UNIVERSITY OF ZULULAND

Molecular cloning, recombinant protein expression and structural characterization of a universal stress G4LZI3 protein from *Schistosoma mansoni*

By

Abiola Fatimah ADENOWO

(201451455)

B.Sc. (Hons) Biochemistry (Lagos State University), M.Sc. Biochemistry (University of Lagos)

Thesis submitted in fulfilment of the requirements for the award of Degree of Doctor of Philosophy (PhD) in the field of Biochemistry

Department of Biochemistry and Microbiology, Faculty of Science and Agriculture

University of Zululand, KwaDlangezwa, KwaZulunatal

Supervisor: Prof. Abidemi Paul KAPPO

March, 2018

ABSTRACT

Human schistosomiasis is a primary tropical disease caused by the larvae of a blood fluke parasite called schistosome. Different species of schistosomes are responsible for various types of schistosomiasis but most of the burden of the disease is attributed to the species *Schistosoma mansoni*. The outcome of *S. mansoni* infection includes damage to vital organs such as the intestine, spleen, bladder, lungs and liver. If left untreated it can also lead to weight loss and growth retardation in children, higher risk of anaemia, infertility, and increase susceptibility to HIV and in extreme cases, cancer. Till date, the frontline drug for the treatment of schistosomiasis is praziquantel which has various limitations and of late, drug resistance. The limitations of this first line drug has made it necessary for alternative interventions for combating the disease in terms of novel drugs and vaccines discovery and development, as well as the design of specific diagnostic tools for early diagnosis.

The genome of *S. mansoni* has been predicted to encode different stress proteins including the Universal Stress Protein (USP). It has been previously hypothesized that the USP of schistosome parasites have shared protein sequence features which can be useful for inferring their biochemical and environmental regulation. Interestingly, proteins that have such shared features can be very useful targets for disease intervention. Universal stress proteins (USPs) act as natural biological defence mechanisms in living organisms. They are overexpressed in stress conditions and assist in overcoming stressors through various mechanisms. USPs has been predicted to play a role in host invasion by pathogens such as *Mycobacterium, Salmonella* and *Klebsiella*; thus may be useful in the treatment of various pathogenic

diseases. The aim of this study is to characterize a novel *Schistosoma mansoni* universal stress protein G4LZI3, which has been hypothesized as a druggable target, as well as a vaccine candidate against the parasite in humans. Bioinformatics analysis of the G4LZI3 gene sequence was done to predict structural information about the protein which served as a guide towards the choice of conditions for further experimental procedure. These experimental procedures were used to determine the purity, the actual molecular weight, secondary structure element and thermostability of the protein. More so, the characterization of the expressed G4LZI3 protein reveals the secondary structural elements of the proteins and above all, a considerable amount of the recombinant protein could be expressed in bacteria and adequately purified. The results showed the possible amenability of the protein to structural determination either by heteronuclear NMR spectroscopy or X-ray crystallography which is vital towards the ultimate goal of developing a new alternative but efficient drug against schistosomiasis.

Keywords

Human schistosomiasis, neglected tropical diseases, schistosomes, *Schistosoma mansoni,* praziquantel, Universal Stress Protein, recombinant protein expression, *in-silico*.

DECLARATION

I declare that ***"*Molecular cloning, recombinant protein expression and structural characterization of a novel universal stress G4LZI3 protein from *Schistosoma mansoni"*** is my own work, it has not been previously submitted to any other institution for any degree or examination. I have adequately referenced and acknowledged all the resources I used.

PhD candidate: Abiola Fatimah Adenowo

Signature: ……

Date: ……

Supervisor: Prof. Abidemi Paul Kappo

Signature: ……

Date: ……

DEDICATION

This thesis is profoundly dedicated to **Almighty Allah**; Lord of infinite mercy and incomparable majesty.

ACKNOWLEDGEMENTS

I give all thanks and glorifications to Allah who out of His mercy made the achievement of this PhD thesis a reality. Indeed, **You** are the **BEST OF PLANNERS.**

I sincerely appreciate the understanding, support, guidance and mentoring of my able supervisor, Prof Abidemi Paul Kappo. Your encouragement and words of advice really made the PhD journey a fruitful one. You are indeed a worthy mentor. I also say thank you to Prof. Andy Opoku for his fatherly concerns and the HOD Biochemistry and Microbiology Department, Prof. Basson for his words of encouragement and support.

I am grateful to Allah for giving me an awesome husband like no other; Dr Abdurrasheed Adetokunbo Abayomi. Your kind is indeed very rare, you encouraged me and supported me throughout this research journey. I appreciate your patience while I was away and for giving me the strength to move on through your words and actions when waning of determination almost set in. My love and respect for you will definitely continue to wax stronger, thank you so much for your love and for believing in me. To my wonderful boys Abdurrahman Abimbola, Abdulhameed Opeyemi and Anas Adebambo; thank you for your patience and endurance while I was away. And to Usamah, Asmau, Muhammad, Abdullah and Mukhtar I appreciate the supports rendered in various ways. I really love you all.

I am really indebted to my parents; Alhaji and Alhaja Afolabi for their care, love and support in all ramification, you were never tired of nurturing me from cradle till now. I appreciate the support, prayers and encouragement of my siblings Toyyibbah and

Khadijah. And my brother Mr Abdurrasheed Afolabi for being there whenever I needed him.

And especially to my elder sister; Mrs Rasheedah Afolabi Abdurrahman and her hubby Dr. Mikail Abdurrahman; words are not enough to appreciate your great assistance, you actually made this PhD training possible by taking care of my children like your own, I know I can never thank you enough.

I appreciate Dr. (Mrs) Ronke Gbemisola Saibu for the link, and Dr. Azeez Ayomide Fatai for the brief but immensely helpful lesson on Bioinformatics. I say thank you Dr. Mutiu Kazeem for his support and encouragement. I appreciate the assistance of Dr. Foluso Osunsanmi for helping out in my admission process and to Dr. Salmon Abdussalaam for his help and homely welcome on getting to University of Zululand.

To my colleagues on the Schistosomiasis project, Dr. Babatunji Oyinloye and Priscila Masamba, I say a very big thank you for your fellowship and cooperation. I can never forget those days trekking together several minutes on the dusty roads of Mafikeng and Venda as well as those days in ‘‘the white house’’. Dr. Oyinloye, our journey together on this PhD started unplanned at the South African Embassy in Lagos, your assistance then and till the end of the journey is very much appreciated.

My gratitude goes to all my Brethren in South Africa; Dr. Abdulhameed Ojugbele, Dr. Ridwan Ayetoro, Mr and Mrs Oni and my sister in faith Rasheedah Adams Ojugbele, you usually give ‘a home away from home’ whenever I come to you in Durban, those were wonderful times spent with you. Thank you all for your supports, prayers and being there for me whenever a need a shelter away from my solitary room in UNIZULU. To Engineer Bashir Bamiro, your technical assistance on the manuscript

is highly appreciated. The role played by my numerous brothers and sisters (M.S.S, Al-Muminat, T.M.C) whose names will fill up the pages if I attempt to make a list is greatly treasured. I say to you all JazakaAllahu Khayran Ketheeran.

Alex Adekiya and Taiwo Aruleba of the BSB group, your timely assistance is immensely appreciated, I see you guys as very great potentials, your dedication to your endeavours and keenness to assist others will certainly take you to very great heights. Mrs Edwina Uzunuigbe, your companionship, words of encouragement and support will always linger in my memory, I wish you success in completion of your PhD research. To all members of the BSB group, Mrs Funmi Akappo, Paul, Thando, Kho, Makhosi, Nthando, Philisiwe, Khuboni and Sbonelo it was nice being with you all.

I appreciate the technical assistance of Prof. A. Shonhai and his research group most especially Dr. Tawanda Zininga and Mr. Stanley Makumire at the University of Venda.

I acknowledge University of Zululand Research Committee for funding this project and National Research Foundation (NRF) South Africa for granting me funding for the third year of my PhD research.

Once more I appreciate the favours of my Creator and Sustainer, truly You are the BEST of PLANNERS.

List of Figures

LIST OF TABLES

ABBREVIATIONS

a. Amino acids - one and three letter codes:

Name of amino acid	Code (three letter)	Code (one letter)
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Asparagine or aspartic acid	Axp	B
Cysteine	Cys	C
Glutamic acid	Glu	E
Glutamic acid	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

b. **Abbreviations and their meaning:**

Acronyms	**Meaning**
®	Registered sign
B	Beta
A	Alpha
M	Micro
~	Approximately
°C	Degrees Celsius
Mg	Microgram(s)
Ml	Microliter(s)
Mm	Micromolar
™	Trade mark sign
A	Absorbance
A_{595}	Absorbance at 595 nm
A_{280}	Absorbance at 280 nm
AA	Amino acid
Abs	Absorbance
APS	Ammonium persulphate
BamHI	*Bacillus amyloliquefaciens*, strain H, I^{st} enzyme
Bp	Base pair
BLAST	Basic local alignment search tool
BSA	Bovine Serum Albumin
C	Carbon atom
CD	Circular dichroism

cDNA	Complementary deoxyribonucleic acid
Camp	Cyclic Adenosine monophosphate
COSY	correlation spectroscopy
CV	Column volume
Da	Dalton
DSC	Differential scanning calorimetry
DTT	Dithiothreitol
E.coli	*Escherichia coli*
EDTA	Ethylenediaminetetra acetic acid
FTIR	Fourier transform infrared
g	Gram(s)
g	Gravitational force
HindIII	*Haemophilus influenzae,* strain d, 3rd enzyme
His6	Hexahistidine
HSP	Heat shock protein
Hz	Hertz
MHz	Megahertz
IPTG	Isopropyl-ß-D-thiogalactopyranoside
IMAC	Immobilized affinity chromatography
Kb	Kilo base pairs
kDa	Kilo Dalton
L	Litres
LB	Luria broth
LC-MS	Liquid chromatography–mass spectrometry

M	Molar
MALDI-TOF	Matrix Assisted Laser Desorption Ionization - Time of Flight.
Min	Minute
ml	Millilitre(s)
mg	Milligram(s)
Mm	Millimolar
mRNA	Messenger Ribonucleic acid
MS	Mass spectrometry
MWCO	Molecular weight cut-off
m/z	Mass/charge ratio
NCBI	National Centre for Biotechnology information
Ni-NTA	Nickel nitrilotriacetic acid
NTD	Neglected tropical diseases
Nm	Nanometer
NMR	Nuclear magnetic resonance
NOESY	Nuclear Overhauser Enhancement Spectroscopy
OD	Optical density
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
pI	Isoelectric point
PMSF	Phenylmethylsulfonyl fluoride
Ppm	Part per million
Rp-HPLC	Reverse phase high performance liquid chromatography

RNA	Ribonucleic acid
rpm	Revolutions per minute
RT	Room temperature
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
TAE	Tris/acetic /EDTA
TE	Tris-EDTA
TEMED	*N,N,N',N'*-Tetra methyl ethylene diamine
Tris	Tris(hydroxylmethyl)aminomethane
USP, Usp	Universal stress protein
UspA	Universal stress protein A
µm	Micrometer
UV-Vis	Ultraviolet-visible
V	Voltage
v/v	Volume to volume ratio
w/v	Weight to volume ratio

General stock solutions, buffers and media based on experimental procedures

Agarose gel electrophoresis reagents and buffers

Ethidium Bromide: 10 mg/ml stock solution was prepared in water, and kept in a dark bottle at 4 °C.

10X TAE: 48.4 g Tris was dissolved in 600 ml of water, 20 ml of 0.5 M EDTA pH 8.0 and 11.44 ml glacial acetic acid was added, volume was made up to 1 L with deionized water.

1X TAE: 10 x TAE diluted 10 folds prior to use.

0.5 M EDTA: 186.1 g of EDTA was dissolved in 800 ml of deionized water, the pH was adjusted to 8.0 with NaOH and volume made up to 1 L.

Preparation of competent cells, transformation

Ampicillin: 100 mg/ml stock solution was made in deionised water and solution was filter-sterilized with a 0.22-micron filter, it was aliquoted and kept at -20 °C.

LB agar plate: 10 g Tryptone powder, 5 g Yeast extract, 5 g NaCl and 14 g bacteriological agar was dissolved in 1 L distilled water and autoclaved. It was left to cool, I ml ampicillin was added and then aseptically poured into petri dishes.

LB broth: 10 g Tryptone powder, 5 g Yeast extract, 5 g NaCl was dissolved in 1 L of distilled water and autoclaved.

0.1 M $MgCl_2$: 10.16 g of $MgCl_2$ was dissolved in 500 ml of distilled water.

0.1 M CaCl2: 2.94 g $CaCl_2$ was dissolved in 500 ml of distilled water.

30 % glycerol stock: 30 ml of 100 % glycerol was mixed with 70 ml of distilled water.

Protein expression

2X YT: 16 g tryptone, 10 g yeast and 5 g NaCl was dissolved in 1 L of distilled water and autoclaved. It was allowed to cool and 4 g glucose of glucose was added to supplement the media after the addition of 1 ml ampicillin.

1 M IPTG: 2.38 g of IPTG was dissolved in 8 ml of distilled H_2O, solution was made up to a final volume of 10 ml with distilled water. It was filter sterilized with a 0.22 µ syringe filter and stored as 1 ml aliquots at -20 °C.

PBS: 8g NaCl, 0.2 g KCl, 1.44 g Na_2HPO_4 and 0.24 g KH_2PO_4 was dissolved in 800 ml of distilled water, the pH was adjusted to 7.4 with HCl and solution was made up to 1 L and then sterilized by autoclaving.

SDS Polyacrylamide gel electrophoresis

2X SDS PAGE Sample Buffer: This was prepared with 10 % SDS, 1 M Tris-HCl pH 6.8, 2 ml Glycerol, 500 µl Bromophenol Blue, 2.5 ml 100 mM DTT made up to 10 ml with distilled water. The buffer was kept at room temperature.

10 % SDS: 10 g of SDS was dissolved in 80 ml of deionized water and then made up to 100 ml.

5X SDS-PAGE Electrophoresis running buffer: 25 mM Tris-HCl, 0.1 % SDS and 250 mM Glycine pH 8.3. The buffer was kept at room temperature and diluted 5-times when needed.

Ammonium Persulphate: 10 % stock solution of this was made in deionised water and solution was kept at 4 °C.

Coomassie Staining Solution: This was constituted with 0.25 g Coomassie Blue R-250, 40 % methanol and 10 % acetic acid in 2 L of deionised water.

Destaining Solution: This was prepared with 10 % (v/v) acetic acid and 40 % methanol (v/v).

Separating Buffer: 1.5 M Tris-HCl was adjusted to pH 8.8 with HCl. The buffer was stored at 4 °C.

Stacking Buffer: 0.5 M Tris-HCl was adjusted to pH 6.8 with HCl. The buffer was stored at 4 °C.

Cell extraction and purification

1 M imidazole: 68.077 g of imidazole was dissolved in 1 L of deionized water and solution was kept at 4 °C.

1 M NaCl: 58.44 g of NaCl was dissolved in 1 L of deionized water and solution was kept at room temperature.

1 M Tris: 121.14 g of Tris powder was dissolved in 800 ml dd$H_2$0, pH was adjusted to 8.0 Solution was made up to 1 L, autoclaved and kept at room temperature.

10 mM PMSF: 0.0174 g of PMSF was dissolved in 10 ml isopropanol, it was aliquoted and stored at -20 °C.

100 µg/ml Lysozyme: 10 g was dissolved in 100 ml of dd$H_2$0, it was aliquoted into 1.5 ml Eppendorf tubes and kept at 4 °C.

Cell lysis buffer: This was made up of 100 µg/ml lysozyme, 10 mM PMSF, 50 mM Tris-HCl, 100 mM NaCl, 10 mM imidazole and kept at 4 °C.

1 M DTT: 1.5 g was dissolved in 8 ml of deionized water, volume was made up to 10 ml and then aliquoted into 1.5 ml Eppendorf tubes and stored at -20 °C.

10 % NaN_3: 10 g was dissolved in 100 ml of deionized water and stored at room temperature.

Buffer A (equilibration and wash buffer): This was made with 50 mM Tris pH 8.0 and 40 mM imidazole. Solution was stored at 4 °C.

Buffer B: This was made with 50 mM Tris pH 8.0, 300 mM imidazole. Solution was stored at 4 °C.

Protein concentration determination (Bradford assay)

Coomassie®protein assay reagent: 1:4 dilutions were made by mixing one-part reagent in 4-part deionized water.

BSA: 2 mg/ml stock solution was made by dissolving 0.02 g of BSA powder in 10 ml of deionized water.

Bacterial strains used

E. coli JM109 cells (Promega)

Genotype: *F´ traD36 proA$^+$B$^+$ lacI^q Δ(lacZ)M15/ Δ(lac-proAB) glnV44 e14$^-$ gyrA96 recA1 relA1 endA1 thi hsdR17*

CHAPTER ONE

GENERAL INTRODUCTION AND LITERATURE REVIEW

1.1 Background of study

Schistosomiasis is a disease caused by helminthic blood-dwelling parasitic worms belonging to the genus *Schistosoma* (Chitsulo *et al.,* 2000). It was first described by Theodore Maximillian Bilharz, a German pathologist. After performing autopsies on diseased Egyptian patients in 1851, he discovered the schistosome worms in the portal system as well as in the bladder. He described the schistosomal eggs with their unusual pointy terminal projection. This prime tropical disease was named bilharziasis after him (Chitsulo *et al.,* 2000; Coon, 2005). Schistosomiasis is an old disease of human which poses boundless economic and public health threats to various countries of the world. Although the global distribution of the disease has changed over many decades due to various successful control and eradication projects, the disease persists and is still endemic in many countries most especially the sub-Sahara Africa (Chitsulo *et al.,* 2000; Engels *et al.,* 2002; Madinga *et al.,* 2015; Bawa *et al.,* 2016). The impact of human schistosomiasis in the sub-Sahara Africa is presented in more detail in chapter two of this thesis.

1.2 Pathology and Morbidity of Schistosomiasis

This parasitic infection has two distinct immunological phases with different symptomatic presentations referred to as acute and chronic schistosomiasis (Caldas *et al.,* 2008; Olveda *et al.,* 2013).

1.2.1 Acute schistosomiasis

Acute schistosomiasis otherwise called Katayama fever is presented as an early symptom of infection by parasitic schistosomes in people without immunity who are exposed to cercariae infested water in endemic areas.

The early manifestation of the infection includes a temporary urticarial rash that could persist for many days especially after primary infection as is seen in migrants and tourists. A febrile stage associated with abdominal/pulmonary symptoms, which may develop within 2-10 weeks of infection, also characterizes the Katayama fever. The symptoms of acute schistosomiasis are suggested to be an outcome of allergic responses to antigens released at the time of larval migration and early oviposition. Symptoms of this condition are often mild and may sometimes be unnoticed. These symptoms include fever, anorexia, diarrhoea as well as arthralgia along with eosinophilia and leucocytosis (Bottieau *et al.,* 2006; Caldas *et al.,* 2008). Immunological disorders are usually established at the initial phase of schistosomal infections. More so, neurological symptoms are seen in some people infected with the disease. The disease does not have a specific treatment but steroids are administered to alleviate symptoms. Animal studies had shown that artemisinin derivatives could prove effective in the early stage of the disease and could be beneficial for treatment (Hatz, 2005).

The Katayama fever is common among expatriates and tourists who are unaware of the dangers of being infected by schistosomiasis parasites in the pools and other water bodies they come into contact with during recreational activities. Majority of cases of the disease recorded at the Western travel clinics were brought in from African countries. Widely known sources of infection in sub-Saharan Africa include Lake Malawi, Lake Volta, Lake Victoria, Niger and Zambezi deltas as well as from lakes in South Africa (Jelinek *et al.,* 1996).

1.2.2 Chronic schistosomiasis

At the chronic phase, mature schistosome infection are related with severe local inflammatory responses to parasite eggs entrapped inside the tissues of the infected person, leading to chronic morbidity, inflammation of some organs, obstructive diseases, and tissue fibrosis (Olveda *et al.,* 2013). The chronic forms of schistosomiasis are discussed below.

1.2.2.1 Intestinal schistosomiasis

Intestinal schistosomiasis is caused by *Schistosoma mansoni*. Eggs of schistosomes moving across the walls of the human intestine instigate pseudopolyposis, tumours of the mucous membrane, microulceration and superficial bleeding. Symptoms of the intestinal schistosomiasis include severe or intermittent abdominal ache and distress, diarrhoea with or without blood and loss of appetite. The features mentioned cannot be ascribed unambiguously to schistosome infection in individuals with multiple infections as is seen in places with high endemicity (Montes *et al.,* 2004; Gryseels *et al.,* 2006). Schistosomal appendicitis is a rare condition in some communities with a prevalence of 0.02 % - 6.30 % representing 28.60 % of the appendicitis occurrence in endemic regions (Elbaz and Esmat, 2013).

1.2.2.2 Hepatic schistosomiasis

Hepatic schistosomiasis, otherwise called schistosomal hepatopathy, is a very rampant type of chronic disease resulting from infection by *S. mansoni*. It is an outcome of the host granulomatous cell-mediated immune reaction to the parasites' soluble egg antigen, which advances to the irreversible periportal fibrosis and eventually causing severe portal hypertension.

The fibrotic portal pipe resulting from the hepatic schistosomal infection resembles clay pipe stems passed through the liver and it is referred to as a Symmers' pipe portal fibrosis. Advanced stages of the disease have presentations such as muscle wasting, hypoalbuminemia, ascites as well as coma (Chevillard *et al.,* 2003; Elbaz and Esmat, 2013).

1.2.2.3 Urinary schistosomiasis

Schistosoma haematobium is responsible for urinary schistosomiasis which is endemic in several regions of the world, but more in the sub-Saharan African nations. An estimate of about 436 million persons are in danger of being infected with urinary schistosomiasis in this region (Deribe *et al.,* 2011). Early symptoms are dysuria, pollakisuria and most importantly haematuria. Later symptoms include proteinuria, calcification of the bladder, uretic obstruction, bacterial infection of the patient's urinary tract, hydronephrosis, as well as renal failure. Structural aberrations of the urinary tract might be found in infected children (Gryseels *et al.,* 2006; Gray *et al.,* 2011). In endemic areas, haematuria is the first signal of schistosomiasis in young children of age range 5-10 years but unfortunately it is usually mistaken to be the onset of menstruation in young girls and beginning of puberty in boys. In extreme cases of infection, the urine can be completely dark in colour. Medical surveillance and post-mortem examinations show that several individuals, especially aged patients, die due to kidney damage caused by schistosomal infection (Gryseels *et al.,* 2006).

1.2.2.4 Female genital schistosomiasis

Eggs of *S. mansoni* and *S. haematobium* can be entrapped in the genital tracts of females causing female genital schistosomiasis (FGS) which is accompanied by

genital itches, pelvic discomfort, spot bleeding, and pain during sex, abnormal vaginal discharges, increased frequency of urination and involuntary discharge of urine. Cases of increased abortion, ectopic pregnancy and infertility have also been reported in women infected with the disease. The peripheral blood mononuclear cells in women infected with *S. mansoni* expressed higher HIV co-receptors, an indication that the wounds around the parasites' eggs deposited in genital mucous membrane is a good access point for the HIV virus (Kjetland *et al.,* 2012).

1.3 Schistosomiasis and co-infection

Schistosomiasis opens the door to other opportunistic infections and devastating health conditions. Schistosomiasis is one of the major sources of pulmonary hypertension globally particularly in regions that has high prevalence of schistosomiasis (Schwart *et al.,* 2000; Lapa *et al.,* 2009). In sub-Saharan Africa *S. haematobium* is a significant source of serious urinary tract diseases and dysuria. An estimated 18 million people undergo pains resulting from bladder wall pathology and 10 million people endure hydronephrosis due to infection by the schistosome species (van Der Werf *et al.,* 2003).

There is an established association between inflammation caused by schistosome infection and bladder squamous carcinoma especially in the Middle East and Africa. The International Agency for Cancer Research classfied *S. haematobium* as a carcinogen. *Schistosoma mansoni* and *S*. *japonicum* are linked to the colorectal, liver and cervical carcinomas (Samaras *et al.,* 2010; De Martel *et al.,* 2012).

Accumulation of schistosome eggs in the bladder instigates extreme inflammatory responses which are associated with the discharge of oxygen-derived free radicals, which help in the generation of carcinogenic N-nitrosamines, and eventually

malignant transformation. The formation of N-nitrosamine might also be related to protracted bacterial infection in the urinary bladder caused by Schistosoma infection (Zaghloul, 2012).

Urogenital schistosomiasis is an important agent for the transmission of Human Immunodeficiency Virus. The occurrence of this form of schistosomiasis in HIV positive patients increases the chances of transmission to their sexual partners. Additionally, it hastens the advancement of HIV in persons who already have HIV through the rapid increase in the viral load (Feldmeier *et al.,* 1994; Mbabazi *et al.,* 2011). Several studies indicate that there is a higher progress of liver fibrosis in individuals co-infected with schistosomiasis and hepatitis-C virus compared with cases of single infection. This eventually leads to progressive liver diseases (Angelico *et al.,* 1997; Halim *et al.,* 1999; El-Awady *et al.,* 2006).

Studies have shown that schistosomiasis can cause greater risk of anaemia, delay in growth, lethargy, faintness, diminishing of memory and mental reasoning in children. These negative outcomes of the disease lead to very low academic performance, thereby restraining the future prospects of such children (Gray *et al.,* 2011; Butler *et al.,* 2012), hence further adding to the socioeconomic problem (Butler *et al.,* 2012). Several studies established a link between anaemia and the three main species of schistosomes afflicting humans (Ajanga *et al.,* 2006; Koukounari *et al.,* 2006; Mahgoub *et al.,* 2009). Various mechanisms have been proposed to explain how schistosome infections lead to anaemia in humans. Anaemia might result due to blood loss during the movement of *S. mansoni* eggs through the human intestinal wall. It was also suggested that autoimmune haemolysis leading to decreased lifespan of red blood cells could be responsible for anaemia in schistosomal infections. High levels of proinflammatory cytokines in the serum due to infection

have also been proposed as likely mechanism for occurrence of anaemia. The pro-inflammatory cytokines lead to loss of appetite (anorexia) and thus a reduced intake of micronutrients needed for blood formation (Friedman *et al.,* 2005).

1.4 Schistosomiasis and pregnancy

Numerous researches on animal models as well as human subjects showed negative effects of Schistosoma infections on pregnancy outcomes. *Schistosoma haematobium* has been associated with inflammation of the placental leading to poor birth-outcome caused by placental incompetency. Substantial *S. mansoni* infection causes a higher threat of anaemia subsequently leading to maternal death or low birth weight offspring. The anaemia might be caused by loss of iron into excretory products. Pro-inflammatory cytokines ensuing from schistosomiasis infection cause anorexia in expectant women leading to reduced weight gain (Friedman *et al.,* 2007).

1.5 Diagnosis of schistosomiasis

Examination of stool and urine with a microscope is commonly used for identifying schistosomiasis. The eggs are easy to identify on microscope because of their unusual size and shape and presence of a terminal spine. It is necessary to wait two months after freshwater contact before checking for the eggs because infected patients require that period to start producing eggs (Gryseels *et al.,* 2006; Gray *et al.,* 2011).

The Kato-Katz parasitological technique is highly recommended for the diagnosis of schistosome infection due to it being quantitative, economical, and simple, but the sensitivity of this technique decreases with low intensity and prevalence of the disease. These methods are not very efficient in regions of low endemicity, in post-

treatment circumstances and in control of transmission (Rabello *et al.,* 2002; Pontes *et al.,* 2003). The Kato-katz method or the faecal thick smears are usually used on the field because they enable quantification of the infection by egg counts which is usually expressed as per gram faeces (Gryseels *et al.,* 2006).

Antibody-based assays are perfectly sensitive but do not have the ability to make a distinction between past and current infections. This method is also relatively non-specific and liable to cross-reactions with other antigens. Additionally, the intensity of the infection is not revealed in antibody titres (Rollinson *et al.,* 2013). Antigen-based assays are highly specific but relatively insensitive (Pontes *et al.,* 2003). They are however very relevant in the diagnosis of migrants, travellers and other people occasionally exposed to infection (Bottieau *et al.,* 2006). These assays are also important for incidence studies in children and in post-control or low transmission settings (Gryseels *et al.,* 2006).

The use of a dipstick for diagnosis of schistosome infection in the circulating anodic antigen (CAA) and circulating cathodic antigen (CCA) present in the blood or urine of infected persons is also available (Rollinson *et al.,* 2013). This method however, is not sensitive to light infections and is therefore not very useful for clinical diagnosis (Coulibaly *et al.,* 2013). It is however a specific, stable and direct method for measurement of worm burdens thus making it an important research tool for therapeutic and epidemiological studies (Gryseels *et al.,* 2006).

Cystoscopy and endoscopy are used to visualise bladder lesions and oesophageal varices in hospital settings. Laparoscopy and wedge bioscopy are useful for revealing the macroscopic and histological appearance of periportal fibrosis or granulomatous inflammation (Hayashi *et al.,* 2000; Gryseels *et al.,* 2006). Computed

tomography scan as well as MRI scan of the brain and spinal cord are also important for the detailed visualization of lesion, especially for neuroschistosomiasis. In patients with acute schistosomiasis, it is necessary to carry-out an abdominal and pelvis CT scan as well as chest radiograph (Jenkins-Holick and Kaul, 2013).

A specific and highly sensitive polymerase chain reaction assay has been discovered to test for schistosoma DNA in faeces or blood at all stages of clinical disease including Katayama fever (Gray *et al.,* 2011). The PCR assay requires further standardization before it can be used for routine diagnosis. Application of PCR assay requires high-tech laboratory infrastructure and a careful handling of samples, kits, and equipment. At the moment it has limited broad-scale field applicability (Rollinson *et al.,* 2013).

1.6 Preventive chemotherapy for schistosomiasis and its associated drawback

1.6.1 Artemisinin derivatives

Artemisinin derivatives are antimalaria agents with an anti-schistosomal action. Details are presented in the publication in Chapter Two of this thesis.

1.6.2 Oxamniquine

Oxamniquine (1,2,3,4-Tetrahydro-2-[(isopropylamino)methyl]-7-nitro-6-quinolinemethanol) is effective for killing *S. mansoni* worms but has no effect on *S. haematobium*. Three oral ingestion each of 20 mg/kg of the drug are required for a cure rate of 80-90 %, this percentage cure rate is influenced by geographic region. It is considered that this drug experiences esterification by a sulfotransferase in sensitive Schistosomal parasites, the ester formed dissociates and give rise to an electrophilic reactant with the capability of alkylating the parasite DNA leading to

inhibition of the synthesis of both DNA and RNA. Resistance to this drug has emerged probably as a result of mutation in the parasite gene encoding the esterifying enzyme (Berge *et al.,* 2011; El Ridi and Tallima, 2013) The drawback of this drug includes its severe side effects, high cost and not being commercially available in some countries, thus making the drug out of reach of the poor (Berge *et al.,* 2011).

1.6.3 Praziquantel

Randomised controlled trials established that praziquantel ($C_{19}H_{24}N_2O_2$), is a safe oral drug for the cure of schistosomiasis triggered by the numerous schistosome species (Doenhoff *et al.,* 2008; Doenhoff *et al.,* 2009). Details are presented in Chapter Two.

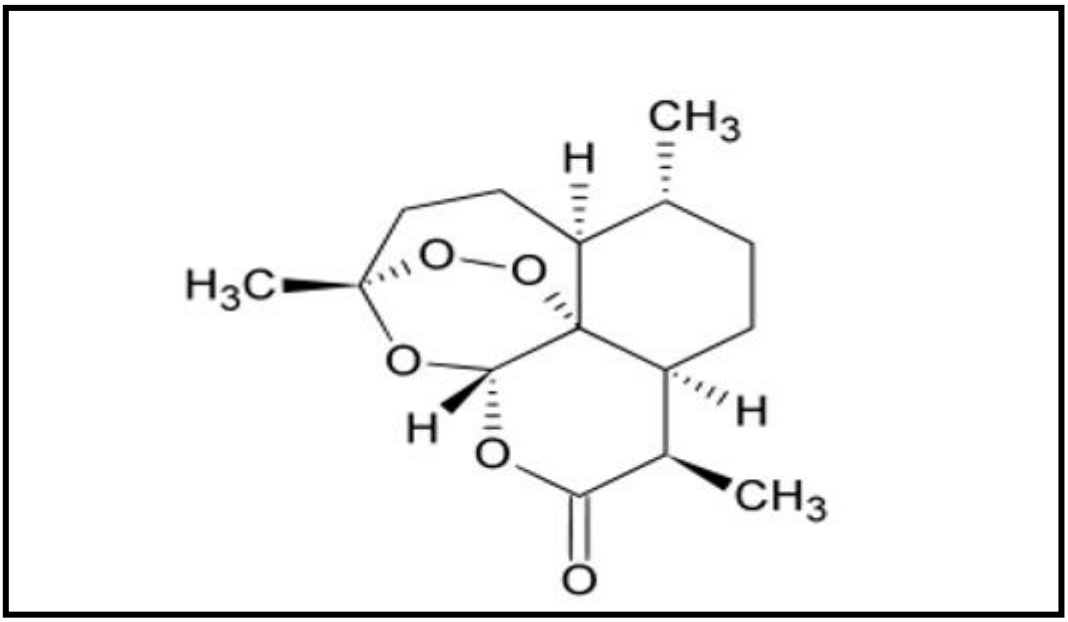


Fig 1.1: The chemical structure of Artemisinin. An antimalarial with anti-schistosomal properties.

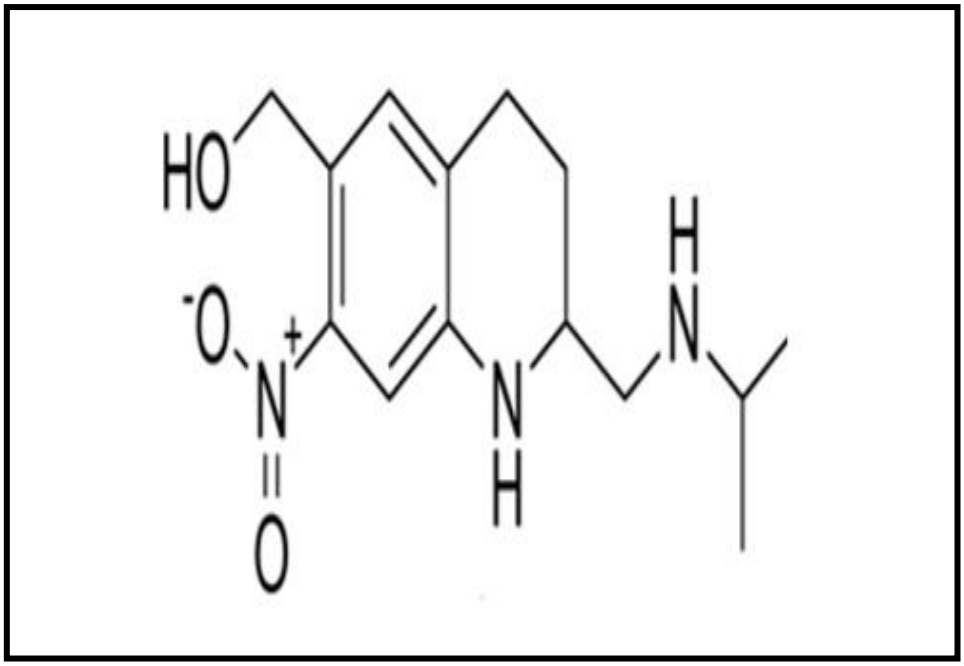


Fig 1.2: The chemical structure of oxamniquine (C15H22O5). A drug effective against the schistosomal worm.

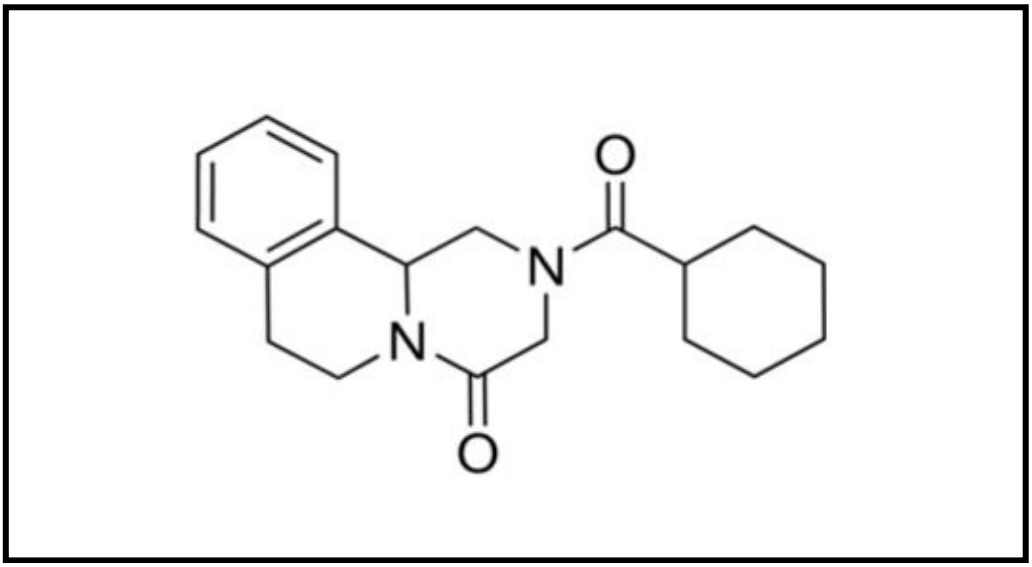


Fig 1.3: The chemical structure of praziquantel ($C_{19}H_{24}N_2O_2$). Currently the most effective drug against all forms of schistosomiasis.

1.7 Mechanism of action of Praziquantel

Despite the fact that praziquantel has been administered in humans for the treatment of various human parasitic diseases including schistosomiasis for more than twenty years, its exact mode of action is yet to be precisely clarified. However, as shown in Fig. 1.4, it has been proposed that praziquantel interrupt calcium ions homeostasis in adult schistosome worms (Aragon *et al.,* 2009). It has been demonstrated that as soon as adult male *Schistosoma mansoni* get in contact with praziquantel *in vitro*, the worm experiences an extreme muscular paralysis along with swift flow of calcium ions, a reduced flow of sodium ions and a slower flow of potassium ions leading to a disturbance of inorganic ion homeostasis in the worm (Aragon *et al.,* 2009). The drug has also been reported to cause rapid disruption of the worm tegument resulting in the uncovering of antigens on the schistosome parasite surface; this was also linked to the interruption of the calcium ion homeostasis (Redman *et al.,* 1996). This structural disorder in the schistosome leads to a detachment of the worm from the human intestine (Aragon *et al.,* 2009).

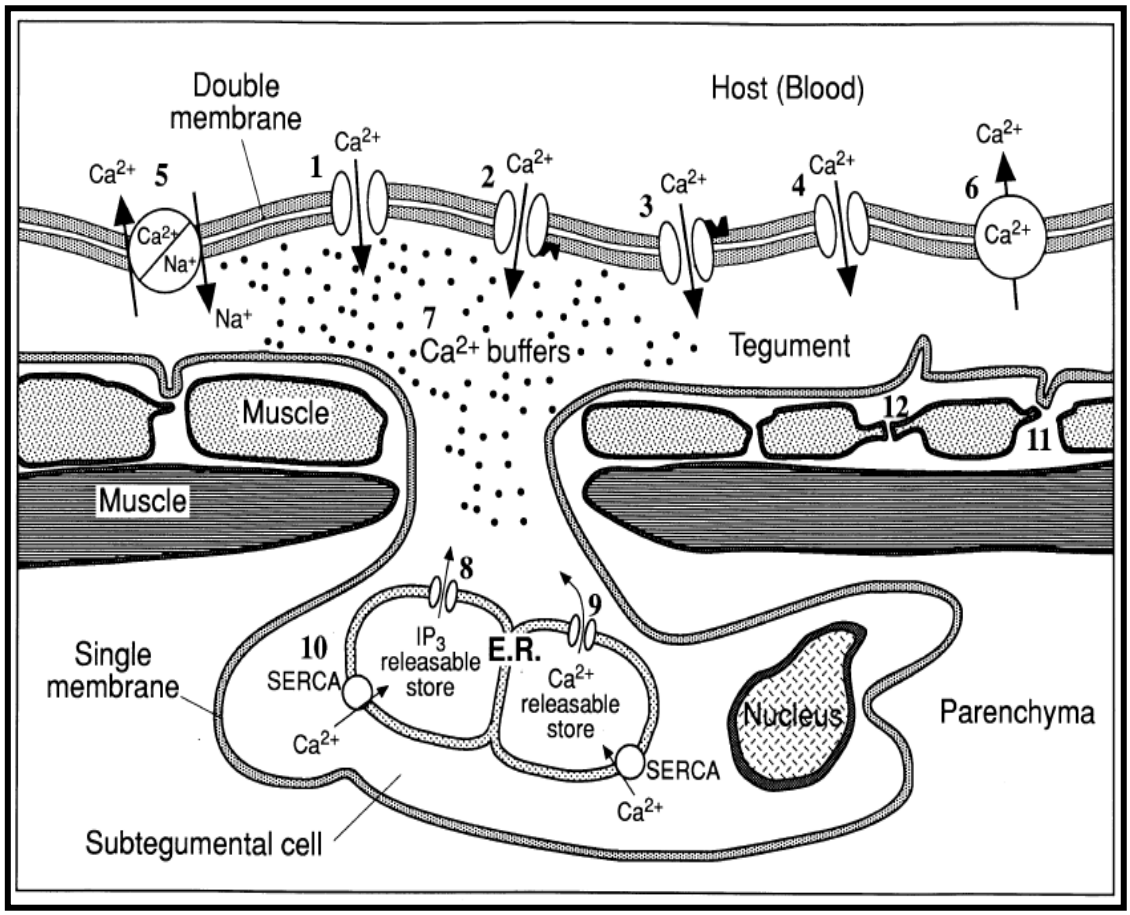


Fig. 1.4 Structure of the body wall of *S. mansoni* showing the likely sites of action of praziquantel and mechanisms of Ca^{2+} transport into and out of the tegument. (Figure taken from Martin, 1997). 1: voltage-activated Ca^{2+} channel. 2: Intracellular messenger activated Ca^{2+} channel. 3: Extracellular receptor operated Ca^{2+} channel. 4: Non-selective cation channel also allowing entry of Ca^{2+}. 5: Na^{+}/Ca^{2+} exchanger. 6: Ca^{2+} ATPase pumping out Ca^{2+}. 7: Intrategumental Ca^{2+} buffers. 8: IP_3 releasable store from the sarcoplasmic reticulum. 9: Ca^{2+}-induced Ca^{2+} release channel (CICR channel). 10: Ca^{2+} ATPase pump: sarcoplasmic endoplasmic reticulum Ca^{2+} (SERCA). 11: Electrical junction between muscle cell and tegument. 12: Electrical junctions between muscle cells. If praziquantel acts like caffeine it would act on site 9.

1.8 Resistance of praziquantel, inability to prevent re-infection and need for alternative interventions

Various studies have revealed the resistance of schistosomes to treatment with praziquantel in various geographical foci (Fallon *et al.,* 1996; Doenhoff *et al.,* 2002; Fenwick and Webster, 2009; Melman *et al.,* 2009). Melman and co-workers (2009) established the existence of the Kenyan isolate of *S. mansoni* from an individual who never got healed from schistosomiasis despite repeated treatment with praziquantel (Melman *et al.,* 2009). Furthermore, Ismail and co-workers (1999) in Egypt studied twelve isolates of schistosome parasite from patients who remained uncured after chemotherapy; they established that these isolates were not responsive to treatment with praziquantel (Ismail *et al.,* 1999). Also, Fallon and co-workers (1996) studied schistosome isolates from Senegal, and discovered that the isolates had reduced susceptibility to treatment with praziquantel compared to isolates from another geographical area (Fallon *et al.,* 1996).

Additionally, it has been established that praziquantel does not prevent re-infection hence, the need for continuous administration of the drug annually in areas of high endemicity (N'Goran *et al.,* 2001; Tchuenté *et al.,* 2013). There is a universal agreement amid professionals in the field that a resilient and continuous decline in the disease spectrum and transmission will be achieved through a long-standing protection by means of vaccination connected with chemotherapy (Siddiqui *et al.,* 2011). The limitations of praziquantel call for the need to research into alternative interventions for the treatment of the disease in term of novel drugs, vaccines as well as specific diagnostic tools for early diagnosis of schistosomiasis.

1.9 Parasitic recombinant proteins as target for disease intervention

Availability of genome sequences of numerous parasites of medical significance has led to exponential advancement in the understanding of the parasite's biology and facilitated the identification of their potential as target in disease intervention (Fernandez-Robledo and Vasta, 2010). Various studies have established the potentials of parasitic stress proteins such as the Heat shock proteins (HSPs) and Small heat shock proteins (sHSP) in various disease interventions, especially diseases involving protein folding (Maloney and Workman, 2002; Sun and MacRae, 2005; Pesce *et al.,* 2010; Shonhai, 2010). An immunological intervention presented as schistosome vaccine is essential for conferring protection against schistosomiasis. A vital strategy in vaccine development for a disease is to identify a molecule critical for the parasites' survival with the anticipation that immunization with this molecule will give protection against the disease (Schechtman *et al.,* 2001). Production of recombinant proteins remains an essential tool in biotechnological research, it is valuable for understanding the pathogenesis of a disease as well as host susceptibility to disease, and can function as vaccines components. Recombinant proteins are peptides with well-known sequences hence, they are useful to increase the specificity of diagnostic tests (Cardoso *et al.,* 2007).

A significant number of biological agents presently in usage are recombinant proteins used as target effector molecules in cancer immunotherapies, as well as those in the management of haemophilia (Sodoyer, 2004). Biophysical and structural characterization of recombinant proteins is a very vital biological prerequisite before recombinant proteins are utilized for the downstream

pharmacological processes like production of vaccines, antibodies as well as new drugs. Modern characterization entails subjecting recombinant proteins to analytical procedures such as circular dichroism, FTIR, UV-visible spectroscopy, dynamic scanning calorimetry, mass spectrometry, NMR spectroscopy, X-ray crystallography amongst others (Hermeling *et al.,* 2005; Bertucci *et al.,* 2011). Advances in development of vaccine for schistosomiasis is discussed further in the second chapter of this thesis.

1.10 Genomic information on the schistosome parasites

The availability of genome sequences of many parasitic organisms of medical significance has facilitated development in the understanding of the parasite's biology thus, enabling the identification of prospective targets for disease intervention. Moreover, a sustained progress in technologies enabling genome sequencing provides great hope for obtaining genomic information on any parasitic organism of interest (Fernandez-Robledo and Vasta, 2010). Notably, the current development of high throughput sequencing in biology witnessed accomplishment of the genomes of the three main schistosome species and other parasite species. These novel sequence resources have significantly extended the usefulness of the schistosome proteomics research, chiefly for the detection of new drug, vaccine and diagnostic biomarkers (Driguez *et al.,* 2016).

Berriman and co-workers (2009) determined the genome sequence of *Schistosoma mansoni* using genome shotgun sequencing. The *S. mansoni* genome sequence was assembled into 5 745 scaffolds, which is greater than 2 kb and totalling 363 Mb. Though 40 % of the *S. mansoni* genome is repetitive, however 50 % was assembled into scaffolds having at least 824.5 kb. More so, 43 % of the genome assembly

which was distributed over 153 scaffolds was unambiguously assigned to chromosomes using florescence *in situ* hybridization (FISH).

The *S. mansoni* chromosome possess seven autosomal and ZW sex determination pairs (Berriman *et al.,* 2009). They also identified 11, 809 putative genes which encodes 13, 197 transcripts. The average gene size was 4.7 kb. Usually, in large introns the average is 1, 692 bp while in smaller exons it is averagely 217 bp. Additionally, the introns in *S. mansoni* genome displayed a significantly biased distribution which has never been discovered in other eukaryotic organisms wherein the 5′ introns are smaller in size compared with the 3' introns. In genes that are multi-exon, the leading set of introns could be very small like 26 bp but the introns near the 3′ end are usually kilobases in length with the largest being 33.8 kb. However, the cause of this observation is uncertain though transcriptional control was suggested. Genes which encode protein kinases, G-protein couple receptors, voltage-gated ion channels, neuropeptides, proteases as well as ligand-gated ion channels are reported to be present in the *S. mansoni* genome. These proteins are considered as vaccine as well as drug targets (Berriman *et al.,* 2009).

Isokpehi and co-workers (2011b) studied the eight genes which encodes the USP domain in the genome of *S. mansoni*. The following locus identifiers are given to the genes; Smp_136870, Smp_043120, Smp_097930, Smp_136890, Smp_001010, Smp_001000, Smp_031300, and Smp_076400 (Isokpehi *et al.,* 2011b). Mbah and co-workers (2013) hypothesized that the USP of schistosome parasites possess some shared protein sequence features that could be helpful for inferring their biochemical as well as environmental regulation, proteins that possess such shared features can be useful targets for disease intervention (Mbah *et al.,* 2013).

The researchers used bioinformatics and visual analytic methods to determine protein sequence length, binding sites for ligands, protein domain length, chemical ligands that are significant biologically, enzymatic regulation, subcellular localization as well as developmental regulation of thirteen schistosome USP sequences (8 for *S. japonicum* and 5 for *S. mansoni*). All the analysed USP sequence have the ATP-binding motif G-2x-G-9x-G(S/T).

1.11 Living organisms and stress

Stress as a concept has been the topic of many scientific discussions since it was first used in scientific research by Selye Hans in 1950. It was initially defined as a non-specific response of an organism to all harmful stimuli. The concept was later redefined to distinguish between stressor and stress response (Koolhaas *et al.,* 2011). In order to survive, all living organisms must uphold a complex homeostasis which is consistently challenged by stressors. These stressors include internal and external antagonistic forces which threaten the existence of the organism while the stress response refers to an organism's reaction towards restoring homeostasis, (Chrousos, 2009; Koolhaas *et al.,* 2011). Examples of stressors include remote temperature, toxic chemicals and gases, nutrient starvation, hydrostatic pressure, high salinity, radioactive substances, drought, heavy metals, chemotherapeutic agents, oxidative agents, nitrosative substances, amino acid analogue, human immune response, and inhibitors of metabolism (Sousa and McKay, 2001; Kerk *et al.,* 2003; Nachin *et al.,* 2005; Mbah *et al.,* 2013).

When living organisms undergo stress, the final effect is protein denaturation and thereafter, the instigation of stress response mechanism. Activation of different signalling pathways inside the cell leads to the up-regulation of genes and proteins

whose activity helps in responding to the various stress situations, thus alleviating the effect of the stress. (Bukau *et al.,* 2000; Mahajan *et al.,* 2005; Wang *et al.,* 2011; Kim *et al.,* 2012). Extreme stress experienced by living organisms induces a loss in the protein native conformation leading to misfolding and aggregation (Somero, 1995; Calamini and Morimoto, 2012). Protein damage is further aggravated by mutations and polymorphism, error prone synthesis as well as post-translational modifications. It must be noted that in order for all proteins to ensure the desired and optimal cellular functioning they must maintain their native conformation. This is achieved by protein homeostasis otherwise called Proteostasis Network (PN). The PN comprises of conserved stress protein pathways which control protein synthesis, protein folding and trafficking as well as degradation to jointly achieve stability and functional attributes of the proteome (Calamini and Morimoto, 2012).

1.12 Stress proteins discovered in living organisms

1.12.1 Heat shock proteins

Heat shock proteins also called heat stress proteins are greatly conserved molecules which are present constitutively or induced in all species of living organisms. Heat shock proteins (HSP) are grouped into families and given nomenclature based on their molecular weight. HSPs are present in different intracellular compartments and function as molecular chaperones in physiological conditions. HSPs make up 5-10 % of the protein constituent of living organisms in normal growth conditions (Pockley, 2003). However, the intracellular concentration of HSPs may be increased by stressors that instigate protein aggregation, unfolding or mis-folding as well as the flow of the newly made non-native proteins. HSPs are not only expressed during temperature rise as the name indicates but also during exposure to other diverse

environmental stressors such as nutritional deficiencies, chemical toxicity, viral infection, oxidative stress and ultraviolet irradiation. Examples include HSP70, HSP90 and sHSPs (Pockley, 2003; Saibil *et al.,* 2013; Mattoo and Goloubinoff, 2014).

A vital heat shock protein is the HSP70; which has been identified in the nucleus, cytosol, matrix of mitochondria and the lumen of the endoplasmic reticulum. Members of the HSP70 family stabilize an unfolded protein by binding to it. Different physiological processes rely on the HSP70 system, these include; processing of nascent protein inside the cytoplasm, in the endoplasmic reticulum and other cellular organelles, translation initiation, import and export of nuclear protein, assembling of ribosome, membrane translocation of secretory proteins, controlling action of regulatory proteins, organization of the cytoskeletal systems, protecting of nuclear structure and response to environmental stressors (Bakau *et al.,* 2000; Young *et al.,* 2003; Bakau, 2005).

HSP90 is one of the most copious constitutively expressed chaperone in the cytosol of eukaryotic organisms. Apart from its capability of maintaining target protein in expected folding state, HSP90 also plays a crucial role in cellular homeostasis, signal transduction pathways as well as cell cycle regulation. More so, this chaperone is of special interest because of the significant role it plays in the regulation of the proteins needed in malignant transformation (Pearl and Prodromou, 2000; Chiosis *et al.,* 2004).

Small heat shock proteins with molecular weights within 15 kDa and 30 kDa occur as complexes in plants, fungi, *Drosophila,* and yeast, and also in vertebrates. In

mammals, they are oligomeric structures having about 32 subunits which correspond to a molecular weight of 800 kDa. They exist in the cytosol of many cells and tissues

even while there is no threat from environmental stressors. The increased expression of sHSPs during heat shock and their protective activity on cellular viability at increased temperatures suggest sHSPs might have a role in the formation or maintenance of the cytosolic protein native conformation (Basha *et al.,* 2012; Haslbeck and Vierling, 2015).

1.12.2 Universal Stress Proteins

The genome of *S. japonicum*, *S. haematobium* as well as *S. mansoni* encode proteins having the Universal Stress Protein (USP) domain with the Pfam accession number; PF00582 (Kvint *et al.,* 2003; Berriman *et al.,* 2009; Young *et al.,* 2012; Mbah *et al.,* 2013). USPs have been discovered in yeast, archaea, bacteria, fungi and plants. Expressions of the USPs are initiated by many environmental stressors, including extreme temperature, toxic chemicals, nutrient starvation, oxidants, DNA damaging agents, heavy metals, un-couplers of the electron transport chain, acids and antibiotics (Nystrom and Neidhart, 1993; Diez *et al.,* 2000; Niachin, 2005; Mbah *et al.,* 2013; Tkaczuk *et al.,* 2013; Masamba *et al.,* 2016). It is suggested that the ancestral function of USPs is signal transduction and nucleotide binding (Isokpehi *et al.,* 2011b).

The universal stress protein domain can exist as a single domain in small USP proteins of approximately 14-15 kDa, as tandem domain found in larger USP proteins with approximately 30 kDa or as one or two USP domains alongside other functional domain (Kerk *et al.,* 2003) like antiporters, amino acid permease, voltage channels as well as protein kinase domains (Nystrom and Neidhardt, 1993; Kerk *et*

al., 2003). The number of *USP* genes occurring in different organisms varies considerably. *Xanthomonas campestris* (pv. Campetris) encodes only one *USP* gene while *Streptomyces coelicolor* (strain A3 (2)) has been predicted to encode as many as twelve (12) *USP* genes (O'Toole and Williams, 2003).

1.12.2.1 The *E. coli* universal stress protein

Although the USPs have been discovered in different species of organisms, evidence concerning their biological and biochemical characteristics are based on studies in *Escherichia coli* (strain K-12), which has its genome encoding six USPs (Gustavsson *et al.,* 2002). The first *E. coli* USPs to be detected was a protein present in the cytoplasm having 144 amino acids, and an up-regulation of this USP was significantly boosted when the bacterium development was suppressed by a huge diversity of stress posed by the environment like starvation of essential nutrients; including sulphates, phosphates, amino acids, carbon as well as nitrogen (Diez *et al.,* 2000; Niachin *et al.,* 2005). Since expression of the Universal stress protein is prompted by a diversity of stress conditions, inactivation of UspA in *E. coli* has been established to harm the existence of *E. coli* under a collection of growth arrest situations (Nystrom and Neidhardt, 1994). Five members of the USP family were discovered in *E. coli* namely: UspC (YecG), UspD (YiiT), UspE (YdaA), UspF (YnaF) as well as UspG (Gustavsson *et al.,* 2002).

Mutation of UspA leads to the altering of its global expression whereas overexpression leads to a global alteration in its pattern of protein expression. It is likely that overexpression of UspA may have an indirect effect on gene expression. Unsuitable expression of UspA in cells growing exponentially resulted in an immediate reduction in the growth rate (Nystrom and Neidhardt, 1996). Mushegian and Koonin (1996), suggested that DNA binding characteristics have a part to play in

stress enduring ability of the UspA of *E. coli*. It has been hypothesized that UspA has a structural resemblance to the transcriptional factor of the MADS-box protein family and the human serum response factor (MefA2) due to sequence match between the DNA binding helix of the two proteins and UspA (Mushegian and Koonin, 1996). However, after determining the three dimensional structure of *Haemophilus influenzae* UspA orthologue, the previous claim has been refuted (Sousa and McKay, 2001).

UspA is an example of genes which respond to growth arrest and their expression increases during a vast number of environmental insults. The protein product of UspA gene is usually abundant during the stationary phase of cells. Mutants of *E. coli* without UspA are unable to survive during growth arrest as well as during the DNA damaging conditions (Nystrom and Neidhardt, 1994; Gustavsson *et al.,* 2002).

Apart from being up-regulated during growth arrest, UspA of *Echerichia coli* is likewise phosphorylated on threonine as well as serine. It is suggested that phosphorylation act like the intracellular switch which stimulates UspA to accomplish its functions in the cell. If the suggestion is confirmed, it would become obvious that aside from USPs expression being up-regulated by growth phase indicators, other factors might control the action of USPs by modulating their phosphorylation condition. UspA experiences phosphorylation *in vitro* by means of ATP and GTP (phosphate donors) while other proteins are absent. Nevertheless, genetic studies elucidated that the phosphorylation of UspA *in vivo* needs an existence of the ribosomal tyrosine phosphoprotein TypA (Freestone *et al.,* 1997; Freestone *et al.,* 1998; O'Toole and Williams, 2003). It is suggested that USPs possess protein-protein binding activities.

A study showed UspA forming homodimers after purification and also associating with larger complexes in native extracts of *E. coli* (Gustavsson *et al.,* 2002; O'Toole and Williams, 2003). This report is agreeable with the discoveries on the USP of *Methanocaldococcus jannaschii*, (MJ0577) which also form homodimers (Zarembinski *et al.,* 1998; O'Toole and Williams, 2003). Additionally, the UspG of *E. coli* has been studied to be in complexes with GroEL, which are isolated from the stationary-phase cultures. The GroEL-UspG interface was established by means of co-immunoprecipitation experimentations. Hence, showing the existence of an experimental proof of the USPs partake in protein-protein interactions. It will be interesting to find out how this information can contribute to the function of USPs (Bochkareva *et al.,* 2002; O'Toole and Williams, 2003).

Findings have shown that adding ATP to GroEL complexes discharges UspG and on the other hand, the occurrence of ATP blocks the interface between UspG and GroEL (Bochkareva *et al.,* 2002). The interface between UspG and GroEL may be observed as exclusively the result of ATP on GroEL, however, the likelihood that the binding of ATP with USPs might affect their protein-protein interactions should not be eliminated outrightly. With inadequate information on the exact function of ATP binding, it might be significant in the functioning of USPs.

More so, the open reading frame directly upstream of UspA in *E. coli* encodes a protein with 111 amino acid residues. This protein was given the name universal stress protein B (UspB) due to its overall receptiveness to diverse starvation and stress situations in addition to the position of its structural gene next to UspA (Farewell *et al.,* 1998; O'Toole and Williams, 2003). At the stationary phase, overexpression of UspB led to cell death whereas mutants of UspB were sensitive to

ethanol but they were not sensitive to exposure to heat in the environment. Farewell and co-workers (1998) predicted that UspB is an integral membrane protein with a minimum of one membrane-spanning domain. However, it is significant to note that UspB was discovered prior to the discovery of the PF00582 family and it has no sequence match to UspA or other members of the USP family (Farewell *et al.,* 1998; O'Toole and Williams, 2003).

Due to the wide variety of resistance properties offered by the USPs of *E. coli*, it is not unexpected to discover them in the genome of a different species including *Schistosoma* species. More so, the exact functions of USPs in stress response have been revealed in few organisms. An example is in the pathogenic organism, *Porphyromonas gingivalis*, where the UspA gene was described to be a significant element in the formation of biofilm proposed as a way the organism react to stress in its environment (Chen *et al.,* 2006).

1.13 Functions of previously discovered USPs in living organisms

Most of the USPs discovered in the *Mycobacterium tuberculosis* which causes tuberculosis are described to be up-regulated in hypoxic situations, proposing that they perform imperative function in the continued existence of the bacteria during growth arrest (Hingley-Wilson *et al.,* 2010). More so, the USP of *M. tuberculosis* Rv2623 is described to control growth of the bacillus due to ATP-binding capacity in the organism (Drumm *et al.,* 2009).

UspA perform a vital function in resistance to metabolic as well as oxidative stress in *Salmonella typhimurium* which is an enteric bacterium. It was reported that the protein levels of UspA in the organism were established to be higher during the stationary phase as well as when the bacteria experiences exposure to rather high

temperatures between 30 °C-42 °C (Liu *et al.,* 2007). Also, USPs have been studied to perform important function during the stationary phase existence of *Pseudomonas aeruginosa* (Schreiber *et al.,* 2006).

The USP genes as well as their signal peptides respond sensitively to the threshold values of different stressors, therefore, the USP genes are expressed during a minimum threshold value while at above the maximum threshold value the living organism goes into extinction from the system. Hence, only organisms having well-armed USP apparatus will be able to survive threshold values which are higher. This means the threshold value of the environmental insult could be a dynamic parameter for the microbial consortium with the appearance or disappearance of a specific organism helping as a biomarker (Mbah and Isokpehi, 2013).

Zarembinski and co-workers (1998) determined the crystal structure of MJ0577 using its selenomethionyl derivatives. MJ0577 is a monomer structure with an open-twisted five-stranded parallel β-sheet having two helices on each side of the sheet. The amino acid residues encompassing the triphosphate binding loop were identified.

In addition to this, sequence alignment of MJ0577 showed the conservation of several ATP binding residues through a huge percentage of proteins categorized as members of USP family (Zarembinski *et al.,* 1998; O'Toole and Williams, 2003).

The MJ0577 possess a nucleotide-binding pocket which is bounded by motifs having incomplete similarities to what is usually found in ATP-binding proteins, however, the sequential arrangement of the motifs differs from other ATP-binding proteins. Hence, the crystal structure of MJ0577 characterizes another family of ATP-binding molecules. It is note-worthy that ATP was not used in the purification or

crystallization processes, therefore, MJ0577 might have scavenged the ATP from its host in the process of overexpression and held it during purification. MJ0577 is predicted to be an ATP-binding molecular switch, it hydrolyses ATP in the presence of *M. jannaschii* crude extract hence, the factor for the hydrolysis of ATP is in *M. jannaschii* (Zarembinski *et al.,* 1998).

Sousa and McKay (2001) studied the crystal structure of a Universal stress protein of *Haemophilus influenza*. The study revealed four UspA protomers in an asymmetric unit which are assigned to two dimers. Based on gel filtration results, the UspA of *H. influenza* shows it as an asymmetric dimer having a molecular weight of 34.5 kDa in solution compared with the calculated dimeric molecular weight of 32.0 kDa. It has a cross-validated estimated coordinate error of 0.17 Á. Within the subunits of each dimer, there is a sequesterion of a sulphate ion beside the side chains of residues Arg28 and His29. The two dimers of *H. influenza* UspA can be superimposed, however they differ noticeably in conformation in two segments of polypeptide thus making the dimers asymmetric (Sousa and McKay, 2001).

The UspA of *H. influenza* has a similarity in protomer fold as well as dimer assembly to the MJ0577 of *Methanococcus jannaschii*. Protomer of MJ0577 is a parallel α/β structure having five β strands, however the UspA protomer possess an analogous structural core but varies from MJ0577 due to its sixth, short and antiparallel β strand on the edge of the sheet distal to the dimer interface and a shorter α2 helix. There is also similarity in the dimer surface in the two proteins; the interface is formed due to an anti-parallel interaction between adjacent β5 strands (Sousa and McKay, 2001).

The UspA of *H. influenza* and MJ0577 differs in ATP binding. The crystal structure of MJ0577 binds ATP tightly despite the non-inclusion of ATP during its purification and

crystallization, whereas, UspA does not bind ATP despite it being grown and kept in 1mM MgATP. Hence, proteins sharing the MJ0577 fold can be classed into two; ATP-binding and non-ATP-binding. Even though the agreed sequence for ATP binding is yet to be defined, it is suggested that the G-2X-G-9X-G(S/T) motif is an important feature for classification. Some proteins have sequence motif and are classified as ATP binding while others do not and are thus classified as non-ATP binding example being the UspA of *H. influenza*. This suggests that the activities of UspA of *H. influenza* will be independent of ATP. The UspA from *H. influenza* shares 68% sequence identity with the UspA of *Escherichia coli* and 12 % identity with MJ0577.

The structural elucidation of this protein offers a basis for structure-function research to clarify the biochemical actions (Sousa and McKay, 2001). Tkaczuk and co-workers presented the crystal structures of the universal stress proteins AF0826 from the euryarchaeal *Archaeoglobus fulgidus* and NE1028 from the bacteria *Nitrosomonas europaea*. The apo-form of NE1028 has a monomer structure with an open twisted five-strand parallel β-sheet having a topology β3-β2- β-β1-β4-β5, fit in by α-helices. This structure resembles the *M. jannashii*, MJ0577 protein which has an ATP bound between β1 and β4. The apo-AF0826 monomer structure is analogous to the apo-NE1028 USP. The AF0826 USP forms an open, twisted, five-strand parallel β-sheet, with two additional β-strands forming a β-hairpin which is inserted in-between β4 and β5. Residues between the β4-β5 region usually form a loop in other USPs discovered (Tkaczuk *et al.,* 2013).

Loukehaich and co-workers (2012) cloned and functionally characterized an annexin-interacting USP in *Solanum pennelli*. They reported that SpUSP was

significantly induced by oxidative stress, dehydration, phytohormone abscisic acid (ABA) and salt stress. They predicted SpUSP to encode a protein with 145 amino acids, a molecular weight of 16.2 kDa and an isoeletric point of 5.9. The secondary structure of SpUSP is predicted to have conserved ATP-binding regions as well as other binding sites. The distribution of alpha-helices and beta-strands in SpUSP are very similar to that of MJ0577 crystal structure (Loukehaich *et al.,* 2012).

Maqbool and colleagues (2009) identified and cloned two USP genes; GUSP1 and GUSP2 of *Gosspyium arboretum* to study their role in drought-related stress. GUSP1 and GUSP2 are predicted to have molecular weights of 18.2 kDa and 19.1 kDa respectively. With the aid of sequence analysis, both USPs are reported to be very similar to the MJ0577-like USPs. Nucleotide sequence of both genes exhibited an 81% similarity with their encoded proteins sharing 75 % amino acid homology. The two proteins have 17-61 % similarity to the previously analysed USPs from various bacteria and plants (Maqbool *et al.,* 2009).

Kim and co-workers recently reported the crystal structure of a USP-like encoded protein At3g01520 from *Arabidopsis thaliana*. The crystal structure consists of three At3g01520 protein dimers with an AMP molecule attached to each protomer. Each At3g01520 protomer includes a central five-stranded, parallel β-sheet (β4-β5-β1-β2-β3), surrounded by five α-helices, it is folded into the typical Rossmann-like α/β-fold, similar to the structures of other described USP family members. At3g01520 was classified as an ATP binding USP due to the bound AMP and a preservation of residues in the ATP-binding loop (Kim *et al.,* 2015).

1.15 Motivation and problem statement

Schistosomiasis, a neglected tropical disease, has continued to cause various degrees of mortality and morbidity in various parts of the world. Although over the years, a lot of efforts have been exerted towards the control and elimination of this enervating disease of poverty, limited progress has been attained. The main breakthrough is the drug praziquantel formulated in the early 1970s which has been administered successfully to treat all forms of schistosomal infections but with many limitations as well as side-effects. More so, reports of the emergence of resistance of the schistosome worm to praziquantel treatment increased globally calls for alternative approaches to deal with the disease through early diagnostics, formulation of vaccine, as well as novel drugs. The genome of *S. mansoni* has been confirmed to encode the USP domain (Berriman *et al.,* 2009; Brindley *et al.,* 2009). These proteins play important roles during unfavourable conditions in many organisms, playing vital role during all the developmental stages of the *S. mansoni* parasite (Isokpehi *et al.,* 2011a, Isokpehi *et al.,* 2011b).

Mbah and co-workers states that schistosomes are able to overcome arrays of stress due to the up-regulation of stress regulating genes, which encode the universal stress proteins (USPs) in its genome (Mbah *et al.,* 2013). Structural characterization of this *S. mansoni* important regulatory protein could give a good lead to schistosomiasis intervention since research has established that recombinant proteins produced from an organism could be useful as druggable targets and are also vital in the production of therapeutic agents to combat diseases.

1.16 Scope of study

Schistosoma mansoni has been chosen for study because more than 90% of schistosomal infections are due to this specie (El-Shehabi *et al.,* 2012). *Schistosoma mansoni* is reputed to be the chief culprit for various forms of schistosomiasis especially intestinal and hepatic schistosomiasis hence, it is often the specie of the organism under study (Gryseels *et al.,* 2006; Gray *et al.,* 2011; Elbaz and Esmat, 2013).

1.17 Research hypothesis

The molecular cloning, production of a recombinant universal stress protein and subsequent structural characterization of the recombinant protein will provide structurally important data about the *S. mansoni* universal stress protein structure. The availability of such data is essentially useful in further studies towards schistosomiasis interventions for the control, treatment or elimination of the disease through the formulation of point-of-care diagnostics, vaccines and new drugs.

1.18 Objectives

1.18.1 Broad objectives

1. To experimentally clone, express and purify a stable, soluble and natively-folded G4LZI3 from *S. mansoni* at appropriate concentration.

2. To preliminarily characterize the purified recombinant G4LZI3 towards further structural elucidation.

1.18.2 Specific objectives

***In-silico* structural analysis of G4LZI3**

i. To retrieve the coding sequence of the universal stress protein, G4LZI3 from NCBI.

ii. To do the nucleotide analysis by blasting the NCBI database for related non-redundant protein sequences from the NCBI.

iii. To carry-out multiple sequence alignment of the G4LZI3 with selected homologous sequences.

iv. To predict the amino acid content of G4LZI3.

v. To predict the physicochemical properties of G4LZI3.

vi. To do the secondary structure prediction.

vii. To carry-out homology modelling and tertiary structure determination of the protein.

viii. To confirm the accuracy of the homology model.

ix. To carry-out 3D structure refinement.

x. To determine the solvent accessibility.

xi. To predict the binding sites on G4LZI3.

xii. To determine the protein subcellular localization.

xiii. To analysis the antigenic determinants/epitopes.

xiv. To carry-out gene ontology (GO) prediction on G4LZI3.

Laboratory experimental procedures toward characterization of G4LZI3.

i. To confirm the integrity of cloned pQE30-G4LZI3 by restriction digest analysis.

ii. To generate Histidine-tagged recombinant G4LZI3 protein from *S. mansoni* by recombinant expression procedures in *E. coli* system.

iii. To develop protein purification procedures for purifying this protein to homogeneity.

iv. To determine the protein concentration.

v. To ascertain the purity and molecular weight by SDS-PAGE analysis and mass spectroscopy.

vi. To do the preliminary structural characterization of the *S. mansoni* G4LZI3 protein.

References

Ajanga A., Lwambo N.J., Blair L., Nyandindi U., Fenwick A. and Brooker S. (2006). *Schistosoma mansoni* in pregnancy and associations with anaemia in northwest Tanzania. *Transactions of the Royal Society of Tropical Medicine and Hygiene,* **100**(1): 59-63.

Angelico M., Renganathan E., Gandin C., Fathy M., Cristina Profil M., Rafai W., De Santa A., Nagi A., Amin G., Capocaccia L., Callea F., Rapicetta M., Badr G. and Rocchi G. (1997). Chronic liver disease in the Alexandria governorate, Egypt: contribution of schistosomiasis and hepatitis virus infections. *Journal of Hepatology,* **26**(2): 236-243.

Aragon A.D., Imani R.A., Blackburn V.R., Cupit P.M., Melman S.D., Goronga T., Webb T., Loker E.S. and Cunningham C. (2009). Towards an understanding of the mechanism of action of praziquantel. *Molecular and Biochemical Parasitology,* **164**(1): 57-65.

Basha E., O'Neill H. and Vierling E. (2012). Small heat shock proteins and α-crystallins: dynamic proteins with flexible functions. *Trends in Biochemical Sciences,* **37**(3): 106-117.

Bawa J.A., Auta T., Msughter I., and Umar Y.A. (2016). Urinary Schistosomiasis among Primary School Children in Dutsin-Ma Town, Katsina State, Nigeria. *Annual Research and Review in Biology*, **10**(1): 1-8.

Berge S.T., Kabatereine N., Gundersen S.G., Taylor M., Kvalsvig J.D., Mkhize-Kwitshana Z., Jinabhai C. and Kjetland E.F. (2011). Generic praziquantel in South Africa: the necessity for policy change to provide cheap, safe and efficacious schistosomiasis drugs for the poor, rural population. *South African Journal of Epidemiology and Infection,* **26**(1): 22-25.

Berriman M., Haas B.J., LoVerde P.T., Wilson R.A., 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., Dillon G.P., Day T.A., Delcher A., DeMarco R., Djikeng A., Eyre T., Gamble J.A., Ghedin E., Gu Y., Hertz-Fowler C., Hirai H., Houston R., Ivens A., Johnston D.A., Lacerda D., Macedo C.D., McVeigh P., Ning Z., Oliveira G., Hirai Y., 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., FraserLiggett C.M., Barrell B.G. and El-Sayed N.M. (2009). The genome of the blood fluke *Schistosoma mansoni*. *Nature,* **460**(7253): 352-358.

Bertucci C., Pistolozzi M. and De Simone A. (2011). Structural characterization of recombinant therapeutic proteins by circular dichroism. *Current Pharmaceutical Biotechnology*, **12**(10): 1508-1516.

Bochkareva E.S., Girshovich A.S. and Bibi E. (2002). Identification and characterization of the *Escherichia coli* stress protein UP12, a putative *in vivo* substrate of GroEL. *European Journal of Biochemistry,* **269**(12): 3032-3040.

Bottieau E., Clerinx J., de Vega M.R., den Enden E.V., Colebunders R., Esbroeck M.V., Vervoort T., Gompel A.V. and den Ende J.V. (2006). Imported Katayama fever:

clinical and biological features at presentation and during treatment. *Journal of Infections,* **52**(5)**:** 339-345.

Brindley P.J., Mitreva M., Ghedin E. and Lustigman S. (2009). Helminth genomics: the implications for human health. *PloS Neglected Tropical Diseases,* **3**(10): E538.

Bukau B., Deuerling E., Pfund C. and Craig E. A. (2000). Getting newly synthesized proteins into shape. *Cell,* **101**(2): 119-122.

Butler S.E., Muok E.M., Montgomery S.P., Odhiambo K., Mwinzi P.M.N., Secor W.E. and Karanja D.M.S. (2012). Mechanism of anaemia in *schistosoma mansoni*–infected school children in Western Kenya*. American Journal of Tropical Medicine and Hygiene,* **87**(5): 862–867.

Calamini B. and Morimoto R.I. (2012). Protein homeostasis as a therapeutic target for diseases of protein conformation. *Current Topics in Medicinal Chemistry,* **12**(22): 2623.

Caldas I.R., Campi-Azevedo A.C., Oliveira L.F.A., Silveira A.M.S., Oliveira R.C. and Gazzinelli G. (2008). Human schistosomiasis mansoni: immune responses during acute and chronic phases of the infection. *Acta Tropica,* **108**(2): 109-117.

Cardoso L.S., Araujo M.I., GóesLucila A.M., Pacífico G., Oliveira R.R. and Oliveira S.C. (2007). Polymyxin B as inhibitor of LPS contamination of *Schistosoma mansoni* recombinant proteins in human cytokine analysis. *Microbial Cell Factories,* **6**:1.

Chen W., Honma K., Sharma A. and Kuramitsu H.K. (2006). A universal stress protein of *Porphyromonas gingivalis* is involved in stress responses and biofilm formation. *FEMS Microbiology Letters,* **264**(1): 15-21.

Chevillard C., Moukoko C.E., Elwali N.E.M., Bream J.H., Kouriba B., Argiro L., Rahoud S., Mergani A., Henri S., Gaudart J., Mohamed-Ali Q., Young H.A. and Dessein A. J. (2003). IFN-γ polymorphisms (IFN-γ+ 2109 and IFN-γ+ 3810) are associated with severe hepatic fibrosis in human hepatic schistosomiasis (*Schistosoma mansoni*). *Journal of Immunology,* **171**(10): 5596-5601.

Chiosis G., Vilenchik M., Kim J., and Solit D. (2004). Hsp90: the vulnerable chaperone. *Drug Discovery Today,* **9**(20): 881-888.

Chitsulo D., Engels A., Montresor and L Savioli. (2000). The global status of schistosomiasis and its control. *Acta Tropica,* **77**: 41–51.

Chrousos G.P. (2009). Stress and disorders of the stress system. *Nature Reviews Endocrinology,* **5**(7): 374-381.

Coon D.R. (2005). Schistosomiasis: Overview of the history, biology, clinicopathology, and laboratory diagnosis. *Clinical Microbiology Newsletter* **27**(21): 163-168.

Coulibaly J.T., N'Goran E.K., Utzinger J., Doenhoff M.J. and Dawson E.M. (2013). A new rapid diagnostic test for the detection of anti-*Schistosoma mansoni* and anti-*Schistosoma haematobium* antibodies. *Parasites and Vectors,* **6**:29.

De Martel C., Ferlay J., Franceschi S., Vignat J., Bray F., Forman D. and Plummer M. (2012). Global burden of cancers attributable to infections in 2008: a review and synthetic analysis. *Lancet Oncology,* **13**(6): 607-615.

Deribe K., Meribo K., Gebre T., Hailu A., Ali A., Aseffa A., and Davey G. (2012). The burden of neglected tropical diseases in Ethiopia, and opportunities for integrated control and elimination. *Parasites and Vectors*, ***5***(1): 240.

Diez A., Gustavsson N. and 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. *Molecular Microbiology,* **36**(6): 1494-1503.

Doenhoff M.J., Hagan P., Cioli D., Southgate V., Pica-Mattoccia L., Botros S., Coles G., Tchuem Tchuente L.A., Mbaye A. and Engels D. (2009). Praziquantel: its use in control of schistosomiasis in sub-Saharan Africa and current research needs. *Parasitology,* **136**(13): 1825-1835.

Doenhoff M.J., Kusel J.R., Coles G.C. and Cioli D. (2002). Resistance of *Schistosoma mansoni* to praziquantel: is there a problem? *Transactions of the Royal Society of Tropical Medicine and Hygiene,* **96**(5): 465-469.

Doenhoff MJ, Cioli D, Utzinger J (2008). Praziquantel: mechanism of action, resistance and new derivatives for schistosomiasis. *Current Opinion on Infectious Diseases* **21**(6): 659-667. Driguez P., McManus D.P. and Gobert G.N. (2016). Clinical implications of recent findings in schistosome proteomics. *Expert Review of Proteomics*, **13**(1), 19-33.

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. and Chan J. (2009). *Mycobacterium tuberculosis* universal stress protein Rv2623 regulates

bacillary growth by ATP-Binding: requirement for establishing chronic persistent infection. *PLoS Pathogen,* **5**: e1000460.

El-Awady M.K., Youssef S.S., Omran M.H., Tabil A.A., El Garf W.T. and Salem A.M. (2006). Soluble egg antigen of *Schistosoma haematobium* induces HCV replication in PBMC from patients with chronic HCV infection*. BioMed Central Infectious Diseases,* **6**(1): 91.

Elbaz T. and Esmat T. (2013). Hepatic and intestinal schistosomiasis: review. *Journal of Advanced Research,* **4**(5): 445-452.

El Ridi A.F., Tallima H.A. (2013). Novel therapeutic and prevention approaches for schistosomiasis: review. *Journal of Advanced Research,* **4**(5): 467-478.

El-Shehabi F., Taman A., Moali, L.S., El-Sakkary N and Ribeiro P. (2012). A novel G protein-coupled receptor of *Schistosoma mansoni* (SmGPR-3) is activated by dopamine and is widely expressed in the nervous system. *PLoS Neglected Tropical Diseases,* **6**(2): e1523.

Engels D., Chitsulo L., Montresor A. and Savioli L. (2002). The global epidemiological situation of schistosomiasis and new approaches to control and research. *Acta Tropica,* **82**(2): 139-146.

Fallon P.G., Tao L.F., Ismail M.M. and Bennett J.L. (1996). Schistosome resistance to praziquantel: fact or artifact? *Parasitology Today,* **12**(8): 316-320.

Farewell A., Kvint K. and Nystrom T. (1998). UspB, a new sigma S-regulated gene in *Escherichia coli* which is required for stationary-phase resistance to ethanol. *Journal of Bacteriology,* **180**: 6140-6147.

Feldmeier H., Krantz I. and Poggensee G. (1994). Female genital schistosomiasis as a risk-factor for the transmission of HIV. *International Journal of STD and AIDS,* **5**(5): 368-372.

Fenwick A. and Webster J.P. (2009). Schistosomiasis: challenges for control, treatment and drug resistance*. Currrent Opinion on Infectious Diseases,* **19**(6): 577-582.

Fernández-Robledo J.A. and Vasta G.R. (2010). Production of recombinant proteins from protozoan parasites. *Trends in Parasitology,* **26**(5): 244-254.

Foret S., Seneca F., de Jong D., Bieller A., Hemmrich G., Augutin R., Hayward D.C., Ball E.E., Bosch T.C.G., Agata K., Hassel M. and Miller D.J. (2011). Phylogenomics reveals an anomalous distribution of USP genes in metazoans. *Molecular Biology and Evolution,* **28**(1): 153-161.

Freestone P., Nystrom T., Trinei M. and Norris V. (1997). The universal stress protein, UspA, of *Escherichia coli* is phosphorylated in response to stasis. *Journal of Molecular Biology,* **274** (3): 318-324.

Freestone P., Trinei M., Clarke S.C., Nystrom T. and Norris V. (1998). Tyrosine phosphorylation in *Escherichia coli*. *Journal of Molecular Biology,* **279**: 1045-1051.

Friedman J.F., Kanzaria H.K., McGarvey S.T. (2005). Human schistosomiasis and anaemia: the relationship and potential mechanisms. *Trends in Parasitology,* **21**(8): 386-392.

Friedman J.F., Mital P., Kanzaria H.K., Olds O.R. and Kurtis J.D. (2007). Schistosomiasis and pregnancy. *Trends in Parasitology,* **23**(4): 159-164.

Gray D.J., Ross A.G., Li Y. and McManus D.P. (2011). Diagnosis and management of schistosomiasis. *British Medical Journal,* **342**: d2651.

Gryseels B., Polman K., Clerinx J. and Kestens L. (2006). Human schistosomiasis. *Lancet,* **368**(9541): 1106-1118.

Gustavsson N., and Nyström T. (2002). The universal stress protein paralogues of *Escherichia coli* are co-ordinately regulated and co-operate in the defence against DNA damage. *Molecular Microbiology*, *43*(1), 107-117.

Halim A.B., Garry R.F., Dash S. and Gerber M.A. (1999). Effect of schistosomiasis and hepatitis on liver disease. *The American Journal of Tropical Medicine and Hygiene,* **60**(6): 915-920.

Haslbeck M. and Vierling E. (2015). A first line of stress defense: small heat shock proteins and their function in protein homeostasis. *Journal of molecular biology*, **427**(7): 1537-1548.

Hatz C. (2005). Schistosomiasis: an underestimated problem in industrialised countries? *Journal of Travel Medicine,* **12**(1)**:** 2.

Hayashi S., Ohtake H. and Koike M. (2000). Laparoscopic diagnosis and clinical course of chronic *Schistosomiasis japonica*. *Acta Tropical,* **77**(1): 133-140.

Hermeling S., Aranha L., Damen J.M.A., Slijper M., Schellekens H., Crommelin D.J. and Jiskoot W. (2005). Structural characterization and immunogenicity in wild-type and immune tolerant mice of degraded recombinant human interferon alpha2b. *Pharmaceutical Research*, **22**(12): 1997-2006.

Hightower L.E. (1991). Heat shock, stress proteins, chaperones, and proteotoxicity. *Cell,* **66**(2): 191-197.

Hingley-Wilson S.M., Lougheed K.E., Ferguson K., Leiva S. and Williams H.D. (2010). Individual *Mycobacterium tuberculosis* universal stress protein homologues are dispensable *in vitro*. *Tuberculosis* **90**(4): 236-244.

Ismail M., Botros S., Metwally A., William S., Farghally A., Tao L.F., Day T.A. and Bennett J.L. (1999). Resistance to praziquantel: direct evidence from *Schistosoma mansoni* isolated from Egyptian villagers. *American Journal of Tropical Medicine and Hygiene,* **60**(6): 932-935.

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-Mclnnis 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. and Oliveira G. (2011a). Developmental regulation of genes encoding universal stress proteins in *Schistosoma mansoni. Gene Regulation and Systems Biology,* **5**: 61-74.

Isokpehi R.D., Simmons S.S., Cohly H.H., Ekunwe S.I., Begonia G.B. and Ayensu W.K. (2011b). Identification of drought-responsive universal stress proteins in viridiplantae. *Bioinformatics and Biology Insights,* **5**: 41.

Jakob U., Gaestel M., Engel K., and Buchner J. (1993). Small heat shock proteins are molecular chaperones. *Journal of Biological Chemistry,* **268**(3): 1517-1520.

Jelinek T., Nothdurft H.D. and Loscher T. (1996). Schistosomiasis in travellers and expatriates. *Journal of Travel Medicine,* **3**(3): 160–164.

Jenkins-Holick D.S. and Kaul T.L. (2013). Schistosomiasis. *Urology and Nursing,* **33**(4): 163-170.

Johnson R.B., Fearon K., Mason T. and Jindal S. (1989). Cloning and characterization of the yeast chaperonin HSP60 gene. *Gene,* **84**(2): 295-302.

Kerk D., Bulgrien J., Smith D.W. and Gribskov M. (2003). Arabidopsis proteins containing similarity to the universal stress protein domain *Plant Physiology,* **132**: 1209-1219.

Kim D. J., Bitto E., Bingman C. A., Kim H. J., Han B. W., and Phillips G. N. (2015). Crystal structure of the protein At3g01520, a eukaryotic universal stress protein-like protein from *Arabidopsis thaliana* in complex with AMP.*Proteins: Structure, Function, and Bioinformatics*, **83**(7): 1368-1373.

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

Kjetland E.F., Leutscher P.D.C. and Ndhlovu P.D. (2012). A review of female genital schistosomiasis. *Trends in Parasitology,* **28**(2): 58-65.

Koolhaas J.M., Bartolomucci A., Buwalda B., De Boer S.F., Flügge G., Korte S.M. and Fuchs E. (2011). Stress revisited: a critical evaluation of the stress concept. *Neuroscience & Biobehavioral Reviews,* **35**(5): 1291-1301.

Koukounari A., Fenwick A., Whawell S., Kabatereine N. B., Kazibwe F., Tukahebwa E. M., Stothard J.R., Donnelly C.A. and Webster J.P. (2006). Morbidity indicators of *Schistosoma mansoni*: relationship between infection and anemia in Ugandan schoolchildren before and after praziquantel and albendazole chemotherapy. *American Journal of Tropical Medicine and Hygiene,* **75**(2): 278-286.

Kvint K., Nachin L., Diez A. and Nystrom T. (2003). The bacterial universal stress protein: function and regulation. *Current Opinion in Microbiology,* **6**(2): 140-145.

Lapa M., Dias B., Jardim C., Fernandes C.J., Dourado P.M., Figuiredo M., Farias A., Tsutsui J., Terra-Filho M., Humbert M., Souza R. (2009). Cardiopulmonary manifestations of hepatosplenic schistosomiasis. *Circulation,* **119**(11): 1518-1523.

Liu W.T., Karavolos M.H., Bulmer D.M., Allaoui A., Hormaeche R.D., Lee J.J. and Khan C. M. (2007). Role of the universal stress protein UspA of *Salmonella* in growth arrest, stress and virulence. *Microbial and Pathogenesis,* **42**(1): 2-10.

Loukehaich R., Wang T., Ouyang B., Ziaf K., Li H., Zhang J., Lu Y. and Ye Z. (2012). SpUSP, an annexin-interacting universal stress protein, enhances drought tolerance in tomato. *Journal of Experimental Botany,* ers220.

Madinga J., Linsuke S., Mpabanzi L., Meurs L., Kanobana K., Speybroeck N., Lutumba P. and Polman K. (2015). Schistosomiasis in the Democratic Republic of Congo: a literature review. *Parasites and Vectors*, **8**(1): 1-10.

Mahajan S., and Tuteja N. (2005). Cold, salinity and drought stresses: an overview. *Archives of Biochemistry and Biophysics,* **444**(2): 139-158.

Mahgoub H.M., Mohamed A.A., Magzoub M., Gasim G.I., Eldein W.N., Ahmed A.A., and Adam I. (2010). *Schistosoma mansoni* infection as a predictor of severe anaemia in school-children in Eastern Sudan. *Journal of Helminthology,* **84**(2): 132-135.

Maloney A. and Workman P. (2002). HSP90 as a new therapeutic target for cancer therapy: the story unfolds. *Expert Opinion on Biological Therapy*, **2**(1): 3-24.

Maqbool A., Zahur M., Husnain T. and Riazuddin S. (2009). GUSP1 and GUSP2, two drought-responsive genes in *Gossypium arboreum* have homology to universal stress proteins. *Plant Molecular Biology Reporter,* **27**(1): 109-114.

Martin R.J. (1997). Modes of action of anthelmintic drugs. *Veterinary Journal,* **154**(1): 11-34.

Masamba P., Adenowo A.F., Oyinloye B.E. and Kappo A.P. (2016). Universal Stress Proteins as New Targets for Environmental and Therapeutic Interventions of Schistosomiasis. *International Journal of Environmental Research and Public Health,* **13**(10): 972.

Mattoo R.U. and Goloubinoff P. (2014). Molecular chaperones are nanomachines that catalytically unfold misfolded and alternatively folded proteins. *Cellular and Molecular Life Sciences*, **71**(17): 3311-3325.

Mayer M.P., and Bukau B. (2005). Hsp70 chaperones: Cellular functions and molecular mechanism. *Cellular and Molecular Life Sciences,* **62**(6): 670-684.

Mbabazi P.S., Andan O., Fitzgerald D.W., Chitsulo L., Engels D. and Downs J.A. (2011). Examining the relationship between urogenital schistosomiasis and HIV infection. *PLoS Neglected Tropical Disease,* **5**: e1396.

Mbah A.N. and Isokpehi R.D. (2013). Application of Universal Stress Proteins in Probing the Dynamics of Potent Degraders in Complex Terephthalate Metagenome. *BioMed Research International,* **2013**: 1-21.

Mbah A.N., Mahmud O., Awofolu O.R. and Isokpehi R.D. (2013). Inferences on the biochemical and environmental regulation of universal stress proteins from

Schistosomiasis parasites. *Advances and Applications in Bioinformatics and Chemistry,* **6**: 15-27.

Melman S.D., Steinauer M.L., Cunningham C., Kubatko L.S., Mwangi I.N., Wynn N.B., Mutuku M.W., Karanja D.M.S., Colley D.G., Black C.L., Secor W.E., Mkoji G.M. and Loker E.S. (2009). Reduced susceptibility to praziquantel among naturally occurring Kenyan isolates of *Schistosoma mansoni*. *PLoS Neglected Tropical Diseases,* **3**(8): e504.

Montes M., White A.C., and Kontoyiannis D.P. (2004). Symptoms of intestinal schistosomiasis presenting during treatment of large B cell lymphoma. *The American journal of tropical medicine and hygiene,* **71**(5): 552-553.

Mushegian A.R. and Koonin E.V. (1996). Sequence analysis of eukaryotic developmental proteins: ancient and novel domains. *Genetics,* **144**(2): 817-828.

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

Negrão-Corrêa D., Mattos A.C.A., Pereira C.A.J., Martins-Souza R.L., and Coelho P.M.Z. (2012). Interaction of *Schistosoma mansoni* sporocysts and hemocytes of *Biomphalaria. Journal of Parasitology Research,* **2012**: 1-6.

Nystrom T. and Neidhardt F.C. (1992). Cloning, mapping and nucleotide sequencing of a gene encoding a universal stress protein in *Escherichia coli*. *Molecular Microbiology,* **6**(21): 3187-3198.

N'Goran E.K., Utzinger J., N'guessan A.N., Müller I., Zamblé K., Lohourignon K.L., Traoré M., Sosthène B.A., Lengeler C. and Tanner M. (2001). Reinfection with Schistosoma haematobium following school-based chemotherapy with praziquantel in four highly endemic villages in Côte d'Ivoire. *Tropical Medicine and International Health*, ***6***(10): 817-825.

Nyström T. and Neidhardt F.C. (1993). Isolation and properties of a mutant of *Escherichia coli* with an insertional inactivation of the *UspA* gene, which encodes a universal stress protein. *Journal of Bacteriology,* **175**(13): 3949-3956.

Nystrom T. and Neidhardt F.C. (1994). Expression and role of the universal stress protein, UspA, of *Escherichia coli* during growth arrest. *Molecular Microbiology,* **11**(3): 537-544.

Nystrom T. and Neidhardt F.C. (1996). Effects of overproducing the universal stress protein, UspA, in *Escherichia coli* K-12. *Journal of Bacteriology,* **178** (3): 927-930.

O'Toole R. and Williams H.D. (2003). Universal stress proteins and *Mycobacterium tuberculosis*. *Research in Microbiology,* **154**(6): 387-392.

Olveda D.U., Li Y., Olveda R.M., Lam A.K., Chau T.N.P., Harn D.A., Williams G.M., Gray D.J. and Ross A.G. (2013). Bilharzia: Pathology, diagnosis, management and control. *Tropical Medicine and Surgery,* **1**: 135.

Pearl L.H., and Prodromou C. (2000). Structure and *in vivo* function of Hsp90. *Current Opinion in Structural Biology,* **10**(1): 46-51.

Pesce E.R., Cockburn I.L., Goble J.L., Stephens L.L. and Blatch G.L. (2010). Malaria heat shock proteins: drug targets that chaperone other drug targets. *Infectious Disorders-Drug Targets,* **10**(3): 47-157.

Pockley A.G. (2003). Heat shock proteins as regulators of the immune response. *Lancet*, **362**(9382): 469-476.

Pontes L.A., Oliveira M.C., Katz N., Dias-Neto E. and Rabello A. (2003). Comparison of a polymerase chain reaction and the Kato-Katz technique for diagnosing infection with *Schistosoma mansoni. American Journal of Medicine and Hygiene,* **68**(6): 652-656.

Rabello A., Pontes L.A. and Dias-Neto E. (2002). Recent advances in the diagnosis of Schistosoma infection: the detection of parasite DNA. *Memorias do Instituto Oswaldo Cruz,* **97**: 171-172.

Redman C.A., Robertson A., Fallon P.G., Modha J., Kusel J.R., Doenhoff M.J. and Martin R.J. (1996). Praziquantel: an urgent and exciting challenge. *Parasitology Today,* **12**(1):14-20.

Rollinson D., Knopp S., Levitz S., Stothard J.R., Tchuem Tchuente., Garba A., Muhammed K.A., Schur N., Person B., Colley D.G. and Utzinger J. (2013). Time to set agenda for schistosomiasis elimination. *Acta Tropical,* **128**(2): 423-440.

Sadhu P.S., Kumar S.N., Chandrasekharam M., Pica-Mattoccia L., Cioli D. and Rao V.J. (2012). Synthesis of new praziquantel analogues: potential candidate for treatment of schistosomiasis. *Bioorganic and Medicinal Chemistry Letters,* **22**(2): 1103-1106.

Saibil H. (2013). Chaperone machines for protein folding, unfolding and disaggregation. *Nature Reviews of Molecular and Cell Biology*, **14**(10): 630-642.

Samaras V., Rafailidis P.I., Mourtzoukou E.G., Peppas G. and Falagas M.E. (2010) Chronic bacterial and parasitic infections and cancer: a review. *Journal of Infections in Developing Countries,* **4(**5): 267-281.

Schreiber K., Boes N., Eschbach M., Jaensch L., Wehland J., Bjarnsholt T., Givskov M., Hentzer M. and Schobert M. (2006). Anaerobic survival of *Pseudomonas aeruginosa* by pyruvate fermentation requires an Usp-type stress protein. *Journal of Bacteriology,* **188**(2): 659-668.

Schechtman D., Tarrab-Hazdai R. and Arnon R. (2001). The 14-3-3 protein as a vaccine candidate against schistosomiasis. *Parasite Immunology*, **23**(4): 213-217.

Schwartz E., Rozenman J., and Perelman M. (2000). Pulmonary manifestations of early schistosome infection among nonimmune travellers. *American Journal of Medicine,* **109**(9): 718-722.

Shonhai A. (2010). Plasmodial heat shock proteins: targets for chemotherapy. *FEMS Immunology and Medical Microbiology*, **58**(1): 61-74.

Siddiqui A.A., Siddiqui B.A. and Ganley-Leal L. (2011). Schistosomiasis vaccines. *Human Vaccines*, **7**(11): 1192-1197.

Sodoyer R. (2004). Expression System for the production of pharmaceuticals. *Biodrugs,* **18**(1): 51-62.

Somero G.N. (1995). Proteins and temperature. *Annual Review of Physiology,* **57**(1): 43-68.

Sousa M.C. and McKay D.B. (2001). Structure of the universal stress protein of *Haemophilus influenzae*. *Structure,* **9**(12): 1135-1141.

Sun Y. and MacRae T.H. (2005). The small heat shock proteins and their role in human disease. *FEBs Journal*, **272**(11): 2613-2627.

Tkaczuk K.L., Shumilin I., Chruszcz M., Evdokimova E., Savchenko A. and Minor W. (2013). Structural and functional insight into the universal stress protein family. *Evolutionary Applications,* **6**(3): 434-449.

Tchuenté L.A.T., Momo S.C., Stothard J.R. and Rollinson D. (2013). Efficacy of praziquantel and reinfection patterns in single and mixed infection foci for intestinal and urogenital schistosomiasis in Cameroon. *Acta tropica*, **128**(2): 275-283.

Toft D.O. (1998). Recent advances in the study of Hsp90 structure and mechanism of action. *Trends in Endocrinology and Metabolism,* **9**(6): 238-243.

van der Werf M.J., de Vlas S.J., Brooker S., Looman C.W., Nagelkerke N.J., and Habbema J.D. (2003). Quantification of clinical morbidity associated with schistosome infection in sub-Saharan Africa*. Acta Tropical,* **86**:125-139.

Wang H., Lei Z., Li X. and Oetting R.D. (2011). Rapid cold hardening and expression of heat shock protein genes in the B-biotype *Bemisia tabaci*. *Environment and Entomology,* **40**(1): 132-139.

Young J.C., Barral J.M. and Hartl F.U. (2003). More than folding: localized functions of cytosolic chaperones. *Trends in Biochemical Sciences,* **28**(10): 541–547.

Young N.D., Jex A.R., Li B., Liu B., Yang L., Xiong Z., Li Y., Cantacessi C., Hall R.S., Xu X., Chen F., Wu X., Zerlotini A., Oliveira G., Hoffmann A., Zhang G., Fang X.,

Kang Y., Campbell B.E., Loukas A., Ranganathan S., Rollinson D., Rinanldi G., Brindley P.J., Yang H., Wang J., Wang J. and Gasser R.B. (2012). Whole-genome sequence of *Schistosoma haematobium*. *Nature Genetics,* **44**(2): 221-225.

Zaghloul M.S. (2012). Bladder cancer and schistosomiasis*. Journal of Egypt National Cancer Institute* **24**:151-159.

Zarembinski T.I., Hung L.W., Mueller-Dieckmann H.J., Kim K.K., Yokota H., Kim R. and Kim S.H. (1998). Structure-based assignment of the biochemical function of a hypothetical protein: a test case of structural genomics. *Proceedings of the National Academy of Sciences,* **95:** 26.

CHAPTER TWO

Publication by Adenowo A.F., Oyinloye B.E., Ogunyinka B.I. and Kappo A.P. (2015). Impact of human schistosomiasis in sub-Saharan Africa. *Brazilian Journal of Infectious Diseases*, 19(2): 196-205.

Article abstract page

BRAZ J INFECT DIS. 2015;19(2):196–205

The Brazilian Journal of
INFECTIOUS DISEASES

www.elsevier.com/locate/bjid

Review article

Impact of human schistosomiasis in sub-Saharan Africa

Abiola Fatimah Adenowo[a], *Babatunji Emmanuel Oyinloye*[a,b], *Bolajoko Idiat Ogunyinka*[a], *Abidemi Paul Kappo*[a,*]

[a] *Biotechnology and Structural Biology (BSB) Group, Department of Biochemistry and Microbiology, University of Zululand, KwaDlangezwa 3886, South Africa*
[b] *Department of Biochemistry, College of Sciences, Afe Babalola University, Ado Ekiti, Nigeria*

ARTICLE INFO

Article history:
Received 28 September 2014
Accepted 12 November 2014
Available online 27 January 2015

Keywords:
Neglected tropical diseases
Praziquantel
Schistosomiasis
Sub-Saharan Africa

ABSTRACT

Schistosomiasis, a neglected tropical disease of poverty ranks second among the most widespread parasitic disease in various nations in sub-Saharan Africa. Neglected tropical diseases are causes of about 534,000 deaths annually in sub-Saharan Africa and an estimated 57 million disability-adjusted life-years are lost annually due to the neglected tropical diseases. The neglected tropical diseases exert great health, social and financial burden on economies of households and governments. Schistosomiasis has profound negative effects on child development, outcome of pregnancy, and agricultural productivity, thus a key reason why the "bottom 500 million" inhabitants of sub-Saharan Africa continue to live in poverty. In 2008, 17.5 million people were treated globally for schistosomiasis, 11.7 million of those treated were from sub-Saharan Africa. This enervating disease has been successfully eradicated in Japan, as well as in Tunisia. Morocco and some Caribbean Island countries have made significant progress on control and management of this disease. Brazil, China and Egypt are taking steps towards elimination of the disease, while most sub-Saharan countries are still groaning under the burden of the disease. Various factors are responsible for the continuous and persistent transmission of schistosomiasis in sub-Saharan Africa. These include climatic changes and global warming, proximity to water bodies, irrigation and dam construction as well as socio-economic factors such as occupational activities and poverty. The morbidity and mortality caused by this disease cannot be overemphasized. This review is an exposition of human schistosomiasis as it affects the inhabitants of various communities in sub-Sahara African countries. It is hoped this will bring a re-awakening towards efforts to combat this impoverishing disease in terms of vaccines development, alternative drug design, as well as new point-of-care diagnostics.

* *Corresponding author.*
E-mail addresses: KappoA@unizulu.ac.za, bidemi2k@gmail.com (A.P. Kappo).
http://dx.doi.org/10.1016/j.bjid.2014.11.004
1413-8670/

2.1 Introduction

Human schistosomiasis otherwise called bilharzia, is a freshwater snail transmitted intravascular debilitating disease resulting from infection by the parasitic dimorphic *Schistosoma* trematode worms, which lives in the bloodstream of humans.[1,2] The

World Health Organization (WHO) regards the disease as a neglected tropical disease, with an estimated 732 million persons being vulnerable to infection world-wide in renowned transmission areas.[3] Steinmann and co-workers documented that over 200 million individuals are infected with this disease from Africa, Asia, and South America.[1] The WHO further estimated that schistosome infections and geohelminths accounts for over 40 % of the world tropical disease burden with the exclusion of malaria.[4] Humans get infected with this disease when they make contact with water contaminated with the skin-penetrating cercariae. Prevalence of schistosomiasis at present, is still high in sub-Saharan Africa. In 2008, 17.5 million people were treated globally for schistosomiasis, 11.7 million of those treated were from sub-Saharan Africa.[3] An approximate of 120 million individuals in sub-Saharan Africa have schistosomiasis-related symptoms while about 20 million undergoes hardship as a result of chronic presentations of the disease.[5]

Schistosomiasis has been successfully eliminated in Japan and Tunisia. Morocco and some Caribbean Islands countries had made significant progress on control of the disease while Brazil, China, and Egypt are taking steps towards elimination of the disease.[6] Schistosomiasis is more rampant in poor rural communities especially places where fishing and agricultural activities are dominant. Domestic activities such as washing clothes and fetching water in infected water exposes women and children to infection.

Recreational activities like swimming and poor hygiene also makes children vulnerable to schistosomiasis.[3] Humans are usually infected by five species of schistosomes, namely *Schistosoma mansoni, Schistosoma haematobium, Schisosoma japonicum, Schistosoma mekongi and Schistosoma intercalatum,* but

the main burden of disease in sub-Saharan Africa is usually attributed to two species, which are, *S. mansoni* and *S. haematobium* and are referred to as the major human schistosomes.[6]

Biomphalaria snails are responsible for the transmission of *S. mansoni* which is the source of hepatic and intestinal schistosomiasis in places like the Arabian countries, South America, and Africa. *Bulinus* snails transmits *S. haematobium* which is the chief cause of urinary schistosomiasis in Africa and the Arab world.[2] The *Biomphalaria* snail comprises of many species including *Bi. alexandrina, Bi. Sudanica, Bi. Pfeifferi and Bi. hoanomphala,* while the genus *Bulinus* comprises of the following species; *Bu. tropicus, Bu. globosus, Bu. truncatus, Bu. forskalli* and *Bu. Africanus.*[7] *Schistosoma japonicum* is spread by the freshwater snail *Oncomelania* and it is responsible for intestinal and hepatosplenic schistosomal infections in Indonesia, Peoples Republic of China, and the Phillipines. It is a zoonotic parasite infecting animals including pigs, dogs, cattle, and rodents. Some other species of schistosome are parasites of different animals, but occassionally infect humans. *Schistosoma mansoni* can also be found in primates and rodents but the main host are human beings.[2]

Matured schistosomes are usually greyish or white worms with a length of 7 mm-20 mm; having a cylindrical shape with two ending suckers, a blind digestive tract, a complex tegument, and reproductive organs. A distinguishing feature in this trematode compared with other trematodes is its existence in two sexes.

The male has a gyneaphoric channel or a groove, wherein it grips the female which is usually longer and thinner. The male and female schistosomes lives as permanently embraced couples in the perivesical venous (in *Schistosoma*

haematobium) or in the mesentric venous (in *Schistosoma mansoni* and *Schistosoma japonicum* species). The schistosomes gets nourishment from the host blood and globulins by means of anaerobic glycolysis and excretes the waste back into the body of the hosts.[2] A female schistosome has the capacity to produce hundreds of eggs in a day as discovered in the African species, and about thousands of eggs per day in the oriental species. An individual ovum possesses miracidium larva with a cilia that produces proteolytic enzymes which aids the eggs to move inside either the lumen of the bladder or into the host intestine.[2]

Subsequently, the parasites eggs are released into the faeces or urine where it remains alive for about seven days. When it gets into freshwater, the miracidium is released from the egg. With the aid of chemical stimuli and light, the miracidium seek the freshwater snail which is its intermediate host. On locating the snail, the miracidium penetrates it and undergo asexual reproduction to produce multicellular sporocytes which develops to cercarial larvae having embryonic suckers as well as a two-branched tail.[2] After 4-6 weeks of infecting the snail, the cercariae leaves the snail and gyrates around for about 72 hours looking out for the skin of a prospective host. Releasing the cercariae from the snail is instigated by light and takes place mainly during the day time. On locating a human host skin, the cercariae burrows into it, migrates into the blood through the liver and lungs and undergoes transformation into a schistosomulae otherwise called young worms.[2]

The schistosomulae matures within 4-6 weeks inside the portal vein, mates, and migrates to their destination, which is either the perivesicular or mesenteric venous plexus) to start the cycle again (Fig. 1). A single infected snail has the potential of shedding thousands of cercariae daily for many months. An adult schistosome has an average lifespan of between three years to five years, but it can as well live for 30

years. A single schistosome pair has a theoretical reproduction potential of up to 600 billion schistosomes.[2] The intermediate freshwater snail inhabit calm or slowly moving freshwater lakes, rivers, ponds or streams. The rate of infection in human increases with the duration of time spent in contaminated water.[8] The use of microscope to examine stool and urine is the gold standard for detection (diagnosis) of schistosomal infection. The schistosome eggs are easily seen and identified on microscopy due to having a peculiar size and shape and possession of a lateral or terminal spine.[8]

2.2 Schistosomiasis: A Neglected Tropical Disease of Poverty

Neglected tropical diseases (NTDs) are generally referred to as a collection of chronic, disabling, and physically disfiguring infectious diseases that are found mostly among the poor rural dwellers and some urban population. Studies have revealed that the NTDs are more prominent in sub-Saharan African communities. NTDs have profound negative effects on development in children, outcome of pregnancy and agricultural productivity, thus a key reason why the ‘‘bottom 500 million’’ inhabitants of sub-Saharan Africa continue to be in poverty. This led to the concern for elimination of the NTDs as a major element of the Millennium Development Goals MDGs.[9]

More than seventeen diseases are referred to as NTDs but seven of them have attracted more attention due to their high prevalence and resistance to control worldwide.[10]

NTDs poses great health as well as economic challenges for the poor dwellers of Africa, Latin America as well as Asia. These diseases are the cause of about 534, 000 deaths annually; an estimated 57 million diability-adjusted life-years (DALYs) are

lost annually due to the NTDs. These neglected tropical diseases put great health, social and financial burden on economies of households and governments.[11]

Schistosomiasis is the second most common NTDs after hookworm in Sub-Saharan Africa. Children and young adults bears most of the burden resulting from this disease in Africa. Sub-Saharan Africa accounts for 93 % (192 million) of the world estimated 207 million cases of schistosomiasis. The highest prevalence of this infection is seen in Nigeria (29 million), which is closely followed by United Republic of Tanzania (19 million), Ghana and Democratic Republic of Congo (15 million) making up the top five countries in Africa with schistosomal infection (Fig. 2). However, under-estimation of the true prevalence of schistosomiasis have been reported; it is proposed that the prevalence of schistosomal-related diseases may be more than 400-600 million globally.[12]

2.3 Prevalence of schistosomiasis in some sub-Saharan Africa countries

S. mansoni is the chief cause of clinical abnormalities such as hepatomegaly, splenomegaly, and periportal fibrosis in various sub-Saharan Africa countries. A study in northern Ethiopia in Alamata district revealed an alarming 73.9 % prevalence of schistosomal infection, with presentation of 3.7 % hepatomegaly, 7.4 % splenomegaly, and 12.3 % periportal fibrosis.[13]

A cross-sectional study carried out among 362 school age children in Machakos district in Kenya showed varying intensities of *S. mansoni* infections and its association with hepatomegaly, it was observed that children with infections exhibited stunted growth and wasted muscles compared with children without infection.[14] Cases of

appendicitis-related to schistosomal infections are reported in various studies. Schistosomal appendicitis caused by *S. mansoni* infection is an unusual complication in endemic communities with 0.02 %-6.3 % prevalence, which represent 28.6 % of severe appendicitis cases in such endemic region.[15]

A study carried out in Nigeria reported that 2.3 % of over 1000 cases of appendicitis have schistosome eggs discovered in histological sections, with 56 % of the cases attributable to *S. mansoni*, 26.0 % to *S. haematobium*, and 19.0 % to coinfection by both species.[16] In another study, 4.2 % of appendicitis cases were classified as schistosomiasis of the appendix.[17] Schistosomiasis due to *S. mansoni* is on top of the list of the causes of pulmonary hypertension worldwide especially in areas where schistoasomiasis is endemic.[18]

A survey of school children aged 5-19 years in Mbita and some Islands close to Lake Victoria in Kenya revealed that the communities are highly endemic for *S. mansoni* infection with a high prevalence of 60.5 %.[19] Another survey covering the inhabitants of Lake Rweru in Rwanda indicated that 21.1% of the population screened had intestinal schistosomiasis. This suggests a likely high prevalence of schistosomiasis in the communities visited. Out of the three communities visited, a community had only one case of infection, which can be attributed to the presence of piped water in the commmunity.[20]

A study carried-out in Yewa North Local Government of Ogun State, Nigeria buttressed the fact that schistosomiasis exists among pregnant women. The study population consisted of pregnant women of age range 15-42 years. A prevalence of 20.8 % of *Schistosoma haematobium* infection was observed, with the younger pregnant women at greater risk of the infection, which is in consonant with a

previous report from Tanzania. It was also observed that the prevalence recorded within pregnant mothers in Yewa North is obviously lower compared with earlier reports from similar population groups due to a taboo in that part of the country restricting pregnant women from visiting natural water bodies.[21]

A cross-sectional study of school children of ages 8-17 years in Sengerema District, Nyamatongo ward of North-West Tanzania, revealed an alarming prevalence of 64.3 % *S. mansoni* infection. This prevalence is remarkably high in spite of previous control efforts in the ward such as mass drug administration.[22] An investigative study of in-school and not-in-school children of ages 6-15 years living in communities along the Tono Irrigation Canal in North Ghana revealed a prevalence of 33.2 % of *S. haematobium* infection and a 19.8 % for *S. mansoni*. An overall schistosomiasis prevalence of 47.7 % was reported. It was also observed that there was higher infection rate in male children compared to their female counterpart, which might be due to longer period of contact with contaminated water.[23]

A study consisting adult male and female subjects residing in the Volta Basin of Ghana revealed a 46.5 % prevalence of urinary schiatosomiasis; this report is a good indication that schistosomal infection prevalence is not transient among adults.[24]

An investigative study carried out in Zenu, a surburb community of the capital city of Ghana, revealed a prevalence of 30.7 % incidence of urinary schistosomiasis, among the study population of school children aged 3-16 years. Demographic analysis showed that the proximity of households to a dam is associated with the prevalence observed among the children.[25] A high prevalence (57.6 %) of urinary schistosomiasis and elevated intensity of *S. haematobium* infection (185 eggs/ml)

was also reported among school-age children in Niakhar District, Senegal. This endemicity could be attributed to the inhabitants' great dependence on backwater and ponds for various domestic and recreational activities.[26]

The report of a pilot study in 2010, in the Eastern Cape Province of South Africa, with school-age students as subject revealed an alarming prevalence of 73.3 %. The authors observed that due to the study population size being small, there is need for further study to determine the extent of schistosomiasis in the area.[27] A nation-wide survey of the prevalence of schistosomal infections and soil helminths in school children from Mozambique reported a prevalence of 47.0 % *S. haematobium* infection and 1.0 % *S. mansoni* infection. Further observation from this study showed infection increases with age, with the age group 10-14 years having the highest prevalence of *S. haematobium* infection. It was also observed that across the districts, male children had the higher prevalence compared to their female counterparts. The high prevalence of schistosomiasis was attributed to inadequate water supply in the districts, poor sanitation and a low level of socioeconomic development in most parts of Mozambique.[28]

A cross sectional study carried out in Zarima town, northwest Ethiopia among 319 elementary school children revealed a prevalence of 37.9 % *S. mansoni infection*.

The researchers observed there was an increase in prevalence compared with previous studies in the area despite school deworming programmes. A study done at Barombi Kotto focus, South West Cameroun, revealed an alarming 69.17 % prevalence of *S. haenatobium* infection confirming the area highly endemic. It was noted that most lakes in the area are habitats for the *B. globosus* intermediate host required for disease transmission.[29] Added to this, another study carried out in

Agboville, Cote d'Ivoire among school age children showed a very high prevalence of 85. 3 % and 53.8 % for *S. haematobium* and *S. mansoni* respectively.[30]

2.4 Schistosomal morbidity and mortality in sub-Saharan Africa

The significance of morbidity and mortality resulting from helminthic infections including schistosomiasis has been grossly underestimated in most developing countries.[12] School-aged children, teenagers, women and young adults are mostly hit with the morbidity and mortality associated with schistosomiais. [31] Population studies of schistosoma-infected children revealed that schistosomiasis can cause retardation of growth, fatigue, weakness, impairment of memory and cognitive reasoning and increased risk of anaemia, leading to poor academic performance, thus limiting the potential of infected children.[8] These negative outcomes in the children adds to the socioeconomic burden of the society.[11]

Though schistosomiasis is rarely fatal, it causes long term morbidity such as anaemia which results from bleeding from the urinary and intestinal tracts due to worm invasion and movements, iron deficiency is also an outcome of the disease sequel to nutritional impairment such as nutrient malabsorption and digestive disorder like diarrhea due to infection.[32]

A study in Daekena, in the Republic of Niger an area endemic for urinary schistosomiasis, revealed 41.7 % of the 174 school children having iron deficiency while 57.7 % exhibit anaemia related iron deficiency.[33] A longitudinal study in Burkina Faso among 1727 Burkinabe children of age range 6-17 years revealed a dramatic increase in haemoglobin concentration of children diagnosed with *S. haematobium* infection after chemotherapy compared with baseline concentration a year earlier.[34]

Clinical observation and autopsy strongly indicates that people, specifically elderly patients, die as a result of schistosomal-induced kidney damage.[2] Urogenital schistosomiasis is a key predisposing agent for Human Immunodeficiency Virus (HIV) transmission. Urogenital schistosomiasis in HIV-positive women increases the ease of transmission to male sexual partners, as well as in transmission from a HIV-positive male to his sex partner. It also hastens the progression of the disease in people already infected with the virus by increasing the plasma concentration of the HIV RNA (viral load).[35] A study among Zimbabwean women showed that women with *S. haematobium eggs* in their pap smear had three times risk of having HIV.[36]

Studies on animal models and human subjects have established adverse consequences of schistosomiasis on pregnancy outcomes. *S. haematobium* infection has been linked to placental inflammation, leading to poor birth-outcomes as a result of placental incompetency. Heavy *S. mansoni* infection have also been linked to higher risk of anaemia that might subsequently lead to maternal mortality or low birth weight (LBW) babies. The anaemia may be due to loss of iron in urine and faecal waste. Proinflammatory cytokines resulting from schistosomiasis infection leads to anorexia or loss of appetite in a pregnant woman which might ultimately result in reduced maternal weight gain and thus lead to a baby with. low birth weight.[37]

An observational study in Tanzania showed that there was a high prevalence of *S. mansoni* infection among the pregnant women under study and the high intensity of infection exposes the women to more risk of anaemia during pregnancy.[38] Another study in Tanzania linked the delivery of low birth weight babies (LBW) to infection with parasitic diseases including schistosomiasis during pregnancy.[39] Study has shown an elevated progression of liver fibrosis in people co-infected with

schistosomiasis and hepatitis-C virus in comparison with cases of single infection, which ultimately led to advanced liver diseases.[40]

2.5 Factors determining the continuous transmission of schistosomiasis in sub-Saharan Africa

The continuous transmission of schistosomiasis in sub-Saharan Africa is attributable to various environmental and socio-economic factors such as:

2.5.1 Climatic changes

There is an established link between climatic changes and infectious disease transmission. Schistosomiasis is a typical example of diseases whose local infection and geographical expansion is influenced by climatic changes and global warming.[41] Mangal and co-workers showed that a rise in ambient temperature from 20 ˚C to 30 ˚C will lead to over tenfold increase in the mean burden of *S. mansoni* infection in endemic areas. Whereas at temperature above 30 ˚C a decrease in the disease burden was observed, probably due to the rise in the death rate of the intermediate snail host. They observed that although an increase in disease burden leads to increased morbidity and mortality, there might be a negligible increase in prevalence of the disease.[17] Rainfall patterns also have effect on the transmission of schistosomiasis; in Senegal, the snail specie *Biomphalaria pfeifferi* is responsible for *S. mansoni* transmission during the raining season, while during the dry season *S. haematobium* infection is transmitted by *Bilunus globosus.*[42]

2.5.2 Proximity to water sources

The schistosome parasite requires an avenue wherein there is direct contact between the molluscan intermediate snail and the final human host for transmission of schistosomiasis to take place [43], an estimated 76 % of the sub-Saharan population

live close to various open water bodies which are infested with the intermediate snail host necessary for the transmission of the disease.[1] Various studies have established a direct association between the intensity of the disease and proximity of infected individuals to natural water source such as lakes, rivers and ponds.[44, 45] A study carried out in Blantyre district in Malawi, showed that children whose school are closer to open water bodies have increased risk of infection,[44] this finding is in consonance with the finding of Clennon and co-workers. [45]

2.5.3 Man-made ecological changes

Ecological changes due to man-made construction of irrigation schemes, reservoirs and dams for agricultural purposes and electricity generation are also responsible for continued transmission of schistosomiasis in some in sub-Saharan African countries.[46] Construction of dams led to remarkable increase in cases of urinary schistosomiasis as experienced in some sub-Saharan Africa countries such as Senegal, Cote d'Ivoire, Ghana, Mali, Namibia and Cameroun.[4] Steinmann and co-workers estimated that 13.6 % (106 million) of people vulnerable to schistosomiasis reside close to irrigation schemes and large dam reservoirs.[1]

2.5.4 Socio-economic factors

Socio-economic factors influencing the continuous transmission of the debilitating disease in sub-Saharan countries include poverty occupational activities, poor sanitation and hygiene and non-availability of portable water for domestic use.[2] A World Bank analysis confirmed that majority of sub-Saharan population survives on between US$1.25-US$2 per day.[9] King 2010, postulated a vicious cycle between poverty and schistosomiasis. He explained that poverty compels an individual to utilizing contaminated water sources for his domestic activities therefore getting infected with the disease, on getting sick due to the infection, he becomes unable to

engage in activities to earn his livelihood and thus, poverty persists.[12] Ugbomoiko and co-workers established the link between schistosomiasis and poverty in their cross-sectional study in two peri-urban communities in Osun State, Nigeria. An alarming 62 % prevalence was recorded among 1023 individuals under study.[47]

An analysis by Esrey and co-workers reported on the role of improved water supply and hygiene on disease transmission and incidence concluded that availability of good water and sanitation are necessary for reducing the incidence and prevalence of schistosomiasis and some other water related diseases.[48] Many inhabitants of sub-Saharan countries has limited access to portable water for domestic use, thus they are left with the option of using natural water bodies such as lakes, rivers, ponds and other water sources contaminated with developmental stages of the schistosome parasite. A study carried out in northern Nigeria showed the link between safe water and intensity of urinary schistosomiasis, a higher of infection (88.57 %) was recorded in a community with only pond as source of water for domestic use in comparison with a neighbouring community with borehole which had a prevalence of 0.59 %.[49]

Occupational activities such as fishing and farming are also risk factors for transmission of the disease. Contact with infected water is a vital factor in transmission of infection; women and children get exposed to infection during activities such as laundry, plate washing, water fetching for domestic use and bathing. Recreational activities such as swimming and diving also exposes an individual to infection.[3]

2.6 Chemotherapy for Schistosomiasis and associated fallout

Several drugs have been used in the treatment of schistosoma infection, notable among them are oxamniquine, metrifonate, Artemisinin deivatives and praziquantel.

Artemisinin is an antimalarial, which have been studied to have antischistosomal properties, with the best outcome with artesunate and artemether. These drugs have the capacity to kill juvenile worms of both *S. mansoni* and *S. japonicum*. Trials studies done in China and Cote d'Ivoire established the efficacy of the derivatives of artemisinin. However, there is widespread concern on the use of these drugs in areas where malaria is endemic due to fear of raising resistance to plasmodium.[50]

Praziquantel, a pyrazinoisoquinoline derivative have been shown by randomised controlled trials to be a very safe oral drug for the treatment of schistosomiasis caused by the various schistosome species.[49] This drug is made of a white crystalline bitter tasting compound which is stable in normal storage condition and insoluble in water. Commercially, the drug is available as a racemate mixture with equal 'laevo' and 'dextro' isomers. The 'laevo' part have schistosomicidal activity *in vivo* and *in vitro*. It is mainly available as 600mg crystalline tablets, but the generally recommended dosage is 40mg/kg body weight in a single dose. A 600mg/5ml syrup is also available for small children.[51] Praziquantel is still the best drug for combating infections from all five species of schistosomes afflicting humans, it has a cure rate of 60-90 % in various epidemiological settings.[52]

Millions of people are given the drug yearly in sub-Saharan African nations courtesy of various mass chemotherapy control programmes especially under the Bill and Mellinda Gates Foundation sponsored Schistosomiasis Control Initiative (SCI).[53] In the year 2002, WHO endorsed the drug safe for treating pregnant women as well as

lactating mothers. However, total cures (100%) have scarcely been recorded in areas that are endemic.[54] Praziquantel acts within an hour of oral ingestion but its actual mechanism of action on the adult worm is not yet known. Laboratory studies have

revealed that the drug causes contraction of tegumental vacuoles making the worm to disengage from the wall of the veins and die. The calcium ion channel of the schistosome worm has been indirectly established as the molecular target of the drug.[8] An hypothesis to elucidate the mode of action of this drug states that it inserts itself into the parasites membrane and produces lipid phase transition thus destabilizing the membrane.[52]

Praziquantel negative effects are very mild and these include nausea, vomitting, abdominal pain, malaise and in serious infections, intense colic and bloody diarrhoea soon after treatment, which may be attributed to worm mass movement and antigen release.[2] Praziquantel does not have effect on young worms both *in vivo* and *in vitro*. This significant deficiency is responsible for poor cure rate as well as treatment failure reported in some studies, especially in areas where transmission rate is very high. To overcome this, two courses of drug administration was advocated which gave a better cure rate.[51] Extensive and intensive usage of the drug gave rise to concerns on the emergence of drug-resistant mutants of the schistosomal parasites and thus, the need for research into evolvement of new antischistosomal drugs.[8]

2.7 Schistosomiasis and cancer

Studies in Africa and the Middle East have been carried out which establish the relationship between inflammation resulting from schistosomiasis and squamous cell carcinoma of the bladder.[55] The International Agency for Cancer Research in

association with World Health Organization categorized *S. haematobium* infection as carcinogenic. *S. mansoni* and *S. japonicum* are associated with cervical, liver as well as colorectal carcinomas.[56] Deposition of parasite ova in the bladder leads to intense inflammatory reactions that are linked to the release of oxygen-derived free radicals, which aids the production of carcinogenic N-nitrosamines, and thus malignant transformation. N-nitrosamine formation has been linked to chronic bacterial infection in the urinary bladder due to schistosomal infections.[57] Groeneveld and co-worker stated in their work that eradication of schistosomiasis as well a timely detection of bladder cancer will give room for efficient management of squamous cell carcinoma of the bladder in African patients.[58]

2.8 Combating Schistosomiasis in sub-Saharan Africa

Preventive chemotherapeutic approach using praziquantel is the most common strategy to control schistosomiasis. It actually leads to decrease in schistosomal-related morbidity but high disease reoccurrence and transmission of infection still persisted. Hence, in order to support the benefits of the drug, it must be administered at regular intervals for limitless period of time to prevent return of morbidity.[8]

The World Health Organization reported that as at 2010, 34.8 million people were treated with praziquantel in 30 countries; the drug was made available through yearly donation of 250 million tablets. The WHO set the goal of controlling schistosomiasis-associated morbidity in all endemic countries by the year 2020. It aimed at eliminating the disease in all endemic countries by the year 2025 and to interrupt transmission of schistosomal infections in non-endemic regions like America, Europe, East Mediterranean, South-East Asia, as well as some selected African countries.[59]

Despite sub-Saharan Africa countries having the highest burden of morbidity and mortality associated with the disease, limited efforts have been made in the control of the disease until the introduction of Schistosomiasis Control Initiatives in 2002 (SCI).[60] The SCI successfully instituted national schistosomiasis control programmes in some sub-Saharan nations like Zambia, Mali, Tanzania, Burkina Faso, Niger and Uganda. However, these programmes were not absolutely successful due to the use of vertical mass drug distribution. The coverage of the programme was ineffective, infants and preschool children were not administered the drug.[61]

More so, below 50.0 % of the high risk population were given the drug while individuals who were not infected with the disease were given due to the focus on treating everyone without case finding. Thus, morbidity reduction was very low and cessation of transmission very limited.[62] Before 2013, over US$150 has been exhausted on schistosomiasis and other NTDs in sub Saharan Africa nations and by 2013, there was an increase after the United State of America announced a fresh funding for a initiative for combating NTDs.[63]

Realistically, most SSA countries cannot afford the estimated 1.2 billion tablets of Praziquantel required to treat 400 million individuals per annum for 5 years at a total cost of US$100 million,[64] thus, schistosomiasis control in these countries cannot be achieved without foreign donor fundings.[65] Over the decades, the approach used in sub-Saharan countries is aimed that reducing morbidity of infection,[61] however, there is dire need for sustainable strategies geared towards transmission control and elimination of this impoverishing disease. Vaccines are known to be vital in disease control and eradication; this was evident in the 1978 eradication of small pox.[66] There is a consensus among researchers in the field that effective control of transmission and long term morbidity control can be achieved through the

introduction of antischistosomal vaccines capable of building human immune responses to parasite invasion.[63, 67]

2.8.1 Improvement in water supply, sanitation and hygiene

Improved sanitation and access to good water supply are key factors in schistosomiasis control and elimination.[31] Advancement in water supply and sanitation play a role in general economic development and growth of a country.[68] Campbell and colleagues [69] opined that there is need to progress from emphasis on treatment of morbidity due to infection towards reduction in exposure to infection through the implementation of Water, Sanitation and Hygiene (WASH) intervention policy. Many authorities in sub-Saharan Africa are not dedicated to making available clean water sources for their communities, and without accessibility to clean water, eradication of schistosomiasis in sub-Saharan Africa will continue to be a mirage.

2.8.2 Educating people on behavioural changes

There is need to educate people, both young and old on the implication of voiding their bladder and faecal waste into water bodies. This behavioural change will go a long way in reducing contamination of water sources and thus lead to reduction in the rate of transmission of schistosomiasis and some other water-transmitted diseases. An intensified effort at educating people on the risk of schistosomal infection and transmission is necessary to assist in achieving positive behavioural changes as regard waste elimination and personal exposure to open water sources.[9] Inhabitants of endemic areas need to be encouraged to reduce water contact as much as possible in a bid to reduce schistosomiasis transmission.[70]

2.8.3 Snail control

Though not in use in sub Saharan Africa, the use of molluscicides reduces snail population but rarely eliminate the snails. This is because the snails recolonise their habitat after a while, so regular and long-term application of the chemical is required. Although, this control method was successfully used to achieve the eradication of this disease in Morocco and Japan, it was also used in control programme strategies in Egypt and China. Toxicity of the chemical to fish as well as other water-dwelling organisms is of great concern ecologically and economically. Alternatives to chemical-based molluscicides include the use of plant-based derivatives (e.g endod) and biological control with snail competitors.[31]

2.9 Advances in schistosomal vaccine candidates

Sub-Saharan African researchers need to be aggressively involved in the on-going search for alternative drugs or antischistosomal vaccines to overcome the inherent

challenges posed by this neglected but destructive disease of poverty. The foregoing are some advances in vaccine discovery. Vaccine candidates for schistosomal infections have been in different phases of development. Some vaccines are in pre-clinical trial phase while some are in phase one and others in phase two clinical trials. Examples include Fatty Acid Binding Protein (FABP), Calpain, Triose-phosphate isomerase, and Tetraspanins, GST, and Paramyosin.[6] The Calpain subunit called Sm-p80 has been experimentally tested for its anti-fecundity and anti-infective efficacy. Experimental mice vaccinated with DNA plasmids with Sm-p80 and interleukin-2 (IL-2) provided 59 % protection, while vaccination using plasmid designed with Sm-p80 and interleukin-12 (IL-12) gave 45 % protection. The study indicated that Sm-p80 is a good vaccine candidate for schistosomiasis and further

advancement and optimization is needed for human trials.[71] Triose-phosphate isomerase (TPI) and tetraspanins are another group of antigens with promising vaccine potentials; trials have been done on buffalos to determine the efficacy levels of both vaccines on their own and fused together with heat-shock protein 70 (Hsp70). The vaccine constructs are SjCTP1-Hsp70 and SjC23-Hsp70; the former was reported to be more successful as it produced a worm burden decrease of 51.2 %, liver egg decrease of 61.5 %, faecal waste egg reduction of 52.1 % and 52.1 % decrease in hatching faecal miracidia.[72]

Pearson and colleagues [73] also succeeded in producing a chimeric vaccine named Sm-TSP-2. Their study indicated this vaccine antigen could be safely used against schistosome adult worms and liver eggs as well as for the hookworm parasite. The 28 kDa GST recombinant protein (Sh28GST) from *S. haematobium* has been experimentally tested as regards safety, immunogenicity as well as toxicity.

The vaccine candidate molecule, paramyosin (rSj97) recombinant fraction has also been discovered to give protection together with immunogenicity in BALB/c mice. The full length recombinant paramyosin protein elicited antibody responses in humans with re-infection resistance.[6] However, till the moment, scientific hurdles and socioeconomic factors remain major challenges towards successful transition of the above mentioned vaccine candidates into forms that could be used for schistosomiasis intervention in sub-Saharan Africa and other regions of the world.[67]

2.10 Conclusion

For many decades, human schistosomiasis has been of great health concern globally and mostly in communities of sub-Saharan Africa. The economic loss, disability and agony caused by the various species of schistosomes cannot be over-

emphasized. Concerted efforts should be made towards the elimination of this disease of poverty in all countries in sub-Saharan Africa. The various authorities in sub-Saharan countries should make provision of clean and safe water available for the citizens. There should be constant awareness programmes to educate the populace on the prominence of sanitation and hygiene in achieving disease control and the effect of schistosomiasis on the quality of life and general well-being of people.

Most importantly, it is hoped that this review will generate new interest among researchers in sub-Saharan Africa towards further scientific discoveries to combat schistosomiasis, in terms of alternative drugs to address re-infection and resistance, vaccine discovery, formulation of point-of-care diagnostics, snail control and other germane discoveries that will serve as guide to voting-off schistosomiasis in sub-Saharan Africa and some other endemic regions of the world.

References

1. Steinmann P, Keiser J, Bos R, Tanner M, Utzinger J. Schistosomiasis and water resources development: systematic review, meta-analysis, and estimates of people at risk. Lancet Infect Dis. 2006; 6:411-25.
2. Gryseels B, Polman K, Clerinx J, Kestens L. Human schistosomiasis. Lancet. 2006; 368:1106-118.
3. World Health Organization. WHO Schistosomiasis fact sheet. 2014. Available from http://www.who.int/mediacentre/factsheets/fs115/en [Accessed 10 April 2014].
4. Olveda DU, Li Y, Olveda RM, Lam AK, PChau TN, Harn DA, et al. Bilharzia: pathology, diagnosis, management and control. Trop Med Surg. 2013; 1:135.
5. Chitsulo L, Engels D, Montresor A, Savioli L. The global status of schistosomiasis and its control. Acta Tropica. 2000; 77(1):41-51.
6. Utzinger J, Raso G, Brooker S, De Savigny D, Tanner M, Ombjerg N, et al. Schistosomiasis and neglected tropical diseases: towards integrated and sustainable control and a word of caution. Parasitology. 2009; 136:1859-874.
7. Ekpo UF, Oluwole AS, Abe EM, Etta HE, Olamiju F, Mafiana CF. Schistosomiasis in infants and pre-school children in sub-Saharan Africa: implication for control. Parasitology. 2012; 139:835- 41.
8. Gray DJ, Ross AG, Li Y, McManus DP. Diagnosis and management of schistosomiasis. BMJ British Medical Journal. 2011; 342: d2651.
9. Hotez PJ, Kamath A. Neglected tropical diseases in sub-Saharan Africa: review of their prevalence, distribution, and disease burden. PloS Negl Trop Dis. 2009; 3:e412.

10. Deribe K, Meribo K, Gebre T, Hailu A, Ali A, Aseffa A, et al. The burden of neglected tropical diseases in Ethiopia, and opportunities for integrated control and elimination. Parasit Vectors. 2012; 5:240.

11. Conteh L, Engels T, Molyneux D. Socioeconomic aspect of neglected tropical diseases. Lancet. 2010; 375:239-47.

12. King CH. Parasites and Poverty: The Case of Schistosomiasis. Acta Trop. 2010; 113:95-104.

13. Abebe N, Erko B, Medhin G, Berhe N. Clinico-epidemiological study of Schistosomiasis mansoni in Waja-Timuga, District of Alamata, northern Ethiopia. Parasit Vectors. 2014; 7(1):158.

14. Corbett EL, Butterworth AE, Fulford AJC, Ouma JH, Sturrock RF. Nutritional status of children with schistosomiasis mansoni in two different areas of Machakos District, Kenya. T Roy Soc Trop Med H. 1992; 86(3):266-73.

15. Elbaz T, Esmat T. Hepatic and intestinal schistosomiasis: review. J Adv Res. 2013; 4:445-52.

16. Badmos KB, Komolafe AO, Rotimi O. Schistosomiasis presenting as acute appendicitus. E Afr Med J. 2007; 83(10):528-32. Gali BM, Nggada HA, Eni EU. Schistosomiasis of the appendix in Maiduguri. Trop Doctor. 2006; 36:162-63.

17. Gali BM, Nggada HA, Eni EU. Schistosomiasis of the appendix in Maiduguri. Trop Doctor. 2006; 36:162-63. Lapa M, Dias B, Jardim C, Fernandes CJ, Dourado PM, Figuiredo M, et al. Cardiopulmonary manifestations of hepatosplenic schistosomiasis. Circulation. 2009; 119:1518-523.

18. Lapa M, Dias B, Jardim C, Fernandes CJ, Dourado PM, Figuiredo M, et al. Cardiopulmonary manifestations of hepatosplenic schistosomiasis. Circulation. 2009; 119:1518-523.

19. Odiere MR, Rawago FO, Obok M, Secor WE, Karanja DM, Mwinzi PN, et al. High prevalence of schistosomiasis in Mbita and its adjacent islands of Lake Victoria, western Kenya. Parasit Vectors. 2012; 5:278.

20. Ruberanziza E, Mupfasoni D, Karibushi B, Kabera M, Karema C, Nyatanyi T, et al. A recent update of schistomiasis mansoni endemicity around Lake Rweru. Rwanda Med J. 2010; 68:5-9.

21. Salawu OT, Odaibo AB. Schistosomiasis among pregnant women in rural communities in Nigeria. Int J Gynaecol Obstet. 2013; 122:1-4.

22. Mazigo HD, Waihenya R, Mkoji GM, Zinga M, Ambrose EE, Jahanpour OF, et al. Intestinal schistosomiasis: prevalence, knowledge, attitude and practices among school children in an endemic area of North Western Tanzania. J Rural Trop Public Health. 2010; 9:53-60.

23. Anto F, Asoala V, Adjuik M, Anyorigiya T, Oduro A, Akazili J, et al. Water contact activities and prevalence of schistosomiasis infection among school-age children in communities along an irrigation scheme in rural Northern Ghana. J Bacteriol Parasitol. 2013; 4:177.

24. Yirenya-Tawiah DR, Annang T, Otchere J, Bentum D, Edoh D, Amoah C, et al. Urinary schistosomiasis among adults in the Volta Basin of Ghana: Prevalence, knowledge and practices. J Trop Med Parasitol. 2013; 34:1-16.

25. Tetteh-Quarcoo PB, Attah SK, Donkor ES, Nyako M, Minamor AA, Afutu E, et al. Urinary schistosomiasis in children-still a concern in part of the Ghanaian capital city. Open J Med Microbiology. 2013; 3:151-58.

26. Senghor B, Diallo A, Sylla SN, Doucoure S, Ndiath MO, Gaayeb L, et al. Prevalence and intensity of urinary schistosomiasis among school children in the district of Niakhar, region of Fatick, Senegal. Parasit Vectors. 2014; 7:5.

27. Meents EF, Boyles TH. Schistosoma haematobium prevalence in school children in the rural Eastern Cape Province South Africa. S Afr J Epidemiol Infect. 2010; 25:28-29.

28. Augusto G, Nalá R, Casmo V, Sabonete A, Mapaco L, Monteiro J. Geographic distribution and prevalence of schistosomiasis and soil-transmitted helminths among school children in Mozambique. Am J Trop Med Hyg 81 Suppl. 2009; 5:799-803.

29. Nkengazong L, Njiokou F, Asonganyi T. Two years impact of single praziquantel treatment on urinary schistosomiasis in the Barombi Kotto focus, South West Cameroon. J Parasit Vector Biol. 2013; 5(6):83-89.

30. Müller I, Coulibaly JT, Fürst T, Knopp S, Hattendorf J, Krauth SJ, et al. Effect of schistosomiasis and soil-transmitted helminth infections on physical fitness of school children in Côte d'Ivoire. PLoS Neglect Trop Dis. 2011; 5(7):e1239.

31. Rollinson D, Knopp S, Levitz S, Stothard JR, Tchuem Tchente LA, Garba A, et al. Time to set agenda for schistosomiasis elimination. Acta Trop. 2013; 128:423-440.

32. Hürlimann E, Yapi RB, Houngbedji CA, Schmidlin T, Kouadio BA, Silué KD, et al. The epidemiology of polyparasitism and implications for morbidity in two rural communities of Côte d'Ivoire. Parasit Vectors. 2014; 7(1):81.

33. Prual A, Daouda H, Develoux M, Sellin B, Galan P, Hercberg S. Consequences of Schistosoma haematobium infection on the iron status of schoolchildren in Niger. Am J Trop Med Hyg. 1992; 47(3):291-97.

34. Koukounari A, Gabrielli AF, Touré S, Bosqué-Oliva E, Zhang Y, Sellin B, et al. Schistosoma haematobium infection and morbidity before and after large-scale administration of praziquantel in Burkina Faso. J Infect Dis. 2007; 196(5):659-69.

35. Mbabazi PS, Andan O, Fitzgerald DW, Chitsulo L, Engels D, Downs JA. Examining the relationship between urogenital schistosomiasis and HIV infection. PLoS Negl Trop Dis. 2011; 5:e1396.

36. Kjetland EF, Ndhlovu PD, Gomo E, Mduluza F, Midzi N, Gwanzura L, et al. Association between genital schistosomiasis and HIV in rural Zimbabwean women. Aids. 2006; 20(4):593-600.

37. Friedman JF, Mital P, Kanzaria HK, Olds OR, Kurtis JD. Schistosomiasis and pregnancy. Trends Parasitol. 2007; 23:159-64.

38. Ajanga A, Lwambo NJ, Blair L, Nyandindi U, Fenwick A, Brooker S. Schistosoma mansoni in pregnancy and associations with anaemia in northwest Tanzania. T Roy Soc Trop Med H. 2006; 100(1):59-63.

39. Dreyfuss ML, Msamanga GI, Spiegelman D, Hunter DJ, Urassa EJ, Hertzmark E, et al. Determinants of low birth weight among HIV-infected pregnant women in Tanzania. Am J Clin Nutr. 2001; 74(6); 814-26.

40. El-Awady MK, Youssef SS, Omran MH, Tabil AA, El Garf WT, Salem AM. Soluble egg antigen of *Schistosoma haematobium* induces HCV replication in PBMC from patients with chronic HCV infection. BMC Infect Dis. 2006; 6:91.

41. Mas-Coma S, Valero MA, Bargues MD. Climate change effects on trematodiases, with emphasis on zoonotic fascioliasis and schistosomiasis. Vet Parasitol. 2009; 163(4):264-80.

42. Ernould JC, Ba K, Sellin B. The impact of the local water development programme on the abundance of the intermediate hosts of schistosomiasis in three villages of the Senegal River delta. Ann Trop Med Parasit. 1999; 93:135-45.

43. Brooker S. Spatial epidemiology of human schistosomiasis in Africa: risk models, transmission dynamics and control. T Roy Soc Trop Med H. 2007; 101(1):1-8.

44. Kapito-Tembo AP, Mwapasa V, Meshnick SR, Samanyika Y, Banda D, Bowie C, et al. Prevalence distribution and risk factors for Schistosoma hematobium infection among school children in Blantyre, Malawi. PLoS Neglect Trop Dis. 2009;3(1):e361

45. Clennon JA, Mungai PL, Muchiri EM, King CH, Kitron U. Spatial and temporal variations in local transmission of schistosoma haematobium in Msambweni, kenya. Am J Trop Med Hyg. 2006; 75(6):1034- 041.

46. Fenwick A, Webster JP, Bosque-Oliva E, Blair L, Fleming FM, Zhang Y, et al. The Schistosomiasis Control Initiative (SCI): rationale, development and implementation from 2002–2008. Parasitology. 2009; 136(13):1719-730.

47. Ugbomoiko US, Ofoezie IE, Okoye IC and Heukelbach J. Factors associated with urinary schistosomiasis in two peri-urban communities in south-western Nigeria. Ann Trop Med Parasitol. 2010; 104(5):409-19.

48. Esrey SA, Potash JB, Roberts L, Shiff C. Effects of improved water supply and sanitation on ascariasis, diarrhoea, dracunculiasis, hookworm infection, schistosomiasis, and trachoma. B World Health Organ. 1991; 69(5):609.

49. Kanwai S, Ndams IS, Kogi E, Abdulkadir JS, Gyam ZG, Bechemagbor A. Cofactors influencing prevalence and intensity of *Schistosoma haematobium*

infection in sedentary Fulani settlements of Dumbi, Igabi LGA, Kaduna State, Nigeria. Sci World J. 2011; 6(2):15-19.

50. Fenwick A, Webster JP. Schistosomiasis: challenges for control, treatment and drug resistance. Curr Opin Infect Dis. 2006; 19:577-82.

51. Doenhoff MJ, Hagan P, Cioli D, Southgate V, Pica-Mattoccia L, Botros S, et al. Praziquantel: its use in control of schistosomiasis in sub-Saharan Africa and current research needs. Parasitology. 2009; 136:1825–835.

52. El-Ridi AF, Tallima HA. Novel therapeutic and prevention approaches for schistosomiasis: review. J Adv Res. 2013; 4:467-78.

53. World Health Organization. WHO Schistosomiasis: progress report 2001-2011 and strategic plan 2012-2020. 2013. Available from http://www.who.int/neglected_disease/en [Accessed 15 April 2014].

54. Doenhoff MJ, Cioli D, Utzinger J. Praziquantel: mechanism of action, resistance and new derivatives for schistosomiasis. Curr Opin Infect Dis. 2008; 21:659-67.

55. Mostafa MH, Helmi S, Badawi AF, Tricker AR, Spiegelhalder B, Preussmann R. Nitrate, nitrite and volatile N-nitroso compounds in the urine of *Schistosoma haematobium* and *Schistosoma mansoni* infected patients. Carcinogenesis. 1994; 15:619-25.

56. Samaras V, Rafailidis PI, Mourtzoukou EG, Peppas G, Falagas ME. Chronic bacterial and parasitic infections and cancer: a review. J Infect Dev Ctries. 2010; 4**(**5):267-81.

57. Zaghloul MS. Bladder cancer and schistosomiasis. J Egypt Natl Canc Inst. 2012; 24:151-59.

58. Groeneveld AE, Marszalek WW, Heyns CF. Bladder cancer in various population groups in the greater Durban area of KwaZulu-Natal, South Africa. British J urology. 1996; 78(2): 205-208.

59. World Health Organization. WHO Schistosomiasis: progress report 2001-2011 and strategic plan 2012-2020. 2013. Available from http://www.who.int/neglected_disease/en [Accessed 15 April 2014].

60. van der Werf MJ, de Vlas SJ, Brooker S, Looman CW, Nagelkerke NJ, Habbema JD, et al. Quantification of clinical morbidity associated with schistosome infection in sub-Saharan Africa. Acta Trop. 2003; 86:125-39.

61. Fenwick A, Webster JP, Bosque-Oliva E, Blair L, Fleming FM, Zhang Y, et al. The Schistosomiasis Control Initiative (SCI): rationale, development and implementation from 2002–2008. Parasitology. 2009; 136(13):1719-730.

62. Gray DJ, McManus DP, Li Y, Williams GM, Bergquist R, Ross AG. Schistosomiasis elimination: lessons from the past guide the future. Lancet Infect Dis. 2010; 10(10):733-36.

63. Bergquist R, Utzinger J, McManus DP. Trick or treat: the role of vaccines in integrated schistosomiasis control. PLoS Neglect Trop Dis. 2008; 2(6):e244.

64. Hotez PJ, Fenwick A. Schistosomiasis in Africa: an emerging tragedy in our new global health decade. PLoS Neglect Trop Dis. 2009; 3(9):e485.

65. Gray DJ, McManus DP, Li Y, Williams GM, Bergquist R, Ross AG. Schistosomiasis elimination: lessons from the past guide the future. Lancet Infect Dis. 2010; 10(10): 733-36.

66. Geddes AA. The history of smallpox. Clin Dermatol. 2006; 24:152-57.

67. Beaumier CM, Gillespie PM, Hotez PJ, Bottazzi ME. New vaccines for neglected parasitic diseases and dengue. Translational Res. 2013; 162(3):144-55.

68. Knopp S, Person B, Ame SM, Mohammed KA, Ali SM, Khamis IS, et al. Elimination of schistosomiasis transmission in Zanzibar: baseline findings before the onset of a randomized intervention trial. PLoS Negl Trop Dis. 2013; 7(10):e2474.

69. Campbell SJ, Savage GB, Gray DJ, Atkinson J-AM, Magalhaes RJS, Nery SV, et al. Water, Sanitation, Hygiene (WASH): A critical component for sustainable soil-transmitted helminth and schistosomiasis control. PLoS Negl Trop Dis. 2014; 8:2651.

70. Stothard JR, Mook P, Mgeni AF, Khamis IS, Khamis AN, Rollinson D. Control of urinary schistosomiasis on Zanzibar (Unguja Island): a pilot evaluation of the educational impact of the Juma na Kichocho health booklet within primary schools. Mem. Inst. Oswaldo Cruz. 2006; 101:119-24.

71. Siddiqui AA, Phillips T, Charest H, Podesta RB, Quinlin ML, Pinkston JR, et al. Enhancement of Smp-80 (large subunit of calpain) induced protective immunity against *Schistosoma mansoni* through co-delivery of interleukin-2 and interleukin-12 in a DNA vaccine formulation. Vaccine. 2003; 21:2882-889.

72. Da'dara AA, Li YS, Xiong T, Zhou J, Williams GM, McManus DP, et al. DNA-based vaccines protect against zoonotic schistosomiasis in water buffalo. Vaccine. 2008; 26:3617-625.

73. Pearson MS, Pickering DA, McSorley HJ, Bethony JM, Tribolet L, Dougall AM, et al. Enhanced protective efficacy of a chimeric form of the schistosomiasis vaccine antigen *Sm*-TSP-2. PLoS Neglect Trop Dis. 2012; 6:e1564.

74. Oyinloye B, Adenowo F, Gxaba N, Kappo A. The promise of antimicrobial peptides for treatment of human schistosomiasis. Curr Drug Targets 2014; 15(9):852-59.

Figure Legends:

Fig. 2.1: Life cycle of Schistosoma 1. Definitive host**,** 2. Schistosome eggs released in urine or faeces of definitive host, 3. Free-swimming miracidia from eggs, 4. Intermediate host, 5. Cercaria (penetrates skin, loses its tail and transforms into schistosomulum), 6. Paired adult worm (schistosomulum migrates to the hepatic portal system; adult worms mature in pairs in the veins surrounding the bladder, intestines or liver. They produced eggs, which majority are eliminated in urine or feces of definitive host to the environment. The eggs hatch liberating miracidia to restart the life cycle.[74]

Fig. 2.2: Map of Africa showing ranking of estimated schistosomiasis burden in sub-Saharan African (SSA) countries. Schistosomiasis prevalence in SSA is documented to be 192 million, which is 93% of the total global prevalence of the disease. A total of 29 million people is infected by this disease in Nigeria, 19 million in Tanzania, 15 million each in Democratic Republic of Congo and Ghana, while Mozambique with 13 million people completes the top five countries with highest prevalence in SSA. This prevalence figures by country is represented by a pie chart as a ratio of the total disease burden in SSA using a normalization figure of 0.0069. The prevalence figures is taken from Hotez and Kamath (2009), while prevalence map was generated using the ArcGIS software.

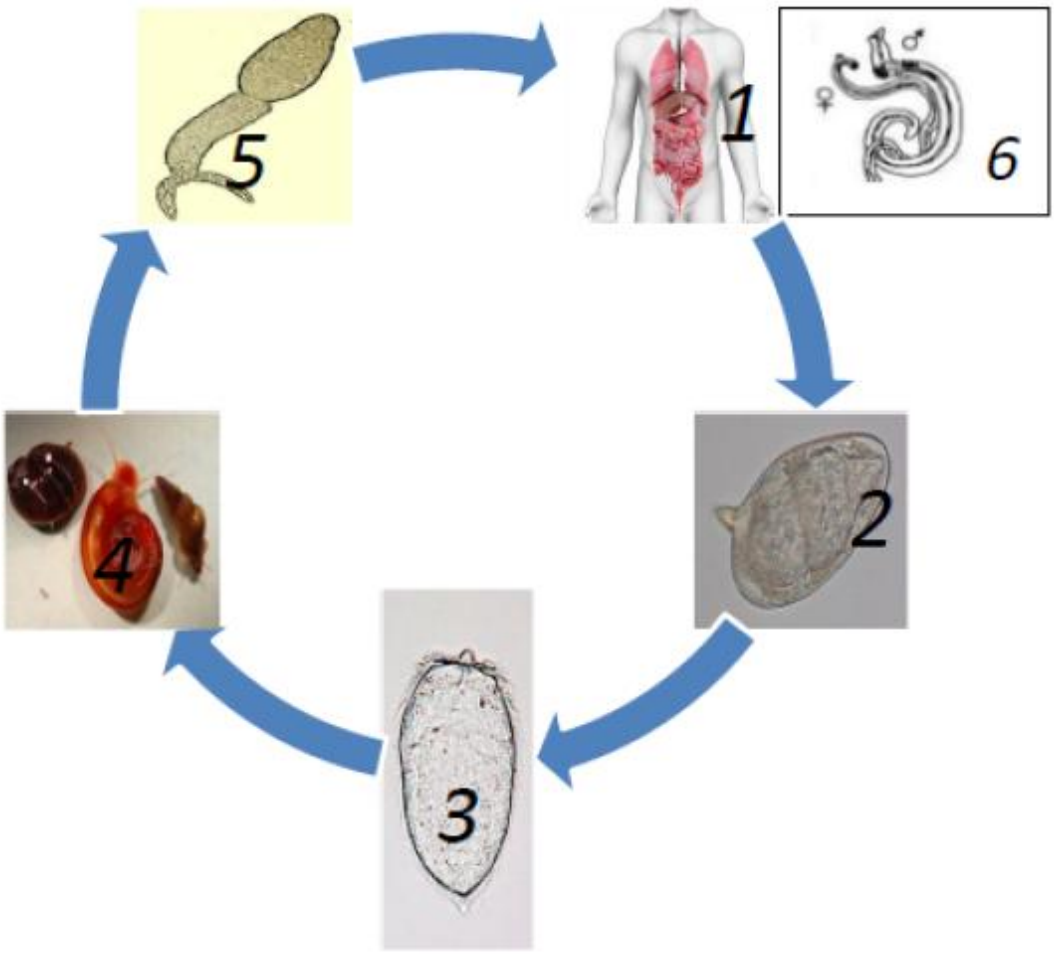


Fig. 2.1: Lifecycle of Schistosome. Developmental stages are shown. The labelling is presented in the figure legend.

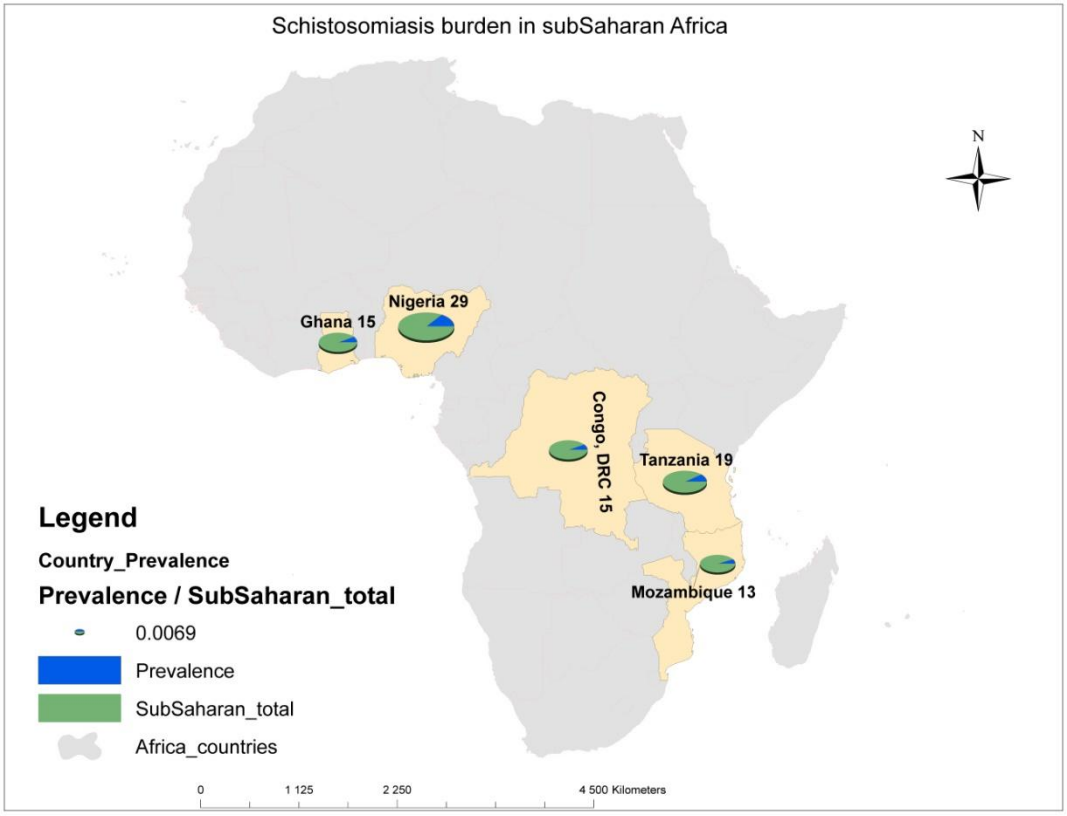


Fig. 2.2: Map showing prevalence of schistosomiasis. Disease still prevalent in sub-Saharan Africa countries.

CHAPTER THREE

STRUCTURAL CHARACTERIZATION OF RECOMBINANT *S. MANSONI* UNIVERSAL STRESS G4LZI3 PROTEIN

3.1 Introduction

Due to the evolution of genomic information mining, numerous databases have been generated which are committed to accommodating genomic data of different species. Assigning structure as well as function to every single gene and gene product (protein) of all organisms epitomises a chief task in the present post-genomic era. Three-dimensional (3D) protein structures are very crucial for the rational design of various kinds of biological experiments such as site-directed mutagenesis and structure-based detection of specific inhibitors. Yet, the amount of proteins characterized structurally is minute in comparison to the number of identified protein sequences. Thousands of experimentally determined protein structures have been deposited in the Protein Data Bank, while millions of protein sequences are deposited in the UniProt protein knowledge database (Westbrook *et al.,* 2003; Bairoch *et al.,* 2005).

Diverse computational approaches for modelling 3D structures of proteins have been evolved to solve the disparity between the large number of protein sequence available and the amount of 3D structures. More so, the 3D structure of protein is better conserved than the sequence and it is relatively easier to identify homologous proteins with a known structure (template) for a particular protein sequence (target) (Arnorld *et al.,* 2006).

Several bioinformatics tools and a collection of online databases are available for achieving the prediction of various protein structure and functions (Kanehisa and Bork, 2003; Yang *et al.,* 2015). In the course of *in-silico* studies, homologous proteins are identified and then matched based on their functional and structural

properties to identify unfamiliar ones. Protein structure prediction with computational analysis can encompass search for similar sequence, multiple sequence alignments, domain identification and characterization, secondary structure prediction, prediction of solvent accessibility, protein fold recognition, building of 3D model and model validation (Edwards and Cottage, 2003; Yang *et al.,* 2015).

Additionally, bioinformatics is also useful in the examination, establishment, and ranking of biological candidate genes. Protein-protein interactions may be used to detect disease-related complexes. This approach puts the prospective disease-causing proteins in a functional context in relation to other identified or unidentified disease-associated proteins; hence systematic research on such complexes can reveal new candidate genes for disease intervention (; Lage *et al.,* 2007; Banasik *et al.,* 2011). Furthermore, the data obtained from *in-silico* studies can be utilized as a guide to laboratory experiments to confirm predicted properties which might subsequently lead to the discovery of novel therapeutic proteins required for disease intervention.

3.2 Materials

The databases, tools and software used for computational analysis in this study are presented in the Table 3.1a and 3.1b.

Table 3.1a: Databases and their respective URL for *in-silico* analysis

Databases	URL
National Center for Biotechnology Information (NCBI)	http://www.ncbi.nlm.nih.gov/
Expert Protein Analysis System (ExPASy)	http://www.expasy.org/
EMBL-EBI	http://www.ebi.ac.uk/
Swiss-model	http://swissmodel.expasy.org
Protein Data Bank (PDB)	http://www.rcsb.org/pdb/home/home.do
Conserved Domain Database (CDD)	https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi
Sable	http://sable.cchmc.org/
GalaxyRefine	http://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE
ASAView	http://www.abren.net/asaview/

Table 3.1b: Software and tools, and their respective URL for *in-silico* analysis

Software and Tools	URL
Basic Local Alignment Search Tool (BLAST)	http://blast.ncbi.nlm.nih.gov/Blast.cgi
Constraint-based Multiple Alignment Tool (COBALT)	www.ncbi.nlm.nih.gov/tools/cobalt/cobalt.cgi
ProtParam	http://web.expasy.org/protparam/
PSI-blast based secondary structure prediction (PSIPRED)	http://bioinf.cs.ucl.ac.uk/psipred/
YASPIN secondary structure prediction	http://www.ibi.vu.nl/programs/yaspinwww/
RAMPAGE	http://mordred.bioc.cam.ac.uk/~rapper/rampage.php
Chimera	http://www.cgl.ucsf.edu/chimera/
PredictProtein (PP)	www.predictprotein.org/
3DLigandSite	http://www.sbg.bio.ic.ac.uk/3dligandsite/3dligand_report
Pymol	https://www.pymol.org/
Antigenic determinant	http://imed.med.ucm.es/Tools/antigenic.pl

3.3 Methods

3.3.1 Retrieval of G4LZI3 sequence

The NCBI database (Geer *et al.,* 2010) was searched to obtain the coding sequence of the Universal Stress Protein G4LZI3. The mRNA sequence consisting of the coding region (exons) without the non-coding regions (introns) was retrieved. The mRNA sequence was thereafter translated into amino acid sequence on ExPASy (Artimo *et al.,* 2012) using the translate tool.

3.3.2 Nucleotide Analysis

The target protein sequence was blasted for non-redundant protein sequences from the NCBI database using the PSI-BLAST tool (Johnson *et al.,* 2008; Boratyn *et al.,* 2013). The megablast option for extremely similar sequences was selected. A graphical summary of the BLAST results as well as a visual representation of the results for the G4LZI3 sequence was obtained. Additionally, homologous sequences that produced significant alignments with the G4LZI3 protein sequence were also derived.

3.3.3 Identification of conserved domain

Conserved domain analysis of the full length G4LZI3 was done on the NCBI Conserved Domain Database (CDD) using the artificial neural network (ANN) for the recognition of domains in protein sequences. The domain recognition was established based on the distribution of BLAST scores for previously computed known domain groups in the NCBI database (Murvai *et al.,* 2001).

3.3.4 Multiple sequence alignment

Multiple sequence alignment of G4LZI3 was done with COBALT on NCBI website. Twenty selected homologous sequences of the query protein (G4LZI3) were aligned with the query sequence in order to rewrite the selected sequences in a way that similar features end up in the same column. The aim behind the multiple sequence alignment is to position nucleotides in the same column due to their similarity based on specific criteria such as functional similarity, structural similarity, sequence similarity and evolutionary similarity (Notredame, 2002; Claverie and Notredame, 2006). In this study, the objective is sequence similarity, this implies that the nucleotides found in the same columns produced alignments with maximum similarity.

3.3.5 Generation of phylogenetic tree

The evolutionary distance between homologous sequences was visualized through the generation of phylogenetic tree. The tree was produced with Blast-pairwise alignment on the NCBI server. The parameters used for generating the tree include; Neighbourhood joining tree method with a maximum sequence difference of 0.85.

3.3.6 Analysis of the physico-chemical properties of G4LZI3

ProtParam (Gasteiger *et al.,* 2005) an online tool on the ExPASy website was used to compute a histogram revealing the percent amino acid composition as well as the physico-chemical properties of G4LZI3 using the protein sequence. Aside the amino acid composition, other physicochemical properties computed with the ProtParam tool include the molecular weight, theoretical isoelectric point, extinction coefficient, atomic composition, estimated half-life, instability index (pI), aliphatic index as well as grand average of hydropathicity (GRAVY).

3.3.7 Secondary structure prediction

Prediction of the secondary structure of G4LZI3 was done using SABLE prediction server, PSIPRED secondary structure prediction tool and YASPIN online server {see appendix for YASPIN G4LZI3 secondary structure prediction} (Jones, 1999; Porollo *et al.,* 2004; Buchan *et al.,* 2013). SABLE prediction server was used to achieve the prediction by scanning the submitted protein sequence against a large sequence database with the use of PSI-BLAST. Thereafter, the profile produced by PSI-BLAST was processed by the neural network secondary structure prediction program. Alpha-helices, beta-strands and coils found in the protein were then presented graphically. PSIPRED is a secondary structure prediction method which integrates two feed-forward neural networks that perform an analysis on output acquired from PSI-BLAST (Position Specific Iterated-BLAST). It is used to predict a protein secondary structure from its primary sequence.

3.3.8 Homology Modelling and 3D Structure Determination

The homology model of G4LZI3 was generated using Swiss-model tool. The Swiss-model workspace is an integrated web-based modelling system used to generate a homology model for the protein. Library of experimentally determined protein structures was searched to identify suitable template. Based on the sequence alignment between the G4LZI3 protein and the template structure, a 3D model for the G4LZI3 protein was generated. Furthermore, the Swiss-model server was also used to estimate the reliability of the built 3D structure through the QMEAN score. Thereafter, the model was viewed with PyMol (The PyMol Molecular Graphics system, Version 1.8 Schrodinger, LLC).

3.3.9 Confirmation of the accuracy of homology model

The predicted 3D homology models of G4LZI3 protein generated with Swiss-model was validated for accuracy using RAMPAGE. The RAMPAGE tool was used to generate Ramachandran plots (Ramachandran *et al.,* 1963; Lovell *et al.,* 2003) which evaluates the protein PDB coordinate models and stereochemistry of the model.

3.3.10 Structure refinement of G4LZI3 3D homology model

Structural refinement of the modelled structure was done using the GalaxyRefine web server for protein structure refinement (Heo *et al.,* 2013). The structure generated was visualised and further analysed with Chimera (Pettersen *et al.,* 2004).

3.3.11 Assessment of solvent accessibility

Accessible surface area of G4LZI3 was estimated with the ASAView server (Shander *et al.,* 2004). The server offers graphical depiction of solvent accessibility of amino acid residues present in proteins with known structures. The 3D model of G4LZI3 was used on the server to compute the solvent accessibility. The absolute surface area (ASA) of individual amino acid residue provided by DSSP (database of secondary structure assignments) program was converted to relative values of ASA. Two types of plots were generated; a Spiral Plot which makes it possible to quickly notice the surface residues in a protein, and a Bar Plot which show relative solvent accessibility of amino acid residues in form of a bar chart. Amino acid residues are organised in the order in which they exist in the original structure.

3.3.12 Analysis of binding sites present on G4LZI3

The binding regions present on the protein were determined computationally on the PredictProtein server with two methods; ISIS2 which predicts protein-protein binding

sites and SomeNA (which predicts polynucleotide binding site). ISIS2 is a machine learning-based procedure which detects interacting residues from protein sequence. It involves using transient protein-protein interfaces from the complexes of experimentally determined 3D structures by combining predicted features with evolutionary information. On the other hand, SomeNA is based on the use of a set of artificial neural network to distinguish between residues of proteins partaking in polynucleotide binding and residues of those that are non-binding (Ofran and Rost, 2007).

3.3.13 Determination of G4LZI3 subcellular localization

LOCTree2 algorithm was used on the PredictProtein server to predict the subcellular localization of G4LZI3. To achieve this, LOCTree2 retrieved a PSI-BLAST profile of G4LZI3 sequence through the PredictProtein pipeline. The profile obtained was utilized for a PSI-BLAST search of a close homologue in the database of proteins that have been experimentally determined. The annotation of the homologue was then transferred to the query protein (G4LZI3) automatically (Goldberg *et al.,* 2012).

3.3.14 Assessment of antigenic determinants/epitopes

The presence and position of antigenic epitopes on G4LZI3 was evaluated with the antigenic peptide prediction tool on the bioinformatics server of Immunomedicine Group of the Universidad Complutense de Madrid (Alvarez *et al.,* 2016). The program predicts the segments within the protein sequence which are expected to be antigenic by stimulating an antibody response. Predictions are based on a method which involve the use of data which reflects the amino acid residues in experimentally determined segmental epitopes. The accuracy of this method for predicting antigenic determnants is 75 % (Kolaskar and Tingankar, 1990; Alvarez *et al.,* 2016).

3.3.15 Gene ontology prediction

Gene ontology (GO) entailed a prediction of the likely molecular function, cellular component and biological process in which G4LZI3 could be involved. The GO prediction was done on the Predictprotein online server. It entails the use of a method known as Metastudent, which involves prediction of Gene Ontology terms through the use of already annotated homologous proteins. Firstly, a PSI-BLAST query was performed on the database of all proteins with known functions. The result was thereafter sent to each base classifier that compiles the GO terms of the BLAST hits, their cut-off score (e-values) as well as sequence identities into a list of predicted GO terms.

Thereafter, the meta-classifier finally gathers all predictions from the base classifiers and pools them together in an optimum way. Metastudent accomplishes its state-of-the-art accuracy through extremely optimized free parameters namely; PSI-BLAST, e-Value cut-off, scoring scheme parameters, number of iterations and by concentrating only on the up-to-date experimentally determined GO term annotations (Hamp *et al.,* 2013).

3.4 Results

3.4.1 Retrieved sequences for G4LZI3

Fig. 3.1 shows the DNA sequence with a total of 184 amino acids retrieved from the NCBI nucleotide data base. The initiation codon is highlighted in red while the stop codon is highlighted in purple. The total number of nucleotide present in the sequence is 55bp. The FASTA format of the protein sequence of G4LZI3 is presented in Fig. 3.2.

NCBI Reference Sequence: Smp_076400

>gi|350645018|emb|CCD60301.1| Universal stress protein G, putative [*Schistosoma mansoni*]

ATGTCAGAAACAGAACAACCTTCGACTTCCGACGGATTAGATATTGGAGAAACGAAAGGAAC
TATATCAATGACTGATGCAACACGTAAAGTATTAATGCCGGTCGATGGAAGTGAACATAGTG
AAAGAGCATTCAATTGGTATATGGATAATGTTATGAAAATAACTGATGGTTTATATCTAGTT
CATATTGTAGAACCATTATCACAAGGATTGAATTATAATCTTGCAAGTAAATCTCCCTCTAT
AAAAGATGATTTTTCAAAACATCTAAATTCACTTGTTGAATCTGGTCGTGCATTACGTGCTA
AATTTTTTACACGTTGTGAAGATAGTGGTTTATCAGCAAGATTTACCATACACGTTGGTACT
AAACCAGGTGAAAATATTGTACGTATCGCCCATGAACATGGAGTTGATTTAGTTATTATTGG
AAATCGTGGTATTGGAACAGTAAAGAGAACATTTTTAGGTAGTGTAAGTGATTATGTTTTAC
ATCATGCCAATGTTCCTGTTATTATAATACCACCACCTAAACATCCAAAGAAAAAATAA

Fig. 3.1: DNA sequence of Universal Stress Protein G4LZI3. The initiation codon is highlighted in red while the stop codon is highlighted in purple.

MSETEQPSTSDGLDIGETKGTISMTDATRKVLMPVDGSEHSERAFNWYMDNVMKI
TDGLYLVHIVEPLSQGLNYNLASKSPSIKDDFSKHLNSLVESGRALRAKFFTRCEDS
GLSARFTIHVGTKPGENIVRIAHEHGVDLVIIGNRGIGTVKRTFLGSVSDYVLHHANV
PVIIIPPPKHPKKK

Fig. 3.2: FASTA format of G4LZI3. The protein sequence in FASTA format is used as starting point for most computational analysis.

3.4.2 BLAST results for the nucleotide sequence

Fig. 3.3 is a graphical representation of the BLAST results for the G4LZI3 sequence; the top red bar indicated the query sequence. Every other bar represents a portion of a different sequence which is similar to the query sequence alongside the region of the sequence where the similarity occurs. Fig. 3.4 shows a segment of the list of homologous sequences that produced significant alignments with the G4LZI3 protein sequence. The e-values as shown in the table are the expected values; it is the number of times the database match might have occurred randomly. The value offers a criterion that gives more objectivity over the percentage-of-similarity or the identity values (Claverie and Notredame, 2006). Hence, a good match can be

regarded as one which is highly unlikely to happen by chance. The lower the e-value the higher the authenticity of the selected sequence.

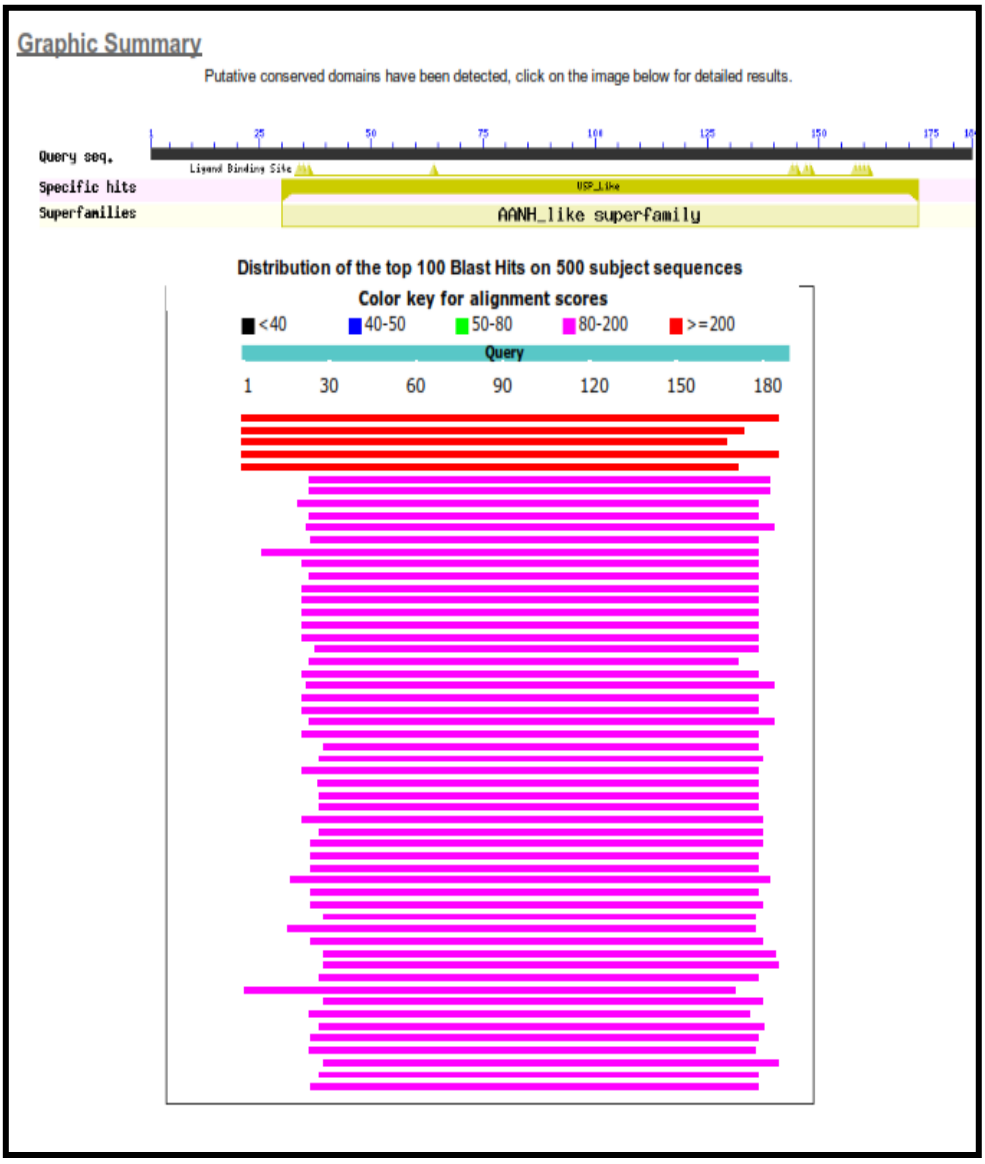
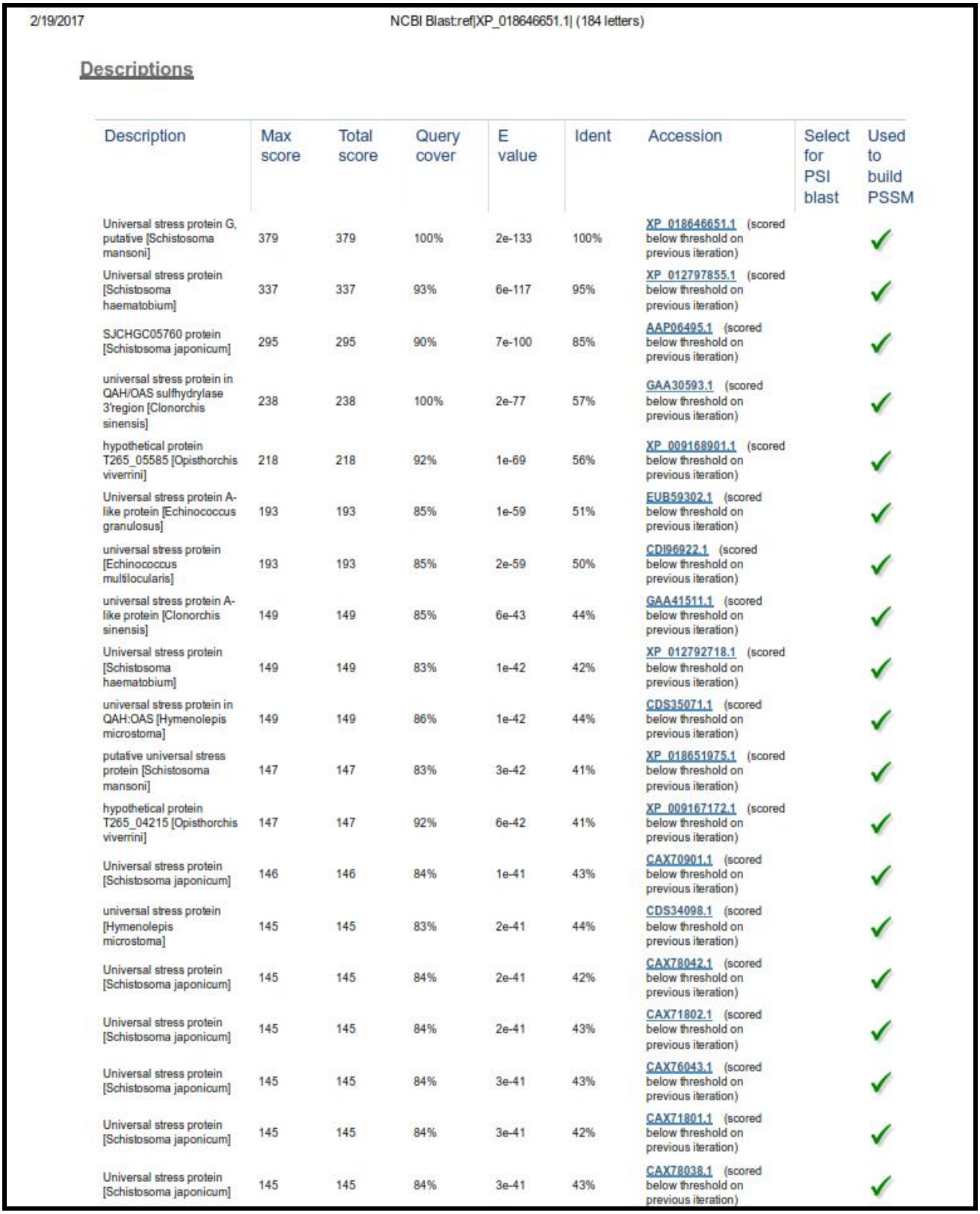


Fig. 3.3: Graphical summary of BLAST results using PSI-Blast iteration. The top red bar represents the query sequence. Other bars represent a portion of other sequences similar to the query.



2/19/2017 NCBI Blast:ref|XP_018646651.1| (184 letters)

Descriptions

Description	Max score	Total score	Query cover	E value	Ident	Accession	Select for PSI blast	Used to build PSSM
Universal stress protein G, putative [Schistosoma mansoni]	379	379	100%	2e-133	100%	XP_018646651.1 (scored below threshold on previous iteration)		✓
Universal stress protein [Schistosoma haematobium]	337	337	93%	6e-117	95%	XP_012797855.1 (scored below threshold on previous iteration)		✓
SJCHGC05760 protein [Schistosoma japonicum]	295	295	90%	7e-100	85%	AAP06495.1 (scored below threshold on previous iteration)		✓
universal stress protein in QAH/OAS sulfhydrylase 3'region [Clonorchis sinensis]	238	238	100%	2e-77	57%	GAA30593.1 (scored below threshold on previous iteration)		✓
hypothetical protein T265_05585 [Opisthorchis viverrini]	218	218	92%	1e-69	56%	XP_009168901.1 (scored below threshold on previous iteration)		✓
Universal stress protein A-like protein [Echinococcus granulosus]	193	193	85%	1e-59	51%	EUB59302.1 (scored below threshold on previous iteration)		✓
universal stress protein [Echinococcus multilocularis]	193	193	85%	2e-59	50%	CDI96922.1 (scored below threshold on previous iteration)		✓
universal stress protein A-like protein [Clonorchis sinensis]	149	149	85%	6e-43	44%	GAA41511.1 (scored below threshold on previous iteration)		✓
Universal stress protein [Schistosoma haematobium]	149	149	83%	1e-42	42%	XP_012792718.1 (scored below threshold on previous iteration)		✓
universal stress protein in QAH:OAS [Hymenolepis microstoma]	149	149	86%	1e-42	44%	CDS35071.1 (scored below threshold on previous iteration)		✓
putative universal stress protein [Schistosoma mansoni]	147	147	83%	3e-42	41%	XP_018651975.1 (scored below threshold on previous iteration)		✓
hypothetical protein T265_04215 [Opisthorchis viverrini]	147	147	92%	6e-42	41%	XP_009167172.1 (scored below threshold on previous iteration)		✓
Universal stress protein [Schistosoma japonicum]	146	146	84%	1e-41	43%	CAX70901.1 (scored below threshold on previous iteration)		✓
universal stress protein [Hymenolepis microstoma]	145	145	83%	2e-41	44%	CDS34098.1 (scored below threshold on previous iteration)		✓
Universal stress protein [Schistosoma japonicum]	145	145	84%	2e-41	42%	CAX78042.1 (scored below threshold on previous iteration)		✓
Universal stress protein [Schistosoma japonicum]	145	145	84%	2e-41	43%	CAX71802.1 (scored below threshold on previous iteration)		✓
Universal stress protein [Schistosoma japonicum]	145	145	84%	3e-41	43%	CAX76043.1 (scored below threshold on previous iteration)		✓
Universal stress protein [Schistosoma japonicum]	145	145	84%	3e-41	42%	CAX71801.1 (scored below threshold on previous iteration)		✓
Universal stress protein [Schistosoma japonicum]	145	145	84%	3e-41	43%	CAX78038.1 (scored below threshold on previous iteration)		✓

Fig. 3.4: A segment of protein sequences that produced significant alignments with G4LZI3. Homologous sequences of G4LZI3 with their accession numbers.

3.4.3 Multiple sequence alignment (MSA)

The alignment result from COBALT is presented in Fig. 3.5a and Fig. 3.5b while Table 3.2 depicts the summary of the homologous sequences that produced the alignment. Multiple sequence alignment with no gaps is coloured blue or red. Wherever 50 % of the sequence consists of gaps, they are presented in grey uppercase while greater than 50 % are in grey lowercase. The red colour indicates highly conserved columns while the blue shows less conservation. The numbers represent the relative entropy threshold that needs to be attained for an alignmentcolumn to be presented in red. A high number shows greater degree of conservation.

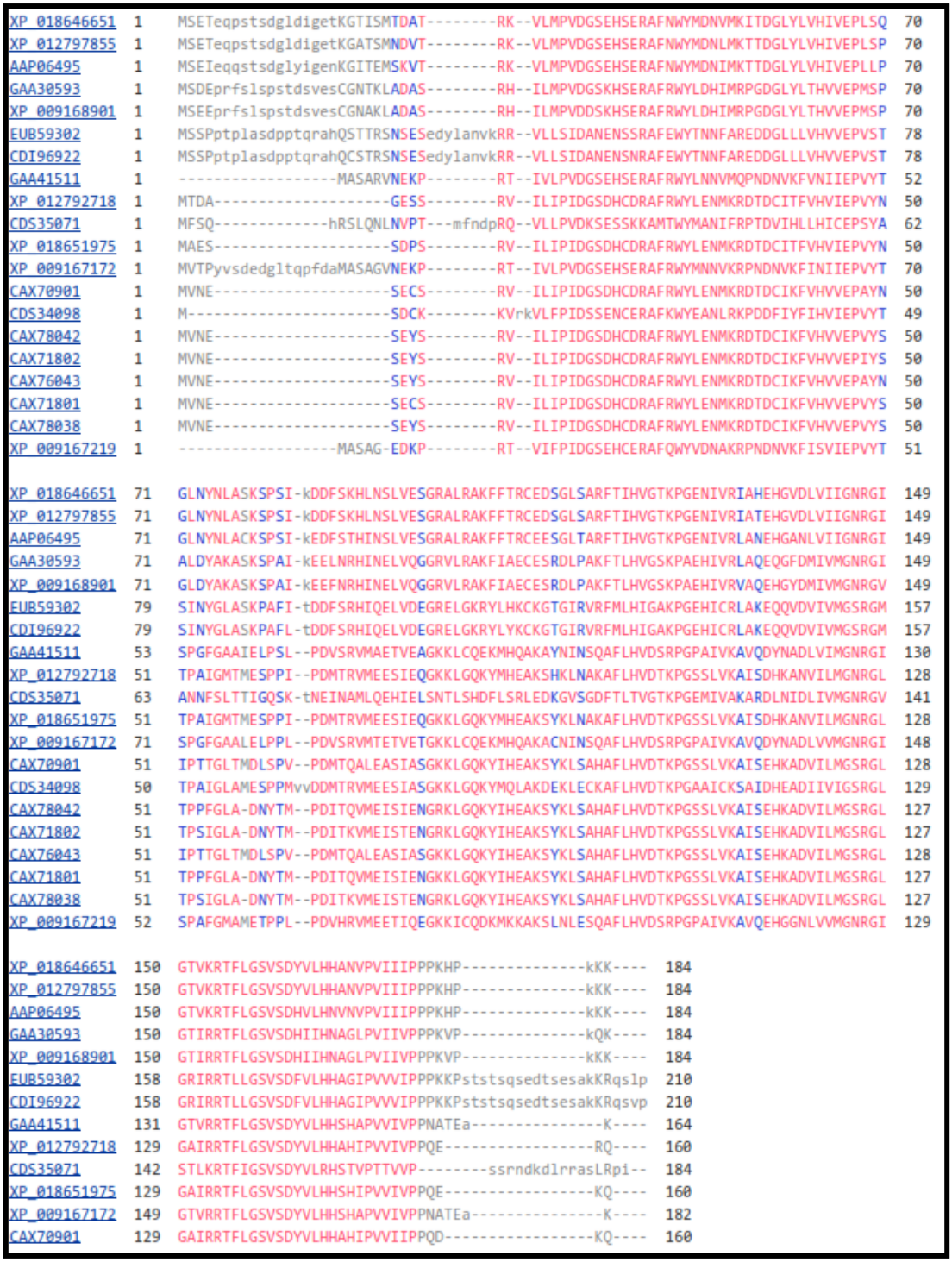


Fig. 3.5a: Multiple sequence alignment of 20 selected homologous protein sequences from NCBI. The red colour indicates highly conserved columns while the blue shows less conservations.

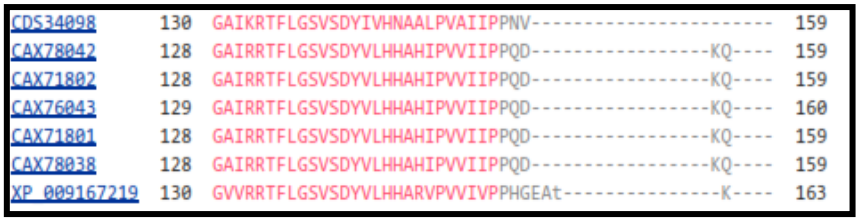


Fig. 3.5b: (continued): Multiple sequence alignment of 20 selected homologous protein sequences from NCBI. The red colour indicates highly conserved columns while the blue shows less conservations.

Table 3.2: List of 20 homologous sequences that produced significant alignment. Shown are the accession number and brief description of the source.

Accession number	Description
XP_018646651.1	Universal stress protein G, putative [Schistosoma mansoni]
XP_012797855.1	Universal stress protein [Schistosoma haematobium]
AAP06495.1	SJCHGC05760 protein [Schistosoma japonicum]
GAA30593.1	Universal stress protein in QAH/OAS sulfhydrylase 3'region [Clonorchis sinensis]
XP_009168901.1	Hypothetical protein T265_05585 [Opisthorchis viverrini]
EUB59302.1	Universal stress protein Alike protein [Echinococcus granulosus]
CDI96922.1	Universal stress protein [Echinococcus multilocularis]
GAA41511.1	Universal stress protein Alike protein [Clonorchis sinensis]
XP_012792718.1	Universal stress protein [Schistosoma haematobium]
CDS35071.1	Universal stress protein in QAH:OAS [Hymenolepis microstoma]
XP_018651975.1	Putative universal stress protein [Schistosoma mansoni]
XP_009167172.1	Hypothetical protein T265_04215 [Opisthorchis viverrini]
CDS34098.1	Universal stress protein [Hymenolepis microstoma]
CAX78042.1	Universal stress protein [Schistosoma japonicum]
CAX71802.1	Universal stress protein [Schistosoma japonicum]
CAX76043.1	Universal stress protein [Schistosoma japonicum]
CAX71801.1	Universal stress protein [Schistosoma japonicum]
CAX78038.1	Universal stress protein [Schistosoma japonicum]
XP_009167219.1	Hypothetical protein T265_13453 [Opisthorchis viverrini]

3.4.4 Analysis of phylogenetic relationship

Phylogenetic tree presented in Fig. 3.6 shows the evolutionary relationship between homologous sequences of G4LZI3. The homologous sequences are shown to be phylogenetically closely related because they form close cluster in a clade.

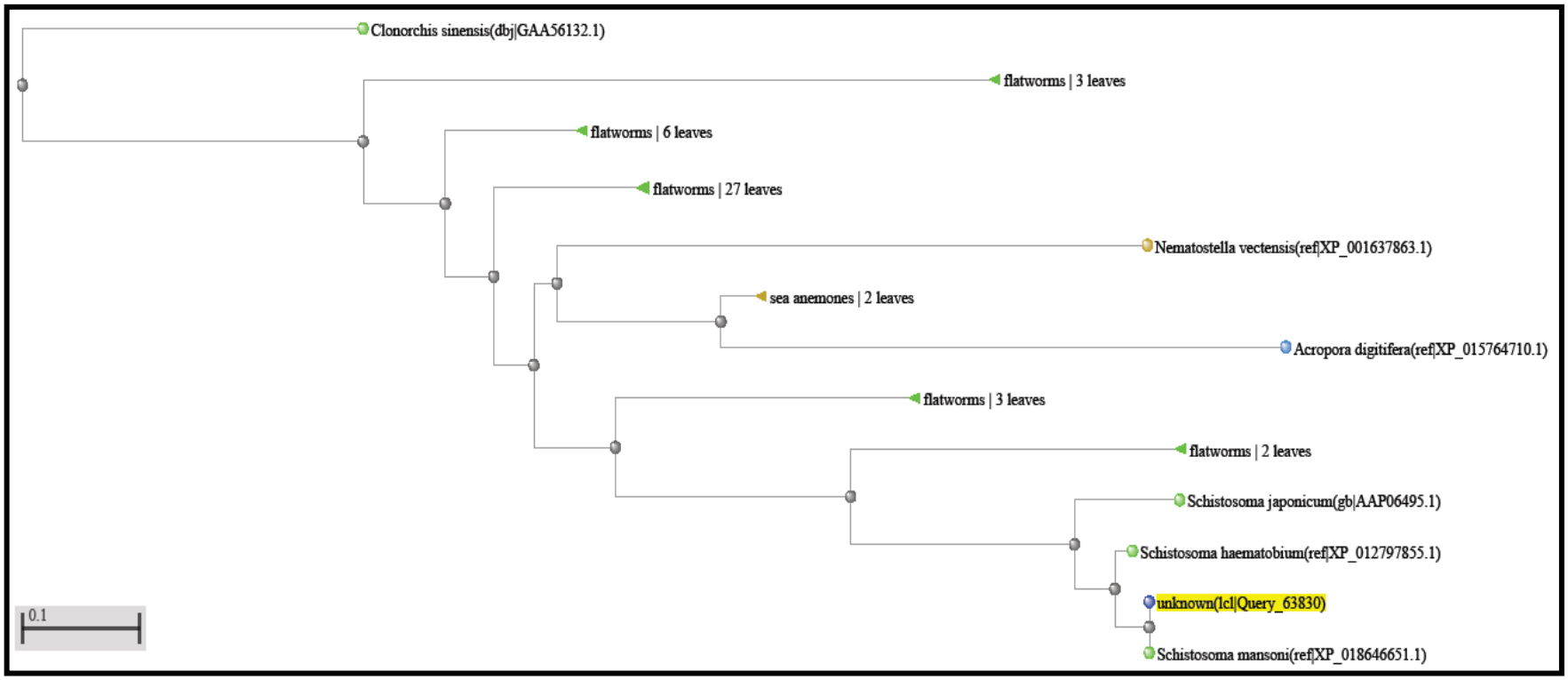


Fig 3. 6: Phylogenetic tree for G4LZI3 homologues. The query sequence, G4LZI3 is highlighted in yellow. The close closely clustered clades shows evolutionary relationship between the homologues.

3.4.5 Identification of conserved domains

Conserved domain analysis done on CDD showed that the protein has the following domains; Usp-like domain, Usp domain, UspA domain, NaH-Antiporter C domain and STK-N domain. Table 3.3 shows the domain hits, their accession numbers, intervals and their e-values. Fig 3.7 depicts a screenshot of the CDD with the domains shown.

Table 3:3: List of domain hits in G4LZI3. The accession number, interval and e-values are also shown.

Name	Accession	Interval	e-value
USP_Like	cd00293	30-172	5.52e-26
Usp	pfam00582	29-172	4.10e-24
UspA	COG0589	29-172	1.00e-18
NaH_Antiporter_C	cd01988	130-172	8.33e-03
STK_N	cd01989	35-168	9.91e-03

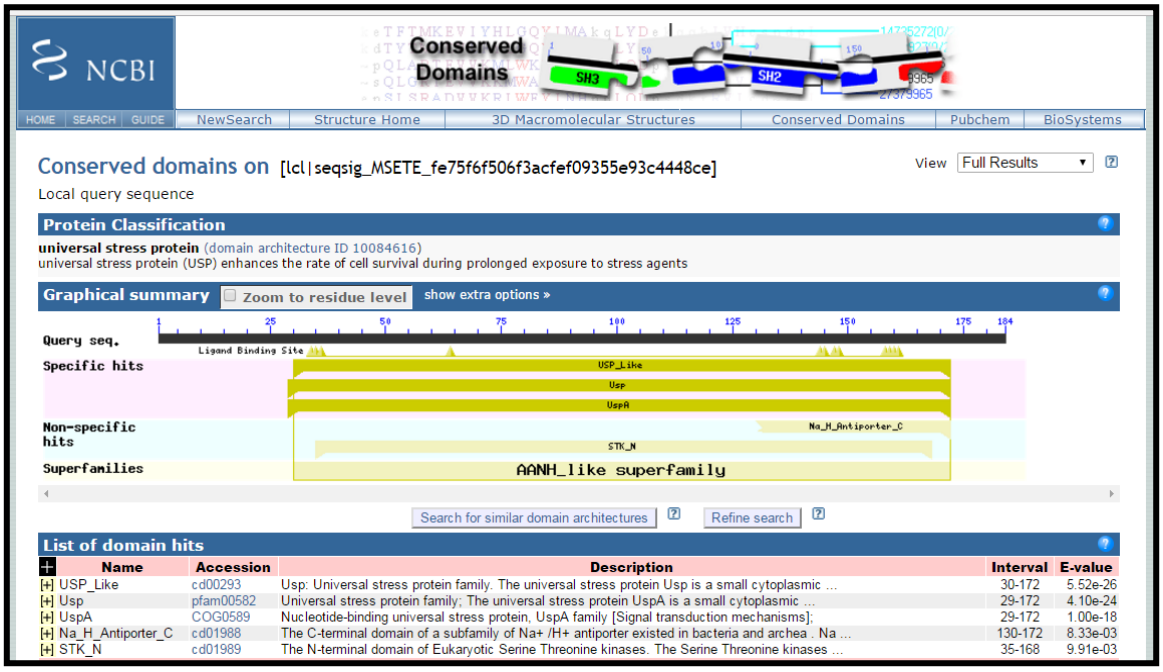


Fig. 3.7: Screen shot of the conserved domain of G4LZI3. Domain analysis of the protein revealed five domain hits

3.4.6 Amino acid composition and physicochemical parameters of G4LZI3

Result from the ExPASy server showed that the G4LZI3 protein sequence consists of 184 amino acid residues with all the 20 amino acids contributing to give the protein an average molecular weight of 20.3 kDa. Analysis suggests that the protein is more hydrophobic due to the amino acid composition; there are more non-polar (hydrophobic) amino acid residues (50.6 %) compared to the polar (hydrophilic) amino acid residues (49.4 %). Fig. 3.8 shows the most abundant amino acid is Serine with 9.2 % abundance trailed by Glycine and Valine both with an abundance of 8.2 %. Tryptophan and Cysteine contributes the lowest predicted abundance 0.5 % each.

The total number of atoms present in G4LZI3 is predicted to be 2861. The predicted molecular formula is $C_{897}H_{1435}N_{253}O_{270}S_{6.}$ Isoelectric point of G4LZI3 is predicted to be 7.92. A pI greater than 7 suggests that the protein is slightly alkaline in nature. The total number of negatively charged residues (Asp + Glu) is 21 while the total number of positively charged residues (Arg + Lys) is 22, indicating that the protein possesses an overall positive net charge, hence if purification to homogeneity is required, it can be accomplished with the application of cation exchange chromatography using a negatively charged column. The protein is predicted to have an instability index of 34.97 suggesting that the protein is relatively stable. The computed half-life is 30 hours in mammalian reticulocytes (*in vitro*), greater than 20 hours in yeast (*in vivo*), and greater than 10 hours in *E. coli* (*in vivo*). The half-life of a protein confirms the instability factor by suggesting how long it can survive without being degraded after it has been expressed in a foreign host. On the whole, this information is very beneficial because it reveals there is a minimal possibility of protein degradation in the course of structure-function studies. The extinction

coefficient is stated to be 11 460 and an estimated

absorbance of 0.1 % (=1 g/l) that is 0.565, assuming all Cys residues are reduced. It must be noted that the low abundance of Tryptophan (0.5 %), Cysteine (0.5 %) as well as Tyrosine (2.2 %) contributed to the low value of extinction coefficient recorded. Additionally, the *in silico* result showed that G4LZI3 protein has a moderately high aliphatic index of 87.34 and a low Grand Average Hydropathy (GRAVY) of -0.305.

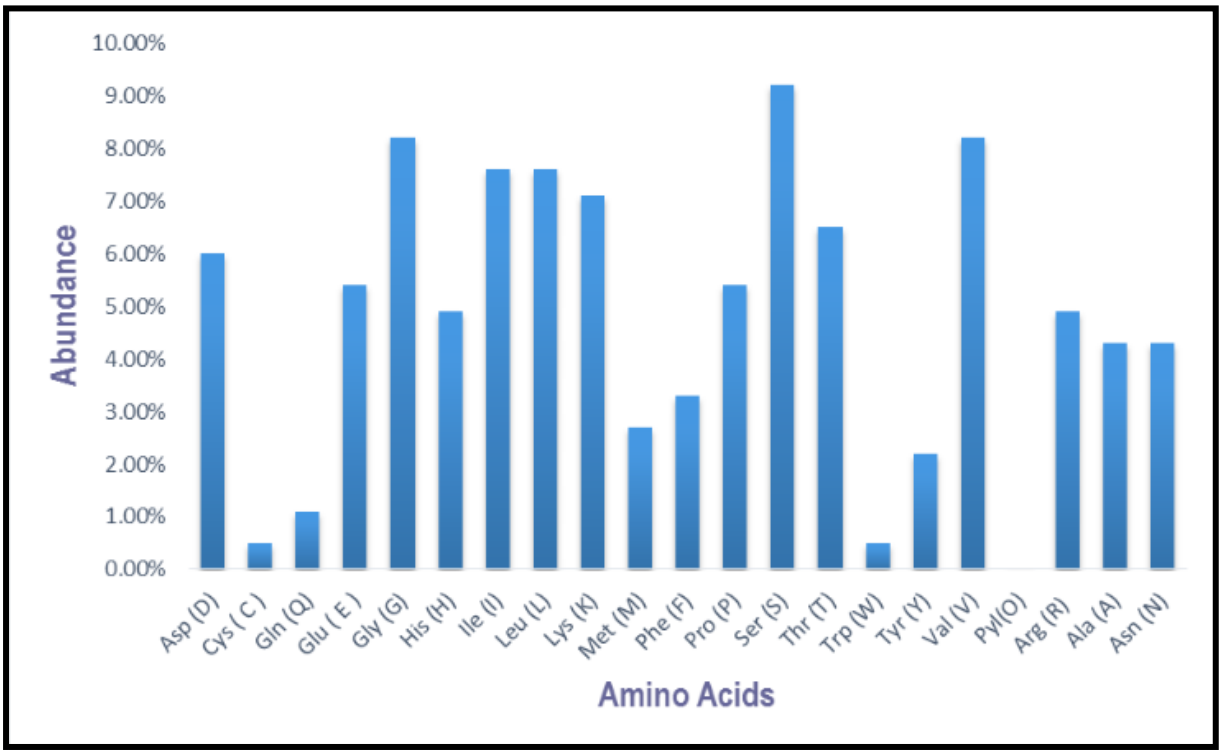


Fig. 3.8: Graphical representation of the abundance of twenty amino acid present in G4LZI3. Serine has the highest abundance while cysteine and tryptophan has the least abundance.

3.4.7 Secondary structure prediction of G4LZI3

Results from the secondary structure prediction using SABLE server indicates that G4LZI3 consists of five α-helices and six β-strands. The graphical representation of the secondary structure prediction is presented in Fig. 3.9. However, the PSIPRED secondary structure prediction in Fig. 3.10 gave a conflicting result in comparison with the SABLE server prediction; PSIPRED prediction shows that there are four helices and five beta-strands in the G4LZI3 protein.

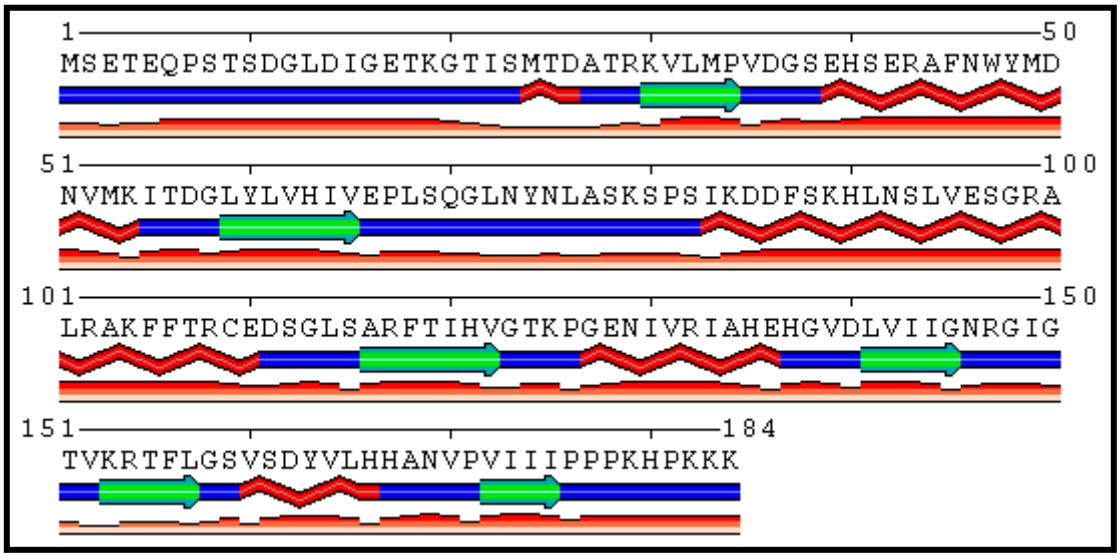


Fig. 3.9: Secondary structure prediction of G4LZI3 done on Sable prediction server..

Five helices and six strands are predicted.

Key: Green arrows represent a helix

Red zig-zaggy lines represent strands

Blue thick lines represent coils

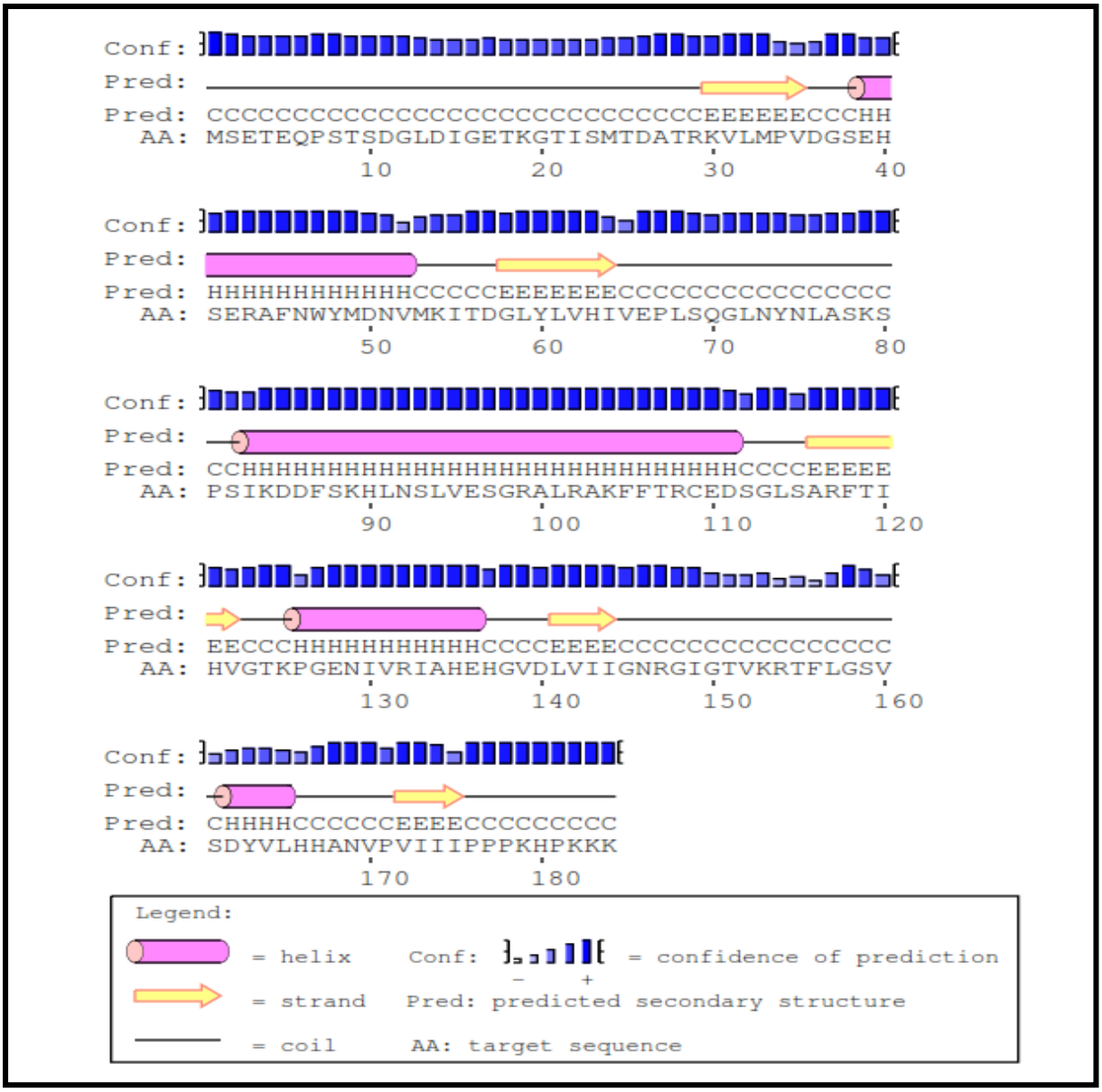


Fig. 3.10: Secondary structure prediction of G4LZI3 using PSIPRED. G4LZI3 is predicted to consist of four helices and five strands.

3.4.8 Homology modelling and surface representation model of G4LZI3

Fig. 3.11 is the tertiary structure of G4LZI3 as predicted by the Swiss-model workplace. The template used to build the model is a homologous universal stress protein MJ0577 from *Methanocaldococcus jannaschii* (Kim *et al.,* 2015). The sequence identity of the homology model with the template is 27.89 %. Fig. 3.12 shows the QMEAN score plot which estimates the reliability of the model; the value estimated is -1.4 which shows that the model is a reliable one. It must be noted that the higher the value of QMEAN score, the more reliable the model. The local estimate plot of the protein 3D model based on the QMEAN score is also shown in Fig. 3.13.

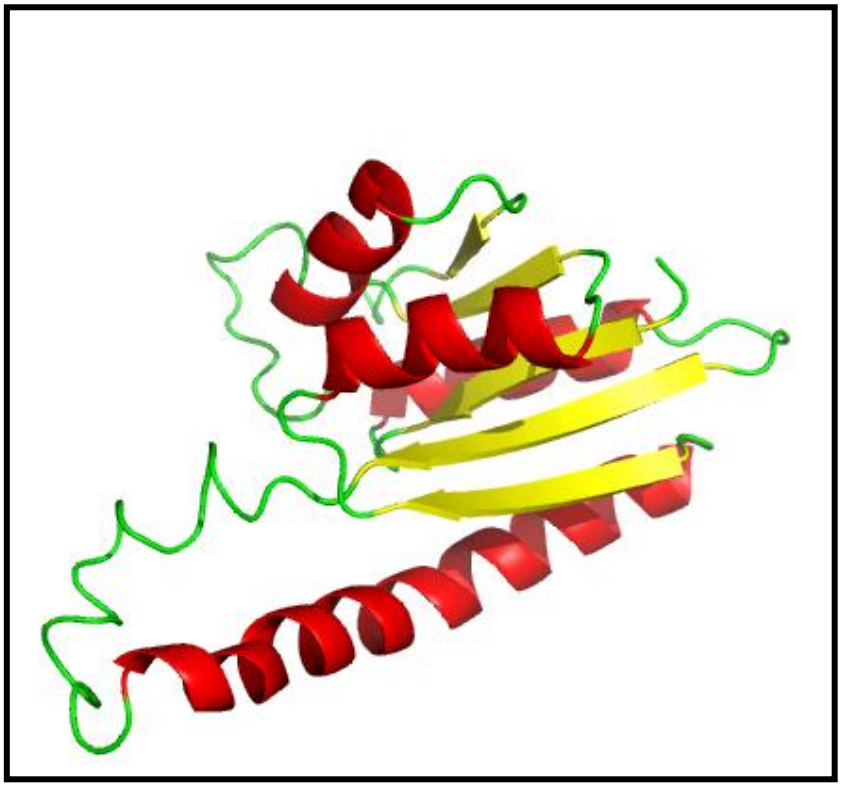


Fig. 3.11: 3D model of Universal Stress Protein G4LZI3 using Swiss-model viewed with PYMOL. The secondary structure elements are made up of four α-helices and five parallel β-sheets.

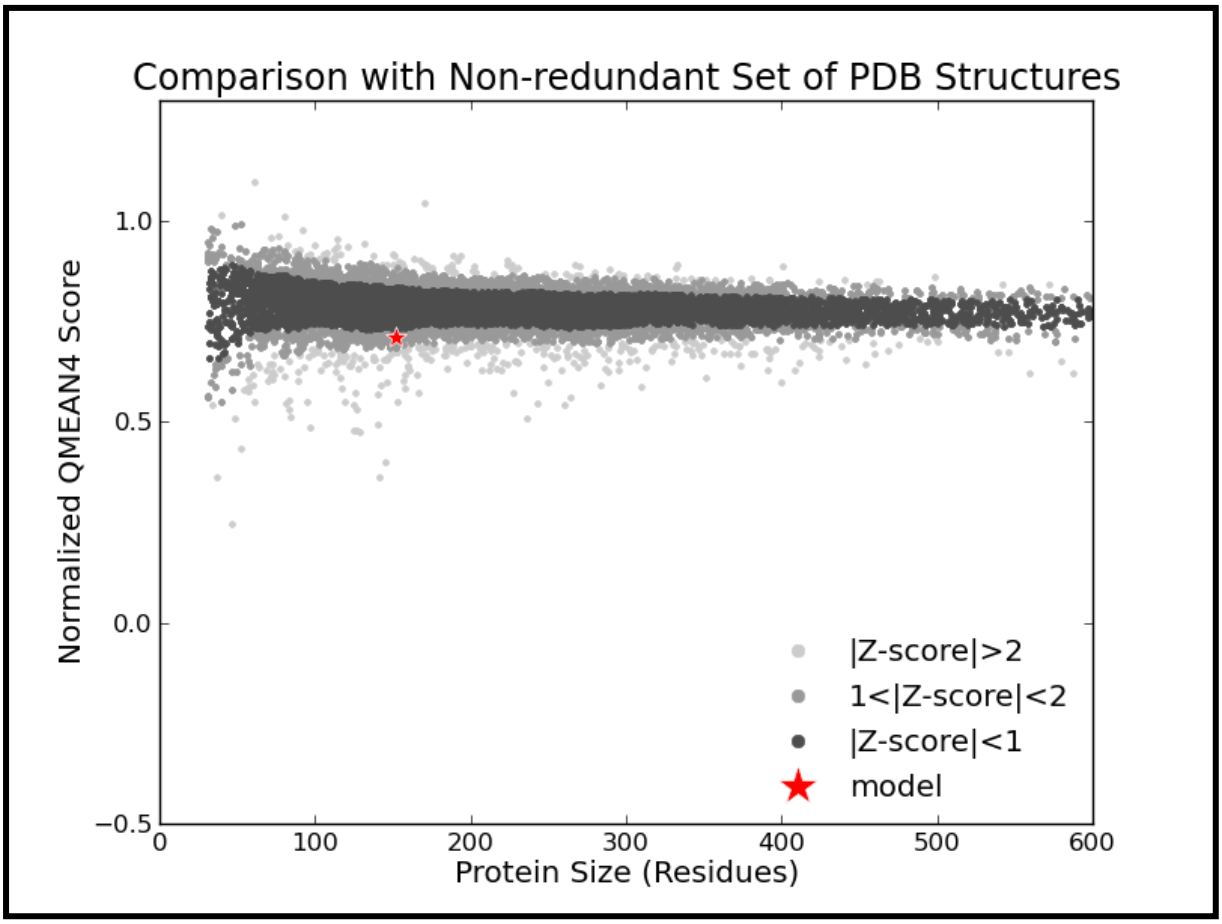


Fig. 3.12: QMEAN score plot of 3D structure of G4LZI3. It depicts the comparison with non-Redundant Sets of PDB structure of homologous proteins (See appendix).

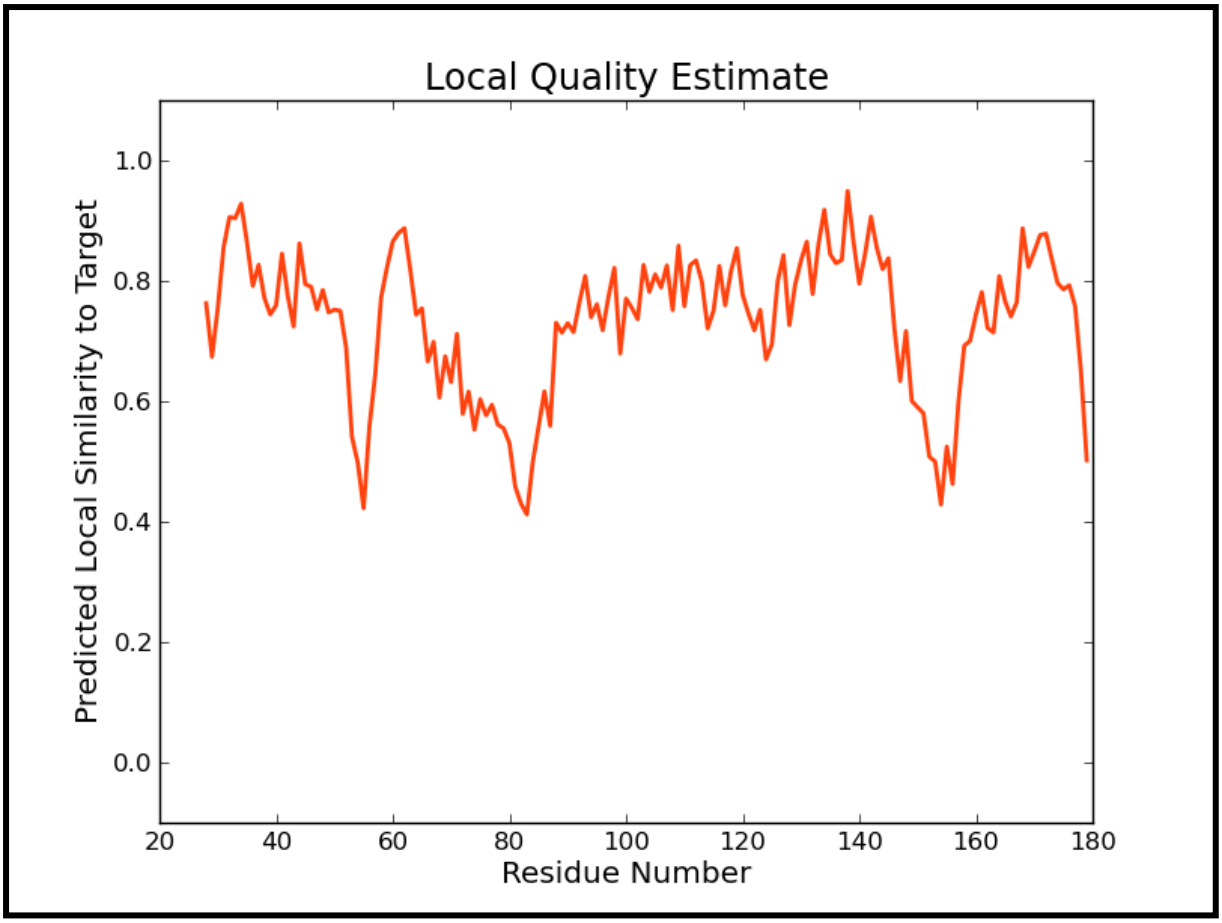


Fig. 3.13: Showing the Local Quality Estimate of G4LZI3 modelled structure by Swiss-model.

3.4.9 Validation of accuracy of modelled structures

The tertiary structure of G4LZI3 built with the Swiss-model tool was validated with Ramachandra plot derived from the RAMPAGE server (Lovell *et al.*, 2002) The Ramachandran plot offers a stress-free way to view the dispersion of torsion angles phi (φ) and psi (ψ) of the amino acid residues in a protein structure. Additionally, it provides an outline of allowed and disallowed regions of torsion angle values, thereby serving as a significant factor in the assessment of the quality of the predicted 3D structure of protein. The Ramachandra plot for the Swiss-model 3D structure in Fig. 3.14a and 3.14b indicates that ~88.0 % of the residues are in the favoured region, ~8.0 % are in the allowed region while ~4.0 % are in the outliner region. The present prediction revealed that a good percentage of the modelled

G4LZI3 amino acid residues are appropriately located, hence the Ramachandra plot ascertained that stereochemical parameters were not violated throughout the model building process therefore, the G4LZI3 homology model can be regarded as a good quality model.

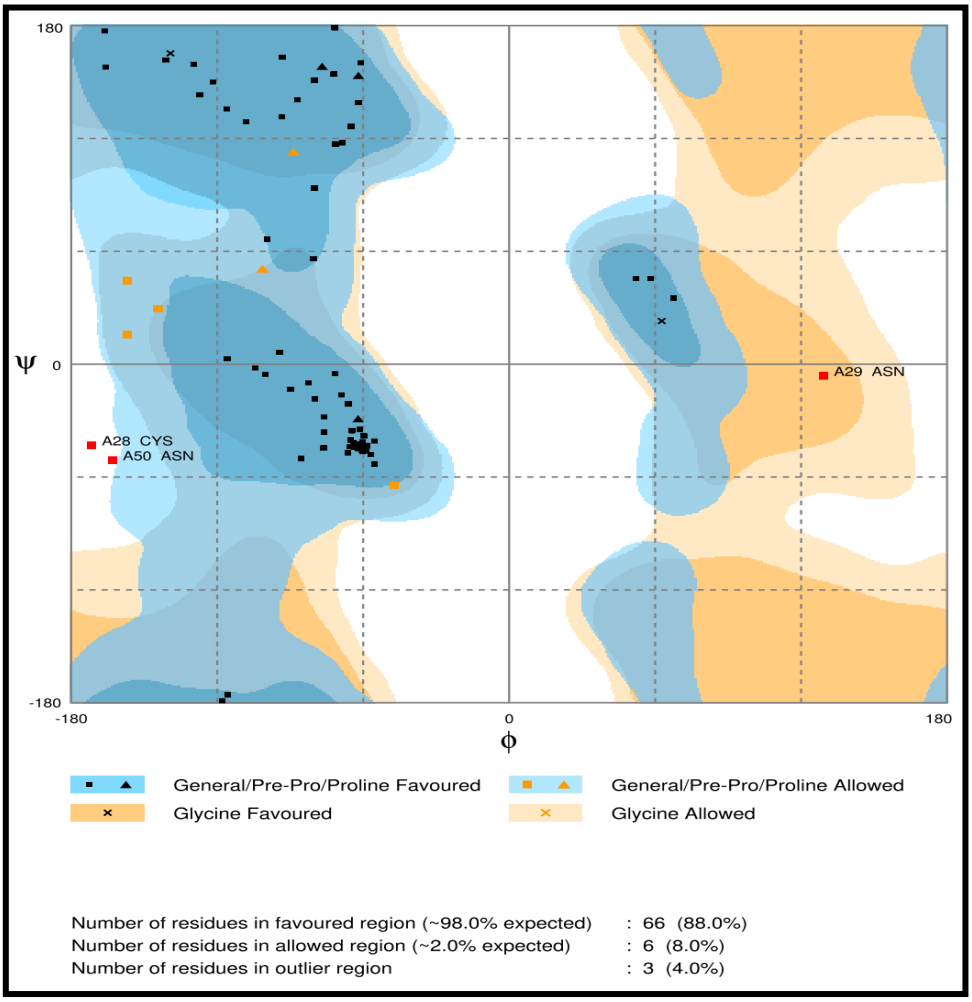


Fig 3. 14a: Ramachandra Plot of modelled Universal Stress Protein G4LZI3 using Swiss model. The figure was generated and visualized using RAMPAGE.

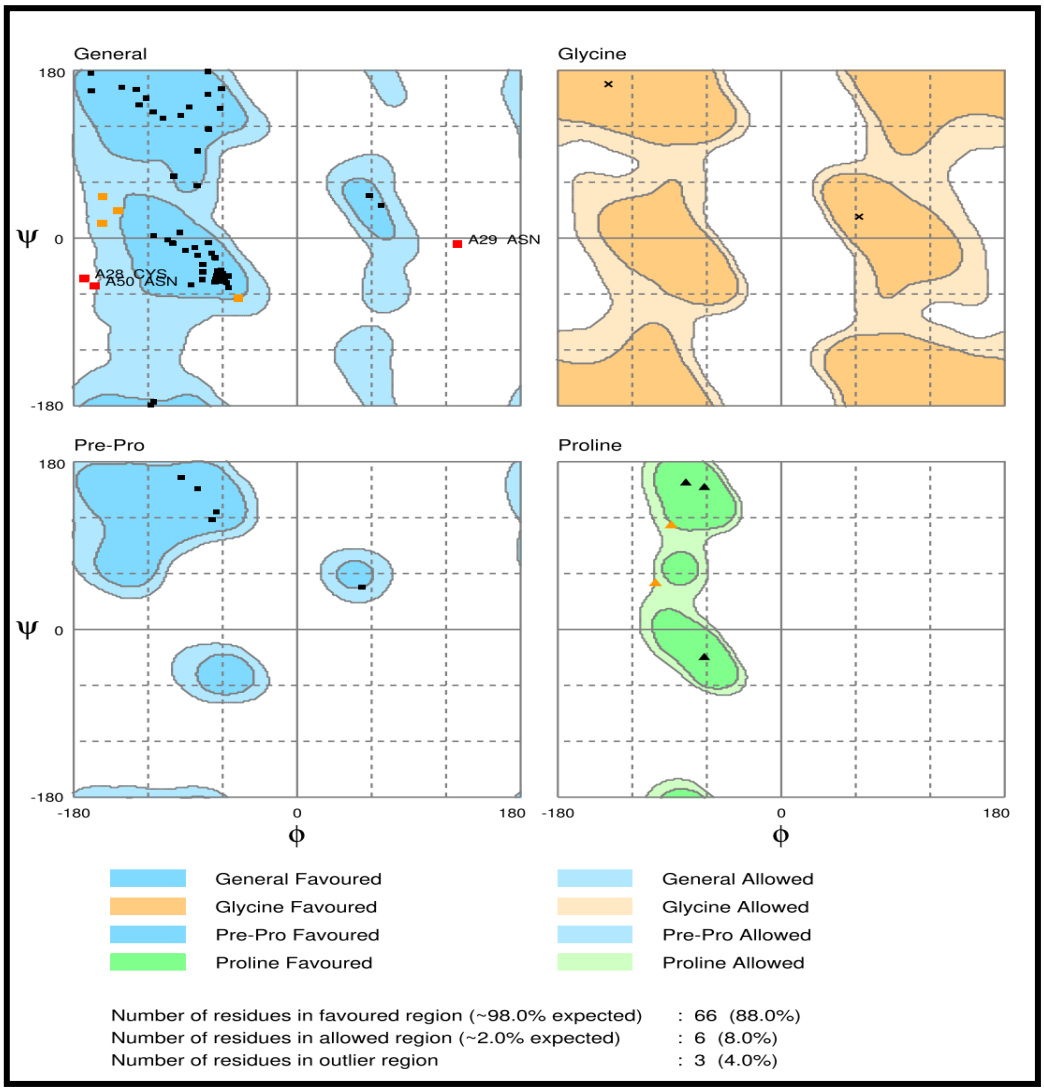


Fig. 3.14b: Favoured and allowed regions in Ramachandran plot of G4LZI3 3D model generated by Swiss-model.

3.4.10 Structure refinement of 3D homology model

Presented in Fig. 3.15 is the structure refined model of *S. mansoni* G4LZI3 3D model done on the GalaxyRefine server. The refined model is expected to be more thermodynamically stable and with better built back bone structures.

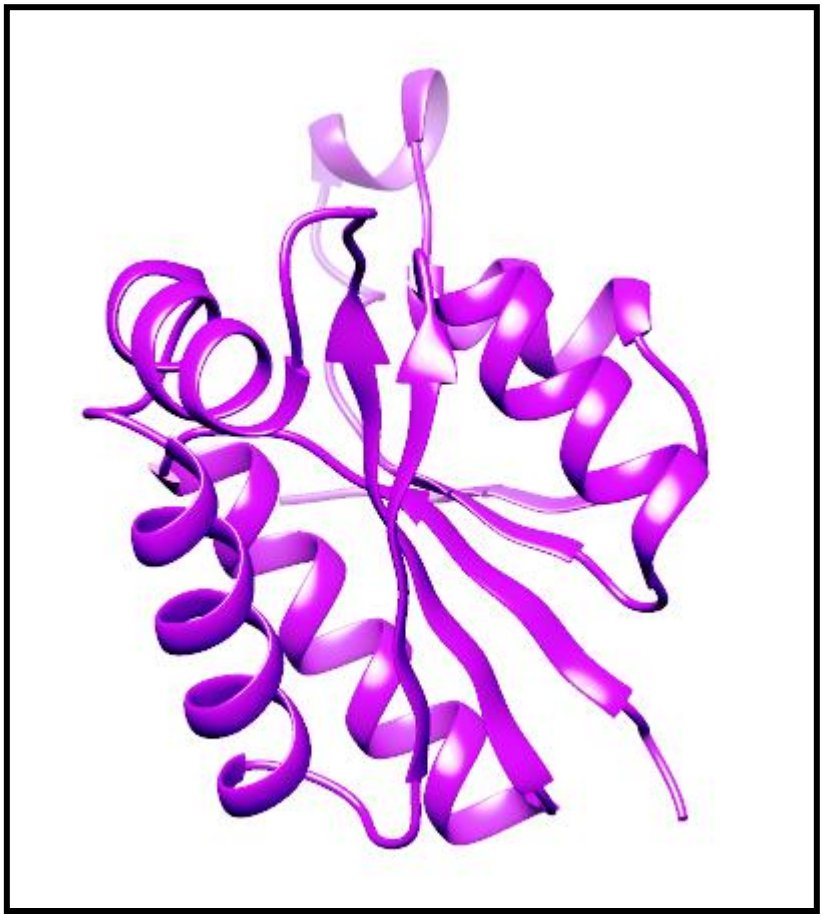


Fig. 3.15: Thermodynamically stable 3D structure of G4LZI3. Structure refinement done with GalaxyRefine.

3.4.11 Solvent accessibility of G4LZI3

Computation of the solvent accessiblility of the protein was done on the ASAview server. Fig. 3.16 shows the colour coded spiral view graphic indicating the various amino acds residues present in G4LZI3 in order of their solvent accessibility. The most accessible residues are located on the outer ring of the spiral graphic plot. The radius of the solid circles standing for each amino acid residue corresponds to the relative solvent accessibility. The bar chart in Fig. 3.17 are generated for G4LZI3 by retaining the order in which the amino acid residues occur in the original input file. It shows the relative solvent accessibility plot in the original order of amino acid residues.

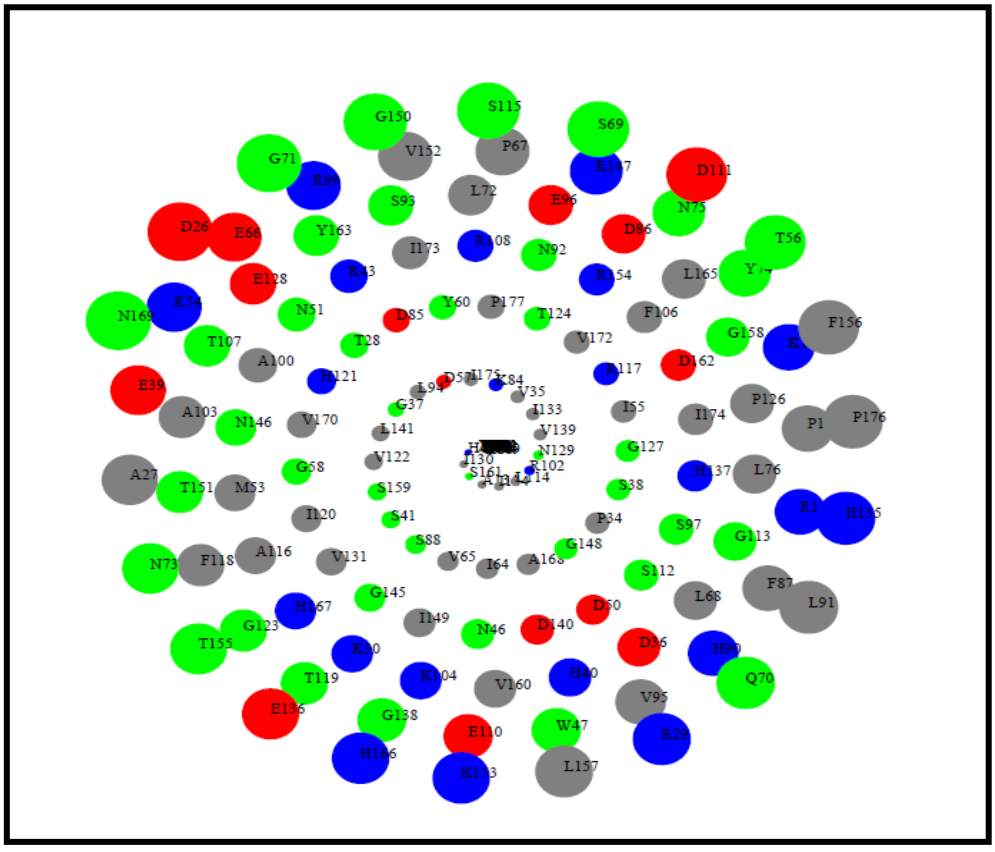


Fig. 3.16: Spiral view of amino acid residues of G4LZI3 in order of solvent accessibility. Most accessible residues feature on the outermost ring.

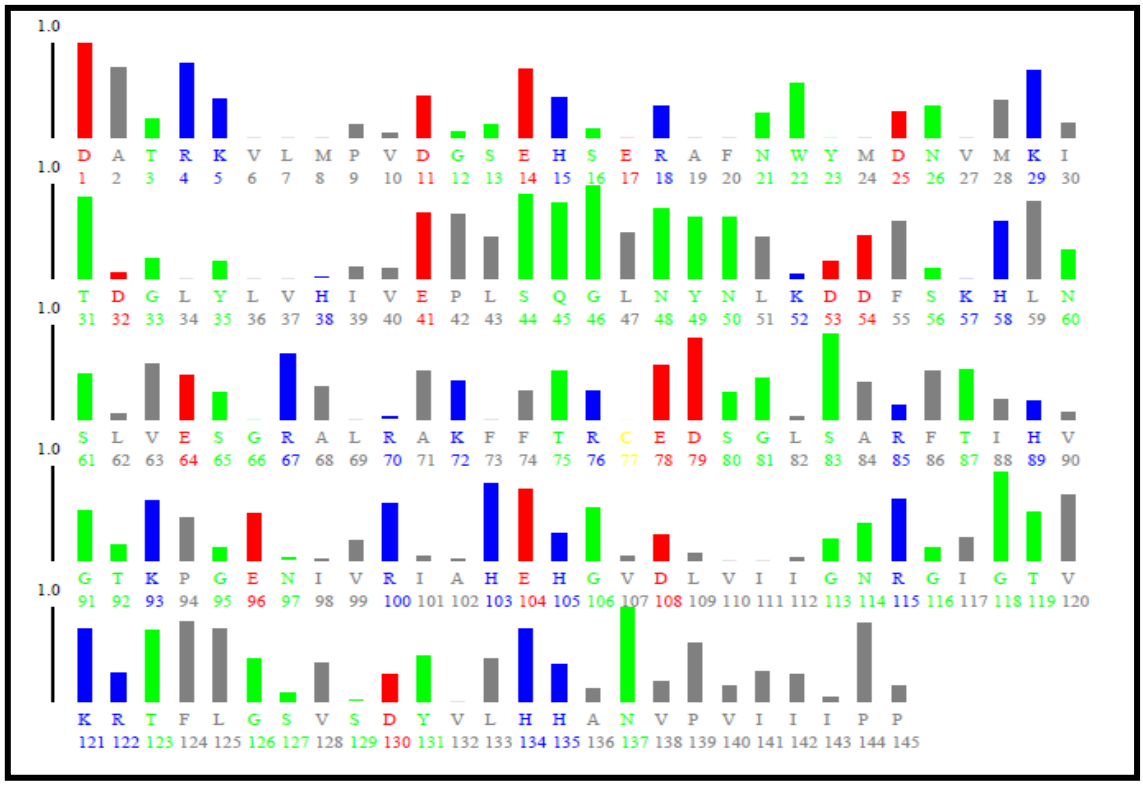


Fig. 3.17: Bar plot of amino acid residues of G4LZI3 showing relative solvent accessibility. The order of residues are retained as they occur in the original input file. Colour coding of residues: Gray-Hyrophobic; red-negatively charged; blue-positively charged; green-polar neutral residues; Yellow-Cysteine with unique properties.

3.4.12 Prediction of binding site of G4LZI3

Fig. 3.18 lays out predicted features corresponding to protein binding regions within G4LZI3. It shows that the protein-protein binding regions on the protein sequence include positions 18, 54, 74, 78, 124 and 147. It is also predicted that the polypeptide-binding site is found on position 146 (Ofran and Rost, 2007).

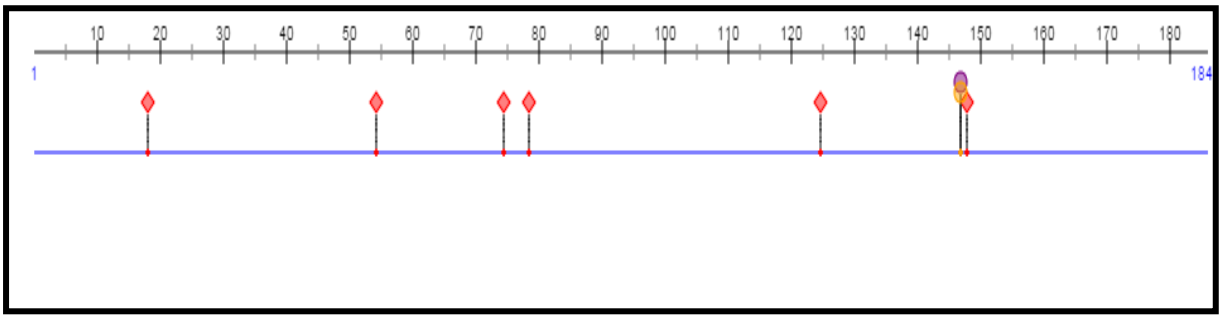


Fig. 3.158: Predicted binding sites on G4LZI3. There are seven binding sites on the protein.

3.4.13 Determination of subcellular localization of G4LZI3

Fig. 3.19 shows a schematic cell with the predicted subcellular localization compartment tinted in green. The predicted localization of G4LZI3 in *Schistosoma mansoni* is the cytoplasm. The prediction confidence is 40, which is considered reliable (Goldberg and Rost 2012).

Fig. 3.19: Schematic representation of subcellular localization of G4LZI3 protein derived fom the PredictProtein online server. The protein is localized in the cytoplasm.

3.4.14 Antigenic determinants on G4LZI3

The antigenic peptide prediction revealed the presence of six predicted peptide epitopes in the G4LZI3 sequence. These epitopes are characterized by specific start positions and end points which are shown in Table 3.4. Results also indicate that the average antigenic propensity of the protein is 1.0252. Fig. 3.20 depicts an antigenic plot for *Schistosoma mansoni* G4LZI3.

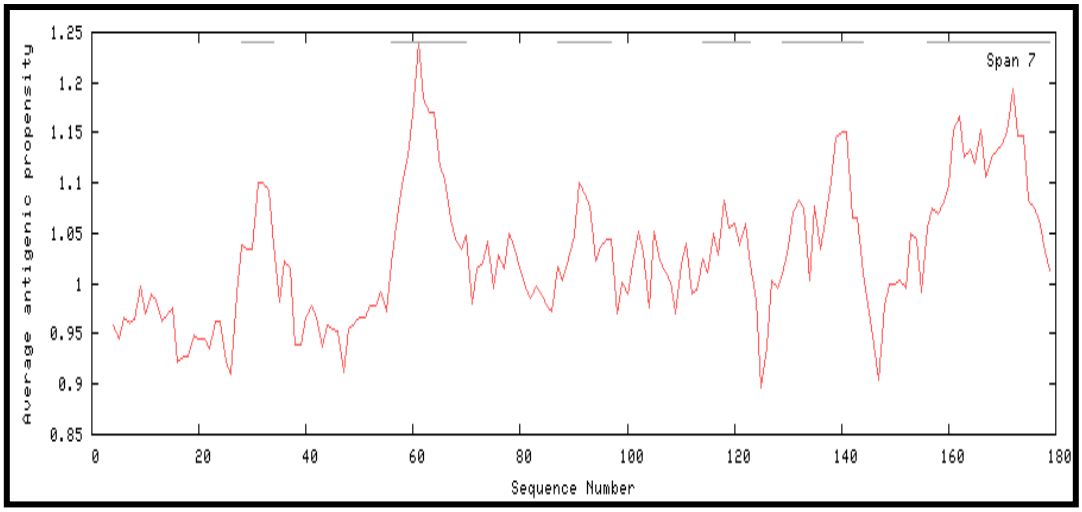


Fig. 3.20: Antigenic plot of G4LZI3. The start and end points within the G4LZI3 sequence are shown.

Table 3.4: Predicted antigenic peptides on G4LZI3. The epitope sequences' start and end positions on the protein sequence are shown.

S/n	Start position	Sequence	End position
1	28	TRKVLMP	34
2	56	TDGLYLVHIVEPLSQ	70
3	87	FSKHLNSLVES	97
4	114	LSARFTIHVG	123
5	129	NIVRIAHEHGVDLVII	144
6	156	FLGSVSDYVLHHANVPVIIIPPPK	179

3.4.15 Gene ontology predictions

GO is a functional grouping scheme which describes three aspects in respect of a gene product; these aspects usually include biological process, molecular function, and cellular component (Ashburner *et al.,* 2000). The result of metastudent consists of a list of Gene Ontology (GO) terms which designate the functions of a protein. These functions are either part of the Biological Process Ontology (BPO) or the Molecular Function Ontology (MFO). Ontology is a rooted graph where individual nodes represent a GO term which links a functional association. Each of the predicted terms is linked with a reliability value of 0-100, defining the chance that the target protein really has the predicted function. It must be noted that the predictions generally have quite a low reliability, however, the output of the prediction is still very useful, to investigate most of the functions of a protein experimentally, one has to assess the predicted functions, not the entire GO.

Table 3.5 shows some selected predicted molecular function predictions for the G4LZI3 protein, which revealed that the protein is likely to possess binding activity, structural molecular activity, transporter activity as well as catalytic activity. Fig 3.21 indicates the biological process prediction of G4LZI3 to be response to stimuli and response to stress. Fig 3.22 is a graphical representation of the cellular component ontology, predicting the protein to be found in either the cytoplasm or the cell membrane (Hamp *et al.,* 2013).

Table 3.5: Some predicted molecular functions of G4LZI3 based on Gene ontology analysis

S/N	GO ID	GO Term	Reliability
1	GO:0005488	Binding	35
2	GO:0035639	Purine ribonucleoside triphosphate binding	26
3	GO:1901363	Heterocyclic compound binding	26
4	GO:0001883	Purine nucleoside binding	26
5	GO:0001882	Nucleoside binding	26
6	GO:0043167	Ion binding	26
7	GO:1901265	Nucleoside phosphate binding	26
8	GO:0017076	Purine nucleotide binding	26
9	GO:0043168	Anion binding	26
10	GO:0097159	Organic cyclic compound binding	26
11	GO:0036094	Small molecule binding	26
12	GO:0005524	ATP binding	26
13	GO:0005515	Protein binding	24
14	GO:0000166	Nucleotide binding	23
15	GO:0019205	Nucleobase-containing compound kinase activity	9
16	GO:0003824	Catalytic activity	9
17	GO:0016772	Transferase activity, transferring phosphorus-containing groups	9
18	GO:0015077	Monovalent inorganic cation transmembrane transporter activity	4
19	GO:0015297	Antiporter activity	4
20	GO:0008324	Cation transmembrane transporter activity	4

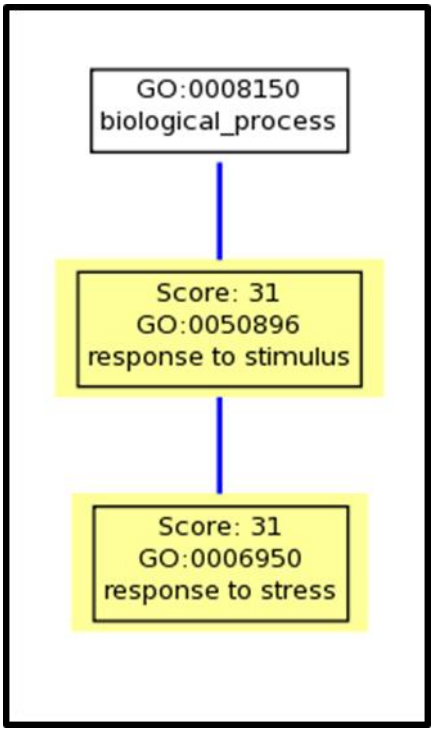


Fig. 3.21: Biological process ontology of G4LZI3. Gene ontology prediction shows that the protein might be implicated in response to stimuli and stress.

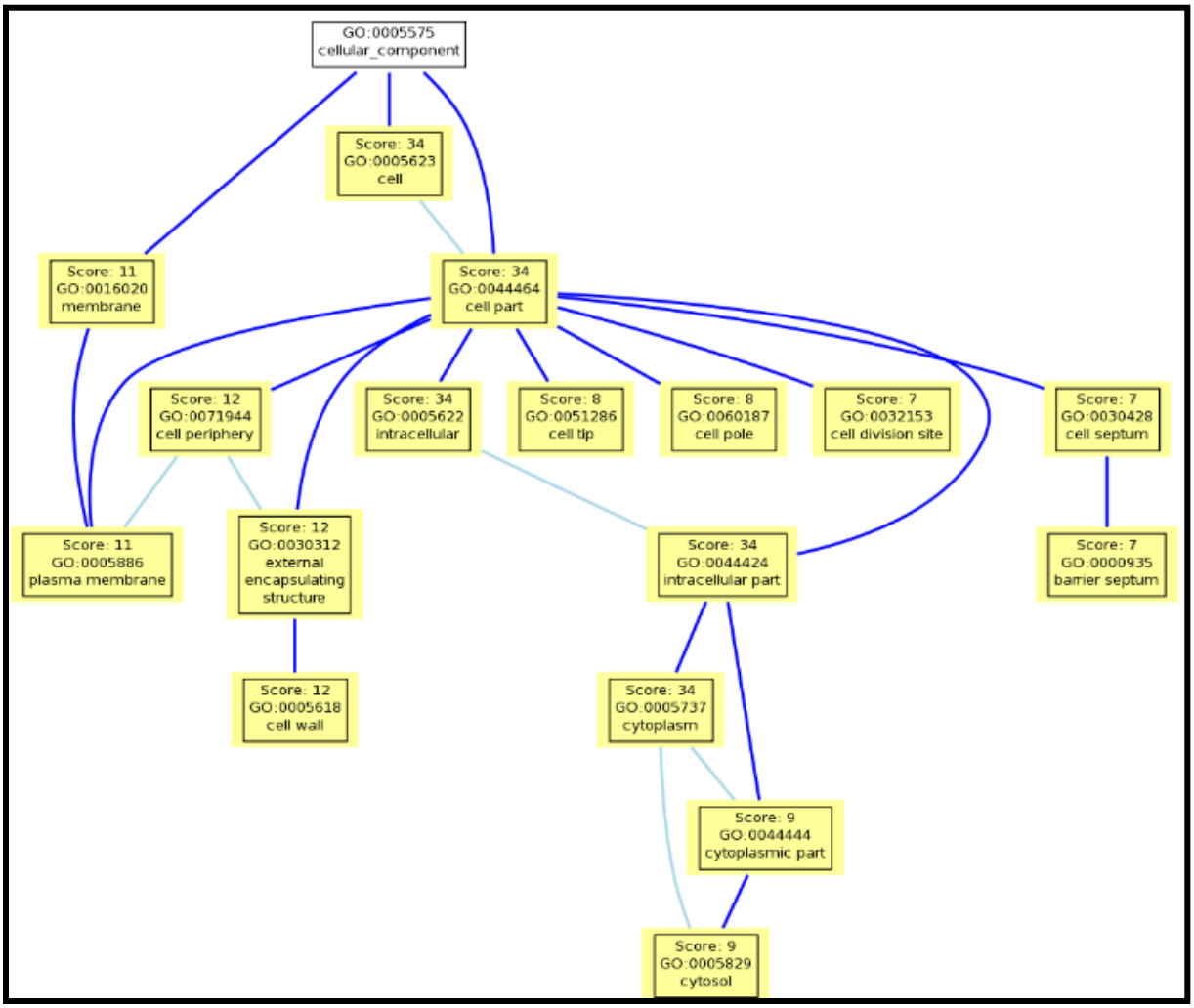


Fig. 3.22: Cellular component ontology of G4LZI3. The protein is predicted be found either in the cytoplasm or the cell membrane.

3.5 Discussion and conclusion

An aspiring aim of proteomics is to determine the structure, molecular function and interaction of all proteins in living cells and organisms. It is expected that the structural elucidation of proteins will make available a better perception of cellular processes and networks at the protein level, and eventually providing a better knowledge of disease mechanisms hence, suggesting novel approaches for disease intervention (Bock and Gough 2001; Bandyopadhyay, 2002). The advent of bioinformatics, genomics, proteomics as well as structural information about novel proteins has made available hundreds of new targets and ligands. *In-silico* techniques offer great potential in target identification, target validation, knowledge of the protein, evolution and phylogeny as well as protein modelling (Singh *et al.,* 2006; Vu *et al.,* 2015). The computational approach that was used in this study for predicting the properties, primary, secondary and tertiary structure of *S. mansoni* universal stress protein, indicated the effectiveness of several bioinformatics tools. These tools were and are necessary in the prediction of structural and functional features, thus aiding experimental analysis of the protein. The result of this study can be explored further towards utilization of this novel protein for possible therapeutic purposes.

Worthy of note is the fact that the protein is made up of five important conserved domains namely; USP_Like, Usp, UspA, NaH_Antiporter_C and STK_N domains as seen in Fig 3.7. Previous studies have established the occurrence of these conserved domains in various organisms. In one previous study, a USP_Like gene was discovered and characterized in a medicinal wild plant species *Calotropis procera* through the *de*

novo assembled genome contigs of the high-throughput sequencing dataset (Shokry *et al.,* 2014).

Interestingly, the UspA domain is a well-known protein structure established to be extensively present in prokaryotic organisms, including bacteria and archaea. Genetic evidence has successively revealed that UspA facilitates survival of cells starved from different variations of nutrients and exposure to toxicity and/or UV light damage. This protein is a Ser and Thr phosphoprotein and it is phosphorylated by the Tyr phosphoprotein TypA. However, the exact biochemical function of UspA is unexplained (Kerk *et al.,* 2003; Siegele, 2005). The Usp domain is a conserved domain made up of 140-160 amino acids; it is common in bacteria, archaea, plants and fungi, however, it is absent in humans, thus making the protein a potential drug target (Hingley-Wilson *et al.,* 2010).

Na^+/H^+ antiporters (NaH_Antiporters) are integral membrane proteins which are present in nearly all cells in all living organisms of the animal kingdom. They are indispensable for the regulation of intracellular Na^+ concentration, regulation of pH, as well as cell volume. Furthermore, these secondary active transporters exchange sodium ions against protons through an interchanging access mechanism yet to be fully understood. Na^+/H^+ antiporters exhibit discrete species-specific transport features and regulatory properties which show a relationship with individual physiological functions (Lentes *et al.,* 2014). The domain STK family is a member of single transmembrane proteins that has an intracellular conserved kinase catalytic domain and extracellular receptor domain on the N-terminal and these have been discovered in several organisms including pollen grains of maize (Zhou *et al.,* 2014).

Generally, the domains predicted to be conserved in G4LZI3 are very important in stress response, transportation within the cells amongst other vital functions. Multiple sequence alignment is an imperative aspect of protein research. Important biological inferences can be obtained from the conservation or variation which occurs between aligned positions, particularly with reference to the structure of one of the aligned sequences. A protein is made up of surface loops and core regions. The quickly evolving subsets of a protein are denoted as the surface loops which are frequently obvious in the multiple sequence alignments and are usually referred to as gap-containing or gap-rich regions (Blouin *et al.,* 2004). Hence, their counter-parts are the core regions called the gap-free regions. The loops are the weaker portion existing on the surface of the proteins which connect to more rigid portions. Conversely, the core regions act as support walls for the proteins (Claverie and Notredame, 2006). The observed sequence variability in the protein is related to the structural inconsistency of the surface loops, therefore making them inclined towards rapid evolutions. On the other hand, the core regions are less disposed to rapid evolutions compared with the surface loops (Blouin *et al.,* 2004).

The MSA results (Fig. 3.5a and 3.5b) obtained showed that the sequences generally contained gap-free regions with only very few gap-rich regions. Therefore, it can be concluded that the selected proteins chiefly contained core regions with few surface loops, implying that the protein is highly conserved since the core protein which is in abundance does not experience rapid evolutions and stays unchanged.

Comparisons of gene sequences in a phylogenetic context are capable of providing meaningful insights into the biology of species, while a phylogenetic viewpoint also

offers foundation for comparative genomics. Furthermore, evolutionary history of living species might be inferred through the phylogenetic analysis of molecular and morphological information (Soltis and Soltis, 2003; Yuan and Zhu, 2014). In the present study, the generated phylogenetic tree revealed closely clustered clades as shown in Fig. 3.6 which is an indication of evolutionary relationships between homologous sequences. The analysis of the level of conservation of the protein shows a high degree of conservation. It must be noted that conservation across species shows that a sequence has been preserved by evolution despite the formation of new and dissimilar species during the course of evolution. Notably, a highly conserved sequence is a sequence which has stayed unaltered within the phylogenetic tree. Additionally, conservation analysis can often reveal the importance of the protein in terms of structure as well as function of protein in different species (Jensen *et al.,* 2003).

Physiochemical properties analysed in this study included number of amino acid residues, molecular mass, theoretical isoelectric point (pI), amino acid and total atom composition, extinction coefficients at 280 nm, instability index and grand average hydropathicity (GRAVY). They are very vital parameters because they immensely influence further applications of the protein under study (Kumar and Bhalla, 2011). It can be inferred from amino acid composition and physicochemical analysis results that the protein is structurally-flexible and that there is good interaction between the protein and water. These results principally suggest that G4LZI3 is relatively reactive in water and other aqueous environments.

Based on the secondary structure prediction results (3.9 and 3.10), it can be inferred that the G4LZI3 protein possesses remarkable stability because of the presence of β-

turns that compensated for the absence of disulfide bridge formation. Additionally, the protein has a high percentage of α-helices which comprise of hydrogen bonds. Hydrogen bonds offer utmost directional interactions which reinforce protein folding, protein structure as well as molecular recognition. Hence, the presence of helices and hydrogen bonds is a major factor that plays a vital part in the protein's stability (Goetz *et al.,* 2014; Pace *et al.,* 2014).

Homology modelling is a procedure used to predict the 3-dimensional structure of a protein whose structure has not been previously solved experimentally. It plays a very important role in the process of structure-based drug discovery. It makes use of accessible high resolution protein structures to yield a model for a similar protein without a known structure (Fiser and Sali, 2003; Bishop *et al.,* 2008; Cavasotto and Phatak, 2009).

In order to produce a homology model, a percentage of primary sequence identity between a previously experimentally produced template structure and the query protein of approximately 20 % or more is essential. Modelling accuracy drops lower than 50 % and declines quickly below 30% identity (Berry and Board, 2017). In the present study, the Swiss-model server was used to generate 3D homology models of G4LZI3. Swiss-model is an automated system for modelling the 3D structure of a protein. The server is characterized by a user-friendly web interface that enables non-specialists to generate 3D models from a simple web-browser without any need for the instalment of complex molecular modelling software. It is one of the most commonly used modelling web servers in the world with over 0.9 million entreaties for protein models yearly (Arnorld *et al.,* 2006; Biasini *et al.,* 2014). The 3D structure of Universal Stress Protein G4LZI3

which was veiwed using Pymol as presented in Fig. 4.11 revealed that the secondary structure elements are made up of four α-helices and five parallel β-sheets. However, the Swiss-model could, in future, be further verified through laboratory experimental procedures for structural elucidation of the three dimensional structure of the protein such as NMR and X-ray crystallography. Further effort was made to validate the accuracy of the built model with the use of Ramachandra plots generated with the Rampage tool (Fig. 3.14a and 3.14b). Analysis of the plot showed that the model is accurate and hence, it was used for further *in-silico* procedures.

Structure refinement of the Swiss-model homology model was done to improve the local structure quality of the protein. It ensures adequate upgrading in backbone structure quality and thermodynamic stability of the three dimensional protein structure (Heo *et al.,* 2013).

Determination of solvent accessibility of the 3D homology model of the protein showed that the polar amino acid residues of the protein are situated on the outermost ring of the spiral plot while the hydrophobic residues are buried in the core of the protein as displayed in Fig. 3.16. Usually, hydrophobic residues have a tendency to be in the core of a protein, in which solvent accessibility is low, while polar residues are normally found on the surface of a protein, where solvent accessibility is high (Moelbert *et al.,* 2004). It has been shown that the surface solvent accessible area of proteins has direct impact on the folding state and stability of proteins. The folded state is stabilized predominantly by the burial and constricted packing of above 80 % of the peptide groups and non–polar side chains (Pace *et al.,* 2004; Carugo, 2013). Hence, G4LZI3 can be concluded to have a good interaction with water and is stable in aqueous environments. This

characteristic of the universal stress protein might be responsible for its ability to be efficiently expressed using recombinant expression technology without the production of undesirable inclusion bodies.

Subcellular localization is a basic functional characteristic of a protein. Fundamentally, cellular functions are usually localized in particular compartments. A prediction of the subcellular localization of new proteins can be used to acquire handy information as regards to their functions. Furthermore, investigating the subcellular localization of proteins is very helpful in unravelling disease mechanisms and for the development of novel drugs and vaccines (Su *et al.,* 2007). In the present study, the G4LZI3 protein is predicted to be localized in the cytoplasmic compartment of the cell. This prediction is based on a simple but reliable method of predicting the localization site of an unknown protein by inheriting subcellular localization from homologous proteins (Lin *et al.,* 2009).

Antigenic determinants, also called epitopes, are structural constituents of an antigen to which antibodies or T-cell receptors bind, it is the portion of an antigen which is recognized by the immune system. *Schistosoma* antigens have been recognized by various studies as potential candidates in development of vaccines against schistosomiasis. These include Sm97-paramyosin, Sm62-IrV5, Sm28-GST, Sm37-GAPDH, Sm14-FABP, paramyosin, Sm14 and Sm23, (Argiro *et al.,* 2000; Bergquist *et al.,* 2002; Al-Sherbiny *et al.,* 2003; Pearce, 2003). Determining the antigenic epitopes in G4LZI3 is thus a worthwhile effort towards further study for actualization of the long-term goal of using the protein as a vaccine candidate.

The gene ontology prediction is in support of previous studies showing that the universal stress protein superfamily comprises a conserved group of proteins implicated

in resistance to stress. Usps support cell survival during persistent exposure to stress and they may also activate a general mechanism required for stress endurance. Additionally, reports also show that the universal stress protein might be involved in transduction signalling (as a molecular switch) as predicted by the gene ontology analysis (Uduwat *et al.,* 2014; De Souza *et al.,* 2016).

Conclusively, the *in-silico* analysis on G4LZI3 presented in this chapter showed that the protein is made up of very important conserved domains which have been implicated in stress response amongst several organisms. The cytoplasm-located protein is also highly conserved and has very intense evolutionary relationships with its homologues. The protein is also relatively stable and has good solvent interaction. The 3D homology model revealed that the protein is well folded and will be well amenable for further structural and functional elucidation. Overall, the data derived from the study shows that the protein has great potential as a drug and vaccine candidate; hence, it is imperative that wet-laboratory procedures be done on the basis of the predictions of these computational studies.

References

Al-Sherbiny M., Osman A., Barakat R., El Morshedy H., Bergquist R. and Olds R. (2003). *In vitro* cellular and humoral responses to *Schistosoma mansoni* vaccine candidate antigens. *Acta Tropica*, **88**(2): 117-130.

Álvarez C.A., Gomez F.A., Mercado L., Ramírez R. and Marshall S.H. (2016). *Piscirickettsia salmonis* imbalances the innate immune response to succeed in a productive infection in a salmonid cell line model. *PloS One*, **11**(10): e0163943.

Argiro L., Kohlstädt S., Henri S., Dessein H., Matabiau V., Paris P., Bourgois A. and Dessein A.J. (2000). Identification of a candidate vaccine peptide on the 37 kDa *Schistosoma mansoni* GAPDH. *Vaccine*, **18**(19): 2039-2048.

Arnold K., Bordoli L., Kopp J. and Schwede T. (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics*, **22**(2): 195-201.

Artimo P., Jonnalagedda M., Arnold K., Baratin D., Csardi G., de Castro E., Duvaud S., Flegel V., Fortier A., Gasteiger E., Grosdidier A., Hernandez C., Ioannidis V., Kuznetsov D., Liechti R., Moretti S., Mostaguir K., Redaschi N., Rossier, G., Xenarios I. and Stockinger H. (2012). ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Research,* **40**:597-603.

Ashburner M., Ball C.A., Blake J.A., Botstein D., Butler H., Cherry J.M., Davis A.P., Dolinski K., Dwight S.S., Eppig J.T. and Harris M.A. (2000). Gene Ontology: tool for the unification of biology. *Nature Genetics*, **25**(1): 25-29.

Bairoch A.M., Apweiler R., Wu C.H., Barker W.C., Boeckmann B., Ferro Rojas S., Gasteiger E., Huang H., Lopez R., Magrane M. and Martin M.J. (2005). The universal protein resource (UniProt). *Nucleic Acids Research*, **33***:*154-159.

Banasik K., Justesen J.M., Hornbak M., Krarup N.T., Gjesing A.P., Sandholt C.H., Jensen T.S., Grarup N., Andersson Å., Jørgensen T. and Witte D.R. (2011). Bioinformatics-driven identification and examination of candidate genes for non-alcoholic fatty liver disease. *PloS one*, **6**(1): 16542.

Bandyopadhyay R. (2002). Predicting protein-protein interactions from primary structure (Doctoral dissertation, Rice University).

Bergquist R., Al-Sherbiny M., Barakat R. and Olds R. (2002). Blueprint for schistosomiasis vaccine development. *Acta Tropica*, **82**(2): 183-192.

Berry C. and Board J. (2017). The use of structural modelling to infer structure and function in biocontrol agents. *Journal of Invertebrate Pathology*, **142**: 23-26.

Biasini M., Bienert S., Waterhouse A., Arnold K., Studer G., Schmidt T., Kiefer F., Cassarino T.G., Bertoni M., Bordoli L. and Schwede T. (2014). SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Research*, **340**.

Bishop A., De Beer T.A. and Joubert F. (2008). Protein homology modelling and its use in South Africa. *South African Journal of Science*, **104**(2): 2-6.

Blouin C., Butt D. and Roger A.J. (2004) Rapid evolution in conformational space: A study of loop regions in a ubiquitous GTP binding domain. *Protein Science: A Publication of the Protein Society,* **13**(3): 608-616.

Bock J.R. and Gough D.A. (2001). Predicting protein–protein interactions from primary structure. *Bioinformatics*, **17**(5): 455-460.

Boratyn G.M., Camacho C., Cooper P. S., Coulouris G., Fong A., Ma N., Madden T.L., Matten W.T., McGinnis S.D., Merezhuk Y., Raytselis Y., Sayers E.W., Tao T., Ye, J. and Zaretskaya I. (2013) BLAST: a more efficient report with usability improvements. *Nucleic Acids Research,* **41**:29-33.

Buchan D.W.A., Minneci F., Nugent T.C.O., Bryson K. and Jones D.T. (2013) Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Research,* **41**: 349-357.

Carugo O. (2013). Solvent accessible surface of globular proteins: how exposed and buried amino acid residues are divided. *OA Biology,* **1**(1):1.

Cavasotto C.N. and Phatak S.S. (2009). Homology modeling in drug discovery: current trends and applications. *Drug Discovery Today*, **14**(13): 676-683.

Claverie J. M. and Notredame C. (2006) Bioinformatics for Dummies. 2nd edition: Wiley Publishing, Inc., p. 456.

De Souza C.S., Torres A.G., Caravelli A., Silva A., Polatto J.M. and Piazza R.M. (2016). Characterization of the universal stress protein F from atypical enteropathogenic *Escherichia coli* and its prevalence in *Enterobacteriaceae*. *Protein Science*, **25**(12): 2142-2151.

Edwards Y.J., and Cottage A. (2003). Bioinformatics methods to predict protein structure and function. *Molecular biotechnology*, **23**(2): 139-166.

Fiser A., and Šali A. (2003). Modeller: generation and refinement of homology-based protein structure models. *Methods in Enzymology*, **374**: 461-491.

Gasteiger E., Hoogland C., Gattiker A., Duvaud S.E., Wilkins M.R., Appel R.D. and Bairoch A. (2005). Protein identification and analysis tools on the ExPASy server. Humana Press, 571-607.

Geer L.Y., Marchler-Bauer A., Geer R.C., Han L., He J., He S., Liu C., Shi W. and Bryant S.H. (2010) The NCBI BioSystems database. *Nucleic Acids Research,* **38**: 492-496.

Goetz G.H., Farrell W., Shalaeva M., Sciabola S., Anderson D., Yan, J., Philippe L. and Shapiro M.J. (2014). High throughput method for the indirect detection of intramolecular hydrogen bonding. *Journal of Medicinal Chemistry*, **57**(7); 2920-2929.

Goldberg T., Hamp T. and Rost B. (2012). LocTree2 predicts localization for all domains of life. *Bioinformatics*, **28**(18): 458-465.

Hamp T., Kassner R., Seemayer S., Vicedo E., Schaefer C., Achten D., Auer F., Boehm A., Braun T., Hecht M. and Heron M. (2013). Homology-based inference sets the bar high for protein function prediction. *BMC bioinformatics*, **14**(3): 7.

Heo L., Park H. and Seok C. (2013). GalaxyRefine: protein structure refinement driven by side-chain repacking. *Nucleic Acids Research*, **41**(1): 384-388.

Hingley-Wilson S.M., Lougheed K.E.A., Ferguson K., Leiva S. and Williams H.D. (2010). Individual Mycobacterium tuberculosis universal stress protein homologues are dispensable in vitro. *Tuberculosis*, **90**(4): 236-244.

Jensen L.J., Ussery D.W. and Brunak S. (2003). Functionality of system components: conservation of protein function in protein feature space. *Genome Research*, **13**(11): 2444-2449.

Johnson M., Zaretskaya I., Raytselis Y., Merezhuk Y., McGinnis S. and Madden T. L. (2008). NCBI BLAST: a better web interface, *Nucleic Acids Research*, 36: W5-W9.

Jones D.T. (1999) Protein secondary structure prediction based on position specific scoring matrices, *Journal of Molecular Biology,* **292**(2): 195-202.

Kanehisa M., and Bork P. (2003). Bioinformatics in the post-sequence era. *Nature Genetics*, **33**: 305-310.

Kerk D., Bulgrien J., Smith D.W. and Gribskov M. (2003). Arabidopsis proteins containing similarity to the universal stress protein domain of bacteria. *Plant Physiology*, **131**(3): 1209-1219.

Kim D.J., Bitto E., Bingman C.A., Kim H.J., Han B.W. and Phillips G.N. (2015). Crystal structure of the protein At3g01520, a eukaryotic universal stress protein-like protein from arabidopsis thaliana in complex with AMP. *Proteins: Structure, Function, and Bioinformatics*, **83**(7): 1368-1373.

Kolaskar A.S. and Tongaonkar P.C. (1990). A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS letters*, **276**(1-2): 172-174.

Kumar N. and Bhalla T.C. (2011). *In-silico* analysis of amino acid sequences in relation to specificity and physiochemical properties of some aliphatic amidases and kynurenine formamidases. *Journal of Bioinformatics and Sequence Analysis*, **3**(6): 116-123.

Lage K., Karlberg E.O., Storling Z.M, Olason P.I. and Pedersen A.G. (2007). A human phenome-interactome network of protein complexes implicated in genetic disorders. *Nature Biotechnology*, **25**: 309–316.

Laskowski R., Rullmann J.A., MacArthur M., Kaptein R. and Thornton J. (1996). AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *Journal of Biomolecular NMR,* **8**(4): 477-486.

Laskowski R.A., MacArthur M.W., Moss D.S. and Thornton J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography,* **26**(2): 283-291.

Lentes C.J., Mir S.H., Boehm M., Ganea C. and Fendler K. (2014). Molecular Characterization of the Na^+/H^+ Antiporter NhaA from Salmonella Typhimurium. *PLoS One* **9**(7): e101575.

Lin H.N., Chen C.T., Sung T.Y., Ho S.Y. and Hsu W.L. (2009). Protein subcellular localization prediction of eukaryotes using a knowledge-based approach. *BMC Bioinformatics*, **10**(15): 8.

Lovell S.C., Davis I.W., Arendall W.B., de Bakker P.I.W., Word J.M., Prisant M.G., Richardson J.S. and Richardson D.C. (2003) Structure validation by Calpha geometry: phi,psi and Cbeta deviation. *Proteins: Structure, Function and Genetics*. **50**: 437-450.

Moelbert S., Emberly E. and Tang C. (2004). Correlation between sequence hydrophobicity and surface-exposure pattern of database proteins. *Protein Science*, **13**(3): 752-762.

Murvai J., Vlahoviček K., Szepesvári C. and Pongor S. (2001). Prediction of protein functional domains from sequences using artificial neural networks. *Genome Research*, **11**(8):1410-1417.

Notredame C. (2002). Recent progress in multiple sequence alignment: a survey. *Pharmacogenomics*, **3**(1): 131-144.

Ofran Y. and Rost B. (2007). ISIS: interaction sites identified from sequence. *Bioinformatics*, **23**(2): 13-16.

Pace C.N., Fu H., Fryar K., Landua J., Trevino S.R., Schell D., Thurlkill R.L., Imur S., Scholtz, J.M., Gajiwala K. and Sevcik J. (2014). Contribution of hydrogen bonds to protein stability. *Protein Science*, **23**(5): 652-661.

Pace C.N., Trevino S., Prabhakaran E. and Scholtz J.M. (2004). Protein structure, stability and solubility in water and other solvents. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **359**(1448): 1225-1235.

Pearce E.J. (2003). Progress towards a vaccine for schistosomiasis. *Acta Tropica*, **86**(2): 309-313. Pettersen E.F., Goddard T.D., Huang C.C., Couch G.S., Greenblatt D.M., Meng E.C. and Ferrin T.E. (2004). UCSF Chimera: a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, **25**(13):1605-1612.

Porollo A. A., Adamczak R. and Meller J. (2004). POLYVIEW: a flexible visualization tool for structural and functional annotations of proteins. *Bioinformatics*, **20**(15): 2460-2462.

Ramachandran G.N., Ramakrishnan C. and Sasisekharan V. (1963) Stereochemistry of polypeptide chain configurations, Journal of Molecular Biology, **7**(1): 95-99.

Schrödinger L: The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC. 2010 available at http://www.pymol.org/citing.

Shander A., Gromiha M., Fawareh H. and Sarai A. (2004). ASA view: solvent accessibility graphics for proteins. *Bioinformatics*, **51***:* 51.

Shokry A.M., Al-Karim S., Ramadan A., Gadallah N., Al Attas S.G., Sabir J.S., Hassan S.M., Madkour M.A., Bressan R., Mahfouz M. and Bahieldin A. (2014). Detection of a Usp-like gene in Calotropis procera plant from the de novo assembled genome contigs of the high-throughput sequencing dataset. *Comptes Rendus Biologies*, **337**(2): 86-94.

Siegele D.A. (2005). Universal stress proteins in Escherichia coli. *Journal of Bacteriology*, **187**(18), 6253-6254.

Singh S., Malik B.K. and Sharma D.K. (2006). Molecular drug targets and structure based drug design: A holistic approach. *Bioinformation*, **1**(8): 314-320.

Soltis D.E. and Soltis P.S. (2003). The role of phylogenetics in comparative genetics. *Plant Physiology*, **132**(4): 1790-1800.

Su E.C.Y., Chiu H.S., Lo A., Hwang J.K., Sung T.Y. and Hsu W.L. (2007). Protein subcellular localization prediction based on compartment-specific features and structure conservation. *BMC bioinformatics*, **8**(1): 330.

Udawat P., Mishra A. and Jha B. (2014). Heterologous expression of an uncharacterized universal stress protein gene (SbUSP) from the extreme halophyte, *Salicornia brachiata*, which confers salt and osmotic tolerance to *E. coli. Gene*, **536**(1): 163-170.

Vu L.A, Quyen P.T.C. and Huong N.T (2015). *In silico* Drug Design: Prospective for Drug Lead Discovery. *International Journal of Engineering Science Invention*, **4**(10): 60-70.

Westbrook J., Feng Z., Chen L., Yang H. and Berman H.M. (2003). The protein data bank and structural genomics. *Nucleic Acids Research*, **31**(1): 489-491.

Yang J., Yan R., Roy A., Xu D., Poisson J. and Zhang Y. (2015) The ITASSER Suite: protein structure and function prediction, *Nature Methods*, 12(1).

Yuan J., Zhu Q. and Liu B. (2014). Phylogenetic and biological significance of evolutionary elements from metazoan mitochondrial genomes. *PloS one*, **9**(1): e84330.

Zhou T., Fan M., Irfan M., Wang H., Wang D., Wang L., Zhang C. and Feng L. (2014). Phylogenetic analysis of STK gene family and Usp domain in maize. *Molecular Biology Reports*, **41**(12): 8273-8284.

CHAPTER FOUR

MOLECULAR CLONING, RECOMBINANT PROTEIN EXPRESSION AND AFFINITY PURIFICATION OF *S. MANSONI* UNIVERSAL STRESS G4LZI3 PROTEIN

4.1 Introduction

Sequel to the *in-silico* study in the preceding chapter (chapter three) which outlined predictions on the primary, secondary and tertiary structure of G4LZI3, this chapter sought to experimentally clone, express and purify a stable, soluble and natively-folded G4LZI3 from *S. mansoni* in appropriate concentration for preliminary structural characterization.

4.2 Methodology

4.2.1 Molecular cloning and restriction analysis of construct

The *S. mansoni* database was searched using the BLAST program on the National Centre for Biotechnology Information (NCBI) to obtain the nucleotide sequence (mRNA) for the novel universal stress protein G4LZI3 as stated in the bioinformatics analysis chapter (Chapter three). The gene sequence obtained was sent to GenScript (USA) for gene synthesis and production of a codon harmonized construct of the G4LZI3 gene. The G4LZI3 gene was cloned into pQE30 (Qiagen) vector using BamHI and HindIII as restriction enzymes.

4.2.2 Preparation of synthesized insert for restriction digest analysis

In order to prepare the synthesized insert for restriction digest analysis, the vial containing the lyophilized pQE30-G4LZI3 plasmid DNA obtained from GenScript was centrifuged at 6 000 x *g* for 1 minute at 4 $^{\circ}$C. The vial was opened and 20 µl of deionized water was added, it was then vortexed for 1 minute and the solution was heated at 50 °C for 15 minutes in order to dissolve the DNA.

4.2.3 Confirmation of integrity of pQE30-G4LZI3 by restriction digest analysis

The integrity of the pQE30-G4LZI3 construct was confirmed by restriction digest analysis, this reaction was carried out using BamH1 and Hind111 as restriction enzymes. The reaction set-up is as shown in Table 4.1. The reagents were mixed in an Eppendorf tube and incubated in a water-bath at 37 °C for 40 minutes. Thereafter, the reaction samples were analysed by electrophoresis on a 1.2 % agarose gel.

Table 4.1: Reaction set up of restriction digest. The pQE30-G4LZI3 construct from Genscript was digested using BamH1 and Hind111

Reagents	Volume (µl)
pQE30-G4LZI3	10
Deionized water	5
Restriction enzyme buffer	2
BamH1	1
Hind111	2
Total	20

4.2.4 Agarose gel electrophoresis

A 1.2% agarose gel was prepared by mixing 1.2 g of agarose gel with 150 ml 1x TAE buffer, this mixture was heated in a microwave oven for about 5 minutes till all agarose had melted. The gel solution was allowed to cool and then 10 µl of ethidium bromide was added by slightly mixing it into the gel solution. The gel was gently poured into a gel rack and a gel comb was inserted into one side of the gel and was allowed to set for about 20 minutes. Thereafter, the comb was detached from the solidified gel and the rack (with the gel in it) was put into a chamber containing 1x TAE gel running buffer. The gel was positioned with the wells towards the negative current. Loading buffer was put in the DNA samples in order to visualize the DNA bands in the gel. A DNA ladder having a mixture of DNA fragments with known sizes, was loaded into the first well

followed by the restriction digest products. The electrophoresis chamber was covered and the gel was run at 80 volts for 45 minutes. Thereafter, the ethidium bromide stained gel was viewed under UV light and photographed.

4.2.5 Preparation of competent *E. coli* JM109 cells for transformation

A single colony of *Escherichia coli* JM109 cells (ThermoFisher Scientific, USA) (previously grown overnight from glycerol stock) was picked from an LB agar plate and was grown in 3 ml LB broth overnight. The next morning 500 µl of the overnight culture was put into 100 ml of LB broth. The growth of the culture was monitored with a spectrophotometer (Shimadzu) until the optical density had reached between 0.45-0.47, the flask was quickly immersed in ice water and swirled. The culture was centrifuged at 5 000 rpm at 4 °C for 5 minutes. Thereafter, the cell pellet was re-suspended in 24 ml of ice cold 0.1 M $MgCl_2$ and incubated on ice for 30 minutes, centrifugation was done for 5 minutes at 5 000 rpm, the supernatant was discarded and the cell was re-suspended in 24 ml ice cold sterile 0.1M $CaCl_2$ by pipetting gently and then incubated on ice for 4 hours. Centrifugation at 5 000 rpm was done and the supernatant was discarded, while the cell pellet was retained. Thereafter, the cells were further re-suspended in 4.3 ml of 0.1 M ice-cold sterile $CaCl_2$. Sterile 30 % glycerol (700 µl) was added and then thoroughly mixed and the competent cells made was dispensed in sterile Eppendorf tubes as 200 µl aliquots, snap freezed and stored at -80 $^{\circ}$C until when needed.

4.2.6 Transformation of *E. coli* JM109 with *S. mansoni* pQE30-G4LZI3 construct

After the confirmation of the plasmid construct by restriction digestion as stated previously in section 4.2.3, 2 µl was mixed with 100 µl of pre-thawed (thawed on ice) competent JM109 cells. This served as the transformation mixture and the tube was

labelled experimental tube. A control tube with only competent cells without plasmid DNA was set up and labelled as control tube. The mixtures in both sets of tubes were incubated on ice for 20 minutes and heat-shocked by incubating at 42 °C for 2 minutes. The mixtures were further incubated on ice for 5 minutes. Afterwards, 900 µl of pre-warmed Luria broth with ampicillin was added to the mixtures and further incubated at 37 °C for 1 hours. After this period of incubation, 50 µl of the transformed cells was plated out on LB agar plates with 100 µg/ml ampicillin to select for positive transformants. Thereafter, the plates were incubated overnight at 37 °C.

4.2.7 Expression screening of colonies from transformed *S. mansoni* pQE30-G4LZI3

Four single colonies were randomly picked from the overnight transformation spread-plate and inoculated into five tubes containing 5 ml 2x YT medium broth with 100 µg/ml ampicillin. The culture tubes were incubated at 37 °C for 4 hours with vigorous shaking. After this incubation period, 1 ml culture was taken from each culture tubes and transferred into clean autoclaved tubes; this was done in duplicate. One set of tubes was labelled un-induced control culture, while 0.5 mM Isopropyl β-D-thiogalactoside (IPTG) was added to the other set of tubes and labelled induced culture. Both sets of culture tubes were incubated further for 2 hours at 37 °C with vigorous shaking. Thereafter, the cells from both sets of cultures was harvested by centrifugation at 13 200 x *g* for 5 minutes. The cell pellets from all the tubes (both induced and un-induced) were re-suspended in 50 µl PBS. Thereafter, 20 µl from each cell suspension was electrophoresed on 16 % SDS-PAGE gel till the dye front had reached the bottom of the

gel. The 3 ml leftover original culture was used to prepare glycerol stocks and stored in 20 °C for further expression.

4.2.8 Large scale expression of pQE30-G4LZI3

Sequel to the successful expression of protein from the screening, 100 µl of the well expressed protein clone was inoculated into 100 ml of 2x YT medium broth to which 100 µg/ml of ampicillin has been previously added and was agitated vigorously overnight at 37 °C. The next morning, the overnight culture was scaled up to 2 L with 2x YT containing 100 µg/ml of ampicillin and the incubation was continued with shaking at 37 °C. The optical density of the cell culture was monitored until it attained 0.4-0.6 OD_{600} (measured using UV spectrophotometer) and then 0.5 mM of IPTG was added to induce protein expression. The incubation of the culture was continued overnight but induction temperature was reduced to 25 °C with vigorous shaking. The following morning, the cells were harvested by centrifugation at 6 000 x *g* for 40 minutes at 4 °C and the harvested cell pellets was kept at -80 °C.

4.2.9 Extraction of pQE30-G4LZI3 His-tagged recombinant protein

The cells harvested from 4.2.8 above was allowed to thaw completely (on ice) then it was re-suspended in cell lysis buffer, 100 µg/ml lysozyme, 1 mM PMSF and 1 mM DTT was added and rocked gently for 1 hour to assist enzymatic action of the lysozyme. Three cycles of freeze-thawing (-80 °C to freeze, 37 °C to thaw) was done to break-up the cells in order to release the recombinant G4LZI3 protein, this was followed by ultra-sonication (3 cycles for 30 seconds each) to fully break-up the cells and further release the recombinant protein. The recombinant protein was extracted by centrifugation of the cell suspension at 10 000 x *g* for 30 minutes at 4 °C. After the centrifugation, the

supernatant which contained a clear total bacterial lysate without cell debris and particulate matter was decanted into a new tube and placed on ice for purification procedure.

4.2.10 Purification of pQE30-G4LZI3 His-tagged recombinant protein using immobilized metal-ion affinity chromatography (IMAC)

A 7 ml His-Select Nickel Affinity Gel (Sigma-Aldrich) was packed onto a column. The affinity gel was first washed with 2 CV of deionized water to eliminate the storage ethanol and then equilibrated with 5 CV of equilibration buffer (buffer A). The bacterial lysate was then loaded onto the column and the flow-through was collected and kept on ice. Thereafter, the column was washed with 7 CV of buffer A and the wash fraction and the last two drops of the wash buffer was collected to ascertain the clearness of the column before elution. Subsequently, the target protein which is expected to have bound to the column was eluted with 7 CV of the elution buffer (buffer B) and the various fractions collected. The fractions collected were electrophoresed on a 16 % SDS PAGE gel.

4.2.11 SDS Polyacrylamide gel electrophoresis

SDS-PAGE was used to determine the molecular weight and purity of the various fractions of the recombinant protein. The casting frame was set up by clamping two glass plates on the casting stand. The separating gel solution was prepared as described in Table 4.2 in a small beaker and the solution was mixed gently but thoroughly and filled into the gap within the glass plates. In order to ensure evenness of the gel surface, water was put on top of the gel until there was an overflow. The separating gel was left to polymerize for 20 minutes and the water on it was discarded.

Thereafter, the stacking gel was prepared as described previously and filled into the glass over the separating gel until there was an overflow and then the well-forming comb was inserted into it ensuring that there was no air trapped in-between. The stacking gel was left to polymerize for 20 minutes. After the gel has polymerized completely, the comb was detached and the glass plate was removed from the casting frame and set up in the buffer chamber. The inner and outer buffer chambers were filled with appropriate amount of running buffer. A 20 µl of the sample was mixed with 20 µl of loading buffer and heated for 10 minutes at 95 $^{\circ}$C.

A 10 µl of heated sample and 5 µl of a standard protein ladder were then loaded into the wells and the gel was run at 150 V for 90 minutes. Thereafter, the gel was soaked in Coomassie staining solution overnight with slight shaking. The next morning, the gel was destained (preparation of destaining solution previously described) for 90 minutes with slight shaking and then band corresponding to target protein was identified by comparing its molecular weight with the standard protein ladder.

Table 4.2: Constituents of a 16% SDS-PAGE gel

	Separating gel	Stacking gel
Deionised water	3.50 ml	2.90 ml
Separating buffer	2.62 ml	-
4x stacking buffer	-	1.26 ml
40% Bisacrylamide	3.48 ml	0.50 ml
Ammonium persulphate	50.00 µl	25.00 µl
10% SDS	100.00 µl	50.00 µl
TEMED	8.00 µl	5.00 µl

4.2.12 Dialysis

Purified protein elution fractions were pooled together and transferred into a Snake Skin Pleated Dialysis tubing (ThermoFisher Scientific) with a molecular weight cut off (MWCO) of 3500 Da and dialyzed for 4 hours at 4 °C in a freshly-prepared buffer (50 mM Tris, pH 8 and 100 mM NaCl with 1 mM DTT) with continuous stirring. The buffer was changed and the dialysis was continued for further 6 hours. Thereafter, 60 ml of dialysed protein was concentrated into 7 ml with Pall Macrosep Centrifugal device with a MWCO of 3.5. The concentrated protein was analysed on SDS-PAGE and subsequently stored at -80 °C after aliquoting into 1.5 ml Eppendorf tubes.

4.2.13 Protein concentration determination

A NanoDrop® ND 2000 spectrophotometer spectrophotometer (Thermo Fisher Scientific) was used to determine the concentration of the protein by evaluating the absorbance at 280 nm. A 2 µl of buffer B was carefully put on the detection surface of the spectrophotometer to blank it, 2 µl of concentrated G4LZI3 was then put on the detection surface and the absorbance of the protein sample was recorded.

4.3 Results

4.3.1 Molecular cloning and recombinant expression of G4LZI3 gene

The NCBI database provided the DNA sequence as well as the protein sequence of the universal stress protein as reported in the *in-silico* chapter. The DNA sequence and FASTA mRNA sequence was used by GeneScipt incorporation to synthesize a codon-harmonized G4LZI3 gene construct which was cloned in the BamH1 and Hind111 sites of the pQE30 cloning vector. The pQE30-G4LZI3 gene construct is presented in Fig. 4.1. Restriction digest analysis done using BamH1 and Hind111 restriction enzymes to

confirm the integrity of the construct from Genscript showed that construct is of the expected size of 555 bp. A 1.2 % agarose gel elecrophoresis showing the restriction digest insert is shown in Fig. 4.2.

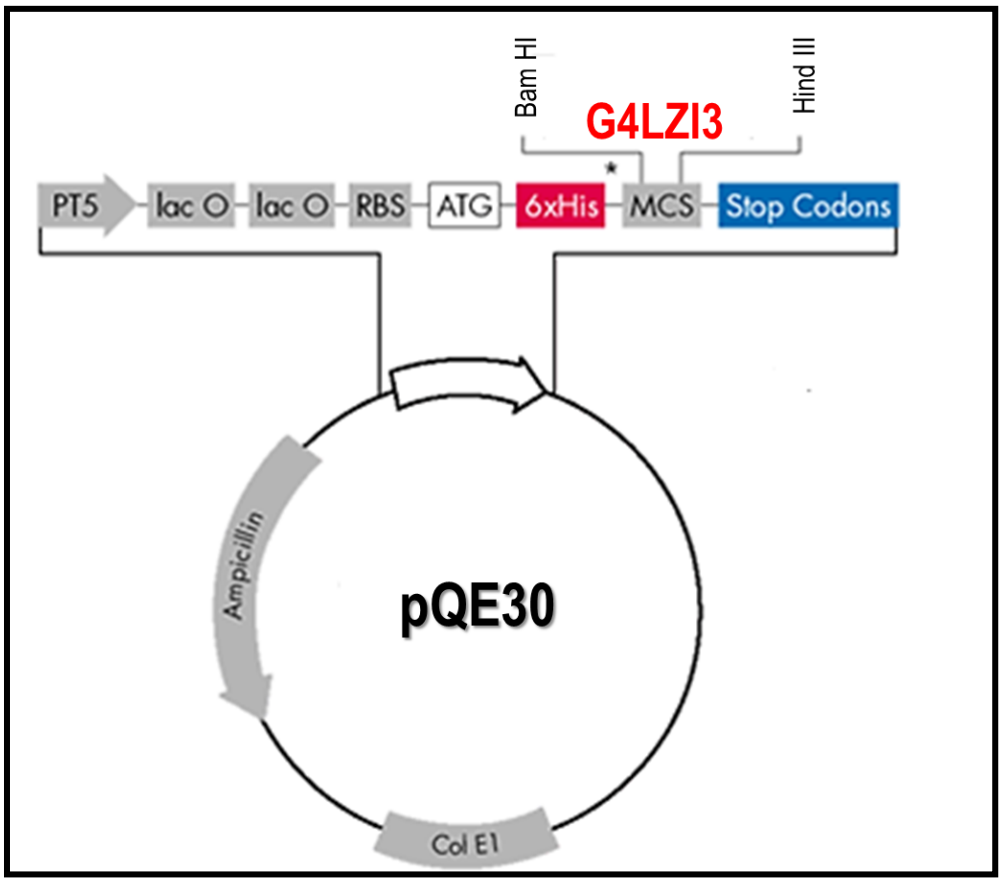


Fig. 4.1: A diagrammatic representation of pQE30-G4LZI3 gene construct. The diagram illustrate the insertion of the G4LZI3 into BamH1 aand Hind111 sites of the pQE30 vector.

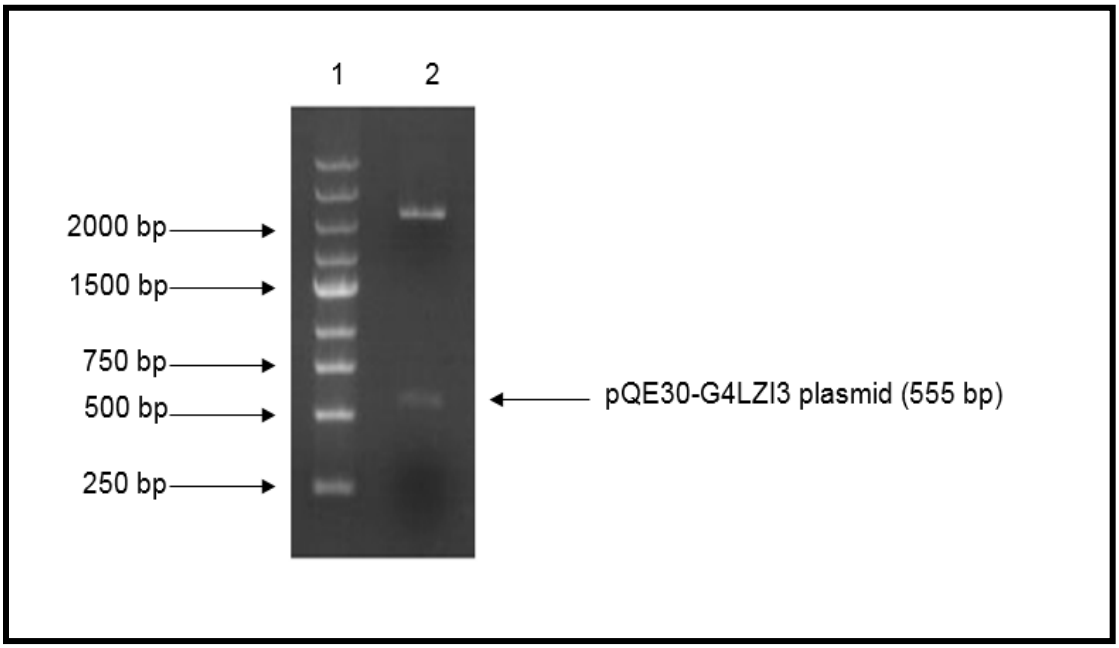


Fig. 4.2: Restriction digest gel for confirmation of G4LZI3 insert fragment. Lane 1 is the DNA marker, lane 2 is the plasmid digested by BamH1 and Hind111.

4.3.2 Transformation and expression screening of pQE30-G4LZI3

E. coli JM109 cells were transformed with pQE30-G4LZI3 construct and the positive transformants produced were screened for the expression of His-G4LZI3 fusion protein. SDS-PAGE analysis of the total bacterial lysates from four colonies which were picked randomly is shown in Fig. 4.3. Lanes 2-9 consist of alternating pairs of induced and un-induced samples. A conspicuous band can be seen in the induced lanes but are absent in the un-induced lanes. The conspicuous bands in the induced lane shows that His-G4LZI3 is well expressed, the bands correspond to a molecular weight of approximately 20.3 kDa.

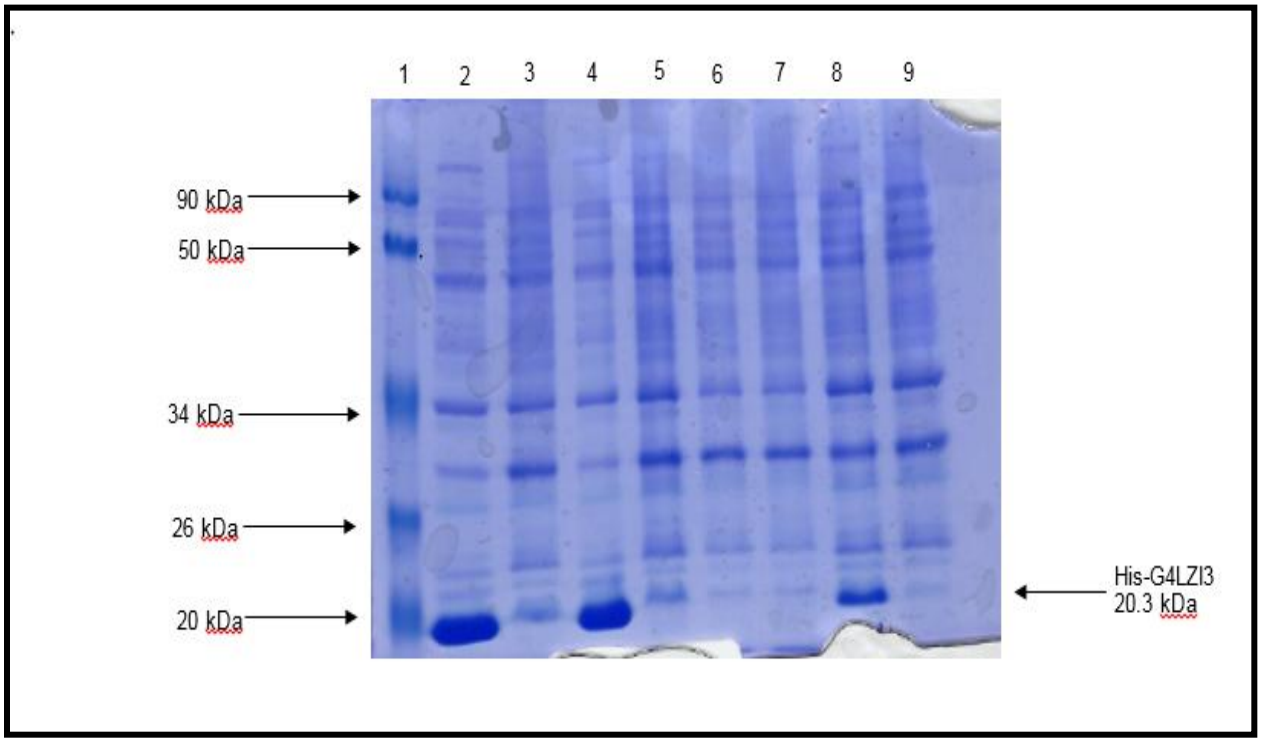


Fig. 4.3: Expression screen of transformed colonies from pQE30-G4LZI3 for protein expression. Lane 1 is the protein molecular weight marker, lanes 2, 4, 6 and 8 indicates total bacterial cell lysates after induction of protein expression with addition of 0.5mM IPTG, while lanes 3, 5, 7 and 9 indicates the corresponding total bacterial cell lysates of the un-induced cultures. The induced lanes (2, 4 and 8) shows bands of approximately 20.3 kDa, corresponding to the His-G4LZI3 fusion protein.

4.3.3 Large scale expression and purification of His-G4LZI3

The large scale expression of His-G4LZI3 was successfully done. After extraction of the crude lysate, a clear soluble protein bacterial lysate without the presence of inclusion body was obtained after centrifugation. Fig 4.4 shows the SDS-PAGE analysis of the soluble protein. The expressed soluble protein was successfully purified in His-Select® Nickel Affinity gel. Fig.4.5 showed the SDS-PAGE analysis of the purification procedure, the elution fractions shows varying concentration of G4LZI3 protein as indicated by the decline in the thickness of the protein bands.

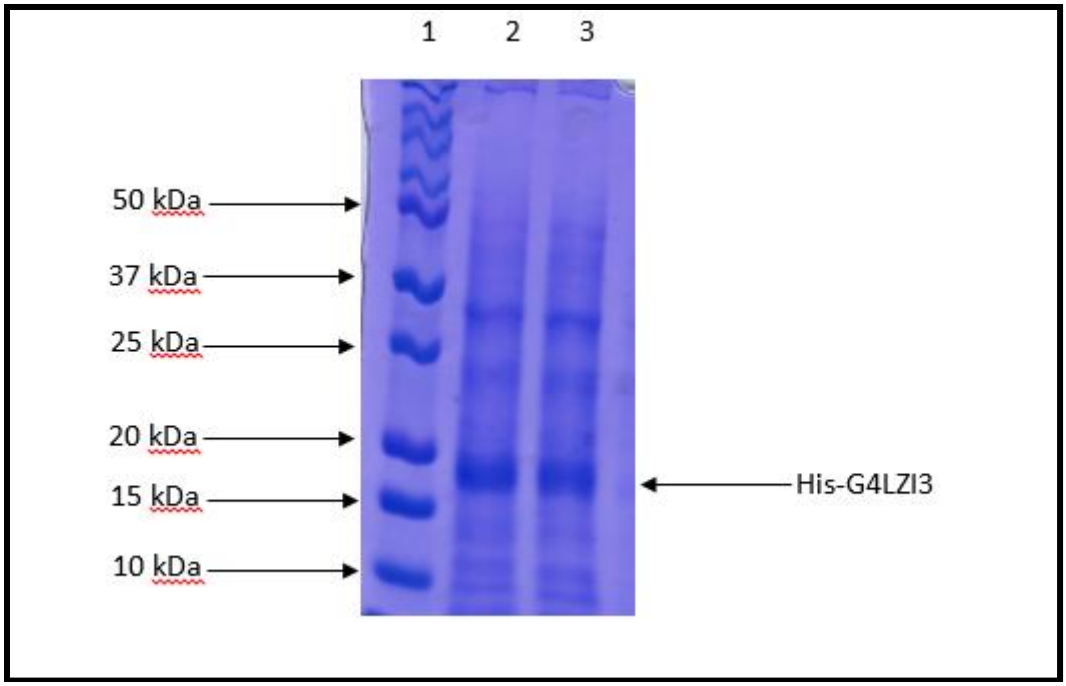


Fig. 4.4: Large scale expression of His-G4LZI3. Lane 1 is the molecular weight maker; lane 2 is loaded with 10 µl of protein sample taken from 100 ml expression volume while lane 3 is loaded with 10 µl of protein sample taken from 200 ml expression volume.

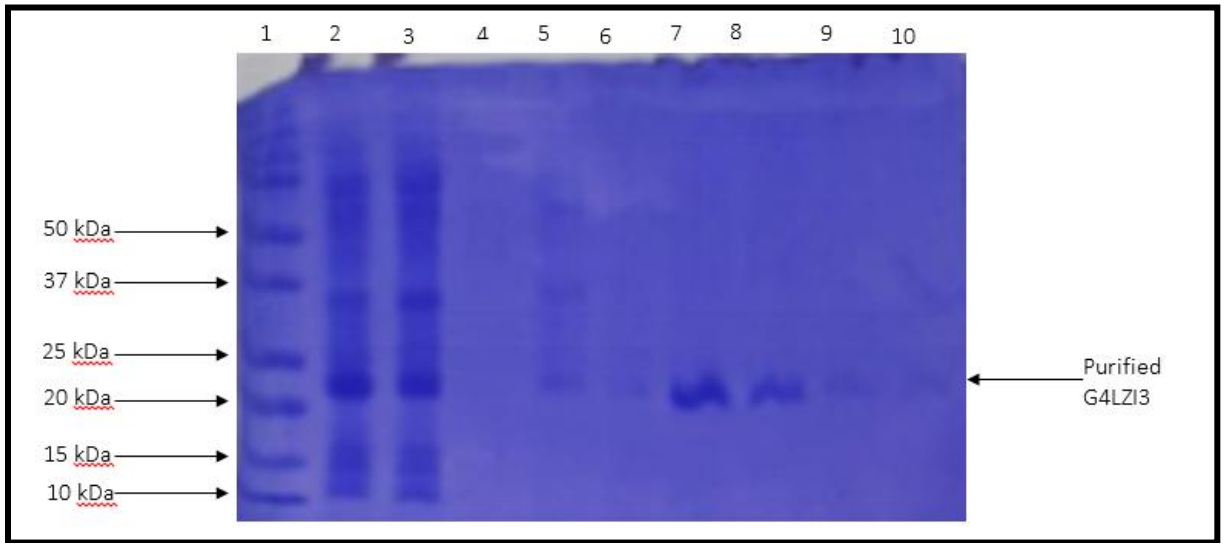


Fig. 4.5: Affinity purification of His-G4LZI3 with His-Select Nickel affinity gel. Lane 1 is the protein marker; lane 2 is the clear bacterial lysate; lane 3 is the flow-through from the affinity column; lane 4 is the second wash fraction; lane 5 is the first wash fraction; lane 6, 7, 8, 9 and 10 are the elution fractions. The decline in the thickness of the bands on the elution fractions from lane 7 down to lane 9 is due to decrease in the concentration of the protein present in the elutions as the protein is eluted from the column.

4.3.4 Dialysis and protein concentration

The first attempt at dialysing the protein resulted in slight precipitation of the purified protein but after ensuring the dialysis was strictly done at 4 °C and the duration of dialysis was reduced (from 24 hours to 6 hours), successful dialysis of the protein was achieved. The dialysis was done to reduce the concentration of imidazole in the purified protein. A 20 ml dialysed protein was successfully concentrated to 3 ml with a Pall Macrosep Centrifugal device with a MWCO of 3.5. Aliquots were made and kept for further use. Fig. 4.6 shows the SDS-PAGE analysis of 10 µl sample of the protein taken from 3 ml concentrated protein.

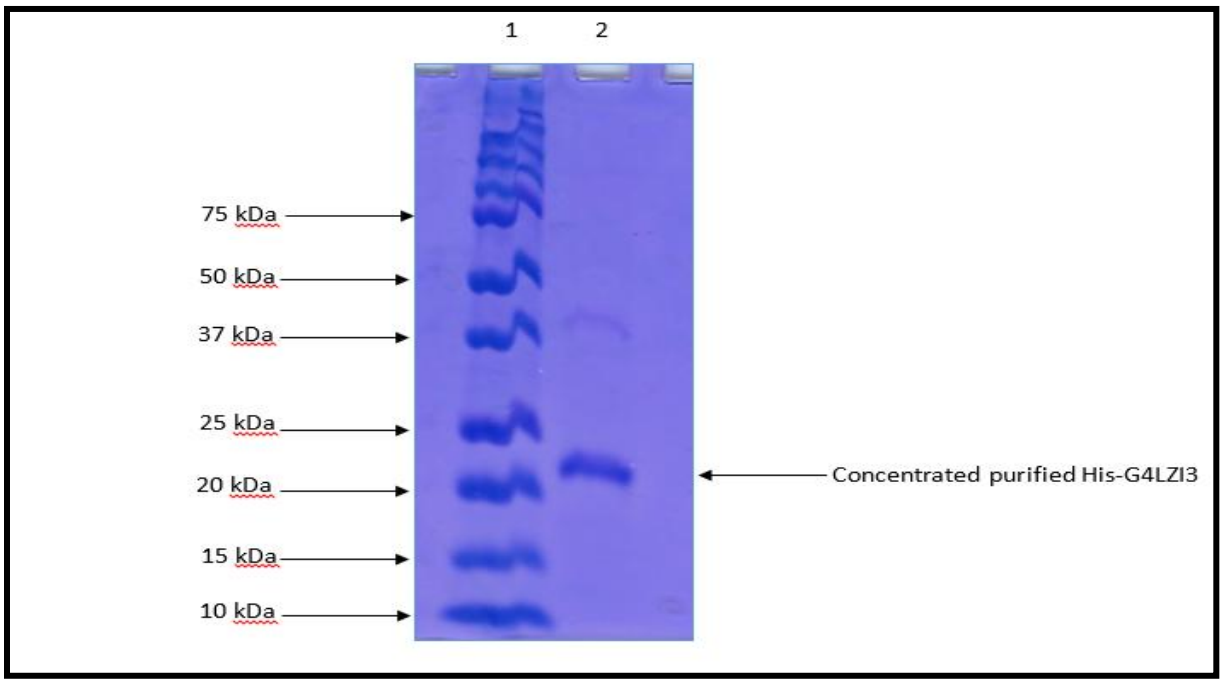


Fig. 4.6: Concentration of the elution fractions. Lane 1 is the molecular weight marker; lane 2 is the 10 µl sample taken from 3 ml concentrated purified protein.

4.3.5 Measurement of protein concentration

Fig. 4.7 shows the concentration of a sample of purified G4LZI3 using NanoDrop® ND 2000 spectrophotometer. The concentration of the protein at 280 nm is 2.00 mg/ml.

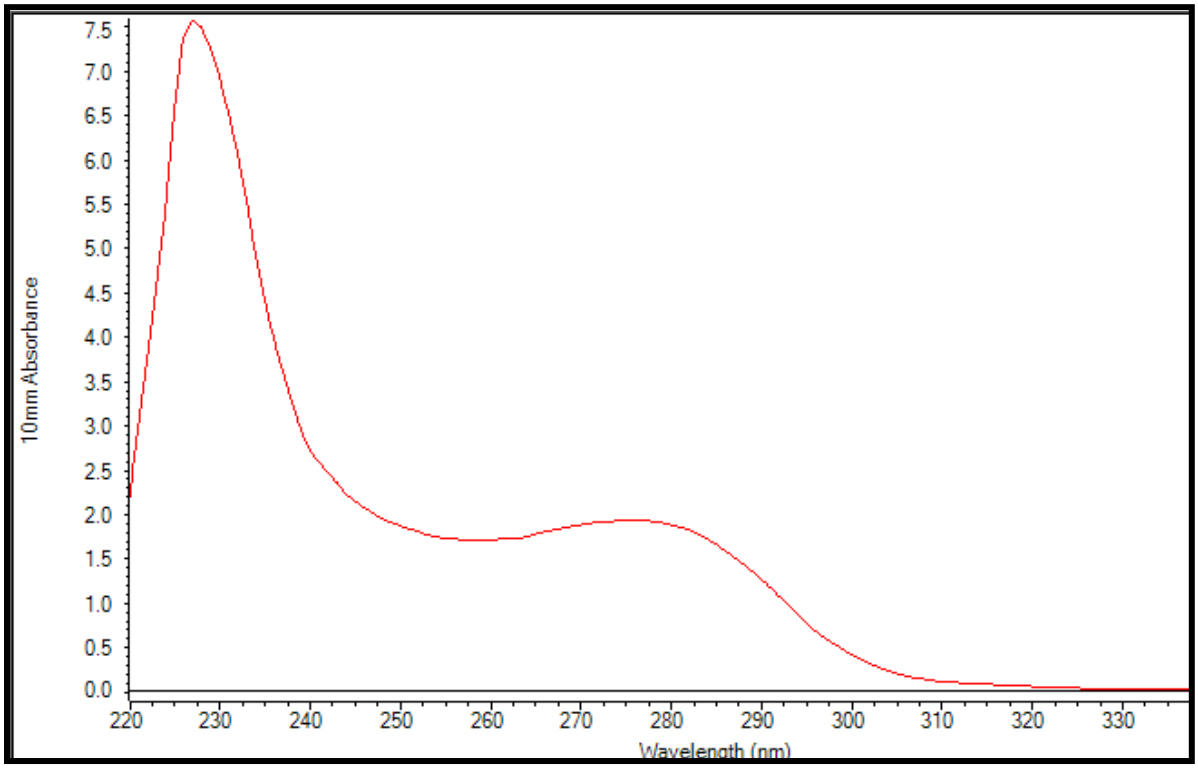


Fig. 4.7: The UV absorption spectrum of concentrated G4LZI3 absorbance at 280 nm (A280 nm) recorded using NanoDrop ® ND1000 Spectrophotometer (NanoDrop Technologies Inc.).

4.4 Discussion

To our knowledge, this is the first report of the successful cloning, purification, recombinant expression and preliminary characterization of the recombinant universal stress protein of *Schistosoma mansoni* parasite, G4LZI3. The capability to express and purify a recombinant protein in sufficient quantity is extremely essential for structural elucidation, biochemical characterization and functional studies which will ultimately lead to the use of such recombinant protein in production of therapeutic agents for combating diseases as well as in industrial processes for the development of commercial goods

To achieve the aim of producing the *S. mansoni* G4LZI3, an expression plasmid, pQE30-G4LZI3, fused to a Histidine tag was successfully cloned. The codon harmonised pQE30-G4LZI3 was cloned into the BamHI and Hind111 cloning sites. Restriction digest analysis showed that the size of the inserted fragment is correct (555 bp) and it is devoid of unexpected bands, this shows that the restriction enzymes (BamH1 and Hind111) were able to successfully cut the G4LZI3 plasmid DNA at the required restriction sites hence preparing it for combination with the *E. coli* host DNA to produce a recombinant DNA. The pQE30-G4LZI3 construct was codon optimized with the sole aim of increasing recombinant protein production by increasing the translation efficiency during the process of expression (Angov *et al.,* 2008; Gustaffson *et al.,* 2012; He *et al.,* 2014; Quax *et al.,* 2015).

A wide range of protein expression systems are obtainable including cell cultures of mammals, bacteria, yeasts, transgenic plants and insects. *E. coli* was chosen for this

study because it possesses the necessary factors for quality of protein yield, speed of production and functionality (Demain and Vaishnav, 2009; Chen *et al.,* 2012; Carneiro *et al.,* 2013). Protein expression is reliant on multistep processes comprising regulation at the level of transcription, mRNA turnover, protein translation, and post-translational modifications resulting in the formation of a stable product (Angov, 2011). The gene map in Fig. 4.1 showed that the expression vector used (pQE30 vector) has a T5 promoter recognised by *E. coli* host RNA polymerase and it is inducible upon addition of IPTG, this is a crucial factor to consider in ensuring successful protein expression. The expression screening step was observed to produce higher level of expression in the induced lane compared with the un-induced, this is attributed to the addition of IPTG to the samples in the induced lanes. Addition of IPTG switches on synthesis of the T5 RNA polymerase by activating the release of tetrameric Lac I from the *lac* operator which subsequently triggers the transcription of the target gene from the T5 promoter that is initiated by T5 RNA polymerase (Sørensen and Mortensen, 2005).

The best expressing protein clone from the expression screening was used to produce the protein at a large scale level. The large scale expression was achieved successfully, several conditions were tried for induction; 1mM IPTG at 37 $^{\circ}$C for 5 hours, 0.5 mM at 37 $^{\circ}$C overnight (results not shown) but the best expression result was obtained by the addition of 0.5 mM IPTG with incubation at 25 $^{\circ}$C overnight as shown in Fig. 4.4. Previous studies have reported the successful heterologous expression of universal stress proteins from several species of organisms such as *E. coli*, *Saliconia brachiata* and *Mycobacterium* in *E. coli* expression system. (Uduwat *et al.,* 2014; de Souza *et al.,* 2016; Jia *et al.,* 2016). Protein extraction and subsequent SDS-PAGE analysis showed

that the protein was present in the supernatant and not in the pellet, this is an indication that G4LZI3 was produced as a soluble protein and not as inclusion bodies or insoluble aggregates. This suggests that overexpression of G4LZI3 with 0.5mM IPTG with incubation at temperature of 25 °C overnight is ideal for achieving protein solubility.

Formation of inclusion bodies in recombinant proteins overexpressed in *E. coli* cells might result from build-up of high concentrations of folding intermediates. However, refolding of target protein from inclusion bodies is achievable but is undesirable due to disadvantages such as reduced recovery yields, the conditions for optimization of refolding requirements for individual target protein as well as the likelihood that the re-solubilisation procedures might alter the integrity of refolded proteins. Hence, exploiting conditions for the overexpression of recombinant proteins in a soluble form is thus a smart alternative compared with *in vitro* refolding procedures (Sørensen and Mortensen, 2005a; Sørensen and Mortensen, 2005b). A strategy to avoid or reduce formation of inclusion body is to decrease cultivation temperature of the host after induction (Gopal and Kumar, 2013). The clear soluble recombinant His-G4LZI3 lysate was successfully purified with Immobilized Metal ion Affinity Chromatography (IMAC) using His-Select® Nickel Affinity gel (Sigma-Aldrich) as can be seen in Fig. 4.5. IMAC is a method centred on the interface between proteins having histidine residues on its exterior with divalent metal ions immobilized through chelating ligand. Proteins with Histidine-tag possess very high affinity in IMAC due to the multiple histidine residues and normally they are the toughest binders amongst every other protein present in a crude extract. It is interesting to note that the imidazole side chain present in the histidine tag have high affinity for chelated metals (Arnau *et al.,* 2006; Block *et al.,* 2008; Young *et al.,* 2012).

The His-tag was not cleaved-off from the protein after purification because generally, small-size fusion tags like histidine, calmodulin-binding peptide (CBP), FLAG and Strep II commonly do not need to be cleaved-off before further downstream applications after purification since these tags have an insignificant effect on the physical characteristics as well as the biological activity of the target protein (Esposito D. and Chatterjee, 2006; Zhao *et al.,* 2013). Sagurthi *et al.,* (2007) successfully purified a small cytoplasmic bacterial protein, universal stress protein UspF (YnaF) from *Salmonella typhimurium* with Ni–NTA affinity chromatography prior to crystallization and preliminary characterization with X-ray diffraction analysis (Sagurthi *et al.,* 2007). Similar success of purification with Ni-NTA affinity column was reported by Jung and co-workers (2015) in their study on the universal stress protein of *Arabidopsis thaliana*, AtUSP (Jung *et al.,* 2015). Our purification outcome and the reports from other purification studies might suggest that Ni-NTA is ideal for purification of recombinant universal stress proteins. Proteins may be characterized by using different structural, spectroscopic, electromagnetic, thermodynamics and biochemical properties which are conferred on them due to their composition. Production of proteins and subsequent structural and functional characterization exploits cumulative information from their fundamental general makeup (primary and secondary elements) and the diversity allowed by such composition (Vedadi *et al.,* 2010; Raynal *et al.,* 2014).

4.5 Conclusion

Reported in this chapter is the successful molecular cloning into a pQE30 cloning vector and recombinant expression in *E. coli* host system followed by an efficient purification of a universal stress protein G4LZI3 from the helminthic disease causing organism

Schistosoma mansoni. We have been able to establish that Histindine tagged G4LZI3 expresses very well as a soluble protein in the *E. coli* host system without fear of the production of inclusion body. Additionally, the Ni-NTA affinity purification system led to a relatively homogenously pure G4LZI3 which can be used to achieve the next objective of preliminarily characterising the protein as regards its structure.

References

Angov E. (2011). Codon usage: nature's roadmap to expression and folding of proteins. *Biotechnology Journal,* **6**(6): 650-659.

Angov E., Hillier C.J., Kincaid R.L. and Lyon J.A. (2008). Heterologous protein expression is enhanced by harmonizing the codon usage frequencies of the target gene with those of the expression host. *PloS One,* **3**(5): e2189.

Arnau J., Lauritzen C., Petersen G.E. and Pedersen J. (2006). Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins. *Protein Expression and Purification,* **48**(1): 1-13.

Block H., Maertens B., Spriestersbach A., Brinker N., Kubicek J., Fabis R., Labahn J. and Schäfer F. (2009). Immobilized-metal affinity chromatography (IMAC): a review. *Methods in Enzymology*, **463**: 439-473.

Carneiro S., Ferreira E.C. and Rocha I. (2013). Metabolic responses to recombinant bioprocesses in Escherichia coli. *Journal of Biotechnology*, **164**(3): 396-408.

Chen R. (2012). Bacterial expression systems for recombinant protein production: E. coli and beyond. *Biotechnology Advances*, **30**(5): 1102-1107.

de Souza C.S., Torres A.G., Caravelli A., Silva A., Polatto J.M. and Piazza R.M. (2016). Characterization of the universal stress protein F from atypical enteropathogenic *Escherichia coli* and its prevalence in *Enterobacteriaceae*. *Protein Science*, **25**(12): 2142-2151.

Demain A.L. and Vaishnav P. (2009). Production of recombinant proteins by microbes and higher organisms. *Biotechnology Advances,* **27**(3): 297-306.

Esposito D. and Chatterjee D.K. (2006). Enhancement of soluble protein expression through the use of fusion tags. *Current Opinion in Biotechnology,* **17**(4): 353-358.

Gopal G.J. and Kumar A. (2013). Strategies for the production of recombinant protein in *Escherichia coli. Protein Journal*, **32**(6): 419-425.

Gustafsson C., Minshull J., Govindarajan S., Ness J., Villalobos A. and Welch M. (2012). Engineering genes for predictable protein expression. *Protein Expression and Purification*, **83**(1): 37-46.

He M., Wu D., Wu J. and Chen J. (2014). Enhanced expression of endoinulinase from *Aspergillus niger* by codon optimization in Pichia pastoris and its application in inulooligosaccharide production. *Journal of Industrial Microbiology and Biotechnology*, **41**(1): 105-114.

Jia Q., Hu X., Shi D., Zhang Y., Sun M., Wang J., Mi K. and Zhu G., (2016). Universal stress protein Rv2624c alters abundance of arginine and enhances intracellular survival by ATP binding in mycobacteria. *Scientific Reports*, ***6***.

Jung Y.J., Melencion S.M.B., Lee E.S., Park J.H., Alinapon C.V., Oh H.T., Yun D.J., Chi Y.H. and Lee S.Y. (2015). Universal Stress Protein exhibits a redox-dependent chaperone function in Arabidopsis and enhances plant tolerance to heat shock and oxidative stress. *Frontiers in Plant Science*, **6**.

Quax T.E., Claassens N.J., Söll D. and van der Oost J. (2015). Codon bias as a means to fine-tune gene expression. *Molecular Cell,* **59**(2): 149-161.

Raynal B., Lenormand P., Baron B., Hoos S. and England P. (2014). Quality assessment and optimization of purified protein samples: why and how? *Microbial Cell Factories,* **13**(1): 180.

Sagurthi, S. R., Panigrahi, R. R., Gowda, G., Savithri, H. S., & Murthy, M. R. N. (2007). Cloning, expression, purification, crystallization and preliminary X-ray diffraction analysis of universal stress protein F (YnaF) from Salmonella typhimurium. *Acta Crystallographica Section F: Structural Biology and Crystallization Communications*, **63**(11): 957-960.

Sørensen H.P. and Mortensen K.K. (2005). Advanced genetic strategies for recombinant protein expression in *E. coli*. *Journal of Biotechnology*, **115**:113-128.

Sørensen H.P. and Mortensen K.K. (2005). Soluble expression of recombinant proteins in the cytoplasm of Escherichia coli. *Microbial Cell Factories,* **4**(1): 1.

Udawat P., Mishra A. and Jha B. (2014). Heterologous expression of an uncharacterized universal stress protein gene (SbUSP) from the extreme halophyte, *Salicornia brachiata*, which confers salt and osmotic tolerance to *E. coli. Gene*, **536**(1): 163-170.

Vedadi M., Arrowsmith C.H., Allali-Hassani A., Senisterra G. and Wasney G.A. (2010). Biophysical characterization of recombinant proteins: a key to higher structural genomics success. *Journal of Structural Biology,* **172**(1): 107-119.

Young C.L., Britton Z.T. and Robinson A.S. (2012). Recombinant protein expression and purification: a comprehensive review of affinity tags and microbial applications. *Biotechnology Journal*, **7**(5): 620-634.

Zhao X., Li G. and Liang S. (2013). Several affinity tags commonly used in chromatographic purification. *Journal of Analytical Methods in Chemistry*, **2013**.

CHAPTER FIVE

PRELIMINARY BIOPHYSICAL CHARACTERIZATION OF PURIFIED RECOMBINANT *S. MANSONI* G4LZI3

5.1 Introduction

The significance of recombinant proteins in biotechnological research is much emphasized due to their diverse use as drug targets, therapeutic agents, investigation of protein-protein interactions, ligand binding, as well as in studying enzyme activity (Howell *et al.,* 2006; Bertucci *et al.,* 2011). Therapeutic recombinant proteins production usually entails series of steps in order to ensure biological quality of the protein since they are targeted for human clinical use. The expectation is that these proteins are properly characterized after expression and purification. Various methods are employed to explicitly confirm the protein structure, protein sequence as well as its activity. These characterization methods will consequently also aid structural genomic efforts (Barnes and Lim, 2007; Faraji *et al.,* 2010). In the present study, preliminary characterization of G4LZI3 was carried out using ultraviolet-visible spectrophotometry, fourier transform infrared spectroscopy, differential scanning calorimetry, mass spectroscopy and fluorescence spectroscopy.

5.2 Methodology

5.2.1 UV-visible spectroscopy

Optical absorption measurement was done on a Varian Cary® 50 UV-Vis spectrophotometer (Agilent Technologies). An aliquot of the purified G4LZI3 in 20mM Tris-HCl buffer (pH 8.0) was placed in a Quartz cuvette of 1 cm path length and the absorbance measured between 180-340 nm wavelength ranges at room temperature.

5.2.2 Fourier transform infrared spectroscopy (FTIR)

A 40 µl of the 0.1 mg/ml re-constituted protein sample was put in the sample cell of a Perkin Elmer FTIR instrument (Perkin-Elmer USA). FTIR spectra were obtained in the frequency range 4000-400 cm^{-1} in absorbance mode. The parameters selected for running the scans were resolution of 8 cm^{-1}, with an average of 15 scans, and no apodization option was selected. The data obtained was analysed with Origin software (MicroCal Software, Inc., MA, USA).

5.2.3 Dynamic scanning calorimetry

A thermal analysis of lyophilized G4LZI3 was done using Perkin-Elmer DSC 6000 (Perkin-Elmer USA). The G4LZI3 was reconstituted in distilled water and filtered thoroughly through 0.1 µm filter. The sample and the PBS reference solution were degassed by continuous stirring under vacuum for 4 minutes at room temperature.

Thereafter, the degassed sample and reference were filled into the sample and reference cells with an air-tight syringe. Heating of the sample and reference was done at 20 °C to 150 °C at a heating rate of 10 °C/min. Thereafter, the differential power compensation needed to maintain the sample cell and reference cell at the same temperature was recorded against temperature. This was used for calculating calorimetric enthalpy and mid-point transition temperature (Tm) necessary for the determination of the thermostability of the protein. Data obtained was interpreted using an Origin software (MicroCal Software, Inc., MA, USA).

5.2.4 Mass spectrometry

The G4LZI3 purified protein was subjected to mass spectrometry to validate the mass of the purified protein seen on SDS-PAGE. Matrix Assisted Laser Desorption/Ionization Time of Flight Mass Spectrometry spectra was acquired using an Ultra-flex time-of-flight mass spectrometer (Bruker Daltonics, Bremen, Germany) which is fitted with a LIFT-MS/MS facility. The sample was mixed with 2, 5- Dihydroxybenzoic acid which serves as a matrix and 50 % acetonitrile. Spectra were acquired in the positive ion mode and internally recalibrated after measurement.

5.2.5 Fluorescence spectroscopy

The fluorescence spectra of the recombinant G4LZI3 was recorded at 0.10 mg/ml in 50 mM Tris–HCl buffer (pH 7.0) at 25 °C using a Perkin Elmer Fluorescence Spectrophotometer (Perkin-Elmer USA) and 0.5 cm X 0.5 cm Quartz cuvettes. Tris-HCl buffer was used as blank and an excitation wavelength of 270 nm was used while the intrinsic fluorescence emission scanning was recorded between 200-450 nm.

5.3 Results

5.3.1 Investigation of purity of G4LZI3 with UV-visible spectroscopy

The UV-Visible curve produced by a sample of recombinant G4LZI3 presented in Fig. 5.1 shows the presence of two peaks at approximately 220 nm and 280 nm. These peaks correspond to the presence of peptide bonds which absorbs at 220 nm as well as the presence of aromatic amino acid residues including Phe, Tyr and Trp which absorbs at 280 nm. Worthy of note is the absence of absorbance peak at 260 nm and the diminishing aborbance after 280 nm, this might be attributed to the absence of

impurities. Typically, nucleic acid a major impurity during protein expression absorbs at 260 nm (Cao *et al.,* 2004; Lv *et al.,* 2015).

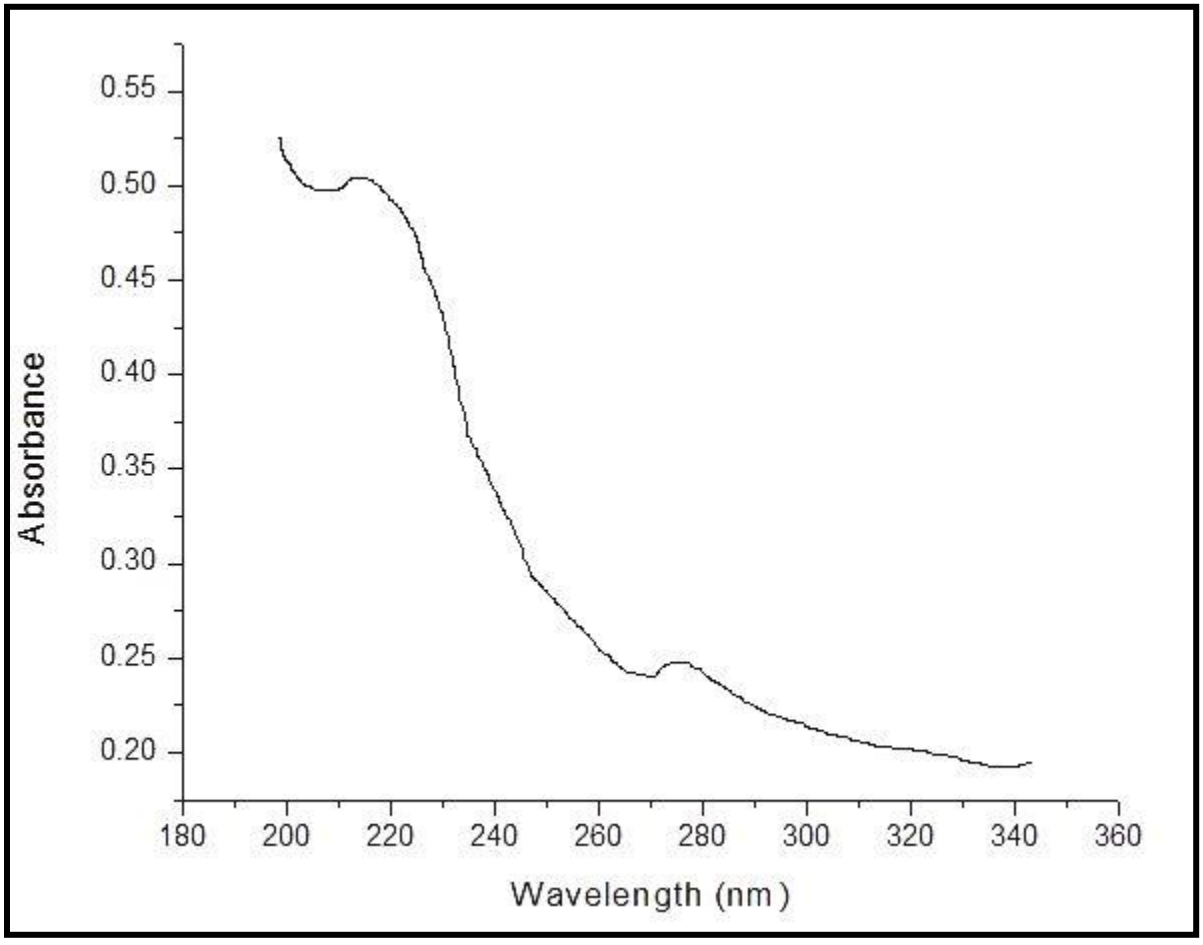


Fig. 5.1: UV-Visible absorbance spectrum of G4LZI3. The spectrum shows peaks at 220 nm and 280 nm are due to the presence of peptide bonds and aromatic amino acid residues respectively.

5.3.2 Analysis of secondary structure and functional groups with Fourier transform infrared spectroscopy (FTIR)

The occurrence and abundance of secondary structure elements and functional groups in the structure of G4LZI3 was analysed using FTIR. The spectrum in Fig. 5.2 shows the presence of three prominent peaks. Peak 1 can be attributed to the amide I band which is the sum of all the contributions produced by the secondary structure of the protein including α-helix, β-sheets, turns and unordered structures (Lv *et al.,* 2015).

The peak at approximately 1650 cm^{-1} is associated with the vibrational stretching of C=O stretching, the observed positioning and shape of the amide I band are used for the analysis of protein secondary structure. It is suggested that the sharp amide I band at 1650 cm^{-1} for G4LZI3 indicates that α-helix is the major element in the protein secondary structure. Beside this, the peak 2 at approximately 1200 cm^{-1} represents the secondary amide III band which is produced by the C=N stretching vibration coupled with the N-H plane bending vibration. The last peak at approximately 3340 cm^{-1} represents the Amide A group and it originates from the resonance between the first overtone of amide II and the N-H stretching vibration (Gallagher, 2009).

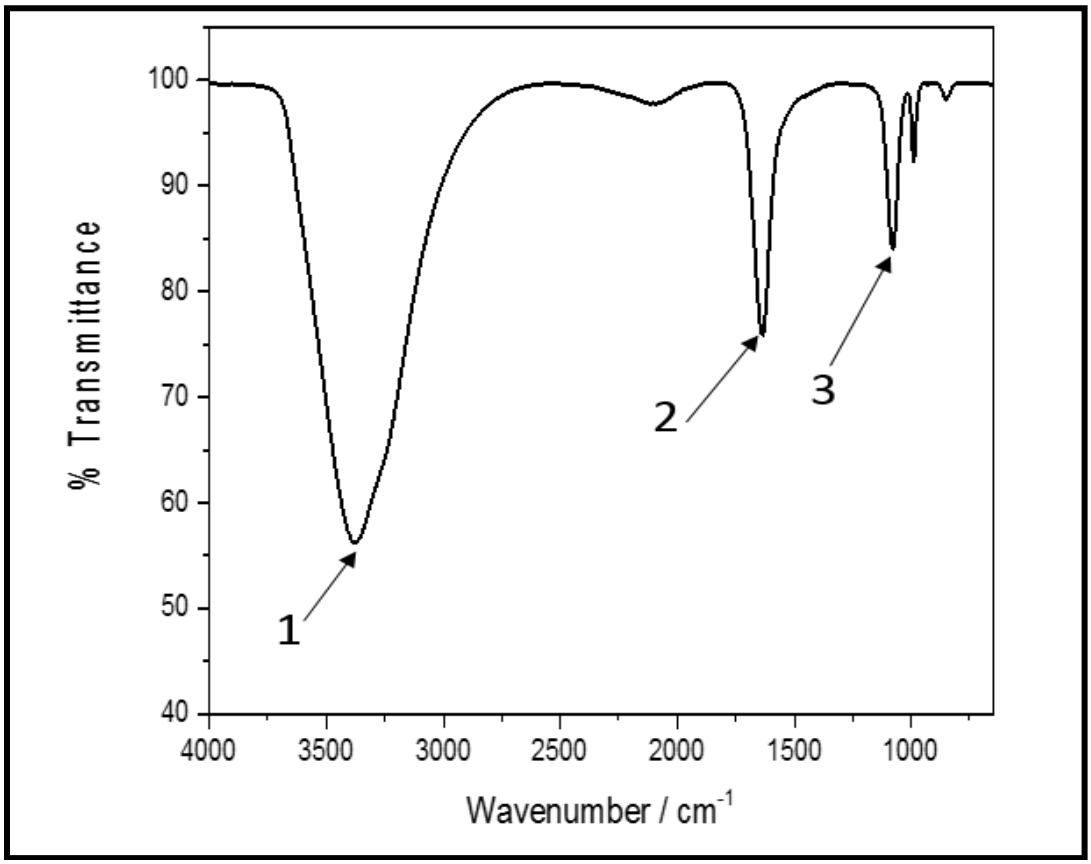


Fig. 5.2: IR spectrum of G4LZI3. Shown are three prominent peaks indicating the secondary structure element and functional groups present in the protein.

5.3.3 Investigation of thermal stability of G4LZI3 with differential scanning calorimetry

Differential scanning calorimetry analysis of G4LZI3 as depicted in Fig. 5.3 showed that the protein has distinctive sharp peak that is extrapolated to a temperature of 110 °C. This temperature designated by Tm is the transition temperature of the protein. Tm is an index of thermostability, usually, the greater the Tm, the more thermodynamically stable the protein is. Proteins that have higher Tm are less liable to unfolding and denaturation at lower temperatures. This result showed that the protein can remain stable at temperatures below 110 °C and becomes irreversibly denatured at higher temperature.

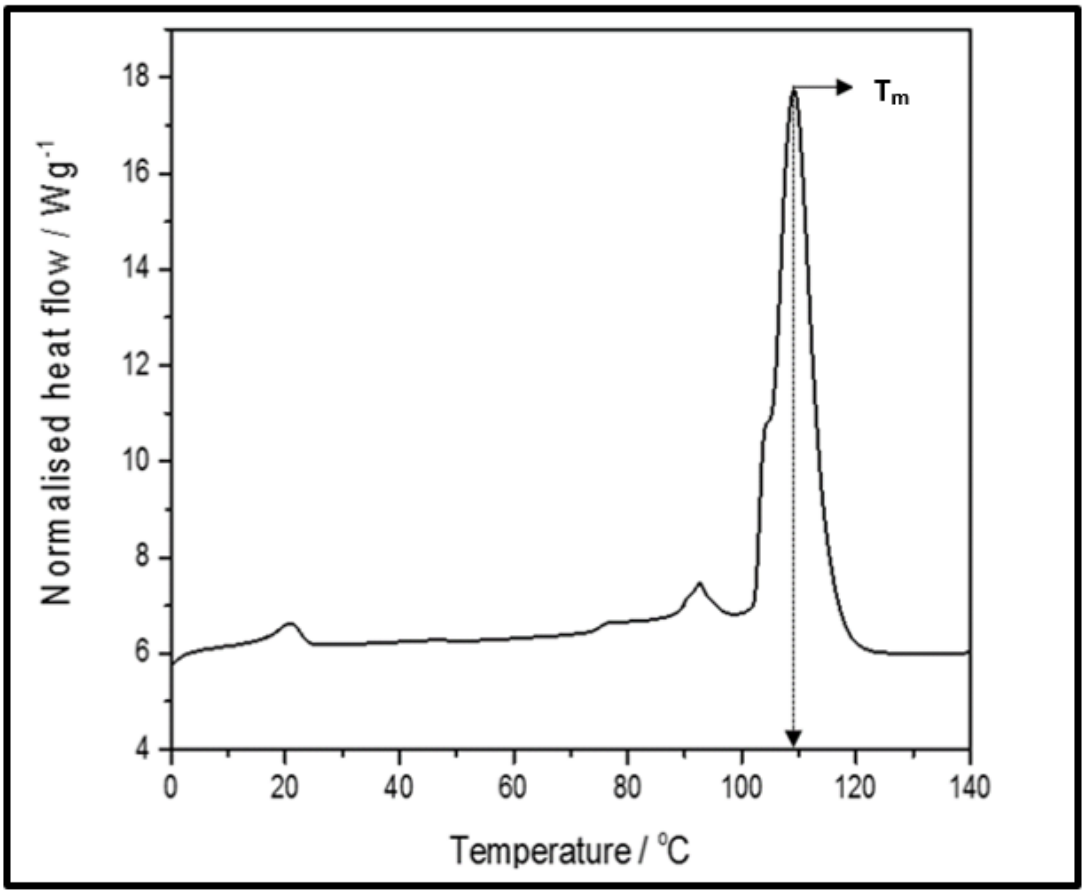


Fig. 5.3: A DSC thermogram of G4LZI3. The protein is shown to possess a transition temperature of approximately 110°C.

5.3.4 Determination of molecular mass of G4LZI3 with mass spectroscopy

Presented in Fig. 5.4 is a mass spectrogram of recombinant G4LZI3 generated using MALDI-TOF procedure. As evident in the spectrogram, the peaks generated does not correspond to the expected molecular weight as predicted by the *in-silico* analysis and most importantly as seen on the various SDS-PAGE analysis which consistently indicated a molecular weight of approximately 20.3 kDa.

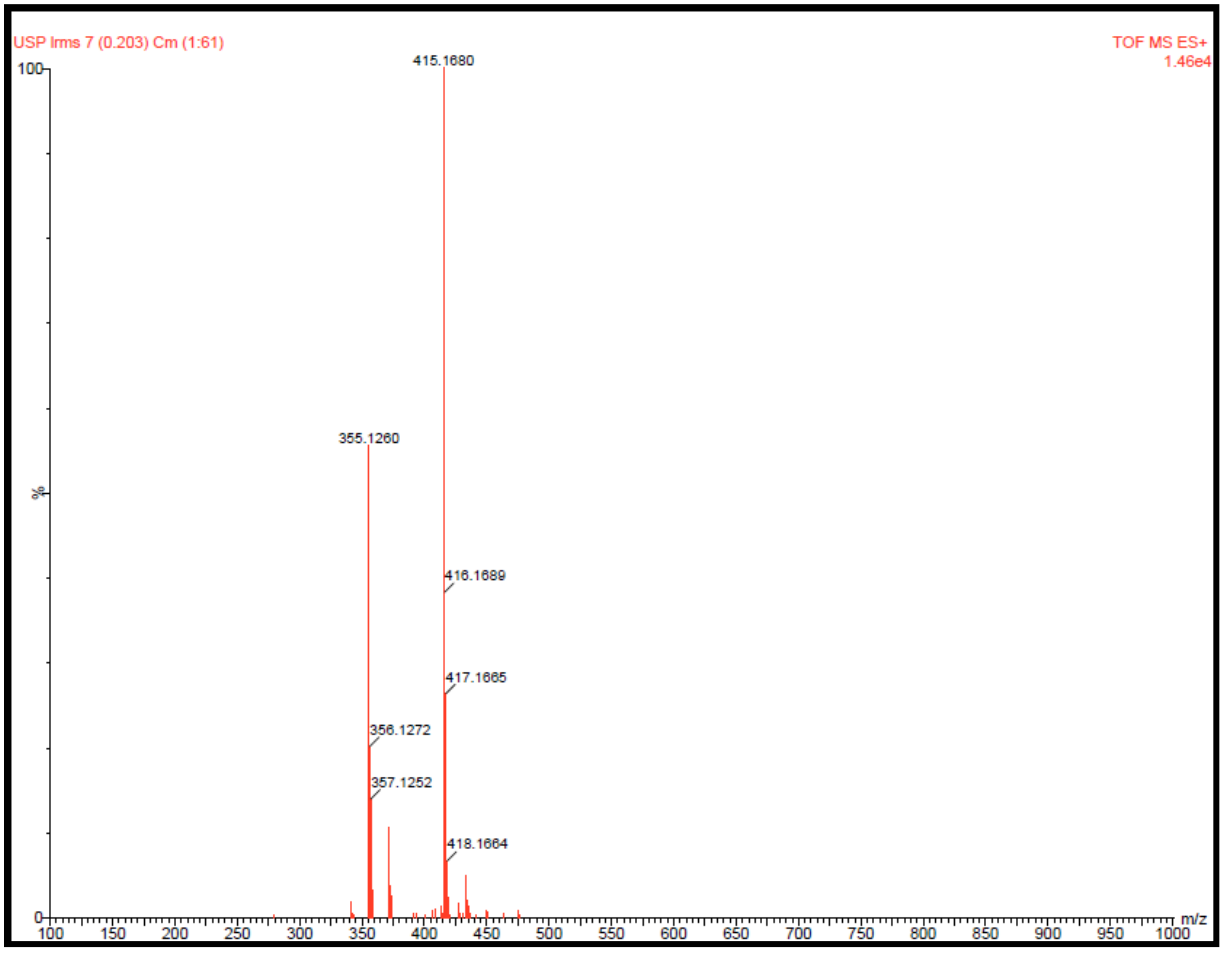


Fig. 5.4: Mass spectrogram of purified G4LZI3. The peaks not in consistency with the *in-silico* prediction from ProtParam and SDS-PAGE analysis.

5.3.5 Assessment of stability and foldedness of G4LZI3 with fluorescence spectroscopy

Fluorescence spectroscopy was done to assess the changes in stability of the protein by observing the intrinsic maximum fluorescence emission intensity of the fluorophores present in the protein. The result represented by the fluorescence spectrum in Fig. 5.5 show that excitation at the 270 nm wavelength recorded approximately 344 nm for the maximum fluorescence emission intensity. This observed emission spectrum is characteristic of the aromatic amino acid residues basically tryptophan, phenylalanine and tyrosine hence, an indication of the stability and folded state of the G4LZI3 protein.

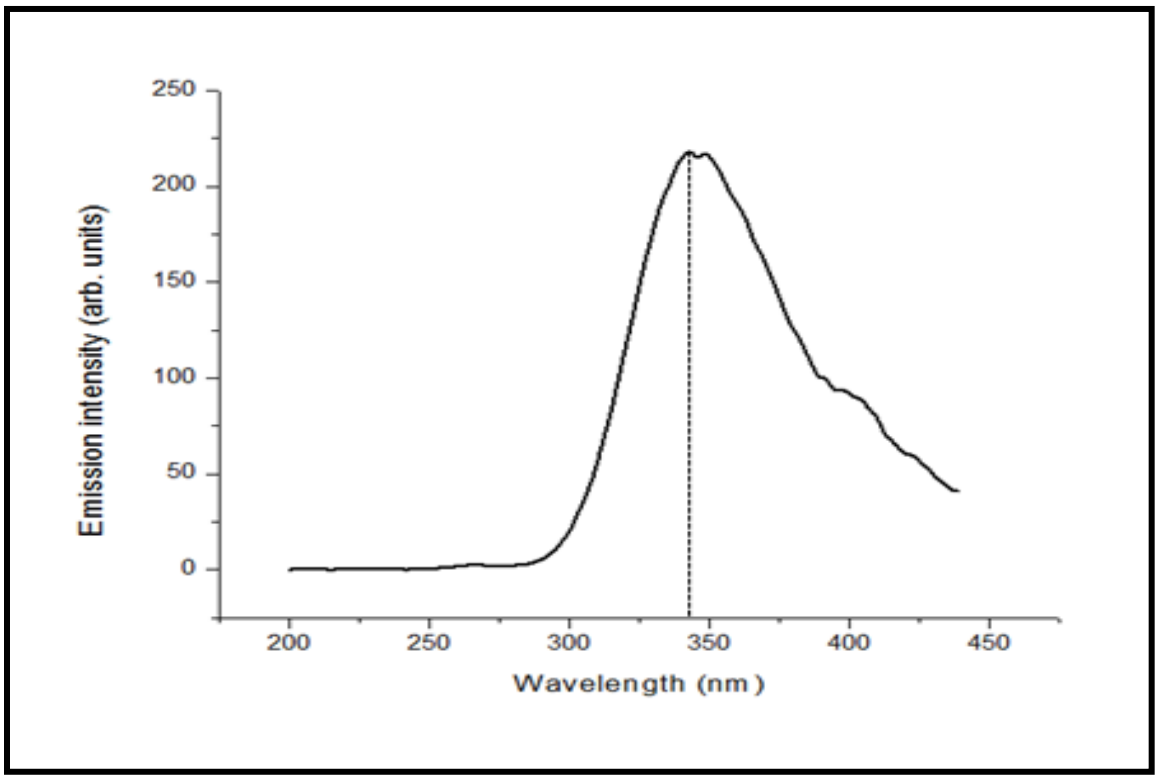


Fig. 5.5: Fluorescence spectroscopy spectra of G4LZI3. The absorbance at approximately 340 nm is due to the fluorophores in the protein.

5.4 Discussion

UV-Vis can be utilized as a sensitive method to assess subtle changes in protein structure as well as to determine the protein concentration and purity. In proteins, certain major chromophores are significant for UV/Vis spectroscopy these include peptide bonds (amide bond), specific amino acid side chains majorly tryptophan, cysteine and tyrosine. It must be noted that the absorption spectrum of a chromophore is partly determined by its chemical structure (Fornander *et al.,* 2014). The UV-visible spectrum of G4LZI3 as seen in Fig. 5.1 confirms the *in silico* prediction that there is a low abundance of aromatic amino acids in the protein; Trp (1 residue), Tyr (4 residues), cysteine (1 residue) and Phe (6 residues). The spectrum also confirms that the protein is free from nucleic acid which are common contaminants during protein expression. Similar pattern of absorption at 280 nm has been reported for other recombinant protein due to the aromatic residues in such proteins (Lv *et al.,* 2015).

FTIR spectroscopy gives information on the secondary structure content of proteins. The secondary structure stipulates consistent polypeptide chain folding arrangements of helices, sheets, coils, and turns that add up to form the tertiary structure. Different techniques are available for the elucidation of the tertiary structure of proteins including prediction centred on the sequence and physicochemical properties of amino acid residues and methods with better precision such as NMR spectroscopy and X-ray diffraction. Fourier transform infrared (FTIR) spectroscopy is a useful tool for characterizing of protein secondary structure, although the precision it offers is not as excellent as that of NMR and X-ray diffraction (Aboul-enein *et al.,* 2014).

FTIR spectroscopy mechanism involve the shining of infrared radiation on a protein sample and recording which wavelengths of radiation in the infrared region of the spectrum are absorbed by the sample. Every compound possesses a characteristic set of absorption bands in the infrared spectrum. Major distinctive bands seen in the infrared spectra of proteins as well as polypeptides are the Amide I and Amide II which arise from the amide bonds attached to the amino acids. The absorption related to the Amide I band causes stretching vibrations of the C=O bond of the amide while absorption linked with the Amide II band leads mainly to bending vibrations of the N-H bond (Gallager, 2009; Bunaciu *et al.,* 2016).

Worthy of note is the fact that FTIR can be used to analyse proteins of up to 150-200 kDa and that it is not influenced or limited by the nature of the environment. However, sensitivity reduces with excessively large protein size (Fabian and Mantele, 2002). Characterization of G4LZI3 with FTIR showed that the protein has more α-helix as component of its secondary structure. This result presented in Fig. 5.2 is in agreement with the secondary structure prediction which revealed that the protein is made up of good amount of helices. Previous studies reported the successful use of FTIR in characterizing the secondary structure elements of recombinant proteins (Gross-Selbeck *et al.,* 2007; Weisman *et al.,* 2010; Lv *et al.,* 2015). However, the percentage of each secondary structure elements present in the protein can further be ascertained with the use of circular dichroism in which circularly polarized light pass through a sample and the difference between absorption of the components of left and right is measured. Different structural elements possess distinctive CD spectra; alpha-helical proteins produce negative bands at 222 nm and 208 nm and a positive band at 193 nm

while proteins with antiparallel beta pleated sheets produce negative bands at 218 nm and positive bands at 195 nm. Disordered proteins produce extremely low ellipticity above 210 nm and negative bands near 195 nm (Greenfield, 2006; Corrêa and Ramos, 2009).

The structural stability of G4LZI3 was assessed with differential scanning calorimetry; the result presented in Fig. 5.3 showed that the protein is thermodynamically stable at temperatures below 110 °C. This result is in consonance with the *in-Silico* study which predicted the instability index to be 34.97; this index implies that the protein is relatively stable. Thermostability is a major factor to consider during the process of production of recombinant proteins for further down-stream applications. A stable and properly folded protein is an outright prerequisite for a successful biotherapeutic. Every biological process relies on proteins that are stable and in the properly folded conformation (Nurul and Azura 2012). The use of DSC in accessing the thermodynamic stability of recombinant proteins have been previously reported (Togashi *et al.,* 2002; Dedova *et al.,* 2004; Lepock, 2005). Mass spectroscopy is a widely-recognized technique employed in eliciting information as regards the molecular masses of diverse molecular species; specifically, it entails separation of molecules based on mass/size ratio of ionized species within an electric field. Mass spectroscopy is an essential analytical procedure for the characterization of recombinant proteins in the field of biotechnology. It involves the analysis of intact protein or peptides derived from enzymatic or chemical digestion using either LC-ESI-MS/MS (liquid chromatography-electrospray ionization tandem mass spectrometry) or MALDI-TOF-MS/MS (matrix-assisted laser desorption/ionization-time of flight tandem mass spectrometry (Gundry *et al*., 2009). It is

observed that the mass spectroscopy result depicted by Fig. 5.4 is not in consonance with the molecular mass of the protein predicted by computational analysis and the mass consistently seen on the SDS-PAGE analysis of the protein, the reason for this is not clear, it might be due to the unexpected proteolysis of the protein sample prior to the analysis or an undetected default in the algorithm used for data analysis, further study needs to be done to clear the ambiguity.

Fluoresence spectroscopy is an analytical method suitable for the analysis of the three-dimensional structure of proteins. The side chains of the aromatic amino acid residues namely; Phe, Tyr and Trp exhibit fluorescence when they are made to go through excitation. The fluorescence is commonly dominated by the Trp side chains due to the fact that it has stronger absorption as well as intensity of the emission than those of Phe or Tyr. There is usually a change in the fluorescence emission spectrum of a protein if there is a change in the structural conformation (Chelius *et al.,* 2010). Hence, the fluorescence spectrum of G4LZI3 as shown in Fig. 5.5 confirmed that the recombinant protein is well structured and will be valuable for further study towards the comprehensive characterization of the protein.

5.5 Conclusion

This chapter has successfully described the preliminary characterization of a novel recombinant universal stress protein, G4LZI3 from *Schistosoma mansoni*. Results from the different analytical procedures carried out has provided valuable information as regards the characteristics of the protein as per its functional group composition, secondary structure composition, foldedness, thermal stability and purity. These pieces information are necessary for further study on the novel protein towards the

determination of its mechanism of action in stress response as well as its utilization in disease intervention either as a drug target, in design of novel diagnostic tool or in vaccine formulation.

References

Aboul-Enein Y., Bunaciu A.A. and Fleschin S. (2014). Evaluation of the protein secondary structures using Fourier Transform Infrared Spectroscopy. *Gazi University Journal of Science*, **27**(1): 637-644.

Barnes S.C.A. and Lim A. (2007). Applications of mass spectrometry for the structural characterization of recombinant protein pharmaceuticals. *Mass Spectrometry Reviews*, **26**(3): 370-388.

Bertucci C., Pistolozzi M. De Simone A. (2011). Structural characterization of recombinant therapeutic proteins by circular dichroism. *Current Pharmaceutical Biotechnology*, **12**(10): 1508-1516.

Bunaciu A.A., Aboul-Enein H.Y. and Hoang V.D. (2016). Vibrational spectroscopy used in milk products analysis: A review. *Food Chemistry*, **196**: 877-884.

Cao S.X. and Zhao Y.F. (2004). Application of molecular absorption spectrophotometric method to the determination of biologic macromolecular structures. *Spectroscopic and Spectrometry Analysis* 24: 1197–1210.

Chelius D., Ruf P., Gruber P., Plöscher M., Liedtke R., Gansberger E., Hess J., Wasiliu M. and Lindhofer H. (2010). May. Structural and functional characterization of the trifunctional antibody catumaxomab. *MAbs:* Taylor and Francis. 2(3): 309-319.

Correa D.H. and Ramos C.H. (2009). The use of circular dichroism spectroscopy to study protein folding, form and function. *African Journal of Biochemistry Research*, **3**(5): 164-173.

Dedova I.V., Nikolaeva O.P., Mikhailova V.V., dos Remedios C.G. and Levitsky D.I. (2004). Two opposite effects of cofilin on the thermal unfolding of F-actin: a differential scanning calorimetric study. *Biophysical Chemistry*, **110**(1): 119-128.

Fabian H. and Mäntele W. (2002). Infrared spectroscopy of proteins. *Handbook of vibrational spectroscopy*.

Faraji F., Mofid M.R., Babaeipour V., Divsalar A. and Dehaghani S.A. (2010). The structural characterization of recombinant human granulocyte colony stimulating factor. *International Journal of Environmental Science and Development*, **1**(1): 15.

Fornander L.H., Feng B., Beke-Somfai T. and Nordén B. (2014). UV transition moments of tyrosine. *The Journal of Physical Chemistry B*, **118**(31): 9247-925.

Gallagher W. (2009). FTIR analysis of protein structure. *Course Manual Chem*, **455**.

Greenfield N.J. (2006). Using circular dichroism spectra to estimate protein secondary structure. *Nature Protocols*, **1(**6): 2876-2890.

Gross-Selbeck S., Margreiter G., Obinger C. and Bayer K. (2007). Fast Quantification of Recombinant Protein Inclusion Bodies within Intact Cells by FTIR Spectroscopy. *Biotechnology Progress*, **23**(3): 762-766.

Gundry R.L., White M.Y., Murray C.I., Kane L.A., Fu Q., Stanley B.A. and Eyk J.E.V. (2009). Preparation of Proteins and Peptides for Mass Spectrometry Analysis in a Bottom-Up Proteomics Workflow. *Current Protocol in Molecular Biology,* **88**: 1-23.

Howell J.M., Winstone T.L., Coorssen J.R. and Turner R.J. (2006). An evaluation of *in vitro* protein-protein interatction techniques: assessing contaminating background proteins. *Proteomics,* **6:** 2050-2069.

Lepock J.R. (2005). Measurement of protein stability and protein denaturation in cells using differential scanning calorimetry. *Methods*, **35**(2): 117-125.

Lv S., Gao J., Liu T., Zhu J., Xu J., Song L., Liang J. and Yu R. (2015). Purification and partial characterization of a new antitumor protein from *Tegillarca granosa*. *Marine Drugs*, **13**(3): 1466-1480.

Nurul A.I. and Azura A. (2012). Differential scanning calorimetry as tool in observing thermal and storage stability of recombinant bromelain. *International Food Research Journal* **19**(2): 727-731.

Togashi M., Kakinuma M., Nakaya M., Ooi T. and Watabe S. (2002). Differential scanning calorimetry and circular dichroism spectrometry of walleye pollack myosin and light meromyosin. *Journal of Agricultural and Food Chemistry*, **50**(17): 4803-4811.

Weisman S., Haritos V.S., Church J.S., Huson M.G., Mudie S.T., Rodgers A.J., Dumsday G.J. and Sutherland T.D. (2010). Honeybee silk: recombinant protein production, assembly and fiber spinning. *Biomaterials*, **31**(9): 2695-2700.

CHAPTER SIX

RESEARCH CONCLUSION AND FUTURE PERSPECTIVES

6.1 Conclusion

There is no gainsaying that much still need to be done in ameliorating the menace attributed to schistosomiasis, especially in the area of emergence of alternative treatment option aside praziquantel, formulation of vaccine and point-of-care diagnostic tools. The increase in the use of recombinant proteins as therapeutic agents necessitates the need to study the universal stress protein of *Schistosoma mansoni* which has been hypothesized as a promising candidate druggable target for disease intervention. Structural characterization of the G4LZI3 protein is an essential step towards actualizing the goal of investigating if it will be useful for schistosomiasis intervention.

Bioinformatics analyses of G4LZI3 showed that the 20.3 kDa protein is structurally flexible, relatively stable, has good interaction with water and other aqueous environment. The secondary structure elements of the protein might be responsible for its stability, hydrophobic nature as well as its stability. Multiple sequence alignment showed that this stress protein is highly conserved in several *Schistosoma* species as well as other organisms. Furthermore, gene ontology prediction confirmed previously stated studies that showed the universal stress protein family are generally implicated in stress response and signal transduction. Data from the *in silico* study provided good guide for the laboratory experimental procedures.

Novel recombinant protein G4LZI3 from *Schistosoma mansoni* was successfully cloned in pQE30 cloning vector and expressed in *E. coli* host system. The expressed protein was successfully purified to homogeneity.

The pure recombinant protein produced was completely free from aggregations and stable. This study showed that the protein expresses relatively well at various incubation conditions. However, a major shortcoming of this study is the failure to produce an isotopically labelled G4LZI3 within the limited timeframe of the research. Preliminary characterization of the purified G4LZI3 showed it to be folded and thus would be amenable to further structural characterization with solution structure NMR and X-ray crystallography. A major determinant for the use of recombinant protein as druggable target and vaccine target include it foldedness, stability, solubility and appropriate physicochemical features. The outcome of this study is an indication that G4LZI3 is likely to be a promising protein as drug and vaccine candidate.

6.2 Recommendation and future perspectives

It is recommended that further effort be made towards producing an isotopically labelled version of G4LZI3 in order to further structurally characterize the protein using heteronuclear Nuclear Magnetic Resonance (NMR) Spectroscopy. Additionally, it will be worthwhile to determine the 3D structure of the protein with X-ray crystallography in order to ascertain the computed 3D homology model of the protein. It is also pertinent that the precise secondary structural elements in the protein be ascertained with circular dichroism spectroscopy.

APPENDIX I: GENERAL CHEMICALS AND ENZYMES

Chemical	Company
40% 37.5:1 Acrylamide:bis-acrylamide solution	Sigma
Acetic acid	Merck
Agarose	Biorad
Ampicillin	Sigma
APS	Biorad
Bacteriological Agar	Merck
Bromophenol Blue	Sigma
BSA	Sigma
1-Butanol	Sigma
Calcium Chloride	Merck
Coomassie®protein assay reagent	Sigma
Disodium Hydrogen Phosphate Heptahydrate	Merck
Dithiothreitol (DTT)	Sigma
Ethanol	Merck
Ethidium Bromide	Biorad

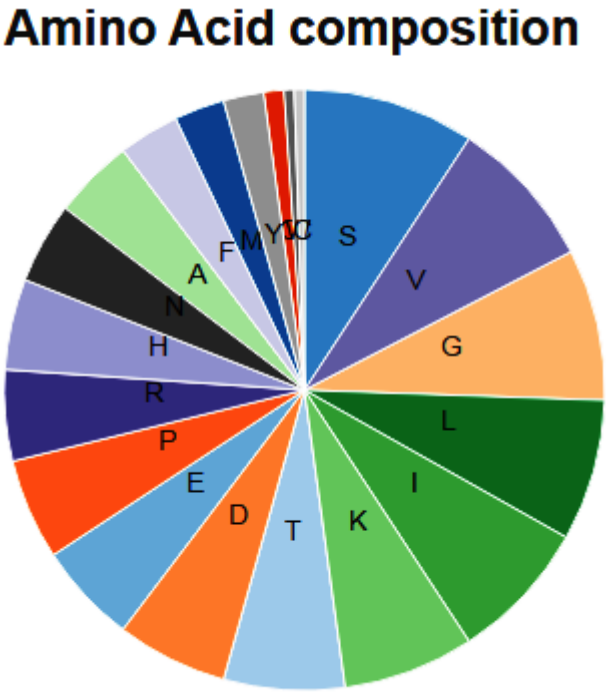
EDTA	Merck
Glacial Acetic Acid	Merck
Glucose	Merck
Glycerol	Sigma
Glycine	Merck
His-Select®Nickel Affinity Gel	Sigma
Hydrochloric Acid	Merck
IPTG	Roche
Imidazole	Merck
Isopropanol	Merck
Lysozyme	Sigma
Magnesium Chloride	Merck
Magnesium Sulphate	Merck
Methanol	Merck
PMSF	Sigma
Potassium Chloride	Merck
Potassium Dihydrogen Phosphate	Merck
Precision Plus Protein™ Dual Color Standards	Biorad

Thermo Scientific Prestained Protein Molecular Weight Marker	Thermo Scientific
Restriction enzymes	Thermo Scientific
Restriction enzyme buffer	Biorad
Sodium Azide	Sigma
Sodium Chloride	Merck
SDS	Sigma
Snakeskin™ pleated dialysis tubing	Thermo Scientific
Sodium Hydroxide	Merck
Sodium dihydrogen phosphate	Merk
TEMED	Sigma
Tris	Merck
Tryptone powder	Merck
Yeast extract	Merck

APPENDIX II: Supplementary *in-silico* analysis results

Evaluation of residues (RAMPAGE VALIDATION OF SWISS MODEL HOMOLOGY MODEL OF G4LZI3

```
Residue [A   9 :LEU] (-156.91,  15.76) in Allowed region
Residue [A  38 :GLU] ( -47.32, -64.37) in Allowed region
Residue [A  49 :PRO] (-101.38,  50.75) in Allowed region
Residue [A  54 :ARG] (-144.28,  29.51) in Allowed region
Residue [A  55 :ASP] (-156.89,  44.46) in Allowed region
Residue [A  63 :PRO] ( -88.67, 112.95) in Allowed region
Residue [A  28 :CYS] (-171.62, -43.16) in Outlier region
Residue [A  29 :ASN] ( 129.21,  -6.26) in Outlier region
Residue [A  50 :ASN] (-163.01, -51.08) in Outlier region
Number of residues in favoured region    (~98.0% expected)  :   66 ( 88.0%)
Number of residues in allowed region     ( ~2.0% expected)  :    6 (  8.0%)
Number of residues in outlier region                        :    3 (  4.0%)
```



Pie chart showing the amino acid composition of G4LZI3 generated from PredictProtein.

```
YASPIN PREDICTION RESULTS

Name: input
Length: 184

*Key
*   AA: Target sequence
* Pred: Predicted secondary structure (H=helix, E=strand, -=coil)
* Conf: Confidence (0=low, 9=high)
*Hconf: Confidence of helix  predictions
*Econf: Confidence of strand predictions
*Cconf: Confidence of coil   predictions

   AA: MSETEQPSTSDGLDIGETKGTISMTDATRKVLMPVDGSEHSERAFNWYMDNVMKITDGLY
 Pred: ----------------------------EEEEEE---HHHHHHHHHHHHHHHH---EEE
 Conf: 962234322468766853564011114455178652769437999999999825589608

Hconf: 000000000000000000000424210000000000099999999999999999000000
Econf: 000000000000000000000000000009999993000000000000000000000999
Cconf: 998999999999999999999465789990000006990000000000000000999000

   AA: LVHIVEPLSQGLNYNLASKSPSIKDDFSKHLNSLVESGRALRAKFFTRCEDSGLSARFTI
 Pred: EEEEEE-----------------HHHHHHHHHHHHHHHHHHHHHHHHHH----EEEEE
 Conf: 998725664333233334350202365788799999999999999998845076126788

Hconf: 000000000000000000072799999999999999999999999999700000000
Econf: 999999000000000000000000000000000000000000000000000000499999
Cconf: 000000999999999999992720000000000000000000000000000299500000

   AA: HVGTKPGENIVRIAHEHGVDLVIIGNRGIGTVKRTFLGSVSDYVLHHANVPVIIIPPPKH
 Pred: EE---HHHHHHHHHH----EEEEEE-----EEEEEE--HHHHHHHH----EEEE-----
 Conf: 85168638999999773882679323464400026621552899426763257860581 5

Hconf: 000009999999999900000000000007000000009999999900000000000
Econf: 995000000000000000399999800000799999400000000000023999970000
Cconf: 004990000000000099960000019999220000059000000009976000029999

   AA: PKKK
 Pred: ----
 Conf: 6339

Hconf: 0000
Econf: 0000
Cconf: 9999
```

EMBL-EBI Services Research Training About us

Sequence search results

Databases > Structure Databases > PDBsum

PDBsum

Sequence:

```
MSETEQPSTSDGLDIGETKGTISMTDATRKVLMPVDGSEHSERAFNWYMDNVMKITDGLYLVHIVEPLSQGLNYNLASKSP
SIKDDFSKHLNSLVESGRALRAKFFTRCEDSGLSARFTIHVGTKPGENIVRIAHEHGVDLVIIGNRGIGTVKRTFLGSVSD
YVLHHANVPVIIIPPPKHPKKK
```

Sequence length: 184 residues.

Your sequence search returned the following 59 hits:

	PDB code	Model	Length	%-tage identity	a.a. overlap	z-score	Ligands	Protein name
1.	1mjh(A)	X-ray 1.70Å	143	29.5%	149	233.0	ATP.	Structure-based assignment of the biochemical function of hypothetical protein mj0577: a test case of structural genomics
2.	3hqm(A)	X-ray 1.90Å	147	30.3%	152	224.3	ATP.	Universal stress protein tead from the trap transporter teaa halomonas elongata
3.	3s3t(A)	X-ray 1.90Å	145	24.5%	147	193.8	ATP, GOL, ACT.	Universal stress protein uspa from lactobacillus plantarum
4.	2z08(A)	X-ray 1.55Å	123	25.2%	147	171.6	ATP.	Crystal structure of uncharacterized conserved protein from thermus thermophilus hb8
5.	2z09(A)	X-ray 1.65Å	124	26.5%	147	171.5	ACP.	Crystal structure of uncharacterized conserved protein from thermus thermophilus hb8
6.	1wiq(A)	X-ray 2.10Å	135	26.7%	150	171.0		Crystal structure of a probable atp binding protein from the themophilus hb8
7.	2z3v(A)	X-ray 1.65Å	137	26.7%	150	170.9	PGO.	Crystal structure of uncharacterized conserved protein from thermophilus hb8
8.	3loq(A)	X-ray 2.32Å	273	26.4%	148	157.8	AMP, ACT.	The crystal structure of a universal stress protein from arc fulgidus dsm 4304
9.	1tq8(A)	X-ray 2.40Å	127	38.9%	54	154.3		Crystal structure of protein rv1636 from mycobacterium tuber h37rv
10.	3hqm(C)	X-ray 1.90Å	135	27.5%	149	153.9	ATP.	Universal stress protein tead from the trap transporter teaa halomonas elongata
11.	2dum(A)	X-ray 2.75Å	143	31.5%	149	148.6		Crystal structure of hypothetical protein, ph0823
12.	5ahw(C)	X-ray 2.15Å	146	23.5%	149	143.5	CMP, SO4, POG.	Crystal structure of universal stress protein msmeg_3811 in complex with camp
13.	3ab7(A)	X-ray 2.52Å	242	28.5%	130	142.7		Crystal structure of the hypothetical tandem-type universal protein ttha0350 from thermus thermophilus hb8
14.	2qm3(A)	X-ray 2.46Å	153	26.3%	137	142.0	AMP.	Crystal structure of an universal stress protein family prot arabidopsis thaliana at3g01520 with amp bound
15.	5ahw(A)	X-ray 2.15Å	125	35.2%	54	140.9	CMP, SO4, POG.	Crystal structure of universal stress protein msmeg_3811 in complex with camp
16.	2qm3(B)	X-ray 2.46Å	145	25.5%	137	131.4	AMP.	Crystal structure of an universal stress protein family prot arabidopsis thaliana at3g01520 with amp bound
17.	3ab8(A)	X-ray 1.70Å	261	23.5%	149	128.7	ATP.	Crystal structure of the hypothetical tandem-type universal protein ttha0350 complexed with atps.
18.	3dlo(A)	X-ray 1.97Å	134	35.8%	53	127.0	UNL, ACT.	Structure of universal stress protein from archaeoglobus ful
19.	3qtb(B)	X-ray 2.10Å	133	35.8%	53	127.0	D5M, ACT.	Structure of the universal stress protein from archaeoglobus in complex with damp
20.	3idf(A)	X-ray 2.00Å	138	25.5%	149	124.3	SO4.	The crystal structure of a usp-like protein from wolinella succinogenes to 2.0a
21.	2jax(A)	X-ray 3.22Å	261	27.4%	168	123.8	ATP, UNX.	Universal stress protein rv2623 from mycobaterium tuberculosis
22.	3dlo(D)	X-ray 1.97Å	135	37.7%	53	123.3	UNL, ACT.	Structure of universal stress protein from archaeoglobus ful
23.	3cis(A)	X-ray 2.90Å	289	25.6%	160	123.1	ATP.	The crystal structure of rv2623 from mycobacterium tuberculosis
24.	3so2(A)	X-ray 1.64Å	137	26.2%	84	118.3		Chlorella dutpase
25.	1k4y(A)	X-ray 2.50Å	501	24.8%	137	118.3	NAG-NAG, NAG-NAG-MAN-MAN-MAN, 4PN.	Crystal structure of rabbit liver carboxylesterase in comple piperidino-piperidine
26.	3fdx(A)	X-	127	24.3%	148	116.3	ATP, FMT.	Putative filament protein / universal stress protein f from pneumoniae.

		ray 1.58Å						
27.	3tnj(A)	X-ray 2.00Å	130	35.3%	51	113.7	AMP.	Crystal structure of universal stress protein from nitrosomo europaea with amp bound
28.	1kqf(B)	X-ray 1.60Å	289	32.8%	64	113.4	SF4, MGD, HEM, CDL.	Formate dehydrogenase n from e. Coli
29.	1kqg(B)	X-ray 2.80Å	289	32.8%	64	113.4	SF4, MGD, CDL, HEM, HQO.	Formate dehydrogenase n from e. Coli
30.	3zqe(A)	X-ray 2.49Å	916	52.2%	46	113.1	ASP, SO4, EDO.	Greater efficiency of photosynthetic carbon fixation due to single amino acid substitution
31.	4bxh(A)	X-ray 2.24Å	929	52.2%	46	113.0	EDO, SO4.	Resolving the activation site of positive regulators in plant phosphoenolpyruvate carboxylase
32.	4bxc(A)	X-ray 2.86Å	929	52.2%	46	113.0	G6P, SO4, EDO, TRS.	Resolving the activation site of positive regulators in plant phosphoenolpyruvate carboxylase
33.	1znv(B)	X-ray 2.00Å	119	27.7%	65	111.9	PO4.	How a his-metal finger endonuclease cole7 binds and cleaves transition metal ion cofactor
34.	4r2k(A)	X-ray 1.97Å	311	25.6%	125	111.7	SO4, EDO, OXD.	Crystal structure of h119a mutant of ydaa (universal stress from salmonella typhimurium
35.	4wt5(A)	X-ray 2.57Å	149	37.5%	56	111.6		ThE C-terminal domain of rubisco accumulation factor 1 from arabidopsis thaliana, crystal form ii
36.	3fh0(A)	X-ray 2.15Å	125	31.1%	74	111.5	ADP, EDO.	Crystal structure of putative universal stress protein kpn_0 atpase
37.	4l7p(A)	X-ray 2.30Å	350	32.9%	70	110.9	M95.	Human artd3 (parp3) - catalytic domain in complex with inhib me0395
38.	4qv2(A)	X-ray 1.80Å	354	32.9%	70	110.8	5ME, DMS.	Human artd3 (parp3) - catalytic domain in complex with inhib me0354
39.	4l7n(A)	X-ray 1.80Å	353	32.9%	70	110.8	1VB.	Human artd3 (parp3) - catalytic domain in complex with inhib sto1542
40.	4qv0(A)	X-ray 1.90Å	354	32.9%	70	110.8	8ME, DMS.	Human artd3 (parp3) - catalytic domain in complex with inhib me0355
41.	4l6z(A)	X-ray 2.00Å	354	32.9%	70	110.8	1DC, DMS.	Human artd3 (parp3) - catalytic domain in complex with inhib sto1168
42.	4l70(A)	X-ray 2.00Å	354	32.9%	70	110.8	1V9, DMS.	Human artd3 (parp3) - catalytic domain in complex with inhib me0352
43.	4l7o(A)	X-ray 2.00Å	354	32.9%	70	110.8	1VD, DMS.	Human artd3 (parp3) - catalytic domain in complex with inhib sto1542
44.	4l7l(A)	X-ray 2.10Å	353	32.9%	70	110.8	1VA, DMS.	Human artd3 (parp3) - catalytic domain in complex with inhib me0368
45.	4l7r(A)	X-ray 2.20Å	353	32.9%	70	110.8	M00.	Human artd3 (parp3) - catalytic domain in complex with inhib me0400
46.	3ce0(A)	X-ray 2.80Å	352	32.9%	70	110.8	P34.	Human poly(adp-ribose) polymerase 3, catalytic fragment in complex with an inhibitor pj34
47.	4l7u(A)	X-ray 2.80Å	352	32.9%	70	110.8	1VC.	Human artd3 (parp3) - catalytic domain in complex with inhib me0398
48.	4qv4(A)	X-ray 1.80Å	357	32.9%	70	110.7	MEJ, DMS.	Human artd3 (parp3) - catalytic domain in complex with inhib me0328
49.	3c4h(A)	X-ray 2.10Å	357	32.9%	70	110.7	DRL, DMS.	Human poly(adp-ribose) polymerase 3, catalytic fragment in c with an inhibitor dr2313
50.	2pa9(A)	X-ray 2.30Å	357	32.9%	70	110.7	GAB, GOL.	Human poly(adp-ribose) polymerase 3, catalytic fragment in complex with an inhibitor 3-aminobenzoic acid
51.	3fhb(A)	X-ray 2.30Å	357	32.9%	70	110.7	GAB.	Human poly(adp-ribose) polymerase 3, catalytic fragment in complex with an inhibitor 3-aminobenzoic acid
52.	3c49(A)	X-ray 2.80Å	357	32.9%	70	110.7	KU8.	Human poly(adp-ribose) polymerase 3, catalytic fragment in complex with an inhibitor ku0058948
53.	3zqb(A)	X-ray 2.71Å	920	47.8%	46	110.6	ASP, EDO, SO4.	Greater efficiency of photosynthetic carbon fixation due to single amino acid substitution
54.	4i0k(A)	X-ray 2.05Å	454	29.5%	61	110.4	EGR, PGE, PG4, PEG.	Tannin acyl hydrolase in complex with ethyl gallate
55.	5ty0(A)	X-ray 2.22Å	378	24.1%	108	110.4	BGC.	2.22 angstrom crystal structure of n-terminal fragment (resi 419) of elongation factor g from legionella pneumophila.
56.	1cau(B)	X-ray 2.30Å	184	20.0%	115	110.2		Determination of three crystal structures of canavalin by mo replacement
57.	1cav(B)	X-ray 2.60Å	184	20.0%	115	110.2		The three-dimensional structure of canavalin from jack bean ensiformis)
58.	1caw(B)	X-ray 2.60Å	184	20.0%	115	110.2		Determination of three crystal structures of canavalin by mo replacement
59.	1cax(B)	X-ray 2.60Å	184	20.0%	115	110.2		Determination of three crystal structures of canavalin by molecular replacement

Note: The above models are ordered by: 1. decreasing z-score, 2. X-ray structures (in order of decreasing resolution), NMR structures, and theoretical models, and 3. PDB code and chain id. Clicking on any PDB code takes you to the corresponding PDBsum page.

Note: The above models are ordered by: 1. decreasing z-score, 2. X-ray structures (in order of decreasing resolution), NMR structures, and theoretical models, and 3. PDB code and chain id. Clicking on any PDB code takes you to the corresponding PDBsum page.