

Primary structures of Sm31/32 diagnostic proteins of *Schistosoma mansoni* and their identification as proteases

Mo-Quen Klinkert, Richard Felleisen, Gisela Link, Andreas Ruppel¹ and Ewald Beck

Zentrum für Molekulare Biologie, and ¹Institut für Tropenhygiene, University of Heidelberg, Heidelberg, F.R.G.

(Received 26 May 1988; accepted 6 October 1988)

We have constructed cDNA clones containing the complete nucleotide sequences coding for two highly antigenic *Schistosoma mansoni* adult worm proteins, Sm31 and Sm32. The predicted amino acid sequence of Sm31 shows significant homology to mouse, rat and human cathepsin B. The nucleotide sequence of Sm32 is identical to that reported by others for *S. mansoni* 'haemoglobinase'. The different nucleotide sequences demonstrate the existence of two different proteolytic enzymes, both of which are synthesised in the form of precursor molecules. Structural homology of the schistosome cathepsin B to the mammalian ones indicates that the mature protein is processed from a propeptide. The calculated molecular weight of haemoglobinase of 47 000 suggests that post-translational processing is also involved in generating an active protease.

Key words: *Schistosoma mansoni*; Nucleotide sequence; Haemoglobinase; Cathepsin B

Introduction

Two *Schistosoma mansoni* adult worm proteins of apparent molecular weights 31 000 and 32 000 (Sm31/32) are highly immunogenic in mice [1] and humans [2,3]. Our interest in these polypeptides arose mainly from their use as antigens for the serodiagnosis of schistosomiasis. In immunoblots performed with total adult worm proteins the native antigens readily reacted with schistosomiasis infection sera, but not with sera from patients infected with other parasites [4].

We have cloned and expressed these antigens in *Escherichia coli*, in order to use them as probes for diagnostic purposes [5]. The cloned proteins

express epitopes present on the native antigens, although the epitopes cloned so far reacted only with some of the patient sera [6].

In this paper, we report the DNA sequences of these naturally immunogenic proteins and demonstrate the homology of Sm31 at the protein level with a mammalian protease, while the nucleotide sequence of Sm32 is shown to be identical to the schistosome 'haemoglobinase'.

Materials and Methods

Materials. Restriction endonucleases and T4 DNA ligase were obtained from Renner, Dannstadt-Schauernheim, F.R.G. Radioactive nucleotides were purchased from Amersham-Buchler, Braunschweig, F.R.G. The plasmid vector pSP65 [7] was used for subcloning schistosome fragments.

Correspondence address: Mo-Quen Klinkert, ZMBH, Universität Heidelberg, Im Neuenheimer Feld 282, D-6900 Heidelberg, F.R.G.

Abbreviations: cDNA, complementary DNA; SDS, sodium dodecyl sulphate; PAGE, polyacrylamide gel electrophoresis.

Note: Nucleotide sequence data reported in this paper have been submitted to the GenBank™ Data Bank with the accession number J03984.

DNA sequencing. Nucleotide sequence analysis was undertaken using the Maxam Gilbert solid phase technique [8] and the dideoxynucleotide chain termination method of Sanger [9]. Different primers were synthesised according to the

phosphoramidite method [10] with an automated oligonucleotide synthesiser (380A9, Applied Biosystems). DNA sequences were processed using the computer programmes of Osterburg et al. [11] and the derived protein sequences were compared to the Dayhoff protein sequence database (EMBL Data Library).

Nucleic acid preparations. DNA and RNA were purified from frozen adult worms as previously described [5]. Radioactive labelling of the DNA was done using random primers purchased from Pharmacia [12].

Blot analysis. For analysis of genomic DNA 5 µg aliquots of schistosome DNA were digested with EcoRI, HindIII and HpaI. The digests were electrophoresed on 0.8% agarose gels and transferred to nylon membranes. RNA was electrophoresed through 1.2% formaldehyde agarose gels [13] and transferred to nylon membrane. Poly(A)⁺ schistosome RNA (2 µg) and total RNA (10 µg) from human liver HepG2 cells and rat HTC cells (kind gifts of Martin Hofmann) were used.

In vitro transcription. Transcription reactions were performed according to Melton et al. [7]. DNA fragments were cloned in pSP65. Linearised DNA templates (5 µg) were transcribed in a total volume of 50 µl containing 40 mM Tris-HCl pH 7.5, 6 mM MgCl₂, 2 mM spermidine, 10 mM dithiothreitol, 50 U RNasin (Promega, Biotec, Madison, WI), 500 µM each of four ribonucleoside triphosphates and 10 U of SP6 polymerase (Boehringer Mannheim, F.R.G.). Following 60 min of incubation at 40°C, the reaction mixture was extracted with phenol and chloroform and precipitated with ethanol. The RNA was redissolved in 40 µl H₂O.

In vitro translations and immunoprecipitation. 10% of the RNA as prepared above, corresponding to 0.5–1 µg RNA, was translated in 40 µl rabbit reticulocyte lysate (Amersham Corp.) in the presence of 50 µCi of [³⁵S]methionine (Amersham Corp.) in a total volume of 50 µl [14]. After various times at 30°C the reaction was stopped by

adding 100 µg ml⁻¹ of pancreatic RNase.

10 µl of the translation products were incubated with 10 µl of antiserum in the presence of 50 µl RIPA (150 mM NaCl, 10 mM sodium phosphate, 1% Nonidet P40, 0.5% sodium deoxycholate, 0.1% sodium dodecyl sulphate (SDS)) for 1 h at room temperature. Immune complexes were precipitated in the presence of 50 µl protein A-Sepharose (Pharmacia) for a further hour at room temperature [15]. Unbound materials were removed by washing twice with RIPA. Immunoprecipitated proteins were solubilised in an equal volume of sample buffer by heating to 95°C for 15 min and analysed by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and autoradiography.

Results

The complete Sm31 sequence. Using the EcoRI insert isolated from the original Sm31 clone, H3, a further clone λ H3-21 was isolated from our λgt10 library. The clone contains a cDNA insert of 1200 base pairs (bp). Digestion with EcoRI yielded two insert fragments of 220 bp and a third fragment of 750 bp. Each was subcloned in plasmid pSP65 and its nucleotide sequence determined. A partial restriction map of the clone, together with the sequencing strategy is shown in Fig. 1A.

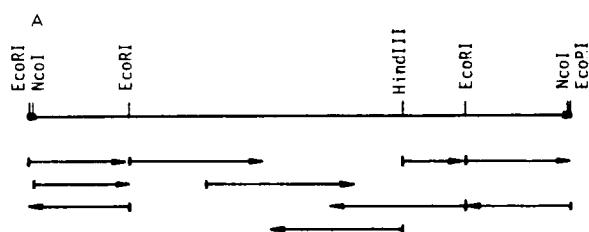
The primary DNA sequence is shown in Fig. 1B. A long open reading frame starts from an initiation triplet at position 39 and extends over most of the nucleotide sequence. The total coding capacity of the sequence starting from the first methionine was calculated to be 38.5 kDa. Since the native protein has an apparent molecular weight of 31 000 in Western blots, it is likely that a precursor protein exists, which is further processed into a smaller polypeptide (see below). The first 17 amino acid residues may represent a signal sequence.

The clone contains a polyadenylation signal (underlined in Fig. 1B) and, approximately 100 nucleotides downstream from the end of the open reading frame, 11 'A' residues, which are taken to represent part of the poly(A) tail and thus the 3' end of the mRNA.

Sequence homology with cathepsin B. In order to establish a possible function for Sm31 in schistosomes, the deduced amino acid sequence was compared to the sequence data bank of Dayhoff. Since the Sm31 clone encodes a polypeptide that

was earlier shown to originate from the gut and was suspected to be a digestive enzyme [16], the inferred amino acid sequence of the cDNA insert was examined for homology with that of other proteases.

The best match was found with liver cathepsin B of the mouse, rat and human [17] (Fig. 2). For the mammalian proteinases a precursor molecule and cleavage mechanisms to generate the mature enzymes have been postulated. The primary structure of the schistosome protein appears to be similar to the mammalian counterparts, in that the existence of precursor forms containing prepeptide and propeptide chains at the N-terminus can



B

1	<u>GGAATTC</u> <u>CATG GATGCATGCA</u> TTCAAGAGTT ATTTGGAC	ATG CTC ACA TCT ATT TTG TGT ATT GCT TCA CTA ATA ACT TTT
		Met Leu Thr Ser Ile Leu Cys Ile Ala Ser Leu Ile Thr Phe
80	CTT GAA GCT CAT ATT TCA GTT AAG AAC GAA AAG TTT GAA CCA TTG TCA GAT GAC ATA ATC TCG TAT ATC AAT GAA	Leu Glu Ala His Ile Ser Val Lys Asn Glu Lys Phe Glu Pro Leu Ser Asp Asp Ile Ile Ser Tyr Ile Asn Glu
155	CAT CCG AAT GCT GGA TGG AGA GCT GAG AAA AGC AAC CGT TTC CAC TCA CTG GAT GAC GCC CGA ATT CAG ATG GGT	His Pro Asn Ala Gly Trp Arg Ala Glu Lys Ser Asn Arg Phe His Ser Leu Asp Asp Ala Arg Ile Gln Met Gly
230	GCA AGA AGA GAA GAA CCA GAT TTG AGG AGG AAG AGA CGT CCC ACA GTT GAC CAT AAT GAT TGG AAT GTG GAA ATT	Ala Arg Arg Glu Glu Pro Asp Leu Arg Arg Lys Arg Arg Pro Thr Val Asp His Asn Asp Trp Asn Val Glu Ile
305	CCA TCC AAT TTC GAT TCC AGG AAG AAA TGG CCT GGT TGT AAA AGT ATC GCA ACC ATT CGT GAT CAA TCA CGA TGT	Pro Ser Asn Phe Asp Ser Arg Lys Lys Trp Pro Gly Cys Lys Ser Ile Ala Thr Ile Arg Asp Gln Ser Arg Cys
380	GGA TCA TGT TGG TCT TTC GGT GCG GTC GAA GCA ATG AGT GAT CGA AGC TGC ATA CAG TCA GGT GGC AAA CAA AAT	Gly Ser Cys Trp Ser Phe Gly Ala Val Glu Ala Met Ser Asp Arg Ser Cys Ile Gln Ser Gly Gly Lys Gln Asn
455	GTC GAA CTG AGT GCT GTT GAT CTT TTG ACC TGT TGC GAA AGT TGT GGA CTT GGA TGT GAA GGT GGT ATC CTT GGT	Val Glu Leu Ser Ala Val Asp Leu Leu Thr Cys Cys Glu Ser Cys Gly Leu Gly Cys Glu Gly Gly Ile Leu Gly
530	CCT GCT TGG GAT TAC TGG GTG AAG GAA GGT ATT GTT ACT GCA AGT TCG AAA GAA AAT CAC ACA GGT TGT GAA CCA	Pro Ala Trp Asp Tyr Trp Val Lys Glu Gly Ile Val Thr Ala Ser Ser Lys Glu Asn His Thr Gly Cys Glu Pro
605	TAT CCA TTT CCG AAA TGT GAA CAT CAT ACC AAA GGA AAG TAT CCT CCA TGT GGT AGT AAA ATC TAC AAC ACT CCT	Tyr Pro Phe Pro Lys Cys Glu His His Thr Lys Gly Lys Tyr Pro Pro Cys Gly Ser Lys Ile Tyr Asn Thr Pro
680	CGT TGT AAA CAA ACT TGT CAG AGA AAA TAT AAA ACC CCG TAC ACC CAA GAC AAG CAT CGA GGT AAA TCA TCC TAC	Arg Cys Lys Gln Thr Cys Gln Arg Lys Tyr Lys Thr Pro Tyr Thr Gln Asp Lys His Arg Gly Lys Ser Ser Tyr
755	AAT GTT AAA AAT GAT GAA AAG GCT ATT CAA AAA GAA ATC ATG AAA TAC GGA CCA GTA GAA GCA AGC TTC ACC GTT	Asn Val Lys Asn Asp Glu Lys Ala Ile Gln Lys Glu Ile Met Lys Tyr Gly Pro Val Glu Ala Ser Phe Thr Val
830	TAT GAA GAT TTT CTA AAT TAC AAA TCT GGA ATC TAT AAA CAT ATT ACT GGA GAA GCA CTT GGT GGG CAC GCT ATA	Tyr Glu Asp Phe Leu Asn Tyr Lys Ser Gly Ile Tyr Lys His Ile Thr Gly Glu Ala Leu Gly Gly His Ala Ile
905	CGT ATT ATC GGA TGG GGT GTG GAA AAC AAG ACT CCT TAT TGG TTG ATT GCG AAT TCG TGG AAT GAA GAT TGG GGT	Arg Ile Ile Gly Trp Gly Val Glu Asn Lys Thr Pro Tyr Trp Leu Ile Ala Asn Ser Trp Asn Glu Asp Trp Gly
980	GAA AAT GGA TAT TTC AGA ATA GTG CGT GGT CGT GAT GAG TGT TCC ATT GAA TCT GAA GTA ATA GCC GGT CGA ATA	Glu Asn Gly Tyr Phe Arg Ile Val Arg Gly Arg Asp Glu Cys Ser Ile Glu Ser Glu Val Ile Ala Gly Arg Ile
1055	AAC TAA TAACACAATC TTGTAACATC ATCACAATAA GTTACTTAGG GTTAATCAAG TCATTACAAA TGTATACTAA TACTTCAAAAT	Asn ***
1141	<u>AAAAGTTT</u> TGTTCTGAAA AAAAAAAGC ATGCATCCAT <u>GGAATTC</u>	

Fig. 1. (A) Partial restriction map and sequencing strategy for *S. mansoni* Sm31 cDNA. The full-length clone H3-21 was constructed in λ gt10. Arrows indicate the direction in which the fragments were sequenced, and their lengths are proportional to the distance sequenced. (B) The complete nucleotide and deduced amino acid sequences of *S. mansoni* Sm31. Boxes are the adaptor sequences containing EcoRI sites used for cloning. The predicted protein starting at the first methionine is 340 amino acid residues long. Asterisks indicate the translation stop codon. A sequence showing the polyadenylation signal upstream of the poly(A) tract is underlined.

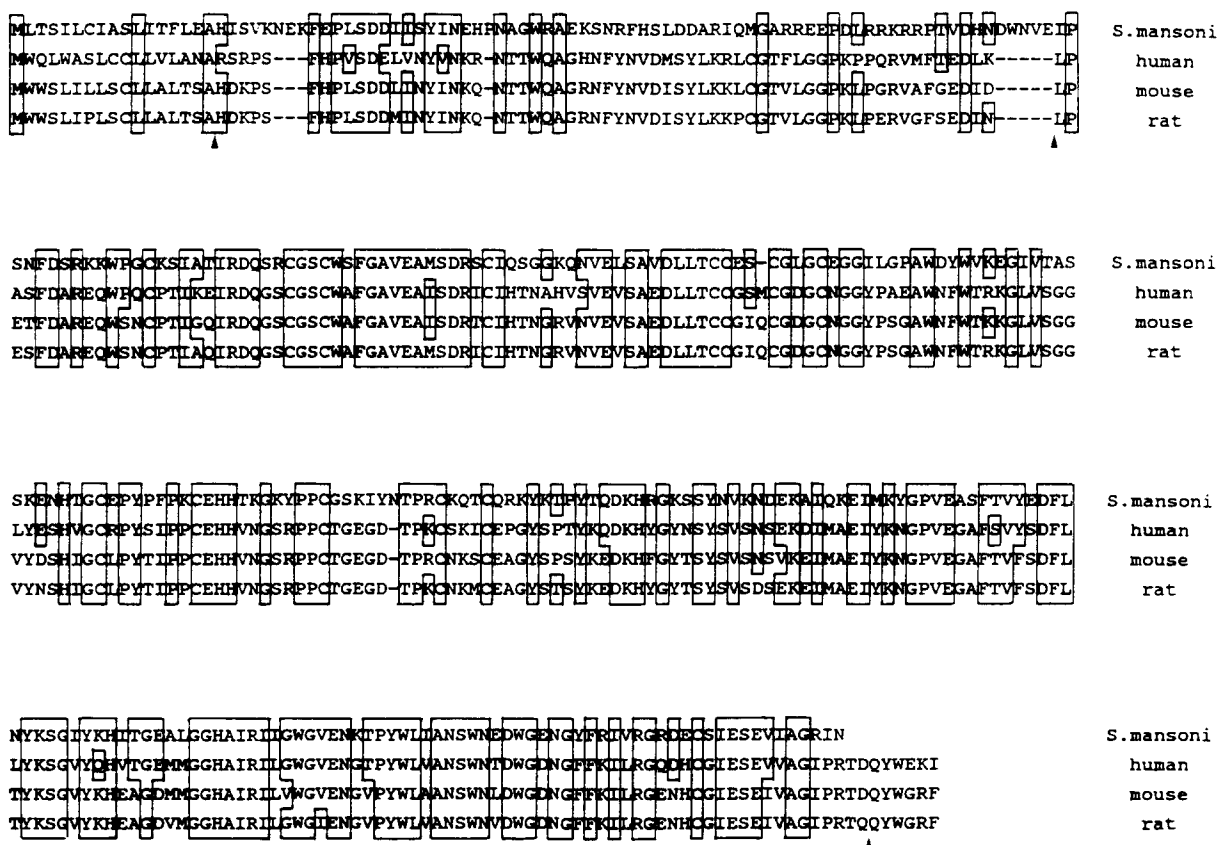


Fig. 2. Homology of the predicted amino acid sequence of *S. mansoni* protein Sm31 with preprocathepsin B of mouse, rat and man. Boxes indicate the amino acid residues of the schistosome sequence that are identical to the mammalian proteases. Arrowheads indicate the potential cleavage sites of the prepeptide and propeptide fragments to generate the mature cathepsin B, as proposed by Chan et al. [17].

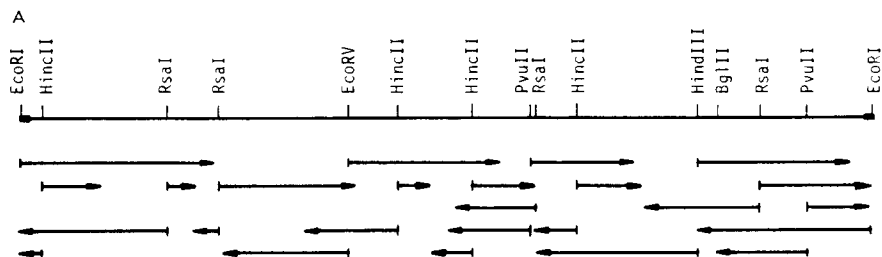


Fig. 3. (A) Partial restriction map and sequencing strategy for *S. mansoni* haemoglobinase. The full-length clone M2-16 was isolated from λ gt10. Arrows depict the length and direction of the sequences. (B) Nucleotide and predicted amino acid sequences of the complete clone M2-16. The nucleotide sequence from position 647 onwards is identical to the published sequence [18], with the exception of 7 nucleotide exchanges indicated above the sequence. Adaptor sequences are boxed. A potential signal peptide sequence (broken line) and polyadenylation signals (solid lines) are shown.

1 GAATTCCTAG GATGCATGGT TTTTTTTT TTACCCCTTG ATTGTGAAT TTTTATTTC CTCGGTGTTA ACCTCATTTC TCCATACCCA
90 TTTGGATT TAACATCTG CTAACATTG TAAATTTTGT TTTTTCCTTT CCCATTGAAA TAACATAGAC AAACGATGCA TTTCTCTAGC
180 CTTTTTTTCA TTGTATGTT TTATTCATTC TCAATATAAA TAACACTATT CGTTGTACA TTATCCATTC TTCCTTTTTT GCGTAAAAAGA
270 AATATTACCA TTACCATTAC TATTATCGTT ATTCCTCGTA TTATTATTAC TATCATCCTT GTTTATATTA TTGTGAGTGT ATGTGTACAT
360 GTTTATGTAT GAGCATGGAG CCTATAATGT GAAGATTCTT GAGTCATTAA TCGAGAAAAG CACAGCAACT TTAGAACTAC ACTACAATGT
450 TGTACTIONTAC TTTTAAAAAT ACATTAANAAC ACCACCACCA TTATTGAT ATG ATG CTA TTC TCT TTA TTT CTT ATT AGC ATC
Met Met Leu Phe Ser Leu Phe Leu Ile Ser Ile
531 TTA CAT ATT CTA TTA GTC AAA TGT CAA TTA GAT ACA AAT TAT GAA GTA TCC GAT GAA ACT GTT AGT GAT AAT AAT
Leu His Ile Leu Leu Val Lys Cys Gln Leu Asp Thr Asn Tyr Glu Val Ser Asp Glu Thr Val Ser Asp Asn Asn
606 AAA TGG GCT GTA TTA GTA GCT GGA TCT AAT GGT TAT CCC AAT TAT AGA CAT CAA GCA GAT GTC TGT CAT GCT TAT
Lys Trp Ala Val Leu Val Ala Gly Ser Asn Gly Tyr Pro Asn Tyr Arg His Gln Ala Asp Val Cys His Ala Tyr
681 CAT GTA TTA CGT TCA AAG GGT ATA AAA CCT GAG CAT ATT ATC ACA ATG ATG TAT GAT GAT ATC GCT TAT AAT TTG
His Val Leu Arg Ser Lys Gly Ile Lys Pro Glu His Ile Ile Thr Met Met Tyr Asp ile Ala Tyr Asn Leu
756 ATG AAT CCA TTT CTT GGG AAA CTT TTT AAT GAT TAT AAN CAT AAA GAT TGG TAT GAA GGA GTG GTG ATA GAT TAT
Met Asn Pro Phe Leu Gly Lys Leu Phe Asn Asp Tyr Asn His Lys Asp Trp Tyr Glu Gly Val Val Ile Asp Tyr
831 CGC GGT AAA AAA C GTC AAC TCG AAA ACT TTT CTG AAA GTT TTG AAG GGA GAT AAA AGC GCT GGT GGG AAA GTT TTG
Arg Gly Lys Lys Val Asn Ser Lys Thr Phe Leu Lys Val Leu Lys Gly Asp Lys Ser Ala Gly Gly Lys Val Leu
906 AAA AGT GGA AAA AAT GAT GAT GTA TTC ATA TAC TTT ACT GAT CAT GGT GCA CCG GGT CTA ATT GCT TTC CCT GAT
Lys Ser Gly Lys Asn Asp Asp Val Phe Ile Tyr Phe Thr Asp His Gly Ala Pro Gly Leu Ile Ala Phe Pro Asp
981 GAT GAA TTA TAT GCC AAA GAA TTT ATG TCA ACG TTG AAA TAC TTA CAT AGT CAC AAA CGT TAT TCA AAA TTG GTG
Asp Glu Leu Tyr Ala Lys Glu Phe Met Ser Thr Leu Lys Tyr Leu His Ser His Lys Arg Tyr Ser Lys Leu Val
1056 ATT TAT ATT GAA GCA AAT GAA TCA GGT TCC ATG TTC CAA CAA ATA TTA CCG TCG AAT CTA AGT ATT TAC GCG ACT
Ile Tyr Ile Glu Ala Asn Glu Ser Gly Ser Met Phe Gln Gln Ile Leu Pro Ser Asn Leu Ser Ile Tyr Ala Thr
1131 ACA GCT GCT AAT CCA ACT GAA TGT AGT TAT AGT ACA TTT TGT GGT GAT CCA ACA ATA ACT ACT TGT CTG GCT GAT
Thr Ala Ala Asn Pro Thr Glu Cys Ser Tyr Ser Thr Phe Cys Gly Asp Pro Thr Ile Thr Act Cys Leu Ala Asp
1206 TTA TAC TCA TAC AAC TGG ATT GTT GAC TCA CAA ACA CAT CAT TTA ACA CAA CGA ACA CTC GAT CAA CAG TAT AAA
Leu Tyr Ser Tyr Asn Trp Ile Val Asp Ser Gln Thr His His Leu Thr Gln Arg Thr Leu Asp Gln Gln Tyr Lys
1281 GAG GTT AAA CGT GAA ACG GAT CTT AGT CAT GTT CAG AGA TAT GGT GAT ACG AGA ATG GGT AAA TTA TAC GTT AGC
Glu Val Lys Arg Glu Thr Asp Leu Ser His Val Gln Arg Tyr Gly Asp Thr Arg Met Gly Lys Leu Tyr Val Ser
1356 GAA TTT CAA GGA AGT CGA GAT AAG TCT TCA T ACT GAG AAC GAT GAA C TCT CCA ATG AAG CCA AGA CAT TCC G ATC GCT
Glu Phe Gln Gly Ser Arg Asp Lys Ser Ser Thr Glu Asn Asp Glu Ser Pro Met Lys Pro Arg His Ser Ile Ala
1431 TCG AGA GAT ATT CCA TTG CAT ACT CTA CAT CGT CAA ATA ATG ATG ACC AAT AAT GCA GAA GAC AAA AGC TTT CTG
Ser Arg Asp Ile Pro Leu His Thr Leu His Arg Gln Ile Met Met Thr Asn Asn Ala Glu Asp Lys Ser Phe Leu
1506 ATG CAA ATT CTT GGT TTG AAA CTG AAG AGA AGA GAT CTC ATT GAG GAT ACT ATG AAA TTA ATA GTG AAA GTA ATG
Met Gln Ile Leu Gly Leu Lys Leu Lys Arg Arg Asp Leu Ile Glu Asp Thr Met Lys Leu Ile Val Lys Val Met
1581 AAT AAT GAA GAA ATA CCA AAT ACC AAG GCA ACT ATC GAT CAA ACA TTG GAT TGT ACA GAA TCA GTC TAT GAA CAG
Asn Asn Glu Glu Ile Pro Asn Thr Lys Ala Thr Ile Asp Gln Thr Leu Asp Cys Thr Glu Ser Val Tyr Glu Gln
1656 TTC AAA AGT AAA TGT TTC ACA CTA CAA CAG GCT CCA GAA GTT GGA GGG CAC TTT TCT ACT CTT TAC AAT TAT TGT
Phe Lys Ser Lys Lys Cys Phe Thr Ala Ceu Gln Gln Ala Pro Glu Val Gly His Phe Ser Thr Leu Tyr Asn Tyr Cys
1731 GCA GAT GGT TAT ACA GCT GAA ACG ATC AAT GAA GCA ATC ATA AAA ATT TGC GGT TAA ATCAAATTG ATCTGTCACT
Ala Asp Gly Tyr Thr Ala Glu Thr Ile Asn Glu Ala Ile Ile Lys Ile Cys Gly ***
1808 GTAATAAACT TGTGCGGTGT GAAATTTTIA ATAAATGATT CTTTCCTCAA AAAAAAAAAA AAAAAAAAAA CATGCATCCA TGGAAATTC

be deduced. Amino acid sequence comparison of the mature cathepsin B of all four species reveals more than 50% identity, as shown by the boxed amino acids. However, the propeptide sequence of the worm protease expresses less homology than the other prosegments. In addition, even though the amino terminal ends are not homologous, they seem to represent signal peptides. The potential cleavage site of the signal peptide is proposed to be between alanine and histidine [17]. A stop codon is present in the schistosome sequence before the position of the putative processing site in the mammalian proteases. Thus the schistosome polypeptide does not appear to be processed at the C-terminus.

Sm32 is identical to haemoglobinase. The cDNA clone M2-16 containing the complete sequence of another gut-located protein was isolated from the λ gt10 library, probed with the original H6 clone and confirmed with another clone M2, which extends approximately 200 bp into the N-terminal end of H6. Fig. 3 shows the sequencing strategy along with the nucleotide sequence. An initial attempt to compare the nucleotide and deduced amino acid sequences with known sequences (Dayhoff data base) yielded no significant homologies. However, our sequence was found to be identical to the one recently published for the schistosome haemoglobinase by Davis et al. [18].

The sequence has a long open reading frame, which encodes approximately 80 amino acids in addition to those published, 19 of which appear to represent a typical signal sequence (Fig. 3B). The potential cleavage site is proposed to be between cysteine and glutamine at positions 19 and 20. Beginning with the start codon a protein comprising more than 400 amino acid residues can be deduced. The calculated molecular weight is 47 000, which is considerably larger than the native haemoglobinase. Davis et al. [18] postulated processing of the precursor protein around the C-terminus to give a 32 kDa protein, by virtue of homology of the deduced amino acids at the N-terminus with data obtained from amino acid sequencing.

The nucleotide sequence contains a homopolymeric run of 11 'T' residues downstream of the 5' adaptor sequence, suggesting that the M2-16 clone

represents a head-to-head fusion of two different cDNAs. This idea is supported by the observation of an open reading frame of approximately 50 amino acids in the opposite direction (towards the 5' end of the cloned cDNA). The beginning of this open reading frame may indicate the point where the two cDNAs have been fused, some 120 nucleotides upstream of the start of the haemoglobinase coding region. This 120 bp sequence contains several stop codons in all possible reading frames in both 5' and 3' directions and is thus an untranslated region rather than a part of a protein reading frame.

Cathepsin B and haemoglobinase transcripts. Poly(A)⁺ RNA from adult worms was analysed by the Northern blot technique. A 1.4 kb transcript was recognized using the H3-21 cDNA insert as probe (Fig. 4, lane 1). Using HepG2 human and HTC rat RNA, no hybridisation signals were detected (data not shown). This finding is not too surprising since there is no strong homology at the nucleotide level with cathepsin B from other organisms for crosshybridisation to occur. The haemoglobinase-specific insert derived from M2-16 hybridised to a slightly larger transcript of approximately 1.5 kb in length (lane 2). This finding is consistent with the results of Davis et al. [18]. In both cases smaller transcripts around 1–1.2 kb were also observed. These RNA species are presumably degradation products.

In vitro translations. RNA transcribed in vitro was used for in vitro translation in rabbit reticulocyte lysate in the presence of [³⁵S]methionine. The labelled products were analysed by SDS-PAGE and autoradiography.

Using the RNA transcribed by pSP65-H3-21, one prominent band of approximately 31 kDa was detected (Fig. 5, lane 1, arrow). On a longer exposure (not shown) an additional protein with apparent molecular weight of 42 000 was also recognised (see also lane 7). While the smaller protein is taken to represent the mature protease, the larger product could serve as a precursor protein. The predicted molecular mass of a gene product starting from the first methionine [19] was calculated to be 38.5 kDa and could be represented by the 42 kDa species. The reason for the slower mi-

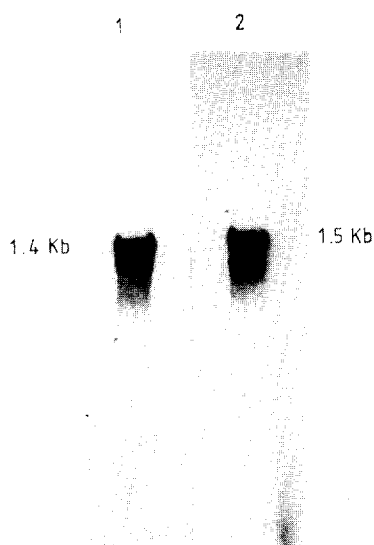


Fig. 4. Northern blot analysis of the cathepsin B and haemoglobinase transcripts of *S. mansoni*. 2 µg of poly(A)⁺ RNA were separated on a 1.2% formaldehyde-agarose gel, transferred to nylon membrane and hybridised to (1) H3-21 cathepsin B-specific and (2) M2-16 haemoglobinase-specific probes. The size of the mRNA was estimated from restriction fragments of single stranded DNA from bacteriophage fd.

gration, however, is not clear. From its relatively weak intensity it would appear that the precursor protein is translated and simultaneously processed into a more stable mature product of 31 kDa. Alternatively the 31 kDa species, rather than being the mature enzyme, could correspond to a protein synthesised from the second methionine in the reading frame. Evidence against this argument is the lack of a eukaryotic translation initiation sequence upstream from this ATG triplet [20].

Immunoprecipitations were carried out in a parallel experiment, the results of which support the existence of a precursor protein and the mature protease (Fig. 5). Both the 31 and 42 kDa products were immunoprecipitated using specific antiserum raised against the MS2-cathepsin B fusion protein, while the 45 kDa band was not immunoprecipitated (lane 8). The 45 kDa protein band is presumably a reticulocyte lysate specified polypeptide (lanes 1–7). It is also present during translation of the lysate in the absence of exogenous RNA (data not shown).

The appearance of the translation products was

followed by translating the RNA for various periods of time. The results demonstrate that the 31 kDa band appeared as early as 10 min (lane 3) and its intensity increased over 60 min (lanes 4–6). A possible explanation as to why the 42 kDa precursor was not observed could be that it is not translated in sufficient quantity in this particular experiment to be visible or it is quickly processed into the mature protease. Treatment of the reticulocyte lysate with RNase after translation removed acyl-tRNA molecules (arrowhead) as performed in lanes 4–6.

The processing of the mammalian preprocathepsin B at amino acid position 88 between glutamine and isoleucine has been postulated [13]. In this work a plasmid was constructed, in which the first 81 amino acids of pSP65-H3-21 were deleted through a convenient HincII site, and replaced by a linker fragment containing a eukaryotic translation initiation consensus sequence including the initiation triplet [20].

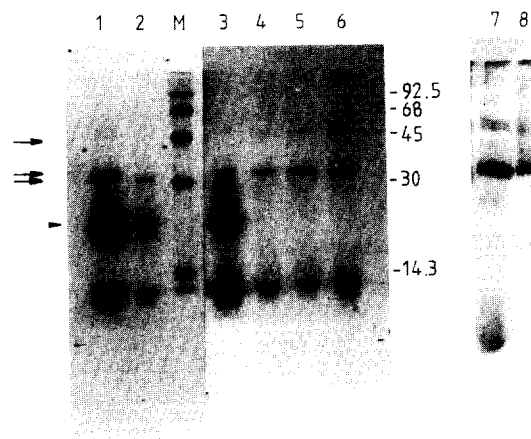


Fig. 5. In vitro translations of cathepsin B. A rabbit reticulocyte extract was incubated with in vitro synthesised RNA and the translation products analysed by SDS-PAGE and autoradiography. On translation of pSP65-H3-21 RNA, a 31 kDa protein was detected (lane 1, arrow). On a longer exposure a second protein of 42 kDa was also observed (see also lane 7). RNA from a deletion mutant directed the synthesis of a slightly smaller protein of 30 kDa (lane 2, arrow). In the same experiment (after exposure for a longer period), the appearance of the 31 kDa protein over a time course was observed (lane 3, 10 min; lane 4, 20 min; lane 5, 30 min; and lane 6, 60 min). Translation was terminated with the addition of RNase (lanes 4–6). In a different experiment the 31 and 42 kDa polypeptides could be specifically immunoprecipitated by anti-cathepsin B antibodies (lane 8) from total translation products (lane 7).

The mRNA is expected to synthesise a product initiating 9 amino acids upstream from the postulated cleavage site and therefore be larger than the mature enzyme. Alternatively the product is processed at the postulated site and thus comigrates with the mature cathepsin B. However, on translation of the mRNA a slightly smaller gene product of 30 kDa was obtained (lane 2). This finding suggests that processing into a mature protease does not take place at the same site postulated for the mammalian enzyme, but is likely to occur further upstream. By comparing the worm sequence with that of the other species, we note that the former is longer and that there are 5 additional amino acids immediately upstream of the putative cleavage site in the sequence (see Fig. 2). This divergence could account for a different cleavage specificity in the schistosome.

Discussion

This report identifies two highly antigenic *S. mansoni* worm proteins, Sm31 and Sm32, as acid proteases. Based on the nucleotide and deduced amino acid sequence data these proteins are homologous to mammalian cathepsin B and schistosome 'haemoglobinase', respectively. Differences in nucleotides between our sequence and the 'haemoglobinase' sequence of Davis et al. [18] may be due to allelic or strain variation.

The presence of an acidic proteolytic enzyme in the digestive tract of *S. mansoni* with a substrate specificity for haemoglobin was demonstrated by Timms and Bueding as early as 1959 [21]. Since then the general term 'haemoglobinase' has been in use for gut-derived schistosomal proteinases. Biochemical evidence suggested that more than one enzyme is present [22–26] and similarities to mammalian cathepsin B [23] as well as to cathepsin D [27] have been observed. In addition, several acid proteases without haemolytic activity but associated with the worm's digestive system have been reported [28]. This report provides conclusive evidence on the basis of the different nucleotide sequences that 'haemoglobinase' and cathepsin B are different enzymes.

Southern blot analysis was performed to estimate the gene copy number of cathepsin B. Using nick translated pSP65-H3-21 DNA to probe

worm genomic DNA digested with different restriction enzymes simple hybridisation patterns characteristic of a single copy gene were observed in each case (data not shown). This finding is consistent with that reported for mammalian cathepsin B.

In vitro translation studies demonstrated that two proteins 31 and 42 kDa in size were synthesized under the direction of RNA from H3-21. Results from immunoprecipitation experiments showed that both proteins were immunoprecipitated using specific antiserum, suggesting that the larger product probably represents the primary unprocessed polypeptide and the smaller one the mature enzyme.

Interestingly, two potential glycosylation sites that are conserved in the mammalian cathepsin B species (amino acid positions 42 and 201, Fig. 2) are not found in the schistosome sequence, but are present elsewhere in the sequence (positions 184 and 299). In addition two potential glycosylation sites have also been located on the haemoglobinase sequence. However, previous work showed that both worm proteases do not stain for carbohydrate. It is therefore reasonable to assume that they are not glycosylated.

Sm31 and Sm32 are among the most immunogenic proteins of adult *S. mansoni* and induce high levels of antibodies already during prepatency in mice [1] and men [2]. Similar results were obtained with *S. japonicum* [3,29] and *S. haematobium* [4]. Immunogenicity of a haemoglobinolytic enzyme following its intradermal injection to Schistosomiasis mansoni patients was previously observed by Senft and Maddison [30]. The lack of cross-reactivity with other schistosome species observed by these authors is in contrast with the high cross-reactivity of Sm31 and Sm32 obtained in our studies, but this discrepancy might be related to the different test systems used. Anti-Sm31/32 antibodies were not observed following vaccination of mice with irradiated cercariae [1] nor after immunisation with a schistosome homogenate [16]. However, vaccination did not preclude antibody formation following a challenge infection [1]. These findings were recently confirmed by the observation of anti-SMw32 proteinase antibodies in infected or in vaccinated-challenged baboons, but not in animals vaccinated with irradiated schistosomula [31].

The identification of Sm31 and Sm32 as two proteolytic enzymes permits a correlation of the immunological and biochemical properties of these proteins. The difficulty in resolving clearly individual schistosome components in the same molecular weight range both in Western blots [1,2] and in enzyme preparations [27,31] argues for the presence of several proteins with similar mobilities. Thus, the possibility that other proteolytic enzymes exist, besides the two described here, cannot be ruled out. An immunogenic proteinase of 30 kDa has also been reported in cercariae [32].

In view of the usefulness of Sm31 and Sm32 as probes in the development of diagnostic methods, we attempted to express all putative antigenic determinants on the proteins as recombinant antigens in *E. coli*. To date, we have only succeeded in fusing sequences encoding the almost full-length proteases to the MS2 polymerase fragment. Preliminary results suggest that the cloning of proteases both as precursors and in active forms in high copy number plasmids has a lethal effect on host cells. Success in the isolation of our full-length haemoglobinase clone in λ ap-

pears to be due to the fusion near the N-terminal start of the haemoglobinase to an unrelated sequence, which contains several AUG codons followed by translation stop signals, such that translation of the haemoglobinase precursor is prevented. Attempts to subclone the segments encoding only haemoglobinase-related sequences by use of various restriction enzymes have so far failed.

The cloning and full-length expression of these proteolytic enzymes are desirable for two reasons. It would be possible to express the proteins in their optimal antigenic forms and to incorporate them into diagnostic tests. In addition the processing or autoprocessing of these proteases could be studied both in vivo and in vitro in more detail.

Acknowledgements

This investigation received financial assistance from the UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases and from the Bundesministerium für Forschung und Technologie (BCT 318-5).

References

- 1 Ruppel, A., Rother, U., Vongerichten, H., Lucius, R. and Diesfeld, H.J. (1985) *Schistosoma mansoni*: Immunoblot analysis of adult worm proteins. *Exp. Parasitol.* 60, 195–206.
- 2 Ruppel, A., Diesfeld, H.J. and Rother, U. (1985) Immunoblot analysis of *Schistosoma mansoni* antigens with sera of schistosomiasis patients: diagnostic potential of an adult schistosome polypeptide. *Clin. Exp. Immunol.* 62, 499–506.
- 3 Ruppel, A., Shi, Y.E., Wei, D.X. and Diesfeld, H.J. (1987) Sera of *Schistosoma japonicum*-infected patients crossreact with diagnostic 31/32 kD proteins of *S. mansoni*. *Clin. Exp. Immunol.* 69, 291–298.
- 4 Idris, M.A. and Ruppel, A. (1988) Diagnostic *M*, 31/32 000 *Schistosoma mansoni* proteins (Sm31/32): Reaction with sera from Sudanese patients infected with *S. mansoni* or *S. haematobium*. *J. Helminthol.* 62, 95–101.
- 5 Klinkert, M.Q., Ruppel, A. and Beck, E. (1987) Cloning of diagnostic 31/32 kDa antigens of *Schistosoma mansoni*. *Mol. Biochem. Parasitol.* 25, 247–255.
- 6 Klinkert, M.Q., Ruppel, A., Felleisen, R., Link, G. and Beck, E. (1988) Expression of diagnostic 31/32 kilodalton proteins of *Schistosoma mansoni* as fusions with bacteriophage MS2 polymerase. *Mol. Biochem. Parasitol.* 27, 233–240.
- 7 Melton, D.A., Krieg, P.A., Rebagliati, M.R., Maniatis, T., Zinn, K. and Green, M.R. (1984) Efficient in vitro synthesis of biologically active RNA and RNA hybridisation probes from plasmids containing a bacteriophage SP6 promoter. *Nucleic Acids Res.* 12, 1035–1056.
- 8 Rosenthal, A., Jung, R. and Hunger, H.D. (1986) Optimized conditions for solid-phase sequencing: simultaneous chemical cleavage of a series of long DNA fragments immobilized on CCS anion-exchange paper. *Gene* 42, 1–9.
- 9 Sanger, F., Nicklen, S. and Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74, 5463–5467.
- 10 Caruthers, M.H. (1982) Chemical synthesis of oligonucleotides using the phosphite triester intermediates. In: *Chemical and Enzymatic Synthesis of Gene Fragments: A Laboratory Manual* (Gasser, H.G. and Lang, A., eds.), pp. 71–79, Verlag Chemie, Weinheim.
- 11 Osterburg, G., Glatting, J.H. and Sommer, R. (1982) Computer programs for the analysis and the management of DNA sequences. *Nucleic Acids Res.* 10, 207–216.
- 12 Feinbert, A.P. and Vogelstein, B. (1983) A technique for radiolabelling DNA restriction endonuclease fragments to a high specific activity. *Anal. Biochem.* 132, 6–13.
- 13 Lehrach, H., Diamond, D., Wozney, J.M. and Boedtker, H. (1977) RNA molecular weight determinations by gel

- electrophoresis under denaturing conditions, a critical reexamination. *Biochemistry* 16, 4793–4751.
- 14 Pelham, H.R.B. and Jackson, R.J. (1976) An efficient mRNA dependent translation system for reticulocyte lysates. *Eur. J. Biochem.* 67, 247–256.
 - 15 Goding, J.W. (1978) Use of staphylococcal Protein A as an immunological reagent. *J. Immunol. Methods* 13, 215–226.
 - 16 Ruppel, A., Breternitz, U. and Burger, R. (1987) Diagnostic M_r 31 000 *Schistosoma mansoni* proteins: requirement of infection, but not immunization, and use of the 'miniblot' technique for the production of monoclonal antibodies. *J. Helminthol.* 61, 95–101.
 - 17 Chan, S.J., San Segundo, B., McCormick, M.B. and Steiner, D.F. (1986) Nucleotide and predicted amino acid sequences of cloned human and mouse preprocathepsin B cDNAs. *Proc. Natl. Acad. Sci. USA* 83, 7721–7725.
 - 18 Davis, A.H., Nanduri, J. and Watson, D.C. (1987) Cloning and gene expression of *Schistosoma mansoni* protease. *J. Biol. Chem.* 262, 12851–12855.
 - 19 Kozak, M. (1978) How do eucaryotic ribosomes select initiation regions in messenger RNA? *Cell* 15, 1109–1123.
 - 20 Kozak, M. (1986) Influences of mRNA secondary structure on initiation by eucaryotic ribosomes. *Proc. Natl. Acad. Sci. USA* 83, 2850–2854.
 - 21 Timms, A.R. and Bueding, E. (1959) Studies of a proteolytic enzyme from *Schistosoma mansoni*. *Br. J. Pharmacol.* 14, 68–73.
 - 22 Sauer, M.C.V. and Senft, A.W. (1972) Properties of a proteolytic enzyme from *Schistosoma mansoni*. *Comp. Biochem. Physiol.* 42B, 205–220.
 - 23 Dresden, M.H. and Deelder, A.M. (1979) *Schistosoma mansoni*: thiol proteinase properties of adult worm 'hemoglobinase'. *Exp. Parasitol.* 48, 190–197.
 - 24 Bogitsh, B.J. and Dresden, M.H. (1983) Fluorescent histochemistry of acid proteases in adult *Schistosoma mansoni* and *Schistosoma japonicum*. *J. Parasitol.* 69, 106–110.
 - 25 Chappell, C.L. and Dresden, M.H. (1986) *Schistosoma mansoni*: proteinase activity of 'hemoglobinase' from the digestive tract of adult worms. *Exp. Parasitol.* 61, 160–167.
 - 26 Lindquist, R.N., Senft, A.W., Petitt, M. and McKerrow, J. (1986) *Schistosoma mansoni*: purification and characterization of the major acidic proteinase from adult worms. *Exp. Parasitol.* 61, 398–404.
 - 27 Bogitsh, B.J. and Kirschner, K.F. (1987) *Schistosoma japonicum*: Immunocytochemistry of adults using heterologous antiserum to bovine cathepsin D. *Exp. Parasitol.* 64, 213–218.
 - 28 Cesari, I.M., Rodriguez, M. and McLaren, D. (1981) Proteolytic enzymes in adult male and female *Schistosoma mansoni*. *Acta Cient. Venez.* 32, 324–329.
 - 29 Shi, Y.E., Dell, R., Diesfeld, H.J., Burger, R., Wei, D.X. and Ruppel, A. (In Press) Diagnostic 31/32 kD schistosoma proteins in *Schistosoma japonicum*: Reactivity in immunoblots with patient sera and monoclonal antibodies. *J. Parasitol. Parasitic Dis.*
 - 30 Senft, A.W. and Maddison, S.E. (1975) Hypersensitivity to parasite proteolytic enzyme in schistosomiasis. *Am. J. Trop. Med. Hyg.* 24, 83–89.
 - 31 Zerda, K.S., Dresden, M.H., Damian, R.T. and Chappell, C.L. (1987) *Schistosoma mansoni*: Anti-SMw32 proteinase response in vaccinated and challenged baboons. *Am. J. Trop. Med. Hyg.* 37, 320–326.
 - 32 Pino-Heiss, S., Petitt, M., Beckstead, J.H. and McKerrow, J.H. (1986) Preparation of mouse monoclonal antibodies and evidence for a host immune response to the preacetabular gland proteinase of *Schistosoma mansoni* cercariae. *Am. J. Trop. Med. Hyg.* 35, 536–543.