Open Access Full Text Article

ORIGINAL RESEARCH

Inferences on the biochemical and environmental regulation of universal stress proteins from Schistosomiasis parasites

Andreas N Mbah[1,2]
Ousman Mahmud[1]
Omotayo R Awofolu[2]
Raphael D Isokpehi[1]

[1]Center for Bioinformatics and Computational Biology, Department of Biology, Jackson State University, Jackson, MS, USA; [2]Department of Environmental Sciences, College of Agriculture and Environmental Sciences, University of South Africa, Pretoria, South Africa

Background: Human schistosomiasis is a freshwater snail-transmitted disease caused by parasitic flatworms of the *Schistosoma* genus. *Schistosoma haematobium*, *Schistosoma mansoni*, and *Schistosoma japonicum* are the three major species infecting humans. These parasites undergo a complex developmental life cycle, in which they encounter a plethora of environmental signals. The presence of genes encoding the universal stress protein (USP) domain in the genomes of *Schistosoma* spp. suggests these flatworms are equipped to respond to unfavorable conditions. Though data on gene expression is available for USP genes, their biochemical and environmental regulation are incompletely understood. The identification of additional regulatory molecules for *Schistosoma.* USPs, which may be present in the human, snail, or water environments, could also be useful for schistosomiasis interventions.

Methods: We developed a protocol that includes a visual analytics stage to facilitate integration, visualization, and decision making, from the results of sequence analyses and data collection on a set of 13 USPs from *S. mansoni* and *S. japonicum*.

Results: Multiple sequence alignment identified conserved sites that could be key residues regulating the function of USPs of the *Schistosoma* spp. Based on the consistency and completeness of sequence annotation, we prioritized for further research the gene for a 184-amino-acid-long USP that is present in the genomes of the three human-infecting *Schistosoma* spp. Calcium, zinc, and magnesium ions were predicted to interact with the protein product of the gene.

Conclusion: Given that the initial effects of praziquantel on schistosomes include the influx of calcium ions, additional investigations are required to (1) functionally characterize the interactions of calcium ions with the amino acid residues of *Schistosoma* USPs; and (2) determine the transcriptional response of *Schistosoma.* USP genes to praziquantel. The data sets produced, and the visual analytics views that were developed, can be easily reused to develop new hypotheses.

Keywords: ATP binding protein, calcium, functional sites, praziquantel, *Schistosoma*, schistosomiasis

Introduction

Human schistosomiasis is a freshwater snail-transmitted disease caused by parasitic flatworms of the *Schistosoma* genus.[1] *Schistosoma haematobium*, *Schistosoma mansoni*, and *Schistosoma japonicum* are the three major species infecting humans. Schistosomiasis has been designated as one of the "neglected tropical diseases" of poverty and is the second most significant tropical disease, after malaria, in public health significance.[2] The large majority of human schistosomiasis and most of the severest disease states are now concentrated in the relatively resource-poor countries of sub-Saharan Africa, contributing to approximately 280,000 deaths per annum.[3] Schistoso-

Correspondence: Raphael D Isokpehi
Center for Bioinformatics and Computational Biology, Department of Biology, 1400 JR Lynch Street, Jackson State University, PO Box 18540, Jackson, MS 39217, USA
Tel +1 601 979 1678
Fax +1 601 979 5853
Email raphael.isokpehi@jsums.edu

15

miasis is also among the severest parasitic diseases targeted, in terms of morbidity and mortality, and has been highlighted for control by the World Health Organization (WHO), with the urinary form highly associated with increased risks for bladder cancer.[2,4] The drug of choice for treatment of schistosomiasis is praziquantel (PZQ), but there is great concern regarding effective treatment in affected communities, due to the potential for parasite resistance to PZQ.[4–6]

The genomes of *S. haematobium*, *S. mansoni*, and *S. japonicum* encode proteins with the universal stress protein (USP) domain (Pfam Identifier: PF00582).[7–9] The USPs are known to function during unfavorable environmental conditions, including the life cycle developmental stages in *Schistosoma* spp.[10–12] Though data on gene expression is available for genes encoding USPs (USP genes), their biochemical and environmental regulation are incompletely understood.[10–19] The identification of additional regulatory molecules for *Schistosoma* USPs, which may be present in the human, snail, or water environments, could also be useful for schistosomiasis interventions. Our hypothesis is that the USPs of *Schistosoma* spp. have shared protein sequence features that can help us infer their biochemical and environmental regulation. Further, in the context of an infectious disease caused by geographically dispersed species of the same genus, proteins that have shared features could be targets for intervention.

The USPs are found in a diverse group of organisms, including archaea, bacteria, yeast, fungi, and plants, and encompass a conserved group of proteins whose expressions are triggered by a variety of environmental insults, including toxic chemicals, drought, and extreme temperature.[20] Genes encoding USPs have not been identified in the human genome, thus making them attractive as drug targets.[21] In a previous report,[12] we analyzed the developmental expression of eight USP genes predicted from the *S. mansoni* genome. In the case of *S. japonicum*, multiple research investigations have detected developmental stage expression of USP genes.[10,11,16]

The *Schistosoma* spp. undergo a complex developmental life cycle that includes multiple morphological stages and transition between hosts. The life cycle developmental forms of *Schistosoma* spp. include egg, miracidium, sporocyst, cercaria, schistosomulum, and adult (male and female). These stages must survive diverse stress conditions. The eggs, miracidia, and cercariae are found outside the human and snail hosts and are thus exposed to the stress conditions associated with the freshwater environment of the snail vectors. The cercariae and sporocysts are found in the snail host

and must respond to toxic substances, causing oxidative and nitrosative stresses in the snail hemocytes.[22,23] In the human host, to develop to the adult form, the schistosomula migrate through multiple organs, including lungs, heart, and liver. These organ systems have defense mechanisms, including production of nitric oxide and hydrogen peroxide, designed to kill the parasite stages.[24,25] In summary, all the developmental stages of the *Schistosoma* spp. are exposed to various stresses. A common aspect of these environment-inducing stresses is that they result in proteins with nonnative conformations.[26] Inducible stress tolerance has increasingly been understood to result from numerous molecular mechanisms, such as heat shock proteins (Hsps) and USPs.[12,27–31]

Adenosine triphosphate (ATP) binding is a biochemical mechanism that regulates the function of USPs through phosphorylation.[32] Members of the USP family can be categorized into two groups based on the presence or absence of the ATP-binding motif G-2x-G-9x-G(S/T) in their amino acid sequence.[20,33] USPs can be phosphorylated on serine and threonine residues by phosphate donors ATP and guanosine triphosphate (GTP), in the absence of other proteins, coupled with an upregulation response to stressors,[34] This observation indicates that in addition to environmental-stressor-mediated regulation, other cellular factors could modulate the activity of USPs by controlling their phosphorylated state.[35] Given the broad range of resistance functions conferred by the USPs, it is not surprising that they are encoded in the genomes of a variety of both pathogens and nonpathogens.[20,36–38]
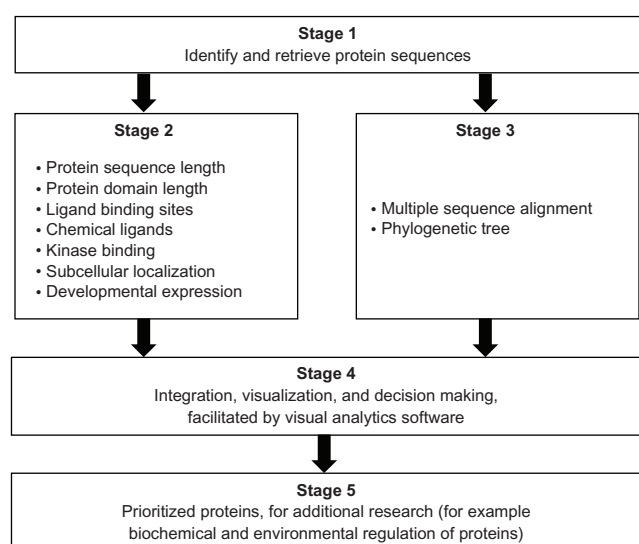
We have determined protein sequence length, protein domain length, ligand-binding sites, biologically relevant chemical ligands, enzymatic regulation, developmental regulation, and subcellular localization for 13 *Schistosoma* USP sequences (five *S. mansoni* and eight *S. japonicum* sequences). Multiple sequence alignment and phylogenetic analysis provided the evolutionary groupings to allow inferences on biochemical and environmental regulation of the proteins. The results from the analyses were integrated and visualized with visual analytics software. Multiple sequence alignment identified conserved sites of aspartate (Asp), glycine (Gly), histidine (His), leucine (Leu), and proline (Pro) residues in all the sequences. These residues could be key residues regulating the function of the USPs of *Schistosoma* spp. We prioritized a group of two 184-amino-acid-long USP sequences (Q86DW2 [*S. japonicum*] and G4LZI3 [*S. mansoni*]) because they had identical values for multiple annotation features. Data visualization revealed the two proteins have identical values for subcellular localization, ligand-binding sites, chemical ligands, and enzymatic

regulation. Specifically, calcium, zinc, and magnesium ions were predicted to interact with the two proteins. Given that the initial effects of PZQ on schistosomes include the influx of calcium ions,[13,39–41] additional investigations are required to (1) functionally characterize the interactions of calcium ions with the amino acid residues of *S.* USPs; and to (2) determine the transcriptional response of *Schistosoma* USP genes to PZQ.

Methods
Overview of bioinformatics and visual analytics methods

A variety of limitations, including costs, preclude the functional characterization of all predicted proteins from a genome sequencing project. The selection of proteins for further research is a decision-making process by a researcher or research team. Thus, we developed a protocol that integrates the visual analytics stages, to facilitate the interaction with the results from sequence analysis and data collection, on a set of USPs from *S. mansoni* and *S. japonicum*. Visual analytics is an iterative process conducted via visual interfaces that

involves collecting information, data preprocessing, knowledge representation, interaction, and decision making.[42–46]

The overview of the bioinformatics and visual analytics methods is summarized in Figure 1. The protocol consists of five stages that start with the protein sequences to be investigated (Stage 1). Two sets of bioinformatics analyses are performed (Stage 2 and Stage 3). Stage 2 consists of analyses done on each sequence, while, in Stage 3, all the sequences are used for multiple sequence alignment and to construct phylogenetic trees. The bioinformatics sequence analyses in Stage 2 determine (1) the protein sequence length; (2) the protein domain length; (3) the ligand-binding sites; (4) chemical ligand binding; (5) kinase binding; and (6) subcellular localization. These analyses can be particularly useful for prioritizing sequences for research on the biochemical and environmental regulation of proteins. Information on the developmental expression of gene transcripts can assist in deciding the choice of life cycle parasite form to investigate. The data on the developmental expression of gene transcripts were obtained from publications. Multiple sequence alignment and phylogenetic trees provided the statistically and evolutionary support for groupings of the sequences.

The prioritization process was done in Stage 4, with the criterion determined by the researchers. In this report, the criterion was to identify pairs of protein sequences (one from *S. mansoni* and another from *S. japonicum*) that share identical annotations, from the following analyses: protein sequence length, ligand-binding sites (amino acid type and amino acid position), and chemical ligands that are predicted to bind. Stage 5 was the product of the prioritization process. In this report, we expect (i) orthologous pairs of protein sequences and (ii) a visualization that provides an integrated view of the shared annotations. We used a visual analytics software package[45] (Tableau 7.0, Tableau Software Inc, Seattle, WA, USA) to perform several visual analytics tasks, including interaction, computing, analysis, integration, and visualization.

Figure 1 Overview of a set of bioinformatics and visual analytics methods used to prioritize protein sequences for further research.
Notes: The core of the prioritization process is a visual analytics stage (Stage 4) that enables the interaction of researcher(s) with the results from the bioinformatics analyses (Stage 2 and Stage 3) of the protein sequences (Stage 1). The evolutionary relatedness of the protein sequences is based on statistically supported groups in a phylogenetic tree that is derived, in turn, from the multiple sequence alignment of all the sequences (Stage 3). Additional evidence for evolutionary relatedness is obtained from the gene synteny on the chromosomal regions. The protocol can be particularly suited for identifying orthologous proteins with shared patterns of sequence and functional annotations (Stage 5). In the context of schistosomiasis parasites from different regions of the world, the identified proteins could be targets for understanding shared biological processes during the life cycle of parasites. Details of each method are available in the Methods section of the article.

Retrieval of protein sequences

Proteins annotated with the USP domain (PF00582) from the *S. mansoni* and *S. japonicum* genomes were identified in Universal Protein Resource (UniProt release 2011_11: http://www.uniprot.org/).[47] Predicted protein sequences were retrieved. The final list of protein sequences for comparative sequence analysis was determined by the revision history of the sequence in the UniProt as well as by entries in GeneDB[48] and SchistoDB.[49]

Conserved domain search for functional sites

The search for amino acid residues that are functionally important was performed using two public servers. A single-sequence input server (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi)[50] and a multiple-sequence input or batch server (http://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi)[50] were used in finding conserved domains for the sequences. The ATP-binding motif residues and other ligand-binding residues were identified and documented for all the USP sequences, including their domain architecture. To facilitate comparison of the functional sites, we constructed a functional site signature for each sequence. The signature is a string of the amino acid letters. In a case where no site is predicted in the 12-letter signature, the position was assigned "X." Therefore, for protein sequence MJ0577, the template for the ligand-binding sites, the signature is PTDVMGHGGSVT. This approach of constructing functional sites has been implemented in previous research on functional sites.[51]

Prediction of chemical ligand and enzymatic regulation

The three-dimensional (3D) chemical ligands were predicted using the 3DLigandSite server (http://www.sbg.bio.ic.ac.uk/3dligandsite).[52] These biologically relevant chemical ligands are potential regulators of the function of the USPs. The 3DLigandSite is a top-performing web server for chemical ligand prediction and provides structural models for unsolved proteins, using protein-structure prediction. The specific kinases were predicted using the NetPhosK 1.0 tool (http://www.cbs.dtu.dk/services/NetPhosK),[53] with a stringent threshold value set at 0.65. The kinase with the highest threshold value was selected.

Prediction of subcellular location

The subcellular locations of all the genes were retrieved from literature, databases or predicted if possible using the server Euk-mPLoc 2.0 (http://www.csbio.sjtu.edu.cn/bioinf/euk-multi-2/).[54] The prediction by Euk-mPLoc2.0 is based on integrating information from gene ontology, functional domain, and evolutionary relationships.

Compilation of developmental expression of genes

The developmental stage expression profiles of the selected *Schistosoma* USP genes were extracted from our previous publication[12] for *S. mansoni*. In the case of *S. japonicum*, expression profiles were from the SjTPdb, an integrated transcriptome and proteome database and analysis platform for *S. japonicum*.[55] The database contains expressed sequence tags (ESTs), EST clusters, and the proteomic dataset for *S. japonicum*.

Prediction of evolutionary relatedness of sequences

Evolutionary relatedness of the selected USPs, from both *S. mansoni* and *S. japonicum*, was determined to ascertain their functional and evolutionary relationship. The sequences were aligned using the ClustalW tool (http://www.ch.embnet.org/software/ClustalW.html),[56] applying the default settings. The evolutionary relationship of the sequences was inferred using the maximum likelihood method,[57] based on the JTT matrix-based model[58] at 1000 bootstrap, with MEGA software version 5 (http://www.megasoftware.net/ [Center for Evolutionary Medicine and Informatics, Tempe, AZ, USA]).[59] The bootstrap test indicated above, at 1000 replications, was used to determine the percentage of the replicate trees, in which the genes clustered together.[60,61]

Visual analytics of datasets

A purpose of Stage 4 of the protocol was to provide an integration and visualization portal for the results of the bioinformatics analysis in Stage 2 and Stage 3 (Figure 1). The visual analytics tasks to be performed on the data sets can be influenced by how the data is organized in the data records (rows) and data fields (columns) in the data source (eg, spreadsheet file and comma delimited file). For Stage 2, each data record had the following data fields: (1) organism; (2) locus tag; (3) UniProt ID; (4) feature; and (5) feature value. The feature field had the following types: protein domain length, protein length, protein domain start position, protein domain end position, ATP-binding motif, kinase type, kinase type score, 3D chemical ligand, ligand-binding amino acid, amino acid and sequence position, developmental expression, and subcellular localization. In the case of the functional site signature data set, each record consisted of the UniProt ID and 12 fields for each of the 12-letter signatures for the USP ligand-binding sites. For the Stage 3 data set (phylogenetic tree groupings), each data record consisted of data fields for (1) organism; (2) UniProt ID; and (3) phylogenetic group. The data sources (in this case, spreadsheet files) were loaded to the visual analytics software for the visual analytics tasks, including the design of the data integration and visualization.

Table I Annotation features for universal stress proteins of *Schistosoma mansoni* and *Schistosoma japonicum*

Organism	UniProt ID	Locus tag	Feature value							Frequency per protein				
			Protein domain length (aa)	Protein length (aa)	Protein domain start position	Protein domain end position	ATP-binding motif	Kinase type	Kinase type score	3D chemical ligand	Ligand binding amino acid	Amino acid and sequence position	Developmental expression	Predicted subcellular localization
S. japonicum	Q5DDH7		148	172	15	163	P	PKC	0.87	5	12	12		1
	Q5DED2		136	160	15	151	P	PKC	0.87	5	12	12		2
	Q5DGI9		146	159	8	154	P	PKC	0.76	5	12	12	1	1
	Q5DH64		122	129	2	124	P	PKC	0.78	4	9	9	6	1
	Q5DHK1		149	172	18	167	P	PKC	0.80	5	12	12	6	5
	Q5DI36		125	133	3	128	P	PKC	0.72	4	9	9	5	1
	Q86DW2		148	184	28	176	P	PKC	0.93	7	12	12	5	3
	Q86DX1		147	155	7	154	P	PKC	0.85	5	12	12	3	3
S. mansoni	C1M0Q2	Smp_097930	148	159	7	155	P	PKA	0.68	5	12	12	1	1
	G4LZI3	Smp_076400	149	184	27	176	P	PKC	0.93	7	12	12	3	3
	G4VSS2	Smp_001000	148	174	15	163	P	PKA	0.77	6	12	12	3	3
	G4VIW9	Smp_043120	148	160	7	155	P	PKC	0.83	6	12	12	7	1
	G4VPM6	Smp_031300	149	160	6	155	P	PKC	0.71	5	12	12	4	2

Note: Sequences of *S. mansoni* have "Smp" in the sequence identifier.
Abbreviations: aa, amino acid; PKA, protein kinase A; PKC, protein kinase C; UniProt, Universal Protein Resource (Apweiler et al).[47]

Results

Data set for visual analytics

The data set analyzed consisted of twelve annotation features for 13 USP sequences (Table 1). All the sequences analyzed contained the ATP-binding motif G-2x-G-9x-G(S/T). We were particularly interested in shared annotations that can help us infer the joint biochemical and environmental regulation of the USPs from the two pathogenic *Schistosoma* spp. The protein sequences consisted of five *S. mansoni* and eight *S. japonicum* sequences. Seven of the annotation features required only one feature value per protein sequence. These features were protein domain length, protein length, protein domain start position, protein domain end position, ATP-binding motif, kinase type, and kinase type score. The five annotation features with variable frequency per protein were 3D chemical ligand, ligand-binding amino acid, amino acid and sequence position, developmental expression, and subcellular localization. This dataset was formatted for visualization and analysis in a visual analytics resource available at http://public.tableausoftware.com/views/schisto_features_usp/feature_per_usp.

The sequences were first grouped by protein length, protein domain length, and ligand-binding sites. Since we intend to conduct additional research on the *Schistosoma* USPs, the primary purpose of these groups was to prioritize protein sequences that have relatively complete and consistent annotation. The additional bioinformatics predictions from the sequences helped to confirm the sequence level observations. The comparison of the other annotation features was done in the context of the evolutionary relatedness predicted by multiple sequence alignment.

Grouping of *Schistosoma* USPs by sequence length

The 13 *Schistosoma* protein sequences were grouped by protein length and domain length (Figure 2). Seven distinct domain lengths (122, 125, 136, 146, 147, 148, and 149 amino acids [aa]) were observed in the sequences compared. Ten of the 13 sequences had domain lengths from 146 aa to 149 aa. The groups with domain length types of 148 aa and 149 aa had sequences from the two species. For the protein sequence length grouping, eight distinct types (129, 133, 155, 159, 160, 172, 174, and 184 aa) were observed. Three groups (159 aa, 160 aa, and 184 aa) had members from both species. There were no groups of sequences from both species that had identical members, by protein sequence and protein domain lengths. However, we observed protein sequences from both species that shared protein sequence length with the difference in domain

Feature	Feature Value	Organism	UniProt ID	Locus Tag
Domain Length (aa)	122	*Schistosoma japonicum*	Q5DH64	
	125	*Schistosoma japonicum*	Q5DI36	
	136	*Schistosoma japonicum*	Q5DED2	
	146	*Schistosoma japonicum*	Q5DGI9	
	147	*Schistosoma japonicum*	Q86DX1	
	148	*Schistosoma japonicum*	Q5DDH7	
			Q86DW2	
		Schistosoma mansoni	C1M0Q2	Smp_097930
			G4V5S2	Smp_001000
			G4VIW9	Smp_043120
	148	*Schistosoma japonicum*	Q5DHK1	
		Schistosoma mansoni	G4LZI3	Smp_076400
			G4VPM6	Smp_031300
Protein length (aa)	129	*Schistosoma japonicum*	Q5DH64	
	133	*Schistosoma japonicum*	Q5DI36	
	155	*Schistosoma japonicum*	Q86DX1	
	159	*Schistosoma japonicum*	Q5DGI9	
		Schistosoma mansoni	C1M0Q2	Smp_097930
	160	*Schistosoma japonicum*	Q5DED2	
		Schistosoma mansoni	G4VIW9	Smp_043120
			G4VPM6	Smp_031300
	172	*Schistosoma japonicum*	Q5DDH7	
			Q5DHK1	
	174	*Schistosoma mansoni*	G4V5S2	Smp_001000
	184	*Schistosoma japonicum*	Q86DW2	
		Schistosoma mansoni	G4LZI3	Smp_076400

Figure 2 Grouping of 13 *Schistosoma* USPs by sequence length.
Notes: The image provides a visual comparison of the protein sequence and USP domain sequence lengths for 13 *Schistosoma* USPs. A visual analytics resource that can be used to interact with the data is available at http://public.tableausoftware.com/views/schisto_features_usp/groupbylength. Sequences of *Schistosoma mansoni* have "Smp" in the sequence identifier.
Abbreviations: aa, amino acid; UniProt, Universal Protein Resource (Apweiler et al);[47] USP, universal stress protein.

length of 1 aa, as in the 184 aa USPs (Q86DW2 and G4 LZI3), or 2 aa, as in the 159 aa USPs (Q5DGI9 and C1M0Q2).

Functional site signatures of *Schistosoma* USP sequences

The Conserved Domain Search tool at the National Center for Biotechnology Information (NCBI) website (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi)[50] uses the fragment A (UniProt Identifiers: Q57997; Y577_METJA) of the USP MJ0577 of *Methanocaldococcus jannaschii* as the template to align the *Schistosoma* USPs for the conserved protein domain search. The search reports the ligand-binding sites (functional sites) including the ATP-binding motif [G-2X-G-9X-G(S/T)] present in the query sequence. A total of 12 ligand-binding sites are predicted for MJ0577. The amino acids and their position are Pro11, Tyr12, Asp13, Val41, Met126, Gly127, His129, Gly130, Gly140, Ser141, Val142, and Thr143. The Asp13 and Val41 are binding sites for adenosine nucleoside.[33] The amino acid sequences from met 126 to Thr143 contain the motif G2xG9xG(S/T), which includes binding sites

Functional site signature	UniProt ID	Locus Tag	Organism
AIDAIGRGGSVS	Q5DDH7		*Schistosoma japonicum*
	Q5DED2		*Schistosoma japonicum*
AIDAIGRGGSVS	G4V5S2	Smp_001000	*Schistosoma mansoni*
PIDVIGRGGSVS	G4VPM6	Smp_031300	*Schistosoma mansoni*
	Q5DHK1		*Schistosoma japonicum*
PIDVMGRGGSVS	G4VIW9	Smp_043120	*Schistosoma mansoni*
	Q5DGI9		*Schistosoma japonicum*
PIDVMGHGGSVT	Y577_METJA		*Methanocaldococcus jannaschii*
PVDIIGRGGSVS	G4LZI3	Smp_076400	*Schistosoma mansoni*
	Q86DW2		*Schistosoma japonicum*
PVDSMGRGGSVS	Q86DX1		*Schistosoma japonicum*
PVDVIGRGGSVS	C1M0Q2	Smp_097930	*Schistosoma mansoni*
XXXVMGRGGSVS	Q5DH64		*Schistosoma japonicum*
	Q5DI36		*Schistosoma japonicum*

Figure 3 Grouping of 13 *Schistosoma* universal stress proteins by functional site signature.
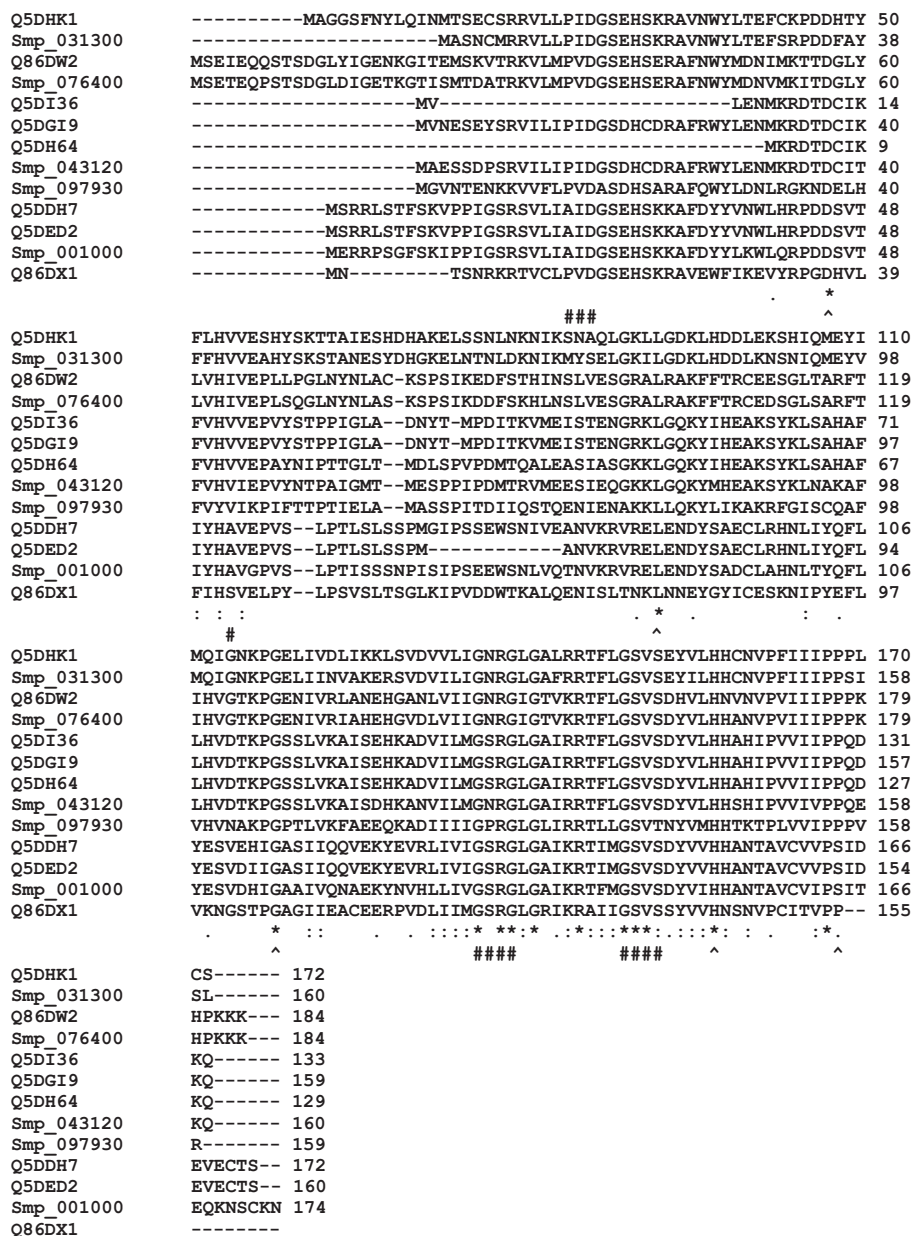**Notes:** The functional site signature is constructed by joining the twelve ligand binding sites known for the ATP-binding USP from *Methanocaldococcus jannaschii* (UniProt [Apweiler et al][47] ID: Y577_METJA). The image provides a visual comparison of the functional site signatures for 13 *Schistosoma* USPs. A visual analytics resource that can be used for interacting with the data is available at http://public.tableausoftware.com/views/schisto_features_usp/groupbylength. Sequences of *Schistosoma mansoni* have "Smp" in the sequence identifier.
Abbreviations: ATP, adenosine triphosphate; UniProt, Universal Protein Resource; USP, universal stress protein.

for the phosphoyrl and ribosyl groups of ATP. Among the 13 *Schistosoma* sequences with ATP-binding motif, eight functional site signatures (AIDAIGRGGSVS, AIDAVGRGGSVS, PIDVIGRGGSVS, PIDVMGRGGSVS, PVDIIGRGGSVS, PVDSMGRGGSVS, PVDVIGRGGSVS, and XXX-VMGRGGSVS) were observed (Figure 3). The last seven letters (GRGGSVS) of the signature, which corresponds to the ATP-binding motif, were identical for all the signatures of the *Schistosoma* USPs. Three shared signatures were observed for the two species. However, only one signature (PVDIIGRGGSVS) had the same members as in the protein length grouping (184 aa: Q86DW2, G4LZI3).

Grouping of *Schistosoma* USP sequences by alignment

A multiple sequence alignment of the 13 sequences was generated by ClustalW (Figure 4). The ligand-binding sites (functional sites) annotated in the Conserved Domain Database are labeled with hashes (#) in Figure 4. The alignment revealed gap positions where amino acid residues were missing. These gaps could explain the differences in lengths reported for the sequences. In the USP functional site signatures, the first three letters were not predicted for sequences Q5DI36 and Q5DH64. The multiple sequence alignment showed where series of gaps were inserted by ClustalW to align the 13 sequences. In addition, using Smp_076400 from *S. mansoni* as the reference sequence, there are conserved sites (denoted by ^) of aspartate (Asp; D), leucine (Leu; L), glycine (Gly; G), histidine (His; H) and proline (Pro; P) residues at positions 57, 101, 127, 166

```
Q5DHK1      ----------MAGGSFNYLQINMTSECSRRVLLPIDGSEHSKRAVNWYLTEFCKPDDHTY 50
Smp_031300  ---------------------MASNCMRRVLLPIDGSEHSKRAVNWYLTEFSRPDDFAY 38
Q86DW2      MSEIEQQSTSDGLYIGENKGITEMSKVTRKVLMPVDGSEHSERAFNWYMDNIMKTTDGLY 60
Smp_076400  MSETEQPSTSDGLDIGETKGTISMTDATRKVLMPVDGSEHSERAFNWYMDNVMKITDGLY 60
Q5DI36      --------------------MV------------------------LENMKRDTDCIK 14
Q5DGI9      --------------------MVNESEYSRVILIPIDGSDHCDRAFRWYLENMKRDTDCIK 40
Q5DH64      ---------------------------------------------MKRDTDCIK 9
Smp_043120  --------------------MAESSDPSRVILIPIDGSDHCDRAFRWYLENMKRDTDCIT 40
Smp_097930  --------------------MGVNTENKKVVFLPVDASDHSARAFQWYLDNLRGKNDELH 40
Q5DDH7      -----------MSRRLSTFSKVPPIGSRSVLIAIDGSEHSKKAFDYYVNWLHRPDDSVT 48
Q5DED2      -----------MSRRLSTFSKVPPIGSRSVLIAIDGSEHSKKAFDYYVNWLHRPDDSVT 48
Smp_001000  -----------MERRPSGFSKIPPIGSRSVLIAIDGSEHSKKAFDYYLKWLQRPDDSVT 48
Q86DX1      -----------MN---------TSNRKRTVCLPVDGSEHSKRAVEWFIKEVYRPGDHVL 39
                                                           .    *
                                 ###                      ^
Q5DHK1      FLHVVESHYSKTTAIESHDHAKELSSNLNKNIKSNAQLGKLLGDKLHDDLEKSHIQMEYI 110
Smp_031300  FFHVVEAHYSKSTANESYDHGKELNTNLDKNIKMYSELGKILGDKLHDDLKNSNIQMEYV 98
Q86DW2      LVHIVEPLLPGLNYNLAC-KSPSIKEDFSTHINSLVESGRALRAKFFTRCEESGLTARFT 119
Smp_076400  LVHIVEPLSQGLNYNLAS-KSPSIKDDFSKHLNSLVESGRALRAKFFTRCEDSGLSARFT 119
Q5DI36      FVHVVEPVYSTPPIGLA--DNYT-MPDITKVMEISTENGRKLGQKYIHEAKSYKLSAHAF 71
Q5DGI9      FVHVVEPVYSTPPIGLA--DNYT-MPDITKVMEISTENGRKLGQKYIHEAKSYKLSAHAF 97
Q5DH64      FVHVVEPAYNIPTTGLT--MDLSPVPDMTQALEASIASGKKLGQKYIHEAKSYKLSAHAF 67
Smp_043120  FVHVIEPVYNTPAIGMT--MESPPIPDMTRVMEESIEQGKKLGQKYMHEAKSYKLNAKAF 98
Smp_097930  FVYVIKPIFTTPTIELA--MASSPITDIIQSTQENIENAKKLLQKYLIKAKRFGISCQAF 98
Q5DDH7      IYHAVEPVS--LPTLSLSSPMGIPSSEWSNIVEANVKRVRELENDYSAECLRHNLIYQFL 106
Q5DED2      IYHAVEPVS--LPTLSLSSPM-----------ANVKRVRELENDYSAECLRHNLIYQFL 94
Smp_001000  IYHAVGPVS--LPTISSSNPISIPSEEWSNLVQTNVKRVRELENDYSADCLAHNLTYQFL 106
Q86DX1      FIHSVELPY--LPSVSLTSGLKIPVDDWTKALQENISLTNKLNNEYGYICESKNIPYEFL 97
                 : : :           . *      . :  .
                     #                    ^
Q5DHK1      MQIGNKPGELIVDLIKKLSVDVVLIGNRGLGALRRTFLGSVSEYVLHHCNVPFIIIPPPL 170
Smp_031300  MQIGNKPGELIINVAKERSVDVILIGNRGLGAFRRTFLGSVSEYILHHCNVPFIIIPPSI 158
Q86DW2      IHVGTKPGENIVRLANEHGANLVIIGNRGIGTVKRTFLGSVSDHVLHNVNVPVIIIPPPK 179
Smp_076400  IHVGTKPGENIVRIAHEHGVDLVIIGNRGIGTVKRTFLGSVSDYVLHHANVPVIIIPPPK 179
Q5DI36      LHVDTKPGSSLVKAISEHKADVILMGSRGLGAIRRTFLGSVSDYVLHHAHIPVVIIPPQD 131
Q5DGI9      LHVDTKPGSSLVKAISEHKADVILMGSRGLGAIRRTFLGSVSDYVLHHAHIPVVIIPPQD 157
Q5DH64      LHVDTKPGSSLVKAISEHKADVILMGSRGLGAIRRTFLGSVSDYVLHHAHIPVVIIPPQD 127
Smp_043120  LHVDTKPGSSLVKAISDHKANVILMGNRGLGAIRRTFLGSVSDYVLHHSHIPVVIVPPQE 158
Smp_097930  VHVNAKPGPTLVKFAEEQKADIIIIGPRGLGLIRRTLLGSVTNYVMHHTKTPLVVIPPPV 158
Q5DDH7      YESVEHIGASIIQQVEKYEVRLIVIGSRGLGAIKRTIMGSVSDYVVHHANTAVCVVPSID 166
Q5DED2      YESVDIIGASIIQQVEKYEVRLIVIGSRGLGAIKRTIMGSVSDYVVHHANTAVCVVPSID 154
Smp_001000  YESVDHIGAAIVQNAEKYNVHLLIVGSRGLGAIKRTFMGSVSDYVIHHANTAVCVIPSIT 166
Q86DX1      VKNGSTPGAGIIEACEERPVDLIIMGSRGLGRIKRAIIGSVSSYVVHNSNVPCITVPP-- 155
              .     *  ::    .  . ::::* **:* .:*:::***.:::*: :  .   :*.
                   ^                 ####        ####    ^        ^
Q5DHK1      CS------ 172
Smp_031300  SL------ 160
Q86DW2      HPKKK--- 184
Smp_076400  HPKKK--- 184
Q5DI36      KQ------ 133
Q5DGI9      KQ------ 159
Q5DH64      KQ------ 129
Smp_043120  KQ------ 160
Smp_097930  R------- 159
Q5DDH7      EVECTS-- 172
Q5DED2      EVECTS-- 160
Smp_001000  EQKNSCKN 174
Q86DX1      -------- 
```

Figure 4 Multiple sequence alignment of the sequences of selected universal stress proteins of *Schistosoma mansoni* and *Schistosoma japonicum*.

Notes: The sequence alignment of the 13 sequences with ATP-binding motif [G2XG9XG(S/T)] was generated using ClustalW (Larkin et al).[56] Sequences of *S. mansoni* have "Smp" in the sequence identifier. The ligand binding sites (functional sites), annotated in the Conserved Domain Database (http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml) are labeled with hashes (#). An observation is that aspartate, leucine, glycine, histidine, and proline residues are conserved in all the sequences (denoted by ^). The conserved positions are 57, 101, 127, 166, and 176 in Smp_076400 from *S. mansoni*. The conserved residues could be common functional sites for biochemical or environmental regulation of *Schistosoma* universal stress proteins. Meaning of alignment symbols: "*", residues in column are identical; ":", conserved substitutions; ".", semiconserved substitutions. A visual analytics resource that can be used to interact with data is available at http://public.tableausoftware.com/views/schisto_features_usp/usp_align.

Abbreviations: ATP, adenosine triphosphate; D, aspartate; G, glycine; H, histidine; L, leucine; P, proline; Uniprot, Universal Protein Resource; USP, universal stress protein.

and 176 (Figure 4). The relationship of the sequences was visualized as a phylogenetic tree (Figure 5). Five groups (A to E) of sequences were observed with each of the five *S. mansoni* sequences assigned to a group. Groups A and E contained multiple *S. japonicum* sequences. The bootstrap statistical support value for the branch of Group C (Smp_076400 [G4 LZI3] and Q86DW2) was 100%. The grouping of sequences by the maximum parsimony model was in agreement with the maximum likelihood model (Figure 6).

Dynamic integration of annotation features for *Schistosoma* USPs

The groupings described in the previous sections were based on the primary amino acid sequences of the USPs. To facilitate dynamic integration and updates of the data
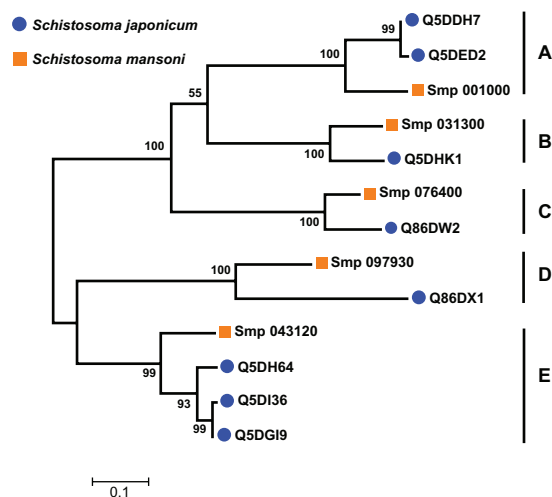
Figure 5 Grouping of 13 *Schistosoma* universal stress protein sequences. The phylogenetic tree was generated with MEGA5 (Tamura et al),[59] using the maximum likelihood method. The 13 *Schistosoma* universal stress protein sequences were clustered in five groups (**A–E**). The numbers near the clades are the statistics from the 1000 bootstrap that support the phylogeny recovery of the clades. A visual analytics resource that can be used to view the image, with other associated data, is available at http://public.tableausoftware.com/views/schisto_features_usp/phylotrees. Sequences of *S. mansoni* have "Smp" in the sequence identifier.

sources, we developed a web-based visual analytics resource. As mentioned previously, the five annotation features with variable frequency per protein were 3D chemical ligand, ligand-binding amino acid, amino acid and sequence position, developmental expression, and subcellular localization. To help guide further research and hypotheses generation, we checked for the phylogenetic groups in which the ligand
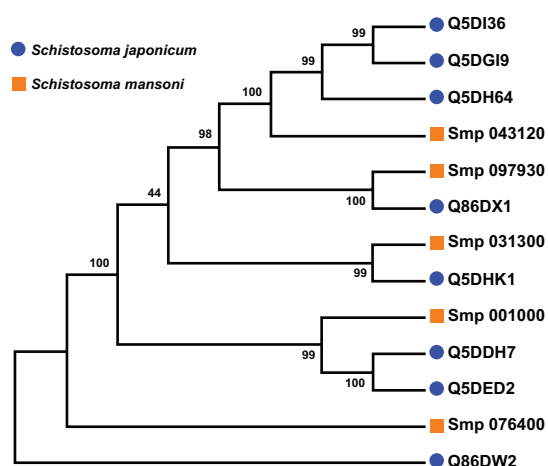
Figure 6 Parsimony test for the phylogeny tree reconstructed for *Schistosoma* universal stress proteins with the maximum likelihood method.
Notes: Maximum parsimony analysis of each of the 1000 bootstrap replications, from maximum likelihood, (Figure 4) determined the percentage of the bootstrap replications in which a particular clade (a node and all of its descendent taxa) was recovered. Those clades, which were recovered close to 100% of the bootstrap replications, indicate confident and statistical support in our analysis. A visual analytics resource that can be used to view the image, with other associated data, is available at http://public.tableausoftware.com/views/schisto_features_usp/phylotrees. Sequences of *S. mansoni* have "Smp" in the sequence identifier.

amino acids were identical in amino acid type and the position of the amino acids. Other features based on the protein sequences that were considered were chemical ligands, kinase type, kinase score, and subcellular localization. The developmental expression feature was not considered in the decision-making process for prioritizing the sequences. This feature is based on extracted data from multiple peer reviewed reports and is incomplete. Nonetheless, the information can assist in directing new research.

Our visual analytics-supported decision-making process prioritized for discussion Group C, which included sequence Q86DW2 from *S. japonicum* and sequence G4LZI3 (Smp_076400) from *S. mansoni* (Figure 7). An integration view for Group D (Q86DX1 and C1M0Q2 [Smp_097930]) is presented to show the differences in patterns of identical annotations when compared with Group C (Figure 8). The biologically relevant chemical ligands predicted to bind to the Group C proteins include four phosphate-containing ligands (adenosine diphosphate [ADP], adenosine monophosphate [AMP], adenosine triphosphate [ATP] and guanosine triphosphate [GTP]) and three metallic ion ligands (calcium [Ca^{2+}], magnesium [Mg^{2+}], and zinc [Zn^{2+}]). The two proteins in the group were also predicted to be (1) localized in the cytoplasm and (2) capable of phosphorylation by phosphokinase C, with a value of 0.93. In the Group C USP genes, there was evidence of gene expression in all the stages by at least one of the genes. The schistomulum stage had the only identical annotation for the developmental gene expression for the two Group C USP genes. A screenshot showing a design that provides an integrated view of the chemical ligands, ligand-binding sites, functional site signature, the presence of ATP-binding motif, kinase type, and kinase score is presented in Figure 9.

Discussion

S. haematobium, *S. japonicum*, and *S. mansoni* are the major human schistosomiasis parasites. These parasites undergo a complex developmental life cycle, in which they encounter a plethora of environmental stressors, such as transition from aerobic to anaerobic environment during the cercarial penetration of the human skin.[62] The presence of genes encoding the USP domain in the genomes of *Schistosoma* spp. suggests these flatworms are equipped to respond to unfavorable conditions that induce USP function.[12,23,25,38] The bioinformatics-based predictions generated a variety of data types, including amino acid functional site, multiple sequence alignment, prediction score, protein domain organization, phylogenetic tree, and sequence length. We used a visual analytics approach
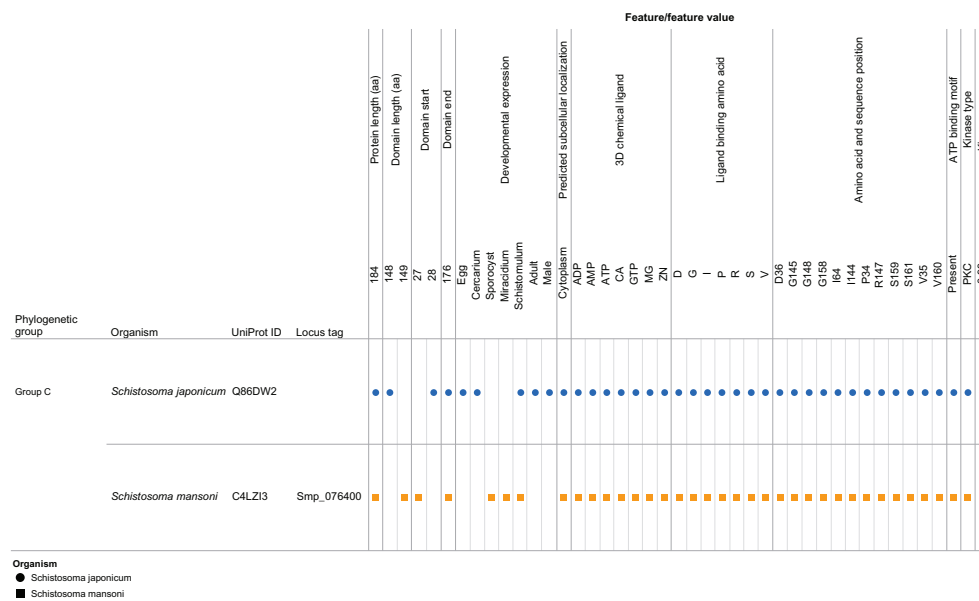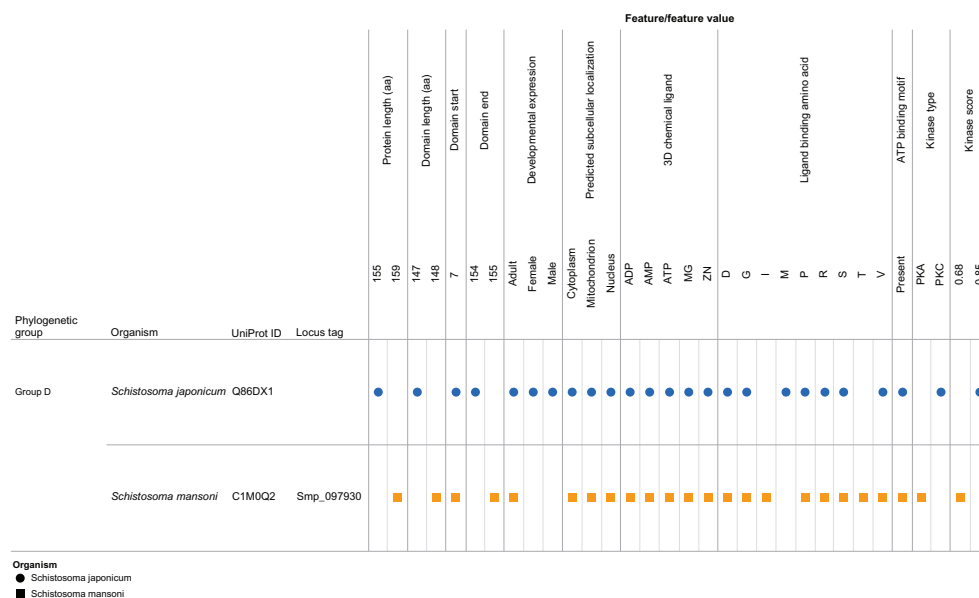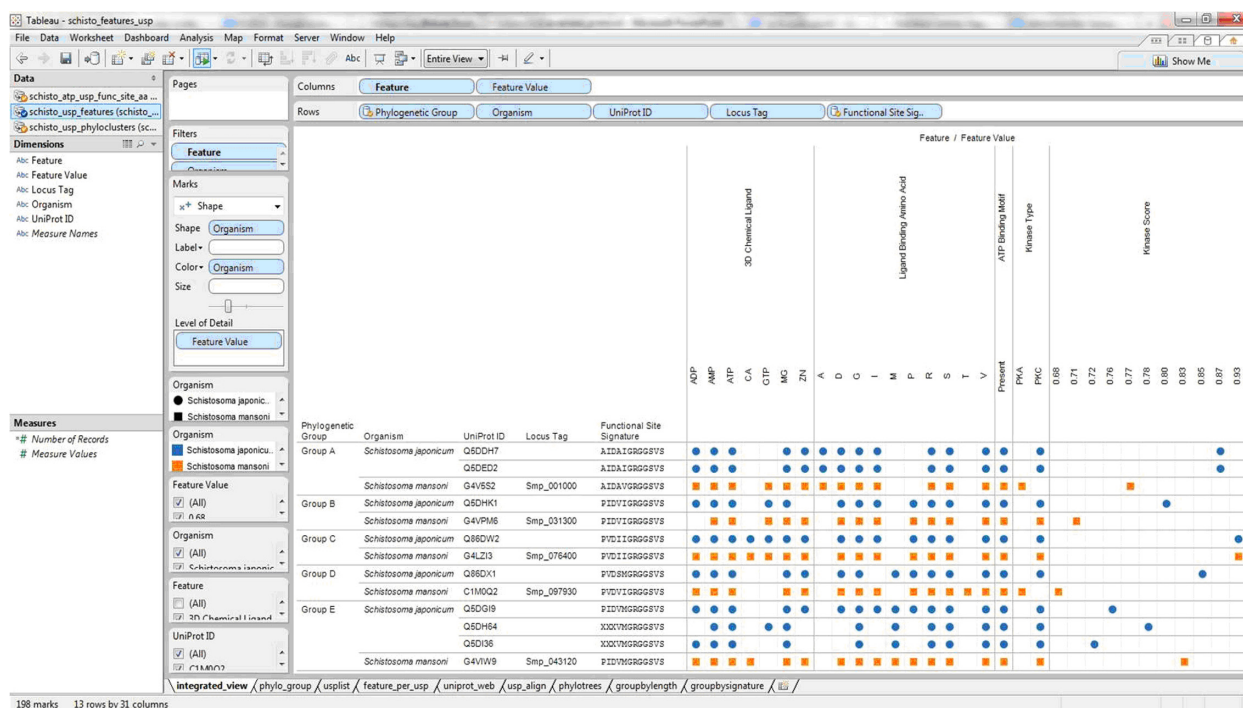
Figure 7 Integration and visualization of the data on the sequence features, evolutionary relatedness, and developmental expression of *Schistosoma* universal stress proteins (Q86DW2 and G4LZI3).

Notes: The integration and visualization design was implemented in the visual analytics software environment (Tableau Software Inc, Seattle, WA, USA). Among the 13 sequences compared, the two 184-amino-acid-long sequences Q86DW2 (Sjp_0058490) and G4LZI3 (Smp_076400) were prioritized for further research. The decision was based on statistical support from the phylogenetic analysis as well as the relatively complete and consistent annotations in the protein sequence length, biologically relevant chemical ligands, and ligand-binding amino acids (amino acid type and amino acid position). A visual analytics resource that can be used to interact with the view is available at http://public.tableausoftware.com/views/schisto_features_usp/phylo_group. Sequences of *S. mansoni* have "Smp" in the sequence identifier.

Abbreviations: ADP, adenosine diphosphate; AMP, adenosine monophosphate; ATP, adenosine triphosphate; CA, calcium; D, aspartate; G, glycine; GTP, guanosine triphosphate; l, isoleucine; Mg, magnesium; PKC, protein kinase C; P, proline; R, arginine; S, serine; UniProt, Universal Protein Resource (Apweiler et al);[47] V, valine; Zn, Zinc.

Figure 8 Integration and visualization of the data on the sequence features, evolutionary relatedness, and developmental expression of *Schistosoma* universal stress proteins (Q86DX1 and C1M0Q2).

Notes: This figure illustrates the decision-making process. In comparison with Q86DW2 and G4LZI3 (Figure 7), the annotations for the protein sequence length, biologically relevant chemical ligands, and ligand-binding amino acids (type and position) were not identical. A visual analytics resource that can be used to interact with the view is available at http://public.tableausoftware.com/views/schisto_features_usp/phylo_group. Sequences of *S. mansoni* have "Smp" in the sequence identifier.

Abbreviations: ADP, adenosine diphosphate; AMP, adenosine monophosphate; ATP, adenosine triphosphate; D, aspartate; G, glycine; GTP, guanosine triphosphate; l, isoleucine; M, methionine; Mg, magnesium; PKA, protein kinase A; PKC, protein kinase C; P, proline; R, arginine; S, serine; T, threonine; UniProt, Universal Protein Resource (Apweiler et al);[47] V, valine; Zn, Zinc.

Figure 9 Design layout and visualization of data sets from the sequence analysis, evolutionary relatedness, and developmental expression of 13 *Schistosoma* universal stress proteins.

Notes: The details of the annotation features are available in the Methods section. The views constructed and data are available for download from an Internet website: http://public.tableausoftware.com/views/schisto_features_usp/integrated_view. The free software Tableau Reader (http://www.tableausoftware.com/products/reader) (Tableau Software Inc) can be used for offline access to the downloaded views and data.

Abbreviations: ADP, adenosine diphosphate; AMP, adenosine monophosphate; ATP, adenosine triphosphate; CA, calcium; D, aspartate; G, glycine; GTP, guanosine triphosphate; I, isoleucine; M, methionine; Mg, Magnesium; PKA, protein kinase A; PKC, protein kinase C; P, proline; R, arginine; S, serine; T, threonine; UniProt, Universal Protein Resource (Apweiler et al);[47] V, valine; Zn, Zinc.

to integrate these data types and to identify orthologous pairs of protein sequences with a protein length of 184 aa USP in *S. mansoni* (Smp_076400) and *S. japonicum* (Locus Tag: Sjp_0058490; UniProt ID: Q86DW2). Gene synteny, obtained from SchistoDB[49] and evolutionary genomics analysis called the *S. mansoni* phylome,[63] indicated that an ortholog (Sha_107834) is encoded in the *S. haematobium* genome. Thus, the genomes of the three major human schistosomiasis parasites encode the 184 aa USP.

Since inferences on chemical and environmental regulation are our interest, we focus the discussion of the results on the findings on the five conserved residues and the chemical ligands predicted to bind to the prioritized protein. All the 13 protein sequences have conserved sites for Asp, Leu Gly, His, and Pro at positions 57, 101, 127, 166, and 176, using Smp_076400 from *S. mansoni* as a reference sequence (Figure 4). These conserved residues did not coincide with any of the predicted ligand-binding sites and could be common functional sites for regulating *Schistosoma* USPs.

The predicted 3D chemical ligands for Smp_076400 and Sjp_0058490 (Q86DW2) included three metal ions Ca^{2+},

Mg^{2+}, and Zn^{2+} (Figure 7). Metal ions are involved in many diverse biochemical reactions,[64] including cellular cofactors for phosphorylation. The UspA protein of *Escherichia coli* undergoes phosphorylation in vitro with its phosphate donors ATP and/or GTP, in the absence of other proteins.[65] The ATP molecule and metallic chemical ligands, such as Mg^{2+} ion, might bind together at the Mg-ATP-binding groove during phosphorylation or ATP-dependent stress-response mechanism.[66,67] The presence of Mg^{2+} ion suggests that it can be an integral and critical component in the reaction.[68–71] This result could be affected if there is any structural conformation in the binding site residues that prevents the Mg^{2+} ion from binding to the ATP molecule at the active groove. The resultant effect might be translated to compromised functional efficiency in binding ATP during phosphorylation and also in keeping the metallic Mg ions unstable in the active groove while it is in contact with the ATP molecule.

Ca^{2+} was predicted to bind to proteins in Group C of the phylogenetic tree (Figures 6 and 7). In *S. mansoni*, Ca^{2+} is considered vital for regulated motor-related activities[72] and also critical for the egg hatching process in fresh water.[73,74]

In the tegument fraction of *S. mansoni*, Ca^{2+} simulated the activity of ATPase in the absence of Mg^{2+}.[75] Further, cyclic adenosine monophosphate and Ca^{2+} work in synergy to regulate the transformation of miracidial to sporocysts.[76] The protein kinase C and Ca^{2+} metabolism regulate the induction of proteolytic enzyme from cercariae, which is vital for modulating the musculature activity of the schistosome.[77,78] A key mechanism for the action of PZQ has been proposed to be the disruption of the Ca^{2+} homeostasis in schistosomes, leading to the large, rapid influx of Ca^{2+} ions into the worm and quick muscular contractions.[41,79–81] Microarray-based transcriptome analysis of the response of the *S. mansoni* PR-1 strain to PZQ has identified genes for cytosolic Ca^{2+} regulation.[82]

Conclusion

S. haematobium, *S. mansoni*, and *S. japonicum* are human parasites that undergo a complex developmental life cycle, in which they encounter a plethora of environmental stressors. Though there are multiple research reports on the developmental regulation of genes encoding USPs in *Schistosoma* spp., knowledge of their biochemical and environmental regulation is still limited. The draft status of the genome sequences of *Schistosoma* spp. also provides possibilities that future revisions could be made to gene prediction and protein annotations. We have used a decision-making strategy, facilitated by visual analytics, to identify USPs in two *Schistosoma* species with shared sequence features and when compared with the other sequences they have relatively complete and consistent annotations. These findings further enabled us to make inferences about the biochemical and environmental regulation of *Schistosoma* USPs. Future research directions could (1) functionally characterize the interactions of Ca^{2+} ions with the amino acid residues of *Schistosoma* USPs; and (2) determine the transcriptional response of *Schistosoma* USP genes to PZQ. The data sets produced, and the visual analytics views developed, can be easily reused to develop new hypotheses.

Acknowledgments

Disclosure

The authors report no conflicts of interest in this work.

References

1. Nahum LA, Mourão MM, Oliveira G. New frontiers in *Schistosoma* genomics and transcriptomics. *J Parasitol Res*. Epub November 21, 2012.
2. Borch M, Kiernan M, Rust K, et al. Schistosomiasis: a case study. *Urol Nurs*. 2009;29(1):26–29.
3. van der Werf MJ, de Vlas SJ, Brooker S, et al. Quantification of clinical morbidity associated with schistosome infection in sub-Saharan Africa. *Acta Trop*. 2003;86(2–3):125–139.
4. Han ZG, Brindley PJ, Wang SY, Chen Z. Schistosoma genomics: new perspectives on schistosome biology and host-parasite interaction. *Annu Rev Genomics Hum Genet*. 2009;10:211–240.
5. Doenhoff M, Hagan P, Cioli D, et al. Praziquantel: its use in control of schistosomiasis in sub-Saharan Africa and current research needs. *Parasitology*. 2009;136(13):1825–1835.
6. Melman SD, Steinauer ML, Cunningham C, et al. Reduced susceptibility to praziquantel among naturally occurring Kenyan isolates of *Schistosoma mansoni*. *PLoS Negl Trop Dis*. 2009;3(8):e504.
7. Berriman M, Haas BJ, LoVerde PT, et al. The genome of the blood fluke *Schistosoma mansoni*. *Nature*. 2009;460(7253):352–358.
8. *Schistosoma japonicum* Genome Sequencing and Functional Analysis Consortium. The Schistosoma japonicum genome reveals features of host-parasite interplay. *Nature*. 2009;460(7253):345–351.
9. Young ND, Jex AR, Li B, et al. Whole-genome sequence of *Schistosoma haematobium*. *Nat Genet*. 2012;44(2):221–225.
10. Gobert GN, McManus DP, Nawaratna S, Moertel L, Mulvenna J, Jones MK. Tissue specific profiling of females of *Schistosoma japonicum* by integrated laser microdissection microscopy and microarray analysis. *PLoS Negl Trop Dis*. 2009;3(6):e469.
11. Gobert GN, Moertel L, Brindley PJ, McManus DP. Developmental gene expression profiles of the human pathogen *Schistosoma japonicum*. *BMC Genomics*. 2009;10:128.
12. Isokpehi RD, Mahmud O, Mbah AN, et al. Developmental regulation of genes encoding universal stress proteins in *Schistosoma mansoni*. *Gene Regul Syst Bio*. 2011;5:61–74.
13. Aragon AD, Imani RA, Blackburn VR, Cunningham C. Microarray based analysis of temperature and oxidative stress induced messenger RNA in *Schistosoma mansoni*. *Mol Biochem Parasitol*. 2008;162(2):134–141.
14. Moertel L, McManus DP, Piva TJ, Young L, McInnes RL, Gobert GN. Oligonucleotide microarray analysis of strain- and gender-associated gene expression in the human blood fluke, *Schistosoma japonicum*. *Mol Cell Probes*. 2006;20(5):280–289.
15. Williams DL, Sayed AA, Bernier J, et al. Profiling *Schistosoma mansoni* development using serial analysis of gene expression (SAGE). *Exp Parasitol*. 2007;117(3):246–258.

16. Liu F, Chen P, Cui SJ, Wang ZQ, Han ZG. SjTPdb: integrated transcriptome and proteome database and analysis platform for *Schistosoma japonicum*. *BMC Genomics*. 2008;9:304.

17. Protasio AV, Tsai IJ, Babbage A, et al. A systematically improved high quality genome and transcriptome of the human blood fluke *Schistosoma mansoni*. *PLoS Negl Trop Dis*. 2012;6(1):e1455.

18. Taft AS, Vermeire JJ, Bernier J, et al. Transcriptome analysis of *Schistosoma mansoni* larval development using serial analysis of gene expression (SAGE). *Parasitology*. 2009;136(5):469–485.

19. Gobert GN, McInnes R, Moertel L, et al. Transcriptomics tool for the human *Schistosoma* blood flukes using microarray gene expression profiling. *Exp Parasitol*. 2006;114(3):160–172.

20. Nachin L, Nannmark U, Nyström T. Differential roles of the universal stress proteins of *Escherichia coli* in oxidative stress resistance, adhesion, and motility. *J Bacteriol*. 2005;187(18):6265–6272.

21. Hingley-Wilson SM, Lougheed KE, Ferguson K, Leiva S, Williams HD. Individual *Mycobacterium tuberculosis* universal stress protein homologues are dispensable in vitro. *Tuberculosis (Edinb)*. 2010;90(4):236–244.

22. Raghavan N, Miller AN, Gardner M, et al. Comparative gene analysis of *Biomphalaria glabrata* hemocytes pre-and post-exposure to miracidia of *Schistosoma mansoni*. *Mol Biochem Parasitol*. 2003;126(2):181–191.

23. Negrão-Corrêa D, Mattos AC, Pereira CA, Martins-Souza RL, Coelho PM. Interaction of *Schistosoma mansoni* Sporocysts and Hemocytes of *Biomphalaria*. *J Parasitol Res*. Epub June 28, 2012.

24. Oswald IP, Eltoum I, Wynn TA, et al. Endothelial cells are activated by cytokine treatment to kill an intravascular parasite, *Schistosoma mansoni*, through the production of nitric oxide. *Proc Natl Acad Sci U S A*. 1994;91(3):999–1003.

25. Loverde PT. Do antioxidants play a role in schistosome host-parasite interactions? *Parasitology Today*. 1998;14(7):284–289.

26. Somero GN. Proteins and temperature. *Annu Rev Physiol*. 1995;57: 43–68.

27. Isokpehi RD, Simmons SS, Cohly HH, Ekunwe SI, Begonia GB, Ayensu WK. Identification of drought-responsive universal stress proteins in viridiplantae. *Bioinform Biol Insights*. 2011;5:41–58.

28. Kim H, Goo E, Kang Y, Kim J, Hwang I. Regulation of universal stress protein genes by quorum sensing and RpoS in *Burkholderia glumae*. *J Bacteriol*. 2012;194(5):982–992.

29. Pineda MC, Turon X, López-Legentil S. Stress levels over time in the introduced ascidian *Styela plicata*: the effects of temperature and salinity variations on hsp70 gene expression. *Cell Stress Chaperones*. 2012;17(4):435–444.

30. Wang H, Lei Z, Li X, Oetting RD. Rapid cold hardening and expression of heat shock protein genes in the B-biotype *Bemisia tabaci*. *Environ Entomol*. 2011;40(1):132–139.

31. Zhang YY, Yang J, Yin XX, Yang SP, Zhu YG. Arsenate toxicity and stress responses in the freshwater ciliate *Tetrahymena pyriformis*. *Eur J Protistol*. 2012;48(3):227–236.

32. Florczyk MA, McCue LA, Stack RF, Hauer CR, McDonough KA. Identification and characterization of mycobacterial proteins differentially expressed under standing and shaking culture conditions, including Rv2623 from a novel class of putative ATP-binding proteins. *Infect Immun*. 2001;69(9):5777–5785.

33. Zarembinski TI, Hung LW, Mueller-Dieckmann HJ, et al. Structure-based assignment of the biochemical function of a hypothetical protein: a test case of structural genomics. *Proc Natl Acad Sci U S A*. 1998;95(26):15189–15193.

34. Freestone P, Nyström T, Trinei M, Norris V. The universal stress protein, UspA, of *Escherichia coli* is phosphorylated in response to stasis. *J Mol Biol*. 1997;274(3):318–324.

35. Freestone P, Trinei M, Clarke SC, Nyström T, Norris V. Tyrosine phosphorylation in *Escherichia coli*. *J Mol Biol*. 1998;279(5):1045–1051.

36. Liu WT, Karavolos MH, Bulmer DM, et al. Role of the universal stress protein UspA of *Salmonella* in growth arrest, stress and virulence. *Microb Pathog*. 2007;42(1):2–10.

37. Raspoet R, Gantois I, Devloo R, et al. *Salmonella enteritidis* universal stress protein (usp) gene expression is stimulated by egg white and supports oviduct colonization and egg contamination in laying hens. *Vet Microbiol*. 2011;153(1–2):186–190.

38. Seifart Gomes C, Izar B, Pazan F, et al. Universal stress proteins are important for oxidative and acid stress resistance and growth of *Listeria monocytogenes* EGD-e in vitro and in vivo. *PLoS One*. 2011;6(9):e24965.

39. Salvador-Recatalà V, Greenberg RM. Calcium channels of schistosomes: unresolved questions and unexpected answers. *Wiley Interdiscip Rev Membr Transp Signal*. 2012;1(1):85–93.

40. Greenberg RM. Are Ca^{2+} channels targets of praziquantel action? *Int J Parasitol*. 2005;35(1):1–9.

41. Coles GC. The effect of praziquantel on *Schistosoma mansoni*. *J Helminthol*. 1979;53(1):31–33.

42. Chabot C. Demystifying visual analytics. *IEEE Comput Graph Appl*. 2009;29(2):84–87.

43. Chang R, Ziemkiewicz C, Green TM, Ribarsky W. Defining insight for visual analytics. *IEEE Comput Graph Appl*. 2009;29(2):14–17.

44. Johnson MO, Cohly HH, Isokpehi RD, Awofolu OR. The case for visual analytics of arsenic concentrations in foods. *Int J Environ Res Public Health*. 2010;7(5):1970–1983.

45. Mackinlay JD, Hanrahan P, Stolte C. Show me: automatic presentation for visual analysis. *IEEE Trans Vis Comput Graph*. 2007;13(6): 1137–1144.

46. Thomas JJ, Cook KA. A visual analytics agenda. *IEEE Comput Graph Appl*. 2006;26(1):10–13.

47. UniProt Consortium. Ongoing and future developments at the Universal Protein Resource. *Nucleic Acids Res*. 2011;39:D214–D219.

48. Hertz-Fowler C, Peacock CS, Wood V, et al. GeneDB: a resource for prokaryotic and eukaryotic organisms. *Nucleic Acids Res*. 2004;32: D339–D343.

49. Zerlotini A, Heiges M, Wang H, et al. SchistoDB: a Schistosoma mansoni genome resource. *Nucleic Acids Res*. 2009;37:D579–D582.

50. Marchler-Bauer A, Lu S, Anderson JB, et al. CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res*. 2011;39:D225–D229.

51. Salsbury FR Jr, Knutson ST, Poole LB, Fetrow JS. Functional site profiling and electrostatic analysis of cysteines modifiable to cysteine sulfenic acid. *Protein Sci*. 2008;17(2):299–312.

52. Wass MN, Kelley LA, Sternberg MJ. 3DLigand Site: predicting ligand-binding sites using similar structures. *Nucleic Acids Res*. 2010;38:W469–W473.

53. Blom N, Sicheritz-Pontén T, Gupta R, Gammeltoft S, Brunak S. Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics*. 2004;4(6):1633–1649.

54. Chou KC, Shen HB. A new method for predicting the subcellular localization of eukaryotic proteins with both single and multiple sites: Euk-mPLoc 2.0. *PLoS One*. 2010;5(4):e9931.

55. Liu F, Lu J, Hu W, Wang SY et al. New perspectives on host-parasite interplay by comparative transcriptomic and proteomic analyses of *Schistosoma japonicum*. *PLoS Pathogens*. 2006;2(4):e29.

56. Larkin M, Blackshields G, Brown NP, et al. Clustal W and Clustal X version 2.0. *Bioinformatics*. 2007;23(21):2947–2948.

57. Tamura K, Nei M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol*. 1993;10(3):512–526.

58. Jones DT, Taylor WR, Thorton JM. The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci*. 1992;8(3): 275–282.

59. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol*. 2011;28(10):2731–2739.

60. Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*. 1985;39(4):783–791.

61. Steel M, Lockhart PJ, Penny D. A frequency-dependent significance test for parsimony. *Mol Phylogenet Evol*. 1995;4(1):64–71.

62. Jolly ER, Chin CS, Miller S, et al. Gene expression patterns during adaptation of a helminth parasite to different environmental niches. *Genome Biol*. 2007;8(4):R65.

63. Silva LL, Marcet-Houben M, Nahum LA, Zerlotini A, Gabaldón T, Oliveira G. The *Schistosoma mansoni* phylome: using evolutionary genomics to gain insight into a parasite's biology. *BMC Genomics*. 2012;13:617.

64. Lu CH, Lin YF, Lin JJ, Yu CS. Prediction of metal ion-binding sites in proteins using the fragment transformation method. *PLoS One*. 2012;7(6):e39252.

65. Freestone P, Nyström T, Trinei M, Norris V. The universal stress protein, UspA, of *Escherichia coli* is phosphorylated in response to stasis. *J Mol Biol*. 1997;274(3):318–324.

66. Buchachenko AL, Kouznetsov DA, Arkhangelsky SE, Orlova MA, Markarian AA. Spin biochemistry: magnetic 24 Mg-25 Mg-26 Mg isotope effect in mitochondrial ADP phosphorylation. *Cell Biochem Biophys*. 2005;43(2):243–251.

67. Buchachenko AL, Kouznetsov DA, Breslavskaya NN, Orlova MA. Magnesium isotope effects in enzymatic phosphorylation. *J Phys Chem B*. 2008;112(8):2548–2556.

68. Luo R, Zhou C, Lin J, Yang D, Shi Y, Cheng G. Identification of in vivo protein phosphorylation sites in human pathogen *Schistosoma japonicum* by a phosphoproteomic approach. *J Proteomics*. 2012;75(3):868–877.

69. Nebl T, Prieto JH, Kapp E, et al. Quantitative in vivo analyses reveal calcium-dependent phosphorylation sites and identifies a novel component of the *Toxoplasma* invasion motor complex. *PLoS Pathog*. 2011;7(9):e1002222.

70. Szabo A, Langer T, Schröder H, Flanagan J, Bukau B, Hartl FU. The ATP hydrolysis-dependent reaction cycle of the *Escherichia coli* Hsp70 system DnaK, DnaJ, and GrpE. *Proc Natl Acad Sci U S A*. 1994; 91(22):10345–10349.

71. Langer T, Lu C, Echols H, Flanagan J, Hayer MK, Hartl FU. Successive action of DnaK, DnaJ and GroEL along the pathway of chaperone-mediated protein folding. *Nature*. 1992;356(6371):683–689.

72. Soares de Moura R, Rozental R, Claudio-da-Silva TS. *Schistosoma mansoni*: effects of bromolysergic acid diethylamide, verapamil, and Ca^{2+}-free solution on the motor activity of the isolated male worm induced by electrical stimulation and oxamniquine. *Exp Parasitol*. 1987;63(2):173–179.

73. Katsumata T, Kohno S, Yamaguchi K, Hara K, Aoki Y. Hatching of *Schistosoma mansoni* eggs is a Ca^{2+}/calmodulin-dependent process. *Parasitol Res*. 1989;76(1):90–91.

74. Katsumata T, Shimada M, Sato K, Aoki Y. Possible involvement of calcium ions in the hatching of *Schistosoma mansoni* eggs in water. *J Parasitol*. 1988;74(6):1040–1041.

75. Cunha V, Noël F. A Mg^{2+}-independent Ca^{2+}-stimulated ATPase activity in the tegument of *Schistosoma mansoni*. *Braz J Med Biol Res*. 1988;21(3):449–451.

76. Kawamoto F, Shozawa A, Kumada N, Kojima K. Possible roles of cAMP and Ca^{2+} in the regulation of miracidial transformation in *Schistosoma mansoni*. *Parasitol Res*. 1989;75(5):368–374.

77. Blair KL, Bennett JL, Pax RA. *Schistosoma mansoni*: evidence for protein kinase-C-like modulation of muscle activity. *Exp Parasitol*. 1988;66(2):243–252.

78. Matsumura K, Mitsui Y, Sato K, Sakamoto M, Aoki Y. *Schistosoma mansoni*: possible involvement of protein kinase C in linoleic acid-induced proteolytic enzyme release from cercariae. *Exp Parasitol*. 1991;72(3):311–320.

79. Doenhoff MJ, Cioli D, Utzinger J. Praziquantel: mechanisms of action, resistance and new derivatives for schistosomiasis. *Curr Opin Infect Dis*. 2008;21(6):659–667.

80. Greenberg RM. Ca^{2+} signalling, voltage-gated Ca^{2+} channels and praziquantel in flatworm neuromusculature. *Parasitology*. 2005; 131 Suppl:S97–S108.

81. Pica-Mattoccia L, Orsini T, Basso A, et al. *Schistosoma mansoni*: lack of correlation between praziquantel-induced intra-worm calcium influx and parasite death. *Exp Parasitol*. 2008;119(3):332–335.

82. Aragon AD, Imani RA, Blackburn VR, et al. Towards an understanding of the mechanism of action of praziquantel. *Mol Biochem Parasitol*. 2009;164(1):57–65.