Nucleic Acids Research, 1993, Vol. 21, No. 16 3643–3646

The gene encoding DNA polymerase α from *Plasmodium falciparum*

John H.White, Brian J.Kilbey*, Erik de Vries[1], Michael Goman, Pietro Alano[2], Sandie Cheesman, Sybil McAleese and Robert G.Ridley+

Institute of Cell and Molecular Biology, University of Edinburgh, Kings Buildings, Mayfield Road, Edinburgh EH9 3JR, UK, [1]Institute of Infectious Diseases and Immunology, Department of Parasitology, University of Utrecht, PO Box 80165, 3508 TD Utrecht, The Netherlands and [2]Laboratorio Biologie Cellulare, Istituto Superiore di Sanita, viale Regina Elene 299, 00161 Rome, Italy

Received May 28, 1993; Revised and Accepted July 12, 1993

GenBank accession no. L18785

ABSTRACT

The gene encoding DNA polymerase α from the human malaria parasite *Plasmodium falciparum* has been sequenced and characterised. The deduced amino acid sequence possesses the seven sequence motifs which characterise eukaryotic replicative DNA polymerases (I – VII) and four of five motifs (A – E) identified in α DNA polymerases. The predicted protein also contains sequences which are reminiscent of *Plasmodium* proteins but absent from other DNA polymerases. These include four blocks of additional amino acids interspersed with the conserved motifs of the DNA polymerases, four asparagine rich sequences and a novel carboxy-terminal extension. Repetitive sequences similar to those found in other malarial proteins are also present. cDNA-directed PCR was used to establish the presence of these features in the ~7kb mRNA. The coding sequence contains a single intron. The gene for DNAPolα is located on chromosome 4 and is transcribed in both asexual and sexual erythrocytic stages of the parasite.

INTRODUCTION

DNA polymerase α (DNAPolα) is an essential component of the chromosomal DNA replication apparatus (1). It performs an essential function in *S.cerevisiae* (2) and catalyses limited DNA replication in the SV40 cell-free replication system (3). Data from a range of eukaryotic cell types indicate a conserved subunit composition; a catalytic subunit (about 180kDa) is associated with two polypeptides with primase activity (55–60kDa and 48–50kDa) and a third subunit (70–75kDa) of uncertain function (4). The DNAPolα/primase complex is believed to be responsible for the initiation of DNA replication and for repeatedly initiating lagging strand synthesis (5). Recently nucleic acid sequences encoding DNAPolα have been isolated from Man (6), *Drosophila melanogaster* (7), *Saccharomyces cerevisiae* (8), *Schizosaccharomyces pombe* (9) and *Trypanosoma brucei* (10). We are currently studying DNA replication in the human malaria parasite *Plasmodium falciparum* in the hope that differences between the parasite and host mechanisms may form the basis for specific drug therapy. Preliminary studies in *P.berghei* and *P.falciparum* suggest that DNA replication in malaria parasites may be a target for specific inhibitors (11). A fragment of this sequence was reported in an earlier paper(12).

MATERIALS AND METHODS

Parasite culture; DNA and RNA isolation and blotting techniques

Strains K1 (13) and 3D7A (14) were employed in this study. Asexual parasites were cultured using established methods (15, 16). Gametocytes were isolated as reported previously (12). The methods used for the isolation of DNA (17) and RNA (18, 19) and for Southern (20) and Northern blotting (18) have been described.

Isolation of the DNAPolα gene

A single 1.0kb *Eco* RI insert encoding a protein with sequence similarity to DNAPolα was reported earlier (12). Overlapping clones were isolated by hybridisation from *Eco* RI and *Dra* I λ genomic libraries and by inverse PCR (21). In the latter case sequence data were verified by comparing the products of three independent reactions.

Gene sequencing and analysis

Gene fragments were subcloned into either pUBS (22) or pUC19 (23) and sequenced by the chain termination method (24) using the Sequenase 2.0 kit (USB). Sequence comparisons were made using the University of Wisconsin Genetics Computer Group (UWGCG) programmes (25).

* To whom correspondence should be addressed

+ Present address: Pharma Division, Preclinical Research, F.Hoffmann-La Roche, CH4002 Basel, Switzerland

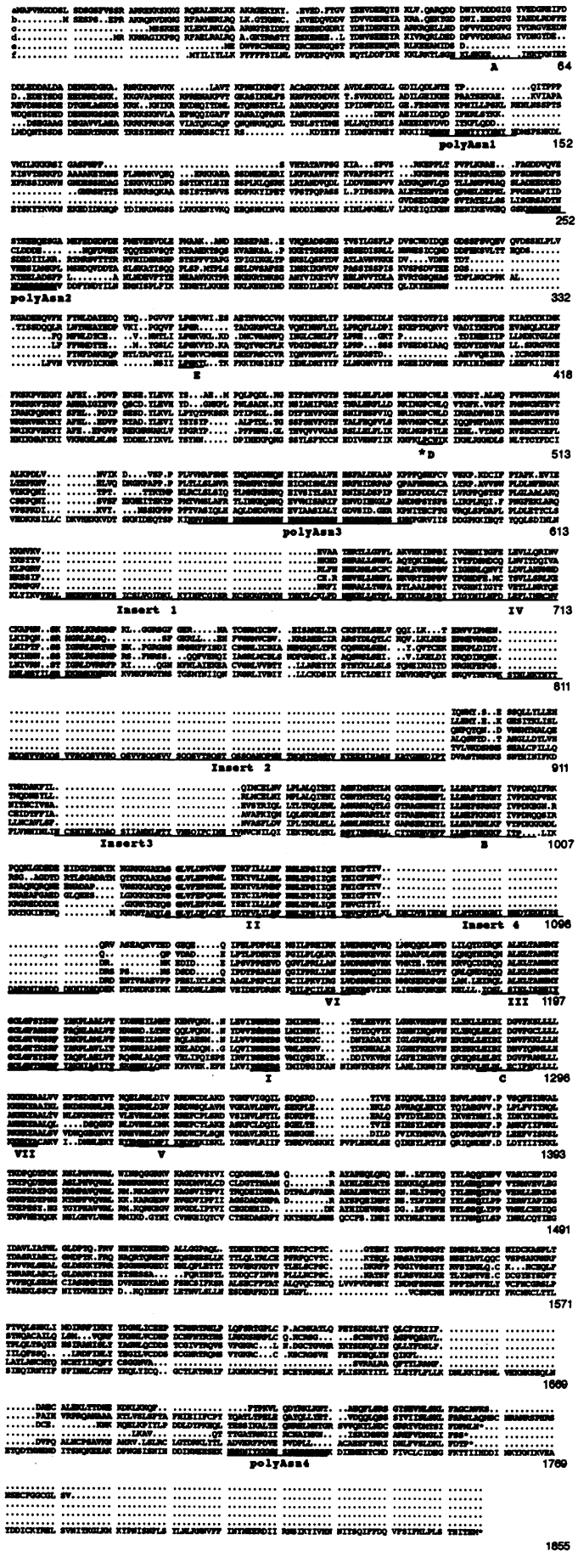

Figure 1. UWGCG pileup analysis comparing the DNA polymerase α sequences from Man (a), *D.melanogaster* (b), *S.cerevisiae* (c), *S.pombe* (d), *T.brucei* (e) and *P.falciparum* (f). Universally conserved residues are indicated in bold type. Conserved motifs I–VII and A–E are underlined. Four asparagine rich motifs and four additional inserts are also indicated. A functionally important residue in motif D is starred. The alignment is numbered according to the *Plasmodium* amino acid sequence.

PCR from cDNA

PolyA$^+$ RNA from strain K1 was used with an AMV Reverse Transcription Kit (Promega) in accordance with the manufacturer's instructions (26). The cDNA generated was checked for contaminating genomic DNA by running a PCR reaction using primers flanking the introns of the *P.falciparum* β-tubulin gene. The only band observed corresponded to the expected size of the spliced product (data not shown).

RESULTS

The DNAPolα gene from *Plasmodium falciparum*

The nucleotide sequence encoding the putative DNAPolα ORF of *P.falciparum* is 5.7kb in length with a single intron of 204bp starting at nucleotide 4143 of the coding sequence. The 1855 residue predicted protein has a molecular mass of ~205 kDa. The sequence was aligned with sequences of the other known DNAPolα proteins (Fig. 1). The seven conserved motifs I to VII which are a constant feature of DNAPolαs are all present in the *P.falciparum* sequence and their order is also conserved. Four of the five DNAPolα-specific motifs, A–E, are also present. In both the *P.falciparum* and *T.brucei* sequences, motif A is absent. The sequence appears, therefore, to encode the DNAPolα of *P.falciparum* (Pf DNAPolα). Co-probing of CHEF chromosome blots with Pf DNAPolα- and DHFR-specific probes (27) coupled with genomic southern analysis (data not shown) show that it exists as a single copy on chromosome 4.

Fig. 1 identifies a number of features unique to Pf DNAPolα: The glycine residue in motif D, which is present in all other DNAPolα sequences described, is replaced by leucine. This residue (493 in the *S.cerevisiae* sequence) has been implicated in the DNAPolα/primase interaction in *S.cerevisiae* (28,29). There are four blocks of inserted amino acids interspersed between the DNAPol-specific motifs. Insert 2 has seven copies of a degenerate six residue repeat (QQSVVS), while Insert 4 has four copies of another degenerate six residue repeat (KNIHSD). Repetitive sequences of this sort are commonly found in malarial proteins (30). There are several asparagine rich tracts; also a common feature of malarial proteins. The sequence ends with a novel C-terminal extension of approximately 74 residues. The putative translation initiation signal is the first ATG of the open reading frame. It is preceded a by a stop codon but is followed by an unexpectedly hydrophobic region including five successive phenylalanines. An alternative possibility is the second ATG codon (residue 66). This would, however, eliminate a partially conserved region (residues 20–58). The putative initiating ATG is present in the mRNA but confirmation of its function awaits the determination of the N-terminal sequence of the purified, unproteolysed protein.

The structure of PfDNAPolα mRNA

Genomic DNA was used to generate the sequence shown in Fig. 1. In order to determine whether or not the features of the

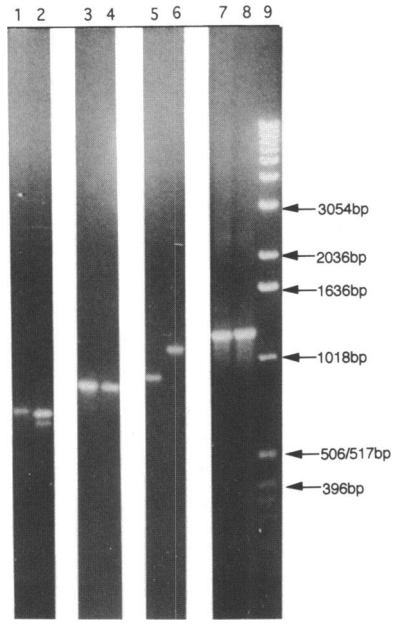

Figure 2. Some examples of the products from cDNA and genomic (g)DNA-directed PCR reactions. The products were separated in 1% agarose. The regions examined here are as follows:Lane 1; cDNA N-terminus, lane 2; gDNA N-terminus). Lane 3; cDNA inserts 2 and 3 , lane 4; gDNA inserts 2 and 3. Lane 5; cDNA intron, lane 6; gDNA intron. Lane 7; cDNA C-terminus, lane 8; gDNA C-terminus and lane 9: BRL 1kb ladder.

predicted gene product (amino acid inserts etc.) are present in the processed message, cDNA-directed PCR was performed using synthetic primers flanking the regions of interest, and the products compared with related products obtained using genomic DNA as the template. Some of these data are presented in Fig. 2. In every case, except that of the anticipated intron, the evidence suggests that the unique features are not processed out of the mature message. The intron *is* removed and provides additional evidence that the cDNA is uncontaminated with genomic DNA. The identity of the PCR products was confirmed by a combination of probing with internal sequences and DNA sequencing. In our earlier report we suggested that sequence encoding insert 2 formed part of an in-frame intron (12). We now know that this result was obtained with only one RNA preparation and that the PCR product could not be subcloned and sequenced. PCR reactions across the same region with subsequent RNA preparations and different primers gave the result shown in Fig. 2. Even so, smaller products have occasionally been obtained with these primers also. When one of these was sequenced, it proved to be a deletion of part of insert 2 together with 3′ adjacent sequence conserved in all DNAPolα sequences so far identified. The ends of the deletion did not correspond with the splicing consensus signals. We conclude that the earlier result was artifactual. The conclusion that insert 2 is encoded in the mature message is further supported by the results of probing Northern blots of RNA prepared from asexual and sexual (gametocyte) parasite stages with sequences encoding insert 2 amino acids (data not shown). In both cases a message of approximately 7kb was detected. The same result was obtained when the blot was stripped and reprobed with a Pf DNAPolα exon probe.

DISCUSSION

The predicted PfDNAPolα has only a 14−17% amino acid identity with DNAPolα proteins from other organisms. This is the most divergent DNAPolα protein reported so far (6−10). Although this is in part due to the extensive additional sequences, the product is even so much less similar to its yeast counterpart than is the case for the DNAPolδ proteins (12). Insertions of blocks of amino acids and polyasparagine tracts are features which have been found in other proteins from *Plasmodium*. C-terminal extensions have also been reported. Repeat sequences were first reported in proteins which were isolated from cDNA libraries screened immunologically (30) and it was suggested that they may play a role in confusing the host's immune system(31). However, they can also occur in genes selected on the basis of protein sequence conservation, as in the present case. The *P.falciparum* DNA-dependent RNA polymerases II and III provide further examples (32,33). RNA polymerase II has five conserved domains (A−E) separated by four variable domains (A′, B′ C′ and D′) which are asparagine rich, with distinct asparagine repeats in three of the variable regions. The variable domains, particularly region C′, also contain acidic repeats. The C-terminal domain contains a heptapeptide repeat conserved in all species, but this is preceded and followed by sequences rich in serine-proline dipeptides. RNA polymerase III from *P.falciparum* has similar features (33). Whether these unusual features hold significance for DNA-de[endent polymerase function or are simply the unavoidable consequence of some aspect of the parasite's DNA metabolism is not yet known but it may become clearer once specific enzymic assay systems are established. At present it is only possible to assay for DNA polymerase activity in semi-purified extracts; assays of the activity of specific gene products have still to be developed. For the same reason it is at present difficult to pursue the significance of the glycine to leucine change in the conserved D motif.

ACKNOWLEDGEMENTS

We would like to thank Graham Harold, Linda McCallum and Lorna McGregor for technical assistance. This work was funded by the Medical Research Council and P.A was supported by E.E.C. Grant STDII TS3CT920116.

REFERENCES

1. Wang, T. S. -F. (1991) *Annu. Rev Biochem.*, **60**, 513−552.
2. Johnson, L. M., Snyder, M., Chang, L. M. S., Davis, R. W. and Campbell, J. L. (1985) *Cell* , **43** 369−377.
3. Murakami, Y., Wobbe, C. R., Weissbach, L., Dean, F. B. and Hurwitz, J. (1986) *Proc. Natl. Acad. Sci. USA* , **83**, 2869−2873.
4. Burgers, P. M. J. (1989) *Prog. Nucleic Acid Res. Mol. Biol.*, 197−245.
5. Tsurimoto, T., Melendy, T. and Stillman, B. (1989) *Nature*, **346** 534−539.
6. Wong, S. W., Wahl, A. F., Yuan, P.-M., Arai, N., Pearson, B. E., Arai, K., Korn, D., Hunkapiller, M. W. and Wang, T. S.-F. (1988) *EMBO J.*, **7**, 37−47.
7. Hirose, F., Yamaguchi, M., Nishida, Y., Masutani, M., Miyazawa, H., Hanaoka, F. and Matsukage, A. (1991) *Nucleic Acids Res.*, **19**, 4991−4998.
8. Pizzagalli, A., Valsasnini, P., Plevani, P. and Lucchini, G. (1988) *Proc, Natl. Acad. Sci. USA*, **85**, 3772−3776.
9. Damagnez, V., Tillit, J., de Recondo, A.-M. and Baldacci, G. (1991) *Mol. Gen. Genetics* , **226**, 182−187.
10. Leegwater, P. A. J., Strating, M., Murphy, N. B., Kooy, R. F., van der Vliet, P. C. and Overdulve, J. P. (1991) *Nucleic Acids Res.*, **19**, 6441−6447.
11. de Vries, E., Stam, J. G., Franssen, F. F. J., van der Vliet, P. C. and Overdulve, J. P. (1991) *Mol Biochem. Parasitol.*, **45**, 223−232.
12. Ridley, R. G., White, J. H., McAleese, S. M., Goman, M., Alano, P., deVries, E., and Kilbey, B. J. (1991) *Nucleic Acids Res.*, **19**, 6731−6736.

13. Thaithong, S. and Beale, G. H. (1981) *Trans, R. Soc. Trop. Med. Hyg.*, **75**, 271–273.
14. Walliker, D., Quakyi, I. A., Wellems, T. E., McCutchan, T. F., Szarfman, A., London, W. T., Corcoran, L. M., Burkot, T. R. and Carter, R. (1987) *Science*, **236**, 1661–1666.
15. Trager, W. and Jensen, J. B. (1976) *Science*, **217**, 254–257.
16. Zolg, J. W., Mcleod, A. J., Dickson, I. H. and Scaife, J. G. (1982) *J. Parasitol.*, **68**, 1072–1080.
17. Goman, M., Langsley, G., Hyde, J. E., Yankovsky, N., Zolg, J. W. and Scaife, J. G. (1982) *Mol Biochem. Parasitol.*, **5**, 391–400.
18. Chomczynski, P. and Sacchi, N. (1987) *Anal. Biochem.*, **162**, 156–159.
19. Delves, C. J., Alano, P., Ridley, R. G., Goman, M., Holloway, S. P., Hyde, J. E. and Scaife, J. G. (1990) *Mol Biochem. Parasitol.*, **43**, 271–278.
20. Delves, C. J., Ridley, R. G., Goman, M., Holloway, S. P., Hyde, J. E. and Scaife, J. G. (1989) *Mol Microbiol.*, **3**, 1511–1519.
21. Ochman, H., Medhora, M. M., Garza, D. and Hartl, D. L. (1990) In PCR Protocols, Academic Press. pp 219–227.
22. Goman, M., Mons, B. and Scaife, J. G. (1991) *Mol Biochem. Parasitol.*, **45** 125, 134.
23. Yanisch-Peron, C., Vieira,J. and Messing, J. (1982). *Gene* **33**, 103–119.
24. Sanger, F., Nicklen, S. and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* , **83**, 5463–5467.
25. Devereux, J., Haeberli, P and Smithies, O. (1984) *Nucleic Acids Res.*, **12**, 387–395.
26. Goodman, H. M. and MacDonald, R. J. (1979) *Meth. Enzymol.*, **68**, 75–90.
27. Triglia,T., Wellems,T.E. and Kemp, D.J. (1992) *Parasitol.Today*, **8**, 225–229
28. Lucchini, G., Mazza, C., Scacheri, E. and Plevani, P. (1988) *Mol. Gen. Genet* ., **212**, 459–465.
29. Lucchini, G., Falconi, M. M., Pizzagalli, A., Aguilera, A., Klein, H. L. and Plevani, P. (1990) *Gene*, **90**, 99–104.
30. Ridley, R. G. (1991) *Biochemical Society Transactions*, **19**, 525–528.
31. Schofield, L. (1991) *Parasitol. Today*, **7**, 99–105.
32. Li, W. B., Bzik, D. J., Gu, H., Tanaka, M., Fox, B. A. and Inselburg, J (1989) *Nucleic Acids Res.*, **17**, 9621–9636.
33. Li, W. B., Bzik, D. J., Tanaka, M., Gu, H., Fox, B. A. and Inselburg, J. (1991) *Mol . Biochem. Parasitol.*, **46**, 229–240.