2017

H-IPSE is a pathogen-secreted host nucleus infiltrating protein (infiltrin) 3 expressed exclusively by the Schistosoma haematobium egg stage

Luke Pennington

Abdulaziz Alouffi

Evaristus Mbanefo

Debalina Ray

David Heery

See next page for additional authors

APA Citation

Pennington, L., Alouffi, A., Mbanefo, E., Ray, D., Heery, D., Jardetzky, T., Hsieh, M., & Falcone, F. (2017). H-IPSE is a pathogen-secreted host nucleus infiltrating protein (infiltrin) 3 expressed exclusively by the Schistosoma haematobium egg stage. *Infection and Immunity,* (). http://dx.doi.org/10.1128/IAI.00301-17

Authors
Luke Pennington, Abdulaziz Alouffi, Evaristus Mbanefo, Debalina Ray, David Heery, Theodore Jardetzky, Michael Hsieh, and Franco Falcone

TITLE

H-IPSE is a pathogen-secreted host nucleus infiltrating protein (infiltrin) expressed exclusively by the *Schistosoma haematobium* egg stage

RUNNING TITLE

H-IPSE nuclear infiltrin from *Schistosoma haematobium*

AUTHORS

Luke F. Pennington[a$], Abdulaziz Alouffi[b$] Evaristus C. Mbanefo[c], Debalina Ray[d], David M. Heery[e], Theodore S. Jardetzky[a], Michael H. Hsieh[c,f] and Franco H. Falcone[b]

[$] Equal contributors

AFFILIATIONS

[a]Stanford University School of Medicine, Stanford, CA, USA;
[b]School of Pharmacy, Division of Molecular Therapeutics and Formulation, University of Nottingham, Nottingham, UK;
[c] Biomedical Research Institute, Rockville, MD, USA
[d]Department of Pathology, University of California, San Francisco, CA,USA
[e]School of Pharmacy, Division of Biomolecular Science and Medicinal Chemistry, University of Nottingham, Nottingham, UK
[f]Department of Urology, The George Washington University, Washington, D.C., USA

CORRESPONDING AUTHOR

Dr Franco H. Falcone. Mailing address: University of Nottingham, School of Pharmacy, Science Road, Boots Science Building, Nottingham, NG7 2RD, United Kingdom. Tel: 44 115 84 66073, Fax: 44 115 29 951 5102; E-mail: franco.falcone@nottingham.ac.uk

ABSTRACT

Urogenital schistosomiasis, caused by the parasitic trematode *Schistosoma haematobium*, affects over 112 million people worldwide. As with *S. mansoni* infections, the pathology in urogenital schistosomiasis is mainly related to the egg stage, which induces granulomatous inflammation of affected tissues. Schistosoma eggs and their secretions have been studied extensively for the related *S. mansoni* organism which is more amenable to laboratory studies. Indeed, we have shown that IPSE/alpha-1 (M-IPSE herein), a major protein secreted from *S .mansoni* eggs, can infiltrate host cells. Although M-IPSE function is unknown, its ability to translocate to their nucleus and bind DNA suggests a possible role in immune modulation of host cell tissues. Whether IPSE homologs are expressed in other Schistosome species has not been investigated.

Here, we describe the cloning of two paralog genes H03-IPSE and H06-IPSE which are the ortholog of M-IPSE, from the egg-cDNA of *S. haematobium*. Using PCR and immunodetection, we confirmed that expression of these genes is restricted to the egg stage and female adult worms, while H-IPSE protein is only detectable in mature eggs but not adults. We show that both H03-IPSE and H06-IPSE proteins can infiltrate HTB-9 bladder cells when added exogenously to culture medium. Monopartite C-terminal NLS motifs conserved in H03-IPSE 'SKRRRKY' and H06-IPSE 'SKRGRKY' NLS motifs, are responsible for targeting the proteins to the nucleus of HTB-9 cells, as demonstrated by site directed mutagenesis and GFP tagging. Thus, *S. haematobium* eggs express IPSE homologs that appear to perform similar functions in infiltrating host cells.

INTRODUCTION

Schistosomes are digenetic blood trematodes, which rely on their egg stage for transmission to the intermediate host, a water snail (1). In order to reach the aquatic environment, the eggs deposited by adult female worms in the blood vessels of their mammalian host have to cross several layers of host tissue before they can reach the lumen of the gut, or, in the case of *S. haematobium*, the bladder. This is a critical step in the life cycle of the parasite and is therefore very likely to have been fine-tuned to the host's immunological and tissue environment during the course of evolution. The microenvironment of the bladder and the gut are histologically and immunologically quite different, and this may be reflected in differences between the molecules produced by the eggs of *S. haematobium* and *S. mansoni* and the underlying mechanisms leading to translocation of eggs across the tissues. Most proteomic studies have concentrated on the egg stage of the more available trematode, *S. mansoni* (2–4)*;* these have identified three major protein components (3) produced by mature eggs: omega-1 (5), kappa-5 (6) and IPSE/alpha-1 (7).

Here, we show that *S. haematobium* expresses multiple variants of a protein homologous to IPSE/alpha-1 in *S. mansoni* (M-IPSE), which we have called H-IPSE, a term which will be used here to collectively describe the different orthologs of M-IPSE in *S. haematobium*. The mRNA expression of H-IPSE is restricted to the egg and female worm stage, but only translated as protein in eggs. H-IPSE shares an important biological activity described for M-IPSE: the ability to be taken up by and translocate to the nucleus of host cells.

MATERIALS AND METHODS

Cloning of IPSE transcripts. Total RNA was isolated from *S. haematobium* eggs from the liver of infected hamsters using TRizol. Following DNAse treatment and inactivation, cDNA was generated from RNA samples using the superscript III first strand cDNA synthesis kit (Invitrogen) with oligo dT primers, or custom oligo dT primers with a 5' anchoring sequence corresponding to 3' RACE reverse primers (See Table S1 in suppl. data). The resulting cDNA was then amplified with either targeted 5' and 3' primers designed from predicted H-IPSE transcripts, or amplified by a nested 3' RACE PCR with a 5' primer targeted to a conserved 5' elements in IPSE using Platinum Taq Supermix (Invitrogen). Gel purified PCR fragments were then treated with Taq polymerase to facilitate cloning into the pCR 2.1-TOPO vector included with the TOPO TA cloning kit (Invitrogen) and all positive colonies from the blue/white assay were sequenced.

Recombinant protein expression in BL21 Star DE3 *E. coli.* The H06-IPSE protein coding sequence lacking the N-terminal signal sequence was cloned into the pET-100D Topo expression vector (Invitrogen) and transformed into BL21 Star DE3 *E. coli*. 1L cell cultures were grown to an OD of 0.6 under ampicillin selection (100 µg/mL), and were induced with 1mM IPTG for 3 hours before harvest. Cell pellets were then suspended in 50 mL of lysis buffer (6M guanidine hydrochloride, 10 mM imidazole, 20 mM sodium phosphate, and 500 mM NaCl at pH 7.8) and treated with EDTA-free protease inhibitor tablets (Pierce). Cells were lysed with three freeze/ thaw cycles and sonicated on ice. Nickel NTA resin purifications were conducted with a binding buffer containing 8M urea, 10 mM imidazole, and PBS at pH 7.4, a wash buffer consisting of 8M urea, 25 mM imidazole, and PBS at a pH of 7.4, and an

elution buffer containing 8M urea, 300 mM imidazole, and PBS at a pH of 7.4. This solution was successively dialyzed against PBS pH 7.4 solutions containing 4.0 M, 2.0 M, and 1.0 M urea over three days before being dialyzed overnight against PBS (see Fig. S4A in suppl. materials). Refolded protein was concentrated to <0.50 mg/mL in a 3.0 kDa cutoff Centricon centrifugal concentrator (EMD Millipore).

Generation of polyclonal IPSE antibodies. Recombinant bacterial derived H06-IPSE protein (obtained as described above) was used to immunize rabbits 4 times over the course of 8 weeks (ProMab Biotechnologies Inc.) (Fig. S1B in suppl. materials). Antibody was precipitated from sera using ammonium sulfate and suspended in PBS with 0.03% sodium azide prior to use.

Cloning of H-IPSE into pTT5 expression vector. To facilitate mammalian expression, codon optimized synthetic IPSE vectors were generated for H03, H06, and M-IPSE (GeneArt Invitrogen). The IPSE variants were PCR amplified from these synthetic constructs. A second insert containing the human VEGF signal sequence, an 8x His tag, and a TEV cleavage site was also amplified. These two fragments were inserted into an EcoRI/NheI-digested pTTVH8G vector (licensed from the Canadian Research Council (8)) by Gibson assembly. Subsequently, during vector optimization, the full IPSE expression cassette with N-terminal signal sequence, tag, and cleavage site was transferred from the pTTVH8G vector to the pTT5 vector by conventional restriction cloning using EcoRI and NotI. The H06 SKAAAKY NLS mutant was cloned by site-directed mutagenesis of H06-pTT5 vectors using Phusion High Fidelity PCR Master mix (Invitrogen), followed by Dpn1 digestion for 1 hr.

RT-PCR and stage specific expression of H-IPSE. Total RNA was isolated from the following *S. haematobium* life cycle stages (obtained from the NIAID Schistosomiasis Resource Center for distribution through BEI Resources, NIAID,

NIH): purified eggs, retrieved from the liver of *S. haematobium* infected hamsters, miracidia, cercariae, schistosomula, adult females, adult males and mixed sex adult worms using the RNAzol kit (MRC, Ohio), following the manufacturer's instructions. Contaminating genomic DNA was removed by DNAse treatment using TURBO DNase (Invitrogen Ambion, USA) and chemical DNAse inactivation, as per manufacturer's instructions. After removal of genomic DNA contaminants, cDNA was obtained by reverse transcription using iScript cDNA synthesis kit (Bio-Rad, USA) as directed by the manufacturer. After reverse transcription, RNA sample concentrations were measured using a Nanodrop ND-1000 spectrophotometer (ThermoScientific Fisher) and adjusted to 500 ng/µL for all samples with molecular grade water. The PCR was performed on a BioRad CFX Connect thermocycler, using the following cycling conditions:

Initial denaturation step [2 min at 94 °C], followed by 35 cycles of denaturation [30 sec at 94 °C], annealing [45 sec at 56 °C] and extension [1 min at 72 °C], and followed by a final extension [5 min at 72 °C]. The polymerase used was Takara Taq polymerase using 2 µL 10x polymerase buffer, 2 µL 10 µM dNTPs, 1 µL each of forward (5'-GCTCACTCTCACCACCATG-3') and reverse (5'-TCCTTCGACGTTTCGATTCAC-3') primers, 2 µL of cDNA template and 11.5 µL molecular grade water in a 20 µL total volume. PCR reactions were then subjected to electrophoretic separation on 1% agarose containing 0.5 µg/mL ethidium bromide in 0.5x Tris Borate EDTA (TBE) buffer. A 100 bp DNA ladder (Promega, USA) was used for sizing of the amplicon. Gels were imaged using a GelDoc XR Molecular Imager (BioRad, USA) and saved as tif files. The oligonucleotide primers used span an intron, discriminating cDNA (295 bp) from genomic DNA (402 bp).

155 **Recombinant protein expression using HEK 293-6E cells.** H03-IPSE and
156 H06-IPSE were recombinantly expressed for uptake and microscopy experiments
157 using the pTT5 HEK293-6E expression platform (L-11565) licensed from the
158 Canadian Research Council (8). A large-scale gene expression workflow was
159 developed using 2L vented shaking flasks. Cells were cultivated in suspension in an
160 incubator at 37 °C in 5% CO_2 humidified atmosphere under constant shaking at a
161 rate of 120 rpm. The medium consisted of Freestyle F17 medium (GIBCO, Rockville,
162 MD) supplemented with 0.1% w/v Kolliphor P-188 (Sigma-Aldrich), with 4 mM L-
163 glutamine and 25 µg/mL G418 (ThermoScientific Fisher). Freestyle 293 medium
164 (GIBCO, Rockville, MD) was used interchangeably with F17 without glutamine
165 supplementation.

166 For transient transfection, 500 mL of cell suspension were mixed with 12.5 mL
167 of medium containing the plasmid DNA and another 12.5 mL of medium for
168 resuspension of linear 25 kdA polyethylenimine (PEI) (PolyPlus). The final DNA
169 amount for each pTT construct was 0.5 mg, and this was with mixed 1.5 mg of PEI
170 (3:1 PEI:DNA ratio) and incubated for 3 minutes at room temperature. The resultant
171 complex was then added to the cells. 2.5 mL of 20% (w/v) Tryptone N1 (TN1,
172 TekniScience Inc., Canada) was added 24 h after transfection.
173
174 H-IPSE wildtype or mutant proteins, secreted into the HEK293-6E serum-free
175 cell culture medium, were harvested 7 days after transfection, followed by protein
176 purification. This supernatant was centrifuged at 2,800 x g for 10 minutes (4°C),
177 followed by 0.22 µm filtration to remove cell debris and aggregates, then purified by
178 immobilized metal affinity chromatography (IMAC) using TALON Superflow cobalt
179 affinity resin (GE Healthcare, Freiburg, Germany) or Ni-NTA Agarose (Qiagen). For

cobalt resin purification the binding buffer consisted of 50 mM sodium phosphate, 300 mM NaCl, pH 7.4; the wash buffer contained 50 mM sodium phosphate, 300 mM NaCl, 5 mM imidazole, pH 7.4, while the elution buffer consisted of 50 mM sodium phosphate, 300 mM NaCl 1, 150 mM imidazole, pH 7.4. Nickel resin purifications were conducted with a binding buffer containing PBS pH 7.4, a wash buffer consisting of PBS and 10mM imidazole at a pH of 7.4, and an elution buffer containing PBS and 300mM imidazole at a pH of 7.4.

SDS-PAGE, Coomassie stain and Western blotting. 15 µL aliquots of the eluted purified protein fractions were separated by SDS-PAGE using Bio-Rad Mini-Protean ready gels (4–20% gradient TGX) in a Mini-Protean electrophoresis cell, as recommended by the manufacturer. The gels were then incubated for half an hour in Instant Blue (Expedeon, Harston, UK) for Coomassie staining, followed by washing in deionised water. For Western Blotting, the gradient TGX gels were transferred to a 0.2 µm nitrocellulose membranes using Trans-Blot® Turbo™ Transfer System, as per the manufacturer's protocol (Bio-Rad). The membranes were blocked by blocking buffer (5% (w/v) dried skimmed milk, 0.01% (v/v) Tween 20 and TBS) with shaking for 1 hour at room temperature. Next, membranes were incubated with the primary mouse anti-His antibody (GE Healthcare) as primary antibody, which was diluted 1:5000, at 4°C overnight, followed by three washes in Tris-buffered saline solution (TBS) containing 1% Tween (T) for 10 minutes each. The membranes were then incubated with anti-mouse IgG (whole molecule), HRP-conjugated antibody (Sigma-Aldrich, UK) as a secondary antibody (1:4000) for one hour at room temperature, followed by washing in the manner described above. Membranes were imaged using a Fuji LAS4000 imager with chemiluminescence-luminol reagent (3 µl

of 30% H_2O_2, Tris/HCl 0.1 mM, pH, 8; 2.5 mM luminol and 400 µM coumaric acid). For detection of native IPSE protein in adult worm extract and egg-derived samples approximately 40 µg of parasite-derived material was loaded per well as determined by each samples' A_{280} using a spectrophotometer. These samples were run on 4-20% ExpressPlus Page gels (Genescript) as directed by manufacturer, and stained with Coomassie Brilliant Blue G250 (BioRad). For western blotting gels were transferred to a 0.22 µm PDVF membrane after preactivation with methanol. The membranes were blocked by blocking buffer (5% (w/v) dried skimmed milk, 0.01% (v/v) Tween-20 and PBS) with shaking for 1 hour at room temperature. Next, blots were incubated with polyclonal anti-H06-IPSE rabbit antibodies at a 1:500 dilution and stained overnight at 4°C. After three washes in PBS-T, blots were then developed with polyclonal HRP conjugated goat anti-rabbit secondary antibody (EMD Millipore) at a dilution of 1:5000. The gels were washed 3 additional time in PBS-T and developed with SignalFire ECL reagent (Cell Signaling).

Cloning of predicted NLS and mutants into pTetra-EGFP. The predicted NLS for each protein was subcloned into the Tetra-EGFP vector, which carries a kanamycin resistance gene (9). This vector encodes four EGFP repeats with a multiple cloning site inserted between the third and fourth EGFP sequence. The nucleotides encoding the predicted NLS in H03-IPSE and H06-IPSE, as well as predicted NLS mutant and the canonical Sv40 NLS, were inserted into pTetra-EGFP using oligonucleotide primers and specific restriction enzymes (see Table S1 in suppl. mat). This leads to a construct, which codes for a tetra-EFGP fusion protein of approximately 113 kDa that due to its large size is completely excluded from the nucleus in the absence of a functional NLS. Initially, 5'-phosphorylated pairs of

matching oligonucleotides were designed to code putative NLSs containing GATC overhangs by using 1 µL of each oligo (100µM) mixed with 98 µL of 10 mM Tris-HCl, 1 mM EDTA, pH 8.0 followed by denaturation at 95°C for 7 minutes, then 3 minutes at 5°C. Next, the double-stranded oligonucleotides and the pTetra-EGFP vector were digested with the restriction enzyme *Bgl*II (New England Biolabs), according to the manufacturer's protocol. The ends of the linearized pTetra-EGFP vector were dephosphorylated with Antarctic Phosphatase (New England BioLabs) to avoid re-ligation to itself, following the manufacturer's instructions. Ligation was performed using 1 µL vector, 3 µL insert, 1 µL 10X Buffer T4-ligase, 4 µL molecular biology grade water and 1 U T4 DNA ligase (Promega) in a 10 µL reaction mixture and incubated overnight at 16°C. DNA sequencing (Source BioScience, UK) using T7 primers determined successful insertion in the correct orientation.

HTB-9 Cell culture and transfection. The human bladder cancer cell line HTB-9 (5637 ATCC) was grown in T75 flasks (Sarstedt, Germany) at 37°C in a humidified 5% CO_2 incubator, with Minimum Essential Medium Eagle (MEM; Sigma-Aldrich) supplemented with 5% heat-inactivated fetal bovine serum (FBS, GIBCO), 2 mM L-glutamine, 100 unit/mL penicillin and 100 µg/mL streptomycin (Sigma-Aldrich, UK). Transient transfections of HTB-9 cells were performed using X-tremeGENE9 DNA transfection reagent (Roche Applied Science, Germany) according to the manufacturer's protocol. Cells were plated onto 5 mg/mL rat-tail collagen I-coated glass cover slips (Invitrogen, UK, 15mm diameter, # 1 thickness) in 6-well plates, and transfected with the different Tetra-EGFP plasmids at 60-70% confluency.

Cell Fixation and Fluorescence Microscopy. One-day after transfection the cells were washed with Dulbecco's phosphate-buffered saline (DPBS, Gibco) and fixed at room temperature for 10-15 min in 4% paraformaldehyde. The cells were then washed three times with DPBS and incubated with 0.5 g/mL Hoechst 33342 stain (Sigma-Aldrich) at room temperature for 8-15 minutes, before washing again three times with DPBS. Slides were mounted with mounting medium (Sigma-Aldrich). The transfected cells were visualized by fluorescence microscopy (EVOS *fl*, Advanced Microscopy Group, USA) or confocal microscopy (LSM510 META, ZEISS, Germany), and analyzed using Zeiss LSM Image Browser software (version 4.2.0.121).

Cellular Uptake of H03-IPSE. HTB9 cells were seeded onto Lab-Tek 8-well chambered cover glass (Nalgene Nunc International) at a density of 5 x 10^5 cells in order to achieve 50-60% confluency after 24 hours. Cells were then incubated with 15-0.40 nM of recombinant proteins (H03-IPSE WT or mut) in serum-free internalization medium (HEPES buffered Ham's F12 medium (SigmaAldrich, UK) containing 10 mM $NaHCO_3$ and 2 mg/mL (bovine serum albumin) (Fraction V Biomol GmbH, Germany), followed by fixation at room temperature for 10-15 min in 4% paraformaldehyde solution. The cells were then washed 4-5 times with DPBS and incubated with 0.5 µg/mL Hoechst 33342 or 5 µM DRAQ5 (Thermo Fisher Scientific) nuclear stain for 15 min and permeabilized with 0.2% Triton X-100 in DPBS for 10 min. The cells were washed 4-5 times with DPBS and incubated separately at room temperature for 30 min with two different primary antibodies diluted 1:5000, either mouse anti-His antibody (GE Healthcare). The cells were then washed thrice and labelled with the secondary antibody, Alexa Fluor 555-conjugated goat anti-mouse

IgG (H+L; Molecular Probes), diluted 1:500 by incubation at room temperature for 30 min, followed by three final washes with DPBS.

RESULTS

Identification of H-IPSE variants. To identify homologs of *S. mansoni* IPSE in the *S. haematobium* genome, we performed a BLAST search at Wormbase.org using the predicted transcript of M-IPSE (Smp_112110.1) (10–12). This analysis identified three paralogs within the *S. haematobium* genome, all with predicted transcripts with high identity to M-IPSE (Gene ID, % amino acid identity: C_00050, 67%: C_00244, 63%: B_00796, 56%). To verify these transcripts we isolated cDNA from *S. haematobium* eggs, and employed two strategies. The first employed 5' and 3' primers designed to amplify transcripts predicted from the C_00244 locus, and the second employed a 3' RACE cloning strategy with a 5' primer targeting highly conserved regions of all H-IPSE variants (Table S1). In total, 14 IPSE transcript sequences were obtained, and 8 Sanger sequencing runs contained data sufficient for unambiguous base calling throughout the ORF`s. These transcripts clustered with two of the three H-IPSE paralogs (Figure 1A), and one transcript from each cluster (H03 and H06) was selected for further study. Sequence variations within these clones suggest that H-IPSE genes are polymorphic (see Figures S1-S3 for supplementary information).

<Figure 1 here>

Alignment of H03-/H06-IPSE with M-IPSE (Fig. 1B) demonstrates conservation of the seven cysteines, known to form three intramolecular disulfide bonds and one intermolecular bond, resulting in a homodimeric structure. Two potential N-linked glycosylation consensus motifs are also present with small variations. A 20 amino acid long N-terminal classical secretory sequence (CSS) is predicted for both H03- and H06-IPSE by SignalP 4.1 (13). To verify the presence of IPSE protein in *S. haematobium* parasite-derived material, we expressed and refolded H06-IPSE from insoluble inclusion bodies, and used this bacterially derived H06-IPSE to generate polyclonal anti-H-IPSE antibodies in rabbits (Figure S1). On western blots, anti H-IPSE antibodies bind a ~40 kDa protein species in both *S. haematobium* egg secreted protein and soluble egg antigens, but not in adult worm extract (Figure 1C). This corresponds to the expected size for dimeric, glycosylated H-IPSE variants.

H03- and H06-IPSE variants have a predicted nuclear localization sequence (NLS). Several algorithms (cNLSMapper (14), pSORTII (15) NLStradamus (16) and NucPred (17)) predict a C-terminal nuclear localization sequence (NLS) in H03-IPSE close to the C-term (data not shown), similarly to the NLS described for M-IPSE (18). Intriguingly, the H06-IPSE paralog carried an R128G variant within the nuclear localization sequence corresponding to the validated NLS (18) in M-IPSE (Fig. 1B). Such variants have also been observed in *S. mansoni* studies. For example ESP3-6, a protein later recognized as an M-IPSE variant, is 97% homologous to the published M-IPSE sequence (GenBank Acc. Nr.: AAK26170.1), but contains an R132L variant within the NLS (ESP3-6 GenBank Acc. Nr.: AF527011). Positively charged amino acids in an NLS are key to its nuclear targeting activity mediated by binding to cytosolic importin-α (19), thus such a

328 replacement will have an impact on the protein's ability to translocate to the nucleus,
329 ranging from less efficient translocation to no translocation at all, depending on the
330 exact position in the NLS (19). Substitutions in the NLS will potentially also have an
331 effect on DNA binding specificity. This is also reflected in the less certain prediction
332 of an NLS in H06-IPSE by cNLSmapper and the other tested algorithms (data not
333 shown).
334
335 **The NLS in H03- and H06-IPSE is able to direct a large fluorescent protein**
336 **to the nucleus of mammalian cells.** Therefore, in order to assess functionality of
337 the predicted NLS in H03-IPSE and the potential impact of the R128G substitution in
338 H06-IPSE, we cloned the oligonucleotides encoding the NLS predicted sequences
339 into the previously described Tetra-EGFP vector (18). The resulting constructs were
340 then transfected into HTB-9 uroepithelial carcinoma cells as a model of host cells
341 relevant to *S. haematobium* infection. The results of the transfection of the Tetra-
342 EGFP vector encoding the NLS sequences of H03-IPSE (SKRRRKY) and H06-IPSE
343 (SKRGRKY), as well as the predicted NLS mutant SAAGAAY (Figure 2) confirm that
344 the H03-IPSE NLS is fully functional, resulting in complete translocation of the large
345 Tetra-EGFP protein into the nucleus. In the H06-IPSE NLS, the presence of an
346 uncharged G in the charged KRRRK H03-IPSE core sequence of the NLS appears
347 to weaken its strength as nuclear targeting signal, as documented by the mixed
348 cytosolic/nuclear localization in contrast to the exclusive nuclear localization with the
349 SKRRRKY H03-IPSE NLS sequence (Fig. 2B). This difference is consistent with the
350 results obtained with different prediction algorithms (not shown). Substitution of
351 Lysine and Arginine with Alanine (SAAGAAY) results in a non-functional NLS, which
352 is no longer able to translocate the Tetra-EGFP protein into the nucleus (Fig. 2A and

B). These results clearly show that while the predicted NLS in H03-IPSE is fully functional, the G substitution in the positively charged core compromises this function at least in part. Further substitutions almost fully ablate NLS functionality.

<Figure 2 here>

Recombinant M-IPSE and H-IPSE can be expressed in high yields in HEK293-6E cells grown in suspension. For subsequent experiments, we cloned and expressed H-IPSE using HEK293-6E cells. This system uses a serum-free medium adapted clone and allows HEK293 cell culture in suspension, enabling high cell densities and recombinant protein yields. Proteins generated using these cells are glycosylated, which more closely parallels glycosylation of native proteins. Recombinant protein can be harvested from culture supernatants after transient transfection and can be purified e.g. via IMAC using the 8xHis-Tag in the construct (Figure 3A).

<Figure 3 here>

Results shown in Figure 3B demonstrate successful expression of H03-IPSE as a mostly dimeric protein of approximately 38-40 kDa molecular weight, with small amounts of monomeric protein of about 20 kDa, in line with what we have previously described for M-IPSE (1,2). The double bands are presumably due to glycosylation variants, which are well described for M-IPSE (21). Purity in eluted fractions after IMAC was high, and did not require any additional purification steps for downstream experiments. In our hands, IPSE proteins produced a range of yields, with H03-IPSE

exhibiting the lowest final yield (~5-10 mg/L) and M-IPSE and H06-IPSE SKAAAKY NLS mutant producing the highest yield (~15-25 mg/L) (Figure S5). Attempts to concentrate protein to higher concentration above 0.5 mg/mL resulted in formation of aggregates, not seen by SDS, but appearing in size-exclusion chromatography.

Recombinant H03-IPSE added exogenously is taken up by HTB-9 uroepithelial cells and efficiently translocates to the nucleus. To mimic *S. haematobium* infection conditions, we added recombinant 8xHis tagged H03-IPSE to the culture medium of proliferating HTB-9 uroepithelial carcinoma cells as described in Materials and Methods. After 24 hours, fixed cells were immunostained with anti-His tag antibody. As shown in Figure 4, this revealed a highly efficient uptake of exogenous H03-IPSE, which was present in HTB-9 nuclei, as revealed by co-staining of nuclear DNA with DRAQ5. The nuclear staining pattern suggests that H03-IPSE was largely excluded from the nucleolar regions. This result indicates that H03-IPSE protein can infiltrate the vast majority of cells with remarkable efficiency and localize to the nucleus. The only cells we observed that did not stain for recombinant H03-IPSE were those actively undergoing mitosis, in which the nuclear membrane is broken down

<Figure 4 here>

The NLS in H03- and H06-IPSE is essential for nuclear translocation, but not for cellular uptake. Next, we compared the ability of both H-IPSE variants with the ability of the NLS Alanine mutant, to gain access to the nuclear compartment of HTB-9 cells when added to cell culture medium. The results, visualized on a fluorescence microscope, are shown in Figure 5. When assessing the subcellular

403 localization of the recombinant molecules with a molecular weight of approximately
404 40 kDa, both H03- and H06-IPSE variants seemed similarly efficient in translocating
405 across the nuclear membrane.
406
407 <Figure 5 here>
408
409 Unlike what was observed for the Tetra-EGFP-NLS constructs (Fig. 2A and B),
410 we did not find any reduction in translocation efficiency in H06-IPSE compared with
411 H03-IPSE, suggesting that these differences may only become apparent with larger
412 proteins (such as Tetra-EGFP) and thus may not be relevant in the context of
413 molecular crosstalk between H-IPSE and the host cells. In contrast, the SKAAAKY
414 NLS H03 mutant, despite potentially being able to cross nuclear pores due to its low
415 molecular weight, remained completely excluded from the nucleus (Fig. 5 H03-IPSE
416 mut). The H03 mutant appears to be located in vacuoles or endosome-like
417 structures, mainly located around the nucleus, rather than diffuse in the cytoplasm.
418 This suggests that an intact NLS might be an important feature needed e.g. for
419 endosomal escape.
420 The lack of uptake of the Ala mutant shows that the NLS in H-IPSE is
421 monopartite and can be described as necessary and sufficient, i.e. no nuclear
422 translocation occurs in its absence, and it is the only NLS in the molecule. This is
423 consistent with the lack of prediction of additional nuclear translocation signals
424 elsewhere in the molecule. These data also demonstrate that the NLS is not required
425 for H-IPSE transport into cells.
426

H-IPSE mRNA is expressed in *S. haematobium* adult females and the egg stage. Next, we investigated the expression of H-IPSE across the different life cycle stages, using conventional RT-PCR. The results are shown in Figure 6. RT-PCR data indicated adult worm cDNA preparations were more contaminated with genomic DNA, in comparison with other life cycles in control experiments; however, DNAse treatment completely removed genomic DNA. Using DNAse-treated samples, RT-PCR indicated expression of H-IPSE mRNA in purified eggs, female adult worms (AdF), and weak expression in mixed gender adult worms (Ad Mix), but none in cercariae (cer), schistosomula (som), miracidia (mir) or male adult worms (AdM).

<Figure 6 here>

Overall, this result is consistent with an expression pattern restricted to the egg stage (since female worms often contain immature eggs), and similar to what was described for M-IPSE (7). Interestingly, despite the detection of H-IPSE transcripts in AdF, we detected no H-IPSE protein by western blot in AWA preparations (Figure 1b).

DISCUSSION

The concurrent presence of a classical secretory sequence (CSS) and a nuclear localization sequence (NLS) on the same protein, two apparently contradictory signals in terms of subcellular targeting, is a rare feature. Only four of 19 tested algorithms correctly identified the presence of a C-terminal, monopartite NLS in H03-IPSE (cNLSmapper (14), PSORT II (15), NLStradamus (16) and

452 NucPred (17)), while most other tested programs predicted a secretory pathway,
453 some unexpectedly also after removal of the signal sequence. Thus prediction of an
454 NLS by algorithms is still insufficient and such predictions need to be verified
455 experimentally. At least three properties need to be fulfilled in order to confirm
456 functionality of an NLS; a) a functional NLS needs to be able to direct the protein to
457 the nucleus, and b) the NLS in isolation should also be able to direct heterologous
458 proteins to the nucleus. Finally, c) mutation of one or several basic amino acids
459 should lead to loss of translocation, or in the case of a single substitution, at least
460 weakening of NLS functionality.
461
462 Our data confirm that the nuclear localization signals present in H03- and H06-
463 IPSE are functional and essential for translocation of IPSE into the nuclei of host
464 cells. This is somewhat surprising, as 40 kDa is well below the known limit for
465 passive diffusion across the nuclear pore complex, which has been described as
466 'quite larger than 60 kDa' (23). A possible explanation for such behavior could be yet
467 to be characterized interactions with cellular structures or soluble proteins in the
468 cytosol, making the resulting complex too large for passive diffusion into the nucleus.
469
470 Kosugi and coauthors described six classes of NLS with different specificities
471 for the binding grooves of the karyopherin importin α (19). Based on their
472 classification, the H03-IPSE SKRRRKY would be considered a Class I classical
473 NLS, characterized by a stretch of at least four consecutive basic amino acids (either
474 K or R). In contrast, the SKRGRKY NLS in H06-IPSE conforms with being a Class II
475 classical NLS with the consensus sequence K(K/R)X(K/R), in which one non-basic
476 amino acid interrupts the adjacency of the basic amino acids found in Class I signals,

reducing the signal to an interrupted sequence of three basic amino acids. The only difference from Kosugi's predicted canonical patterns is that both H03- and H06-IPSE appear to possess one additional basic residue in their NLS. Both classes would be predicted to bind to the large major binding pocket of Importin α (19).

Having shown that the NLS in H-IPSE is fully functional, and bearing in mind that the sequence cannot be functional within the schistosome eggs themselves, as the N-terminal CSS will target the protein for secretion well before the N-terminal NLS is synthesized, the key question is what the biological function of such a protein might be. The restriction of its expression to a single stage of the parasitic life cycle (the egg stage), which is in line with the M-IPSE homolog's restricted expression (7), suggests a specialized function needed only during a specific phase of egg embryogenesis, or a function needed to govern the very important interaction with the host cells and tissues. The former is unlikely due to the secretory nature of the molecule. To further explore this possibility, we first need to summarize the fate of eggs after oviposition by the female schistosomes. Newly deposited eggs do not have the complex morphology found in mature eggs; this has been described in detail for *S. mansoni* by Neill (24), Ashton (25), Jurberg (26) and their coauthors. Fewer details are available regarding *S. haematobium*. Immature eggs are smaller and characterized by the absence of the two envelopes surrounding the miracidia in mature eggs: the outer envelope (Reynold's layer, RL) and the inner envelope (von Lichtenberg's layer, vLL). The RL is enriched with tightly packed rough endoplasmic reticulum structures and is therefore thought to be a major site of protein synthesis. Under the electron microscope, the RL appears richly surrounded by granular materials in a 1μm wide space underneath the egg shell (25). Eggs deposited *in vitro*

by *ex vivo* worms have been shown to take about a week to fully develop into mature, infective miracidia (27), but development in the host is likely to be more rapid.

The production of the M-IPSE in *S. mansoni* eggs has been clearly shown to occur in the subshell area within the fully formed RL and vLL. This has been demonstrated by immunohistochemical staining with a monoclonal antibody to M-IPSE, and by in situ hybridization with labeled antisense transcripts of full-length M-IPSE cDNA (7). M-IPSE can also be seen in contact with the tissues around the eggs (confirming that it is secreted by the eggs) (20) and has been also seen inside surrounding host cells (7). Thus, the emerging picture is that immature eggs initially do not produce IPSE, but that this protein is produced as eggs mature while migrating through the tissues, releasing it into the tissues, where it is able to enter host cells. Inside host cells, it rapidly (in less than half an hour (18)) translocates to the nucleus, where it binds DNA (F. Falcone, unpublished data) with yet-to-be described downstream effects. It is very clear however, that translocation across the host tissues, in order for the eggs to reach the lumen of the bladder or gut, is an event of paramount importance in the life cycle of the parasite. Hence, it can be assumed that molecules secreted exclusively by more mature forms of this life stage may be involved in the egress process, and have evolved under high evolutionary pressure.

The first step after oviposition is the escape of eggs from the venules in which they were deposited; *in vitro*, human cells obtained from umbilical venous endothelial cells (HUVECs) rapidly overgrow eggs directly oviposited onto a monolayer (within 4 hours) and a similar behavior is seen when eggs are inserted into umbilical veins (28). More recently, de Walick and coauthors demonstrated deposition of van Willebrand factor and other prothrombic plasma proteins onto the eggshell of *S.*

mansoni (29). However, such steps occur immediately after oviposition, hence IPSE and other molecules secreted only after a few days of egg maturation cannot play a role in this initial process. Indeed, it was reported that this process was slower in mature eggs obtained from infected rodent livers in comparison with freshly deposited eggs, which may depend on the presence of uterine secretions covering the freshly deposited eggs (28). Once the eggs have reached the perivascular tissues, it takes another few days during which they need to cross several robust layers, including the submucosa, the outer muscularis mucosa, consisting of circular and longitudinal muscle, and the inner mucosa, complete with basal membrane and a very tight epithelial cell layer, before they can reach the lumen. A seemingly impossible journey?

It is well-accepted that the granulomatous reaction induced by the eggs plays a key role in this process (30, 31); much of the past research has focused on the interactions between immune cells and the schistosome eggs. In *S. mansoni*, eggs have been proposed to exploit gut lymphoid structures known as Peyer's patches as a preferential route of egress into the gut lumen (32), however a comparable route is not available to *S. haematobium* in the bladder. This leads to the question whether in addition to the inflammatory granulomatous response involving immune cells, any direct interactions with non-immune cells, such as fibroblasts, muscle cells or epithelial cells are also involved in facilitating egg translocation. In this context, it is interesting to note that in *S. haematobium*-infected animals, uroplakins and claudins involved in epithelial tight junction formation are downregulated after bladder exposure to eggs (33). Whether this downregulation, which is likely to aid egg egress by disrupting the integrity of the epithelium is due to the effects of H-IPSE or other egg-derived components, remains to be established. The nuclear translocation and

direct effects on gene transcription of H-IPSE are currently under active investigation in our laboratories.

Perhaps the most surprising result was to see to what extent an exogenously added parasitic molecule was able to enter host cells. This uptake does not appear to be very selective for specific cell types or animal species, as we have seen uptake of M-IPSE in human Huh-7, U2-OS and hamster CHO cell lines (18) as well as HUVEC and human monocyte derived dendritic cells (Falcone et al., unpublished data). H-IPSE was taken up by human HTB-9, Huh-7 and monkey Cos-7 cell lines. This raises the question as to whether or which receptors are involved. For M-IPSE, the uptake mechanism has been shown to involve the carbohydrate residues on the protein and C-type lectin receptors such as the mannose receptor, the dendritic cell-specific ICAM3-grabbing non-integrin (DC-SIGN), and a macrophage galactose-type lectin and the mannose receptor (34). Similar receptors have been shown to be involved in the uptake of other Schistosome molecules such as omega-1 (5) and kappa-5 (34). Uptake of a secreted molecule by dendritic cells and macrophages has also been shown for *S. japonicum* Sj16 (35), but there is no information regarding the receptors involved in this process; the expression of Sj16 in *E. coli* used for the described experiments however suggests that protein glycosylation does not play a role in uptake of this molecule.

Thus secretion of a molecule by a parasitic life stage which is in intimate contact with host tissues, and its subsequent uptake by host cells, may be a more common feature in the host-parasite relationship than hitherto assumed, at least as far as trematodes are concerned. We would like to propose the term ‘infiltrin’ to

577 denote the ability of such molecules to enter host cells in the course of crossing
578 several barriers (the cell membrane, presumably the endosomal membrane, and in
579 some cases, the nuclear membrane). The archetypal nuclear infiltrins, characterized
580 by the simultaneous presence of a classical secretory and a nuclear localization
581 signal (CSS/NLS) signal (11), would be M-IPSE and H-IPSE, while the archetypal
582 cytosolic infiltrin would be omega-1 (5). The ability of exogenous polypeptides to
583 enter human cells crossing biological membranes is not a new finding. This was
584 shown for the first time for trans-activating transcriptional activator (TAT) of human
585 immunodeficiency virus 1 (HIV-1) in 1988 (36) and the 60 amino acid peptide
586 encoded by antennapedia gene homeobox in *Drosophila* (37). However, with the
587 exception of our previous report (18), such a principle has not been described for
588 molecules secreted by macroparasites, which are too large to enter host cells. HIV-1
589 Tat can also carry heterologous proteins across the cell membrane (38), a process
590 now understood to be mediated via a caveolin-dependent uptake route (39).
591 Interestingly, both HIV-1 Tat and *Drosophila* antennapedia homeobox peptide exhibit
592 DNA binding activities, which are also predicted in silico for H-IPSE. Whether H-
593 IPSE has similar properties is also under investigation. We believe that these
594 observations make a compelling case warranting more in-depth studies of parasitic
595 infiltrins and their potential roles as pathogen-derived nuclear transcription factors.
596
597 Finally, it needs to be noted that H-IPSE's ability to enter host cells is not
598 dependent on an intact NLS, as the H03-IPSE NLS mutant, as well as the previously
599 described M-IPSE NLS mutants (18), are also able to enter mammalian cells. The
600 same is true for HIV-1 Tat, where the regions responsible for cellular uptake and
601 nuclear translocation are distinct (40, 41). Our data suggest that in the absence of an

intact NLS, H-IPSE is able to enter host cells, but remains trapped in endosome-like vesicles with a perinuclear distribution. Whether NLS mutants retain their ability to bind host DNA remains to be established.

Taken together, we suggests that nuclear infiltrins, by acting e.g. as transcription factors, might play a central role in controlling the host-parasite relationship at the molecular level.

REFERENCES

1. Gryseels B, Polman K, Clerinx J, Kestens L. 2006. Human schistosomiasis. Lancet 368:1106–18.
2. Dewalick S, Bexkens ML, van Balkom BWM, Wu Y-P, Smit CH, Hokke CH, de Groot PG, Heck AJR, Tielens AGM, van Hellemond JJ. 2011. The proteome of the insoluble *Schistosoma mansoni* eggshell skeleton. Int J Parasitol 41:523–32.
3. Mathieson W, Wilson RA. 2010. A comparative proteomic study of the undeveloped and developed *Schistosoma mansoni* egg and its contents: the miracidium, hatch fluid and secretions. Int J Parasitol 40:617–28.
4. Cass CL, Johnson JR, Califf LL, Xu T, Hernandez HJ, Stadecker MJ, Yates JR, Williams DL. 2007. Proteomic analysis of *Schistosoma mansoni* egg secretions. Mol Biochem Parasitol 155:84–93.
5. Fitzsimmons CM, Schramm G, Jones FM, Chalmers IW, Hoffmann KF, Grevelding CG, Wuhrer M, Hokke CH, Haas H, Doenhoff MJ, Dunne DW. 2005. Molecular characterization of omega-1: A hepatotoxic ribonuclease from

627 *Schistosoma mansoni* eggs. Mol Biochem Parasitol 144:123–127.
628 6. Schramm G, Hamilton JV, Balog CIA, Wuhrer M, Gronow A, Beckmann S,
629 Wippersteg V, Grevelding CG, Goldmann T, Weber E, Brattig NW, Deelder
630 AM, Dunne DW, Hokke CH, Haas H, Doenhoff MJ. 2009. Molecular
631 characterisation of kappa-5, a major antigenic glycoprotein from *Schistosoma*
632 *mansoni* eggs. Mol Biochem Parasitol 166:4–14.
633 7. Schramm G, Gronow a, Knobloch J, Wippersteg V, Grevelding CG, Galle J,
634 Fuller H, Stanley RG, Chiodini PL, Haas H, Doenhoff MJ. 2006. IPSE/alpha-1:
635 a major immunogenic component secreted from *Schistosoma mansoni* eggs.
636 Mol Biochem Parasitol 147:9–19.
637 8. Delafosse L, Xu P, Durocher Y. 2016. Comparative study of polyethylenimines
638 for transient gene expression in mammalian HEK293 and CHO cells. J
639 Biotechnol 227:103–111.
640 9. Beetz C, Brodhun M, Moutzouris K, Kiehntopf M, Berndt A, Lehnert D, Deufel
641 T, Bastmeyer M, Schickel J. 2004. Identification of nuclear localisation
642 sequences in spastin (SPG4) using a novel Tetra-GFP reporter system.
643 Biochem Biophys Res Commun 318:1079–1084.
644 10. Howe KL, Bolt BJ, Cain S, Chan J, Chen WJ, Davis P, Done J, Down T, Gao
645 S, Grove C, Harris TW, Kishore R, Lee R, Lomax J, Li Y, Muller H-M,
646 Nakamura C, Nuin P, Paulini M, Raciti D, Schindelman G, Stanley E, Tuli MA,
647 Van Auken K, Wang D, Wang X, Williams G, Wright A, Yook K, Berriman M,
648 Kersey P, Schedl T, Stein L, Sternberg PW. 2016. WormBase 2016:
649 expanding to enable helminth genomic research. Nucleic Acids Res 44:D774-
650 80.
651 11. Howe KL, Bolt BJ, Shafie M, Kersey P, Berriman M. 2016. WormBase

652 ParaSite - a comprehensive resource for helminth genomics. Mol Biochem
653 Parasitol.
654 12. Young ND, Jex AR, Li B, Liu S, Yang L, Xiong Z, Li Y, Cantacessi C, Hall RS,
655 Xu X, Chen F, Wu X, Zerlotini A, Oliveira G, Hofmann A, Zhang G, Fang X,
656 Kang Y, Campbell BE, Loukas A, Ranganathan S, Rollinson D, Rinaldi G,
657 Brindley PJ, Yang H, Wang J, Wang J, Gasser RB. 2012. Whole-genome
658 sequence of *Schistosoma haematobium*. Nat Genet 44:221–5.
659 13. Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0:
660 discriminating signal peptides from transmembrane regions. Nat Methods
661 8:785–786.
662 14. Kosugi S, Hasebe M, Tomita M, Yanagawa H. 2009. Systematic identification
663 of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by
664 prediction of composite motifs. Proc Natl Acad Sci U S A 106:10171–10176.
665 15. Nakai K, Horton P. 1999. PSORT: A program for detecting sorting signals in
666 proteins and predicting their subcellular localization. Trends Biochem Sci.
667 16. Nguyen Ba AN, Pogoutse A, Provart N, Moses AM. 2009. NLStradamus: a
668 simple Hidden Markov Model for nuclear localization signal prediction. BMC
669 Bioinformatics 10:202.
670 17. Brameier M, Krings A, MacCallum RM. 2007. NucPred - Predicting nuclear
671 localization of proteins. Bioinformatics 23:1159–1160.
672 18. Kaur I, Schramm G, Everts B, Scholzen T, Kindle KB, Beetz C, Montiel-Duarte
673 C, Blindow S, Jones AT, Haas H, Stolnik S, Heery DM, Falcone FH. 2011.
674 Interleukin-4-inducing principle from *Schistosoma mansoni* eggs contains a
675 functional C-terminal nuclear localization signal necessary for nuclear
676 translocation in mammalian cells but not for its uptake. Infect Immun 79:1779–

1788.

19. Kosugi S, Hasebe M, Matsumura N, Takashima H, Miyamoto-Sato E, Tomita M, Yanagawa H. 2009. Six classes of nuclear localization signals specific to different binding grooves of importin alpha. J Biol Chem 284:478–85.

20. Schramm G, Falcone FH, Gronow A, Haisch K, Mamat U, Doenhoff MJ, Oliveira G, Galle J, Dahinden C a, Haas H. 2003. Molecular characterization of an interleukin-4-inducing factor from *Schistosoma mansoni* eggs. J Biol Chem 278:18384–92.

21. Wuhrer M, Balog CIA, Catalina MI, Jones FM, Schramm G, Haas H, Doenhoff MJ, Dunne DW, Deelder AM, Hokke CH. 2006. IPSE/alpha-1, a major secretory glycoprotein antigen from schistosome eggs, expresses the Lewis X motif on core-difucosylated N-glycans. FEBS J 273:2276–2292.

22. Shatkay H, Höglund A, Brady S, Blum T, Dönnes P, Kohlbacher O. 2007. SherLoc: High-accuracy prediction of protein subcellular localization by integrating text and protein sequence data. Bioinformatics 23:1410–1417.

23. Wang R, Brattain MG. 2007. The maximal size of protein to diffuse through the nuclear pore is larger than 60 kDa. FEBS Lett 581:3164–3170.

24. Neill PJG, Smith JH, Doughty BL, Kemp M. 1988. The ultrastructure of the *Schistosoma mansoni* egg. Am J Trop Med Hyg 39:52–65.

25. Ashton PDD, Harrop R, Shah B, Wilson RA. 2001. The schistosome egg: development and secretions. Parasitology 122:329–338.

26. Jurberg AD, Gonçalves T, Costa TA, Mattos ACA, Pascarelli BM, Manso PPA, Ribeiro-Alves M, Pelajo-Machado M, Peralta JM, Coelho PMZ, Lenzi HL. 2009. The embryonic development of *Schistosoma mansoni* eggs: Proposal for a new staging system. Dev Genes Evol 219:219–234.

27. Michaels RM, Prata A. 968. Evolution and characteristics of *Schistosoma mansoni* eggs laid in vitro. J Parasitol 54:921–930.

28. File S. 1995. Interaction of schistosome eggs with vascular endothelium. J Parasitol 81:234–8.

29. deWalick S, Hensbergen PJ, Bexkens ML, Grosserichter-Wagener C, Hokke CH, Deelder AM, de Groot PG, Tielens AGM, Van Hellemond JJ. 2014. Binding of von Willebrand factor and plasma proteins to the eggshell of *Schistosoma mansoni*. Int J Parasitol 44:263–268.

30. Doenhoff MJ, Hassounah O, Murare H, Bain J, Murare H. 1986. The schistosome egg granuloma: Immunopathology in the cause of host protection or parasite survival? Trans R Soc Trop Med Hyg 80:503–514.

31. Pearce EJ. 2005. Priming of the immune response by schistosome eggs. Parasite Immunol.

32. Turner JD, Narang P, Coles MC, Mountford AP. 2012. Blood Flukes Exploit Peyer's Patch Lymphoid Tissue to Facilitate Transmission from the Mammalian Host. PLoS Pathog 8:e1003063.

33. Ray D, Nelson TA, Fu CL, Patel S, Gong DN, Odegaard JI, Hsieh MH. 2012. Transcriptional Profiling of the Bladder in Urogenital Schistosomiasis Reveals Pathways of Inflammatory Fibrosis and Urothelial Compromise. PLoS Negl Trop Dis 6:e1912.

34. Meevissen MHJ, Driessen NN, Smits HH, Versteegh R, van Vliet SJ, van Kooyk Y, Schramm G, Deelder AM, Haas H, Yazdanbakhsh M, Hokke CH. 2012. Specific glycan elements determine differential binding of individual egg glycoproteins of the human parasite *Schistosoma mansoni* by host C-type lectin receptors. Int J Parasitol 42:269–277.

35. Sun X, Yang F, Shen J, Liu Z, Liang J, Zheng H, Fung M, Wu Z. 2016. Recombinant Sj16 from *Schistosoma japonicum* contains a functional N-terminal nuclear localization signal necessary for nuclear translocation in dendritic cells and interleukin-10 production. Parasitol Res 115:4559–4571.

36. Frankel AD, Pabo CO. 1988. Cellular uptake of the tat protein from human immunodeficiency virus. Cell 55:1189–1193.

37. Joliot A, Pernelle C, Deagostini-Bazin H, Prochiantz A. 1991. Antennapedia homeobox peptide regulates neural morphogenesis. Proc Natl Acad Sci U S A 88:1864–8.

38. Fawell S, Seery J, Daikh Y, Moore C, Chen LL, Pepinsky B, Barsoum J. 1994. Tat-mediated delivery of heterologous proteins into cells. Proc Natl Acad Sci U S A 91:664–668.

39. Ferrari A, Pellegrini V, Arcangeli C, Fittipaldi A, Giacca M, Beltram F. 2003. Caveolae-mediated internalization of extracellular HIV-1 Tat fusion proteins visualized in real time. Mol Ther 8:284–294.

40. Efthymiadis A, Briggs LJ, Jans DA. 1998. The HIV-1 tat nuclear localization sequence confers novel nuclear import properties. J Biol Chem 273:1623–1628.

41. Kuppuswamy M, Subramanian T, Srinivasan A, Chinnadurai G. 1989. Multiple functional domains of Tat, the trans-activator of HIV-1, defined by mutational analysis. Nucleic Acids Res 17:3551–61.

ACKNOWLEDGMENTS

We would like to thank Dr Christian Beetz for the kind donation of the TetraEGFP plasmid and Dr Hilary Collins for help with confocal microscopy. This work was funded by an NIH R56 grant (R56AI119168) awarded to MHH and TSJ. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Figure Legends

Figure 1: *S. haematobium* expresses multiple forms of H-IPSE. A) Amino acid sequences of predicted H-IPSE paralogs, M-IPSE, and sequenced transcripts from egg cDNA were globally aligned using a Blosum62 cost matrix and a tree was built using the neighbor-joining method in Geneious 7.1.4. Scale bar represents amino acid substitutions per site. H06 and H03 variants chosen for expression are highlighted in green, and their respective amino acid identity and identity to M-IPSE are shown. All variants identified through 3' RACE cloning are denoted with asterisks. B) Alignment of amino acid sequences of H03-IPSE (top row) and H06-IPSE (middle row) with the homolog in *S. mansoni* (IPSE/alpha-1, named M-IPSE here, bottom row). These H-IPSE clones retain ~63-68% amino acid identity and several previously features described in M-IPSE (20), including: a 20 amino acid classical signal sequence, seven cysteine residues involved in disulfide bonds, two N-linked glycosylation consensus motifs, and a predicted nuclear localization sequence (data not shown). Residues colored in green are identical, residues in yellow share properties (e.g. hydrophobicity, polarity), and residues in red lack similarity. C) SDS-PAGE gel (left) and western blot with anti-H06-IPSE antiserum (right). Lanes contain parasite-derived adult worm antigen (AWA), egg secretory protein (ESP), or soluble egg antigen (SEA).

782 **Figure 2:** Effect of multiple amino acid substitutions on NLS in H-IPSE. A) The
783 nucleotides encoding the H03/H06-IPSE nuclear localization sequence (SKRRRKY
784 and SKRGRKY, respectively) were inserted into the pTetra-EGFP construct (2,3).
785 pTetra-EGFP encodes a tetrameric EGFP construct resulting in the expression of a
786 fluorescent protein which due to its size (>100 kDa) is excluded from the nucleus in
787 the absence of a functional NLS (Tetra-EGFP) or imported into the nucleus in the
788 presence of a functional NLS (canonical SV40 NLS, H03/H06-IPSE NLS). Nuclei
789 were stained with DAPI and green fluorescence measured with the GFP light cube
790 on an EVOS *fl* microscope, 24 hours after transfection. Bar is 100 µm. B)
791 Comparison of wild-type H06-IPSE, H03-IPSE and H03-IPSE mutant NLS effect on
792 nuclear localisation of Tetra-EGFP fusion protein. One hundred transfected HTB9
793 cells were evaluated under the EVOS *fl* microscope for each transfection and the
794 percentage of cells displaying exclusive nuclear fluorescence, as opposed to
795 cytosolic only or mixed cytosolic/nuclear localization, recorded. Positive control:
796 Sv40 canonical NLS sequence; negative control: unmodified Tetra-EGFP vector
797 (Tetra-EGFP).
798

Figure 3: Expression of M-IPSE and H-IPSE in HEK293-6A cells. A) Schematic diagram of pTT5 H03/06-IPSE expression cassette. eCMV=Cytomegalovirus (CMV) enhancer sequence; pCMV=CMV promoter; TPL=tripartite leader sequence from adenovirus; eMLP=enhancer element from the adenovirus major late promoter (MLP); hVEGF=human vascular endothelial growth factor signal sequence; 8xHis=octahistidine tag; TEV=Tobacco Etch Virus protease cleavage site; STOP: stop codon; pA: β-globin polyadenylation signal. B) Coomassie-stained 4-20% SDS-PAGE gradient gel and C) Western Blotting of recombinant H03-IPSE (and M-IPSE, used as comparison) expressed in HEK 293SF-3F6 cells and purified by IMAC from serum-free culture supernatant, and run under non-reducing (NR) or reducing (R) conditions.

Figure 4: Recombinant H03-IPSE is taken up by HTB-9 host cells and translocates to the nucleus. HTB-9 cells, incubated for 24 hours with 0.40 nM recombinant H03-IPSE, were stained with 5 µM DRAQ5 nuclear stain for 15 minutes at room temperature, followed by staining with a mouse anti-His antibody and Alexa Fluor® 555 conjugated Goat anti-Mouse IgG (H+L) as secondary antibody. The right column shows the overlay of the two channels. The uptake in HTB-9 cells was visualized by confocal microscopy. The primary anti-His antibody was omitted in the control lane.

Figure 5: Fluorescence microscopy of HTB-9 cells incubated with recombinant H03-IPSE (NLS: SKRRRKY), H06-IPSE (NLS: SKRGRKY) or H03-IPSE mutant (NLS: SKAAAKY). HTB-9 cells were stained with Hoechst 33342 nuclear stain for 15 minutes at room temperature, followed by staining with a mouse anti-His antibody and Alexa Fluor® 555-conjugated Goat anti-Mouse IgG (H+L) as secondary antibody. The right column shows the overlay of the two channels. The primary anti-His antibody was omitted in the control lane. Bar size is 100 µm.

Figure 6: Stage-specific expression of H-IPSE mRNA. RT-PCR results for H-IPSE obtained from cDNAs, prepared by reverse transcription of DNAse-treated RNA of various life stages of *S haematobium.* Ladder: 100 basepair (Bp) DNA ladder; egg: *S. haematobium* egg cDNA; mir: miracidial cDNA; cer: cercarial cDNA; som: *in vitro* mechanically transformed schistosomula cDNA; Ad, F, M mixed cDNA from female, male or mixed adult worms, respectively. ShTub: control housekeeping gene, *S. haematobium* tubulin.