The aspartic proteinase from the rodent parasite *Plasmodium berghei* as a potential model for plasmepsins from the human malaria parasite, *Plasmodium falciparum*

Michelle J. Humphreys[a], Richard P. Moon[b], Annette Klinder[a], Sylvia D. Fowler[a], Katharina Rupp[b], Daniel Bur[b], Robert G. Ridley[b], Colin Berry[a],*

[a] *Cardiff School of Biosciences, Cardiff University, P.O. Box 911, Cardiff CF1 3US, UK*
[b] *Hoffmann-La Roche, Pharmaceuticals Division, Pharma Research Pre-Clinical, Basel, Switzerland*

Abstract The gene encoding an aspartic proteinase precursor (proplasmepsin) from the rodent malaria parasite *Plasmodium berghei* has been cloned. Recombinant *P. berghei* plasmepsin hydrolysed a synthetic peptide substrate and this cleavage was prevented by the general aspartic proteinase inhibitor, isovaleryl pepstatin and by Ro40-4388, a lead compound for the inhibition of plasmepsins from the human malaria parasite *Plasmodium falciparum*. Southern blotting detected only one proplasmepsin gene in *P. berghei*. Two plasmepsins have previously been reported in *P. falciparum*. Here, we describe two further proplasmepsin genes from this species. The suitability of *P. berghei* as a model for the in vivo evaluation of plasmepsin inhibitors is discussed.
© 1999 Federation of European Biochemical Societies.

Key words: Aspartic proteinase; Plasmepsin; Malaria; *Plasmodium berghei*; *Plasmodium falciparum*

1. Introduction

The aspartic proteinases known as plasmepsins have been officially recognised by the World Health Organisation as suitable targets for the design of novel chemotherapeutic compounds for the treatment of malaria [1]. Enzymes of this class initiate the haemoglobin breakdown pathway that provides intra-erythrocytic malaria parasites with nutritional resources [2] and inhibition of their activity results in death of the malarial parasite [3,4]. As drugs are being designed to block the action of the malarial aspartic proteinases specifically, an essential safety component of the drug development process is the need for rigorous testing of such compounds for toxicity. The availability of an animal model, such as the rodent malaria parasite *Plasmodium berghei*, in which the potency and toxicity of lead compounds in vivo could be assessed, would thus be invaluable. In this study, we describe the cloning, expression and analysis of a plasmepsin from *P. berghei* and consider its relationship to its counterpart enzymes from other *Plasmodium* species, particularly the most dangerous human pathogen, *Plasmodium falciparum*.

2. Materials and methods

2.1. Cloning of the gene encoding the precursor of P. berghei plasmepsin (PbPM)

A library of *Hin*dIII-digested *P. berghei* genomic DNA cloned into pUC8 was a kind gift from Dr J. Dalton (Dublin City University). A clone encoding part of the PbPM sequence was identified by probing this library with the *P. falciparum* proplasmepsin II gene that was radiolabelled by random primed synthesis using the Megaprime kit (Amersham International) according to the manufacturer's instructions. The sequence of the *P. berghei* inserted DNA was determined and was found to correspond to the 3′ end of a gene encoding an aspartic proteinase. In order to clone further sequences 5′ to this initial fragment, a polymerase chain reaction (PCR)-based strategy was employed. A reverse primer CB33 (gcgagcatgcggatccTTAGTT-TTTTGCAACTGCAAAACC) was designed from the sequence of the 3′ end of the gene (capital letters) including its stop codon and which incorporated *Sph*I and *Bam*HI restriction sites (lower case letters) to facilitate subsequent cloning experiments. A degenerate forward primer CB30 (CWKCWAATKTATGGGTNCC) was designed to correspond to a region of high DNA sequence conservation between the genes encoding *P. falciparum* proplasmepsin I and proplasmepsin II. The CB30 and CB33 primers were used in a PCR amplification with *P. berghei* total genomic DNA as template and the approximately 880 bp product of this reaction was cloned into pUC18 using the Sureclone PCR cloning kit (Pharmacia). The inserted *P. berghei* sequence was determined and a further specific oligonucleotide (GGTGTACCATCTTTTTCATATG) was designed. This primer was used in conjunction with the vector-based reverse primer (CAGGAAACAGCTATGAC) in a PCR reaction to amplify the 5′ end of the proplasmepsin gene from the original, pooled *Hin*dIII library. The resulting 815 bp fragment was cloned using the Sureclone kit as above and the inserted DNA was sequenced.

2.2. Southern blot analysis

A ∼1.1 kb fragment released by *Bam*HI digestion from the plasmid which encodes semi-proplasmepsin I from *P. falciparum* [3] was radiolabelled and used as a probe to screen for the presence of proplasmepsin genes in *P. berghei*. Total *P. berghei* DNA was digested with the enzymes *Dra*I, *Ssp*I, *Eco*RI and *Hin*dIII and fragments were separated on a 0.8% agarose gel prior to Southern blotting. The blot was washed under low stringency conditions (2×SSC, 55°C) prior to autoradiography.

2.3. Production of wild-type and mutant forms of semi-proPbPM

A segment of the proPbPM gene encoding a truncated form of the precursor zymogen (semi-proPbPM: the sequence from the codon for residue 77 in the propart to the stop codon of proPbPM) was amplified from *P. berghei* total DNA template using the 5′ primer CB42 (ccggaattcggaTCCGAATATTTAACAATTCG) and the 3′ primer CB33 above. The amplified fragment was digested with *Bam*HI, cloned into pET3a and its sequence was determined to ensure that no mutations had arisen during PCR. A mutant form carrying a Lys-Ile substitution at residue 111p in the propart was produced by an overlapping PCR method as follows. Two complementary oligonucleotide primers were synthesised (spanning nucleotides 487–512 in

*Corresponding author. Fax: (44)-29-20874116.
E-mail: berry@cf.ac.uk

the *P. berghei* sequence) containing a single nucleotide substitution (underlined) to introduce the desired mutation. These primers were: MH5, GAATCATTCA<u>T</u>ATTATTAAAATCAGG (top strand); MH6, CCTGATTTTAATAAT<u>A</u>TGAATGATTC (bottom strand). PCR was used to amplify the 5′ end of the mutant gene using primer CB42 and the bottom strand primer MH6. The 3′ end of the gene was amplified using the top strand primer MH5 and CB33. Finally, the full-length mutated gene was generated in a third PCR in which the products of the first two reactions were purified, mixed and amplified using primers CB42 and CB33. The resulting mutant gene was cloned into the *Bam*HI site of pET3a and the inserted DNA was sequenced to ensure that the desired mutation had been introduced and that the rest of the sequence remained unaltered.

Wild-type and mutant constructs were transformed separately into *Escherichia coli* BL21(DE3)pLysS. Gene expression and production and purification of the two forms of recombinant semi-proPbPM were carried out essentially as described previously for the counterpart protein from *P. falciparum* [3].

2.4. Activation of wild-type and mutant precursors

Wild-type and mutant forms of semi-proPbPM were incubated for 1 h at 37°C at a variety of pH values. The buffers used were 0.1 M sodium acetate (pH 3.0–4.4) or 0.1 M sodium acetate (pH 4.5–5.5), with the ionic strength adjusted where necessary to 0.1 M by the addition of NaCl. The samples were then subjected to sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) to determine whether a reduction in molecular weight, consistent with autoactivation by removal of the propart, had occurred. Samples were also prepared for N-terminal sequencing by separation on SDS-PAGE (12% gel) prior to blotting onto a polyvinylidene difluoride membrane. The membrane was then stained with amido black and the appropriate band was excised and subjected to Edman degradation using a 494 Procise model sequencer equipped with a narrow bore HPLC system (Perkin Elmer) [5,6].

2.5. Proteinase assays

Aliquots of each form of semi-proPbPM zymogen were mixed with 0.1 volume of 1 M sodium acetate buffer, pH 4.6, and incubated at room temperature for 30 min prior to measurement in duplicate of proteinase activity using the fluorogenic peptide substrate EDANS-Asp-Arg-Leu-Glu-Arg-Thr-Phe*Leu-Ser-Phe-Thr-Thr-Asp-Arg-DAB-SYL at a final concentration of 0.5 μM in 100 μl of 100 mM sodium acetate buffer, pH 4.6, in a 96 well microtitre plate. Fluorescence was monitored (excitation at 350 nm, detection at 560 nm) in a Cytofluor 2350 plate reader (Millipore AG, Switzerland) after 225 min. Control reactions were performed using substrate in the absence of enzyme and the absorbance values were subtracted from those obtained in the presence of enzyme. Samples were also pre-incubated with the inhibitors isovaleryl pepstatin (IVP) or Ro40-4388 [3] (50 μM final concentration) for 5 min before reactions were started by the addition of the substrate. Since the substrate was dissolved in 50% (v/v) DMSO and inhibitor stocks were dissolved in 100% (v/v) DMSO, the concentration of DMSO in all assays was standardised to a final concentration of 1.5% (v/v).

2.6. Molecular modelling of PbPM

A three dimensional model of mature PbPM was constructed, using the crystal structure of human cathepsin D [7] as the template structure, on a Silicon Graphics Indy R4600 workstation using Moloc, the Hoffmann-La Roche in-house molecular modelling package [8]. The quality of the model was checked with a program developed by Luethy et al. [9].

2.7. Database searches and sequence alignments

The proPbPM protein sequence was used in BLAST searches [10] to search various databases for other proplasmepsin sequences. Preliminary sequence data for *P. falciparum* chromosome 14 were obtained from The Institute for Genomic Research website (www.tigr.org). Sequencing of chromosome 14 was part of the International Malaria Genome Sequencing Project and was supported by awards from the Burroughs Wellcome Fund and the U.S. Department of Defense. Other proplasmepsin sequences (together with the sequences of zymogens of an aspartic proteinase-like protein, HAP, from *P. falciparum* [11] and of an aspartic proteinase from the protozoan parasite *Eimeria acervulina*) were derived from the Swiss-Prot and EMBL databases (Table 1). Sequences were aligned initially using the Pileup program in the GCG package before minor modifications were made manually from a knowledge of aspartic proteinase structures.

3. Results and discussion

3.1. Sequence and activation

The DNA sequence encoding *P. berghei* proplasmepsin has been deposited in the EMBL database under accession number AJ223308 (Table 1). It predicts a protein of 450 amino acids (Fig. 1) containing all the features typical of an aspartic proteinase but with a much longer propart (124 residues) than archetypal zymogens such as pepsinogen (59 amino acids including the signal peptide). A truncated form of the *P. berghei* zymogen (semi-proPbPM) with a shortened propart, similar in length to that of pepsinogen, was produced in *E. coli*. However, the wild-type form of the truncated precursor did not undergo autoactivation when incubated over a range of acidic pH values (results not shown). In order to facilitate autoactivation of semi-proPbPM, it was therefore necessary to introduce a suitable self-processing site between the propart and the mature enzyme regions of the precursor. The wild-type protein sequence was changed from Phe(110p)-Lys(111p)-Leu(112p)-Leu(113p)-Lys(114p) to Phe(110p)-Ile(111p)-Leu-(112p)-Leu(113p)-Lys(114p) (p denotes that these are propart residues) in order to introduce an isoleucine residue into the P_2 position of the putative $\sim P_2$-P_1*P_1'-P_2' $\sim$ autoactivation site (nomenclature of Schechter and Berger [12]) since it was predicted from the model of the PbPM active site that this would be accommodated more efficiently than the wild-type lysine. A comparable mutation introduced into semi-proplasmepsin I from *P. falciparum* resulted in successful autoactiva-

Table 1
Database entries used in sequence alignment

Source organism	Protein	Accession no.	Reference
P. berghei	ProPbPM	AJ223308[a]	This study
P. falciparum	Proplasmepsin I	X75787[a]/P39898[b]	[4]
P. falciparum	Proplasmepsin II	L10740[a]/P46925[b]	[13]
P. falciparum	Proplasmepsin III	AL008970[a]	[17]
P. falciparum	Proplasmepsin IV	Chromosome 14, contig >138[c]	Unpublished
P. falciparum	ProHAP	AJ009990[a]	[11]
E. acervulina	EaAP	Z24676[a]	[14]

The alignment shown in Fig. 1 was generated from sequences for the indicated proteinase using accession numbers for EMBL database, Swiss-Prot and The Malaria Genome Project contig number.
[a]EMBL database.
[b]Swiss-Prot.
[c]The Malaria Genome Project contig number.

```
                                         40p                        60p
ProPbPM      MEYSEKESNYSNGLMRNGSAFGHLKFDNIK-SFKIQKKFQALYFIIFICIIGSIFAYMVGTNYYSTKTIDIDKII
ProPfPM I    AL  I   -DF SAFAK E  VNSST N NMKTW    R  I  VFF LL T AL Y LID- VLFP NKK NE M
ProPfPM II   DITV  HDFKH FIKSN T DG NI  S NKK    G  I  VLL CSVMCGL Y VYE- VWLQRDNEMNE L
ProPfPM IV   ALTV  EEF  T IK A    DR LG L -NL      L F  L L VL T VF FFLI - F --SHRKLYQV
Pro HAP      NLTI  EDFTNTF K EES NTFRVTKV R-WNAKRL KI FVTV  VLA GFSY IFE- FVFQ NRK NH
ProPfPM III  ----------------------------------------MP H  LY LI-- CVL HICPIH LN FKND--
Ea AP        -------------------------------------------MRSLLVVA --L GCSSFAPTDARHRFLSETL

                                         110p⇓      120p
ProPbPM      ANSEYLTIRAKIERPRDKLFKKVMNKNVSNYIKESFKLLKSGLLKKEHL--TKYEDSI-----ELDQSLGLSFIG
ProPfPM I    NT  KHVI GFS  NSH RIM T KQHRLK       L FF T  TQ PH --GNAG  V-----T NDVANVMYY
ProPfPM II   K    H   GF V NAH RIL TIKTHKLK       VNF N   T TNY --GSSN N ----- VDFQNIM Y
ProPfPM IV   K TKHT  GF D  H  VLSS LKNKL T V       FF   YAQ GY --GSEN    ----- DVAN M Y
Pro HAP      KT K S VGFN  NSY R M TIKEHKLK       V FNK  T  SY --GSEF NV----- KDLANVLSF
ProPfPM III  E EKG-SLNIPLGKENNLF N---EIKLE RF NNI GYIQNVQ FHY MEKNKPNVLSYIQED LNFHNSQ  A
Ea AP        EEP D-VMLKTADLHTNL REPP TIKLD RY FTGLGELVSQ IDH TTMGSVGS GTMARQK LNYHNSQYF


ProPbPM      TAELGNNKQSFSFILDTGSANLWVPSKECKSGGCAYKHRYDSSASNTYEK------DGTPVSILYGSGGIKGFFS
ProPfPM I    E QI D   K A  F        AQ NTI   KT NL   NK K      ------  K EMN V  TVS
ProPfPM II   D  V D Q P T          VK TTA  LT  L    K R      ------  K EMN V  TVS
ProPfPM IV   EGQI T   P M  F        VN D I  ST  L   A   KS   ------  K E S    TVR Y
Pro HAP      E K   D G K N LFH A S V     IK T ES ES NH    K K      ------ D   KLTSKA T S I
ProPfPM III  DIGV  PP V KVVF     S  AI  TK IK      SHKKFNPNK R FT NLKNNQESVYTY Q  T TSILEQ
Ea AP        EIKI TPGRR VVVF     S     AA  EK     PHEKF PKY S FSPIRSLTG PAVAF Q  T ACVLRMG


ProPbPM      NDMFTIGHHTIPYKFIEVTETDDLEPIYTASPFDGILGLGWKSLAVGNVE-------PVIVEMKKTGQIENAVFS
ProPfPM I    K IV  ANLSF        D NGF  A  LGQ    V      D SI S D------- V  L NQNK  Q   T
ProPfPM II   K LV V NLSL       ID NGF  T   T         D SI S D------- IV  L NQNK    L T
ProPfPM IV   K VISL DLSL        DA       SG E        D SI SID------- V  L  QNK D  L T
Pro HAP      K LV   KLSV     M  IVGF  F SE DV  VF     D SI SID------- Y  L TQNK  Q  Y
ProPfPM III  Y DVYLKGLK KHQC GLAIEES H -FSDL      V   FSDPDFRSQN---KYAS L ETI   QNLLKRNI
Ea AP        R IVE  GIKV NQA GLAVEESTH -FADL     LV    F--PD SGE GLPSSAL IVDQ V EKVLDRN

                                                                         215
ProPbPM      FYLPEADDSVGYFTIGGIEESF-YTGD-VTYEKLTNESYWQINL------DVDFGIVALNNANAIVDSGTSAITA
ProPfPM I        FD KHK  L      DR - E Q-L     NHDL    VD ------ LH  NLTVEK T        S
ProPfPM II       VH KHT FL      R - E P-L     NHDL     T ------ AHV NIM EK  C         V
ProPfPM IV       VH KH   L      SD - E P-L     NHDL   D ------ IH  KYVMQK   V      T
Pro HAP      I   PENKNK  L      R -FD P-LN    NHDLM VD ------  H  N SSKK  V L  A  V  V
ProPfPM III      V KKLEKS AI F KANKKYTVE KSIEWFPVISLY  E    LDIQLSHKNLFLCESKKCR AI T S L  G
Ea AP        V MS DINRP EISF AADPKYTFA HTPKWFPVISLD  E G HGMKINGKS VCEKRGCR A  T S L  G


ProPbPM      PSDFLNKFLNSIMSIPVPFIPLRIVLCHDTN-LPTLKFRSKS-----TEYTIEPKHYLIELDPV--G-GICAISL
ProPfPM I      TE     FEGLDVVKI  L  Y TT NNPK-    E   AT-----NV  L  EY  QQIFDF-- ISL MV I
ProPfPM II     T       M QNLDV K  L FYVT NNSK-    FE T EN-----GK  L  EY  QHIED -- P L MLNI
ProPfPM IV     TS      FRDMNV K  L  YVTT DNDD-    E H RN-----NK  L  EF MDP SDI--DPAL MLYI
Pro HAP        TE F Q VE ASVFK   LS YVTT GN K-    EY  PN-----KV  L   Q  EP ENI--FSAL MLNI
ProPfPM III      T IQPL EK NLER--------D SNKES  IIS VL NVEGKEITLDFM ED I  EGDTENNTLE V GI
Ea AP            SVI PLIKALNVAE---------N SNLGT    T VL DIYGRLVNFSL RD VV ELDARGNPNN  AGF


             290 ****
ProPbPM      IDVDIDANT---FILGDVFFKKYYTIFDYDNSRVGFAVAKN
ProPfPM I      P  LNK  ---     P MR  F V    HT    L  KKL
ProPfPM II     GL FPVP ---     P MR  F V    HS  I L  KNL
ProPfPM IV   LP    D   ---     P MR  F V   EKES      L
Pro HAP      VPI LEK   --- V   P MR  F VY   HT    L  L
ProPfPM III  MPL VPPPRGPI  F NS IR       N HKLI LIE NHNF
Ea AP        MAM VP PRGPL V  NS IR    S   R HMM   MR NHEGSGPLIKGYPSSAPSVSASCLVAASAAAFALSLF
```

Fig. 1. Alignment of the deduced amino acid sequence of the plasmepsins from *P. berghei* with those of other aspartic proteinases. Sequences listed in Table 1 were aligned as described in Section 2. The proPbPM sequence is given in full. In the other sequences, spaces indicate identity to the equivalent residue in proPbPM, hyphens indicate gaps introduced to optimise the alignment. ProPfPM I, II, III and IV = the *P. falciparum* proplasmepsins I [4], II [13] and the new proplasmepsins III and IV identified in this study. ProHAP = the histo-aspartic protein precursor [11] and EaAP is the precursor of the aspartic proteinase from *E. acervulina*. The location of the lysine residue 111p that was mutated in semi-proPbPM is marked with an ⇓. The polyproline loop residues 292–295 (pepsin numbering) are marked with *. Other residues mentioned in the text are numbered according to the pig pepsin numbering system.
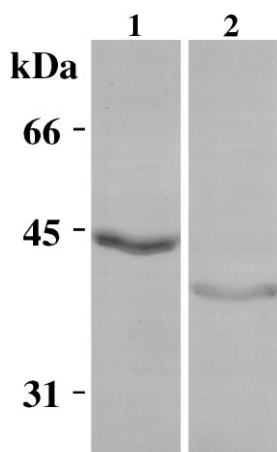
Fig. 2. Autoactivation of Lys111pIle mutant semi-proPbPM. Samples of purified, recombinant mutant semi-proPbPM were analysed by SDS-PAGE under reducing conditions before (lane 1) and after (lane 2) incubation in 0.1 M sodium acetate buffer, pH 4.6 for 1 h at 37°C.

tion [3]. The mutated semi-proPbPM was found to have an M_r of 43 kDa on SDS-PAGE under reducing conditions. The N-terminal sequence determined for this precursor band was Ala-Ser-Met-Thr-Gly-Gly-Gln-Gln-Met-Gly-Arg-Gly~, which corresponds to the N-terminal leader sequence introduced from the pET3a vector. When aliquots of this protein were incubated at pH 4.6, the mutant semi-proPbPM exhibited a shift in size from 43 to 37 kDa (Fig. 2). N-terminal sequencing of the 37 kDa band gave equal amounts of two overlapping sequences which could be resolved by reference to the predicted amino acid sequence of proPbPM. The first, Leu-Lys-Ser-Gly-Leu-Leu-Lys-Lys-Glu-His-Leu-Thr~ (amino acids 113p–124p; Fig. 1), represents cleavage at the intended Leu(112p)*Leu(113p) bond in the mutant activation sequence (~Ile(111p)-Leu(112p)*Leu(113p)-Lys(114p)-Ser(115p)-Gly(116p)~). The second sequence, Ile-Leu-Leu-Lys-Ser-Gly-Leu-Leu-Lys-Lys-Glu-His~ is two residues longer than the first and indicates that activation had also occurred to an equal extent at another cleavage point created by the mutation of residue 111p to isoleucine in the sequence ~Phe(110p)*Ile(111p)-Leu(112p)-Leu(113p)-Lys(114p)-Ser-(115p)-Gly(116p)~.

3.2. Activity of mature PbPM

Wild-type semi-proPbPM was inactive when incubated with the fluorogenic substrate EDANS-Asp-Arg-Leu-Glu-Arg-Thr-Phe*Leu-Ser-Phe-Thr-Thr-Asp-Arg-DABSYL both before and after exposure to acidic conditions (not shown). Similarly, the Lys111pIle mutant precursor (43 kDa) did not show any activity against this peptide prior to exposure to the acidic

Table 2
Substrate cleavage and inhibition of the plasmepsin from *P. berghei*

Inhibitor added (50 μM)	Activity (arbitrary fluorescence units)
None	21
Pepstatin	1
Ro40-4388	1

Samples of mature PbPM were incubated for 225 min at pH 4.6 with EDANS-Asp-Arg-Leu-Glu-Arg-Thr-Phe*Leu-Ser-Phe-Thr-Thr-Asp-Arg-DABSYL as substrate, either alone or in the presence of 50 μM inhibitors IVP or Ro40-4388 [3].

activation conditions. In contrast, the mature PbPM (37 kDa) that resulted from autoactivation of mutant semi-proPbPM readily hydrolysed the fluorogenic substrate (Table 2) and at a rate comparable to that measured previously for plasmepsin I from *P. falciparum* (Moon, unpublished). Substrate cleavage was prevented not only by the general aspartic proteinase inhibitor IVP, but also by the compound Ro40-4388 (Table 2) which we have reported previously to be able to kill *P. falciparum* maintained in human red blood cells in culture [3]. Thus, the aspartic proteinase from *P. berghei* appears to have a susceptibility to inhibition and an activity which are similar to its counterpart enzyme(s) from the major human pathogen *P. falciparum*. On this basis, *P. berghei* appears worthy of consideration as a model species for evaluation of inhibitor potency and toxicity. However, two plasmepsins have been characterised previously from *P. falciparum*, plasmepsin I [4] and plasmepsin II [13], together with a third gene encoding a closely related protein, HAP (which shares >60% identity in the mature region with plasmepsin I and plasmepsin II but lacks one of the active site aspartic acid residues necessary for the activity of the aspartic proteinase class of enzymes [11]). Southern blot analysis of *P. berghei* DNA (Fig. 3) revealed only one hybridising band in each lane, indicating that this parasite species contains only one gene encoding a proplasmepsin.
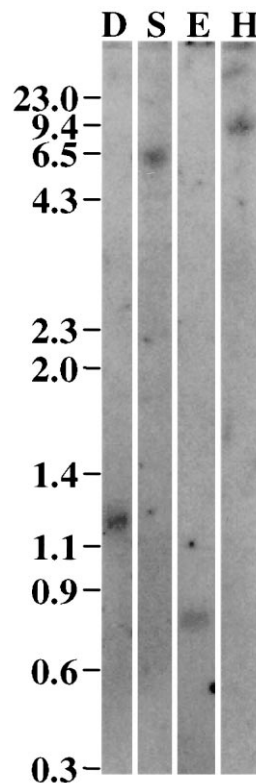

Fig. 3. Southern blot of *P. berghei* DNA. Total *P. berghei* genomic DNA was cut with the restriction enzymes *Dra*I (D), *Ssp*I (S), *Eco*RI (E) or *Hin*dIII (H) and the products were separated in a 0.8% agarose gel. After transfer, the blot was hybridised with plasmepsin I probe DNA and washed as described in Section 2. Marker sizes are shown in kb.

3.3. New plasmepsins in P. falciparum

With the realisation that *P. berghei* contrasts with *P. falciparum* in containing only one aspartic proteinase (-like) gene, we used the proPbPM sequence in BLAST searches of the ever-expanding databases derived from the *P. falciparum* genome sequencing projects. In this way, we established the presence of genes encoding two yet further aspartic proteinases (plasmepsins III and IV) in *P. falciparum*.

Fig. 1 shows a multiple alignment of the sequence of proPbPM with the four aspartic proteinase zymogens from *P. falciparum*, in addition to *P. falciparum* proHAP and the precursor of an aspartic proteinase from the parasite *E. acervulina* (accession numbers and references for all sequences are listed in Table 1). ProPbPM clearly differs from proHAP since the latter lacks one of the active site Asp residues necessary for the mechanism of action of an aspartic proteinase [11]. In addition, proPbPM has only a low level of identity (25%) with proplasmepsin III. The proplasmepsin III precursor is itself quite distinct from the other *Plasmodium* proteins aligned in Fig. 1 since it has ≤25% identity to any of the others and has the sequence ∼Asp-215-Thr-216-Gly-217∼ flanking the catalytic Asp(215) residue. In contrast, plasmepsins I, II and IV and the plasmepsin from *P. berghei* all have an ∼Asp-Ser-Gly∼ sequence for this hallmark motif (Fig. 1; [4,13]). The proplasmepsin III sequence is most closely related (42% identity) to the precursor of an aspartic proteinase from another apicomplexan parasite, *E. acervulina* [14]. Both have considerably shorter proparts (∼80 residues) than proplasmepsins I, II, IV and the proPbPM identified in this study (Fig. 1). They appear to have a signal peptide, as observed in all vertebrate/fungal aspartic proteinase precursors [15], whereas the proparts of proplasmepsins I, II and IV and proPbPM all contain a membrane-spanning stretch of 20 hydrophobic amino acids (residues 38p–59p) which has been implicated in the type II membrane anchoring of these zymogens [16]. In addition, inspection of the locus of the proplasmepsin III gene within the completed sequence of chromosome 3 from *P. falciparum* reveals that the coding exons are interrupted by no less than 14 introns. This is in itself uncommon in *P. falciparum* but is totally unprecedented with regard to the other aspartic proteinase genes from *P. falciparum* or *P. berghei*. These coding sequences are not interrupted by any introns whatsoever.

In contrast, proplasmepsin IV from *P. falciparum* appears to be closely related to the previously identified proplasmepsins I and II and indeed to the proplasmepsin from *P. berghei* described in this study (50–55% identity in each case). Distinction amongst these enzymes is possible, however, by examination, for example, of the sequence of residues 292–295 (Fig. 1) which has frequently been referred to as the 'polyproline loop' in vertebrate/fungal aspartic proteinases because of the predominance of Pro residues at this location in many enzymes. Plasmepsin II (Fig. 1) has the sequence ∼FPVP∼, whereas plasmepsin I has ∼LNKN∼ containing a basic residue, whilst plasmepsin IV has ∼IDDN∼ with two acidic Asp residues. In the PbPM sequence of ∼IDAN∼, there is only a single acidic residue in this region. Since the polyproline loop influences the nature of substrate residues accommodated within the S_2' and S_3' subsites of the active site cleft of aspartic proteinases, these sequence variations in this region of plasmepsins I, II, IV and PbPM are likely to confer subtly different substrate specificities on each of these enzymes. Thus, on this basis alone, it is not possible to categorise the

plasmepsin from *P. berghei* as the direct counterpart of any one of the *P. falciparum* proteinases. Furthermore, since the rodent malaria parasite appears to have only one gene encoding a proplasmepsin, it seems that *P. berghei* does not represent an ideal model system for the in vivo analysis of inhibitors targeted against the battery of plasmepsins in the dangerous human pathogen, *P. falciparum*. A study of the localisation of PbPM and elucidation of the expression patterns of its gene throughout the life-cycle stages of the rodent parasite will be necessary to determine whether PbPM is able by itself to fulfil all of the functions that apparently require the attentions of four closely related plasmepsins (and HAP) within *P. falciparum*. From the results presented here, *P. berghei* may have a different dependence on aspartic proteinases from its most deadly human counterpart, *P. falciparum*.

References

[1] World Health Organisation (1996) *P. falciparum* Proteinases Inhibitor Development, World Health Organisation, Geneva.

[2] Goldberg, D.E., Slater, A.F.G., Beavis, R., Chait, B., Cerami, A. and Henderson, G.B. Hemoglobin degradation in the human malaria pathogen *Plasmodium falciparum*: a catabolic pathway initiated by a specific aspartic protease, (1991) J. Exp. Med. 173, 961–969.

[3] Moon, R.P., Tyas, L., Certa, U., Rupp, K., Bur, D., Jaquet, C., Matile, H., Loetscher, H.-R., Grueninger-Leitch, F., Kay, J., Dunn, B.M., Berry, C. and Ridley, R.G. Expression and characterisation of plasmepsin I from *Plasmodium falciparum*, (1997) Eur. J. Biochem. 244, 552–560.

[4] Francis, S.E., Gluzman, I.Y., Oksman, A., Knickerbocker, A., Mueller, R., Bryant, M.L., Sherman, D.R., Russell, D.G. and Goldberg, D.E. Molecular characterisation and inhibition of a *Plasmodium falciparum* aspartic hemoglobinase, (1994) EMBO J. 13, 306–317.

[5] Hewick, R.M., Hunkapillar, M.W., Hood, L. and Dreyer, W.J. A gas-liquid, solid phase peptide and protein sequenator, (1981) J. Biol. Chem. 256, 7990–7999.

[6] Matsudira, P. Sequence from picomole quantities of proteins electroblotted on to polyvinylidene difluoride membranes, (1987) J. Biol. Chem. 262, 10035–10038.

[7] Baldwin, E.T., Bhat, T.N., Gulnik, S., Hosur, M.V., Sowder, R.C., Cachau, R.E., Collins, J., Silva, A.M. and Erickson, J.W. Crystal structures of native and inhibited forms of human cathepsin D: implications for lysosomal targeting and drug design, (1993) Proc. Natl. Acad. Sci. USA 90, 6796–6800.

[8] Gerber, P.R. and Mueller, K. MAB, a generally applicable molecular force field for structure modelling in medicinal chemistry, (1995) J. Comput. Aided Mol. Des. 9, 251–268.

[9] Luethy, R., Bowie, J.U. and Eisenberg, D. Assessment of protein models with three-dimensional profiles, (1992) Nature 356, 83–85.

[10] Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. Basic local alignment search tool, (1990) J. Mol. Biol. 215, 403–410.

[11] Berry, C., Humphreys, M.J., Matharu, P., Granger, R., Horrocks, P., Moon, R.P., Certa, U., Ridley, R.G., Bur, D. and Kay, J. A distinct member of the aspartic proteinase gene family from the human malaria parasite *Plasmodium falciparum*, (1999) FEBS Lett. 447, 149–154.

[12] Schechter, I. and Berger, A. On the size of the active site in proteases. I. Papain, (1967) Biochem. Biophys. Res. Commun. 27, 157–162.

[13] Dame, J.B., Reddy, G.R., Yowell, C.A., Dunn, B.M., Kay, J. and Berry, C. Sequence, expression and modelled structure of an aspartic proteinase from the human malaria parasite *Plasmodium falciparum*, (1994) Mol. Biochem. Parasitol. 64, 177–190.

[14] Laurent, F., Bourdieu, C., Kaga, M., Chilmonczyk, S., Zgrebski, G., Yvore, P. and Pery, P. Cloning and characterization of an *Eimeria acervulina* sporozoite gene homologous to aspartyl proteinases, (1993) Mol. Biochem. Parasitol. 62, 303–312.

[15] Koelsch, G., Mares, M., Metcalf, P. and Fusek, M. Multiple functions of proparts of aspartic proteinase zymogens, (1994) FEBS Lett. 343, 6–10.

[16] Francis, S.E., Banerjee, R. and Goldberg, D.E. Biosynthesis and maturation of the malaria aspartic hemoglobinases plasmepsins I and II, (1997) J. Biol. Chem. 272, 14961–14968.

[17] Bowman, S., Lawson, D., Basham, D., Brown, D., Chillingworth, T., Churcher, C.M., Craig, A., Davies, R.M., Devlin, K., Feltwell, T., Gentles, S., Gwilliam, R., Hamlin, N., Harris, D., Holroyd, S., Hornsby, T., Horrocks, P., Jagels, K., Jassal, B., Kyes, S., McLean, J., Moule, S., Mungall, K., Murphy, L., Oliver, K., Quail, M.A., Rajanream, M.-A., Rutter, S., Skelton, J., Squares, S., Sulston, J.E., Whitehead, S., Woodward, J.R., Newbold, C. and Barrel, B.G. The complete nucleotide sequence of chromosome 3 of *Plasmodium falciparum*, (1999) Nature 400, 532–538.