MOLECULAR REGULATION OF UNIVERSAL STRESS PROTEINS IN ENVIRONMENTALLY MEDIATED SCHISTOSOMIASIS PARASITES

SUMMARY

Human schistosomiasis popularly known as bilharzias in many regions of Africa is a freshwater snail-transmitted disease caused by parasitic flatworms known as schistosomes. The growth and development of schistosomes typically requires developmental stages in multiple hosts and transmission stages in freshwater. These life cycle environments present a plethora of stressors. Certain gene families including heat shock proteins (HSPs/Hsps) and universal stress proteins (USPs) help schistosomes to respond to unfavourable conditions.

The availability of genomes sequences information for *Schistosoma japonicum*, *Schistosoma mansoni* and *Schistosoma haematobium* provide unique research resources to apply bioinformatics analysis of its associated USPs to predict regulatory features from sequence analysis. The objectives of the research were to (i) Infer the biochemical and environmental regulation of universal stress proteins of *Schistosoma* species; (ii) Identify biological function relevant protein sequence and structure features for prioritized universal stress proteins from *Schistosoma* species; (iii) Determine the distinctive structural features of a predicted regulator of *Schistosoma* adenylate cyclase activity that has possible influence on the functioning of universal stress proteins.

The findings revealed that (i) schistosomes USPs are hydrophilic and very reactive in the water environment or in aqueous phase, which seems

adaptive with their immediate environment and developmental stages; (ii) The functions of Smp_076400 and Sjp_0058490 (Q86DW2) are regulated by conserved binding site residues and metallic ions ligands (Ca2+, Mg2+ and Zn2+), particularly Ca2+ predicted to bind to both USPs; (iii) The *S. mansoni* life cycle and stress resistance pathway protein (Smp_059340.1) is regulated by Ser53, Thr188, Gly210 and Asp207 residues. The overall scope has highlighted the role of bioinformatics in predicting exploitable regulatory features of schistosome universal stress proteins and biological pathways that might lead to identification of putative functional biomarkers of common environmental diseases. The findings of this research can be applicable to other areas of environmental health and environmental genomics.

KEYWORDS: adenylate cyclase, ATP binding protein, calcium, chemical ligands, environmental stressors, biomolecular regulators, functional sites, praziquantel, *Schistosoma*, schistosomiasis and universal stress proteins.

MOLECULAR REGULATION OF UNIVERSAL STRESS PROTEINS IN ENVIRONMENTALLY MEDIATED SCHISTOSOMIASIS PARASITES

by

ANDREAS NJI MBAH

submitted in accordance with the requirements for the degree of

DOCTOR OF PHILOSOPHY

in the subject

ENVIRONMENTAL SCIENCES

at the

UNIVERSITY OF SOUTH AFRICA

SUPERVISOR: DR R D ISOKPEHI

CO-SUPERVISOR: PROF O R AWOFOLU

SEPTEMBER 2013

MOLECULAR REGULATION OF UNIVERSAL STRESS PROTEINS IN ENVIRONMENTALLY MEDIATED SCHISTOSOMIASIS PARASITES

By

ANDREAS NJI MBAH

AIBMS, MSc

Master of Science (MSc) in Medical Genetics with Bioinformatics (Double Master Degree), Institute of Cancer Research Brunel University, Uxbridge, London, UK (2004)

Bachelor of Science (B.Sc. Hon's) in Medical Microbiology, University of Calabar, Calabar, Nigeria (1996)

Professional Biomedical Scientist Certification (Grade AIBMS) and Membership Institute of Biomedical Science of United Kingdom, Registration No. 00373780 (2004)

DECLARATION

I declare that this thesis titled **"MOLECULAR REGULATION OF UNIVERSAL STRESS PROTEINS IN ENVIRONMENTALLY MEDIATED SCHISTOSOMIASIS PARASITES"** is my own work carried out at the Department of Environmental Sciences, University of South Africa, and all the sources that I have used or quoted have been indicated and acknowledged by means of complete references.

Signature & Date  **09/9/2013**

Student: Mr Andreas Nji Mbah **STUDENT NO: 4898-921-5**

DEDICATION

This research is dedicated to my lovely wife Comfort Ashunombi Nji-Mbah and my friend Tita Teboh Tadius Joviale.

ABSTRACT

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* species suggests that these flatworms are equipped to respond to unfavourable environmental conditions. The USPs encompass a conserved group of proteins that are found in archaea, eubacteria, yeast, fungi and plants.

Though data on gene expression is available for *Schistosoma* USP genes, their biochemical and environmental regulation are incompletely understood. Adenosine TriPhosphate (ATP) is a known regulator of functioning of USPs through phosphorylation. In schistosomiasis parasites, an enzyme adenylate cyclase catalyzes the conversion of ATP to cyclic Adenosine MonoPhosphate (cAMP). Adenylate cyclase and cAMP modulate stress resistance and virulence in prokaryotic and eukaryotic pathogens. Since ATP is a regulator of USP functioning, understanding the regulators of adenylate cyclase could provide new insights on biological pathways for regulating the growth, survival and infectivity of schistosomiasis parasites. Interestingly, a Guanosine TriPhosphate (GTP) protein with sequence identifier Smp_059340 has been identified in the *Schistosoma mansoni* genome and predicted to have adenylate cyclase-stimulating activity. Though this *Schistosoma* GTP-

binding protein is annotated as a drug target for schistosomiasis, knowledge on amino acid residues for interaction with adenylate cyclase is limited.

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. The availability of the genome sequences of *Schistosoma* species provides opportunities for bioinformatics analysis of its associated USPs to predict regulatory features from sequence analysis.

The overall research goal is to predict the functioning of universal stress proteins of schistosomiasis parasites during unfavourable environmental conditions. The purpose of the research therefore, is to identify sequence features and chemical ligands that are associated with the functioning of universal stress proteins in schistosomiasis parasites. The hypothesis of the research is that the application of bioinformatics methods will help predict functional residues and chemical ligands that are associated with the universal stress proteins of schistosomiasis parasites during unfavourable environmental conditions.

The objectives designed to test the hypothesis are: (1) Infer the biochemical and environmental regulation of universal stress proteins of *Schistosoma* species; (2) Identify biological function relevant protein sequence and structure features for prioritized universal stress proteins from *Schistosoma* species; and (3) Determine the distinctive structural protein features of a predicted regulator of *Schistosoma* adenylate cyclase activity that has possible influence on the functioning of universal stress proteins.

Overall, the methods for the research involved the deployment of sequence analysis and structural bioinformatics tools to accomplish the objectives. Visual analytics approach were integrate to the methods to enhance decision making on the results obtain from the bioinformatics tools. A total of 13 *Schistosoma* USP sequences and one *Schisitosoma* GTP binding protein were analyzed. Objective 1 was achieved by developing a protocol that integrates visual analytics stages to facilitate the interaction with the results from sequence analysis and data collection on a set of universal stress proteins from *Schistosoma mansoni* and *Schistosoma japonicum*. In objective 2 and objective 3, a suite of sequence analysis and structural bioinformatics tools were used to determine protein physicochemical properties, homology models of protein structures and protein domain organization.

Multiple sequence alignment identified conserved sites that could be key residues regulating the function of the 13 universal stress proteins of *Schistosoma* species (five *S. mansoni* and eight *S. japonicum* sequences). The computation of physicochemical parameters showed that schistosome USPs had high aliphatic index values and conspicuously very low grand average hydropathy (GRAVY). These observations indicate that these USPs are very reactive in the water environment or in aqueous phase, which seems consistent with their immediate environment and developmental stages.

A 184 amino acid long USP sequence present in the three human *Schistosoma* species was prioritized based on overall agreement of bioinformatics results. Thus two sequences (Q86DW2 [*S. japonicum*] and G4LZI3 [*S. mansoni*]) were further investigated to gain insight functional

residues and chemical ligands that could modulate *Schistosoma* USP function. Both had more hydrophilic amino acids which are usually located at the surface and active sites of proteins. The hydrophilic amino acids could be contributing to the reactivity and life cycle adaptation in aqueous environment. The predicted 3D chemical ligands for Q86DW2 and G4LZI3 modulating their functions included three metallic ions of Ca^{2+}, Mg^{2+} and Zn^{2+}. The ligand binding sites of both G4LZI3 and Q86DW2 were made up in part by the same residues of the ATP binding motif (Gly145, Arg147, Gly148, Gly158, and Ser159). The adaptive structural conformation predicted, indicates possible functional efficiency in binding ATP during phosphorylation and stress response mechanism. The conserved binding sites and the metallic ions could be the key features regulating the molecular mechanism of stress response by schistosomiasis parasites.

Homology modeling and conserved domain analyses of *S. mansoni* GTP-binding and adenylate cyclase stimulating protein (Smp_059340) predicted that the key residues for interaction (stimulation and inhibition) with adenylate cyclase. In particular, the residues Ser53, Thr188, Gly210 and Asp207 were identified for further research on the function of the GTP-binding protein.

The bioinformatics analysis has proved vital in predicting exploitable regulatory features of the stress response relevant proteins from *Schistosoma* species. This research revealed that the physicochemical properties of USPs could be related to their aqueous environment of developmental stages. The key features modulating schistosome USP could be the conserved binding site

residues and the metallic ions, particularly calcium ion (Ca^{2+}) predicted for homologs in *S. japoncum* and *S. mansoni*. The functional regulatory residues predicted from the Smp_058340.1 protein provides a basis for the design of experiments such as site–directed mutagenesis to determine impacts of residues on adenylate cyclase activity particularly (i) developmental processes and (ii) stress resistance.

Given that the initial effects of praziquantel on schistosomes include influx of calcium ions, this research proposed additional investigations to (i) functional characterize the interaction of calcium ions with amino acid residues of *Schistosoma* USPs; and (ii) determine the transcriptional response of *Schistosoma* USP genes to praziquantel. The datasets produced and the visual analytics views developed can be easily reused to develop new hypotheses. However, because the distinctive characteristic features are predictions, additional research may be required to confirm the predictions. This research, in its overall scope has highlighted the role of bioinformatics in elucidating exploitable regulatory features and biological pathways that could lead to the identification of putative functional biomarkers of common environmental diseases. The findings of this research can be applicable to other areas of environmental health and environmental genomics.

KEYWORDS: adenylate cyclase, ATP binding protein, calcium, chemical ligands, environmental stressors, biomolecular regulators, functional sites, praziquantel, *Schistosoma*, schistosomiasis and universal stress proteins

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS AND ACRONYMS

Amino acids, one and three letter codes

Amino acid	Three letter code	One letter code
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Asparagine or aspartic acid	Asx	B
Cysteine	Cys	C
Glutamic acid	Glu	E
Glutamine	Gln	Q
Glutamine or glutamic acid	Glx	Z
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

Other abbreviations used

AA	Amino acid
Acc	Accuracy
ADP	Adenosine diphosphate
AMP	Adenosine monophosphate
ATP	Adenosine triphosphate
ATP-BP	Adenosine triphosphate- Binding Protein
BLAST	Basic Local Alignment Search tool
B	Beta
C	Carbon atom
cDNA	Complimentary deoxyribonucleic acid
Ca or CA	Calcium atom
CASP	Critical Assessment of techniques for protein Structure Prediction
cAMP	cyclic Adenosine MonoPhosphate

CDD	Conserved Domain Database
CLUSTALW	Multiple sequence alignment program for DNA or proteins
CD-HIT	Cluster Database at High Identity with Tolerance
COG database	Clusters of orthologous group Database
DNA	Deoxyribonucleic acid
DnaJ	Heat shock protein J, chaperone protein DnaJ
3D	3 Dimension
E.coli	Escherichia coli
ESTs	Expressed Sequence Tags
Euk-mPLoc 2.0	Eukaryotic protein subcellular localization multiple site prediction server
FASTA	A text based format for representing either nucleotide sequences or peptide sequences using single letter codes
FMN	Flavin Mononucleotide
FPR	False positive rate
FP	False positive
FN	False negative
GRAVY	Grand average hydropathy
Gαs	G protein alpha -s- subunit
GDP	Guanine diphosphate
GTP	Guanine triphosphate
GNTD	Global Neglected Tropical Disease
HMMs	Hidden Markov Models
Hsps	Heat shock proteins
Hsp70	Heat shock proteins 70
Hsp40	Heat shock proteins 40
H. influenza	Haemophilus influenzae
H	Hydrogen atom
H202	Hydrogen peroxide
IMG/M	Integrated Microbial Genomes/Metagenomes
KNN	K-nearest neighbor algorithms
LIBSVM	A library of support vector machine
ATPase	Enzyme that catalyze the conversion of ATP to ADP and a free phosphate ion
LTR	Long terminal repeat
mRNA	Messenger Ribonucleic acid
MEGA5	Molecular Evolutionary Genetics Analysis tool version 5
Mg or MG	Magnesium atom
MCC	Mathews Correlation Coefficient
M. tuberculosis	Mycobacterium tuberculosis
MSA	Multiple Sequence Alignment
ML	Maximum likelihood
NTD	Neglected Tropical Disease
NCBI	National Center for Biotechnology Information
NOS	Nitrogen species
N	Nitrogen atom
O	Oxygen
PZQ	Praziquantel

Pi	Reactive phosphate or free phosphate ion
PKA	Protein Kinase A
PKA-R	Protein Kinase A regulatory unit
PKA-C	Protein Kinase A catalytic unit
PCR	Polymerase Chain Reaction
PSSM	Position-specific score matrice
Pfam database	Protein families database
PSI-BLAST	Position Specific terative- Basic Local Alignment Search tool
PRK database	Protein K(c)lusters database
PISCES server	Protein Sequence Culling Server
PROFEAT serve	Protein Features server
pI	Isoelectric point of proteins
rRNA	ribosomal Ribonucleic acid
ROS	Reactive oxygen species
ROC curve	Receiver Operating Characteristics curve
RPS-BLAST	Reverse PSI-BLAST
S. japonicum	*Schistosoma japonicum*
S. mansoni	*Schistosoma mansoni*
S. hematobium	*Schistosoma hematobium*
Smp_001000	*Schistosoma mansoni* protein_001000
SchistoDB	*Schistosoma* Database
SMART	Simple Modular Architecture research tool
S	Sulfur atom
SOPMA tool	Self-Optimized Prediction Method with Alignment
SjTPdb	Integrated transcriptome and proteome database and analysis platform for Schistosoma japonicum
SVM	Support Vector Machine
SOPM	Self-Optimized Prediction Method
TMPRED server	Protein Transmembrane Region and Orientation Prediction server
TMbase	Database of transmembrane proteins and their helical membrane-spanning domains
TP	True positive
TN	True negative
TIGRFAM database	A database of protein families designed to support manual and automated genome annotation
USP, Usp	Universal stress protein
UspA	Universal stress protein A
UV	Ultraviolet
UniProtKB	Universal Protein Knowledgebase
WHO	World Health Organization
ɣ	gamma
ZN or Zn	Zinc atom

ACKNOWLEDGEMENTS

The undertaking of a Ph.D. relies on the help and encouragement of many. Therefore, I wish to recognize these individuals without whom this dissertation would not have come to pass, and – perhaps more importantly – without whom it simply would not have been such an enjoyable experience.

First and foremost, I would like to extend my thanks to my dedicated mentors Dr. Raphael. D. Isokpehi, Prof. O. R. Awofolu and Prof S.M. Moja. They are brilliant scientists and very amiable men. Dr. Isokpehi has always given me the necessary freedom to pursue my scientific interests and consistently supported me with his invaluable advice. The inputs from Dr. Isokpehi, Prof. S.M. Moja and Prof. Awofolu are truly appreciated. Dr. Isokpehi has time and again shown his dedication to the role as a committed supervisor by his encouragements, his expert advice, and by sharing his expert insights and ideas. Dr. Isokpehi brought unique perspectives to my research, enriching it greatly.

In addition, I would like to thank colleagues for their support. The following individuals: Angelique Lee, Lee Campbell, Ousman Mahmud, Shaneka Simmons, Dr. Wellington K. Ayensu, Dominique Smith-McInnis, Baraka S. Williams, Charnia Hall and Christina Bernard, all contributed to the advancement of this science in one way or the other. My appreciation also goes to the enthusiastic learning attitude of all my trainee students at the Center for Bioinformatics & Computational Biology (CBCB) at the Jackson State University, Mississippi, USA. They contributed immensely by providing a

pleasant and vibrant atmosphere for conducting research. These students repeatedly transferred considerable parts of their impressive knowledge during our interesting conversations and training sessions. I also consider it a rare privilege to have been able to work with such enthusiastic and determined students.

I am indebted to my friends, Matilda O. Johnson, Samson Emeakpor, Akame Roland, Nkafu Julius, Kons Nkafu, Achu Waziri and Kafui Edusei for their support throughout the research period. They gave me their unreserved support and confidence. They always provided me unwavering love and encouragement, thank you for believing in me.

One of the great things about being a scientist is that you get the opportunity to build collaborations and friendships with bright and engaging people from all over the world. Consequently, I would like to thank Dr. Oyekanmi Nashiru from the University of Ibadan, Nigeria, Dr Henri Kanga and Dr Emmanuel Nfor an expert in inorganic toxicology from the University of Buea, Cameroon. Their expertise and unfailing sense of humor have enabled the creation of exciting synergies in our research focus today and for posterity.

Life is of course more than work itself, but for a researcher that distinction sometimes seems to fade. My family and my family-in-laws have repeatedly helped me across this fine line with great interest in my work, progress and commitment to science. They will be remembered with smiles of appreciation and all deserve a heart touching and sincere ‘thank you’ here.

It is important to note that most people will only remember the first and last persons mentioned in the acknowledgements. It is for this reason only that

I mention here at the very end my wife Comfort Ebob Nji-Mbah, who has stood by me through all these years, displaying endless and almost inhuman patience and understanding. For consistently being the light of my life and the dearest friend imaginable, I would like to conclude by extending my most cordial thanks to her for being there for me.

CHAPTER ONE

INTRODUCTION

1.1. Overview

Human schistosomiasis popularly known as bilharzias in many regions of Africa is a freshwater snail-transmitted disease caused by parasitic flatworms known as schistosomes (Gobert et al., 2009b, Jolly et al., 2007). The major *Schistosoma* species that infect humans are *Schistosoma japonicum*, *Schistosoma mansoni* and *Schistosoma haematobium* (Gryseels, 2012).

Schistosomiasis is an ancient disease reported in the pharaonic Egypt and has been found in human remains over 2000 years old from China (Adamson, 1976, Zhou et al., 2005). Schistosomiasis is estimated to infect about 210 million people in 76 countries within Africa, Asia, the Middle East and South America, despite strenuous control efforts (Steinmann et al., 2006) and is second only to malaria in public health significance (Borch et al., 2009).

The *Schistosoma* spp. undergo a complex developmental life cycle that includes multiple morphological stages and transition between hosts (Figure 1). 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 (Chai et al., 2006, Grabe and Haas, 2004). 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 (Negrao-Correa et al., 2012, Raghavan et al., 2003). 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 (Loverde, 1998, Oswald et al., 1994). In summary, all the developmental stages of the *Schistosoma* species are exposed to various stresses.

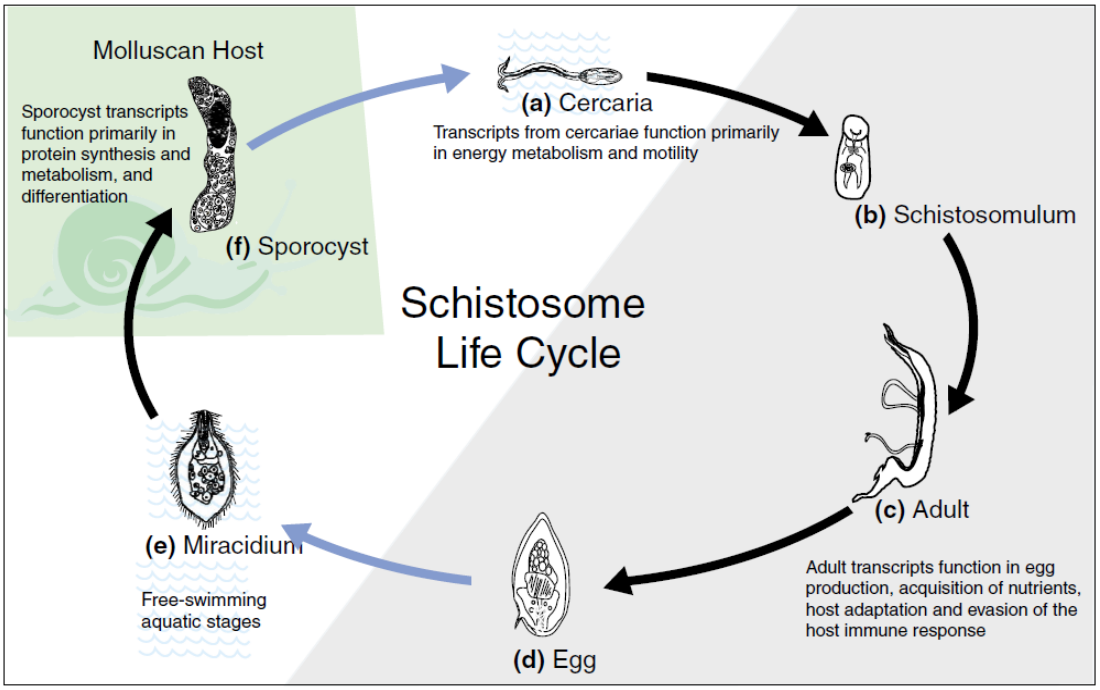


Figure 1: Life cycle of *Schistosoma* species

The life cycle of *Schistosoma* species is complex, with morphologically distinct stages occupying several ecological niches. The infective cercariae stage **(a)** swim in fresh water to find and then infect the mammalian host. After host invasion, cercariae transform into schistosomula **(b)** and adapt to survival in the host bloodstream. Then schistosomula mature into adult male or female schistosomes **(c)**, which pair and produce eggs **(d)**. Eggs are excreted from the host. In fresh water, the eggs hatch into miracidia **(e)**, which infect a snail host and develop into sporocysts **(f).** Daughter sporocysts generate infectious cercariae, completing the life cycle (Jolly et al., 2007) .

A common aspect of these environment-inducing stresses is that they result in proteins with nonnative conformations (Somero, 1995). Inducible stress tolerance has increasingly been understood to result from numerous molecular mechanisms involving proteins such as heat shock proteins (Hsps) and Universal Stress Proteins (Isokpehi et al., 2011a, Isokpehi et al., 2011b, Kim et al., 2012, Pineda et al., 2012, Wang et al., 2011, Zhang et al., 2012). The schistosomes may respond to the environmental stressors by inducing the expression of stress response proteins through the activation of various intracellular signaling pathways.

Genes that encode the Universal Stress Protein (USP) domain, commonly referred to as Universal Stress Proteins, are found in diverse group of organisms including archaea, eubacteria, yeast, fungi and plants (Isokpehi et al., 2011a, Isokpehi et al., 2011b, Kerk et al., 2003, Tkaczuk et al., 2013). The accession identifier in the Protein Family (Pfam) database (http://pfam.sanger.ac.uk) is PF00582. The protein family encompasses a conserved group of proteins whose expressions are triggered by a large variety of environmental insults including toxic chemicals, drought, and extreme temperature (Kerk et al., 2003, Nachin et al., 2008, Sousa and McKay, 2001).

Binding of Adenosine triphosphate (ATP) through phosphorylation is a biochemical mechanism that regulates the function of USPs (Zarembinski et al., 1998). 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 (Sousa and McKay, 2001, Zarembinski et al.,

1998). The 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 (Gustavsson et al., 2002, Kvint et al., 2003). 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 (Freestone et al., 1997). 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 (Nachin et al., 2005, Schreiber et al., 2006, Schweikhard et al., 2010, Seifart Gomes et al., 2011).

The adenylate cyclase in the schistosomiasis parasites regulates ATP conversion to cAMP to power developmental processes and other biological activities in the worm. The biological response of cAMP is one of the main environmental sensing machineries associated with the stress response in a wide range of unicellular eukaryotes such as *Trypansosoma spp*, *Plasmodium spp* and other organisms (Kaushal et al., 1980). The cAMP dependent activity is ideally suited to the complex environment that pathogenic *Schistosoma* experiences. In *Mycobacterium smegmatis*, a universal stress protein (MSMEG_4207) acts as an acetylation substrate for a cAMP-regulated acetyltransferase (MSMEG_5458) (Nambi et al., 2010). Further, because of the tight binding affinity between the USP and the enzyme, the USP has been proposed as a regulator of the cAMP-regulated acetyltransferase. Additionally, acetylation of the USP occurred only in the presence of cAMP. Thus, cAMP

could be an important molecule in the biology of schistosomes especially in modulating the activity of USPs.

Adenylate cyclase and cyclic AMP modulate stress resistance and virulence in many prokaryotic and eukaryotic pathogens (Kohut et al., 2010, Mattila et al., 2009, Mazzucchelli and Sassone-Corsi, 1999, Donovan et al., 2013, Salmon et al., 2012, Hanawa et al., 2013). Therefore, to gain a better understanding of the mechanism regulating the growth, survival and infectivity of schistosomes in their immediate hostile environments, studies aimed at unraveling control points in adenylate cyclase pathway regulatory protein are needed.

A *Schistosoma mansoni* Guanine nucleotide binding protein with sequence identifier Smp_059340 has been predicted to have adenylate cyclase-stimulating activity (Berriman et al., 2009). This protein has been listed as an attractive drug target (Crowther et al., 2010). The possible link of Smp_059340 to availability of cAMP for the functioning of universal stress proteins of schistosomes makes research on this GTP-binding protein more compelling.

1.2. Motivation

1.2.1. Problem Statement

The genomes of *S. haematobium*, *S. mansoni*, and *S. japonicum* encode proteins with the universal stress protein (USP) domain (Pfam Identifier: PF00582) (Berriman et al., 2009, *Schistosoma japonicum* Genome Sequencing Consortium, 2009, Young et al., 2012). The USPs are known to function during unfavourable environmental conditions, including the life cycle

developmental stages in *Schistosoma* species (Gobert et al., 2009a, Gobert et al., 2009b, Isokpehi et al., 2011a). Though data on gene expression is available for genes encoding USPs (USP genes) (Aragon et al., 2008, Gobert et al., 2006, Gobert et al., 2009a, Gobert et al., 2009b, Gobert et al., 2010, Isokpehi et al., 2011a, Liu et al., 2008, Moertel et al., 2006), 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.

1.2.2. Rationale for the Research

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 (Borch et al., 2009). 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 (van der Werf et al., 2003). Schistosomiasis 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 (Borch et al., 2009, Han et al., 2009).

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 (Doenhoff et al., 2008, Doenhoff et al., 2009, Melman et al., 2009).

This research will highlight the role of bioinformatics in elucidating the biological regulatory mechanisms of USPs of schistosomiasis parasites at the molecular level during environmental stress response. If this approach succeeds in predicting a model that reasonably represents the biochemical reality, the payoff is large in both formulating models for environmental control of schistosomiasis and the identification of putative new functional biomarkers with possible novel drug targets and discovery. The contributive expert knowledge of this research is expected to have application to the less investigated subject of environmental toxicogenomics and other aspects of environmental health.

1.3. Research Goal, Purpose, Hypothesis and Objectives

The overall research goal is to predict the functioning of universal stress proteins of schistosomiasis parasites during unfavourable environmental conditions.

The purpose of the research therefore, is to identify sequence features and chemical ligands that are associated with the functioning of universal stress proteins in schistosomiasis parasites.

The hypothesis of the research is that the application of bioinformatics methods will help predict functional residues and chemical ligands that are associated with the universal stress proteins of schistosomiasis parasites during unfavourable environmental conditions.

The objectives designed to test the hypothesis are:

(1) Infer the biochemical and environmental regulation of universal stress proteins of *Schistosoma* species.

(2) Identify biological function relevant protein sequence and structure features for prioritized universal stress proteins from *Schistosoma* species.

(3) Determine the distinctive structural protein features of a predicted regulator of *Schistosoma* adenylate cyclase activity that has possible influence on the functioning of universal stress proteins.

CHAPTER TWO

LITERATURE REVIEW

2.1. Environmental Stressors for Living Organisms

A fundamental characteristic of a living organism is responsiveness to its immediate environment. When this environment is hostile or unfavourable for survival the organism must move or adapt to the environmental condition. These unfavourable conditions include extreme of temperatures, exposure to toxic chemicals, toxic gases, hydrostatic pressure, radioactive substances, toxic substances of human origin, nutrient starvation, drought, high salinity, chemotherapeutic agents, amino acid analogue, transitional heavy metals, oxidant injury and inhibitors of energy metabolisms (Gobert et al., 2009b, Kerk et al., 2003, Nachin et al., 2008, Sousa and McKay, 2001). The net effect of stress to the organism is to denature its proteins and induce a stress response mechanism (Feder and Hofmann, 1999).

Activation of various intracellular signaling pathways will result in the expression of genes and protein activity that assist in response to stress conditions (Kim et al., 2012, Pineda et al., 2012, Wang et al., 2011). Sufficiently intense stressors will induce loss of the native conformations of proteins (Somero, 1995). Organisms might tolerate these inducible stresses through numerous mechanisms including the expression of heat shock proteins (Hsps) and universal stress proteins (USPs) (Kim et al., 2012, Pineda et al., 2012, Wang et al., 2011).

There are other mechanisms associated with cellular stress responses such as production of polyols and trehalose, modifications of the saturated cell membrane lipids (homeoviscous adaptation), compensatory expression of isozymes or allozymes of significant enzymes, metabolic arrest and radical scavengers (superoxide dismutase, glutathione system, cytochrome P450) (Feder and Hofmann, 1999). Therefore the unambiguous attribution of stress resistance to a particular set of stressors requires more than a correlative evidence (Shilova et al., 2006), implying that nature will seldom undergo a unique stress at any time (Feder and Hofmann, 1999).

In a geographical habitat, the organisms present might be exposed to temperature variations that might vary between −100 °C to greater than 100 °C. In addition, there are other environmental parameters such as extremes of chemical and gas concentration, food and water availability, hydrostatic pressure, radiation and other toxic substances of human and non-human origin, suggesting that stress proteins expression can be regarded as a common phenomenon in life (Feder and Hofmann, 1999, Feder and Krebs, 1997). Some organisms respond to these immediate environmental changes by movement and/or other behaviors which enable them to seek out favourable microhabitats (Thomson et al., 2010).

2.2. Parasitic Environment and Its Related Stressors

Proteins for response to stress during host-parasite interactions have been investigated from both clinical and biological perspectives (Feder and Hofmann, 1999). These parasites express stress response proteins as cellular

defense mechanisms, adapting them to diverse environmental stressors during their developmental stages (Van der Ploeg et al., 1985).

The expressions of Hsps in parasite life cycle differ both in quantity and type of genes that the perturbation induces. In *Trypanosoma brucei* , mRNA transcripts for Hsp70 and Hsp83 have been shown to gain a 100 fold elevation when the parasite leaves the tsetse fly and enters a mammalian host (Van der Ploeg et al., 1985). While in *S. mansoni* the aquatic snails release cercariae in fresh water which can express two heat-inducible proteins that are not present in other stages (Neumann et al., 1993), but limited research have been conducted on universal stress proteins expressed at various life cycle stages (Isokpehi et al., 2011a).

Developmental regulation of the expression of Heat Shock Proteins (Hsps) are well documented in pathogens such as parasitic nematode (van Leeuwen, 1995), the malaria-causing organisms (*Plasmodium* species) (Syin and Goldman, 1996)*, Borrelia burgdorferi*, the etiological agent of Lyme disease (Carroll et al., 2001), the protist *Leishmania* (Hubel et al., 1997), *Trypanosoma cruzi* (Giambiagi-deMarval et al., 1996) and *Theileri* (Daubenberger et al., 1997).

Other organisms investigated include *Histoplasma capsulatum* which might experience temperature shift during host infection leading to the expression of Hsps (Maresca, 1995). The parasitic *Eimeria* expresses Hsp90 during their infective life cycle stages and infect diverse hosts such as marine fish, poultry and cattle; all having wide body temperature variations (Feder and Hofmann, 1999). The specificity of infection is conspicuous at the species

level. The *Eimeria bovis* is known to have the cattle as its exclusive host (Clark et al., 1996). The evolutionary relationship of *Eimeria* Hsps and host species is not clear (Feder and Hofmann, 1999). However as nature permits, ectothermic parasites suffer internal temperature shifts when their host body temperature varies as documented for parasites of reptiles, fish and other organisms (Joseph et al., 2010).

2.3. The Developmental Stages and Its Stressors

At different developmental stages such as gametogenesis, embryogenesis and metamorphosis some organisms may express striking characteristics ranging from an outright stress response to no stress response at different developmental stages (Feder and Hofmann, 1999, Dix, 1997). These stress response patterns are mostly parallel with the enhanced stress resistance noticed with the developmental stages under enormous stress or in conditions such as dormancy or diapauses (Feder and Hofmann, 1999).

A common theme is that some stress response proteins are not expressed in early embryogenesis or late gametogenesis due to the harmful effect of some stress response proteins to the developing cell (Gagliardi et al., 1995, Ovakim and Heikkila, 2003). Thus the parental Hsp mRNAs can circumvent the absence of gametic or embryonic heat shock gene expression (Gordon et al., 1997). In other cases this absence presumably presents a significant problem for the continuous development in a stressful environment (Michel et al., 1992). There is evidence that the common effects of stress to the early developmental stages are outright death or phenocopying (Feder and

Hofmann, 1999, Welte et al., 1995), a process which can be minimized by Hsps.

The developmental expressions of stress response proteins are diverse, with some plants seed requiring extreme and challenging conditions before germination. However, the seeds can clearly undergo developmental regulation of stress response proteins within the embryos in response to environmental stressors (Wehmeyer et al., 1996). Some investigators have considered whether these patterns of expression are amplified or modified in species and ecotypes which naturally encounter very challenging stress regimes (Helm et al., 1989). This evidence correlates with fungal spores expressing Hsp under a regulated developmental program (Sanchez et al., 1992).

The commonly studied animal case is the encysted brine-shrimp (*Artemia*) whose embryo undergoes developmental arrest and survives for years in the absence of water or oxygen by accumulating enormous amount of small Hsps (Liang et al., 1997) and trehalose (Clegg and Jackson, 1992). Immature *Drosophila melanogaster* can be under threat when they feed on harmful necrotic fruits under sunlight (Clegg and Jackson, 1992, Krebs RA, 1997). Their developmental stages express large amounts of Hsps to tolerate natural heat stress (Krebs RA, 1997, Flannagan et al., 1998). During seasonal variations such as winter, some flies express large amounts of Hsps for cold adaption and then uses ubiquitin to degenerate flight muscles after nuptial flight (Davis et al., 1994).

2.4. Universal Stress Proteins (USPs) Family

The Universal Stress Proteins (USPs; Pfam accession number PF00582) encompass a conserved group of proteins that are found in archaea, eubacteria, yeast, fungi and plants (Kvint et al., 2003). Their expressions are triggered by a large variety of environmental stressors (Kvint et al., 2003, Nachin et al., 2005). The population of USPs present in the different organisms varies considerably (Kvint et al., 2003). For example, *Escherichia coli* have five small USPs and one tandem-type, whereas *Streptomyces coelicolor* has eight USPs, five of which are tandem-type USPs. *Halobacterium sp*. NRC-1 has one of the largest numbers of USP genes per bacterial genome (14 USP genes). Though the USPs are widely distributed among various species, information relating to their physiological and biochemical properties have been derived from their study in *E. coli* (strain K-12), whose genome encodes 6 USPs (Gustavsson et al., 2002).

The first member of the *E. coli* USPs to be discovered was a small cytoplasmic protein of 144 amino acids whose expression is greatly enhanced when the bacterium growth is inhibited by a large variety of immediate environmental stress conditions such as starvation of the following nutrients: carbon, nitrogen, phosphate, sulfate and the required amino acid. Also the presence of a variety of toxic agents including heavy metals, oxidants, acids, antibiotics, heat shock, DNA damage, phosphate, uncouplers of the electron transport chain, polymyxin, cycloserine, ethanol, and antibiotics have been documented (Diez et al., 2000, Nystrom and Neidhardt, 1992). The participation of this protein in a multitude of starvation events and stressors

has earned it the title of universal stress protein A, or UspA (Nystrom and Neidhardt, 1992).

Considering that the expression of USPs is induced by a wide variety of stressors, inactivation of UspA was found to impair the survival of *E. coli* under a range of diverse growth arrest conditions (Nystrom and Neidhardt, 1994). There have been 5 more members added to the USP family in *E. coli* K-12, these include: UspC (YecG), UspD (YiiT), UspE (YdaA), UspF (YnaF), and UspG (YbdQ or UP12) (Gustavsson et al., 2002). Much interest has developed in unraveling the molecular mechanisms underlying the apparent general "stress endurance" activities exhibited by USPs.

The DNA binding properties have been suggested as having a role to play in the stress endurance of UspA (Mushegian and Koonin, 1996), however, the evidence for this is still circumstantial. There is no mutation effect of UspA on protein expression patterns during steady-state growth, but starvation-induced growth arrest. Mutation of UspA will alter its global expression, while over-expression of UspA causes global changes in the pattern of protein expression (Nystrom and Neidhardt, 1996). This finding linked UspA with a direct role in gene regulation. However, it is also possible that inactivation or overexpression of UspA may affect global gene expression indirectly. For example, inappropriate expression of UspA in exponentially growing cells results in an instant decrease in their growth rate (Nystrom and Neidhardt, 1996) which, in itself, would affect global gene expression.

The UspA and other USP mutants of *E. coli* are sensitive to DNA damage caused by UV exposure (Gustavsson et al., 2002). This may indeed

point to a protective mechanism mediated by specific or non-specific binding of USPs to DNA. However, it is also possible that an indirect process involving other proteins mediate such DNA protection. Finally, it was initially predicted that UspA has structural similarity to human serum response factor (MefA2) and transcriptional regulators of the MADS-box protein family based on the sequence similarity between UspA and the DNA binding helix of these proteins (Mushegian and Koonin, 1996). After the elucidation of the three-dimensional structure of the UspA orthologue of *Haemophilus influenzae*, the attribution of UspA with the DNA binding domain of these proteins has been discounted (Sousa and McKay, 2001).

Further research is required to elucidate whether the USPs bind DNA and if so, what is the nature and function of such binding? With the possibility that USPs have protein–protein binding activities, one report described findings whereby UspA was found to form homodimers when purified and also to associate with even larger (50 kDa) complex in native *E. coli* protein extracts (Gustavsson et al., 2002). These findings agree with similar findings within the bacterium *Methanocaldococcus jannaschii* USP, MJ0577, which has been reported to form homodimers (Zarembinski et al., 1998). Also the UspG of *E. coli* has been found in complexes with GroEL isolated from stationary-phase cultures (Bochkareva et al., 2002). This GroEL-UspG interaction was confirmed by co-immunoprecipitation experiments and provided experimental evidence of USPs taking part in protein-protein interactions.

Coupled with its upregulation of response due to growth arrest, UspA of *E. coli* also undergoes phosphorylation on serine and threonine residues (Freestone et al., 1997). The phosphorylation is seen as the intracellular switch which activates UspA to perform its cellular functions. If this is the case, it seems that as well as having its expression regulated by growth phase indicators, other cellular factors could modulate the activity of USPs by controlling their phosphorylation state. UspA undergoes phosphorylation in vitro with its phosphate donors ATP and GTP in the absence of other proteins (Freestone et al., 1997). Evidence from genetic studies proved that, the *in vivo* phosphorylation depends on the presence of ribosomal tyrosine phosphoprotein TypA (Freestone et al., 1998).

Experiments conducted with the USP MJ0577 have shown that it binds tightly to ATP (Zarembinski et al., 1998). It was noticed that MJ0577 on its own cannot hydrolyse ATP and it was concluded that an additional cellular factor was required for ATP hydrolysis. The 3 dimensional structure of MJ0577 with bound ATP has been resolved and the residues which make up the triphosphate binding loop identified. A sequence alignment of MJ0577 revealed the conservation of many of the ATP binding residues across a large proportion of proteins classified as members of the universal stress protein family (Zarembinski et al., 1998). When UspA of *H. influenzae* was crystallized in the presence of ATP, no bound nucleotide was found in its structure, which was in contrast to MJ0577 (Sousa and McKay, 2001). Superimposition of UspA on the MJ0577 backbone, revealed the replacement of glycine residues in the triphosphate binding loop with glutamine and methionine. It was

concluded that the presence of the latter more bulky residues in the triphosphate binding loop would impair ATP binding.

The combination of this structural and sequence alignment data has led to the suggestion that one essential element for ATP binding is a G-2X-G-9X-G(S/T) motif which is present at residue positions 127 to 141 of MJ0577 (Zarembinski et al., 1998). Variation in sequence of this motif seen in UspA of *H. influenzae* has been suggested to account for the apparent lack of ATP binding in UspA of *H. influenzae* (Sousa and McKay, 2001). This motif variation suggests that members of this protein family will segregate into two groups, based on whether or not they bind ATP. By implication, one subset has ATP dependent function, while another subset undergoes ATP-independent activities.

Conserved residues which interact with the ATP molecule exist in other positions on MJ0577, in particular, aspartate at position 13 and valine at position 41 (Zarembinski et al., 1998). It has been found that addition of ATP to GroEL complexes releases UspG and conversely; the presence of ATP blocks the interaction of UspG with GroEL (Bochkareva et al., 2002). This interaction may be regarded as solely effect of ATP on GroEL, but the possibility that ATP binding with USPs could influence their protein-protein interactions cannot be ruled out completely. With limited knowledge on the precise role of ATP binding, it may well prove to be an important function of USPs. The open reading frame immediately upstream of UspA in *E. coli* encodes a protein of 111 amino acid residues. This protein was named universal stress protein B (UspB) because of its general responsiveness to different starvation and stress

conditions and the location of its structural gene beside UspA (Farewell et al., 1998). In stationary phase, overexpression of UspB causes cell death while UspB mutants are sensitive to ethanol but not to environmental heat exposure. UspB is predicted to be an integral membrane protein having at least one membrane-spanning domain (Farewell et al., 1998). It is important to point out, however, that UspB was discovered before the emergence of the USP family and it bears no sequence similarity to UspA or other members of the PF00582 family. Members of this ancient and conserved family of proteins are found in all forms of life and can be induced by a variety of environmental stresses (Diez et al., 1997). However, the roles of USP proteins in microbial pathogenesis are incompletely understood.

Given the broad range of resistance functions conferred by the USPs of *E. coli*, it is not surprising to find them in a variety of both pathogenic and non-pathogenic genome of other organisms, including the genome of *Schistosoma* species. The specific roles of USPs in stress response have been elucidated in a number of organisms. In the periodontal pathogen *Porphyromonas gingivalis*, UspA was reported to be an important factor in biofilm formation. Biofilm formation is suggested to be a form of *P. gingivalis* response to environmental stress (Chen et al., 2006).

UspA functions in metabolic and oxidative stress resistance in the enteric bacterium *Salmonella typhimurium*. The protein levels of UspA in *Salmonella typhimurium* were found to be elevated in the stationary phase and when the organism is exposed to high temperatures (30°C to 42°C) (Liu et al., 2007). While under the exact types of survival conditions thought to exist in a

cystic fibrosis lung, USPs have been reported to play an important role in the stationary phase survival of the opportunistic pathogen *Pseudomonas aeruginosa* (Schreiber et al., 2006).

The majority of USPs found in the causative agent of tuberculosis, *Mycobacterium tuberculosis* are reported to be induced in hypoxic conditions, suggesting that they play an important role in the survival of the organism during growth arrest (Hingley-Wilson et al., 2010). The USP Rv2623 has been reported to regulate bacillary growth through its ATP-binding capacity in *M. tuberculosis* (Drumm et al., 2009). In the halophilic proteobacterium (*Halomonas elongate)*, USP TeaD regulates the internal ectoine concentration of the organism upon hyperosmotic stress (Schweikhard et al., 2010). Ectoine is an active ingredient of skin care products and functions as a superior moisturizer (Graf et al., 2008) .

The USP domain may appears as a single domain in small USP proteins (~14-15 kDa), as a tandem domain in larger USP proteins (~30 kDa), or as one or two USP domains fused together with other functional domains (Nachin et al., 2005). Functional domains commonly fused to USP domains include antiporter, voltage channels, amino acid permeases, and protein kinase domains. Members of the USP gene family can be categorized into two groups based on the presence of ATP binding motif in their amino acid residues (Nachin et al., 2005). ATP-binding is a molecular mechanism to regulate the function of universal stress proteins (Zarembinski et al., 1998).

Recently, Foret and co-authors (Foret et al., 2011) conducted phylogenomic and *in situ* expression investigations of USP genes distributed in

diverse metazoans genomes. They observed that (i) a unique phylogenomic pattern that reflects at least five independent losses and multiple independent expansions (ii) presence of the USP gene in the common metazoan ancestor and (iii) spatial expression of *Hydra* USP genes in the endodermal epithelium, a highly potent chemical barrier for protection against intruding microbes.

2.5. Genomic Information on *Schistosoma* species

Two decades ago, the lack of genomic information prompted an international meeting to discuss the need for schistosome genome-related research. This gave birth to the *Schistosoma* Genome Network (http://www.genedb.org/Homepage/Smansoni) (Berriman et al., 2009), (http://www.genedb.org/Homepage/Sjaponicum) (*Schistosoma japonicum* Genome Sequencing Consortium, 2009) and initiated the global gene discovery projects for both *S. japonicum* and *S. mansoni*. Over the past decade, this initiative had extended to other helminth-specific genome sequences due to the improvement in techniques for obtaining biological material, extracting RNA and DNA, constructing complimentary DNA (cDNA)/whole genome shotgun libraries and also other advances in DNA sequencing such as second generation sequencing machines and the concomitant decreasing cost (Blaxter et al., 1999). The helminth genomics in general began with the generation and analysis of transcribed sequences (Expressed Sequence Tags) (ESTs) (Franco et al., 1995), which has proved to be a very rapid and cost-effective route to discover genes in other eukaryotes.

Genomic information of other helminth parasites can be exploited in investigating the genome of schistosomes. For example, the subset of genes

evaluated with gene ontology programs for hookworm can provide insights into cellular and metabolic pathway functioning of these parasites (Mitreva et al., 2005). Furthermore, potential targets for interventions could be identified by applying a hierarchy of considerations including a matrix of biological, expression, and phenotypic data (McCarter, 2004) or by performing a pan-phylum analysis to identify conserved parasite-specific genes whose selective targeting will have low or no toxicity to the host (Wasmuth et al., 2008) or genes that have diverged enough from their host counterpart, resulting in altered or absent functions (Wang et al., 2009). By April 2009, there were about 550,000 nematode ESTs and 450,000 platyhelminth ESTs in the dbEST division of GenBank, excluding those from the model nematode *Caenorhabditis elegans* (Brindley et al., 2009). Of these, 60% were from parasites of humans and closely related animal pathogens used to study human infections.

Schistosoma japonicum and *S. mansoni* are members of the Lophotrochozoa. The Lophotrochozoa is a large taxon that includes about 50% of all metazoan phyla including the mollusks, annelids, brachiopods, nemerteans, bryozoans, playthelminths, and others (Dunn et al., 2008). The genome sequence of *S. mansoni* was determined by whole genome shotgun approach and assembled into 5,745 scaffolds greater than 2 kb, of total size 363 megabases (Mb) pairs, encoding at least 11,809 genes with an average gene size of 4.7 kilobases (kb) (Brindley et al., 2009). Although 40% of this genome is repetitive, 50% percent is assembled into scaffolds of at least 824.5 kb. The *Schistosoma mansoni* genome is made up of eight chromosomes (7 autosomal, plus ZW sex determination pairs) which constitute 43% of the

genome as a whole (Short et al., 1979). Genes encoding protein kinases, proteases, neuropeptides, G-protein couple receptors, ligand-gated ion channels and voltage-gated ion channels which are considered drug and vaccine targets are reported to be present in the genome of *S. mansoni* (Berriman et al., 2009).

The genome of *S. japonicum* contains 397 megabase pairs with 13,469 protein-coding genes identified. The genome is also arrayed on eight pairs of chromosomes, seven pairs of autosomes and one pair of sex chromosomes (*Schistosoma japonicum* Genome Sequencing Consortium, 2009). Genes encoding proteases, transporters, including apolipoproteins, low-density lipoprotein receptor, scavenger receptor, fatty-acid-binding protein, ATP-binding-cassette transporters and cholesterol esterase have been found in the genome of *S. japonicum* (*Schistosoma japonicum* Genome Sequencing Consortium, 2009).

Schistosome Expressed Sequence Tags (ESTs) have many applications. They can be used to annotate schistosome genomes to determine alternative splicing, verify open reading frames and confirm exon/intron and gene boundaries. They are also valuable in functional genomics to design probes for gene expression microarray experiments (Farias et al., 2011) and to provide putative protein sequence information for proteomics methods (Liu et al., 2006) amongst other applications. Quantitative analysis of ESTs (transcriptomics), including Serial Analysis of Gene Expression (SAGE), can identify transcripts that are either over- or under-represented by comparison to other transcripts in various schistosome life

cycle stages or tissues (Gobert et al., 2010). Sequences associated with Expressed Sequence Tags (ESTs), Serial Analysis of Gene Expression (SAGE) Tags and microarray probes are functional genomics tools appropriate for dissection of gene functions (Han et al., 2009). These genomics tools can provide expression data to facilitate forward and reverse genetic approaches including RNA Interference (RNAi) (Mourao et al., 2009).

The availability of the draft genome sequence of *S. mansoni* has led to the development of microarray slides based on oligonucleotides from gene sequences, which can be used to investigate the developmental expression of *S. mansoni* genes. For example, there is a comprehensive microarray dataset composed of experimental series of genes expressed during the life cycle stages using a Puerto Rican strain of *S. mansoni* (Fitzpatrick et al., 2009). However SAGE has been used for gene-expression analysis in a many organisms such as *Rattus norvegicus* (Madden et al., 1997), *Saccharomyces cerevisiae* (Velculescu et al., 1997), *Caenorhabditis elegans* (Jones et al., 2001), *Drosophila melanogaster* (Jasper et al., 2001), *Cryptococcus neoformans* (Steen et al., 2002) and many others. Analysis on human parasites include *Plasmodium falciparum* (Patankar et al., 2001), *Giardia lamblia* (Palm et al., 2005) and *Toxoplasma gondii* (Radke et al., 2005). In the year 2007, the first *Schistosoma* SAGE-library was prepared from the adult stage of the parasitic flatworm *Schistosoma mansoni* to view its transcriptome (Ojopi et al., 2007).

2.6. Universal Stress Proteins of Schistosomaisis parasites

Isokpehi et al. 2011 investigated the eight genes encoding the Universal Stress Protein domain (Pfam accession number PF00582) in the draft *Schistosoma mansoni* genome (Isokpehi et al., 2011a). The locus identifiers for the genes are Smp_076400, Smp_097930, Smp_031300, Smp_043120, Smp_001010, Smp_001000, Smp_136890, and Smp_136870. Also 10 USP genes with Universal Protein Resource (UniProt) identifiers: Q86DX1, Q5DED2, Q5DHK1, Q5D136, Q5DDH7, Q5DH64, Q5DG19, Q86DW2, Q5DGK3 and Q5BTE6) are encoded by the *S. japonicum* genome. The Universal Stress Protein sequence and domain information extracted from the preliminary data analysis are shown in Table 1 and Table 2 below.

The *S. mansoni* USP genes have been analyzed in the context of a phylogenomic analysis of metazoan USPs (Foret et al., 2011), in addition to the rich information for schistosome genome, as integrated in the SchistoDB database (http://www.schistodb.net). The information include the functional annotation of gene expression data documented on various developmental stages from high-throughput Expressed Sequences Tags (ESTs), Serial Analysis of Gene Expression (SAGE) and microarray methods.

These gene expression data present rich genomic resources to elucidate regulation of *S. mansoni* USP genes. Combining the evidence for developmental expression based on public domain EST and SAGE, it is evident that there is transcription of USP genes in at least one of the life cycle stages of the helminth (Figure 2) (Isokpehi et al., 2011a). In microarray experiments to determine the mode of action of praziquantel (PZQ) (Aragon et

al., 2009), there was induction of genes annotated with the Gene Ontology 'antioxidant category'. Thus treatment with PZQ may result in a molecular response similar to that observed when schistosomes undergo oxidative stress.

Table 1: Accession Identifiers for *Schistosoma mansoni* genes encoding universal stress protein domain

Locus Tag	Old UniProtKB ID	New UniProtKB ID	ATP binding motif	No. of USP domain	USP Domain coordinate	Protein length (aa)	USP Domain length (aa)
Smp_001000	C4PWX3	G4V5S2	Present	Single	15 to 163	174	148
***Smp_001010**	C4PWX4	Deleted	deleted	Deleted	deleted	deleted	deleted
Smp_031300	C4Q5U1	G4VPM6	Present	Single	6 to 155	160	149
Smp_043120	C4Q8U4	G4VIW9	Present	Single	7 to 155	160	148
Smp_076400	C4QHX3	G4LZI3	Present	Single	27 to 176	184	149
Smp_097930	C1M0Q2	C1M0Q2	Present	Single	7 to 155	159	148
****Smp_136870A**	C4Q4M5	G4VHD1	Absent	domain A	12 to 130	**	118
****Smp_136870B**	C4Q4M5	G4VHD1	Absent	domain B	137 to 286	**	149
Smp_136890	C4Q4M7	G4VHC9	Absent	Single	13 to 128	132	115

Notes: The seven *Schistosoma mansoni* Universal stress protein domains coordinates are displayed including their new UniProtKB IDs from January 2012. The *Smp_001010 was deleted from UniProtKB database as of December 2011 and will not be included in the analysis. # The USP gene **Smp_136870 is the only sequence with double domain architecture indicated with domain fragments A and B and combine protein length of 290 aa.

Gobert and co-authors using gene expression microarrays determined the profile of developmental gene expression in the different life cycle stages of *S. japonicum*. In another investigation a combined laser microdissection microscopy and microarray analysis approach was used to define the expression sites of *S. japonicum* genes (Gobert et al., 2009a).

The gene encoding the USP protein Q5DED2 was differentially expressed in the egg and miracidia life cycle stages. The gene encoding the USP protein Q5DGI9 was found to be differentially expressed in the ovary and vitellarium of the adult female of *S. japonicum.* Additionally, Q5DGI9 was differentially expressed between the Philippine (SJP) and Chinese (SJC)

strains of *S. japonicum* (Moertel et al., 2006). In another microarray analysis of *S. japonicum* and *S. mansoni* transcriptomes; Q5DGI9 was also differentially expressed (Gobert et al., 2006). Analysis of gene expression profiles for 8 USPs of *S. japonicum* (Q5DHK1, Q86DX1, Q86DW2, Q5DGK3, Q5BTE6, Q5DH64, Q5DED2, and Q5DGI9) from SjTPdb (Liu et al., 2008) (Figure 3); revealed the expression of USPs in all the lifecycle stages examined except the miracidium.

Table 2: Accession Identifiers for *Schistosoma japonicum* genes encoding universal stress protein domain

UniProtKB ID	ATP binding motif	No. of USP domain	USP domain coordinate	Protein length (aa)	Domain length (aa)
Q5DHK1	Present	Single	18 to 167	172	149
Q86DX1	Present	Single	7 to 154	155	147
Q5DI36	Present	Single	3 to 128	133	125
Q5DDH7	Present	Single	15 to 163	172	148
Q86DW2	Present	Single	28 to 176	184	148
Q5DED2	Present	Single	15 to 151	160	136
Q5DGI9	Present	Single	8 to 154	159	146
Q5DGK3	Absent	Single	13 to 162	166	149
Q5DH64	Present	Single	2 to 124	129	122

Notes: The nine *Schistosoma japonicum* universal stress protein domains coordinate information and their ATP binding functionalities. Their accession IDs have not change since deposited in UniProtKB database as from 2009. All species USP genes have single domain coordinates. The Q5BTE6 will not be included in the study because it lacks a proper USP domain architecture.

In the developmental egg stage only Q86DX1 was not expressed. Q5DH64 and Q5DGI9 were expressed in all lifecycle stages except the miracidium. Q5DED2 was only expressed in the egg stage while Q86DX1 was expressed in the all mature stages of development (adult, male and female stages).

Figure 2: Developmental stage expression of *S. mansoni* USP genes.

Red represents expression being detected and green represents no evidence of expression. The SchistoDB IDs are indicated on the right side. Mir: Miracidia, Spo: Sporocyst, Cer: Cercariae, Sch: Schistosomula, Mal: Male, Fem: Female, and Adu: Adult (Isokpehi et al., 2011a).

In a microarray experiment from the developmental stages from *S. japonicum*, the induction of the antioxidant thioredoxin occurs during egg and adult male stages (Gobert et al., 2009a). These results support the need to further evaluate and quantify the transcription regulation of USPs of schistosomiasis parasites in response to oxidative and other environmental stressors during its developmental stages. Therefore in addition to developmental stages in the human host, interventions directed to life cycle stages in the environment as well as in the freshwater snail intermediate hosts could provide additional methods to control human schistosomiasis.

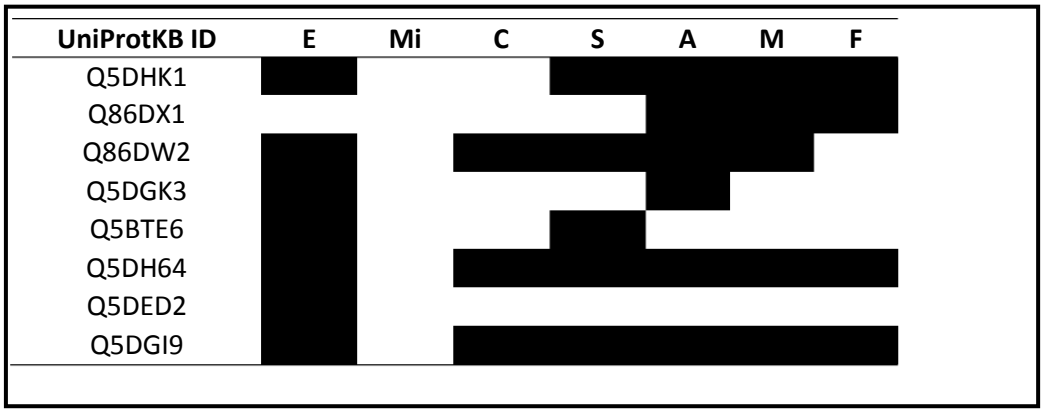


Figure 3: Developmental stage expression of *S. japonicum* USP genes

E: Egg, Mi: Miracidium, C: Cercariae, S: Schistosomulum, A: Adult, M: Male, F: Female. Data was obtained from SjTPdb. Black and white boxes represent the presence or absence of gene expression of the 8 USP genes of *S. japonicum* during the various stages of development in the lifecycle respectively (Liu et al. 2008).

2.7. The cAMP-dependent Protein Kinase A (PKA) in *Schistosoma*

Many eukaryotes, protozoans and helminth parasites extensively use protein kinases to control cellular functions; consequently these kinases may serve as possible novel targets for anti-parasitic drugs (Dissous et al., 2007, Doerig et al., 2008). Possible protein kinase targets in parasitic infections are the cyclic guanosine monophosphate- (cGMP-) dependent protein kinases (PKGs) of *Toxoplasma* (Donald et al., 2002), *Eimeria* (Gurnett et al., 2002) and *Plasmodium* cyclic adenosine monophosphate- (cAMP-) dependent protein kinase (PKA) (Knockaert et al., 2000).

Their inhibitions have elaborated significant anti-parasitic effect both *in vivo* and *in vitro* (Gurnett et al., 2002, Syin and Goldman, 1996). The cAMP-protein kinase (PKA) is regulated by cyclic nucleotide acting as second messengers, produced by purine nucleotide cyclases. Not only the kinase

domains could be potential drug targets, the regulatory domains could also be attractive targets as well (Taylor et al., 2008).

The PKA shows a notable and consistent difference with other protein kinases across the highly divergent taxonomical group (Shoji et al., 1981, Uhler et al., 1986, Walsh et al., 1968). In PKA, the regulatory and catalytic activities are controlled by separate gene transcripts namely PKA-R(CBN) and PKA-C respectively (Knighton et al., 1991). In other kinases the catalytic nucleotide binding (CNB) sites and catalytic domains are usually conjugated in the same polypeptide. Therefore, the inactive conformation of PKA is a heterotetramer of two PKA-R and two PKA-C subunits, whereas in the other kinases such as PKG, homodimers are the inactive conformation (Diller et al., 2001).

The possibility of having both the catalytic and regulatory functions of PKA under the control of two gene products can contribute to functional diversification of PKA (Su et al., 1995). Thus different PKA-C and PKA-R isoforms will be able to combine and producing holoenzymes with variant functions (Taylor et al., 1990). There are many pka-c and pka-r genes in the mammalian genomes, which give rise to different holoenzymes formation (Scully and Ward-Booth, 1995).

Cyclic nucleotide-dependent kinases have been extensively characterized in a variety of eukaryotic organisms and several parasites as mentioned above but yet, there are scanty data available on the role of these kinases in the biology of schistosome universal stress proteins. The adenylyl cyclase also called adenylate cyclase and PKA inhibitors have a negative

effect on the movement of miracidial in a dosage dependent manner, indicating a role for PKA in miracidial locomotion in the environment (Matsuyama et al., 2004). Similarly, agonists of adenyly cyclase were found to prohibit the development of miracidium to mother sporocyst, while chemicals that decrease cAMP levels trigger this transformation (Kawamoto et al., 1989). These findings implicate cAMP and PKA as key players in the larval stage development of schistosomes.

Apart from acting as signal transducing enzymes, protein kinases could also be targeted as novel chemotherapeutics for the schistosomes and other parasitic pathogens (Dissous et al., 2007, Doerig, 2004). The cAMP-dependent kinases (PKAs) are positioned as the major mediators of cellular cAMP signaling in eukaryotic cells. They contribute massively to diverse biological processes such as gene expression, apoptosis, tissue differentiation and cellular proliferation (Taylor et al., 1990). Part of this function occurs through phosphorylation of protein substrates at serine/threonine residues (Taylor et al., 2005). However there are some data available on the characterization of the PKA catalytic domain in *S. mansoni* (Swierczewski and Davies, 2009).

PKA is required for the viability of the adult schistosome *in vitro*. The inhibition of the expressed gene results in the death of the parasite. The PKA-C subunit has a complete ATP binding site containing the motif Gly–2X–Gly–9X - Gly–X–Val) and a serine/ threonine kinase active site containing the motif Arg - Asp–Asp– Leu–Lys–X–X–Asn (Hanks, 2003). Active PKA is expressed in adult *S. mansoni* and the adult schistosomes also express regulatory

proteins that control PKA activation through cAMP, such as adenyly cyclase and PKA regulatory subunits (Swierczewski and Davies, 2009).

The PKA gene is developmentally regulated and its transcripts were detected in all the life cycle stages of *S. mansoni* with the cercariae and adult females expressing the highest level of transcripts. Quantitative PCR analysis of the extra-mammalian stages of the life cycle shows that PKA-C transcripts were expressed at the highest levels in sporocysts and cercariae. The cercariae show more than five-fold higher expression compared to the sporocysts (Swierczewski and Davies, 2009) (Figure 4A). A conclusion from these observations is that the PKA-C expression is required throughout the parasitic life cycle.

The intra-mammalian stages show that adult females have the highest level of expression. It was found to be close to 20-fold higher compared with adult males and the newly transformed schistosomula (Figure 4B). Furthermore, information gleaned through chemical inhibition analysis show that PKA is an essential signaling component in the life cycles of a number of eukaryotic pathogens, including *Plasmodium falciparum* (Syin et al., 2001) , *Leishmania major* (Siman-Tov et al., 1996) and *Giardia lambia* (Abel et al., 2001) suggesting that PKA inhibitors may have anti-parasitic applications.

2.8. The cAMP and cAMP dependent Protein KinaseA (PKA): As an Environmental Sensor and Cytoprotector in Parasites

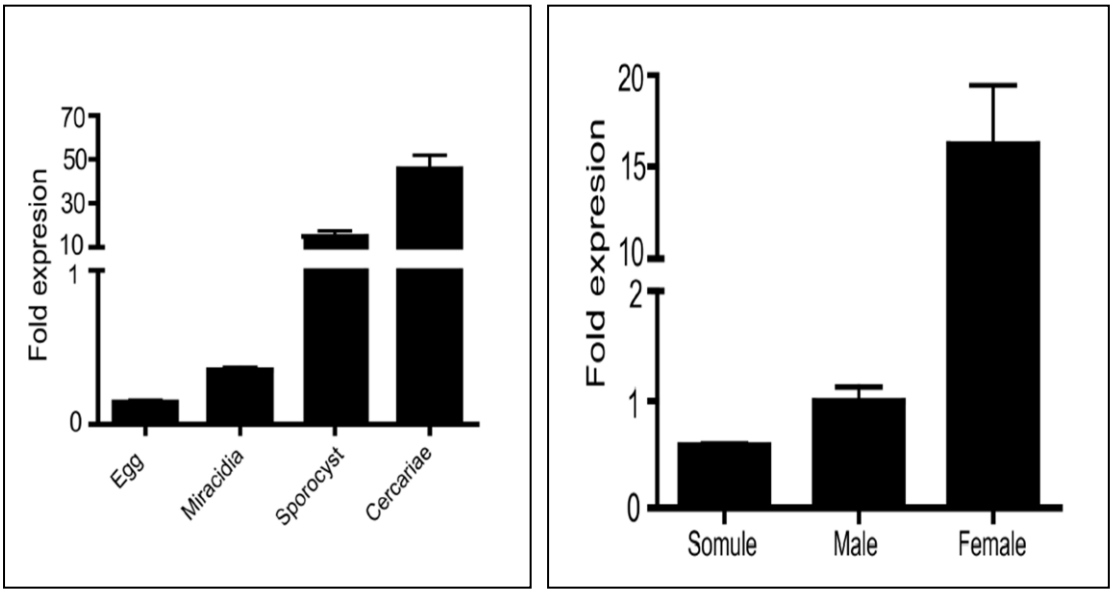
Together with a substantial alteration of nutrient availability and other stressors, parasites must adapt to new conditions of temperature and pH (37 °C and pH 5.5) (Bhattacharya et al., 2008). In most parasitic infections during

phagocytosis by macrophages, the organisms suffer another line of stress caused by the respiratory burst of the macrophages (Bhattacharya et al., 2008). This first line of defense is responsible for producing reactive oxygen species (ROS) and nitrogen species (NOS) (Gantt et al., 2001, Zhang et al., 2012) as well as cytokines and chemokines which are involved in inflammatory cell activation.

Many intercellular parasites bypass this defense by modulating their own biology and the host environment to survive successfully in the host (Bhattacharya et al., 2008). Some parasites when exposed to toxic pro-oxidant as in macrophages, survive by converting to other intracellular parasitic stages (Bhattacharya et al., 2008). This stress response strategy has been seen with *Leishmania* which converts into the intracellular amastigote. In schistosome infection, the molecular mechanism by which the parasitic stage circumvents the toxic effect of ROS and NOS is scanty (Sayed et al., 2006, Williams et al., 2006).

Some studies have identified various genes in parasites responsible for circumventing the toxic effects of reactive oxygen species. In *Leishmania* superoxide dismutase, peroxidoxin and trypanothione reductase have been shown to contribute antioxidant defense against ROS, NOS and their intermediates (Miller et al., 2000, Zhang et al., 2012). Therefore any disruption of these genes will render the parasite susceptible to macrophage action (Barr and Gedamu, 2001, Ghosh et al., 2003, Plewes et al., 2003, Tovar et al., 1998). In schistosomes the triggers for the antioxidant defense action are unknown, while in many lower taxa organisms, environmental stimuli are

critical in directing the biology of parasites. Thus, exposures to environmental stressors such as pH and temperature changes have been related to resistance against oxidative damage in parasitic organisms (Zarley et al., 1991).



A **B**

Figure 4: Developmental stage expression of *S. mansoni* protein kinase.

A: *Schistosoma mansoni* protein kinase A catalytic subunit expression in the larval stages of *S. mansoni.* Relative PKA-C mRNA levels in *S. mansoni* larval stages as determined by quantitative PCR. The cercariae show more than five-fold higher expression compared to the sporocysts PM:19707280 (Swierczewski and Davies 2009)

B: *Schistosoma mansoni* protein kinase A catalytic subunit (PKA-C) mRNA expression in the intra-mammalian stages of *S. mansoni.* A) Relative PKA-C mRNA levels in adult male and female *S. mansoni* worms and schistosomula as determined by quantitative PCR (Swierczewski and Davies 2009). The intra-mammalian stages show that adult females have the highest level of expression. It was found to be close to 20-fold higher compared with adult males and the newly transformed schistosomula(Swierczewski and Davies, 2009) .

The biological response of cyclic Adenosine Monophosphate (cAMP) is one of the main environmental sensing machineries associated with the stress response in a wide range of unicellular eukaryotes including *Trypansosoma spp*, *Plasmodium spp* and other organisms (Kaushal et al., 1980). *Plasmodium*

falciparum is capable of synthesizing its own cAMP through adenylate cyclase without stimulation from the mammalian adenylate cyclase (AC) activator Forskolin or the heteromeric G protein activator AlF4. The protein kinase A (PKA) functions in the conductance of anions across the host cell membrane of *Plasmodium* infected red blood cells (RBC).

Also the PKA-R (PKA regulatory subunit) maybe involved in the activation of anion conductance channel in *P. falciparum* infected RBC (Egee et al., 2002, Merckx et al., 2008). In other organisms like *Entamoeba histolytica*, activation of PKC or cAMP-dependent protein kinase (cAMP-PKA) signaling pathways trigger the phosphorylation of proteins involved in actin re-arrangements necessary for adhesion and movement (Smith et al., 1998). Also cAMP-response elements could play an important role in regulating actin expression and organization of the signaling processes activated during tissue invasion. In Trypanosome differentiation, cAMP causes differentiation from the long slender form to the short stumpy form, in which the concentration of cAMP decreases (Mancini and Patton, 1981). Although cAMP signaling and control have been identified in several organisms as discussed above, the knowledge on cAMP signaling in schistosomes is limited.

2.9. Overview of Relevant Bioinformatics Tools

According to Lascombe and co-authors “Bioinformatics is conceptualizing biology in terms of macromolecules (in the sense of physical-chemistry) and then applying "informatics" techniques (derived from disciplines such as applied maths, computer science, and statistics) to understand and organize the information associated with these molecules, on a large-scale”

(Luscombe et al., 2001). Bioinformatics is by nature cross-disciplinary and interdiscplinary (Gupta, 2009, Kim, 2013). Topics in bioinformatics research have been categorized into the following categories: (1) Genome analysis; (2) Sequence analysis; (3) Phylogenetics; (4) Structural bioinformatics; (5) Gene expression; (6) Genetic and population analysis; (7) Systems biology; (8) Data and text mining; (9) Databases and ontologies and (10) Bioimage Informatics (Peng et al., 2012, Valencia and Bateman, 2005).

Based on the above definition of bioinformatics, the primary purpose of bioinformatics tools could be described as to enable the understanding and organizing the information associated with biological macromolecules, on a large-scale. Bioinformatics tools have been developed to align with the categories presented above. For this research, a set of bioinformatics tools have been assembled to accomplish the research objectives in sequence analysis and structural bioinformatics categories. The tools include those for sequence retrieval, multiple sequence aligned, predictions of protein structure as well as protein domain architecture (Table 3). Additionally, computational tools for visual discovery (exploration, mining and analyzes) of data from the bioinformatics tools are included in the toolset for the research.

Table 3: Overviewed Description of Research Selected Bioinformatics Tools

Tool	Purpose	Parameter	Input	Output	Version
Tableau Sofware (Chabot, 2009, Chang et al., 2009, Johnson et al., 2010, Thomas and Cook, 2006)	Collect information, data processing, knowledge representation , interaction, visualization and decision making	The user define the parameter of the data	Numerical and non-numerical data	Numerical and non-numerical data, tables, and statistical plots	Tableau Software 8, March 2013
Universal Protein Resource (UniProt) (UniProt, 2011)	Retrieve the the *Schistosoma* USP sequences.	Many parameters. User choose according to objective.	Many input options. The research uses the query key word option for USP from both *S. mansoni* and *S. japonicum.*	List of USP protein sequences from both *Schstosoma* species.	UniProt release 2011_11
SchistoDB www.schistodb.net (Zerlotini et al., 2009)	Collect gene expression data documented on its developmental stages for *S. mansoni.*	Expressed Sequences Tags (ESTs); Serial Analysis of Gene Expression (SAGE) and microarray.	The Gene ID e.g Smp_043030 was used in querying the database.	Developmental stages expression pattern.	SchistoDB, version 3.0 July 2012
NCBI Conserved Domain Search Tool (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) (Marchler-Bauer et al., 2011)	Search for amino acid residues that are functionally important.	Uses RPS-BLAST, a variant of PSI-BLAST, to quickly scan a set of pre-calculated position-specific scoring matrices (PSSMs) with a protein query.	Protein amino acid sequences in FASTA format.	The result are presented as an annotation of protein domains on the user query sequence and can be visualized as domain multiple sequence alignment. Specific hit are used to associate the high confidence query and conserved domain.	Last revised March 2013
3DLigandSite server at (http://www.sbg.bio.ic.ac.uk/3dligandsite) (Wass et al., 2010)	3DLigandStie is an automated method for the prediction of ligand binding sites	Parameter is set to automate.	Protein sequence or a protein structure	The submitted sequences structure is used to search structural library to identify homologous structures with bound ligands. Ligands will be superimposed onto the protein structure to predict a ligand	(Wass et al., 2010)

Tool	Purpose	Parameter	Input	Output	Version
				binding site.	
NetPhosK 1.0 tool (http://www.cbs.dtu.dk/services/NetPhosK) (Blom et al., 2004)	Predict USP specific kinases	Prediction without filtering (fast)	Protein Sequence in FASTA format	The output contains four columns: Column 1: position of the residue being analyzed Column 2: the residue Column 3: the predicted kinase (above the threshold ,set at 0.65) Column 4: output score (value in the range [0.000-1.000])	The server was last modified Thursday 25th April 2013 12:53:52 GMT
Euk-mPLoc 2.0 (http://www.csbio.sjtu.edu.cn/bioinf/euk-multi-2/) (Chou and Shen, 2010)	Used to identify eukaryotic proteins within 22 cellular	Based on integrating information from gene ontology, functional domain and evolutionary relationships	Input the eukaryotic protein sequence in Fasta format	Query protein and predicted location(s)	Euk-mPLoc server has been updated to 2.0 version, for the 1.0 version, access (Chou and Shen, 2010)
SjTPdb, (http://function.chgc.sh.cn/sj-proteome/index.htm) (Liu et al., 2008)	Collect gene expression data documented on its developmental stages for *S. japonicum*	An integrated transcriptome and proteome database and analysis platform. Annotated for sequence similarity, structural features, functional ontology, genomic variations and expression patterns across developmental stages and tissues including the tegument and eggshell	Search carried using *S. japonicum* USP geneIDs, based on the developmental stage of the life cycle.	Developmental stages expression pattern	
ClustalW tool (http://www.ch.embnet.org/software/ClustalW.html) (Larkin et al., 2007, Goujon et al., 2010)	To generate multiple sequence alignments	Default settings	Protein Sequence in FASTA format	Clustal format	ClustalW tool, Version 2
MEGA software (http://www.megasoftware.net/) version 5 (Tamura et al., 2007,	Reconstructing the evolutionary histories of	Maximum Likelihood (ML) based on the JTT matrix-based	Multiple sequence alignment from	Estimates of phylogenetic trees, substitution parameters, and	MEGA software, version 5

Tool	Purpose	Parameter	Input	Output	Version
Tamura et al., 2011)	species using USP sequences	model at 1000 Bootstrap	ClustalW. v.2	rate variation among sites.	
ProtParam tool (http://www.expasy.ch/cgi-bin/protparam). (Gasteiger et al., 2003)	Computing various physical and chemical parameters	Compute parameters include theoretical isoelectric point (Ip), molecular weight, net positive and negative residues, extinction coefficient, half-life, instability index, aliphatic index and grand average hydropathy (GRAVY)	Protein can either be specified as a Swiss-Prot/TrEMBL accession number or ID or in the form of a raw sequence	Computer parameter values	Expasy sites and tools are reviewed Constantly
SwissModel server (http:// swissmodel.expasy.org/)	Determine three-dimensional (3D) model structure of the proteins	Used Automatic model mode with PDB template 1mjh (Chain B).	Target protein amino acid sequence in FASTA format	Model 3D structure including model statistics	The server is constantly revised at (http:// swissmodel.expasy.org/)
RAMPAGE server	Ramachandran plot analysis using to assess the quality of predicted models based on the phi – psi torsion angles of all the residues	The Ramachandran compare residues in the favored , allowed and the outlier regions	Model structure in PDB file	Graphs and tables show the statistics of residue distribution to assess model quality	The server is constantly updated at http://mordred.bioc.cam.ac.uk/~rapper/rampage.php
The Rasmol tool(http://www.openrasmol.org/)	Used in visualizing the modeled 3D structures and the distribution of the secondary structures.	All structural parameters in the tool needed was used to achieve objective.	Model structure in PDB file	Model 3D structure visualization. Including structural statistics.	The Rasmol tool at (http://www.openrasmol.org/) remain a popular structural visualization tools, and is constantly reviewed
The SOPMA tool (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html)	Secondary structure prediction and computation.	Default parameter	Protein amino acid sequences in FASTA format	Different secondary protein constituents	The tool server is review constantly
The TMPRED server (http://www.ch.embnet.org/software/TMPRED	Predict transmembrane protein	Default parameter	Protein amino acid sequences in	Determine transmembrane protein orientation	The tool server is review

Tool	Purpose	Parameter	Input	Output	Version
_form.html	topography		FASTA format		constantly
Pepwheel tool (http://emboss.bioinformatics.nl/cgi-bin/emboss/pepwheel	transmembrane helices region visualization and analysis	Default parameter	Transmembrane helices region sequence	helical wheel plots	Pepwheel is included in EMBOSS 2.7 suit. It is constantly reviewed.

2.11. Physicochemical Properties of Amino Acids

The physicochemical property of a protein is determined by types of the 20 naturally occurring amino acids that are present in the primary sequence (Kawashima et al., 2008, Vaidya et al., 2012). The amino acid present determines the types of functional groups that influence the structure and the functional characteristics of the protein. Understanding protein structures are crucial, because they are involved in every cellular activity. Several physicochemical parameters can be calculated for a protein sequence including molecular weight, theoretical pI, amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index, and grand average of hydropathicity (GRAVY) (Song and He, 2012, Gasteiger et al., 2003).

Hydrophobicity always plays a vital role in tertiary structure formation and behavior of a protein molecule. It has been deduced that computing the different physicochemical parameters and their interrelation will aid tremendously in the future of protein science (Gasteiger et al., 2003). There are numerous online tools for predicting physicochemical properties of proteins, but most of them are separate software, which is inconvenient and time consuming. The ProPAS tool is standalone for computing physicochemical properties of protein but it consists of only three components

(Wu and Zhu, 2012). There are many such web based tools for calculating one or few parameters not cited here.

Considering the complexity of bioinformatics and the multiple tools available there is urgent need for a simple and interactive “ONE STOP TOOL” for computing various physicochemical properties and linking structure and function of proteins from their primary sequences. The Protparam is a standalone tool which estimates many basic physicochemical properties of a polypeptide on the basis of its amino acid sequence (http://expasy.org/tools/protparam.html) (Table 4). The Protparam will be used for computing the physicochemical properties needed in this research.

Table 4: Purpose and Interpretation of Physicochemical Properties of Proteins

Physicochemical properties	Purpose	Interpretation	Reference
Amino acid composition	Determine the composition of each amino acid residue per sequence	Predicts the hydrophilicity or hydrophobicity index of proteins by comparing polar and non-polar amino acid residue. The hydrophobic residues are located in the core of most proteins and they help in stabilizing the proteins through van der Waal interactions. The hydrophilic residues are located mostly at the surface and active sites of proteins, where they interact with other polar residues or with water molecule	(Berezovsky and Trifonov, 2001).
Theoretical pI	The Isoelectric point (pI) indicates the pH at which the protein surface is covered with charge	The isoelectric point (PI) >7 indicates proteins are basic in nature. Less 7 indicate acidic in nature. At a given pI proteins are stable and compact. This parameter is useful for developing buffer systems for purifying proteins using isoelectric focusing techniques.	(Chen et al., 1993) (Sillero and Maldonado, 2006).
Number of negative charge residues	Number of negatively charged residues (Asp + Glu)	Contributing factor to the overall charge	
Number of positive charge residues	Number of positive charged residues (Arg + Lys)	Contributing factor to the overall charge	
Extinction coefficient	Estimate the molar extinction coefficient of a protein from knowledge of its amino acid composition	The computed protein concentration and the extinction coefficient can be important in the quantitative analysis of the protein-protein and protein-ligand interaction in solutions	(Stoscheck, 1990)
Half life	Estimated half-life of protein based on amino acid composition	Contribute to the stability of protein	(Varshavsky, 1997) .
Instability index	Predicts instability index from amino acid composition	A protein of instability index < 40 is considered as stable, while those with values > 40 are unstable	(Guruprasad et al., 1990)
Aliphatic index	To determine thermostability of a protein	The aliphatic index (AI) of a protein is the relative volume occupied by the aliphatic side chains (alanine, valine, isoleucine and leucine) and is taken as contributor to the increase thermal stability of globular proteins. Proteins with low thermal stability turn to be more structurally flexible.	(Ikai, 1980)
Grand average hydropathicity (GRAVY)	Computed the average hydropathy values of all the amino acids per in the sequence.	The very low GRAVY Index indicates good interaction between protein and water	(Kyte and Doolittle, 1982)

CHAPTER THREE

METHODS

Objective 1:

3.1. 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, a protocol was developed that integrates visual analytics stages to facilitate the interaction with the results from sequence analysis and data collection on a set of universal stress proteins from *Schistosoma mansoni* and *Schistosoma japonicum*. Visual analytics is an iterative process conducted via visual interfaces that involves collecting information, data preprocessing, knowledge representation, interaction, and decision-making (Chabot, 2009, Chang et al., 2009, Johnson et al., 2010, Thomas and Cook, 2006).

The overview of the bioinformatics and visual analytics methods is summarized in Figure 5. The protocol consists of 5 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 phylogenetic tree.

The bioinformatics sequence analyses in Stage 2 determine (i) protein sequence length, (ii) protein domain length, (iii) ligand binding sites, (iv)

chemical ligand binding, (v) kinase binding, (vi) 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. Data on developmental expression of gene transcripts were obtained from publications. Multiple sequence alignment and phylogenetic trees provide the statistically and evolutionary support for groupings of the sequences.

The prioritization process was done in Stage 4 with the criterion determined by the researcher. 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 is the product of the prioritization process. This report expects (i) orthologous pairs of protein sequences and (ii) a visualization that provides an integrated view of the shared annotations. The visual analytics software package Tableau Software (Mackinlay et al., 2007), was used to perform several visual analytics tasks including interaction, computing, analysis, integration and visualization.

3.2. Retrieval of Protein Sequences

Proteins annotated with the universal stress protein domain (PF00582) from the *Schistosoma mansoni* and *Schistosoma japonicum* genomes were identified in Universal Protein Resource [UniProt] (UniProt release 2011_11) (http://www.uniprot.org/) (UniProt, 2011). Predicted protein sequences were

retrieved. The final list of protein sequences for comparative sequence analysis was determined by revision history of the sequence in the UniProt as well as entries in GeneDB and SchistoDB (Hertz-Fowler et al., 2004, Zerlotini et al., 2009).

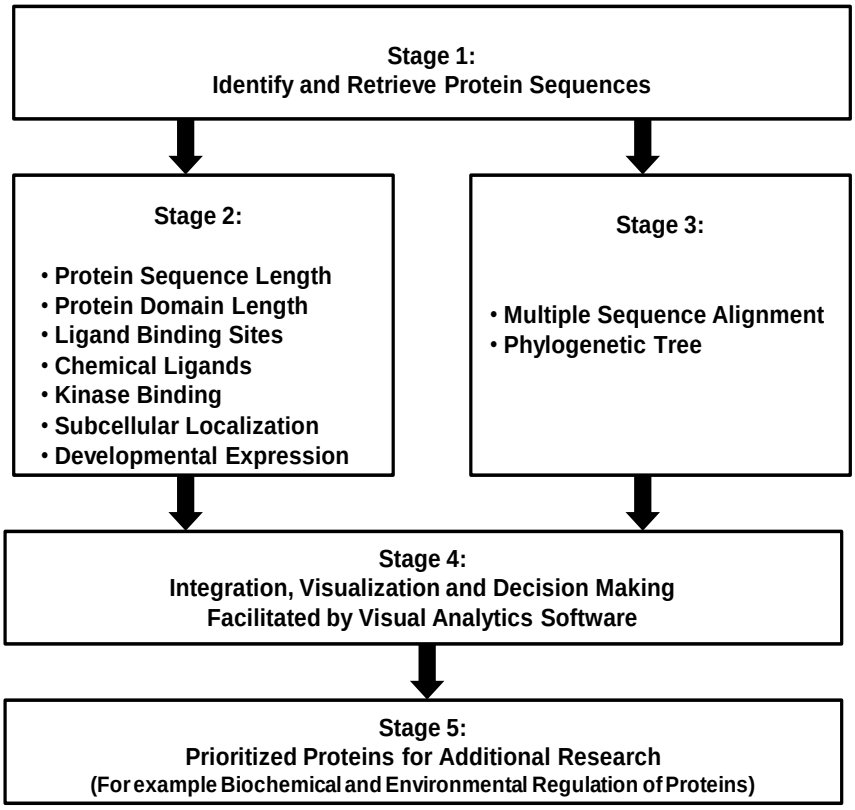


Figure 5: Overview of a set of bioinformatics and visual analytics methods to prioritize protein sequences for further research

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 stages (Stage 2 and Stage 3) on the protein sequences (Stage 1). The evolutionary relatedness of protein sequences is based on statistically supported groups in a phylogenetic tree derived from multiple sequence alignment of all the sequences (Stage 3). Additional evidence for evolutionary relatedness is obtained from gene synteny on 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 process during life cycle of parasites. Details of each method are available in Methods section of the article.

3.3. Conserved Domain Search for Functional Sites

The search for amino acid residues that are functionally important was performed using two public servers (Marchler-Bauer et al., 2011). The single

sequence input server at (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) and the multiple sequence input or the batch server at http://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi 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 a functional site signature for each sequence was constructed. The signature is a string of the amino acid letters. In a case were 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 (Salsbury et al., 2008).

The conserved domain public server is a protein annotation resource that consists of a collection of well-annotated multiple sequence alignment models for ancient domains and full length proteins. These are available as position-specific score matrices (PSSMs) for fast identification of conserved domains in protein sequences via RPS-BLAST. The conserved domain content includes NCBI–curated domains, which uses 3D structure information explicitly to define domain boundaries and provide insights into sequence/structure/function relationships, as well as domain models imported from a number of source databases (Pfam, SMART, COG, PRK and TIGRFAM).

3.4. Prediction of Chemical Ligand and Enzymatic Regulation

The three-dimensional (3D) chemical ligands were predicted using the 3DLigandSite server at (http://www.sbg.bio.ic.ac.uk/3dligandsite) (Wass et al., 2010). These biologically relevant chemical ligands are potential regulators of the function of the USPs. The 3DLigandSite is a top performing web server for chemical ligands prediction and provides structural models for unsolved proteins using protein-structure prediction. The specific kinases were predicted using NetPhosK 1.0 tool (http://www.cbs.dtu.dk/services/NetPhosK) (Blom et al., 2004) with a stringent threshold value set at 0.65. The kinase with the highest threshold value was selected. This server is updated constantly.

3.5. Prediction of Subcellular Location

Information of subcellular locations of proteins is important for in-depth studies of cell biology. It is very useful for proteomics, system biology and drug development. The subcellular locations of all the proteins 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/). This server was selected because of the ability to identify eukaryotic proteins within 22 cellular locations (Chou and Shen, 2010). The prediction by Euk-mPLoc2.0 is based on integrating information from gene ontology, functional domain and evolutionary relationships.

3.6. Compilation of Developmental Expression of Genes

The developmental stage expression profiles of the selected *Schistosoma* USP genes were extracted from previous publication (Isokpehi et al., 2011a), for *S. mansoni*. Furthermore, the functional annotations integrated

into the relational database SchistoDB (http://www.schistodb.net) which include gene expression data documented on its developmental stages from high throughput Expressed Sequences Tags (ESTs); Serial Analysis of Gene Expression (SAGE) and microarray methods were also investigated. These gene expressions data present rich genomic resources to elucidate the developmental regulation of *S. mansoni* USP genes. In the case of *S. japonicum*, expression profiles were from the SjTPdb, an integrated transcriptome and proteome database and analysis platform for *Schistosoma japonicum* (Liu et al., 2008).

The database contains Expressed Sequence Tags (ESTs), EST clusters, and the proteomic dataset for *S. japonicum*. The core of the database is the 8,420 *S. japonicum* proteins translated from the EST clusters, which are well annotated for sequence similarity, structural features, functional ontology, genomic variations and expression patterns across developmental stages and tissues including the tegument and eggshell. The database was queried by search based on the developmental stage of the life cycle, and an integrated search for more specific information. All the useful information based on developmental stage expression of *S. japonicum* USP genes were extracted and documented.

3.7. Prediction of Evolutionary Relatedness of Sequences

Evolutionary relatedness of the selected universal stress proteins from both *S. mansoni* and *S. japonicum* was determined to ascertain their functional and evolutionary relationship. The sequences were aligned using ClustalW tool (http://www.ch.embnet.org/software/ClustalW.html) Version 2 (Larkin et al.,

2007, Goujon et al., 2010), applying the default settings to generate multiple sequence alignments.

Multiple Sequence Alignment (MSA) is the alignment of three or more biological sequences (protein or nucleic acid). From the output, homology can be inferred and the evolutionary relationships between the sequences investigated. Protein multiple sequence alignment is a key bioinformatics method for protein structure and function prediction, biological evolution analysis, phylogeny inference and other common tasks in sequence analysis. A variety of alignment algorithms in this field have achieved great success. CLUSTALW is still the most popular alignment tool to date.

Comparative analysis of molecular sequence data is essential for reconstructing the evolutionary histories of species and inferring the nature and extent of selective forces shaping the evolution of genes and species. The Maximum Likelihood (ML) analysis has been widely used for reconstructing phylogeny. The evolutionary relationship of the sequences was inferred using the Maximum Likelihood (ML) (Tamura and Nei, 1993), method based on the JTT matrix-based model (Potts et al., 2013, Binder et al., 2013, Methven et al., 2013, Jones et al., 1992) at 1000 Bootstrap using the MEGA software (http://www.megasoftware.net/) version 5, (Tamura et al., 2007, Tamura et al., 2011).

3.8. Visual Analytics of Datasets

A purpose of Stage 4 of the protocol was to provide an integration and visualization portal for results of bioinformatics analysis in Stage 2 and Stage 3 (Figure 5).The visual analytics tasks to be performed on datasets can be

influenced by how the data is organized in the data records (rows) and data fields (columns) in the data source (e.g. spreadsheet file and comma delimited file). For Stage 2, each data record had the following data fields (i) Organism; (ii) Locus Tag; (iii) UniProt ID; (iv)Feature and (v) 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 Subcelluar Localization.

In the case of the Functional Site signature dataset, each record consisted of the UniProt ID and 12 fields for each of the 12-letter signature for the USP ligands binding sites. For Stage 3 dataset (phylogenetic tree groupings) each data record consists of data fields for (i) Organism; (ii) UniProt ID and (iii) Phylogenetic Group. The data sources (in this case, spreadsheet files) were loaded to the visual analytics software for visual analytics tasks including design of the data integration and visualization.

Objective 2: Identify biological function relevant protein sequence and structure features for prioritized universal stress proteins from *Schistosoma* species.

3.9. Computation of Amino acid Content and Physicochemical Parameters

The amino acid compositions of the 13 sequences were computed using the ProtParam tool (http://www.expasy.ch/cgi-bin/protparam). ProtParam is a tool for computing various physical and chemical parameters (physico-

chemical properties) (Aslanzadeh and Ghaderian, 2012, Kumar et al., 2012, Zikmanis and Kampenusa, 2012, Mbah et al., 2012) from a given protein sequence stored in Swiss-Prot/TrEMBL or for a user entered sequence. No additional information is required about the protein under consideration. The computed parameters include theoretical isoelectric point (Ip), molecular weight, net positive and negative residues, extinction coefficient, half-life, instability index, aliphatic index and grand average hydropathy (GRAVY) .The percentages of hydrophobic and hydrophilic residues in each gene was calculated from the primary structure analysis and visualized.

3.10. Conserved Domain Search, Homology Modeling and Visualization of 3D Structure

The possible conserved domains regulating the functional mechanism of both [Q86DW2 [*S. japonicum*] and G4LZI3 [*S. mansoni*] proteins were analyzed using the NCBI public server at http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi. Their functional units and domain residues were identified and documented. The three-dimensional (3D) structure of both protein were modeled using the PDB template 1mjh (Chain B). The three-dimensional (3D) structure of the proteins were determined using SwissModel server (http:// swissmodel.expasy.org/)and the Phyre2 server (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=news) (Kelley and Sternberg, 2009). The quality of the model was evaluated with Ramachandran plot data, based on the phi–psi torsion angles of all the residues in the model using DeepView–Swiss-PdbViewer (http://spdbv.vital-it.ch/). The Ramachandran plot obtained from DeepView was further assessed using Ramachandran plot 2 assessment server

(http://dicsoft1.physics.iisc.ernet.in/rp/). The Rasmol tool (http://www.openrasmol.org/) was used in visualizing the modeled 3D structures and the distribution of the secondary structures. The three-dimensional (3D) LigandSite residues and the predicted ligand of the proteins were determined using 3DLigandSite server at (http://www.sbg.bio.ic.ac.uk/3dligandsite) (Wass et al., 2010). The Rasmol tool was further employed in locating the positions of the ligand binding sites and the ligand on the 3D structure with particular attention to the regulatory points for both Q86DW2 [*S. japonicum*] and G4LZI3 [*S. mansoni*].

Objective 3: Determine the distinctive structural protein features of a predicted regulator of *Schistosoma* adenylate cyclase activity that has possible influence on the functioning of universal stress proteins.

3.11. Sequence Retrieval, Amino Acid and Physicochemical Analysis

The *Schistosoma mansoni* Smp_059340.1 (UniProt ID: C4QDC7|C4QDC7_SCHMA; Entrez Gene ID: Smp_059340) reviewed sequence was retrieved from Swiss-Prot database (http://expasy.org/sprot/). The amino acid composition of the sequence was computed using the ProtParam tool (http://www.expasy.ch/cgi-bin/protparam).The ProtParam tool was also used to compute the physicochemical parameters such as theoretical isoelectric point (Ip), molecular weight, total number of positive and negative residues, extinction coefficient, half-life, instability index, aliphatic index and grand average hydropathy (GRAVY) .The percentages of hydrophobic and hydrophilic residues were calculated from the primary structure analysis and

the hydrophobicity plot was done using both Hopp –Woods and Kyte - Doolottle scale for possible antigenicity.

3.12. Prediction of Secondary Structure Elements and Conserved Domains

The SOPMA tool (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html) was used for the secondary structure prediction. The TMPRED server (http://www.ch.embnet.org/software/TMPRED_form.html performed the identification of possible transmembrane helices regions. The predicted transmembrane helices region was visualized and analyzed using helical wheel plots generated by the program Pepwheel (http://emboss.bioinformatics.nl/cgi-bin/emboss/pepwheel), included in the EMBOSS 2.7 suit. The possible conserved domain search was carried out using the public server at http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi for the Smp_059340.1 and the individual domains present were analyzed and binding residues compared within the domains.

3.13. Homology Modeling and Visualization of 3D structure

The three-dimensional (3D) structure of Smp_059340.1 encoded protein was modeled using the PDB template 1TL7C (chain C).The 3D structure of Smp_059340.1 protein was generated using the following servers; SwissModel server (http:// swissmodel.expasy.org/) and the Phyre2 server (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=news). The quality of the model was evaluated with Ramachandran plot data, based on the phi–psi torsion angles of all the residues in the model using Ramachandran plot2

assessment server (http://dicsoft1.physics.iisc.ernet.in/rp/) and RAMPAGE server at http://mordred.bioc.cam.ac.uk/~rapper/rampage.php . The Rasmol tool (http://www.openrasmol.org/) was used in visualizing the modeled 3D structures and to identify any possible "SS" bonds including the distribution of the secondary structures. The three-dimensional (3D) LigandSite residue(s) and the predicted ligand (s) of Smp_059340.1 protein were determined using 3DLigandSite server at (http://www.sbg.bio.ic.ac.uk/3dligandsite). The Rasmol tool was further used in locating the positions of the conserved domains on the 3D structure and particularly the keys residues involved in regulatory mechanism. With Rasmol tool, the actual locations of all the domains were mapped on the protein 3D and folded structure.

CHAPTER FOUR

RESULTS

Objective 1: Infer the Biochemical and Environmental Regulation of the *Schistosoma* Universal Stress Proteins

4.1. Dataset for Visual Analytics

The dataset analyzed consisted of 12 annotation features for 13 universal stress protein sequences (Table 5). All the sequences analyzed contain the ATP-binding motif (G-2x-G-9x-G(S/T). This research is particularly interested in shared annotations that can help infer joint biochemical and environmental regulation of the universal stress proteins from the two pathogenic *Schistosoma* species. 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 5 annotation features with variable frequency per protein were 3D Chemical Ligand, Ligand Binding Amino Acid, Amino Acid and Sequence Position, Developmental Expression and Subcelluar Localization. This dataset was formatted for visualization and analyze 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 the researcher intends to conduct

additional research on the *Schistosoma* USPs, the primary purpose of these groups is to prioritize those 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.

Table 5: Annotation features for universal stress proteins of *Schistosoma mansoni* and *Schistosoma japonicum*

Organism	UniProt ID	Locus Tag	Feature Value							Feature Frequency			
			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
Schistosoma japonicum	Q5DDH7		148	172	15	163	P	PKC	0.87	5	12	12	
	Q5DED2		136	160	15	151	P	PKC	0.87	5	12	12	1
	Q5DGI9		146	159	8	154	P	PKC	0.76	5	12	12	6
	Q5DH64		122	129	2	124	P	PKC	0.78	4	9	9	6
	Q5DHK1		149	172	18	167	P	PKC	0.80	5	12	12	5
	Q5DI36		125	133	3	128	P	PKC	0.72	4	9	9	
	Q86DW2		148	184	28	176	P	PKC	0.93	7	12	12	5
	Q86DX1		147	155	7	154	P	PKC	0.85	5	12	12	3
Schistosoma mansoni	C1M0Q2	Smp_097930	148	159	7	155	P	PKA	0.68	5	12	12	1
	G4LZI3	Smp_076400	149	184	27	176	P	PKC	0.93	7	12	12	3
	G4V5S2	Smp_001000	148	174	15	163	P	PKA	0.77	6	12	12	3
	G4VIW9	Smp_043120	148	160	7	155	P	PKC	0.83	6	12	12	7
	G4VPM6	Smp_031300	149	160	6	155	P	PKC	0.71	5	12	12	4

4.2. Grouping of *Schistosoma* Universal Stress Proteins by Sequence Length

The 13 *Schistosoma* protein sequences were grouped by protein length and domain length (Figure 6). Seven distinct domain lengths (122, 125, 136, 146, 147, 148 and 149 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, 8 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 length of 1 aa as in the 184 aa USPs (Q86DW2 and G4LZI3) or 2 aa as in 159 aa USPs (Q5DGI9 and C1M0Q2).

4.3. Functional Site Signatures of *Schistosoma* Universal Stress Proteins Sequences

The Conserved Domain Search tool at the National Center for Biotechnology Information (NCBI) website uses the fragment A (UniProt Identifiers: Q57997; Y577_METJA) of the universal stress protein MJ0577 of *Methanocaldococcus jannaschii* as template to aligned 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 acid 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 (Zarembinski et al., 1998) .

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
	149	*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 6: Grouping of 13 *Schistosoma* universal stress proteins by sequence length

The image provides a visual comparison of the protein sequence and universal stress protein (USP) domain sequence lengths for 13 *Schistosoma* universal stress proteins. A visual analytics resource to interact with data is available at http://public.tableausoftware.com/views/schisto_features_usp/groupbylength

The amino acid sequences from Met126 to Thr143 contain the motif G2xG9xG(S/T), which includes binding sites 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 XXXVMGRGGSVS) were observed (Figure 7). The last 7 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).

Functional Site Signature	UniProt ID	Locus Tag	Organism
AIDAIGRGGSVS	Q5DDH7		*Schistosoma japonicum*
	Q5DED2		*Schistosoma japonicum*
AIDAVGRGGSVS	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 7: Grouping of 13 *Schistosoma* universal stress proteins by functional site signature

The functional site signature is constructed by joining the 12 ligand binding sites known for the ATP binding universal stress protein from *Methanocaldococcus jannaschii* (UniProt ID: Y577_METJA). The image provides a visual comparison of the functional site signatures for 13 *Schistosoma* universal stress proteins. A visual analytics resource to interact with data is available at http://public.tableausoftware.com/views/schisto_features_usp/groupbylength

4.4. Grouping of *Schistosoma* Universal Stress Proteins Sequences by Alignment

A multiple sequence alignment (MSA) of the 13 sequences was generated by ClustalW (Figure 8). The ligand binding sites (functional sites) annotated in the Conserved Domain Database are labeled with hashes (#) in Figure 8. The alignment revealed gap positions where amino acid residues are missing. These gaps could explain the differences in lengths reported for the sequences. In the USP functional sites signatures, the first three letters where 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, 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 and 176 using Smp_076400 from *S. mansoni* as the reference sequence (Figure 8). The relationship of the sequences was visualized as a phylogenetic tree (Figure 9). Five groups (A to E) of sequences were observed with each of the five *Schistosoma mansoni* sequences assigned to a group. Group A and E contain multiple *S. japonicum* sequences. The bootstrap statistical support value for the branch of Group C (Smp_076400 [G4LZI3] and Q86DW2) was 100%. The grouping of sequences by the maximum parsimony model was in agreement with the maximum likelihood model (Figure 10).

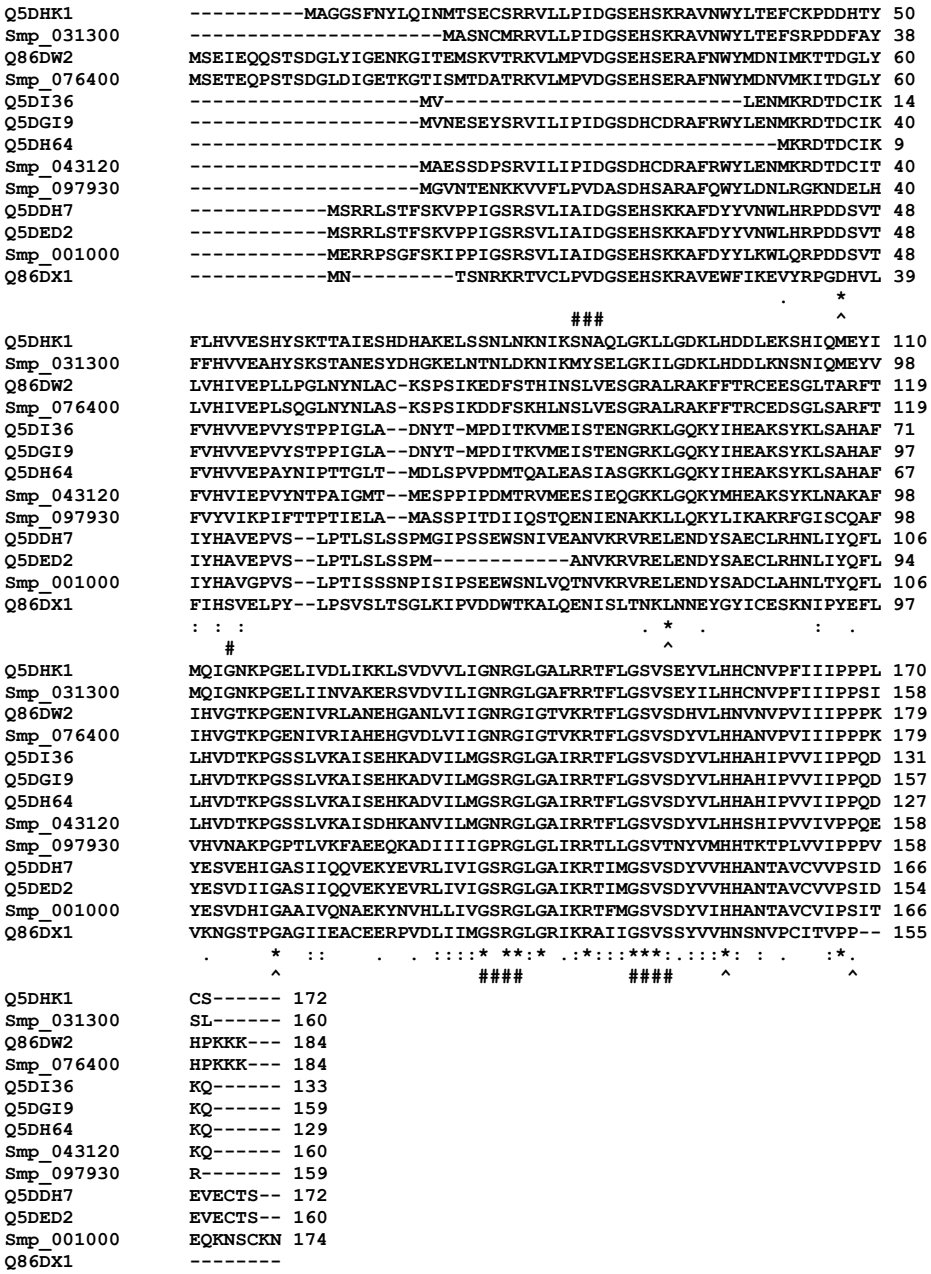


Figure 8: Multiple sequence alignment of the sequences of selected universal stress proteins of *Schistosoma mansoni* and *Schistosoma japonicum*

The sequence alignment of the 13 sequences with ATP binding motif [G2XG9XG(S/T)] was generated using ClustalW. Sequences of *Schistosoma 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 (Asp; D), leucine (Leu; L), glycine (Gly; G), histidine (His; H), and proline (Pro; P) 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; and ".", semi-conserved substitutions. A visual analytics resource to interact with data is available at http://public.tableausoftware.com/views/schisto_features_usp/usp_align

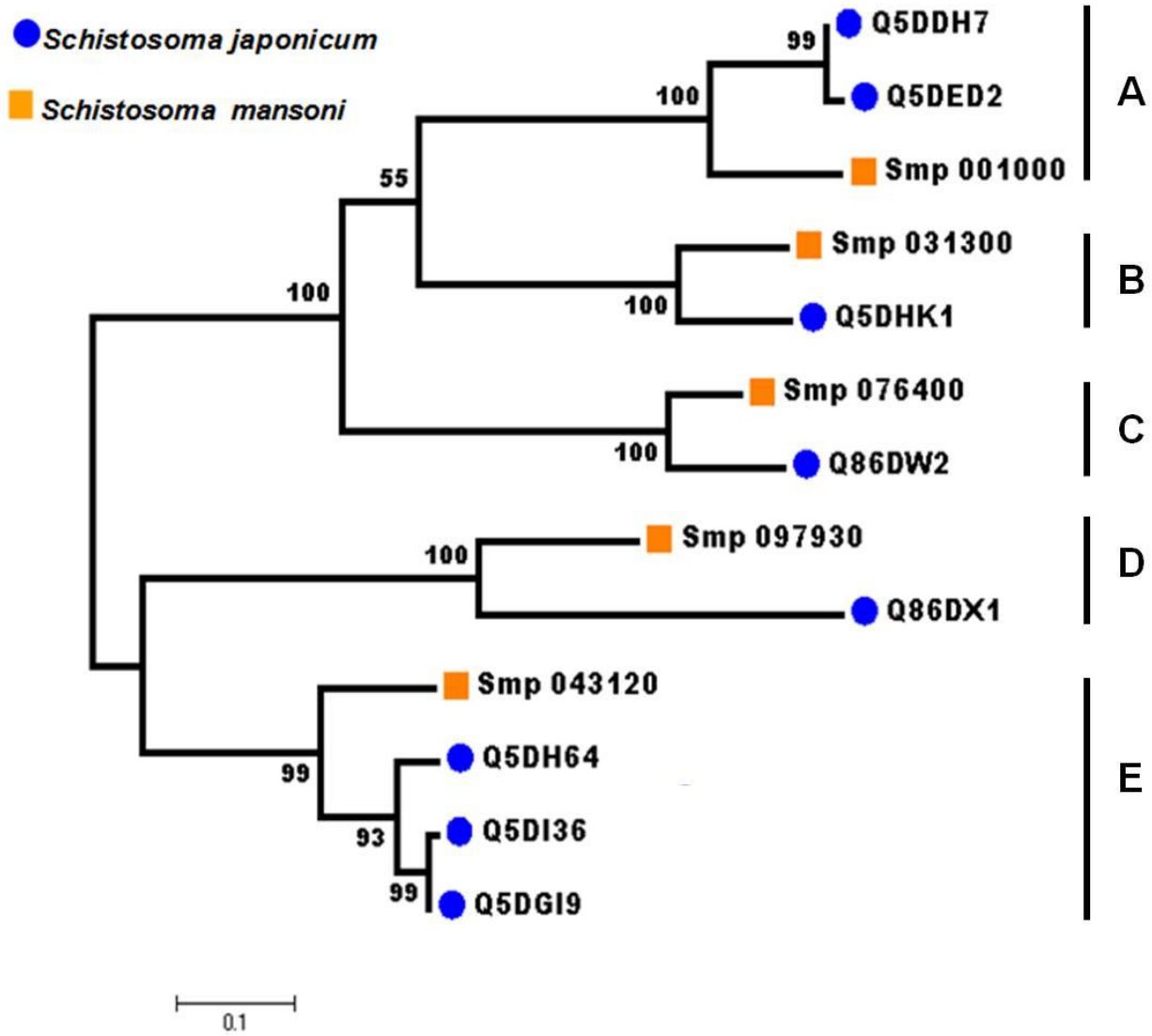


Figure 9: Grouping of 13 *Schistosoma* universal stress protein sequences

The phylogenetic tree was generated with MEGA5 using the maximum likelihood method. The 13 *Schistosoma* universal stress protein sequences clustered in five groups (A-E). The numbers near the clades are the statistics from the 1000 bootstrap that support the phylogeny recovering of the clades. A visual analytics resource to view the image with other associated data is available at http://public.tableausoftware.com/views/schisto_features_usp/phylotrees

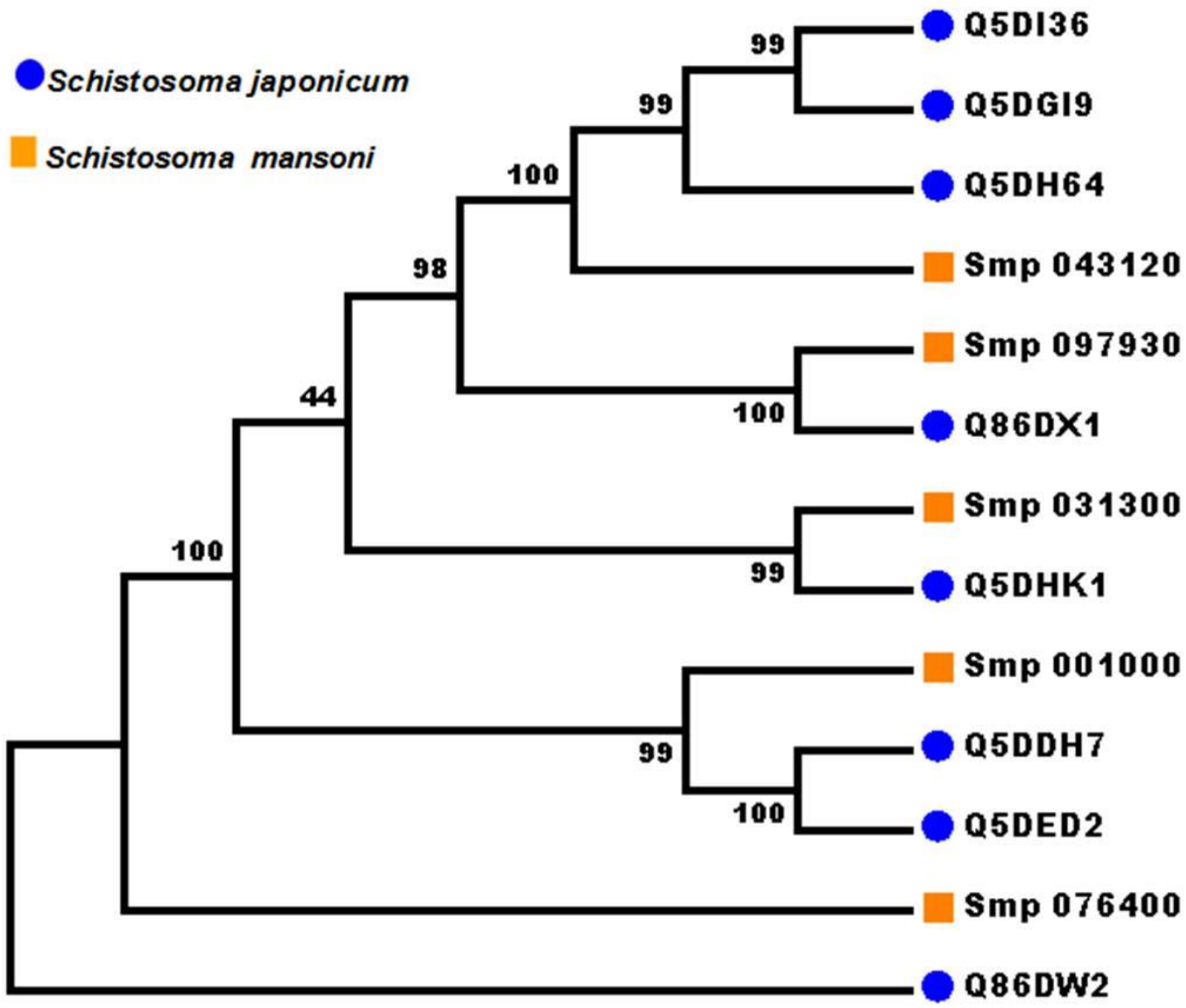


Figure 10: Parsimony test for the phylogeny tree reconstructed for *Schistosoma* universal stress proteins with the maximum likelihood method

Maximum Parsimony analysis on each of the 1000 bootstrap replications from maximum likelihood (Figure 9) determines 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 to view the image with other associated data is available at http://public.tableausoftware.com/views/schisto_features_usp/phylotrees

4.5. Dynamic Integration of Annotation Features for *Schistosoma* Universal Stress Proteins

The groupings described in the previous sections are based on the primary amino acid sequences of the universal stress proteins. To facilitate dynamic integration and updates of data sources a web-based visual analytics

resource was developed. As mention previously, the 5 annotation features with variable frequency per protein were 3D Chemical Ligand, Ligand Binding Amino Acid, Amino Acid and Sequence Position, Developmental Expression and Subcelluar Localization. To help guide further research and hypotheses generation, the research checked for the phylogenetic groups in which the ligand amino acids are identical in amino acid type and the position of the amino acids. Other features based on the protein sequences that have been considered are chemical ligands, kinase type, kinase score, and subcelluar 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. Nonetheless, the information can assist in directing new research.

The 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 11). An integration view for Group D (Q86DX1 and C1M0Q2 (Smp_097930) is presented to show the differences in patterns of identical annotations when compared to Group C (Figure 12). The biologically relevant chemical ligands predicted to bind to the Group C proteins include 4 phosphate containing ligands (ADP, AMP, ATP and GTP) and 3 metallic ion ligands (calcium, magnesium and zinc). The two proteins in the group were also predicted to be (i) localized in the cytoplasm and (ii) capable of phosphorylation by phosphokinase C (PKC) 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 gene. The schistomulum stage had the only identical annotation for 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, presence of ATP binding motif, kinase type and kinase score is shown in Figure 13.

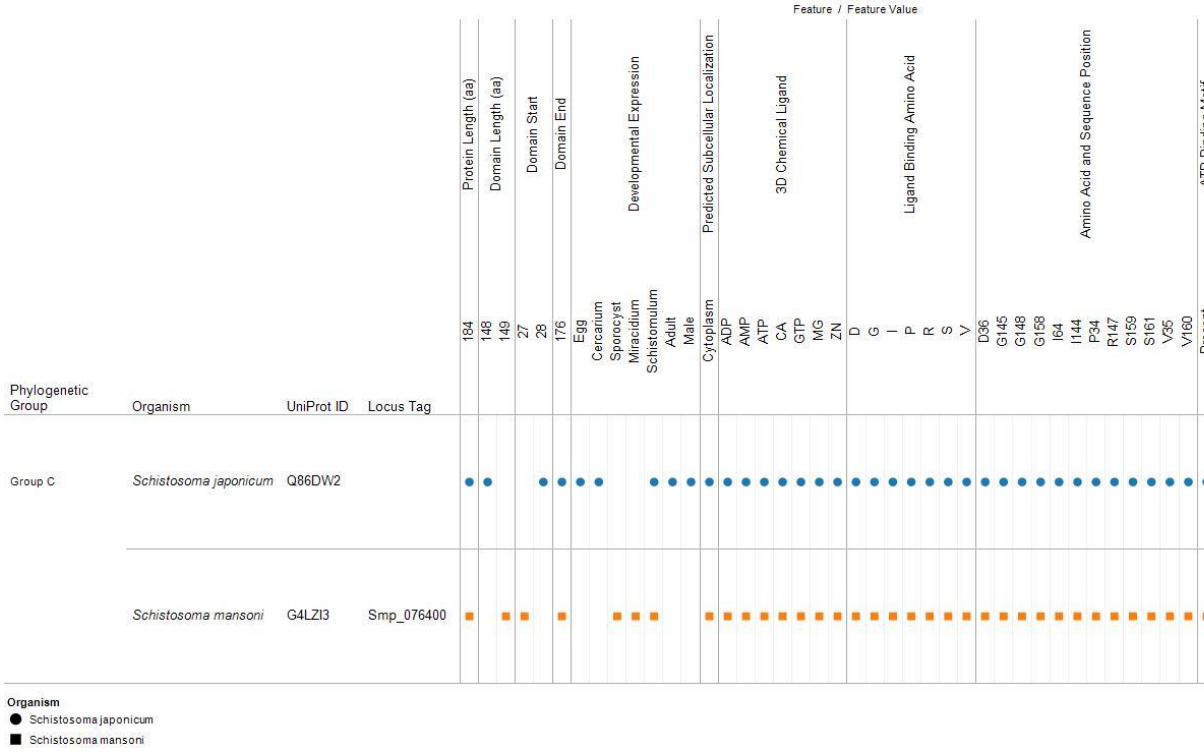


Figure 11: Integration and visualization of data on sequence features, evolutionary relatedness and developmental expression of *Schistosoma* universal stress proteins (Q86DW2 and G4LZI3)

The integration and visualization design was implemented in the visual analytics software environment. A visual analytics resource to interact with the view is available at http://public.tableausoftware.com/views/schisto_features_usp/phylo_group

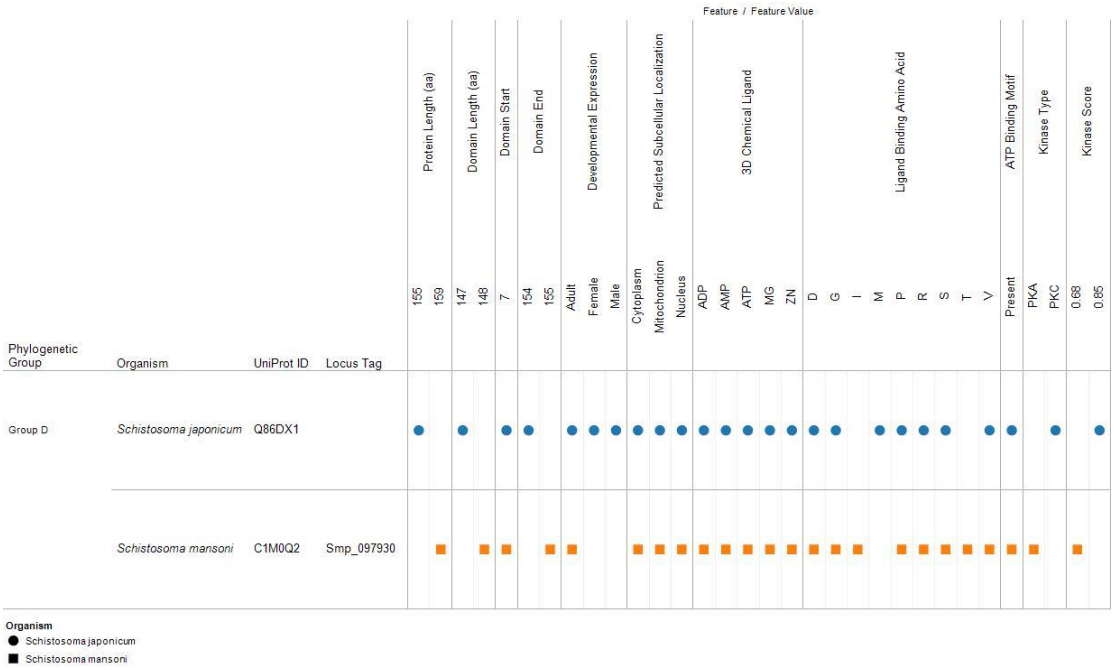


Figure 12: Integration and visualization of data on sequence features, evolutionary relatedness and developmental expression of *Schistosoma* universal stress proteins (Q86DX1 and C1M0Q2)

This figure is to illustrate the decision making process. In comparison to Q86DW2 and G4LZI3 (Figure 7), the annotations for protein sequence length, biologically relevant chemical ligands and ligand binding amino acids (type and position) were not identical. A visual analytics resource to interact with the view is available at http://public.tableausoftware.com/views/schisto_features_usp/phylo_group

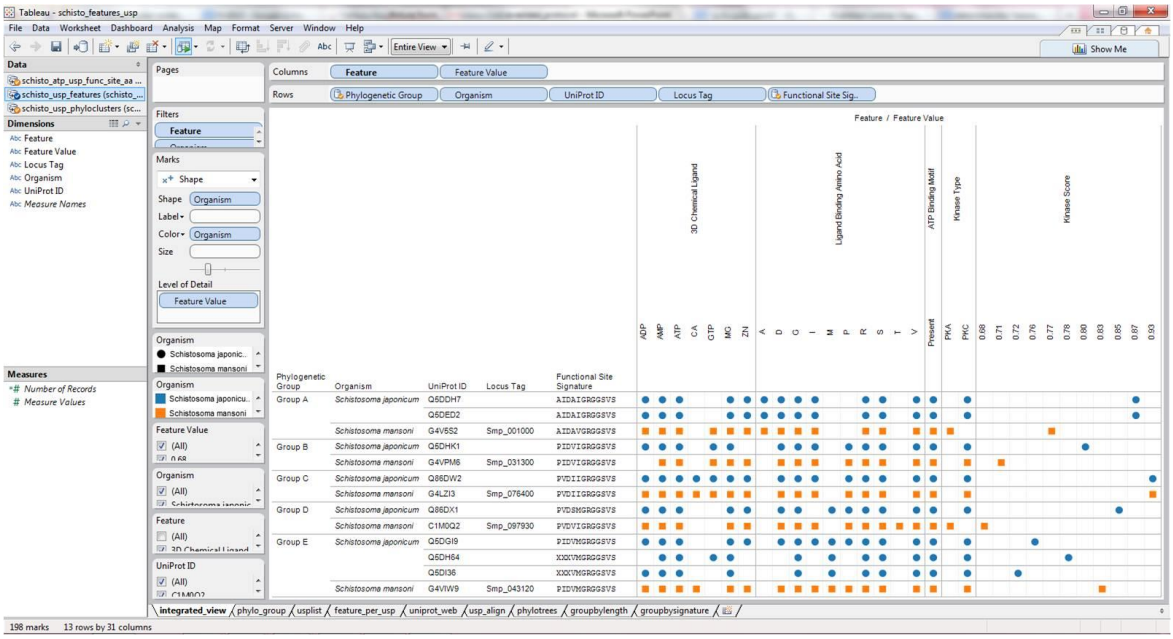


Figure 13: Design layout and visualization of datasets from sequence analysis, evolutionary relatedness and developmental expression of 13 *Schistosoma* universal stress proteins.

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

To facilitate additional exploration of the data without the need to access the internet, a packaged workbook file is available as a Supplemental File. The free software Tableau Reader (http://www.tableausoftware.com/products/reader) can be used to view the supplemental file.

Objective 2: Identify biological function relevant protein sequence and structure features for prioritized universal stress proteins from *Schistosoma* species.

4.6. Physicochemical Characterization of *Schistosoma* USPs

Primary structure analysis shows that the schistosome USP contain more serine followed by isoleucine, valine, leucine and lysine in that order, while there were very few tryptophan followed by cysteine, glutamine, phenylalanine, methionine and tyrosine than the rest of the other amino acid residues (Table 6). Also there were almost equal amounts of proline, alanine and glycine among the USP genes. The isoelectric point (pI) indicates that the schistosome USPs are more basic in nature as most of their pI values were above 7, with those from *S. japonicum* highly basic than those from *S. mansoni*. The following proteins from *S. japonicum* were acidic*:* Q5DHK1, Q5DDH7 and Q5DED2; and for *S. mansoni* only Smp_031300 was acidic (Table 7). In *S. japonicum,* the net surface charge distributions were as follows: (1) positive charge: (Q86DX1, Q5DI36, Q86DW2 and Q5DH64); (2) negative charge: (Q5DHK1, Q5DDH7, Q5DED2) and (3) neutral: (Q5DGI9). For *S. mansoni* charges were distributed as follows (1) positive charge: (Smp_076400 and Smp_097930, (2) negative charge: (Smp_031300) and (3) neutral charge: (Smp_001000 and Smp_043120).

Table 6: Amino Acid Content of Universal Stress Proteins from *Schistosoma* species

UniProt ID	*Schistosoma japonicum*								*Schistosoma mansoni*				
	Q5DHK1	Q86DX1	Q5DI36	Q5DDH7	Q86DW2	Q5DED2	Q5DGI9	Q5DH64	G4V5S2	G4VPM6	G4VIW9	G4LZI3	C1M0Q2
Ala	3.5	3.2	6.0	5.2	3.8	5.6	5.7	8.5	6.3	4.4	5.6	4.3	6.9
Arg	3.5	5.2	3.8	5.8	4.9	6.2	5.0	3.1	4.6	5.0	5.0	4.9	4.4
Asn	5.8	6.5	2.3	3.5	6.5	3.1	2.5	0.8	5.7	7.5	3.1	4.3	5.0
Asp	5.2	3.2	6.0	4.1	3.3	5.0	6.9	6.2	4.6	5.6	6.2	6.0	3.8
Cys	2.3	2.6	0.8	1.7	1.1	1.9	1.3	0.8	1.7	1.2	1.2	0.5	0.6
Gln	2.3	0.6	2.3	1.7	1.1	1.9	1.9	3.3	2.9	1.2	2.5	1.1	3.8
Glu	5.8	7.1	4.5	7.6	7.1	6.2	5.0	3.1	5.2	6.2	5.6	5.4	4.4
Gly	6.4	7.1	6.0	4.7	8.2	4.4	5.7	6.2	5.2	6.2	5.6	8.2	5.0
His	5.8	2.6	6.0	4.1	4.3	3.8	5.7	6.2	4.0	4.4	5.0	4.9	3.1
Ile	7.6	9.0	8.3	8.7	8.2	8.8	8.8	7.8	8.6	8.8	8.1	7.6	9.4
Leu	12.2	7.7	7.5	7.0	8.2	7.5	6.9	8.5	6.9	9.4	6.2	7.6	8.8
Lys	7.0	5.8	8.3	3.5	7.1	3.8	6.9	8.5	5.2	6.2	6.9	7.1	8.2
Met	2.3	1.3	3.8	1.7	2.7	1.9	3.1	3.1	1.1	3.1	5.0	2.7	1.9
Phe	2.9	1.9	2.3	1.7	3.3	1.9	2.5	2.3	2.3	4.4	2.5	3.3	4.4
Pro	4.1	7.1	6.0	4.7	5.4	4.4	6.7	6.2	5.2	3.8	6.9	5.4	6.9
Ser	9.3	9.7	7.5	13.4	7.6	12.5	8.2	8.5	12.1	8.8	8.1	9.2	4.4
Thr	3.5	4.5	5.3	3.5	6.5	3.8	4.4	5.4	4.0	2.5	4.4	6.5	6.9
Trp	0.6	1.3	0.0	1.3	0.5	0.6	0.6	0.0	1.1	0.6	0.6	0.5	0.6
Tyr	3.5	3.9	3.8	4.7	2.2	5.0	4.4	3.1	4.6	4.4	3.1	2.2	2.5
Val	6.4	9.7	9.8	11.6	8.2	11.9	8.8	8.5	8.6	6.2	8.1	8.2	8.8

Note: The values are in %. Schistosome USPs contain more serine followed by isoleucine, valine, leucine and lysine in that order, while there were very few tryptophan followed by cysteine, glutamine, phenylalanine, methionine and tyrosine than the rest of the other amino acid residues.

The computed instability indices of these proteins were conspicuously above 40, except Q5DI36, Q5DGI9, and Q5DH64 from *S. japonicum* and Smp_076400 from S*. mansoni* (Table 7)*.* The USP from *S. japonicum* were predicted to be highly unstable based on their computed instability indices (as they have higher values above 40 (Table 7). Most of the proteins have very moderate extinction coefficients except Q5DI36 and Q5DH64 with relative low

extinction coefficients values. All the proteins had a predicted half-life of greater than 30. These three parameters (instability indices, extinction coefficient and half-life values) account for the stability of most proteins. These values infer that the USPs are relatively unstable in their native state. Their relative high aliphatic index values and conspicuously very low grand average hydropathy (GRAVY) indicate that *Schistosoma* USP genes are very reactive in the water environment or aqueous phase (Mbah et al., 2012) (Table 7), which seems consistent with their immediate environment and developmental stages.

Table 7: Physicochemical Parameters of *Schistosoma* USP Genes

USPID	pI	Net Charge	Extinction Coefficient (M-1 cm-1)	Instability Index	Aliphatic Index	GRAVY
Q5DHK1	6.79	Negative	14690	42.07	99.13	-0.208
Q86DX1	7.68	Positive	20190	41.6	96.71	-0.155
Q5DI36	8.67	Positive	7450	32.43	95.94	-0.147
Q5DDH7	5.95	Negative	23045	55.75	100.17	-0.048
Q86DW2	8.87	Positive	11585	54.16	91.03	-0.264
Q5DED2	6.29	Negative	17545	52.56	103.44	0.013
Q5DGI9	7.13	Neutral	16055	37.77	92.52	-0.23
Q5DH64	9.04	Positive	5960	36.42	96.74	-0.074
Smp_001000	7.05	Neutral	23045	41.88	91.84	-0.224
Smp_031300	6.65	Negative	16055	43.97	93.19	-0.261
Smp_043120	7.1	Neutral	13075	57.8	85.25	-0.298
Smp_076400	7.92	Positive	11460	34.97	87.34	-0.305
Smp_097930	9.64	Positive	11460	41.44	103.58	-0.036

Note: Their computed instability indices indicate that the USPs are very unstable in their native state. Their relative high aliphatic index and conspicuously very low grand average hydropathy (GRAVY) indicate that *Schistosoma* USP are very reactive in water environment. In terms of charge distribution, USPs with PI >7 are positive charge while with those with pI<7 are negative charge. The pI value of 7 is taken as neutral.

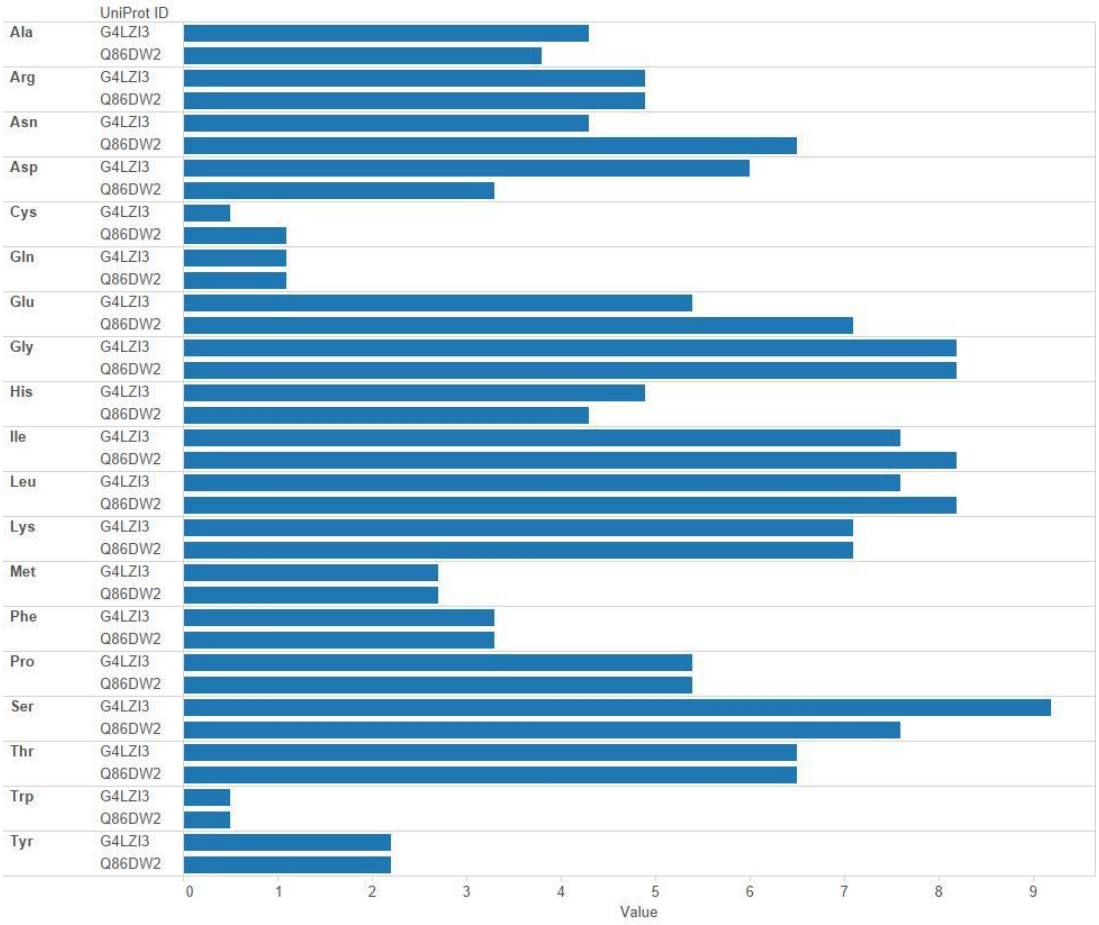


Figure 14: Comparative amino acid content of prioritized universal stress proteins from *S. japonicum* (Q86DW2) and *S. mansoni* (G4LZI3).

4.7. Structural and Functional Relationship of Prioritized *Schistosoma* USPs

The two USPs with sequence length of 184 aa were prioritized for structural and functional characterization. The conserved domain search shows that Smp_076400 and Q86DW2 are made up of single USP domains with ligand binding sites (Figure 15). The ligand binding sites of both Smp_076400 and Q86DW2 are made up in part by the same residues of ATP binding motif (Gly145, Arg147, Gly148, Gly158, and Ser159), using the ATP binding motif signature (G-2X-9x-G(S/T) (Sousa and McKay, 2001, Zarembinski et al., 1998).

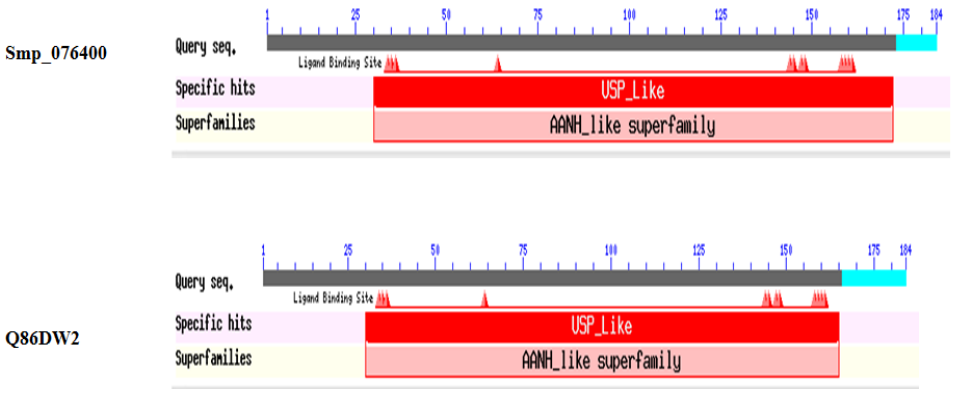


Figure 15: Conserved domain of prioritized USPs (Smp_076400 and Q86DW2)

Note: The conserved domains of both proteins are made up of a single USP domain with ligand binding site constituting ATP binding motif residues Gly145, Arg147, Gly148, Gly158 and Ser159using the ATP binding motif signature (G-2X-9x-G(S/T).

The crystal structures of both Smp_076400 and Q86DW2 have not yet been determined. This research presents the 3D homology model structures of both USPs using 1mjh (chain B) as template using the SwissModel server. The statistics of the targets and template showing their comparison was derived from the SwissModel server (Table 8).The quality of the model was measured using the QMEAN4 score. QMEAN stands for Qualitative Model Energy Analysis. It is a composite scoring function describing the major geometrical aspects of protein structures (Benkert et al., 2008) . QMEAN server is available at: http://swissmodel.expasy.org/qmean. The QMEAN4 score measure the global score of the whole model reflecting the predicted model reliability range from 0 to 1 using statistical potential terms only. The modeled structures had QMEAN4 score values of 0.57 and 0.58 respectively (Table 8), suggesting that they are good model and biological informative.

Table 8: Modeling Statistics of Smp_076400 and Q86WD2 derived from SwissModel server

Model information	Smp_076400	Q86WD2
Modeled residue aligned range	28 to 179	26 to 179
Target sequence length	184	184
Template sequence length	162	162
Template	1mjhB (1.70 Å)	1mjhB (1.70 Å)
Sequence identity	26.32%	24.03%
Target aligned coverage	82.60%	81.52%
E value	4.50e-26	6.40e -27
Model quality	QMEAN4 score: 0.57	QMEAN4 score: 0.58
Ligand in template	ATP: 1	ATP: 1

Note: Residue 28 to 179 of Smp_076400 was aligned with the template covering 82.60% of its entire sequence length with 26.32% sequence identity with a QMEAN4 score value of 0.57. For Q86WD2, the QMEAN4 score was 0.58. Overall data suggest the models could be of good quality and biological informative.

The qualities of the models were further evaluated using Ramachandran plot from RAMPAGE server (Figure 16). The Ramachandran plot has been widely used for assessing the quality of predicted models based on the phi – psi torsion angles of all the residues (Gopalakrishnan et al., 2007, Hooft et al., 1997, Kelley and Sternberg, 2009, Porter and Rose, 2011, Zhou et al., 2011). The Ramachandran plot analysis indicates that most residues are in the favoured region (total 143residues [Smp_076400 (95.3%) and Q86DW2 (94.1%)] with very few residues in the outlier region (2 residues (1.3%) and (3 residues (2.0%) respectively (Table 9).

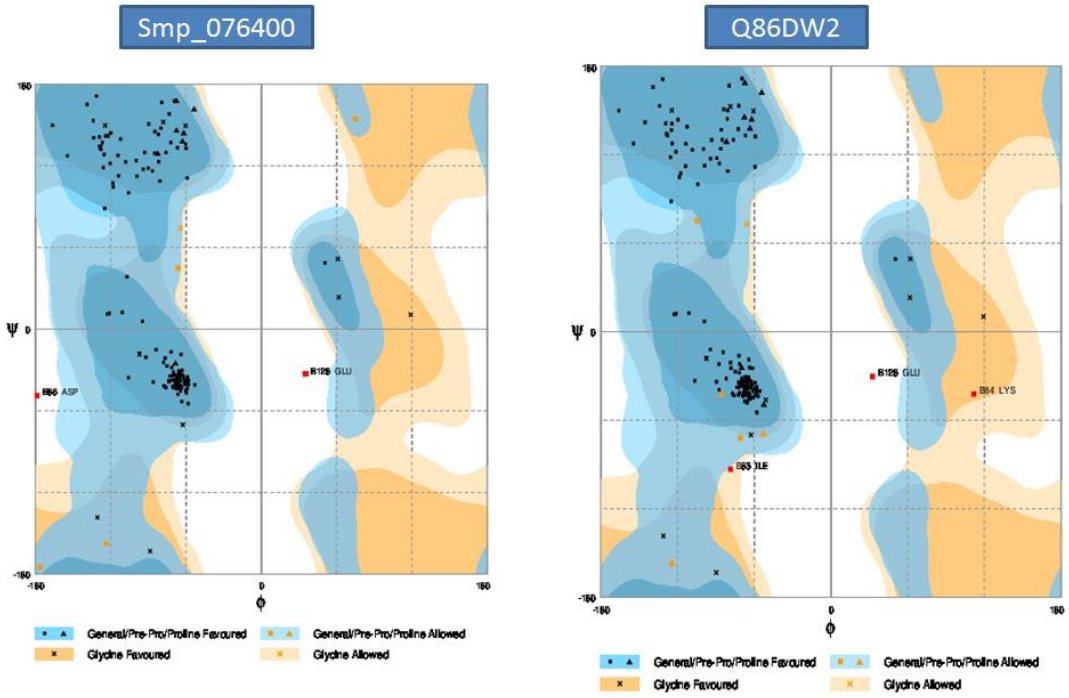


Figure 16: Model quality evaluation using Ramachandran plot based on RAMPAGE server.

Ramachandran plot based on RAMPAGE server: It shows the distribution of the whole model residues into various energy regions based on the phi–psi torsion angles. It shows more model residues are in the favored region (total 143residues [Smp_076400 (95.3%) and Q86DW2 (94.1%)] with very few residues in the outlier region (2 residues (1.3%) and (3 residues (2.0%) respectively.

Residues in the high energy outlier region are considered unstable and can compromise the quality of a modeled structure (Maccallum et al., 1995). A good quality model is expected to have over 90% of the residues in the core region (Kelley and Sternberg, 2009, Porter and Rose, 2011, Zhou et al., 2011). This indicates that both Smp_076400 and Q86WD2 predicted models with more residues in the core region could be of good quality and biological informative (Table 9).

Table 9: Evaluation of model residues from Smp_076400 and Q86DW2

Evaluation of model residues from Smp_076400	Evaluation of model residues from Q86DW2
Allowed region	*Allowed region*
[B 53 :MET] (-123.55,-157.72)	[B 28 :THR] (-104.75, 75.02)
[B 68 :LEU] (-66.34, 44.88)	[B 53 :MET] (-124.29,-157.24)
[B 73 :ASN] (-176.33,-174.74)	[B 69 :LEU] (-70.72, -72.12)
[B 85 :ASP] (75.06, 154.28)	[B 70 :PRO] (-52.85, -68.90)
[B 127 :GLY] (-64.35, 74.14)	[B 80 :SER] (-85.63, -42.41)
	[B 127 :GLY] (-65.79, 73.11)
Outlier region	*Outlier region*
[B 86 :ASP] (-178.46, -48.70)	[B 83 :ILE] (-78.48, -93.19)
[B 128 :GLU] (35.43, -32.83)	[B 84 :LYS] (111.90, -42.19)
Number of residues in favoured region (~98.0% expected): 143 (95.3%)	[B 128 :GLU] (32.76, -30.30)
Number of residues in allowed region (~2.0% expected): 5 (3.3%)	Number of residues in favoured region (~98.0% expected): 143 (94.1%)
Number of residues in outlier region: 2 (1.3%)	Number of residues in allowed region (~2.0% expected): 6 (3.9%)
	Number of residues in outlier region: 3 (2.0%)

4.8. Structural and Functional Regulation of Smp_076400 and Q86WD2 USPs

The 3D ligand site analysis of Smp_076400 shows that the following 23 residues Pro34, Val35, Asp36, Ser38, His40, Ser41, Val62, His63, Ile64, Glu66,Pro126, Ile144, Gly145, Asn146, Arg147, Gly148, Ile149, Gly150, Thr151, Gly58, Ser159, Val160, Ser161are responsible for binding eight ligands (GTP, ADP, MG, CTP, AMP, CA, ATP, ZN) (Figure 15 and Figure 17). These 23 residues contain the ATP binding motif (G-2X-9X-G(S/K) of Smp_076400 USP. The Q86WD2 protein has the following 3D ligand binding sites; Gln7, Ser8, Thr9, Ser10, Asp11, Gly12, Leu13, Tyr14, Pro34, Val35, Gly37, Ser38, Glu39, His40, Ser41, Glu66, Ile144, Gly145, Asn146, Arg147, Gly148, Ile149, Gly150, Thr151, val160, Ser161. These site residues binds to

the following eight chemicals as ligands (GTP, MG, AMP, ATP, ZN, ADP, ADX, CA).The Smp_076400 USP was prioritized further to relates its structural features to functional regulation. The ATP motif residues are depicted black and are constituted by Gly145, Arg147, Gly148, Gly158 and Ser159; while the other 18 binding residues are illustrated in green (Figure 17B and D).

The presentation of these residues on the secondary structure and their binding to the predicted chemical ligands are shown in Figure 17C. The ATP binding motif (Black) is located on the helix, sheet and the loop, while the other binding residues (Green) are located on the same type of secondary structures (Figure 17B). The 3D binding residues architecture on the folded structure shows that the ATP binding residues (Black) are oriented parallel to each other at the entrance of the active groove, while the other binding residues (Green) are situated deeper in the groove where they bind to other environmental chemical ligands such as metallic Mg2+, Zn2+ and Ca2+ ions (Figure 17D and Figure 17E).

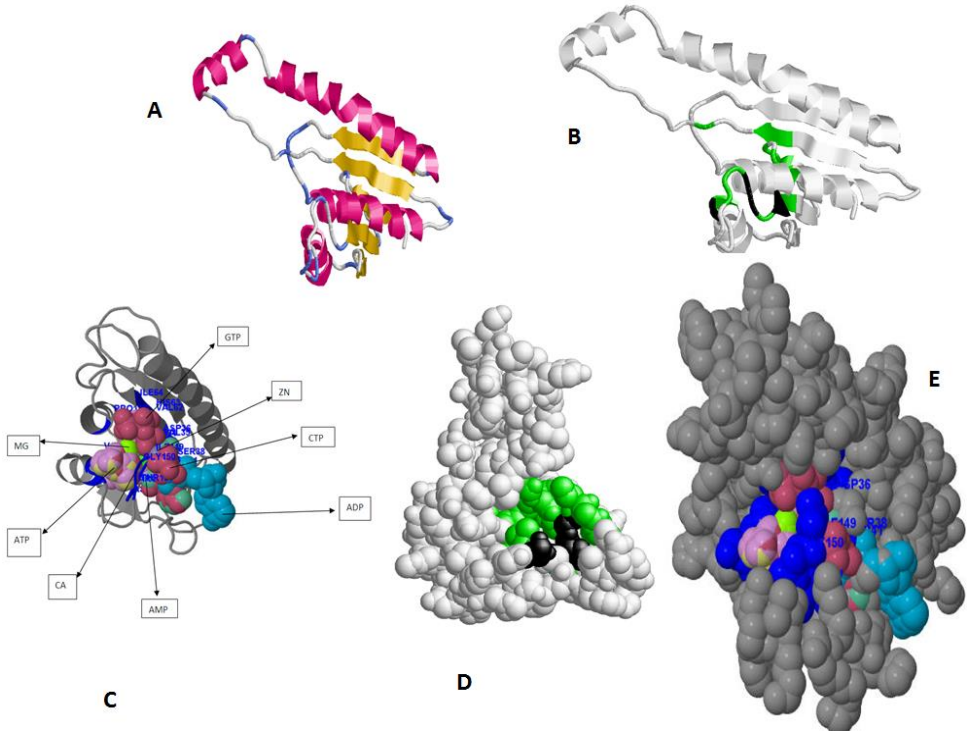


Figure 17: Structural and functional features of USP Smp_076400

A: The secondary structure components composed of 5 helices, 5sheets and 16 turns. The helices are represented in red; sheets are yellow and turns are depicted in blue. The white or grayish colored regions near the turns are other secondary structures.

B: The ATP binding motif (Black) is located on the helix, sheet and the loop; the other binding residues (Green) are located on the secondary structures.

C: The presentation of binding residues on the secondary structure and their binding on predicted ligands.

D: The 3D binding residues architecture on the folded structure shows the exposed ATP binding residues (Black) oriented parallel to each other at the entrance of the active groove, while the other binding residues (Green) are situated deeper in the groove where they bind to other ligands like metallic Mg^{2+}, Ca^{2+} and Zn^{2+} ions.

E: This adaptive structural conformation is efficient in ATP binding and keeping the metallic ligands very stable in the active grove while in contact with the ATP molecule during phosphorylation through a series of hydrogen bonds. The blue colored residues are the combined binding sites including the ATP binding motif residues.

Objective 3: Determine the distinctive structural protein features of a predicted regulator of *Schistosoma* adenylate cyclase activity that has possible influence on the functioning of universal stress proteins.

4.9. Amino Acid Content and Physicochemical Parameters

The analysis of amino acid content and physicochemical parameters suggests that Smp_059340.1 is hydrophilic due to the presence of high polar amino acid residues (50.7%) against non-polar (hydrophobic) amino acids residues (34.95%) in the sequence (Table10). Therefore, the Smp_059340.1 can be described as moderately hydrophilic. The protein is made up of 379 amino acid residues, with all the 20 amino acids present in the primary structure, constituting an average molecular weight of 44045.5 Da. The analysis indicates that there are more Leu, Glu, Lys, Arg, Ile and Ser in that order (Table 10).The atomic composition (6191) consists of 1961 carbons (C), 3093 hydrogen (H), 547 nitrogen (N), 573 oxygen (O) and 17 sulfur atoms with a molecular formula of $C_{1961}H_{3093}N_{547}O_{573}S_{17}$ (Table 11). The 17 sulfur atoms were constituted by the 11 cysteine and the 6 methionine residues present in the primary structure.

The computed pI was 8.51 (pI>7), indicating that this protein is basic in nature. The ProtParam extinction coefficient at a wavelength of 280nm measured in water is favourable because proteins are able to absorb strongly at this wavelength than other substances which may be commonly found in the solution. The extinction coefficient for this protein was computed with respect

to Cys, Trp and Tyr because they were all present in the primary structure. The estimated half-life of this protein with Met as the N-terminal of the sequence was 30 (>20) hours. The predicted instability index was 48.15 and 82.17 for aliphatic index with a very low GRAVY Index of -0.458. The Hopp-Woods scale identified three regions on this polypeptide predicted to be highly hydrophilic. The hydrophilic regions are shown to have peak values greater than 0 (Figure 18), an indication that Smp_059340.1 could be drug target for schistosomiasis.

Table 10: Amino acid composition of Smp_059340.1 computed using ProtParam server

Amino acid	Composition (%)	Hydrophilic (%)	Hydrophobic (%)
Ala	6.1		6.1
Arg	7.4	7.4	
Asn	5.8	5.8	
Asp	5.8	5.8	
Cys	2.9		
Gln	3.2	3.2	
Glu	8.2	8.2	
Gly	4.2		4.2
His	2.1	2.1	
Ile	7.2		7.2
Leu	9.2		9.2
Lys	7.9	7.9	
Met	1.6		
Phe	5.3		
Pro	2.9		2.9
Ser	6.1	6.1	
Thr	4.2	4.2	
Trp	1.1		
Tyr	3.7		
Val	5.3		5.3
Total	**100**	**50.7**	**34.9**

Note: The composition of each amino acid residue is indicated in percentage. The composition of hydrophilic amino acids is 50.7% while hydrophobic amino acids constitute 34.9%. The protein can be described as moderately hydrophilic.

Table 11: Physicochemical properties of Smp_059340.1 computed using ProtParam server

ProtParam parameters	Values
No of amino acids	379
Molecular weight	44045.5Da
Theoretical pI	8.51
Number of negative charge residues	53
Number of positive charge residues	58
Formula	$C_{1961}H_{3093}N_{547}O_{573}S_{17}$
Extinction coefficient	43485M-1cm-1
Estimated half life	30 hours
Instability index	48.15
Aliphatic index	82.17
Grand average of hydropathicity(GRAVY)	-0.458
Total number of atoms	6191

Note: The physicochemical parameters define the protein chemical and physical properties in its native state. Smp_059340.1 has a net positive charge and is basic in nature (pI>7).

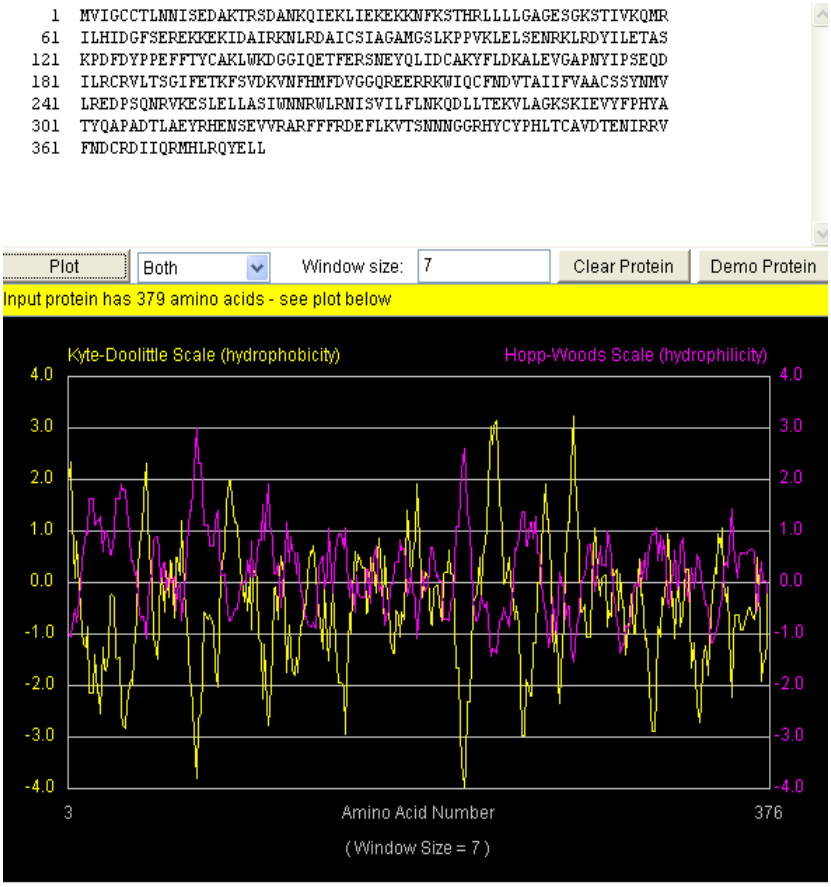


Figure 18: The hydropathy plot for *Schistosoma mansoni* protein Smp_059340.1

The yellow plot is the Kyte-Doolittle hydrophobicity plot. Sections of the plot with high values > 0.0 are highly hydrophobic or membrane spanning segments. The magenta plot is the Hopp-Wood hydrophilicity plot. Higher values above>0.0 predict rich charge exposed regions with potential antigenic site. Smp_059340.1 gene shows potential antigenic sites with values ≥ 2. Above the plot is the Smp_059340.1 amino acids sequence.

4.10. Secondary Structure Characterization

The SOPMA tool at Expasy (http://expasy.org/tools/) was used for predicting the secondary structures (Table 12). The data indicates that this protein has mixed secondary structures made up of alpha helix (Hh) (54.62%), extended strand (Ee) (11.35%), beta turn (Tt) (4.75%) and random coil (Cc) (29.29%). The high alpha helix content may be due to the rich alanine and other hydrophobic alpha helix residues like Phe, Leu and Ile content of the protein (Table 12). The TMPRED server at Expasy (http://expasy.org/tools/) predicted possible transmembrane helices, with two possible models of

orientations in consideration (Table 13). The first model which is strongly preferred based on a total score of 1345, suggest that the N–terminus of the protein is inside and the region starts from the amino acid residue 218 - 237 spanning a total length of 20 amino acid residues in an inside → outside (i →o) orientation. The alternative model suggest an outside → inside (o → i) orientation spanning 23 amino acid residues from position 219- 241 with a lower score of 854. According to the prediction accuracy of the server the protein has transmembrane helices in an outside → inside orientation. The transmembrane regions are rich in hydrophobic amino acids while the hydrophilic regions are predicted as potential antigenic segments.

Table 12: Secondary structure composition for *Schistosoma mansoni* protein Smp_059340.1

Secondary structures	Representation	Composition (%)
Alpha helix	Hh	54.62
Extended stand	Ee	11.35
Beta turn	Tt	4.75
Random coil	Cc	29.29

Note: Secondary structure composition in percentages for Smp_059340.1 computed using SOPMA server at Expasy (http://expasy.org/tools/). The secondary structure elements are composed of four units (alpha helix, extended strand, beta turns and random coil). The structure is made up of more alpha helices and fewer beta turns.

Table 13: Possible transmembrane region and orientation for *Schistosoma mansoni* protein Smp_059340.1

Parameter	inside → outside (i→o)	outside → inside (o→i)
Length in amino acid position	218 - 237(20 residues)	219 – 241 (23 residues)
Score	1345	854
Orientation preference	Strong (++)	Weak (+)

Note: Possible transmembrane region and orientation identified by TMPRED server at Expasy (http://expasy.org/tools/). Two transmembrane orientations were predicted. The protein strongly preferred inside → outside (i →o) orientation base on a total score of 1345 with 20 amino acid residues spanning residue 218 – 237. The alternative model suggest an outside → inside (o → i) orientation spanning 23 amino acid residues from position 219- 241 with a lower score of 854.

The transmembrane regions are rich in hydrophobic amino acids and this region is well indicated with the Kyte and Dolittle average hydrophobicity plot of the protein residues (Figure 18), in which all the points were shown to have values above the 0 mark line. On the same plot (Figure 18) the Hopp – Woods scale predicts the potential antigenic regions of the protein. The hydrophilic regions have values greater than 0 and they are predicted as potential antigenic regions. On the Kyte – Dolittle plot; regions with values less than 0 are hydrophilic in nature. Hence the antigenic regions of proteins are hydrophilic. The transmembrane helices predicted with TMPRED server were visualized with the Pepwheel tool (Figure 19) using the 20 amino acids in an inside →outer orientation. The residues in the blue square (I, L, V) and the violet residues (C, A, W, F, Y) represent the hydrophobic amino acids while those in diamond (red) are the acidic residues (polar residues) (N, T, S, D, Q).

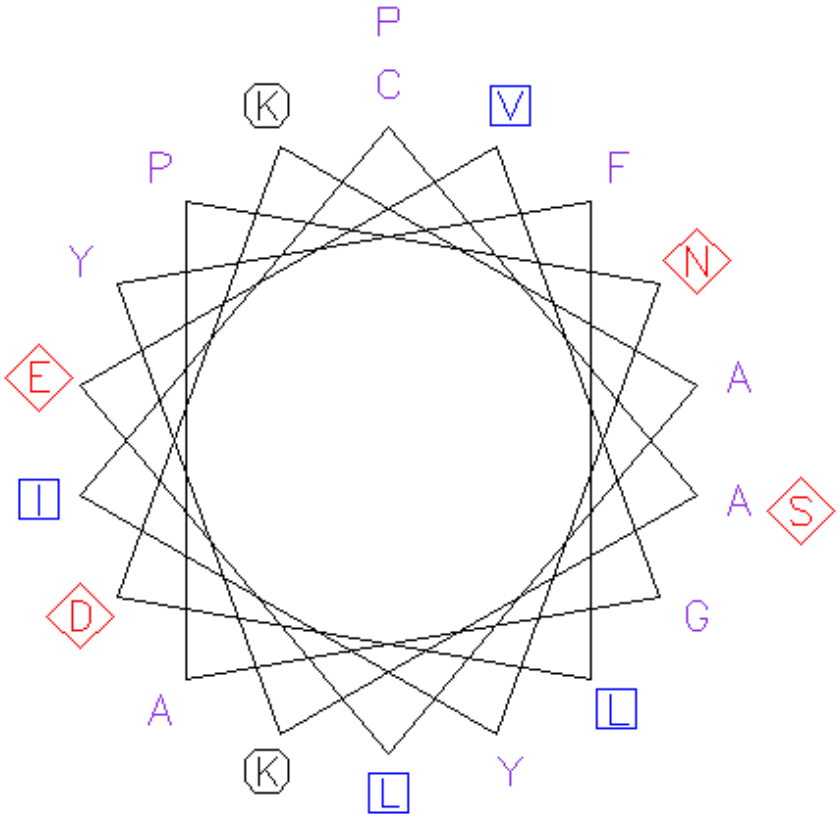


Figure 19: Transmembrane helices plot for *Schistosoma mansoni* Smp_059340.1.

The transmembrane helices were predicted with TMPRED server at Expasy (http://expasy.org/tools/). The popwheel displays the 20 amino acids transmembrane sequence in a helical representation as if looking down the axis of the helix. This representation is useful for highlighting amphipathicity and other properties of residues around a helix. The aliphatic residues in blue square (I, L, V) and the violet residues (C, A, W, F, Y) are hydrophobic, while hydrophilic residues are marked with diamond (red). The positive charge residues (both K) are with octagonal shape.

4.11. Conserved Domain and Functional Analysis of Smp_059340.1

The Conserved Domain search at the National Center for Biotechnology Information (NCBI) website revealed 12 functional units and their full residues (Table 14). The conserved domain search for Smp_059340.1 revealed 12 functional domains scattered within its amino acids sequence length (Figure 20, Table 14). Close observation on Table 14 shows that these functional units share many residues in common, suggesting that the Smp_059340.1 protein functions as a highly interconnected network of functional units. The next

sections provide descriptions of the structural and functional domain interactions to the overall regulatory mechanism of the adenylyl cyclase pathway that controls the general developmental processes of the schistosome.

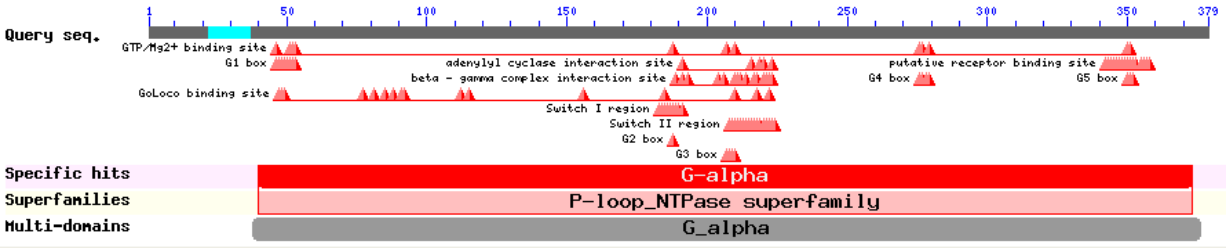


Figure 20: Visualization of the 12 functional domains of *Schistosoma mansoni* Smp_059340.1 protein

The Conserved Domain Search Tool (CD-Search http://www.ncbi.nlm.nih.gov/cdd/) was used to determine functional domains in Smp_059340.1. The position and span of each domain unit across the protein are shown. The positioning indicates interdomain connections in a probable complex network. The 12 domain units regulate the function of the Smp_059340.1 protein.

4.12. The Smp_059340.1 Modeled 3D Structural Analysis and Verification

The crystal structure of *S. mansoni* C4QDC7_SCHMA (Smp_059340.1) has not yet been determined; this research presents the 3D homology model structure of the protein using 1TL7C (chain C) as template with 93% of the residues modeled at >90% confidence level as analyzed using the Phyre2 server (Figure 21A). The statistics of Smp_059340.1 (target) and template showing their comparison was derived from the SwissModel server (Table 15). Residue 39 to 373 of Smp_059340.1 was aligned with the template covering 88.00% of its entire sequence length. Irrespective of the length difference between the template and the target, they both share 70.00% sequence identity.

Table 14: Function of the 12 domains and their constituted residues for *Schistosoma mansoni* protein Smp_059340.1

Conserved domain	Component residue(s)	Function
Mg2+/GTP binding site	G46, G51, K52, S53, T188, D207, G210, N276, K277, D279, C350, A351	The Mg2+ interacts with alpha subunits of the G proteins in the present of GTP to form a complex from which nucleotide dissociate slowly
Adenylyl cyclase interaction site.	I191, R216, I219, Q220, N223	Stimulates the adenylyl cyclase by binding to it
Beta – gamma complex interaction site	S189, I191, E193, H204, F206, G210, Q211, R212, E214, K217, W218, Q220, C221, F222, N223	The beta/gamma complex can also stimulate and inhibit the adenyly cyclase activity, but no model of its function had been established
GoLoco binding site	A47, G48, E49, D77, K81, D85, C88, A91, G92, R112, I115, I156, R185, G210, W218, F222	It binds $G\alpha_{i/o}$/GDP complex and prevents the spontaneous release of GDP by Gα, thus acting as a guanine nucleotide dissociation inhibitor (GDI)
Putative receptor binding site	H342, Y343, C344, Y345, P346, H347, L348, T349, C350, A351, V352, D353, E355, N356, I357, R358	This binds the protein to the serotonin receptor (Smp_12673)
Switch I region	R183, C184, R185, V186, L187, T188, S189, G190, I191	Molecular switches (called the effector loop)
Switch II region	V208, G209, G210, Q211, R212, E213, E214, R215, R216, K217, W218, I219, Q220, C221, F222, N223, D224	Molecular switches
G1 box motif	G46, A47, G48, E49, S50, G51, K52, S53	GTP binding signature
G2 box motif	T188	Involved in Mg2+ coordination
G3 box motif	D207, V208, G209, G210	It links the subsites for binding of Mg2+ and the γ phosphate of GTP (GTP γ-phosphate- binding site)
G4 box motif	N276, K277, Q278, D279	It recognizes the guanine ring
G5 box motif	C350, A351, V352	Buttress the guanine base recognition site

Note: In the *Schistosoma mansoni* protein Smp_059340.1, each of the residues contributing to the function of each domain in the complex network is presented. Most domains shared residues with each other making the network very complex at atomistic level.

The quality of the model was measured using the QMEAN4 score. QMEAN stands for Qualitative Model Energy Analysis. It is a composite scoring function describing the major geometrical aspects of protein structures (Benkert et al., 2008). QMEAN server is available at: http://swissmodel.expasy.org/qmeanThe QMEAN4 score measure the global score of the whole model reflecting the predicted model reliability range from 0 to 1 using statistical potential terms only. The modeled Smp_059340.1 structure had QMEAN4 score value of 0.68, suggesting that it could be a good model and biological informative.

The Swiss model server estimate per residue inaccuracy was visualized on the secondary structural components. The color gradient ranges from blue (more reliable region) to red (potential unreliable regions) (Figure 21B). More of the unreliable regions were located around the loops and random coils which are usually the most difficult parts to predict. The sequence alignment of the template and Smp_059340.1 (target) generated by HMM-HMM matching using the Phyre2 server shows that the alpha helices and beta strands were predicted with approximate better confidence values than the random coil (Figure 22A).

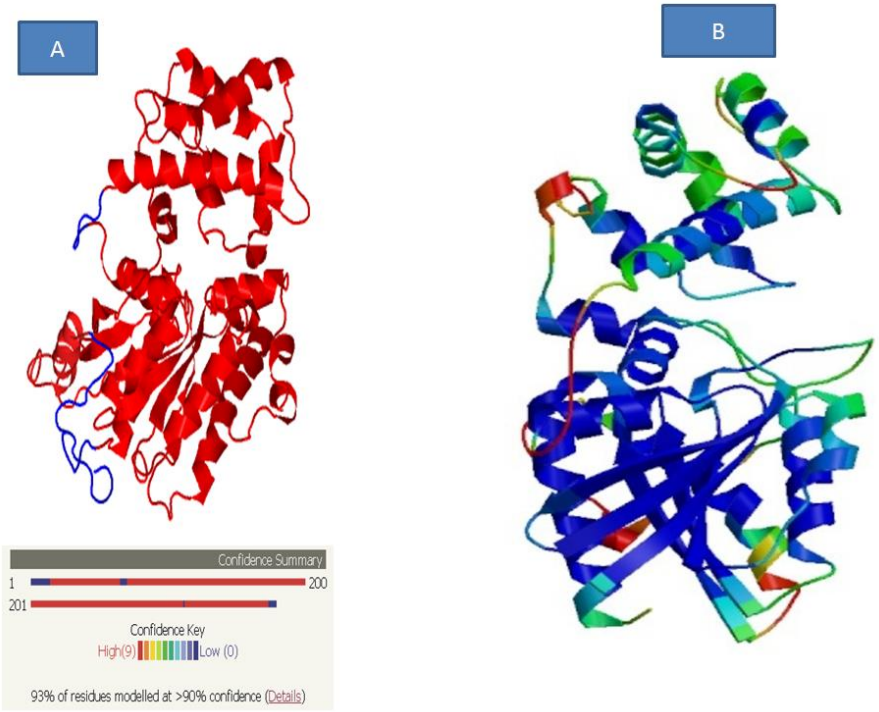


Figure 21: Visualization of the confidence level of the modeled residues and the error regions in the 3D modeled structure of Smp_059340.1.

Image A. The 3D modeled structure of Smp_059340.1 using 1TL7C (chain C) as template with 93% of the residues modeled at >90% confidence level analyzed with Phyre / Phyre2 server and confidence level range from red (high) to low (blue).

Image B. Residues with high score (good structure, highly conserved, deeply buried) are colored blue (low B-factor) while residue with low scores (poor structure, variable, exposed) are colored red (high B-factor). The intermediate values may range from green to yellow to orange. This indicates that the model has problems with the loops and alpha helix but not affecting the beta sheet structures.

Table 15: Modeling Statistics of Smp_059340.1 derived from SwissModel server

Model information	Value
Modeled residue aligned range	39 to 373
Target sequence length	378
Template sequence length	402
Template	1TL7C (2.80 Å)
Sequence identity	70.00%
Target aligned coverage	88.00%
E value	1.15e-133
Model quality	QMEAN4 score: 0.68
Ligand (s) in template	GSP:1, MG:1

Note: Residue 39 to 373 of Smp_059340.1 was aligned with the template covering 88.00% of its entire sequence length with 70.00% sequence identity. QMEAN4 score value of 0.68, suggests the model could be of good quality and can be biological informative.

Further comparison between the template and Smp_059340.1 (target) using structural superposition helped to establish the regions of the model, which agree and disagree (Figure 22B). As with every model the agree regions are trustworthy while disagree regions are generally treated with caution when making biological inferences (Kelley and Sternberg, 2009) .

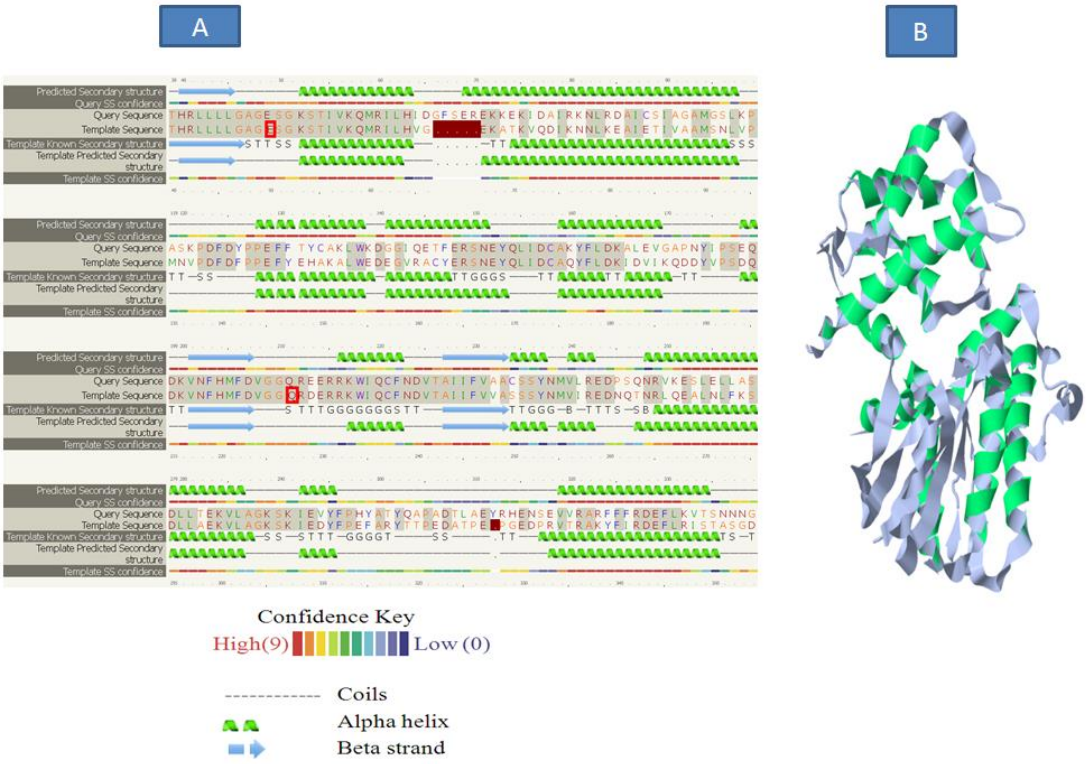


Figure 22: Sequence alignment and structural superposition of Smp_059340.1 (target) and the template

Image A. Alignment generated by Phyre2 server based on HMM-HMM matching shows that the alpha helices and beta strands were predicted with approximate better confidence values than the random coil. Identical residues in the alignment are highlighted with a grey background. Green helices represent α-helices, blue arrows indicate β-strands and faint lines indicate coil. The 'SS confidence' line indicates the confidence in the prediction, with red being high confidence and blue low confidence. The orange, yellow and green indicate a weak prediction.

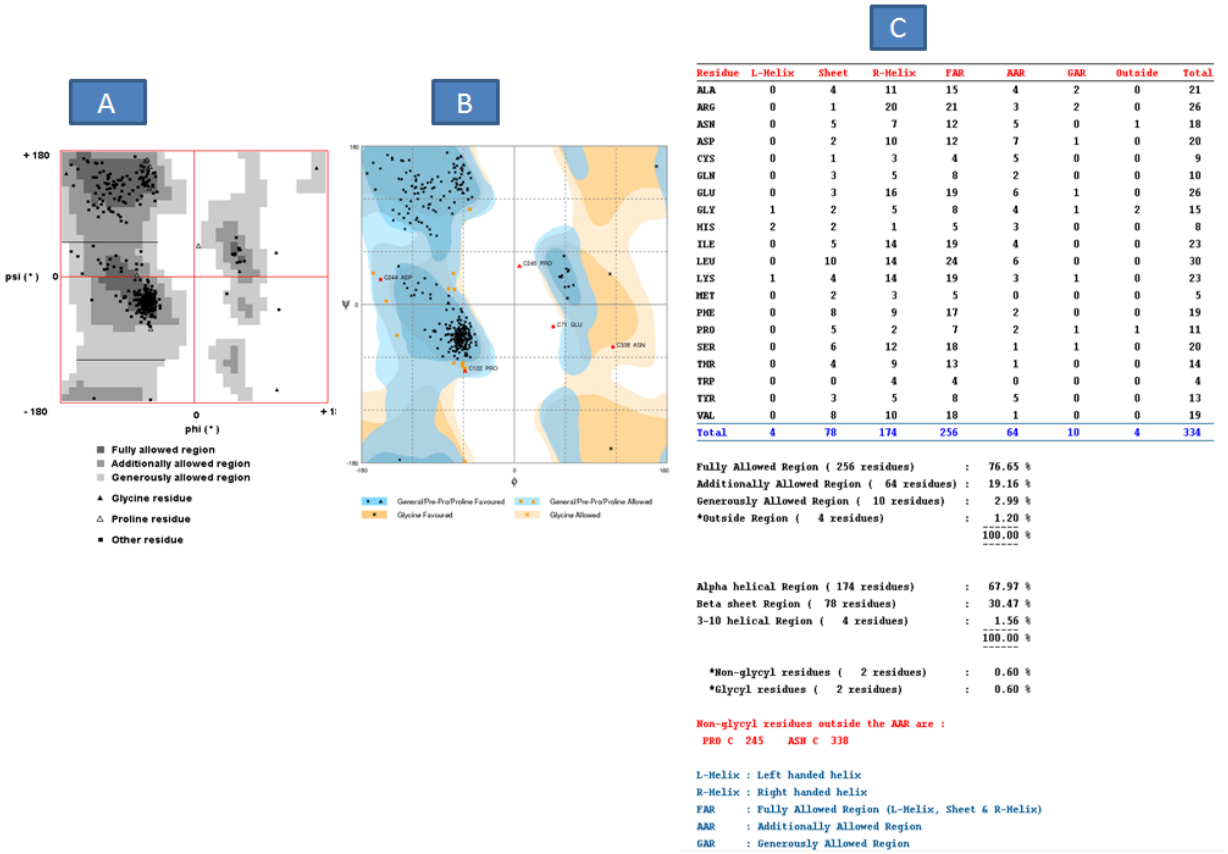
Image B. Comparison between the template and Smp_059340.1 (target) (Green colored) using structural superposition to establish the regions of the models which agree and disagree. The agree regions are trustworthy while disagree regions are generally treated with caution when making biological inference.

The research further evaluate the quality of the model using Ramachandran plot on the web (2.0) (Gopalakrishnan et al., 2007) (Figure 23A) and RAMPAGE server (Lovell SC, 2003) (Figure 23B). The Ramachandran plot had been widely used for assessing the quality of predicted models based on the phi – psi torsion angles of all the residues (Gopalakrishnan et al., 2007, Hooft et al., 1997, Kelley and Sternberg, 2009, Porter and Rose, 2011, Zhou et al., 2011). The fully allowed and additionally allowed regions from Ramachandran plot on the web (2.0) together are

equivalent to the favoured (core) region of the RAMPAGE server. Also the generously allowed region from Ramachandran plot on the web (2.0) is equivalent to the allowed region in RAMPAGE server (Figure 23A and 23B). The Ramachandran plot analysis indicates that 95.72% of the model residues (Fully allowed (76.56%) + additionally allowed(19.16%)) are located in the core or favoured region, 2.99% in the allowed region and 1.2% in the outside region also known as outlier or high energy region (Figure 23C).

Residues in the high energy region are considered unstable and can compromise the quality of a modeled structure (Maccallum et al., 1995). A good quality model is expected to have over 90% of the residues in the core region (Gopalakrishnan et al., 2007, Kelley and Sternberg, 2009, Porter and Rose, 2011, Zhou et al., 2011). This indicates that the Smp_059340.1 predicted model with 320 residues in the core region could be of good quality and biological informative.

The Rasmol tool was used for visualizing the modeled tertiary structure of the Smp_059340.1 protein. The secondary structure components are composed of 20 helices, 8 sheets, 35 turns and 3158 hydrogen bonds. The helices are represented in red; sheets are yellow and turns are depicted in blue. The white or grayish colored regions near the turns are other secondary structures (Figure 24).



Residue	L-Helix	Sheet	R-Helix	FAR	AAR	GAR	Outside	Total
ALA	0	4	11	15	4	2	0	21
ARG	0	1	20	21	3	2	0	26
ASN	0	5	7	12	5	0	1	18
ASP	0	2	10	12	7	1	0	20
CYS	0	1	3	4	5	0	0	9
GLN	0	3	5	8	2	0	0	10
GLU	0	3	16	19	6	1	0	26
GLY	1	2	5	8	4	1	2	15
HIS	2	2	1	5	3	0	0	8
ILE	0	5	14	19	4	0	0	23
LEU	0	10	14	24	6	0	0	30
LYS	1	4	14	19	3	1	0	23
MET	0	2	3	5	0	0	0	5
PHE	0	8	9	17	2	0	0	19
PRO	0	5	2	7	2	1	1	11
SER	0	6	12	18	1	1	0	20
THR	0	4	9	13	1	0	0	14
TRP	0	0	4	4	0	0	0	4
TYR	0	3	5	8	5	0	0	13
VAL	0	8	10	18	1	0	0	19
Total	4	78	174	256	64	10	4	334

Fully Allowed Region (256 residues) : 76.65 %
Additionally Allowed Region (64 residues) : 19.16 %
Generously Allowed Region (10 residues) : 2.99 %
*Outside Region (4 residues) : 1.20 %
100.00 %

Alpha helical Region (174 residues) : 67.97 %
Beta sheet Region (78 residues) : 30.47 %
3-10 helical Region (4 residues) : 1.56 %
100.00 %

*Non-glycyl residues (2 residues) : 0.60 %
*Glycyl residues (2 residues) : 0.60 %

Non-glycyl residues outside the AAR are :
PRO C 245 ASN C 338

L-Helix : Left handed helix
R-Helix : Right handed helix
FAR : Fully Allowed Region (L-Helix, Sheet & R-Helix)
AAR : Additionally Allowed Region
GAR : Generously Allowed Region

Figure 23: Model quality evaluation using Ramachandran plot

Image A. Ramachandran plot based on Ramachandran plot on the web (2.0) server: More of the model residues are in the fully allowed and additional allowed regions (total 320 residues (95.72%)) with very few residues in the outside region (4 residues (1.2%)). The 95.72% indicates that the model is stable and expected to be of good quality based on the phi–psi torsion angles of all the residues in the model.

Image B. Ramachandran plot based on RAMPAGE server: This confirms and validates the Ramachandran plot on the web (2.0) server. It shows more model residues are in the favored region (total 316 residues (94.6%)) with very few residues in the outlier region (5 residues (1.5%)). It also shows that PRO 245 and ASN 338 residues are the non-glycyl residues in the additional allowed region (AAR).

Image C: Complete Ramachandran plot analysis: It shows the distribution of the whole model residues into various energy regions and on the α-helix, β-strand or coil based on the phi–psi torsion angles of all the residues in the model.

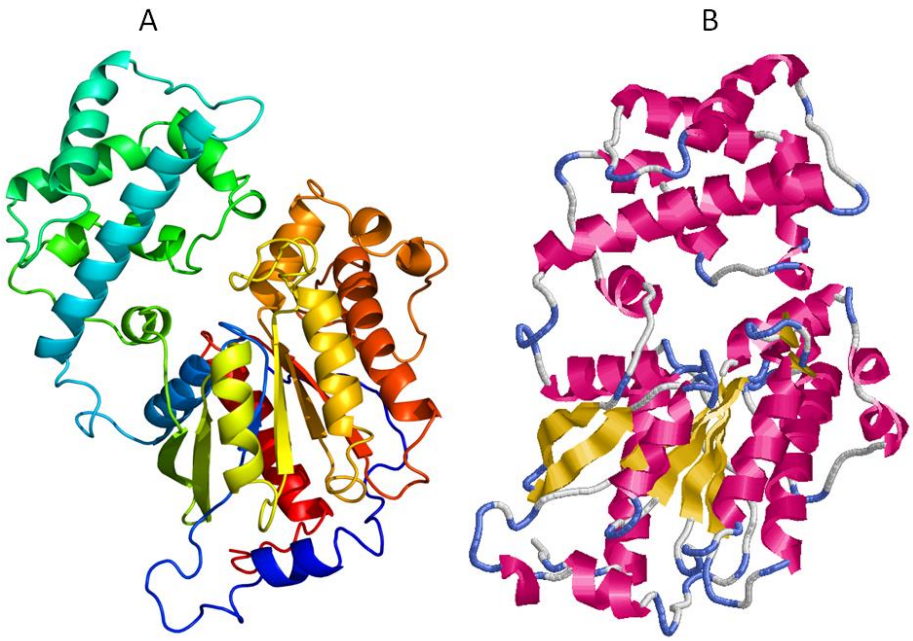


Figure 24: Homology model of *Schistosoma mansoni* protein Smp_059340.1

The image A is the model colored in rainbow N → C terminus. Image B is Rasmol tool visualization of model. The helices are represented in red; beta sheets are yellow and turns are depicted in blue. The white or grayish colored regions near the turns are other secondary structures. The secondary structure components are composed of 20 helices, 8 sheets, 35 turns and 3158 hydrogen bonds.

4.13. Mg2+ ion/GTP Binding Site Domain

The distribution of the GTP/Mg2+ ion complex binding site residues (Green) indicates that they are located on the loops, helices and the β sheets (Figure 25A). On the protein folded structure these residues are situated in the active site (fold) of the protein that extends to the outer side through a hollow cavity cutting across the protein internal structure (Figure 25B). The Mg2+ ion/GTP binding site shares four common residues with G1 box motif at Gly46, Gly51, Lys52 and Ser53 (Table 15) and also the Mg2+ ion/GTP binding site converge at two plastic regions of switch1 (G2 box motif-Thr188) and switch2 (Gly210) (Table 14). The Mg2+ ion/GTP binding site also share some conserved residues with G3 box motif at Asp207, and with G4 box motif at residue Asn276, Lys277 and Asp279 (Table 14).

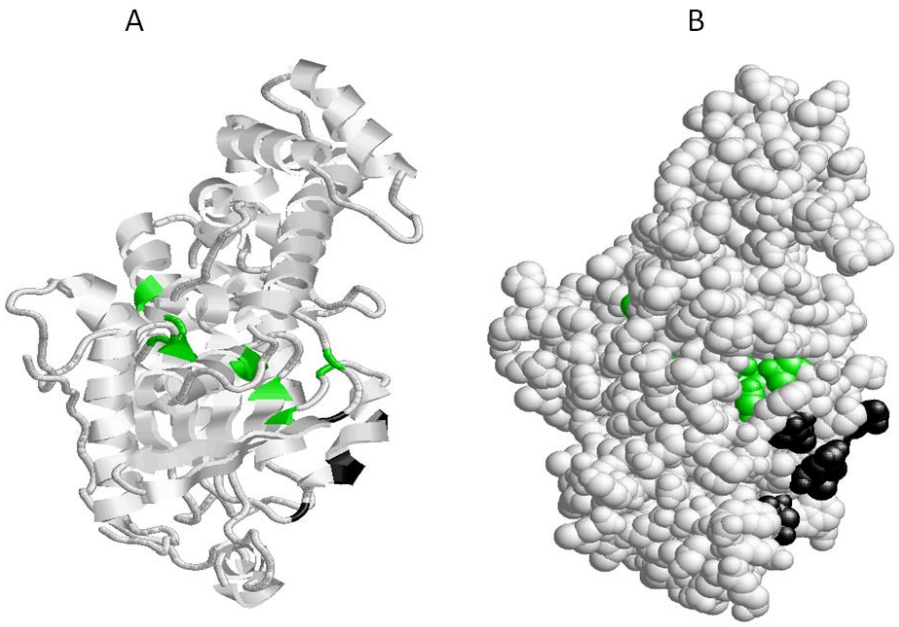


Figure 25: The distribution of GTP/Mg2+ ion complex binding site an adenylyl cyclase interaction site residues on the 3-dimensional structure of *Schistosoma mansoni* Smp_059340.1.

Image A. The GTP/Mg^{2+} ion complex binding site residues (Gly46, Gly51, Lys52, Ser53, Thr188, Asp207, Gly210, Asn276, Lys277, Asp279, Cys350, Ala351) (Green) are located on the loops, helices and the β sheets. The adenylyl cyclase interaction site residues (Ile191, Arg216, Ile219, Gln220, and Asn223) (black) are located on the loops and helices**.**

Image B. On the folded structure the GTP/Mg2+ ion complex binding site residues are situated in the active site (fold) of the protein that extends to the outer side through a hollow cavity cutting across the protein internal structure. The adenylyl cyclase interaction site residues are localized on the surface of the protein.

4.14. Adenylyl cyclase interaction site domain

The adenylyl cyclase interaction site consists of 5 residues made up of Ile191, Arg216, Ile219, Gln220, and Asn223 (black) (Figure 25A), which binds and stimulates the adenyly cyclase activities. The distribution of the adenylyl cyclase interaction site residues (Black) indicates that they are located on the loops and helices only (Figure 25A). On the protein folded structure these residues are localized on the surface of the protein (Figure 25A). This approximate surface location might be a natural structural orientation to provide efficiency in recognition and binding of adenylyl cyclase enzyme

leading to stimulation of the adenylyl cyclase pathway during the various developmental phases of *S. mansoni.* The adenylyl cyclase interaction site shared 4 out of its 5 residues (Arg216, Ile 219, Gln220 and N223) with the switch 2 region and three residues (Ile191, Gln220 and Asn223) with the beta-gamma complex binding site (Table 14). This interaction indicates that the adenylyl cyclase interaction site is totally coupled to the beta-gamma complex and it's almost completely embedded in the switch2 region. As such the adenyly cyclase binding site might be contributing to the conformation changes that occur during the functioning of switch 2 region.

4.15. Beta-gamma Complex Interaction Site and GoLoco Binding Site Domains

The beta-gamma residues (Green) are located on the helix, loop and β sheets and it shares common residues (Gly210, Trp218 and Phe222) (Blue) with GoLoco binding site on the loop and helix (Figure 26A). Within the protein folded structure, the beta –gamma complex site (Green) is positioned on the surface (Figure 26B), in close proximity to the location of GTP/Mg2+ ion complex binding site through residue Gly210. With adenylyl cyclase interaction site the beta –gamma complex site interconnect at residue Ile191, Gln220 and Asn223 while with GoLoco binding site, the beta–gamma complex site shares residue Gly210, Trp218 and Phe222 shown in blue (Figure 26B).

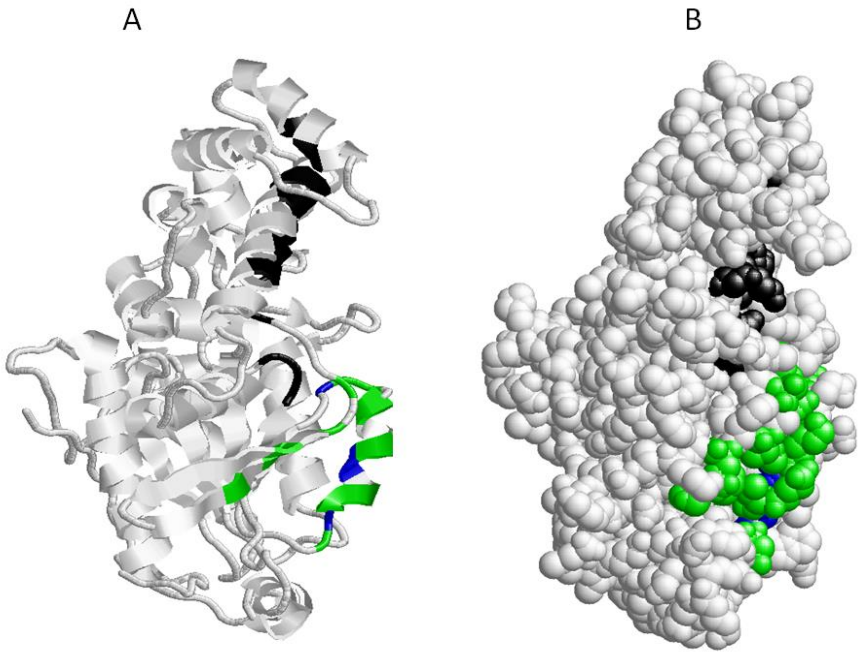


Figure 26: The distribution of beta-gamma an GoLoco binding site residues on the 3-dimensional structure of *Schistosoma mansoni* Smp_059340.1

Image A. Beta-gamma residues (Ser189, Ile191, Glu193, His204, Phe206, Gly210, Gln211, Arg212, Glu214, Lys217, Trp218, Gln220, Cys221, Phe222, Asn223) (Green) are located on the helix, loop and β sheets. The GoLoco binding site residues (Ala47, Gly48, Glu49, Asp77, Lys81, Asp85, Cys88, Ala91, Gly92, Arg112, Ile115, Ile156, Arg185, Gly210, Trp218, Phe222) (Black) are located on the helices and the loops. The common residues shared between both domains (Gly210, Trp218 and Phe222) are depicted blue.

Image B. On the folded structure the beta–gamma complex site (Green) is positioned on the surface in close proximity to the location of GTP/Mg2+ ion complex binding site at residue Gly210 and adenylyl cyclase interaction site(Ile191,Gln220 and Asn223) (Figure 5). The GoLoco binding site is situated exposed on the surface with extension into the Mg2+ ion/GTP complex site at Gly210 (Figure 25).

The GoLoco binding site (Black) has residues located on the helices and the loops (Figure 26A). It is situated in close proximity to the location of switch I, switch II and the beta-gamma binding site (Figure 26B) where it can influence their activities independently due their common shared residues as explained above. The GoLoco binding site also shares conserved residues with Mg2+ ion/GTP complex site and G3 box motif.

4.16. Putative Receptor Binding Site Domain

The putative receptor conserved residues (Red) are positioned on a loop and helix (Figure 27A). On the protein folded structure the receptor binding site (Red) stretches from within the inner active cavity were it shares two residues with the GTP/Mg2+ ion complex binding site at Cys350 and Ala351 to the surface of the protein (Figure 27B). It also shares all the 3 conserved residues (Cys350, Ala351 and Val352) found in G5 box motif. Hence the G5 box motif is part of the putative receptor binding site and function as a buttress factor for guanine base recognition site.

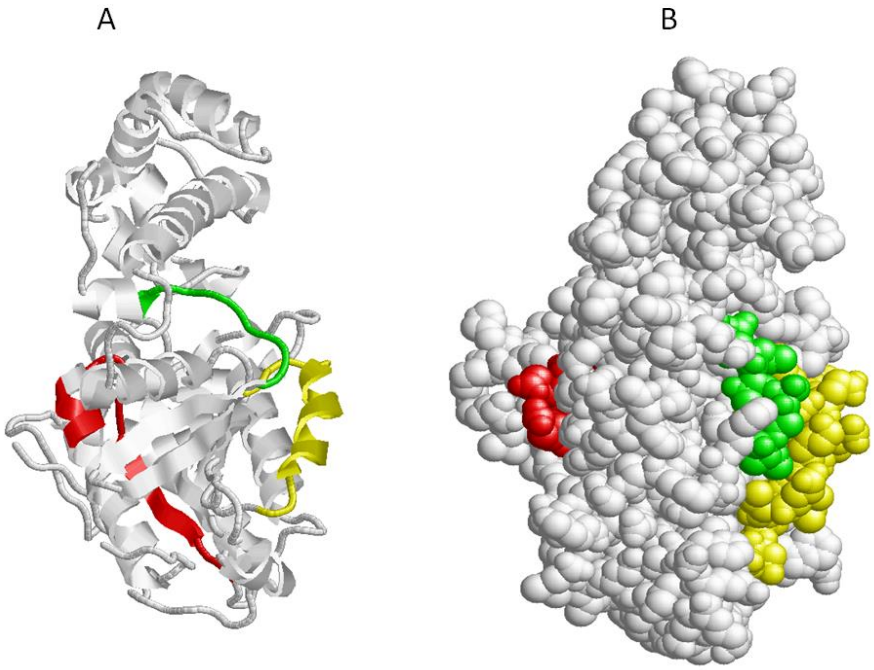


Figure 27: Distribution of putative receptor, switch1 and switch2 residues in *Schistosoma mansoni* Smp_059340.1

Image A. The putative receptor conserved residues (Red) are positioned on the loop and helix; switch1 residues (Green) are located on the loop and helix while switch2 residues (Yellow) are also situated on the loop and helix.

Image B: On the folded structure the putative receptor conserved residues (Red) stretches from within the inner active cavity were it shares two residues with the GTP/Mg2+ ion complex binding site at Cys350 and Ala351 to the surface of the protein. Both switch1 (Green) and switch2 (Yellow) regions are orientated on the surface close to each other. Switch1 is shorter than the switch2 region by 8 residues. These regions shares conserved residues with GTP/Mg2+ ion complex, adenylyl cyclase interaction site and beta–gamma complex, GoLoco binding site as shown above.

4.17. Switch I and Switch II Region Domains

The switch I residues (Green) are located on the loop and helix, while switch II residues (Yellow) are also situated on the loop and helix and both molecular switches cycle between active state (GTP bound) and inactive state (GDP bound) (Figure 27A).

The analysis shows that both switches are located on the outer surface of the protein with switch I shorter than the switch II region by 8 residues (Table 14, Figure 27B). These regions share conserved residues with GTP/Mg2+ ion complex, adenylyl cyclase interaction site, beta–gamma complex, GoLoco binding site and the G2 box motif (Table 14) as explained above. On the folded protein structure the switch1 region (Green) is orientated on the surface, very close to the switch2 region (Yellow) (Figure 27B). The switch2 is another region responsible for the conformation changes between the GTP and the GDP as stated above. This switch2 region shares conserved residues with GTP/Mg2+ ion complex, adenylyl cyclase interaction site, beta–gamma complex, GoLoco binding site and the G3 box motif (Table 14). It was deduced that both the adenyly cyclase binding site (4 out of 5 residues) (Arg216, Ile219, Gln220 and Asn223) and the G3 box motif (3 out of 4 residues) (Val208, Gly209, Gly210) are completely integrated into the switch2 region in terms of their shared residues.

4.18. G1-G5 box Motif Domains

The G protein superfamily folds in different variations depending on its nucleotide binding folds (Brandeen, 1980). There are five polypeptide loops that form the guanine nucleotide-binding site and all have well conserved

elements in their domain. They are used for defining the G protein superfamily. These loops are labeled G1 box-G5 box (Jurnak et al., 1980). All the five loops were present in this protein structural view (Figure 28A). The G1 box motif (GxxxxGK(S/K) (Red) has been mentioned above in connection with Mg2+ ion/GTP binding site and it is located on the loop, sheet and the helix (Figure 28A). It is used for defining GTP binding proteins in general (Jurnak et al., 1980). The G2 box motif (Green) is located on the loop and contains only one residue Thr188. Only the Thr188 is conserved throughout the superfamily, but surrounding residues are conserved within the families. The G2 box motif is part of GTP/Mg2+ ion complex and switch1 region.

The G3 box motif (Black) is also known as the DxxG consensus sequence (Colicelli, 2010, Field and Kellogg, 1999, Pandit and Srinivasan, 2003) and situated on the loop and sheet which function in linking the sub-sites for binding Mg2+ ion with the γ phosphate of the of GTP. It also shares some common conserved residues with the Mg2+ ion/GTP complex, beta-gamma complex, GoLoco binding site and the switch2 region (Table 14) as mentioned above. The G3 box motif is also known as the Walker B motif (Walker et al., 1982) and forms an integral part of the switch2 region. The G4 box motif (NKXD) (Yellow) is localized on the loop, helix and sheet. The main function of the G4 box motif is in recognizing the guanine ring.

The G5 box motif ([C/S]A[K/L/T])(Blue) functions as a booster to guanine base recognition site. It shares one conserved residue (Ala351) with the Mg2+ ion/GTP binding site and all its 3 residues with the putative receptor binding site. Hence it is part of the putative receptor binding site (Table 14).

The protein folded structure indicates that all the five loops (G1 box-G5 box) are located in the active fold (Figure 28B). The G2 box motif (Green) is embedded in the active cavity of the protein very close to G3 box motif (Black) with G4 box motif (Yellow) situated very close to G5 box motif (Blue).

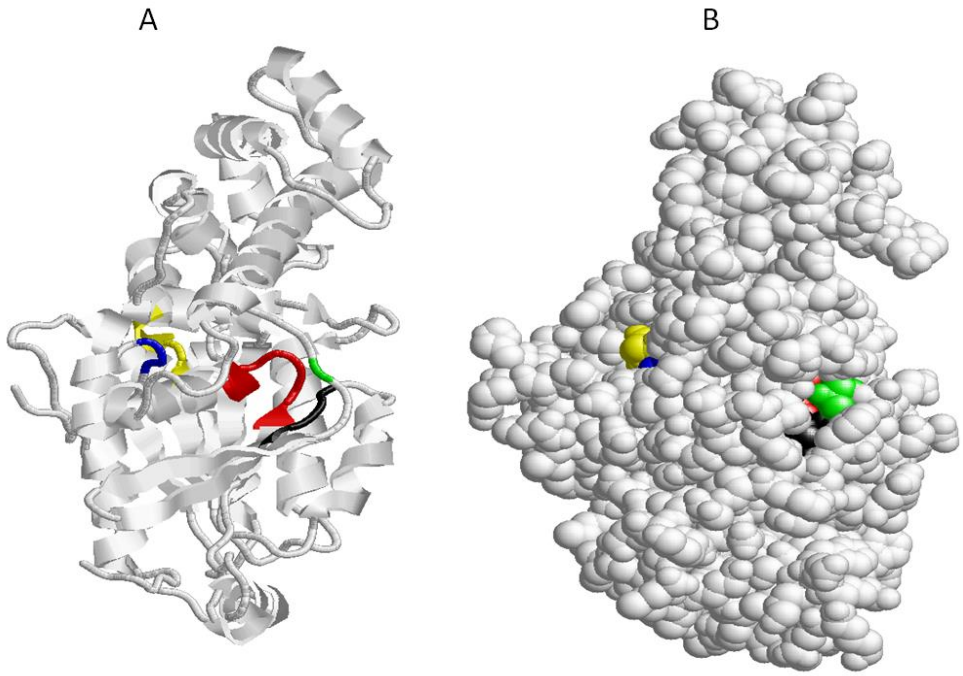


Figure 28: Distribution of G box motifs labeled G1-G5 residues of *Schistosoma mansoni* Smp_059340.1.

Image A: The G1 box motif (Red) is located on the loop, sheet and the helix while the G2 box motif (Green) is located on the loop and contains only one residue Thr188. The G3 box motif (Black) is situated on the loop and sheet; G4 box motif (Yellow) on the loop, helix and sheet and G5 box motif (Blue) is located on the loop.

Image B: On the folded structure all five G motifs are located in the active site were of GTP binding. The G1 box motif share residue with Mg2+ ion/GTP binding site, G2 box motif (Green) share with Mg2+ ion/GTP complex and the switch1, G3 box motif (Black) shares with Mg2+ ion/GTP complex, beta-gamma complex, GoLoco binding site and the switch2 region. The G5 box motif shares residue (Ala351) with the Mg2+ ion/GTP binding site and all its 3 residues with the putative receptor binding site. The G2 box motif (Green) is embedded in the active cavity of the protein very close to G3 (Black) with G4 box motif (Yellow) situated very close to G5 box motif (Blue).

4.19. Ligand Binding Site

The predicted ligand binding analysis shows that at the GTP/Mg2+ ion complex site, the Mg2+ ion binds to Ser53 and the metallic ligand is well embedded in the active cavity (Figure 29). This ligand binding indicates that

irrespective of the residues coordinating the GTP binding at GTP/Mg2+ ion complex site, Ser53 could be considered crucial in regulating the adenyly cyclase pathway in *S. mansoni* developmental cycles.

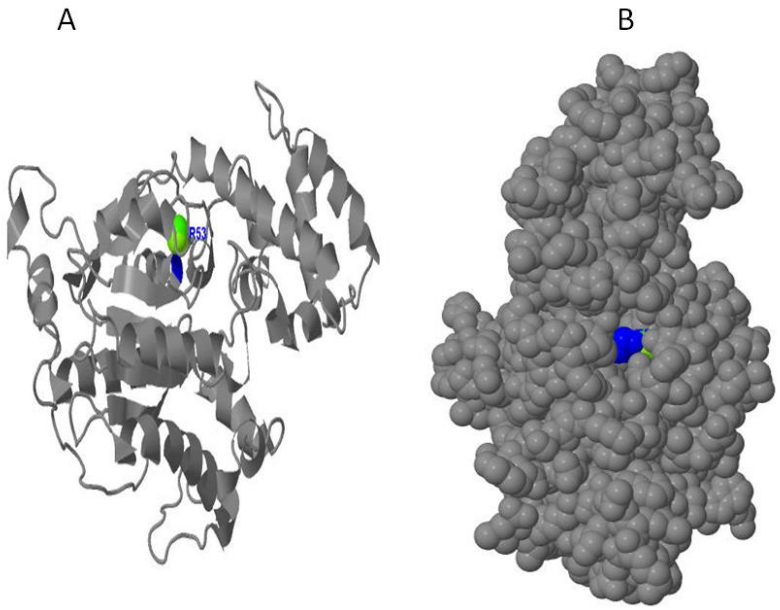


Figure 29: Binding of Mg2+ ion on *Schistosoma mansoni* Smp_059340.1 GTP/Mg2+ ion complex site

Image A: The binding of Mg2+ ions on the secondary structure with only Ser53 (blue) located on the helix is involved in binding.

Image B: The Mg2+ ion is embedded deeper in the cavity binding with Ser53 and regulated by the G box motifs (Figure11) and other domains sharing residues with GTP/Mg2+ ion complex site.

CHAPTER FIVE

DISCUSSION

Objective 1

5.1. Infer the Biochemical and Environmental Regulation of the *Schistosoma* Universal Stress Proteins

Schistosoma haematobium, *Schistosoma japonicum* and *Schistosoma 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 cercarial penetration of the human skin (Jolly et al., 2007). The presence of genes encoding the universal stress protein (USP) domain in the genomes of *Schistosoma* species suggests these flatworms are equipped to respond to unfavourable conditions that induce USP function (Isokpehi et al., 2011a, Loverde, 1998, Negrao-Correa et al., 2012, Seifart Gomes et al., 2011).

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. The research used a visual analytics approach to integrate these data types and to identify orthologous pair 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 (Zerlotini et al., 2009) and evolutionary genomics analysis called the

S. mansoni phylome (Silva et al., 2012) indicate that an ortholog (Sha_107834) is encoded in the *Schistosoma 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 of interest, the research focus the discussion of the results on the findings on the 5 conserved residues and the chemical ligands predicted to bind to the prioritized protein. All the 13 protein sequences have conserved sites for aspartate (Asp; D), leucine (Leu; L), glycine (Gly; G), histidine (His; H), and proline (Pro; P) at positions 57, 101, 127, 166 and 176 using Smp_076400 from *S. mansoni* as a reference sequence (Figure 8). These conserved residues do not coincide with any of the predicted ligand binding sites. These residues could be common functional sites for regulating the Schistosoma universal stress proteins.

The predicted 3D chemical ligands for Smp_076400 and Sjp_0058490 (Q86DW2) included three metal ions Ca^{2+}, Mg^{2+} and Zn^{2+} (Figure 11). Metal ions are involved in a many diverse biochemical reactions (Lu et al., 2012) 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 (Freestone et al., 1997). The ATP molecule and metallic chemical ligands such as Mg ion might bind together at the Mg-ATP binding groove during phosphorylation or ATP dependent stress response mechanism (Buchachenko et al., 2005, Buchachenko et al., 2008).

The presence of Mg ion suggests that it can be an integral and critical component in the reaction (Luo et al., 2012, Nebl et al., 2011, Szabo et al., 1994, Langer et al., 1992). This result could be affected if there is any structural conformation in the binding site residues which affects the Mg2+ ion from binding to the ATP molecule at the active groove. The resultant effect might be translated in compromising the 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.

Calcium ion (Ca^{2+}) was predicted to bind to proteins in the Group C of the phylogenetic tree (Figure 9 and Figure 12). In *S. mansoni*, Ca^{2+} is considered vital for regulated motor related activities (Soares de Moura et al., 1987) and also critical for the egg hatching process in fresh water (Katsumata et al., 1989, Katsumata et al., 1988) . In the tegument fraction of *S. mansoni*, Ca^{2+} simulated the activity of ATPase in the absence of Mg^{2+} (Cunha and Noel, 1988). Further, cAMP and Ca^{2+} work in synergism to regulate the transformation of miracidial to sporocysts (Kawamoto et al., 1989). The protein kinase C (PKC) and Ca^{2+} metabolism regulates the induction of proteolytic enzyme from cercariae, which is vital for modulating the musculature activity of the schistosome (Blair et al., 1988, Matsumura et al., 1991). A key mechanism of action of praziquantel (PQZ) has been proposed to be disruption of calcium homeostatis in schistosomes leading to large rapid influx of calcium ions into the worm and quick muscular contractions (Coles, 1979, Doenhoff et al., 2008, Greenberg, 2005, Pica-Mattoccia et al., 2008). Microarray-based transcriptome analysis of the response of *S. mansoni* PR-1 to praziquantel has identified genes for cytostolic calcium regulation (Aragon et al., 2009).

Objective 2

5.2. Identify biological function relevant protein sequence and structure features for prioritized universal stress proteins from *Schistosoma* species.

5.2.1. Physicochemical Characterization of *Schistosoma* USPs

Primary structure analysis shows that the schistosome USPs contain more hydrophilic amino acids and can be describe as moderately hydrophilic. The hydrophobic residues are located in the core of most proteins and they help in stabilizing the proteins through *van der Waal* interactions (Berezovsky and Trifonov, 2001). The hydrophilic residues are located mostly at the surface and active sites of proteins, where they interact with other polar residues or with water molecule. This contributes to the reactivity of the schistosome USPs. This could be regarded as an adaptation to its life cycle stages in aqueous environment.

The isoelectric point (pI) indicates that the schistosome USPs are more basic in nature as most of their pI values were above 7.The Isoelectric point (pI) indicates the pH at which the protein surface is covered with charge (Chen et al., 1993). In *S. japonicum,* the charge distributions were as follows: (1) positive charge: (Q86DX1, Q5DI36, Q86DW2 andQ5DH64); (2) negative charge: (Q5DHK1, Q5DDH7, Q5DED2) and (3) neutral: (Q5DGI9). For *S. mansoni* charges were distributed as follows (1) positive charge: (Smp_076400 and Smp_097930, ; (2) negative charge: (Smp_031300) and (3) neutral charge: (Smp_001000 and Smp_043120). At a given pI proteins are stable and compact, thus this parameter will be useful for developing buffer

systems for purifying schistosome universal stress proteins using isoelectric focusing techniques (Sillero and Maldonado, 2006).

The computed instability indices of the USPs were conspicuously above 40, however USPs from *S. japonicum* were predicted to be highly unstable. This could be because USPs are not housekeeping genes. They are transcript to circumvent stressor. All the genes showed conspicuous higher values of half-life. These three parameters (instability indices, extinction coefficient and half-life values) account for the stability of most proteins. These values infer that the USP genes are relatively unstable in their native state. Their relative high aliphatic index values and conspicuously very low grand average hydropathy (GRAVY) indicate that schistomome USP genes are very reactive in the water environment or aqueous phase (Mbah et al., 2012), which seems consistent with their immediate environment and developmental stages.

5.2.2. Structural and Functional Relationship of Prioritized *Schistosoma* USPs

The Smp_076400 USP was prioritized to relate its structure to functional regulation. The predicted 3D chemical ligands for Smp_076400 and Sjp_0058490 (Q86DW2) included three metal ions Ca^{2+}, Mg^{2+} and Zn^{2+}. Metal ions are involved in a many diverse biochemical reactions (Lu et al., 2012) including cellular cofactors for phosphorylation. The ATP binding motif might be essential for phosphorylating the USP gene during posttranslational modification reaction and during ATP dependent stress response mechanism. The UspA protein of *Escherichia coli* undergoes phosphorylation in vitro with its phosphate donors ATP and/or GTP in the absence of other proteins

(Freestone et al., 1997). This indicates that ATP binding might be a regulator controlling the diverse specialized functions of schistosome stress response gene family (Nachin et al., 2005, Zarembinski et al., 1998) . The ADP and AMP are product molecules during USP phosphorylation. The ATP molecule and metallic chemical ligands such as Mg ion might bind together at the Mg-ATP binding groove during phosphorylation or ATP dependent stress response mechanism (Buchachenko et al., 2005, Buchachenko et al., 2008).The presence of Mg^{2+} and other metallic ions as chemical ligands indicate that they are important in the phosphorylation of USP genes as evident in other proteins (Buchachenko et al., 2005, Buchachenko et al., 2008, Kuznetsov et al., 2006).

This result could be affected if there is any structural conformation in the binding site residues which affects the Mg^{2+} ion from binding to the ATP molecule at the active groove. The resultant effect could be translated in compromising the 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. The exposed- orientation of ATP binding residues on the protein surface makes them accessible to the incoming free ATP molecules. These conserved binding sites could be the key residues regulating the molecular mechanism of stress response in the schistosome genome.

The ligand binding site has the conserved Gly residue, which was also found to be common among both prioritized USP genes. Gly residues have been shown to contribute to the formation of hydrogen bonds with the gamma-phosphate of nucleotide triphosphates (Milburn et al., 1990). The adaptive

structural conformation indicates functional efficiency in binding ATP during phosphorylation and stress response mechanism. It also helps to keep the metallic ions very stable in the active grove while in contact with ATP molecules during phosphorylation, by utilizing a network of hydrogen bonds. This finding suggests that phosphorylation in the presence of $Mg^{2+,}$ Ca^{2+} and Zn^{2+} ions can be regarded as the critical rate determining step in the functional mechanism of schistosome USP genes.

Calcium ion (Ca^{2+}) was predicted to bind to both Smp_076400 and Sjp_0058490 (Q86DW2). In *S. mansoni*, Ca^{2+} is considered vital for regulated motor related activities (Soares de Moura et al., 1987) and also critical for the egg hatching process in fresh water (Katsumata et al., 1989, Katsumata et al., 1988). In the tegument fraction of *S. mansoni*, Ca^{2+} simulated the activity of ATPase in the absence of $Mg2^{+}$ (Cunha and Noel, 1988). Further, cAMP and Ca^{2+} work in synergism to regulate the transformation of miracidial to sporocysts (Kawamoto et al., 1989). The protein kinase C (PKC) and Ca^{2+} metabolism regulates the induction of proteolytic enzyme from cercariae, which is vital for modulating the musculature activity of the schistosome (Blair et al., 1988, Matsumura et al., 1991).

A key mechanism of action of praziquantel (PQZ) has been proposed to be disruption of calcium homeostatis in schistosomes leading to large rapid influx of calcium ions into the worm and quick muscular contractions (Blair et al., 1988, Doenhoff et al., 2008, Greenberg, 2005, Pica-Mattoccia et al., 2008). Microarray-based transcriptome analysis of the response of *S. mansoni* PR-1 to praziquantel has identified genes for cytostolic calcium regulation (Aragon et

al., 2009) . This mechanism might be the hallmark in the developmental and functional regulation of the USP from schistosomes and could be considered applicable to other organisms possessing universal stress response proteins.

Objective 3

5.3. Determine the distinctive structural protein features of a predicted regulator of *Schistosoma* adenylate cyclase activity that has possible influence on the functioning of universal stress proteins.

5.2.1. Physicochemical Characterization of Smp_059340

The Smp_059340.1 protein identified to regulate adenyly cylase pathway in *Schistosona mansoni* development can be described as moderately hydrophilic and basic in nature (Table 14). The hydrophobic residues are usually found in the core of most proteins and they help in stabilizing the proteins through the numerous *van der Waal* interactions (Berezovsky and Trifonov, 2001) . The hydrophilic residues are located mostly at the surface active sites of the proteins, where they interact with other polar residues in the protein or with water molecule. The Isoelectric point (pI) indicates the pH at which the protein surface is covered with charge (Sillero and Maldonado, 2006), and the net charge of Smp_059340.1 is positive. The high number of positive charged residues (Arg + Lys = 58) against the total number of negatively charged residues (Asp + Glu = 53) is the main contributing factor to the positive charge.

At a given pI, proteins are stable and compact, thus this parameter will be useful for developing buffer systems for purification of this protein by isoelectric focusing techniques (Sillero and Maldonado, 2006). The high

extinction coefficient of 43485 M^{-1} cm^{-1} at 280nm wavelength computed for Smp_059340.1 was due to the individual contributions of Cys (2.9%), Trp (1.1%) and Tyr (3.7%) concentrations respectively. This observation suggests that Smp_059340.1 protein can be analyzed using UV spectrum assay protocol (Stoscheck, 1990). The computed protein concentration and the extinction coefficient can be important in the quantitative analysis of the protein-protein and protein-ligand interaction of this protein in solutions (Gill and von Hippel, 1989, Stoscheck, 1990)

The estimated half-life of this protein with Met as the N-terminal of the sequence was 30 (>20) hours. The high concentrations of Ala (6.1%), Leu (9.2%) and Val (5.3 %) may be contributing to the stability of this protein (Bachmair et al., 1986, Gonda et al., 1989). The half-life of these 3 mentioned residues had been well documented in the mammalian with values of Ala (4.4 hours), Leu (5.5 hours) and Val (100 hour). In other organisms, these residues are also contributing to the protein stability. In yeast, the half-life were the same (>20 hours) for both Ala and Val, also both were the same in *Escherichia coli* with values >10 (Varshavsky, 1997). The instability index prediction using the ProtParam tool indicates that Smp_059340.1 may be unstable with a value of 48.15. This parameter was computed from the impact of dipeptides in the protein sequence (Guruprasad et al., 1990), however Smp_059340.1 is shown to have high aliphatic index, half-life and the large amount of hydrogen atoms (3093) to form hydrogen bonds. Such hydrogen bonds are known to impact significant stability to protein making them resistance to degradation (Chen et al., 1993). Therefore it appears that the formation of hydrogen bonds in Smp_059340.1 may override the impact of dipeptides. The calculated value is

a measure of protein stability in a test tube (Guruprasad et al., 1990). A protein of instability index < 40 is considered as stable, while those with values > 40 are unstable (Guruprasad et al., 1990).

The aliphatic index (AI) of a protein is the relative volume occupied by the aliphatic side chains (alanine, valine, isoleucine and leucine) and is taken as contributor to the increase thermal stability of globular proteins. The aliphatic index computed for Smp_059340.1 was 82.17 using a formulated rule (Ikai, 1980). This high aliphatic index indicates that Smp_059340.1 can be stable within a wide range of temperature. Proteins with low thermal stability turn to be more structurally flexible.

The Grand Average of Hydropathy (GRAVY) is the computed sum of hydropathy values of all the amino acids, divided by the number of residues in the sequence (Kyte and Doolittle, 1982) . The very low GRAVY Index (-0.458) of Smp_059340.1 indicates that there is good interaction between this protein and water. The Hopp-Woods scale identified three regions on this polypeptide predicted to be highly hydrophilic. The exposed hydrophilic regions are exposed on the surface and may possibly represent antigenic sites (Hopp and Woods, 1981), indicating that this protein can serve as a possible drug target for schistosomiasis.

5.3.2. Interactions of Domains Specific Amino Acids at the Smp_059340.1 Protein Active Sites

The presence of conserved domains active sites within the guanine nucleotide-binding proteins (G proteins) is responsible for many important biological functions within the cellular process of an organism (Pandit and

Srinivasan, 2003, Colicelli, 2004) . These functions range from signal transduction, protein synthesis, cell proliferation, protein targeting, membrane trafficking and secretion including cell skeletal organization and movement (Yamanaka et al., 1986, Gilman, 1987, Birnbaumer, 1990). The sharing of functional unit residues among the domains suggests that the Smp_059340.1 functions as a highly interconnected network of functional domains as will be explained below.

5.3.3. Mg2+ ion/GTP Binding Site Interaction

The distribution of the GTP/Mg2+ ion complex binding site residues and their orientation to the active fold may be seen as native adaptation for efficient binding of the GTP molecule. The Mg2+ ion interacts with alpha subunits of the G proteins in the presence of GTP to form a complex from which nucleotides dissociate slowly at multiple sites and higher concentrations of Mg2+ ion had been documented to promote the dissociation of GDP (Higashijima et al., 1987).

The dissociation of nucleotides at multiple sites could be accounted by the stretch of the active fold observed in Figure 25A. The GTP is bounded as a complex with Mg2+ ion through the coordination of one oxygen atom from the γ-phosphate and these Gαs proteins are known to be unstable in the absence of a bounded nucleotide (Ferguson et al., 1986).

The binding sites for both Mg2+ ion and GTP are strongly linked together. It was found that Mg2+ ion/GTP binding site shares four common residues with G1 box motif at residue Gly46, Gly51, Lys52 and Ser53 (Table 14). The G1 box motif is one of the five polypeptide loops that form the

guanine nucleotide-binding site with well conserved elements in the domain. It is also called the Walker A motif and had been used for defining the G protein superfamily and their ability to bind guanine nucleotide such as GTP (Jurnak et al., 1980). The shared residues between Mg2+ ion/GTP binding site and G1 box motif could be accounted for the binding of GTP at the Mg2+ ion/GTP binding site domain.

Mg2+ ion/GTP binding site converged at two plastic regions of switch I (G2 box motif-Thr188) and switch II (Gly210) (Table 14). The coupling between GTP- γ-phosphate site and the Mg2+ ion may be accounted for by the rigidity of the switch I/switch II interface (Hepler et al., 1993). The key residue could be Thr188 which binds the magnesium ions of the Mg–GTP complex, making it to sense the present of the γ- phosphate found in GTP molecules. The Mg2+ ion/GTP binding site also shares some conserved residues with G3 box motif at Asp207 (Table 14). The conserved Asp207 residue in G3 box motif also coordinates the Mg2+ ion through a water molecule. This coordination could also be crucial for the tight linkage of both the Mg2+ ion and the GTP binding sites. In the GTP-Mg2+ ion complex, the ligand bonds are contributed by both the β and γ phosphates of the GTP molecule and also by Thr188 and the water molecules.

The Asp207 in the G3 box motif binds to one of the water molecules in the process and also there is possible hydrogen bond formation with the α-phosphate oxygen atom from the GTP molecule. The Mg2+ ion/GTP binding site also shares some common residues with G4 box motif (Asn276, Lys277 and Asp279) (Table 14). The G4 box motif functions in the recognition of the

guanine ring suggesting that, the Mg-GTP complex binding site can be regarded as the initiating point for the adenyly cyclase metabolic pathway. The methylene group of the common Lys277 residue of both GTP-Mg2+ ion complex binding site and the G4 motif might be providing the hydrophobic surface that lies over the purine ring (Gille et al., 2002) during the binding process.

5.3.4. Beta-gamma Complex and GoLoco Binding Site Interaction Site

On the protein folded structure, the beta –gamma complex site is positioned on the surface in close proximity to the location of GTP/Mg2+ ion complex binding site, adenylyl cyclase interaction site and GoLoco binding site. This closeness helps both the beta-gamma complex and the GoLoco site in influencing the activity of GTP/Mg2+ ion complex binding site, adenylyl cyclase interaction site and switch II region respectively. However, the beta-gamma complex can stimulate and inhibit adenylyl cyclase activity (Weitmann et al., 2001). This inhibition could be through the common shared residues (Ile191, Gln220 and Asn223) with the beta-gamma complex, although no model for this function had been established. The beta-gamma complex also shares Gly210 with G3 box motif.

There is possible hydrogen bond formation between the γ-phosphate of GTP and the main chain amide of the conserved Gly210 in the G3 box. The stimulation or inhibition of the adenylyl cyclase activity may be linked to Gly210 residue, considering its function during the binding of the GTP/Mg^{2+} ion complex. The beta-gamma complex also shares two residues (Ser189 and Ile189) with switch I and 10 residues (Gly210, Gln211, Arg212, Glu214,

Lys217, Trp218, Gln220, Cys221, Phe222 and Asn223) with switch II region (Table 14). The residue sharing suggests that the beta-gamma complex could be influencing the conformation changes exhibited by both switches, which could affect the adenylyl cyclase activity. The GoLoco binding site is known to act as a guanine nucleotide dissociation inhibitor (GDI) (Willard et al., 2004) and this could account on how it could be inhibiting the adenylyl cyclase activity.

The beta-gamma complex interaction site shares conserved residues with many other domains (Table 14). It functions had been extensively covered in part in the sections above on how it might influence the activity of the other domains together with the GoLoco motif binding site. The binding of extracellular hormone (5HT) on the *S. mansoni* serotonin receptor (Smp_126730) initiates the ejection of GDP from the G alpha subunit and initiates the binding of GTP to the G-alpha subunit (Weitmann et al., 1999). This binding causes the disassociation of G-alpha subunit from the G-alpha/beta-gamma complex. Within the protein folded structure the beta–gamma complex site (Green) is positioned on the surface suggesting that, the beta-gamma complex would also be in close proximity to the location of switch I and switch II where it might influence their functions. This closeness could help both the beta-gamma complex and the GoLoco sites in influencing the activity of both switches through their shared residues.

The GoLoco binding site (Black) has residues located on the helices and the loops (Figure 26A). It is situated in close proximity to the location of switch I, switch II and the beta-gamma binding site (Figure 26B) where it could

influence their activities independently due their common shared residues as explained above. The GoLoco binding site also shares conserved residues with Mg2+ ion/GTP complex site and G3 box motif. The GoLoco binding site is also known as G protein regulatory (GPR) motif (Takesono et al., 1999) and it is known to prevent the spontaneous release of GDP by Gαs, thus acting as a guanine nucleotide dissociation inhibitor (GDI) (Willard et al., 2004). This research has discussed in the previous sections how this domain influences the activities of other domains with shared residues and it will be looking at its impact with the rest of the uncovered individual domains.

5.3.5. Putative Receptor, Switch 1 and Switch 2 Interaction Sites

On the protein folded structure the receptor binding site stretches from within the inner active cavity were it shares two residues with the GTP/Mg2+ ion complex binding site at Cys350 and Ala351 to the surface of the protein (Figure 27B). It also shares all the 3 conserved residues (Cys350, Ala351 and Val352) found in G5 box motif. Hence the G5 box motif is part of the putative receptor binding site and function as a buttress factor for guanine base recognition site. Ala351 could be contributing to this buttress factor function due to its presence also at the GTP/Mg2+ ion complex binding site.

The structural conformity of the putative receptor might be contributing to the binding stability of the Gαs (Smp_059340.1) protein, the GTP molecule and to the serotonin receptor (Smp_126730). Considering the function, conserved shared residues and the location of putative receptor binding site, this domain can be very crucial in the continuous flow of signal in the adenylyl cyclase pathway.

The switch1 and switch2 are both molecular switches cycling between active state (GTP bound) and inactive state (GDP bound) and switch I is shorter than switch II by 8 residues. These regions shares conserved residues with GTP/Mg^{2+} ion complex, adenylyl cyclase interaction site, beta–gamma complex, GoLoco binding site and the G2 box motif as explained above. In switch I the transitional conformation changes mainly involve flips of some peptide units and reorientation of some side chains (Milburn et al., 1990, Schlichting et al., 1990).

The key residue could be the Thr188 that coordinates the magnesium ions of the Mg–GTP complex which can thus sense the present of the γ phosphate (Kraulis et al., 1994). Therefore the conformation transition in switch I region is manifested in reorientation by the conserved residue (Thr188) in both Mg^{2+} ion ligand and the G2 box motif (Thr188), which is also an integral residue of switch I region. The switch1 region has also been identified as the GAP (GTP activation protein) binding region (Cales et al., 1988, Adari et al., 1988). This region may be responsible for the tight coupling of the Mg^{2+} ion /GTP complex through their shared Thr188 residue (Hepler et al., 1993).

As explained in the previous paragraphs the function of switch I region during conformation changes between GTP and GDP can be compromised by both the beta-gamma complex and the GoLoco binding sites through their common shared conserved residues (Table 14). On the folded protein structure the switch1 region (Green) is orientated on the surface, very close to the switch II region (Yellow) (Figure 27). As mentioned above, the beta-gamma

complex or the GoLoco binding sites share common conserved residues with both switch I and switch II regions. Therefore this close proximity of switch I and switch II region suggests that the conformational changes in both switches might be coupled and also the functionality of both switches might be influenced simultaneously by either the beta-gamma complex or the GoLoco binding sites.

The switch2 shares conserved residues with GTP/Mg2+ ion complex, adenylyl cyclase interaction site, beta-gamma complex, GoLoco binding site and the G3 box motif. It was deduced that both the adenylyl cylase binding site (4 out of 5 residues) and the G3 box motif (3 out of 4) residues are complete integrated into the switch 2 region in terms of the their shared residues. There is considerable conformational flexibility in this region, especially in the GDP bound state (Schlichting et al., 1990).

As mentioned above the adenyly cyclase binding site may be contributing to this conformation changes. The same plausible reason as explained above for switch1 region, the conformational changes between GTP and GDP can be greatly hampered by both the beta-gamma complex and the GoLoco binding sites through the common shared conserved residues.

The switch2 influences the tight coupling of the Mg^{2+} ion/GTP complex through the shared Gly210 with G3 motif (DxxxD) near the N – terminal end of the switch2 helix (Greasley et al., 1995) . In this superfamily, the binding energy of GTP is normally used in stabilizing the switch regions, so that a conformation is produce, which favour its association with an effector. The switch2 region is relatively mobile at the effector binding region, particularly in

the GTP bound form. The Gly220 in the switch 2 and DxxD (G3 box motif) (Table 14) might be the determining point for the reorientation and partial refolding of the helical region of switch 2 (Milburn et al., 1990) due to its contribution in forming hydrogen bond with the gamma-phosphate as mentioned above (Milburn et al., 1990).

However, there is a slight difference in structural position between switch1 and switch 2 regions (Figure 27A). Both are located on the protein surface, but the switch 2 (Yellow) residues are oriented outwards, while the switch 1 (Green) form part of the active site (fold) of the protein, and it is well localized at the gateway of the active site of the protein. The Thr188 shared between the Mg2+ ion/GTP complex and switch1 region may account for difference in active site location. The Mg2+ ion/GTP binding site residues are located in the active cavity of the protein (Figure 25B).

5.3.6. G1-G5 box Motif Interactions Sites

The G1 box motif is use for defining GTP binding proteins in general (Jurnak et al., 1980). Only the G2 box residue Thr188 is conserved throughout the superfamily, but surrounding residues are conserved within the families. As mentioned above its main function is to coordinates the magnesium ions of the Mg – GTP complex, which can thus sense the present of the γ- phosphate. The function of the Thr188 had been elaborated extensively in both the Mg^{2+} ion/GTP complex and the switch1 sections above. The conserved Asp207 residue in G3 box motif coordinates the Mg^{2+} ion through a water molecule and the conserved Gly220 in the switch 2 and G3 box motif determines the reorientation and partial refolding of the helical region of switch 2 (Milburn et

al., 1990) by forming hydrogen bond with the γ-phosphate (Milburn et al., 1990). The G3 box motif is also known to form an integral part of the switch 2 region. All the five loops (G1-G5) are located in the active fold with G4 box motif situated very close to G5 box motif. The plausible reason may be because the G5 box motif buttress the guanine base recognition site. The residue Ala351 may be the booster component in guanine site recognition. The location of the five loops in the cavity indicates their importance in the binding of GTP molecules (Figure 28B).

CHAPTER SIX

CONCLUSION AND RECOMMENDATIONS

6.1: CONCLUSION

Objective 1

6.1.1: Infer the Biochemical and Environmental Regulatory Features of *Schistosoma* Universal Stress Proteins

Schistosoma haematobium, *Schistosoma mansoni* and *Schistosoma 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 universal stress proteins in *Schistosoma* species, knowledge of their biochemical and environmental regulation is still limited. The draft status of the genome sequences of *Schistosoma* species also provides possibilities that future revisions could be made to gene prediction and protein annotations. This research has used a decision making strategy facilitated by visual analytics to identify universal stress proteins in two *Schistosoma* species with shared sequence features that are in comparison with other sequences relatively complete and consistent annotation. These findings enable further inferences on the biochemical and environmental regulation of *Schistosoma* USPs. Future research directions could (i) determine the transcriptional response of *Schistosoma* genes for universal stress proteins to praziquantel and (ii) functional characterize the interactions of calcium ions with Schistosoma universal stress proteins. The datasets produced and the visual analytics views developed can be easily reused to develop new hypotheses.

Objective 2

6.1.2. Identify biological function relevant protein sequence and structure features for prioritized universal stress proteins from *Schistosoma* species.

Primary structure analysis shows that the schistosome USPs contain more serine followed by isoleucine, valine, leucine and lysine in that order, while there were very few tryptophan followed by cysteine, glutamine, phenylalanine, methionine and tyrosine than the rest of the other amino acid residues. Schistosome USPs contain more hydrophilic amino acids and can be describe as moderately hydrophilic. The hydrophilic residues are located mostly at the surface active sites of proteins. This could contributes to the reactivity of the USPs as well as contribute to life cycle adaptation in aqueous environment. The USPs investigated were more basic in nature as most of their isoelectric point (pI) values were above 7, with those from *S. japonicum* highly basic than those from *S. mansoni*. The computed instability indices of the USPs were conspicuously above 40, however USPs from *S. japonicum* were predicted to be highly unstable. Their relative high aliphatic index values and conspicuously very low grand average hydropathy (GRAVY) indicate that schistomome USPs are very reactive in the water environment or aqueous phase, which seems consistent with their immediate environment and developmental stages.

The conserved domain search shows that Smp_076400 and Q86DW2 are made up of single USP domains with ligand binding sites. The ligand binding sites of both Smp_076400 and Q86DW2 are made up in part by the

same residues of the ATP binding motif (Gly145, Arg147, Gly148, Gly158, and Ser159). The predicted 3D chemical ligands for Smp_076400 and Sjp_0058490 (Q86DW2) included three metal ions of Ca^{2+}, Mg^{2+} and Zn^{2+}. The presence of Mg^{2+} and other metallic ions as chemical ligands indicate that they are important in the phosphorylation of USP as evident in other proteins. The ATP binding motif might be essential for phosphorylating the USP gene during posttranslational modification reaction and during ATP dependent stress response mechanism.

The ATP binding could be a regulator controlling the diverse specialized functions of schistosome stress response gene family. The exposed-orientation of ATP binding residues on the protein surface makes them accessible to the incoming free ATP molecules. These conserved binding sites might be the key residues regulating the molecular mechanism of stress response in the schistosome genome.

The adaptive structural conformation indicates functional efficiency in binding ATP during phosphorylation and stress response mechanism. Calcium ion (Ca^{2+}) was predicted to bind to both Smp_076400 and Sjp_0058490 (Q86DW2). In *S. mansoni*, Ca^{2+} is considered vital for regulated motor related activities and also critical for the egg hatching process in fresh water. Microarray-based transcriptome analysis of the response of *S. mansoni* PR-1 to praziquantel has identified genes for cytostolic calcium regulation. This mechanism might be the hallmark in the developmental and functional regulation of the USP genes in the schistosome genomes and can be consider applicable to other genomes expressing USP genes.

Objective 3

6.1.4. Determine the distinctive structural protein features of a predicted regulator of *Schistosoma* adenylate cyclase activity that has possible influence on the functioning of universal stress proteins

This research has elucidated the distinctive characteristic features of the *Schistosoma mansoni* Smp_059340.1 protein, a predicted drug target, from its primary structure to how these features contribute to the functioning and regulation of the developmental process in the schistosome's life cycle. However because the distinctive characteristic features are predictions, they should be considered with caution. The protein is basic and moderately hydrophilic carrying a net positive charge with possible antigenic properties on its surface. Smp_059340.1 is predicted to be stable within a wide range of temperature with high possibility of it been analyzed using UV spectrum assay protocol.

Further, Smp_059340.1 is a soluble protein consisting of mixed secondary structure features with the transmembrane helices having an inside → outside orientation. The quality of the modeled Smp_059340.1 structure measured in terms of reliability using the QMEAN4 score was 0.68 in range of 0 to 1 using statistical potential terms only and the Ramachandran plot analysis indicates that 95.72% (320 residues) of the model residues are located in the core or favoured region. As such the predicted model can be assumed to be of good quality and biological informative. With respect to the reliability of the modeled protein structure, the key features controlling GTP binding and effective pathway regulations were found to be Thr188, Gly210,

Asp207 and Ser53. Analysis of the protein domain organization revealed that the pathway regulated by this protein is very complex and the domains are interconnected through shared conserved residues regulating the network.

6.2: RECOMMENDATIONS

Cyclic Adenosine MonoPhosphate (cAMP) has a very critical and pivotal function in cell differentiation, cell movement and stress response in several organisms including *Dictyostellium* and *Trypanosoma* (Ott et al. 2000). Therefore, cAMP might be functioning as both an environmental sensor and a cytoprotector to parasites. This implies the cAMP could be valuable in the biology of schistosome universal stress proteins. However no studies or data are available for the roles of cAMP and PKA in the functional expression of the USPs in the schistosome genome. Therefore it is of importance to understand the molecular mechanisms of cAMP and cAMP-dependent PKA in the survival and infectivity of schistomes in their hostile environment, through the expression of universal stress protein (USPs).

Considering the outcome from this research, I recommend also further work on (i) functional characterize the interaction of calcium ions with amino acid residues of *Schistosoma* USPs; and (ii) determine the transcriptional response of *Schistosoma* USP genes to praziquantel. The datasets produced and the visual analytics views developed can be easily reused to develop new hypotheses.

Finally, the results from modeling can guide molecular biologists choice of gene-gene interactions that are most promising for further investigation. Model predictions may provide useful information as molecular biology has increasingly moved away from considering isolated genes towards a pathway-based approach. Nevertheless, models can be used as platforms to test hypotheses that may be experimentally difficult or expensive.

REFERENCES

ABEL, E. S., DAVIDS, B. J., ROBLES, L. D., LOFLIN, C. E., GILLIN, F. D. & CHAKRABARTI, R. 2001. Possible roles of protein kinase A in cell motility and excystation of the early diverging eukaryote *Giardia lamblia*. *J Biol Chem,* 276**,** 10320-9.

ADAMSON, P. B. 1976. Schistosomiasis in antiquity. *Med Hist,* 20**,** 176-88.

ADARI, H., LOWY, D. R., WILLUMSEN, B. M., DER, C. J. & MCCORMICK, F. 1988. Guanosine triphosphatase activating protein (GAP) interacts with the p21 ras effector binding domain. *Science,* 240**,** 518-21.

ARAGON, A. D., IMANI, R. A., BLACKBURN, V. R. & CUNNINGHAM, C. 2008. Microarray based analysis of temperature and oxidative stress induced messenger RNA in *Schistosoma mansoni*. *Mol Biochem Parasitol,* 162**,** 134-41.

ARAGON, A. D., IMANI, R. A., BLACKBURN, V. R., CUPIT, P. M., MELMAN, S. D., GORONGA, T., WEBB, T., LOKER, E. S. & CUNNINGHAM, C. 2009. Towards an understanding of the mechanism of action of praziquantel. *Mol Biochem Parasitol,* 164**,** 57-65.

ASLANZADEH, V. & GHADERIAN, M. 2012. Homology modeling and functional characterization of PR-1a protein of *Hordeum vulgare* subsp. *Vulgare*. *Bioinformation,* 8**,** 807-11.

BACHMAIR, A., FINLEY, D. & VARSHAVSKY, A. 1986. In vivo half-life of a protein is a function of its amino-terminal residue. *Science,* 234**,** 179-86.

BARR, S. D. & GEDAMU, L. 2001. Cloning and characterization of three differentially expressed peroxidoxin genes from *Leishmania chagasi*. Evidence for an enzymatic detoxification of hydroxyl radicals. *J Biol Chem,* 276**,** 34279-87.

BENKERT, P., TOSATTO, S. C. & SCHOMBURG, D. 2008. QMEAN: A comprehensive scoring function for model quality assessment. *Proteins,* 71**,** 261-77.

BEREZOVSKY, I. N. & TRIFONOV, E. N. 2001. Van der Waals locks: loop-n-lock structure of globular proteins. *J Mol Biol,* 307**,** 1419-26.

BERRIMAN, M., HAAS, B. J., LOVERDE, P. T., WILSON, R. A., DILLON, G. P., CERQUEIRA, G. C., MASHIYAMA, S. T., AL-LAZIKANI, B., ANDRADE, L. F., ASHTON, P. D., ASLETT, M. A., BARTHOLOMEU, D. C., BLANDIN, G., CAFFREY, C. R., COGHLAN, A., COULSON, R., DAY, T. A., DELCHER, A., DEMARCO, R., DJIKENG, A., EYRE, T., GAMBLE, J. A., GHEDIN, E., GU, Y., HERTZ-FOWLER, C., HIRAI, H., HIRAI, Y., HOUSTON, R., IVENS, A., JOHNSTON, D. A., LACERDA, D., MACEDO, C. D., MCVEIGH, P., NING, Z., OLIVEIRA, G., OVERINGTON, J. P., PARKHILL, J., PERTEA, M., PIERCE, R. J., PROTASIO, A. V., QUAIL, M. A., RAJANDREAM, M. A., ROGERS, J., SAJID, M., SALZBERG, S. L., STANKE, M., TIVEY, A. R., WHITE, O., WILLIAMS, D. L., WORTMAN, J., WU, W., ZAMANIAN, M., ZERLOTINI, A., FRASER-LIGGETT, C. M., BARRELL, B. G. & EL-SAYED, N. M. 2009. The genome of the blood fluke *Schistosoma mansoni*. *Nature,* 460**,** 352-8.

BHATTACHARYA, A., BISWAS, A. & DAS, P. K. 2008. Role of intracellular cAMP in differentiation-coupled induction of resistance against oxidative damage in *Leishmania donovani*. *Free Radic Biol Med,* 44**,** 779-94.

BINDER, M., JUSTO, A., RILEY, R., SALAMOV, A., LOPEZ-GIRALDEZ, F., SJOKVIST, E., COPELAND, A., FOSTER, B., SUN, H., LARSSON, E., LARSSON, K. H., TOWNSEND, J., GRIGORIEV, I. V. & HIBBETT, D. S. 2013. Phylogenetic and phylogenomic overview of the Polyporales. *Mycologia*.

BIRNBAUMER, L. 1990. Transduction of receptor signal into modulation of effector activity by G proteins: the first 20 years or so. *FASEB J,* 4**,** 3178-88.
BLAIR, K. L., BENNETT, J. L. & PAX, R. A. 1988. *Schistosoma mansoni*: evidence for protein kinase-C-like modulation of muscle activity. *Exp Parasitol,* 66**,** 243-52.
BLAXTER, M., ASLETT, M., GUILIANO, D. & DAUB, J. 1999. Parasitic helminth genomics. Filarial Genome Project. *Parasitology,* 118 Suppl**,** S39-51.
BLOM, N., SICHERITZ-PONTEN, T., GUPTA, R., GAMMELTOFT, S. & BRUNAK, S. 2004. Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics,* 4**,** 1633-49.
BOCHKAREVA, E. S., GIRSHOVICH, A. S. & BIBI, E. 2002. Identification and characterization of the *Escherichia coli* stress protein UP12, a putative in vivo substrate of GroEL. *Eur J Biochem,* 269**,** 3032-40.
BORCH, M., KIERNAN, M., RUST, K., BARON, B., SIMMONS, B., HATTALA, P., DAVEY, A., YOVANOVICH, J., SHAYDER, D., WASILEWSKI, A. & LAFARO, V. E. 2009. Schistosomiasis: a case study. *Urol Nurs,* 29**,** 26-9.
BRANDEEN, C. I. 1980. Relation between structure and function of alpha/beta-proteins. *Q Rev Biophys,* 13**,** 317-38.
BRINDLEY, P. J., MITREVA, M., GHEDIN, E. & LUSTIGMAN, S. 2009. Helminth genomics: The implications for human health. *PLoS Negl Trop Dis,* 3**,** e538.
BUCHACHENKO, A. L., KOUZNETSOV, D. A., ARKHANGELSKY, S. E., ORLOVA, M. A. & MARKARIAN, A. A. 2005. Spin biochemistry: magnetic 24Mg-25Mg-26Mg isotope effect in mitochondrial ADP phosphorylation. *Cell Biochem Biophys,* 43**,** 243-51.
BUCHACHENKO, A. L., KOUZNETSOV, D. A., BRESLAVSKAYA, N. N. & ORLOVA, M. A. 2008. Magnesium isotope effects in enzymatic phosphorylation. *J Phys Chem B,* 112**,** 2548-56.
CALES, C., HANCOCK, J. F., MARSHALL, C. J. & HALL, A. 1988. The cytoplasmic protein GAP is implicated as the target for regulation by the ras gene product. *Nature,* 332**,** 548-51.
CARROLL, J. A., EL-HAGE, N., MILLER, J. C., BABB, K. & STEVENSON, B. 2001. Borrelia burgdorferi RevA antigen is a surface-exposed outer membrane protein whose expression is regulated in response to environmental temperature and pH. *Infect Immun,* 69**,** 5286-93.
CHABOT, C. 2009. Demystifying visual analytics. *IEEE Comput Graph Appl,* 29**,** 84-7.
CHAI, M., MCMANUS, D. P., MCINNES, R., MOERTEL, L., TRAN, M., LOUKAS, A., JONESA, M. K. & GOBERT, G. N. 2006. Transcriptome profiling of lung schistosomula,in vitro cultured schistosomula and adult *Schistosoma japonicum*. *Cell Mol Life Sci,* 63**,** 919-29.
CHANG, R., ZIEMKIEWICZ, C., GREEN, T. M. & RIBARSKY, W. 2009. Defining insight for visual analytics. *IEEE Comput Graph Appl,* 29**,** 14-7.
CHEN, W., HONMA, K., SHARMA, A. & KURAMITSU, H. K. 2006. A universal stress protein of *Porphyromonas gingivalis* is involved in stress responses and biofilm formation. *FEMS Microbiol Lett,* 264**,** 15-21.
CHEN, Y. W., FERSHT, A. R. & HENRICK, K. 1993. Contribution of buried hydrogen bonds to protein stability. The crystal structures of two barnase mutants. *J Mol Biol,* 234**,** 1158-70.
CHOU, K. C. & SHEN, H. B. 2010. A new method for predicting the subcellular localization of eukaryotic proteins with both single and multiple sites: Euk-mPLoc 2.0. *PLoS One,* 5**,** e9931.
CLARK, T. G., ABRAHAMSEN, M. S. & WHITE, M. W. 1996. Developmental expression of heat shock protein 90 in *Eimeria bovis*. *Mol Biochem Parasitol,* 78**,** 259-63.

CLEGG, J. S. & JACKSON, S. A. 1992. Aerobic heat shock activates trehalose synthesis in embryos of *Artemia franciscana*. *FEBS Lett,* 303**,** 45-7.
COLES, G. C. 1979. The effect of praziquantel on *Schistosoma mansoni*. *J Helminthol,* 53**,** 31-3.
COLICELLI, J. 2004. Human RAS superfamily proteins and related GTPases. *Sci STKE,* 2004**,** RE13.
COLICELLI, J. 2010. Signal transduction: RABGEF1 fingers RAS for ubiquitination. *Curr Biol,* 20**,** R630-2.
CROWTHER, G. J., SHANMUGAM, D., CARMONA, S. J., DOYLE, M. A., HERTZ-FOWLER, C., BERRIMAN, M., NWAKA, S., RALPH, S. A., ROOS, D. S., VAN VOORHIS, W. C. & AGUERO, F. 2010. Identification of attractive drug targets in neglected-disease pathogens using an in silico approach. *PLoS Negl Trop Dis,* 4**,** e804.
CUNHA, V. M. & NOEL, F. 1988. A Mg2+-independent Ca2+-stimulated ATPase activity in the tegument of *Schistosoma mansoni*. *Braz J Med Biol Res,* 21**,** 449-51.
DAUBENBERGER, C., HEUSSLER, V., GOBRIGHT, E., WIJNGAARD, P., CLEVERS, H. C., WELLS, C., TSUJI, N., MUSOKE, A. & MCKEEVER, D. 1997. Molecular characterisation of a cognate 70 kDa heat shock protein of the protozoan *Theileria parva*. *Mol Biochem Parasitol,* 85**,** 265-9.
DAVIS, W. L., JACOBY, B. H. & GOODMAN, D. B. 1994. Immunolocalization of ubiquitin in degenerating insect flight muscle. *Histochem J,* 26**,** 298-305.
DIEZ, A., GUSTAVSSON, N. & NYSTROM, T. 2000. The universal stress protein A of *Escherichia coli* is required for resistance to DNA damaging agents and is regulated by a RecA/FtsK-dependent regulatory pathway. *Mol Microbiol,* 36**,** 1494-503.
DIEZ, A. A., FAREWELL, A., NANNMARK, U. & NYSTROM, T. 1997. A mutation in the ftsK gene of *Escherichia coli* affects cell-cell separation, stationary-phase survival, stress adaptation, and expression of the gene encoding the stress protein UspA. *J Bacteriol,* 179**,** 5878-83.
DILLER, T. C., MADHUSUDAN, XUONG, N. H. & TAYLOR, S. S. 2001. Molecular basis for regulatory subunit diversity in cAMP-dependent protein kinase: crystal structure of the type II beta regulatory subunit. *Structure,* 9**,** 73-82.
DISSOUS, C., AHIER, A. & KHAYATH, N. 2007. Protein tyrosine kinases as new potential targets against human schistosomiasis. *Bioessays,* 29**,** 1281-8.
DIX, D. J. 1997. Hsp70 expression and function during gametogenesis. *Cell Stress Chaperones,* 2**,** 73-7.
DOENHOFF, M. J., CIOLI, D. & UTZINGER, J. 2008. Praziquantel: mechanisms of action, resistance and new derivatives for schistosomiasis. *Curr Opin Infect Dis,* 21**,** 659-67.
DOENHOFF, M. J., HAGAN, P., CIOLI, D., SOUTHGATE, V., PICA-MATTOCCIA, L., BOTROS, S., COLES, G., TCHUEM TCHUENTE, L. A., MBAYE, A. & ENGELS, D. 2009. Praziquantel: its use in control of schistosomiasis in sub-Saharan Africa and current research needs. *Parasitology,* 136**,** 1825-35.
DOERIG, C. 2004. Protein kinases as targets for anti-parasitic chemotherapy. *Biochim Biophys Acta,* 1697**,** 155-68.
DOERIG, C., BILLKER, O., HAYSTEAD, T., SHARMA, P., TOBIN, A. B. & WATERS, N. C. 2008. Protein kinases of malaria parasites: an update. *Trends Parasitol,* 24**,** 570-7.
DONALD, R. G., ALLOCCO, J., SINGH, S. B., NARE, B., SALOWE, S. P., WILTSIE, J. & LIBERATOR, P. A. 2002. *Toxoplasma gondii* cyclic GMP-dependent kinase: chemotherapeutic targeting of an essential parasite protein kinase. *Eukaryot Cell,* 1**,** 317-28.
DONOVAN, G. T., NORTON, J. P., BOWER, J. M. & MULVEY, M. A. 2013. Adenylate cyclase and the cyclic AMP receptor protein modulate stress

resistance and virulence capacity of uropathogenic *Escherichia coli*. *Infect Immun,* 81**,** 249-58.
DRUMM, J. E., MI, K., BILDER, P., SUN, M., LIM, J., BIELEFELDT-OHMANN, H., BASARABA, R., SO, M., ZHU, G., TUFARIELLO, J. M., IZZO, A. A., ORME, I. M., ALMO, S. C., LEYH, T. S. & CHAN, J. 2009. *Mycobacterium tuberculosis* universal stress protein Rv2623 regulates bacillary growth by ATP-Binding: requirement for establishing chronic persistent infection. *PLoS Pathog,* 5**,** e1000460.
DUNN, C. W., HEJNOL, A., MATUS, D. Q., PANG, K., BROWNE, W. E., SMITH, S. A., SEAVER, E., ROUSE, G. W., OBST, M., EDGECOMBE, G. D., SORENSEN, M. V., HADDOCK, S. H., SCHMIDT-RHAESA, A., OKUSU, A., KRISTENSEN, R. M., WHEELER, W. C., MARTINDALE, M. Q. & GIRIBET, G. 2008. Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature,* 452**,** 745-9.
EGEE, S., LAPAIX, F., DECHERF, G., STAINES, H. M., ELLORY, J. C., DOERIG, C. & THOMAS, S. L. 2002. A stretch-activated anion channel is up-regulated by the malaria parasite *Plasmodium falciparum*. *J Physiol,* 542**,** 795-801.
FAREWELL, A., KVINT, K. & NYSTROM, T. 1998. uspB, a new sigmaS-regulated gene in *Escherichia coli* which is required for stationary-phase resistance to ethanol. *J Bacteriol,* 180**,** 6140-7.
FARIAS, L. P., TARARAM, C. A., MIYASATO, P. A., NISHIYAMA, M. Y., JR., OLIVEIRA, K. C., KAWANO, T., VERJOVSKI-ALMEIDA, S. & LEITE, L. C. 2011. Screening the *Schistosoma mansoni* transcriptome for genes differentially expressed in the schistosomulum stage in search for vaccine candidates. *Parasitol Res,* 108**,** 123-35.
FEDER, M. E. & HOFMANN, G. E. 1999. Heat-shock proteins, molecular chaperones, and the stress response: evolutionary and ecological physiology. *Annu Rev Physiol,* 61**,** 243-82.
FEDER, M. E. & KREBS, R. A. 1997. Ecological and evolutionary physiology of heat shock proteins and the stress response in *Drosophila*: complementary insights from genetic engineering and natural variation. *EXS,* 83**,** 155-73.
FERGUSON, K. M., HIGASHIJIMA, T., SMIGEL, M. D. & GILMAN, A. G. 1986. The influence of bound GDP on the kinetics of guanine nucleotide binding to G proteins. *J Biol Chem,* 261**,** 7393-9.
FIELD, C. M. & KELLOGG, D. 1999. Septins: cytoskeletal polymers or signalling GTPases? *Trends Cell Biol,* 9**,** 387-94.
FITZPATRICK, J. M., PEAK, E., PERALLY, S., CHALMERS, I. W., BARRETT, J., YOSHINO, T. P., IVENS, A. C. & HOFFMANN, K. F. 2009. Anti-schistosomal intervention targets identified by lifecycle transcriptomic analyses. *PLoS Negl Trop Dis,* 3**,** e543.
FLANNAGAN, R. D., TAMMARIELLO, S. P., JOPLIN, K. H., CIKRA-IRELAND, R. A., YOCUM, G. D. & DENLINGER, D. L. 1998. Diapause-specific gene expression in pupae of the flesh fly *Sarcophaga crassipalpis*. *Proc Natl Acad Sci U S A,* 95**,** 5616-20.
FORET, S., SENECA, F., DE JONG, D., BIELLER, A., HEMMRICH, G., AUGUSTIN, R., HAYWARD, D. C., BALL, E. E., BOSCH, T. C., AGATA, K., HASSEL, M. & MILLER, D. J. 2011. Phylogenomics reveals an anomalous distribution of USP genes in metazoans. *Mol Biol Evol,* 28**,** 153-61.
FRANCO, G. R., ADAMS, M. D., SOARES, M. B., SIMPSON, A. J., VENTER, J. C. & PENA, S. D. 1995. Identification of new *Schistosoma mansoni* genes by the EST strategy using a directional cDNA library. *Gene,* 152**,** 141-7.
FREESTONE, P., NYSTROM, T., TRINEI, M. & NORRIS, V. 1997. The universal stress protein, UspA, of *Escherichia coli* is phosphorylated in response to stasis. *J Mol Biol,* 274**,** 318-24.

FREESTONE, P., TRINEI, M., CLARKE, S. C., NYSTROM, T. & NORRIS, V. 1998. Tyrosine phosphorylation in *Escherichia coli*. *J Mol Biol,* 279**,** 1045-51.
GAGLIARDI, D., BRETON, C., CHABOUD, A., VERGNE, P. & DUMAS, C. 1995. Expression of heat shock factor and heat shock protein 70 genes during maize pollen development. *Plant Mol Biol,* 29**,** 841-56.
GANTT, K. R., GOLDMAN, T. L., MCCORMICK, M. L., MILLER, M. A., JERONIMO, S. M., NASCIMENTO, E. T., BRITIGAN, B. E. & WILSON, M. E. 2001. Oxidative responses of human and murine macrophages during phagocytosis of *Leishmania chagasi*. *J Immunol,* 167**,** 893-901.
GASTEIGER, E., GATTIKER, A., HOOGLAND, C., IVANYI, I., APPEL, R. D. & BAIROCH, A. 2003. ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res,* 31**,** 3784-8.
GHOSH, S., GOSWAMI, S. & ADHYA, S. 2003. Role of superoxide dismutase in survival of *Leishmania* within the macrophage. *Biochem J,* 369**,** 447-52.
GIAMBIAGI-DEMARVAL, M., SOUTO-PADRON, T. & RONDINELLI, E. 1996. Characterization and cellular distribution of heat-shock proteins HSP70 and HSP60 in *Trypanosoma cruzi*. *Exp Parasitol,* 83**,** 335-45.
GILL, S. C. & VON HIPPEL, P. H. 1989. Calculation of protein extinction coefficients from amino acid sequence data. *Anal Biochem,* 182**,** 319-26.
GILLE, A., LIU, H. Y., SPRANG, S. R. & SEIFERT, R. 2002. Distinct interactions of GTP, UTP, and CTP with G(s) proteins. *J Biol Chem,* 277**,** 34434-42.
GILMAN, A. G. 1987. G proteins: transducers of receptor-generated signals. *Annu Rev Biochem,* 56**,** 615-49.
GOBERT, G. N., MCINNES, R., MOERTEL, L., NELSON, C., JONES, M. K., HU, W. & MCMANUS, D. P. 2006. Transcriptomics tool for the human *Schistosoma* blood flukes using microarray gene expression profiling. *Exp Parasitol,* 114**,** 160-72.
GOBERT, G. N., MCMANUS, D. P., NAWARATNA, S., MOERTEL, L., MULVENNA, J. & JONES, M. K. 2009a. Tissue specific profiling of females of *Schistosoma japonicum* by integrated laser microdissection microscopy and microarray analysis. *PLoS Negl Trop Dis,* 3**,** e469.
GOBERT, G. N., MOERTEL, L., BRINDLEY, P. J. & MCMANUS, D. P. 2009b. Developmental gene expression profiles of the human pathogen *Schistosoma japonicum*. *BMC Genomics,* 10**,** 128.
GOBERT, G. N., TRAN, M. H., MOERTEL, L., MULVENNA, J., JONES, M. K., MCMANUS, D. P. & LOUKAS, A. 2010. Transcriptional changes in *Schistosoma mansoni* during early schistosomula development and in the presence of erythrocytes. *PLoS Negl Trop Dis,* 4**,** e600.
GONDA, D. K., BACHMAIR, A., WUNNING, I., TOBIAS, J. W., LANE, W. S. & VARSHAVSKY, A. 1989. Universality and structure of the N-end rule. *J Biol Chem,* 264**,** 16700-12.
GOPALAKRISHNAN, K., SOWMIYA, G., SHEIK, S. S. & SEKAR, K. 2007. Ramachandran plot on the web (2.0). *Protein Pept Lett,* 14**,** 669-71.
GORDON, S., BHARADWAJ, S., HNATOV, A., ALI, A. & OVSENEK, N. 1997. Distinct stress-inducible and developmentally regulated heat shock transcription factors in *Xenopus* oocytes. *Dev Biol,* 181**,** 47-63.
GOUJON, M., MCWILLIAM, H., LI, W., VALENTIN, F., SQUIZZATO, S., PAERN, J. & LOPEZ, R. 2010. A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res,* 38**,** W695-9.
GRABE, K. & HAAS, W. 2004. Navigation within host tissues: cercariae orientate towards dark after penetration. *Parasitol Res,* 93**,** 111-3.
GRAF, R., ANZALI, S., BUENGER, J., PFLUECKER, F. & DRILLER, H. 2008. The multifunctional role of ectoine as a natural cell protectant. *Clin Dermatol,* 26**,** 326-33.

GREASLEY, S. E., JHOTI, H., TEAHAN, C., SOLARI, R., FENSOME, A., THOMAS, G. M., COCKCROFT, S. & BAX, B. 1995. The structure of rat ADP-ribosylation factor-1 (ARF-1) complexed to GDP determined from two different crystal forms. *Nat Struct Biol,* 2**,** 797-806.

GREENBERG, R. M. 2005. Ca2+ signalling, voltage-gated Ca2+ channels and praziquantel in flatworm neuromusculature. *Parasitology,* 131 Suppl**,** S97-108.

GRYSEELS, B. 2012. Schistosomiasis. *Infect Dis Clin North Am,* 26**,** 383-97.

GUPTA, S. 2009. Bioinformatics—Research Applications. *In:* FULEKAR, M. H. (ed.) *Bioinformatics: Applications in Life and Environmental Sciences.* Springer Netherlands.

GURNETT, A. M., LIBERATOR, P. A., DULSKI, P. M., SALOWE, S. P., DONALD, R. G., ANDERSON, J. W., WILTSIE, J., DIAZ, C. A., HARRIS, G., CHANG, B., DARKIN-RATTRAY, S. J., NARE, B., CRUMLEY, T., BLUM, P. S., MISURA, A. S., TAMAS, T., SARDANA, M. K., YUAN, J., BIFTU, T. & SCHMATZ, D. M. 2002. Purification and molecular characterization of cGMP-dependent protein kinase from Apicomplexan parasites. A novel chemotherapeutic target. *J Biol Chem,* 277**,** 15913-22.

GURUPRASAD, K., REDDY, B. V. & PANDIT, M. W. 1990. Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. *Protein Eng,* 4**,** 155-61.

GUSTAVSSON, N., DIEZ, A. & NYSTROM, T. 2002. The universal stress protein paralogues of *Escherichia coli* are co-ordinately regulated and co-operate in the defence against DNA damage. *Mol Microbiol,* 43**,** 107-17.

HAN, Z. G., BRINDLEY, P. J., WANG, S. Y. & CHEN, Z. 2009. *Schistosoma* genomics: new perspectives on schistosome biology and host-parasite interaction. *Annu Rev Genomics Hum Genet,* 10**,** 211-40.

HANAWA, T., YONEZAWA, H., KAWAKAMI, H., KAMIYA, S. & ARMSTRONG, S. K. 2013. Role of *Bordetella pertussis* RseA in the cell envelope stress response and adenylate cyclase toxin release. *Pathog Dis*.

HANKS, S. K. 2003. Genomic analysis of the eukaryotic protein kinase superfamily: a perspective. *Genome Biol,* 4**,** 111.

HELM, K. W., PETERSEN, N. S. & ABERNETHY, R. H. 1989. Heat shock response of germinating embryos of wheat : effects of imbibition time and seed vigor. *Plant Physiol,* 90**,** 598-605.

HEPLER, J. R., KOZASA, T., SMRCKA, A. V., SIMON, M. I., RHEE, S. G., STERNWEIS, P. C. & GILMAN, A. G. 1993. Purification from Sf9 cells and characterization of recombinant Gq alpha and G11 alpha. Activation of purified phospholipase C isozymes by G alpha subunits. *J Biol Chem,* 268**,** 14367-75.

HERTZ-FOWLER, C., PEACOCK, C. S., WOOD, V., ASLETT, M., KERHORNOU, A., MOONEY, P., TIVEY, A., BERRIMAN, M., HALL, N., RUTHERFORD, K., PARKHILL, J., IVENS, A. C., RAJANDREAM, M. A. & BARRELL, B. 2004. GeneDB: a resource for prokaryotic and eukaryotic organisms. *Nucleic Acids Res,* 32**,** D339-43.

HIGASHIJIMA, T., FERGUSON, K. M., STERNWEIS, P. C., SMIGEL, M. D. & GILMAN, A. G. 1987. Effects of Mg2+ and the beta gamma-subunit complex on the interactions of guanine nucleotides with G proteins. *J Biol Chem,* 262**,** 762-6.

HINGLEY-WILSON, S. M., LOUGHEED, K. E., FERGUSON, K., LEIVA, S. & WILLIAMS, H. D. 2010. Individual *Mycobacterium tuberculosis* universal stress protein homologues are dispensable in vitro. *Tuberculosis (Edinb),* 90**,** 236-44.

HOOFT, R. W., SANDER, C. & VRIEND, G. 1997. Objectively judging the quality of a protein structure from a Ramachandran plot. *Comput Appl Biosci,* 13**,** 425-30.
HOPP, T. P. & WOODS, K. R. 1981. Prediction of protein antigenic determinants from amino acid sequences. *Proc Natl Acad Sci U S A,* 78**,** 3824-8.
HUBEL, A., KROBITSCH, S., HORAUF, A. & CLOS, J. 1997. *Leishmania major* Hsp100 is required chiefly in the mammalian stage of the parasite. *Mol Cell Biol,* 17**,** 5987-95.
IKAI, A. 1980. Thermostability and aliphatic index of globular proteins. *J Biochem,* 88**,** 1895-8.
ISOKPEHI, R. D., MAHMUD, O., MBAH, A. N., SIMMONS, S. S., AVELAR, L., RAJNARAYANAN, R. V., UDENSI, U. K., AYENSU, W. K., COHLY, H. H., BROWN, S. D., DATES, C. R., HENTZ, S. D., HUGHES, S. J., SMITH-MCINNIS, D. R., PATTERSON, C. O., SIMS, J. N., TURNER, K. T., WILLIAMS, B. S., JOHNSON, M. O., ADUBI, T., MBUH, J. V., ANUMUDU, C. I., ADEOYE, G. O., THOMAS, B. N., NASHIRU, O. & OLIVEIRA, G. 2011a. Developmental regulation of genes encoding universal stress proteins in *Schistosoma mansoni*. *Gene Regul Syst Bio,* 5**,** 61-74.
ISOKPEHI, R. D., SIMMONS, S. S., COHLY, H. H., EKUNWE, S. I., BEGONIA, G. B. & AYENSU, W. K. 2011b. Identification of drought-responsive universal stress proteins in viridiplantae. *Bioinform Biol Insights,* 5**,** 41-58.
JASPER, H., BENES, V., SCHWAGER, C., SAUER, S., CLAUDER-MUNSTER, S., ANSORGE, W. & BOHMANN, D. 2001. The genomic response of the *Drosophila* embryo to JNK signaling. *Dev Cell,* 1**,** 579-86.
JOHNSON, M. O., COHLY, H. H., ISOKPEHI, R. D. & AWOFOLU, O. R. 2010. The case for visual analytics of arsenic concentrations in foods. *Int J Environ Res Public Health,* 7**,** 1970-83.
JOLLY, E. R., CHIN, C. S., MILLER, S., BAHGAT, M. M., LIM, K. C., DERISI, J. & MCKERROW, J. H. 2007. Gene expression patterns during adaptation of a helminth parasite to different environmental niches. *Genome Biol,* 8**,** R65.
JONES, D. T., TAYLOR, W. R. & THORNTON, J. M. 1992. The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci,* 8**,** 275-82.
JONES, S. J., RIDDLE, D. L., POUZYREV, A. T., VELCULESCU, V. E., HILLIER, L., EDDY, S. R., STRICKLIN, S. L., BAILLIE, D. L., WATERSTON, R. & MARRA, M. A. 2001. Changes in gene expression associated with developmental arrest and longevity in *Caenorhabditis elegans*. *Genome Res,* 11**,** 1346-52.
JOSEPH, S. J., FERNANDEZ-ROBLEDO, J. A., GARDNER, M. J., EL-SAYED, N. M., KUO, C. H., SCHOTT, E. J., WANG, H., KISSINGER, J. C. & VASTA, G. R. 2010. The Alveolate *Perkinsus marinus*: biological insights from EST gene discovery. *BMC Genomics,* 11**,** 228.
JURNAK, F., MCPHERSON, A., WANG, A. H. & RICH, A. 1980. Biochemical and structural studies of the tetragonal crystalline modification of the *Escherichia coli* elongation factor Tu. *J Biol Chem,* 255**,** 6751-7.
KATSUMATA, T., KOHNO, S., YAMAGUCHI, K., HARA, K. & AOKI, Y. 1989. Hatching of *Schistosoma mansoni* eggs is a Ca2+/calmodulin-dependent process. *Parasitol Res,* 76**,** 90-1.
KATSUMATA, T., SHIMADA, M., SATO, K. & AOKI, Y. 1988. Possible involvement of calcium ions in the hatching of *Schistosoma mansoni* eggs in water. *J Parasitol,* 74**,** 1040-1.
KAUSHAL, D. C., CARTER, R., MILLER, L. H. & KRISHNA, G. 1980. Gametocytogenesis by malaria parasites in continuous culture. *Nature,* 286**,** 490-2.
KAWAMOTO, F., SHOZAWA, A., KUMADA, N. & KOJIMA, K. 1989. Possible roles of cAMP and Ca2+ in the regulation of miracidial transformation in *Schistosoma mansoni*. *Parasitol Res,* 75**,** 368-74.

KAWASHIMA, S., POKAROWSKI, P., POKAROWSKA, M., KOLINSKI, A., KATAYAMA, T. & KANEHISA, M. 2008. AAindex: amino acid index database, progress report 2008. *Nucleic Acids Res,* 36**,** D202-5.
KELLEY, L. A. & STERNBERG, M. J. 2009. Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc,* 4**,** 363-71.
KERK, D., BULGRIEN, J., SMITH, D. W. & GRIBSKOV, M. 2003. Arabidopsis proteins containing similarity to the universal stress protein domain of bacteria. *Plant Physiol,* 131**,** 1209-19.
KIM, H., GOO, E., KANG, Y., KIM, J. & HWANG, I. 2012. Regulation of universal stress protein genes by quorum sensing and RpoS in *Burkholderia glumae*. *J Bacteriol,* 194**,** 982-92.
KIM, J. H. 2013. Translational bioinformatics has now come of age: TBC 2012 collection. *BMC Med Genomics,* 6 Suppl 2**,** I1.
KNIGHTON, D. R., ZHENG, J. H., TEN EYCK, L. F., ASHFORD, V. A., XUONG, N. H., TAYLOR, S. S. & SOWADSKI, J. M. 1991. Crystal structure of the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase. *Science,* 253**,** 407-14.
KNOCKAERT, M., GRAY, N., DAMIENS, E., CHANG, Y. T., GRELLIER, P., GRANT, K., FERGUSSON, D., MOTTRAM, J., SOETE, M., DUBREMETZ, J. F., LE ROCH, K., DOERIG, C., SCHULTZ, P. & MEIJER, L. 2000. Intracellular targets of cyclin-dependent kinase inhibitors: identification by affinity chromatography using immobilised inhibitors. *Chem Biol,* 7**,** 411-22.
KOHUT, G., OLAH, B., ADAM, A. L., GARCIA-MARTINEZ, J. & HORNOK, L. 2010. Adenylyl cyclase regulates heavy metal sensitivity, bikaverin production and plant tissue colonization in *Fusarium proliferatum*. *J Basic Microbiol,* 50**,** 59-71.
KRAULIS, P. J., DOMAILLE, P. J., CAMPBELL-BURK, S. L., VAN AKEN, T. & LAUE, E. D. 1994. Solution structure and dynamics of ras p21.GDP determined by heteronuclear three- and four-dimensional NMR spectroscopy. *Biochemistry,* 33**,** 3515-31.
KREBS RA, F. M. 1997. Deleterious consequences of Hsp70 overexpression in *Drosophila melanogaster* larvae. *Cell Stress Chaperones. 1997 Mar;2(1):60-71.*
KUMAR, N. V., RANI, M. E., GUNASEELI, R., KANNAN, N. D. & SRIDHAR, J. 2012. Modeling and structural analysis of cellulases using *Clostridium thermocellum* as template. *Bioinformation,* 8**,** 1105-10.
KUZNETSOV, D. A., ALIAUTDIN, R. N., MARKARIAN, A. A., BERDIEVA, A. G., KHASIGOV, P. Z., GATAGONOVA, T. M., KTSOEVA, S. A. & ORLOVA, M. A. 2006. [The effect of magnesium pool isotopy on reactivation of mitochondrial ATP synthesis suppressed by 1-methyl-nicotine amide]. *Biomed Khim,* 52**,** 146-52.
KVINT, K., NACHIN, L., DIEZ, A. & NYSTROM, T. 2003. The bacterial universal stress protein: function and regulation. *Curr Opin Microbiol,* 6**,** 140-5.
KYTE, J. & DOOLITTLE, R. F. 1982. A simple method for displaying the hydropathic character of a protein. *J Mol Biol,* 157**,** 105-32.
LANGER, T., LU, C., ECHOLS, H., FLANAGAN, J., HAYER, M. K. & HARTL, F. U. 1992. Successive action of DnaK, DnaJ and GroEL along the pathway of chaperone-mediated protein folding. *Nature,* 356**,** 683-9.
LARKIN, M. A., BLACKSHIELDS, G., BROWN, N. P., CHENNA, R., MCGETTIGAN, P. A., MCWILLIAM, H., VALENTIN, F., WALLACE, I. M., WILM, A., LOPEZ, R., THOMPSON, J. D., GIBSON, T. J. & HIGGINS, D. G. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics,* 23**,** 2947-8.
LIANG, P., AMONS, R., CLEGG, J. S. & MACRAE, T. H. 1997. Molecular characterization of a small heat shock/alpha-crystallin protein in encysted *Artemia* embryos. *J Biol Chem,* 272**,** 19051-8.

LIU, F., CHEN, P., CUI, S. J., WANG, Z. Q. & HAN, Z. G. 2008. SjTPdb: integrated transcriptome and proteome database and analysis platform for *Schistosoma japonicum*. *BMC Genomics,* 9**,** 304.

LIU, F., LU, J., HU, W., WANG, S. Y., CUI, S. J., CHI, M., YAN, Q., WANG, X. R., SONG, H. D., XU, X. N., WANG, J. J., ZHANG, X. L., ZHANG, X., WANG, Z. Q., XUE, C. L., BRINDLEY, P. J., MCMANUS, D. P., YANG, P. Y., FENG, Z., CHEN, Z. & HAN, Z. G. 2006. New perspectives on host-parasite interplay by comparative transcriptomic and proteomic analyses of *Schistosoma japonicum*. *PLoS Pathog,* 2**,** e29.

LIU, W. T., KARAVOLOS, M. H., BULMER, D. M., ALLAOUI, A., HORMAECHE, R. D., LEE, J. J. & KHAN, C. M. 2007. Role of the universal stress protein UspA of *Salmonella* in growth arrest, stress and virulence. *Microb Pathog,* 42**,** 2-10.

LOVELL SC, D. I., ARENDALL WB 3RD, DE BAKKER PI, WORD JM, PRISANT MG, RICHARDSON JS, RICHARDSON DC 2003. Structure validation by Calpha geometry: phi,psi and Cbeta deviation. *Proteins. 2003 Feb 15;50(3):437-50.*

LOVERDE, P. T. 1998. Do antioxidants play a role in schistosome host-parasite interactions? *Parasitol Today,* 14**,** 284-9.

LU, C. H., LIN, Y. F., LIN, J. J. & YU, C. S. 2012. Prediction of metal ion-binding sites in proteins using the fragment transformation method. *PLoS One,* 7**,** e39252.

LUO, R., ZHOU, C., LIN, J., YANG, D., SHI, Y. & CHENG, G. 2012. Identification of in vivo protein phosphorylation sites in human pathogen *Schistosoma japonicum* by a phosphoproteomic approach. *J Proteomics,* 75**,** 868-77.

LUSCOMBE, N. M., GREENBAUM, D. & GERSTEIN, M. 2001. What is bioinformatics? A proposed definition and overview of the field. *Methods Inf Med,* 40**,** 346-58.

MACCALLUM, P. H., POET, R. & MILNER-WHITE, E. J. 1995. Coulombic attractions between partially charged main-chain atoms stabilise the right-handed twist found in most beta-strands. *J Mol Biol,* 248**,** 374-84.

MACKINLAY, J., HANRAHAN, P. & STOLTE, C. 2007. Show me: automatic presentation for visual analysis. *IEEE Trans Vis Comput Graph,* 13**,** 1137-44.

MADDEN, S. L., GALELLA, E. A., ZHU, J., BERTELSEN, A. H. & BEAUDRY, G. A. 1997. SAGE transcript profiles for p53-dependent growth regulation. *Oncogene,* 15**,** 1079-85.

MANCINI, P. E. & PATTON, C. L. 1981. Cyclic 3',5'-adenosine monophosphate levels during the developmental cycle of *Trypanosoma brucei brucei* in the rat. *Mol Biochem Parasitol,* 3**,** 19-31.

MARCHLER-BAUER, A., LU, S., ANDERSON, J. B., CHITSAZ, F., DERBYSHIRE, M. K., DEWEESE-SCOTT, C., FONG, J. H., GEER, L. Y., GEER, R. C., GONZALES, N. R., GWADZ, M., HURWITZ, D. I., JACKSON, J. D., KE, Z., LANCZYCKI, C. J., LU, F., MARCHLER, G. H., MULLOKANDOV, M., OMELCHENKO, M. V., ROBERTSON, C. L., SONG, J. S., THANKI, N., YAMASHITA, R. A., ZHANG, D., ZHANG, N., ZHENG, C. & BRYANT, S. H. 2011. CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res,* 39**,** D225-9.

MARESCA, B. 1995. Unraveling the secrets of *Histoplasma capsulatum*. A model to study morphogenic adaptation during parasite host/host interaction. *Verh K Acad Geneeskd Belg,* 57**,** 133-56.

MATSUMURA, K., MITSUI, Y., SATO, K., SAKAMOTO, M. & AOKI, Y. 1991. *Schistosoma mansoni*: possible involvement of protein kinase C in linoleic acid-induced proteolytic enzyme release from cercariae. *Exp Parasitol,* 72**,** 311-20.

MATSUYAMA, H., TAKAHASHI, H., WATANABE, K., FUJIMAKI, Y. & AOKI, Y. 2004. The involvement of cyclic adenosine monophosphate in the control of schistosome miracidium cilia. *J Parasitol,* 90**,** 8-14.

MATTILA, J., BREMER, A., AHONEN, L., KOSTIAINEN, R. & PUIG, O. 2009. *Drosophila* FoxO regulates organism size and stress resistance through an adenylate cyclase. *Mol Cell Biol,* 29**,** 5357-65.
MAZZUCCHELLI, C. & SASSONE-CORSI, P. 1999. The inducible cyclic adenosine monophosphate early repressor (ICER) in the pituitary intermediate lobe: role in the stress response. *Mol Cell Endocrinol,* 155**,** 101-13.
MBAH, A. N., KAMGA, H. L., AWOFOLU, O. R. & ISOKPEHI, R. D. 2012. Drug target exploitable structural features of adenylyl cyclase activity in *Schistosoma mansoni*. *Drug Target Insights,* 6**,** 41-58.
MCCARTER, J. P. 2004. Genomic filtering: an approach to discovering novel antiparasitics. *Trends Parasitol,* 20**,** 462-8.
MELMAN, S. D., STEINAUER, M. L., CUNNINGHAM, C., KUBATKO, L. S., MWANGI, I. N., WYNN, N. B., MUTUKU, M. W., KARANJA, D. M., COLLEY, D. G., BLACK, C. L., SECOR, W. E., MKOJI, G. M. & LOKER, E. S. 2009. Reduced susceptibility to praziquantel among naturally occurring Kenyan isolates of *Schistosoma mansoni*. *PLoS Negl Trop Dis,* 3**,** e504.
MERCKX, A., NIVEZ, M. P., BOUYER, G., ALANO, P., LANGSLEY, G., DEITSCH, K., THOMAS, S., DOERIG, C. & EGEE, S. 2008. *Plasmodium falciparum* regulatory subunit of cAMP-dependent PKA and anion channel conductance. *PLoS Pathog,* 4**,** e19.
METHVEN, A., ZELSKI, S. & MILLER, A. 2013. A molecular phylogenetic assessment of the genus *Gyromitra* in North America. *Mycologia*.
MICHEL, F., JAEGER, L., WESTHOF, E., KURAS, R., TIHY, F., XU, M. Q. & SHUB, D. A. 1992. Activation of the catalytic core of a group I intron by a remote 3' splice junction. *Genes Dev,* 6**,** 1373-85.
MILBURN, M. V., TONG, L., DEVOS, A. M., BRUNGER, A., YAMAIZUMI, Z., NISHIMURA, S. & KIM, S. H. 1990. Molecular switch for signal transduction: structural differences between active and inactive forms of protooncogenic ras proteins. *Science,* 247**,** 939-45.
MILLER, M. A., MCGOWAN, S. E., GANTT, K. R., CHAMPION, M., NOVICK, S. L., ANDERSEN, K. A., BACCHI, C. J., YARLETT, N., BRITIGAN, B. E. & WILSON, M. E. 2000. Inducible resistance to oxidant stress in the protozoan *Leishmania chagasi*. *J Biol Chem,* 275**,** 33883-9.
MITREVA, M., MCCARTER, J. P., ARASU, P., HAWDON, J., MARTIN, J., DANTE, M., WYLIE, T., XU, J., STAJICH, J. E., KAPULKIN, W., CLIFTON, S. W., WATERSTON, R. H. & WILSON, R. K. 2005. Investigating hookworm genomes by comparative analysis of two *Ancylostoma* species. *BMC Genomics,* 6**,** 58.
MOERTEL, L., MCMANUS, D. P., PIVA, T. J., YOUNG, L., MCINNES, R. L. & GOBERT, G. N. 2006. Oligonucleotide microarray analysis of strain- and gender-associated gene expression in the human blood fluke, *Schistosoma japonicum*. *Mol Cell Probes,* 20**,** 280-9.
MOURAO, M. M., DINGUIRARD, N., FRANCO, G. R. & YOSHINO, T. P. 2009. Phenotypic screen of early-developing larvae of the blood fluke, *Schistosoma mansoni*, using RNA interference. *PLoS Negl Trop Dis,* 3**,** e502.
MUSHEGIAN, A. R. & KOONIN, E. V. 1996. Sequence analysis of eukaryotic developmental proteins: ancient and novel domains. *Genetics,* 144**,** 817-28.
NACHIN, L., BRIVE, L., PERSSON, K. C., SVENSSON, P. & NYSTROM, T. 2008. Heterodimer formation within universal stress protein classes revealed by an in silico and experimental approach. *J Mol Biol,* 380**,** 340-50.
NACHIN, L., NANNMARK, U. & NYSTROM, T. 2005. Differential roles of the universal stress proteins of *Escherichia coli* in oxidative stress resistance, adhesion, and motility. *J Bacteriol,* 187**,** 6265-72.
NAMBI, S., BASU, N. & VISWESWARIAH, S. S. 2010. cAMP-regulated protein lysine acetylases in mycobacteria. *J Biol Chem,* 285**,** 24313-23.

NEBL, T., PRIETO, J. H., KAPP, E., SMITH, B. J., WILLIAMS, M. J., YATES, J. R., 3RD, COWMAN, A. F. & TONKIN, C. J. 2011. Quantitative in vivo analyses reveal calcium-dependent phosphorylation sites and identifies a novel component of the *Toxoplasma* invasion motor complex. *PLoS Pathog,* 7**,** e1002222.

NEGRAO-CORREA, D., MATTOS, A. C., PEREIRA, C. A., MARTINS-SOUZA, R. L. & COELHO, P. M. 2012. Interaction of *Schistosoma mansoni* Sporocysts and Hemocytes of *Biomphalaria*. *J Parasitol Res,* 2012**,** 743920.

NEUMANN, S., ZIV, E., LANTNER, F. & SCHECHTER, I. 1993. Regulation of HSP70 gene expression during the life cycle of the parasitic helminth *Schistosoma mansoni*. *Eur J Biochem,* 212**,** 589-96.

NYSTROM, T. & NEIDHARDT, F. C. 1992. Cloning, mapping and nucleotide sequencing of a gene encoding a universal stress protein in *Escherichia coli*. *Mol Microbiol,* 6**,** 3187-98.

NYSTROM, T. & NEIDHARDT, F. C. 1994. Expression and role of the universal stress protein, UspA, of *Escherichia coli* during growth arrest. *Mol Microbiol,* 11**,** 537-44.

NYSTROM, T. & NEIDHARDT, F. C. 1996. Effects of overproducing the universal stress protein, UspA, in *Escherichia coli* K-12. *J Bacteriol,* 178**,** 927-30.

OJOPI, E. P., OLIVEIRA, P. S., NUNES, D. N., PAQUOLA, A., DEMARCO, R., GREGORIO, S. P., AIRES, K. A., MENCK, C. F., LEITE, L. C., VERJOVSKI-ALMEIDA, S. & DIAS-NETO, E. 2007. A quantitative view of the transcriptome of *Schistosoma* mansoni adult-worms using SAGE. *BMC Genomics,* 8**,** 186.

OSWALD, I. P., ELTOUM, I., WYNN, T. A., SCHWARTZ, B., CASPAR, P., PAULIN, D., SHER, A. & JAMES, S. L. 1994. 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,* 91**,** 999-1003.

OVAKIM, D. H. & HEIKKILA, J. J. 2003. Effect of histone deacetylase inhibitors on heat shock protein gene expression during *Xenopus* development. *Genesis,* 36**,** 88-96.

PALM, D., WEILAND, M., MCARTHUR, A. G., WINIECKA-KRUSNELL, J., CIPRIANO, M. J., BIRKELAND, S. R., PACOCHA, S. E., DAVIDS, B., GILLIN, F., LINDER, E. & SVARD, S. 2005. Developmental changes in the adhesive disk during *Giardia* differentiation. *Mol Biochem Parasitol,* 141**,** 199-207.

PANDIT, S. B. & SRINIVASAN, N. 2003. Survey for g-proteins in the prokaryotic genomes: prediction of functional roles based on classification. *Proteins,* 52**,** 585-97.

PATANKAR, S., MUNASINGHE, A., SHOAIBI, A., CUMMINGS, L. M. & WIRTH, D. F. 2001. Serial analysis of gene expression in *Plasmodium falciparum* reveals the global expression profile of erythrocytic stages and the presence of anti-sense transcripts in the malarial parasite. *Mol Biol Cell,* 12**,** 3114-25.

PENG, H., BATEMAN, A., VALENCIA, A. & WREN, J. D. 2012. Bioimage informatics: a new category in Bioinformatics. *Bioinformatics,* 28**,** 1057.

PICA-MATTOCCIA, L., ORSINI, T., BASSO, A., FESTUCCI, A., LIBERTI, P., GUIDI, A., MARCATTO-MAGGI, A. L., NOBRE-SANTANA, S., TROIANI, A. R., CIOLI, D. & VALLE, C. 2008. *Schistosoma mansoni*: lack of correlation between praziquantel-induced intra-worm calcium influx and parasite death. *Exp Parasitol,* 119**,** 332-5.

PINEDA, M. C., TURON, X. & LOPEZ-LEGENTIL, S. 2012. Stress levels over time in the introduced ascidian *Styela plicata*: the effects of temperature and salinity variations on hsp70 gene expression. *Cell Stress Chaperones,* 17**,** 435-44.

PLEWES, K. A., BARR, S. D. & GEDAMU, L. 2003. Iron superoxide dismutases targeted to the glycosomes of *Leishmania chagasi* are important for survival. *Infect Immun,* 71**,** 5910-20.

PORTER, L. L. & ROSE, G. D. 2011. Redrawing the Ramachandran plot after inclusion of hydrogen-bonding constraints. *Proc Natl Acad Sci U S A,* 108**,** 109-13.

POTTS, A. J., HEDDERSON, T. A. & GRIMM, G. W. 2013. Constructing phylogenies in the presence of intra-individual site polymorphisms (2ISPs) with a focus on the nuclear ribosomal cistron. *Syst Biol*.

RADKE, J. R., BEHNKE, M. S., MACKEY, A. J., RADKE, J. B., ROOS, D. S. & WHITE, M. W. 2005. The transcriptome of *Toxoplasma gondii*. *BMC Biol,* 3**,** 26.

RAGHAVAN, N., MILLER, A. N., GARDNER, M., FITZGERALD, P. C., KERLAVAGE, A. R., JOHNSTON, D. A., LEWIS, F. A. & KNIGHT, M. 2003. Comparative gene analysis of *Biomphalaria glabrata* hemocytes pre-and post-exposure to miracidia of *Schistosoma mansoni*. *Mol Biochem Parasitol,* 126**,** 181-191.

SALMON, D., VANWALLEGHEM, G., MORIAS, Y., DENOEUD, J., KRUMBHOLZ, C., LHOMME, F., BACHMAIER, S., KADOR, M., GOSSMANN, J., DIAS, F. B., DE MUYLDER, G., UZUREAU, P., MAGEZ, S., MOSER, M., DE BAETSELIER, P., VAN DEN ABBEELE, J., BESCHIN, A., BOSHART, M. & PAYS, E. 2012. Adenylate cyclases of *Trypanosoma brucei* inhibit the innate immune response of the host. *Science,* 337**,** 463-6.

SALSBURY, F. R., JR., KNUTSON, S. T., POOLE, L. B. & FETROW, J. S. 2008. Functional site profiling and electrostatic analysis of cysteines modifiable to cysteine sulfenic acid. *Protein Sci,* 17**,** 299-312.

SANCHEZ, Y., TAULIEN, J., BORKOVICH, K. A. & LINDQUIST, S. 1992. Hsp104 is required for tolerance to many forms of stress. *EMBO J,* 11**,** 2357-64.

SAYED, A. A., COOK, S. K. & WILLIAMS, D. L. 2006. Redox balance mechanisms in *Schistosoma mansoni* rely on peroxiredoxins and albumin and implicate peroxiredoxins as novel drug targets. *J Biol Chem,* 281**,** 17001-10.

SCHISTOSOMA JAPONICUM GENOME SEQUENCING CONSORTIUM 2009. The *Schistosoma japonicum* genome reveals features of host-parasite interplay. *Nature. 2009 Jul 16;460(7253):345-51. doi: 10.1038/nature08140.*

SCHLICHTING, I., ALMO, S. C., RAPP, G., WILSON, K., PETRATOS, K., LENTFER, A., WITTINGHOFER, A., KABSCH, W., PAI, E. F., PETSKO, G. A. & ET AL. 1990. Time-resolved X-ray crystallographic study of the conformational change in Ha-Ras p21 protein on GTP hydrolysis. *Nature,* 345**,** 309-15.

SCHREIBER, K., BOES, N., ESCHBACH, M., JAENSCH, L., WEHLAND, J., BJARNSHOLT, T., GIVSKOV, M., HENTZER, M. & SCHOBERT, M. 2006. Anaerobic survival of *Pseudomonas aeruginosa* by pyruvate fermentation requires an Usp-type stress protein. *J Bacteriol,* 188**,** 659-68.

SCHWEIKHARD, E. S., KUHLMANN, S. I., KUNTE, H. J., GRAMMANN, K. & ZIEGLER, C. M. 2010. Structure and function of the universal stress protein TeaD and its role in regulating the ectoine transporter TeaABC of *Halomonas elongata* DSM 2581(T). *Biochemistry,* 49**,** 2194-204.

SCULLY, C. & WARD-BOOTH, R. P. 1995. Detection and treatment of early cancers of the oral cavity. *Crit Rev Oncol Hematol,* 21**,** 63-75.

SEIFART GOMES, C., IZAR, B., PAZAN, F., MOHAMED, W., MRAHEIL, M. A., MUKHERJEE, K., BILLION, A., AHARONOWITZ, Y., CHAKRABORTY, T. & HAIN, T. 2011. 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,* 6**,** e24965.

SHILOVA, V. Y., GARBUZ, D. G., MYASYANKINA, E. N., CHEN, B., EVGEN'EV, M. B., FEDER, M. E. & ZATSEPINA, O. G. 2006. Remarkable site specificity of local transposition into the Hsp70 promoter of *Drosophila melanogaster*. *Genetics,* 173**,** 809-20.

SHOJI, S., PARMELEE, D. C., WADE, R. D., KUMAR, S., ERICSSON, L. H., WALSH, K. A., NEURATH, H., LONG, G. L., DEMAILLE, J. G., FISCHER, E. H. & TITANI, K. 1981. Complete amino acid sequence of the catalytic subunit of bovine cardiac muscle cyclic AMP-dependent protein kinase. *Proc Natl Acad Sci U S A,* 78**,** 848-51.

SHORT, R. B., MENZEL, M. Y. & PATHAK, S. 1979. Somatic chromosomes of *Schistosoma mansoni*. *J Parasitol,* 65**,** 471-3.

SILLERO, A. & MALDONADO, A. 2006. Isoelectric point determination of proteins and other macromolecules: oscillating method. *Comput Biol Med,* 36**,** 157-66.

SILVA, L. L., MARCET-HOUBEN, M., NAHUM, L. A., ZERLOTINI, A., GABALDON, T. & OLIVEIRA, G. 2012. The *Schistosoma mansoni* phylome: using evolutionary genomics to gain insight into a parasite's biology. *BMC Genomics,* 13**,** 617.

SIMAN-TOV, M. M., ALY, R., SHAPIRA, M. & JAFFE, C. L. 1996. Cloning from *Leishmania major* of a developmentally regulated gene, c-lpk2, for the catalytic subunit of the cAMP-dependent protein kinase. *Mol Biochem Parasitol,* 77**,** 201-15.

SMITH, A., WARD, M. P. & GARRETT, S. 1998. Yeast PKA represses Msn2p/Msn4p-dependent gene expression to regulate growth, stress response and glycogen accumulation. *EMBO J,* 17**,** 3556-64.

SOARES DE MOURA, R., ROZENTAL, R. & CLAUDIO-DA-SILVA, T. S. 1987. *Schistosoma mansoni*: effects of bromolysergic acid diethylamide, verapamil, and Ca2+-free solution on the motor activity of the isolated male worm induced by electrical stimulation and oxamniquine. *Exp Parasitol,* 63**,** 173-9.

SOMERO, G. N. 1995. Proteins and temperature. *Annu Rev Physiol,* 57**,** 43-68.

SONG, J. & HE, Q. F. 2012. Bioinformatics analysis of the structure and linear B-cell epitopes of aquaporin-3 from *Schistosoma japonicum*. *Asian Pac J Trop Med,* 5**,** 107-9.

SOUSA, M. C. & MCKAY, D. B. 2001. Structure of the universal stress protein of *Haemophilus influenzae*. *Structure,* 9**,** 1135-41.

STEEN, B. R., LIAN, T., ZUYDERDUYN, S., MACDONALD, W. K., MARRA, M., JONES, S. J. & KRONSTAD, J. W. 2002. Temperature-regulated transcription in the pathogenic fungus *Cryptococcus neoformans*. *Genome Res,* 12**,** 1386-400.

STEINMANN, P., KEISER, J., BOS, R., TANNER, M. & UTZINGER, J. 2006. Schistosomiasis and water resources development: systematic review, meta-analysis, and estimates of people at risk. *Lancet Infect Dis,* 6**,** 411-25.

STOSCHECK, C. M. 1990. Quantitation of protein. *Methods Enzymol,* 182**,** 50-68.

SU, Y., DOSTMANN, W. R., HERBERG, F. W., DURICK, K., XUONG, N. H., TEN EYCK, L., TAYLOR, S. S. & VARUGHESE, K. I. 1995. Regulatory subunit of protein kinase A: structure of deletion mutant with cAMP binding domains. *Science,* 269**,** 807-13.

SWIERCZEWSKI, B. E. & DAVIES, S. J. 2009. A schistosome cAMP-dependent protein kinase catalytic subunit is essential for parasite viability. *PLoS Negl Trop Dis,* 3**,** e505.

SYIN, C. & GOLDMAN, N. D. 1996. Cloning of a *Plasmodium falciparum* gene related to the human 60-kDa heat shock protein. *Mol Biochem Parasitol,* 79**,** 13-9.

SYIN, C., PARZY, D., TRAINCARD, F., BOCCACCIO, I., JOSHI, M. B., LIN, D. T., YANG, X. M., ASSEMAT, K., DOERIG, C. & LANGSLEY, G. 2001. The H89

cAMP-dependent protein kinase inhibitor blocks *Plasmodium falciparum* development in infected erythrocytes. *Eur J Biochem,* 268**,** 4842-9.
SZABO, A., LANGER, T., SCHRODER, H., FLANAGAN, J., BUKAU, B. & HARTL, F. U. 1994. The ATP hydrolysis-dependent reaction cycle of the *Escherichia coli* Hsp70 system DnaK, DnaJ, and GrpE. *Proc Natl Acad Sci U S A,* 91**,** 10345-9.
TAKESONO, A., CISMOWSKI, M. J., RIBAS, C., BERNARD, M., CHUNG, P., HAZARD, S., 3RD, DUZIC, E. & LANIER, S. M. 1999. Receptor-independent activators of heterotrimeric G-protein signaling pathways. *J Biol Chem,* 274**,** 33202-5.
TAMURA, K., DUDLEY, J., NEI, M. & KUMAR, S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol,* 24**,** 1596-9.
TAMURA, K. & NEI, M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol,* 10**,** 512-26.
TAMURA, K., PETERSON, D., PETERSON, N., STECHER, G., NEI, M. & KUMAR, S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol,* 28**,** 2731-9.
TAYLOR, S. S., BUECHLER, J. A. & YONEMOTO, W. 1990. cAMP-dependent protein kinase: framework for a diverse family of regulatory enzymes. *Annu Rev Biochem,* 59**,** 971-1005.
TAYLOR, S. S., KIM, C., CHENG, C. Y., BROWN, S. H., WU, J. & KANNAN, N. 2008. Signaling through cAMP and cAMP-dependent protein kinase: diverse strategies for drug design. *Biochim Biophys Acta,* 1784**,** 16-26.
TAYLOR, S. S., KIM, C., VIGIL, D., HASTE, N. M., YANG, J., WU, J. & ANAND, G. S. 2005. Dynamics of signaling by PKA. *Biochim Biophys Acta,* 1754**,** 25-37.
THOMAS, J. J. & COOK, K. A. 2006. A visual analytics agenda. *IEEE Comput Graph Appl,* 26**,** 10-3.
THOMSON, R. L., TOMAS, G., FORSMAN, J. T., BROGGI, J. & MONKKONEN, M. 2010. Predator proximity as a stressor in breeding flycatchers: mass loss, stress protein induction, and elevated provisioning. *Ecology,* 91**,** 1832-40.
TKACZUK, K. L., I, A. S., CHRUSZCZ, M., EVDOKIMOVA, E., SAVCHENKO, A. & MINOR, W. 2013. Structural and functional insight into the universal stress protein family. *Evol Appl,* 6**,** 434-49.
TOVAR, J., CUNNINGHAM, M. L., SMITH, A. C., CROFT, S. L. & FAIRLAMB, A. H. 1998. Down-regulation of *Leishmania donovani* trypanothione reductase by heterologous expression of a trans-dominant mutant homologue: effect on parasite intracellular survival. *Proc Natl Acad Sci U S A,* 95**,** 5311-6.
UHLER, M. D., CARMICHAEL, D. F., LEE, D. C., CHRIVIA, J. C., KREBS, E. G. & MCKNIGHT, G. S. 1986. Isolation of cDNA clones coding for the catalytic subunit of mouse cAMP-dependent protein kinase. *Proc Natl Acad Sci U S A,* 83**,** 1300-4.
UNIPROT, C. 2011. Ongoing and future developments at the Universal Protein Resource. *Nucleic Acids Res,* 39**,** D214-9.
VAIDYA, M., PATEL, M. & PANCHAL, H. 2012. ProPhyC: Protein PhysicoChemical properties calculator. *International Journal of Computer Science and Management Research,* 1**,** 908-983.
VALENCIA, A. & BATEMAN, A. 2005. Increasing the impact of Bioinformatics. *Bioinformatics,* 21**,** 1.
VAN DER PLOEG, L. H., GIANNINI, S. H. & CANTOR, C. R. 1985. Heat shock genes: regulatory role for differentiation in parasitic protozoa. *Science,* 228**,** 1443-6.

VAN DER WERF, M. J., DE VLAS, S. J., BROOKER, S., LOOMAN, C. W., NAGELKERKE, N. J., HABBEMA, J. D. & ENGELS, D. 2003. Quantification of clinical morbidity associated with schistosome infection in sub-Saharan Africa. *Acta Trop,* 86**,** 125-39.
VAN LEEUWEN, M. A. 1995. Heat-shock and stress response of the parasitic nematode *Haemonchus contortus*. *Parasitol Res,* 81**,** 706-9.
VARSHAVSKY, A. 1997. The N-end rule pathway of protein degradation. *Genes Cells,* 2**,** 13-28.
VELCULESCU, V. E., ZHANG, L., ZHOU, W., VOGELSTEIN, J., BASRAI, M. A., BASSETT, D. E., JR., HIETER, P., VOGELSTEIN, B. & KINZLER, K. W. 1997. Characterization of the yeast transcriptome. *Cell,* 88**,** 243-51.
WALKER, J. E., RUNSWICK, M. J. & SARASTE, M. 1982. Subunit equivalence in *Escherichia coli* and bovine heart mitochondrial F1F0 ATPases. *FEBS Lett,* 146**,** 393-6.
WALSH, D. A., PERKINS, J. P. & KREBS, E. G. 1968. An adenosine 3',5'-monophosphate-dependant protein kinase from rabbit skeletal muscle. *J Biol Chem,* 243**,** 3763-5.
WANG, H., LEI, Z., LI, X. & OETTING, R. D. 2011. Rapid cold hardening and expression of heat shock protein genes in the B-biotype *Bemisia tabaci*. *Environ Entomol,* 40**,** 132-9.
WANG, Z., MARTIN, J., ABUBUCKER, S., YIN, Y., GASSER, R. B. & MITREVA, M. 2009. Systematic analysis of insertions and deletions specific to nematode proteins and their proposed functional and evolutionary relevance. *BMC Evol Biol,* 9**,** 23.
WASMUTH, J., SCHMID, R., HEDLEY, A. & BLAXTER, M. 2008. On the extent and origins of genic novelty in the phylum *Nematoda*. *PLoS Negl Trop Dis,* 2**,** e258.
WASS, M. N., KELLEY, L. A. & STERNBERG, M. J. 2010. 3DLigandSite: predicting ligand-binding sites using similar structures. *Nucleic Acids Res,* 38**,** W469-73.
WEHMEYER, N., HERNANDEZ, L. D., FINKELSTEIN, R. R. & VIERLING, E. 1996. Synthesis of small heat-shock proteins is part of the developmental program of late seed maturation. *Plant Physiol,* 112**,** 747-57.
WEITMANN, S., SCHULTZ, G. & KLEUSS, C. 2001. Adenylyl cyclase type II domains involved in Gbetagamma stimulation. *Biochemistry,* 40**,** 10853-8.
WEITMANN, S., WURSIG, N., NAVARRO, J. M. & KLEUSS, C. 1999. A functional chimera of mammalian guanylyl and adenylyl cyclases. *Biochemistry,* 38**,** 3409-13.
WELTE, M. A., DUNCAN, I. & LINDQUIST, S. 1995. The basis for a heat-induced developmental defect: defining crucial lesions. *Genes Dev,* 9**,** 2240-50.
WILLARD, F. S., KIMPLE, R. J., KIMPLE, A. J., JOHNSTON, C. A. & SIDEROVSKI, D. P. 2004. Fluorescence-based assays for RGS box function. *Methods Enzymol,* 389**,** 56-71.
WILLIAMS, D. L., SAYED, A. A., RAY, D. & MCARTHUR, A. G. 2006. *Schistosoma mansoni* albumin, a major defense against oxidative damage, was acquired by lateral gene transfer from a mammalian host. *Mol Biochem Parasitol,* 150**,** 359-63.
WU, S. & ZHU, Y. 2012. ProPAS: standalone software to analyze protein properties. *Bioinformation,* 8**,** 167-9.
YAMANAKA, G., ECKSTEIN, F. & STRYER, L. 1986. Interaction of retinal transducin with guanosine triphosphate analogues: specificity of the gamma-phosphate binding region. *Biochemistry,* 25**,** 6149-53.
YOUNG, N. D., JEX, A. R., LI, B., LIU, S., YANG, L., XIONG, Z., LI, Y., CANTACESSI, C., HALL, R. S., XU, X., CHEN, F., WU, X., ZERLOTINI, A., OLIVEIRA, G., HOFMANN, A., ZHANG, G., FANG, X., KANG, Y., CAMPBELL, B. E., LOUKAS, A., RANGANATHAN, S., ROLLINSON, D.,

RINALDI, G., BRINDLEY, P. J., YANG, H., WANG, J., WANG, J. & GASSER, R. B. 2012. Whole-genome sequence of *Schistosoma haematobium*. *Nat Genet,* 44**,** 221-5.

ZAREMBINSKI, T. I., HUNG, L. W., MUELLER-DIECKMANN, H. J., KIM, K. K., YOKOTA, H., KIM, R. & KIM, S. H. 1998. Structure-based assignment of the biochemical function of a hypothetical protein: a test case of structural genomics. *Proc Natl Acad Sci U S A,* 95**,** 15189-93.

ZARLEY, J. H., BRITIGAN, B. E. & WILSON, M. E. 1991. Hydrogen peroxide-mediated toxicity for *Leishmania donovani chagasi* promastigotes. Role of hydroxyl radical and protection by heat shock. *J Clin Invest,* 88**,** 1511-21.

ZERLOTINI, A., HEIGES, M., WANG, H., MORAES, R. L., DOMINITINI, A. J., RUIZ, J. C., KISSINGER, J. C. & OLIVEIRA, G. 2009. SchistoDB: a Schistosoma mansoni genome resource. *Nucleic Acids Res,* 37**,** D579-82.

ZHANG, Y. Y., YANG, J., YIN, X. X., YANG, S. P. & ZHU, Y. G. 2012. Arsenate toxicity and stress responses in the freshwater ciliate *Tetrahymena pyriformis*. *Eur J Protistol,* 48**,** 227-36.

ZHOU, A. Q., O'HERN, C. S. & REGAN, L. 2011. Revisiting the Ramachandran plot from a new angle. *Protein Sci,* 20**,** 1166-71.

ZHOU, X. N., WANG, L. Y., CHEN, M. G., WU, X. H., JIANG, Q. W., CHEN, X. Y., ZHENG, J. & UTZINGER, J. 2005. The public health significance and control of schistosomiasis in China--then and now. *Acta Trop,* 96**,** 97-105.

ZIKMANIS, P. & KAMPENUSA, I. 2012. Relationships between kinetic constants and the amino acid composition of enzymes from the yeast *Saccharomyces cerevisiae* glycolysis pathway. *EURASIP J Bioinform Syst Biol,* 2012**,** 11.