Sexual stage-specific expression of a third calcium-dependent protein kinase from \textit{Plasmodium falciparum}^{1}

Ji-Liang Li \textsuperscript{a,},*, David A. Baker \textsuperscript{b}, Lynne S. Cox \textsuperscript{a}

\textsuperscript{a} Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK
\textsuperscript{b} Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK

Received 3 August 1999; received in revised form 15 November 1999; accepted 31 January 2000

Abstract

A third calcium-dependent protein kinase (CDPK) gene has been isolated from the human malaria parasite \textit{Plasmodium falciparum} by vectorette technology. The gene consists of five exons and four introns. The open reading frame resulting from removal of the four introns encodes a protein of 562 amino acid residues with a predicted molecular mass of 65.3 kDa. The encoded protein, termed PfCDPK3, consists of four distinct domains characteristic of a member of the CDPK family and displays the highest homology (46% identity and 69% similarity) to PfCDPK2, the second CDPK of \textit{P. falciparum}. The N-terminal variable domain is rich in serine/threonine and lysine and contains multiple consensus phosphorylation sites for a range of protein kinases. The catalytic domain possesses all conserved motifs of the protein kinase family except for the highly conserved glutamic acid residue in subdomain VIII, which is replaced by a glutamine residue. The sequence of the junction domain comprising 31 amino acid residues is less conserved. The calmodulin-like regulatory domain contains four EF-hand calcium-binding motifs, each consisting of a loop of 12 amino acid residues which is flanked by two \( \alpha \)-helices. Southern blotting of genomic DNA digests showed that the \textit{Pfcdpk3} gene is present as a single copy per haploid genome. A 2900 nucleotide transcript of this gene is expressed specifically in the sexual erythrocytic stage, indicating that PfCDPK3 is involved in sexual stage-specific events. It is proposed that PfCDPK3 may serve as a link between calcium and gametogenesis of \textit{P. falciparum}. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Malaria; Calcium-dependent protein kinase; Gametogenesis; \textit{Plasmodium falciparum}

Abbreviations: CDPK, calcium-dependent protein kinase; PfCDPK1, the first \textit{P. falciparum} calcium-dependent protein kinase; PfCDPK2, the second \textit{P. falciparum} calcium-dependent protein kinase; PfCDPK3, the third \textit{P. falciparum} calcium-dependent protein kinase

\* Corresponding author. Fax: +44-1865-222431; E-mail: lij@icrf.icnet.uk

1 Nucleotide sequence data reported in this paper are available in the GenBank, EMBL and DDJB databases under the accession number AF106064.

The resistance of \textit{Plasmodium falciparum} to drugs and the resistance of mosquitoes to insecticides have resulted in a resurgence of malaria in many parts of the world. Therefore, there is an urgent need for development of effective vaccines and new anti-malarial drugs. The identification of new targets for vaccine and drug development will be facilitated by a better understanding of the cellular and molecular processes at different stages of the parasite. The sexual erythrocytic stage is functionally very distinct from the asexual stage. It is responsible for trans-
mission of the parasite to the mosquito, where exflagellation and fertilisation occur sequentially. Therefore, it is conceivable that more active biochemical and molecular events could be involved in the sexual stage than in the asexual stage. However, the number of sexual stage-specific proteins identified so far is very limited. Most of the sexual stage-specific genes isolated encode membrane proteins which are being assessed for their potential as transmission blocking vaccine candidates. Little is known about their biological function in the sexual stage. We are interested in the signal transduction pathways involved in the sexual differentiation of *P. falciparum*. The first step toward this goal is to identify the components, mainly protein kinases and phosphatases, of the signal transduction pathways. Recently, we have reported several sexual stage-specific genes encoding either protein serine/threonine phosphatases [1–3] or protein kinases (Li et al., unpublished). In this report, we describe the molecular cloning and characterisation of another novel gene encoding a third *P. falciparum* calcium-dependent protein kinase (CDPK) (PfCDPK3). PfCDPK3 is expressed specifically in the sexual stage, indicating that it may be important in regulating the processes of sexual stage development such as gametogenesis.

In the *P. falciparum* tag database, the nucleotide sequences of two DNA fragments (tag 0487m3 and tag 0532m3) were found encoding the same protein sequence with a high homology to the catalytic do-

Table 1

<table>
<thead>
<tr>
<th>Segment</th>
<th>Size (bp)</th>
<th>A+T content (%)</th>
<th>Junction</th>
<th>Boundary sequence of exon/intron</th>
</tr>
</thead>
<tbody>
<tr>
<td>S' Untranslated</td>
<td>253</td>
<td>85.4 (216/253)</td>
<td>A/I</td>
<td>ATTGAGCATG:gtataaaaa</td>
</tr>
<tr>
<td>Exon A</td>
<td>978</td>
<td>73.3 (717/978)</td>
<td>I/B</td>
<td>ttatatag:GTAAAGAAGG</td>
</tr>
<tr>
<td>Intron I</td>
<td>218</td>
<td>84.4 (184/218)</td>
<td>B/II</td>
<td>ATTATCCAG:ttttttatat</td>
</tr>
<tr>
<td>Exon B</td>
<td>492</td>
<td>71.5 (352/492)</td>
<td>II/C</td>
<td>ttctttttag:AAATATTAT</td>
</tr>
<tr>
<td>Intron II</td>
<td>174</td>
<td>93.1 (162/174)</td>
<td>C/III</td>
<td>ATTGCCCCAC:gtataaaaa</td>
</tr>
<tr>
<td>Exon C</td>
<td>75</td>
<td>56.0 (42/75)</td>
<td>III/D</td>
<td>ttttttttag:ATACTTATA</td>
</tr>
<tr>
<td>Intron III</td>
<td>118</td>
<td>86.4 (102/118)</td>
<td>D/IV</td>
<td>CGACGGAAAG:gtatactcataa</td>
</tr>
<tr>
<td>Exon D</td>
<td>99</td>
<td>63.6 (63/99)</td>
<td>IV/E</td>
<td>ttttttttag:ATTGATTTC</td>
</tr>
<tr>
<td>Intron IV</td>
<td>101</td>
<td>90.1 (91/101)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exon E</td>
<td>42</td>
<td>78.6 (33/42)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
main of the CDPK family. Nucleotide sequence analysis revealed that the A+T content and codon usage of the tag sequences are typical of the coding region of *P. falciparum* genes. To isolate a full-length of the gene, two specific primers, CD1 (to obtain further sequence in the 3′ direction) and CD2 (to obtain further sequence in the 5′ direction), were constructed on the basis of the tag sequences and used in PCR to screen vectorette libraries [4]. Two fragments (CD1-TaqI and CD2-DraI) (Fig. 1) were obtained and sequenced. The sequence data permitted construction of the CD3 and CD4 primers and subsequent screening of vectorette libraries. Two overlapping PCR products (CD4-RsaI and CD4-HincII) were amplified with CD4, both containing a putative ATG start codon, whereas only one fragment (CD3-DraI) was obtained with CD3. However, the sequence data of CD3-DraI made it possible to construct the CD5 primer that gave rise to the CD5-AluI fragment. Based on the sequence of CD5-AluI, the CD7 primer was designed to produce the CD7-DraI fragment. Analysis of the sequence revealed a putative TAA stop codon in CD7-DraI. In order to confirm the sequence obtained from the overlapping fragments, a pair of primers, CDBP and CDEP, were used to amplify the full-length gene from genomic DNA and the PCR product was sequenced in both strands (see Fig. 1).

The sequence derived from overlapping fragments consists of 2554 bp and contains five exons, four introns and a 5′ untranslated region (Fig. 2). The proposed coding region of the gene starts with an ATG codon at nucleotide 254 and terminates with a TAA codon at nucleotide 2551. The sequence and codon usage in the coding region are typical for a *P. falciparum* gene. The A+T contents of the 5′ flanking (253 bp) and four putative intron non-coding regions are characteristically higher than those of the exon coding regions (Table 1). The four proposed introns, ranging from 101 to 218 nucleotides in length, interrupt the coding region from..
the middle to the C-terminus. Intron I is located in the middle of the gene, corresponding to subdomain X of the catalytic domain of the deduced kinase, whereas introns II, III and IV are located towards the 3' end of the gene, corresponding to the calmodulin-like domain of the kinase. The highly conserved dinucleotides GT and AG, found at eukaryotic intron boundaries [5], define these intervening sequences. The sequences around the 5' splice sites of these introns (GTA/TA/TT) show more conserved nucleotides, while the sequences around the 3' splice sites (TA/TTAG) are fairly consistent with the consensus sequence found at the 3' end of eukaryotic introns [5]. Long runs of poly(AT), poly(T) and poly(A) are present in the introns. A four nucleotide repetitive sequence, consisting of GTAT, occurs within intron II and repeats nine times. To verify the existence of these introns, reverse transcription (RT-) PCRs [2] were performed and the products sequenced (see Figs. 1 and 2). The data confirmed the precise exon-intron boundaries.

The open reading frame resulting from removal of the four introns encodes a protein of 562 amino acids (see Fig. 2) with a predicted molecular mass of approximately 65.3 kDa. Database searches revealed that the amino acid sequence of PfCDPK3 shares 50-69% similarity and 32-46% identity with kinases in the CDPK family. PfCDPK3 has the highest homology to PfCDPK2 (46% identity, 69% similarity), the second CDPK of \textit{P. falciparum} [6], but only 40% identity and 64% similarity to PfCDPK1 (PfCPK), the first CDPK of the malaria parasite \textit{P. falciparum} [7]. Fig. 3A shows a sequence alignment of all three \textit{P. falciparum} CDPKs. PfCDPK3 is composed of four distinct domains characteristic of a member of the CDPK family: (1) a variable N-terminal segment, (2) a highly conserved protein kinase catalytic domain, (3) a junction domain, and (4) a calmodulin-like domain. The N-terminal domain, consisting of 116 amino acid residues, is not related to any previously described protein serine/threonine kinase and has several interesting features. Firstly, it is rich in serine/threonine (14% (16/116)) and lysine (24% (28/116)). Secondly, the N-terminal region contains multiple potential phosphorylation sites for a range of known protein kinases [8], suggesting that the PfCDPK3 activity may also be regulated by reversible phosphorylation of the N-terminal segment. Thirdly, in contrast to those of the known kinases in the CDPK family, the N-terminal segment is among the larger extensions of the CDPKs yet described. The kinase catalytic domain of PfCDPK3 is composed of 264 residues and contains all 11 conserved subdomains of the protein kinase family. PfCDPK3 has almost all of the characteristic features of a kinase [9,10]. These include (corresponding to the residue numbers for the bovine cAPK-\(\alpha\) catalytic subunit): the glycine loop G50-X-G52-X-X-G55, forming part of the ATP-binding site; D166, N171 and D184, which are also thought to be a sequence motif implicated in ATP-binding; the triad composed of the side chain of K72, D184 and E91, that is close to the ATP \(\gamma\)-phosphate and plays a critical role in recognition of the phosphate; the catalytic loop R165-D166-X-X-X-N171, involved in catalysis and in guiding the peptide substrate into the proper orientation for catalysis; D220 and R280, involved in the stabilisation of kinases; and A206, which is diagnostic of the catalytic domain of protein kinases. In addition, PfCDPK3 also contains almost invariant amino acids corresponding to F185, G186, W222 and G225 whose functions are still unclear. The se-
quences (DIKPEN) in subdomain VI and (GTPYY-VAPQ) in subdomain VIII indicate that PfCDPK3 is a serine/threonine kinase rather than a tyrosine kinase [11]. However, a highly conserved E residue in the APE motif of subdomain VIII is replaced by a Q residue in PfCDPK3. A similar change was also found in PfCDPK2 [6]. It is not clear how this change would influence the kinase activity. The junction domain of PfCDPK3 comprises 31 amino acid residues and separates the catalytic and calmodulin-like domains. Sequence comparison of all known protozoan CDPKs showed less conservation in this region for PfCDPK3 (Fig. 3B). The junction region is thought to contain an autoinhibitory motif which acts as a pseudosubstrate, interacting with the normal substrate-binding region of the enzyme to block entry of the substrate and, therefore, inhibiting kinase activity [12,13]. The binding of Ca$^{2+}$ to the calmodulin-like domain is believed to release the pseudosubstrate from the active site of the catalytic domain by a mechanism that involves binding of the calmodulin-like domain to the autoinhibitory motif, therefore activating the kinase [14,15]. However, the basic-X-X-S/T consensus motif recognised by CDPKs [16] does not exist in the junction region of PfCDPK3. The sequences of the three P. falciparum kinases in the junction region are identical in only four of 31 residues, and all differ in at least 22 of 31 residues from the highly conserved junction sequences of plant CDPKs. This divergence of junction sequences within P. falciparum is similar to the situation in Paramecium [17] but is in sharp contrast to that in Arabidopsis, in which the junction region is strongly conserved; many Arabidopsis CDPKs are more than 90% identical in this region [18]. The LRVI sequence is thought to mediate an intramolecular interaction between the calmodulin-like domain and the junction region of the Arabidopsis CPK1 (AK1) [14]. However, only the I residue exists in PfCDPK3, although there are two residues (L and I) conserved in PfCDPK2 (see Fig. 3B). The calmodulin-like regulatory domain of PfCDPK3 consists of 151 residues and contains four putative EF-hand (helix-loop-helix) calcium-binding motifs, which are very similar to the EF-hands of calmodulin and other calcium-binding proteins. Each EF-hand of PfCDPK3 contains nearly all the requirements of such calcium-binding sites: six oxygen-containing ligands at positions 1, 3, 5, 7, 9 and 12, an invariant G residue at position 6 and a conserved aliphatic residue at position 8. The only exception is the highly conserved E residue at position 12 in the first EF-hand, which is replaced by a Q residue (see Fig. 3A). The refinement of the crystal structure of calmodulin has delineated the crucial role of this conserved E residue in calcium-binding [19]. In PfCDPK1, mutation of the conserved E residue to either K or Q in EF-hand 1 is deleterious and dramatically reduces the sensitivity of the Ca$^{2+}$-induced conformational change and the Ca$^{2+}$-dependent activation [20]. Therefore, it would be of great interest to examine whether the replacement in EF-hand 1 of PfCDPK3 would influence the Ca$^{2+}$-binding and the Ca$^{2+}$-dependent enzyme activity. Each of the four calcium-binding sites is flanked by residues predicted to form $\alpha$-helices (data not shown), as expected in a calcium-binding EF-hand. The presence of four EF-hands suggests that calcium would directly bind to PfCDPK3 and regulate its activity.

To investigate the structural organisation of the

Fig. 4. Southern blot analysis of the Pfcdpk3 gene. Four $\mu$g of genomic DNA from P. falciparum clone 3D7A was digested with restriction enzymes, electrophoresed on a 1.0% agarose gel, transferred onto a nylon membrane and probed with the CDB'-CD2 fragment of Pfcdpk3. Lanes 1–6 correspond to digests with AccI, BamHI, BclI, EcoRI, EcoRV and HindIII. The sizes of 1 kb DNA markers are given in kb to the left.
Pfcdpk3 gene in the P. falciparum genome, 3D7A genomic DNA was digested with a number of restriction enzymes and analysed by Southern blotting. Hybridisation of the CDB'-CD2 (see Fig. 1) probe revealed a single band in digests with AccI, BamHI, BclI and EcoRV, respectively, consistent with the restriction map (Fig. 4), suggesting strongly that Pfcdpk3 is encoded by a single copy gene in the parasite genome. However, two bands (one predominant band and the other faint) were detected in digests with EcoRI and HincII, respectively, contradictory to the restriction map (lanes 4 and 6 in Fig. 4).

The intensity of the faint bands decreased with higher stringency washing conditions (data not shown), indicating the presence of CDB'-CD2-related gene(s) in the P. falciparum genome. Indeed, two genes encoding PfCDPK1 and PfCDPK2, respectively, have recently been isolated from P. falciparum [6,7]. Whether the extra bands detected on the Southern blot represent these or other related gene(s) is unknown. Therefore, it is concluded that CDPKs exist as a multigene family in P. falciparum. A similar situation has been found in some plant species [21] and in the ciliated protozoan Paramecium [17]. So far, at least 20 CDPKs have been found in Arabidopsis thaliana [18,21], nine in maize [21,22], three in rice [23,24], three in soybean [25,26] and three in Paramecium tetraurelia [17]. CDPKs have also been isolated from other plants including mungbean, carrot, sweet potato and zucchini [21] and protozoans such as Eimeria tenella and Eimeria maxima [27] and Toxoplasma gondii (accession no. AF043629) as well as the unicellular algae Chlamydomonas moewusii [21].

In order to obtain some information on how Pfcdpk3 mRNA levels are regulated during parasite development and differentiation, a Northern blot containing equal quantities of total RNA prepared from cultures enriched in stage III to stage V gametocytes and from mixed asexual erythrocytic stages was probed with the CDB'-CD2 fragment. A single transcript of approximately 2900 nucleotides in size was detected only in the lane containing the sexual stage RNA, migrating between the 28S and 18S ribosomal RNA bands (Fig. 5A). The result suggests that PfCDPK3 is involved in sexual stage-specific events. To exclude the possibility that Pfcdpk3 could crossreact with the other two related genes (Pfcdpk1 and Pfcdpk2) on the Northern blotting, the same blot was hybridised with the PCR fragment (equivalent to the coding region of the CDB'-CD2 fragment for PfCDPK3) of Pfcdpk1 and Pfcdpk2, respectively (Fig. 5B,C). Interestingly, PfCDPK1 expressed not only in the asexual stage as described previously [7] but also in the sexual stage (see Fig. 5B), whereas PfCDPK2 expressed predominantly in the sexual stage (see Fig. 5C) rather than in the asexual stage [6]. In maize, it has been shown that a CDPK is specifically and developmentally expressed in pollen and required for germination and pollen tube growth [22]. In rice, expression of a CDPK gene has also been reported to be spatially and temporally regulated during seed development [23]. Interestingly, treatment of Plasmodium berghei and P. falciparum with Ca²⁺ antagonists such as TMB-8 (an inhibitor of intracellular Ca²⁺ release) and W-7 (a calmodulin inhibitor) strongly inhibited exflagellation, but
EGTA (a Ca²⁺ chelator) and nicardipine (a Ca²⁺ channel inhibitor) had no effect, indicating that mobilisation of the parasite internal resources of Ca²⁺ is a prerequisite for exflagellation [28]. It has also been shown that DNA synthesis and axoneme formation in male gametocytes may be regulated by Ca²⁺/calmodulin [29]. Taken together, we propose that PfCDPK3 (probably PfCDPK2 as well) may serve as a link between Ca²⁺ and gametogenesis of *P. falciparum*. Identification of the upstream regulators and downstream substrates of PfCDPK3 will afford new insight on the regulatory mechanisms of sexual stage-specific processes. In addition, CDPK does not seem to exist in vertebrates, PfCDPK3 may, therefore, represent a promising target for development of new anti-malarial drugs.

Acknowledgements

We are very grateful to Mr. Richard J. Carrier and Mr. Phillip J. Howard for their critical reading of the manuscript. This work was supported in part by an equipment grant from the Royal Society (to L.S. Cox). D.A. Baker was supported by the Wellcome Trust. J.-L. Li was supported by the Cancer Research Campaign.

References


