), which is repeated 52 times in mouse (4), 42 times in Drosophila (6), 27 times in yeast (3), and is absent only in the *T. brucei* RPII subunit (7,8). The CTD structure is specific for the RPII subunit and is essential for cell viability (5,10). The CTD may function in initiation of transcription by mediating phosphorylation and destabilization of histone and DNA interactions, thus facilitating transcription through nucleosomes (3,4,11,12,13). Alternatively, the CTD may function as a receptor for essential transcription factors, or to anchor the RNA polymerase II to a structure within the nucleus (14).

Little is known about transcription in *P. falciparum*, or its regulation. RNA synthesis in *P. falciparum*, during the 48 hour intraerythrocytic growth cycle *in vitro* (15), starts within 6 hr after infection of red blood cells (RBC), rapidly rises during the early trophozoite stage and peaks during the schizont stage (16). Many species of the parasite's proteins, including most of the known parasite antigens, are synthesized in synchronously infected cultures during the trophozoite and schizont stages [25−42 hr after infection of RBC] (16).

We are initiating a molecular approach to the study of transcription in *P. falciparum* with this report of the isolation, sequence, and structure of the gene that encodes the RPII

subunit of *P. falciparum*. This is the first characterization of a transcription factor for *P. falciparum*. We describe the determination of the RPII gene copy number, its chromosomal location, its expression during erythrocytic growth, and identify conserved regions of the protein that are separated by novel enlarged variable regions.

MATERIALS AND METHODS

Cells, DNA, and vectors

P. falciparum strains FCR3 (17) and Honduras-1 (18) were used in the present studies. Parasites were grown in RPMI 1640 medium (15,18) prepared with HEPES buffer (pH 7.2), 0.2% sodium bicarbonate, 10% heat-inactivated human type A, Rh^+ fresh frozen plasma, penicillin (100 IU/ml), streptomycin (100 mg/ml), and gentamycin (20 mg/ml). α-Amanitin was purchased from Boehringer Mannheim Biochemicals. Honduras-1 DNA was isolated from asynchronous cultures of parasites by lysis in a GuITC solution (4 M guanidium isothiocyanate, 5 mM sodium citrate (pH 7.0), 0.1 M β-mercaptoethanol, and 0.5% Sarkosyl) and centrifugation to separate the RNA by pelleting. The DNA was collected and was purified by banding in CsCl (19). The cloning vectors λgt11, λZAP, and pBluescript were obtained from Stratagene, Inc.

Oligonucleotides

Four oligonucleotide probes, A, B, C, and D were synthesized and used in hybridization experiments. The sequence and location of the probes were as follows: probe A (a mixture containing 16,384 35-mers) 5′-GGA(orT)C(orG)AATAA(orT)G(orC)AA(orT)GGA(orT)G(orC)AA(orT)GTA(or T)GGA(orT)G(orC)AATAA(orT)G(orC)AA(orT)GG (based on the consensus heptapeptide repeat region); probe B (a 30-mer) 5′-ACCTCTATCATTAGCAA CTAAGTGTTCTTC (nucleotide 4860 to 4831); probe C (a 30-mer) 5′-TCCAAACAAAGCCGATGTATTAGAATCACC (nucleotide 5190 to 5161); and probe D (a 22 mer) 5′-TTCTACAGAATGATCACATGCT (nucleotide 462 to 441). The oligonucleotides were treated in concentrated NH_4OH at 55°C for 5 hr, lyophylized, and purified by NENSORB™ PREP (DuPont). The oligonucleotides were end-labeled according to standard procedures (19).

Isolation of cDNA and genomic DNA clones

The cDNA library in λgt11 was constructed from trophozoite and schizont stage Honduras-1 mRNA (20). Honduras-1 genomic EcoRI and XbaI libraries were constructed in λgt11 and λZAP as previously described (21). The DraI genomic DNA library was constructed in pBluescript. Honduras-1 genomic DNA was digested to completion with DraI, electrophoresed on a 1.0% agarose gel, and DNA of 400 to 1000 bp was collected and purified (22). The DraI digested size-fractionated DNA was blunt-end ligated into SmaI digested pBluescript. End-labeled probe A was hybridized to replica filters of λ phage plaques in 6× SSC at 42°C and washed in 1× SSC at 45°C. All of the clones isolated, C1, C3, and C8, gave very strong hybridization signals. A specific 169 bp DpnI restriction fragment from the 5′ part of cDNA clone, C8, was cut out of a 1.0% agarose gel and oligo-labeled to a specific activity of 1×10^9 cpm/μg (23). Filters of cDNA phage plaques were hybridized with the 169 bp DpnI probe in 6× SSC at 42°C and washed in 1× SSC at 60°C. All of the 14 cDNA clones isolated gave strong hybridization signals and were overlapping cDNA clones. Probe B and C were end-labeled and used to screen filters of genomic DNA containing phage plaques by hybridization in 6× SSC at 42°C and washing in 1× SSC at 50°C. Probe D was end-labeled and used to screen filters containing colonies from the DraI genomic DNA library by hybridization in 6× SSC at 37°C and washing

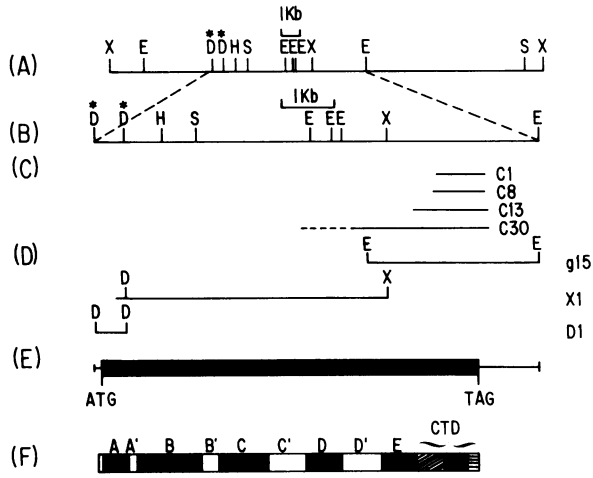


Figure 1. Restriction maps, cDNA, and genomic DNA clones of the *P. falciparum* RPII gene. Restriction sites shown (A) are E = EcoRI, H = HindIII, S = SpeI, X = XbaI, and D = DraI. Only two of the twenty DraI sites are shown (*). Enlarged restriction map encompassing the RPII subunit gene (B). Restriction sites are the same as in (A). Location of cDNA clones C1, C8, C13, and C30 (C). The dashed line on cDNA C30 indicates a region of sequence non-colinearity with the RPII gene due to a double ligation. Location of genomic DNA clones g15, X1, and D1 (D). Location of the RPII subunit open reading frame (E). The ATG start, and TAG termination codons are shown. Domains of the RPII subunit (F). The five conserved regions, A, B, C, D, and E, of the RPII subunit are indicated as solid blocks. The enlarged variable regions, A′, B′ C′, and D′, are indicated as open blocks. The CTD is represented as three domains (see text), where the middle region (solid block) represents the heptapeptide repeat domain.

in 1× SSC at 45°C. All of the clones isolated with probes A, B, C, D, and the 169 bp DpnI fragment were determined to be RPII gene clones by DNA sequence analysis.

DNA sequencing and computer analysis

The cDNA and genomic DNA clones were sequenced as previously described (24). Both DNA strands were completely sequenced using the dideoxy chain termination method (25). The complete DNA sequences were established using the DNA Inspector II programs (Textco Inc., Lebanon, NH). Amino acid sequence comparisons with other RNA polymerase subunits was done by a dot matrix analysis using the MacGene Plus™ program. The prediction of secondary structure of the RPII subunits was performed using the IBI-Pustell programs (International Biotechnologies Incorporated), and the MacGene Plus™ programs.

Northern blot analysis

Total *P. falciparum* Honduras-1 RNA was prepared as previously described (20) from synchronized cultures of trophozoite and schizont stage parasites. Poly(A^+) RNA was purified from total RNA using an oligo d(T)-cellulose column (19). Total poly(A^+) RNA [mRNA] (4 μg per lane) was electrophoresed in a 1.2% agarose-formaldehyde gel, blotted to Zetabind nylon membrane (CUNO, Inc), and hybridized to oligo-labeled cDNA C1 as previously described (24,26). Washing was done in 0.1× SSC at 65°C. cDNA C1 is specific for the RPII gene.

Pulsed field gradient (PFG) electrophoresis of chromosomes

PFG electrophoresis (27) was performed using a contour clamped homogeneous electric

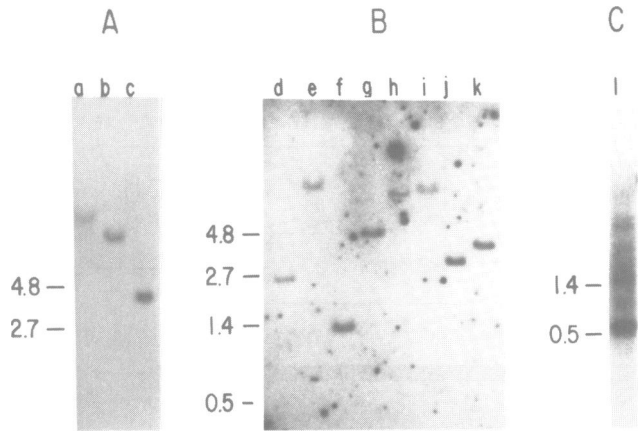


Figure 2. Genomic southern analysis of the *P. falciparum* RPII gene. *P. falciparum* genomic DNA (5 μg per lane) was digested with restriction enzymes SpeI (a), XbaI (b and e), EcoRI (c), BclI (d), XbaI/BclI (f), XbaI/HindIII (g), XbaI/EcoRI (h), XbaI/ClaI (i), XbaI/BglII (j), XbaI/SpeI (k), and DraI (l). The digested DNA was electrophoresed on a 1% agarose gel, blotted to nylon, and hybridized with probe A (A), a mixture of probe B and C (B), or probe D (C). Each probe was hybridized in 6× SSC at 42°C, and was washed in 1× SSC at 45°C (probe A), 50°C (probe B and C), or 45°C (probe D). The sizes of the molecular weight markers are indicated in kb.

field (CHEF) apparatus (28). Agarose blocks containing parasite chromosomes were inserted into a low melting point agarose gel (0.6%) as described (29). Briefly, the electrophoresis was done in 0.75×TBE (67 mM Tris, 67 mM boric acid, and 1.5 mM EDTA, pH 8.0) at 14°C for 48 hr using a 3 min pulse time, and an 8 V/cm field strength. Separated chromosomes were transferred to a Zeta probe nylon membrane (BioRad), and hybridized with the RPII gene specific probe, cDNA C1, in 6× SSC at 42°C, and was washed in 1× SSC at 65°C.

RESULTS AND DISCUSSION

Cloning and sequencing of the P. falciparum RPII gene

A 35-base oligonucleotide mixture (probe A) was synthesized, based on the consensus carboxyl-terminal heptapeptide repeat sequence of the eukaryotic RPII subunits, and used to detect the *P. falciparum* RPII gene. A cDNA library was screened with probe A and clones, C1, C8 and C13 were isolated and sequenced. C1, C8, and C13 cDNA inserts were 800 bp, 1 kb and 820 bp, respectively (Fig. 1C.), and contained overlapping DNA sequences. Fourteen additional cDNA clones were selected with a 169 bp DpnI fragment derived from the 5′ region of cDNA C8. The largest cDNA clone isolated, C30, contained a 3.5 kb cDNA insert (Fig. 1C.).

The difficulty of isolating large cDNA inserts compelled us to begin selecting clones from genomic DNA libraries. This approach is reasonable in *P. falciparum* because relatively few of the characterized protein coding genes contain introns. In addition, *P. falciparum* introns are always short sequences (107 to 430 bp) that are easily identified by their characteristic structure(s) (24). A unique 3.8 kb genomic EcoRI fragment hybridized with probe A (Fig. 2A.). Six λgt11 genomic EcoRI clones, including clone g15 (Fig. 1D.), were selected with probe A and contained the same 3.8 kb EcoRI insert. The DNA

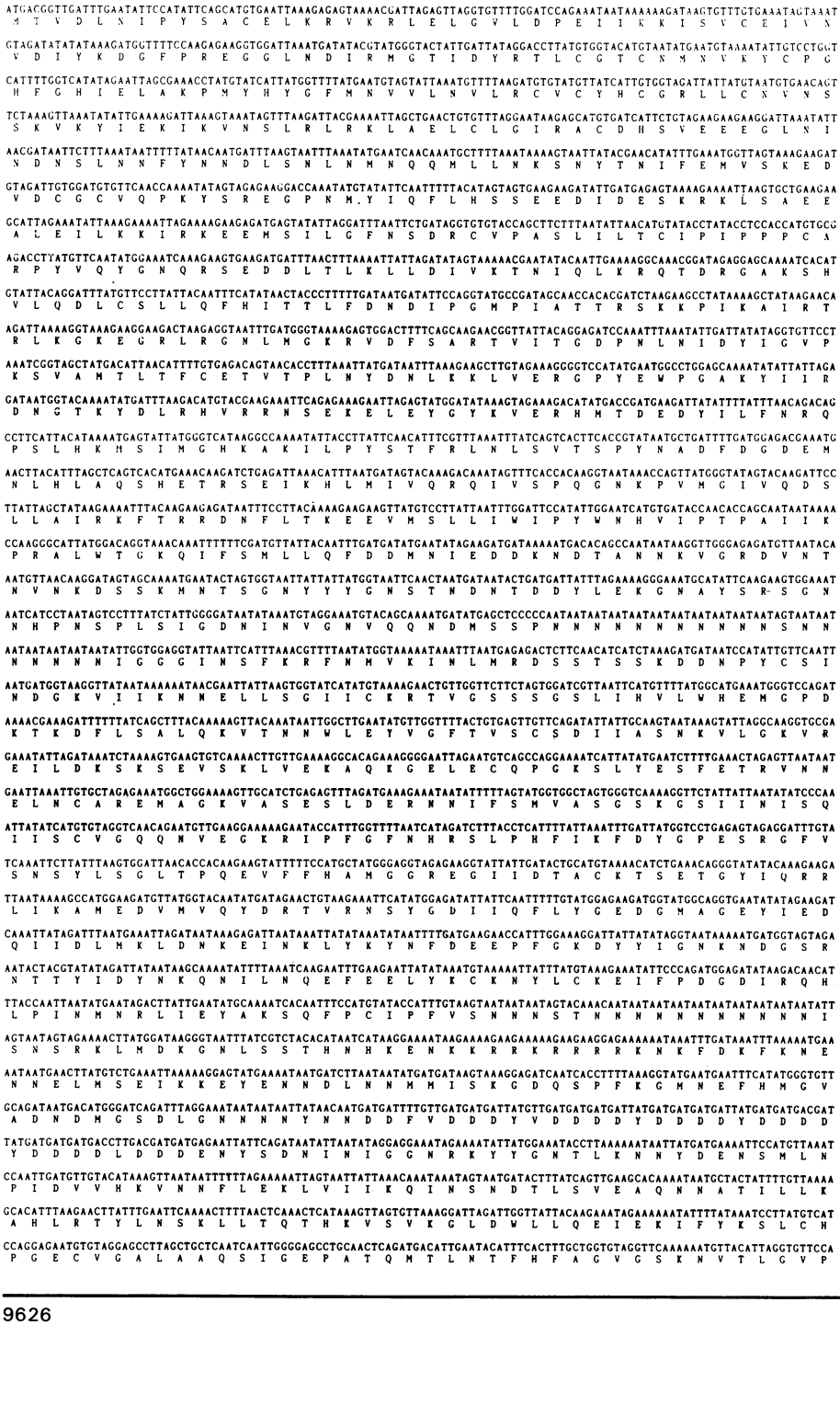
sequences of clone g15 and the cDNA clones were colinear (Fig. 1C. and 1D.). Oligonucleotide probes B and C, derived from the 5′ region of clone g15, hybridized with a 12 kb genomic XbaI fragment (Fig. 2B.). Probe B and C selected a clone, X1, from a XbaI genomic DNA library. Clone X1 contained the 3′, 5.3 kb, sequence of the original 12 kb fragment, and had deleted the 5′, 6.7 kb (Fig. 1D.) [Gene flanking sequences of *P. falciparum* commonly delete in recombination deficient *E. coli* (21).]. Oligonucleotide probe D, derived from the 5′ region of clone X1, identified a 580 bp genomic DraI fragment (Fig. 2C.). Probe D selected clone D1, which contained 580 bp (Fig. 1D.), from a DraI genomic DNA library.

The restriction maps of HindIII, EcoRI, SpeI, and XbaI for the genomic DNA clones spanning clones D1, X1, and g15 are shown (see Fig. 1A. and 1B.). The overlapping genomic DNA clones (g15, X1, and D1) were sequenced and collectively represented 8631 bp (Fig. 3.) [The 5′ 80 bp of clone D1 are not shown]. A large open reading frame (Fig. 1E.) begins with the ATG at bp 1 and ends at the TAG at bp 7357 (Fig. 3.), and the deduced protein contains 2452 amino acids. The distribution of A + T content was typical for a *P. falciparum* coding region (30), as the 5′ and 3′ flanking sequences contained higher A + T content [90% for the 5′ flanking region (data not shown) and 83% for the 3′ region] than the region encoding the long open reading frame (72%). Introns were not considered to be present in the long open reading frame based on the observations that all characterized *P. falciparum* introns (i) have a minimum A + T content of 85%, (ii) have no intron-length open reading frame, (iii) are between 107 to 430 bp long, and (iv) have a highly conserved 5′ and 3′ junction sequence (24).

Amino acid sequence comparisons with other RNA polymerase subunits

Amino acid comparisons of the 2452 amino acid *P. falciparum* protein with various RNA polymerase largest subunits from different species was done by a dot matrix analysis (Fig. 4.). Because the two RPII subunits of *T. brucei*, RPIIA and RPIIB, differ by only 4 amino acid substitutions over 1765 amino acids (7), only the RPIIA subunit was compared. The homology of the 2452 amino acid *P. falciparum* protein to the mouse, Drosophila, yeast, and *T. brucei* RPII subunits was significantly higher than the homology to *E. coli* RNA polymerase (β' subunit), and the largest subunit of yeast RPI and RPIII. This observation, along with the presence of the heptapeptide repeat in the CTD of the *P. falciparum* protein (Fig. 3.), shows that the gene we characterized is a form of the *P. falciparum* RPII subunit. The homologous sequences of the *P. falciparum* RPII subunit and the other RPII subunits were partitioned into only 5 colinear regions (A through E), followed by the CTD. The heptapeptide repeat is absent in both RPII genes of *T. brucei* (7,8). Each conserved region (A through E) of the other RPII subunits was shifted to the carboxyl-terminus (right) when compared with the *P. falciparum* RPII subunit. The shifts show that the *P. falciparum* RPII subunit is the largest characterized RPII subunit, and that it contains an enlarged variable region [regions A′ to D′ (Fig. 1F. and Fig. 4.)] between each of the conserved regions. The longest isolated cDNA clone, C30, included most of variable region D′ (Fig. 1C), demonstrating that this variable region is expressed. The presence of the enlarged variable regions helps to clearly define the conserved regions of the eukaryotic RPII subunits. Previous comparisons indicated that the RPII subunits could be divided into 6, 7, or 8 conserved domains (6,7,31), depending on the specific criteria used for the interpretation of those alignments.

The five conserved regions (A to E) consisted of 148, 416, 312, 239, and 195 amino acids, respectively (Fig. 5.). The conserved regions (total of 1310 amino acids) of the



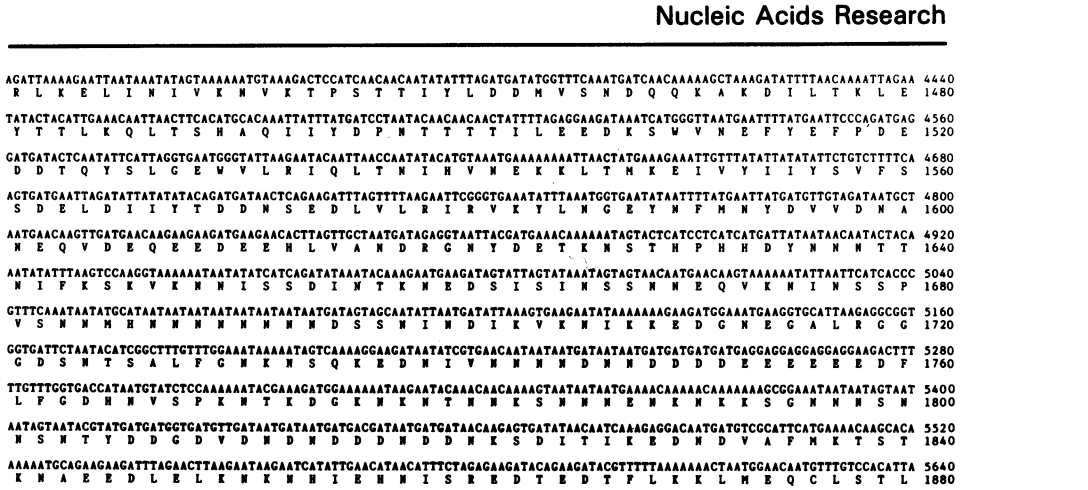


Figure 3. Nucleotide and predicted amino acid sequence of the *P. falciparum* RPII gene. The complete nucleotide sequence of the coding and 3′ flanking portion of the RPII gene is shown. The ATG start codon begins at nucleotide 1 and the RPII gene open reading frame ends at nucleotide 7357. Notable features of the RPII gene nucleotide and amino acid sequence are detailed in the text according to the nucleotide and amino acid numbering shown. The RNA Pol II sequence has been deposited with the EMBL data library (accession number X16561).

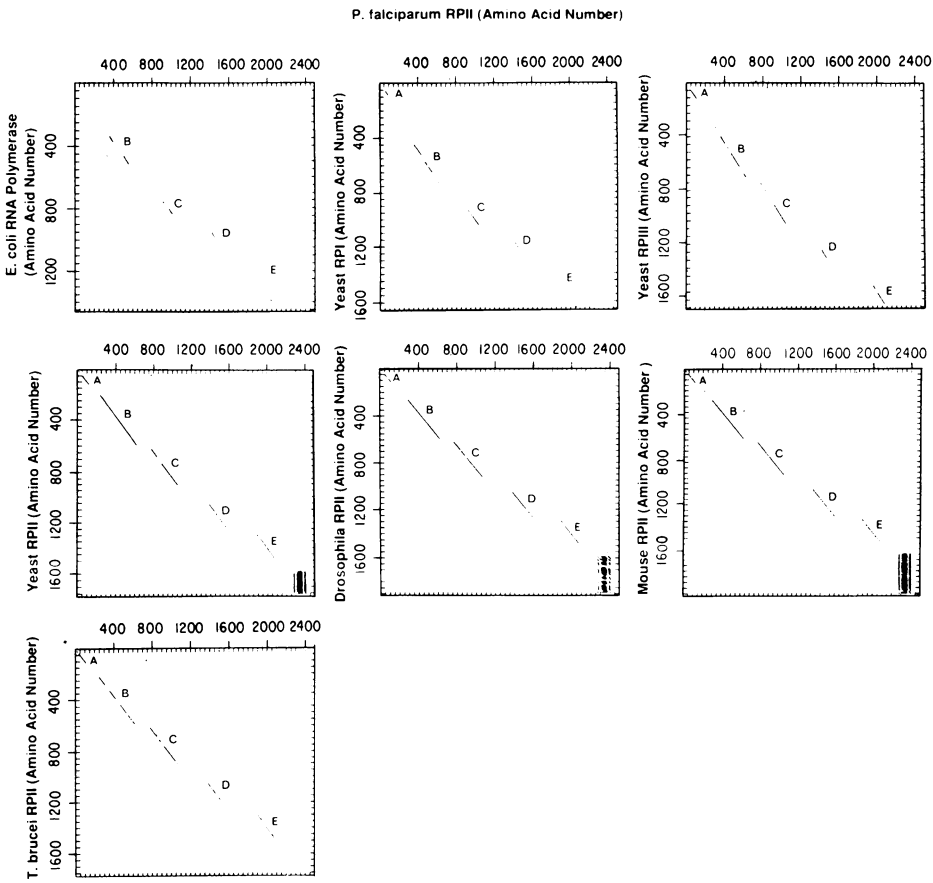


Figure 4. Amino acid comparison of the *P. falciparum* RPII subunit with other RNA polymerase subunits. The complete amino acid sequence of the *P. falciparum* RPII subunit (2452 amino acids), represented by the horizontal axis (top), was aligned with the amino acid sequence of other RNA polymerase largest subunits, represented by the vertical axis (left). The subunits aligned to the P. falciparum RPII subunit were *E. coli* RNA polymerase (β' subunit) (1407 amino acids), yeast RPI large subunit (1664 amino acids), yeast RPIII large subunit (1460 amino acids), yeast RPII subunit (1726 amino acids), Drosophila RPII subunit (1896 amino acids), mouse RPII subunit (1932 amino acids), and *T. brucei* RPII subunit (RPIIA allele) (1765 amino acids). A dot was plotted in the matrix if at least 9, or more, amino acid residues matched when windows of 15 amino acids were compared between the two sequences. The five colinear and conserved regions of the RPII subunits are marked as regions, A, B, C, D, and E. The unmarked vertical boxes at the CTD in some of the comparisons represents a region of homology between the RPII subunits that contain the conserved heptapeptide repeat domain. The different vertical sizes of the CTD homology box shows that the heptapeptide repeat domains are of different sizes.

P. falciparum RPII subunit shared 54% amino acid homology with mouse, 52% with Drosophila, 49% with yeast, and 42% with *T. brucei* RPII subunits (Fig. 5.). The protozoan *P. falciparum* RPII subunit thus showed more homology to higher eukaryotic forms than to either the lower eukaryote yeast, or the protozoan *T. brucei*. The same homology gradient was previously observed for the parasite dihydrofolate reductase-thymidylate synthase (DHFR-TS) gene (21).

Notable sequences and structures in the conserved regions A to E

The DNA-dependent RNA polymerases require tightly bound Zn atoms, or a divalent metal atom such as cobalt, for activity (32). An amino acid sequence thought to form a zinc finger structure that can fold about a Zn atom was present in the largest yeast RPI, RPII, and RPIII subunits as a consensus sequence Cys-X_2-Cys-X_9-His-X_2-His (2). This sequence was found in conserved region A of the *P. falciparum* RPII subunit (position 68 to 84), and was conserved in the mouse (4) and Drosophila RPII subunits (6) (Fig. 5.). The last histidine residue was, however, replaced with a tyrosine residue in the *T. brucei* RPII subunit (7).

The sequence, Tyr-Asn-Ala-Asp-Phe-Asp-Gly-Asp-Glu-Met-Asn (position 511 to 521), was found in conserved region B. This sequence motif is the longest conserved region in all eukaryotic RPI, RPII, and RPIII largest subunits (31).

A sequence motif predicted to form a helix-turn-helix structure is found in region B (position 378 to 411) of the *P. falciparum* RPII subunit (Fig. 5.). This conserved structure has been described in *E. coli* DNA polymerase I and T7 DNA polymerase (3), and is probably involved in DNA binding.

A common amino acid loop, Gly-X_4-Gly-Lys, is found in purine nucleotide binding, or processing, proteins such as elongation factor Tu (EF-Tu) (33). Gly-X_4-Gly forms a relatively large loop in EF-Tu (14). A similar sequence motif, Gly-Asn-Leu-Met-Gly-Lys (position 371 to 376), was found in region B of all the RPII subunits (Fig. 5.) and it may also form a loop for purine nucleotide binding.

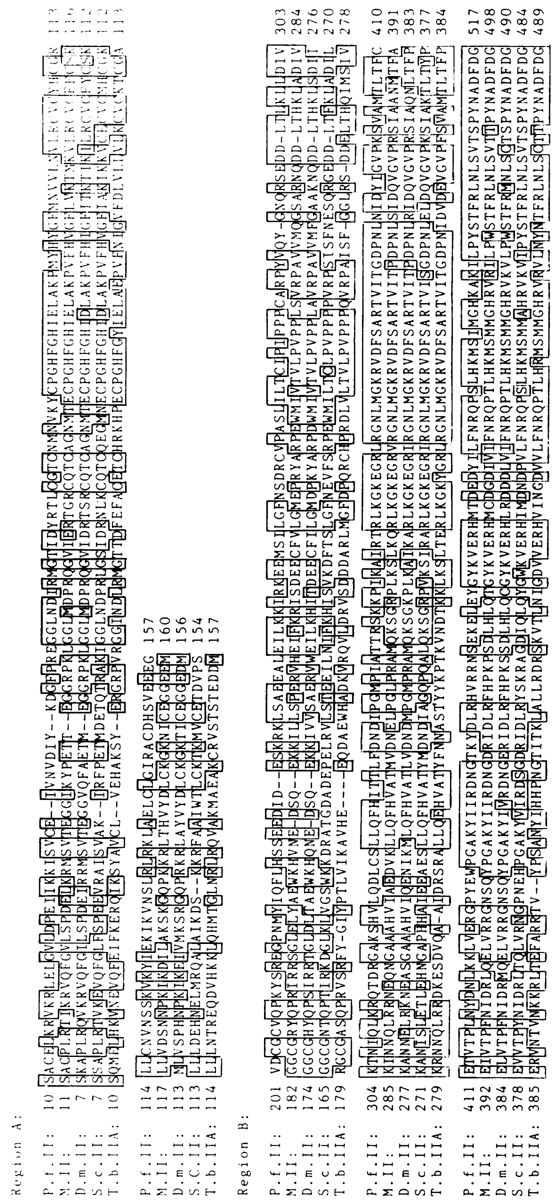
The DD-domain consists of two aspartate residues followed by at least five uncharged residues (34). DD-domains are present in RNA-dependent RNA, and DNA polymerases. A DD-domain is found in the β' subunit of *E. coli* RNA polymerase and in conserved region B (starting at position 293) of the RPII subunits (Fig. 5.) (14). A second *P. falciparum* DD-domain was found in the unique carboxyl-terminal extension (position 2433) (Fig. 3.).

Notable sequences and structures in the variable regions A' to D'

The regions, A' to D', that separate the conserved regions have little overall amino acid homology when compared to the corresponding regions of other RPII subunits, and we refer to them as variable regions. The variable regions do, however, contain short sequences that are conserved among species. Each of the *P. falciparum* variable regions were significantly enlarged in comparison to the variable regions of other RPII subunits (Fig. 4.), and consisted of 43, 136, 278, and 284 amino acid residues, respectively (Fig. 3. and Fig. 5.). The enlarged variable regions could contain unique regulatory domains involved in the control of stage-specific, or species-specific, gene transcription during the developmentally complex life cycle of the parasite.

The variable region C' contained 5 leucine repeats beginning at position 1093, with an interruption of one aspartate (Fig. 3.), and this domain may form a leucine zipper (35,36,37). This structure is thought to facilitate protein-protein interaction (35). Similar sequences have been observed in the corresponding RPII subunit variable domain in Drosophila, yeast, *T. brucei*, and the mouse RPII subunit.

A remarkable basic domain was found in region C' (position 1182 to 1193) that contained 11 basic amino acids out of 12 residues. The DNA binding domain of the transcription/replication factor CTF/NF1 contains a high density of basic amino acids thought to form an α-helical structure (38,39). The basic domain present in region C' could form an α-helical structure that probably interacts with DNA in the transcriptional complex.



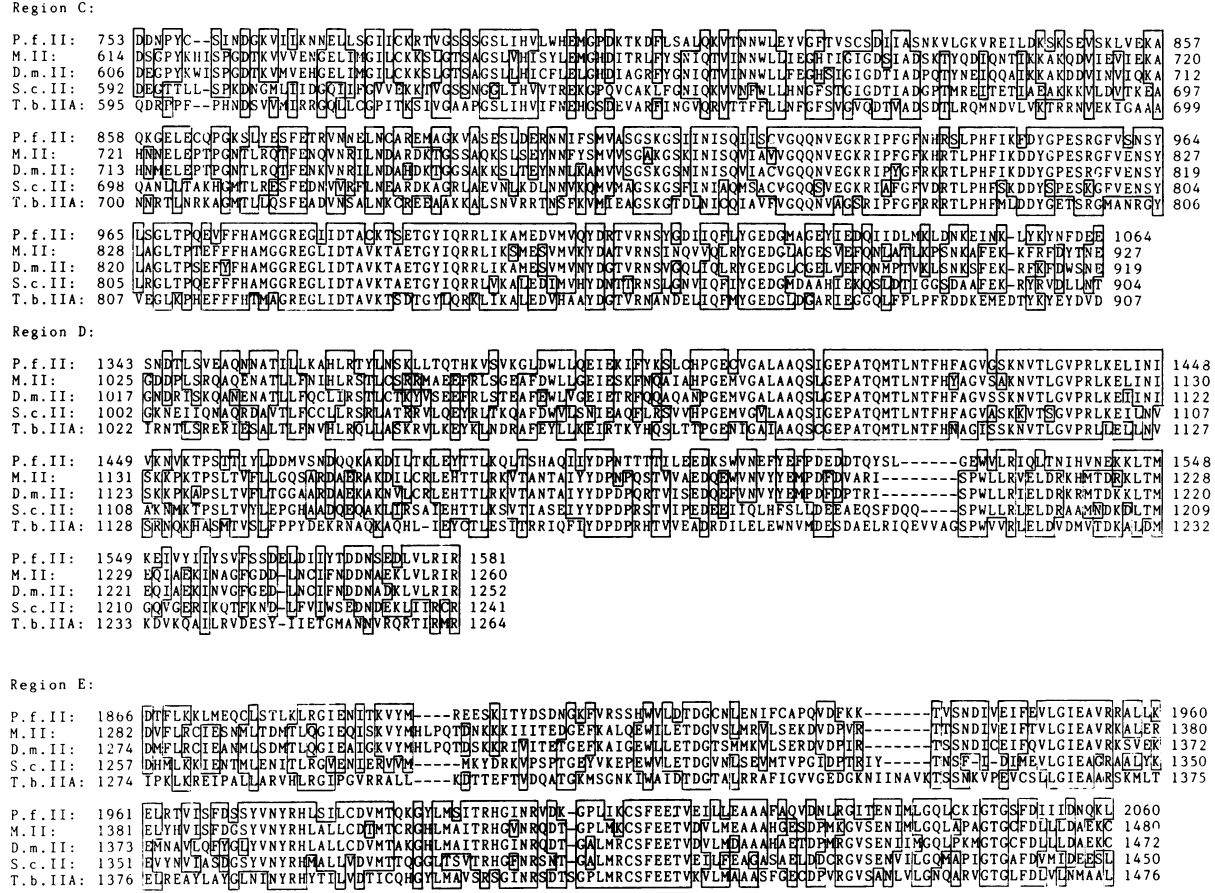

Region C:
P.f.II: 753 857
M.II: 614 720
D.m.II: 606 712
S.c.II: 592 697
T.b.IIA: 595 699
P.f.II: 858 964
M.II: 721 827
D.m.II: 713 819
S.c.II: 698 804
T.b.IIA: 700 806
P.f.II: 965 1064
M.II: 828 927
D.m.II: 820 919
S.c.II: 805 904
T.b.IIA: 807 907
Region D:
P.f.II: 1343 1448
M.II: 1025 1130
D.m.II: 1017 1122
S.c.II: 1002 1107
T.b.IIA: 1022 1127
P.f.II: 1449 1548
M.II: 1131 1228
D.m.II: 1123 1220
S.c.II: 1108 1209
T.b.IIA: 1128 1232
P.f.II: 1549 1581
M.II: 1229 1260
D.m.II: 1221 1252
S.c.II: 1210 1241
T.b.IIA: 1233 1264
Region E:
P.f.II: 1866 1960
M.II: 1282 1380
D.m.II: 1274 1372
S.c.II: 1257 1350
T.b.IIA: 1274 1375
P.f.II: 1961 2060
M.II: 1381 1480
D.m.II: 1373 1472
S.c.II: 1351 1450
T.b.IIA: 1376 1476

Figure 5. Amino acid alignments of the five conserved regions in the RPII subunits. The amino acid alignments of the conserved regions, A, B, C, D, and E, is shown for *P. falciparum* (P.f.II), mouse (M.II), Drosophila (D.m.II), *Saccharomyces cerevisiae* (S.c.II), and *T. brucei* (T.b.II) (allele RPIIA). Amino acids were boxed when three out of five amino acids matched in the vertical column. The amino acid location is given in the left and right column.

The composition of regions C′ and D′ was enriched in acidic amino acids (21%). Four extremely acidic domains were found in C′ (position 1258 to 1290) and D′ position 1602 to 1612, 1746 to 1759, and 1806 to 1820). The longest acidic domain, in region C′, contained a novel Tyr-Asp-Asp-Asp-Asp repeat sequence.

The variable regions C′ and D′ contained high overall charge densities with 54 acidic and 42 basic residues in C′ (out of 278 residues), and 64 acidic and 38 basic residues in D′ (out of 284 residues). Because charge clusters are apparently associated with functional domains of many cellular transcription factors and regulatory proteins (40), we propose that the enlarged variable domains C′ and D′ could provide specific and unique regulatory functions to the P. falciparum RPII subunit.

The notable sequence, Asn-Lys-X-Asp, was found in both variable domains C′ (position 1074 to 1077) and D′ (position 1821 to 1824). This sequence motif is found in *E. coli* EF-Tu, as well as in other elongation factors, and forms a loop for guanine base interaction (33). This sequence motif is unique to the *P. falciparum* RPII subunit.

The variable regions in the *P. falciparum* RPII subunit were extremely asparagine rich (24%) compared to the conserved regions (5%). Asparagine repeats were located in regions B′ (position 707 to 725), C′ (position 1144 to 1159), and D′ (position 1687 to 1694). Two glutamine rich transcriptional activation domains in the transcription factor Sp1 (41), unlike other charged activation domains (42,43), are largely devoid of charged residues in their primary sequence. Amide moieties of the glutamine side chains may participate in hydrogen bonding to the RNA polymerase, or another component of the transcriptional complex (41). The asparagine repeats in the *P. falciparum* RPII subunit could be similarly involved in an interaction with another component of the transcriptional complex via hydrogen bonding.

```
 1    Y  S  P  S  -  -  -
 2    Y  S  P  T  S  P  T Y N A N N A Y
 3    Y  S  P  T  S  P  K N Q N D Q M N V N S Q
 4    Y  N  V  M  S  P  V
 5    Y  S  V  T  S  P  K
 6    Y  S  P  T  S  P  K
 7    Y  S  P  T  S  P  K
 8    Y  S  P  T  S  P  K
 9    Y  S  P  T  S  P  K
10    Y  S  P  T  S  P  K
11    Y  S  P  T  S  P  K
12    Y  S  P  T  S  P  K
13    Y  S  P  T  S  P  V A Q N I
14    A  S  PN Y  S  P  -
15    Y  S  I  T  S  P  K
16    F  S  P  T  S  P  A
17    Y  S  I  S  S  P  V

( Followed by 68 amino acid residues)
```

Figure 6. The *P. falciparum* RPII subunit heptapeptide repeat domain. The *P. falciparum* heptapeptide repeat domain is represented by 17 highly conserved repeats with the concensus sequence Tyr-Ser-Pro-Thr-Ser-Pro-Lys. The repeat region corresponds to amino acids 2247 to 2384 (Fig. 3.).

Remarkably, certain amino acid residues were not found in some of the variable regions. Cysteine was not found in variable region A′, B′, and D′, but it was found in each conserved region, and in variable region C′. Tryptophan was absent from all of the variable regions, but it was found in conserved region B, C, D, and E.

Structure of the carboxyl-terminal domain

A remarkable feature of the eukaryotic RPII subunits is the presence of a heptapeptide repeat (consensus Tyr-Ser-Pro-Thr-Ser-Pro-Ser) in the CTD. The *P. falciparum* RPII subunit contains a small cluster of 17 highly conserved heptapeptide repeats (position 2247 to 2384) (Fig. 6.). The amino acid at position 7 of the repeat unit in *P. falciparum* contained a lysine rather than a serine, which is present in all other RPII subunits with the repeats. This same replacement was found on the carboxyl-terminal side of the repeat element in the mouse and hamster repeats (5), where the serine residue tended to be replaced by a charged amino acid (particularly lysine).

The CTD of the *P. falciparum* RPII subunit contains 392 amino acids that extend from the 3′ side of conserved region E (position 2061) to the carboxyl-terminal amino acid

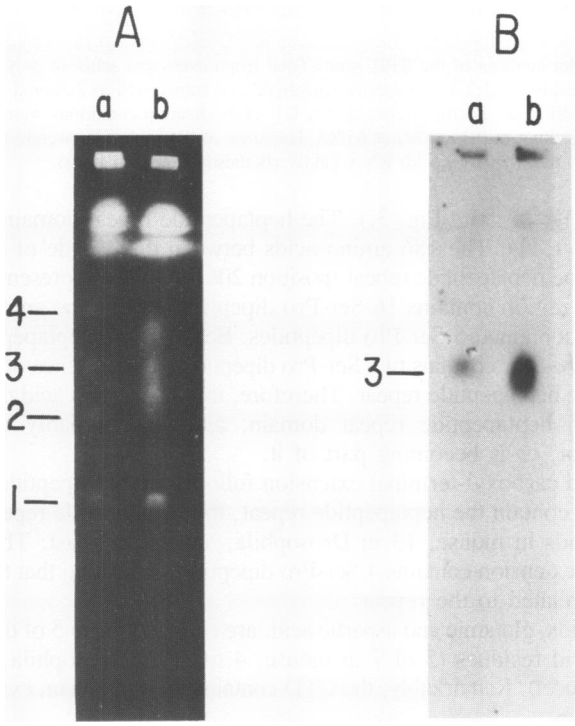


Figure 7. Chromosome assignment of the *P. falciparum* RPII gene. Chromosomes 1, 2, 3, and 4 of *P. falciparum* strains FCR3 (a) and Honduras-1 (b) were separated by pulsed field gradient gel electrophoresis and stained with ethidium bromide (A). The electrophoresis parameters used for this gel were intended to resolve chromosomes 1 to 4, only (28). The chromosomal DNA was transferred to nylon and hybridized with the RPII gene specific probe, cDNA C1, in 6× SSC at 42°C, and was washed in 1× SSC at 65°C (B). The chromosome number assignments are shown.

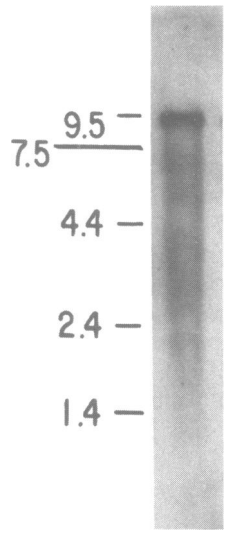


Figure 8. Northern blot analysis of the RPII gene. Total trophozoite and schizont poly(A^+) RNA (4 μg per lane) was electrophoresed on a 1.2% agarose-formaldehyde gel, transferred to Zetabind nylon membrane, and hybridized with the RPII gene specific probe, cDNA C1. Hybridization conditions were previously described (25), and washing was done in 0.1× SSC at 65°C. The sizes of the molecular weight markers are shown in kb, The markers were the 0.24 to 9.5 kb RNA ladder (Bethesda Research Labs).

(position 2452) (Fig. 1F. and Fig. 3.). The heptapeptide repeat domain is located in the middle part of the CTD. The 186 amino acids between the 3′ side of conserved region E and the start of the heptapeptide repeat (position 2061 to 2246) represents another domain of the CTD. This region contains 16 Ser-Pro dipeptides, while the amino-terminal 2060 amino acids only contained 5 Ser-Pro dipeptides. Because one heptapeptide repeat, Tyr-*Ser-Pro*-Thr-*Ser-Pro*-Ser, contains two Ser-Pro dipeptides, the presence of Ser-Pro indicates a relatedness to the heptapeptide repeat. Therefore, this 186 amino acid domain represents a highly diverged heptapeptide repeat domain, and was probably once part of the heptapeptide repeat, or is becoming part of it.

A 68 amino acid carboxyl-terminal extension follows the heptapeptide repeat. In other RPII subunits that contain the heptapeptide repeat, the heptapeptide repeat is followed by only 10 amino acids in mouse, 13 in Drosophila, and 10 in yeast. The 68 amino acid carboxyl-terminal extension contains 4 Ser-Pro dipeptides indicating that the extension may also be distantly related to the repeat.

Acidic amino acids, glutamic and aspartic acid, are enriched for in 5 of the last 9 carboxyl-terminal amino acid residues (5 of 7 in mouse, 4 of 12 in Drosophila, 2 of 6 in yeast, and 6 of 7 in T. brucei). Remarkably, the CTD contains no tryptophan, cysteine, or arginine amino acid residues.

The abundant serine residues in the CTD were previously suggested to provide phosphorylation sites used for regulating RPII activity (10). A protein kinase, CTD kinase, that phosphorylates *in vitro* the CTD of the RPII subunit from mouse cells was recently isolated (44). The CTD kinase phosphorylates one or more serine residues in the CTD

heptapeptide repeat. The purification of CTD kinase now provides direct evidence that the CTD of the RPII subunit may serve as a phosphorylation domain which modulates RPII activity.

Chromosomal location and copy number of the RPII gene

Recently, we clearly separated and determined the sizes of 13 DNA bands, or chromosomes, by pulsed field gradient gel (PFG) electrophoresis of *P. falciparum* strain FCR3 (29). cDNA C1, which is specific for the RPII gene, was hybridized to Southern blot of separated chromosomes 1 to 4, and the RPII gene was located on chromosome 3, based on the number assignment of FCR3 chromosomes, and on a Honduras-1 chromosome that corresponded in size to chromosome 3 of FCR3 (Fig. 7.).

Based on our more precise measurements of the size of the malaria chromosomes by PFG electrophoresis, and an estimated total haploid genome content of 2.64×10^7 bp (29), the copy number of the RPII gene in both Honduras-1 and FCR3 was found to be one copy per parasite (data not shown).

Expression of the RPII gene

Northern blot analysis of a mixture of trophozoite and schizont stage mRNA revealed a single species of RPII mRNA of 9 kb by hybridization to the RPII gene specific probe, cDNA C1 (Fig. 8.). Because of the single copy nature of the RPII subunit gene we conclude that the same RPII gene must be active throughout the life cycle of the parasite both in man and mosquitoes.

We determined the sensitivity of parasite growth to α-amanitin in vitro, and were unable to detect a growth inhibitory effect of α-amanitin at concentrations as high as 20 μg/ml. Because RPII subunit α-amanitin sensitivity should be measured by an in vitro transcription assay, which is not yet developed for this organism, we do not know if the RPII subunit is in fact resistant, or if the cell is impermeable to the drug.

ACKNOWLEDGEMENTS

This work was funded by NIH grants #5R01 AI-20437 and #5R01 AI-22038 to J.I., and #5R29 AI-26651 to D.J.B. The authors are grateful to Thomas Ciardelli, the Structural Biology Laboratory at Dartmouth, and the Dartmouth Molecular Genetics Center for synthesizing the oligonucleotides used in this work.

REFERENCES

1. Sentenac, A. (1985) CRTC Crit. Rev. Biochem. **18**, 31–90.
2. Memet, S., Gouy, M., Marck, C., Sentenac, A., and Buhler, J.-M. (1989) J. Biol. chem. **263**, 2830–2839.
3. Allison, L. A., Moyle, M., Shales, M., and Ingles, C. J. (1985) Cell **42**, 599–610.
4. Ahearn, J. M., Jr., Bartolomei, M. S., West, M. L., Cisek, L. J., and Corden, J. L. (1987) J. Biol. Chem. **262**, 10695–10705.
5. Allison, L. A., Wong, J. K.-C., Fitzpatrick, V. D., Moyle, M., and Ingles, C. J. (1988) Mol. Cell. Biol. **8**, 321–329.
6. Jokerst, R. S., Weeks, J. R., Zehring, W. A., and Greenleaf, A. L. (1989) Mol. Gen. Genet. **215**, 266–275.
7. Smith, J. L., Levin, J. R., Ingles, C. J., and Agabian, N. (1989) Cell **56**, 815–827.
8. Evers, R., Hammer, A., Kock, J., Jess, W., Borst, P., Memet, S., and Cornelissen, A. W. C. A. (1989) Cell **56**, 585–597.
9. Ovchinnikov, Y. A., Monastyrskaya, G. S., Gubanov, V. V., Guryev, S. O., Salomatina, I. S., Shvaea, T. M., Lipkin, V. M., and Sverdlov, E. D. (1982) Nucl. Acids Res. **10**, 4035–4044.
10. Zehring, W. A., Lee, J. M., Weeks, J. M., Jokerst, R. S., and Greenleaf, A. L. (1988) Proc. Natl. Acad. Sci. USA **85**, 3698–3702.

11. Corden, J. L., Cadena, D. L., Ahearn, J. M., and Dahmus, M. E. (1985) Proc. Natl. Acad. Sci. USA **82**, 7934–7938.
12. Ingles, C. J., Moyle, M., Allison, L. A., Wong, J. K.-C., Archambault, J., and Freisen, J. D. (1987) Mol. Cell Biol. **52**, 383–393.
13. Nonet, M., Sweetser, D., and Young, R. A. (1987) Cell **50**, 909–915.
14. Cornelissen, A. W.C.A., Evers, R., and Kock, J. (1988) Oxford Surveys Euk. Gen. **5**, 91–131.
15. Trager, W., and Jensen, J. B. (1976) Science **193**, 673–675.
16. Gritzmacher, C. A., and Reese, R. T. (1984) J. Bacteriol. **160**, 1165–1167.
17. Banyal, H. S., and Inselburg, J. (1985) Am. J. Trop. Med. Hyg. **34**, 1055–1064.
18. Inselburg, J. (1983) J. Parasitol. **69**, 584–591.
19. Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory).
20. Horii, T., Bzik, D. J., and Inselburg, J. (1988) Mol. Biochem. Parasitol. **30**, 9–18.
21. Bzik, D. J., Li, W.-B., Horri, T. and Inselburg, J. (1987) Proc. Natl. Acad. Sci. USA **84**, 8360–8364.
22. Dretzen, G., Bellard, P., Sassone-Corsi, P., and Chambon, P. (1981) Anal. Biochem. **112**, 295–298.
23. Feinberg, A. P., and Vogelstein, B. (1983) Anal. Biochem. **132**, 6–13.
24. Li, W.-B., Bzik, D. J., Horii, T., and Inselburg, J. (1989) Mol. Biochem. Parasitol. **33**, 13–26.
25. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. USA **74**, 5463–5467.
26. Bzik, D. J., Li, W.-B., Horii, T., and Inselburg, J. (1988) Mol. Biochem. Parasitol. **30**, 279–288.
27. Schwartz, D. C., and Cantor, C. R. (1984) Cell **37**, 67–75.
28. Chu, G., Vollrath, D., and Davis, R. W. (1986) Science **234**, 1582–1585.
29. Gu, H., Inselburg, J., Bzik, D. J., and Li, W.-B. (1989) Exp. Parasitol. (in press).
30. Weber, J. L. (1987) Gene **52**, 103–109.
31. Memet, S., Saurin, W., and Sentenac, A. (1988) J. Biol. Chem. **263**, 10048–10051.
32. Wu, F. Y.-H., and Wu, C.-W. (1981) In Advances in Inorganic Biochemistry, Vol. 3, G. L. Eichhorn and L. G. Marzilli, (eds.), Elsevier/North-Holland, New York, pp. 143–166.
33. La Cour, T. F. M., Nyborg, J., Thirup, S., and Clark, B. F. C. (1985) EMBO J. **4**, 2385–2388.
34. Zavriev, S. K., and Borisova, O. V. (1987) [translated from] Mol. Biol. **21**, 229–241.
35. Landschulz, W. H., Johnson, P. F., and McKnight, S. L. (1988) Science **240**, 1759–1764.
36. Kouzarides, T., and Ziff, E. (1988) Nature **336**, 646–651.
37. Struhl, K. (1989) TIBS **14**, 137–140.
38. Santoro, C., Mermod, N., Andrews, P. C., and Tjian, R. (1988) Nature **334**, 218–224.
39. Mitchell, P. J., and Tjian, R. (1989) Science **245**, 371–378.
40. Brendel, V., and Karlin, S. (1989) Proc. Natl. Acad. Sci. USA **86**, 5698–5702.
41. Courey, A. J., and Tjian, R. (1988) Cell **55**, 887–898.
42. Ma, J., and Ptashne, M. (1987) Cell **48**, 847–853.
43. Hope, I. A., and Struhl, K. (1986) Cell **46**, 885–894.
44. Cisek, L. J., and Corden, J. L. (1989) Nature **339**, 679–684.

This article, submitted on disc, has been automatically converted into this typeset format by the publisher.