

Isolation and characterization of serine protease gene(s) from *Perkinsus marinus*

Gwynne D. Brown, Kimberly S. Reece*

Virginia Institute of Marine Science, PO Box 1346, Gloucester Point, Virginia 23062, USA

ABSTRACT: This study reports the first serine protease gene(s) isolated from *Perkinsus marinus*. Using universal primers, a 518 bp subtilisin-like serine protease gene fragment was amplified from *P. marinus* genomic DNA and used as a probe to screen a λ -phage *P. marinus* genomic library; 2 different λ -phage clones hybridized to the digoxigenin(DIG)-labeled subtilisin-like gene fragment. Following subcloning and sequencing of the larger DNA fragment, a 1254 bp open reading frame was identified and later confirmed, by 5' and 3' random amplification of cDNA ends (RACE) and northern blot analysis, to contain the entire coding-region sequence. Sequence analysis of the 3' RACE results from 2 isolate cultures, VA-2 (P-1) and LA 10-1, revealed multiple polymorphic sites within and among isolates. We identified 2 different types of cDNA clones with 95.53% nucleotide sequence similarity, suggesting the possibility of 2 closely related genes within the *P. marinus* genome. Southern blot analysis of genomic DNA from 12 genetically distinct *P. marinus* isolate cultures revealed 2 different banding patterns among isolates.

KEY WORDS: *Perkinsus marinus* · Subtilisin · Serine protease

Resale or republication not permitted without written consent of the publisher

INTRODUCTION

Interest in serine protease genes of pathogenic organisms has developed due to the importance of serine proteases in parasite evasion of host defense mechanisms (Chaudhuri et al. 1989), invasion of host tissue (Banyal et al. 1981, Blackman et al. 1998, Hackett et al. 1999), parasite metabolism and parasite growth (McKerrow et al. 1993). Serine proteases detected in the extracellular products (ECP) of *in vitro* *Perkinsus marinus* cultures (La Peyre et al. 1995) have been implicated in the progression of Dermo, a devastating disease of the eastern oyster *Crassostrea virginica*. Secretion of low molecular weight serine proteases was found to be greater following growth in media supplemented with extract from *C. virginica* oyster homogenate than in media supplemented with extract from the more resistant oyster species *C. gigas* and *C. ariakensis* (MacIntyre et al. 2003). Lytic factors present in ECP have also been found to increase infection intensity, degrade fibronectin and laminin (La Peyre et al. 1996), and suppress some immune functions (Gar-

reis et al. 1996, Tall et al. 1999). There are no reports, however, of any serine protease gene sequences isolated from *P. marinus*.

The use of degenerate primers in the polymerase chain reaction (PCR) to amplify serine protease gene fragments (Sakanari et al. 1989, Blackman et al. 1998) has facilitated the isolation of serine protease genes from multiple organisms including *Neospora caninum* (Louie & Conrad 1999), *Plasmodium* species (Blackman et al. 1998, Hackett et al. 1999), and *Trypanosoma* spp. (Sakanari et al. 1989, Morty et al. 1999). The subtilisin-like (sub)genes from *Neospora caninum* (*ncsub1*) (Louie & Conrad 1999, Louie et al. 2002), *Toxoplasma gondii* (*tgsub1*) (Miller et al. 2001), and *P. falciparum* (*pfsb1*, Blackman et al. 1998; and *pfsb2*, Hackett et al. 1999) have been implicated in host invasion. The proteins encoded by these genes are all found in the apical ends of their respective parasites. Most are associated with micronemes, one of the main secretory vesicles of apicomplexans, and are secreted at or around the time of host cell invasion. Although the substrates for each have not been identified, possi-

*Corresponding author. Email: kreece@vims.edu

ble candidates include other secreted parasite proteins involved in host invasion, parasite cell-surface membrane proteins, and host cell-surface membrane proteins (Blackman et al. 1998, Barale et al. 1999, Hackett et al. 1999, Sajid et al. 2000, Miller et al. 2001, Louie et al. 2002).

In this study, primers designed by Sakanari et al. (1989) and Blackman et al. (1998) were utilized in PCR amplifications with *Perkinsus marinus* genomic DNA in an attempt to identify a serine protease gene. Chymotrypsin-like and subtilisin-like serine protease primers were selected because of the widespread occurrence of both enzyme types in nature (Barrett et al. 1998) and the role that serine proteases may play in Dermo infection. Here we report the identification of the first serine protease gene(s) from *P. marinus*. We identified 2 sequence types with 95.5% similarity that encode proteins of the subtilisin-like serine protease family.

MATERIALS AND METHODS

Amplification of a serine protease gene fragment.

Degenerate 'universal' primers (Sakanari et al. 1989, Elvin et al. 1993, Blackman et al. 1998), designed from conserved regions present in most serine proteases, were used to amplify genomic DNA from *Perkinsus marinus* in PCR. The forward primer 5'-ACA GAA TTC TGG GTN GTN CAN GCN GCN CAY TG-3' and the reverse primer 5' -ACA GAA TTC ARN GGN CCN CCN SWR TCN CC-3' were designed by Sakanari et al. (1989) to amplify chymotrypsin-like serine proteases. The forward primer 5'-CAY GGI ACI CAY GTI GCI GG-3' and the reverse primer 5'-CCI GCI ACR TGI GGI GTI GCC AT-3' were designed by Blackman et al. (1998) from conserved regions common to subtilisin-like serine proteases. Approximately 10 ng DNA from a York River isolate of *P. marinus*, VA 2 (P-1), was used in 50 µl reactions containing 0.5 mM deoxyribonucleotide triphosphate (dNTP), 3.5 or 2.0 mM MgCl₂, 1× reaction buffer, 50 pmol of each relevant primer and 2.5 U of *Taq* polymerase (Perkin-Elmer). Amplifications were performed using a PTC-200 Peltier Thermal Cycler (MJ Research) using the following parameters: 94°C for 2 min followed by 2 cycles of 94°C for 2 min, 25°C for 1 min and 72°C for 2 min; followed by 40 cycles of: 94°C for 2 min, 55°C for 2 min and 72°C for 2 min, with a final extension cycle at 72°C for 10 min. PCR products were analyzed on a 2% agarose gel with a 1 kb DNA ladder (Invitrogen Corporation) as a size standard.

Shotgun cloning and sequencing. Following removal of primers with PCR Select III Spin columns (Eppendorf 5'), amplified fragments were ligated into the TA

cloning vector pCR 2.1 (Invitrogen Corporation) and used to transform *Escherichia coli* INFαF' according to the manufacturer. Following blue/white screening, selected clones were cultured in 2× YT media containing ampicillin, and the plasmid DNA was extracted using PERFECTprep plasmid DNA purification Kit (Eppendorf 5'). DNA was quantified using a DynaQuant 200 (Hoefer, Pharmacia Biotech). Inserts were sequenced in both directions using the thermo sequenase infrared-labeled primer cycle sequencing-kit with 7-deazadGTP (Amersham Corporation) in conjunction with M13 forward and reverse primers (LI-COR) on a LI-COR 4200 automated sequencer. All sequences were subjected to BLASTN and BLASTX analysis (Altschul et al. 1990) against GenBank to determine whether they were similar to serine protease genes.

Probe design and screening of *Perkinsus marinus* λ-phage library. The insert of 1 clone containing a 518 bp fragment that demonstrated sequence similarity to serine protease genes of the subtilisin family was selected for probe development. The PERFECTprep (Eppendorf 5') plasmid DNA purification product from this clone was digested with *EcoRI* and run out on a 1% agarose gel. The 518 bp band was excised from the gel and subjected to the Gene Clean Spin Kit (Bio 101) following the manufacturer's protocol, and 10 to 300 ng of DNA containing the insert fragment were labeled with digoxigenin (DIG) using the random primed labeling kit and DIG-High Prime (Roche).

The DIG-labeled serine probe was used to screen a previously developed *Perkinsus marinus* genomic library comprised of DNA from multiple isolates (Reece et al. 1997a). Preparation and utilization of host strain XL1-Blue cells, λ-phage plating and plaque lifts were performed following the protocol of Sambrook et al. (1989). Hybridization procedures followed the DIG system protocol (Boehringer Mannheim Biochemica 1995) using manufacturer's reagents and membranes (Roche). Phage stocks of isolated plaques were stored at 4°C. Phage DNA was isolated by a plate lysate method using the Lambda DNA Purification Kit according to the manufacturer's protocol (Stratagene).

Southern blot analysis. Phage DNA was digested in separate reactions with the restriction enzymes *SallI*, *SstI*, and *XhoI* overnight at 37°C, electrophoresed on a 1% agarose gel, and a Southern blot was performed according to Sambrook et al. (1989). Following blotting, membranes were fixed by UV cross-linking (Stratagene). Hybridization and post-hybridization detection were performed according to the DIG system protocol (Boehringer Mannheim Biochemica 1995).

Subcloning and sequencing. DNA fragments identified by Southern blot analysis to contain serine protease gene fragments were cloned into KS⁺ Bluescript (Stratagene). Prior to cloning, the KS⁺ Bluescript vector

was digested with *Sa*I and treated with phosphatase according to Sambrook et al. (1989). DNA from identified phage clones was digested with *Sa*I and electrophoresed on a 1% agarose gel. The desired fragments were isolated using the Gene Clean Spin Kit (Bio 101) following the manufacturer's protocol. Fragments were ligated to the vector using T4 DNA ligase at 14°C overnight. Transformation of DH5 α -competent cells (Invitrogen Corporation) was performed according to the manufacturer's protocol. Following blue-white screening, PERFECTprep (Eppendorf 5') plasmid DNA isolations were performed for those samples containing inserts. Inserts of the expected size were sequenced using bi-direction sequencing as described above. Because of the large size (~4 kb) of the insert, internal sequencing primers were designed in order to sequence the entire insert (forward primer Sub2R700 [5'-AAG GTG ACA GTC GGT GTG AA-3'] and reverse primer Sub1F800 [5'-CAT CGT GGT TCA AGC GTT-3']). BLAST searches and CLUSTALW (Thompson et al. 1994) alignments were performed to confirm relationship with serine proteases and the presence of the 518 bp probe sequence.

Genomic Southern blot analysis. Genomic Southern blots were performed on the 12 genetically distinct cultured isolates of *Perkinsus marinus* (Reece et al. 2001) listed in Table 1. Genomic DNA was isolated as described previously. DNA was quantified using a spectrophotometer to ensure that an equivalent amount (330 ng) from each isolate was used for each independent digestion. Digests, with either *Sa*I or *Nci*I, occurred overnight at 37°C. Southern blots, hybridization and detection techniques were identical to those described above.

3' RACE. Total RNA was isolated from each isolate using the TRIzol reagent system (Invitrogen Corporation) (Simms et al. 1993) based on the acid:phenol:guanidinium isothiocyanate RNA extraction protocol of Chomczynski & Sacchi (1987). We converted 5 μ g of total RNA from the York River, Virginia isolate VA 2 (P-1) and from the Louisiana isolate LA 10-1 to first-strand cDNA and target cDNA was amplified according to the manufacturer's protocol (Invitrogen Corporation). All components, with the exception of RNA and gene-specific primers, were provided with the 3' RACE System for Rapid Amplification of cDNA Ends Kit (Invitrogen Corporation). We performed 2 control reactions, 1 for each isolate, without reverse transcriptase to determine whether an amplification product was of genomic or cDNA origin.

Amplifications of target cDNA from the first-strand cDNA reactions above were performed with 2 separate gene specific primers in conjunction with the abridged universal amplification primer (AUAP) provided in the 3' RACE System for Rapid Amplification of cDNA Ends

Table 1. *Perkinsus marinus* isolates examined in present study. Isolate designation according to Reece et al. (2001; designations in parentheses used by original researchers). Composite genotypes based on RFLP analysis of 8 polymorphic loci (Reece et al. 2001).

Isolate	Origin	Composite genotype
MA 3-9	Wareham River, MA	3
MA 2-11 ^a	Nantucket Sound, MA	1
MA 1-1	Cotuit Bay, MA	2
NJ 3-1	Delaware Bay, NY	5
VA-2 (P-1) ^a	York River, VA	9
VA-5 (HVA 18) ^a	Lynhaven River, VA	8
SC 3-2 ^a	Crabhaul Creek, SC	3
SC 2-4	Clambank Creek, SC	1
LA 23-7	Mozambique Point, LA	4
LA 10-1 ^a	Grand Terre, LA	4
LA 8-11	Bay Tambour, LA	9
LA 5-2	Grand Terre, LA	3

^aUsed in northern blot analysis

Kit (Invitrogen Corporation). The gene-specific primers were designed from the serine protease gene sequence determined previously, primer SERRTF1 (5'-AGG AAG CAG TTC GCT CAG TCA G-3'), and primer SERRTF2 (5'-TCC TCT TCT CGG CAC CAA TGA C-3'). PCR was performed with a Biometra Tgradient thermocycler (Whatman Biometra). Optimal PCR conditions were as follows: 1 cycle at 94°C for 2 min followed by 35 cycles of 94°C for 2 min, 67.5°C for 1 min, 72°C for 1 min 45 s, followed by 1 cycle at 72°C for 10 min and then maintained at 4°C.

Appropriate length-amplification products were identified and cloned into the plasmid vector pCR2.1 using the TA cloning kit (Invitrogen Corporation). Plasmid DNA was isolated using the Plasmid miniPrep kit (Qiagen), and sequenced as described above.

5' RACE. Total RNA was isolated from *Perkinsus marinus* isolates VA 2 (P-1) and LA 10-1 as described above. Specific RNA sequences were converted to first-strand cDNA using the gene specific primer SERGSP1 (5'-CAC AGA GCA TAG AAG GAA TCG-3'). We added 5 μ g of sample RNA to each reaction following the protocol and reaction conditions provided with the 5' RACE System for Rapid Amplification of cDNA Ends Kit, Version 2 (Invitrogen Corporation). cDNA was then purified using the GLASSMAX DNA Isolation Kit (Invitrogen Corporation), and terminal deoxynucleotidyl transferase (TdT) tailing of purified cDNA samples was accomplished following kit instructions.

Specific amplifications of the tailing reaction were performed using the abridged anchor primer (AAP) and PCR reagents provided by the manufacturer (Invitrogen Corporation) and a second nested gene-specific primer, SERGSP2 (5'-ACC GCC ATA CGA

ATG TGA GG 3'). PCR was performed using a PTC-200 Peltier Thermal Cycler (MJ Research). Conditions were as follows: 1 cycle at 94°C for 2 min followed by 35 cycles of 94°C for 1 min, 65°C for 1 min, 72°C for 1 min 30 seconds, followed by 1 cycle at 72°C for 10 min and then maintained at 4°C.

Re-amplification of a 1:100 dilution from the above PCR reaction was performed using a third primer, SERGSP3 (5'-GCG ATA CCA AGG CTG TTA TTA GTC-3') and the AUAP following manufacturer's protocol (Invitrogen Corporation). PCR conditions were as follows: 1 cycle at 94°C for 2 min followed by 35 cycles of 94°C for 1 min, 62°C for 1 min, 72°C for 1 min 30 s, followed by 1 cycle at 72°C for 10 min and then maintained at 4°C.

PCR reactions were analyzed on a 1.5% agarose gel. Appropriate length-amplification products were isolated using the Concert Rapid Gel Extraction System (Invitrogen Corporation), cloned and sequenced as described above. The 3' RACE and 5' RACE cDNA clone nucleotide and amino acid sequences were aligned to the genomic sequences reported above using CLUSTALW (Thompson et al. 1994) in the MacVector Sequencing Analysis Package (Oxford Molecular) and examined for polymorphisms.

Phylogenetic analysis. The amino acid sequences corresponding to nucleotides 985 to 1841 (numbering according to C1A15L23) of 2 subtilisin-like sequence types from *Perkinsus marinus* (C1A15L23 [*P. marinus* subt.A] and 3pP1F2B1 [*P. marinus* subt. B]) were aligned to several subtilisin and subtilisin-like amino acid sequences identified by BLAST search, using CLUSTALW (Thompson et al. 1994) and the BLOSUM 30 matrix with a pairwise open-gap penalty of 10, extend-gap penalty of 0.1 and multiple alignment open-gap penalty of 10, with an extend-gap penalty of 0.05 with 40% delayed divergence. Phylogenetic analysis was performed using PAUP*4.0b 10 (Swofford 2002). A neighbor-joining tree was generated using mean character differences.

Northern blot analysis. The same plasmid clone containing the 518 bp serine protease gene fragment selected for DNA probe development was used for RNA probe labeling. In addition, an RNA probe for actin was developed from an actin clone previously isolated by Reece et al. (1997b). Linearization of plasmid DNA was accomplished by overnight digestion with *KpnI* at 37°C. Following digestion, DNA was purified using phenol:chloroform, precipitated with 2.5 vol of 95% ethanol at -80°C for 2 h and spun at 4°C, 10 000 × *g* for 20 min. Pelleted DNA was resuspended in diethyl procarbonate (DEPC)-treated water and quantified using the DynaQuant 200 (Hoefer, Pharmacia Biotech). NTP labeling mixture, transcription buffer, RNase inhibitor and RNA Polymerase T7 were

added to the purified template, according to the manufacturer's protocol (Roche). The reaction was terminated by addition of 0.2 M EDTA and stored at -20°C until used.

We cultured 5 isolates of *Perkinsus marinus* representing different genotypes (1, 3, 4, 8 and 9; Table 1) according to Reece et al. (2001) for use in northern blot analysis. Briefly, cells from each isolate were passed through a 25-gauge needle twice, washed 2 times with 1:1 Dulbecco's modified Eagle's (DME-HAMS) medium supplemented with 5% fetal bovine serum, and counted using a Bright-Line hemocytometer (Reichert). We seeded 25 cm² culture flasks (Corning) with 1 × 10⁶ *P. marinus* cells per ml in 6 ml of DME-HAMS medium supplemented with 5% fetal bovine serum. Flasks were incubated at 27°C, in the presence of 5% CO₂. After 18 d, cells were pelleted and total RNA was isolated as described above. mRNA was isolated from total RNA using MessageMaker mRNA Isolation System (Life Technologies) (Simms 1995).

Equivalent amounts of total RNA (1 to 10 µg) and mRNA (100 to 500 ng) were electrophoresed on 1% agarose formaldehyde gels in MOPS (3-N-morpholino-2-hydroxypropane sulfonic acid) buffer. Following electrophoresis, gels were equilibrated in 20× saline sodium nitrate (SSC) and RNA was transferred to a nylon membrane (Roche) overnight by capillary action (Lehrack et al. 1977, Goldberg 1980) in 15 × SSC at 4°C. Membranes were fixed using a UV-crosslinker (Stratagene). Prehybridization, hybridization, post-hybridization washes, and probe detection were performed following manufacturer's protocol (Boehringer Mannheim Biochemica 1995). An actin gene probe labeled with digoxigenin was used separately as a control housekeeping gene under the same conditions. mRNA banding patterns for each isolate were compared with those present in other isolates to reveal for changes in band intensity and for variations in mRNA species present.

RESULTS

Detection of a serine protease gene

Multiple amplification products were observed following PCR using degenerate 'universal' subtilisin serine protease primers (Blackman et al. 1998) with genomic DNA from the *Perkinsus marinus* VA 2 (P-1) isolate. No products were detected in those lanes containing PCR reactions with chymotrypsin-like serine protease primers (Sakanari et al. 1989). Following 'shotgun' cloning and sequencing of the subtilisin primer amplification products, BLASTX analysis of a 475 bp cloned fragment designated 1_3sub3.5 (518 bp with primers included) (GenBank Accession No.

AY340234), showed similarity to the amino acid sequence of subtilisin-like serine proteases.

Several λ -phage plaques were identified containing phage particles with sequences that hybridized to the DIG-labeled DNA probe developed from 1_3sub3.5. Southern blots of restriction digests of DNA isolated from these λ -phage particles with *SalI*, followed by hybridization to the 1_3sub3.5 probe, revealed 2 different size bands of either approx. 4 or 3 kb, depending on the phage clone. The 4 kb band present in Phage Clone C1A was successfully subcloned into KS⁺ Bluescript. Sequence analysis of the inserts from 2 plasmid subclones, C1A15L2 and C1A15L3, revealed the presence of a 1254 bp open reading frame containing the 518 bp probe sequence (Fig. 1, Nucleotide Positions 1104 to 1621 in the alignment, GenBank Accession No. AY340222). Alignment of the probe sequence with the identical sequences from C1A15L2 and C1A15L3, designated C1A15L23 (GenBank Accession No. AY340222), identified differences at 12 of 475 nucleotide positions (following removal of primers), showing 97.5% similarity among sequences. Although 10 amino acid sites were involved, because of genetic code degeneracy, only 5 amino acids were different among translated sequences (Fig. 1).

Genomic Southern blot analysis

Southern blot analysis of *SalI* and *NciI* restriction digests of genomic DNA from the 12 *Perkinsus marinus* isolates listed in Table 1, followed by hybridization with the DIG-labeled 1_3sub3.5 DNA probe, revealed differences in banding patterns among isolates (Fig. 2). Genomic Southern blots of isolates with a composite genotype of 1, 5 or 9 showed 2 bands between 3000 and 4000 bp following digestion with *SalI*, and 3 bands from approx. 1100 to 1600 bp following digestion with *NciI*. Only 1 band was detected from both blots for the remaining isolates (Composite Genotypes 2, 3, 4 and 8).

3' and 5' RACE

We amplified 2 different PCR products of the expected sizes (approx. 1000 and 1100 bp) following 3' RACE using the gene-specific primers SERRTF1 (F1) and SERRTF2 (F2), respectively, in conjunction with the primer AUAP (gel not shown). Two additional smaller bands were detected from reactions using the F1 primer, and 1 additional smaller band was detected from the F2 primer reaction. There were no apparent differences in banding patterns between VA 2 (P-1) and LA 10-1. No bands were detected in negative control lanes, demonstrating that the amplification products were of cDNA origin.

cDNA from the 5' RACE amplification reactions using the abridged anchor primer and the nested gene-specific primer SERGSP2 were very faint, requiring additional amplification. An additional nested amplification of VA 2 (P-1) cDNA from the above reaction, which used the third gene-specific primer, SERGSP3, produced 3 strong cDNA bands of approx. 800, 350 and 250 bp following agarose gel electrophoresis and staining with ethidium bromide. There was no evidence of DNA contamination. The 800 bp cDNA fragment selected for cloning was the approximate band size predicted from the location of the putative methionine start codon and the primer SERGSP3.

Sequence analysis of cloned 3' and 5' RACE products (GenBank Accession Nos. AY340223 to AY340233) confirmed that the 1254 bp open reading frame detected in the C1A15L3 and C1A15L2 genomic DNA clones contained the entire coding region of the serine protease gene, without the presence of introns. CLUSTALW (Thompson et al. 1994) analysis of the 3' RACE products revealed the presence of 32 polymorphic nucleotide sites in addition to the 12 identified previously by sequence alignment of the probe sequence and C1A15L23 (Fig. 1; GenBank Accession No. AY340222). Specific polymorphisms were detected in multiple DNA and cDNA clones from different amplification reactions, confirming that the nucleotide differences were not a result of *Taq* errors. In total, 41 codons in the 3' RACE region (nucleotide positions 874 to 1954 in the alignment) were affected, resulting in 19 amino acid variations (Fig. 1); 16 of the rare codons appeared in the 3pP1F2B1 and 3pP1F2B4 cDNA clones (designated *Perkinsus marinus* subt. B) from the 3' RACE clones, and 3 rare codon changes were present in C1A15L23 (designated *P. marinus* subt. A). This represented a 95% nucleotide sequence similarity and a 94.6% amino acid sequence similarity between the 2 types over the 857 nucleotide region analyzed. Within the Type B sequence, 1 of these amino acid changes occurred within the conserved region surrounding the histidine catalytic site and 3 amino acid changes occurred within the conserved oxyanion hole region (Fig. 3). Differences were also identified in the 5' RACE products (nucleotide positions 403 to 1119 in the alignment), resulting in 6 amino acid changes from the 11 nucleotide differences (Fig. 1).

With the exception of 3pP1F2B1 and 3pP1F2B4, few serine protease gene sequences were identical (Fig. 1; GenBank Accession No. AY340222), and more than 3 different sequences were obtained from each isolate culture. Frequently, a rare nucleotide observed in one DNA clone would be observed with a different combination of rare nucleotides in a different DNA clone,

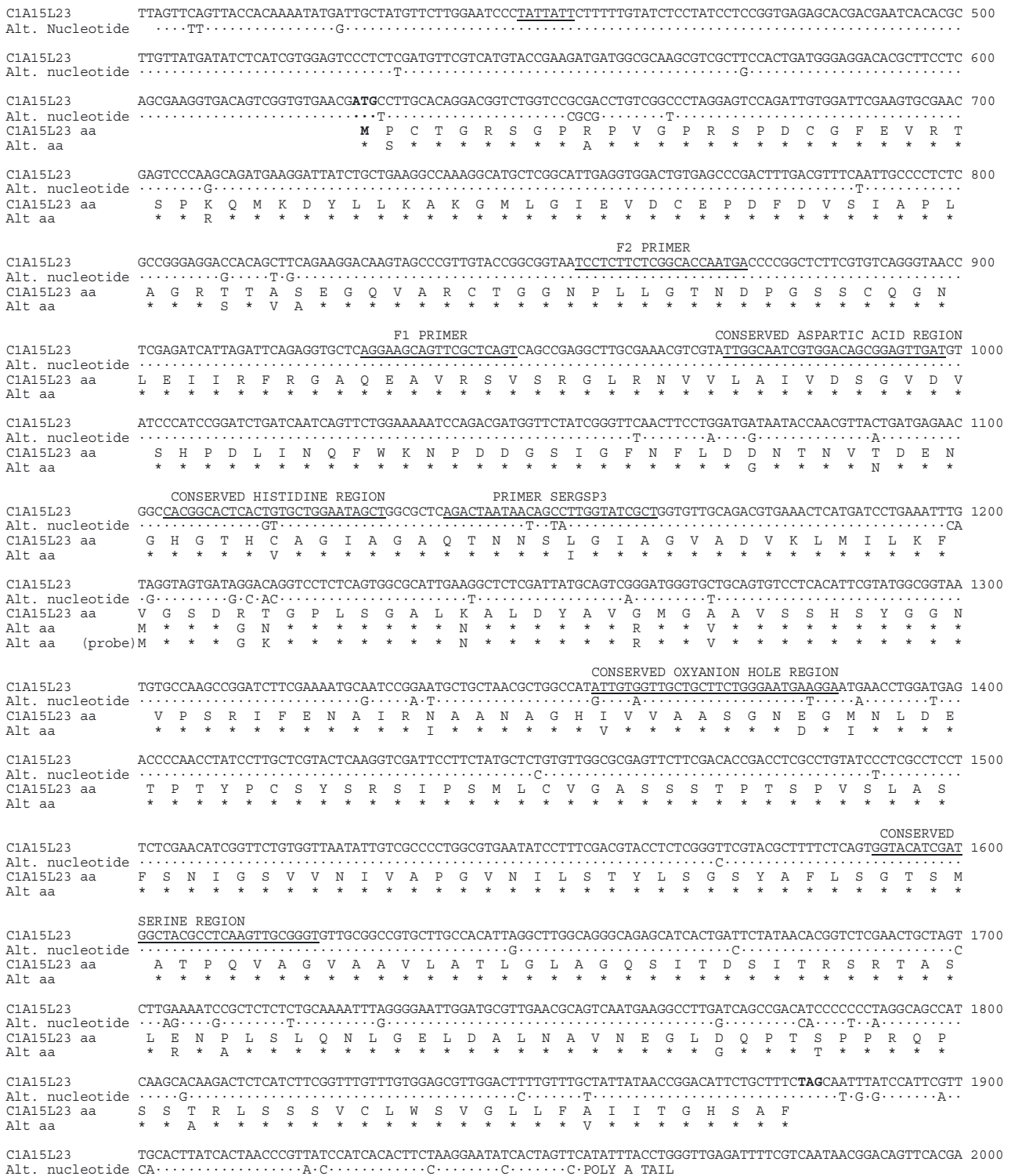


Fig. 1. *Perkinsus marinus*. Partial nucleotide sequence of genomic λ -phage sub-clone C1A15L23 sequence and alternate nucleotides found in either the probe sequence, 3' RACE or 5' RACE clones. (.) Positions of nucleotide identity compared to C1A15L23 sequence. (*) Positions of amino acid identity compared to the predicted C1A15L23 amino acid sequence with alternate amino acids shown below. Putative start and stop codons are in boldface. Primer sequences and conserved regions common to subtilisin-like serine proteases are underlined. GenBank Accession Nos. for individual complete sequences and associated alignment file are AY340222-AY340234

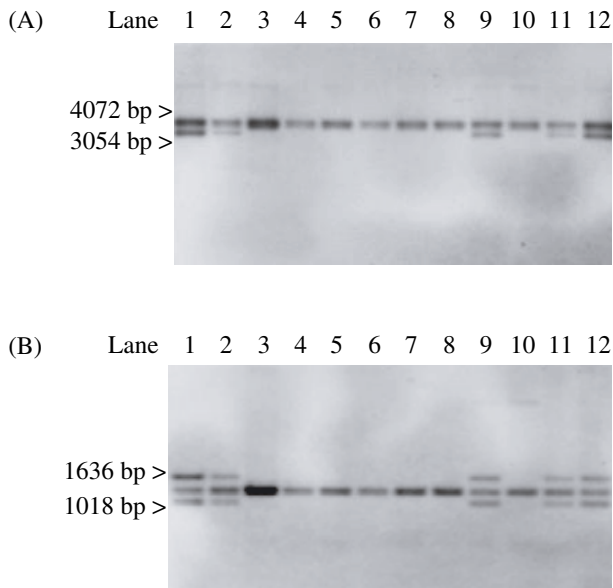


Fig 2. *Perkinsus marinus*. Genomic Southern blot analysis of 330 ng of genomic DNA from each of the 12 isolates following independent digestions with (A) *SalI* and (B) *NciI* demonstrating 2 variations in hybridizing band patterns with the 1_3sub3.5 DNA probe. Lane order is same for both blots, with genotype designation (Reece et al. 2001) in parentheses. Lane 1: MA 2-11 (1); Lane 2: SC2-4 (1); Lane 3: MA 1-1 (2); Lane 4: LA 5-2 (3); Lane 5: MA 3-9 (3); Lane 6: SC 3-2 (3); Lane 7: LA 23-7 (4); Lane 8: LA 10-1 (4); Lane 9.: NJ 3-1 (5); Lane 10: VA-1 (8); Lane 11: VA-2 (9); Lane 12: LA 8-11 (9)

regardless of whether the DNA clone was from VA 2 (P-1) or LA 10-1. For example, the probe sequence generated from the VA 2 (P-1) isolate resembles C1A15L23, lacking the rare nucleotides G and T present in 3pP1F2B1 and 3pP1F2B4 at positions 1116 and 1117 respectively, but then aligns with 3pP1F2B1 and 3pP1F2B4 until nucleotide 1217. With the exception of 1 nucleotide at position 1260, the remainder of the probe sequence resembles C1A15L23.

	(1)	(2)	(3)	(4)
Perk.subt.A	972 LAIVDSGVDVSH	1104 HGTHCAGIAGA	1590 GTSMATPQVAG	1350 GHIIVVAASGNEG
Perk.subt.B	LAIVDSGVDVSH	HGTHVAGIAGA	GTSMATPQVAG	GHVVVAASGNAG
Thermitase	IAIVDTGVQSNH	HGTHCAGIAAA	GTSMATPHVAG	GSVVVAAAGNAG
Subtilisin	VAVIDSGIDSSH	HGTHVAGTVAA	GTSMASPHVAG	GVVVVAAAGNEG
NcSUB1	VCVVDVTGIDYEH	HGTHCAGIAGA	GTSMAAPALSG	GIIHIAAAGNSG

Fig 3. *Perkinsus marinus*. Subtilisin-like serine protease amino acid alignment of conserved regions. Numbering according to C1A15L23 genomic DNA nucleotide sequences. Amino acids in boldface represent conserved catalytic residue, conserved oxyanion hole residue, or changes in conserved regions present in *P. marinus*. Perk.subt.A: C1A15L23 genomic DNA subclones; Perk.subt.B: 3' RACE clones 3pP1F2B1 and 3pP1F2B4; Thermitase: *Thermoactinomyces vulgaris* (P04072); Subtilisin: *Bacillus amyloliquefaciens* (P00782); NcSUB1: *Neospora caninum* (AAFO4257); (1): aspartic acid (D) catalytic region; (2): histidine (H) catalytic region; (3): serine (S) catalytic region; (4): conserved asparagine (N) of oxyanion hole

The phylogenetic tree generated following neighbor-joining analysis of multiple subtilisin-like serine proteases, placed the 2 *Perkinsus marinus* sequence types, *P. marinus* subt. A (C1A15L23) and *P. marinus* subt. B (3pP1F2B1 and 3pP1F2B4), as a sister group to subtilisin-like amino acid sequences from *Toxoplasma gondii*, *Neospora caninum*, and *Plasmodium falciparum* (Fig. 4).

Northern blot analysis

Only 1 band, approximately 1500 to 1600 nucleotides, was detected among all 5 isolates examined following northern blot analysis of total RNA or mRNA and hybridization to the 1_3sub3.5 DIG-labeled RNA probe. The band size was in agreement with the expected mRNA length based on examination of the gene structure from the 3'-5' RACE results. Comparison of gene expression among isolates was inconclusive.

DISCUSSION

Subtilisin-like gene(s) from *Perkinsus marinus*

This report comprises the first description of a serine protease gene(s) from *Perkinsus marinus*. The cloned gene(s) encode a polypeptide of 416 amino acids with an approximate molecular weight of 47 840 Da. Prediction of the final functional protein form *in vivo*, however, is difficult since many proteases are produced as zymogens and require further post-translational processing prior to activation. Significant post-translational processing was observed for 2 subtilisin-like proteases produced by *Plasmodium falciparum*, PfSUB-1 (Blackman et al. 1998) and PfSUB-2 (Hackett et al. 1999), and is suspected for the subtilisin-like serine protease from *Neospora caninum* (Louie & Conrad 1999). In addition, glycosylation will also affect the final molecular weight.

Although the function and proteolytic activity of the protein or proteins coded for by this gene(s) have not yet been characterized, the cloned gene sequences showed similarity to the subtilisin-like serine protease gene family. The 4 conserved amino acid residues common to all subtilisin-like serine proteases, the catalytic triad residues aspartic acid (D), histidine (H), and serine (S) and the

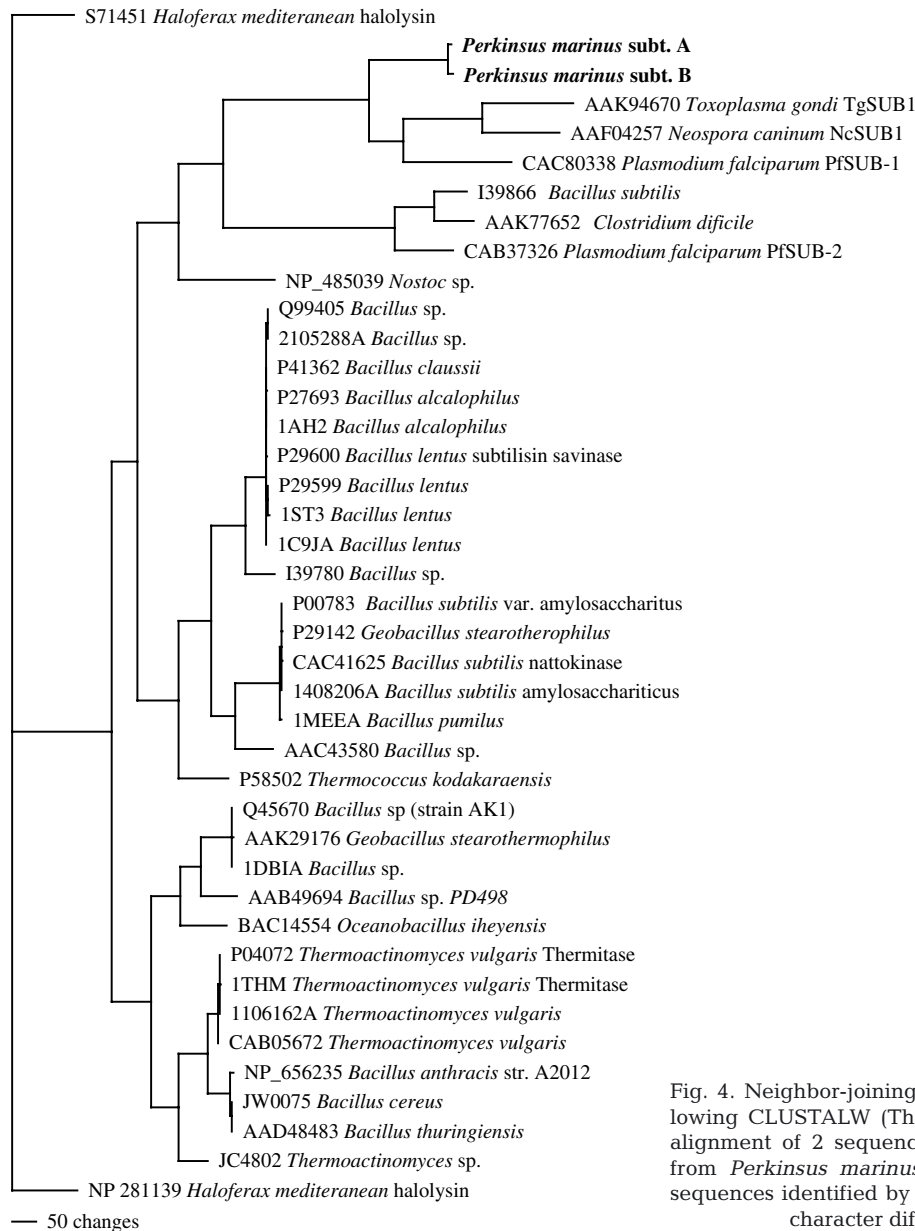


Fig. 4. Neighbor-joining tree generated following CLUSTALW (Thompson et al. 1994) alignment of 2 sequence types determined from *Perkinsus marinus* and subtilisin-like sequences identified by BLAST, using mean character differences

conserved asparagine (N) of the oxyanion hole, were present in both amino acid sequences translated from the different *Perkinsus marinus* nucleotide sequences. In addition, the catalytic residues occurred in the order D/H/S, which is different from that found in chymotrypsin-like serine proteases (H/D/S).

Further analysis of the amino acids surrounding the catalytic sites showed that the subtilisin-like gene(s) present in *Perkinsus marinus* encode a protein(s) most similar to those in the subtilisin subfamily S8A (Barrett et al. 1998). Members of this subfamily possess a serine (or threonine) following the conserved aspartic acid residue, a histidine 3 amino acids following the catalytic histidine residue and a methionine immediately

after the catalytic serine residue (Barrett et al. 1998). These were all present in both *P. marinus* amino acid sequences at the catalytic sites (Fig. 3). Although the third subfamily of subtilisins, S8C, also contains these conserved amino acid sequences, the *P. marinus* catalytic regions did not contain additional amino acids found only in S8C (Barrett et al. 1998).

Phylogenetic analysis grouped the 2 *Perkinsus marinus* sequence types with subtilisin-like serine proteases from *Toxoplasma gondii* (TgSUB1), *Neospora caninum* (NcSUB1) and *Plasmodium falciparum* (PfSUB-1) (Fig. 4). Although the function of Perk. subtyp. A and/or Perk subtyp. B is unknown, all 3 of the above proteins are candidates for involvement with host-cell invasion

by the parasites (Blackman et al. 1998, Louie & Conrad 1999, Sajid et al. 2000, Miller et al. 2001, Louie et al. 2002). PfSUB-1 is concentrated in dense granules at the apical end of the invasive *P. falciparum* merozoites, and is secreted approximately at the time of erythrocyte invasion. Similarly, TgSUB1 is found within the micronemes of *T. gondii*, secretory vesicles located in the apical region of the parasite that secrete their contents early during the invasion process (Miller et al. 2001). Although the specific location of the subtilisin-like protein from *N. caninum* has not been confirmed, characterization of NcSUB1 with rabbit anti-N54 suggests that the protein is also located within micronemes at the apical end of the parasite. The antibody also bound to a major secreted protein of *N. caninum* likely to be the mature form of the NcSUB1 precursor (Louie & Conrad 1999, Louie et al. 2002).

***Perkinsus* subtilisin-like gene(s), polymorphism or more than one gene?**

Genomic Southern blot analysis of the subtilisin-like gene(s) from *Perkinsus marinus* revealed multiple bands from some isolates (Composite Genotypes 1, 5 and 9), whereas only 1 band was observed from others (Composite Genotypes 2, 3, 4 and 8) (Fig. 2). The additional bands could represent additional genes present in only a few of the isolates, or, simply, different *SalI* or *NciI* restriction enzyme digestion sites within the gene(s). Reece et al. (2001) demonstrated genetic variation among *P. marinus* isolates by restriction fragment length polymorphism (RFLP) analysis, and at least some *P. marinus* isolates appear to be diploid (Reece et al. 1997a, 2001). Those isolates showing only 1 band following Southern blot analysis with both *SalI* and *NciI* (MA 1-1, LA 5-2, MA 3-9, SC 3-2, LA 23-7, LA 10-1 and HVA 18) appear to be homozygous at this locus. The presence of 2 different fragments from the 2 types of λ -phage clones and the presence of 2 or 3 bands on the genomic Southern blots indicate a possible heterozygous state.

The presence of 3 bands in the genomic Southern blots following *NciI* digestion is complicated by the observation that more than 2 sequences occur within 1 isolate (Fig. 1). It is possible that recombination among heterozygous isolates has produced an additional restriction site. In addition, it is possible that the 2 bands observed in the *SalI* Southern blots represent 2 separate genes instead of 2 separate alleles, 1 of which is not present in the other isolates. In this case, the *NciI* digest would demonstrate that the center band represented the gene common among all isolates and the upper and lower bands represent different alleles of a second gene. The addition or lack of restriction enzyme site(s), however, has not been demonstrated.

The possible existence of 2 very similar genes is supported by the different amino acid sequence types predicted from nucleotide sequences generated by 3' RACE. There was only 95% amino acid similarity within the coding region identified within the 3' RACE clones 3pP1F2B1 and 3pP1F2B4 (designated Perk. subt. B), in comparison to the genomic λ -phage subclone C1A15L23 (designated Perk. subt. A). Interestingly, the C1A15L23 sequence encoded a cysteine residue 4 amino acids from the catalytic histidine residue (Fig. 3), which is in a homologous location to a histidine residue responsible for the dependence of thiol activation in the cuticle degrading protein, chymoelastase, from *Metarhizium anisopliae* (Rawlings & Barrett 1994). The large number of predicted amino acid changes between the 2 sequences (19), including the possible requirement for thiol activation in the 'A' amino acid sequence but not in the 'B' amino acid sequence, supports the existence of 2 separate genes. Although only 1 band was detected by northern blot analysis, both sequences would encode an mRNA of similar size.

In addition, few of the nucleotide sequences generated, either from the probe, the genomic λ -phage subclone C1A15L23, or by RACE, were the same. The occurrence of multiple nucleotide sequences, that is more than 2 sequences from individual isolates, could easily be explained if 2 genes exist. Alternatively, the observance of multiple sequence types may also be a result of recombination, either real or a PCR artifact produced by strand-jumping of the *Taq* polymerase among different sequences for the same gene (Pääbo et al. 1989).

Although any discussion regarding the function of the 2 subtilisin-like serine protease sequence types is speculative, it is noteworthy that phylogenetic analysis grouped the *Perkinsus marinus* subtilases with proteins of apicomplexan parasites believed to be involved with invasion of host cells. Future work is necessary to determine the processing and the final site of the mature protein, including whether or not it is secreted within the host. Additional studies with other clonal cultures of different genotypes will be also be necessary to determine whether recombination is truly a factor, either among 2 different alleles of the same gene and/or among alleles of different genes. In addition, future work confirming the presence of polymorphic sites in the 5' end of the gene(s) and their relationship to nucleotide and amino acid differences already identified should help to further distinguish among sequence types, determine whether all sequences contain the same translation initiation and transcription start sites, and ultimately whether the 2 sequence types are 2 different genes.

Acknowledgements. The work on this project was partly funded by a Melbourne Carriker Grant and VIMS Graduate Research Grants. The authors would also like to thank Dr. W. MacIntyre for his assistance. VIMs contribution #2573.

LITERATURE CITED

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410
- Banyal HS, Misra GC, Gupta CM, Dutta GP (1981) Involvement of malarial proteases in the interaction between the parasite and host erythrocyte in *Plasmodium knowlesi* infections. *J Parasitol* 67(5):623–626
- Barale J-C, Blisnick T, Fujioka H, Alzari PM, Aikawa M, Braun-Breton C, Langley G (1999) *Plasmodium falciparum* subtilisin-like protease 2, a merozoite candidate for the merozoite surface protein 1–42 maturase. *Proc Natl Acad Sci USA* 96:6445–6450
- Barrett AJ, Rawlings ND, Wuessner JF (1998) Handbook of proteolytic enzymes. Academic Press, London, p 284–366
- Blackman MJ, Fujioka H, Stafford WHL, Sajid M and 5 others (1998) A subtilisin-like protein in secretory organelles of *Plasmodium falciparum* merozoites. *J Biol Chem* 273(36):23398–23409
- Boehringer Mannheim Biochemica (1995) The DIG system user's guide for filter hybridization. Boehringer, Mannheim
- Chaudhuri G, Chaudhuri M, Pan A, Chang KP (1989) Surface acid protease (gp 63) of *Leishmania mexicana*. A metalloenzyme capable of protecting liposome-encapsulated proteins from phagolysosomal degradation by macrophages. *J Biol Chem* 264:1483–1489
- Chomczynski P, Sacchi N (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem* 162:156
- Elvin CM, Whan V, Riddles PW (1993) A family of serine protease genes expressed in adult buffalo fly (*Haematobia irritans exigua*). *Mol Gen Genet* 240:132–139
- Garreis KA, La Peyre JF, Faisal M (1996) The effects of *Perkinsus marinus* extracellular products and purified proteases on oyster defense parameters in vitro. *Fish Shellfish Immunol* 6:581–597
- Goldberg DA (1980) Isolation and partial characterization of the *Drosophila* alcohol dehydrogenase gene. *Proc Natl Acad Sci USA* 77:5794–5798
- Hackett F, Sajid M, Withers-Martinez C, Grainger M, Blackman M (1999) PfSUB-2: a second subtilisin-like protein in *Plasmodium falciparum* merozoites. *Mol Biochem Parasitol* 103(2):113–119
- La Peyre JF, Schafhauser DY, Rizkalla EH, Faisal M (1995) Production of serine proteases by the oyster pathogen *Perkinsus marinus* (Apicomplexa) in vitro. *J Eukaryot Microbiol* 42(5):544–551
- La Peyre JF, Yarnall HA, Faisal M (1996) Contribution of *Perkinsus marinus* extracellular products in the infection of eastern oysters (*Crassostrea virginica*). *J Invertebr Pathol* 68:312–313
- Lehrack H, Diamond D, Wozney JM, Boedtker H (1977) RNA molecular weight determinations by gel electrophoresis under denaturing conditions, a critical reexamination. *Biochemistry* 16:4743–4751
- Louie K, Conrad PA (1999) Characterization of a cDNA encoding subtilisin-like serine protease (NC-p65) of *Neospora caninum*. *Mol Biochem Parasitol* 103(2):211–223
- Louie K, Nordhausen R, Robinson TW, Barr BC, Conrad PA (2002) Characterization of *Neospora caninum* protease, NcSUB1 (NC-p65), with rabbit anti-N54. *J Parasitol* 88(6):1113–1119
- MacIntyre EA, Earnhart CG, Kaattari SL (2003) Host oyster tissue extracts modulate invitro protease expression and cellular differentiation in the protozoan parasite, *Perkinsus marinus*. *Parasitology* 126:293–302
- McKerrow J, Sun E, Rosenthal P, Bouvier J (1993) The proteases and pathogenicity of parasitic protozoa. *Annu Rev Microbiol* 47:821–853
- Miller SA, Binder EM, Blackman MJ, Carruthers VB, Kim K (2001) A conserved subtilisin-like protein TgSUB1 in microneme organelles of *Toxoplasma gondii*. *J Biol Chem* 276(48):45341–45348
- Morty RE, Lonsdale-Eccles JD, Morehead J, Caler EV, Mentele R and 5 others (1999) Oligopeptidase B from *Trypanosoma brucei*, a new member of an emerging group of serine oligopeptidases. *J Biol Chem* 274(37):26149–26156
- Pääbo S, Higuchi RG, Wilson AC (1989) Ancient DNA and the polymerase chain reaction. *J Biol Chem* 264(17):9709–9712
- Rawlings N, Barrett A (1994) Families of serine proteases. *Method Enzymol* 244:19–60
- Reece KS, Bushek D, Graves JE (1997a) Molecular markers for population genetic analysis of *Perkinsus marinus*. *Mol Mar Biol Biotechnol* 6(3):197–206
- Reece KS, Sidall ME, Burrenson EM, Graves JE (1997b) Phylogenetic analysis of *Perkinsus* based on actin gene sequences. *J Parasitol* 83(3):417–423
- Reece KS, Bushek D, Hudson K, Graves JE (2001) Geographic distribution of *Perkinsus marinus* genetic strains along the Atlantic and Gulf Coasts of the USA. *Mar Biol* 139:1047–1055
- Sajid M, Withers-Martinez C, Blackman MJ (2000) Maturation and specificity of *Plasmodium falciparum* subtilisin-like protease-1, a malaria merozoite subtilisin-like serine protease. *J Biol Chem* 275(1):631–641
- Sakanari J, Staunton E, Eakin A, Craik C, McKerrow J (1989) Serine proteases from nematode and protozoan parasites: isolation of sequence homologs using generic molecular probes. *Proc Natl Acad Sci USA* 86:4863–4867
- Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual, 2nd edn., Cold Spring Harbor Laboratory Press, New York, NY
- Simms D (1995) mRNA isolation for high-quality cDNA. *Focus*, New York, 17(2):39–42
- Simms D, Cizdziel PE, Chomczynski P (1993) TRIzol: a new reagent for optimal single-step isolation of RNA. *Focus* New York, 15:99–102
- Swofford DL (2002) PAUP*4.0. Phylogenetic analysis using parsimony (and other methods). Sinauer Associates, Sunderland, MA
- Tall BD, La Peyre JF, Bier JW, Miliotis MD, Hanes DE, Kothary MH, Shah DB, Faisal M (1999) *Perkinsus marinus* extracellular protease modulates survival of *Vibrio vulnificus* in eastern oyster (*Crassostrea virginica*) hemocytes. *Appl Environ Microbiol* 65(9):4261–4263
- Thompson JD, Higgins DG, Gibson TJ (1994) Improving the sensitivity of progressive multiple sequence management, analysis, and homology determination. *Nucleic Acids Res* 22:4673–4680