Identification of a cDNA encoding an active asparaginyl endopeptidase of Schistosoma mansoni and its expression in Pichia pastoris

Conor R. Caffrey, Mary A. Mathieu, Alan M. Gaffney, Jason P. Salter, Mohammed Sajid, Kimberley D. Lucas, Christopher Franklin, Matthew Bogyo, James H. McKeever

Abstract
Asparaginyl endopeptidases, or legumains, are a recently identified family of cysteine-class endopeptidases. A single gene encoding a Schistosoma mansoni asparaginyl endopeptidase (a.k.a. Sm32 or schistosome legumain) has been reported, but by sequence homology it would be expected to yield an inactive product as the active site C197 had been replaced by N. We now describe a new S. mansoni gene in which C197 is present. Both gene products were expressed in Pichia pastoris. Autocatalytic processing to fully active C197 Sm32 occurred at acid pH. In contrast, N197 Sm32 was not processed and this is consistent with the hypothesis that C197 is essential for catalysis. This was confirmed by mutation of N197 to C and re-expression in Pichia. The availability of recombinant active Sm32 allows detailed analysis of its catalytic mechanism and its function(s) in the biology of this important human parasite.

© 2000 Federation of European Biochemical Societies.

Key words: Asparaginyl endopeptidase; Legumain; Sm32; Recombinant expression; Schistosoma

1. Introduction
Asparaginyl endopeptidases or legumains are a novel family of cysteine-class endopeptidases with members reported in plants [1], mammals [2] and in the parasitic blood fluke Schistosoma spp. [3]. Asparaginyl endopeptidases are proposed to belong to the CD clan of cysteine endopeptidases along with caspases, clostripain and gingipains [4]. They have a specificity for an asparaginyl residue in the P1 position of the substrate (terminology of Schechter and Berger [5]) for hydrolysis and possess an approximately 15 kDa extension C-terminal of the catalytic domain [1–3]. In plants, legumains process seed proteins and C-terminal regions of the endopeptidase: forward: 5′-ATACCT-GAGAAAAGACATGATACAAATTATGAAGTATCC-3′ and reverse: 5′-ATCGGCGCGCTTAACCGCAAATTTTTATGATT-GCTCT-3′. For eventual cloning into the P. pastoris expression vector pPIC Zα A, the forward primer incorporated Xhol endonuclease and Kex 2 endopeptidase sites (both underlined) immediately 5′ of the gene-specific sequence. The reverse primer incorporated a transcription termination codon and a NorI site (both underlined) immediately 3′ of the gene-specific sequence. Amplification reactions were performed with Pwo polymerase (Boehringer) with 35 cycles of 94°C, 55°C and 72°C, each for 1 min. PCR products were purified (Qiagen) and digested with Xhol and NorI. pPIC Zα A was similarly digested and ligated to the PCR products. For sequencing, the same PCR products were amplified using the above PCR conditions and cloned into pCR-Blunt II-TOPO according to the manufacturer’s (Invitrogen) instructions. A unique ApH III restriction site was found for C197 Sm32 (see below) and was used to distinguish such clones from those of N197 Sm32. The signal peptide sequence of C197 Sm32 was obtained from individual Blunt II-TOPO clones that had been generated using a forward primer directed to the signal peptide (5′-ATGATGC-TGTGAATCAGG-3′) and the above-described reverse primer. PCR conditions to generate such clones were as described above. The entire open reading frame (ORF) of C197 Sm32 was sequenced in both directions. To generate the N197C mutant Sm32, site-directed mutagenesis was performed by sequential PCR steps [13] using the forward primer 5′-TTATATTGGAACGATTGTAATCAGG-3′ (bases encoding the mutation are underlined) and the appropriate anti-paralle reverse primer. This mutagenesis resulted in the same ApH III site as found in C197 Sm32. All recombinant plasmids were propagated in DH5α cells (Life Technologies) and purified (Qiagen).

*Corresponding author. Fax: (1)-415-750 6947. E-mail: jmck@cgl.ucsf.edu

1 The EMBL Nucleotide Sequence Database accession number for C197 Sm32 is AJ250582.

Received 17 November 1999; received in revised form 21 December 1999

Edited by Takashi Gojobori
2.2. Expression of Sm32 in *P. pastoris*

Procedures to express Sm32 in the X33 strain of *P. pastoris* were as detailed by the manufacturer (Invitrogen). Briefly, recombinant pPIC Zα A (10 µg) was linearized with SacI and purified. *P. pastoris* was electroporated at 1.5 kV and 129 µF in electroporation cuvettes (2 mm gap; BTX). *Pichia* colonies growing under zeocin (Invitrogen) selection were picked for expansion first in 10 ml of yeast extract-peptone-dextrose medium containing 100 µg/ml zeocin and then in 200–400 ml of the same medium. Expression of recombinant protein was induced by incubation of *Pichia* for 48 h in buffered-minimal medium containing 1% methanol as the sole carbon source. Culture medium was clarified through a 0.45 µm filter, lyophilized and stored at 4°C.

![Fig. 1. Nucleotide and deduced amino acid sequence of C197 Sm32. Active site histidine and cysteine residues are presented in bold; potential glycosylation sites are underlined and arrows [14] indicate the putative N-termini of the pro-domain, mature protease and C-terminal extension.](image-url)
Lyophilized culture media containing recombinant Sm32 were resuspended in 0.1 M citrate-phosphate, pH 6.8 to a volume one-tenth of that of the original. Complete buffer exchange was accomplished using PD10 desalting columns (Pharmacia). Solutions were concentrated 10 times in Centricon 10 centrifugation units (Millipore) and stored at \(-20^\circ\)C.

2.3. Detection and activation of recombinant Sm32, and endopeptidase assays

Recombinant Sm32 was detected by immunoblotting after SDS-PAGE (4-12% gradient gels; Novex) using a monospecific rabbit antiserum developed to recombinant N197 Sm32 expressed in *Escherichia coli* (Ewald Beck, Biochemisches Institut, Universität Giessen, Germany). For activation experiments, enzyme solution was incubated for different time periods at 37\(^\circ\)C in the presence of 2 mM dithiothreitol (DTT) and no further additions (pH was measured to be 6.8) or 0.2 M sodium acetate pH 4.5 (pH was adjusted using a one-tenth volume of 2 M sodium acetate pH 4.5). The following endopeptidase inhibitors (final concentrations were 14 M unless otherwise indicated) were tested for their ability to prevent activation of Sm32: iodoacetic acid (IA; 5 M), E-64, morpholinourea-Phe-homoPhe-vinyl sulfone phenyl, PMSF (2 mM) leupeptin, pepstatin A and 1,10-phenanthroline (2 mM). Endopeptidase activity was monitored by hydrolysis of benzoyl (Z)-Ala-Ser-Asn-7-amido-4-methylcoumarin (NMeC; custom synthesis, Enzyme Systems Products). In our hands, this substrate is degraded by Sm32 approximately twice as fast as the more commonly used Z-Ala-Ala-Asn-NMeC (Caffrey, unpublished observation). Assays were performed in black 96-well microtiter plates and contained 5-30 »l enzyme solution preincubated for 10 min in 100 »l 0.1 M citrate-phosphate, 2 mM DTT, pH 6.8. Substrate (stocks in dimethyl sulfoxide) in 100 »l of the same buffer (10 M) was then added and the reaction allowed to proceed for 20-30 min at room temperature. Fluorescence was measured in a Lab-systems Fluoroskan II plate reader at excitation and emission wavelengths of 355 and 460 nm, respectively.

3. Results and discussion

3.1. A second gene encodes Sm32 with cysteine in the active site

Heretofore, only one complete gene sequence has been identified for Sm32 (a.k.a. *S. mansoni* asparaginyl endopeptidase or schistosome legumain) [3,11]. This was predicted to encode an inactive enzyme as the active site C197 had been replaced by an asparagine residue [4]. Consistent with this prediction, our preliminary attempts to obtain endopeptidase activity from an N197 clone expressed in *P. pastoris* had failed. Subsequent isolation and sequencing of a number of clones derived from *S. mansoni* mRNA identified a second gene for Sm32 where C197 was present and the entire ORF is pre-

---

**Fig. 2.** Amino acid sequence alignment of C197 Sm32 with N197 Sm32 [11] and human legumain [2]. Amino acid residues differing between the two schistosome sequences are indicated by asterisks; areas where all three sequences are identical are shaded.
sented in Fig. 1. C197 Sm32 has 97.2 and 41.0% identity to the previously published N197 Sm32 [11] and human legumain [2], respectively (Fig. 2). C197 Sm32 differs from N197 Sm32 by 12 amino acid residues. Also, C197 Sm32 has two potential N-glycosylation sites whereas N197 Sm32 has three, one of which, interestingly, is found at position 197.

3.2. Expression of Sm32 forms in P. pastoris and activation of C197 Sm32

Both N197 and C197 Sm32 were expressed and detectable in Pichia cultures using an anti-N197 Sm32 monospecific rabbit serum (Fig. 3). In fresh preparations of the N197 Sm32, a single immunoreactive protein was detected at 50 kDa (Fig. 3a). For C197 Sm32, predominant protein bands at 50 and 47 kDa were detected with minor species at 42 and 39 (Fig. 3b). Very little hydrolysis (<5%) of the fluorogenic substrate Z-Ala-Ser-Asn-NMec could be detected in fresh C197 Sm32 preparations (Fig. 4). Significant endopeptidolytic activity was only evident on incubation of C197 Sm32 preparations at pH 4.5 and correlated with the appearance of a final protein product at 34 kDa (Fig. 3b, 2 h time point onwards). Prolonged incubation of C197 Sm32 at pH 6.8 did not evoke increased hydrolysis of the fluorogenic substrate from the 5% basal level (Fig. 4) nor conversion of the major 50 and 47 kDa proteins (not shown), thus indicating the acid pH requirement for conversion to the fully active enzymatic form. In contrast to C197 Sm32, N197 Sm32 was incapable of being activated, either at pH 4.5 (Fig. 3) or at pH 6.0 (not shown), and did not hydrolyze Z-Ala-Ser-Asn-NMec (Fig. 4). These results support the prediction based on sequence homology [4] that Sm32 with N rather than C at position 197 would be most likely inactive.

3.3. Activation of Sm32 requires Cys at position 197 and is an autocatalytic event

Confirmation that the endopeptidolytic activity of Sm32 requires Cys at position 197 was obtained by recombinant expression of N197 Sm32 where the active site Asn at 197 had been mutated to a Cys. Like C197 Sm32, the N197C mutant was capable of being activated, but only at pH 4.5 (Figs. 4 and 5) and not at pH 6.8 (not shown). Also, increased hydrolysis of Z-Ala-Ser-Asn-NMec was only associated with the appearance of the final protein product which resolved at 37 kDa (Fig. 5, 2 h time point and onwards). Tests with endopeptidase inhibitors formally eliminate the possibility that an endopeptidase(s) other than Sm32 in the Pichia culture medium activates Sm32 in trans. Only the general cysteine endopeptidase inhibitor IA blocked activation of both C197 (Fig. 3b) and N197C mutant Sm32 (Fig. 5). Importantly, inhibitors of papain-like cysteine endopeptidases such as E-64, morpholino-Phe-homoPhe-vinyl sulfone phenyl (Fig. 3b) and leupeptin (not shown) did not prevent conversion to the final enzyme forms. E-64 has previously been shown to be relatively ineffective against asparaginyl endopeptidase activity in schistosome extracts [15] and pig legumain [2]. Likewise, inhibitors of serine (PMSF and leupeptin), aspartic (pepstatin A) and metallo-endopeptidases (1,10-phenanthroline) were ineffective (not shown).

3.4. The function of the C-terminal extension in Sm32

The immunoblot and enzymatic experiments herein demonstrate that maximal endopeptidolytic activity of C197 and N197C mutant Sm32 is associated with the autocatalytic proc-
cessing of an approximately 50 kDa protein species to about 35 kDa. Consistent with plant and animal legumains [1–3], loss of the C-terminal extension and possibly of the N-terminal pro-domain would account for this conversion. Given the specificity of asparaginyl endopeptidases like Sm32 for Asn at P1, the processing sites for removal of the N-terminal pro-domain and the C-terminal extension of Sm32 as previously proposed [14] (Fig. 1) must now be re-evaluated. The identification of a 50 kDa Sm32 precursor and a 35 kDa mature protein agrees with the detection of native asparaginyl endopeptidase species in S. mansoni extracts [16]. It is not clear, however, whether the conversion to the fully active enzyme in the worm is a cis or trans event. The C-terminal extension may perform a function similar to that of the N-terminal pro-region of papain-like endopeptidases such as cathepsin B and L [17] by inhibiting the active site until activity is required by the cell.

Acknowledgements: This research was supported by a Sandler Program in Basic Sciences Award to C.R.C. and J.M.K. The authors are indebted to Dr. Ewald Beck of the Biochemisches Institut, Universität Giessen, Germany, for the production of the monospecific rabbit anti-Sm32 serum and to Dr. Fred Engelbrecht of the Abteilung für Parasitologie, Hygiene Institut, Heidelberg Universität, Germany, for its supply. C.R.C. gratefully acknowledges the support of Dr. Andreas Ruppel of the Abteilung für Tropenhygiene, Hygiene Institut, Heidelberg Universität, Germany. J.M.K. is a Burroughs Wellcome Scholar in Molecular Parasitology.

References