

Identification of a cDNA encoding an active asparaginyl endopeptidase of *Schistosoma mansoni* and its expression in *Pichia pastoris*¹

Conor R. Caffrey^a, Mary A. Mathieu^a, Alan M. Gaffney^a, Jason P. Salter^a,
Mohammed Sajid^a, Kimberley D. Lucas^a, Christopher Franklin^a, Matthew Bogyo^b,
James H. McKerrow^{a,*}

^aTropical Disease Research Unit at the Department of Pathology, UCSF, VAMC, 4150 Clement St.-113B, San Francisco, CA 94121, USA

^bDepartment of Biochemistry and Biophysics, UCSF, San Francisco, CA 94121, USA

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

Edited by Takashi Gojobori

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been reported, but by sequence homology it would be expected to yield an inactive product as the active site C197 had been replaced by N. We now describe a new *S. mansoni* gene in which C197 is present.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 precursor proteins to their final forms [6,7]. In mammals, the related asparaginyl endopeptidases have a key function in the initial processing of antigens for MHC class II presentation [8] and in the inhibition of osteoclast formation and bone resorption [9].

The asparaginyl endopeptidase of *S. mansoni* (a.k.a. Sm32) is an important immunodiagnostic antigen for schistosomiasis [10]. A single Sm32 gene in both the *S. mansoni* and *S. japonicum* species has been identified [3,11,12]. However, on the

basis of sequence homology with other asparaginyl endopeptidases, the product of this Sm32 gene would not be proteolytically active as the crucial active site cysteine residue at position 197 had been replaced by an asparagine (N197 Sm32) [4]. We now identify a second distinct gene for Sm32 containing the requisite C197 residue (C197 Sm32). Using the *Pichia pastoris* protein expression system, we have expressed recombinant, active C197 Sm32. We further show by site-directed mutagenesis (N197C) that N197 Sm32 is rendered proteolytically active and demonstrate that autocatalysis at acid pH leads to production of mature, active Sm32, accompanied by removal of the C-terminal extension and possibly of the N-terminal pro-domain.

2. Materials and methods

2.1. Cloning and sequencing of C197 Sm32, and site-directed mutagenesis

Poly(A)⁺ mRNA was isolated (Pharmacia) from 6-week-old *S. mansoni* worms of mixed sex and converted to single-stranded cDNA using AMV Reverse Transcriptase as described by the manufacturer (Life Technologies). PCR amplification of the C197 Sm32 gene was accomplished with primers designed to encompass the pro- and C-terminal regions of the endopeptidase: forward: 5'-ATACTC-GAGAAAAGACAATTAGATACAAATTATGAAGTATCC-3' and reverse: 5'-AATGCGGCCGCTTAACCGCAAATTTTTATGATTGCT-3'. For eventual cloning into the *P. pastoris* expression vector pPIC Zα A, the forward primer incorporated *Xho*I 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 *Not*I 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 *Xho*I and *Not*I. 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 *Afl*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-TATTCTCTTTATTCTTA-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'-TTATATTGAAGCATGGAATCAGG-3' (bases encoding the mutation are underlined) and the appropriate anti-parallel reverse primer. This mutation resulted in the same *Afl*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.

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 Z α A (10 μ g) was linearized with *Sac*I and purified. *P. pastoris* was electroporated at 1.5 kV and 129 Ω in electroporation cuvettes (2 mm gap; BTX). *Pichia* colonies growing under zeocin (Invitrogen) selec-

tion were picked for expansion first in 10 ml of yeast extract–peptone–dextrose medium containing 100 μ g/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 μ m filter, lyophilized and stored at 4°C.

```

1/1                               31/11
ATG ATG CTA TTC TCT TTA TTT CTT ATT AGC ATC TTA CAT ATT CTA TTA GTC AAA TGT CAA
M M L F S L F L I S I L H I L L V K C Q
61/21                               91/31
TTA GAT ACA AAT TAT GAA GTA TCC GAT GAA ACT GTT AGT GAT AAT AAT AAA TGG GCT GTA
L D T N Y E V S D E T V S D N N K W A V
121/41                               151/51
TTA GTA GCT GGA TCT AAT GGT TAT CCC AAT TAT AGA CAT CAA GCA GAT GTC TGT CAT GCT
L V A G S N G Y P N Y R H Q A D V C H A
181/61                               211/71
TAT CAT GTA TTA CGT TCA AAG GGT ATA AAA CCT GAG CAT ATT ATC ACA ATG ATG TAT GAT
Y H V L R S K G I K P E H I I T M M Y D
241/81                               271/91
GAT ATC GCT TAT AAT TTG ATG AAT CCA TTT CCT GGG AAA CTT TTT AAT GAT TAT AAC CAT
D I A Y N L M N P F P G K L F N D Y N H
301/101                              331/111
AAA GAT TGG TAT GAA GGA GTG GTG ATA GAT TAT CGC GGT AAA AAA GTC AAC TCG AAA ACT
K D W Y E G V V I D Y R G K K V N S K T
361/121                              391/131
TTT CTG AAA GTT TTG AAG GGA GAT AAA AGC GCT GGT GGG AAA GTT TTG AAA AGT GGA AAA
F L K V L K G D K S A G G K V L K S G K
421/141                              451/151
AAT GAT GAT GTA TTC ATA TAC TTC ACT GAT CAT GGT GCA CCG GGT CTA ATT GCT TTC CCT
N D D V F I Y F T D H G A P G L I A F P
481/161                              511/171
GAT GAT GAA TTA TAT GCT AAA CAA TTT ATG TCA ACA TTG AAA TAC TTA CAT AGT CAC AAA
D D E L Y A K Q F M S T L K Y L H S H K
541/181                              571/191
CGT TAT TCA AAA TTG GTG ATT TAT ATT GAA GCA TGT GAA TCA GGT TCC ATG TTC CAA CGA
R Y S K L V I Y I E A C E S G S M F Q R
601/201                              631/211
ATA TTA CCG TCG AAT CTA AGT ATT TAT GCG ACT ACA GCT GCT AGT CCA ACT GAA TCT AGT
I L P S N L S I Y A T T A A S P T E S S
661/221                              691/231
TAT GGT ACA TTT TGT GAT GAT CCA ACA ATA ACT ACT TGT CTG GCT GAT TTA TAC TCA TAC
Y G T F C D D P T I T T C L A D L Y S Y
721/241                              751/251
GAC TGG ATT GTT GAC TCA CAA ACA CAT CAT TTA ACA CAA CGA ACA CTC GAT CAA CAG TAT
D W I V D S Q T H H L T Q R T L D Q Q Y
781/261                              811/271
AAA GAG GTT AAA CGT GAA ACG AAT CTT AGT CAT GTT CAG AGA TAT GGT GAT ACG AGA ATG
K E V K R E T N L S H V Q R Y G D T R M
841/281                              871/291
GGT AAA TTA CAC GTT AGT GAA TTT CAA GGA AGT CGA AAG TCT TCA ACT GAG AAC GAT
G K L H V S E F Q G S R D K S S T E N D
901/301                              931/311
GAA CCT CCA ATG AAA CCA AGA CAT TCC ATC GCT TCG AGA GAT ATT CCA TTG CAT ACT CTA
E P P M K P R H S I A S R D I P L H T L
961/321                              991/331
CAT CGT CAA ATA ATG ATG ACC AAT AAT GCA GAA GAC AAA AGC TTT CTG ATG CAA ATT CTT
H R Q I M M T N N A E D K S F L M Q I L
1021/341                             1051/351
GGT TTG AAA CTG AAG AGA AGA GAT CTC ATT GAG GAT ACT ATG AAA TTA ATA GTG AAA GTA
G L K L K R R D L I E D T M K L I V K V
1081/361                             1111/371
ATG AAT AAT GAA GAA ATA CCA AAT ACC AAG GCA ACT ATC GAT CAA ACA TTG GAT TGT ACA
M N N E E I P N T K A T I D Q T L D C T
1141/381                             1171/391
GAA TCA GTC TAT GAA CAG TTC AAA AGT AAA TGT TTC ACA CTA CAA CAG GCT CCA GAA GTT
E S V Y E Q F K S K C F T L Q Q A P E V
1201/401                             1231/411
GGA GGG CAC TTT TCT ACT CTT TAC AAT TAT TGT GCA GAT GGT TAT ACA GCT GAA ACG ATC
G G H F S T L Y N Y C A D G Y T A E T I
1261/421
AAT GAA GCA ATC ATA AAA ATT TGC GGT
N E A I I K I C G

```

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.

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 evince 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

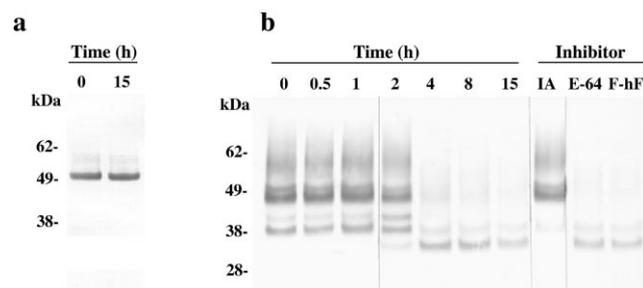


Fig. 3. Conversion of C197, but not N197 Sm32, to the final active enzyme product. As described in Section 2, Sm32 proteins, expressed in *P. pastoris* supernatants, were incubated at 37°C in the presence of DTT at pH 4.5 for different time periods. After SDS-PAGE and blotting onto PVDF membrane, supernatants were reacted with a monospecific anti-N197 Sm32 rabbit serum. a: N197 Sm32 after incubation at pH 4.5 for 0 and 15 h. b: as for a, but with C197 Sm32; inhibitors were incubated for 15 h with C197 Sm32. IA, iodoacetic acid; F-hF, morpholinourea-Phe-homoPhe-vinyl sulfone phenyl.

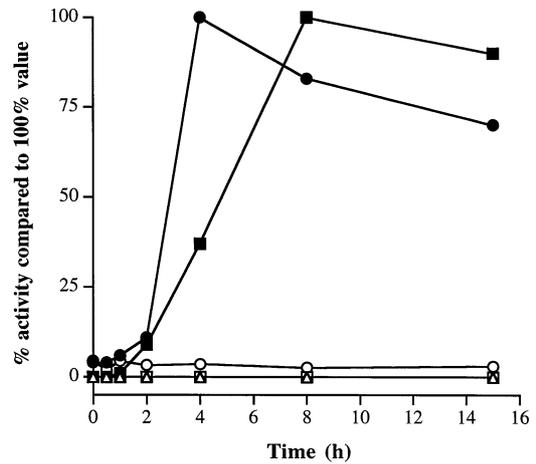


Fig. 4. Conversion to final Sm32 products at acid pH is associated with increased endopeptidolytic activity. As described for the above experiments (Section 2 and Fig. 3) N197, C197 or N197C mutant Sm32 was incubated at pH 6.8 or 4.5 for different time periods for up to 15 h. Endopeptidolytic activity of various Sm32 preparations was measured using the fluorogenic substrate Z-Ala-Ser-Asn-NMec. Incubation at pH 4.5; N197 (△), C197 (●) and N197C mutant (■). Incubation at pH 6.8; C197 (○) and N197C mutant (□).

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-

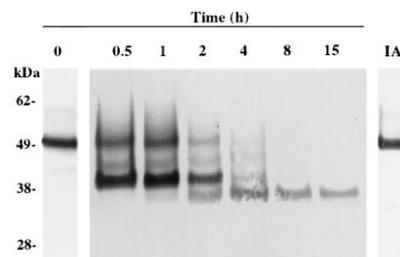


Fig. 5. Mutating N197 to C restores the ability of N197 Sm32 to convert to the final active Sm32 product. Experimental design was as for C197 Sm32 (legend to Fig. 3 and Section 2). IA, iodoacetic acid.

essing 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

- [1] Ishii, S., Abe, Y., Matsushita, H. and Kato, I. (1990) *J. Protein Chem.* 9, 294–295.
- [2] Chen, J.-M., Dando, P.M., Rawlings, N.D., Brown, M.A., Young, N.E., Stevens, R.A., Hewitt, E., Watts, C. and Barrett, A.J. (1997) *J. Biol. Chem.* 272, 8090–8098.
- [3] Davis, A.H., Nanduri, J. and Watson, D.C. (1987) *J. Biol. Chem.* 262, 12851–12855.
- [4] Chen, J.-M., Rawlings, N.D., Stevens, R.A.E. and Barrett, A.J. (1998) *FEBS Lett.* 441, 361–365.
- [5] Schechter, I. and Berger, A. (1967) *Biochem. Biophys. Res. Commun.* 27, 157–162.
- [6] Hara-Nishimura, I., Inoue, K. and Nishimura, M. (1991) *FEBS Lett.* 294, 89–93.
- [7] Jung, R., Scott, M.P., Nam, Y.-W., Beaman, T.W., Bassüner, R., Saalbach, I., Müntz, K. and Nielsen, N.C. (1998) *Plant Cell* 10, 343–357.
- [8] Manoury, B., Hewitt, E.W., Morrice, N., Dando, P.M., Barrett, A.J. and Watts, C. (1998) *Nature* 396, 695–699.
- [9] Choi, S.J., Reddy, S.V., Devlin, R.D., Menaa, C., Chung, H., Boyce, B.F. and Roodman, G.D. (1999) *J. Biol. Chem.* 274, 27747–27753.
- [10] Ruppel, A., Shi, Y.E., Wei, D.X. and Diesfeld, H.J. (1987) *Clin. Exp. Immunol.* 69, 291–298.
- [11] Klinkert, M.-Q., Felleisen, R., Link, G., Ruppel, A. and Beck, E. (1989) *Mol. Biochem. Parasitol.* 33, 113–122.
- [12] Merckelbach, A., Hasse, S., Dell, R., Eschlbeck, A. and Ruppel, A. (1994) *Trop. Med. Parasitol.* 45, 193–198.
- [13] Cormack, B. (1999) in: *Short Protocols in Molecular Biology* (Ausubel, A., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K., Eds.), pp. 8–21, John Wiley, New York.
- [14] Dalton, J.P. and Brindley, P.J. (1998) in: *Handbook of Proteolytic Enzymes* (Barrett, A.J., Rawlings, N.D. and Woessner, J.F., Eds.), pp. 749–754, Academic Press, London.
- [15] Dalton, J.P., Hola-Jamriska, L. and Brindley, P.J. (1995) *Parasitology* 111, 575–580.
- [16] El Meanawy, M.A., Aji, T., Phillips, N.F.B., Davis, R.D., Salata, R.A., Malhotra, I., McClain, D., Aikawa, M. and Davis, A.H. (1990) *Am. J. Trop. Med. Hyg.* 43, 67–78.
- [17] Cygler, M. and Mort, J.S. (1997) *Biochimie* 79, 645–652.