A male-specific (cysteine-rich) protein of *Oesophagostomum dentatum* (Strongylida) with structural characteristics of a serine protease inhibitor containing two trypsin inhibitor-like domains

P. R. BOAG[1,3]†, S. RANGANATHAN[2], S. E. NEWTON[1] *and* R. B. GASSER[3]*

[1] *Victorian Institute of Animal Science, Attwood, Victoria 3049, Australia*
[2] *Bioinformatics Centre, National University of Singapore, Singapore*
[3] *Department of Veterinary Science, The University of Melbourne, Werribee, Victoria 3030, Australia*

(*Received 22 April 2002; revised 3 July 2002; accepted 3 July 2002*)

SUMMARY

A cDNA was isolated from an adult male *Oesophagostomum dentatum* gene library by screening with a male-specific, partial expressed sequence tag (EST) probe identified previously using a differential display technique. The full-length cDNA of 642 bp included 5′ and 3′ untranslated regions of 44 and 121 nucleotides, respectively, and encoded a predicted protein with a putative 18 amino acid signal sequence and a mature polypeptide of 14·7 kDa comprising ~ 15% cysteine residues. The amino acid sequence showed similarity with a number of proteins from *Caenorhabditis elegans*, parasitic nematodes, insects and amphibia, all of which contain a trypsin inhibitor-like cysteine-rich domain. A 3-dimensional structure model constructed for the *O. dentatum* protein (designated OdmCRP) inferred that it is composed of 2 domains, each with 5 disulfide bonds, which are indicative of the *Ascaris* family of serine protease inhibitors. These findings indicate that OdmCRP, with 2 structural domains relating to functionally active sites, is a new member of this inhibitor family.

Key words: gender-specific expression, *Oesophagostomum dentatum*, serine protease inhibitor, parasitic nematode, structural model, trypsin inhibitor-like.

INTRODUCTION

Investigation of molecular reproductive processes in parasitic nematodes has important implications for understanding fundamental aspects, such as male–female attraction, gametogenesis and sperm–egg recognition. Also, an understanding of the molecular interplay between male and female nematodes could lead to novel approaches for parasite control, through interruption of the life-cycle and hence transmission. This is of major relevance, particularly given the current problems with anthelmintic resistance in parasitic nematodes of veterinary importance (Jackson & Coop, 2000).

The porcine nodule worm, *O. dentatum* (Nematoda: Strongylida), represents a useful model system for studying fundamental aspects of reproductive biology (Christensen *et al*. 1996; Christensen, 1997). It has a short (21-day) and direct life-cycle, produces large numbers of progeny and can be readily maintained as a laboratory line (Talvik *et al*. 1997). Importantly, uni-sex or mixed-sex infections can be established experimentally by non-invasive, rectal transplantation into naïve pigs (Christensen, Grøndahl-Nielsen & Nansen, 1996), offering the potential for detailed studies of mating behaviour and sexual maturation *in vivo*. For these reasons, we recently investigated gender-specific gene expression in *O. dentatum* using RNA arbitrarily-primed PCR to identify genes which are expressed during the development and sexual maturation of this parasite (Boag *et al*. 2000). Ten male-specific and 2 female-specific ESTs (expressed in 4th-stage larvae and adults) were isolated, cloned and sequenced. One of the male-specific ESTs (Od153) had similarity with several *Caenorhabditis elegans* proteins predicted to contain 2 trypsin inhibitor-like domains. The present study reports the cloning and sequencing of the full-length cDNA representing Od153 (designated *Odmcrp*), recombinant expression of the predicted mature peptide, and development of a 3-dimensional structural model for the mature peptide. The findings suggest that this protein is related to a large family of tight-binding, small serine protease inhibitors (SSPIs) found in a diverse range of organisms, including parasitic nematodes.

* Corresponding author: Department of Veterinary Science, The University of Melbourne, 250 Princes Highway, Werribee, Victoria 3030, Australia. Tel: +61 3 9731 2283. Fax: +61 3 9731 2366. E-mail: robinbg@unimelb.edu.au
† Present address: Center for Blood Research and Department of Pathology, Harvard Medical School, Boston, USA.

Parasitology (2002), **125**, 445–455.
DOI: 10.1017/S0031182002002329 Printed in the United Kingdom

MATERIALS AND METHODS

Parasite material

O. dentatum was propagated in helminth naïve pigs as described previously (Boag *et al.* 2000). Worms were collected from the large intestines of experimentally infected pigs 15 (L4) and 22 (adult) days post-infection, and the 2 sexes separated prior to freezing in liquid nitrogen.

Isolation of RNA, and construction and screening of a male-specific O. dentatum *cDNA library*

Total RNA was extracted from 22-day-old male *O. dentatum* using Tripure™ (Roche Molecular Biochemicals) and mRNA purified using Poly(A)Pure (Ambion), according to the manufacturers' instructions. A cDNA library was constructed using a Superscript cDNA synthesis kit (Life Technologies) by priming with oligo-dT. The double-stranded cDNA was cloned into λ ZapII and subsequently packaged (Stratagene) as per the manufacturer's instructions. The library contained $\sim 500\,000$ independent clones and was amplified once prior to screening. Approximately 2×10^5 plaque-forming units from the amplified λ ZapII library were screened using a PCR product amplified from the clone Od153 (Accession no. AB029052) (Boag *et al.* 2000) and random-labelled (Promega) with α-^{32}P dCTP (Geneworks). Plaque lifts on to positively-charged nylon membranes (Boehringer-Mannheim) were performed in duplicate, and the membranes pre-hybridized at 65 °C for 3–5 h in Church buffer (0·5 M $NaHPO_4$, 1% (w/v) bovine serum albumin, 1 mM EDTA, 7% (w/v) SDS) (Church & Gilbert, 1984) containing 100 μg/ml of denatured sonicated herring sperm DNA. The denatured probe was added to the hybridization solution and incubated overnight at 65 °C. The filters were washed once in 2 × SSC (0·3 M NaCl, 30 mM Na citrate, pH 7·0) and twice in 0·1 × SSC, 0·1% (w/v) SDS for 20 min at 60 °C. Filters were then exposed to SuperRX X-ray film (Fuji) for 48 h at −70 °C using an intensifying screen. Six positive clones were identified and, after secondary screens, plasmids rescued by *in vivo* excision, according to the instructions of the manufacturer of the λ ZapII vector. Plasmid DNA was purified (Qiagen) and then sequenced using PRISM™ Big Dye terminator chemistry in an automated sequencer (Applied Biosystems).

Computer-based analyses

Nucleotide sequences were compared with the GenBank non-redundant databases using the BLAST 2.0 (Altschul *et al.* 1997) suite of programs from the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/BLAST/), the Sanger Centre (www.sanger.ac.uk/Projects/Celegans/) and the Parasite Genome database(www.ebi.ac.uk/parasites/parasite_blast_server.html). Protein patterns were identified by searching the PROSITE database (Bairoch, 1993) (www.expasy.ch/tools/scnpsit1.html) and Pfam database (Bateman *et al.* 2000) (www.pfam.wustl.edu/hmmsearch.shtml), and the potential signal sequence identified using SignalP V2.0 (Nielsen *et al.* 1997) at the Center for Biological Sequence Analysis (www.cbs.dtu.dk/services/SignalP/). Protein sequence alignments were performed using Clustal W (Thompson, Higgins & Gibson, 1994). The computer program Modeller (Sali & Blundell, 1993) was used to generate structural models for OdmCRP, specifying the formation of 5 disulfide bridges in each putative domain, which were displayed as MOLSCRIPT cartoons (Kraulis, 1991). The expression pattern of selected *C. elegans* genes was identified from the micro-array data available at the website http://cmgm.stanford.edu/~kimlab/wmdirectorybig.html.

Cloning, expression and isolation of the recombinant protein, rOdmCRP

A 411 bp fragment was amplified from the *Odmcrp* clone by PCR using primers (incorporating appropriate restriction endonuclease sites) designed to nucleotides 99–116 (forward primer 5′ CCGGGATCCGCATTACCACAGGTGCGG 3′) and 489–509 (reverse primer 5′ GATGCTCGAGTTACTTCGTCGATTTTTTTGCA 3′). The PCR product was digested with *Bam*HI and *Xho*I, gel-purified (Qiagen) and cloned in frame into the 6xHistidine-tagged vector pET30A (Novagen). The correct reading frame was confirmed by sequencing, and the construct transformed into the BL21 (DE3)pLysS strain of *Escherichia coli*. Recombinant protein expression was induced in log-phase cultures by addition of isopropyl-β-D-thiogalactopyranoside (IPTG) to 1 mM. The cells were incubated at 24 °C for 2 h, harvested by centrifugation at 3000 ***g***, resuspended in binding buffer (500 mM NaCl, 20 mM Tris–HCl, pH 7·4, 5 mM imidazole) and lysed by sonication. The cell sonicate was then centrifuged at 15 000 ***g*** and the supernatant fraction containing the recombinant protein subjected to 2 rounds of affinity chromatography using nickel nitrilo-triacetic acid (Ni-NTA) matrix (Qiagen). The columns were washed with 20 column volumes of binding buffer and the protein eluted with binding buffer containing 50–200 mM imidazole, followed by dialysis against Tris–HCl, pH 7·8, for 16 h at 4 °C and storage at −70 °C.

Production of antisera

Polyclonal antiserum to purified recombinant OdmCRP (rOdmCRP) was raised in 6-month-old, female New Zealand White rabbits by subcutaneous

https://doi.org/10.1017/S0031182002002329 Published online by Cambridge University Press

injection of 100 μg of protein in Freund's Incomplete Adjuvant (Sigma). A pre-bleed was taken from each rabbit. Three injections were given at monthly intervals, and serum was collected from whole blood by centrifugation at 3500 ***g*** and stored at −20 °C.

Antiserum was pre-absorbed to remove antibodies to bacterial proteins by mixing with an equal volume of sonicated BL21 *E. coli* lysate at 4 °C for 2 h, and the insoluble material removed by centrifugation at 13000 ***g***. To affinity purify antibodies to rOdmCRP, the protein was subjected to SDS–PAGE and transferred to polyvinylidene fluoride (PVDF) membranes (as described below). The membranes were stained with Ponceau S, and the region containing the protein was excised and incubated with pre-absorbed rabbit antiserum diluted 1:5 in phosphate-buffered saline (PBS) (0·17 M NaCl, 3·3 mM KCl, 9 mM Na_2HPO_4, 1·8 mM KH_2PO_4, pH 7·2) at room temperature for 60 min. The membranes were washed extensively with PBS–0·5 % (v/v) Tween-20 (PBS-T) and once with borate buffer (0·1 M boric acid, pH 8, 0·5 M NaCl, 0·05 % Tween). The bound affinity-purified antibody was eluted with 0·1 M glycine–HCl (pH 2·6), 0·15 M NaCl and then immediately neutralized with half of the elution volume of 1 M Tris–HCl, pH 8·0 (Harlow & Lane, 1988).

Protein extraction and Western blot analysis

L4s or adult males of *O. dentatum* were homogenized in a hand-held glass homogenizer in ice-cold PBS and incubated at 4 °C for 1 h. The lysate was centrifuged at 15000 ***g*** and the supernatant fraction containing the PBS-soluble proteins stored at −70 °C. Protein concentrations were determined using a Micro-BCA protein estimation kit (Pierce), and samples (5 μg per lane) were resolved by reducing SDS–PAGE on Tris–tricine 10–20 % (w/v) linear-gradient polyacrylamide gels, with a 4 % (w/v) stacking gel (Novex) under standard conditions (Harlow & Lane, 1988). Where required, gels were stained with SimplyBlue (Novex). For Western blots, proteins were transferred electrophoretically to PVDF membranes (Millipore Waters) using the Mini-Protean-Blot system (Bio-Rad). Membranes were blocked with 5 % (w/v) non-fat milk powder in PBS-T (PBS-TB) for a minimum of 30 min at 22–24 °C. Membranes were incubated overnight at 4 °C with affinity-purified rabbit antibodies to rOdmCRP (1:5 dilution). Bound rabbit antibody was detected using a sheep anti-rabbit IgG conjugated to alkaline phosphatase (Silenus Laboratories, AMRAD). Antibody complexes were detected using 5-bromo-4-chloro-3-indolyl phosphate/nitroblue tetrazolium (BCIP/NBT) (Promega) diluted in 100 mM NaCl, 5 mM $MgCl_2$, 100 mM Tris (pH 9·5) as a substrate. Recombinant OdmCRP was detected using a monoclonal antibody against the 6xHistidine tag (Clontech) diluted 1:5000 in PBS-TB, and bound primary antibody was detected using a sheep anti-mouse IgG conjugated to alkaline phosphatase at a dilution of 1:1000 (Silenus Laboratories, AMRAD). Molecular weights were estimated with reference to protein standards 3–98 kDa (Novex) and/or 10–130 kDa (Benchmark Prestained Protein Ladder, Life Technologies).

RESULTS

Characterization of Odmcrp

We have previously reported the isolation of a 322 bp partial cDNA clone (designated Od153) by RNA arbitrarily-primed PCR, which specifically recognized on Northern blot a transcript in L4s and in adult males of *O. dentatum* that was absent from female worms (Boag *et al.* 2000). To obtain a full-length cDNA clone, the radio-isotope labelled insert from clone Od153 was used as a probe to screen a 22-day-old adult male *O. dentatum* cDNA library, and 6 positive clones were identified. The 'largest' cDNA clone contained an insert of 642 nucleotides (Accession no. AF399936), which was smaller than the estimated mRNA size of 900 nucleotides (Boag *et al.* 2000). The method of 5′-Rapid Amplification of cDNA Ends (5′-RACE) (Frohman, Dush & Martin, 1988) was carried out using gene-specific primers, which confirmed that a full-length cDNA had been cloned (data not shown). The full-length clone contained a 5′-untranslated region (UTR) of 44 bp (which did not include a nematode splice leader sequence; Blaxter & Liu, 1996), a 3′-UTR of 121 bp with a polyadenylation signal (AATAAA) 14 bp upstream of the poly-A tail and a single open reading frame (ORF) encoding a predicted polypeptide of 154 amino acids (Fig. 1). This polypeptide had a putative 18 amino acid signal sequence and a predicted mature protein of 136 residues, including 20 cysteines (~15 %), and contained 2 trypsin inhibitor-like (TIL) cysteine rich domains (Pfam: PF01826) between amino acids 8–59 and 64–131 (numbering refers to the mature protein). Within these domains, an EGF-like cysteine pattern (PROSITE: PS01186, C-x-C-x(2)-[GP]-[FYW]-x(4,8)-C) could be identified in the N-terminal portion, while a second region at the C-terminus also contained an EGF-like pattern, with the exception of a difference in the spacing of the final cysteine residue in the motif. Sequence alignment of the amino and carboxyl halves of OdmCRP revealed 21 % identity and 46 % similarity between the 2 domains (Fig. 2B).

Homology searches

Database searches with the predicted peptide sequence of OdmCRP revealed greatest similarity with 3 predicted *C. elegans* proteins, C25E10.7 (26 % identity and 54 % similarity over 153 amino acids), C25E10.8 (32 % identity and 53 % similarity over

https://doi.org/10.1017/S0031182002002329 Published online by Cambridge University Press

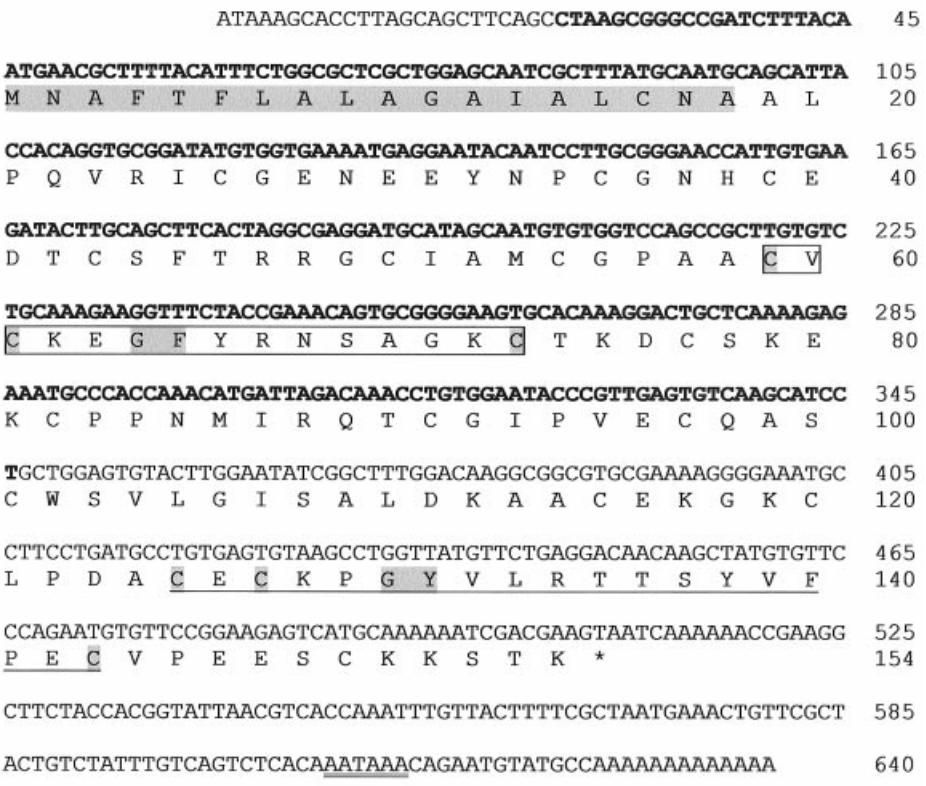

Fig. 1. Nucleotide and predicted amino acid sequences for the *Oesphagostomum dentatum* cDNA clone *Odmcrp*. The predicted signal sequence (shaded), the TAA translation termination signal (asterisk) and the putative polyadenylation signal sequence (double underlined) are indicated. The nucleotide sequence determined from the original EST Od153 (Accession no. AB029052) is shown in bold, and the amino acid residues representing PROSITE PS01186 EGF-like domain signature 2 (C-x-C-x[2]-[GP]-[FYW]x[4,8]-C) are boxed, with key residues shaded. Amino acids forming the variant EGF-like domain are underlined, with key residues shaded.

141 amino acids) and C25E10.9 (32 % identity and 54 % similarity over 140 amino acids). The 3 *C. elegans* proteins all contain 2 EGF-like cysteine patterns within 2 TIL domains (PFAM: PF01826) (Fig. 2B), and are arranged contiguously in the *C. elegans* genome. An additional gene, CE2510.10, is also contiguous in this region and shares the TIL domain, but the overall similarity is lower (16 % identity and 39 % similarity over 152 amino acids). Proteins OdmCRP, C25E10.8 and C25E10.9 contain predicted signal sequences and, together with C25E10.7, all contain 20 cysteine residues, arranged in a spatially conserved manner in the mature protein (Fig. 2A).

Significant sequence similarity was also detected between OdmCRP and several members of the *Ascaris*-type SSPI, chymotrypsin/elastase inhibitors (C/E-1 and C/E-4) from *Ascaris suum* (see Huang *et al*. 1994), a trypsin inhibitor (ATI) from *A. lumbricoides* (see Peanasky *et al*. 1984), an elastase (ASPI-1; PDB: 1EAI) inhibitor from *Anisakis simplex* (see Nguyen *et al*. 1999), a chymotrypsin/cathepsin G inhibitor (AMCI-1; PDB: 1CCV) from the honeybee, *Apis mellifera* (see Cierpicki, Bania & Otlewski, 2000), and a trypsin/thrombin inhibitor (BSTI) from the European frog, *Bombina bombina* (see Mignogna *et al*. 1996). There was also similarity between OdmCRP and the anti-coagulant proteins from *Ancylostoma caninum* (see Cappello *et al*. 1995, 1996; Stassens *et al*. 1996) and *A. ceylanicum* (Accession no. AAK81733), which are related members of the *Ascaris*-type SSPIs. These inhibitors also have 10 cysteine residues, which are arranged in a spatially similar manner, and align with both the N- and C-terminal cysteine domains of OdmCRP and the 3 proteins (C25E10.7, C25E10.8 and C25E10.9) predicted for *C. elegans* (Fig. 2C). The C-terminal domain of OdmCRP contains 3 regions which have insertions of 2–6 amino acids relative to the other members of the *Ascaris* family of inhibitors and the predicted *C. elegans* proteins. Similarity to the OdmCRP peptide was also identified in ESTs from *A. suum*, *A. ceylanicum*, *Strongyloides stercoralis* and *Trichuris muris* (Accession nos. AW165769, AF172653, BE581269 and BF169285, respectively), which all contain either 1 or 2 TIL domains, indicating that this domain is relatively widely distributed throughout the phylum Nematoda.

Recombinant protein expression and Western blot analysis

SDS–PAGE and Western blot analyses of *E. coli*-expressed and purified rOdmCRP revealed the

https://doi.org/10.1017/S0031182002002329 Published online by Cambridge University Press

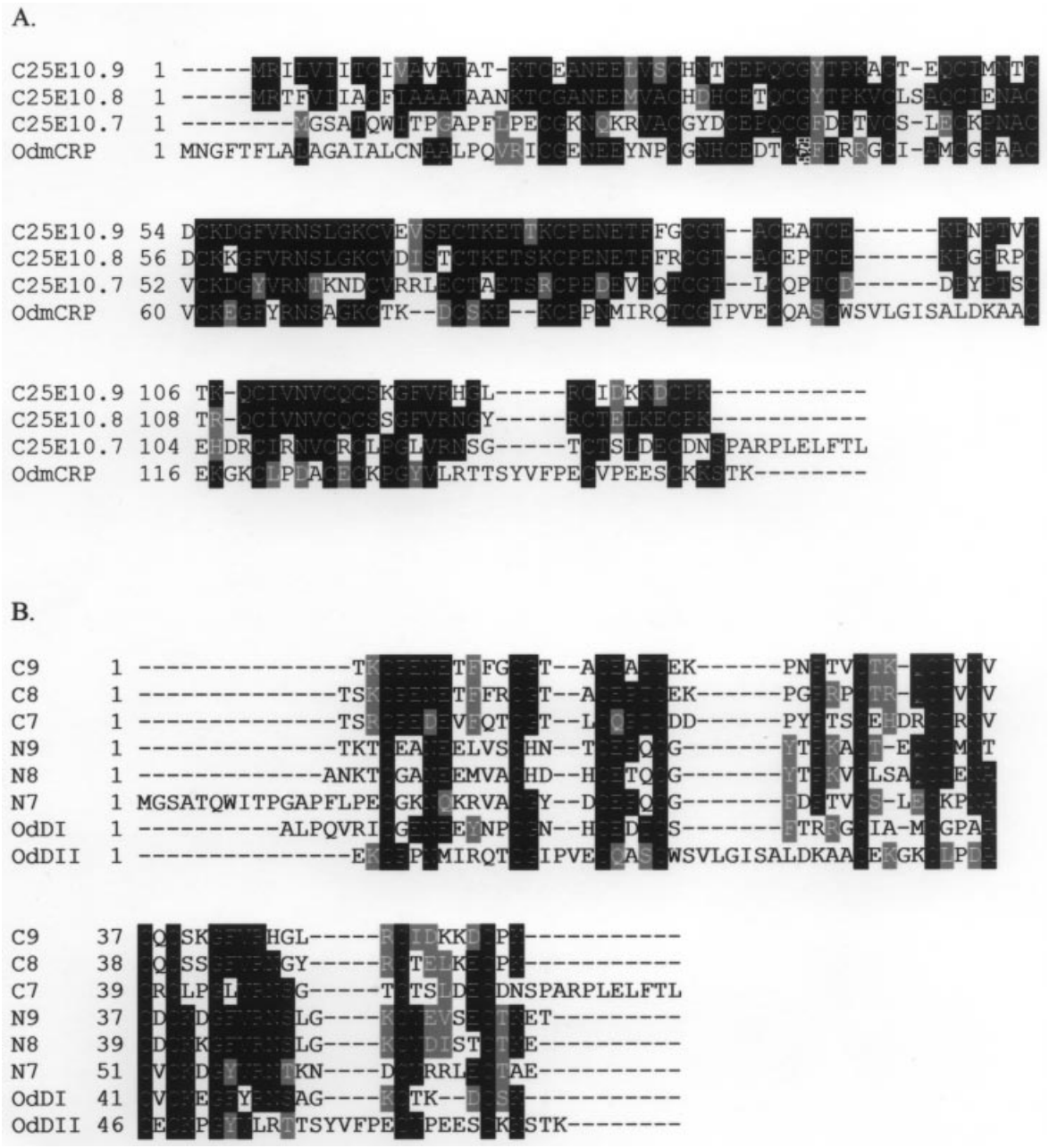

Fig. 2. For legend see overleaf.

presence of a protein of ~ 21 kDa, which is in agreement with the calculated molecular weight of the mature protein plus the vector-derived residues. Western blot analysis of PBS-soluble protein extracts from L4s or adults probed with anti-rOdmCRP antibodies detected a single, dominant band of ~ 15 kDa, consistent with the expected size of the mature protein, in adults but not in L4s (Fig. 3). Normal rabbit (pre-bleed) serum did not bind to parasite extracts or the recombinant protein. Thus, it appears that the mature protein is not cleaved into single domains *in vivo*, but is maintained in the 2-domain form. Low-level expression of the *Odmcrp* transcript has been detected in the L4 stage (Boag *et al*. 2000), but the accumulation of protein may be below the level detectable by Western blot.

Three-dimensional structure prediction for OdmCRP

The 3-dimensional protein structures of several members of the *Ascaris* family of tight-binding SSPIs have been determined (Grasberger, Clore & Gronenborn, 1994; Huang *et al*. 1994; Cierpicki *et al*. 2000). Using this information, a homology-based structural model was developed for OdmCRP, with domain I modelled on the structure (1CCV) of the *A. mellifera* chymotrypsin/cathepsin G inhibitor 1 (AMCI-1; 41 % identity and 51 % similarity over aligned segments) with a root mean square deviation of 0·73 Å over 47/63 aligned Cα positions (Cierpicki *et al*. 2000). Domain II was modelled on the *A. suum* chymotrypsin/elastase inhibitor pairs (1EAI; 27 % identity and 38 % similarity over aligned segments)

https://doi.org/10.1017/S0031182002002329 Published online by Cambridge University Press

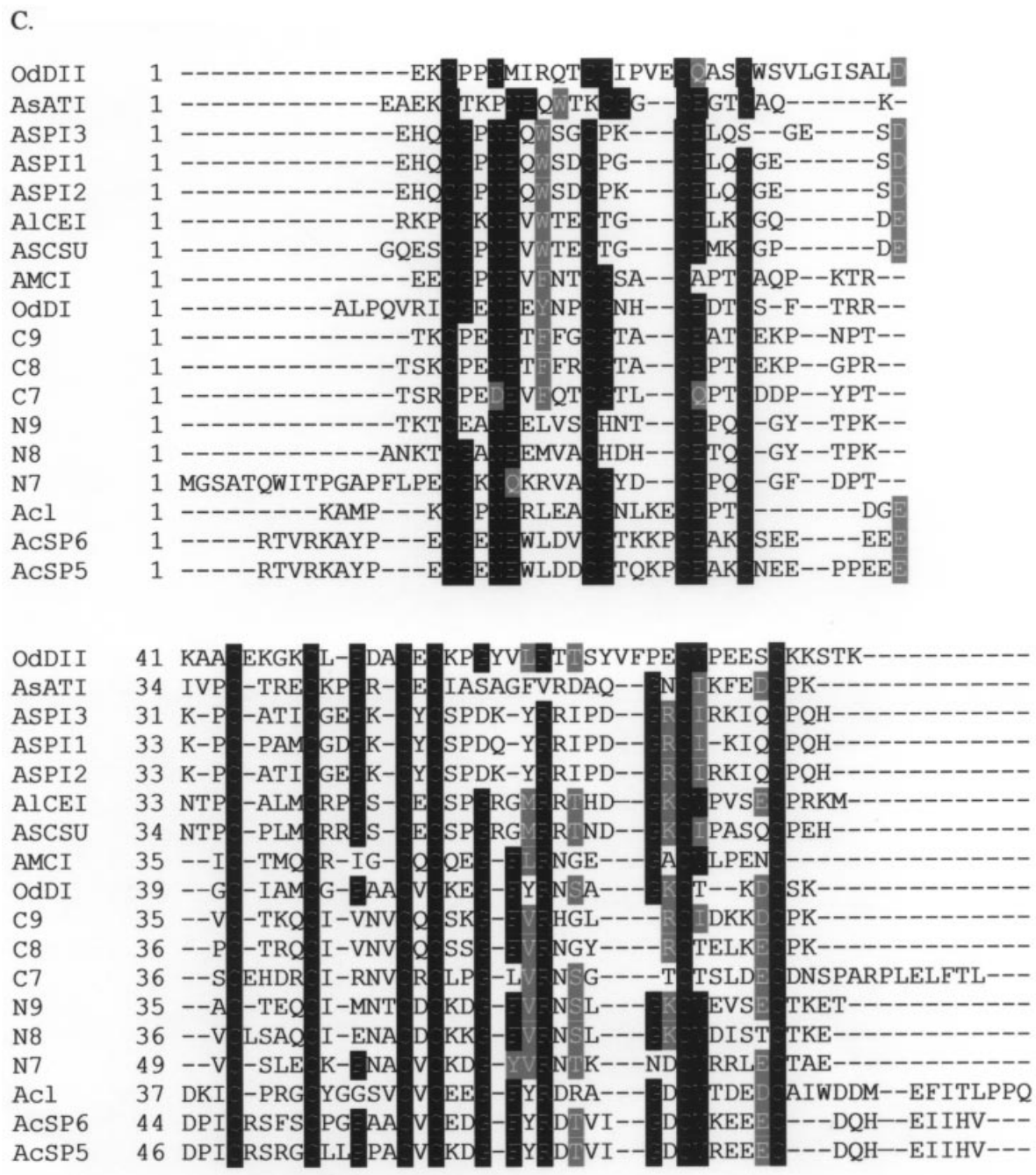

Fig. 2. Alignment of the predicted amino acid sequence of OdmCRP with proteins from other nematodes. Identical residues are shaded in black, non-radical substitutions are shaded grey. (A) OdmCRP aligned with 3 *Caenorhabditis elegans* predicted proteins (C25E10.7, C25E10.8 and C25E10.9). (B) Alignment of domains I and II of Od-mCRP (OdDI and OdDII, respectively) with the N- and C-terminal domains of the 3 *C. elegans* predicted proteins C25E10.7, C25E10.8 and C25E10.9 (N7-N9 and C7-C9, respectively). (C) Alignment of domains I and II of Od-mCRP with the *Ascaris suum* trypsin inhibitor (AsATI), *Anisakis simplex* serine protease inhibitors (ASPI3, ASPI1 and ASPI2), *Ascaris lumbricoides* chymotrypsin/elastase inhibitor (ALCEI), *Ascaris suum* chymotrypsin/elastase inhibitor 1 (ASCSU), *Apis mellifera* (honeybee) chymotrypsin inhibitor (AMCI), *Ancylostoma ceylanicum Ascaris*-type serine protease inhibitor (Acl), and *A. caninum* anti-coagulant precursors (AcSP6 and AcASP5) (Accession nos. P19398, AAC61299, AAC61297, AAC61298, S08572, P07851, P56682, AAD51336, AAC47081 and AAC47082, respectively). All alignments were carried out using CLUSTAL W, and shading was applied using BOXSHADE.

with a root mean square deviation of 0·81 Å over 50/75 aligned Cα positions (Huang *et al*. 1994). Despite low overall amino acid similarity between these SSPIs, each domain of the predicted *O. dentatum* protein was inferred to form the 5 disulfide bridges which provide the majority of the structural stability for this family of inhibitors (Fig. 4A) (Grasberger *et al*. 1994). The exposed reactive loop of OdmCRP domain I has 3 hydrophobic amino acid residues flanked by cysteine residues (CIAMC),

https://doi.org/10.1017/S0031182002002329 Published online by Cambridge University Press

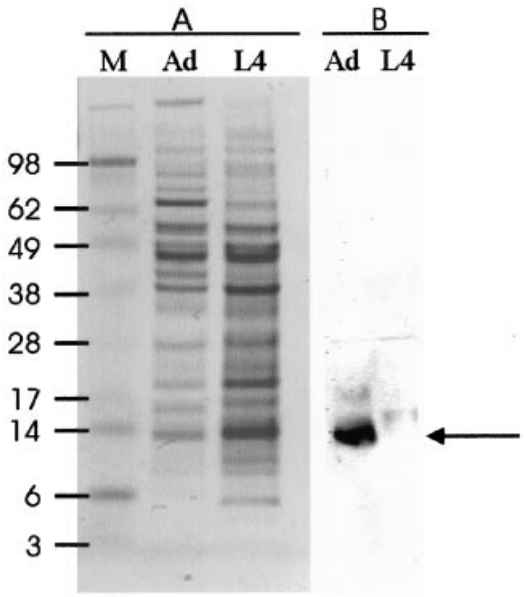

Fig. 3. SDS–PAGE and Western blot analyses of 4th-stage larvae and adults of *Oesophagostomum dentatum* males. (A) Stained SDS–PAGE gel. (B) Western blot probed with affinity-purified antibody to protein rOdmCRP. The presence of the 'native' parasite OdmCRP at ~ 15 kDa is indicated by an arrow. (M) Molecular weight markers (kDa); (Ad) adult male PBS extract; (L4) 4th-stage male larval PBS extract.

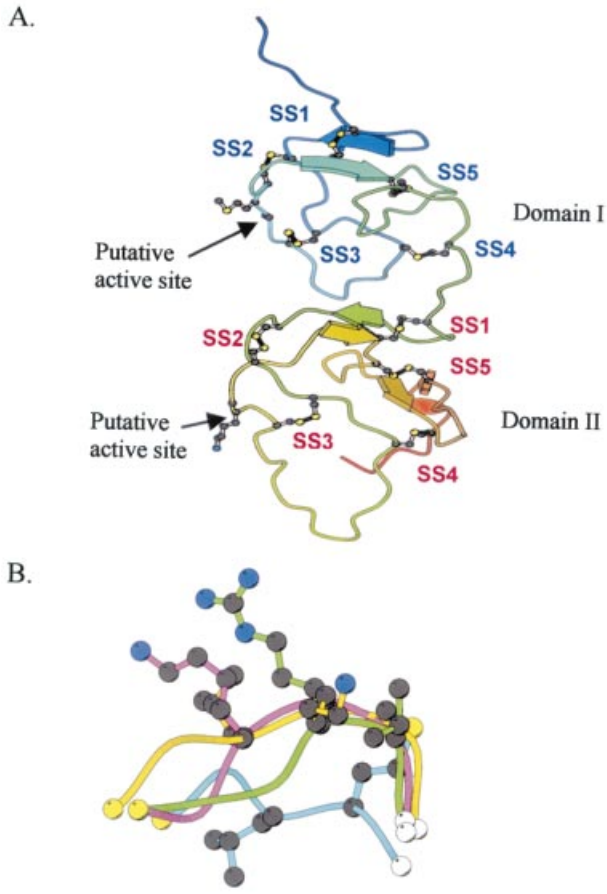

Fig. 4. Three-dimensional structure of OdmCRP predicted by modelling, shown as a MOLSCRIPT cartoon. (A) The N-terminus is blue and the colour spectrum changes through to red at the C-terminus. Heavy atoms of the disulfide bridges (labelled SS1–SS5; domain I in blue, and domain II in red) and the side-chains of the residues forming the putative active sites are shown in ball-and-stick representation, coloured by atom (oxygen in red; nitrogen in blue; sulfur in yellow; carbon in white). Arrows represent β-strands. (B) Overlay of the reactive loop of OdmCRP-C (magenta) with those of *Ascaris suum* (cyan), *Ancylostoma caninum* (green) and *Locusta migratoria* (gold), shown as a MOLSCRIPT Cα trace. The cysteine residues at the start and end of each loop are indicated by yellow and white spheres. Heavy atoms of the side-chains of residues forming the active sites of *A. suum*, *A. caninum* and *L. migratoria*, and the predicted active site of OdmCRP are shown in a ball-and-stick representation, coloured by atom (nitrogen in blue; sulphur in yellow; carbon in black).

which is consistent with the inhibitors of chymotrypsin/elastase from *A. suum* (CPLMC and CALMC), and of elastase from *A. simplex* (CPAMC). Other SSPIs also have 3 residues in their reactive sites (i.e. BSTI and AMCI-1), but the amino acid characteristics (charge and hydrophobicity) are variable (Table 1). Comparison of the structure of the *A. mellifera* chymotrypsin/cathepsin G inhibitor with the OdmCRP domain I model indicated that the side-chain groups of the reactive loop have a similar orientation, and may interact directly with the cognate serine protease (data not shown). Based on comparison of the reactive sites of other SSPIs with characterized reactive bonds (Grasberger *et al.* 1994; Huang *et al.* 1994; Stassens *et al.* 1996), the scissile bond of this domain is predicted to be between the alanine and methionine residues (CIA/MC) (Table 1). The predicted reactive loop of OdmCRP domain II contains 4 amino acids flanked by cysteine residues (CEKGKC), the same number as for the *A. caninum* anticoagulant proteins (Stassens *et al.* 1996). Database searches for other serine inhibitors with 4 amino acids in their reactive loop identified a chymotrypsin/elastase inhibitor from the locust, *Locusta migratoria* (LCLI-I: CTRKGC and LCLI-II: CTLKAC; Boigegrain *et al.* 1992). Interestingly, the scissile bonds differ between the hookworm and locust inhibitors, with CXXX/XC and CXX/XXC being present, respectively. Overlaying models of the reactive loop of domain II with the locust LCLI-II: NMR structure (Mer *et al.* 1996) (0·61 Å over 5/6 loop Cα positions) and hookworm AcAPc2 inhibitors (0·73 Å over 4/6 loop Cα positions) (Fig. 4B) suggests that there is greater similarity between domain II and the locust inhibitor, with 3 consecutive residues overlapping, including the reactive bond L/K. Overlap between domain II and AcAPc2 is restricted to 2 non-consecutive residues and does not include the reactive bond R/V, which may indicate that the domain II scissile bond of OdmCRP may be shifted

https://doi.org/10.1017/S0031182002002329 Published online by Cambridge University Press

Table 1. Reactive site of OdmCRP compared with selected SSPIs and *Caenorhabditis elegans* predicted proteins

(Experimentally determined scissile bond positions are indicated by /. Proteins containing 2 domains are separated into individual domains and are labelled 'N' for amino-terminal end, and 'C' for carboxyl-terminal end.)

Protein	Reactive loop	Target serine protease	Species
C/E 1	C P L/M C	Chymotrypsin/elastase	*Ascaris suum*
C/E 4	C A L/M C	Chymotrypsin/elastase	*A. suum*
AMCI-1	C T M/Q C	Chymotrypsin/cathepsin G	*Apis mellifera*
ASPI-1	C P A/M C	Elastase	*Anisakis simplex*
ASPI-2	C A T I C	Unknown	*A. simplex*
ATI	C T R/E C	Trypsin	*A. lumbricoides*
BSTI	C D K K C	Trypsin/thrombin	*Bombina bombina*
TSTCI	C T R Q C	Trypsin	*Trichuris suis*
AF172653.1	C P R G C	Unknown	*Ancylostoma ceylanicum*
OdmCRP-N	C I A M C	Unknown	*Oesophagostomum dentatum*
C25E10.9-C	C T K Q C	Unknown	*Caenorhabditis elegans*
C25E10.9-N	C T E Q C	Unknown	*C. elegans*
C25E10.8-C	C T R Q C	Unknown	*C. elegans*
C25E10.7-N	C S L E C	Unknown	*C. elegans*
AcAP5	C R S R/G C	Factor Xa	*A. caninum*
AcAP6	C R S F/S C	Factor Xa	*A. caninum*
AcAPc2	C L V R/V C	Factor VIIA/tissue factor	*A. caninum*
OdmCRP-C	C E K G K C	Unknown	*O. dentatum*
C25E10.8-N	C L S A Q C	Unknown	*C. elegans*
C25E10.7-C	C E H D R C	Unknown	*C. elegans*
LCLI-I	C T R K G C	Chymotrypsin, elastase	*Locusta migratoria*
LCLI-II	C T L/K A C	Chymotrypsin, elastase	*L. migratoria*

by 1 residue from CEKG/KC to CEK/GKC, compared with the hookworm inhibitors. An alternative possibility is that the orientation of the binding loop is different in the enzyme-bound state, as seen from the highly flexible canonical loop structure in the case of the anticoagulants from species of *Ancylostoma*.

DISCUSSION

A novel male-specifically expressed cDNA of *O. dentatum* containing an EGF-like cysteine pattern was isolated, sequenced and the predicted encoded protein characterized. This protein contained a putative 18 amino acid signal sequence, resulting in a mature peptide of 14·7 kDa with 2 TIL domains (Pfam: PF01826), 1 containing an EGF-like cysteine pattern, and the second an EGF-like cysteine pattern with aberrant spacing for the final cysteine in the motif. The mature peptide had similarity to several *C. elegans* predicted proteins which contain the 2 TIL domain/EGF-like cysteine motifs and to several SSPIs of the *Ascaris* family, all of which contain a single TIL domain.

The TIL motif occurs in 24 different putative sequences of *C. elegans*, in either 1-, 2- or 3-domain forms (Costanzo *et al*. 2001). Based on Southern blot analysis, it appears to be part of a multi-gene family in *O. dentatum* (data not shown), and is also present in ESTs from a range of other nematodes, in both 1- and 2-domain forms. Their apparent wide distribution in nematodes and the variation in their domain structures suggest that proteins containing the TIL motif are associated with a range of biological pathways involving interactions with specific serine proteases. Gene-silencing experiments using double-stranded RNA interference (RNAi) were conducted in *C. elegans*, targetting the predicted protein CE2510.7, but no phenotypic effect could be observed (data not shown). Other RNAi experiments have also failed to produce observable phenotypes for other genes containing the TIL domain (Fraser *et al*. 2000; Gonczy *et al*. 2000), suggesting that they are either non-essential for nematode survival, or that there is redundancy in their function.

The SSPIs of the *Ascaris* family have been isolated from several other nematode species (Nguyen *et al*. 1999) as well as from the honeybee, *A. mellifera* (see Bania, Stachowiak & Polanowski, 1999) and the frog, *Bombina bombina* (see Mignogna *et al*. 1996). These proteins range in size from 56 to 65 amino acids and have limited overall sequence conservation, with 30–40% identity between family members. However, the relative positions of the 10 cysteine residues are strictly conserved. The 3-dimensional structures have been solved by X-ray crystallography and nuclear magnetic resonance spectroscopy (NMR) for several of the *Ascaris* inhibitors (Grasberger *et al*. 1994; Huang *et al*. 1994; Cierpicki *et al*. 2000). The 10 cysteine residues found in the *Ascaris* family of inhibitors are all involved in the 5 disulfide bridges

(cysteines 1–7, 2–6, 3–5, 4–10 and 8–9) which define them. These disulfide bridges are primary contributors to the 3-dimensional structure of the protein, as there is limited secondary structure. The *Ancylostoma* anticoagulants are slightly larger in size than the *Ascaris* inhibitors, but all contain the 10 cysteine residues in a spatially conserved pattern, and the reactive site of the inhibitors has been demonstrated to be in the same relative position as in other inhibitors within this family (Stassens *et al*. 1996). Structural predictions for protein OdmCRP were made with reference to the *A. mellifera* chymotrypsin inhibitor (AMCI-I) and the *A. suum* chymotrypsin/elastase inhibitor (1EAI) as templates, and indicate that both domains form 5 disulfide bonds and have active sites with similar structural features to other SSPIs. The domain I putative reactive loop contains 3 hydrophobic residues, which is consistent with chymotrypsin and elastase inhibitors of *A. suum* and *A. simplex*. Additionally, the position of the amino acid sidegroups in the reactive loop is similar with respect to the honey bee AMCI-I, which may suggest a similar substrate for this domain. The domain II putative active loop contained 4 residues between cysteines, the same number as in the *A. caninum* anticoagulants and the locust *L. migratoria* SSPI (Mer *et al*. 1996), although, interestingly, the residues are not highly conserved in terms of amino acid characteristics. Although overall similarity is greater between domain II and the hookworm inhibitors, the reactive loop is predicted to have greater similarity to the locust SSPI, which suggests that the reactive bond is CEK/GKC. The structural model predicts that the reactive loops of the 2 TIL domains of OdmCRP may be capable of independent functions, such that the same molecule could play a dual inhibitory role. Dual function has been observed in the 2-domain, 4-disulfide core protein human mucous proteinase inhibitor (SLPI), which has both anti-tryptic and anti-chymotryptic/elastase inhibitory activity (Grütter *et al*. 1988).

Inhibition assays using rOdmCRP and porcine pancreatic elastase, bovine pancreatic chymotrypsin and trypsin did not demonstrate any inhibitory activity. This may reflect a highly specific interaction between the reactive loop of the inhibitor and its target protease, or may suggest that the complex 2-domain structure may not be folded correctly when expressed in *E. coli*. Expression of either the 2-domain OdmCRP or each single domain in insect cells using recombinant baculovirus produced insoluble and, therefore, incorrectly folded protein (data not shown). Recently, 2 putative elastase inhibitors (ASPI-1 and ASPI-2) of *A. simplex* were expressed as recombinant proteins in *Pichia pastoris* (see Nguyen *et al*. 1999). The 2 proteins differ in amino acid sequence within their reactive loops, while the remaining residues are almost identical (96·5%). However, only 1 recombinant protein displayed inhibitory activity, which emphasizes the specificity of the reactive loop to the inhibitory activity.

Interestingly, the *O. dentatum* OdmCRP is most similar to 2 (C25E10.7 and C25E10.8) of the 3 *C. elegans* predicted proteins which share the 2-domain structure, and have predicted reactive loops with 3 and 4 amino acids flanked by cysteine residues. The significance of differing amino acid numbers in the reactive loop is currently unknown, but it is possible that the 2 domains have different substrate specificities, perhaps allowing the proteins to interact with a range of different substrates or regulatory molecules.

The SSPIs characterized previously from parasitic nematodes, which all have a single inhibitory domain, are thought to relate to feeding (Stassens *et al*. 1996), to protection from proteolytic attack in the host (Hawley & Peanasky, 1992), or to modulation of the host's immune system (Rhoads *et al*. 2000). For example, the anticoagulants of *Ancylostoma* species (Cappello *et al*. 1995, 1996; Stassens *et al*. 1996) are potent inhibitors of factor Xa and factor VIIa, both of which are key enzymes involved in blood coagulation. Hence, the ability of these molecules to inhibit blood coagulation is thought to aid in the blood feeding activity of hookworms (Cappello *et al*. 1993). However, the predicted peptide OdmCRP and the 3 predicted *C. elegans* proteins, C25E10.7, C25E10.8 and C25E10.10, all have 2 TIL domains. Expression of these genes in *C. elegans* appears to be relatively low, as no corresponding ESTs have yet been identified. However, genome-wide expression profiling of *C. elegans* has recently revealed that these genes have greater expression levels in males than in hermaphrodites and, for C25E10.10, expression is more abundant in the germline of hermaphrodites producing sperm than those producing oocytes (Reinke *et al*. 2000). This is consistent with expression of OdmCRP exclusively in male *O. dentatum*.

Serine proteases occur in a wide range of metazoans and function in biological processes, such as protein processing and activation of pro-hormones and zymogens (Otlewski, Krowarsch & Apostoluk, 1999). Although vital, these proteinases are potentially harmful (to nematodes) if their proteolytic activity is not tightly controlled by regulatory mechanisms, including the availability of specific inhibitors (Otlewski *et al*. 1999). It seems unlikely that a gender-specific SSPI is involved in protection from 'host attack', or aids in feeding, because both male and female *O. dentatum* reside in the same proteolytic environment and would be expected to feed on the same food source. It is more likely that OdmCRP is involved in interacting with endogenous serine proteases associated specifically with biological processes in the male worm, possibly relating

to reproductive maturation or behaviour. For example, a role for proteases in spermatid development has been demonstrated *in vitro* for both *C. elegans* and *A. suum*, where the addition of a protease solution is sufficient to initiate the terminal differentiation of spermatids to spermatazoa (Ward, Hogan & Nelson, 1983; Sepsenwol & Taft, 1990). Therefore, it is possible that the activity of an endogenous protease in the spermatid is modulated by interaction with a specific inhibitor.

It appears that nematodes contain many serine protease inhibitors of both the 1- and 2-domain forms of TILs, although their 'direct' ligands remain to be identified. Efforts should focus on identifying new members of the 2-domain serine protease inhibitors in *O. dentatum*, identifying which tissues contain these proteins and determining with which proteins they interact.

The technical assistance of the staff, particularly Michael Larsen and Niels-Peter Hansen, of the Danish Centre of Experimental Parasitology and the Victorian Institute of Animal Science is gratefully acknowledged. We thank the late Peter Nansen for his support during the initial stages of this research. We are also grateful to Mark Sandeman and Hua Ren (School of Agriculture, La Trobe University) for valuable discussion on protease inhibition assays, and Vern Bowles and Nunzio Mancuso (Center for Animal Biotechnology, The University of Melbourne) for discussions about protein purification. The pigs and rabbits used in this study were treated in accordance with animal ethics laws of Denmark (experimental animal permit no. 1994-101-115), and the Australian Code of Practice for the Care and Use of Animals for Scientific Purposes. P.R.B. was supported by a joint University of Melbourne/Novartis Animal Health Australia scholarship. Project support was provided through the Australian Research Council (grant no. A00102973), Novartis Animal Health Australasia, the Danish National Science Foundation, Agricultural Victoria and the Department of Natural Resources and Environment of Victoria.

REFERENCES

ALTSCHUL, S. F., MADDEN, T. L., SCHAFFER, A. A., ZHANG, J., ZHANG, Z., MILLER, W. & LIPMAN, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**, 3389–3402.

BAIROCH, A. (1993). The PROSITE dictionary of sites and patterns in proteins, its current status. *Nucleic Acids Research* **21**, 3097–3103.

BANIA, J., STACHOWIAK, D. & POLANOWSKI, A. (1999). Primary structure and properties of the cathepsin G/chymotrypsin inhibitor from the larval hemolymph of *Apis mellifera*. *European Journal of Biochemistry* **262**, 680–687.

BATEMAN, A., BIRNEY, E., DURBIN, R., EDDY, S. R., HOWE, K. L. & SONNHAMMER, E. L. (2000). The Pfam protein families database. *Nucleic Acids Research* **28**, 263–266.

BLAXTER, M. & LIU, L. (1996). Nematode spliced leaders-ubiquity, evolution and utility. *International Journal for Parasitology* **26**, 1025–1033.

BOAG, P. R., NEWTON, S. E., HANSEN, N.-P., CHRISTENSEN, C. M., NANSEN, P. & GASSER, R. B. (2000). Isolation and characterisation of sex-specific transcripts from *Oesophagostomum dentatum* by RNA arbitrarily-primed PCR. *Molecular and Biochemical Parasitology* **108**, 217–224.

BOIGEGRAIN, R. A., MATTRAS, H., BREHELIN, M., PAROUTAUD, P. & COLETTI-PREVIERO, M. A. (1992). Insect immunity: two proteinase inhibitors from hemolymph of *Locusta migratoria*. *Biochemical and Biophysical Research Communications* **189**, 790–793.

CAPPELLO, M., CLYNE, L. P., MCPHEDRAN, P. & HOTEZ, P. J. (1993). *Ancylostoma* factor Xa inhibitor: partial purification and its identification as a major hookworm-derived anticoagulant *in vitro*. *Journal of Infectious Diseases* **167**, 1474–1477.

CAPPELLO, M., HAWDON, J. M., JONES, B. F., KENNEDY, W. P. & HOTEZ, P. J. (1996). *Ancylostoma caninum* anticoagulant peptide: cloning by PCR and expression of soluble, active protein in *E. coli*. *Molecular and Biochemical Parasitology* **80**, 113–117.

CAPPELLO, M., VLASUK, G. P., BERGUM, P. W., HUANG, S. & HOTEZ, P. J. (1995). *Ancylostoma caninum* anticoagulant peptide: a hookworm-derived inhibitor of human coagulation factor Xa. *Proceedings of the National Academy of Sciences, USA* **92**, 6152–6156.

CHRISTENSEN, C. M. (1997). The effect of three distinct sex ratios at two *Oesophagostomum dentatum* worm population densities. *Journal of Parasitology* **83**, 636–640.

CHRISTENSEN, C. M., BARNES, E. H., NANSEN, P. & GRØNDAHL-NIELSEN, C. (1996). Growth and fecundity of *Oesophagostomum dentatum* in high-level infections and after transplantation into naive pigs. *Parasitology Research* **82**, 364–368.

CHRISTENSEN, C. M., GRØNDAHL-NIELSEN, C. & NANSEN, P. (1996). Non-surgical transplantation of *Oesophagostomum dentatum* to recipient pigs via rectal intubation. *Veterinary Parasitology* **65**, 139–145.

CHURCH, G. M. & GILBERT, W. (1984). Genomic sequencing. *Proceedings of the National Academy of Sciences, USA* **81**, 1991–1995.

CIERPICKI, T., BANIA, J. & OTLEWSKI, J. (2000). NMR solution structure of *Apis mellifera* chymotrypsin/cathepsin G inhibitor-1 (AMCI-1): structural similarity with *Ascaris* protease inhibitors. *Protein Science* **9**, 976–984.

COSTANZO, M. C., CRAWFORD, M. E., HIRSCHMAN, J. E., KRANZ, J. E., OLSEN, P., ROBERTSON, L. S., SKRZYPEK, M. S., BRAUN, B. R., HOPKINS, K. L., KONDU, P., LENGIEZA, C., LEW-SMITH, J. E., TILLBERG, M. & GARRELS, J. I. (2001). YPD, PombePD and WormPD: model organism volumes of the BioKnowledge library, an integrated resource for protein information. *Nucleic Acids Research* **29**, 75–79.

FRASER, A. G., KAMATH, R. S., ZIPPERLEN, P., MARTINEZ-CAMPOS, M., SOHRMANN, M. & AHRINGER, J. (2000). Functional genomic analysis of *C. elegans* chromosome I by systematic RNA interference. *Nature, London* **408**, 325–330.

FROHMAN, M. A., DUSH, M. K. & MARTIN, G. R. (1988). Rapid production of full-length cDNAs from rate transcripts: amplification using a single gene-specific oligonucleotide primer. *Proceedings of the National Academy of Sciences, USA* **85**, 8998–9002.

GONCZY, P., ECHEVERRI, G., OEGEMA, K., COULSON, A., JONES, S. J., COPLEY, R. R., DUPERON, J., OEGEMA, J., BREHM, M., CASSIN, E., HANNAK, E., KIRKHAM, M., PICHLER, S., FLOHRS, K., GOESSEN, A., LEIDEL, S., ALLEAUME, A. M., MARTIN, C., OZLU, N., BORK, P. & HYMAN, A. A. (2000). Functional genomic analysis of cell division in *C. elegans* using RNAi of genes on chromosome III. *Nature*, *London* **408**, 331–336.

GRASBERGER, B. L., CLORE, G. M. & GRONENBORN, A. M. (1994). High-resolution structure of *Ascaris* trypsin inhibitor in solution: direct evidence for a pH-induced conformational transition in the reactive site. *Structure* **2**, 669–678.

GRÜTTER, M. G., FENDRICH, G., HUBER, R. & BODE, W. (1988). The 2·5 Å X-ray crystal structure of the acid-stable proteinase inhibitor from human mucous secretions analysed in its complex with bovine alpha-chymotrypsin. *EMBO Journal* **7**, 345–351.

HARLOW, E. & LANE, D. (1998). *Antibodies : A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

HAWLEY, J. H. & PEANASKY, J. R. (1992). *Ascaris suum*: are trypsin inhibitors involved in species specificity of *ascarid* nematodes? *Experimental Parasitology* **75**, 112–118.

HUANG, K., STRYNADKA, N. C., BERNARD, V. D., PEANASKY, R. J. & JAMES, M. N. (1994). The molecular structure of the complex of *Ascaris* chymotrypsin/elastase inhibitor with porcine elastase. *Structure* **2**, 679–689.

JACKSON, F. & COOP, R. L. (2000). The development of anthelmintic resistance in sheep nematodes. *Parasitology* **120**, S95–S107.

KRAULIS, P. J. (1991). MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *Journal of Applied Crystallography* **24**, 946–950.

MER, G., HIETTER, H., KELLENBERGER, C., RENATUS, M., LUU, B. & LEFEVRE, J. F. (1996). Solution structure of PMP-C: a new fold in the group of small serine proteinase inhibitors. *Journal of Molecular Biology* **258**, 158–171.

MIGNOGNA, G., PASCARELLA, S., WECHSELBERGER, C., HINTERLEITNER, C., MOLLAY, C., AMICONI, G., BARRA, D. & KREIL, G. (1996). BSTI, a trypsin inhibitor from skin secretions of *Bombina bombina* related to protease inhibitors of nematodes. *Protein Science* **5**, 357–362.

NGUYEN, T. T., QASIM, M. A., MORRIS, S., LU, C. C., HILL, D., LASKOWSKI, M. JR. & SAKANARI, J. A. (1999). Expression and characterization of elastase inhibitors from the ascarid nematodes *Anisakis simplex* and *Ascaris suum*. *Molecular and Biochemical Parasitology* **102**, 79–89.

NIELSEN, H., ENGELBRECHT, J., BRUNAK, S. & VON HEIJNE, G. (1997). Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Engineering* **10**, 1–6.

OTLEWSKI, J., KROWARSCH, D. & APOSTOLUK, W. (1999). Protein inhibitors of serine proteinases. *Acta Biochimica Polonica* **46**, 531–565.

PEANASKY, R. J., BENTZ, Y., HOMANDBERG, G. A., MINOR, S. T. & BABIN, D. R. (1984). The isoinhibitors of chymotrypsin/elastase from *Ascaris lumbricoides*: the reactive site. *Archives of Biochemistry and Biophysics* **232**, 135–142.

REINKE, V., SMITH, H. E., NANCE, J., WANG, J., VAN DOREN, C., BEGLEY, R., JONES, S. J., DAVIS, E. B., SCHERER, S., WARD, S. & KIM, S. K. (2000). A global profile of germline gene expression in *C. elegans*. *Molecular Cell* **6**, 605–616.

RHOADS, M. L., FETTERER, R. H., HILL, D. E. & URBAN, J. F. JR. (2000). *Trichuris suis*: a secretory chymotrypsin/elastase inhibitor with potential as an immunomodulator. *Experimental Parasitology* **95**, 36–44.

SALI, A. & BLUNDELL, T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *Journal of Molecular Biology* **234**, 779–815.

SEPSENWOL, S. & TAFT, S. J. (1990). In vitro induction of crawling in the amoeboid sperm of the nematode parasite, *Ascaris suum*. *Cell Motility and the Cytoskeleton* **15**, 99–110.

STASSENS, P., BERGUM, P. W., GANSEMANS, Y., JESPERS, L., LAROCHE, Y., HUANG, S., MAKI, S., MESSENS, J., LAUWEREYS, M., CAPPELLO, M., HOTEZ, P. J., LASTERS, I. & VLASUK, G. P. (1996). Anticoagulant repertoire of the hookworm *Ancylostoma caninum*. *Proceedings of the National Academy of Sciences*, *USA* **93**, 2149–2154.

TALVIK, H., CHRISTENSEN, C. M., JOACHIM, A., ROEPSTORFF, A., BJØRN, H. & NANSEN, P. (1997). Prepatent periods of different *Oesophagostomum* spp. isolates in experimentally infected pigs. *Parasitology Research* **83**, 563–568.

THOMPSON, J. D., HIGGINS, D. G. & GIBSON, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**, 4673–4680.

WARD, S., HOGAN, E. & NELSON, G. A. (1983). The initiation of spermatogenesis in the nematode *Caenorhabditis elegans*. *Developmental Biology* **98**, 70–79.