Identification of Novel Proteases and Immunomodulators in the Secretions of Schistosome Cercariae That Facilitate Host Entry*\[S\]

Rachel S. Curwen‡, Peter D. Ashton, Shobana Sundaralingam, and R. Alan Wilson

Schistosomiasis, caused by parasitic helminths, remains a serious human disease in the tropics. Cercariae of *Schistosoma mansoni* infect their hosts by direct skin penetration, aided by secretions from acetabular and head glands. Both proteolytic and immunomodulatory properties have been ascribed to the released material, but to date only five isoforms of elastase and one putative anti-inflammatory protein (Sm16) have been cloned. We analyzed secretions from mechanically transformed cercariae by two-dimensional electrophoresis. An average gel image was created and compared with a separation of soluble larval extract, revealing a less complex spot pattern in the secretions with 60% of the spots matched to the larval extract. Subsequent tandem mass spectrometric analysis identified 48 spots from the released material, representing approximately 80% of its normalized volume. Twenty-nine of these are likely to originate in the vesicles, and 18 are likely to originate in the cytosol of the glands (the latter class being present due to holocrine secretion); one is unknown. The vesicular proteins were significantly more enriched than the cytosolic proteins in the released material when compared with the larval extract. A novel metalloproteinase (termed SmPepM8) was the second most abundant constituent after three isoforms of cercarial elastase. In addition, a dipeptidyl peptidase IV (SmDPP IV) was discovered but in much smaller quantity. A new serine protease inhibitor (SmSerp_c) was also prominent. Along with Sm16, four potential immunomodulators were identified, three with similarity to venom allergens (SmSCP_a, _b, and _c) and one with homology to the potassium channel blockers in scorpion venom (SmKK7). Interrogation of the expressed sequence tag database found transcripts encoding the majority of vesicular proteins present solely in the intramolluscan stages of the life cycle. Distinct patterns of radiolabel incorporation suggested three separate origins for the vesicular proteins. All the novel constituents merit investigation as vaccine candidates, and the potential immunomodulators merit investigation as therapeutic agents. *Molecular & Cellular Proteomics* 5:835–844, 2006.

Schistosomiasis is an important parasitic disease in many parts of the tropics, second only to malaria as a source of morbidity (1, 2) and mortality (2). *Schistosoma mansoni*, one of the three principal species of blood flukes infecting humans, inhabits the blood vessels of the hepatic portal vasculature. Infection occurs when waterborne cercaria larvae penetrate the skin after emergence from a snail intermediate host. The larvae must then migrate via the vasculature to the portal system where the worms mate and the females deposit eggs in the blood vessels of the gut wall. The eggs pass to the gut lumen and continue the life cycle or travel downstream to lodge in the liver where they initiate the pathology that is the hallmark of the disease.

The newly emerged cercaria possesses pre- and postacetabular glands (3) and a head gland (4), which are implicated in the invasion process. Synthesis of gland contents occurs during cercarial development from a germ ball in the daughter sporocyst located in the snail’s hepatopancreas, so that all materials are in place to effect rapid host entry upon their emergence from the snail. Contact with host skin stimulates secretion from the glands (5), the contents of which provide adhesion to (6), and proteolysis of (7), the outer layers of the skin. Simultaneously the cercarial tail detaches, initiating transformation to the schistosomulum larva, a process that involves shedding of the cercarial surface membranes and their replacement with preformed material from subtegumental cell bodies (8). Some functional properties of the gland secretions have been characterized, including proteolytic activity (e.g. Ref. 9) and immunomodulatory capacity (e.g. Ref. 10). Furthermore a small number of constituents have been cloned, namely several isoforms of a serine protease, termed cercarial elastase (SmCE) (e.g. Ref. 11), and Sm16, a putative anti-inflammatory protein (e.g. Ref. 12).

Cercarial proteins released into the skin should be accessible to the immune system and thus could serve as vaccine candidates. This is an attractive proposition because the skin is the parasite’s first point of contact with the host. Proteomics, which permits characterization of complex mixtures of
 proteins, provides an ideal approach to characterize the larval secretions. It has been applied recently to S. mansoni cercariae (13), identifying the SmCE isoforms, proteins associated with calcium function, and cytosolic proteins such as heat shock protein chaperones. We report here our detailed analysis of cercarial secretions, using 2-DE,^1 MS/MS, and biosynthetic labeling, that significantly extends the findings of the recent study. The generation of a large expressed sequence tag (EST) database for S. mansoni, including 17,810 ESTs derived from gemballs (14), together with the release of the draft genome sequence (available at www.schistodb.org) has assisted greatly in our identification of several novel molecules with putative proteolytic or immunomodulatory properties.

**EXPERIMENTAL PROCEDURES**

Processing of Cercariae and Their Secretions—A Puerto Rican isolate of S. mansoni was maintained by routine passage through albino Biomphalaria glabrata snails and MF1 outbred mice. Shedding of cercariae from snails was stimulated by exposure to bright light and spot volume normalization (each spot being expressed as a normalizing factor). For each sample, cercariae were released from gemballs with putative proteolytic or immunomodulatory properties.

**FUNCTIONS OF PROTEINS**

The abbreviations used are: 2-DE, two-dimensional electrophoresis; RP, released protein; aldolase, fructose-1,6-bisphosphate aldolase; ATP:GK, ATP:guanidine kinase; CE, cercarial elastase; DPP IV, dipeptidyl peptidase IV; EST, expressed sequence tag; FaBP, fatty acid-binding protein or Sm14; GST28, 28-kDa GST; NCBI, National Center for Biotechnology Information; NV, normalized volume; S.A., specific activity; SCAP, soluble cercarial proteins; SCP, sperm-coating protein; Sm, S. mansoni; TPI, triose-phosphate isomerase; bis-Tris, 2-[bis(2-hydroxyethylamino)-2-(hydroxymethyl)propane-1,3-diol.

^1 The abbreviations used are: 2-DE, two-dimensional electrophoresis; RP, released protein; aldolase, fructose-1,6-bisphosphate aldolase; ATP:GK, ATP:guanidine kinase; CE, cercarial elastase; DPP IV, dipeptidyl peptidase IV; EST, expressed sequence tag; FaBP, fatty acid-binding protein or Sm14; GST28, 28-kDa GST; NCBI, National Center for Biotechnology Information; NV, normalized volume; S.A., specific activity; SCAP, soluble cercarial proteins; SCP, sperm-coating protein; Sm, S. mansoni; TPI, triose-phosphate isomerase; bis-Tris, 2-[bis(2-hydroxyethylamino)-2-(hydroxymethyl)propane-1,3-diol.

in normalized volume (NV) between matched spots to be calculated.

In-gel Digestion and Mass Spectrometric Analysis—Spots chosen for analysis were manually excised from BioSafe Cooamassie- (Bio-Rad) or MS-compatible silver (17)-stained 2-DE gels containing 250–500 μg of 0–3-h RP. Following in-gel digestion with trypsin (0.2 μg in 40 μl of 20 mM ammonium bicarbonate at 37 °C for 18 h), 2 × 1 μl of eluted peptides were sequentially spotted onto a MALDI target plate, air-dried, and followed by 0.6 μl of matrix (a saturated solution of α-cyanohydroxyquinolinic acid (Sigma) in 50% acetonitrile, 0.1% TFA diluted to half-strength with the same solvent). Where necessary remaining peptides were desalted and concentrated using a C18 ZipTip (Millipore). A 4700 Proteomics Analyzer with TOF-TOF Optics (Applied Biosystems, Framingham, MA) calibrated using CalMix2 (Applied Biosystems) was used to obtain MS and MS/MS spectra in positive reflector mode. For each target spot, a mass spectrum was collected between m/z 700 and 4000; where present, two trypsin autolysis peaks were used as an additional, internal calibrant. The 15 most intense peaks from each spectra were automatically selected for MS/MS, ignoring masses in the “exclusion list” (e.g. common trypsin autolysis peaks and keratin contamination (18)). Peptide CID was performed at a collision energy of 1 keV with air as the collision gas.

Database Searching—Matches to the MS/MS data were retrieved using GPS Explorer software (Applied Biosystems) underpinned by Mascot version 1.9 (www.matrixscience.com). The National Center for Biotechnology Information (NCBI) non-redundant protein database was searched first, and then a compilation of clusters, termed “SchistoCDS,” was searched. The latter comprises all publicly available S. mansoni ESTs (at NCBI dbEST and The Welcome Trust Sanger Institute, ftp.sanger.ac.uk/pub/pathogens/Schistosoma/mansonii) and gene sequences predicted from the assembled schistosome genome using three different heuristics (Snap, PHAT, and GlimmerHMM at www.schistodb.org). Search parameters allowed for one missed tryptic cleavage site, the carbamidomethylation of cysteine, and the possible oxidation of methionine; precursor and product ion mass error tolerance was ±0.1 Da. A protein was considered positively identified if the GPS total ion score confidence interval exceeded 99.9%. Identified proteins were classified as vesicular in origin when their full-length sequences possessed a signal peptide as determined by SignalP (19) or by inference from their nearest homolog as determined by BLAST. PSORT II (20) was used to determine whether proteins contained any transmembrane domains. Where possible, the pattern of life cycle stage expression for each identified protein was determined by reference to the ESTs in SchistoCDS. Where matches were made to the NCBI database, BLAST analysis was performed to identify the relevant EST cluster in SchistoCDS.

Extension of Partial Gene Sequences Using Genomic Data—Where possible, when MS data retrieved hits to partial EST sequences or incomplete gene predictions, the coding sequence of the gene was extended by manual inspection of the genome (www.schistodb.org). The genomic contig was retrieved and edited in Artemis (21). Briefly where multiple partial sequences were found, these were concatenated, ensuring that the intron splice donor and acceptor sequences surrounding the exons conformed to those expected in S. mansoni. The putative coding sequence was then searched against the NCBI non-redundant database using BLAST. Where gaps were indicated in the query, the sequence retrieved by BLAST was itself searched against the genomic assembly to pinpoint any missing exons, which were then added. This process was repeated until no more exons could be found without disrupting the correct intron splice donor and acceptor sites and the single open reading frame.

**Protein Fragment Analysis**—To decide whether the protein fragments identified in the 0–3-h RP were potentially due to the action of SmCE-1a and -1b, the full amino acid sequence of each fragment
protein was assessed for the presence of “PL,” the preferred residues N-terminal to the scissile bond (positions P1 and P2 counting outward (11)). The program Compute pI/Mw (available at www.expasy.ch) was used to predict the size and pI of theoretical polypeptide fragments resulting from cleavage at this point. To assess whether each potential cleavage site was likely to be accessible to SmCE-1a in the folded protein, the S. mansoni amino acid sequence was modeled onto the three-dimensional structure of the most similar full-length protein available using Cn3D 4.1 software (available from www.ncbi.nlm.nih.gov/Structure/CN3D/cn3d.shtml). Uncomplexed rabbit actin (Protein Data Bank code 1J6Z) was used for modeling S. mansoni actin 2 (UniProt accession number P53471), the mature form of Leishmania major leishmanolysin (Protein Data Bank code 1LML) was used for SmPepM8, and creatine kinase from Limulus polyphemus (Protein Data Bank code 1RL9) was used for ATP:guanidine kinase (ATP:GK) (UniProt accession number P16641).

Metabolic Labeling of 0–3-h RPs—To identify proteins synthesized in the developing sporocyst, infected snails were exposed to radiolabeled amino acids (22) using 1000 Ci/mmol [35S]Cys/Met (Pro-mix L™, Amersham Biosciences) instead of 75Se-methionine. Three, 5, 7, and 9 days after labeling, 0–3-h RPs were collected as described above, and the processed sample was counted in a liquid scintillation counter (PerkinElmer Life Sciences) to determine cpm/μg of protein.

The incorporation of radiolabel per parasite was assessed by counting 10 aliquots of 10 cercariae with background contributions from radioactivity in the water (e.g. from labeled snail mucous) subtracted.

Detection of Metabolically Labeled Proteins and Calculation of “Specific Activity”—Labeled 0–3-h RPs were separated by 2-DE minigel; the preparation and method was as for the large gels except a 125-μl final volume was applied to each 7-cm IPG-strip (Amersham Biosciences), and an approximate total of 14-kV-h focusing was achieved on an IPGphor (Amersham Biosciences). Second dimension separation used NuPage Novex 4–12% bis-Tris ZOOM gels and MES running buffer (Invitrogen). Minigels were stained with SYPRO Ruby, and the images were captured as described above before drying onto cellulose film (Perbio Science UK Ltd., Northumberland, UK). Dried gels were exposed to Fuji storage phosphor screens (Raytek Scientific Ltd., Sheffield, UK) for 3 days before images were captured using a Molecular Imager FX. Minigel images were analyzed using the Evolution software with the volume of each spot expressed as a sum of the pixel intensities within the spot boundary after “mode of non-spot” background subtraction. This value was used to represent the “amount of protein” in a given spot. Storage phosphorimages were manipulated and quantified using QuantityOne software (Bio-Rad) to determine the “amount of radioactivity” associated with each spot expressed as the sum of pixel values in a designated area following “lowest on boundary” background subtraction. These values were used to calculate the specific activity (S.A.) of selected spots.

RESULTS

Proteins of a Vesicular Origin Are Highly Enriched in the 0–3-h RPs—Analysis of the three replicate 0–3-h RP preparations produced an average gel image with a relatively simple pattern of 144 spots (Fig. 1). The 50 most abundant spots, based on their NV and accounting for 79% of the total protein on the gel, were selected for MS analysis (specimen spectrum, Supplemental Fig. 1). The 44 identities obtained represented 16 separate proteins, seven of which were present as two or more spots (Table I). SignalP classified seven proteins as vesicular in origin, but none contained transmembrane domains on the basis of PSORT II analysis, which also designated the remaining nine proteins as cytosolic in origin. The vesicular proteins accounted for 54% of the normalized spot volume (or 69% of the selected spot volume) with two proteases alone, SmCE and SmPepM8 (see below for details), comprising 34.4 and 12.8% NV, respectively. The remaining
| Spot no. | Normalized volume | Putative ID\(^a\) | NCBI accession no. | Peptide count\(^b\) | BLAST expect score | Species matched\(^c\) | Origin\(^d\) | -Fold increase\(^e\) | Expression pattern\(^f\) |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 8.467 | SmCE | AAC46967 | 6 | 0.00E–109 | Sm | V | 14.9 | G |
| 2 | 5.459 | SmCE | AAM43943 | 1 | 1.00E–109 | Sh | V | 5.5 | G |
| 3 | 5.316 | SmCE | AAC46967 | 4 | 0 | Sm | V | 6.0 | G |
| 4 | 4.237 | SmPepM8 | Snap13515/6/8 | 2/3/2 (NS) | 6.00E–11 | Ag | V | 9.0 | —\(^g\) |
| 5 | 3.855 | Not identified | — | — | — | — | — | — | — |
| 6 | 3.334 | SmSerpin_c | Snap06756 | 4 | 3.00E–11 | Sj | C | 12.3 | — |
| 7 | 3.214 | SmCE | Sm09202 | 3 | 1.00E–131 | Sm | V | 7.8 | G |
| 8 | 2.759 | Sm16 | AAF75550 | 3 | 0 | Sm | V | 2.4 | G |
| 9 | 2.741 | Cyclophilin_1 | Q26565 | 1 | 0 | Sm | C | 10.7 | E Mi G C L A Ma F |
| 10 | 2.665 | SmCE | AAM43943 | 1 | 1.00E–109 | Sh | V | 4.4 | G |
| 11 | 2.3 | SmPepM8 | Snap13515/6/8 | 2/3/3 (NS) | 4.00E–12 | Hs | V | 10.6 | — |
| 12 | 2.283 | SmPepM8 | Snap13515/6/8 | 3/2/1 (NS) | 5E–11 | Ag | V | 6.7 | — |
| 13 | 2.208 | SmCE | AAM43943 | 1 | 1.00E–109 | Sh | V | 2.8 | G |
| 14 | 1.915 | SmSCP_a | Snap09319 | 1 | 2.00E–20 | Sj | V | 25.4 | G |
| 15 | 1.567 | SmPepM8 | Snap13515/6/8 | 3/2/2 (NS) | 6E–11 | Ag | V | 15.4 | — |
| 16 | 1.559 | SmSerpin_c | Snap06756 | 4 | 3.00E–11 | Sj | C | 9.9 | — |
| 17 | 1.478 | GST28 | P09792 | 8 | 0 | Sm | C | 2.3 | E Mi G C L A Ma F |
| 18 | 1.444 | Not identified | — | — | — | — | — | — | — |
| 19 | 1.23 | Thioredoxin | AAL79841 | 3 | 0 | Sm | C | 1.8 | G C L |
| 20 | 1.208 | SmCE | AAM43943 | 1 | 1.00E–109 | Sh | V | 5.1 | — |
| 21 | 1.204 | SmCE | AAM43943 | 2 | 1.00E–109 | Sh | V | 15.6 | — |
| 22 | 0.939 | SmCE | AAM43943 | 2 | 1.00E–109 | Sh | V | 4.2 | G |
| 23 | 0.88 | SmCE | AAM43943 | 2 | 0 | Sm | V | — | — |
| 24 | 0.87 | Actin | AAD38855 | 4 | 0 | Sb | C | — | E Mi G C L A Ma F |
| 25 | 0.849 | SmPepM8 | Snap13515/6/8 | 3/2/2 | 5.00E–11 | Ag | V | 4.3 | — |
| 26 | 0.834 | Not identified | — | — | — | — | — | — | — |
| 27 | 0.834 | SmCE | A28942 | 4 | 0 | Sm | V | — | — |
| 28 | 0.804 | GST28 | P09792 | 5 | 0 | Sm | C | 1.5 | E Mi G C L A Ma F |
| 29 | 0.799 | SmPepM8 | Snap13518 | 2 | 4.00E–12 | Hs | V | — | — |
| 30 | 0.776 | Immunophillin | Snap20098 | 1 | 7.00E–26 | Sm | C | — | — |
| 31 | 0.772 | SmCE | A28942 | 4 | 0 | Sm | V | 57.2 | G |
| 32 | 0.761 | SmPepM8 | Snap13515/6/8 | 1 (NS)/1/1 | 4E–12 | Hs | V | 22.9 | — |
| 33 | 0.717 | SmCE | Sm09202 | 3 | 1.00E–131 | Sm | V | — | — |
| 34 | 0.668 | Cyclophilin_2 | Q26551 | 2 | 0 | Sm | C | 3.1 | Mi G C L A |
| 35 | 0.621 | Sm16 | AAF75550 | 1 | 0 | Sm | V | 7.0 | G |
| 36 | 0.611 | SmSCP_b | Snap11344 | 1 | 0.009 | Rn | V | — | — |
| 37 | 0.595 | Aldolase | AAA57567 | 4 | 0 | Sm | C | — | E Mi G C L A Ma F |
| 38 | 0.593 | TPI | A38233 | 3 | 0 | Sm | C | 1.0 | E G C L A M F |
| 39 | 0.558 | SmKK7 | Sm12916 | 1 | 0.19 | Bm | V | — | C L |
| 40 | 0.546 | SmCE | AAC46967 | 1 | 0 | Sm | V | — | G |
| 41 | 0.524 | Cyclophilin_1 | Q26565 | 1 | 0 | Sm | C | 1.2 | E Mi G C L A Ma F |
| 42 | 0.502 | Cyclophilin_1 | Q26565 | 3 | 0 | Sm | C | 6.8 | E Mi G C L A Ma F |
| 43 | 0.499 | Cyclophilin_1 | Q26565 | 1 | 0 | Sm | C | 4.4 | E Mi G C L A Ma F |
| 44 | 0.484 | Not identified | — | — | — | — | — | — | — |
| 45 | 0.459 | FaBP | AAL15461 | 3 | 0 | Sm | C | 1.3 | E Mi G C L A Ma F |
| 46 | 0.436 | SmSCP_c | Snap04450 | 3 | 3.00E–04 | Sj | V | — | — |
| 47 | 0.408 | Actin | AAC78682 | 4 | 0 | Pm | C | −1.9 | E Mi G C L A Ma F |
| 48 | 0.402 | Not identified | — | — | — | — | — | 1.9 | — |
| 49 | 0.397 | Not identified | — | — | — | — | — | — | — |
| 50 | 0.382 | GST28 | P09792 | 6 | 0 | Sm | C | 1.3 | E Mi G C L A Ma F |
| 51 | 0.378 | ATP:GK | P16641 | 6 | 0 | Sm | C | — | G L Ma |
| 52 | 0.262 | SmCE | Sm09618 | 6 | 0 | Sm | V | — | G |
| 53 | 0.218 | SmDPP IV | Sm01575/11230 | 2/3 | 3.00E–25 | Mrn | V | — | E L/E L |
| 54 | 0.131 | Unknown | Snap25393 | 3 | — | None | U | 4.4 | — |

\(^a\) Note that at this stage of analysis it was impossible to determine which SmCE isoform was identified for each spot, and hence all are termed SmCE. Sequences for two different, previously characterized cyclophilins were matched; they have been arbitrarily designated 1 and 2 to aid distinction between them.
proteins with vesicular locations are all potential immuno-modulators. Three novel proteins, each containing a sperm-coating protein (SCP) domain (pfam accession number PF00188) and with homology to venom allergens, were identified. These were designated SmSCP_a, _b, and _c and represented 1.91, 0.61, and 0.44% NV, respectively. A further novel protein (0.56% NV) had homology to potassium channel inhibitors found in scorpion venom (named SmKK7 after its nearest homolog in Mesobuthus martensii, BmKK7). The final vesicular protein is the well characterized, putative anti-inflammatory protein Sm16 (3.38% NV). Of the cytosolic proteins identified (25% NV in total), a novel serpin (named SmSerpin_c as two serpins are already sequenced from S. mansoni; 4.89% NV) and several cyclophilins were dominant, whereas the remainder (actin, 28-kDa GST (GST28), fatty acid-binding protein (FaBP), triose-phosphate isomerase (TPI), aldolase, and thioredoxin) are ubiquitous throughout the schistosome life cycle.

A separation of SCAP (of which 0–3-h RPs should be a subset) revealed a much more complex pattern of 1314 spots (Supplemental Fig. 2). Software comparison of the two images showed 60% of the 0–3-h RP spots matched to those in the SCAP (data not shown). Normalized volumes for matched spots were obtained from each gel, and the -fold difference in the 0–3-h RPs, relative to the SCAP, was calculated (Table I). Due to the lower complexity of the 0–3-h RPs, all but one of the normalized spot volumes (actin, spot 47) were higher in this preparation than in the SCAP, but the -fold increase of vesicular protein spots (mean ± S.E., 11.95x ± 2.93) was significantly greater than that of the cytosolic protein spots (mean ± S.E., 4.18x ± 1.21; two-sample t test with pooled variances, t = 2.45, adjusted degrees of freedom = 24, p = 0.02). There was a small number of exceptions within the cytosolic group, the two SmSerpin_c spots (spots 6 and 16) and a cyclophilin_1 spot (spot 9) whose -fold increase approached that of the vesicular proteins. Of the unidentified proteins with unknown origin, two fell into the vesicular category and one fell into the cytosolic category on the basis of their -fold increase. Four spots, not among the top 50, were highly enriched in the 0–3-h RP sample and were therefore subjected to MS analysis. Two proved to be of vesicular origin, one an apparent high molecular weight version of SmCE (spot 68) and the other (spot 74) homologous to dipeptidyl peptidase IV (see below for details). One cytosolic protein (spot 57, ATP:GK) was found at an unexpected gel position, and a hit (spot 97) was made to an S. mansoni EST with no homology.

When the expression profile for each protein was determined by reference to SchistoCDS, two clear patterns were evident (Table I). Proteins of vesicular origin had a restricted expression profile (e.g. SmCE, Sm16, and SmSCP_a), whereas those of cytosolic origin were present in virtually all stages (e.g. cyclophilins, actin, aldolase, FaBP, ATP:GK, and TPI). The cytosolic protein thioredoxin showed a more restricted expression pattern, whereas SmKK7 and SmDPP IV were detected in only two life cycle stages. Expression profiling was not possible where identification was made only to gene predictions from the genome.

Further Characterization of Novel Proteins—MS data were searched against predictions of coding regions derived from the S. mansoni genomic assembly. Of the seven spots designated SmPepM8, six hit three separate Snap predictions, each with homology to proteins with Peptidase_M8 domains (PF01547). Manual inspection of the genomic data identified 16 putative exons (Fig. 2A) with the full-length sequence containing an entire Peptidase_M8 domain and a perfect metalloprotease metal-binding site; the eight distinct peptide hits were distributed throughout the sequence (Fig. 2B). A BLAST search with the full-length sequence gave an expect score of 2.00E–74 to an Anopheles gambiae sequence (NCBI accession number XP_311944) that also contains the Peptidase_M8 domain and an expect score of 9.00E–24 to the Leishmania donovani leishmanolysin (gp63) surface metalloprotease (A45621). PSORT II failed to detect a transmembrane domain at the C terminus of the predicted schistosome molecule. The inclusion of an EST contig (Sm09573) in the extended SmPepM8 sequence enabled us to determine that transcripts were present only in the germball stage.

Two EST contigs were hit for spot 74, both having homology to dipeptidyl peptidase IV; thus the protein was termed SmDPP IV. Manual inspection of the genomic data showed that they are arranged in tandem with no missing exons. The exon pattern was largely replicated by data from three independent Snap predictions (Fig. 2C). A molecular mass of

---

b The number of fragmentation spectra that matched each identity is designated by the peptide count. Where more than one sequence was hit and subsequently shown to be part of the same gene, the accession number and peptide coverage for all hits are shown. NS, the Mascot score for the peptides matching this sequence was not significant.

c Species abbreviations are as follows: Ag, A. gambiae; Bm, M. martensii; Hs, Homo sapiens; Mm, Mus musculus; Pm, Penaeus monodon; Rn, Rattus norvegicus; Sb, Suillus bovinus; Sh, S. haematobium; Sj, S. japonicum; Sm, S. mansoni.

d Origin abbreviations are as follows: V, vesicular; C, cytosolic; U, unknown.

e -Fold increase indicates the difference in normal volume between matched spots from 0–3-h RPs and SCAP.

f Expression pattern abbreviations indicate the life cycle stage from which mRNA was obtained for sequencing: E, egg; Mi, miracidium; G, germball; C, cercaria; L, lung stage schistosomulum; A, adult; Ma, adult male; F, adult female. Letter in bold indicates the stage at which expression was highest on the basis of relative EST abundance. Dashes indicate where the expression pattern could not be determined as identification was made only to gene predictions from the genome.

g —, not determined.
approximately 90 kDa is predicted from the 2-DE gel in agreement with that of the nearest homolog in *Drosophila melanogaster* (AAF49712; BLAST expect score of 2.00E−63), but the inferred coding sequence predicts a protein of only 67 kDa (Fig. 2D), and its N terminus could not be obtained. Nevertheless two domains, dipeptidyl peptidase IV N-terminal region (PF00930) and Peptidase_S9 (PF00326) were evident in common with the *Drosophila* protein.

It is probable that the complete coding regions of SmSerpin_c and SmKK7 were present in the EST contigs hit by MS data. This is borne out by the presence of the entire Serpin domain (PF00079) in Snap prediction 06756, whereas the theoretical pI (9.32) and molecular mass (44 kDa) of the coding region correspond to those observed on the 2-DE gel (Fig. 1). No specific domain is available to aid further characterization of SmKK7, but comparison with its nearest homologs shows that features of those molecules are conserved (Fig. 2E). It was not possible to perform the gene reconstruction procedure for the three SCP domain-containing proteins using the current, incomplete version of the *S. mansoni* genome assembly.

Three SmCE Isoforms Are Present in the 0–3-h RPs—Fourteen gel spots in the 0–3-h RPs were identified as SmCE
and appeared at the expected molecular weight, although only five genes encoding schistosome cercarial SmCEs have been described to date (Supplemental Fig. 3A). To discover whether individual spots corresponded to separate isoenzymes, we determined the theoretical masses of diagnostic tryptic peptides for the five predicted amino acid sequences; between 18 and 41 peptides were novel for each. A peptide mass fingerprint produced for each of the 14 spots revealed the presence of between two and 10 masses ostensibly corresponding to selected diagnostic peptides from all five isoenzymes. However, fragmentation of these peptides and Mascot searches with the resulting data confirmed the presence of only SmCE-1a, -1b, and -2a (Supplemental Fig. 3B).

**Fragments of Proteins in the 0–3-h RPs May Result from SmCE Action—** Three 0–3-h RP spots, identified as actin, SmPepM8, and ATP:GK, were present in gel positions that were not consistent with their predicted pl and below their expected molecular mass. However, parasites were cultured in the absence of protease inhibitors (to preserve viability) the anomalous spots could result from proteolytic cleavage. We located PL residues, the cut site for SmCE, in single PL site would give fragments of molecular mass 29.5 kDa, pI 5.5 and molecular mass 12.5 kDa, pI 5.05 (Fig. 3), the theoretical masses and isoelectric points of the fragments that would result from cleavage. Digestion of actin at the single PL cut site (enolase).

**DISCUSSION**

We undertook an in-depth analysis of the proteins released during transformation of *S. mansoni* cercariae into schistosomula. We are confident that the 0–3-h RP preparation, generated by mechanical induction of transformation, faithfully represents the gland contents rather than proteins leaking from dying and dead parasites. Viable larvae are produced, 95% of which develop to the lung stage in vitro and are capable of normal migration and maturation if surgically transplanted into mice (23). This contrasts with an alternative method involving exposure of cercariae to skin lipids in water planted into mice (23). This contrasts with an alternative method involving exposure of cercariae to skin lipids in water that leads to very high mortality levels (up to 100% (24)) with an inevitable artifactual contribution of somatic proteins to the preparation. It may explain the marked differences in the findings reported here with those of the recent “shotgun” proteomics study by Knudsen *et al.* (13). In addition, we also searched the *S. mansoni* genome database, and many of our identifications are derived from that rather than EST databases. Significant enrichment for vesicular proteins was achieved in our 0–3-h RPs, including the known SmCE isoforms and Sm16, compared with a soluble protein extract of cercariae.

The function of the 0–3-h RPs is to facilitate host entry, and there are numerous accounts of protease activity in the cercarial secretions (11). The SmCE variants were the most abundant proteins with three of the five isoforms (SmCE-1a, -1b, and -2a) identified as in previous investigations (11, 13). We did not find cysteine proteases, reported as present on the basis of immunocytochemical localization (25), but we added two further proteases, SmPepM8 and SmDPP IV. SmPepM8, a metalloprotease of the leishmanolysin family, was the second most abundant constituent, a surprising finding because proteases of this class have only been detected in older schistosomula (26). Membrane-bound leishmanolysin enhances passage of the protozoan parasite *Leishmania* across the skin extracellular matrix and is also capable of degrading basement membrane components (27). These structures are encountered by *S. mansoni* larvae traversing the skin, so a
similar role for SmPepM8 seems likely. Although we found no C-terminal membrane anchor in its putative full-length sequence, secreted forms of homologous leishmanolysin play a role in host invasion by *Leishmania* (28). The second protease DPP IV is primarily a membrane-bound protease (29), but no transmembrane domain was found in the *S. mansoni* version. However, an anchoring domain could be missing from our putative sequence that lacks the N-terminal region. The functions of DPP IV in other species imply that it may be involved in skin invasion by *S. mansoni*. Thus, the mammalian version, located within "invadopodia," assists cell migration through connective tissue during wound repair (30) and cancer metastasis (31), whereas a secreted form of DPP IV in *Aspergillus fumigatus* has a potential role in lung invasion by this fungus (32). The identification of these two proteases suggests that elastases are not the sole agents used by cercariae to invade the skin.

Proteins with immunomodulatory potential were the second most common class identified. The 0–3-h RPs have been demonstrated to drive Th2 responses via dendritic cells and the T-cells they polarize (33), but the molecules responsible are unknown. The three proteins containing SCP domains with their similarity to venom allergens (34) could be involved, but their function in any organism has yet to be elucidated. The crystal structure of one such protein from *Necator americanus* shows it bears similarities to chemokines, perhaps explaining its immunomodulatory potential (35). SmKK7, another protein identified, has homology to a scorpion toxin that binds to and blocks K^+^ channels; the activity of the scorpion toxin is dependent on the position of several key amino acids within a three-dimensional structure that is stabilized by three disulfide bridges (36). The level of conservation between these important residues in the schistosome and scorpion proteins suggests a similar function for the schistosome molecule. Because T-cell activation is known to be regulated by K^+^ channels (37), it is plausible that the invading schistosome modulates the immune response of the host by this mechanism. It is notable that some SCP domain-containing proteins may also block voltage-gated K^+^ channels (38).

Eleven cytosolic proteins were identified in the 0–3-h RPs, probably the result of holocrine secretion by the cercarial acetabular glands in which the entire cell contents are forced though the ducts by muscular contraction. Most of the gland cell is occupied by vesicles with cytoplasm confined to the periphery (6), explaining the low abundance of cytosolic proteins compared with those in the whole cercarial extract, with two exceptions. The first, SmSerpin_c serine protease inhib-

---

**Fig. 4.** A, incorporation of [[35]S]Cys/Met into intact parasites (cpm/cercaria) and into 0–3-h RPs (cpm/μg of protein). B, two-dimensional separation of 0–3-h RPs metabolically labeled “in-snail” 3–9 days previously visualized by SYPRO Ruby and phosphorimaging. Gel spots are designated as follows: 1, SmPep8; 2, SmSerpin_c; 3, TPI; 4, SmSCP_b; 5, Sm16; 6, SmSCP_c; 7, actin; 8, enolase; 9, SmCE. C, specific activity (amount of radioactivity per unit of protein in a spot) was calculated at the time points after labeling using data acquired from analysis of the images in A as amount of radioactivity/amount of protein (units are arbitrary for both terms).
itor, is unlikely to be needed for protection of the gland cells because its potential SmCE targets are expressed as inactive proenzymes, but it could limit damage to the host or parasite upon release. The relative abundance of the second protein, cyclophilin, may be explained by a high demand for its peptidyl-prolyl cis-trans isomerase activity during vesicle content synthesis. Although GST28 was not enriched in the gland cells compared with the whole cercarial extract, its proposed prostaglandin-D₂ synthase activity (39) provides a biochemical basis for the observed eicosanoid production during skin penetration (40).

Three broad patterns of 0–3-h RP synthesis could be categorized after in-snail pulse labeling with [³⁵S]Cys/Met. The height of each specific activity peak is dependent on the rate of protein synthesis during the short labeling period with a single dominant peak indicating a narrow window. Vesicular proteins, packaged soon after translation, would not be expected to turn over and so should retain a high S.A., whereas this is less likely for the cytosolic proteins. Because proteins within a vesicle have been synthesized at the same time, the three patterns of incorporation could reflect distinct cellular compartments, perhaps corresponding to the pre- and postacetabular glands and the head gland. Alternatively they could represent three vesicle types within a given gland (6, 41). Only immunolocalization of the various products we identified will answer this question.

We believe that our study highlighted the principle proteins used by the parasite to enter and establish itself in the host. Many of the vesicular proteins are glycosylated, and we will report on the composition of the glycan residues in a separate publication. Our next major task is to clone and express the various proteins to determine their functional properties in vitro and in vivo. This information may provide both potential candidates for a S. mansoni vaccine and novel immunomodulatory compounds with therapeutic value.

Acknowledgment—We thank Dr. Patricia Coulson for critical reading of the manuscript and assistance with the resubmission.

* This work was supported by the Wellcome Trust (to R. S. C. and S. S.) and the Biotechnology and Biological Sciences Research Council (to P. D. A.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

† To whom correspondence should be addressed: Dept. of Biology (Area 5), P. O. Box 373, York YO10 5YW, UK. Tel.: 44-1904-328593; Fax: 44-1904-328599; E-mail: rsc105@york.ac.uk.

REFERENCES

1. King, C. H., Dickman, K., and Tisch, D. J. (2005) Reassessment of the cost of chronic helminthic infection: a meta-analysis of disability-related out-
comes in endemic schistosomiasis. Lancet 365, 1561–1569
2. van der Werf, M. J., de Vlas, S. J., Brooker, S., Looman, C. W., Nagelkerke, N. J., Habbema, J. D., and Engels, D. (2003) Quantification of clinical morbidity associated with schistosomiasis infection in sub-Saharan Africa. Acta Trop. 86, 125–139
3. Dorsey, C. H., Cousin, C. E., Lewis, F. A., and Stanton, M. A. (2002) Ultrastructural structure of the Schistosoma mansoni cercaria. Micron 33, 279–323
4. Crabtree, J. E., and Wilson, R. A. (1985) Schistosoma mansoni: an ultrastructural examination of skin migration in the hamster cheek pouch. Parasitology 91, 111–120
5. Haas, W., Diekhoff, D., Koch, K., Schmauss, C., and Loy, C. (1997) Schistosoma mansoni cercariae: stimulation of acetal protein gland secretion is adapted to the chemical composition of mammalian skin. J. Parasitol. 83, 1079–1088
6. Stirewalt, M. A. (1974) Schistosoma mansoni: cercaria to schistosomula. Adv. Parasitol. 12, 115–182
7. McKerron, J. H., Jones, P., Sage, H., and Pino-Heiss, S. (1985) Proteinases from invasive larvae of the trematode parasite Schistosoma mansoni degrade connective-tissue and basement-membrane macromolecules. Biochim. J. 231, 47–51
8. Skelly, P. J., and Shoemaker, C. B. (2000) Induction cues for tegument formation during the transformation of Schistosoma mansoni cercariae. Int. J. Parasitol. 30, 625–631
9. McKerron, J. H., Pino-Heiss, S., Lindquist, R., and Werb, Z. (1985) Purification and characterization of an elastinolytic proteinase secreted by cercariae of Schistosoma mansoni. J. Biol. Chem. 260, 3703–3707
10. Ramaswamy, K., Salafsky, B., Potluri, S., He, Y. X., Li, J. W., and Shibuya, T. (1995) Secretion of an anti-inflammatory, immunomodulatory factor by schistosomula of Schistosoma mansoni. J. Inflamm. 46, 13–22
11. Salter, J. P., Choe, Y., Albrecht, H., Franklin, C., Lim, K. C., Craig, C. S., and McKerron, J. H. (2002) Cercarial elastase is encoded by a functionally conserved gene family across multiple species of schistosomes. J. Biol. Chem. 277, 24618–24624
12. Rao, K. V., and Ramaswamy, K. (2000) Cloning and expression of a gene encoding Sm16, an anti-inflammatory protein from Schistosoma mansoni. Mol. Biochem. Parasitol. 108, 101–108
13. Knudsen, G. M., Medzhradszky, K. F., Lim, K. C., Hansell, E., and McKerron, J. H. (2005) Proteomic analysis of Schistosoma mansoni cercarial secretions. Mol. Cell. Proteomics 4, 1862–1875
14. Verjovski-Almeida, S., DeMarco, R., Martins, E. A., Guimaraes, P. E., Ojopi, E. P., Paquola, A. C., Piazza, J. P., Nishiyama, M. Y., Jr., Kitajima, J. P., Adamson, R. E., Ashton, P. D., Bonaldo, M. F., Coulon, P. S., Dillon, G. P., Farias, L. P., Gregorio, C., Stukart, G. C., Soares, M. B., Gargioni, C., Kawano, T., Duprat, V., Madeira, A. M., Wilson, R. A., Menck, C. F., Setubal, J. C., Leite, L. C., and Dias-Neto, E. (2003) Transcriptome analysis of the acelolome human parasite Schistosoma mansoni. Nat. Genet. 35, 148–157
15. Perinolho-Pinto, F. J., Gazzinelli, R. S., Mota-Santos, T. A., Figureiredo, E. A., and Pellegrino, J. (1974) Schistosoma mansoni: defined system for stepwise transformation of cercaria to schistosomule in vitro. Exp. Parasitol. 36, 360–372
16. Curwen, R. S., Ashton, P. D., Johnston, D. A., and Wilson, R. A. (2004) The Schistosoma mansoni soluble proteome: a comparison across four life-cycle stages. Mol. Biochem. Parasitol. 138, 57–66
17. Yan, J. X., Watt, R., Berkelman, T., Harry, R. A., Westbrook, J. A., Wheeler, C. H., and Dunn, M. J. (2000) A modified silver staining protocol for visualization of proteins compatible with matrix-assisted laser desorption/ionization and electrospray ionization-mass spectrometry. Electrophoresis 21, 3666–3672
18. Ding, Q., Xiao, L., Xiong, S., Jia, Y., Que, H., Guo, Y., and Liu, S. (2003) Unmatched masses in peptide mass fingerprints caused by cross-contamination: an updated statistical result. Proteomics 3, 1313–1317
19. Bennett, J. D., von Heijne, G., and Bairoch, A. (2004) Improved prediction of signal peptides: SignalP 3.0. J. Mol. Biol. 340, 783–795
20. Nakai, K., and Horton, P. (1999) PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. Trends Biochem. Sci. 24, 34–36
21. Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajendream, M. A., and Barrell, B. (2000) Artemis: sequence visualization and anno-

Molecular & Cellular Proteomics 5.5 843
Proteomic Identification of *S. mansoni* Cercarial Secretions

22. Wilson, R. A., and Coulson, P. S. (1986) *Schistosoma mansoni*: dynamics of migration through the vascular system of the mouse. *Parasitology* **92**, 83–100

23. Harrop, R., and Wilson, R. A. (1993) Protein synthesis and release by cultured schistosomula of *Schistosoma mansoni*. *Parasitology* **107**, 265–274

24. McKerrow, J. H., Keene, W. E., Jeong, K. H., and Werb, Z. (1983) Degradation of extracellular matrix by larvae of *Schistosoma mansoni*. *Parasitology* **92**, 83–100

25. Dalton, J. P., Clough, K. A., Jones, M. K., and Brindley, P. J. (1997) The cysteine proteinases of *Schistosoma mansoni*. *Parasitology* **114**, 105–112

26. McKerrow, J. H., Keene, W. E., Jeong, K. H., and Werb, Z. (1983) Degradation of extracellular matrix by larvae of *Schistosoma mansoni*. *Parasitology* **114**, 105–112

27. McGwire, B. S., Chang, K. P., and Engman, D. M. (2003) Migration through the extracellular matrix by the parasitic protozoan *Leishmania* is enhanced by surface metalloprotease gp63. *Infect. Immun.* **71**, 1008–1010

28. McGwire, B. S., O'Connell, W. A., Chang, K. P., and Engman, D. M. (2002) Extracellular release of the glycosylphosphatidylinositol (GPI)-linked *Leishmania* surface metalloprotease, gp63, is independent of GPI phospholipolysis. Implications for parasite virulence. *J. Biol. Chem.* **277**, 8802–8809

29. Engel, M., Hoffmann, T., Wagner, L., Wermann, M., Heiser, U., Kiefersauer, R., Huber, R., Bode, W., Demuth, H. U., and Brandstetter, H. (2003) The crystal structure of dipeptidyl peptidase IV (CD26) reveals its functional regulation and enzymatic mechanism. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 5063–5068

30. Gheri, G., Dong, H., Goldstein, L. A., Yeh, Y., Hakkinen, L., Larjava, H. S., and Chen, W. T. (2002) Regulation of fibroblast migration on collagenous matrix by a cell surface peptidase complex. *J. Biol. Chem.* **277**, 29231–29241

31. Chen, W. T., and Kelly, T. (2003) Seprase complexes in cellular invasiveness. *Cancer Metastasis Rev.* **22**, 259–269

32. Beauvais, A., Monod, M., Wyniger, J., Debeaupuis, J. P., Grouzmann, E., Brakch, N., Svab, J., Hovanessian, A. G., and Latge, J. P. (1997) Dipeptidyl-peptidase IV secreted by *Aspergillus fumigatus*, a fungus pathogenic to humans. *Infect. Immun.* **65**, 3042–3047

33. Jenkins, S. J., and Mountford, A. P. (2005) Dendritic cells activated with products released by schistosome larvae drive Th2-type immune responses, which can be inhibited by manipulation of CD40 costimulation. *Infect. Immun.* **73**, 395–402

34. King, T. P., Moran, D., Wang, D. F., Kochoumian, L., and Chait, B. T. (1990) Structural studies of a hornet venom allergen antigen 5, Dol m V and its sequence similarity with other proteins. *Protein Seq. Data Anal.* **3**, 263–266

35. Asojo, O. A., Goud, G., Dhar, K., Loukas, A., Zhan, B., Deumic, V., Liu, S., Borkstahl, G. E., and Hotez, P. J. (2002) Structural studies of a hornet venom allergen antigen 5, Dol m V and its sequence similarity with other proteins. *Protein Seq. Data Anal.* **3**, 263–266

36. Korolikova, Y. V., Bocharov, E. V., Angelo, K., Maslennikov, I. V., Grinenko, O. V., Lipkin, A. V., Nosyrev, E. D., Pluzhnikov, K. A., Olesen, S. P., Arseniev, A. S., and Grishin, E. V. (2002) New binding site on common molecular scaffold provides HERRG channel specificity of scorpion toxin BeKm-1. *J. Biol. Chem.* **277**, 43104–43109

37. Panty, G., Varga, Z., and Gaspar, R. (2004) Ion channels and lymphocyte activation. *J. Biol. Chem.* **277**, 55–66

38. Yamazaki, Y., Kojima, H., Sugiyama, Y., Motoyoshi, K., Wada, T., Hishinuma, S., Mita, M., and Morita, R. (2002) Cloning and characterization of novel snake venom proteins that block smooth muscle contraction. *Eur. J. Biochem.* **269**, 2708–2715

39. Herve, M., Angel, V., Pinzar, E., Wintjens, R., Faveeuw, C., Narumiya, S., Capron, A., Urate, Y., Capron, M., Riveau, G., and Trottein, F. (2003) Pivotal roles of the parasite PGD2 synthase and of the host D prostanoid receptor 1 in schistosome immune evasion. *Eur. J. Immunol.* **33**, 2764–2772

40. Fusco, A. C., Salafsky, B., and Kevin, M. B. (1985) *Schistosoma mansoni*: eicosanoid production by cercariae. *Exp. Parasitol.* **59**, 44–50

41. Dorsey, C. H. (1976) *Schistosoma mansoni*: description of the head gland of cercariae and schistosomules at the ultrastructural level. *Exp. Parasitol.* **39**, 444–459