Echinococcus granulosus antigen 5 is closely related to proteases of the trypsin family

Carmen LORENZO*, Gustavo SALINAS*, Andreina BRUGNINI*, Christer WERNSTEDT†, Ulf HELLMAN† and Gualberto GONZÁLEZ-SAPIENZA*[1]

*Cátedra de Inmunología, Facultad de Química, UDELAR, Instituto de Higiene, Av. Navarro 3051, piso 2, 11600 Montevideo, Uruguay, and †Ludwig Institute for Cancer Research, Uppsala Branch, 751 24 Uppsala, Sweden

Antigen 5 (Ag5) is a dominant secreted component of the larval stage of *Echinococcus granulosus*, and is highly immunogenic in human infections. Although the diagnostic value of Ag5 has been thoroughly evaluated, there has been little progress in its molecular characterization and the understanding of its biological role. In the present study, the *Ag5* gene was cloned by reverse transcription-PCR on the basis of the amino acid sequences of tryptic fragments. The nucleotide sequence indicates that Ag5 is synthesized as a single polypeptide chain that is afterwards processed into single disulphide-bridged 22 and 38 kDa subunits. Whereas the 22 kDa component contains a highly conserved glycosaminoglycan-binding motif that may help to confine Ag5 in the host tissue surrounding the parasite, the 38 kDa subunit is closely related to serine proteases of the trypsin family. The sequences in the vicinity of the active-site histidine, aspartic acid and serine residues, and critical cysteine residues involved in disulphide formation, are well conserved, but the catalytic serine residue is replaced by threonine. Since there are no significant chemical differences between the Oγ atoms of these residues, we performed a series of enzymic assays to find out whether Ag5 is a catalytic molecule. Neither proteolytic activity nor binding to protease inhibitors could be detected using the native purified antigen. Thus it may be possible that Ag5 possesses a highly specific physiological substrate or, more likely, that trypsin-like folding has been recruited to fulfil novel functions.

Key words: catalytic triad, hydatid disease, serine protease, threonine.

INTRODUCTION

Antigen 5 (Ag5) is one of the major antigens of *Echinococcus granulosus*, the causative agent of cystic hydatidosis. Its relevance in diagnosis has been evident since the earliest analysis of the diagnostic value of different parasite components. In fact, the antigen took its name after the work of Capron et al. [1], who identified a relevant precipitation line in immunoelectrophoresis experiments, designated 'arc 5', and proposed its use as a reliable marker of hydatid infection. The diagnostic value of arc 5 was promoted further by the work of other authors [2,3], which popularized its use. Despite the fact that the diagnostic use of Ag5 was shown at a later stage to exhibit problems of specificity and sensitivity [4–7], the formation of arc 5 is still widely used in routine serological diagnosis, particularly as a confirmation technique. The fact that Ag5 is a major component of hydatid cyst fluid suggests its relevance as a key molecule in the biology of *E. granulosus*. This may involve the development and preservation of the metacestode in the intermediate host, as well as the successful infection of the definitive host, where it may contribute to the establishment phase of the protoscoleces. Indeed, Ag5 is a secreted molecule, the presence of which in hydatid cyst fluid is only evident after the larval stage of the parasite has developed the capacity to produce brood capsules and protoscoleces. By this time, the metacestode is fully infective if eaten by a suitable carnivore and, at this point, Ag5 becomes a dominant parasite component of the hydatid cyst fluid. However, in spite of its diagnostic value and potential relevance in *E. granulosus* biology, the biological role of Ag5 is almost completely unknown. An old paper describing esterase activity associated with arc 5 [8], and a speculative paper about the similarity of the N-terminal amino acid sequence of Ag5 to those of serine proteases of the trypsin family [9], are the only studies published concerning the putative function of this molecule.

Progress in the molecular characterization of Ag5 has been limited. After initial studies of the antigen as a thermolabile component that was eluted in both the void and the 60–70 kDa fraction in gel-filtration experiments [8], significant advances in the biochemical characterization of Ag5 were made with the isolation and analysis of the antigen with monoclonal antibodies. In SDS/polyacrylamide gels, Ag5 occurs as a major component of 67 kDa and a relatively minor component of 57 kDa [10]; under reducing conditions, Ag5 consists of 38 and 22 kDa subunits [6]. The antigen is a glycoprotein containing a substantial amount of complex N-linked oligosaccharides without terminal *N*-acetyl-D-glucosamine residues, and lacking high-mannose carbohydrates [11]. Apart from the N-terminal amino acid sequences reported for the 38 kDa subunit [12], there have been no significant advances in the molecular characterization of the antigen. Different groups, using antibodies reactive against

Abbreviations used: Ag5, antigen 5; AMC, 7-amino-4-methyl coumarin; F5, Ag5-enriched fraction; mAb, monoclonal antibody; MALDI–TOF, matrix-assisted laser-desorption ionization–time-of-flight; RT-PCR, reverse transcription-PCR; 38s–MBP, recombinant 38 kDa subunit fused to the C-terminus of maltose-binding protein.

[1] To whom correspondence should be addressed (e-mail ggonzal@fq.edu.uy).

The nucleotide sequence data of the gene encoding Ag5 has been deposited in DDBJ, EMBL, GenBank® and GSDB Nucleotide Sequence Databases under the accession number AY052477.

the 38 kDa subunit, isolated partial cDNA sequences from *E. granulosus* libraries [13] and its homologous sequence from *E. multilocularis* [14]. These cDNAs, however, were shown recently [15] to code for P-29, a molecule immunologically related to, but distinct from, Ag5. The cross-reactivity that occurs between P-29 and Ag5 appears to have accounted for previous failures in the cloning of Ag5 by immunoscreening, and highlighted the need for alternative cloning strategies. In the present study, we report an alternative strategy for the isolation of the *Ag5* gene by reverse transcription-PCR (RT-PCR) using degenerate primers designed from the amino acid sequence of internal fragments of the protein. This allowed the cloning of the full-length gene of Ag5, which encodes a single polypeptide chain that is afterwards processed into the 38 and 22 kDa subunits. In addition to the impact that this may have on the diagnostic use of the antigen, the search for similarity against the databases yielded interesting information about the possible function of the two Ag5 subunits, which we also discuss.

MATERIALS AND METHODS

Parasite antigens

Hydatid cysts were collected from cattle killed in abattoirs in Uruguay. The hydatid cyst fluid and protoscoleces were aspirated aseptically from fertile cysts; the protoscoleces were then decanted, and the supernatant was supplemented with 500 mg/l NaN_3/1 mM EDTA and kept at −20 °C until used. An Ag5-enriched fraction (F5) was prepared by anion-exchange of the hydatid cyst fluid, followed by depletion of antigen B by immunoadsorption with an anti-(antigen B) monoclonal antibody (mAb) immobilized on agarose, as described previously [16]. Ag5 was purified from the F5 fraction by immunochromatography using the mAb 1D1 described below.

mAb 1D1 reactive against Ag5 was prepared as follows: a mouse was immunized with 20 μg of F5 in Freund's complete adjuvant intraperitoneally, and boosted after 3 weeks with the same antigen in Freund's incomplete adjuvant. After 3 days, the mouse was killed and the splenocytes were fused with SP2/0 cells. Techniques used for cell hybridization, culturing and cloning of hybrid cell lines were carried out essentially as described in [17]. The selection of cultures producing mAb reactive against F5 was performed by ELISA and Western blotting. The reactivity of mAb 1D1 against Ag5 was confirmed further by both SDS/PAGE and amino acid sequencing of the immunopurified antigen.

Amino acid sequencing

Fragmentation of the different Ag5 components was performed by digestion *in situ* after SDS/PAGE, essentially as described by Hellman et al. [18]. Briefly, immunopurified Ag5 was run on an SDS/12 % (w/v) polyacrylamide gel under reducing and non-reducing conditions. The proteins were Coomassie-Blue stained, and then the 67, 38 and 22 kDa bands were excised from the gel, washed and dried under nitrogen. Once rehydrated, endoproteinase Lys-C or trypsin was added, and the digestions were incubated overnight at 30 °C. After several extraction steps with 0.1 % (w/v) trifluoroacetic acid/60 % (v/v) acetonitrile, the supernatant was run on a μRPC C2/C18 SC 2.1/10 column, operated in the SMART System (Pharmacia, Uppsala, Sweden). Elution was performed by a linear gradient from 0 to 40 % acetonitrile in 0.065 % (w/v) trifluoroacetic acid. Selected peaks were sequenced in an Applied Biosystems Protein Sequencer (Model 470a; Applied Biosystems, Foster City, CA, U.S.A.). Similarly, the N-terminal amino acid sequences of the 22 and 57 kDa components of Ag5 were obtained from material blotted on to PVDF (Bio-Rad).

Isolation of the full-length cDNA encoding Ag5

Total RNA was isolated from freshly isolated *E. granulosus* protoscoleces using Trizol reagent (Gibco BRL) according to the manufacturer's instructions and reverse-transcribed using oligo-dT primer and Superscript reverse transcriptase (Life Technologies), also following the manufacturer's instructions. The cDNA obtained was used subsequently as a template for PCR reactions, following standard techniques. Sense and antisense primers were designed to PCR-amplify a fragment of the cDNA coding for Ag5. Sense oligonucleotides (5′-AAGGCYGC-NAAYTAYAAYGC-3′ and 5′-TTYCAYCAYGAYGAYGA-AAA-3′, corresponding to the amino acid sequences of KAAN-YNA and FHDDEN respectively) and the antisense oligonucleotides [5′-ATTYTCRTCRTCRTGRTGAAA-3′ and 5′-ATTYT-CRTCRTCRTGRTGGAA-3′ (derived from FHDDEN), and 5′-GCRTTRTARTTRGCNGCTTT-3′ and 5′-GCRTTRTAR-TTRGCNGCCTT-3′ (derived from KAANYNA)] were combined in four sets of PCR reactions. Only one of the PCR reactions yielded a product that was isolated (GeneClean kit; BIO 101, Inc., Vista, CA, U.S.A.), cloned into pGEM T Easy (Promega) and sequenced using Ampli*Taq* FS according to standard protocols on an ABI Prism 377 DNA sequencer. The 3′ end of the *Ag5* gene was PCR-amplified using an Ag5 sense oligonucleotide (5′-GAACGTCGACCACAGCGACC-3′) in combination with an oligo-dT. The PCR product was isolated, cloned and sequenced as described above. The 5′-end of Ag5 was also obtained by RT-PCR, using the Ag5 antisense primer (5′-CAAGCTTCATTAGACTGCGTAGCGGTTGATCC-3′) and the sense oligonucleotide (5′-CACCGTTAATCGGTCCT-TACC-3′) derived from the sequence of the 36 nt exon identified as a splice leader found at the 5′ end of some *E. granulosus* mRNA species [19]. The PCR product was isolated, cloned and sequenced as described above.

Proteolytic and inhibitor-binding assays

The proteolytic assays were performed using a panel of synthetic substrates coupled with 7-amino-4-methyl coumarin (AMC). This panel included a wide range of specific substrates for different enzymes (trypsin, factor X, elastase, chymotrypsin and thrombin) of the trypsin family: L-Arg-AMC, DL-Arg-AMC, Phe-Arg-AMC, Gly-Pro-Arg-AMC, Gly-Gly-Arg-AMC, Ile-Glu-Gly-Arg-AMC, L-Phe-AMC, Ala-Ala-Pro-Phe-AMC, L-Ala-AMC, Ala-Pro-Ala-AMC, Ala-Ala-Ala-AMC, L-Lys-AMC and Ala-Phe-Lys-AMC. Reactions containing 10 nM Ag5 or control protease, 10 mM $CaCl_2$ and 1 μM substrate were set up in a final volume of 1 ml at different pH values (buffers: 100 mM sodium acetate, pH 5.0; 50 mM sodium phosphate, pH 7.2; and 50 mM Tris/HCl, pH 8.6). After incubation at 37 °C for 60 min, fluorescence was measured (excitation wavelength, 355 nm; emission wavelength, 460 nm).

The interaction between Ag5 and three inhibitors of the serine protease family was analysed in microtitre plates. Wells were coated overnight at 4 °C with 100 μl/well of 20 μg/ml immunopurified Ag5, porcine trypsin (Sigma) or porcine elastase (Sigma) in PBS. After blocking with PBS/1 % (w/v) BSA and washing, decreasing concentrations of soya-bean trypsin inhibitor, antithrombin III or α_1-antiprotease conjugated to biotin were added and incubated for 1 h at 37 °C. After washing, 100 μl of streptavidin coupled with horseradish peroxidase (diluted 1:2000 in PBS/1 % BSA; Sigma) were added and incubated for 1 h at

37 °C. After washing, the peroxidase activity was developed with 2,2′-azinobis-(3-ethylbenzothiazoline-6-sulphonic acid) ('ABTS')/H_2O_2, and absorbance at 415 nm was measured.

Recombinant 38 kDa subunit

The 38 kDa subunit was amplified by PCR using the following primers: sense (5′-GGATCCGAATTCATTCTTGGTGGAAA-AAGCGCCA-3′) and antisense (5′-CCAAGCTTCATTAGAC-TGCGTAGCGGTTGATCC-3′); the PCR product was then purified and digested with *Eco*RI and *Hin*dIII and cloned between the corresponding sites in the pMAL-c2X vector (New England Biolabs, Beverly, MA, U.S.A.). Electrocompetent XL-1Blue *Escherichia coli* cells were transformed with the construct, and the recombinant 38 kDa subunit fused to the C-terminus of maltose-binding protein (38s–MBP) was purified from the cultured cells on amylose-agarose (New England Biolabs) following standard procedures.

Binding of Ag5 to heparin

Heparin-acrylic beads (30 μl; Sigma) equilibrated with 20 mM Tris/HCl, pH 7.4/30 mM $CaCl_2$ were incubated with 10 μg of Ag5 in the same buffer for 30 min, with or without pre-incubation of the antigen with dithiothreitol (100 μM), or with recombinant 38s–MBP. The beads were washed twice with 20 mM Tris/HCl, pH 7.4, containing 30 mM $CaCl_2$ and 100 mM NaCl, and then treated with 20 μl of SDS sample buffer and run on an SDS/15 % polyacrylamide gel under reducing conditions.

Reactive-thiol-groups analysis

Ag5 (5 μg), ovalbumin (5 μg; Sigma) and myoglobin (5 μg; Sigma) were incubated with 1 μl of 10 mM biotinyl-3-maleimido-propionamidyl-3,6-dioxaoctanedimaine (Pierce) in PBS, pH 7.2/0.1 % SDS for 20 min at room temperature. The samples were then subjected to electrophoresis on SDS/12 % polyacrylamide gels under reducing conditions, and transferred to a nitrocellulose membrane (pore size 0.45 μm; Bio-Rad). After blocking with PBS/1 % BSA, the membrane was incubated with ExtrAvidin alkaline phosphatase conjugate (Pierce) and revealed using 5-bromo-4-chloro-3-indolyl phosphate and Nitro Blue Tetrazolium (Sigma) as substrate.

Detection of calcium binding

Ag5, trypsin and BSA were biotinylated using biotinamido-caproate *N*-hydroxysuccinimide (Sigma) and then immobilized on to 10 μl of MagnaBind™ Streptavidin Beads (Pierce). The beads were washed with 10 mM imidazole/HCl, 60 mM KCl, pH 7.0, and were then incubated for 15 min with 26 μl of the same buffer containing 0.05 μCi of $^{45}Ca^{2+}$ in the presence of increasing concentrations of non-radioactive $CaCl_2$. The beads were washed three times with the imidazole buffer, and calcium ions were eluted from the beads by incubation with 50 μl of 10 mM EGTA for 10 min. The β-emission of the supernatant was measured with 1 ml of LSC Cocktail, Ultima Gold (Sigma) in a 1450 MicroBeta Trilux, Wallac counter.

RESULTS

Amino acid sequencing of Ag5 internal fragments

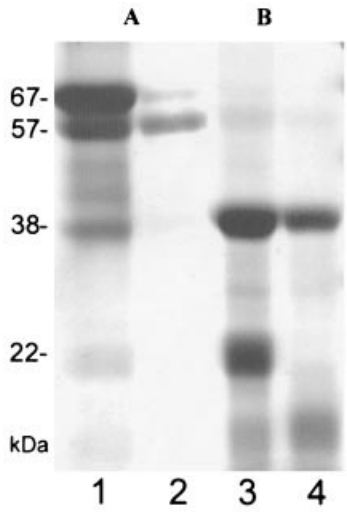
Figure 1(A) illustrates the different components of immuno-purified Ag5 on SDS/PAGE under reducing and non-reducing conditions. As described previously, Ag5 is detected by SDS/PAGE mainly as a 67 kDa band, which dissociates on disulphide-bond reduction into subunits of approximately 38 and 22 kDa. In addition to the 67 kDa band, a 57 kDa component is also found on non-reducing SDS/polyacrylamide gels. The relative amount of this smaller component may vary between batches, as can be seen in Figure 1(A), lanes 1 and 2. The 57 kDa component arises from the main 67 kDa form of Ag5 in a process that involves the removal of the C-terminal portion of the 22 kDa subunit. Evidence of this is the absence of the 22 kDa subunit in reduced samples of the 57 kDa component (lane 4 in Figure 1A) and the N-terminal sequencing of the 57 kDa band (see below).

The Ag5 components were digested *in situ* after SDS/PAGE with trypsin and Lys-C, and the digests were analysed by matrix-assisted laser-desorption ionization–time-of-flight (MALDI–TOF) MS. The peptide fragments were then separated by reversed-phase chromatography and relevant fractions, selected by comparative analysis of the elution profiles and the MS spectra, were sequenced. The amino acid sequences of these fractions are summarized in Figure 1(B). The fragment AANY-XASEGLTRK overlapped the N-terminal sequence of the Ag5 38 kDa subunit described by Zhang and McManus [12].

The 22 and 38 kDa subunits of Ag5 are encoded by a single gene

Initially, a gene fragment was cloned by RT-PCR from *E. granulosus* protoscoleces using degenerate primers designed from amino acid sequences of the 38 kDa internal fragments. The amino acid sequences underlined in Figure 1(B) were chosen to minimize degeneracy. For each region, sense and antisense primers were designed and a set of four PCR reactions were set up to amplify a fragment of Ag5 cDNA from an *E. granulosus* library. Only one combination of primers yielded a PCR product, which was subsequently cloned into pGEM-T Easy vector and sequenced. The fact that several of the internal fragments, shown in Figure 1(B), could be assigned to the deduced amino acid sequence of the product cloned strongly indicates that a fragment of the Ag5-encoding gene has been isolated. On the basis of this new information, gene-specific primers were synthesized and the 5′- and 3′-ends were amplified using the strategy described in the Materials and methods section. The full-length cDNA was 1616 nt long (GenBank® AY052477) containing an 1452 bp open reading frame from an initiator ATG codon at position 38 (Figure 2). The preceding 36-nt-long untranslated region was identical with the spliced leader sequence recently reported to be present in approx. 20 % of *E. granulosus* mRNAs [19]. The deduced amino acid sequence of 485 residues is also shown in Figure 2. It was possible to assign all the internal fragments obtained from the different components of Ag5 to this sequence, including those of the 22 kDa subunit. This not only confirmed that the Ag5 encoding gene was cloned, but also indicated that Ag5 is synthesized as a single polypeptide chain that is afterwards processed into two subunits.

The alignment of the deduced amino acid sequence of Ag5 with that of the proteolytic internal fragments indicates that the 22 kDa subunit is generated by processing of the N-terminal region of the original polypeptide chain (pro-Ag5), whereas the 38 kDa component corresponds to the C-terminal portion. Direct N-terminal amino acid sequencing of the blotted 22 kDa component yielded the sequence LELTLDPDELVKAQRE, which is found in positions 21–36 of the deduced sequence (shown in Figure 2). The preceding sequence corresponds to a hydrophobic leader peptide of 20 amino acids (italicized in Figure 2), which is in agreement with the fact that Ag5 is a secreted protein. The C-terminal end of the 22 kDa component was deduced owing to the availability of the internal fragment LPYS; this component



Sequence	Component	Digested with:
AANYXASEGLTRLK	67, 57 and 38 kDa	Lys-C
SRPISK	57 and 38 kDa	Lys-C
RFHHDDENEAQI	38 kDa	trypsin
RPQRPTSLNEIR	38 kDa	trypsin
YAV	67 and 38 kDa	trypsin
YIAEERPXEDK	67 and 22 kDa	Lys-C
LPYS	67 kDa	trypsin
PSEELTRMQEK	22 kDa	Lys-C
PSRRRWK	67 and 22 kDa	Lys-C
ENIYDGWSR	22 kDa	trypsin
RRRPTFFNPF	38 kDa	trypsin

Figure 1 SDS/PAGE analysis of immunopurified Ag5, and amino acid sequences of internal fragments of the 22, 38 and 67 kDa bands

(**A**) Coomassie Blue staining of immunopurified *E. granulousus* Ag5 run on an SDS/12% polyacrylamide gel under non-reducing (lanes 1 and 2) and reducing conditions (lanes 3 and 4). Lanes 2 and 4 correspond to a different batch of Ag5 in which the 57 kDa band was the main component. Molecular-mass markers are shown on the left. (**B**) Amino acid sequences of internal fragments of the 22, 38 and 67 kDa components of Ag5 after Lys-C or trypsin digestion. Sequences are shown using the one-letter code, where 'X' is an undetermined residue. The regions from which sense and antisense degenerate primers were designed for cloning of a fragment of Ag5 cDNA are shown underlined.

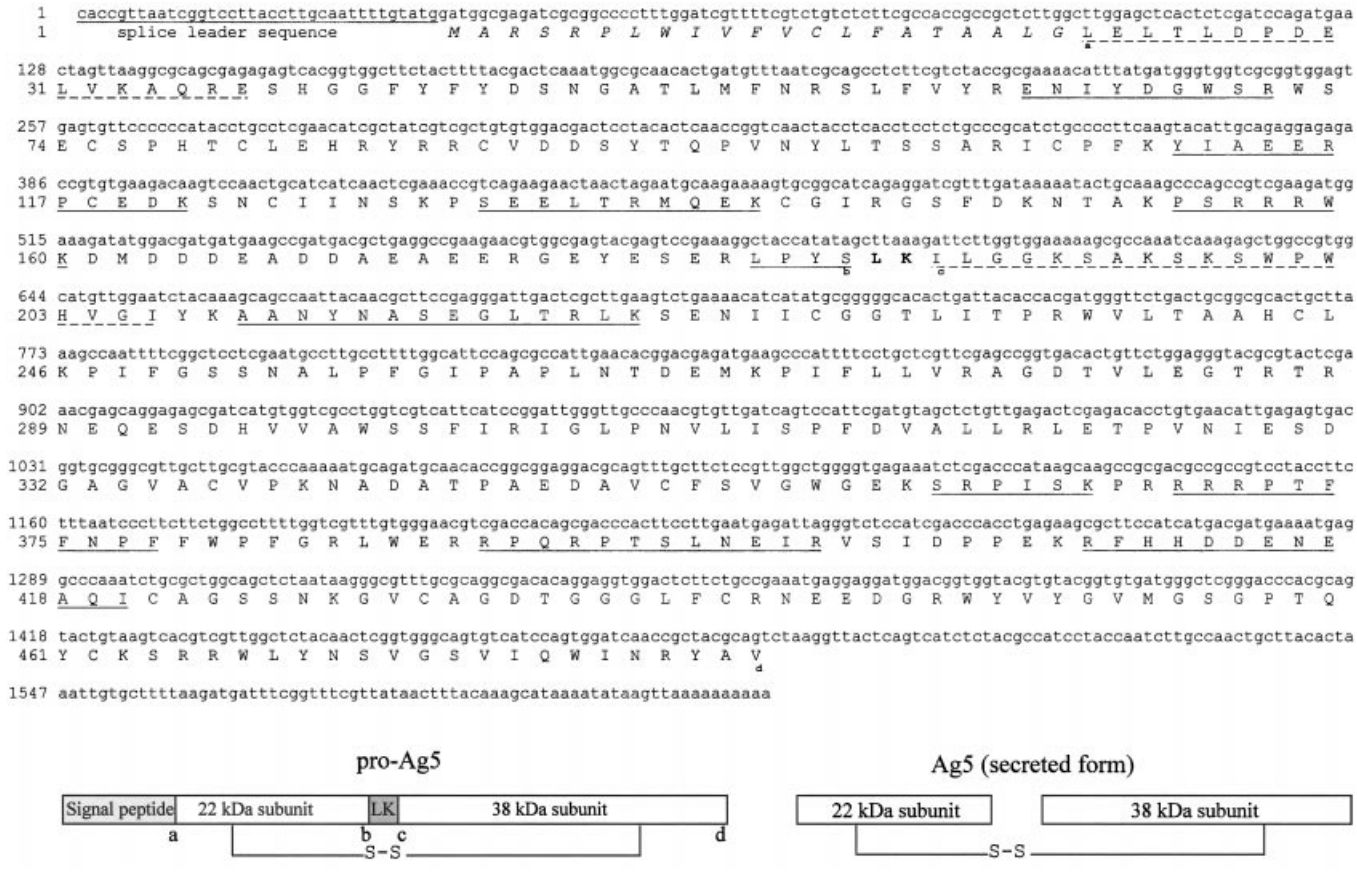


Figure 2 Nucleotide and amino acid deduced sequences from the *E. granulosus* Ag5 cDNA

The complete deduced sequence shows a hydrophobic N-terminal region of 20 amino acids (in italics) that shows characteristics of a signal peptide. The emboldened letters **a** and **b**, and **c** and **d**, in the panel below the sequence, indicate the N- and C-termini of the 22 and 38 kDa subunits respectively. The intervening dipeptide, Leu-Lys (**LK**), removed after chain processing is shown in bold. The sequences corresponding to the proteolytic fragments reported in Figure 1(B) are underlined. The N-terminal amino acid sequences of the 22 and 38 kDa subunits obtained by Edman degradation are underlined by a broken line. The pro- and secreted forms of Ag5 are represented schematically at the bottom of the Figure.

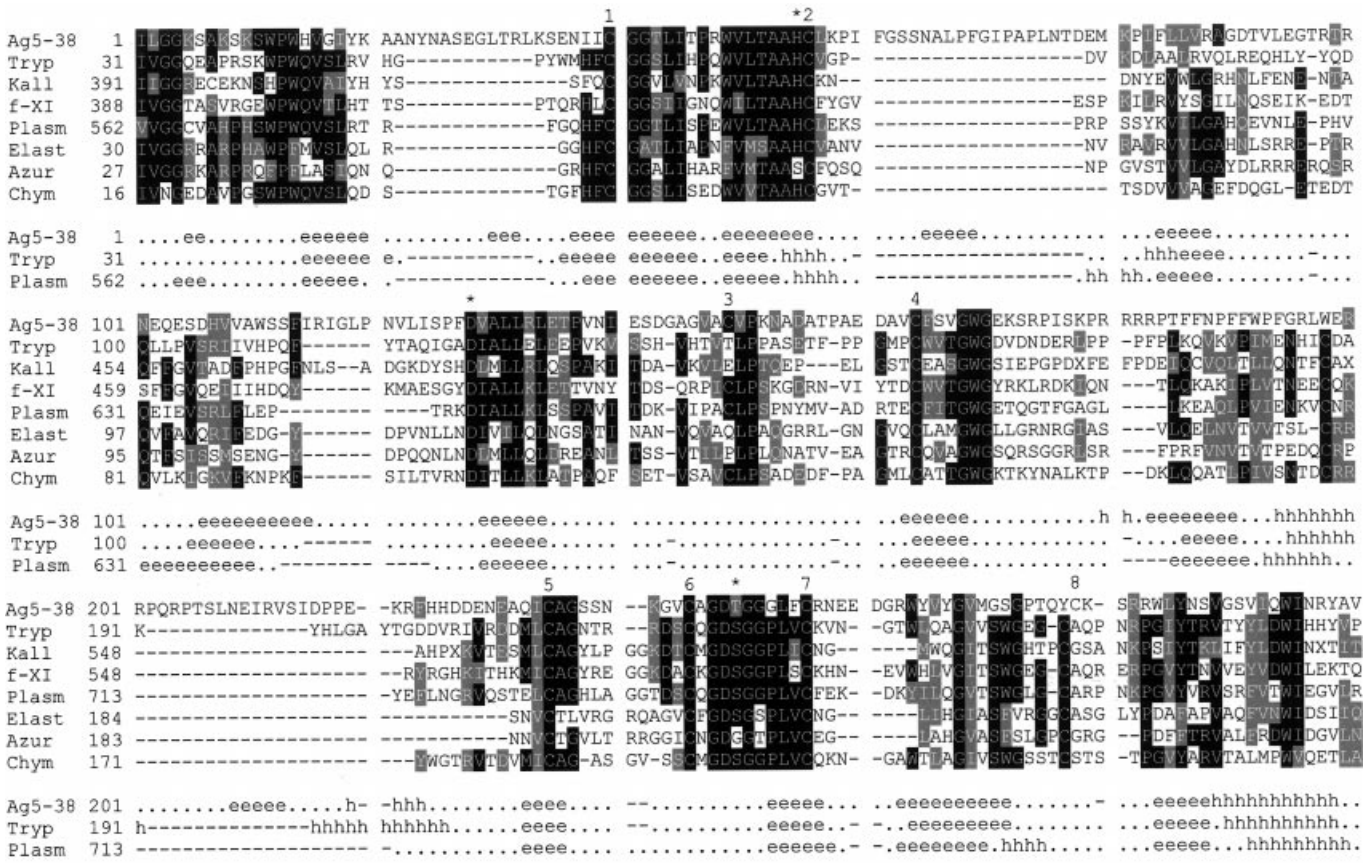


Figure 3 Amino acid sequence alignment of the Ag5 38 kDa subunit with representative enzymes of the trypsin family that yielded significant BLAST score

The amino acid sequence of Ag5 was aligned with enzymes of the trypsin family, including human tryptase (Tryp), porcine kallikrein (Kall), human coagulation factor XI (f-XI), human plasminogen (Plasm), human neutrophil elastase (Elast), human azurocidin (Azur) an inactive member of the family, and porcine chymotrypsin-B (Chym). Alignment was performed using the ClustalW World-Wide-Web Service at the European Bioinformatics Institute. Shading is used to indicate identity (black) or similarity (grey), and dashes represent gaps inserted to maximize the alignment. The positions of the residues of the catalytic triad are marked with asterisks, and the eight cysteine residues of the Ag5 38 kDa subunit are indicated by numbers (1–8) above the sequence. The secondary structure prediction of Ag5, and the secondary structure of tryptase and plasminogen obtained from their three-dimensional structures (Protein Data Bank: 1AOL and 1QRZ respectively) are shown underneath the sequence alignment, i.e. sheets are represented by 'e', α-helix by 'h' and other structure by '.'.

(Figure 1B) differs by two amino acids from the theoretically predicted tryptic fragment, LPYSLK. The accuracy of the LPYS fragment sequence was confirmed by MALDI–TOF-MS determination (calculated mass, 478.55 Da; measured mass, 478.24 Da). Furthermore, when the 38 kDa N-terminal sequence (ILGGKSAKSKSWPWHVGI) reported by Zhang and McManus [12] was considered, it became evident that the mature form of Ag5, consisting of 22 and 38 kDa subunits, appears after removal of the Leu-Lys dipeptide (shown in bold in Figure 2). The calculated relative molecular masses of the deduced amino acid sequences of the 22 and 38 kDa subunits are 19556 and 33008 respectively. The reported glycosylation may account for the variation observed with regard to the electrophoretic mobility. In fact, a putative N-glycosylation site is predicted for each of the subunits, located at residues Asn^{55} and Asn^{215}. The possibility of glycosylation at residue Asn^{215} is consistent with the lack of signal at that position during the sequencing of the fragment AANYXASEGLTRLK by Edman degradation (Figure 1B).

Regarding the 57 kDa form of the Ag5 molecule, as mentioned before, this arises by fragmentation of the 22 kDa component. We attempted to localize this cleavage site by N-terminal sequencing of the 57 kDa component, which yielded the subsequences ILGGKS and LELTLD. These sequences represent the N-termini of the 38 and 22 kDa subunits, showing that, after cleavage, the N-terminal region of the 22 kDa component remains bound to the bigger subunit, while the C-terminal half is removed. Since this latter fragment is lost during the immunopurification of Ag5, this could not be clarified further.

The 38 kDa subunit is similar to serine proteases

The amino acid sequence of the 38 kDa subunit shows high similarity to that of serine proteases of the trypsin family; specifically, with the major neutral proteases of mast cells, tryptases. High BLAST scores were also observed with kallikreins, coagulation factors, plasminogens, neutrophil proteases and pancreatic enzymes. Figure 3 shows the amino acid alignment of the Ag5 38 kDa component with serine proteases representative of these subfamilies. In addition to the overall similarity of their primary structure, main elements of the predicted secondary structure [20] appeared to be well conserved, as is shown by comparison with the secondary structure of human tryptase and plasminogen. However, Ag5 lacks Ser^{195} (chymotrypsinogen numbering), a key residue of the catalytic triad, which is substituted with threonine, a result that was obtained consistently after sequencing of three clones from different PCR reactions.

YDGWSRWSPCSPHTCLEH---RYRRC Ag5 22 kDa component (45-68)
MSPWSEWSQCDP--CLRQMF-RSRSI C9 complement component
WSPWSEWTSCST-SCGNGIQQRGRSC thrombospondin 1
WSHWSPWSSCSV-TCGVGNITRIRLC thrombospondin 2
LSPWSEWSDCSV-TCGKG--MRTRQR VSGP/F-spondin
WSLWSTWAPCSV-TCSEGSQLRYRRC properdin
NSLSTEWSPCSV-TCGNGIQVRIKPG *P. falciparum* circumsporozoite protein

Figure 4 Alignment of residues 45 to 68 of the Ag5 22 kDa subunit with similar sequences from other proteins that bear a cell-adhesive motif

This motif is found in a diverse group of proteins, including thrombospondin, the terminal complement proteins, neural adhesion molecules (spondin) and the malaria circumsporozoite protein. Boxes are used to denote the three characteristic regions of the motif, i.e. the tryptophan-rich, the cysteine-rich and the basic-residues-rich regions (located on the left, centre and right respectively). Highly conserved residues are shown in bold, and gaps are denoted by dashes. VSGP/F, vascular smooth muscle cell growth-promoting factor.

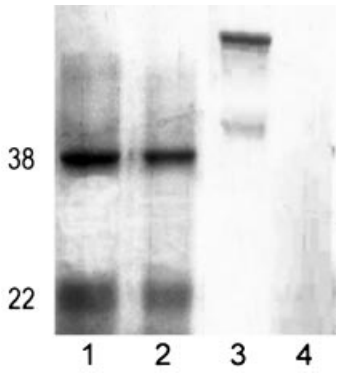


Figure 5 Binding of the Ag5 subunits to heparin analysed by SDS/PAGE

Lanes were loaded as follows: lane 1, native Ag5; lane 2, pre-reduced native Ag5 eluted from heparin–Sepharose; lane 3, 38s–MBP; lane 4, eluate of heparin beads incubated with 38s–MBP. The position of the 38 and 22 kDa subunits are indicated.

The possibility that, in spite of this substitution, Ag5 might still be a functional protease was studied by: (a) zymograms on SDS/polyacrylamide gels containing standard substrates (casein and gelatin); (b) using a panel of synthetic protease substrates representative of chymotrypsin-like serine proteases; and (c) searching for putative physiological substrates, analysing variations in the electrophoretic pattern of host-tissue homogenates (bovine liver) after incubation with Ag5. No evidence of peptidase activity was found with any of these substrates under different pH conditions. In addition, the conservation of other features of the catalytic pocket, such as binding to protein inhibitors, was also examined, but none of the inhibitors investigated (soya-bean trypsin inhibitor, antithrombin III or α_1-antiprotease) reacted with Ag5 (results not shown).

Since an old report described that the precipitation line associated with Ag5 possessed esterase activity [8], we assayed the enzymic activity of the native Ag5, using α-naphtol acetate, α-naphtol butyrate and β-naphtol acetate as substrates [21]. However, the electrophoresed purified antigen did not show esterase activity in non-denaturing gels (results not shown).

The 22 kDa subunit possesses a putative cell-adhesive motif

In the case of the 22 kDa component, no significant similarity was observed with other published protein sequences in databases investigated, except for a highly conserved cell-adhesive motif [22] corresponding to residues 45 to 68 of the 22 kDa subunit (Figure 4). All the features of this motif (namely, a tryptophan-containing region, a middle cysteine-containing region and a downstream portion with basic residues) are present in this subunit. The binding of the 22 kDa subunit to heparin was studied using heparin–acrylic beads (Figure 5). Adsorption of the native antigen or the reduced antigen to the beads was complete (the antigen was undetectable in the supernatants; results not shown), and both components of Ag5 were eluted from the beads. However, the 38s–MBP did not bind to heparin, showing that the retention of this component is more likely to be due to its non-covalent interaction with the 22 kDa component, even in the absence of the disulphide bonds.

Another striking feature of the 22 kDa subunit is the presence of three twin glutamic acid residues and a stretch of highly acidic residues in the C-terminal portion (amino acids 161 to 174), which suggests that Ag5 may bind calcium ions. This was studied using ^{45}Ca, and trypsin and BSA, as positive and negative controls respectively. Figure 6 shows that Ag5 binds ^{45}Ca at ion concentrations significantly below the physiological values, and, similarly to the situation with trypsin, this binding can be displaced completely by competition with an excess of non-radioactive ion. The fact that the 38 kDa subunit lacks the equivalent glutamate residues of loop 70–80 that form the high-affinity calcium-binding site of trypsin suggests that calcium binding is provided by the 22 kDa component. In the non-catalytic subunit of vitamin K-dependent serine proteases of the coagulation cascade, the binding of Ca^{2+} is mediated by adjacent γ-carboxyglutamic (Gla) acid residues. The possibility of a Gla-rich module in Ag5 was investigated by sequencing two peptidic fragments containing twin glutamic acid residues (Figure 1B). In both cases, the result of Edman degradation and subsequent mass determination unequivocally showed that these residues were not carboxylated.

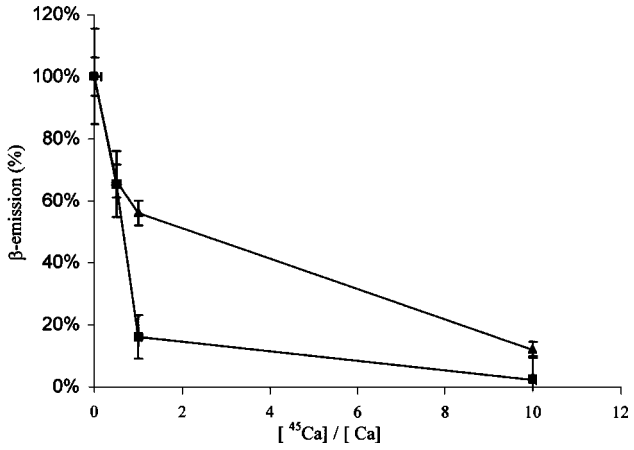


Figure 6 Analysis of calcium-ion binding

Antigen 5 (triangles) and trypsin (squares) adsorbed on to magnetic beads were incubated with fixed amounts of ^{45}Ca ion in the presence of increasing concentrations of the non-radioactive ion. The values for BSA (negative control) were identical with those of the background and are not included in the Figure.

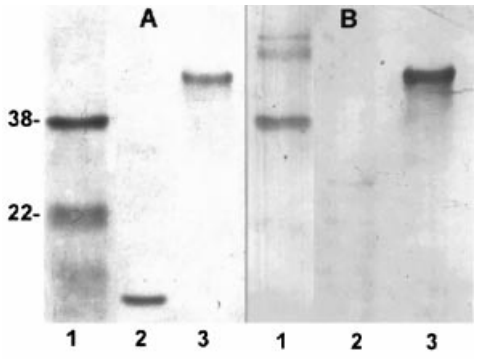


Figure 7 Analysis of free thiol groups in the Ag5 molecule

The Figure shows SDS/PAGE (**A**) and Western blot (**B**) analysis: lanes 1, Ag5; lanes 2, myoglobin (negative control); and lanes 3, ovalbumin (positive control). The three proteins were reacted with biotinyl-3-maleimidopropionamidyl-3,6-dioxaoctanedimaine (maleimide-activated biotin) under denaturing conditions, and were then run under reducing conditions. Biotin derivatization in the blotted samples was detected with a streptavidin–peroxidase conjugate. The positions of the 38 and 22 kDa subunits are shown on the left.

The 38 and 22 kDa subunits are bound by a single disulphide bridge

The 38 and 22 kDa subunits of Ag5 contain eight and seven cysteine residues respectively. In the case of the 38 kDa subunit, by comparison with the disulphide-bond pattern of the trypsin family fold, there might be three conserved disulphide bonds (between cysteine residues 1 and 2, 4 and 7 and 6 and 8), whereas cysteine residues 3 and 5 might be unpaired and, therefore, serve as candidates for the formation of disulphide bonds with the 22 kDa subunit. Whether one or both cysteine of these residues is involved in disulphide formation was analysed by biotin labelling of thiol-reactive groups (Figure 7). This analysis shows that there is a free thiol group in the 38 kDa subunit, and thus a single disulphide bond between the subunits. Cysteine residue 3 is the most probable candidate to be involved in inter-chain disulphide formation, since the equivalent cysteine residue fulfils this function in similar proteases. No reactive thiol groups were observed in the 22 kDa subunit, which indicates that three intra-chain S–S bonds must exist in this subunit.

DISCUSSION

A remarkable characteristic of the Ag5 molecule is the strong similarity of the encoded 38 kDa subunit with serine proteases of SA (chymotrypsin-like) clans [23]. Highest similarities are observed at the N-terminal extension, within conserved regions that contain the residues of the catalytic triad ($His^{57}/Asp^{102}/Ser^{195}$), and at the C-terminal region, which contains critical residues of the S1 pocket. In addition, the main elements of its secondary structure and half cystines are well conserved. Furthermore, the fact that Ag5 is composed of two subunits that are bridged by a single disulphide bond is highly reminiscent of the architecture of other serine proteases of the trypsin family, such as blood-clotting proteases. Nevertheless, the 38 kDa subunit has a larger size due to the occurrence of three insertions at approximately positions 21 to 37, 63 to 80 and 200 to 222. According to the predicted structure of the antigen, these insertions occur in connecting loops between β-strands 1 and 2 and strands 3 and 4 of domain 1, and between strands 2 and 3 of domain 2. The insertion in the loop region 3–4 occurs several residues downstream of the catalytic histidine residue. The other insertions do not affect the loop regions that contain the other catalytic residues, or those residues that are instrumental for the formation of the oxyanion hole (residues 193 to 195) or the specificity pocket (residues 189, 214 to 216 and 226). In this way, these insertions would increase the size of the domains without distorting the formation of the β barrels or the basic elements of the catalytic machinery.

Perhaps the most important difference between Ag5 and other serine proteases is that the serine residue of the catalytic triad is substituted by threonine. This was a main point of concern during this study. The catalytic activity of trypsin-like proteases is provided by a charge-relay system, where the nucleophilicity of Ser^{195} is highly increased by the interaction with His^{57} and Asp^{102}, which 'pull off' the proton of Ser^{195}. The composition of this triad is highly conserved in chymotrypsin-like serine proteases, and there are examples of proteins that clearly belong to this family, but are inactive due to substitutions of these residues. They include proteins with diverse functions, such as haptoglobulin [24], azurocidine [25] and protein Z [26], in which the catalytic triad has been modified to Lys/Asp/Ala, Ser/Asp/Gly and Lys/Asp/Asp respectively. Comparatively, the single $Ser^{195} \rightarrow Thr$ substitution observed in Ag5 is certainly a less dramatic alteration. Serine and threonine differ in conformation and bulk, but there are not significant chemical differences between their Oγ atoms. Actually, catalytic threonine residues are found in different hydrolytic enzymes, such as the proteasome catalytic subunit [27], asparaginase [28] and aspartyl glucosaminidase [29]. In addition, other elements that contribute to the stabilization of the negatively charged carbonyl oxygen in the transition state, notably Gly^{193}, are also conserved in Ag5. However, no signs of enzymic activity were detected with the proteolytic assays performed in this study.

Owing to the fact that most of the trypsin-like proteases are synthesized and secreted as inactive zymogens, which are subsequently activated by proteolysis, we considered the possibility that some activation step could be needed. In the case of Ag5, we speculated that the minor 57 kDa component, which is almost absent in fresh native Ag5 preparations, could represent the active form of the antigen. Short incubation of native immunopurified Ag5 with low concentrations of trypsin quantitatively convert the 67 kDa Ag5 molecule into a 57 kDa form, with electrophoretic mobility identical with that of the 57 kDa minor component of Ag5 (results not shown). Although this demonstrated the existence of a peptide bond that is particularly sensitive to proteolytic cleavage in the 22 kDa subunit of Ag5 (Figure 1), it does not appear to result in the activation of the Ag5 molecule, since the cleaved 57 kDa form of Ag5 was also inactive in our proteolytic assays.

Regarding the 22 kDa subunit, the heparan sulphate proteoglycans and calcium-binding sites found in this component seem to provide binding targets for the Ag5 molecule. These would target the antigen and ensure its localization in the host tissue surrounding the metacestode, or otherwise, the mucosal epithelium of the *E. granulosus* definitive host.

In conclusion, the overall profile characterized for the Ag5 molecule is highly similar to that found for other members of the trypsin family, with a small subunit that provides interactions with cell surfaces and the extracellular matrix, and a large subunit that possesses many structural features that are highly similar to those found in the catalytic domain characteristic of this family. The single amino-acid substitution in the catalytic triad ($Ser^{195} \rightarrow Thr$) is the most intriguing feature of this molecule, and to our knowledge, there is no other example of such a conserved substitution in this family. Although this change may render Ag5 inactive, the need for a specific activation step or the

existence of a highly specific physiological substrate cannot be ruled out. In any case, whatever the functionality of this cestode molecule may be, it constitutes an interesting case in the evolution of serine proteases, where the trypsin-like structure might have been adapted to carry out new biological functions. We hope that this work will provide a basis for further studies that allow the identification and understanding of these functions.

This work was supported by the Comisión Sectorial de Investigación Científica (CSIC; Universidad de la República, Uruguay), the Zaffaroni Foundation (Palo Alto, CA, U.S.A.), and grant #D43 TW005718 from the Fogarty International Center of the United States National Institutes of Health. We wish to thank Dr Claudio Benech for his help with the calcium-binding assays.

REFERENCES

1 Capron, A., Vernes, A. and Biguet, J. (1967) Le diagnostic immuno-électrophorétique de l'hydatidose, pp. 27–40, SIMEP, Lyon
2 Pozzuoli, R., Musinai, P., Arru, E., Patrono, C. and Piantelli, M. (1974) *Echinococcus granulosus*: evaluation of purified antigens immunoreactivity. Exp. Parasitol. **35**, 52–60
3 Musiani, P., Piantelli, M., Lauriola, L., Arru, E. and Pozzuoli, R. (1978) *Echinococcus granulosus*: specific quantification of the two most immunoreactive antigens in hydatid fluids. J. Clin. Pathol. **31**, 475–478
4 Gottstein, B., Witassek, F. and Eckert, J. (1986) Neues zur Echinokokkose. Med. Wochenschr. **116**, 810–817
5 Shepherd, J. and McManus, D. (1987) Specific and cross-reactive antigens of *Echinococcus granulosus* hydatid cyst fluid. Mol. Biochem. Parasitol. **25**, 143–154
6 Lightowlers, M., Liu, D., Haralambous, A. and Rickard, M. (1989) Subunit composition and specificity of the major cyst fluid antigens of *Echinococcus granulosus*. Mol. Biochem. Parasitol. **37**, 171–178
7 Barbieri, M., Fernandez, V., Gonzalez, G., Luaces, V. M. and Nieto, A. (1998) Diagnostic evaluation of a synthetic peptide derived from a novel antigen B subunit as related to other available peptides and native antigens used for serology of cystic hydatidosis. Parasite Immunol. **20**, 51–61
8 Bout, D., Fruit, J. and Capron, A. (1974) Purification d'un antigène spécifique du liquide hydatique. Ann. Immunol. (Paris) **125**, 775–788
9 Willis, A. C., Diaz, A., Nieto, A. and Sim, R. B. (1997) *Echinococcus granulosus* antigen 5 may be a serine proteinase. Parasite Immunol. **19**, 385
10 Di Felice, G., Pini, C., Afferni, C. and Vicari, G. (1986) Purification and partial charcterization of the major antigen of *Echinococcus granulosus* (antigen 5) with monoclonal antibodies. Mol. Biochem. Parasitol. **20**, 133–142
11 March, F., Enrich, C., Mercader, M., Sanchez, M., Muñoz, C., Coll, P. and Prats, G. (1991) *Echinococcus granulosus*: antigen characterization by chemical treatment and enzymatic deglycosylation. Exp. Parasitol. **73**, 433–439
12 Zhang, L. H. and McManus, D. P. (1996) Purification and N-terminal amino acid sequencing of *Echinococcus granulosus* antigen 5. Parasite Immunol. **18**, 597–606
13 Facon, B., Chamekh, M., Dissous, C. and Capron, A. (1991) Molecular cloning of an *Echinococcus granulosus* protein expressing and immunogenic epitope of antigen 5. Mol. Biochem. Parasitol. **45**, 233–240
14 Siles-Lucas, M., Gottstein, B. and Felleisen, R. S. (1998) Identification of a differentially expressed *Echinococcus multilocularis* protein Em6 potentially related to antigen 5 of *Echinococcus granulosus*. Parasite Immunol. **20**, 473–481
15 Gonzalez, G., Spinelli, P., Lorenzo, C., Hellman, U., Nieto, A., Willis, A. and Salinas, G. (2000) Molecular characterization of P-29, a metacestode-specific component of *Echinococcus granulosus* which is immunologically related to, but distinct from, antigen 5. Mol. Biochem. Parasitol. **105**, 177–184
16 Gonzalez, G., Nieto, A., Fernandez, C., Orn, A., Wernstedt, C. and Hellman, U. (1996) Two different 8 kDa monomers are involved in the oligomeric organization of the native *Echinococcus granulosus* antigen B. Parasite Immunol. **18**, 587–596
17 Gonzalez, G., Orn, A., Cazzulo, J. J. and Gronvik, K. O. (1994) Production and characterization of monoclonal antibodies against the major cysteine proteinase of *Trypanosoma cruzi*. Scand. J. Immunol. **40**, 389–394
18 Hellman, U., Wernstedt, C., Gonez, J. and Heldin, C. H. (1995) Improvement of an ''In-Gel'' digestion procedure for the micropreparation of internal protein fragments for amino acid sequencing. Anal. Biochem. **224**, 451–455
19 Brehm, K., Jensen, K. and Frosch, M. (2000) mRNA *trans*-splicing in the human parasitic cestode *Echinococcus multilocularis*. J. Biol. Chem. **275**, 38311–38318
20 Rost, B. and Sander, C. (1993) Prediction of protein secondary structure at better than 70% accuracy. J. Mol. Biol. **232**, 584–599
21 Manchenko, G. P. (1994) Handbook of Detection of Enzymes on Electrophoretic Gels, pp. 178–179, CRC Press, Inc., Boca Raton
22 Gantt, S. M., Clavijo, P., Bai, X., Esko, J. D. and Sinnis, P. (1997) Cell adhesion to a motif shared by the malaria circumsporozoite protein and thrombospondin is mediated by its glycosaminoglycan-binding region and not by CSVTCG. J. Biol. Chem. **272**, 19205–19213
23 Lesk, A. M. and Fordham, W. D. (1996) Conservation and variability in the structures of serine proteinases of the chymotrypsin family. J. Mol. Biol. **258**, 501–537
24 Wassell, J. (2000) Haptoglobin: function and polymorphism. Clin. Lab. **46**, 547–552
25 Morgan, J. G., Sukiennicki, T., Pereira, H. A., Spitznagel, J. K., Guerra, M. E. and Larrick, J. W. (1991) Cloning of the cDNA for the serine protease homologous CAP37/azurocidin, a microbicidal and chemotactic protein from human granulocytes. J. Immunol. **147**, 3210–3214
26 Ichinose, A., Takeya, H., Espling, E., Iwanaga, S., Kisiel, W. and Davie, E. W. (1990) Amino acid sequence of human protein Z, a vitamin K-dependent plasma glycoprotein. Biochem. Biophys. Res. Commun. **172**, 1139–1144
27 Seemuller, E., Lupas, A., Stock, D., Lowe, J., Huber, R. and Baumeister, W. (1995) Proteasome from *Thermoplasma acidophilum*: a threonine potease. Science **268**, 579–582
28 Palm, G. J., Lubkowski, J., Derst, C., Schleper, S., Rohm, K. H. and Wlodawer, A. (1996) A covalently bound catalytic intermediate in *Escherichia coli* asparaginase: crystal structure of a Thr-89-Val mutant. FEBS Lett. **390**, 211–216
29 Oinonen, C., Tikkanen, R., Rouvinen, J. and Peltonen, L. (1995) Three-dimensional structure of human lysosomal aspartylglucosaminidase. Nat. Struct. Biol. **2**, 1102–1108

Received 5 September 2002/27 September 2002; accepted 1 October 2002
Published as BJ Immediate Publication 1 October 2002, DOI 10.1042/BJ20021402