

**Structural Determination of the novel Universal Stress  
Protein G4LZI3 from *Shistosoma mansoni* using X-ray  
Crystallography**

**Priscilla Masamba**

**201454125**



A thesis submitted in fulfillment of the requirements for the award of Philosophiae Doctor  
(Biochemistry) in the Department of Biochemistry and Microbiology, Faculty of Science and  
Agriculture, University of Zululand

**Supervisor: Prof. Abidemi Paul Kappo**

**April 2020**

**Structural determination of the novel Universal Stress Protein G4LZI3 from *Schistosoma manson*i using X-ray Crystallography**

**Priscilla Masamba**

April 2020

**Keywords:**

Bioinformatics

Crystals

G4LZI3

Mr\_Rosetta

*PHENIX*

Schistosomiasis

Structure

Universal Stress Protein (USP)

X-ray Crystallography

# ABSTRACT

---

## **Structural determination of the novel Universal Stress Protein G4LZI3 from *Schistosoma mansoni* using X-ray Crystallography**

Priscilla Masamba

PhD (Biochemistry) thesis, Department of Biochemistry and Microbiology, Faculty of Science and Agriculture, University of Zululand

*Schistosoma mansoni* is a parasitic multicellular helminth of the Schistosomatidae family and is one of the three main causative agents responsible for the multifactorial disease schistosomiasis, a debilitating and chronic vascular infection of paramount socio-economic importance and international public health concern. Recent estimates approximate the disease to account for the loss of 70 million disability-adjusted life years (DALYs) thus acknowledging schistosomiasis a more debilitating infection than malaria and a huge competitor with HIV/AIDS and TB. Mortality rates result in nothing less than 250 000 deaths per annum in sub-Saharan Africa alone where more than 90% of all schistosome infections stem from. Treatment for the disease has now become an issue of gross concern worldwide without the development of a vaccine as well as growing evidence of drug resistant strains against Praziquantel, which has been the backbone of treatment for all schistosome strains for over three decades. Furthermore, the elusiveness surrounding the mechanism of action of Praziquantel (PZQ), its lack of success in killing larval schistosomes 2-4 weeks post-infection as well as its failure to disrupt the transmission cycle of

the worm, has thereby caused wide interest in the design and development of alternative treatment against schistosomiasis.

It has been proposed that the cascade of events encompassing the transition of the worm throughout its developmental cycle, leading to the various changes in its morphological and chemical composition allowing the worm evade the various defense mechanism of humans, as well as its tolerance to sudden environmental changes and conditions of stress, is to a large extent due to the over-expression of heat shock proteins (HSPs) and universal stress proteins (USPs). This current study therefore focuses on the G4LZI3 USP whose over-expression has been displayed throughout all developmental forms of the parasite during conditions of stress and has therefore been identified as a ‘lead’ molecule in the design of alternative treatment against schistosomiasis. However, the mechanism of action of USPs is not well understood. Therefore, the main aim of this study was to determine the structure and infer biological function of this important druggable protein.

Before the structure was solved, a wide range of polyphenols, which have been proposed as potential inhibitors were docked onto the homology model of the protein and subjected to molecular dynamics (MD) simulations. Results revealed strong interaction between two polyphenols namely curcumin and catechin, suggesting the inhibitory potential of these polyphenols. The ability of polyphenols to scavenge oxidative stress associates these compounds with the prevention of diseases or infections that are linked to ROS production, which is one of the features of schistosomiasis. Hence, investigating polyphenols as potential nutraceuticals for schistosomiasis may be an interesting venture towards alternative treatment.

The structure of the G4LZI3 protein was thereafter solved using X-ray crystallography; the protein consists of 4 parallel  $\beta$ -sheets and 3  $\alpha$ -helices with the presence of a short  $3_{10}$  helix positioned close to the C-terminal of the protein. A generated Ramachandran plot showed that though there is still room for improvement, the solved structure is a relatively good model with 92% of the residues present in the favored and allowed regions. Inferred function points towards the protein binding ATP for regulation or even acting as an ATPase.

Finally, another set of molecular dynamic simulations were conducted on small molecule inhibitors docked onto the solved structure of the G4LZI3 protein using a pharmacophore generated from curcumin. Two of the molecules identified exhibited inhibitory activity against the protein by demonstrating future studies would therefore consider synthesizing these compounds for both *in vivo* and *in vitro* studies towards alternative anti-schistosomids.

# DECLARATION

I declare that “**Structural determination of the novel Universal Stress Protein G4LZI3 from *Schistosoma mansoni* using X-ray Crystallography**” is my own work and it has not been submitted for any degree or examination in any other university. All source used in this document have been indicated and acknowledged by complete references.

**Priscilla Masamba**

**April 2020**

Signed .....

# ACKNOWLEDGEMENTS

**To my Loving Family:** You're everything a family can be and more. What would I ever do without you? I am indebted to you forever and this thesis is a dedication to all of you. Thank you mum and daddy for always believing in me and allowing me to be the best I can ever be. For your countless support in all areas of my life throughout this journey is really hard to even comprehend. I have not always been the most brilliant child but you pushing me forward all the way, is what has shaped me today. Thank you for putting me in place and giving me space to shine. I am so proud to call you my parents and I hope that I have made you proud too.

Yaya Linda, Magalie and Sarah....you guys are amazing. I would never wish for another set of crazy 'soeurs' like you in the whole world. Thanks Yaya for the spontaneous long hour calls. Though far away, you are always close to my heart. Thank you Magz, for sharing this incredible journey with me. We have been stuck together, frustrated together and I pray that we will rejoice with our PhD certificates in our hands together. Sarah, I always have and always will label you as my personal hero. I look at your life and see God's goodness, which gives me strength to face another day because I know if you can do it, I definitely can too. Thank you for being you and for cross-checking the language on my thesis too. PS: Fallon, here we come!

**To Prof. Abidemi Paul Kappo:** I think pages and pages of words will not be enough to thank you for all you have done for me. You have gone beyond being my supervisor and greatest mentor to being a best friend, an honest voice and driving force in my life. Eight years ago when you first walked into my second year class I had no idea that our destinies will still be tied together the way they are. Thank you for taking a chance on me, even when I did not look like it. I have been to places and learnt so much from and through you. Thank you for making me part

of your family. I thank your beautiful and humble wife, Dr Christie Kappo and your two lovely kids, Daniel and Daniella, for embracing me as one of their own. Thank you Prof for mentoring me throughout this process, allowing me express my crazy self, never dampening my spirit and being a tremendous support system. I pray that wherever your children find themselves, they too shall also receive favor through the goodness you lent towards me.

**To the Aaron Klug Centre of Imaging and Analysis (UCT):** This project will not have been possible or completed without Prof Trevor Sewell. Thank you for allowing me to work in your lab and for making it your personal responsibility to hold my hand throughout the structure-solving process. I appreciate all your patience and diligence. Thank you so much Brandon for being the person you are. All your patience and advice even when it seemed as if I didn't know what I was doing saw me through some of my toughest times. I also would like to thank Kaylene, Phillip and Seun for all the help they rendered to me during all my stays in Cape Town.

**To Sahabne Ullah:** They say if you are friends for more than 7 years then you will be friends forever. You have been my friend for over 10 years. We don't see or talk to each other as often as we would like (ours is a long distance relationship lol) but we always hit it where we left off. It's been a friendship with no drama, jealousy or petty arguments but instead one of support, love, and encouragement in being the best people we can be. As a new chapter in my life begins, I thank God that after all this time you can be a part of it. Thank you for being there and being a friend indeed. I love you.

**To Jonathan Onyema Enyinnah (JOE):** Our friendship has been an interesting one because I never thought that I would learn so much from you. It is an undeniable fact that you became one of my very best friends on this part of the world. The wisdom that brews out of you amazes me

each time. Thank you for teaching me the non-academic side of life. Thank you for your constant encouragement and uplifting words. I really can't wait to see what the future holds for you because beyond reasonable doubt, you are going far. You're such a bright and lively young fellow and through this thesis I encourage you to go for your dreams and never give up. Current position does not mean forever destination. I always see you going higher and becoming a voice that can never be silenced. I hope this dedication will stand as a testament that we knew each other when the world knows your name one day.

**To Lloyd Mabonga:** You have been a voice of reason and a wonderful person with such bright ideas. Thank you for passing some of your wisdom onto me and I hope to see you on stage with me in our red gowns.

**To Dr Geraldene Munsamy:** Honestly, the time you showed up was just right. Thank you for all you input in this research work. Two whole chapters in this thesis would not have been if it were not for your tireless effort. Thank you.

**To the BSB Group:** Thank you for the good and bad times we shared, it has all shaped me to be the person I am now. I have learnt so much about life and teamwork through my interactions with everyone of you. I first acknowledge Kho and Makhosi who I have been with through this journey. Thank you for all the times we have shared and I wish you all the best in your future endeavors. I extend my gratitude to Ndibho, I remember the time you gave me some protein when I had run out and through that gift I was able to get my last set of crystals. Thank you so much for that as well as your constant smile and cheery spirit. I also say a big thanks to Amandla for being a lovely friend and always showing up when I need help with one thing or the other. To Lindi, Cuma, Lungani and Amanda, thanks for being such great colleagues; working with

everyone of you has been a pleasure. To Madam Boss (Sanele)...when I grow up I want to be like you! I know you have been through a lot in life but amongst it all you still remain standing. You are an inspiration to young ladies and are a testament to how strong women can be. I'm sure you are not aware but I salute and respect you so much.

**To the RCCG Christ Assembly Family:** Thank you all for the years we have spent together, before this journey began up until now. Special thanks to Pst Henock and his family, you always pull through for me whenever I'm stuck and I thank you for all your prayers over my life. Thank you to the Christ Assembly RCCG Choir for being my happy place. Str Lesego, Ole, Toby, Zinhle, Joy, Esther, Jacko, Happuck, Str Sphiwe, Str Buhle and Br Siphos....I love you all so much. The opportunity to minister with you guys was a blessing I will cherish for the rest of my days.

I extend my thanks to **Dr Chinaza Uleanya** for editing this document. God bless you sir.

I extend my thanks to both the **University of Zululand** and the **National Research Fund** for allowing me the opportunity to conduct my research at this institution as well as the financial support I received to pursue my studies.

Finally, I will always acknowledge the person who opened the door to me doing this degree and introducing me to my destiny. **Mr. Romanus Kwaye**, I wish you were here to see how much I've grown and how far I've come, but I know you're watching down from heaven and that you're proud of me. Thank you for that talk that we had which opened my eyes not to conform to the status quo. Thank you for going the extra mile to make sure that I made the best of a situation that looked dim. I will always remember you and may you continue to rest in perfect peace.

Last but never ever the least.....**To my God and Savior**.....I give You my thanks all my days.  
Thank you for being the driving force of my life and the wind underneath my wings. Thank you  
that my life and my future is in Your Hands and that this is just the beginning of new things for  
me. Merci Jésus!

# ABBREVIATIONS

|         |                                                 |
|---------|-------------------------------------------------|
| 1D      | One dimensional                                 |
| 2D      | Two dimensional                                 |
| 3D      | Three dimensional                               |
| AIDS    | Acquired Immunodeficiency Syndrome              |
| AMP     | Adenosine Monophosphate                         |
| APS     | Ammonium persulphate                            |
| ATP     | Adenosine Triphosphate                          |
| BAP     | Biotin Acceptor Peptide                         |
| BCS     | Biopharmaceutics Classification System          |
| cAMP    | Cyclic Adenosine Monophosphate                  |
| CEQ     | Catechol Estrogen Quinones                      |
| CF      | Cystic Fibrosis                                 |
| Coot    | Crystal Object-Oriented Toolkit                 |
| Cryo-EM | Cryo-electron Microscopy                        |
| CSBDD   | Computational Structure-Based Drug<br>Discovery |
| DALYs   | Disability-adjusted Life Years                  |

|                |                                              |
|----------------|----------------------------------------------|
| <i>DISTL</i>   | Diffraction Image Screening Tool and Library |
| DNA            | Deoxyribonucleic acid                        |
| DTT            | Dithiothreitol                               |
| <i>E. coli</i> | <i>Escherichia coli</i>                      |
| EDTA           | Ethylenediaminetetracetic acid               |
| Fig            | Figure/s                                     |
| FPLC           | Fast Protein Liquid Chromatography           |
| GAFF           | General Amber Force Field                    |
| GBD            | Global Burden of Disease                     |
| GDA            | Generic Data Acquisition                     |
| GSH            | Glutathione                                  |
| GTP            | Guanosine Triphosphate                       |
| GUI            | Graphical Interface User                     |
| His            | Histidine                                    |
| HIV            | Human Immunodeficiency Virus                 |
| HMM            | Hidden Markov Models                         |
| hr             | Hour/s                                       |
| HSPs           | Heat Shock Proteins                          |

|               |                                                             |
|---------------|-------------------------------------------------------------|
| HSQC          | Heternonuclear Single Quantum Coherence                     |
| HTS           | High Throughput Screening                                   |
| ID            | Identity Document                                           |
| IFN- $\gamma$ | Interferon gamma                                            |
| IgE           | Immunoglobulin E                                            |
| IL            | Interleukin                                                 |
| IMAC          | Immobilized Metal Affinity Chromatography                   |
| IPTG          | Isopropyl $\beta$ -D-thiogalactopyranoside                  |
| ISPyB         | Information System for Protein<br>Crystallography Beamlines |
| kDa           | Kilodalton                                                  |
| LB            | Lysogeny Broth                                              |
| LBVS          | Ligand-Based Virtual Screening                              |
| LPO           | Lipid Peroxidation                                          |
| MAD           | Multiple-Wavelength Anomalous Dispersion                    |
| MD            | Molecular Dynamics                                          |
| MDA           | Mass Drug Administration                                    |
| MGL           | Molecular Graphics Laboratory                               |

|         |                                                       |
|---------|-------------------------------------------------------|
| min     | Minute/s                                              |
| MIR     | Multiple Isomorphous Replacement                      |
| MM      | Molecular Mechanics                                   |
| MM/GBSA | Molecular Mechanics/Generalized Born<br>Surface Area  |
| MMP     | Matrix Metalloprotease                                |
| MM/PBSA | Molecular Mechanics Poisson-Boltzmann<br>Surface Area |
| MOPS    | 3-( <i>N</i> -Morpholino)propane sulphonic acid       |
| MR      | Molecular Replacement                                 |
| MSA     | Multiple Sequence Alignment                           |
| MWCO    | Molecular Weight Cut-Off                              |
| NCBI    | National Centre for Biotechnology Information         |
| Ni-NTA  | Nickel-nitrilotriacetic acid                          |
| NMR     | Nuclear Magnetic Resonance                            |
| NO      | Nitric Oxide                                          |
| NTDs    | Neglected Tropical Diseases                           |
| PAGE    | Polyacrylamide Gel Electrophoresis                    |

|               |                                                                   |
|---------------|-------------------------------------------------------------------|
| PDB           | Protein Data Bank                                                 |
| PDB           | Protein Data Bank                                                 |
| PEG           | Polyethylene glycol                                               |
| <i>PHENIX</i> | Python-based Hierarchical ENvironment for Integrated Xtallography |
| PMSF          | Phenylmethanesulphonyl Fluoride                                   |
| ProSA         | Protein-Structure Analysis Web                                    |
| PZQ           | Praziquantel                                                      |
| R&D           | Research and Development                                          |
| RAMPAGE       | Ramachandran Plots using the Richardsons' data                    |
| RMSD          | Root Mean Square Deviation                                        |
| RMSF          | Root Mean Square Fluctuation                                      |
| RNA           | Ribonucleic acid                                                  |
| RNS           | Reactive Nitrogen Species                                         |
| RoG           | Radius of Gyration                                                |
| ROS           | Reactive Oxygen Species                                           |
| RT-PCR        | Reverse Transcriptase-Polymerase Chain Reaction                   |

|              |                                        |
|--------------|----------------------------------------|
| SAD          | Single-Wavelength Anomalous Dispersion |
| SANS         | Small Angle Neutron Scattering         |
| SAS          | Small Angle Scattering                 |
| SASA         | Solvent Accessible Surface Area        |
| SAXS         | Small-Angle X-ray scattering           |
| SBDD         | Structure-Based Drug Discovery         |
| SDS          | Sodium Dodecyl Sulphate                |
| SEC          | Size Exclusion Chromatography          |
| SOD          | Superoxide Dismutase                   |
| SSA          | Sub-Saharan Africa                     |
| SWAP         | Schistosome Adult Worm Antigen         |
| TEM          | Transmission Electron Microscopy       |
| TFG- $\beta$ | Transformation Growth Factor $\beta$   |
| Th           | T helper                               |
| TNF          | Tumor Necrosis Factor                  |
| TSB          | Transformation Storage Buffer          |
| TSBG         | Transformation Storage Buffer Glucose  |
| USPs         | Universal Stress Proteins              |

|       |                                    |
|-------|------------------------------------|
| VD-HD | Vapor Diffusion Hanging Drop       |
| VD-SD | Vapor Diffusion Sitting Drop       |
| VEGF  | Vascular Endothelial Growth Factor |
| VLS   | Virtual Ligand Screening           |
| WASH  | Water, Sanitation and Hygiene      |
| WHO   | World Health Organization          |
| XRD   | X-ray Crystallography              |

# Table of Contents

|                                                                               |    |
|-------------------------------------------------------------------------------|----|
| ABSTRACT.....                                                                 | 2  |
| DECLARATION .....                                                             | 5  |
| ACKNOWLEDGEMENTS .....                                                        | 6  |
| ABBREVIATIONS .....                                                           | 11 |
| LIST OF TABLES .....                                                          | 23 |
| LIST OF FIGURES .....                                                         | 24 |
| CHAPTER 1: INTRODUCTION .....                                                 | 27 |
| 1.1 Overall Background of the Study.....                                      | 27 |
| 1.2 Problem Statement .....                                                   | 31 |
| 1.3 Aim.....                                                                  | 32 |
| 1.4 Objectives.....                                                           | 32 |
| 1.5 Chapter Delineation.....                                                  | 33 |
| CHAPTER 2: LITERATURE REVIEW .....                                            | 39 |
| 2.1 Introduction .....                                                        | 39 |
| 2.2 Neglected Tropical Diseases and Schistosomiasis.....                      | 40 |
| 2.3 Geographical Distribution of Schistosomiasis .....                        | 43 |
| 2.4 Lifecycle of Schistosomes.....                                            | 45 |
| 2.5 Signs, Symptoms and Immunobiology of Schistosomiasis .....                | 47 |
| 2.5.1 Acute and Chronic Schistosomiasis.....                                  | 48 |
| 2.5.2 Advanced stage of Schistosomiasis and other associated infections ..... | 49 |
| 2.6 Diagnostics, Therapies and Control .....                                  | 50 |
| 2.7 Praziquantel.....                                                         | 55 |
| 2.8 Universal Stress Proteins.....                                            | 58 |
| 2.9 Structure of Usps.....                                                    | 59 |
| 2.10 The Universal Stress Proteins of <i>E. coli</i> .....                    | 60 |
| 2.11 USPs in Intracellular Human Pathogens .....                              | 64 |
| 2.12 USPs and Schistosomes .....                                              | 65 |
| 2.13 Methods for 3D protein structure determination.....                      | 69 |
| 2.13.1 Cryogenic Electron Microscopy (Cryo-EM) .....                          | 69 |

|                                                                  |                                                                        |     |
|------------------------------------------------------------------|------------------------------------------------------------------------|-----|
| 2.13.2                                                           | Small-angle Scattering (SAS) .....                                     | 72  |
| 2.13.3                                                           | Nuclear Magnetic Resonance vs X-ray Crystallography .....              | 73  |
| 2.14                                                             | Protein production strategies .....                                    | 76  |
| 2.15                                                             | The Empirical Task of Protein Crystallization .....                    | 79  |
| 2.15.1                                                           | Vapor Diffusion .....                                                  | 82  |
| 2.15.2                                                           | Crystallization by Batch Diffusion .....                               | 85  |
| 2.15.3                                                           | Other Crystallization Methods .....                                    | 87  |
| 2.16                                                             | Principles of X-ray Crystallography .....                              | 90  |
| 2.16.1                                                           | What is a crystal? .....                                               | 90  |
| 2.16.2                                                           | X-ray Diffraction: Bragg's Law .....                                   | 93  |
| 2.16.3                                                           | The Phase Problem .....                                                | 95  |
| 2.16.4                                                           | Electron Density Maps .....                                            | 100 |
| 2.17                                                             | Structure-Based Drug Discovery .....                                   | 101 |
| 2.18                                                             | Computational Chemistry .....                                          | 102 |
| 2.19                                                             | Structure Prediction Methods .....                                     | 103 |
| 2.20                                                             | Protein Simulations .....                                              | 104 |
| 2.20.1                                                           | Molecular Mechanics .....                                              | 104 |
| 2.20.2                                                           | Molecular Dynamics .....                                               | 105 |
| CHAPTER 3: CYTOKINES, OXIDATIVE STRESS AND SCHISTOSOMIASIS ..... |                                                                        | 130 |
| 3.1                                                              | Introduction .....                                                     | 130 |
| 3.2                                                              | Immunology of the Acute and Chronic Stages of Schistosomiasis .....    | 132 |
| 3.2.1                                                            | Interleukin 4 .....                                                    | 136 |
| 3.2.2                                                            | Interleukin 10 .....                                                   | 137 |
| 3.2.3                                                            | IFN- $\gamma$ .....                                                    | 137 |
| 3.2.4                                                            | TNF- $\alpha$ .....                                                    | 138 |
| 3.3                                                              | Paradox of the Granuloma .....                                         | 139 |
| 3.4                                                              | ROS Production in Schistosomiasis .....                                | 141 |
| 3.5                                                              | Oxidative stress, Inflammatory Cytokines and Genetic Instability ..... | 143 |
| CHAPTER 4: MATERIALS AND METHODS .....                           |                                                                        | 155 |
| 4.3                                                              | Preparation of Competent <i>E. coli</i> cells .....                    | 155 |
| 4.3.1                                                            | Preparation of DH5 $\alpha$ <i>E. coli</i> cells .....                 | 155 |

|                                                                                                                                                                                    |     |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| 4.3.2 Preparation of JM109 <i>E. coli</i> cells.....                                                                                                                               | 155 |
| 4.3.3 Preparation of M15 <i>E. coli</i> cells .....                                                                                                                                | 156 |
| 4.4. Transformation of <i>E. coli</i> cells with pQE30-G4LZI3 Construct .....                                                                                                      | 156 |
| 4.5. Recombinant Production of the Universal Stress G4LZI3 Protein .....                                                                                                           | 157 |
| 4.5.1 Small Scale Expression Screen.....                                                                                                                                           | 157 |
| 4.5.2 Large Scale Protein Expression .....                                                                                                                                         | 158 |
| 4.5.3 Cell Extraction and Lysis .....                                                                                                                                              | 158 |
| 4.5.4 Affinity Purification of G4LZI3 USP.....                                                                                                                                     | 159 |
| 4.5.5 Purification of the G4LZI3 Protein by Size Exclusion Chromatography .....                                                                                                    | 160 |
| 4.5.6 Sodium dodecyl sulfate polyacrylamide gel (SDS-PAGE).....                                                                                                                    | 161 |
| 4.6 Determination of Protein Structure by X-ray Crystallography .....                                                                                                              | 162 |
| 4.6.1 Protein Crystallization Trials.....                                                                                                                                          | 162 |
| 4.6.2 Sample Preparation and Shipping.....                                                                                                                                         | 163 |
| 4.6.3 Crystal Mounting and Data collection.....                                                                                                                                    | 164 |
| 4.6.4 Structure Software and Data processing.....                                                                                                                                  | 165 |
| CHAPTER 5: CLOSING THE GAP: AN ATOMISTIC STRUCTURAL AND FUNCTIONAL<br>PERSPECTIVE OF <i>S. MANSONI</i> UNIVERSAL STRESS G4LZI3 PROTEIN IN COMPLEX<br>WITH PHENOLIC COMPOUNDS ..... | 170 |
| Introduction.....                                                                                                                                                                  | 171 |
| Materials and Methods.....                                                                                                                                                         | 173 |
| Homology Modeling of G4LZI3.....                                                                                                                                                   | 173 |
| Ligand Acquisition and Preparation.....                                                                                                                                            | 173 |
| Molecular Docking.....                                                                                                                                                             | 173 |
| Molecular Dynamic (MD) Simulations.....                                                                                                                                            | 178 |
| Post-Dynamic Analysis .....                                                                                                                                                        | 178 |
| Binding Free Energy Calculations .....                                                                                                                                             | 178 |
| Results and Discussion .....                                                                                                                                                       | 179 |
| Homology Modeling and Molecular Docking .....                                                                                                                                      | 179 |
| Assessment of system stability.....                                                                                                                                                | 180 |
| Protein structural flexibility .....                                                                                                                                               | 181 |
| Binding of polyphenols induce structural compactness.....                                                                                                                          | 183 |

|                                                                                                                                                              |     |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Hydrogen Bond Profile .....                                                                                                                                  | 183 |
| Binding free energy calculations .....                                                                                                                       | 184 |
| Per-residue energy decomposition .....                                                                                                                       | 185 |
| Conclusion .....                                                                                                                                             | 188 |
| References .....                                                                                                                                             | 188 |
| <b>CHAPTER 6: RECOMBINANT EXPRESSION, PURIFICATION AND STRUCTURE DETERMINATION OF THE UNIVERSAL STRESS G4LZI3 PROTEIN.....</b>                               |     |
| 6.1 Introduction .....                                                                                                                                       | 192 |
| 6.2 Recombinant Expression and Purification of pQE30-G4LZI3 Protein .....                                                                                    | 192 |
| 6.2.1 Transformation and Small Scale Expression Screen .....                                                                                                 | 192 |
| 6.2.2 Heterologous Over-expression of G4LZI3 Protein .....                                                                                                   | 193 |
| 6.2.3 Protein Purification by Affinity and Size-Exclusion Chromatography .....                                                                               | 196 |
| 6.3 Investigation of Universal Stress G4LZI3 Protein Structure by X-ray Crystallography ..                                                                   | 197 |
| 6.3.1 Protein Crystallization Trials of the G4LZI3 USP .....                                                                                                 | 197 |
| 6.3.2. Crystal Diffraction and Data Collection .....                                                                                                         | 206 |
| 6.3.3 Structure Determination and Refinement .....                                                                                                           | 211 |
| <b>CHAPTER 7: IDENTIFICATION OF POSSIBLE INHIBITORS AGAINST G4LZI3 PROTEIN USING PHARMACOPHORE MODELING TOWARDS THE DESIGN OF NEW SCHISTOSOMICIDES .....</b> |     |
| 7.1 Abstract .....                                                                                                                                           | 226 |
| 7.2 Introduction .....                                                                                                                                       | 227 |
| 7.3 Materials and Methods .....                                                                                                                              | 229 |
| 7.3.1 Molecular Docking .....                                                                                                                                | 229 |
| 7.3.2 Pharmacophore Model Generation and Screening .....                                                                                                     | 230 |
| 7.3.3 Molecular Dynamic Simulations .....                                                                                                                    | 230 |
| 7.4 Results and Discussion.....                                                                                                                              | 231 |
| 7.4.1 Pharmacophore Generation .....                                                                                                                         | 231 |
| 7.4.2 Molecular Dynamics and Post Molecular Dynamic Simulations .....                                                                                        | 237 |
| 7.5 Conclusion.....                                                                                                                                          | 241 |
| <b>CHAPTER 8: GENERAL DISCUSSION AND CONCLUSION .....</b>                                                                                                    |     |
| <b>APPENDIX I: General Stock Solutions, Buffers and Media .....</b>                                                                                          |     |
| <b>APPENDIX II: Bacterial Strains Used .....</b>                                                                                                             |     |

|                                                                                 |     |
|---------------------------------------------------------------------------------|-----|
| APPENDIX III: Morpheus Screen Constituents.....                                 | 257 |
| APPENDIX IV: I03 MX Beamline Specifications.....                                | 261 |
| APPENDIX V: Best 15 matches against G4LZI3 model derived from Dali Server ..... | 262 |
| APPENDIX IV: General Chemicals and Reagents .....                               | 263 |

# LIST OF TABLES

|                                                                                                       |     |
|-------------------------------------------------------------------------------------------------------|-----|
| Table 4.1: Summarized Buffer Composition of a 15% SDS PAGE gel.....                                   | 162 |
| Table 1: 3D Structures and Molecular Docking Scores of Polyphenol Compounds.....                      | 174 |
| Table 2: Residue fluctuation of all systems during the 100ns MD simulation measured in Angstroms..... | 182 |
| Table 3: MM/GBSA-based binding free energy profile analysis.....                                      | 185 |
| Table 6.1: Progression of Crystal Growth during Crystallization.....                                  | 204 |
| Table 6.2: Crystallographic Data Statistics.....                                                      | 209 |
| Table 6.3: Refinement Statistics.....                                                                 | 220 |

# LIST OF FIGURES

|                                                                                                                     |     |
|---------------------------------------------------------------------------------------------------------------------|-----|
| Fig. 2.1: Global Distribution of Schistosomiasis.....                                                               | 44  |
| Fig. 2.2: Lifecycle of the Schistosome Worm.....                                                                    | 46  |
| Fig. 2.3: The 6 <i>E. coli</i> USP proteins and their associated functions.....                                     | 63  |
| Fig. 2.4: Phylogenetic analysis of the 6 <i>E. coli</i> USP proteins.....                                           | 63  |
| Fig 2.5: Structure of the MJ0577 USP (PDB ID: 1MJH) from <i>Methanococcus jannaschii</i> .....                      | 68  |
| Fig. 2.6: Simplified Illustration of the Dubochet's Vitrification Method.....                                       | 71  |
| Fig. 2.7a: Structure Determination by X-ray crystallography.....                                                    | 75  |
| Fig. 2.7b: Structure Determination by NMR.....                                                                      | 75  |
| Fig 2.8: Protein Crystallization Phase Diagram.....                                                                 | 80  |
| Fig 2.9: Hanging and Sitting Drop Vapor Diffusion Crystallization Experiments.....                                  | 84  |
| Fig. 2.10: Batch and Microbatch Experiments.....                                                                    | 86  |
| Fig. 2.11: Crystallization Technique within a Single Vessel.....                                                    | 86  |
| Fig. 2.12: Crystallization by Dialysis.....                                                                         | 88  |
| Fig. 2.13: The Macro and Microseeding Crystallization Approach.....                                                 | 88  |
| Fig. 2.14: Representation of a unit cell to the crystal.....                                                        | 92  |
| Fig 2.15a: Constructive vs. Destructive Interference.....                                                           | 94  |
| Fig 2.15b: Visual of Bragg's Law.....                                                                               | 94  |
| Fig 4.1: Affinity Purification Strategy.....                                                                        | 160 |
| Fig. 4.2: Size Exclusion Purification Strategy.....                                                                 | 161 |
| Fig. 4.3: The mosquito liquid handling crystal Nanolitre Protein Crystallization Robot TTP<br>Labtech Ltd (UK)..... | 166 |
| Fig. 4.4: The LEICA Z16 APO microscope.....                                                                         | 166 |

|                                                                                                                       |     |
|-----------------------------------------------------------------------------------------------------------------------|-----|
| Fig. 4.5: X-ray generator (Rigaku MSC Micromax 007 dector.....                                                        | 167 |
| Fig. 4.6: IO3 MX beamline, Diamond Light Source Synchrotron.....                                                      | 168 |
| Figure 1. Diagrammatic representation of the sequence and homology model of the G4LZI3 protein.....                   | 180 |
| Figure 2. Graph showing Root mean square fluctuation of the C $\alpha$ backbone of the G4LZI3....                     | 181 |
| Figure 3.Root mean square fluctuation of G4LZI3.....                                                                  | 182 |
| Figure 4.Radius of gyration of G4LZI3 in complex with a range of selected polyphenol inhibitors.....                  | 183 |
| Figure 5. Stabilizing hydrogen bonds of the G4LZI3.....                                                               | 184 |
| Figure 6. Conformational changes of G4LZI3 in association with bound Curcumin and Catechin over 100ns simulation..... | 186 |
| Figure 7.Per-residue energy decomposition of G4LZI3.....                                                              | 187 |
| Fig. 6.1: Small scale expression profile of colonies transformed with pQE30-G4LZI3.....                               | 195 |
| Fig. 6.2: Large scale expression of the G4LZI3 USP in <i>E. coli</i> cells.....                                       | 198 |
| Fig. 6.3: Affinity chromatography of the G4LZI3 protein.....                                                          | 199 |
| Fig 6.4: Removal of residual proteins by size exclusion chromatography.....                                           | 200 |
| Fig 6.5: Gel filtration purification gels of the G4LZI3 protein.....                                                  | 201 |
| Fig. 6.6: Crystal Diffraction of Dataset 1.....                                                                       | 207 |
| Fig. 6.7: Crystal Diffraction of Dataset 2.....                                                                       | 208 |
| Fig. 6.8: Starting Solution obtained by Molecular Replacement.....                                                    | 213 |
| Fig. 6.9: Model obtained from AutoBuild.refine.....                                                                   | 214 |
| Fig. 6.10: Final Model of the G4LZI3 Structure.....                                                                   | 216 |

|                                                                                            |     |
|--------------------------------------------------------------------------------------------|-----|
| Fig. 6.11: Superimposition and Sequence Alignment of 2GM3 and the solved G4LZI3 model..... | 217 |
| Fig 6.12: Ramachandran Plot of the final G4LZI3 model.....                                 | 221 |
| Fig 7.1: Generated Pharmacophore model from ZINC PHARMER.....                              | 233 |
| Fig. 7.2: Model of G4LZI3 in complex with Inhibitor I.....                                 | 234 |
| Fig. 7.3: Model of G4LZI3 in complex with Inhibitor II.....                                | 235 |
| Fig. 7.4: Model of G4LZI3 in complex with Inhibitor III.....                               | 236 |
| Fig. 7.5: Graph showing Root mean square deviation of G4LZI3.....                          | 238 |
| Fig. 7.6: Graph showing Root mean square fluctuation of G4LZI3.....                        | 238 |
| Fig. 7.7: Radius of gyration profile of G4LZI3 .....                                       | 240 |
| Fig. 7.8: Hydrogen Bond Profile of G4LZI3.....                                             | 240 |

# CHAPTER 1: INTRODUCTION

## 1.1 Overall Background of the Study

Parasitic worms infect over a third of the world's population and schistosomes are not an exception. Among the many infectious diseases plaguing today's generation, schistosomiasis, according to the World Health Organization (WHO), is one of the most prevailing and persistent parasitic and ancient neglected tropical scourges of socioeconomic importance and public health significance (Tebeje *et al.*, 2016). Although it has been eradicated successfully in areas that make up most of the Caribbean and in regions such as: Morocco, Tunisia, Iran, Japan, Puerto Rico, Venezuela and Mauritius (Inobaya *et al.*, 2014). They further state that close to 42 of mostly tropical and subtropical countries of Africa still remain heavily condemned to schistosome servitude. More than 200 million people are affected by the disease, which accounts for the loss of 1.53 million disability-adjusted life years (DALYs) of which 85% come from sub-Saharan Africa, a continent that represents 13% of the world's total population (Sah *et al.*, 2015; Sacolo *et al.*, 2018). In the year 2011, at-risk populations in South America, Africa and Asia combined were estimated at 779 million, while an approximate of 207 million were infected with schistosomiasis (Trainor-Moss and Mutapi, 2016). Moreover, these figures are steadily escalating. Current re-estimates of schistosome-associated DALYs stands at 20 times higher than previous assumptions (Inobaya *et al.*, 2014). Also, it is estimated that one out of 30 people are infected with schistosomiasis worldwide (Thomassen Morgas *et al.*, 2010). In South Africa, an estimated 4 million people are assumed to be infected with high levels of the disease persistent in the KwaZulu-Natal region (Magaisa *et al.*, 2015).

Human schistosomiasis is principally caused by 5 species out of the total 25 known from the *Schistosoma* genus: *S. intercalatum* and *S. mekongi*, which are rare and under-recognized species that were discovered in 1934 in the rain forests of Central Africa and 1978 in Cambodia and the Laos People Democratic Republic of the Western Pacific respectively. *S. haematobium* was discovered in 1852 to be responsible for urinary schistosomiasis, which affects many parts of Africa, the Middle East and India. *S. japonicum* results in hepatic schistosomiasis that was discovered in 1904 and is localized in Asian countries including the Philippines, China and Indonesia. Lastly, *S. mansoni*, the causative agent of the intestinal form of the disease was discovered in 1907 with infection generally localized in Africa, the Caribbean and the Middle East, Venezuela and Brazil (Inobaya *et al.*, 2014; Sah *et al.*, 2015; Sacolo *et al.*, 2018; Miraj *et al.*, 2017). The magnitude and continuous transmission of schistosomiasis is, to a large extent, highly contingent on various socioeconomic factors that include inadequate access to safe water and substandard sanitation practices, hence, the disease is rampant in impoverished settings and constantly allied to poverty, especially in low-and middle-income endemic countries (King, 2010; Sokesi *et al.*, 2016). Transmission of the disease, which is considered water-borne, occurs during contact with cercariae-infested waters that have been infiltrated by intermediate snails infected by miracidia. These are released from parasitic eggs that have been expelled via human excreta.

Most global strategies set in place to combat schistosomiasis emphasize the use of multi-sectoral and integrated approaches that encourage water sanitation and hygiene (WASH) practices. In the same vein, repeated Mass Drug Administration (MDA) with Praziquantel (PZQ) has been adopted by the majority of all affected countries as the most effective, cheapest and readily accessible method of schistosomiasis control for over three decades, especially in the absence of

a vaccine (Chan *et al.*, 2013). Although high cure rates have been credited to this drug, schistosome resistance and high re-infection rates have also been observed (El Ridi and Tallima, 2013). With inadequate enlightenment of the mechanism of action of PZQ despite years of research, clinically relevant and alternative treatment for schistosomiasis is a fast approaching and unquestionable goal for the near future (Huang *et al.*, 2012).

The G4LZI3 Universal stress protein has been identified as an important druggable protein that is over-expressed during conditions of stress throughout all developmental forms of the schistosome parasite (Mbah *et al.*, 2013). It has been proposed that a cascade of events encompassing the transition of the worm throughout its developmental cycle and leading to the various changes in its morphological and chemical composition, allows the worm to evade the various defense tactics put in place by both the human and snail host. The worm's tolerance to sudden environmental changes and conditions of stress is, to a large extent, due to the over-expression of heat shock proteins (HSPs) and universal stress proteins (USPs) (Gobert *et al.*, 2015). During sudden environmental changes and conditions of stress such as exposure to toxic chemicals, extreme temperatures, drought and salinity, USPs are up-regulated and assist the worm to withstand and survive such during its complex developmental cycle (Isokpehi *et al.*, 2011). The cercariae and sporocysts lifecycle stages are exposed to toxic substances released by host snails, which result in oxidative and nitrosative stress (Negrão-Corrêa *et al.*, 2012). Additionally, defense mechanisms in the form of nitric oxide and hydrogen peroxide are produced by the human host to eliminate the schistosomulae larval stage (LoVerde, 1998). However, the over-production of USPs during each of these unique circumstances ensures the survival of the parasite. However, the mechanism of action of USPs is also still not well

understood. Thus, it is vital to understand how these proteins function, particularly because their presence in humans is nonexistent (Mbah *et al.*, 2013).

Information on proteins is key to understanding various processes and mechanisms in biology and ultimately of life. How bio-molecules operate is largely dependent on their conformation, the changes associated with various biological processes, the correct and native folding state of the protein as well as protein dynamics. Comprehending protein functional diversity requires determination of the three dimensional structure of the protein in utmost detail. Such information offers opportunities for structure-based drug design, hence, holds one of the keys in achieving this objective. Since the 1990s, drug design based on protein structures has been an important feature in the drug discovery process (Hillisch, 2008). Not only are novel drugs developed with speed and cost-effectiveness; potential drug molecules already in existence can be improved, the potency and selectivity of these compounds can be increased and because most drugs are actually proteins that play crucial roles in disease pathways, the mode of action can be identified and explained, thereby generating structural information that can be exploited for disease interventions (Brown and Shotton, 2015).

Modern advances in structural biology have in recent times seen the extensive use of nuclear magnetic resonance (NMR) spectroscopy, X-ray crystallography, cryo-electron microscopy (Cryo-EM) and small-angle X-ray scattering (SAXS) as deciphering methods for 3-D structure determination (Náray-Szabó and Perczel, 2014). The majority of most solved structures however, are attributed to either X-ray diffraction or NMR spectroscopy methods with each exhibiting their own advantages and disadvantages. Since the 1950's till date, over 80 000 X-ray solved protein structures, many of which are of biomedical importance, have been deposited in the Protein Data Bank (PDB), while structures solved by NMR lay claim to less than 10 000,

proving that the former is a convenient method for providing detailed information and structural clarification on proteins and other biological macromolecules and is therefore regarded as one of the most preferred methods for determining the structure of complexes (Pugh, 2005). For this project, the X-ray crystallography route was employed as the method of choice.

## 1.2 Problem Statement

After hookworm infections, schistosomiasis is ranked the second most common neglected Tropical Disease of Poverty in the sub-Saharan region, responsible for serious aggressive clinical deformities and mortality rates mostly among children, teenagers and women (Adenowo *et al.*, 2015). Despite its existence from antiquity, the disease is still endemic to more than 78 countries and is accountable for the deaths of 280 000 to 500 000 individuals per annum, taking first place as the “most prevalent pathology caused by worms worldwide” (da Silva *et al.*, 2017). Devoid of any vaccines for schistosomiasis, or any helminth infection, repeated mass drug administration with Praziquantel has been the most efficient and partially effective port of call for schistosomiasis control in most developing countries mostly due to its cost-effectiveness, ready accessibility, easy-to-use nature, minimal associated side effects and high cure rates (Chan *et al.*, 2013). However, its continued use for mass populations for over three decades has unfortunately resulted in documented reports of decreased drug efficacy and schistosome resistance, notwithstanding extremely high re-infection rates, which the drug fails to address (El Ridi and Tallima, 2013; LoVerde *et al.*, 2004). Additionally, its mechanism of action, regardless of the many years invested in research, still remains obscure (Aragon *et al.*, 2009). Hence, the design of alternative treatment regimen that can target specific stages during the developmental cycle of the schistosome is rapidly becoming a matter of urgency. This study centers on a group of proteins known as Universal Stress Proteins (USPs) that have been hypothesized as a possible

solution regarding this venture. The use of X-ray crystallography has, for several years, been used as a powerful and principal tool for experimentally determining and obtaining macromolecular structures. It is in turn a vital and imperative part of the structure-based drug discovery process which involves docking, assessment of the fit and evaluation of the various conformations of the formed complexes.

### **1.3 Aim**

The aim of this study therefore, was to determine the structure of the Universal Stress G4LZI3 Protein by X-ray crystallography and employ computational biophysics in the hit identification of small molecule inhibitors that can serve as putative anti-schistosomes.

### **1.4 Objectives**

The objectives of this study were:

- To model the G4LZI3 protein computationally using MODELLER
- To optimize a library of natural polyphenols using Avogadro Molecular Editor
- To perform molecular docking using MGL tools and AutoDock Vina
- To perform molecular dynamic simulations by ANTECHAMBER
- To determine the stability and conformational flexibility of the docked complexes using the AMBER 14 suit
- To calculate binding free energies using the Molecular Mechanics/Generalized Born Surface Area (MM/GBSA) method
- To recombinantly express sufficient amounts of soluble G4LZI3 protein
- To purify the G4LZI3 protein to homogeneity using Ni-NTA Affinity purification and Size Exclusion Chromatography

- To crystallize the protein using the vapor diffusion sitting drop technique
- To diffract formed crystals using the I03 MX beamline (Diamond Light Source, UK)
- To acquire data acquisition from the ISPyB database via the SynchWeb interface
- To construct an electron density map by molecular replacement using MR\_Rosetta
- To perform model building and structure refinement using PHENIX and Coot softwares.

## 1.5 Chapter Delineation

This first chapter lays the foundation of the project and highlights the main reasons why this particular research was undertaken. Chapter 2 delves into extensive details of the literature surrounding the research. The epidemiology of schistosomiasis, the various forms of treatment for the infection and the disease's possible association with USPs is reviewed, laying ground as to why the structure of these proteins could be the answer to structure-based drug design for schistosomiasis. Universal stress proteins are the biomolecules of interest in this study and their possible mechanism of action in schistosomes are thereby reviewed. Various methods for the acquisition of protein structures are then described laying great emphasis on x-ray crystallography, an invaluable tool in modern biology that reveals the many secrets held by proteins and nucleic acids. It does this by providing a unique understanding into protein function and clarifying the complexities surrounding the inner workings of cells as well as the many mechanisms that drive numerous life processes. The laws and concepts that drive x-ray crystallography are initiated by the subtopics "Protein Production Strategies" and "The Empirical Task of Protein Crystallization", which decipher the pros and cons associated with the technique. The chapter ends by briefly unpacking the intricacies surrounding structure-based drug design. Chapter 3 unpacks schistosome immunology and explains how oxidative stress plays a role in schistosome survival within the human host. Chapter 4 reports on the wetlab experimental

methods used in achieving the aim and objectives of this study. Chapters 5–7 present the results determined from the analyses of the binding of inhibitory natural compounds to the G4LZI3 protein through to the determination of its structure and finally to the generation of a pharmacophore for the G4LZI3 protein. Chapter 8 is a general discussion of all the results and concludes the thesis by displaying future prospects concerning this project. It should be noted with absolute joy and pride that the determined structure from this study serves as the first of all universal stress proteins to be solved in schistosomes. All supplementary data and results in the form of tables and diagrams, as well as the companies from where all the general chemicals and reagents used in this project were purchased are all displayed in Appendixes I to IV.

## References

- Adenowo AF, Oyinloye BE, Ogunyinka BI and Kappo AP (2015). Impact of human schistosomiasis in sub-Saharan Africa. *Brazilian Journal of Infectious Diseases* **19**(2): 196-205.
- Aragon AD, Imani RA, Blackburn VR, Cupit PM, Melman SD, Goronga T, Webb T, Loker ES and Cunningham C (2009). Towards an understanding of the mechanism of action of praziquantel. *Molecular and biochemical parasitology* **164**(1): 57-65.
- Brown DG and Shotton EJ (2015). Diamond: shedding light on structure-based drug discovery. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **373**(2036): 20140468.
- Chan JD, Zarowiecki M and Marchant JS (2013). Ca<sup>2+</sup> channels and praziquantel: a view from the free world. *Parasitology international* **62**(6): 619-628.
- Da Silva VBR, Campos BRKL, De Oliveira JF, Decout J-L and De Lima MDCA (2017). Medicinal chemistry of antischistosomal drugs: Praziquantel and oxamniquine. *Bioorganic & medicinal chemistry* **25**(13): 3259-3277.
- El Ridi RA and Tallima Ha-M (2013). Novel therapeutic and prevention approaches for schistosomiasis. *Journal of Advanced Research* **4**(5): 467-478.
- Gobert GN, Nawaratna SK, Harvie M, Ramm GA and Mcmanus DP (2015). An ex vivo model for studying hepatic schistosomiasis and the effect of released protein from dying eggs. *PLoS neglected tropical diseases* **9**(5): e0003760.
- Hillisch A (2008). Protein structure-based drug design: applications, limitations and future developments. *Chemistry Central Journal* **2**(S1): S15.

- Huang H-H, Rigouin C and Williams DL (2012). The redox biology of schistosome parasites and applications for drug development. *Current pharmaceutical design* **18**(24): 3595-3611.
- Inobaya MT, Olveda RM, Chau TN, Olveda DU and Ross AG (2014). Prevention and control of schistosomiasis: a current perspective. *Research and reports in tropical medicine* **2014**(5): 65.
- Isokpehi RD, Mahmud O, Mbah AN, Simmons SS, Avelar L, Rajnarayanan RV, Udensi UK, Ayensu WK, Cohly HH and Brown SD (2011). Developmental regulation of genes encoding universal stress proteins in *Schistosoma mansoni*. *Gene regulation and systems biology* **5**: GRSB. S7491.
- King CH (2010). Parasites and poverty: the case of schistosomiasis. *Acta tropica* **113**(2): 95-104.
- Loverde PT (1998). Do antioxidants play a role in schistosome host–parasite interactions? *Parasitology Today* **14**(7): 284-289.
- Loverde PT, Carvalho-Queiroz C and Cook R (2004). Vaccination with antioxidant enzymes confers protective immunity against challenge infection with *Schistosoma mansoni*. *Memórias do Instituto Oswaldo Cruz* **99**: 37-43.
- Magaisa K, Taylor M, Kjetland EF and Naidoo PJ (2015). A review of the control of schistosomiasis in South Africa. *South African Journal of Science* **111**(11-12): 1-6.
- Mbah AN, Mahmud O, Awofolu OR and Isokpehi RD (2013). Inferences on the biochemical and environmental regulation of universal stress proteins from Schistosomiasis parasites. *Advances and applications in bioinformatics and chemistry: AABC* **6**: 15.

- Miraj CG, Fan W-H and Feng B (2017). Current Management Progress and Future Direction of Schistosomiasis: A Review. *American Journal of Biology and Life Sciences* **5**(5): 39-46.
- Náray-Szabó G and Perczel A (2014). Protein structure and dynamics. *International Journal of Terraspace Science and Engineering* **6**(1): 7-16.
- Negrão-Corrêa D, Mattos A, Pereira C, Martins-Souza R and Coelho P (2012). Interaction of *Schistosoma mansoni* sporocysts and hemocytes of *Biomphalaria*. *Journal of parasitology research* **2012**.
- Pugh D (2005). Biomolecular NMR spectroscopy in South Africa: the first five years: NRF/Royal Society programme. *South African Journal of Science* **101**(9-10): 421-429.
- Sacolo H, Chimbari M and Kalinda C (2018). Knowledge, attitudes and practices on Schistosomiasis in sub-Saharan Africa: a systematic review. *BMC infectious diseases* **18**(1): 46.
- Sah VK, Wang L, Min X, Rizal R, Feng Z, Ke Z, Deng M, Li L and Li H (2015). Human schistosomiasis: a diagnostic imaging focused review of a neglected disease. *Radiology of Infectious Diseases* **2**(3): 150-157.
- Sokesi T, Malama K, Masaninga F, Vwalika B, Kachimba J, Mufunda J and Songolo P (2016). Lessons learnt from the implementation of mass drug Administration for Schistosomiasis and Soil-Transmitted Helminths in Lusaka Province, Zambia. *Medical Journal of Zambia* **43**(2): 82-87.
- Tebeje BM, Harvie M, You H, Loukas A and Mcmanus DP (2016). Schistosomiasis vaccines: where do we stand? *Parasites & vectors* **9**(1): 528.

Thomassen Morgas DE, Kvalsvig JD, Taylor M, Gundersen S and Kjetland EF (2010). Schistosomiasis and water-related practices in school girls in rural KwaZulu-Natal, South Africa. *Southern African Journal of Epidemiology and Infection* **25**(4): 30-33.

Trainor-Moss S and Mutapi F (2016). Schistosomiasis therapeutics: whats in the pipeline? *Expert Review of Clinical Pharmacology* **9**(2): 157-160.

# CHAPTER 2: LITERATURE REVIEW

## 2.1 Introduction

‘Proteins’, a word which emanates from the Greek word ‘protos’ meaning ‘first’, are ubiquitous organic bio-molecules or compounds that occur in all organisms, which perform an extensive spectra of biological activities. Proteins are thermodynamically and kinetically responsible for all biochemical processes that occur and are therefore liable for coordinating and regulating the birth, growth, maturation, division, reproduction, nutrition, excretion and in due course, the death of all organisms (Vekilov, 2016). DNA and RNA molecules are heavily dependent on proteins for the replication and transfer of genetic material required for species survival. Moreover, their shapes and morphological constructions dictate numerous metabolic functions, which may lead to a diseased state if altered (Vasudevan *et al.*, 2013). X-ray crystallography provides insight, not only into the composition of biomolecules, but also into their various interactions with other compounds and their roles in biological mechanisms. This technique therefore forms the chief cornerstone of all molecular biology studies and provides an imperative foundation for rationale drug design and development (Ennifar, 2013). However, a fair amount of experimental techniques are available for determining the atomic structures of macromolecules, which is imperative in comprehending the biological and molecular mechanisms of infections and diseases (Adiyaman and McGuffin, 2019). The following section gives a background on schistosomiasis, the disease of interest in this study and one which has and continues to claim the lives of thousands of people every year while leaving millions disabled and incapable of normal living. This study hopes to contribute to current efforts focused on abolishing this neglected tropical disease of poverty.

## 2.2 Neglected Tropical Diseases and Schistosomiasis

The Neglected Tropical Diseases (NTDs) are an infection and parasite-causing group of illnesses responsible for 90% of the world's global health disease burden (Remme *et al.*, 2002). Setting them apart from other groups or categories of sicknesses is the sad fact that they are “neglected” diseases, as their term suggests, affecting millions of poor and vulnerable communities. Regrettably, their unfortunate plights are hardly taken into recognition and are basically ignored especially by the pharmaceutical sector, mainly owing to little turnover for profit in spite of various investments that have been poured into NTD research. Exacerbating this challenge is the very low priority that endemic countries themselves award NTDs (Morel, 2003). In 1975, the UNDP/World Bank/WHO Special Program for Research and Training in Tropical Disease (commonly referred to as TDR) was forged as an approach for analyzing research opportunities and needs for advancing NTD research for affected countries (Remme *et al.*, 2002). This has however done little to stop the global spread of these infections. Development and industrialization has always accompanied globalization and as the world is slowly tending towards a global village, the emergence and re-emergence of NTDs has now more than ever been encouraged. Man-made factors that include demographics, health care accessibility and availability, food production, migration and travel as well as susceptibility to infection, ecological and environmental changes have all contributed to the up-trajectory of these infections (Mackey and Liang, 2012). Diseases are accompanied by perpetual poverty and NTDs are a perfect example of such. Negative pregnancy outcomes and agricultural productivity, poor development and cognitive impairment in children, malnutrition, blindness, physical disability and disfigurement as well as social stigma and loss of status all occur as a direct result of NTDs, making it a significant reason why “the bottom 500 million” of SSA – the poorest one sixth of

the world's population – will continue to remain in dire poverty (Adenowo *et al.*, 2015; Feasey *et al.*, 2009; Mitra and Mawson, 2017; Wainwright *et al.*, 2019).

Over 1 billion people worldwide are afflicted by NTDs, which roughly equates up to 20% of the global population, the bulk of which are concentrated in sub-Saharan Africa (Addisu *et al.*, 2019; Knaul *et al.*, 2019). In the region, up to 51% and 73% of the population survive on less than US\$1.25 and US\$2 per day respectively (Mitra and Mawson, 2017). The term “Neglected Tropical Diseases” was created and introduced into literature in the year 2003 by Hotez and co-workers in efforts to spark political interest, increase donor funding and shift attention towards NTD research and development (R&D) (Addisu *et al.*, 2019). However, the current focus of R&D is largely committed toward the “big three” that is, malaria, HIV/AIDS and tuberculosis. A hefty 57% of NTDs on the other hand receive no attention. It was established in the year 2015 by the WHO that US\$18 billion (~US\$3.6 billion yearly) would be required for NTD elimination by the year 2020, however, less than US\$200 million per year has been allocated towards this goal; a 94% yearly shortfall (Addisu *et al.*, 2019). Thus, a blueprint for the elimination and control of the seven most prevalent NTDs (hookworm infections, schistosomiasis, trichuriasis, lymphatic filariasis, ascariasis, onchocercariasis and trachoma) known as the Global Network for Neglected Tropical Diseases, serves to deliver key sustainable strategies to address disease burdens in some of the most poor and affected countries in the world (Ebrahim, 2009). Vector control, mass drug administration and intense health care education all fall under part of their agenda. However, drug resistance has and continues to be an inevitable risk in regards to mass drug administration. Hence, research is still needed, not only to assess current strategies also to develop new ones.

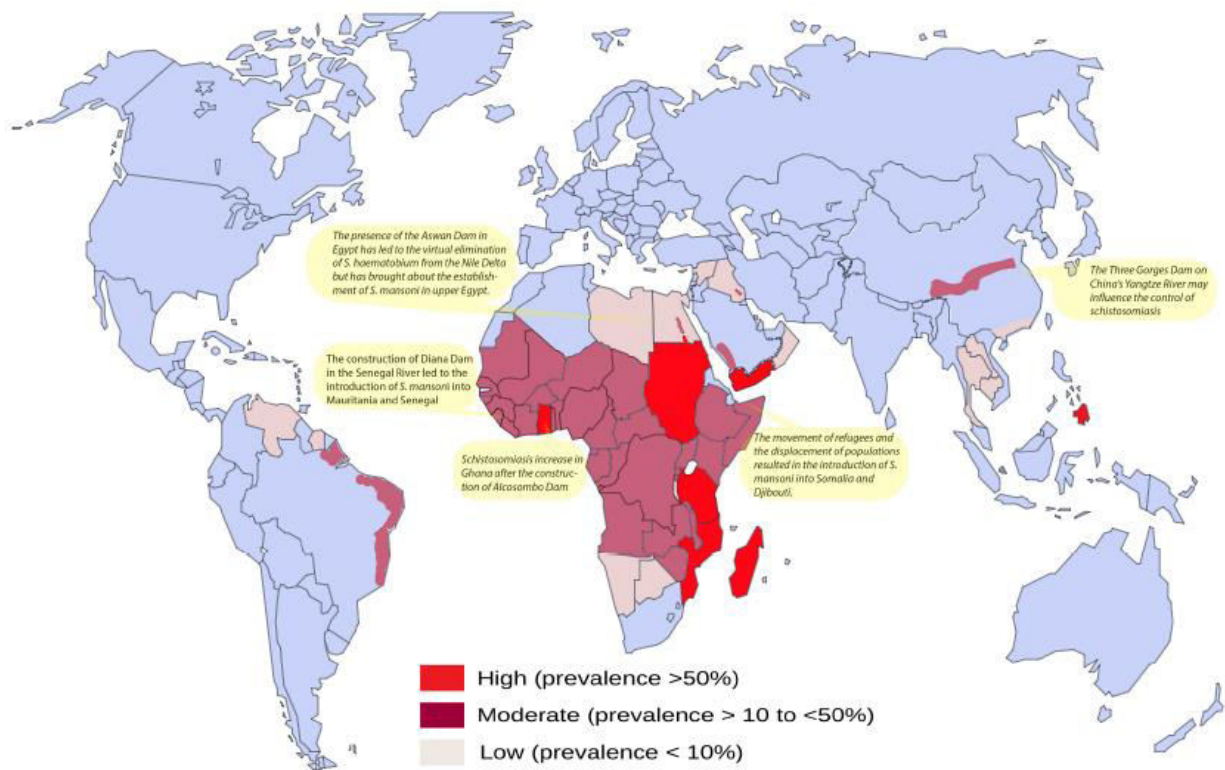
Helminth infections – lymphatic filariasis, soil-transmitted helminthiasis and schistosomiasis to be precise – are the most prevalent of the NTDs. These three account for up to 78% and 87% of

the annual mortality and global disease burden caused by NTDs respectively in terms of disability-adjusted life years (DALYs) (Lammie *et al.*, 2006; Merrifield *et al.*, 2016). Schistosomiasis has been referred to as a “silent killer” and is still regarded as the most important water-based disease from a global health standpoint (Utzinger *et al.*, 2011; Woelfle *et al.*, 2011). After malaria, schistosomiasis is the most potent and death-related disease caused by a parasite in tropical and subtropical regions of the world, infecting more than 252million people with 90% resident in SSA (Merrifield *et al.*, 2016; Thomas *et al.*, 2015). More than 250 000 deaths are attributed to schistosomiasis and although this may not be in millions, the disease is associated with extensive morbidity (Chitsulo *et al.*, 2000). An estimated 120 million out of those infected are considered asymptomatic, while 20 million display severe cases of clinical disease. In the year 2010, the Global Burden of Disease (GBD2010) estimated that no less than 3.3million DALYs were lost to schistosomiasis alone, while other studies have suggested that this may be a gross underestimation due to the fact that some factors were not taken into consideration such as pain and inflammation, malnutrition and chronic morbidities (Merrifield *et al.*, 2016). This is a hefty rise from the 1.7 million DALYs lost which was estimated by the GBD in the year 2001, 84% of which were lost from SSA (Sibomana, 2010). Although the disease has been in existence from antiquity and still continues to burden many regions in Africa, countries such as Japan and Tunisia have managed to successfully eradicate it while the Caribbean Islands and Morocco are moving closer to the elimination goal. Additionally, China, Egypt and Brazil have made extensive developments in the current control of schistosomiasis (Utzinger *et al.*, 2009). However, the disease is still a relevant public health issue that plagues 76 countries of the world with continents such as Africa included, Asia and the South and Central Americas (Fig 2.1) (de Oliveira *et al.*, 2013). Unfortunately, the disease has now seen its spread in areas where it was

not present before. The urinary form of the disease, whose incidence in Europe was last recorded in Portugal during the 1950s, has recently been detected in the French island Corsica, Spain and Portugal by the presence of the worm's intermediate hosts (*Bulinus contortus*, *B. truncatus* and *Planorbarius metidjensis*) (Berry *et al.*, 2014). Evidence suggests that transmission of the disease occurred via the Cava River southeast of Corsica. In addition to this, other schistosomiasis outbreaks have been recorded in Germany and Italy with the belief that the disease may have been spread as early as 2011 but was only detected in May 2014 (Boissier *et al.*, 2015). This has raised the alarm on potential risk factors that were previously ignored that may contribute to the spread of schistosomiasis.

## 2.3 Geographical Distribution of Schistosomiasis

Parasitic blood-dwelling trematodes belonging to the Phylum *Platyhelminthes* cause schistosomiasis and three species are largely responsible for the disease: *Schistosoma mansoni* and *Schistosoma japonicum*, which cause hepato-intestinal disease and *Schistosoma haematobium*, which causes the urogenital form of the illness (Clerinx and Van Gompel, 2011). Tropical and sub-tropical regions of sub-Saharan Africa, the Caribbean, South America and the Middle East are mostly affected by *S. mansoni* infection, while *S. japonicum* is mainly housed by Asian countries. *S. haematobium* infections occur in parts of North and sub-Saharan Africa, India and the Middle East (Fig. 2.1) (Elbaz and Esmat, 2013). *S. mekongi* and *S. intercalatum* also parasitize humans but are rare and less prevalent species that are sparsely distributed in the Western Pacific and also endemic to Central Africa, respectively (Nour, 2010).

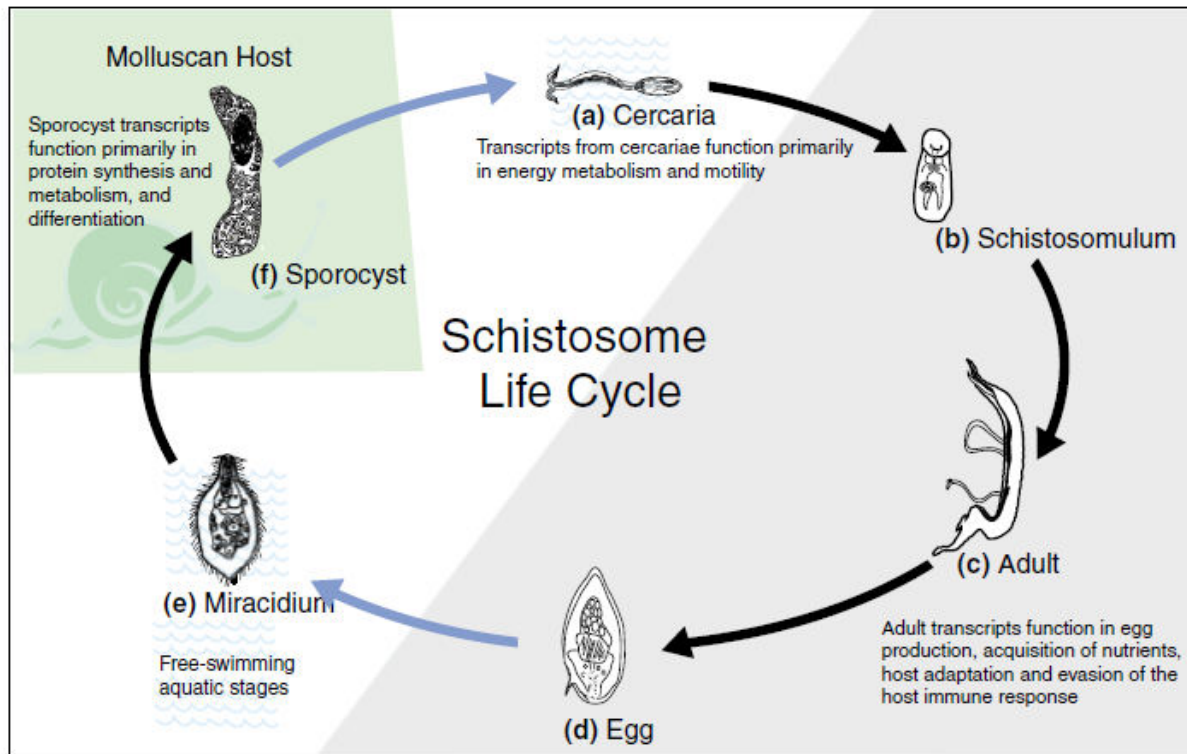


**Fig 2.1: Global Distribution of Schistosomiasis.** Over 252million people in over 76 countries are infected with the disease with the largest burden present in SSA. The yellow bubbles indicate examples of incidences that have led to the rise in schistosomiasis in the highlighted regions (Butrous, 2019)

Schistosomiasis is endemic to South Africa with an estimated 2.5 million infected individuals (Appleton and Kvalsvig, 2006; Magaisa *et al.*, 2015). High levels of infection come from the Pietermaritzburg area and districts such as uMkhanyakude and Ugu of the KwaZulu-Natal province and areas of the Eastern Cape such as Port St Johns and Uitenhage (Magaisa *et al.*, 2015; Johnson and Appleton, 2005; Kabuyaya *et al.*, 2017). Rural-urban migration, the growth of informal settlements and crowding, coupled with poor hygiene standards consequentially increase schistosomiasis transmission in these areas. *S. haematobium* is more prevalent in these areas resulting in various complications such as **Female Genital Schistosomiasis (FGS)**, which in turn predisposes females to HIV/AIDS. It is anticipated that efforts by the WHO to eliminate schistosomiasis by the year 2020 in endemic areas will in turn hopefully result in the elimination of this morbid infection in South Africa by the year 2025 (Magaisa *et al.*, 2015).

## 2.4 Lifecycle of Schistosomes

Schistosomes are digenetic dioecious trematodes that reproduce asexually via a freshwater intermediate snail host of the *Biomphalaria* spp and sexually through a vertebrate definitive host. In general, all schistosomes follow the same developmental lifecycle (Fig. 2.2) exhibiting both parasitic and free-living forms as is similar to all other trematodes. Once eggs produced from adult worms reach freshwater and open water sources, miracidia form under the most favorable conditions and through ciliary movement, locate and penetrate a snail intermediate host specific to each schistosome specie (Collins III *et al.*, 2011). Aquatic *Biomphalaria* spp. is imperative for the transmission of *S. mansoni*. *B. glabrata*, *B. tenagophila* and *B. straminea*, which are most susceptible to *S. mansoni* infection (Negrão-Corrêa *et al.*, 2012).



**Fig 2.2: Lifecycle of the Schistosome Worm.** The complex life cycle of the worm contains morphologically distinct life forms whose infective stage known as cercariae (a) begin the cycle by swimming in freshwater to locate and penetrate mammals to transform into schistosomulae (b). These adapt to survive within the bloodstream where they mature into adult schistosomes (c) that migrate and parasitize the mesenteric and vesical venous plexuses of the host where they mate and lay eggs (d). Once the eggs are excreted from the host through urine and feces into freshwater, miracidia (e) are formed that develop into sporocysts (f) once they have parasitized a snail host. The lifecycle is complete when the sporocysts develop asexually and are released back into the watery environment as cercariae (a) (Jolly *et al.*, 2007).

After two rounds of asexual reproduction, the infective stage of the worm (cercariae) is released from the snail and loses its tail to transform into schistosomulae once it comes into contact with human skin, thus parasitizing the lungs, systemic vasculature and hepatic portal system. By this time the pair would have migrated to the mesenteric and vesical venous plexuses whose migration is highly dependent on the species an individual is infected with: *S. mansoni* worms migrate to the superior mesenteric vein, *S. haematobium* to the pelvic venous plexus and *S. japonicum* to the inferior mesenteric vein where they feed on blood, develop into adults and mate (Protasio *et al.*, 2012; Weerakoon *et al.*, 2015). Laid eggs are then shed in the stools and urine of infected individuals thereby continuing the cycle. The worm generally survives for 3–5 years but this may well prolong for up to 30 years with one pair producing up to 600 billion schistosomes in its lifetime (Gryseels *et al.*, 2006).

## **2.5 Signs, Symptoms and Immunobiology of Schistosomiasis**

The clinical features and primary etiology of schistosomiasis occur more as a result of granulomatous reactions caused by deposited eggs in host tissue or primary sites such as the liver rather than by the presence of worms, which lead to a number of immune-fibrotic responses (Gobert *et al.*, 2015). However, the infected host has to deal with different lifecycle stages of the worm: cercariae which penetrate the host's skin, migrating schistosomulae and developed worms and eggs laid by adult pairs. All these stages exhibit several antigenic moieties, which in turn stimulate different humoral and cellular immune responses (Colley *et al.*, 2014). The disease is characterized by two main and distinct conditions: acute and chronic schistosomiasis, which under severe circumstances may deteriorate into the advanced form of the disease.

### ***2.5.1 Acute and Chronic Schistosomiasis***

Acute schistosomiasis, also referred to as Katayama fever, is a toxemic disease that is uncommon amongst populations in endemic areas, but is more characteristic in those who have not been previously exposed to the illness and are visiting such areas for the first time. Primary infection is caused by cercariae that enter the host via a percutaneous route by the use of proteolytic enzymes, exhibiting symptoms that first present as a common rash otherwise referred to as cercarial dermatitis then followed by anorexia, diarrhoea, fever, cough and arthralgias combined with eosinophilia and leukocytosis (Caldas *et al.*, 2008; Jenkins *et al.*, 2005). A dominant T helper 1 (T<sub>H</sub>1)-like response reflects in the host once exposed to migrating immature worms 3–5 weeks after infection. In the plasma are peripheral-blood mononuclear cells (PBMCs), which produce large quantities of tumour-necrosis factor (TNF), interleukin-1 (IL-1) and interleukin-6 (IL-6) (Pearce and MacDonald, 2002). Other pro-inflammatory cytokines and mediators with immunoregulatory functions such as IL-12 and macrophage inflammatory protein alpha (MIP $\alpha$ ), as well as IL-10, prostaglandins (PG) E<sub>2</sub> and D<sub>2</sub> are also produced (Jenkins *et al.*, 2005). Although very rare, death may also result from Katayama fever (Conlon, 2005).

Chronic schistosomiasis results in a shift from the T<sub>H</sub>1 response to a polarized T<sub>H</sub>2 response, which is accompanied by elevated IL-4, IL-5, IL-13 and IL-10 interleukin levels (Araújo *et al.*, 2004). The conversion of a T<sub>H</sub>1 to T<sub>H</sub>2 response is imperative as the absence of such would result in hepatic inflammation coupled with hepatic cellular injury and necrosis (Gause *et al.*, 2003). The strong T<sub>H</sub>2 response is primarily induced by egg antigens and its prolonged state results in chronic morbidity and hepatic fibrosis (Pearce and MacDonald, 2002). Additional studies have shown that the T<sub>H</sub>2 response is actually ‘modified’ by high IL-10 levels and immunoglobulin (Ig) GE, antibody (Ab) isotypes and low IgE reactions (Capron *et al.*, 2005). On

the other hand however, it has been suggested that this same response promotes naturally-acquired resistance to schistosomiasis re-infection amongst those living in endemic areas (Pearce and MacDonald, 2002; Araújo *et al.*, 2004). Granulomatous and pathological lesions form around the ova, which have been deposited around “primary sites” such as the lower urinary tract (*S. haematobium*), leading to terminal haematuria and increased micturition and dysuria, or in the distal colon (*S. mansoni*, *S. japonicum*, *S. mekongi* and *S. intercalatum*), resulting in polyp formation, congestion, edema, abdominal pain or cramping, bloody diarrhoea and dysentery (Barsoum *et al.*, 2013). These primary sites get targeted first for species survival and are of the earliest and worst pathology. The liver is the main victim due to the movement of eggs via the bloodstream to secondary target sites, which include the kidneys, eyes, skin and genital area, which may predispose individuals to sexually transmitted diseases, cause infertility in women due to the formation of lesions on the ovaries and fallopian tubes, while affecting the spermatic chord, epididymis and prostate in men (Clerinx and Van Gompel, 2011).

### ***2.5.2 Advanced stage of Schistosomiasis and other associated infections***

In rare cases, the acute and chronic stages of schistosomiasis may degenerate into extremely - severe clinical outcomes. Neurological symptoms or periovular granulomas in the central nervous system (neuroschistosomiasis) may develop during Katayama fever and may also disappear once the syndrome has been treated or has subsided (Lambertucci, 2010; Lambertucci *et al.*, 1998). Due to the normally under-diagnosed nature of neuroschistosomiasis, the condition can degenerate into a very serious disorder that may manifest in the form of Acute Schistosomal Encephalopathy (ASE), Pseudotumoral Encephalic Schistosomiasis (PES) or Spinal Cord Schistosomiasis (SCS) where symptoms include a simple headache, which can then extend to weakness in the limbs (eventually leading to the inability to walk) or paraesthesia, seizures and

sensory disturbances, motor deficits, vomiting, ataxia, visual abnormalities, altered mental status, bladder dysfunction, constipation and sexual impotence (Ferrari and Moreira, 2011).

Female genital schistosomiasis (FGS) or vaginal schistosomiasis and bladder cancer are other complications that arise from schistosomal infection. *S. mansoni*, *S. haematobium* and *S. japonicum* worms deposited and trapped in the fallopian tubes, vagina and/or especially in the cervix result in sandy patches, lesions and bleeding (Kjetland *et al.*, 2012). This predisposes the infected to sexually transmitted diseases (STDs) and viruses such as human papillomavirus (HPV) and HIV/AIDS, genital itching, abnormal discharges, open wounds, spontaneous abortions and a decrease in fertility.

The World Health Organization and the International Agency for Research on Cancer have concurrently categorized *S. haematobium* as a Group I Carcinogen due to recent reports on the disease predisposing infected individuals to carcinomas such as squamous bladder cancer (Gryseels *et al.*, 2006; Oyinloye *et al.*, 2015). The deposition of eggs in the bladder results in histopathological changes and sandy patches. Additionally, long-term and extensive schistosomal colitis increases risk to colorectal cancer, although evidence of association between the two disorders is still unclear and unproven (Al-Mashat *et al.*, 2001; Coon, 2005).

## **2.6 Diagnostics, Therapies and Control**

Detection of eggs in the stool and urine of infected individuals by microscopic means is the gold standard for diagnosing schistosomiasis, which is recognized by their distinct and peculiar lateral and terminal spine, shape and size (Adenowo *et al.*, 2015). Current parasitological methods include point of contact (POC) tests such as the circulating cathodic or anodic antigen (POC-CCA or POC-CAA) tests as well as the Kato-Katz thick smear or Kato technique (Secor and

Montgomery, 2015). A number of concentration techniques and serological tests are also used and these include Merthiolate-Iodine Concentration Technique (MIFC), Merthiolate-Formaldehyde Concentration (MFCT), Formaldehyde-Ether, Miracidia Hatching Technique (MHT), Circumoval Precipitin Test (COPT), and Indirect Hemagglutination Assay (IHA) (Olveda *et al.*, 2013). However, not all these methods are sensitive enough to detect the presence of eggs; therefore, PCR, PCR-ELISA and Western Blot Analysis may be used as alternatives, even though these techniques are not always suitable for every field.

There have been a successful number of anti-schistosomal pharmacological agents, which were used at various points in time for the treatment of schistosomiasis, most of which were specie-dependent. Antimony potassium tartrate (APT) is regarded as the oldest recorded form of schistosome treatment, dating back to 1605 (Trainor-Moss and Mutapi, 2016). Drug uptake results in females dislodging from their male counterparts and human blood vessels as well suppressed ovulation. APT is no longer used due to its toxic side effects and difficult administration by injection but is instead employed as an emetic for victims of poisoning and for trypanosomiasis, leishmaniasis and anti-microbial treatment (Trainor-Moss and Mutapi, 2016).

In 1948, Lucanthone was introduced as the first *S. mansoni* non-antimonial drug that required conversion by the host into an active metabolite, the hydroxymethyl analogue of lucanthone. This was then named Hycanthone, a drug that displayed activity both in vitro and upon parenteral and oral administration (Cioli *et al.*, 1993). Several propositions have been made concerning its mechanism of action but the intense inhibition of nucleic acid synthesis is one unignorable feature worms display once exposed to the drug. Hycanthone is a frame-shift mutagen whose carcinogenic and teratogenic effects have been observed in animals. Hence its use has been discontinued although its effects in humans have not clearly been identified.

Metrifonate was an organophosphorus cholinesterase inhibiting insecticide developed by Bayer in 1952 that later showed activity against adult *S. haematobium* parasites. Muscular contractions are triggered by disrupted nerve signals and synaptic connections. The multiple doses that are required for effective treatment and its reduced efficacy in comparison to alternative drugs such as Praziquantel has discontinued its use but it is alternatively applied as an insecticide, veterinary anti-helminthic drug and is currently proposed and investigated as a therapy for Alzheimer's disease (Trainor-Moss and Mutapi, 2016).

Oltipraz (OPZ), a synthetic dithiolthione, was mostly used to target *S. mansoni*, *S. haematobium* and *S. intercalatum* infections. Its unclear mechanism of action assumes to decrease glutathione synthetase (GSS) levels while simultaneously exposing the parasite to reactive oxygen intermediates, resulting in its expulsion from the host system (Aruleba *et al.*, 2018). Slow mechanism of the drug, which takes more than two months, its induced photosensitivity as well as its frequent side effects have removed OPZ from the market for schistosome treatment.

Oxaminiquine, a 1,2,3,4-tetrahydro-2-(isopropylamino)methyl(-)-7-nitro-6-nitro-quinoline methanol developed by Pfizer in the 1960s, acts solely on *S. mansoni* male worms (Trainor-Moss and Mutapi, 2016; El Ridi and Tallima, 2013). It has been suggested that vulnerable schistosomes possess a sulfo-transferase causing the drug to undergo esterification. Detachment of the ester generates as electrophilic reaction capable of alkylating schistosome DNA, in turn causing RNA and DNA inhibition. Currently oxaminiquine is only used in Brazil and nowhere else due to the emergence of drug resistant foci, proposed to have appeared due to mutating genes that encode the esterifying enzyme.

Oral administration of the Niridazole anti-schistosomal agent displays activity against *S. mansoni*, *S. japonicum* and *S. haematobium* parasites. The covalent bonding of [14C]-niridazole results in nitro-reduction and bond-formation to parasite macromolecules. The worms are therefore eliminated via a decrease in glycogen levels, which impedes on glucose and lactate uptake (Aruleba *et al.*, 2018). In addition, the drug destroys the worm's female reproductive system and suppresses delayed intolerance. Abandonment of the drug is attributed to its severe side effects.

Mefloquine is an aryl-amino-quinoline that is primarily used as an anti-malarial drug but has also shown efficacy against juvenile and adult schistosomes. Significant reductions in schistosomulae and adult *S. mansoni* and *S. japonicum* worm burdens have been observed following single dose administration (You *et al.*, 2014). Other studies have suggested that this effect may possibly be due to the drug's role as a glycolysis inhibitor. It is now being investigated whether mefloquine treatment is valuable to patients that possess mixed schistosome and malarial infections as it is possible that this may only be effective in cases where the *Plasmodium* parasite is resistant to mefloquine, thus awarding the drug superiority for schistosome treatment since both mature and immature worms would be eliminated during treatment. This proposal however, calls for extensive study and research.

In the same vein, artemisinin is a sesquiterpene lactone, potent in anti-malarial treatment. Of particular interest however, are its derivatives namely artesunate, artemether, arteether and dihydroartemisinin that have been reported in several human and animal experimental studies to possess anti-schistosomal activity (Aruleba *et al.*, 2018). Discovered in the 1990s, it was observed that multiple doses of artesunate and artemether over frequent periodic intervals provided increased efficacy and schistosomulae elimination rates of up to 97% (Olveda *et al.*,

2014). It has also been shown that combination therapies of artemisinin derivatives and PZQ result in higher protection levels than with the use of PZQ alone. Although therapy does not prevent re-infection, infection burden is reduced drastically and complications that include anaemia as well as liver and spleen enlargement are resolved and reversed respectively. However, concerns have also been raised on the potential for *Plasmodium* resistance especially in areas endemic to malaria.

The use of molluscicides for snail control has been considered an effective tool in conjunction with chemotherapy for many integrative approaches especially in countries such as Morocco and Japan, which were successful in eliminating schistosomiasis, while their use has had significant impact in Ghana, Tanzania, Venezuela, Zimbabwe, Philippines, Madagascar and Brazil (Augusto *et al.*, 2018; McCullough *et al.*, 1980; Rollinson *et al.*, 2013). Other countries such as Egypt and The Peoples Republic of China have also followed suite in adopting this strategy and are currently on track in achieving total elimination of the disease. Yurimin, sodium pentachlorophenate, N-tritylmorpholine and nicrotilide were the four compounds previously recognized as the most effective but their use has been discontinued for various reasons (McCullough *et al.*, 1980). Niclosamide (Bayluscide®) is a chemical compound identified and recommended by the WHO for snail control, which has shown promising activity against *Oncomelania* snails and cercariae both in the field and laboratory. However, this molluscicide as with other compounds bears disadvantages of high toxicity especially to fish as well as human cell lines and low water-solubility (Wang *et al.*, 2016). In addition, the product's high costs, associated carcinogenic and genotoxic effects on non-target organisms such as other animals and plants, failure to interrupt re-colonization and the development of drug resistant *Bulinus* and *Biomphalaria* species have all played a role in the limited use of this compound in large scale

control interventions today (Augusto *et al.*, 2018). In light of these circumstances, the face of schistosomiasis control and treatment drastically changed in the 1970s when Praziquantel (PZQ) was introduced as a new, effective, safe and cost-efficient chemotherapeutic drug for treating all schistosome infections.

## **2.7 Praziquantel**

Since the mid-1920s, Kikuth and co-workers pioneered the vigorous search for anti-schistosomal treatment (Alsaqabi and Lotfy, 2014). It was at the laboratory of Bayer, Germany that the underlying structure of Praziquantel (PZQ) – the pyrazino isoquinoline ring - was noted for its anti-parasitic activity. These labs in association with E Merck, Germany discovered various pyrazino isoquinoline compounds, which were screened for anthelmintic activity. In 1977 the activity of PZQ in animals, known as EMBAY at the time, was published followed closely by the publishing of human studies in 1978. In close cooperation with the WHO, successful clinical trials were conducted on the three major strains and in 1983 Shin Poong, a Korean company instigated the drug's low prices due the competitive markets they created through the development of a novel process for PZQ development that resulted in significant reductions in its production. This has still allowed China and Korea to remain the main and largest producers of PZQ raw materials (Alsaqabi and Lotfy, 2014). Eventually producers such as EIPICO began PZQ production in Egypt in 1987 under license from Shin Poong, paving the way for other producers all over the world (Cioli and Pica-Mattoccia, 2003).

The white and bitter crystalline powder is the only WHO recommended drug for schistosomiasis treatment and is composed of two equal 'lacvo' and 'dextro' isomeric parts of which the former possesses anti-schistosomal activity (Kabuyaya *et al.*, 2017; Doenhoff *et al.*, 2008). Cure rates of up to 90% and 90–95% in egg reductions have been recorded attributed to the high levels of

therapeutic dosages linked to PZQ, ranging from 150, 500 to 600 mg based on patient body weight (da Silva *et al.*, 2017). Thus the drug is labeled a Group II Biopharmaceutics Classification System (BCS) member due to its low and high solubility and permeability levels respectively, which are both a direct cause of its low dissolution rates and the major rationale behind the prescribed amounts of treatment. This problematic characteristic renders the production of new formulations a great challenge, hence more research is required to not only increase PZQ solubility but also improve doses for treatment. On the other hand, more authors point toward the tolerability and efficacy of PZQ with the gastrointestinal gut (through the GYP450 3A4 enzyme) absorbing over 80% of the drug and effects occurring approximately 15mins after oral ingestion (Inobaya *et al.*, 2014; Nascimento *et al.*, 2016; Siqueira *et al.*, 2017). Unfortunately, this is only observed 6–7 weeks subsequent to infection due to a tremendous increase in the parasite's susceptibility levels (Siqueira *et al.*, 2017). Despite its high levels of tolerability, PZQ does come with its own share of side effects, which are rather mild and quite short-lived (lasting for about 24 hours), which include rash, dizziness, drowsiness, vomiting, headache, abdominal pain, nausea and tremors (Nascimento *et al.*, 2016).

Schistosomes undergo both physiological and morphological changes once in contact with the PZQ. Though still unclear, studies have suggested that the drug imposes antigen exposure, which produces muscle contraction and paralysis; this in turn has a negative effect on inorganic ion transport by causing an increase in the entry of calcium ions into the schistosomal worm and the opposite in potassium and sodium ion influxes (Aragon *et al.*, 2009). It has been suggested that the sodium potassium pump, which is connected to monovalent and bivalent cation flow (mostly that of calcium) leading to increased muscular movement, contraction, paralysis and eventually death (Siqueira *et al.*, 2017). In addition, the tetanic contractions induced by PZQ interaction

results in the exposure of vacuoles on the parasite's tegument, causing the detachment of worms from the venous walls of the host's tissues leading to muscle function failure and modifications in the tegument structure, leading to death (Siqueira *et al.*, 2017; McManus *et al.*, 2018). Other consequences also include the inhibition of eggs produced by females.

For over 30 years, world schistosomiasis treatment and control programs have relied heavily on this drug and this has raised a drug resistance concern. Decreased susceptibility to the drug has already been reported, both in mouse experiments and in the field (Brindley and Hotez, 2013; Cioli *et al.*, 2014; Fallon *et al.*, 1996; Gouveia *et al.*, 2018; Melman *et al.*, 2009). Other concerns as mentioned before, pertain to the fact that the drug is not sensitive to all stages of the worm, particularly those that are 2–4 weeks old (Aragon *et al.*, 2009; Aragon *et al.*, 2008). Additionally, the drug does not stop individuals from re-infection (Siddiqui *et al.*, 2011). These reasons therefore make it clear that new anti-schistosomal drugs and protective vaccines are needed. In a bid to contribute to efforts towards such, a group of proteins known as Universal Stress Proteins (USPs) have been hypothesized as a possible solution regarding this venture.

Environmental stress is inescapable and living organisms are in a constant battle to curb the manifold consequences that accompany these predicaments. Cellular and genomic integrity may be compromised, in turn damaging cellular structures and interfering with cellular function, leading to the death of the organism (Flick and Kaiser, 2012). More-so, environmental and cellular stress can instigate changes in protein conformation leading to the denaturation and misfolding of proteins. It is a well-established fact that the native state of a protein is vital for it to perform its biological function. Exposure of the organism to various stresses results in a conformational-homeostatic shift toward more “aggregation-prone states” where misfolded proteins whose hydrophobic regions are exposed interact with others leading to abnormal

protein-protein interactions (Weids *et al.*, 2016). In several organisms, universal stress proteins are induced in response to a myriad of environmental insults for their survival (Bangera *et al.*, 2015). Due to the fact that they are a highly conserved group of proteins, they have recently sparked interest as potential novel drug targets. The following sections elucidate the characteristics of this special class of biological macromolecules.

## 2.8 Universal Stress Proteins

The multi-gene universal stress protein (USP) family makes up a highly evolutionarily conserved group of proteins whose members are spread across bacterial, plant, archaea, some invertebrate, yeast, protist and fungal genomes (Vollmer and Bark, 2018; Wang *et al.*, 2017). The expression of USPs provide protective and regulatory effects to the afore-mentioned organisms for either their survival or adaptation to biotic and abiotic slurs such as drought, high temperatures, salinity, toxic chemicals and antibiotics, heavy metals, un-couplers of the electron transport chain and DNA damaging agents such as oxidants (Vollmer and Bark, 2018; Adenowo, 2018; Masamba *et al.*, 2016). Such environmental issues triggers the production of reactive oxygen species (ROS) such as singlet oxygen ( $^1\text{O}_2$ ), hydrogen peroxide ( $\text{H}_2\text{O}_2$ ), hydroxyl radical ( $\cdot\text{OH}$ ) and superoxide anion ( $\text{O}_2^-$ ), which in turn activate downstream signaling processes that include transcription factors, phosphatases, molecular chaperones, protein kinases and defense-related molecules or proteins that include USPs (Jung *et al.*, 2015). The complexities of this system in conjunction with cytokines produced within the schistosome parasite will be discussed in detail in the next chapter of this thesis. Despite the widespread insight into the regulation and physiology of USPs through various foundational studies, the exact molecular and biochemical mechanisms of these proteins have still not been identified. This section summarizes current

information on USPs such as their proposed mechanisms of action and highlights their importance in various organisms as well as in schistosomes.

## 2.9 Structure of Usps

The USP domain is a highly conserved 140–160 amino acid-containing domain that may exist in one of either three forms: as a single 14–15kDa domain, as a tandem domain of approximately 30kDa within larger USP proteins or as two USP domains fused together with catalytic domains that include voltage channels, antiporters, protein kinase domains or amino acid permeases (Adenowo, 2018; Hingley-Wilson *et al.*, 2010). USPs are categorized according to their ability to bind ATP; UspA and UspA-like proteins do not bind ATP while Usps of the FG-type bind ATP (Banerjee *et al.*, 2015). Usp domains are almost always of the FG type and these include Usps F (*ynaF*) and G (*ybdQ*) (Class II) as the name suggests while Usps A, C (*yecG*) and D (*yiiT*) (Class I) form part of the UspA subfamily (Heermann *et al.*, 2009). It has been proposed that this domain in turn falls under the  $\alpha/\beta$  domain referred to as the **HIGH**-signature proteins, **Usp**-like domains and **PP**-ATPase (HUP) domain, which is an ancient and diverse group of proteins believed to have evolved from UspA-like proteins by losing their ATPase and nucleotide-binding functions (Kvint *et al.*, 2003). The domains present in this group are mostly small domains assembled as Rossmann fold-like structures consisting of three layers containing five parallel  $\beta$  sheets with two flanking  $\alpha$  helices on each side (Bangera *et al.*, 2015). Class III and IV are composed of the two UspE (*ydaA*) units that is, E1 and E2 (Jung *et al.*, 2015; Shokry *et al.*, 2014; Tkaczuk *et al.*, 2013).

The ATP-binding site is characterized by the conserved G-2X-G-9X-G-(S/T) motif and proteins that bind (UspFG) and do not bind ATP (UspA) in respect to substrate specificities despite possessing significant substrate similarities (Tkaczuk *et al.*, 2013; Melencion *et al.*, 2017;

Siegele, 2005). The UspA of *Haemophilus influenza* is 68% and 12% homologous to *E. coli* UspA and the MJ0577 Usp from *Methanococcus janaschi* respectively. Similarities have been observed in its tertiary  $\alpha/\beta$  fold within the asymmetric dimer, which is similar to that of MJ0577, where adjacent  $\beta 5$  strands result in an anti-parallel interaction forming the interface of both proteins (Adenowo, 2018). Among the structural differences notable between the two proteins include their ATP-binding activity. A study by Sousa and McKay confirmed that UspA does not bind ATP despite the use of 1mM MgATP included in crystal growth experiments. Tight ATP binding has however been observed in MJ0557 experiments despite the exclusion of ATP from purification and crystallization experiments. The stimulation of ATP hydrolysis by MJ0577 is suggestive of the protein functioning as a cell factor-dependent ATPase, which sheds more light on the biochemical activities of bacterial Usps which still require more identification (Sousa and McKay, 2001).

## **2.10 The Universal Stress Proteins of *E. coli***

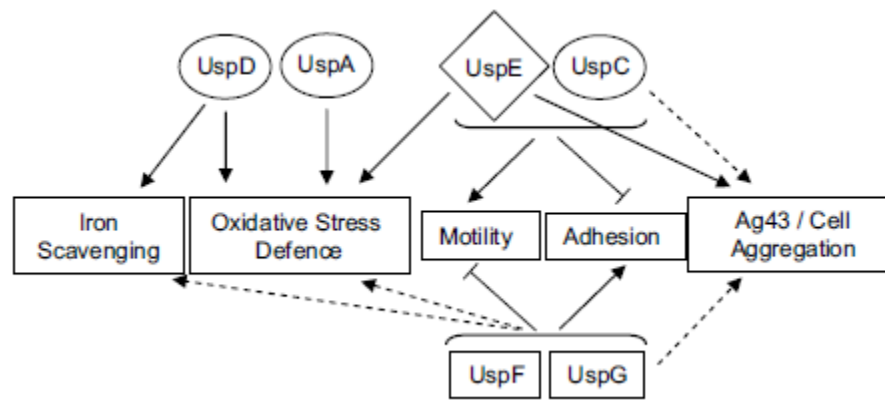
Universal stress protein A was first discovered over 25 years ago in 1992 by Nyström and Neidhardt in *E. coli* and was first referred to as C13.5 due to its appearance as a 13.5-kDa cytoplasmic protein on a 2-dimensional Isoelectric Focusing (IEF) PAGE-gel (Vollmer and Bark, 2018; Wang *et al.*, 2017; Kvint *et al.*, 2003; Bradley, 2011). The name later changed to USP since it appeared to have an almost universal responsiveness to a wide variety of stresses that include DNA and membrane damage, heat shock, heavy metal stress and other toxic agents such as acids, oxidants, antibiotics and the inhibition of growth caused by sulphur, nitrogen, phosphate or carbon nutrient starvation (Freestone *et al.*, 1997). Only cold shock and the tetracycline antibiotic do not exhibit UspA response. Later studies soon discovered that the production of UspA is more growth phase than growth rate-dependent owing to the fact that

growth-arrested cells shift whatever resources remain after depletion towards their maintenance rather than to growth and proliferation (Freestone *et al.*, 1997; Persson *et al.*, 2007). Thus, the response elicited by the UspA protein is one that is non-specific and highly abundant once cell growth ceases (Nyström and Neidhardt, 1994). The importance of the protein for cell survival has been demonstrated in a number of ways: 1) production of mutant *E. coli* cells lacking UspA, show increased susceptibility to death or impairment when exposed to complete and prolonged growth inhibition stresses such as extended carbon starvation, H<sub>2</sub>O<sub>2</sub>, carbonylcyanide-chlorophenylhydrazone (CCCP), CdCl<sub>2</sub> and dinitrophenol (DNP) exposure as well as osmotic shock; 2) the over-expression of UspA results in the blockage of growth-arrested cells; and 3) mutations in the *uspA* gene exerts an effect on gene expression timing, which in turn accelerates gene expression pattern changes and protein synthesis (Nyström and Neidhardt, 1994; Diez *et al.*, 2000; Nyström and Neidhardt, 1993).

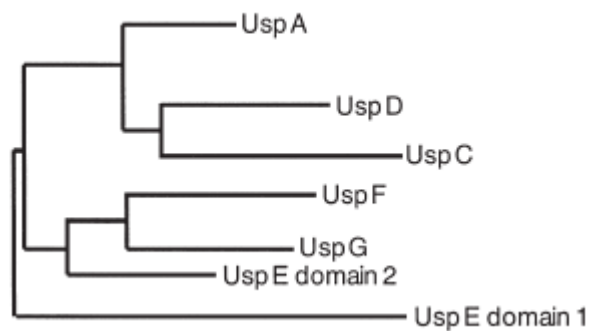
Transcription of the *uspA* gene is as a result of a  $\sigma^{70}$ -dependent promoter, which is turned off rapidly once a nutritional up-shift is observed in growth-arrested bacteria (Bradley, 2011; Nyström and Neidhardt, 1993). Due to the fact that the gene is incorporated in the FadR regulon suggests its potential role in fatty acid or membrane lipid metabolism since FadR represses and activates genes involved in fatty acid degradation and biosynthesis respectively (Kvint *et al.*, 2003; Freestone *et al.*, 1997). Positive regulation of the *uspA* gene is via alarmone guanosine tetraphosphate or ppGpp, a key nucleotide that regulates numerous stationary phase-induced genes, which it does through the  $\beta$ -subunit of RNA polymerase (Kvint *et al.*, 2003; Persson *et al.*, 2007). Subsequent analysis of the *E. coli* genome led to the discovery of five more *usp* family genes in bacteria named *uspC*, *uspD*, *uspE*, *uspF* and *uspG*, which are all associated with adhesion, motility and resistance to oxidative stress (Fig 2.3 and Fig 2.4) (Wang *et al.*, 2017;

Nachin *et al.*, 2005). Together with UspA, UspD not only defends against agents responsible for superoxide generation but is also a regulator of intracellular iron. UspE and UspD on the other hand, are essential for cellular motility, which takes place at the expense of cell adhesion through *uspE* and *uspD* genes, whose link to cell adhesion has been demonstrated via electron microscopy. However, UspF and UspD exhibit the exact opposite of UspE and UspD effects (Nachin *et al.*, 2005; Kim *et al.*, 2012; Loukehaich *et al.*, 2012). All this prove the evolvement of Usp to exhibit various physiological roles within the cell for its survival, defense and escape from cellular stress (Nachin *et al.*, 2005).

The UspA protein is a serine and threonine autophosphorylating protein whose phosphorylation is highly dependent on the TypA tyrosine phosphoprotein, which is activated in response to stasis (Freestone *et al.*, 1997; Nachin *et al.*, 2005). A possible explanation for the importance of this feature is the ability for UspA to regulate the detrimental effects of other proteins in growth-arrested cells. It has been suggested that UspA in coercion with the o591 tyrosine phosphoprotein degrades such proteins by phosphorylation tagging (Freestone *et al.*, 1997). Other studies have also suggested that the increases in UspA phosphorylation may be the factor responsible for the acidification of a minimum of six proteins during the early stages of growth arrest in cells (Freestone *et al.*, 1997; Nyström and Neidhardt, 1996). Other studies have suggested that the absence of this protein increases sensitivity when cells are exposed to mitomycin C and UV irradiation since it forms part of the RecA-dependant DNA protein and repair system (Liu *et al.*, 2007).



**Fig. 2.3: The 6 *E. coli* USP proteins and their associated functions.** The circular and square shapes around the names signify that the indicated USPs belong to Class I and II respectively while the diamond shape represents class III and IV. The forward and T-shaped arrows on the other hand represent the positive and negative effects of the proteins performing a specific function respectively while the brackets illustrate the involvement of both the proteins in the processes (Nachin *et al.*, 2005).



**Fig. 2.4: Phylogenetic analysis of the 6 *E. coli* USP proteins.** The analysis indicates the subfamilies of the proteins as they are grouped based on the sequence similarities in their domains (Gustavsson *et al.*, 2002).

## 2.11 USPs in Intracellular Human Pathogens

A total of 9001 Usp sequences as of the year 2001 were recorded and described in the Pfam database, while a hefty 1306 bacterial species constitute 7655 USP domains from the above mentioned figure (Bradley, 2011). *Halobacteriaceae* represents the specie with the most expressed *usp* genes due to the large challenges it faces with regards to osmoregulation while organisms such as *Natrialba magadii* ATCC 43099 harbor 54 Usps consisting 66 domains. Apart from bacteria, several Usps have been noted to play significant roles in the promotion and persistence of infectious diseases. The PA4352 protein is one that belongs to the Usp superfamily, which contains two Usp domains in tandem from *Pseudomonas aeruginosa*: the dominant pathogen of the cystic fibrosis (CF) lung infection (Boes *et al.*, 2006). Knockout mutant experiments exhibited a decrease in pathogen viability and sensitivity to carbonyl cyanide *m*-chlorophenylhydrazone, which is the chemical uncoupler of the electron transport chain. The study showed that the protein aids in anaerobic energy stress survival, which is important for prolonged infection.

Structural and functional studies of two Usps known as YdaA (UspE) and YnaF (UspF) from *Salmonella typhimurium*, which causes salmonellosis, reported possible functions of the proteins in chloride sensing, ATP hydrolysis and roles in membrane alterations (Bangera *et al.*, 2015). Tuberculosis is an acute pulmonary disease caused by the bacteria *Mycobacterium tuberculosis*. Cyclic AMP (cAMP) is actively released by the bacteria and studies of the Rv1636 Usp show that it binds to cAMP with higher affinity and efficiency than to ATP. This suggests that mycobacterial Usps may act as ‘sinks’ for the secondary messenger for its availability in downstream effectors during alterations in ATP levels within the cell (Banerjee *et al.*, 2015). Additional studies showed the inability of Rv2623-deficient guinea pigs and mice that unable to

establish chronic tuberculosis infection and attributed the ATP-binding ability of the protein as a signaling intermediate for the promotion of persistent infection (Drumm *et al.*, 2009). Other pathogens that involve Usps in their pathways to aid with stress-related responses and for the promotion of pathogenesis include *Porphyromonas gingivalis*, which causes subgingival plaque, *Acinetobacter naumannii*, which is responsible for hospital-acquired infections such as ventilator-associated pneumonia, bacteremia and soft-tissue infection and *Listeria monocytogenes*, which causes food-borne infections especially among the elderly and immune-compromised individuals (Chen *et al.*, 2006; Elhosseiny *et al.*, 2015; Gomes *et al.*, 2011).

## 2.12 USPs and Schistosomes

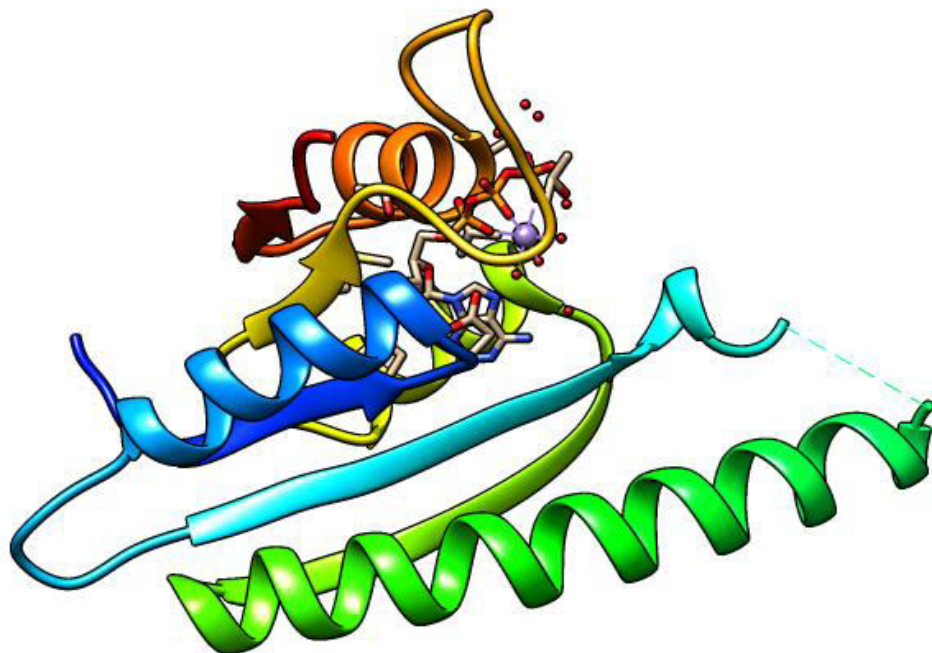
Genomics analyses using various bioinformatics tools by Isokpehi and colleagues identified eight genes that encode the USP domain with six sequences indicating that their functions are regulated by ATP due to the presence of ATP-binding motif residues. Further analyses by Reverse Transcriptase-Polymerase Chain Reaction (RT-PCR) showed that transcription of the genes were present in no less than one developmental life stage of the worm. The up-regulation of six of these were however identified within the miracidium, suggesting possible specialized functions in response to oxidative stress in the form of H<sub>2</sub>O<sub>2</sub> produced by snail hepatocytes (Isokpehi *et al.*, 2011). Subsequent bioinformatics studies by Mbah and co-workers show the length of protein sequences and domains, enzymatic and developmental, ligand binding sites and potential biologically relevant and regulatory interacting ligands as well as sub-cellular localization of five and eight *S. mansoni* and *S. japonicum* USP sequences respectively. Attention was however drawn to two 184-amino-acid-long USP sequences; Q86DW2 from *S. japonicum* and G4LZI3 from *S. mansoni* as they both possessed the same enzymatic regulation, ligand binding sites and chemical ligands as well as the same sub-cellular localization (Mbah *et*

*al.*, 2013). The study inferred the interaction of three putative ligands: calcium, magnesium and zinc cations with the two proteins and the presence of conserved sites that could be instrumental in the regulation of *Schistosoma* Usps.

Determining the function of a protein can pose as a very big challenge especially in the absence of characterization. Identifying homologues of the protein of interest may be a strategy for inferring function while designing experiments that achieve this objective. However, associating function and homology is quite complex and difficult to achieve as sequence similarity does not always equate to function similarity (Pearson, 2013). Nevertheless, function can still be inferred, especially from significant or excess similarities. The MJ0577 USP (PDB ID: 1MJH), a homologue that shares 28% similarity with the G4LZI3 protein, comes from the *Methanococcus jannaschii* hyperthermophile. Structure shows that the protein contains bound ATP, which is suggestive that function may include the protein being an ATP-mediated molecular switch or an ATPase (Fig. 2.5) (Zarembinski *et al.*, 1998). ATP was not included in growth media so the bound ATP may have occurred as a result of the protein scavenging ATP from its host without its release during purification or hydrolysis. A nucleotide-binding pocket is present, whose motifs are similar to those found in ATP-binding proteins, although their spacing and arrangements are significantly different to ATP-binding proteins. This categorizes the structure of the protein under a different family of ATP-binding molecules. Anchored to the ATP is a divalent cation, thought to be either magnesium or manganese. Structure comparison conducted via the Dali server (<http://ekhidna2.biocenter.helsinki.fi/dali/>) revealed the existence of several proteins exhibiting similar folds despite the lack of sequence similarity. This however, provided little to no functional information. Biochemical experiments on the other hand, showed that for ATP hydrolysis to take place, MJ0577 requires the presence of soluble molecules such as

ATPase-activating proteins such as those analogous to GTP-activating protein for Ras GAP. This suggests that the protein may function as a molecular switch within certain cellular processes (Zarembinski *et al.*, 1998).

Interestingly, USP-encoding genes have not been identified in humans making them potent and attractive drug targets. Therefore, this current study aimed at determining the three-dimensional structure of the G4LZI3 protein for the potential development of novel structure-based drug schistosomicides. The following section briefly describes a number of the various methods that may be employed to fulfill this quest. However, more emphasis is laid on nuclear magnetic resonance spectroscopy and single crystal X-ray diffraction commonly known as macromolecular X-ray crystallography.



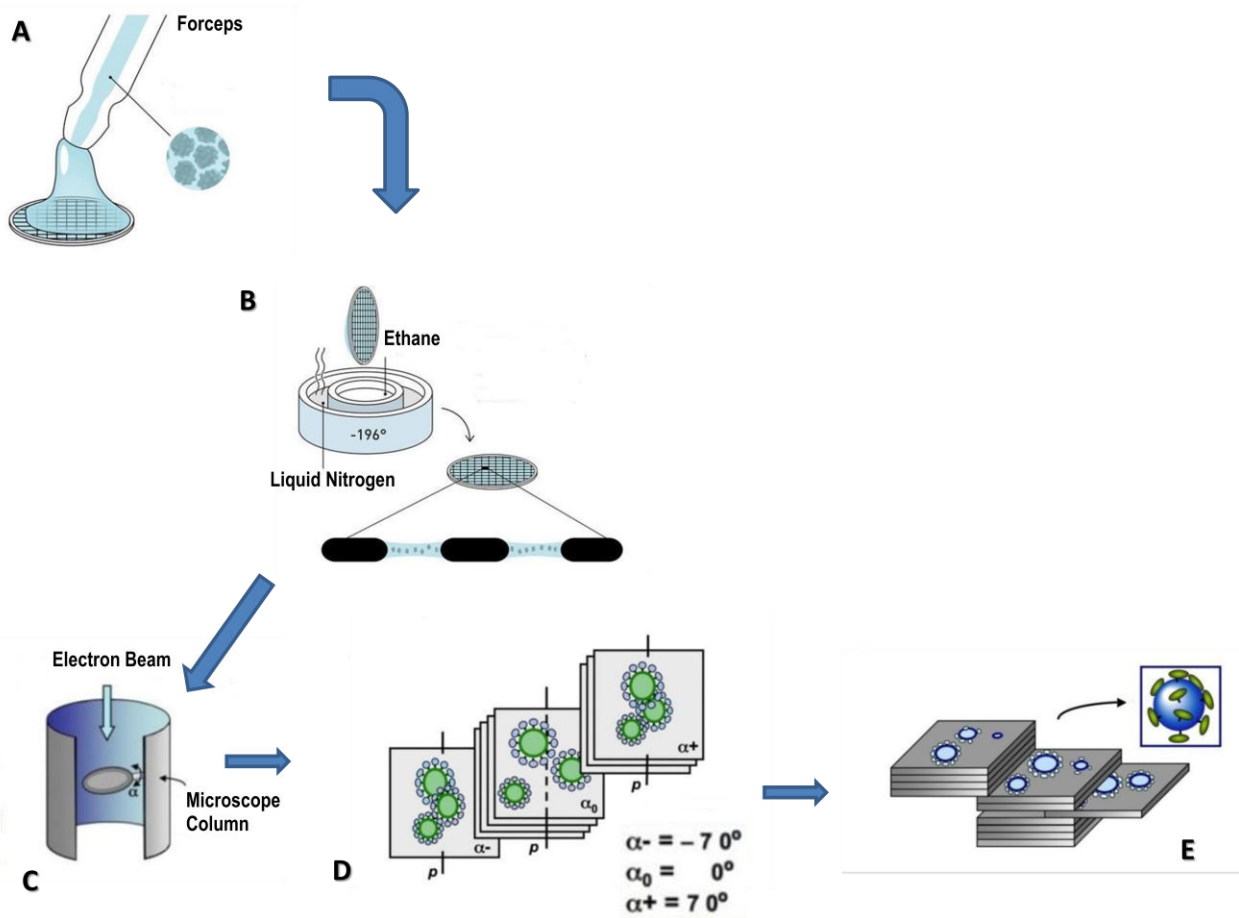
**Fig. 2.5: Structure of the MJ0577 USP (PDB ID: 1MJH) from *Methanococcus jannaschii*.** Structure shows a bound ATP molecule and shared homology with the G4LZI3 protein suggests the latter may incorporate ATP to carry out part of its function. Structure accessed from the Protein Data Bank database.

## 2.13 Methods for 3D protein structure determination

### 2.13.1 Cryogenic Electron Microscopy (Cryo-EM)

Over the years, cryo-EM has undergone an excellent number of advances in its technique and has been viewed as a “resolution revolution” that has observed a steady increase in the acquisition of near-atomic resolution structures of proteins ranging even up to 4Å (Koning *et al.*, 2018). Between the years 2011 to 2016, the number of articles published in association with the Electron Microscopy Data Bank nearly tripled from only 36 deposited structures in 2014 to a hefty 114 and 240 structures in the years 2015 and 2016 respectively (Earl *et al.*, 2017). EM allows for the three dimensional determination or reconstruction of individually purified monodisperse macromolecules greater than 100 kDa such as whole cells, organelles or tissue sections at cryogenic temperatures and the visualization of these assemblies using a transmission electron microscope (Carroni and Saibil, 2016; Lyumkis, 2019). Preliminary visualization and high contrast images may involve the use of solutions containing heavy metal salt to negatively stain samples before drying them (Carroni and Saibil, 2016). This however, may result in low resolution images because of the introduction of artifacts and/or stain granules to the sample during staining. Hence, samples are now thinly blotted onto an electron microscopy grid, dipped in liquid ethane and subjected to rapid freezing temperatures of below -185°C in a high pressure chamber, forming a ‘shield’ of vitreous ice around the sample (Vénien-Bryan *et al.*, 2017). Entrapment by the layer of ice does not only provide a level of protection against damage that may result from the intensity of electron beam scattering, also preserves the characteristics of the biological structure in a vitreous state without crystallizing the water. The sample is additionally sheltered against artifacts and heavy metals associated with stains. However, care is taken to ensure that the ice is not thick enough to distort the particles (Fig. 2.6).

The samples are exposed to low dose electron beams of  $10\text{--}20\text{ e}/\text{\AA}^2$  due to the high sensitivity of the samples to electron damage and for the preservation of resolution (Carroni and Saibil, 2016). This results in thousands of projection images that are combined and averaged. Using *ad hoc* software, the final map is constructed based on the accuracy with which the particle orientations were determined. The unique challenges associated with EM are its preference to macromolecules larger than 250 kDa because of its low signal-to-noise ratio, poor contrast of resulting images, as well as the movement of the electron beam during exposure as this also contributes to poor image quality. The latter results from a break in covalent bonding, which results in the formation of gaseous molecules such as hydrogen, methane, nitrogen and oxygen that escape the vitreous shield, and in turn move the sample and degrading high resolution data (Vénien-Bryan *et al.*, 2017).



**Fig. 2.6: Simplified Illustration of the Dubochet's Vitrification Method.** (A) Preparation begins by transferring ~3uL of material to a TEM grid made out of perforated carbon metal mesh, which is blotted using filter paper to remove any excess sample. (B) The sample is dipped into liquid ethane, forming a thin film of vitreous ice across the holes of the film layer. (C) The sample is then positioned perpendicular to the electron beam and can be rotated at angles between  $-70^\circ$  to  $70^\circ$  by the specimen holder. (D) Images are generated and recorded with 'p' representing the orientation of each image. (E) The final 3D reconstruction map is produced. The three dimensional model (blue insert) can then be extracted from the map for additional analysis (Broadwith, 2017; Fogarty *et al.*, 2011).

### ***2.13.2 Small-angle Scattering (SAS)***

Small-angle scattering applications of X-rays and neutrons (BioSAXS and BioSANS) are novel techniques that incorporate X-ray and neutron scattering of atoms within a biological macromolecule or aggregates of molecules for 3-dimensional determinations (Trehella, 2016). SAS takes into account the fact that the cell is a congested and erratic environment where biological macromolecules such as proteins, polysaccharides, nucleic acids and lipids are not in a rigid state but are constantly interacting with one another over spatiotemporal distribution in numerous conformational states that are imperative for their various functions (Vestergaard, 2016). Hence, SAS experiments require no specific sample preparations and are performed either in a liquid or experimentally relevant state.

When X-rays and neutrons pass through matter, two sets of waves are generated: plane waves that exhibit both amplitude and wavelength, and secondary wavelengths, which are produced from interactions or ‘scatterings’ of the wavelets, with each atom giving rise to a diffraction pattern that can either interfere constructively or destructively (small angle scattering) (Jacques and Trehella, 2010). This begets data that is of low-resolution but possesses high precision information based on the molecule’s size and shape. Although excellent purity as well as high concentration and solubility of the sample is a prerequisite for the technique as is the case with all other structure determination methods, the characteristic advantage of SAS lies in the speed at which data is collected (this can literally be in seconds) and the manner with which sample characterization is conducted (which is done almost immediately) (Mertens and Svergun, 2010). Additionally, factors such as little to no radiation damage that allows for the collection of multiple data sets which can be compared and analyzed individually, low milligram requirements in non-crystalline solutions, speedy and efficient sample mounting and the development of more

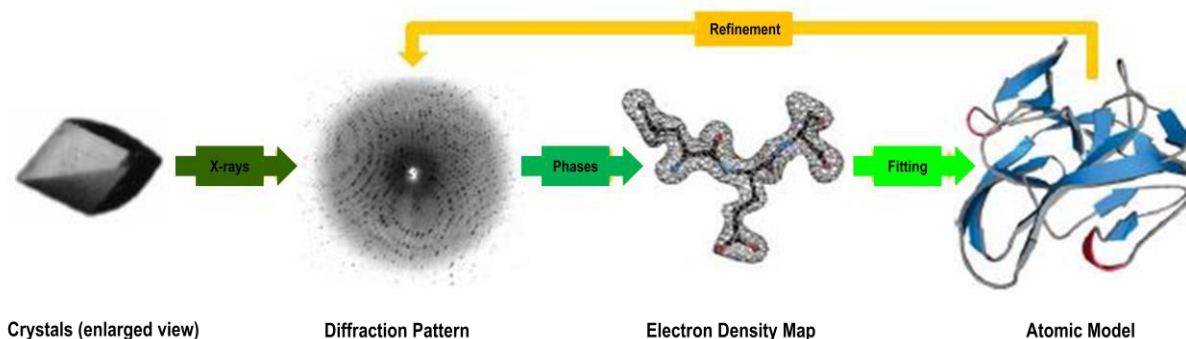
user-friendly and automated software packages, makes SAS a fantastic complementary tool to other determination methods for model development (Vestergaard, 2016; Mertens and Svergun, 2010; Petoukhov and Svergun, 2006).

### ***2.13.3 Nuclear Magnetic Resonance vs X-ray Crystallography***

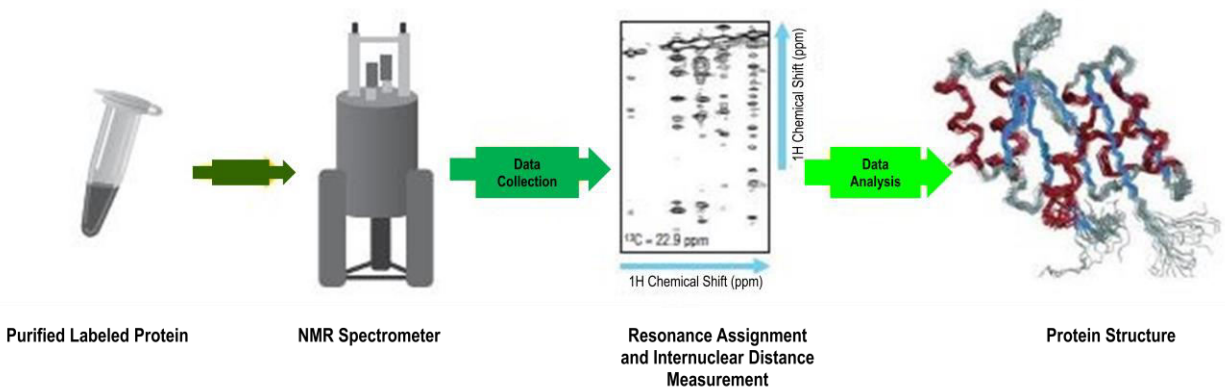
The principal and most powerful techniques responsible for the majority of experimental determination of atomic three dimensional structures of proteins are single crystal x-ray crystallography (XRD) and solution Nuclear Magnetic Resonance Spectroscopy (NMR). Both techniques determine the atomic coordinates of atoms within a molecule using their own distinct physical processes. XRD makes use of x-rays that are bombarded against a protein crystal without losing energy. An interference pattern is produced when atoms within the crystal scatter the rays, giving rise to an electron density map, which is in fact a contour map that visualizes the positions of all non-hydrogen atoms in the molecules within the crystal (Fig. 2.7a). Analysis of the electron density map is conducted with the use of semi-automated computational software. XRD relies heavily on the production of well-ordered crystals but conditions such as pH and ionic strength required for the crystallization process may in fact distort the structure, causing its appearance in solution to differ significantly from that in a rigid and packed crystalline state (Brandt, 2000).

NMR on the other hand, bypasses the challenges and artifacts associated with the crystallization process by studying the protein in solution. Structure determination may take place by one of two approaches: 1). Electromagnetic resonances are absorbed by the various nuclei present in a protein at various frequencies because of the different electromagnetic settings that are produced as a result of the 3-D structure of the protein. Therefore, the nuclear magnetic resonances of carbons and nitrogens that are labeled isotopically within the molecule can be measured. 2). By

unsettling the resonances of various atoms and observing the response of each resonance, internuclear distances can be calculated. Ultimately the coordinates of all bonded and non-bonded close contacts in the form of distances between atomic nuclei, including hydrogen, will be produced (Fig. 2.7b) (Brandt, 2000; Guss and King, 2000; UAH, 2004). NMR is usually considered when proteins are abundant in high and soluble concentrations in the absence of aggregates. Additionally, it is advantageous in situations where proteins struggle to crystallize as is commonly the case with membrane proteins (Brandt, 2000; UAH, 2004). The technique is however limited by the small proteins and protein complexes that it can solve (less than 50kDa), due to the spectral complexity of larger proteins (Billeter, 1992).



**Fig. 2.7a: Structure Determination by X-ray crystallography.** The production of crystals is imperative to execute any x-ray crystallographic experiment. A synchrotron is often used to source x-rays that are bombarded against the crystal to produce a diffraction pattern that generates an electron density map for the different parts of the crystal. From the map, a model of the protein is constructed and compared to the actual diffraction pattern. The model is refined to a point of no convergence where the difference between the actual diffraction pattern and what was calculated is reduced until correspondence between the model and reality is approximately the same (UAH, 2004).



**Fig. 2.7b: Structure Determination by NMR.** Highly concentrated and isotopically-labeled protein is placed in a magnetic field where the sample is subjected to radio frequency pulses. Based on the environment of the nuclei or atoms within the protein, radio frequency radiation is released and recorded in the form of NMR spectrums. The positions of the atoms will be calculated from the chemical shifts and used to generate various models which will be compared. The difference between the models will be calculated or measured and used to determine structure quality (UAH, 2004).

## 2.14 Protein production strategies

For the structural biologist, producing high-quality samples in milligram quantities is top priority. It is imperative for crystallographic application that the protein not only be stable and properly folded, but be of at least 95% purity and devoid of any contaminants that would degrade, modify or destabilize the protein before or during the crystallization or structure determination process (Kim *et al.*, 2008). To facilitate and circumvent some of these technical challenges, the use of cloning vectors manipulated by an inducible T7 RNA polymerase promoter is commonly employed for the over-expression of recombinant and soluble protein. For this reason, the *Escherichia coli* bacterium has been exploited for years as the most extensively used and well established expression system due to its cost effectiveness, ready accessibility and easy-to-use nature (Edwards *et al.*, 2000). Additionally, its talent in producing large amounts of protein has given it a competitive edge over other expression systems for decades. In T7 systems, protein expression may be stimulated either through the manipulation of carbon sources or through the use of the chemical isopropyl- $\beta$ -D-thio-galactoside (IPTG) inducer (Consortium, 2008). In some cases, however, bacteria may fail to execute post translational modifications such as glycosylation, which may result in improperly folded, insoluble or unsuccessful production of protein. In such cases, a variety of possible strategies need exploring. Examples include converting to alternative strains of *E. coli* compatible with the chosen promoter system and the use of an appropriate antibiotic-selection marker. Ampicillin over the years has been the go-to option but is now gradually being replaced by more stable chemicals such as: kanamycin, carbenicillin and chloramphenicol.

During the protein expression process, a vast amount of proteins apart from the protein of interest are also produced, resulting in a complex mixture of proteins, cells and tissues. Hence,

procedures that involve the isolation of the protein of interest from non-protein parts and undesired protein are non-negotiable. The inclusion of short affinity peptide tags to the target protein as a fusion such as the N-terminal 6-His tag (which is the most preferred), drastically facilitates purification without necessarily affecting the biological or biochemical activity of the protein (Consortium, 2008). Also, it has become apparent over the years that the presence of affinity tags actually enhances solubility, aids proper folding and increases yield of the target protein (Waugh, 2005). Numerous tags have been recorded in literature, all bearing their own sets of pros and cons. These include the all popular glutathione S-transferase (GST), maltose-binding protein (MBP), 6X-Histidine affinity tags as well as other less common tags such as FLAG octapeptide, biotin acceptor peptide (BAP) and thioredoxin. The most frequently used among the bunch has by far been the His-tag, whose purification strategy is based on immobilized metal affinity chromatography (IMAC) that uses the Ni(II)-nitrilotriacetic acid (Ni-NTA) resin as the chromatographic matrix. Advantages of this system include its ability to purify protein under denaturing conditions, the capacity it possesses to undergo several cycles of regeneration, its inexpensive resin in comparison to other purification systems and its placid elution conditions (Waugh, 2005). Equilibration buffer containing low concentrations of imidazole is first applied to the column resulting in the immobilization of imidazole to the metal ion ( $\text{Ni}^{2+}$  or  $\text{Cu}^{2+}$ ) (Spriestersbach *et al.*, 2015). The protein sample is loaded next with histidine proteins binding to the column and displacing imidazole. Binding buffer is poured down the column for a washing step and the bound proteins are released by gradient or step elutions of 100 to 500mM imidazole concentrations (Healthcare, 2015). Depending on the elution strategy, one of two chromatograms may result in either two peaks: the first representing non-specific binding proteins present in the lysate and the later peak representing the histidine-tagged target protein,

or one peak corresponding to a slightly pure protein of interest, which can be purified further using size exclusion chromatography (also known as gel filtration) as the last polishing step.

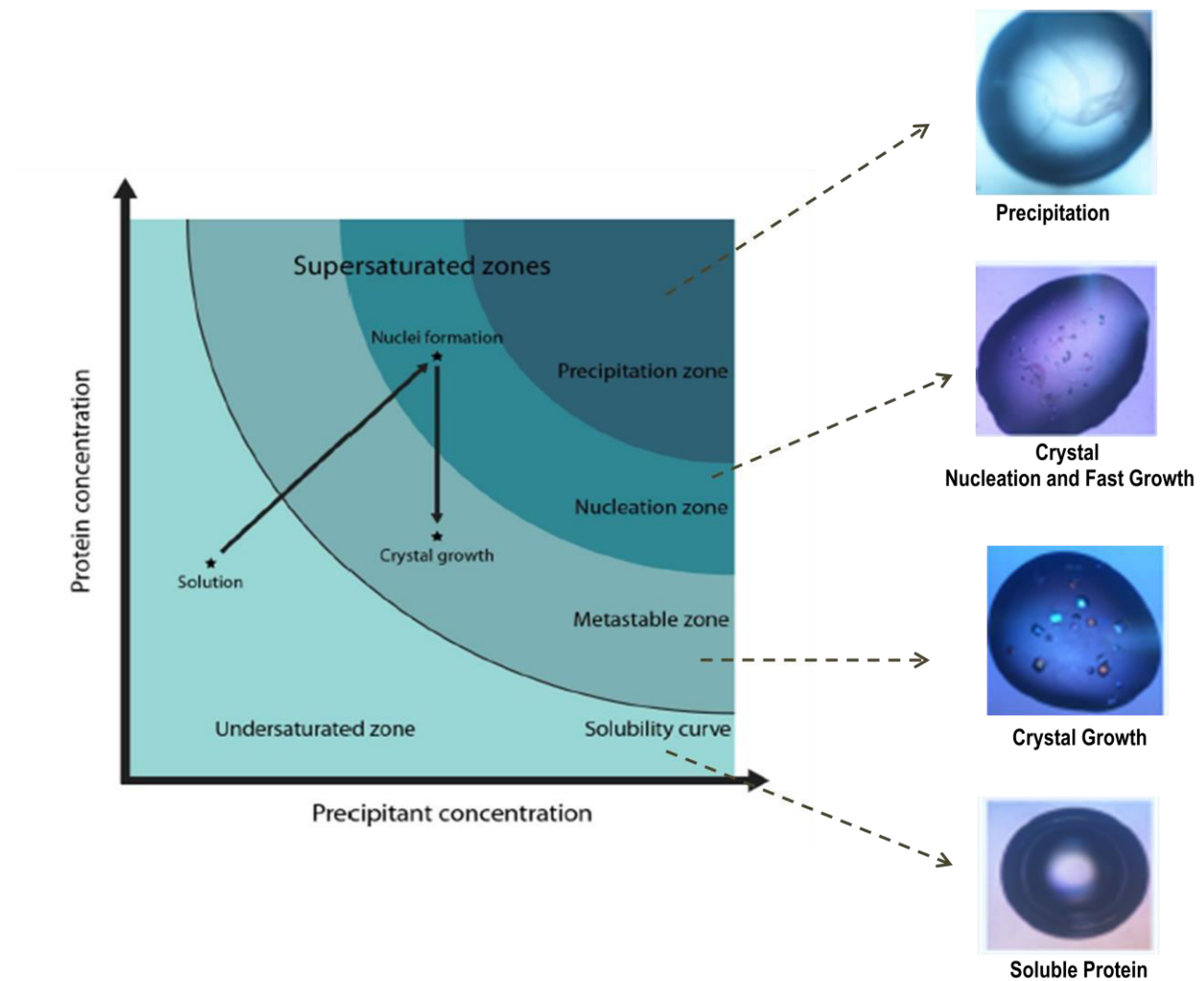
Ion exchange and size exclusion chromatography (SEC) emerge as additional purification steps for the removal of persistent contaminants. Gel filtration holds the advantage of separating impurities and large aggregates that may elute close to the void volume (Kim *et al.*, 2008). The principle of the system lies in the disparity between the flow delays of the various molecular weights or hydrodynamic radiuses of the different contaminants and this is referred to as an “inverse molecular sieving” mechanism (Mojsiewicz-Pieńkowska, 2014; Striegel, 2008). The protein sample is applied to a column whose stationary phase is packed with heteroporous material made up of cross-linked polymeric beads. Particles small in size permeate the beads while larger particles in a process known as ‘exclusion’, diffuse through the spaces in solution between the beads, eluting out first and producing the first peak on the chromatogram. Smaller molecules gambol between the mobile and stationery phase, occasionally entering the beads and eluting next at an intermediate position while the smallest prolong their journey through the column and exit last.

The crystallization of macromolecules in nature is an extremely rare event. During early biology and biochemistry days, it was generally accepted that only diamond, calcite and quartz were the crystal properties of the mineral world until the accidental discovery of hemoglobin protein crystals in 1840 (Vekilov, 2016). Even after several years, the process of producing diffractable crystals is regarded more of an art than a science (Gavira, 2016). This is mainly due to the countless unpredictable and complex nucleation and growth parameters required for well-formed crystals. Hence, the process is deemed more of “trial and error” with hardly any straightforward or rational methods.

## 2.15 The Empirical Task of Protein Crystallization

The crystallization process can be summed up into three main steps: nucleation, growth and termination of growth. A highly concentrated sample is subjected to a state of supersaturation, which is slowly forced out of solution whilst simultaneously avoiding non-specific aggregation (Jaeger, 2004). Properties such as pH, type of buffer, precipitant concentration and ionic strength, saturation and concentration of the sample itself, temperature, detergents and organic additives all play different roles in crystal formation. The nucleation stage is the most difficult of all the stages to address, both theoretically and experimentally because it is the transition phase of the protein molecule from a disordered to an ordered state that results in partially ordered crystalline intermediates referred to as critical nuclei. Crystallization is best described and illustrated by the phase diagram (Fig. 8). A non-equilibrium solution of mother liquor containing the macromolecule (protein) at a concentration high above its solubility limit is first combined and referred to as a supersaturated state.

A precipitate or protein nucleus is formed by the equilibrium reached between the excess energy gained (thermodynamic contribution) and the rate at which this energy has been attained (kinetic contribution). Formation of the solid state may be encouraged by modifying either the chemical activity or properties of an under-saturated solution or that of the macromolecule itself to prevent or reduce its solubility.



**Fig 2.8: Protein Crystallization Phase Diagram.** The solubility curve separates the phase space into regions of undersaturation (where crystals dissolve) and supersaturation (where crystallization of protein occurs). The supersaturation region is further divided into three levels. The protein precipitates (Precipitation Zone) and protein concentration decreases after the formation of nuclei (Nucleation/Labile Zone) to induce slow crystal growth process (Metastable Zone) to cause the protein crystal to attain its highest degree of order (Khajepour, 2012; Russo Krauss *et al.*, 2013)

This may be achieved by:

- i. Altering the pH of the protein, which will in turn alter the ionization state of surface amino acid residues,
- ii. Supplementing the solution with salt, which would alter the chemical activity of the water,
- iii. Adding bridging ions or adjusting the pH to increase protein molecule attraction for one another,
- iv. Optimizing the interactions between solvent and the macromolecules by the addition of polymers or ions.

Once nuclei form, the energy barrier is crossed allowing conditions to enter the zone of supersaturation. Formation of stable nuclei in this zone will continue growing slowly in an orderly fashion, precipitating larger but fewer crystals while depleting the solution of its nutrients resulting in a decrease in supersaturation until the system regains equilibrium (Gavira, 2016; McPherson and Gavira, 2014; Smatanová, 2002). It is therefore obvious that the success of nucleation and growth during the crystallization process critically hinges on the supersaturation variable. Several techniques are available to drive this process as has been described in the list below but the most common are vapor and batch diffusion.

**List of various approaches used in achieving protein crystal growth** (McPherson and Gavira, 2014; Leunissen, 2001):

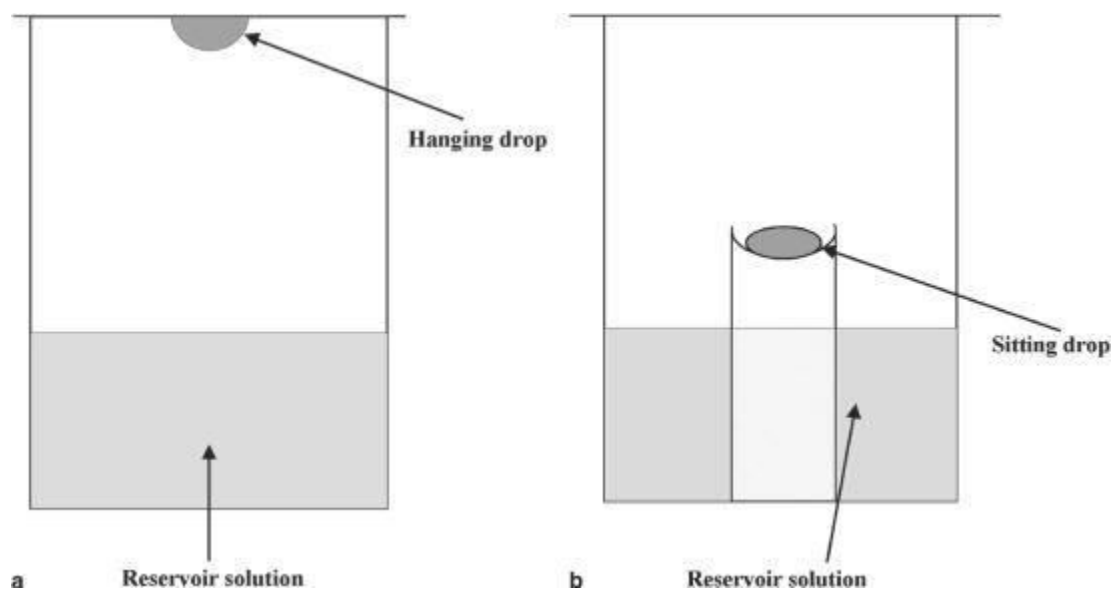
- Bulk dialysis
- Bulk crystallization
- Controlled evaporation

- Vapor diffusion
- Vapor diffusion in hanging drops
- Vapor diffusion on plates and slides (sitting drops)
- Vapor diffusion by microscopic seeding
- Macroseeding
- Batch method
- Liquid bridge
- Microbatch under oil
- Microdialysis
- Sequential extraction
- pH-induced crystallization
- Counter diffusion in capillaries
- Concentration by evaporation
- Free interface diffusion
- Crystallization by effector addition
- Temperature-induced crystallization
- Electrodialysis against tap water

### ***2.15.1 Vapor Diffusion***

This crystallization method is the most frequently used and can be applied in one of either two ways: the “hanging drop” (VD-HD) or “sitting drop” (VD-SD) method (Smatanová, 2002). A sample containing precipitant or crystallizing reagent, buffer and additives such as salt and protein solution are mixed into a drop and either suspended under a silicone-coated cover slip (hanging drop) or placed on an elevated post (sitting drop) above a liquid reservoir containing

buffer precipitant and salt at least double the concentration of the crystallization cocktail (Fig. 2.9). Equilibrium is achieved by diffusion of water from the droplet through air-gap separation into the reservoir due to the difference in the drop and reservoir concentrations. As diffusion takes place, the protein concentration in the sample increases by dehydration-driven reduction until the precipitant concentration is approximately the same as the reservoir concentration and crystals begin to form (Gavira, 2016; Smatanová, 2002; Leunissen, 2001) Both SD and HD methods hold the advantage of being cost effective experiments to set up. However, choice of method mostly depends on the ease of fishing crystals, which the latter provides since the crystals are on a siliconized glass surface and multiple experiments can actually be performed in a single reservoir, thereby increasing the chances of crystal formation. SD crystals tend to adhere to the plastic surface of the well, making the crystals difficult to remove, thus placing their integrity at risk. On the other hand, this method is more time efficient, easier to set into stable sitting drops with organic reagents and detergents and is best in aiding the nucleation process.



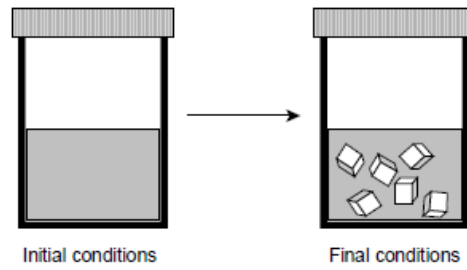
**Fig. 2.9: Hanging and Sitting Drop Vapor Diffusion Crystallization Experiments.** A low saturated drop containing protein solution is suspended over a highly concentrated reservoir where sample concentration increases as water leaves the drop to the reservoir for the formation of nuclei and subsequent crystal growth. Equilibration is attained once the reservoir and reagent concentration correspond to each other. Diagram taken from (Ilari and Savino, 2008).

On a general note, the vapor diffusion technique offers the benefits of being able to influence conditions such as the supersaturation and equilibration rate by altering the distance between the drop and reservoir, enclosing the top of the reservoir with an oil barrier or by diluting or concentrating the reservoir to vary concentrations in the sample without having to interact with the drop. In spite of this, the technique has its own share of drawbacks in that the sample still remains in its highly supersaturated state even after nucleation, resulting in additional simultaneous nucleation and crystal growth. The continuous change in conditions causes the trial and error processes that are so commonly associated with crystallography.

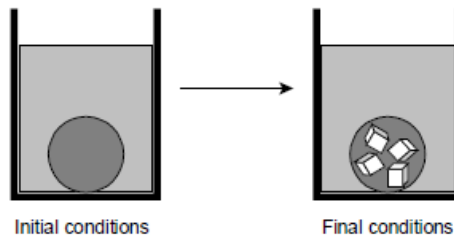
### ***2.15.2 Crystallization by Batch Diffusion***

Batch crystallization, which may include the “micro-batch under-oil method” tries as much as possible to avoid diffusion of water and desiccation of the drop by mixing protein and precipitant cocktail into a drop whose amount may range from 1  $\mu\text{L}$  to 10  $\mu\text{L}$  (Gavira, 2016; Dessau and Modis, 2011). The drop is covered with an inert oil layer such as paraffin or silicone. This protects the sample from airborne contamination and is a method more popularly used for crystallizing small molecules. The main advantages of this method include direct activation of the protein into the nucleation zone, tighter control of the crystallization conditions and better crystal quality. The crystallization conditions remain the same throughout the experiment, enabling the crystallographer to know the exact conditions that were applied if crystallization occurs. This may be disadvantageous on the other hand in that when crystallization does not occur, the crystallographer has to guess which conditions are required for success, which will involve disturbing the droplet (Fig. 2.10).

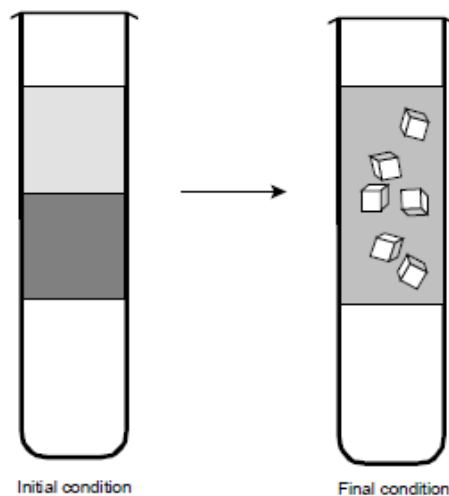
### Batch crystallization



### Microbatch crystallization



**Fig. 2.10: Batch and Microbatch Experiments.** Supersaturated protein solutions in batch experiments are sealed inside vials and left unperturbed for crystal growth. In microbatch experiments, both protein and precipitant are submerged in inert oil where concentration of the solution decreases for equilibrium to be obtained (Leunissen, 2001).



**Fig. 2.11: Crystallization technique within a single vessel.** Protein and precipitant solution are placed within a vial and as mixing between the solutions occur, crystal growth is encouraged as the bulk of the solution becomes supersaturated whilst preventing the crystal from dissolving (Leunissen, 2001).

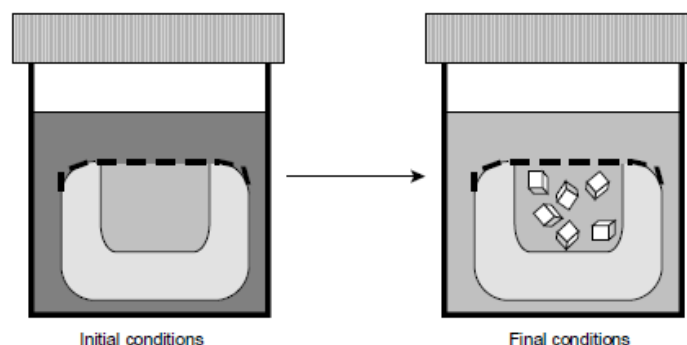
### **2.15.3 Other Crystallization Methods**

#### **2.15.3.1 Free Interface Diffusion**

Alternative crystallization methods include free interface diffusion where the sample and the precipitant diffuse into each other due to their close contact, eventually resulting in crystal formation (Fig. 2.11). It is expected that the ‘regions’ or ‘interfaces’ where the solutions will become supersaturated from precipitant penetration will result in nuclei development. Small nuclei will disappear and dissolve back into solution while larger ones will continue to grow. It is best that microgravity conditions are used for this technique due to the possibility of an eventually formed protein crystal sinking into the high precipitant concentration zone, allowing it to grow too fast and be of poor quality. Hence, the “liquid bridge” technique can be used as a modification of this method. It uses free diffusion by using a fine needle to draw a thin liquid bridge between a drop of the sample and mother liquor that are respectively placed in close proximity to each other on a glass slide. Evaporation is prevented by sealing the drops with cover slips. Diffusion occurs “across the bridge” resulting in crystallization.

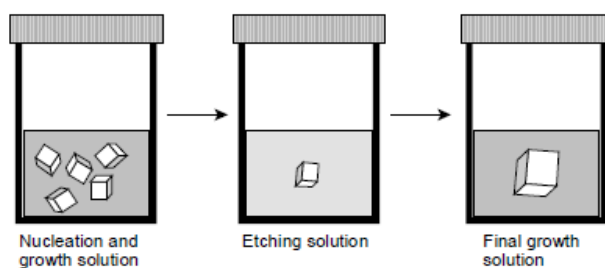
#### **2.15.3.2 Dialysis Crystallization**

Another successful method is dialysis and this technique retains constant protein concentration throughout the experiment, similar to the batch method. The difference lies in its use of microdialysis buttons or short permeable thick-walled capillary tubes to contain the sample (Gandhi *et al.*, 2018). The apparatus is suspended or immersed in a concentrated organic solvent or salt, allowing crystals to form by precipitation or in water, only when it is known that the protein will not denature or aggregate (Fig. 2.12).

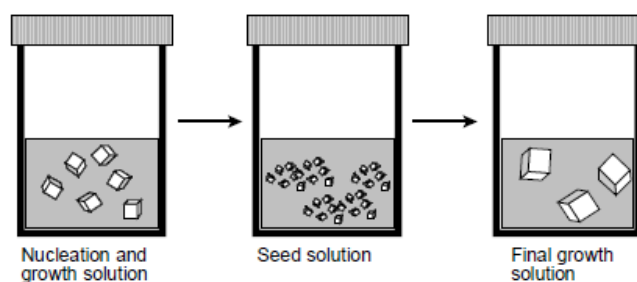


**Fig. 2.12: Crystallization by Dialysis.** Protein molecules are subjected to a fixed volume, causing protein concentration to remain constant. Dialysis via a semi-permeable membrane changes the protein solution concentration resulting in nucleation and growth (Leunissen, 2001).

### Macroseeding



### Microseeding



**Fig. 2.13: The Macro and Microseeding Crystallization Approach.** Formed crystals, either single (macroseeding) or multiple small ones crushed to make a seed stock (microseeding), are transferred to a precipitant solution, sometimes containing protein, with the intent of growing larger crystals. The seeds act as a “ready-made template” and hence skip the spontaneous nucleation step (Leunissen, 2001; Bergfors, 2007).

Through this technique the desired precipitant strength for the crystal can be achieved and one parameter at a time can be controlled since protein concentration remains constant. However, protein concentration might be the limiting factor for crystal formation, which may in the end count as a drawback.

### **2.15.3.3 Seeding**

One of the most eminent and versatile crystallization strategies for optimizing crystal growth is through seeding (Ali *et al.*, 2013). This method is typically applied:

- i. To increase crystal size,
- ii. To reduce crystal showers,
- iii. When crystals do not grow individually but in clusters,
- iv. When spontaneous nucleation takes too long,
- v. For reproducibility of experiments when unpredictable nucleation occurs (Bergfors, 2007).

Seeding is a potent tool used to decouple nucleation from the growth phase. For nucleation to occur, supersaturated conditions are ideal while the growth of crystals requires lower levels. Thus, the conditions for nucleation are simply not enough to support further crystal growth. In seeding, crystals that have reached the nucleation stage are transferred to a medium that supports crystal growth. A supersaturated drop is prepared and a ‘seed’ or small crystal is introduced into the drop for the growth of larger single crystals. This may be achieved through micro- or macro-seeding (Fig. 2.13).

Micro-seeding involves the transfer of a solution of small nuclei, which can be taken from a seed stock while macro-seeding entails the transfer of a single crystal to the growth medium. The latter is however, more labour intensive as it requires that the parent crystal be washed multiple times in unsaturated solution to remove particles or any macromolecules that may have stuck or even poisoned the crystal surface (Bergfors, 2003). Nevertheless, it is more beneficial in the long-run especially for increasing crystal size. Micro-seeding has its own pitfalls in that the amount of crystals transferred to the drop can be very difficult to control. Hence, different ways have been devised to try and circumvent these difficulties. Seed stock can be prepared by crushing or pulverizing crystalline particles using seed rods or beads, vortexing the crystals in mother liquor or by tissue homogenization. Serial dilutions of the seed stock can also be prepared to distribute the nuclei while a pipette can also be used to aliquot a small amount of the solution. A seeding wand dipped in the seed stock can additionally be streaked or stirred into the drop for the transfer of nuclei.

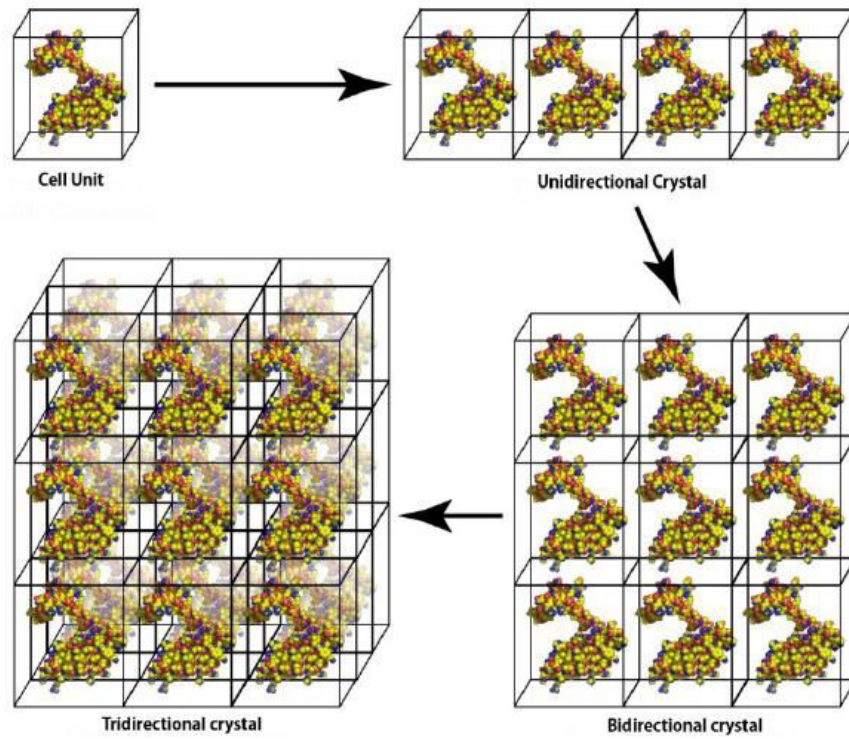
## **2.16 Principles of X-ray Crystallography**

As is the case with every biological technique, knowing the underlying principles is imperative in understanding why and how a study was conducted. This section reviews the main aspects and basic theory behind macromolecular x-ray crystallography.

### ***2.16.1 What is a crystal?***

Due to the fact that proteins are ‘mobile’ because of their numerous and constantly modified orientations and rotating bonds within living organisms, determining their structure in such a state poses a challenge. In general, over 99% of all x-rays that strike against a molecule either pass through without scattering or do scatter but with very weak diffraction (Khajepour, 2012; Gandhi *et al.*, 2018; Mattsson, 2011). Crystallography overcomes this difficulty by subjecting the

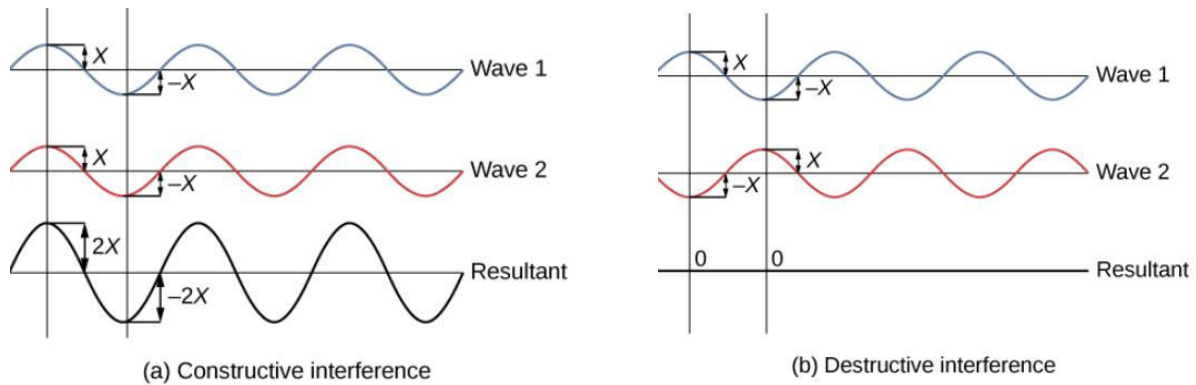
molecule to a ‘frozen’ or ‘fixed’ state. The term crystallography stems from the Greek words *crystallon*, which means “frozen or cold drop” and *grapho* meaning ‘write’. A crystal in basic term is therefore a highly ordered and unique three dimensional arrangement of a high number of molecules in a solid or crystalline state (Khajepour, 2012). The smallest parts of the crystal are referred to as “cell units”, which are identical repeating units (atoms, ions etc.) that form throughout the crystal in several definite arrays of alignments with points known as ‘lattices’ (Fig. 2.14) (Mattsson, 2011). Stacking of the unit cells allows them to fill all space but this in turn restricts their shape. For this reason, only 7 crystal systems, which are characterized by their symmetry based on either the size or shape of the unit cell (point symmetry) or positioning of the lattice points (translational symmetry), is possible. These are named the Triclinic, Monoclinic, Orthorhombic, Tetragonal, Trigonal, Hexagonal and Cubic crystal systems. Point symmetry rotates around the center of a pattern at  $90^\circ$ , which results in an unchanged pattern that maps back onto it while Translational Symmetry moves the pattern from one position to another in a linear direction by a certain amount causing the same pattern to occur again. Hence, the total set of these symmetries is mathematically described as the Space Group of the crystal, which in essence, describes the internal structure and symmetry of the crystal. When x-rays interact with a rotating crystal, Bragg reflections and only a few crystal lattice points are diffracted based on the crystal’s orientation. Hence, various exposures at different orientations is imperative for 3D construction (Khajepour, 2012). Reflections result in image peaks, which are based on the crystal lattice points that in turn correspond to each space group (x-ray crystallography). Therefore, crystals with a high symmetry are preferable as they would require less diffraction images to cover the crystal lattices or points.



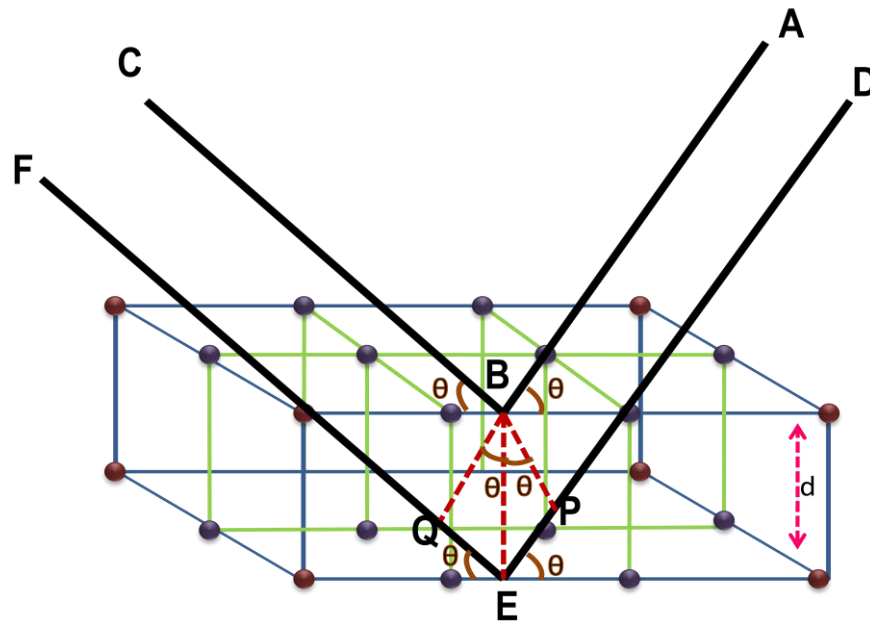
**Fig. 2.14: Representation of a unit cell to the crystal.** The smallest parts of the crystal are unit cells that are stacked upon each other to form planes of lattices, which are in turn arranged to constitute the 3-dimensional crystal (Khajepour, 2012).

### **2.16.2 X-ray Diffraction: Bragg's Law**

X-rays are defined as waves of accelerating electrons or electromagnetic radiation, which are normally generated from synchrotron storage rings for the purpose of x-ray crystallography experiments using electrons that travel with the speed of light (Gandhi *et al.*, 2018). Elastic scattering occurs due to the 'scattering' and production of secondary spherical waves caused by striking x-rays against a crystal (scatterer). Bearing in mind that crystals are ordered arrays of atoms, striking x-rays against a well-ordered crystal will also produce a regular array of waves. Most generated waves will undergo a phenomenon known as destructive interference where both positive and negative wave peaks possessing the same maximum displacement overlap, eventually canceling themselves out and resulting in a low total amplitude (Fig. 2.15a). A few waves however, possess peaks and troughs that correspond to one another resulting in increased amplitude (constructive interference). These interferences are responsible for the intensities of spots observed in diffraction patterns and well-ordered crystals allow for increased constructive interference, which is directly proportional to spot intensity (Khajepour, 2012). This phenomenon is described by Bragg's Law:  $2d\sin\theta = n\lambda$  (Fig. 2.15b)



**Fig 2.15a: Constructive vs. Destructive Interference.** Amplitudes of waves add up and increase in phase (constructive interference) while destructive interference contains waves that cancel out each other because they add up  $180^\circ$  out of phase (PhysicsLibreTexts, 2019).



**Fig 2.15b: Visual of Bragg's Law.** X-rays that incidentally strike a crystal surface at angle  $\theta$  are scattered back also at angle  $\theta$ , producing a path difference  $d$  that equates to a whole number of wavelength  $n$ , leading to constructive interference.

The internal dimensions of the crystal lattice, which contains arranged parallel planes can be described by a set of integers (h,k,l) known as Miller indices. Once radiation in the form of x-rays hit against the crystal, Wave 1 (AB) diffracts and travels along BC while some radiation bypasses the first lattice plane to produce Wave 2 (DE), which also diffracts and travels along EF. However, Wave 2 travels an extra distance from PE to QE and because these distances are equal, Wave 2 essentially travels QE twice or vice versa (2PE). The angles of deflection are described as  $\theta$  or angle QBE. Distance BE is the spacing between the two parallel planes of the lattice (d) with Miller indices h,k,l ( $d_{hkl}$ ). Hence, distance QE is equal to  $d_{hkl}\sin\theta$ . However, because Wave 2 has travelled twice QE further than Wave 1, this can be described as multiple integers of the wavelength ( $n\lambda$ ) where  $n = 1,2,3,4\dots$ etc. This results in constructive interference, which in turn causes the reinforcement of waves and intensity of spots. Hence,  $QE + PE = 2d\sin\theta = n\lambda$ . This is known as Bragg's Law (Fig. 2.15b).

### 2.16.3 The Phase Problem

The corresponding regular array of light spots in the diffraction pattern formed as a result of crystal exposure to x-rays is referred to as the reciprocal lattice. Widely-spaced atoms within a crystal results in closely-arranged spots within the reciprocal lattice with this phenomenon also varying with direction, while the addition of atoms without changing the crystal's position results in spots with varying intensities. These intensities are directly related to the 'density' or 'spacings' of atoms within the unit cell, which can be described mathematically by the Fourier Transform equation (1):

$$p(xyz) = \frac{1}{V} \sum_{hkl} \underbrace{|F_{hkl}|}_{\text{Amplitudes}} \exp(i\alpha_{hkl}) \underbrace{(-2\pi i h k + k y + l z)}_{\text{Phases}}, \quad (1)$$



Real Space  
Direct Space



Reciprocal Space

where amplitude of the wave  $F_{hkl}$  increases with the square root of the spot intensities. To calculate the electron density of a position or point within the unit cell (xyz) will require the summation of all the xyz point contributions ( $\sum_{hkl}$ ) of the unit cell's planes ( $F_{hkl}$ ), whose amplitudes are directly proportional to the electrons present in the plane.  $V$  represents the unit cell volume, while  $\alpha_{hkl}$  is the phase associated with the unit cell's plane ( $F_{hkl}$ ). Consequently, determining the contents of the unit cell determines the diffraction pattern and determining crystal scattering (diffraction), ascertains the contents of the unit cell (Cowtan, 2001). The electron density map accomplishes the former by describing the average contents of the unit cells of the crystal. For this to be achieved however, both the amplitude and the phase of the diffracted waves are important. The amplitude (spot intensities) can be deduced experimentally but the phase, which is a measure on how far a wave has gone during its cycle or its relative positions, unfortunately cannot be detected during data collection because this property gets lost during the diffraction experiment (Gandhi *et al.*, 2018; Taylor, 2003). This is termed “The Phase Problem”.

Solving the phase problem is in itself a problem due to the fact that there is no direct relationship between the phases and amplitudes of waves. The only relationship between the two entities is through the reconstruction of the electron density map (Taylor, 2003). It is simply not enough to gather information on the amplitude only as the phase is also needed for the development of the electron density map, which works out the points in the molecule by working backwards from the small waves emitted from the sample to the larger and more complex waves produced on the detector (i.e. from diffraction to atomic positions). The electron density is determined by the inverse Fourier transform of the diffraction pattern produced from both the amplitude and phase of the waves (Náray-Szabó and Perczel, 2014). Hence, solving the phase problem becomes the rate-determining step in deciphering the 3D structure after the acquisition of protein crystals

(Pusey *et al.*, 2005). Various methods have therefore been derived to assist in this regard, which are discussed below.

### **2.16.3.1 Patterson Methods**

It is possible to solve the structures of small molecules with limited complexity without phase information through what is known as the Patterson Methods, which assumes or sets all the phases to zero by summing  $F(hkl)$  to  $F^2(hkl)$  to produce the Patterson function (2) (Patterson, 1934):

$$p(uvw) = \frac{1}{V} \sum_h \sum_k \sum_l |F_{hkl}|^2 \cos[2\pi(hu + kv + lw)] \quad (2)$$

Additionally, it may be defined as the Fourier transform of observed amplitudes (intensities) (3) (Grosse-Kunstleve and Schneider, 2007):

$$\text{Patterson} \propto FT^{-1}(I) \quad (3)$$

The Patterson function calculates the atomic spacing in the molecule and gives rise to a Patterson map that indicates peaks corresponding to every inter-atomic vector in the structure (Cowtan, 2001). The position of the heavy atoms can therefore be deduced from the location of the Patterson peaks. This method however, is only limited to structures containing 20–50 atoms due to overlapping inter-atomic vectors.

### **2.16.3.2 Multiple Isomorphous Replacement (MIR)**

The larger the protein, the more atoms it will contain, thus, lowering the resolution of the diffraction pattern. Multiple Isomorphous Replacement (MIR) is a “substructure method” that introduces heavy atoms into the macromolecule without changing its structure but ensuring that the heavy atom crystal (derivative) contains the same lattice properties as the native crystal thus making it isomorphous (Cowtan, 2001; Lieberman *et al.*, 2013). The method seeks to locate the

positions of heavy atoms in the protein via the various scatterings produced by the introduced atom so as to calculate the magnitude and phase. Introduction of these heavy atoms (e.g. Pt, Au, Ag, Pb, Os, Ir) either by co-crystallization or soaking, produces a notable difference in the respective diffraction patterns (intensities) of the native and heavy-atom (derivative) crystal. The patterns between the samples are compared and a stronger structure factor ( $F_{hkl}$ ) emitted from the heavy-atom crystal is considered constructive interference and vice-versa. However, comparison of diffraction patterns from multiple crystals bound to various heavy atoms then becomes necessary to reduce ambiguity and noise (Cowtan, 2001). This information is then used to uniquely and accurately calculate the phases of the native crystal with correct handedness (Lieberman *et al.*, 2013; Sayre, 2002).

### ***2.16.3.3 Anomalous Dispersion***

Multiple or Single Wavelength Dispersion (MAD or SAD) are other substructure methods that present themselves as popular alternatives to the MIR experiment. Every atom possesses a few “absorption edges” where the amplitude and phase varies with wavelength. Varying the wavelength factor makes it possible to determine phase information similar to the MIR method if the positions of the atoms responsible for scattering are known (Cowtan, 2001). This is through the differences in intensities exhibited by symmetry-related reflections known as Bijvoet pairs (Lieberman *et al.*, 2013). In the SAD experiment, these differences are used to solve the phase problem, while MAD phasing compares the intensities collected just above the “absorption edge” and at regions on the crystal that have been exposed to fewer doses of energy. MAD techniques require the use of only one crystal hence, holding an advantage over the MIR method. However, the two can be combined and complemented to produce greater phasing power.

#### 2.16.3.4 *Molecular Replacement (MR)*

Molecular replacement (MR) stands as an additional and the most common or attractive phasing method used for structure determination with more than 70% of all structures solved by x-ray crystallography employing MR as the method of choice due to the fact that it is cheap, fast and highly automated (Evans and McCoy, 2008). This strategy makes use of a similar structure's phases (sequence identity > 25%; r.m.s. deviation of < 2.0Å between the  $\alpha$ C atoms of the unknown structure and homology model) to produce a starting model or a significant portion of it (Mattsson, 2011; Taylor, 2003). By assuming that sequence similarity implies structure similarity, the structure of the homology model is positioned in various orientations within the crystal's unit cell (Sayre, 2002). Patterson functions are then explored (4):

$$X_2 = [C]X_1 + d \quad (4)$$

which states that position  $X_1$  observed in molecule 1 are present at position  $X_2$  of the second molecule (Messerschmidt, 2007). Model structure factor magnitudes are deemed optimal once the translation of the model is oriented correctly relative to the origin of the crystal's unit cell (Taylor, 2003). The phases of the obtained structure factors are then used to calculate Fourier syntheses to generate electron density maps for model building, correctness and refinement. The MR method also has its own shortcomings in that this approach is not feasible when the search model is significantly different or too incomplete from the crystal structure because one cannot then assume that the optimal position is determined and this may lead to the production of an incorrect model position. Hence, about four factors need to be taken into consideration when the method is undertaken: choosing and improving the models, scoring each determined orientation and position, searching for solutions, and implementing certain computer tricks to speed up

calculations (Evans and McCoy, 2008). Although these factors are in essence independent from each other, poor choices may result in the development of a poor model or no solution at all.

Computer programs designed for MR allow a process that would otherwise be slow and extremely tedious for the crystallographer, become faster and more efficient as more searches may be made thereby increasing the number of MR trials conducted. Automation additionally makes room for less satisfactory models, data or crystal symmetry to be improved through the use of algorithms while exploring all choices available in the structure solution. Phaser is one such software that contains novel phasing algorithms by making use of maximum likelihood and multivariate statistics (McCoy *et al.*, 2007). Alternative software which incorporate a few of its features include *AMoRe*, *MOLPREP*, *SHARP*, *MLPHARE* and *BP3* among others (Pusey *et al.*, 2005; McCoy *et al.*, 2007; Taylor, 2010). *phenix.mr\_rosetta* is another automated tool that combines Phenix and Rosetta crystallographic and structure modeling algorithms to generate and evaluate models. It is an approach that is ideally taken when data resolution is low or when models disagree with the target. With this technique, more complete and accurate models are generated with better model and electron density quality (DiMaio, 2017; Terwilliger *et al.*, 2012).

#### ***2.16.4 Electron Density Maps***

Interpretation of diffraction from XRD determination experiments is represented by the electron density map, which is a depiction of model-data fit based on the combination of experimental data and the structural model, averaging the contents of the crystal's unit cells. Due to the phase problem, estimates of initial phases by MR first produces ill-defined and noisy maps that undergo improvement once crystallographic model building begins, either manually or by the use of automated software (Lamb *et al.*, 2015). Interpretation of the electron density map may contain

errors, especially in cases with low resolution data resulting in continuous regions of density. Refinement aims to position the structure into the electron density as best as possible and these adjustments are performed manually until the difference in phase estimates can no longer be observed or made, which at this point the refinement is stopped and the structure considered solved. Resulting structures can then be used to screen for small low-molecular weight molecules, which can act against the clinically relevant macromolecule for the development of novel drugs (Chilingaryan *et al.*, 2012).

## **2.17 Structure-Based Drug Discovery**

The initiative in using structures obtained by XRD for drug discovery surfaced more than 30 years ago but was almost abandoned completely by pharmaceutical companies as the strategy was considered both time-consuming and extremely costly hence, it was regarded more as an academic venture (Blundell 2006). However, a drastic transformation in drug design occurred almost a decade ago, changing perspectives which have currently led to the structures of key proteins now used throughout the drug discovery process. Traditionally, protein structure was used in lead optimization and the identification and selection of druggable or tractable drug targets, which were generally performed using homology models. It was not until more recently that x-ray crystallography has been employed for virtual screening or the detection of chemical fragments. Structure-based virtual screening involves the ‘docking’, sterical fitting or computational interaction of small bioactive and interactive molecules onto the protein of interest and the assessment of the fit by considering the various conformations the protein and ligand may take followed by the scoring of the interactions taking into account electrostatics, hydrogen bonding, polar interactions, correct prediction of main chain flexibility, water binding and free binding energy (Bonetta 2016). In addition, SBDD also involves the optimization of other

parameters that include metabolic stability, distribution, dissolution, toxicity, absorption, plasma protein binding, pharmaceutical properties as well as cost effectiveness (Davies 2003).

## 2.18 Computational Chemistry

Computational chemistry in basic term refers to a mathematical method that has been developed significantly well enough to be automated for use on a computer to describe and solve chemistry problems (Shih, 2009; Young, 2001). It presents itself as a novel tool that may be used to model molecular systems for the synthesis of compounds, which would otherwise require labor intensive procedures, several raw materials, generate a whole lot of toxic waste and involve the expense of precious time. In addition, clarity to problems are obtained and understood more clearly as some solutions can arrive more easily than through the experimental. In the pharmaceutical industry, catalytic reactions and interactions of potential drugs by docking candidate molecules into the active site of a candidate can be explored (Lewars, 2003). Through this strategy, the scientific world has experienced a notable paradigm shift, so much so that most engage in computational studies before the commencement of any major research project. Although it does not replace the experimental part of any scientific discovery due to the fact that the development of novel treatment requires laboratory tests, computational chemistry poses as a cost effective, environmentally safe and speedy approach for the study of molecular and material science (Lewars, 2016). A few of the main tools used under this method include molecular mechanics, molecular dynamics and density functional calculations. In the case of proteins however, computational chemistry incorporates bioinformatics for understanding the three-dimensional configuration of the protein and this begins with structure prediction. Present methods include homology or comparative modeling, *ab initio* or *de novo* prediction and fold recognition or threading.

## 2.19 Structure Prediction Methods

Characterizing a protein sequence in terms of its function is one of the most challenging aspects in biology (Eswar *et al.*, 2006). Homologous sequences that have a similarity score of at least 25% to 50% are taken into consideration. The technique relies on that fact that changes in the amino acid sequence of a protein will result in fewer structure variations hence, evolutionary-related proteins possess similar structures (Vyas *et al.*, 2012). This method is therefore used as a helpful tool in drug design and function annotation, prediction of ligand binding sites and in analyzing specificity of potential substrates. It is a multi-step process that first involves the identification of a template, or in other words, a known protein structure whose sequence similarity is closest to the target sequence. Various approaches may be undertaken by satisfying spatial constraints via the assumption that the correlating distances of the target protein and template are the same or by using the conserved loops and regions of the protein with a degree of flexibility. Loops and side chains are predicted and constructed where the former is based on a variety of factors such as loop length, conformational space and backbone construction, while the latter is modeled by *ab initio* simulations, side chains ‘borrowed’ from the template or a combination of the two strategies. Finally, model refinement entails the use of molecular mechanics or dynamics to correct errors that may have occurred during model building by making changes in the target sequence through the insertion, deletion or substitution of amino acids. Accuracy of the backbone conformation and corresponding side chains are greatly improved whilst relaxing the structure in a realistic “all-atom force field”. A variety of software are currently available that automate the processes of building a complete three-dimensional structure with the most popular, accurate and reliable among the sea of available tools being MODELLER and SwissModel (Eswar *et al.*, 2006; Vyas *et al.*, 2012; Webb and Sali, 2014).

## 2.20 Protein Simulations

Important and specific questions that experiments may find difficult to answer on actual systems can be answered by protein simulations (Karplus and McCammon, 2002). Computer simulations basically act as a bridge between the prediction and verification of experiments and in the testing and/or optimization of models. Laboratory work should cover these factors to some extent but occasional incidents do arise which are relatively too complex for theoretical logic or just cannot be observed experimentally. The characterization of atomic features that include dynamics, mobility, energetics and structure of a biomolecular system are examples of factors limited by experimental measuring techniques (Van Gunsteren *et al.*, 1994). Molecular Dynamics (MD) and Molecular Mechanics (MM) stand as powerful tools in such cases for the structural analysis and prediction of complex biological macromolecules.

### 2.20.1 Molecular Mechanics

Molecular mechanics or Force-Field Methods is almost always applied to Computational Structure-Based Drug Discovery (CSBDD) via the use of classical or Newtonian Laws (Vanommeslaeghe and Guvench, 2014). The technique reproduces molecular behaviour by solving energy equations that demonstrate the total energy within a molecular model demonstrated as a collection of mass or atoms and bonds in the form of balls and springs respectively (Lewars, 2016; Jakubowski, 2019). The interactions between these masses are described as potential energy functions in the following equation (5), which is the sum of bonded and unbonded energy (Bowen and Allinger, 1991; Sherill, 2016):

$$E_{steric\ energy} = E_{str} + E_{bend} + E_{tors} + E_{vdw} + E_{el} + E_{cross\ steric} \quad (5)$$

This includes stretching, bending, torsion, van der Waals, electrostatics and cross terms (Jakubowski, 2019; Sherill, 2016). The energy of the system can therefore be calculated once the

energy needed to stretch and bend the ‘balls’ and ‘springs’ is known. Additionally, geometry optimization of the molecule can be calculated by assuming that any shift from its ‘ideal’ geometry results in energy increase. The goal is to therefore change the geometry of the molecule so as to reach its lowest energy potential (Lewars, 2016; Bowen and Allinger, 1991).

### ***2.20.2 Molecular Dynamics***

Molecular Dynamics (MD) looks at how solvent plays a determining role in the classical equations of motions, Newton’s Second Law of motion to be precise. It takes particular note on the motions acting within a molecule as a function of time subsequent to the application of energy in the form of high temperatures to the system (Karplus and McCammon, 2002; Jakubowski, 2019). Molecular mechanics calculations are often used to determine the energy of the system in a time-dependent manner (Young, 2001). The time factor is however the major limitation of MD simulations as they require short time steps to sample the interconversions of the protein conformational changes that take place within the system (Jakubowski, 2019; Dror, 2017). Although major advances in the technology have greatly improved computer time, this still remains a challenge.

Additionally, structure-based drug design approaches are frequently used by numerous pharmaceutical companies for the development of novel drugs. A number of important drugs such as the HIV protease inhibitors saquinavir, amprenavir (Agesenase®) and nelfinavir (Viracept®) as well as flu antiviral drugs such as zanamivir (Relenza®) and oseltamivir (Tamiflu®) have come into existence through the contribution of structural information. The birthing of these discoveries through the applications of protein crystallography has not only broadened the impact of protein structure in understanding medicine and biology, but has also opened a can of worms into solving the many economic and social struggles of various societies

of our time (Blundell, 2017). Most importantly, it has successfully assisted in unlocking avenues for developing treatment for numerous disease disorders that are rare, genetic or infectious of which the latter is currently a major concern in several developing countries of the world today.

## References

- Addisu A, Adriaensen W, Balew A, Asfaw M, Diro E, Djirmay AG, Gebree D, Seid G, Begashaw H and Harries AD (2019). Neglected tropical diseases and the sustainable development goals: an urgent call for action from the front line. *BMJ global health* **4**(1): e001334.
- Adenowo AF, Oyinloye BE, Ogunyinka BI and Kappo AP (2015). Impact of human schistosomiasis in sub-Saharan Africa. *Brazilian Journal of Infectious Diseases* **19**(2): 196-205.
- Adenowo AF. (2018). *Molecular cloning, recombinant protein expression and structural characterization of a universal stress G4LZI3 protein from Schistosoma mansoni*. University of Zululand.
- Adiyaman R and McGuffin LJ (2019). Methods for the Refinement of Protein Structure 3D Models. *International journal of molecular sciences* **20**(9): 2301.
- Ali M, Said Z, Rahman R, Chor A, Basri M and Salleh A (2013). Capillary-Seeding Crystallization and Preliminary Crystallographic Analysis of a Solvent-Tolerant Elastase from *Pseudomonas aeruginosa* Strain K. *International journal of molecular sciences* **14**(9): 17608-17617.
- Al-Mashat F, Sibiany A, Radwi A, Bahadur Y, Al-Radi A, Meir H and Ahmed GE (2001). Rectal cancer associated with schistosomiasis: report of two cases and review of the literature. *Annals of Saudi Medicine* **21**(1-2): 65-67.
- Alsaqabi SM and Lotfy WM (2014). Praziquantel: A review. *Veterinary Science & Technology* **5**(5): 1.

- Appleton C and Kvalsvig J (2006). A school-based helminth control programme successfully implemented in KwaZulu-Natal. *Southern African Journal of Epidemiology and Infection* **21**(2): 55-67.
- Aragon AD, Imani RA, Blackburn VR and Cunningham C (2008). Microarray based analysis of temperature and oxidative stress induced messenger RNA in *Schistosoma mansoni*. *Molecular and biochemical parasitology* **162**(2): 134-141.
- Aragon AD, Imani RA, Blackburn VR, Cupit PM, Melman SD, Goronga T, Webb T, Loker ES and Cunningham C (2009). Towards an understanding of the mechanism of action of praziquantel. *Molecular and biochemical parasitology* **164**(1): 57-65.
- Araújo MI, Hoppe BS, Medeiros Jr M and Carvalho EM (2004). *Schistosoma mansoni* infection modulates the immune response against allergic and auto-immune diseases. *Memórias do Instituto Oswaldo Cruz* **99**: 27-32.
- Aruleba R, Adekiya T, Oyinloye B, Masamba P, Mbatha L, Pretorius A and Kappo A (2018). PZQ therapy: how close are we in the development of effective alternative anti-schistosomal drugs? *Infectious disorders drug targets*.
- Augusto RDC, Mello-Silva D and Correa CC (2018). Phytochemical Molluscicides and Schistosomiasis: What We Know and What We Still Need to Learn. *Veterinary sciences* **5**(4): 94.
- Banerjee A, Adolph RS, Gopalakrishnapai J, Kleinboelting S, Emmerich C, Steegborn C and Visweswariah SS (2015). A universal stress protein (USP) in mycobacteria binds cAMP. *Journal of Biological Chemistry* **290**(20): 12731-12743.
- Bangera M, Panigrahi R, Sagurthi S, Savithri H and Murthy M (2015). Structural and functional analysis of two universal stress proteins YdaA and YnaF from *Salmonella typhimurium*:

- possible roles in microbial stress tolerance. *Journal of structural biology* **189**(3): 238-250.
- Barsoum RS, Esmat G and El-Baz T (2013). Human schistosomiasis: clinical perspective: review. *J Adv Res* **4**(5): 433-444.
- Bergfors T (2003). Seeds to crystals. *Journal of structural biology* **142**(1): 66-76.
- Bergfors T (2007). Succeeding with seeding: some practical advice *Evolving Methods for Macromolecular Crystallography*: Springer pp. 1-10.
- Berry A, Moné H, Iriart X, Mouahid G, Aboo O, Boissier J, Fillaux J, Cassaing S, Debuissou C and Valentin A (2014). Schistosomiasis haematobium, Corsica, France. *Emerging Infectious Diseases* **20**(9): 1595.
- Billeter M (1992). Comparison of protein structures determined by NMR in solution and by X-ray diffraction in single crystals. *Quarterly reviews of biophysics* **25**(3): 325-377.
- Blundell TL (2017). Protein crystallography and drug discovery: recollections of knowledge exchange between academia and industry. *IUCrJ* **4**(4): 308-321.
- Boes N, Schreiber K, Härtig E, Jaensch L and Schobert M (2006). The *Pseudomonas aeruginosa* universal stress protein PA4352 is essential for surviving anaerobic energy stress. *Journal of bacteriology* **188**(18): 6529-6538.
- Boissier J, Moné H, Mitta G, Bargues MD, Molyneux D and Mas-Coma S (2015). Schistosomiasis reaches Europe. *The Lancet infectious diseases* **15**(7): 757-758.
- Bowen JP and Allinger NL (1991). Molecular mechanics: The art and science of parameterization. *Reviews in computational chemistry*: 81-97.

Bradley D (2011). The Universal Stress Proteins of Bacteria.

Brandt M (2000). Protein structure determination. [https://www.rose-hulman.edu/~brandt/Chem330/Protein\\_structure\\_determ.pdf](https://www.rose-hulman.edu/~brandt/Chem330/Protein_structure_determ.pdf).

Brindley PJ and Hotez PJ. (2013). Break out: urogenital schistosomiasis and *Schistosoma haematobium* infection in the post-genomic era: Public Library of Science.

Broadwith P (2017). Explainer: What is cryo-electron microscopy. <https://www.chemistryworld.com/news/explainer-what-is-cryo-electron-microscopy/3008091.article>.

Butrous G (2019). Schistosome infection and its effect on pulmonary circulation. *Global cardiology science & practice* **2019**(1).

Caldas IR, Campi-Azevedo AC, Oliveira LFA, Silveira AMS, Oliveira RC and Gazzinelli G (2008). Human schistosomiasis mansoni: immune responses during acute and chronic phases of the infection. *Acta tropica* **108**(2-3): 109-117.

Capron A, Riveau G, Capron M and Trottein F (2005). Schistosomes: the road from host–parasite interactions to vaccines in clinical trials. *Trends in parasitology* **21**(3): 143-149.

Carroni M and Saibil HR (2016). Cryo electron microscopy to determine the structure of macromolecular complexes. *Methods* **95**: 78-85.

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

- Chilingaryan Z, Yin Z and Oakley AJ (2012). Fragment-based screening by protein crystallography: successes and pitfalls. *International journal of molecular sciences* **13**(10): 12857-12879.
- Chitsulo L, Engels D, Montresor A and Savioli L (2000). The global status of schistosomiasis and its control. *Acta tropica* **77**(1): 41-51.
- Cioli D and Pica-Mattoccia L (2003). Praziquantel. *Parasitology research* **90**(1): S3-S9.
- Cioli D, Pica-Mattoccia L and Archer S (1993). Drug resistance in schistosomes. *Parasitology Today* **9**(5): 162-166.
- Cioli D, Pica-Mattoccia L, Basso A and Guidi A (2014). Schistosomiasis control: praziquantel forever? *Molecular and biochemical parasitology* **195**(1): 23-29.
- Clerinx J and Van Gompel A (2011). Schistosomiasis in travellers and migrants. *Travel Medicine and Infectious Disease* **9**(1): 6-24.
- Colley DG, Bustinduy AL, Secor WE and King CH (2014). Human schistosomiasis. *The Lancet* **383**(9936): 2253-2264.
- Collins Iii JJ, King RS, Cogswell A, Williams DL and Newmark PA (2011). An atlas for *Schistosoma mansoni* organs and life-cycle stages using cell type-specific markers and confocal microscopy. *PLoS neglected tropical diseases* **5**(3): e1009.
- Conlon CP (2005). Schistosomiasis. *Medicine* **33**(8).
- Consortium SG (2008). Protein production and purification. *Nature methods* **5**(2): 135.

- Coon DR (2005). Schistosomiasis: overview of the history, biology, clinicopathology, and laboratory diagnosis. *Clinical Microbiology Newsletter* **27**(21): 163-168.
- Cowtan K (2001). Phase Problem in X-ray Crystallography, and Its Solution. *e LS*.
- Da Silva VBR, Campos BRKL, De Oliveira JF, Decout J-L and De Lima MDCA (2017). Medicinal chemistry of antischistosomal drugs: Praziquantel and oxamniquine. *Bioorganic & medicinal chemistry* **25**(13): 3259-3277.
- De Oliveira RB, Senger MR, Vasques LM, Gasparotto J, Dos Santos JPA, De Bittencourt Pasquali MA, Moreira JCF, Silva Jr FP and Gelain DP (2013). Schistosoma mansoni infection causes oxidative stress and alters receptor for advanced glycation endproduct (RAGE) and tau levels in multiple organs in mice. *International journal for parasitology* **43**(5): 371-379.
- Dessau MA and Modis Y (2011). Protein crystallization for X-ray crystallography. *JoVE (Journal of Visualized Experiments)* (47): e2285.
- Diez A, Gustavsson N and Nyström 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.
- Dimaio F (2017). Rosetta structure prediction as a tool for solving difficult molecular replacement problems *Protein Crystallography*: Springer pp. 455-466.
- Doenhoff MJ, Cioli D and Utzinger J (2008). Praziquantel: mechanisms of action, resistance and new derivatives for schistosomiasis. *Current Opinion in Infectious Diseases* **21**(6): 659-667.

Dror R (2017). Molecular Dynamics Simulation.  
<https://web.stanford.edu/class/cs279/lectures/lecture4.pdf>.

Drumm JE, Mi K, Bilder P, Sun M, Lim J, Bielefeldt-Ohmann H, Basaraba R, So M, Zhu G and Tufariello JM (2009). Mycobacterium tuberculosis universal stress protein Rv2623 regulates bacillary growth by ATP-Binding: requirement for establishing chronic persistent infection. *PLoS Pathogens* **5**(5): e1000460.

Earl LA, Falconieri V, Milne JL and Subramaniam S (2017). Cryo-EM: beyond the microscope. *Current opinion in structural biology* **46**: 71-78.

Ebrahim G. (2009). Neglected Tropical Diseases: Oxford University Press.

Edwards AM, Arrowsmith CH, Christendat D, Dharamsi A, Friesen JD, Greenblatt JF and Vedadi M (2000). Protein production: feeding the crystallographers and NMR spectroscopists. *Nature Structural & Molecular Biology* **7**(11s): 970.

El Ridi RA and Tallima Ha-M (2013). Novel therapeutic and prevention approaches for schistosomiasis. *Journal of Advanced Research* **4**(5): 467-478.

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

Elhosseiny NM, Amin MA, Yassin AS and Attia AS (2015). Acinetobacter baumannii universal stress protein A plays a pivotal role in stress response and is essential for pneumonia and sepsis pathogenesis. *International Journal of Medical Microbiology* **305**(1): 114-123.

Ennifar E (2013). X-ray crystallography as a tool for mechanism-of-action studies and drug discovery. *Current pharmaceutical biotechnology* **14**(5): 537-550.

- Eswar N, Webb B, Marti-Renom MA, Madhusudhan M, Eramian D, Shen MY, Pieper U and Sali A (2006). Comparative protein structure modeling using Modeller. *Current protocols in bioinformatics* **15**(1): 5.6. 1-5.6. 30.
- Evans P and McCoy A (2008). An introduction to molecular replacement. *Acta Crystallographica Section D: Biological Crystallography* **64**(1): 1-10.
- Fallon P, Tao L, Ismail M and Bennett J (1996). Schistosome resistance to praziquantel: fact or artifact? *Parasitology Today* **12**(8): 316-320.
- Feasey N, Wansbrough-Jones M, Mabey DC and Solomon AW (2009). Neglected tropical diseases. *British Medical Bulletin* **93**(1): 179-200.
- Ferrari TCA and Moreira PRR (2011). Neuroschistosomiasis: clinical symptoms and pathogenesis. *The Lancet Neurology* **10**(9): 853-864.
- Flick K and Kaiser P. (2012). *Protein degradation and the stress response*. Paper presented at the Seminars in Cell and Developmental Biology.
- Fogarty KH, Zhang W, Grigsby IF, Johnson JL, Chen Y, Mueller JD and Mansky LM (2011). New insights into HTLV-1 particle structure, assembly, and Gag-Gag interactions in living cells. *Viruses* **3**(6): 770-793.
- Freestone P, Nyström 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.
- Gandhi K, Kumar A, Gosewade S, Kaushik R and Lal D (2018). X-Ray Crystallography and Its Applications in Dairy Science: A Review. *Research & Reviews: Journal of Dairy Science and Technology* **2**(1): 21-33.

- Gause WC, Urban Jr JF and Stadecker MJ (2003). The immune response to parasitic helminths: insights from murine models. *Trends in Immunology* **24**(5): 269-277.
- Gavira JA (2016). Current trends in protein crystallization. *Archives of biochemistry and biophysics* **602**: 3-11.
- Gobert GN, Nawaratna SK, Harvie M, Ramm GA and Mcmanus DP (2015). An ex vivo model for studying hepatic schistosomiasis and the effect of released protein from dying eggs. *PLoS neglected tropical diseases* **9**(5): e0003760.
- Gomes CS, Izar B, Pazan F, Mohamed W, Mraheil MA, Mukherjee K, Billion A, Aharonowitz Y, Chakraborty T and Hain T (2011). Universal stress proteins are important for oxidative and acid stress resistance and growth of *Listeria monocytogenes* EGD-e in vitro and in vivo. *PloS one* **6**(9): e24965.
- Gouveia MJ, Brindley PJ, Gärtner F, Costa JM and Vale N (2018). Drug repurposing for schistosomiasis: combinations of drugs or biomolecules. *Pharmaceuticals* **11**(1): 15.
- Grosse-Kunstleve R and Schneider T (2007). Structure Determination in Isomorphous Replacement and Anomalous Diffraction Experiments. *Methods in Molecular Biology* **364**: 197 - 214.
- Gryseels B, Polman K, Clerinx J and Kestens L (2006). Human schistosomiasis. *The Lancet* **368**(9541): 1106-1118.
- Guss JM and King GF (2000). Macromolecular structure determination: comparison of crystallography and NMR. *Encyclopedia of Life Sciences*.

- Gustavsson N, Diez A 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.
- Healthcare G (2015). Protein Purification Methods. *GE Healthcare Bio-Sciences AB, Uppsala Sweden*: 1-12.
- Heermann R, Weber A, Mayer B, Ott M, Hauser E, Gabriel G, Pirch T and Jung K (2009). The universal stress protein UspC scaffolds the KdpD/KdpE signaling cascade of *Escherichia coli* under salt stress. *Journal of molecular biology* **386**(1): 134-148.
- Hingley-Wilson S, Loughheed K, Ferguson K, Leiva S and Williams H (2010). Individual *Mycobacterium tuberculosis* universal stress protein homologues are dispensable in vitro. *Tuberculosis* **90**(4): 236-244.
- Ilari A and Savino C (2008). Protein Structure Determination by X-Ray Crystallography. In Keith JM (Ed.), *Bioinformatics: Data, Sequence Analysis and Evolution*. Totowa, NJ: Humana Press pp. 63-87.
- Inobaya MT, Olveda RM, Chau TN, Olveda DU and Ross AG (2014). Prevention and control of schistosomiasis: a current perspective. *Research and reports in tropical medicine* **2014**(5): 65.
- Isokpehi RD, Mahmud O, Mbah AN, Simmons SS, Avelar L, Rajnarayanan RV, Udensi UK, Ayensu WK, Cohly HH and Brown SD (2011). Developmental regulation of genes encoding universal stress proteins in *Schistosoma mansoni*. *Gene regulation and systems biology* **5**: GRSB. S7491.
- Jacques DA and Trehwella J (2010). Small-angle scattering for structural biology—Expanding the frontier while avoiding the pitfalls. *Protein science* **19**(4): 642-657.

- Jaeger J (2004). Macromolecular Structure Determination by X-ray Crystallography. *Nature Encyclopedia of Life Sciences*.
- Jakubowski H (2019). Introduction to Molecular Mechanics and Molecular Dynamics. [https://chem.libretexts.org/Courses/CSU\\_Chico/CSU\\_Chico%3A\\_CHEM\\_451 -  
Biochemistry I/CHEM\\_451\\_Test/03%3A\\_Lipid\\_Structure/3.3%3A\\_Dynamics\\_of Mem  
brane Lipids/Molecular Mechanics and Dynamics/A. Introduction to Molecular Mec  
hanics and Molecular Dynamics](https://chem.libretexts.org/Courses/CSU_Chico/CSU_Chico%3A_CHEM_451_-_Biochemistry_I/CHEM_451_Test/03%3A_Lipid_Structure/3.3%3A_Dynamics_of_Membrane_Lipids/Molecular_Mechanics_and_Dynamics/A_Introduction_to_Molecular_Mechanics_and_Molecular_Dynamics).
- Jenkins S, Hewitson J, Jenkins G and Mountford A (2005). Modulation of the host's immune response by schistosome larvae. *Parasite immunology* **27**(10-11): 385-393.
- Johnson C and Appleton C (2005). Urban schistosomiasis transmission in Pietermaritzburg, South Africa. *Southern African Journal of Epidemiology and Infection* **20**(3): 103-110.
- Jolly ER, Chin C-S, Miller S, Bahgat MM, Lim K, Derisi J and Mckerrow JH (2007). Gene expression patterns during adaptation of a helminth parasite to different environmental niches. *Genome Biology* **8**(4): R65.
- Jung YJ, Melencion SMB, Lee ES, Park JH, Alinapon CV, Oh HT, Yun D-J, Chi YH and Lee SY (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**: 1141.
- Kabuyaya M, Chimbari MJ, Manyangadze T and Mukaratirwa S (2017). Efficacy of praziquantel on *Schistosoma haematobium* and re-infection rates among school-going children in the Ndumo area of uMkhanyakude district, KwaZulu-Natal, South Africa. *Infectious diseases of poverty* **6**(1): 83.

- Karplus M and Mccammon JA (2002). Molecular dynamics simulations of biomolecules. *Nature Structural & Molecular Biology* **9**(9): 646.
- Khajepour MYH. (2012). *Amélioration et automatisation des étapes de préparation des cristaux de protéines à la diffraction aux rayons X*. Grenoble.
- 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.
- Kim Y, Bigelow L, Borovilos M, Dementieva I, Duggan E, Hatzos C, Joachimiak G, Li H, Maltseva N and Mulligan R (2008). High-throughput protein purification for X-ray crystallography and NMR. *Advances in protein chemistry and structural biology* **75**: 85-105.
- Kjetland EF, Leutscher PD and Ndhlovu PD (2012). A review of female genital schistosomiasis. *Trends in parasitology* **28**(2): 58-65.
- Knaul FM, Rodriguez NM, Arreola-Ornelas H and Olson JR (2019). Cervical cancer: lessons learned from neglected tropical diseases. *The Lancet Global Health* **7**(3): e299-e300.
- Koning RI, Koster AJ and Sharp TH (2018). Advances in cryo-electron tomography for biology and medicine. *Annals of Anatomy-Anatomischer Anzeiger* **217**: 82-96.
- Kvint K, Nachin L, Diez A and Nyström T (2003). The bacterial universal stress protein: function and regulation. *Current opinion in microbiology* **6**(2): 140-145.
- Lamb AL, Kappock TJ and Silvaggi NR (2015). You are lost without a map: Navigating the sea of protein structures. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* **1854**(4): 258-268.

- Lambertucci JR (2010). Acute schistosomiasis mansoni: revisited and reconsidered. *Memórias do Instituto Oswaldo Cruz* **105**(4): 422-435.
- Lambertucci JR, Rayes AaM, Serufo JC, Gerspacher-Lara R, Brasileiro-Filho G, Teixeira R, Antunes CMDf, Goes AM and Coelho PMZ (1998). Schistosomiasis and associated infections. *Memórias do Instituto Oswaldo Cruz* **93**: 135-139.
- Lammie PJ, Fenwick A and Utzinger J (2006). A blueprint for success: integration of neglected tropical disease control programmes. *Trends in parasitology* **22**(7): 313-321.
- Leunissen M (2001). Protein crystallization. *Nijmegen University* **44**.
- Lewars E (2003). Computational chemistry. *Introduction to the theory and applications of molecular and quantum mechanics*: 318.
- Lewars EG (2016). An Outline of What Computational Chemistry Is All About *Computational Chemistry*: Springer pp. 1-8.
- Lieberman RL, Peek ME and Watkins JD (2013). Determination of soluble and membrane protein structures by X-ray crystallography *Electron crystallography of soluble and membrane proteins*: Springer pp. 475-493.
- Liu W-T, Karavolos MH, Bulmer DM, Allaoui A, Hormaeche RDCE, Lee JJ and Khan CA (2007). Role of the universal stress protein UspA of Salmonella in growth arrest, stress and virulence. *Microbial 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* **63**(15): 5593-5606.

- Lyumkis D (2019). Challenges and opportunities in cryo-EM single-particle analysis. *Journal of Biological Chemistry* **294**(13): 5181-5197.
- Mackey TK and Liang BA (2012). Rebalancing brain drain: exploring resource reallocation to address health worker migration and promote global health. *Health Policy* **107**(1): 66-73.
- Magaisa K, Taylor M, Kjetland EF and Naidoo PJ (2015). A review of the control of schistosomiasis in South Africa. *South African Journal of Science* **111**(11-12): 1-6.
- Masamba P, Adenowo A, Oyinloye B and Kappo A (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.
- Mattsson J (2011). Documentation of protein crystallization techniques and characterization of crystals.
- Mbah AN, Mahmud O, Awofolu OR and Isokpehi RD (2013). Inferences on the biochemical and environmental regulation of universal stress proteins from Schistosomiasis parasites. *Advances and applications in bioinformatics and chemistry: AABC* **6**: 15.
- Mccoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC and Read RJ (2007). Phaser crystallographic software. *Journal of applied crystallography* **40**(4): 658-674.
- Mccullough F, Gayral P, Duncan J and Christie J (1980). Molluscicides in schistosomiasis control. *Bulletin of the World Health Organization* **58**(5): 681.
- Mcmanus DP, Dunne DW, Sacko M, Utzinger J, Vennervald BJ and Zhou X-N (2018). Schistosomiasis (Primer). *Nature Reviews: Disease Primers*.

- McPherson A and Gavira JA (2014). Introduction to protein crystallization. *Acta Crystallographica Section F: Structural Biology Communications* **70**(1): 2-20.
- Melencion SMB, Chi YH, Pham TT, Paeng SK, Wi SD, Lee C, Ryu SW, Koo SS and Lee SY (2017). RNA chaperone function of a universal stress protein in Arabidopsis confers enhanced cold stress tolerance in plants. *International journal of molecular sciences* **18**(12): 2546.
- Melman SD, Steinauer ML, Cunningham C, Kubatko LS, Mwangi IN, Wynn NB, Mutuku MW, Karanja DM, Colley DG and Black CL (2009). Reduced susceptibility to praziquantel among naturally occurring Kenyan isolates of *Schistosoma mansoni*. *PLoS neglected tropical diseases* **3**(8): e504.
- Merrifield M, Hotez PJ, Beaumier CM, Gillespie P, Strych U, Hayward T and Bottazzi ME (2016). Advancing a vaccine to prevent human schistosomiasis. *Vaccine* **34**(26): 2988-2991.
- Mertens HD and Svergun DI (2010). Structural characterization of proteins and complexes using small-angle X-ray solution scattering. *Journal of structural biology* **172**(1): 128-141.
- Messerschmidt A (2007). *X-ray crystallography of biomacromolecules: a practical guide*: John Wiley & Sons.
- Mitra AK and Mawson AR (2017). Neglected tropical diseases: epidemiology and global burden. *Tropical medicine and infectious disease* **2**(3): 36.
- Mojsiewicz-Pieńkowska K (2014). Size Exclusion Chromatography a Useful Technique For Speciation Analysis of Polydimethylsiloxanes *Green Chromatographic Techniques*: Springer pp. 181-202.

- Morel CM (2003). Neglected diseases: under-funded research and inadequate health interventions. *EMBO reports* **4**(S1): S35-S38.
- 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.
- Náray-Szabó G and Perczel A (2014). Protein structure and dynamics. *International Journal of Terraspace Science and Engineering* **6**(1): 7-16.
- Nascimento LS, Ebert JP, Teles FV and Cruz DB (2016). Update on *Schistosomiasis mansoni*: Epidemiology, Pathogenesis, Diagnosis, Treatment and Prevention. *JSM Gastroenterology and Hepatology* **4**(4): 1068.
- Negrão-Corrêa D, Mattos A, Pereira C, Martins-Souza R and Coelho P (2012). Interaction of *Schistosoma mansoni* sporocysts and hemocytes of *Biomphalaria*. *Journal of parasitology research* **2012**.
- Nour NM (2010). Schistosomiasis: health effects on women. *Reviews in Obstetrics and Gynecology* **3**(1): 28.
- Nyström T and Neidhardt FC (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.
- Nyström T and Neidhardt FC (1994). Expression and role of the universal stress protein, UspA, of *Escherichia coli* during growth arrest. *Molecular microbiology* **11**(3): 537-544.
- Nyström T and Neidhardt FC (1996). Effects of overproducing the universal stress protein, UspA, in *Escherichia coli* K-12. *Journal of bacteriology* **178**(3): 927-930.

Olveda DU, Li Y, Olveda RM, Lam AK, Chau TN, Harn DA, Williams GM, Gray DJ and Ross AG (2013). Bilharzia: pathology, diagnosis, management and control. *Tropical medicine & surgery* **1**(4).

Olveda DU, Olveda RM, Mcmanus DP, Cai P, Chau TN, Lam AK, Li Y, Harn DA, Vinluan ML and Ross AG (2014). The chronic enteropathogenic disease schistosomiasis. *International Journal of Infectious Diseases* **28**: 193-203.

Oyinloye B, Adenowo A and Kappo A (2015). Reactive oxygen species, apoptosis, antimicrobial peptides and human inflammatory diseases. *Pharmaceuticals* **8**(2): 151-175.

Patterson AL (1934). A Fourier series method for the determination of the components of interatomic distances in crystals. *Physical Review* **46**(5): 372.

Pearce EJ and Macdonald AS (2002). The immunobiology of schistosomiasis. *Nature Reviews Immunology* **2**(7): 499.

Pearson WR (2013). An introduction to sequence similarity (“homology”) searching. *Current protocols in bioinformatics* **42**(1): 3.1. 1-3.1. 8.

Persson Ö, Valadi Å, Nyström T and Farewell A (2007). Metabolic control of the Escherichia coli universal stress protein response through fructose-6-phosphate. *Molecular microbiology* **65**(4): 968-978.

Petoukhov MV and Svergun DI (2006). Joint use of small-angle X-ray and neutron scattering to study biological macromolecules in solution. *European Biophysics Journal* **35**(7): 567-576.

Physicslibretexts (2019). Young's Double-Slit Interference. [https://phys.libretexts.org/Bookshelves/University\\_Physics/Book%3A\\_University\\_Physics](https://phys.libretexts.org/Bookshelves/University_Physics/Book%3A_University_Physics)

- Protasio AV, Tsai IJ, Babbage A, Nichol S, Hunt M, Aslett MA, De Silva N, Velarde GS, Anderson TJ and Clark RC (2012). A systematically improved high quality genome and transcriptome of the human blood fluke *Schistosoma mansoni*. *PLoS neglected tropical diseases* **6**(1): e1455.
- Pusey ML, Liu Z-J, Tempel W, Praissman J, Lin D, Wang B-C, Gavira JA and Ng JD (2005). Life in the fast lane for protein crystallization and X-ray crystallography. *Progress in biophysics and molecular biology* **88**(3): 359-386.
- Remme JH, Blas E, Chitsulo L, Desjeux PM, Engers HD, Kanyok TP, Kayondo JFK, Kioy DW, Kumaraswami V and Lazdins JK (2002). Strategic emphases for tropical diseases research: a TDR perspective. *Trends in parasitology* **18**(10): 421-426.
- Rollinson D, Knopp S, Levitz S, Stothard JR, Tchuente L-aT, Garba A, Mohammed KA, Schur N, Person B and Colley DG (2013). Time to set the agenda for schistosomiasis elimination. *Acta tropica* **128**(2): 423-440.
- Russo Krauss I, Merlino A, Vergara A and Sica F (2013). An overview of biological macromolecule crystallization. *International journal of molecular sciences* **14**(6): 11643-11691.
- Sayre D (2002). X-ray crystallography: The past and present of the phase problem. *Structural Chemistry* **13**(1): 81-96.
- Secor WE and Montgomery SP (2015). Something old, something new: is praziquantel enough for schistosomiasis control? *Future medicinal chemistry* **7**(6): 681-684.

Sherill D (2016). Introduction to Molecular Mechanics.  
<http://vergil.chemistry.gatech.edu/courses/chem6485/pdf/molmech-lecture.pdf>.

Shih H-W (2009). Fundamentals of Computational Chemistry: A Guide for Organic Chemists.  
[https://www.princeton.edu/chemistry/macmillan/group-meetings/HWS\\_computational.pdf](https://www.princeton.edu/chemistry/macmillan/group-meetings/HWS_computational.pdf).

Shokry AM, Al-Karim S, Ramadan A, Gadallah N, Al Attas SG, Sabir JS, Hassan SM, Madkour MA, Bressan R and Mahfouz M (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 biologiques* **337**(2): 86-94.

Sibomana L. (2010). *Association of Schistosomiasis prevalence with socio-demographic status measures in sub-Saharan Africa*. University of Pittsburgh.

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

Siegele DA (2005). Universal stress proteins in Escherichia coli. *Journal of bacteriology* **187**(18): 6253-6254.

Siqueira LDP, Fontes DaF, Aguilera CSB, Timóteo TRR, Ângelos MA, Silva LCPBB, De Melo CG, Rolim LA, Da Silva RMF and Neto PJR (2017). Schistosomiasis: Drugs used and treatment strategies. *Acta tropica* **176**: 179-187.

Smatanová IK (2002). Crystallization of biological macromolecules. *Materials Structure* **9**(1): 14-15.

Sousa MC and McKay DB (2001). Structure of the universal stress protein of Haemophilus influenzae. *Structure* **9**(12): 1135-1141.

- Spriestersbach A, Kubicek J, Schäfer F, Block H and Maertens B (2015). Purification of His-tagged proteins *Methods in Enzymology* (Vol. 559.): Elsevier pp. 1-15.
- Striegel AM (2008). Size-exclusion chromatography: smaller, faster, multi-detection, and multi-dimensions. *Analytical and bioanalytical chemistry* **390**(1): 303-305.
- Taylor GL (2003). The phase problem. *Acta Crystallographica Section D: Biological Crystallography* **59**(11): 1881-1890.
- Taylor GL (2010). Introduction to phasing. *Acta Crystallographica Section D: Biological Crystallography* **66**(4): 325-338.
- Terwilliger TC, Dimaio F, Read RJ, Baker D, Bunkóczy G, Adams PD, Grosse-Kunstleve RW, Afonine PV and Echols N (2012). phenix. mr\_rosetta: molecular replacement and model rebuilding with Phenix and Rosetta. *Journal of structural and functional genomics* **13**(2): 81-90.
- Thomas CM, Fitzsimmons CM, Dunne DW and Timson DJ (2015). Comparative biochemical analysis of three members of the *Schistosoma mansoni* TAL family: Differences in ion and drug binding properties. *Biochimie* **108**: 40-47.
- Tkaczuk KL, A. 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.
- Trainor-Moss S and Mutapi F. (2016). Schistosomiasis therapeutics: whats in the pipeline? : Taylor & Francis.
- Trewhella J (2016). Small-angle scattering and 3D structure interpretation. *Current opinion in structural biology* **40**: 1-7.

Uah (2004). Structure Determination. *New Science Press*.  
[http://www3.uah.es/farmamol/New\\_Science\\_Press/nsp-protein-5.pdf](http://www3.uah.es/farmamol/New_Science_Press/nsp-protein-5.pdf)

Utzinger J, N'goran EK, Caffrey CR and Keiser J (2011). From innovation to application: social–ecological context, diagnostics, drugs and integrated control of schistosomiasis. *Acta tropica* **120**: S121-S137.

Utzinger J, Raso G, Brooker S, De Savigny D, Tanner M, Ørnbjerg N, Singer B and N'goran E (2009). Schistosomiasis and neglected tropical diseases: towards integrated and sustainable control and a word of caution. *Parasitology* **136**(13): 1859-1874.

Van Gunsteren W, Luque F, Timms D and Torda A (1994). Molecular mechanics in biology: from structure to function, taking account of solvation. *Annual review of biophysics and biomolecular structure* **23**(1): 847-863.

Vanommeslaeghe K and Guvench O (2014). Molecular mechanics. *Current pharmaceutical design* **20**(20): 3281-3292.

Vasudevan DM, Sreekumari S and Vaidyanathan K (2013). *Textbook of biochemistry for medical students*: JP Medical Ltd.

Vekilov PG (2016). Nucleation of protein crystals. *Progression in Crystal Growth and Characterization of Materials* **62**: 136-154.

Vénien-Bryan C, Li Z, Vuillard L and Boutin JA (2017). Cryo-electron microscopy and X-ray crystallography: complementary approaches to structural biology and drug discovery. *Acta Crystallographica Section F: Structural Biology Communications* **73**(4): 174-183.

Vestergaard B (2016). Analysis of biostructural changes, dynamics, and interactions–small-angle X-ray scattering to the rescue. *Archives of biochemistry and biophysics* **602**: 69-79.

- Vollmer AC and Bark SJ (2018). Twenty-five years of investigating the universal stress protein: function, structure, and applications *Advances in Applied Microbiology* (Vol. 102.): Elsevier pp. 1-36.
- Vyas V, Ukawala R, Ghate M and Chintha C (2012). Homology modeling a fast tool for drug discovery: current perspectives. *Indian journal of pharmaceutical sciences* **74**(1): 1.
- Wainwright E, Evans D, Rotondo L, Pou B, Yevstigneyeva V, Zoerhoff KL, Ottesen EA and Reithinger R (2019). The Elimination of Neglected Tropical Diseases (NTDs): A Case Study Exemplifying How Foreign Assistance Funding Can Be Catalytic in Reducing the Burden of Major Global Health Conditions. *Clinical Infectious Diseases*.
- Wang W, Qin Z, Zhu D, Wei Y, Li S and Duan L (2016). Synthesis, bioactivity evaluation, and toxicity assessment of novel salicylanilide ester derivatives as cercaricides against *Schistosoma japonicum* and molluscicides against *Oncomelania hupensis*. *Antimicrobial agents and chemotherapy* **60**(1): 323-331.
- Wang X-F, Su J, Yang N, Zhang H, Cao X-Y and Kang J-F (2017). Functional characterization of selected universal stress protein from *Salvia miltiorrhiza* (SmUSP) in *Escherichia coli*. *Genes* **8**(9): 224.
- Waugh DS (2005). Making the most of affinity tags. *Trends in biotechnology* **23**(6): 316-320.
- Webb B and Sali A (2014). Comparative protein structure modeling using MODELLER. *Current protocols in bioinformatics* **47**(1): 5.6. 1-5.6. 32.
- Weerakoon KG, Gobert GN, Cai P and Mcmanus DP (2015). Advances in the diagnosis of human schistosomiasis. *Clinical microbiology reviews* **28**(4): 939-967.

- Weids AJ, Ibstedt S, Tamás MJ and Grant CM (2016). Distinct stress conditions result in aggregation of proteins with similar properties. *Scientific reports* **6**: 24554.
- Woelfle M, Seerden J-P, De Gooijer J, Pouwer K, Oliaro P and Todd MH (2011). Resolution of praziquantel. *PLoS neglected tropical diseases* **5**(9): e1260.
- You H, Stephenson RJ, Gobert GN and Mcmanus DP (2014). Revisiting glucose uptake and metabolism in schistosomes: new molecular insights for improved schistosomiasis therapies. *Frontiers in genetics* **5**: 176.
- Young DC. (2001). Computational chemistry: A John Wiley & Sons.
- Zarembinski TI, 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): 15189-15193.

# Chapter 3: Cytokines, Oxidative Stress and Schistosomiasis

## 3.1 Introduction

The complexities of the inter-relationship between the mammalian host and the schistosome parasite bear significant bearing on the human immune response and the redox network. It is a well-known fact that the aggressive entry of helminth parasites into the human host system is primarily for its colonization and long-term survival through immunomodulation of the immune system. Schistosomes have successfully developed, over eons of evolutionary time, the ability to stealthily plague the host and evade detection, not necessarily by multiplying in number or fluctuating antigen homeostasis, but by causing tolerance of the immune system to parasite antigens instead, resulting in the host's inability to eliminate these parasites (Coakley *et al.*, 2016). Many of the pathologies and a vast bulk of the morbidity resulting from schistosome infection is mainly attributed to immune response-based inflammation against eggs caught in between host tissues. Hypothetically speaking, schistosomiasis control and elimination should be effortless due to the parasite's multiple but treatable lifecycle stages. However, association of the parasite with the human host has significantly augmented its persistence over time (Cai *et al.*, 2016). Regardless of the countless antigen molecules tested for vaccine development, none have been successful till date. Wilson and Coulson (2006) rightly declare "...schistosomes are not stupid. They have had tens of millions of years to evolve mechanisms that help them survive immune attack from the mammalian host, even an attack orchestrated by our vaccine strategies. Schistosomes are only a formidable adversary that won't be easily beaten" (Wilson and Coulson, 2006). To a great extent most of the schistosome's acquired immunity is highly attributed to

antibody-dependent mechanisms. The longevity of schistosomes within the human host, like so many other helminths, is accompanied by a myriad of immune implications, which the parasite has been able to successfully manipulate and modulate for its own survival. While down-regulation of the immune response not only awards the parasite the opportunity to develop and thrive in the host, a dual effect is also presented as evolution of the host-parasite interaction limits severe pathology by modulation of over-vigorous immunopathology (Jenkins *et al.*, 2005). It has been the school of thought for a long period of time that immune modulation is almost always induced by eggs produced from adult worms but growing evidence now supports the idea that immune regulation actually begin from the point of infection all throughout the stages the host is exposed to. Biomolecules secreted by skin-penetrating cercariae, migrating schistosomulae, larval and adult worms together with their eggs possess the ability to modulate both innate and adaptive immune responses, bestowing the schistosome parasite evasion tactics from the host's defense mechanisms. These actions include the down-regulation and up-regulation of pro- and anti-inflammatory cytokines respectively as well as switches between the Th1 and Th2 immune response (Janssen *et al.*, 2016).

While the host battles with issues pertaining to immunity, the presence of the granuloma, which is so typically associated with schistosome infection immunity is in addition highly associated with oxidative damage through the generation of free radicals. The production of reactive oxygen species (ROS) is an inevitable process of cellular aerobic metabolism but may be aggravated due to and as a result of ionizing radiation, UV light, exercise, pollutants as well as various physiological disorders such as HIV/AIDS, cancer, diabetes and helminth infections (Brindley and Loukas, 2017; Ivanov *et al.*, 2016). Free radicals are extremely reactive molecules formed by unmatched electrons in the valence electron shell that have the ability to react with or generate

other free radicals by subtracting an electron from neighboring molecules, causing their instability. This creates a rippling effect where incomplete molecules seek for the ‘missing’ electron in an attempt to stabilize themselves but simultaneously cause the irreversible damage and modification of various cellular components such as membranes and receptors (Chiumiento and Bruschi, 2009). Whilst in the human host, both the host and schistosome parasite experience immense oxidative stress from the release of free radicals either generated by the human host through the immune response or from the parasite itself via respiration resulting in the production of ferrous iron and toxic heme through the consumption and breakdown of hemoglobin (Sayed *et al.*, 2006). Surprisingly the schistosome parasite are poorly equipped to handle oxidative stress but despite these unfavourable circumstances the worm has devised means to persistently survive and occupy the human host for even up to three decades without experiencing any severe effects (Huang *et al.*, 2012; Simeonov *et al.*, 2008). Broadening and expanding our knowledge in these areas creates several opportunities for anti-schistosome therapeutic development. Hence this review/chapter discusses the various strategies the schistosome exploits to establish its production, growth and thriving ability.

### **3.2 Immunology of the Acute and Chronic Stages of Schistosomiasis**

In its infancy, schistosomiasis is considered an asymptomatic disease that is due to a toxemic and allergic reaction or systemic hypersensitivity toward migrating and maturing larvae while its acuteness is more of a result of worm burden and immune response to parasite antigens (Han *et al.*, 2009; Siqueira *et al.*, 2017). The complex migratory lifecycle of the schistosome has been described in great detail in literature and is initiated by the cutaneous penetration of free-swimming multicellular infective cercariae into the human mammalian host aided by a number of environmental signals such as light-dark contrast, motion, chemical and thermal gradients

(Han *et al.*, 2009), which concurrently lose their bifurcated tails to develop into schistosomulae. Contact with cercariae-infested waters for a minimum of 1-5mins is all that is required for infection, which is typical especially among young adolescents in endemic regions (Siqueira *et al.*, 2017; Colley *et al.*, 2014). Cercarial dermatitis, commonly referred to as Katayama fever, is a form of a rash that may appear 24hrs after of exposure and disappear again within a few hours. This is referred to as the acute phase and is mostly associated with the Type 1 T helper (Th1) response and high levels of tumour necrosis factor (TNF- $\alpha$ ) and interferon gamma (IFN- $\gamma$ ) (Araújo *et al.*, 2004). However, continuous exposure exhibits the production of other pro-inflammatory cytokines that include IL-1 $\beta$ , IL-12, IL-10, IL-6, and MIP1 $\alpha$  (Jenkins *et al.*, 2005). The schistosome infective agent dictates varying symptoms that include dry cough and angio-oedema, headache, fever, urticaria and maculopapular pruritic skin eruption described above as cercarial dermatitis with first time non-immune travelers to schistosome-endemic areas being the most affected (Jauréguiberry *et al.*, 2010).

These symptoms may however, digress into a number of more serious and complicated clinical sequelae. It is the intention, from the schistosome's perspective, that the eggs be expelled through urine and feces based on the various schistosome specie locations. Target sites include the rectum and distal colon for *S. japonicum*, *S. mansoni*, *S. intercalatum* and *S. mekongi* parasites, while the bladder is the location of choice for *S. haematobium* worms (Barsoum *et al.*, 2013). However, this is not always the case as venous blood currents result in the flow of eggs in the opposite direction; preventing environmental escape and causing entrapment (Colley and Secor, 2014). For endogenous populations, the clinical and epidemiological picture of this disease is quite different. Chronic schistosomiasis is well-established among endemic residents, whose intensity of infection depends on continuous exposure to infested waters and the

individual's immune response to infection, resulting in either the intestinal or urinary form of the disease. Infection may be acquired from as early as 5 years and below, peaking at 8–15 years of age and persisting at moderate levels throughout adulthood (Colley *et al.*, 2014; Gryseels *et al.*, 2006; King, 1991). Intestinal schistosomiasis results in sporadic abdominal pain, per-rectal bleeding, chronic diarrhea caused by ulceration, mucosal hyperplasia, abscess formation and polyposis due to the formation of granuloma along the intestinal wall and liver (Weerakoon *et al.*, 2015). Poor regulation of the immune response causes chronic patients to subsequently develop hepatosplenic periportal fibrosis commonly known as Symmer's pipe stem fibrosis where scarring results in enlarged liver veins that have the appearance of pipe-stems in a cross section, which in turn affects hepatocellular function and results in portal hypertension and splenomegaly (Colley *et al.*, 2014; King, 1991). Urinary schistosomiasis emerges as a consequence of lesions present on the bladder wall resulting in bloody urine (haematuria), chronic pain, anaemia, pollakisuria, proteinuria and dysuria (Gryseels *et al.*, 2006). Haematuria is the most common and widespread lower tract symptom that is more than often regarded as a sign of puberty in young boys and girls of many endemic regions. Bladder malignances, calcification and fibrosis of the urinary tract resulting in obstructive uropathy are also observed (Weerakoon *et al.*, 2015). Further complications include the development of inflammatory lesions in female genitalia resulting in sandy patches that not only cause pain but also increase infertility rates, spontaneous abortion and risk to HIV transmission (Colley *et al.*, 2014). Growing evidence additionally implicates *S. haematobium* as a Group 1 definitive biological precursor to bladder cancer occurrences (Ishida and Hsieh, 2018). It has been suggested that schistosome eggs release vascular endothelial growth factor (VEGF) proteins that contribute to the incidence of bladder cancer. Inflammation induced by *S. haematobium* infection has been

associated with the activation of carcinogenic metabolites produced by aromatic amines and polycyclic aromatic hydrocarbons (Oyinloye *et al.*, 2015).

Additionally, initiation of cancer instigated by DNA disruption and mutational changes has also been linked to schistosome-induced inflammation. The Th2 type response typified by the up-regulation of IL-4, IL-5 and particularly IL-10 is then stimulated by egg antigens while simultaneously exhibiting decreased levels of IFN- $\gamma$  (Araújo *et al.*, 2004). In addition, the egg-induced response releases eosinophils, IgE, IL-9, IL-13, IL-21, basophils and mast cells that possess the ability to release other Th2-type cytokines (Anthony *et al.*, 2007; Capron *et al.*, 2005; Capron and Capron, 1994). Unexpectedly, human resistance to subsequent schistosome infections has been linked to the Th2 response. Epidemiological studies have established that the production of certain Immunoglobulin E (IgE) antibodies, eosinophils, IL-4 and IL-5 cytokines after Praziquantel (PZQ) treatment, plays a role in resistance while IgG4 is associated with susceptibility to infection (Colley and Secor, 2014; McManus *et al.*, 2018). It is suggested that the presence of dead worms within the human host may cross-react with antigens released from susceptible or migrating larvae thereby inducing protective parasite specific IgE, IL-5 and eosinophilia responses resulting in stronger protective responses, a trait mostly observed among infected adults than in younger children. However, despite successful vaccination trials in both experimental and reservoir hosts, there is still no vaccine available for schistosomiasis. As sparsely mentioned above, the mechanisms that span around the symptoms manifested during chronic schistosomiasis are mainly provoked by the production of certain signature cytokines. IL-4 is a key inducer in the pathogenesis of schistosomiasis as a stimulator of granuloma formation while IL-5 is involved in the development of liver fibrosis. IL-10 is actively involved in the down-regulation of both Th1 and Th2 responses, IFN- $\gamma$  is a potent activator of

inflammatory cells such as macrophages and simultaneously provides a protective role against periportal fibrosis, whereas TNF- $\alpha$  is involved in disease aggravation (Silveira-Lemos *et al.*, 2013). These various cytokine profiles are expounded in detail below.

### **3.2.1 Interleukin 4**

Basophils are innate cells that are large producers of IL-4 that are stimulated by the Th2 response in the presence of circulating antigen-specific Immunoglobulin E (McSorley *et al.*, 2013). Studies conducted by Schramm *et al* identified IPSE (interleukin-4 inducing principle from *Schistosoma* eggs), a cationic glycoprotein previously discovered as the egg antigen alpha 1 and confirmed to be the same molecule that is produced and secreted exclusively from the sub-shell area of *S. mansoni* eggs. Release of the protein into an environment filled with recruited inflammatory cells, triggers the degranulation and activation of basophils to release large amounts of IL-4 and histamine (Schramm *et al.*, 2006; Kaur *et al.*, 2011). Due to the fact that the Th2 response is predominantly egg-dependent, IL-4 is therefore considered a potent role-player in initiating this reaction. Knockout experiments in mice noted substantial decrease in the granulomatous response in IL-4 receptor-alpha knockout (IL4ra<sup>-/-</sup>) mice as compared to the wild type (Mbanefo *et al.*, 2019). Despite this, IL-4 production is linked to potential acquired immunity to schistosomiasis. Schistosome adult worm antigen (SWAP) levels increase significantly in resistant adults while an increase in IgG4, IgG2 and IgM serum levels on the other hand, correspond to increase susceptibility by blocking IgE-mediated killing of schistosomes. IL-4 in conjunction with IL-5 is responsible for IgE and eosinophil regulation and production, where the increase in their production is associated with the increase in subsequent acquired immunity (Medhat *et al.*, 1998).

### 3.2.2 Interleukin 10

IL-10 is an anti-inflammatory cytokine produced by T regulatory cells widely associated with the induction of soluble egg antigens (SEA) thus regulating Th1 and Th2 response (Lundy and Lukacs, 2013). In association with transformation growth factor  $\beta$  (TGF- $\beta$ ), IL-10 functions to decrease T-helper cell activity and thus suppress granulomatous pathology and inflammation in chronic schistosomiasis (Costain *et al.*, 2018). Although it has been shown to produce mixed Th1 and Th2 responses resulting in severe liver damage and increased morbidity levels in mice during acute infection, double IL-4<sup>-/-</sup>IL-10<sup>-/-</sup> and IL-12<sup>-/-</sup>IL-10<sup>-/-</sup> knockout studies showed the protective effect of IL-10 against tissue damage and highly polarized immune response effects (Hesse *et al.*, 2004). The expression of IL-10 mostly occurs after the secretion of other pro-inflammatory cytokines such as IL-1 $\beta$ , IL-12, IL-18 and TNF whose production is inhibited due to suppression of macrophage and cell activity by IL-10, thereby suggesting its role in decreasing their other cytokine effects (Jenkins *et al.*, 2005; Hesse *et al.*, 2004). It has been reported that deficiencies in IL-10 may lead to the severe pathogenesis observed in chronic schistosomiasis (Caldas *et al.*, 2008). In addition, IL-10 is involved in the down-regulation and inhibition of MHC class II, co-stimulatory molecules and the production and proliferation of T cell cytokines respectively (Hesse *et al.*, 2004). All this shows the IL-10 cytokine's role is a complex one that is incorporated in various immune cells whose responses are stimulated by various mechanisms.

### 3.2.3 IFN- $\gamma$

Protection against fibrosis is attributed to the anti-fibrogenic and profibrotic cytokine IFN- $\gamma$ , which blocks hepatic stellate cells from producing extracellular molecules while simultaneously augmenting the expression of matrix metalloproteinase (MMP) (Mutengo *et al.*, 2018). Mouse experiments exhibited an overall decrease in liver fibrosis, resultant from the inhibition of

lipocyte activation, which in turn led to a reduction in collagen I mRNA (Chevallard 2016). However, NO and other inflammatory mediators are produced by the activation of macrophages induced by IFN- $\gamma$ . This was observed in the reduction of NO levels and granuloma formation in infected IL-4<sup>-/-</sup> IFN- $\gamma$  neutralized mice (La Flamme *et al.*, 2001a).

### **3.2.4 TNF- $\alpha$**

Contrasting studies have associated high levels of TNF- $\alpha$  with disease severity and other cytokines. Some observations have highlighted the expression of TNF- $\alpha$  as dependent on the production of other cytokines such as IFN- $\gamma$ . In addition it has also been regarded as a protector against egg-induced liver pathology due to the increase in its levels in the emergence of little or no fibrosis (Mutengo *et al.*, 2018). Studies have shown increases in TNF- $\alpha$  levels in mice with hepatoesplenic disease, which is commonly associated with the stimulation of peripheral blood mononuclear cells (PBMCs) whose induction produces TNF- $\alpha$  (de Jesus *et al.*, 2004; Adewusi *et al.*, 1996). Human studies have also demonstrated the elevation of serum TNF- $\alpha$  levels in patients infected with urinary schistosomiasis as well as individuals having carcinoma of the urinary bladder with and without schistosomiasis (Raziuddin *et al.*, 1993).

One of the most apparent and distinct clinical features and hallmarks of schistosome infection is the formation of the granuloma, a collective mass of collagen-bound inflammatory cells that encapsulate mature parasite eggs with a wall during the chronic phase of infection. Soluble egg antigens are released from viable larvae within schistosome eggs that stimulate granuloma development (Lundy and Lukacs, 2013). TNF $\alpha$  and IL-1 inflammatory cytokines and chemokines are produced from antigen presenting cells such as macrophages and epithelial cells, which give rise to neutrophils, monocytes, lymphocytes and granulomas. Granulomas are masses of tissue rich in esinophils, macrophages, lymphocytes and fibroblasts that are instigated by

CD4<sup>+</sup> T-helper cells, which categorically differentiate into Th1 or Th2 subgroup cells while releasing TH1-type cytokines, IFN $\gamma$  and IL-2 (Brunet *et al.*, 1998). A characteristic type-1 cytokine response overridden by a domineering type-2 response drives granuloma formation and evidence of this was observed in IL-12 vaccinated mice made to induce a CD4<sup>+</sup> Th1 response, which developed smaller lesions compared to non-vaccinated Th2-induced mice, while the production of excessive type-1 cytokine levels was associated with reduced fibrosis (Wynn *et al.*, 2004). Induction of the Th1 response to produce of IL-12, TNF and IFN- $\gamma$  stimulates a cell-mediated response, attracting immune cells to the place of granuloma formation. The packing of cells around the ova results in the organization and maturation of the granuloma, commencing the Th2 response (Olveda *et al.*, 2014). As the late stages of granuloma completion occur, fibroblasts are produced and replace cells with collagen deposits, which results in fibrosis, the major cause of schistosome morbidity and mortality. The presence of fibrous tissue results in the onset of hepatosplenic schistosomiasis, causing the increase of TNF $\alpha$  and IFN $\gamma$  levels while IL-5 cytokines, which are released by the Th2 type response is decreased (Schwartz and Fallon, 2018). The twofold heterogeneity of the Th1 and Th2 response proffers benefits to the host once it elapses beyond the ability to evade persistent infection but must still protect against tissue damage during progression from the acute to the chronic stage (Barron and Wynn, 2011).

### **3.3 Paradox of the Granuloma**

Granulomas are pathogenic in their own right as explained above, but they also avert hepatic and tissue damage by impounding and providing physical impediment against proteolytic molecules, enzymes and toxins released by the egg (Brunet *et al.*, 1998; Ross *et al.*, 2012). It is now envisaged that the inflammatory response linked to granuloma formation also seems to perform a host-protective effect during schistosome infection. Antigens from the schistosome parasite are

not only potent stimulators of the immune system; they are also hepatotoxic, causing damage to target tissues, such that mostly involve the liver (Hams *et al.*, 2013). Granulomatous lesions prevent contact between these toxins and the cells of the liver. A study by Fallon and co-workers showed that IL-4-deficient mice displayed higher mortality rates in comparison to the IL-3-deficient mouse population. In the former, suppression of the Th2 response resulted in the inability of mice to regulate the excretion of eggs into the gut lumen causing their entrapment in the intestine and resulting in the continuous deposition of hepatic collagen. This in turn led to severe intestinal inflammation and consequently death. In contrast, a normal Th2 response and formation of granuloma was observed in IL-3 deficient mice but was accompanied by a decrease in mortality rates. This study showed that cytokines produced in response to the Th2 effect reduces the damage elicited onto the intestine during schistosome infection (Fallon *et al.*, 2000).

At the heart of granuloma formation and fibrosis is chronic inflammation aggravated by the release of soluble egg antigens from entrapped eggs in host tissue. Full blown inflammation is largely driven by IL-17 released from the pro-inflammatory Th17 cell group, which is also in turn largely responsible for the very severe immunopathological consequences associated with schistosome infection (Yepes *et al.*, 2015). Macrophages, eosinophils and neutrophils, which are inflammatory cells actively involved in chronic schistosomiasis, are potent sources of reactive oxygen species (ROS); crucial ingredients to tissue damage, tumour progression and cancer (Rosin *et al.*, 1994). Studies in infected mice showed a significant decrease in Vitamins C and E and catalase levels in the liver while superoxide dismutase, body weight and lipid peroxides were observed after in the mice after various weeks of infection (Rizk *et al.*, 2006). From this study, it was concluded that oxidative stress results in the generation of free radicals in an attempt to eradicate worms. Additionally, an increase in nitric oxide, peroxynitrite and superoxide levels

coupled with hepatocyte enlargement and a decrease in antioxidant levels have also been observed in other experiments (La Flamme *et al.*, 2001b). The following sections break down the associations between oxidative stress and cytokines in the host-parasite relationship and how this may be considered as targets for the development of novel schistosomicides.

### 3.4 ROS Production in Schistosomiasis

The vertebrate human host operates as a potent source of oxidants. Professional phagocytes are cells of the innate immune system capable of producing large amounts of bactericidal ROS, which coupled with redox-active and toxic by-products, factor into the schistosome's wrestle against oxidative stress (Perbandt *et al.*, 2014). Through phagocytosis, also referred to as oxidative or respiratory burst, phagocytic cells (leukocytes) form a cytotoxic defense against helminths, parasitic fungi, *Mycobacteria* and protozoa (Datta *et al.*, 2000). Phagosomal membrane-bound nicotinamide adenine dinucleotide phosphate (NADPH) oxidase enzyme is activated thereby reacting and consuming molecular oxygen to produce superoxide anion ( $O_2^-$ ). The superoxide anion forms the basis for additional ROS generation in the form of hydrogen peroxide ( $H_2O_2$ ), a weak oxidizing agent, through the conversion of  $O_2^-$  by superoxide dismutase (SOD) (Bryan *et al.*, 2012). Through the Fenton reaction, ferrous iron ( $Fe^{2+}$ ) reacts with  $H_2O_2$  to form the highly reactive hydroxyl radical ( $OH^\bullet$ ), hydroxyl ion ( $OH^-$ ) and ferric iron  $Fe^{3+}$ . The latter will in turn react with  $O_2^-$  to be reduced back into  $Fe^{2+}$  allowing for its conversion to  $H_2O_2$  and  $OH^-$  again (Haber-Weiss pathway) (Bogdan *et al.*, 2000). Phagosome cytotoxicity is increased through phagosome fusion caused by hypohalous acids such as hypochloride ( $OCl^-$ ) derived from  $H_2O_2$ .  $OCl^-$  may additionally react with  $H_2O_2$  and amine, producing singlet oxygen ( $^1O_2$ ), which in turn forms the basis for other free radicals. Superoxide may also react with nitric oxide (NO) to produce peroxynitrite ( $ONOO^-$ ), which may further react with carbonyl

compounds to form organic peroxy radicals (Bryan *et al.*, 2012). Through the stimulation of cytokines and other immunological reactions, macrophage NOS in the form of NOS2, macNOS and iNOS, which is in fact never produced within resting cells, is produced (Bogdan *et al.*, 2000). Other implications of ROS produced by immune cells include NADPH oxidase released by phagocytes. The  $O_2$  in reactive oxygen is transformed in a bid to attack microorganisms. This however results in the accumulation of inflammatory cells such as macrophages and lymphocytes, which leads to the formation of the granuloma (Chiumiento and Bruschi, 2009). Thus the production of ROS promotes schistosome infection progression. A study conducted by El-Sokkary and co-workers discovered and confirmed that the production of free radicals is actually an underlying factor in disease progression and the pathological changes that are associated with schistosomiasis. Oxidative processes in schistosome-infected mice were investigated using the antioxidant melatonin. Untreated mice observed an increase in lipid peroxidation (LPO) and nitric oxide (NO) levels, while SOD activity, glutathione (GSH) and Vitamin E levels in the spleen, kidney and liver saw a decrease in levels. Granulomas were present in the liver with an associated presence of eosinophils as well as the manifestation of necrotic cells and megakaryocytes in the kidney and spleen of the mice respectively (El-Sokkary *et al.*, 2002). Similar studies conducted by Gharib and colleagues showed that the presence of parasitic eggs stimulates endogenous eosinophil peroxidase release while a decrease in the activities of glutathione peroxidase, superoxide dismutase as well as catalase was observed (Gharib *et al.*, 1999).

### 3.5 Oxidative stress, Inflammatory Cytokines and Genetic Instability

Parasitic diseases that are associated with inflammation are now becoming topics of great interest especially because of possible links to angiogenesis and tumor formation. An estimated 20% of cancers arise due to inflammation induced by chronic infection and even without cancer formation, there is marked increase in the expression of cytokines in the tumor microenvironment (Grivennikov and Karin, 2011). In addition, oxidative stress has now been increasingly implicated in such processes, sometimes initiated directly by pro-inflammatory cytokines or their products.

Tumor necrosis factor (TNF- $\alpha$ ) is considered one of the underlying and major factors that drive inflammatory processes, which have for a long time been linked strongly to cancer development. TNF- $\alpha$  possesses the ability to activate NF-kappa B through the recruitment of TNF- $\alpha$  receptor-associated death domain (TRADD). This in turn activates I $\kappa$ B kinase (IKK), leading to its phosphorylation, ubiquitination and degradation, eventually activating NF-kappa B. Its activation expresses inflammatory genes that include chemokines, anti-apoptotic proteins, cyclooxygenase-(2) and inducible nitric oxide synthase (iNOS), which suppress apoptosis induced by TNF- $\alpha$  (Sethi *et al.*, 2008).

In association with IL-6, TNF- $\alpha$  activates other oncogenic transcription factors, which include STAT-3 and AP-1 respectively. These influence the major stages of tumor development: initiation, promotion, progression and metastasis (Grivennikov and Karin, 2011). The activation of these transcription factors have been identified in over 50% of all cancer and induce resistance to apoptosis in fully transformed and pre-malignant cells, thus increasing proliferation and tumor growth.

Nitric oxide is a lipophilic free radical produced via the NO synthase (NOS) enzymatic reaction catalyzed as a result of the conversion of oxygen and arginine to citrulline. In schistosomiasis cases, inducible NOS (iNOS) is expressed most notably in response to IFN- $\gamma$  and other pro-inflammatory cytokines such as TNF- $\alpha$  and IL-1 $\beta$  (Brunet *et al.*, 1999). It has been awarded as a host protective molecule by regulating inflammatory responses and mediating the cytotoxic pathway responsible for killing 2-week-old schistosomulae *in vitro* by activated macrophages (Brunet *et al.*, 1999; Oswald, 2002). Despite its beneficial roles however, NO has been found to stimulate disease process by forming toxic radicals that induce cell apoptosis and damage.

Oxidative stress in schistosomiasis is accompanied by mediated site-specific tissue damage such as the liver especially at sites of granulomatous inflammation due to the production of ROS by macrophages and eosinophils (Soliman and El-Shenawy, 2003). Lipid peroxidation occurs, damaging antioxidant systems set in place for ROS scavenging; this in turn results in peroxidative impairment of lipids found within the liver (Al-Olayan *et al.*, 2016). Apart from this oxidative stress in general stimulates pathways that damage cell structure and function thus linking it to the initiation and progression of cancer by inducing DNA mutations and damage and causing cell proliferation and genome instability (Reuter *et al.*, 2010). *S. haematobium* has been mentioned severally to be classified as a Group 1 carcinogen by the World Health Organization and the International Agency for Research on Cancer due to recent reports on the disease predisposing infected individuals to carcinomas such as squamous bladder cancer (Gryseels *et al.*, 2006; Oyinloye *et al.*, 2014). ROS released by inflammatory cells, which include macrophages and neutrophils cause DNA-strand breaks, sister chromatid exchanges as well as mutations.

One of the notable ways in how oxidative stress in schistosomiasis causes DNA damage can be explained by the estrogen-DNA adduct mediated pathway. Estrogen-like metabolites secreted by *S. haematobium* worms and eggs are highly genotoxic as a result of catechol estrogens such as 2-hydroxy(OH)E1(E2) and 4-OHE1(E2) oxidizing to form catechol estrogen quinones (CEQ) and semiquinones such as E1(E2)-3,4Q (Botelho *et al.*, 2015; Santos *et al.*, 2014). This in turn results in redox cycling and production of ROS, which react with DNA. Depurinating adducts form resulting in the generation of apurinic sites, which have the potential to become mutagenic over several cell generations, eventually becoming cancerous (Ishida and Hsieh, 2018; Santos *et al.*, 2014). In addition to cancer, mutations generated by apurinic sites are linked to autosomal recessive infertility.

In comparison to their human host, schistosomes are poorly equipped to cope with oxidative stress. The worms lack catalase and their GPx proteins exhibit low activity for quenching hydrogen peroxide, making them more susceptible to hydrogen peroxide than their host (Huang *et al.*, 2012). Therefore, the redox network in schistosomes are considered good target for candidate drugs. Chapter 5 proposes a possible feat in this direction.

## References

- Adewusi OI, Nix NA, Lu X, Colley DG and Secor WE (1996). Schistosoma mansoni: relationship of tumor necrosis factor- $\alpha$  to morbidity and collagen deposition in chronic experimental infection. *Experimental Parasitology* **84**(2): 115-123.
- Al-Olayan EM, El-Khadragy MF, Alajmi RA, Othman MS, Bauomy AA, Ibrahim SR and Moneim AEA (2016). Ceratonia siliqua pod extract ameliorates Schistosoma mansoni-induced liver fibrosis and oxidative stress. *BMC Complementary and Alternative Medicine* **16**(1): 434.
- Anthony RM, Rutitzky LI, Urban Jr JF, Stadecker MJ and Gause WC (2007). Protective immune mechanisms in helminth infection. *Nature Reviews Immunology* **7**(12): 975.
- Araújo MI, Hoppe BS, Medeiros Jr M and Carvalho EM (2004). Schistosoma mansoni infection modulates the immune response against allergic and auto-immune diseases. *Memórias do Instituto Oswaldo Cruz* **99**: 27-32.
- Barron L and Wynn TA (2011). Macrophage activation governs schistosomiasis-induced inflammation and fibrosis. *European Journal of Immunology* **41**(9): 2509-2514.
- Barsoum RS, Esmat G and El-Baz T (2013). Human schistosomiasis: clinical perspective. *Journal of Advanced Research* **4**(5): 433-444.
- Bogdan C, Röllinghoff M and Diefenbach A (2000). Reactive oxygen and reactive nitrogen intermediates in innate and specific immunity. *Current Opinion in Immunology* **12**(1): 64-76.

- Botelho MC, Alves H, Barros A, Rinaldi G, Brindley PJ and Sousa M (2015). The role of estrogens and estrogen receptor signaling pathways in cancer and infertility: the case of schistosomes. *Trends in parasitology* **31**(6): 246-250.
- Brindley PJ and Loukas A (2017). Helminth infection–induced malignancy. *PLoS Pathogens* **13**(7): e1006393.
- Brunet LR, Beall M, Dunne D and Pearce E (1999). Nitric oxide and the Th2 response combine to prevent severe hepatic damage during *Schistosoma mansoni* infection. *The Journal of Immunology* **163**(9): 4976-4984.
- Brunet LR, Dunne D and Pearce E (1998). Cytokine interaction and immune responses during *Schistosoma mansoni* infection. *Parasitology Today* **14**(10): 422-427.
- Bryan N, Ahswini H, Smart N, Bayon Y, Wohler S and Hunt JA (2012). Reactive oxygen species (ROS)—a family of fate deciding molecules pivotal in constructive inflammation and wound healing. *Eur Cell Mater* **24**(249): e65.
- Cai P, Gobert GN, You H and Mcmanus DP (2016). The Tao survivorship of schistosomes: implications for schistosomiasis control. *International journal for parasitology* **46**(7): 453-463.
- Caldas IR, Campi-Azevedo AC, Oliveira LFA, Silveira AMS, Oliveira RC and Gazzinelli G (2008). Human schistosomiasis mansoni: immune responses during acute and chronic phases of the infection. *Acta tropica* **108**(2-3): 109-117.
- Capron A, Riveau G, Capron M and Trottein F (2005). Schistosomes: the road from host–parasite interactions to vaccines in clinical trials. *Trends in parasitology* **21**(3): 143-149.

- Capron M and Capron A (1994). Immunoglobulin E and effector cells in schistosomiasis. *SCIENCE-NEW YORK THEN WASHINGTON*:-: 1876-1876.
- Chiumiento L and Bruschi F (2009). Enzymatic antioxidant systems in helminth parasites. *Parasitology research* **105**(3): 593.
- Coakley G, Buck AH and Maizels RM (2016). Host parasite communications—messages from helminths for the immune system: parasite communication and cell-cell interactions. *Molecular and biochemical parasitology* **208**(1): 33-40.
- Colley DG and Secor W (2014). Immunology of human schistosomiasis. *Parasite immunology* **36**(8): 347-357.
- Colley DG, Bustinduy AL, Secor WE and King CH (2014). Human schistosomiasis. *The Lancet* **383**(9936): 2253-2264.
- Costain AH, Macdonald AS and Smits HH (2018). Schistosome egg migration: mechanisms, pathogenesis and host immune responses. *Frontiers in immunology* **9**.
- Datta K, Sinha S and Chattopadhyay P (2000). Reactive oxygen species in health and disease. *National Medical Journal of India* **13**(6): 304-310.
- De Jesus AR, Magalhaes A, Miranda DG, Miranda RG, Araújo MI, De Jesus AA, Silva A, Santana LB, Pearce E and Carvalho EM (2004). Association of type 2 cytokines with hepatic fibrosis in human *Schistosoma mansoni* infection. *Infection and Immunity* **72**(6): 3391-3397.
- El-Sokkary GH, Omar HM, Hassanein A-FM, Cuzzocrea S and Reiter RJ (2002). Melatonin reduces oxidative damage and increases survival of mice infected with *Schistosoma mansoni*. *Free Radical Biology and Medicine* **32**(4): 319-332.

- Fallon PG, Richardson EJ, McKenzie GJ and McKenzie AN (2000). Schistosome infection of transgenic mice defines distinct and contrasting pathogenic roles for IL-4 and IL-13: IL-13 is a profibrotic agent. *The Journal of Immunology* **164**(5): 2585-2591.
- Gharib B, Abdollahi OMS, Dessein H and De Reggi M (1999). Development of eosinophil peroxidase activity and concomitant alteration of the antioxidant defenses in the liver of mice infected with *Schistosoma mansoni*. *Journal of Hepatology* **30**(4): 594-602.
- Grivennikov SI and Karin M (2011). Inflammatory cytokines in cancer: tumour necrosis factor and interleukin 6 take the stage. *Annals of the Rheumatic Diseases* **70**(Suppl 1): i104-i108.
- Gryseels B, Polman K, Clerinx J and Kestens L (2006). Human schistosomiasis. *The Lancet* **368**(9541): 1106-1118.
- Hams E, Aviello G and Fallon PG (2013). The schistosome granuloma: friend or foe? *Frontiers in immunology* **4**: 89.
- Han Z-G, Brindley PJ, Wang S-Y and Chen Z (2009). *Schistosoma* genomics: new perspectives on schistosome biology and host-parasite interaction. *Annual Review of Genomics and Human Genetics* **10**: 211-240.
- Hesse M, Piccirillo CA, Belkaid Y, Prufer J, Mentink-Kane M, Leusink M, Cheever AW, Shevach EM and Wynn TA (2004). The pathogenesis of schistosomiasis is controlled by cooperating IL-10-producing innate effector and regulatory T cells. *The Journal of Immunology* **172**(5): 3157-3166.
- Huang H-H, Rigouin C and Williams DL (2012). The redox biology of schistosome parasites and applications for drug development. *Current pharmaceutical design* **18**(24): 3595-3611.

- Ishida K and Hsieh MH (2018). Understanding urogenital schistosomiasis-related bladder cancer: an update. *Frontiers in medicine* **5**.
- Ivanov AV, Valuev-Elliston VT, Ivanova ON, Kochetkov SN, Starodubova ES, Bartosch B and Isaguliant MG (2016). Oxidative stress during HIV infection: mechanisms and consequences. *Oxidative Medicine and Cellular Longevity* **2016**.
- Janssen L, Santos S, Lorranna G, Muller HS, Vieira ARA, De Campos TA and De Paulo Martins V (2016). Schistosome-derived molecules as modulating actors of the immune system and promising candidates to treat autoimmune and inflammatory diseases. *Journal of immunology research* **2016**.
- Jauréguiberry S, Paris L and Caumes E (2010). Acute schistosomiasis, a diagnostic and therapeutic challenge. *Clinical Microbiology and Infection* **16**(3): 225-231.
- Jenkins S, Hewitson J, Jenkins G and Mountford A (2005). Modulation of the host's immune response by schistosome larvae. *Parasite immunology* **27**(10-11): 385-393.
- Kaur I, Schramm G, Everts B, Scholzen T, Kindle KB, Beetz C, Montiel-Duarte C, Blindow S, Jones AT and Haas H (2011). Interleukin-4-inducing principle from *Schistosoma mansoni* eggs contains a functional C-terminal nuclear localization signal necessary for nuclear translocation in mammalian cells but not for its uptake. *Infection and Immunity* **79**(4): 1779-1788.
- King CH (1991). Acute and chronic schistosomiasis. *Hospital Practice* **26**(3): 117-130.
- La Flamme AC, Patton EA and Pearce EJ (2001a). Role of gamma interferon in the pathogenesis of severe schistosomiasis in interleukin-4-deficient mice. *Infection and Immunity* **69**(12): 7445-7452.

- La Flamme AC, Patton EA, Bauman B and Pearce EJ (2001b). IL-4 plays a crucial role in regulating oxidative damage in the liver during schistosomiasis. *The Journal of Immunology* **166**(3): 1903-1911.
- Lundy SK and Lukacs NW (2013). Chronic schistosome infection leads to modulation of granuloma formation and systemic immune suppression. *Frontiers in immunology* **4**: 39.
- Mbanefo EC, Fu C-L, Ho CP, Le L, Ishida K and Hsieh MH (2019). Interleukin-4 signaling plays a major role in urogenital schistosomiasis-associated bladder pathogenesis. *bioRxiv*: 746552.
- Mcmanus DP, Dunne DW, Sacko M, Utzinger J, Vennervald BJ and Zhou X-N (2018). Schistosomiasis (Primer). *Nature Reviews: Disease Primers*.
- Mcsorley HJ, Hewitson JP and Maizels RM (2013). Immunomodulation by helminth parasites: defining mechanisms and mediators. *International journal for parasitology* **43**(3-4): 301-310.
- Medhat A, Shehata M, Bucci K, Mohamed S, Dief ADE, Badary S, Galal H, Nafeh M and King CL (1998). Increased interleukin-4 and interleukin-5 production in response to *Schistosoma haematobium* adult worm antigens correlates with lack of reinfection after treatment. *Journal of Infectious Diseases* **178**(2): 512-519.
- Mutengo MM, Mduluza T, Kelly P, Mwansa JC, Kwenda G, Musonda P and Chipeta J (2018). Low IL-6, IL-10, and TNF- $\alpha$  and high IL-13 cytokine levels are associated with severe hepatic fibrosis in *Schistosoma mansoni* chronically exposed individuals. *Journal of parasitology research* **2018**.

- Olveda DU, Olveda RM, Mcmanus DP, Cai P, Chau TN, Lam AK, Li Y, Harn DA, Vinluan ML and Ross AG (2014). The chronic enteropathogenic disease schistosomiasis. *International Journal of Infectious Diseases* **28**: 193-203.
- Oswald IP (2002). Nitric oxide in schistosomiasis *Nitric Oxide and Infection*: Springer pp. 343-360.
- Oyinloye B, Adenowo A and Kappo A (2015). Reactive oxygen species, apoptosis, antimicrobial peptides and human inflammatory diseases. *Pharmaceuticals* **8**(2): 151-175.
- Oyinloye B, Adenowo F, Gxaba N and Kappo A (2014). The promise of antimicrobial peptides for treatment of human schistosomiasis. *Current Drug Targets* **15**(9): 852-859.
- Perbandt M, Ndjonga D and Liebau E (2014). Protective mechanisms of helminths against reactive oxygen species are highly promising drug targets. *Current Medicinal Chemistry* **21**(15): 1794-1808.
- Raziuddin S, Masihuzzaman M, Shetty S and Ibrahim A (1993). Tumor necrosis factor alpha production in schistosomiasis with carcinoma of urinary bladder. *Journal of Clinical Immunology* **13**(1): 23-29.
- Reuter S, Gupta SC, Chaturvedi MM and Aggarwal BB (2010). Oxidative stress, inflammation, and cancer: how are they linked? *Free Radical Biology and Medicine* **49**(11): 1603-1616.
- Rizk M, Fayed T, Badawy M and El-Regal NS (2006). Effect of different durations of *Schistosoma mansoni* infection on the levels of some antioxidants in mice. *Med. J. Islam. World Acad. Sci* **16**: 25-34.

- Rosin MP, Anwar WA and Ward AJ (1994). Inflammation, chromosomal instability, and cancer: the schistosomiasis model. *Cancer Research* **54**(7 Supplement): 1929s-1933s.
- Ross AG, Mcmanus DP, Farrar J, Hunstman RJ, Gray DJ and Li Y-S (2012). Neuroschistosomiasis. *Journal of Neurology* **259**(1): 22-32.
- Santos J, Gouveia MJ, Vale N, De Lurdes Delgado M, Goncalves A, Da Silva JMT, Oliveira C, Xavier P, Gomes P and Santos LL (2014). Urinary estrogen metabolites and self-reported infertility in women infected with *Schistosoma haematobium*. *PloS one* **9**(5): e96774.
- Sayed AA, Cook SK and Williams DL (2006). Redox balance mechanisms in *Schistosoma mansoni* rely on peroxiredoxins and albumin and implicate peroxiredoxins as novel drug targets. *Journal of Biological Chemistry* **281**(25): 17001-17010.
- Schramm G, Gronow A, Knobloch J, Wippersteg V, Grevelding C, Galle J, Fuller H, Stanley R, Chiodini P and Haas H (2006). IPSE/alpha-1: a major immunogenic component secreted from *Schistosoma mansoni* eggs. *Molecular and biochemical parasitology* **147**(1): 9-19.
- Schwartz C and Fallon PG (2018). *Schistosoma* “eggs-iting” the host: granuloma formation and egg excretion. *Frontiers in immunology* **9**.
- Sethi G, Sung B and Aggarwal BB (2008). TNF: a master switch for inflammation to cancer. *Frontiers in Bioscience* **13**(2): 5094-5107.
- Silveira-Lemos D, Fernandes Costa-Silva M, Cardoso De Oliveira Silveira A, Azevedo Batista M, Alves Oliveira-Fraga L, Soares Silveira AM, Barbosa Alvarez MC, Martins-Filho OA, Gazzinelli G and Corrêa-Oliveira R (2013). Cytokine pattern of T lymphocytes in acute Schistosomiasis mansoni patients following treated praziquantel therapy. *Journal of parasitology research* **2013**.

- Simeonov A, Jadhav A, Sayed AA, Wang Y, Nelson ME, Thomas CJ, Inglese J, Williams DL and Austin CP (2008). Quantitative high-throughput screen identifies inhibitors of the *Schistosoma mansoni* redox cascade. *PLoS neglected tropical diseases* **2**(1): e127.
- Siqueira LDP, Fontes DaF, Aguilera CSB, Timóteo TRR, Ângelos MA, Silva LCPBB, De Melo CG, Rolim LA, Da Silva RMF and Neto PJR (2017). Schistosomiasis: Drugs used and treatment strategies. *Acta tropica* **176**: 179-187.
- Soliman MF and El-Shenawy NS (2003). Evaluation of the protective effect of two antioxidative agents in mice experimentally infected with *Schistosoma mansoni*: haematological and histopathological aspects. *Pak. J. Biol. Sci* **6**: 887-897.
- Weerakoon KG, Gobert GN, Cai P and Mcmanus DP (2015). Advances in the diagnosis of human schistosomiasis. *Clinical microbiology reviews* **28**(4): 939-967.
- Wilson RA and Coulson PS (2006). Schistosome vaccines: a critical appraisal. *Memórias do Instituto Oswaldo Cruz* **101**: 13-20.
- Wynn TA, Thompson RW, Cheever AW and Mentink-Kane MM (2004). Immunopathogenesis of schistosomiasis. *Immunological Reviews* **201**(1): 156-167.
- Yepes E, Varela-M RE, López-Abán J, Rojas-Caraballo J, Muro A and Mollinedo F (2015). Inhibition of granulomatous inflammation and prophylactic treatment of schistosomiasis with a combination of edelfosine and praziquantel. *PLoS neglected tropical diseases* **9**(7): e0003893.

# Chapter 4: Materials and Methods

## 4.3. Preparation of Competent *E. coli* cells

### 4.3.1 Preparation of DH5 $\alpha$ *E. coli* cells

Agar plates without antibiotics were streaked with glycerol stocks of DH5 $\alpha$  cells using an inoculation loop and grown overnight at 37°C. Colonies were picked with a yellow tip, inoculated in 5 mL LB broth and grown overnight at 37°C. A 500  $\mu$ L measure of the overnight culture was inoculated in 500 mL LB broth and grown for 4 hrs until the OD<sub>600</sub> reached 0.5. The cells were chilled in ice for 10 mins and centrifuged at 2000 Xg for 10mins at 4°C. The supernatant was discarded and the pellet re-suspended in 50 mL ice cold TFB1 Transformation Buffer. The cells were kept on ice with shaking for 20–30 mins to re-suspend the cells. The mixture was centrifuged at 2000 Xg for 10 mins at 4°C, the supernatant discarded and re-suspended in 25 mL ice cold TFB1 Transformation Buffer. The cells were once again kept on ice with shaking for 20–30 mins and the mixture aliquoted in 1.5 mL eppendorfs and stored in -80°C.

### 4.3.2 Preparation of JM109 *E. coli* cells

Colonies were picked from agar plates streaked with glycerol stocks of JM109 cells that were left to grow overnight at 37°C without antibiotics and incubated in 5 mL LB broth at the same temperature overnight. The culture was scaled up to 50 mL with LB broth without antibiotics and left at 37°C until the OD<sub>600</sub> 0.4–0.6 was obtained. Centrifugation of the culture ensued at 5000 rpm at 4°C for 10 mins. The supernatant was discarded and the pellet re-suspended in 10 ml ice cold 0.1 M MgCl<sub>2</sub>. The mixture was incubated on ice for 30 mins, spun down at 4000 rpm, the supernatant discarded and the pellet re-suspended in 10 mL ice cold 0.1 M CaCl<sub>2</sub>.

This was kept on ice for 4–5 hours to increase the competency of the cells, which were centrifuged again at 4000 rpm. The pellets were again re-suspended in 3 mL ice cold 0.1 M  $\text{CaCl}_2$  and 30% glycerol respectively and the competent cells were aliquoted and stored at  $-80^\circ\text{C}$  for future use.

#### **4.3.3 Preparation of M15 *E. coli* cells**

Cells from M15 glycerol stocks were streaked onto Nutrient Agar plates without antibiotics and grown at  $37^\circ\text{C}$  overnight. Colonies were picked and grown in 5 mL LB without antibiotics overnight and 1.5 mL of the culture was scaled up with 50 mL LB the next morning. The cells were grown to the early log phase ( $\text{OD}_{600}$  0.3–0.6). The cells were centrifuged at 4000 rpm at  $4^\circ\text{C}$  for 5–10 mins and the supernatant discarded. The cells were re-suspended in 5 mL ice cold Transformation Storage Buffer and incubated on ice for 20 mins with the addition of 15% glycerol. The cells were aliquoted in ice cold eppendorf tubes and stored in  $-80^\circ\text{C}$ .

#### **4.4. Transformation of *E. coli* cells with pQE30-G4LZI3 Construct**

The full length gene sequence encoding the G4LZI3 protein was cloned into a pQE30 protein expression vector containing an N-terminal 6xHis tag using BamH1 and HindIII restriction enzymes and the resultant construct was purchased from Genescript. Transformation was performed using 100  $\mu\text{L}$  of thawed JM109 and DH5 $\alpha$  competent cells and 2  $\mu\text{L}$  of the pQE30-G4LZI3 construct in a pre-chilled eppendorf tube, which was allowed to incubate on ice for 15 mins respectively. A control tube was also prepared with the addition of competent cells only. Heat shocking then followed at  $42^\circ\text{C}$  for 45 s to create pores in the cell membrane thus increasing membrane permeability and allowing for rapid uptake of the plasmid DNA. The tubes were transferred back on ice for 5 mins followed by the addition of 900  $\mu\text{L}$  of warm LB broth and then incubated in a shaker for 2 hrs at  $37^\circ\text{C}$ . The transformation mixtures were centrifuged at

5000 rpm for 5 mins and the pellets re-suspended in 100  $\mu$ L of warm LB. The cells were then spread on pre-prepared ampicillin-agar plates and incubated overnight for 18 hours at 37°C.

Transformation of the M15 cells was done by mixing 100 ng of the pQE30-G4LZI3 construct with 200  $\mu$ L of the cells. The mixture was incubated on ice for 30 mins – 2 hrs to increase transformation efficiency followed by the addition of 800  $\mu$ L TSBG. Incubation at 37°C for 1 hr ensued afterwards with shaking at 225 rpm. The cells were then spun down at the highest speed for 5 mins and 50–100  $\mu$ L of the transformation mixture was then spread on nutrient agar plates containing 50  $\mu$ g/mL kanamycin and 100  $\mu$ g/mL ampicillin and left incubating at 37°C overnight.

## **4.5. Recombinant Production of the Universal Stress G4LZI3 Protein**

### ***4.5.1 Small Scale Expression Screen***

Several colonies were randomly selected from the transformation plates and transferred into tubes containing 5 ml YT broth and 100  $\mu$ g/ml ampicillin (50  $\mu$ g/mL kanamycin and 100  $\mu$ g/mL ampicillin for the DH5 $\alpha$  cells). These were allowed to grow for 4hrs with shaking at 37°C while observing that the OD<sub>600</sub> reached 0.4–0.6. Test cultures were set up by taking 1ml of the sample and inducing it with 0.5 mM Isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG); these were grown further for 2 hrs. Control tubes were also set up by taking 1 ml samples and also incubating them for 2 hrs without the addition of IPTG. The remaining 3 ml cultures were made into glycerol stocks by aliquoting equal amounts of the transformation mixture and 30% glycerol, which were stored in -80°C for future use. After two hours, the samples were centrifuged at 5000 rpm for 5 mins and the pellets re-suspended in 50  $\mu$ L 2X sample buffer. These were heated at 100°C for

10 mins and 20  $\mu$ L analyzed on 15% sodium dodecyl sulfate polyacrylamide gel (SDS-PAGE) electrophoresis.

#### ***4.5.2 Large Scale Protein Expression***

Glycerol stocks of the best expressing clone, based on the SDS-PAGE gel results, were used to prepare starter cultures of 100  $\mu$ L in 100 ml LB broth containing 100  $\mu$ g/ml ampicillin and allowed to grow overnight for 18hrs at 37°C. The following morning, 30 ml of the cultures was diluted with 2 L of LB broth and incubated at 37°C until the OD<sub>600</sub> reached between 0.4–0.6. To the culture, 0.5 mM IPTG was added to induce protein expression. The induction temperature was lowered to 25°C and the cells were left to grow shaking overnight for 18 hrs. The cultures were spun down using 250 mL Beckman Coulter™ centrifuge bottles and the cells were then harvested by centrifugation at 20 000 Xg for 20 mins. The resultant pellets were recovered and weighed followed by the re-suspension of each gram of wet pellet in 5 mL cell lysis buffer. A 100  $\mu$ g/mL of lysozyme and 1 mM of phenylmethylsulfonyl fluoride (PMSF) was then added to the mixtures and the resuspensions were then stored at -80°C until required for purification. To analyze protein expression, a 15% SDS-PAGE gel was used. Protein expression was done repeatedly to cultures up to 20 L volumes to produce sufficient amounts of protein.

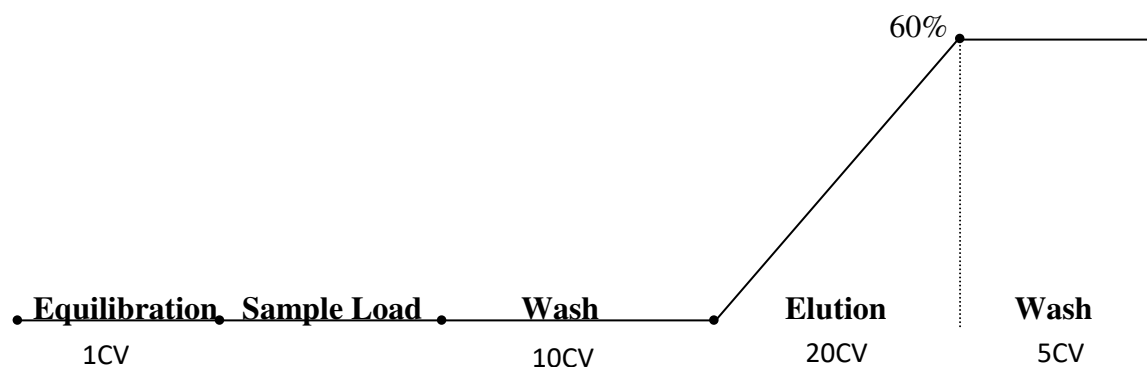
#### ***4.5.3 Cell Extraction and Lysis***

The frozen re-suspended large scale pellets were allowed to thaw on ice for 2–3 hrs and cOmplete EDTA-free Protease inhibitor cocktail tablets (Merck, Darmstadt, Germany), a tablet for each 50 ml sample, were crushed and added to the cell suspensions. The tablets were dissolved in the samples by gentle inversion of tubes. The cells were then homogenized for 4mins with 15 s pulses for every 50 ml sample at a power output of 6.0 whilst simultaneously keeping the sample immersed in an ice bath to prevent excessive heating. This method was used

to disrupt and break the cell walls for the release of the G4LZI3 protein. The cell suspensions were then centrifuged for 30 mins at 20 000 Xg using Beckman Coulter polycarbonate aluminium tubes and the total bacterial lysate separated from the pellet. All resultant lysate was collected, filtered through a 0.45 µm Millipore™ syringe filter and transferred to a schott bottle, which was kept on ice.

#### ***4.5.4 Affinity Purification of G4LZI3 USP***

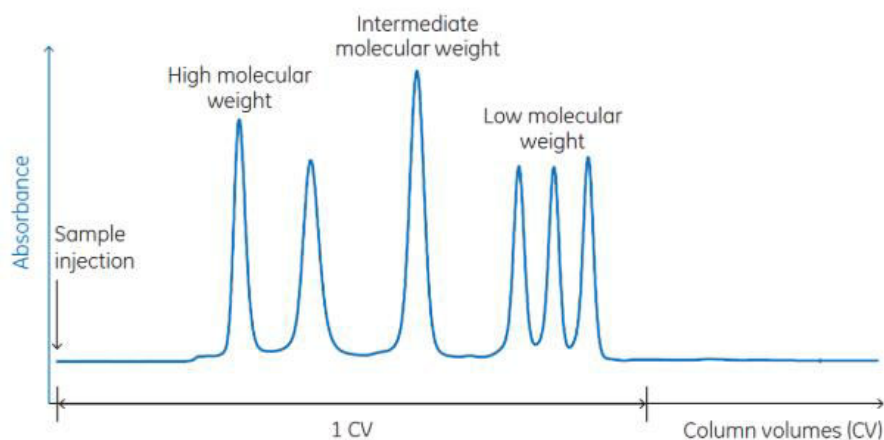
The ÄKTA Prime FPLC system (GE Healthcare Life Sciences) and GE ÄKTA explorer, which are one-step liquid chromatography systems, were used for purification of the G4LZI3 6xHis-tagged fusion protein. A HisTrap FF column packed with Ni-Sepharose® beads was first washed with 5 column volumes (CV) of ethanol, filtered water and elution buffer containing a high concentration imidazole to prevent the binding of all unwanted protein, which may have previously stuck to the column. The beads were thereafter equilibrated with 5CV of equilibration buffer and the lysate loaded. A 10CV washing step was done with equilibration buffer containing low concentrations of imidazole to inhibit the possibility of any unspecific binding. A step-up gradient of 60% concentration was performed with 20 CV of elution buffer followed by a final washing step at 100% concentration with 5 CV of equilibration buffer (Fig 4.1). The G4LZI3 protein was therefore eluted using an isocratic flow of the elution buffer at 5 mL/min. Fractions corresponding to the peak produced by the elution step were then analyzed on an 15% SDS-PAGE gel to determine the extent at which purification had taken place.



**Fig 4.1: Strategy Adopted for Affinity Purification of the G4LZI3 Protein.** The column was first equilibrated before loading the total bacterial lysate followed by a 10CV washing step. The protein was eluted using 20CV of elution buffer containing 400 mM Imidazole.

#### ***4.5.5 Purification of the G4LZI3 Protein by Size Exclusion Chromatography***

After analyzing the SDS-PAGE gel, the fractions were pooled and concentrated down to a final volume of 10 mL using a 3 000MWCO 50 mL Amicon Stirred Cell. The sample was then filtered through a 0.22  $\mu$ m Millipore™ syringe filter and subjected to gel filtration on a HiLoad Superdex75 prep grade (pg) column of size 16 mm diameter and 600 mm bed height (16/600) as well as the HiLoad Superdex200 prep grade (pg) column of size 26 mm diameter and 600 mm bed height (26/600) (GE Healthcare Sciences), both running at a 0.5 ml/min flow rate. The column was equilibrated overnight for 6hrs with 1CV of SEC Storage Buffer. The protein was eluted in 375  $\mu$ L fractions, which were collected based on the resulting chromatogram profile (Fig. 4.2). The fractions were analyzed on a 15% SDS PAGE gel for analysis. The fractions were pooled and concentrated down to a final volume of 3 mL with a concentration of 27 mg/mL using a 10 000MWCO Amicon Ultra-15 Centrifugal Filter Unit. The protein was kept at 4°C until required for crystallization trials. All protein samples that were not utilized were flash frozen in liquid nitrogen and stored at -80°C.



**Fig. 4.2: Strategy Adopted for Size Exclusion Purification of the G4LZI3 Protein.** Purification columns were equilibrated overnight before injecting the partially purified samples. The protein was purified to homogeneity by separating impurities from the protein of interest based on the molecular weights and hydrodynamic radiuses of the various contaminants.

#### ***4.5.6 Sodium dodecyl sulfate polyacrylamide gel (SDS-PAGE)***

SDS PAGE gels were prepared as is presented in Table 2.1 below for the analysis of successful protein expression or the estimation of protein purity. For protein expression, 5  $\mu$ L of the samples were mixed with 50  $\mu$ L 2X sample application buffer, heated at 95°C for 10 mins and 5, 10, 15 and 20 L volumes were loaded onto the gels. For purified samples, 30  $\mu$ L of the purified fractions were collected and mixed with 10  $\mu$ L 4X sample application buffer where loading of 5  $\mu$ L was typically used.

All samples were electrophoresed onto 15% SDS polyacrylamide separating (pH8.8) gels with 5% stacking (pH6.8) gels. The gels were ran at a voltage of 150 V/cm with a current of 30 mA using the PowerPac Basic Power Supply (Bio-Rad) system. Protein expression products were identified by comparison of mass/size with a standard protein ladder (Precision Plus Protein™

Dual Xtra Standard, Biorad) whose sizes range from 5–250 kDa. Gels were stained in 5 mins by incubating the gels at room temperature in AcquaStain Protein Gel Dye (Bulldog Bio).

**Table 4.1: Summarized Buffer Composition of a 15% SDS PAGE gel**

|                                           | Separating Gel<br>(10 mL) | Stacking Gel (5 mL) |
|-------------------------------------------|---------------------------|---------------------|
| <b>Deionized Water</b>                    | 3.550                     | 2.960               |
| <b>40% Bis/Acrylamide 37.5:1</b>          | 3.750                     | 0.500               |
| <b>Separating Buffer</b>                  | 2.600                     | –                   |
| <b>Stacking Buffer</b>                    | –                         | 0.540               |
| <b>10% Ammonium Persulphate<br/>(APS)</b> | 0.100                     | 0.040               |
| <b>TEMED</b>                              | 0.006                     | 0.008               |

## 4.6 Determination of Protein Structure by X-ray Crystallography

### 4.6.1 Protein Crystallization Trials

The concentrated protein sample was diluted into various concentrations (10 mg/ml, 15 mg/ml, 17 mg/ml, 20 mg/ml and 24 mg/ml) using the SEC storage buffer for crystallization trials using the sitting drop diffusion technique. The **Morpheus® Crystallization** screen (Appendix III) which contains 96 conditions of low-molecular weight ligands covering a range of various pH, salt additives, PEGs (polyethylene glycol) and precipitants, was used to set up the experiments for crystal formation. 1ml measure of each precipitant condition was first aliquoted into a D-well precipitation block, which was constantly kept at 4°C for easy access and periodic use. From this block, 50µL of each precipitant condition was aliquoted into a 96-well plate.

With the help of a robot known as the mosquito liquid handling crystal Nanolitre Protein Crystallization Robot TTP Labtech Ltd (UK) (Fig 4.3), the different concentrated protein

samples were added to each plate, sealed with tape and left at various temperatures (room temperature, 15°C and 20°C).

The plates were monitored daily using the LEICA Z16 APO microscope (Fig 4.4) for any crystals and successful hits were fished out and tested to determine if the formed crystals were that of salt or protein. The latter were then used as seed stock to generate more. This was made by transferring the crystal drop into a sterile 500 µL eppendorf tube and re-suspending the crystals in precipitant buffer. The seed stock was then used to set up more sets of plates with 20 mg/ml protein at room temperature.

#### ***4.6.2 Sample Preparation and Shipping***

The plates were examined for the presence of well-structured crystals and selected. These were picked using 22 mm Hampton CrystalCap<sup>TM</sup> HT SPINE pins of various loop sizes according to the crystals' dimensions. The loops used are purposed for SPINE style cap cryocrystallography systems which attach to magnetic CrystalCap SPINE Vials and goniometer bases. Crystals were then flash frozen in liquid nitrogen, taking care to avoid the accumulation of excessive condensation or ice. The pins were injected directly into Unipucks (Universal pucks) that were cooled in liquid nitrogen along with its base plate. Once all positions were filled with crystals, the puck was aligned with the base plate and pushed into place. The lids were then secured firmly with four Philips head screws and stored in liquid nitrogen until ready for the synchrotron visit. The pucks were packed into dewars 'charged' with liquid nitrogen which were labeled with codes and registered on the ISPyB (<https://ispyb.diamond.ac.uk>) and SynchWeb (<https://diamondlightsource.github.io/SynchWeb/>) databases. These were then shipped to the Diamond Light Source Synchrotron in the United Kingdom via DHL. Data was collected and pre-processed remotely.

### ***4.6.3 Crystal Mounting and Data collection***

The latest NoMachine Client software for free remote desktop access was downloaded and installed. A new connection was created by using the NX Protocol. The hostname was set to nx.diamond.ac.uk and the port adjusted to 4000. Password Authentication without the use of proxy was selected and the name “Connection to nx.diamond.ac.uk” was given to the new connection. The connection was opened and Federal ID and password filled in with the creation of a new virtual desktop. To observe live feed of the robot webcams and sample positions, ISPyB was logged into and the appropriate beamline status selected for viewing. Data was collected using Generic Data Acquisition (GDA) software.

The sample was loaded via the Diamond BART robot sample changer on the IO3 MX beamline, which encompasses a standard and working wavelength of 0.976Å (12.7keV) and 0.6–2.48Å range respectively. Mounting was done from the ‘Samples’ perspective. After loading, the crystal was centered by clicking on the OAV crystal image and rotating the crystal +/- 90° or +/- 180° manually or by using the extra centering tools that employ 3-click or automated centering. Since the majority of all the crystals selected were relatively small, the diffraction scan grid function was accessed from the “Grid Scan” perspective. This assisted in detecting the positioning of the crystal within the loop as well as x-ray centering especially on the best portions of the crystal. The x-ray grid scans were additionally used for automatic centering of the crystal sample. Once the sample was in focus, a 2-dimensional grid scan was constructed over the entirety of the loop. The image was scored by Bragg coordinates followed by the construction of a linear grid scan perpendicular to the best diffracting position of the previous 2D scan. Data was collected automatically across 360°. An x-ray fluorescence spectrum (MCA) was collected for detecting elements within the sample by generating peaks corresponding to energies

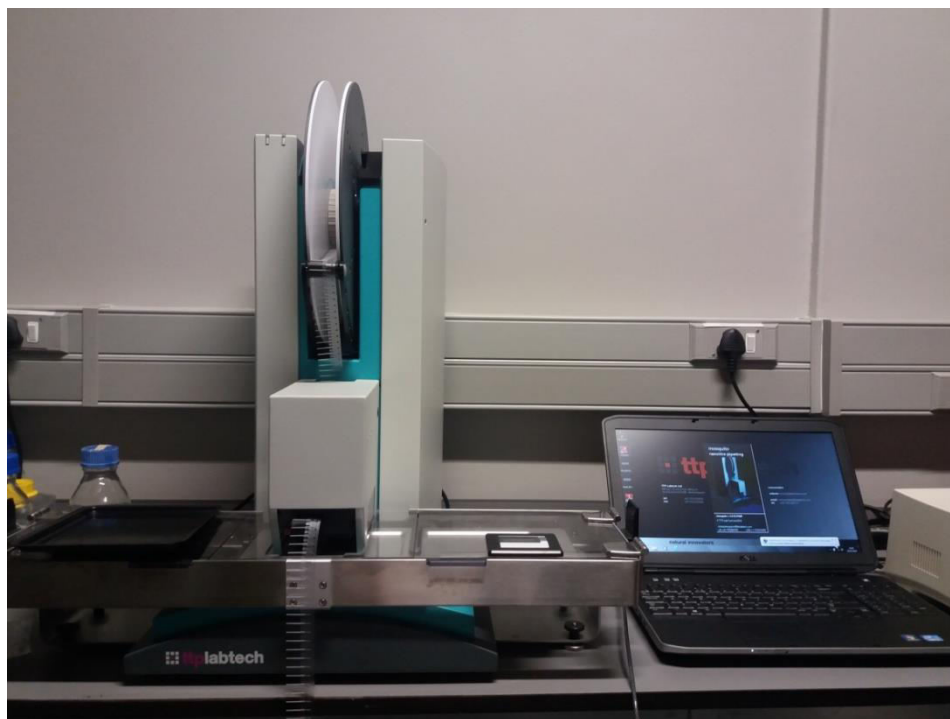
released by light emitted from the sample caused by incident x-ray beams. All data collected were reduced to fast\_dp and xia2 files via the Automatic Software Pipeline.

#### ***4.6.4 Structure Software and Data processing***

Datasets of the best diffractions were downloaded from the ISPyB database while another set of compulsory files were produced by various servers for automated molecular replacement using MR\_Rosetta (Terwilliger *et al.*, 2012). Fragment files were generated via the Robetta fragment server from the University of Washington (<http://robetta.bakerlab.org/fragmentsubmit.jsp>). The amino acid sequence of the G4LZI3 protein was pasted into the form and an email address provided for notification of job completion, which provided two downloadable fragment files (aat000\_03\_05 and aat000\_09\_05) were downloaded. The HHpred server (<https://toolkit.tuebingen.mpg.de/tools/hhpred>) was accessed to detect protein homology and alignment between various search models by HMM-HMM comparison (Söding, 2004).

The amino acid sequence of the G4LZI3 protein was pasted in the HHpred. form and the job submitted to generate a single .hhr alignment file. These, in addition to the amino acid sequence of the G4LZI3 protein in the form of a text file as well as a homologue were all used as input files and template files to run MR-ROSETTA.

Manual building of the resulting density was performed in Coot and supplemented by automated building using Autobuild in *PHENIX* (Adams *et al.*, 2011). The structure was refined in both Coot and *PHENIX* and solutions visualized and superimposed using Chimera (Pettersen *et al.*, 2004). The homologues chosen was the crystal structure of hypothetical protein, PHO823 from *Pyrococcus horikoshii* (PDB ID: 2DUM), which shares 31.06% homology with the G4LZI3 protein.



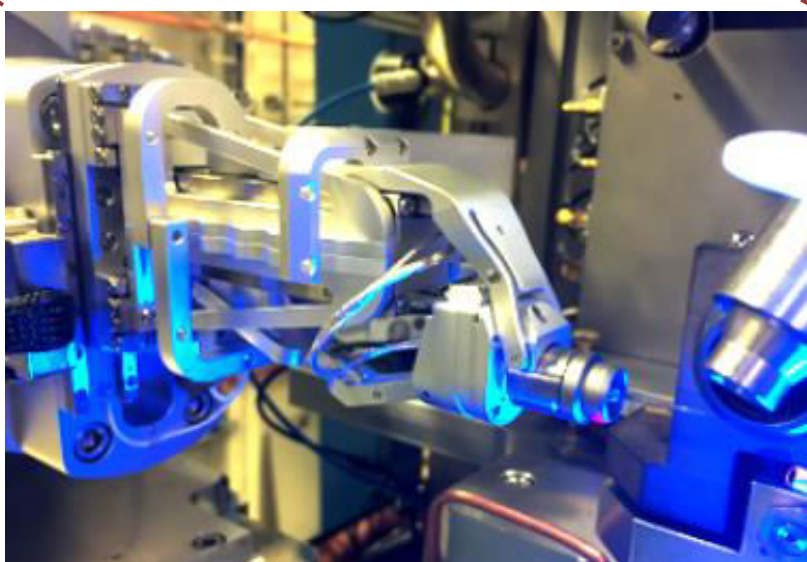
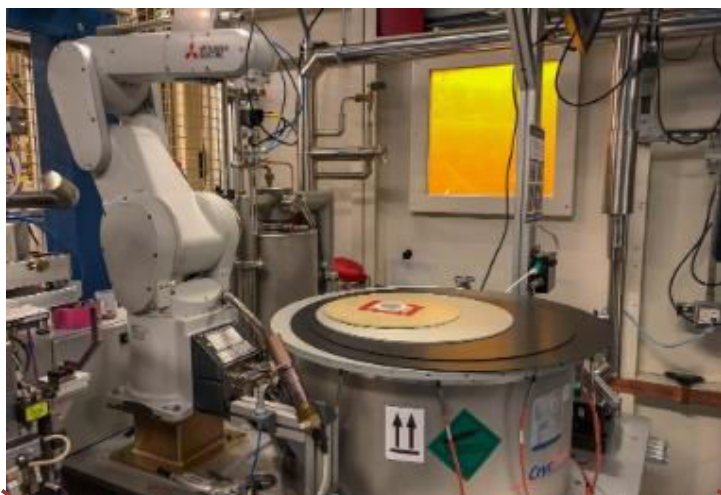
**Fig. 4.3: The mosquito liquid handling crystal Nanolitre Protein Crystallization Robot TTP Labtech Ltd (UK).** All crystal plates were set up using this equipment.



**Fig. 4.4: The LEICA Z16 APO microscope.** Crystals were viewed under this microscope which is attached to the SCHOTT KL 1500 LCD Fiber Optic Light Source.



**Fig. 4.5: X-ray generator (Rigaku MSC Micromax 007 dector).** Crystals were diffracted for the detection of salt or protein crystals



**Fig. 4.6: IO3 MX beamline, Diamond Light Source Synchrotron.** Image shows the BART robot exchanger with the attached SmarGon goniometer. Image taken from <https://www.diamond.ac.uk/Instruments/Mx/IO3.html> [Accessed 1/12/2019]S.

## References

- Adams PD, Afonine PV, Bunkóczi G, Chen VB, Echols N, Headd JJ, Hung L-W, Jain S, Kapral GJ and Kunstleve RWG (2011). The Phenix software for automated determination of macromolecular structures. *Methods* **55**(1): 94-106.
- Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC and Ferrin TE (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of Computational Chemistry* **25**(13): 1605-1612.
- Söding J (2004). Protein homology detection by HMM–HMM comparison. *Bioinformatics* **21**(7): 951-960.
- Terwilliger TC, Dimaio F, Read RJ, Baker D, Bunkóczi G, Adams PD, Grosse-Kunstleve RW, Afonine PV and Echols N (2012). phenix. mr\_rosetta: molecular replacement and model rebuilding with Phenix and Rosetta. *Journal of structural and functional genomics* **13**(2): 81-90.

# **Chapter 5: Closing the gap: an atomistic structural and functional perspective of *S. mansoni* universal stress G4LZI3 protein in complex with phenolic compounds**

Manuscript submitted to: Biomolecules (biomolecules-663034)

# Closing the gap: an atomistic structural and functional perspective of *S. mansoni* universal stress G4LZI3 protein in complex with phenolic compounds

<sup>1</sup>Priscilla Masamba,<sup>1</sup>Geraldene Munsamy and <sup>1,2\*</sup>Abidemi Paul Kappo

<sup>1</sup>Biotechnology and Structural Biochemistry Group, Department of Biochemistry and Microbiology, University of Zululand, KwaDlangezwa 3886, South Africa; presh4u@rocketmail.com (PM);

geraldenemunsamy@gmail.com (GM); <sup>2</sup>Molecular Biophysics and Structural Biology (MBSB) Group, Department of Biochemistry, Faculty of Science, University of Johannesburg, Kingsway Campus, Auckland Park 2006, South Africa; crystalloppaul@gmail.com (APK)

\* Correspondence: Prof Abidemi Paul Kappo, Molecular Biophysics and Structural Biology (MBSB) Group, Department of Biochemistry, Faculty of Science, University of Johannesburg, Kingsway Campus, Auckland Park 2006, South Africa; crystalloppaul@gmail.com Tel: +27 72 084 3110, Fax: +27 11 559 2605

**Abstract:** For decades, Praziquantel has been the undisputed drug of choice for all schistosome infections, but rising concerns due to unelucidated mechanism of action of the drug and unavoidable reports of emerging drug resistant strains has necessitated the need for alternative treatment drug. Moreover, the current apprehension has been reinforced by the total dependence on the drug for treatment hence, the search for novel and effective anti-schistosomal drugs. This study made use of bioinformatic tools to determine the structural binding of the Universal G4LZI3 stress protein (USP) in complex with ten polyphenol compounds, thereby highlighting the effectiveness of these recently identified ‘lead’ molecules in the design of novel therapeutics targeted against schistosomiasis. Upregulation of the G4LZI3 USP throughout the schistosome multifaceted developmental cycle sparks interest in its potential role as a druggable target. The integration of *in silico* tools provides an atomistic perspective into the binding of potential inhibitors to target proteins. This study therefore, implemented the use of molecular dynamic (MD) simulations to provide functional and structural insight into key conformational changes upon binding of G4LZI3 to these key phenolic compounds. Post-MD analyses revealed unique structural and conformational changes in the G4LZI3 protein in complex with curcumin and catechin respectively. These systems exhibited the highest binding energies, while the major interacting residues conserved in all the complexes provides a route map for structure-based drug design of novel compounds which could possess enhanced inhibitory potency and minimal toxicity against the G4LZI3 protein. This study suggests an alternative approach for the development of anti-schistosomal drugs using natural compounds.

**Keywords:** G4LZI3; polyphenols; praziquantel; schistosomiasis

## Introduction

A significant number of the world’s most impoverished, prejudiced, marginalized and vulnerable populations especially among the African, Asian and Latin American tropical and subtropical regions of the world still experience the devastating effects of infectious neglected tropical diseases [1,2]. Schistosomiasis is no stranger to this group of illnesses and stands as one of the most chronic and debilitating parasitic infections that is commonly associated with limited access to health care services and clean drinking water, inadequate nutrition, poor sanitation and sub-standard social behaviours,

which altogether bear negative effects on child development, pregnancy outcomes and agricultural productivity, thereby posing major hindrances to socio-economic growth [3-5]. Over 207 million people in more than 76 countries of the world carry the brunt of infection, while a staggering 779 million are at possible risk of infection with most cases emanating from sub-Saharan Africa [6,7]. This disease is accompanied by a myriad of momentous and life-threatening chronic complications that include liver fibrosis, cirrhosis and bladder calcification, muscle weakness, visual impairment, severe kidney damage, microscopic bleeding, portal and pulmonary hypertension, cognitive and physical under-development in children, as well as closely associated ties to bladder cancer and HIV/AIDS [8-10].

For decades, the pyrazino-isoquinoline derivative, PZQ has and still remains the sole line of chemotherapeutic treatment against schistosomiasis and has taken precedence in countless disease-control programs, particularly in developing countries as it is readily available, relatively affordable and possesses high levels of tolerability and efficacy. Minimal and short-term side effects are displayed by the drug and high egg reduction and cure rates of up to 95% are expected after treatment [7,11,12]. In addition to these minor setbacks are more serious concerns about PZQ; the most alarming being reports on reduction of the drug's efficacy and schistosome resistance due to drug-selective pressure [12]. Amidst this, are other disadvantages such as the drug inactiveness against early schistosome stages, failure of the drug to reverse the chronic complications that accompany the disease and the fact that even after decades of research, little is known about the exact mechanism of action of the drug [13]. Hence, an urgent need for the discovery of alternative treatment is required and a keen interest has been sparked in studying the effects of natural compounds for schistosome treatment.

Secondary metabolites from plants are gradually gathering attention as possible drug leads for schistosomiasis treatment and control [14]. A class of such compounds is the polyphenols, which are ubiquitous plant organic compounds shown to exhibit beneficial effects in the treatment, prevention and control of cardiovascular and neurodegenerative disorders, obesity and cancer [7,11,12,15,16]. Additionally, polyphenols also possess the ability to scavenge free radicals and reduce inflammation [17]. Schistosome infections, and as seen in all other chronic infections caused by helminth parasites, trigger the activation of host immune cells (immunomodulation) that in turn generate reactive oxygen species and oxidative damage via the production of pro-inflammatory secretions such as leukocytes and eosinophils [18-20]. A study by Aanandhi and co-workers demonstrated molecular simulations of the hypothetical *in-silico* inhibition of the Universal Stress Protein (USP) Rv1636 strain from *Mycobacterium tuberculosis* by a select group of natural polyphenols [21]. This current study adopted a similar approach in determining possible inhibitors of the G4LZI3 USP from *Schistosoma mansoni* using a range of polyphenols. In several organisms including schistosomes, USPs are induced in response to a myriad of environmental insults such as drought, high temperatures, salinity, toxic chemicals and antibiotics, heavy metals, un-couplers of the electron transport chain and DNA damaging agents such as oxidants to provide protective and regulatory effects that aid their survival [3,22-24].

A study by Mbah and colleagues highlighted and prioritized the G4LZI3 and Q86DW2 USPs from *S. mansoni* and *S. japonicum* respectively as attractive 'lead' compounds for future schistosomiasis interventions [25]. However, the exact molecular and biochemical mechanism of this class of proteins is yet to be elucidated despite widespread insight into the regulation and physiology of USPs through various studies. This current study therefore integrated the use of *in silico* tools to gain an atomistic perspective into the binding of a selective range of polyphenols to the G4LZI3 USP based on their potential inhibitory activity against universal stress proteins. In the absence of any USP experimental structures derived from schistosomes, the structure of the G4LZI3 protein was predicted computationally and used for molecular docking. The best five complexes were selected for post-MD analyses and subsequent binding free energy calculations. This study suggests an alternative approach for the development of anti-schistosomal drugs that makes use of natural compounds with high efficacy.

## Materials and Methods

### *Homology Modeling of G4LZI3*

The amino acid sequence for the Universal Stress G4LZI3 protein from *Schistosoma mansoni* was accessed from NCBI (Accession number AB023721.2) [26], followed by the generation of predictive 3D structure using the built-in version of modeller in Chimera [27]. A Ramachandran plot for the analyses of bond angles and torsional strain was generated using RAMPAGE [28]. Results showed that 98% of all residues were in the favoured regions, while 2% of all residues were in the allowed regions. This left two outliers, none of which formed part of the active site of the protein. The fold reliability was assessed using the ProSA-web server, which predicts the energy profile of a structure based on its Z-score. The Z-score of the predicted model of G4LZI3 was -5.4 [29].

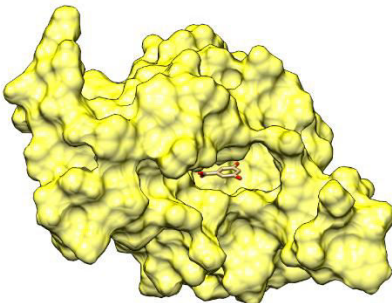
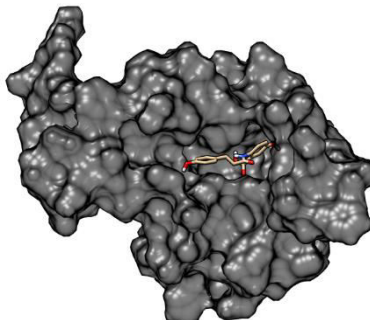
### *Ligand Acquisition and Preparation*

The 3-D structures of the polyphenol compounds listed in Table 1 were obtained from PubChem [30]. These compounds were optimized using the Avogadro molecular editor [31]. The G4LZI3 model was prepared for docking using the Chimera software [32].

### *Molecular Docking*

The Molecular docking software utilized in this study included Raccoon [33], Autodock Graphical user interface supplied by MGL tools[34] and AutoDockVina [35] with default docking parameters. Prior to docking, Gasteiger charges were added to the compounds and the non-polar hydrogen atoms were merged to carbon atoms. The compounds were then docked into the binding pocket of G4LZI3 by defining the grid box with 1 Å spacing and 46 × 60 × 56 size pointing in x, y and z directions. The docking scores for each complex is presented in Table 1 and the top five complexes with the most negative binding score were then subjected to molecular dynamic simulations.

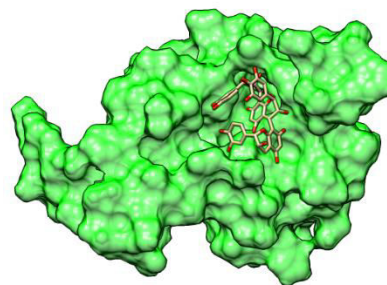
Table 1: 3D Structures and Molecular Docking Scores of Polyphenol Compounds.

| <i>Compound</i>          | <i>Molecular formula</i> | <i>3D structure</i>                                                                  | <i>Docking score</i> |
|--------------------------|--------------------------|--------------------------------------------------------------------------------------|----------------------|
| <i>A-resorcylic acid</i> | $C_7H_6O_4$              |   | -5.9                 |
| <i>Avenanthramide A</i>  | $C_{16}H_{13}NO_5$       |  | -7.4                 |

---

*Catechin*

$C_{15}H_{14}O_6$

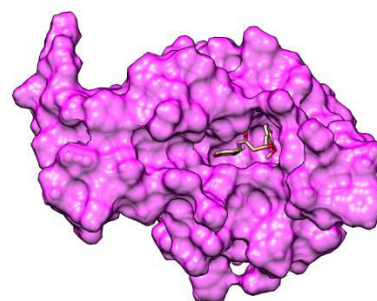


-9.4

---

*Curcumin*

$C_{21}H_{20}O_6$

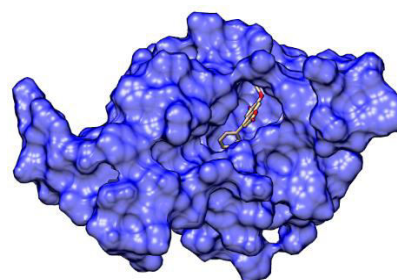


-7.9

---

*Dalbergin*

$C_{16}H_{12}O_4$



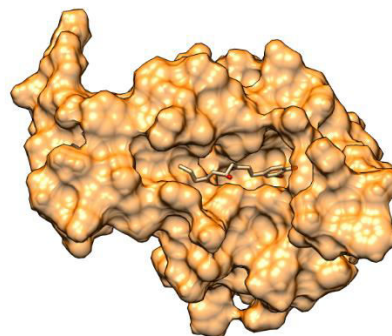
-8.8

---

---

*Gingerol*

$C_{17}H_{26}O_4$

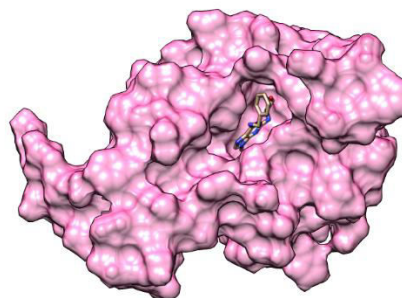


-6.1

---

*Olomoucine*

$C_{15}H_{18}N_6O$

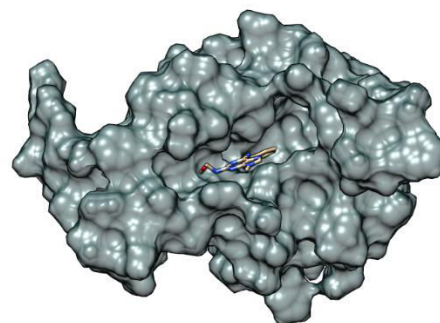


-6.0

---

*Pterostilbene*

$C_{16}H_{16}O_3$



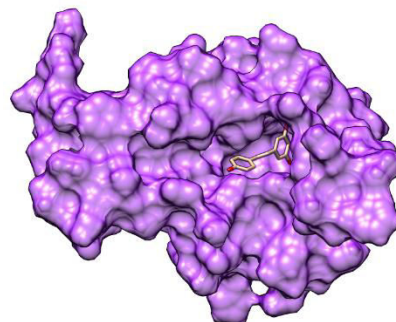
-6.2

---

---

*Resveratrol*

$C_{14}H_{12}O_3$

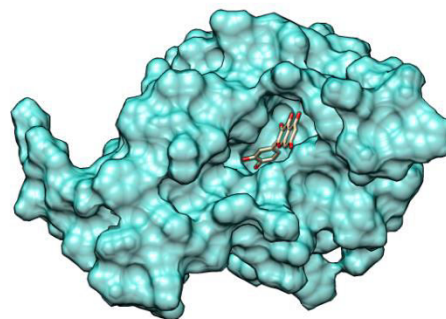


-7.0

---

*Quercetin*

$C_{15}H_{10}O_7$



-8.4

---

### *Molecular Dynamic (MD) Simulations*

Molecular dynamic (MD) simulations provide a robust tool to explore the physical movements of atoms and molecules, thus providing insight into the dynamic evolution of biological systems. The MD simulations were performed using the GPU version of the PMEMD engine provided within the AMBER 18 package [36]. ANTECHAMBER was used to generate atomic partial charges for the compounds by utilizing General Amber Force Field (GAFF) procedures. The Leap module of AMBER 18 suspended each system implicitly within an orthorhombic box of TIP3P water molecules ensuring all atoms were within 10Å of any box edge. Leap also allowed for the addition of Na<sup>+</sup> or Cl<sup>-</sup> counter ions for neutralization of all six systems. An initial minimization of 20000 steps was carried out with an applied restraint potential of 500kcal/mol. This was followed by a full minimization of 10000 steps carried out using the conjugate gradient algorithm in the absence of restraints. A gradual heating MD simulation from 0K to 300K was run for 50ps, to ensure that all systems maintained a fixed number of atoms and volume. The solutes within the system were exposed to a potential harmonic restraint of 10kcal/mol and collision frequency of 1ps. Following heating, an equilibration step estimating 500ps for each of the systems was conducted with the operating temperature being kept constant at 300K. Additional features such as the number of atoms and pressure were also kept constant, mimicking an isobaric-isothermal ensemble. The systems pressure was maintained at 1 bar using the Berendsen barostat. The total time for the MD simulation conducted was 100ns. In each simulation, the SHAKE algorithm was employed to constrict the bonds of hydrogen atoms. The step size of each simulation was 2fs and an SPFP precision model was used. The simulations coincided with isobaric-isothermal ensemble, with randomized seeding, constant pressure of 1bar, a pressure-coupling constant of 2ps, a temperature of 300K and Langevin thermostat with collision frequency of 1ps.

### *Post-Dynamic Analysis*

The coordinates of the free enzyme and bound complexes were then saved after every 1ps and the trajectories were analyzed using the CPPTRAJ module employed in AMBER 14 suit. The Root Mean Square Deviation (RMSD) and thermodynamic energies of each system were thereafter investigated.

### *Binding Free Energy Calculations*

The Molecular Mechanics/GB Surface Area method (MM/GBSA)[37] was employed to estimate the binding free energy of each system. This was averaged over 20000 snapshots extracted from the 20ns trajectory. The free binding energy ( $\Delta G$ ) computed by this method for each molecular species (complex, ligand and receptor) can be represented as:

$$\Delta G_{\text{bind}} = G_{\text{complex}} - G_{\text{receptor}} - G_{\text{ligand}} \quad (1)$$

$$\Delta G_{\text{bind}} = E_{\text{gas}} + G_{\text{sol}} - TS \quad (2)$$

$$E_{\text{gas}} = E_{\text{int}} + E_{\text{vdw}} + E_{\text{ele}} \quad (3)$$

$$G_{\text{sol}} = G_{\text{GB}} + G_{\text{SA}} \quad (4)$$

$$G_{\text{SA}} = \gamma \text{SASA} \quad (5)$$

The term  $E_{\text{gas}}$  denotes the gas-phase energy, which consists of the internal energy ( $E_{\text{int}}$ ), Coulomb energy ( $E_{\text{ele}}$ ) and the van der Waals energies ( $E_{\text{vdw}}$ ). The  $E_{\text{gas}}$  was directly estimated from the FF14SB force field terms. Solvation free energy ( $G_{\text{sol}}$ ) was estimated from the energy contribution from the polar states

( $G_{GB}$ ) and non-polar states (G). The non-polar solvation energy ( $G_{SA}$ ) was determined from the solvent accessible surface area (SASA), using a water probe radius of 1.4 Å, whereas the polar solvation ( $G_{GB}$ ) contribution was estimated by solving the GB equation. The terms S and T denote the total entropy of the solute and temperature respectively.

## Results and Discussion

Efforts directed towards the elimination of schistosomiasis in numerous endemic countries have failed dismally for several years. The infection causes the deaths of over 200 000 individuals per year, while scarring survivors with permanent disabilities and chronic life-threatening diseases, 90% of whom are resident in sub-Saharan Africa [38]. Typically associated with schistosome infections is the presence of granulomatous inflammations that result in scarring and fibrosis triggered by immune responses to parasite eggs that unintentionally become wedged in host tissues such as the liver, intestine and genital areas [39,40]. The host response to the parasite serves as a potent producer of Reactive Oxygen Species (ROS) resulting in oxidative stress [41]. This in turn may lead to protein, nucleic acid, lipid and membrane damage once an imbalance between the normal production of ROS and the antioxidant system has been reached, causing harm to the host. In most diseased states, the overproduction of ROS not only results in oxidative stress but also the release of inflammatory signals which in turn play a pathogenic role in the manifestation of inflammatory diseases such as schistosomiasis [42].

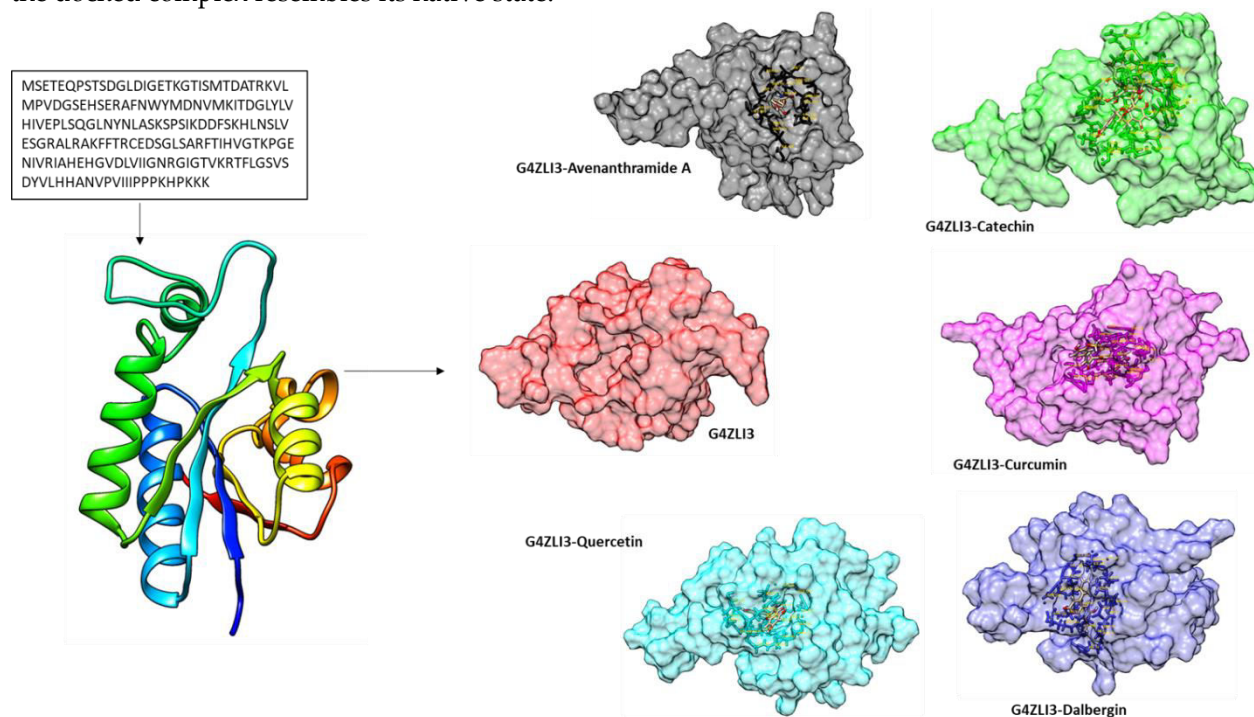
Over the years, various drugs such as oxamniquine, hycanthone and metrifonate were used for the treatment of schistosomiasis but their use has however, been discontinued due to specie selectivity, severe side effects and drug resistance [43]. Praziquantel has been used consistently for treatment against all schistosome strains but has also exhibited drug resistant traits, which have sparked the need to develop new schistosomicides that are not only safe but also highly effective. For the past decade, interest has grown immensely into the search for compounds with anti-parasitic activity from natural sources for drug development [44]. Polyphenols possess a wide variety of biological activities including the unique ability to reduce inflammatory effects by interacting and scavenging ROS/RNS [42]. Traditionally, high-throughput screening (HTS) has been the norm for the identification of biologically active compounds but the introduction of automated (*in silico*) techniques has steadily increased efficacy, cost efficiency and speed in the drug discovery process. Identification of disease proteins remains essential in the drug discovery process as it provides vital information regarding the mechanism of action and effectiveness of a drug molecule [45]. Hence, this study took a particular interest in the G4LZI3 USP, a protein over-expressed during the schistosome developmental cycle that helps the worm to survive the myriad of unfavourable environmental conditions through a variety of unknown mechanisms [24]. Therefore, this study aims to correlate the potential inhibitory activity of selected polyphenol compounds against the G4LZI3 USP with the aid of computational tools.

### *Homology Modeling and Molecular Docking*

In the absence of experimental structures, homology modeling presents itself as an innovative and alternative tool that is not only trustworthy and reliable to a considerable extent, but is also time efficient and cost-effective for the generation of three dimensional protein models [45]. Homology modeling forms the basis for an optimal and effective drug discovery and design process by providing insight into the structural mechanisms of a protein, which involves the interaction between the potential drug and protein, as well as its associated molecular functions. A systematic search of the PDB database was made with the amino acid sequence of the G4LZI3 protein obtained from NCBI. Templates were selected based on those with high similarities of at least 30%, taking into account that the accuracy of predicted models is directly proportional to the increase in similarity as a result of fewer alignment errors [46]. The G4LZI3 model was generated using modeller and the structure accuracy determined by a Ramachandran plot,

which is a 2-dimensional graphic representation of the model that base its analysis on the presence of torsion angles at specific regions within the map [47]. Results indicated that 98% of all residues were present in the favoured regions, while 2% were in the allowed region, leaving two outliers that did not form part of the active site (Fig S1). This result showed the predicted model was of good quality structure and was further corroborated by a Z-score of -5.4 (Fig S2). This parameter not only displays the fold reliability of the protein but in turn predicts the energy profile of the model.

The field of drug discovery has in recent years been aided greatly by the docking computational technique, especially in areas that incorporate lead optimization, drug after-effect predictions and virtual screening. The technique is regarded as a potent complementary tool that screens prospective drug targets and evaluates biomolecular interactions formed by protein-protein or protein-ligand complexes [48]. Docking characterizes the binding affinity of a ligand to a target molecule and the energy released from the complex system [49]. The polyphenols as listed in Table 1 were docked onto the G4LZI3 model using Raccoon [33], AutodockVina [35] and Autodock Graphical User [34] after the preparation of the protein using Chimera software [32] (Figure 1). The docked Avenanthramide A, Catechin, Curcumin, Quercetin and Dalbergin-G4LZI3 complexes were considered the best as they possessed the least energy functions of -7.4, -9.4, -7.9, -8.4 and -8.8 respectively (Figure 1). These complexes were then selected and subjected to MD simulations. The lower the Z-score or binding energy, the closer it can be assumed that the docked complex resembles its native state.

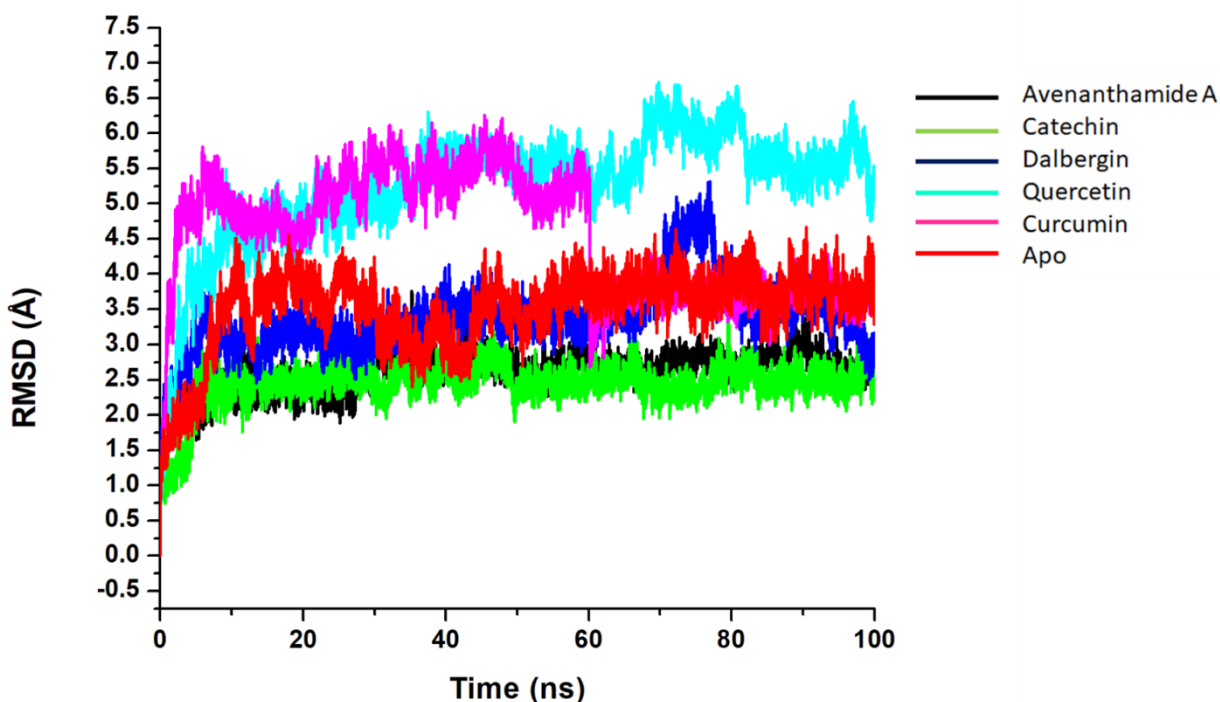


**Figure 1.** Diagrammatic representation of the sequence and homology model of the G4LZI3 protein. The docked complexes of the top five systems were subjected to molecular dynamic simulations.

#### *Assessment of system stability*

Root mean square deviation (RMSD) analysis reveals information that can be used to measure the stability or stereochemical variability across a set of structural models; this parameter being complementary to the crystallographic B factor [50-52]. RMSD analysis was performed to validate the stability of all the systems (Figure 2) and to ensure accuracy of successive post-dynamic analyses. The RMSD value is in addition an indicator of atom mobility during MD simulation [53]. Therefore, the

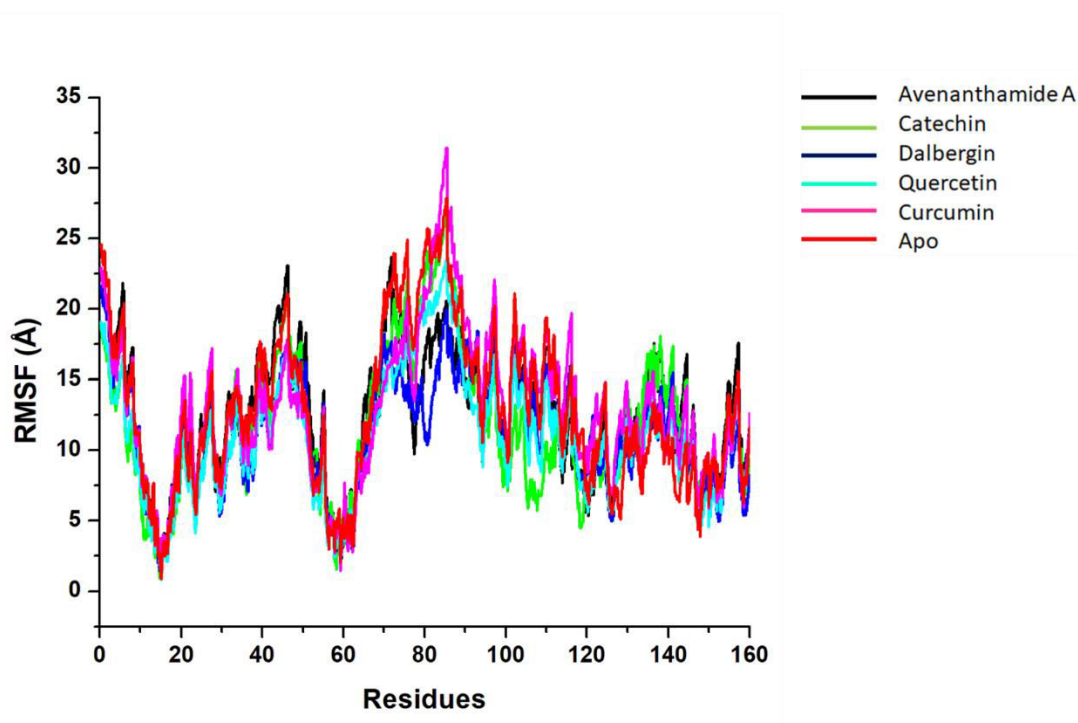
higher the value the higher the mobility and equally vice versa. As observed in Figure 2, all systems reached a state of stability and the initial fluctuation observed in the complexes may be attributed to the adjustment of the ligands as they position themselves within the active site during simulation. The average RMSD of the G4LZI3 protein in complex with Avenanthramide A, Catechin, Curcumin, Dalbergin and Quercetin were 2.61 Å, 2.74 Å, 4.55 Å, 3.44 Å and 5.23 Å respectively (Figure 2). The G4LZI3-Curcumin and G4LZI3-Quercetin complexes exhibited the highest fluctuation followed closely by the 3.55 Å aposystem. System stability of the other docked polyphenols suggests increased stability in ligand binding within the hydrophobic pocket of the G4LZI3 protein [54]. Further analysis was performed to decipher the unique structural versatility observed in these systems.



**Figure 2.** Graph showing Root mean square fluctuation of the C $\alpha$  backbone of the G4LZI3 apo protein (red) as well G4LZI3 in complex with Avenanthramide A (black); Catechin (green); Dalbergin (dark blue); Quercetin (cyan) and Curcumin (magenta).

#### *Protein structural flexibility*

To effectively quantify the flexibility observed in the C $\alpha$  backbone of the G4LZI3 protein, Root Mean Square Fluctuation (RMSF) calculations were performed. Proteins undergo conformational changes even in their free and unbound state. These minor conformations can resemble functional states such as the structure that a protein adopts when it binds to another biological molecule [55,56].



**Figure 3.**Root mean square fluctuation of G4LZI3 apo protein (red) as well G4LZI3 in complex with Avenanthramide A (black); Catechin(green); Dalbergin (dark blue); Quercetin (cyan) and Curcumin (magenta).

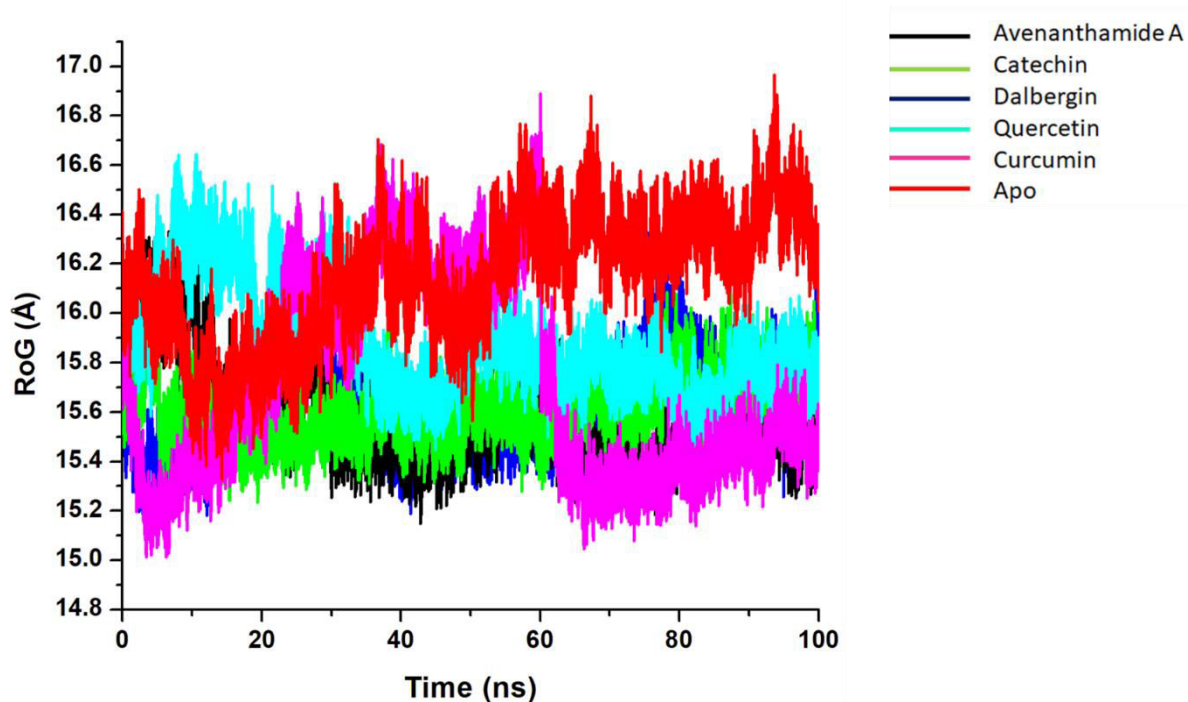
The ‘perturbations’ that affect the populations of the protein ensemble are mediated by modifications, mutations and interactions with other biomolecules and may have a direct impact on related biological mechanisms [57]. Intuitively, the molecular interaction that occurs during the binding of a protein to other molecules may actively restrict the intrinsic flexibility of the binding region in a protein and in its binding target, which results in a significant loss in flexibility as can be observed in Figure 3. The G4LZI3 protein in complex with Catechin, Dalbergin and Quercetin systems exhibited the lowest average residue fluctuations (Figure 3 and Table 2), which may be due to conformational changes, or the tight binding of the ligand, thus effectively reducing the fluctuation of the C $\alpha$  backbone of the protein. Decreased fluctuation in residues ~10–20, ~55–65 and 110–130 highlights the interaction in the binding of each ligand to G4LZI3 protein. The energy contribution of each of these amino acid residues of the G4LZI3 protein was further expounded.

Table 2: Residue fluctuation of all systems during the 100ns MD simulation measured in Angstroms.

|                       | Average | Maximum | Minimum | $\Delta^a$ |
|-----------------------|---------|---------|---------|------------|
| <b>Avenanthramide</b> | 12.36   | 23.68   | 0.85    | 22.83      |
| <b>Catechin</b>       | 11.67   | 27.04   | 0.93    | 26.11      |
| <b>Dalbergin</b>      | 11.35   | 21.63   | 0.08    | 21.55      |
| <b>Quercetin</b>      | 10.86   | 23.51   | 0.81    | 22.70      |
| <b>Curcumin</b>       | 13.09   | 31.43   | 0.09    | 23.42      |
| <b>Apo</b>            | 12.41   | 27.88   | 0.90    | 26.98      |

### *Binding of polyphenols induce structural compactness*

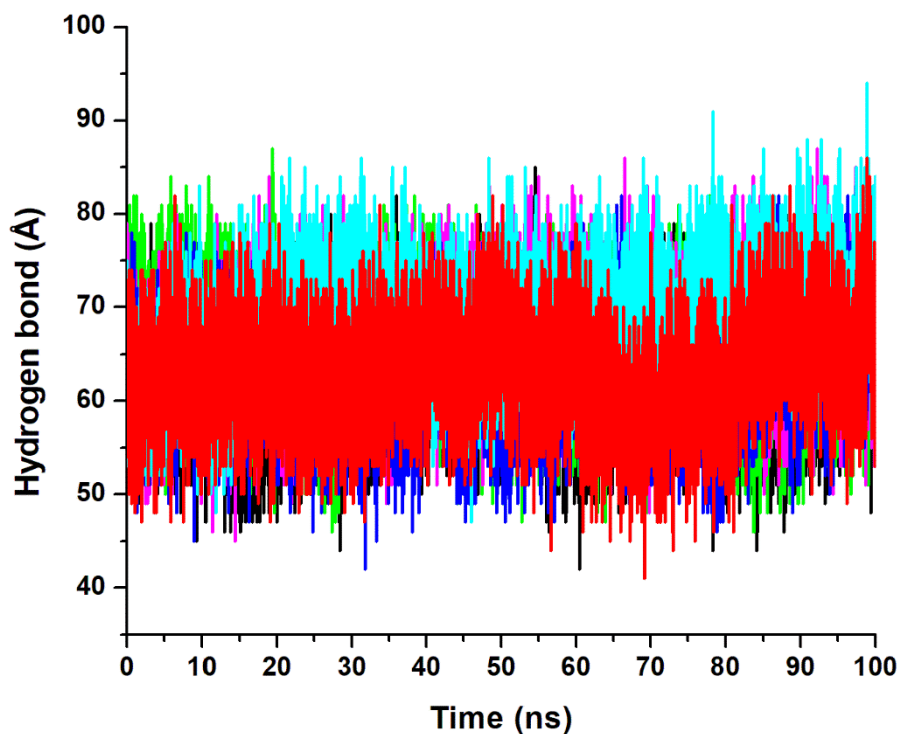
The radius of gyration (RoG) indicates the compactness of the tertiary structure of a protein and this is denoted by the root mean square distance between the centre of gravity and the edges of an object[54,58]. All bound systems exhibited a RoG ranging between  $\sim 15.58$ - $15.89\text{\AA}$ , whilst the G4LZI3apo protein exhibited aRoG of  $16.16\text{\AA}$  (Figure 4). Observation of the figure indicates an increase in the RoG score of the bound systems during the first  $\sim 50\text{ns}$  of the simulation, followed by a maintained decrease in globularity of the bound systems for the remainder of the simulations. The RoG results indicated the bound systems were highly compact in comparison to the apo system, which was the least compact system. The decrease in the RoG of the bound systems is suggestive of ligand binding, resulting from the formation of hydrogen bonds and electrostatic interactions playing a critical role in the binding of each ligand within the catalytic domain, and simultaneously reducing the flexibility of the protein. Interestingly, RMSF and RoG profiles for the G4LZI3 apo protein are consistent with the resultant RMSD profile. Curcumin however, exhibited a compact system during the first  $50\text{ns}$  followed by a sharp decrease in globularity at  $\sim 60\text{ns}$  indicating an increase in the protein compactness resulting in structural stability induced by the binding of Curcumin to the G4LZI3 protein, thus corresponding to its RMSD value.



**Figure 4.**Radius of gyration of G4LZI3 in complex with a range of selected polyphenol inhibitors.

### *Hydrogen Bond Profile*

Figure 5 represents the hydrogen bonds present in each system, with those bound displaying a greater number of hydrogen bonds as compared to the apo system. With the formation of hydrogen bonds comes protein stability [53]. At the start of the simulation, there was an increase in the flexibility of the protein as a larger catalytic space becomes available. Once the ligand binds, the bound systems reach a state of stability facilitated by the formation of hydrogen bond interactions.



**Figure 5.** Stabilizing hydrogen bonds of the G4LZI3 apo protein (red) as well the apo protein in complex with Avenanthramide A (black); Catechin (green); Dalbergin (dark blue); Quercetin (cyan) and Curcumin (magenta).

#### *Binding free energy calculations*

The G4LZI3-polyphenol complexes were ranked based on their binding affinities to the G4LZI3 USP using MM-PBSA calculations, a valuable binding free energy determining tool [59]. MM/PBSA calculations were performed to calculate binding free energies for the five different systems during 100ns of each MD simulation trajectory, which revealed specific binding to G4LZI3. The high relative binding free energies observed in all systems are linked to the increased number of non-covalent interactions observed between each ligand within the active site, resulting in restricted mobility of the interacting amino acid residues. Although all complexes display relatively high binding free energies, the G4LZI3-Catechin complex exhibited the highest binding energy of -55.86kcal/mol, followed closely by the G4LZI3-Curcumin complex with a binding energy of -40.70 kcal/mol.

Figure 6 shows the change in conformation of the G4LZI3 protein upon binding of curcumin and catechinduring 100ns of simulation. The high binding entropy observed in these systems may be as a direct result of the conformational entropy brought about by the unique flexibility change upon the binding of each ligand. Prominent intermolecular interactions observed in all complexes originated from van der Waals and electrostatic forces (Table 3). The favourable electrostatic interactions are neutralized by the polar solvation effect resulting in a much lower electrostatic contribution. As such, large van der Waals interactions contribute to the overall total binding of G4LZI3 in both complexes. Furthermore, the conformational changes associated with the G4LZI3-catechin complex may be as a result of a high number of interacting residues due to its bulky structure. The G4LZI3-Curcumin complex on one hand

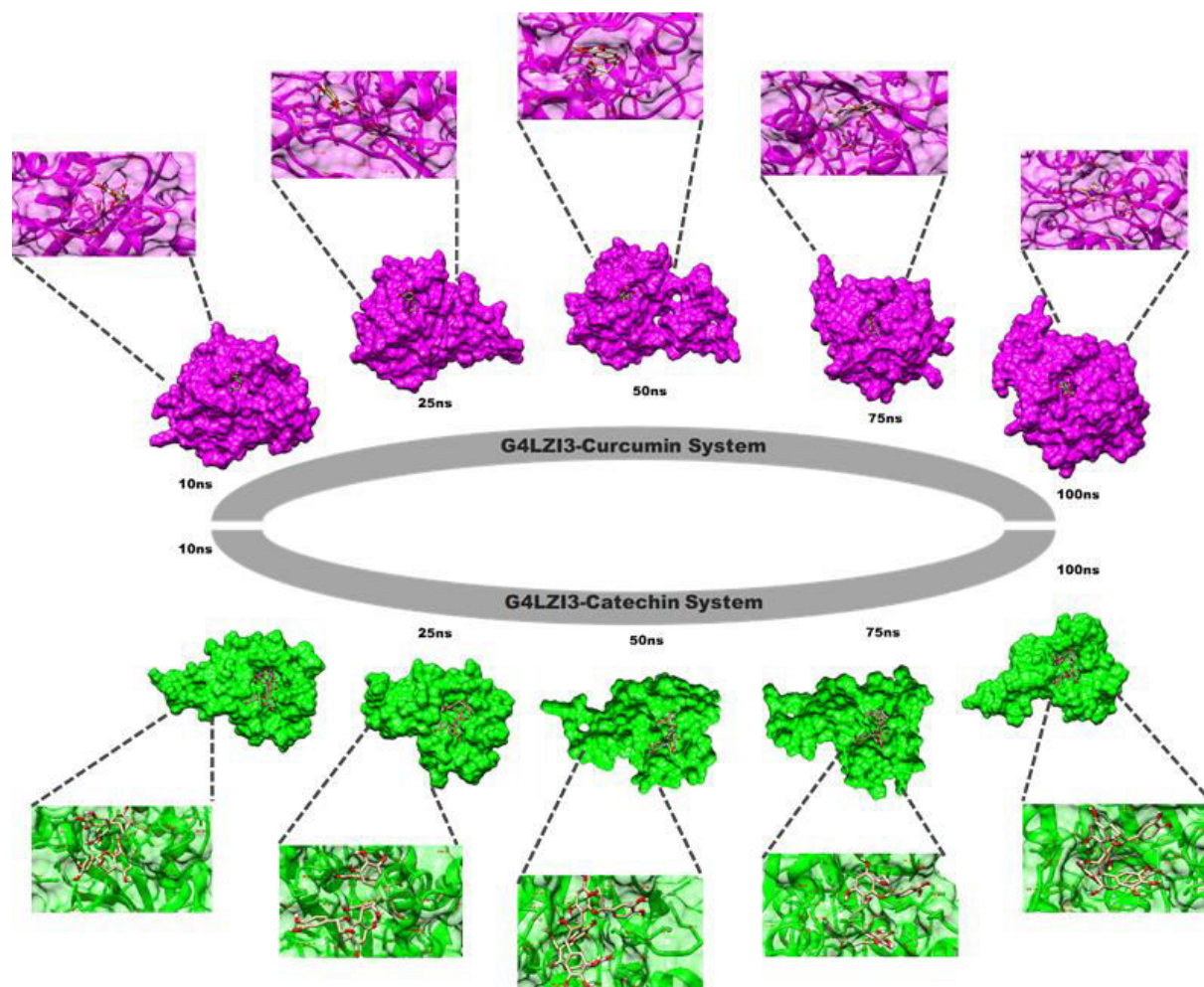
observably displays a significant conformational change 50ns after the bound complex was formed, maintaining a closed structure in its bound state there afterwards. This is in perfect corroboration with the RoG profile of the complex, which portrayed a highly compact and structurally stable complex ~60ns into the simulation analysis. This may be as a result of the high number of hydrogen bonds and electrostatic interactions that occurred, which did not allow the protein assume any other conformational shape. These results suggest improved drug design should focus on increasing van der Waals interactions, while simultaneously decreasing the net electrostatic repulsion upon binding. However, optimization of the net electrostatic interaction remains a difficult task, as the greater the direct electrostatic interactions are between an inhibitor and a protein, the more likely an increase in polar solvation penalty.

**Table 3.** MM/GBSA-based binding free energy profile analysis.

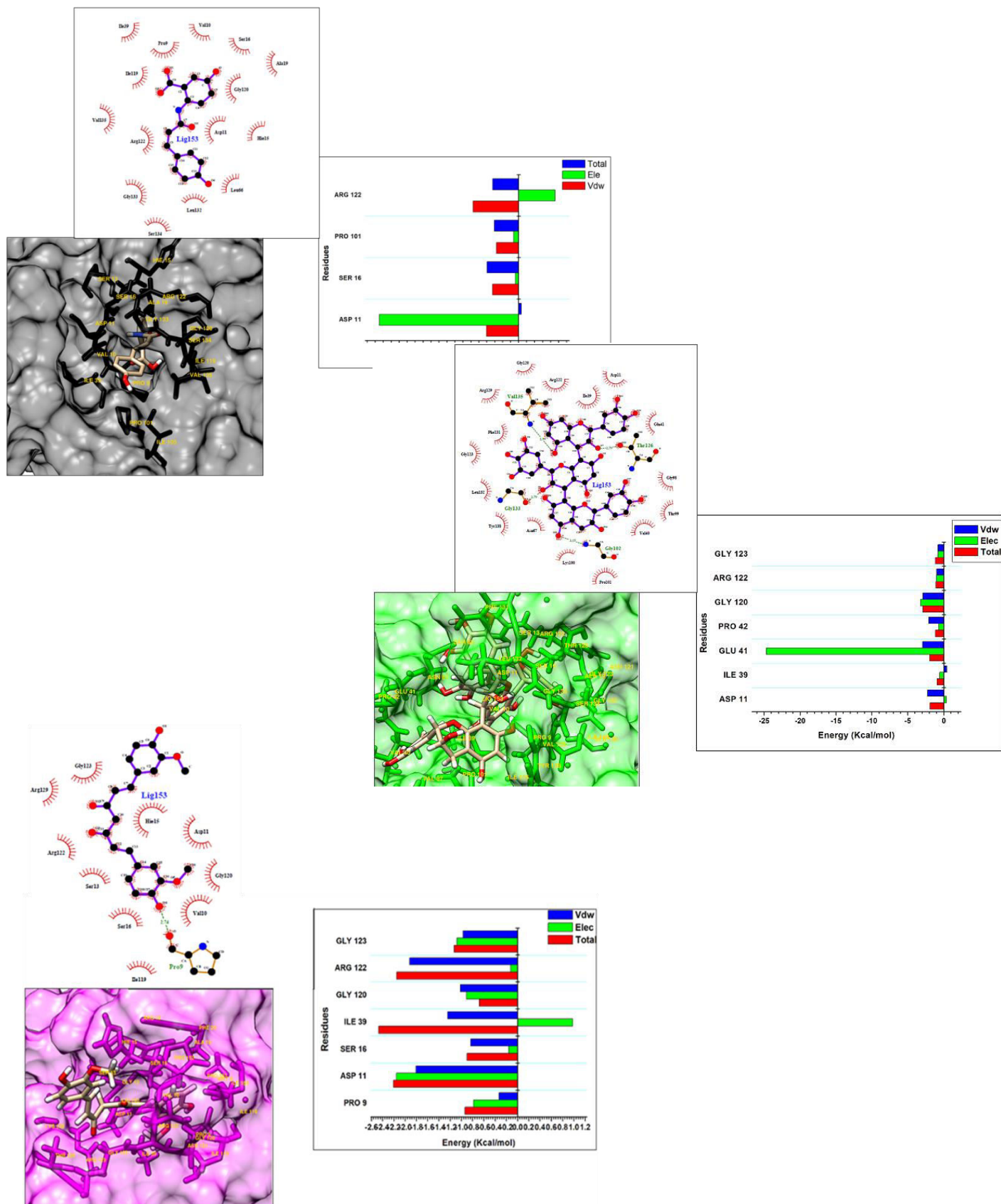
| Systems        | Energy components<br>(kcal/mol) |                  |                  |                  |                   |
|----------------|---------------------------------|------------------|------------------|------------------|-------------------|
|                | $\Delta E_{vdw}$                | $\Delta E_{ele}$ | $\Delta G_{gas}$ | $\Delta G_{sol}$ | $\Delta G_{biod}$ |
| Avenanthramide | -35.48±0.16                     | -24.35±0.43      | -59.84±0.41      | 32.59±0.33       | -27.24±0.16       |
| Catechin       | -55.11±0.11                     | -85.86±0.35      | -140.97±0.33     | 85.11±0.22       | -55.86±0.15       |
| Curcumin       | -49.15±0.08                     | -21.11±0.18      | -70.26±0.20      | 29.56±0.15       | -40.70±0.09       |
| Dalbergin      | -37.84±0.07                     | -21.53±0.14      | -59.37±0.17      | 24.93±0.08       | -34.44±0.10       |
| Quercetin      | -33.06±0.09                     | -37.85±0.36      | -70.90±0.31      | 35.75±0.19       | -35.15±0.14       |

#### *Per-residue energy decomposition*

It was interesting to note that although binding of the compounds occurred at varying spatial landscapes, there was a conservation of interacting residues at the active site among all complexes. Figure 7 depicts the binding energies of various amino acids within the best two and least-favourable complex. Overall, the major interacting residues conserved in all complexes were Asp11; Ile39; Gly120; Arg122 and Gly123. Close-up images displaying these residues can be accessed in the Supplementary information (S4). The lesser the value in kJ/mol, the more involved that particular amino acid is in the binding of the ligand to the complex, thus showing that Asp11 was the highest energy contributing residue in all complexes, forming a hydrogen bond interaction with each ligand [60]. A combination of hydrogen bond interactions (Figure 5) along with hydrophobic interactions (Figure 6) plays an essential role in stabilizing the binding of each ligand within the active site. Identifying these crucial residues and their respective interactions provides basis for further structure-based drug design of new compounds, which could possess inhibitory potency against the G4LZI3 protein. This could also allow for the construction of a pharmacophore model for screening new hits from large databases.



**Figure 6.**Conformational changes of G4LZI3 in association with bound Curcumin and Catechin over 100ns simulation.



**Figure 7.** Per-residue energy decomposition of G4LZI3 in complex with (a) Avenathramide A (black); (b) Catechin (green); and (c) Curcumin (magenta).

## Conclusion

Schistosomiasis is an ancient and major helminthic disease that has seen several abortive attempts for its long-term elimination and control. This study predicted the structure of the G4LZI3 USP and the conformational changes associated with the interaction of the protein to a range of polyphenolic compounds. Relatively high binding free energies were displayed among all the systems, which displayed a high level of compactness and stability, although the G4LZI3-Catechin and G4LZI3-Curcumin exhibited more distinct features. Key conserved interacting residues within the complexes were highlighted and were proposed for the design of new anti-schistosomal agents. A number of studies *in vitro* have been done on curcumin and have elucidated its potential as such. Therefore, more studies should not only be directed towards its mechanism of action as well as that of other polyphenols, but should also focus on substantiating the results revealed in this study.

**Acknowledgments:** Research reported in this article was supported by the South African Medical Research Council (SAMRC) through funding received from the South African National Treasury. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the South African Medical Research Council. Dr. Munsamy's research assistantship was funded by the SAMRC, while the Priscilla Masamba obtained a doctoral bursary from the South African National Research Foundation (NRF). More so, the support of the University of Zululand Research Committee to Priscilla Masamba is greatly acknowledged.

**Author Contributions:** P.M. and G.M. performed the experiments, analyzed the data and wrote the draft of the paper; A.P.K. conceived and designed the experiments, as well as wrote the final copy of the paper.

**Conflicts of Interest:** Authors declare no conflict of interest.

## References

1. Barry, J.A. Social sciences research on infectious diseases of poverty: too little and too late? *PLoS neglected tropical diseases* **2014**, 8, e2803, doi: 10.1371/journal.pntd.0002803.
2. Karunamoorthi, K.; Almalki, M.J.; Ghailan, K.Y. Schistosomiasis: A neglected tropical disease of poverty: A call for intersectoral mitigation strategies for better health. *Journal of Health Research and Reviews* **2018**, 5, 1, doi: 10.4103/jhrr.jhrr\_92\_17.
3. Adenowo, A.F.; Oyinloye, B.E.; Ogunyinka, B.I.; Kappo, A.P. Impact of human schistosomiasis in sub-Saharan Africa. *Brazilian Journal of Infectious Diseases* **2015**, 19, 196-205, doi: 10.1016/j.bjid.2014.11.004.
4. Hajissa, K.; Muhajir, A.E.M.; Eshag, H.A.; Alfadel, A.; Nahied, E.; Dahab, R.; Ali, S.M.; Mohammed, M.; Gaafar, M.; Mohamed, Z. Prevalence of schistosomiasis and associated risk factors among school children in Um-Asher Area, Khartoum, Sudan. *BMC research notes* **2018**, 11, 779, doi: 10.1186/s13104-018-3871-y.
5. Weifeng, G.; Joseph, W.; Zhisheng, D.; Yumin, Z.; Wei, H. Advances in diagnosis of schistosomiasis. *Microbiology: Current Research* **2018**, 2, 14 - 19 doi: 10.4066/2591-8036.18-404.
6. Allam, G. Immunomodulatory effects of curcumin treatment on murine schistosomiasismansoni. *Immunobiology* **2009**, 214, 712-727, doi: 10.1016/j.imbio.2008.11.017.
7. McManus, D.P.; Dunne, D.W.; Sacko, M.; Utzinger, J.; Vennervald, B.J.; Zhou, X.-N. Schistosomiasis (Primer). *Nature Reviews: Disease Primers* **2018**, doi: 10.1038/s41572-018-0013-8.
8. Nascimento, L.S.; Ebert, J.P.; Teles, F.V.; Cruz, D.B. Update on *Schistosomiasismansoni*: Epidemiology, Pathogenesis, Diagnosis, Treatment and Prevention. *JSM Gastroenterology and Hepatology* **2016**, 4, 1068, doi: 10.1016/s0001-706x(01)00086-9.
9. Oyinloye, B.; Adenowo, A.; Kappo, A. Reactive oxygen species, apoptosis, antimicrobial peptides and human inflammatory diseases. *Pharmaceuticals* **2015**, 8, 151-175, doi: 10.3390/ph8020151.
10. Poggensee, G.; Feldmeier, H. Female genital schistosomiasis: facts and hypotheses. *Actatropica* **2001**, 79, 193-210, doi: 10.1016/s0001-706x(01)00086-9.
11. Chan, J.D.; Zarowiecki, M.; Marchant, J.S. Ca<sup>2+</sup> channels and praziquantel: a view from the free world. *Parasitology international* **2013**, 62, 619-628, doi: 10.1016/j.parint.2012.12.001.

12. Wang, W.; Wang, L.; Liang, Y.-S. Susceptibility or resistance of praziquantel in human schistosomiasis: a review. *Parasitology research* **2012**, 111, 1871-1877, doi: 10.1007/s00436-012-3151-z.
13. Vale, N.; Gouveia, M.J.; Rinaldi, G.; Brindley, P.J.; Gärtner, F.; da Costa, J.M.C. Praziquantel for schistosomiasis: single-drug metabolism revisited, mode of action, and resistance. *Antimicrobial agents and chemotherapy* **2017**, 61, e02582-02516, doi: 10.1128/AAC.02582-16.
14. de Moraes, J. Natural products with antischistosomal activity. *Future medicinal chemistry* **2015**, 7, 801-820, doi: 10.4155/fmc.15.23.
15. Cory, H.; Passarelli, S.; Szeto, J.; Tamez, M.; Mattei, J. The role of polyphenols in human health and food systems: A mini-review. *Frontiers in nutrition* **2018**, 5, doi: 10.3389/fnut.2018.00087.
16. Vauzour, D.; Rodriguez-Mateos, A.; Corona, G.; Oruna-Concha, M.J.; Spencer, J.P. Polyphenols and human health: prevention of disease and mechanisms of action. *Nutrients* **2010**, 2, 1106-1131, doi: 10.3390/nu2111106.
17. Zhang, H.; Tsao, R. Dietary polyphenols, oxidative stress and antioxidant and anti-inflammatory effects. *Current Opinion in Food Science* **2016**, 8, 33-42, doi: 10.1016/j.cofs.2016.02.002.
18. Aloho-Bekong, D.; Agbor, G.; Nkengazong, L.; Tchente, R.N.; Asonganyi, T. Oxidative stress in school children co-infected with *Schistosomamansoni* and soil-transmitted helminths. *Health Sciences and Diseases* **2013**, 12, doi:
19. de Oliveira, R.B.; Senger, M.R.; Vasques, L.M.; Gasparotto, J.; dos Santos, J.P.A.; de BittencourtPasquali, M.A.; Moreira, J.C.F.; Silva Jr, F.P.; Gelain, D.P. *Schistosomamansoni* infection causes oxidative stress and alters receptor for advanced glycationendproduct (RAGE) and tau levels in multiple organs in mice. *International journal for parasitology* **2013**, 43, 371-379, doi: 10.1016/j.ijpara.2012.12.006.
20. Hewitson, J.P.; Grainger, J.R.; Maizels, R.M. Helminthimmunoregulation: the role of parasite secreted proteins in modulating host immunity. *Molecular and biochemical parasitology* **2009**, 167, 1-11, doi: 10.1016/j.molbiopara.2009.04.008.
21. Aanandhi, M.V.; Bhattacharjee, D.; George, P.S.G.; Ray, A. Natural polyphenols down-regulate universal stress protein in *Mycobacterium tuberculosis*: An in-silico approach. *Journal of advanced pharmaceutical technology & research* **2014**, 5, 171, doi: 10.4103/2231-4040.143036.
22. Bangera, M.; Panigrahi, R.; Sagurthi, S.; Savithri, H.; Murthy, M. Structural and functional analysis of two universal stress proteins YdaA and YnaF from *Salmonella typhimurium*: possible roles in microbial stress tolerance. *Journal of structural biology* **2015**, 189, 238-250, doi: 10.1016/j.jsb.2015.01.003.
23. Vollmer, A.C.; Bark, S.J. (2018) Twenty-five years of investigating the universal stress protein: function, structure, and applications, In *Advances in applied microbiology*, pp 1-36, Elsevier.
24. Masamba, P.; Adenowo, A.; Oyinloye, B.; Kappo, A. Universal stress proteins as new targets for environmental and therapeutic interventions of Schistosomiasis. *Int. J. Environ. Res. Public Health* **2016**, 13, 972, doi: 10.3390/ijerph13100972.
25. Mbah, A.N.; Mahmud, O.; Awofolu, O.R.; Isokpehi, R.D. Inferences on the biochemical and environmental regulation of universal stress proteins from Schistosomiasis parasites. *Advances and applications in bioinformatics and chemistry: AABC* **2013**, 6, 15, doi: 10.2147/AABC.S37191.
26. (NCBI), N.C.f.B.I. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information. **1988**, doi: <https://www.ncbi.nlm.nih.gov/>
27. Biasini, M.; Bienert, S.; Waterhouse, A.; Arnold, K.; Studer, G.; Schmidt, T.; Kiefer, F.; Cassarino, T.G.; Bertoni, M.; Bordoli, L. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* **2014**, 42, W252-W258, doi: 10.1093/nar/gku340.
28. Chen, V.B.; Arendall, W.B.; Headd, J.J.; Keedy, D.A.; Immormino, R.M.; Kapral, G.J.; Murray, L.W.; Richardson, J.S.; Richardson, D.C. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D: Biological Crystallography* **2010**, 66, 12-21, doi: 10.1107/S0907444909042073.
29. Wiederstein, M.; Sippl, M.J. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* **2007**, 35, W407-W410, doi: 10.1093/nar/gkm290.

30. Kim, S.; Thiessen, P.A.; Bolton, E.E.; Chen, J.; Fu, G.; Gindulyte, A.; Han, L.; He, J.; He, S.; Shoemaker, B.A. PubChem substance and compound databases. *Nucleic Acids Res.* **2015**, *44*, D1202-D1213, doi: 10.1093/nar/gkv951.
31. Hanwell, M.D.; Curtis, D.E.; Lonie, D.C.; Vandermeersch, T.; Zurek, E.; Hutchison, G.R. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *J. Cheminform.* **2012**, *4*, 17, doi: 10.1186/1758-2946-4-17.
32. Pettersen, E.F.; Goddard, T.D.; Huang, C.C.; Couch, G.S.; Greenblatt, D.M.; Meng, E.C.; Ferrin, T.E. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605-1612, doi: 10.1002/jcc.20084.
33. Ramírez, D.; Caballero, J. Is it reliable to use common molecular docking methods for comparing the binding affinities of enantiomer pairs for their protein target? *International journal of molecular sciences* **2016**, *17*, 525, doi: 10.3390/ijms17040525.
34. Sanner, M.F. The Python interpreter as a framework for integrating scientific computing software-components. *Scripps Res. Inst.* **2008**, *26*, 1-12, doi:
35. Trott, O.; Olson, A.J. AutoDockVina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2010**, *31*, 455-461, doi: 10.1002/jcc.21334.
36. Nair, P.C.; Miners, J.O. Molecular dynamics simulations: from structure function relationships to drug discovery. *In silico pharmacology* **2014**, *2*, 4, doi: 10.1186/s40203-014-0004-8.
37. Ylilauri, M.; Pentikäinen, O.T. MMGBSA as a tool to understand the binding affinities of filamin-peptide interactions. *J. Chem. Inf. Model.* **2013**, *53*, 2626-2633, doi: 10.1021/ci4002475.
38. Barsoum, R.S.; Esmat, G.; El-Baz, T. Human schistosomiasis: clinical perspective: review. *J Adv Res* **2013**, *4*, 433-444, doi: 10.1016/j.jare.2013.01.005.
39. Huang, H.-H.; Rigouin, C.; Williams, D.L. The redox biology of schistosome parasites and applications for drug development. *Current pharmaceutical design* **2012**, *18*, 3595-3611, doi: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3638776/pdf/nihms458708.pdf>
40. Pearce, E.J.; MacDonald, A.S. The immunobiology of schistosomiasis. *Nature Reviews Immunology* **2002**, *2*, 499, doi: 10.1038/nri843.
41. Oke, T.T.; Moskovitz, J.; Williams, D.L. Characterization of the methionine sulfoxide reductases of *Schistosomamansoni*. *The Journal of parasitology* **2009**, *95*, 1421, doi: 10.1645/GE-2062.1.
42. Hussein, A.; Rashed, S.; El Hayawan, I.; El-Sayed, R.; Ali, H. Evaluation of the Anti-schistosomal Effects of Turmeric (*Curcuma longa*) Versus Praziquantel in *Schistosomamansoni* Infected Mice. *Iranian journal of parasitology* **2017**, *12*, 587, doi:
43. El Ridi, R.A.; Tallima, H.A.-M. Novel therapeutic and prevention approaches for schistosomiasis. *Journal of Advanced Research* **2013**, *4*, 467-478, doi: 10.1016/j.jare.2012.05.002.
44. Magalhães, L.G.; Machado, C.B.; Morais, E.R.; de Carvalho Moreira, É.B.; Soares, C.S.; da Silva, S.H.; Da Silva Filho, A.A.; Rodrigues, V. In vitro schistosomicidal activity of curcumin against *Schistosomamansoni* adult worms. *Parasitology research* **2009**, *104*, 1197-1201, doi: 10.1007/s00436-008-1311-y.
45. Munsamy, G.; Soliman, M.E. Homology modeling in drug discovery-an update on the last decade. *Letters in Drug Design & Discovery* **2017**, *14*, 1099-1111, doi: 10.2174/1570180814666170110122027.
46. Jacobson, M.; Sali, A. Comparative protein structure modeling and its applications to drug discovery. *Annu Rep Med Chem* **2004**, *39*, 259-274, doi: 10.1016/S0065-7743(04)39020-2.
47. Kumar, P.; Arya, A. Ramachandran Plot: A simplified approach. [https://www.researchgate.net/publication/330158666\\_Ramachandran\\_plot-\\_A\\_simplified\\_approach/statsdoi](https://www.researchgate.net/publication/330158666_Ramachandran_plot-_A_simplified_approach/statsdoi): [https://www.researchgate.net/publication/330158666\\_Ramachandran\\_plot-\\_A\\_simplified\\_approach/stats](https://www.researchgate.net/publication/330158666_Ramachandran_plot-_A_simplified_approach/stats)
48. Rawal, K.; Khurana, T.; Sharma, H.; Verma, S.; Gupta, S.; Kubba, C.; Strych, U.; Hotez, P.J.; Bottazzi, M.E. An extensive survey of molecular docking tools and their applications using text mining and deep curation strategies. *PeerJ Preprints* **2019**, *7*, e27538v27531, doi: 10.7287/peerj.preprints.27538v1.
49. Rauf, M.A.; Zubair, S.; Azhar, A. Ligand docking and binding site analysis with pymol and autodock/vina. *International Journal of Basic and Applied Sciences* **2015**, *4*, 168, doi: 10.14419/ijbas.v4i2.4123.

50. Pandey, B.; Grover, A.; Sharma, P. Molecular dynamics simulations revealed structural differences among WRKY domain-DNA interaction in barley (*Hordeum vulgare*). *BMC genomics* **2018**, *19*, 132, doi: 10.1186/s12864-018-4506-3.
51. Pitera, J.W. Expected distributions of root-mean-square positional deviations in proteins. *The Journal of Physical Chemistry B* **2014**, *118*, 6526-6530, doi: 10.1021/jp412776d.
52. Schreiner, W.; Karch, R.; Knapp, B.; Ilieva, N. Relaxation estimation of RMSD in molecular dynamics immunosimulations. *Computational and Mathematical Methods in Medicine* **2012**, *2012*, doi: 10.1155/2012/173521.
53. Fu, Y.; Zhao, J.; Chen, Z. Insights into the Molecular Mechanisms of Protein-Ligand Interactions by Molecular Docking and Molecular Dynamics Simulation: A Case of Oligopeptide Binding Protein. *Computational and mathematical methods in medicine* **2018**, *2018*, doi: 10.1155/2018/3502514.
54. Verma, S.; Grover, S.; Tyagi, C.; Goyal, S.; Jamal, S.; Singh, A.; Grover, A. Hydrophobic interactions are a key to MDM2 inhibition by polyphenols as revealed by molecular dynamics simulations and MM/PBSA free energy calculations. *PloS one* **2016**, *11*, e0149014, doi: 10.1371/journal.pone.0149014.
55. Ainsley, J.; Mulholland, A.J.; Black, G.W.; Sparagano, O.; Christov, C.Z.; Karabancheva-Christova, T.G. Structural Insights from Molecular Dynamics Simulations of Tryptophan 7-Halogenase and Tryptophan 5-Halogenase. *ACS Omega* **2018**, *3*, 4847-4859, doi: 10.1021/acsomega.8b00385.
56. Condic-Jurkic, K.; Subramanian, N.; Mark, A.E.; O'Mara, M.L. The reliability of molecular dynamics simulations of the multidrug transporter P-glycoprotein in a membrane environment. *PloS one* **2018**, *13*, e0191882, doi: 10.1371/journal.pone.0191882.
57. Lawrenz, M.; Wereszczynski, J.; Amaro, R.; Walker, R.; Roitberg, A.; McCammon, J.A. Impact of calcium on N1 influenza neuraminidase dynamics and binding free energy. *Proteins: Structure, Function, and Bioinformatics* **2010**, *78*, 2523-2532, doi: 10.1002/prot.22761.
58. Lobanov, M.Y.; Bogatyreva, N.; Galzitskaya, O. Radius of gyration as an indicator of protein structure compactness. *Mol. Biol.* **2008**, *42*, 623-628, doi: 10.1134/S0026893308040195.
59. Hou, T.; Wang, J.; Li, Y.; Wang, W. Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. *J. Chem. Inf. Model.* **2010**, *51*, 69-82, doi: 10.1021/ci100275a.
60. Yang, B.; Lin, S.-J.; Ren, J.-Y.; Liu, T.; Wang, Y.-M.; Li, C.-M.; Xu, W.-W.; He, Y.-W.; Zheng, W.-H.; Zhao, J. Molecular Docking and Molecular Dynamics (MD) Simulation of Human Anti-Complement Factor H (CFH) Antibody Ab42 and CFH Polypeptide. *International journal of molecular sciences* **2019**, *20*, 2568, doi: 10.3390/ijms20102568.



© 2019 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

# **Chapter 6: Recombinant Expression, Purification and Structure Determination of the Universal Stress G4LZI3 Protein**

## **6.1 Introduction**

The main aim of this study was to determine the three-dimensional structure of the Universal Stress G4LZI3 Protein for the purpose of identifying potential small molecule inhibitors that could provide leads towards the biological function of this protein. This chapter describes how heterologous expression of the protein was accomplished. It also shows how purification of the protein to homogeneity was performed, how protein crystals were derived and diffraction data collected for the construction of the G4LZI3 model. The determination of this model serves as a milestone achieved in investigations initiated by several researchers in understanding the contributions rendered by USPs to fundamental biological processes in schistosomes and how they may be targeted for novel drug development. Currently, no crystal structures are available for schistosome USPs making this the first.

## **6.2 Recombinant Expression and Purification of pQE30-G4LZI3 Protein**

### ***6.2.1 Transformation and Small Scale Expression Screen***

The pQE30-G4LZI3 expression construct was transformed into DH5 $\alpha$ , JM109 and M15 cells. The process of transformation causes bacterial cells that have been subjected to a state of competency to engulf exogenous genetic matter through their membrane walls. Although this phenomenon may occur spontaneously in nature, bacteria should be capable of taking up foreign

DNA from their environment (Schutz, 2014). Due to the fact that DNA is hydrophilic, natural internalization into bacteria is generally rare. Hence, several procedures have been devised that create pores in the cell membrane. Chemical treatment with high concentrations of dimethyl sulfoxide (DMSO) and divalent cations such as calcium chloride ( $\text{CaCl}_2$ ), magnesium chloride ( $\text{MgCl}_2$ ), manganese chloride ( $\text{MnCl}_2$ ) and rubidium chloride ( $\text{RbCl}$ ) have proven the most effective (Asif *et al.*, 2017). Electroporation through introduction of an electric field as well as heat shock, which causes depolarization of the membrane resulting in a decrease in the negative potential within the cell's interior. This in turn triggers the flow of DNA into the cell. Incubation of the cells on ice restores the potential of the membrane, while the addition of growth media to the cells results in cell recovery due to immense exposure to conditions of stress.

Positive transformants from the DH5 $\alpha$  transformation plates were used for miniprep, while the expression profile of the protein was screened from the other cells. Colonies grown against the appropriate antibiotics were incubated in small volumes of LB broth and alternating induced and un-induced samples were electrophoresed onto 15% SDS-PAGE gels for analysis as is shown in Fig 6.1a and b for both JM109 and M15 cells respectively. The expected molecular weight of the protein computed by the ExPASy online server (<https://web.expasy.org/protparam/>) is 20.3kDa and evidence that expression was successful in both cell lines is depicted by the presence of a prominent band at the approximate size. Although both cell lines successfully produced the protein on a small scale, JM109 cells visibly expressed more protein. However, both cell lines were tested for large scale expression.

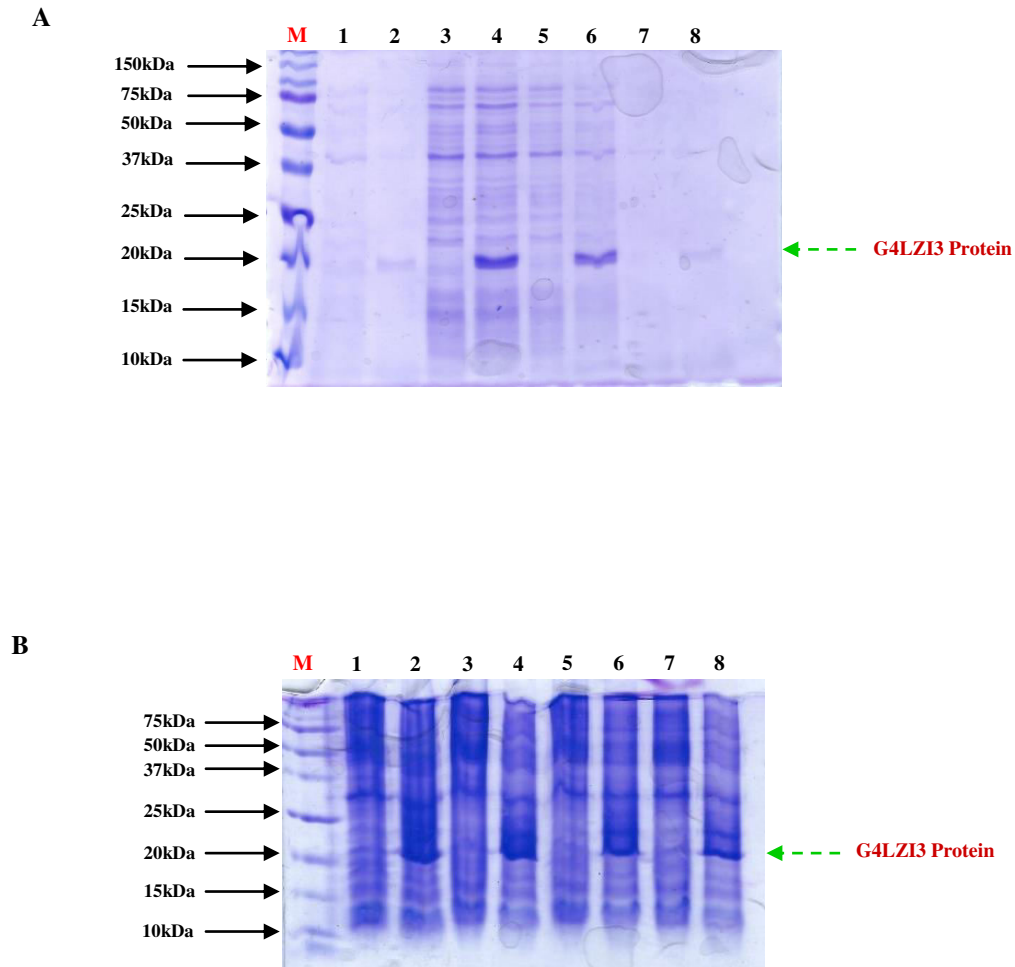
### **6.2.2 Heterologous Over-expression of G4LZI3 Protein**

Colony 1 and 2 from the M15 and JM109 respective expression screens were chosen for large scale expression. An overnight culture consisting 100  $\mu\text{L}$  glycerol stock and 100 mL LB with the

appropriate antibiotics to prevent bacterial contamination were scaled up to 2 L cultures the following morning. After the OD<sub>600</sub> 0.4–0.6 was reached, both cultures were induced by the addition of 0.5 mM IPTG for protein expression. Induction timing was essential for the cells to attain the relevant density as protein expression is most ideal when the majority of cells are viable and healthy and this typically occurs during the exponential growth phase.

Most bacterial strains used primarily for protein expression incorporate promoters which are activated by supplementing growth media with inducers such as IPTG. One such popular induction system is DE3 lysogen, which is incorporated in bacterial strains such as BL21 and JM109. The expression of RNA polymerase (RNAP) by the T7 promoter is controlled by the *lac* repressor, which when induced by IPTG, releases RNAP for gene transcription. In addition, pQE-30 vectors utilize the phage T5 promoter that lacks the *lacI* gene and this problem may be solved by the use of strains such as JM109 that have the *lacI*<sup>q</sup> gene or co-transforming the cell line of choice and the pQE-30 vector with pREP4. Strains such as M15 not only support the T5 promoter but also encompass pREP4, which carries the *lacI* gene that produces the lac repressor for gene transcription and the prevention of leaky expression.

The identities of the expression products were analyzed by a 15% SDS-PAGE gel and bands were visualized after 5 mins following the addition of AcquaStain Protein Gel Dye (Bulldog Bio). Figures 6.2a and b depict successful protein expression for both cell lines. However, as was apparent with the expression screens, JM109 once again visibly exhibited higher protein yields. The gels ran exhibit successive progression of protein expression in the soluble fraction. Cultures of up to 20 L were then set up for the over-expression of sufficient amounts of the G4LZI3 protein.



**Fig. 6.1: Small scale expression profile of colonies transformed with pQE30-G4LZI3.** In (A) expression screening was performed in JM109 cells while in (B) this was done in M15 cells. For both gels Lane 1 shows protein ladder, even-numbered lanes indicate induction of protein expression with 0.5 mM IPTG while odd-numbered lanes depicts total bacterial cell lysates of cultures without induction by IPTG. In (A) two colonies within the induced cultures showed clear expression of the G4LZI3 protein at 20.3 kDa which corresponds in size to the predicted molecular weight of the protein while in (B) all induced colonies expressed the G4LZI3 protein at its predicted molecular weight.

### ***6.2.3 Protein Purification by Affinity and Size-Exclusion Chromatography***

Purification of the soluble fractions to homogeneity was performed by IMAC (Immobilized Metal Affinity Chromatography) affinity and size exclusion chromatography. Affinity chromatography was conducted using the GE Healthcare AKTA FPLC Purifier systems with the use of Ni-Sepharose beads. Prior to subjecting the protein to the column, cocktail inhibitors were added to the lysis buffer re-suspended cells for the prevention of proteolysis. Cell lysis was then conducted by sonification to disrupt the cell wall for the release of the protein and cytoplasm into the supernatant. The resin was washed thoroughly for the removal of non-specific binding proteins and the protein was eluted with high concentrations of imidazole. The histidine tag possesses affinity for certain metals and forms covalent bonds with immobilized ones such as nickel thus retaining the tagged protein (Schutz, 2014). The addition of competitive ions such as imidazole releases the bound protein which may then be recovered. SDS PAGE gel analysis represented by Fig 6.3 indicates obvious elution of the G4LZI3 protein by clearly visible tight and strong bands appearing at 20.3kDa. Fig 6.3A shows a chromatogram of the step gradient beginning at 0%B, the protein eluting at 60%B and until the final 100%B. The presence of bands at approximately 11kDa is possible breakdown or degradation products of the protein.

Semi-pure fractions corresponding to the protein peak were pooled together, concentrated down to 15 mL and subjected to gel filtration chromatography. This technique was used as a final polishing step for purification of the protein as well as a useful procedure for exchanging the buffer of the sample into a buffer that would be suitable for crystallization. The resulting chromatogram shows acceptable separation between the protein and any high molecular weight residues that were still present (Fig. 6.4A). An overlay of all the purification peaks performed on size exclusion showed that the result was consistent (Fig. 6.4B). The purity of the protein was

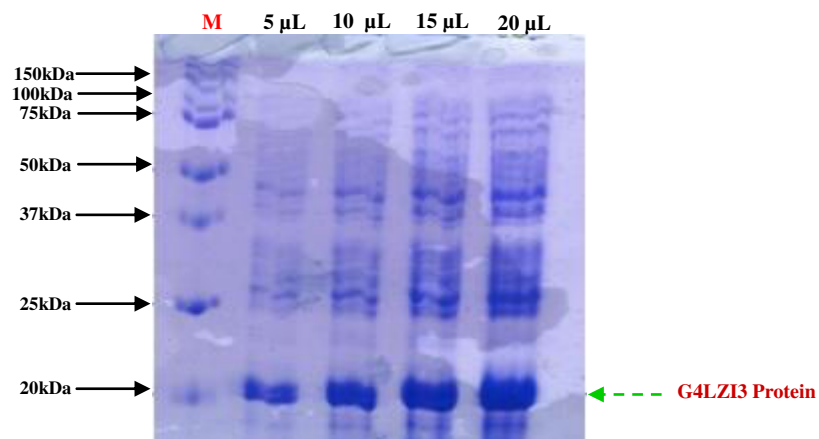
assessed by running the eluted peaks on an SDS PAGE gel revealing that the main peak corresponded to the molecular weight of the protein (Fig 6.5A). The protein was eluted in storage buffer consisting 150 mM Tris and 150 mM NaCl, concentrated down to 27 mg/mL (Fig. 6.5B) and stored until required for crystallization trials.

## **6.3 Investigation of Universal Stress G4LZI3 Protein Structure by X-ray Crystallography**

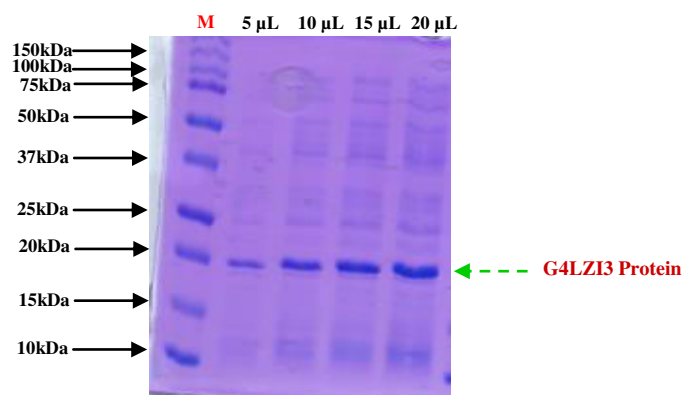
### ***6.3.1 Protein Crystallization Trials of the G4LZI3 USP***

Initial crystals were obtained by sitting drop method in a 1:1 protein precipitate buffer ratio with 10 mg/mL of protein (Table 6.1i) by the vapor-diffusion using the sitting drop technique with the **Morpheus**<sup>®</sup> crystallization screen (Appendix III). Vapor diffusion experiments are the most popular crystallization methods and strive to dehydrate the experimental drop while causing an increase in non-volatile solute concentrations within the crystal drop (Luft *et al.*, 2011). Most drops were clear, while those containing crystals were present in the first wells and had a mosquito-like appearance (Table 6.1 ii). Clear drops often mean one of two theories, either the solution is completely under-saturated or within the metastable supersaturated zone (Luft *et al.*, 2011). These drops were however, simple to distinguish because they displayed no sign of liquid-liquid phase separation also known as precipitation and were in no close proximity to conditions that yielded protein crystals. Therefore, other plates were set up with an increase in protein concentration.

**A**

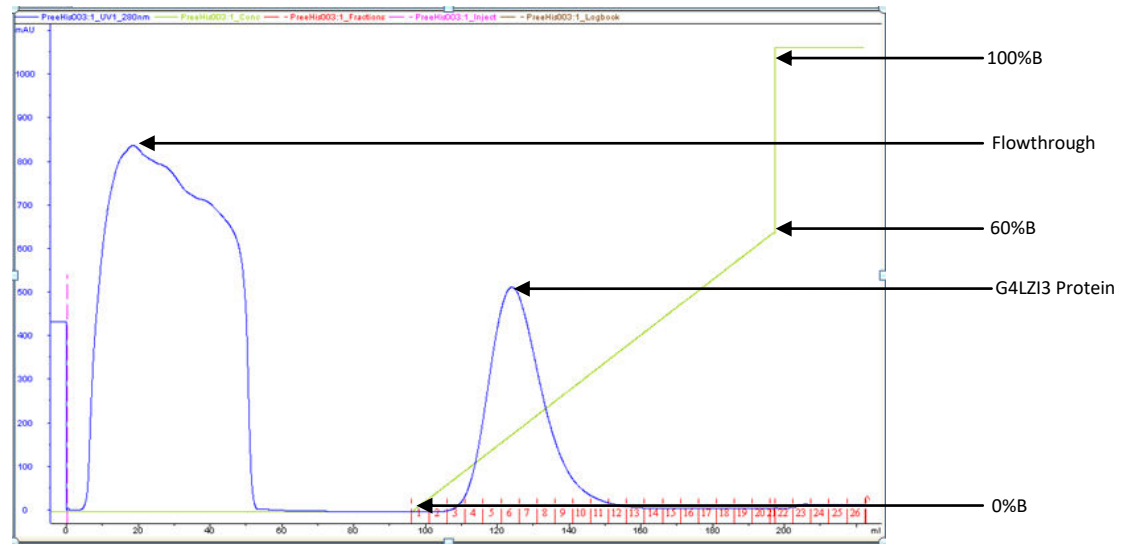


**B**

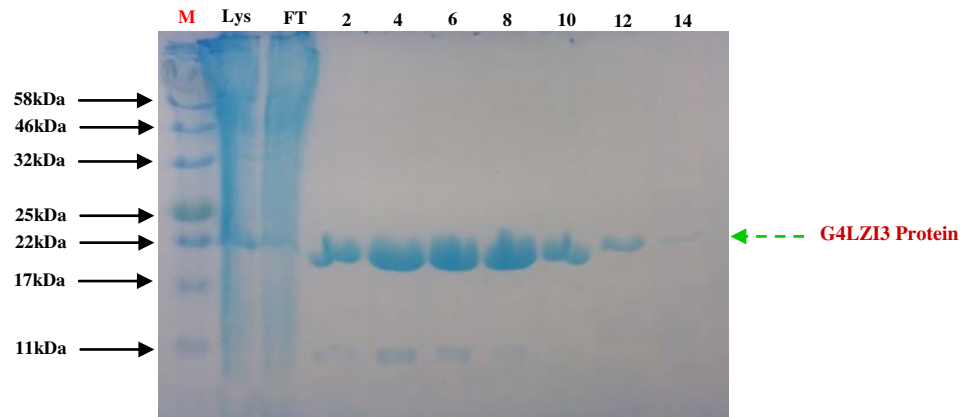


**Fig. 6.2: Large scale expression of the G4LZI3 USP in *E. coli* cells.** In (A) expression was performed in JM109 cells while in (B) expression was performed in M15 cells both in 2 litres respectively. M represents the protein ladder. Run on the gel is the total bacterial cell suspension before cell lysis. The lanes show progressive expression of the protein from 5  $\mu$ L to 20  $\mu$ L. Expression of the protein was more apparent in JM109 cells.

**A**

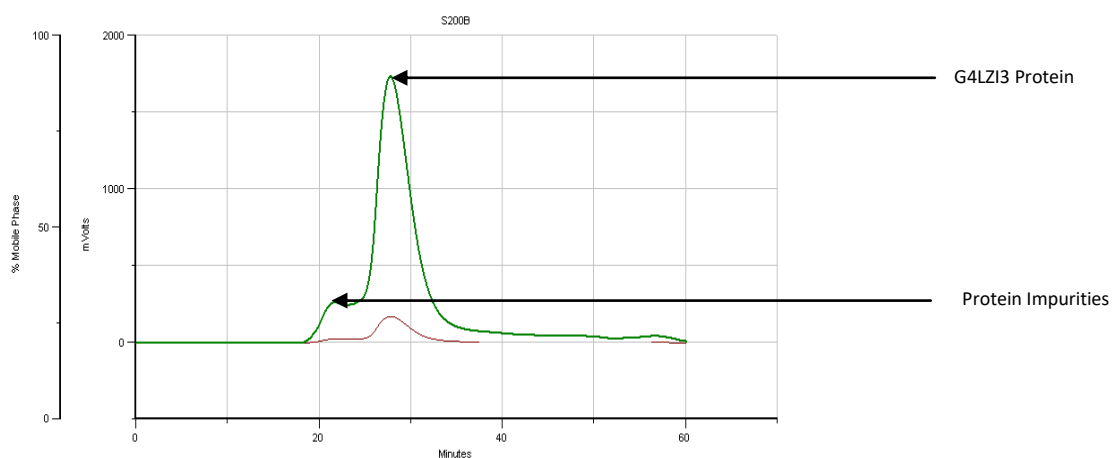


**B**

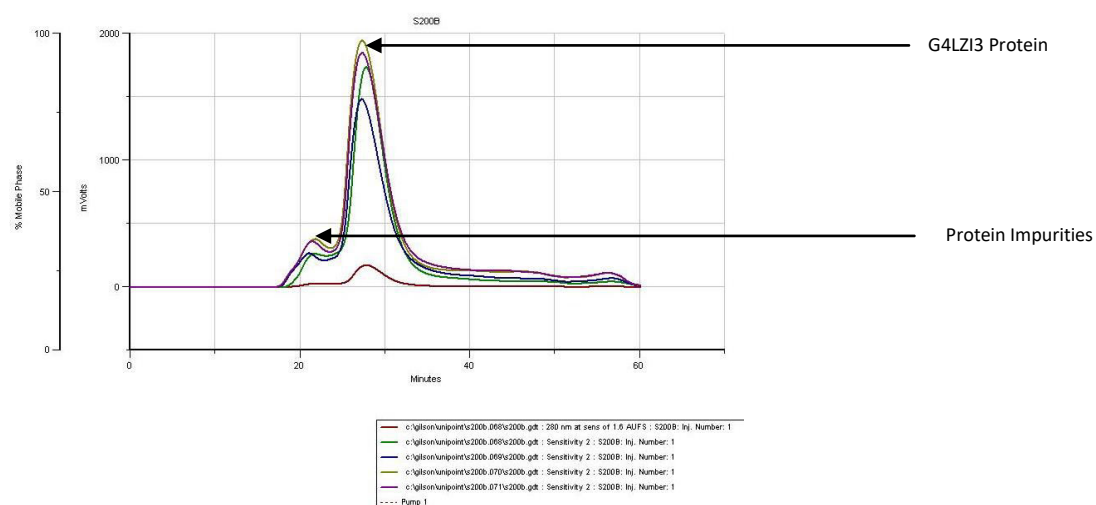


**Fig. 6.3: Affinity chromatography of the G4LZI3 protein.**(A) Chromatography showing Ni-NTA chromatography profile of the G4LZI3 protein. Illustrated are two major peaks; the first representing impurities flown out of the column while the second well-resolved peak represents the protein eluted at 60%B gradient. **(B)** SDS-PAGE gel analysis of the fractions below the protein peak as is shown in the chromatogram. Lane 1 is the molecular weight marker. Lane 2 shows the impure total bacterial cell lysate present in the soluble fraction while Lane 3 is the flowthrough. Lane 4 to Lane 10 represents alternating numbers of the fractions under the protein peak and tight bands corresponding to the expected molecular weight of the protein appear on the gel.

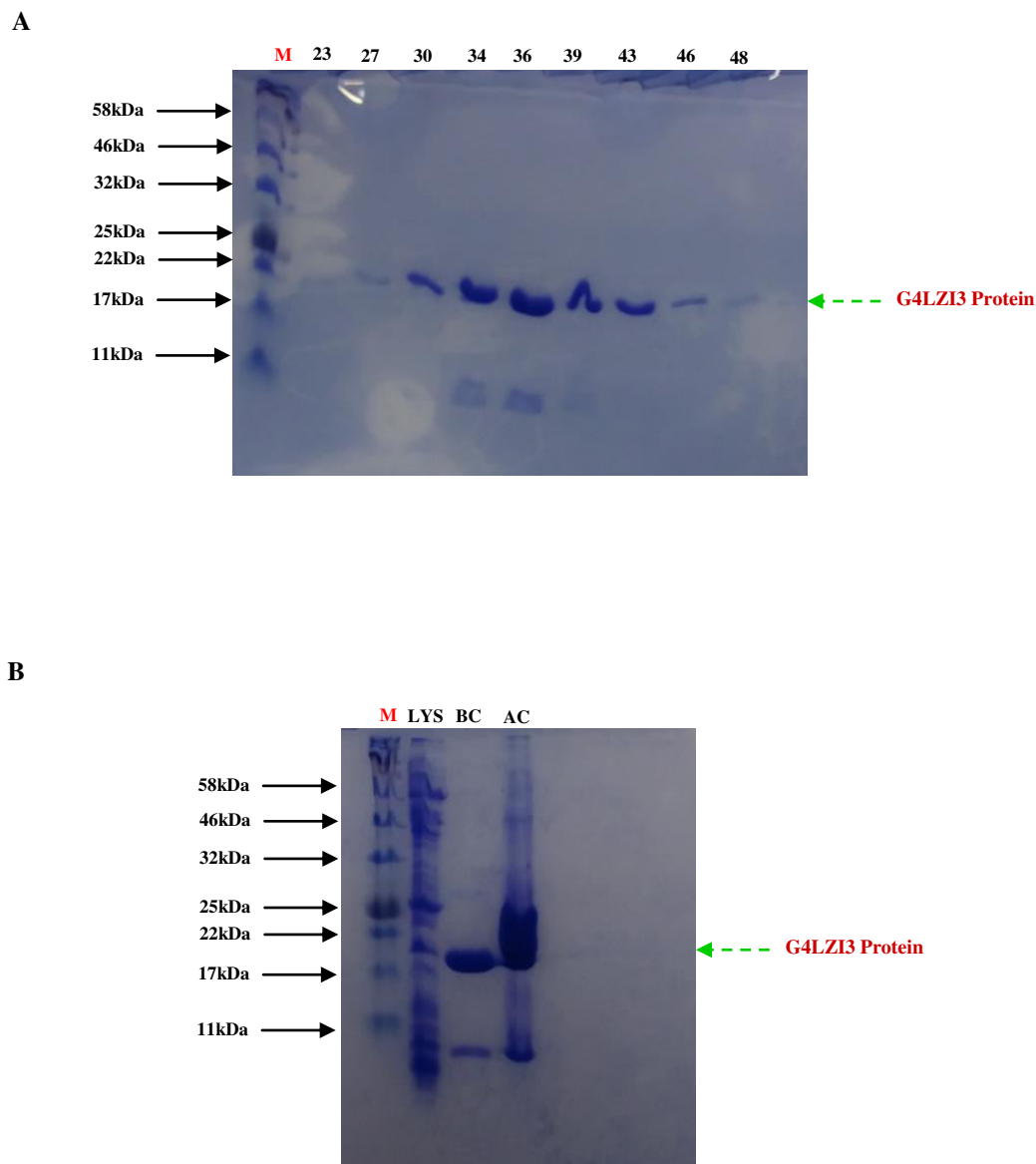
**A**



**B**



**Fig 6.4: Removal of residual proteins by size exclusion chromatography.**(A)Chromatogram showing purification of the G4LZI3 protein by separation technique. The small peak corresponds to impurities still present within the protein while the major peak corresponds to the G4LZI3 protein. (B) An overlay of all the size exclusion chromatograms depicting consistent results.



**Fig 6.5: Gel filtration purification gels of the G4LZI3 protein.** (A) SDS-PAGE gel analysis of the corresponding peak of the size exclusion chromatogram. Lane 1 represents the protein molecular weight ladder while the remaining lanes corresponding to the main peak. (B) SDS-PAGE gel analysis of the concentrated protein. Lane 1 represents the protein molecular weight marker; Lane 2 shows the total bacterial cell lysate before affinity purification; Lanes 3 shows the size exclusion purified sample before concentration; Lane 4 show shows the size exclusion purified sample concentrated down to 2mg/mL.

The next trial was set using 25 mg/mL of protein. Precipitation was observed in almost all conditions and produced no crystals. This was an indication of excess protein concentration. Hence, attention was focused on protein concentrations between 11 mg/mL and 24 mg/mL of protein. Plates set with 12 mg/mL yielded the first batch of protein crystals (Table 6.1 iii) at 16°C with conditions D11. The hit contained tiny crystals which were too minute for fishing and diffracting. Hence, a seed stock was made so it may be used to increase crystal quality and size. It was done by micro-seeding. The drops containing crystals were picked using a pipette. Crystals were crushed and re-suspended in 20 µL mother liquor in an eppendorf tube. This produced a 6 mg/mL stock solution, which was stored at 4°C whenever required. Through this technique, the point of nucleation has been reached resulting in the emergence of an ordered lattice and causing the added solute or protein to attract and form crystals around the lattice, eventually detaching from the seeds once they have obtained a considerable size (Moyo, 2017). It is for this reason that the sitting drop technique presents itself as excellent technique for optimization and screening experiments due to the fact that it favors the nucleation process.

A 2 µL measure of seed stock and 5 µL of 15 mg/mL, 17 mg/mL, 20 mg/mL and 24 mg/mL protein were used to set up a whole array of crystal plates. From all the conditions, the 15 mg/mL and 17 mg/mL condition yielded the best crystals under the condition D11, which consists of 1.12 M Alcohols (0.2 M 1,6-Hexanediol; 0.2 M 1-Butanol; 0.2 M 1,2-Propanediol; 0.2 M 2-Propanol; 0.2 M 1,4-Butanediol; 0.2 M 1,3-Propanediol), 0.1 M Buffer System 3 pH8.5 (1.0 M Tris (base); BICINE), 50% (v/v) Precipitant Mix 3 (40% (v/v) Glycerol; 20% (w/v) PEG4000) and condition H1 composed of 0.1 M Amino Acids (0.2 M DL-Glutamic acid monohydrate; 0.2 M DL-Alanine; 0.2 M Glycine ; 0.2 M DL-Lysine monohydrochloride; DL-Serine), 0.1 M Buffer System 1 pH6.5 (1.0 M Imidazole; MFS monohydrate acid), 50% (v/v) Precipitant Mix 1

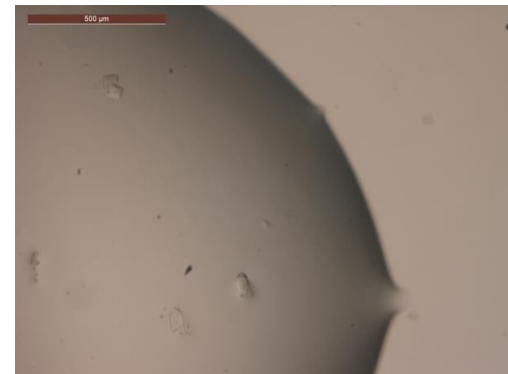
(40% (v/v) PEG500 MME, 20% (w/v) PEG20000) (Table 6.1 iv). Consistently, showers of crystals that were observed in individual drops, were of 3-dimensional shape and of considerable size (Table 6.1 v). Attempts to increase the sizes further still however, yielded the same results.

**Table 6.1: Progression of Crystal Growth during Crystallization**

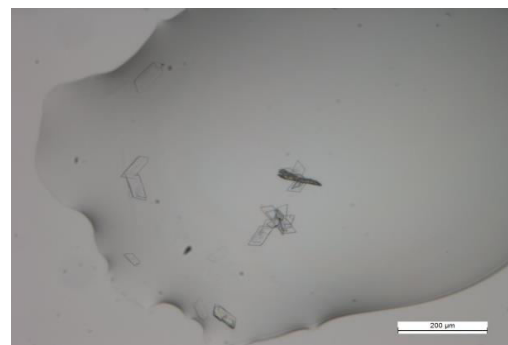
**Crystal Status**

**Images**

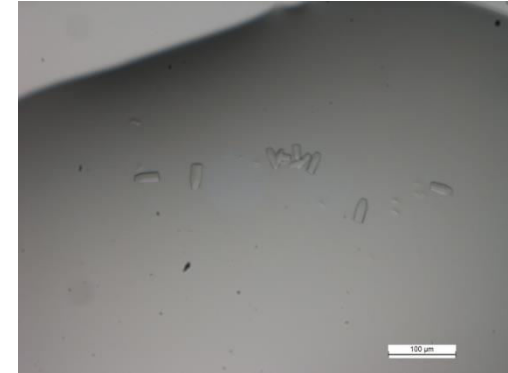
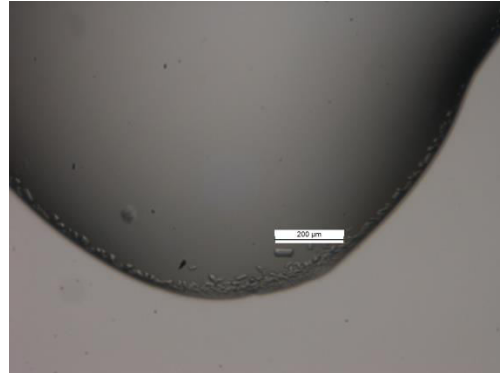
i. Presence of clear drops indicating low protein concentrations



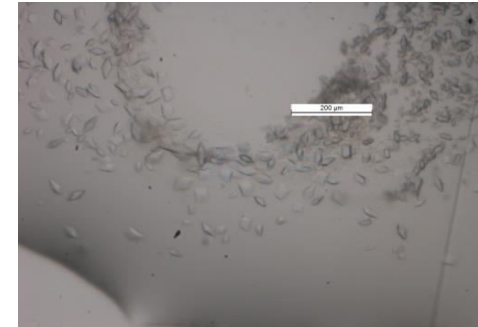
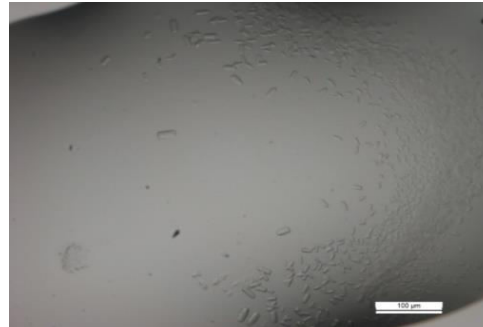
ii. Appearance of mosquito-like crystals within the first few wells



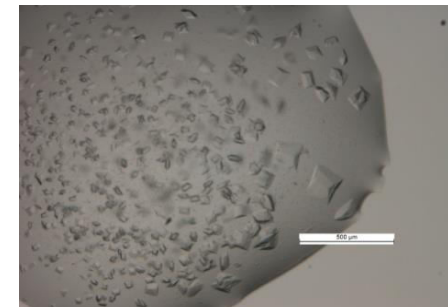
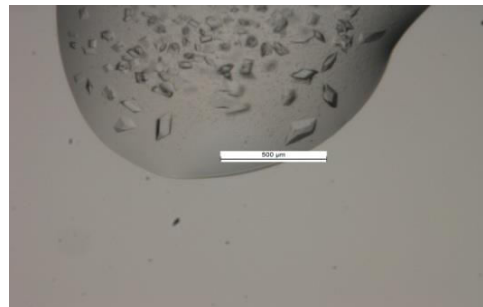
iii. Crystals used for micro-seeding



iv. Progression of seed growth



v. Final crystals used to acquire diffraction datasets

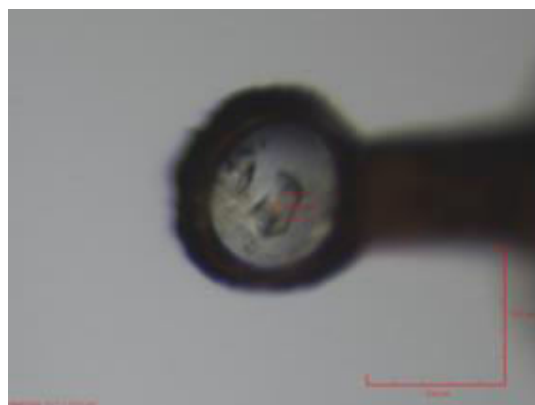
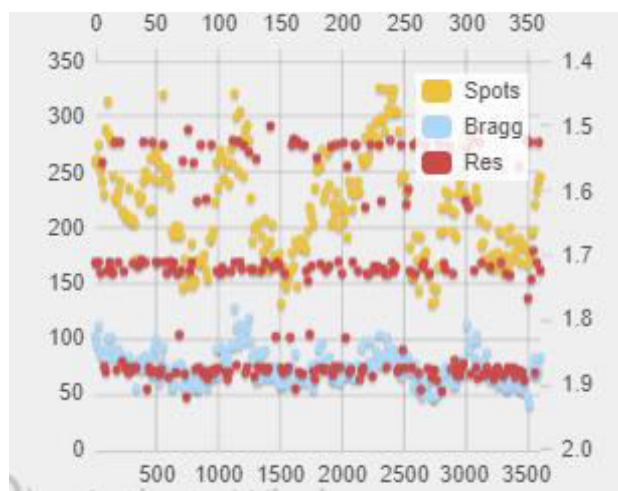
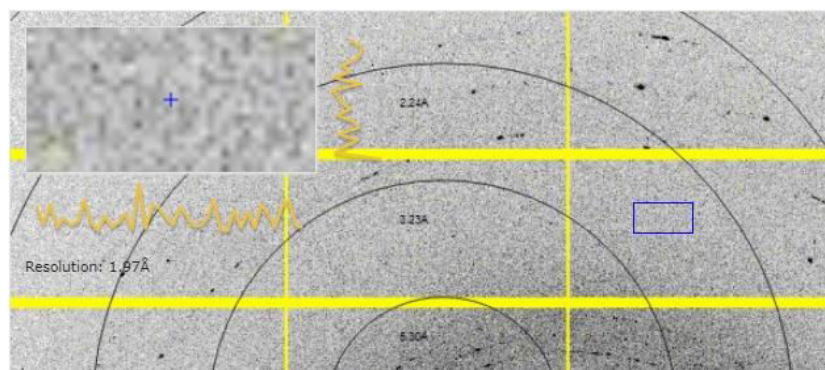


Images show the progression of crystal growth until the final sets of crystals were obtained. The crystals shown in Image (v) were fished out and sent for analysis to the Diamond Light Source synchrotron in the UK.

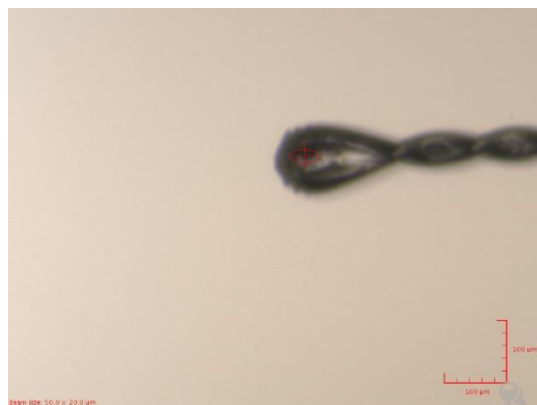
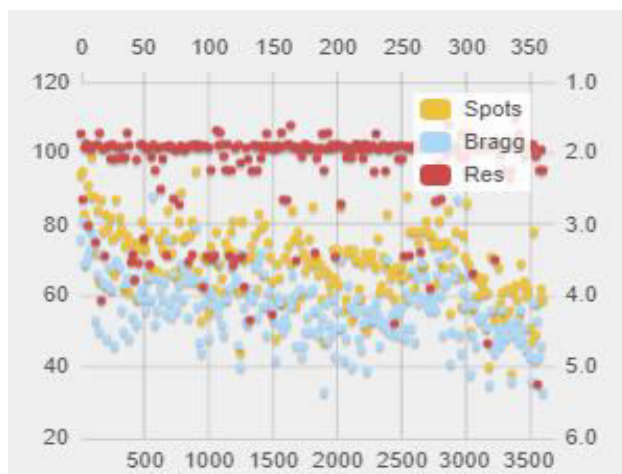
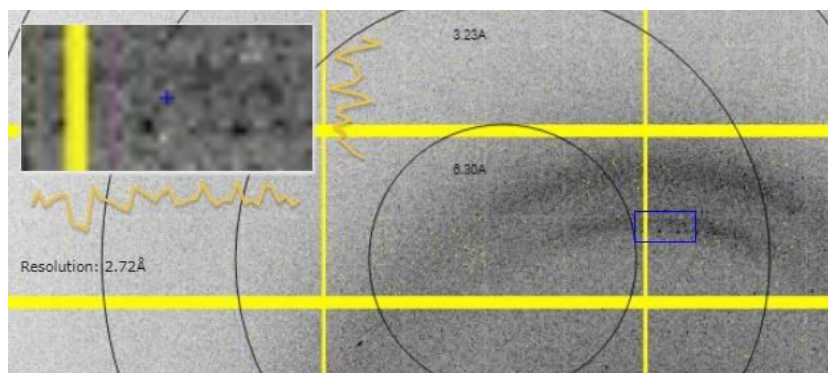
Crystals were then selected, fished out using cryo-loops, flash frozen with liquid nitrogen and packed into dewars. Unfortunately, fishing the crystals were a challenge because almost all were adhering to the bottom of the plate, were very fragile and difficult to remove without damaging them, a feature typically associated with the sitting drop technique (Thorne *et al.*, 2010). Hence, mother liquor was constantly added to the drop to allow the crystals to float for easier access. The remaining crystals within the drop or within the same conditions were also fished out and diffracted in-house to ensure that all crystals being sent for data collection were protein, not salt. Following this verification, the dewars were packed and shipped to the Diamond Light Source (UK) for the data collection via remote access.

### **6.3.2. Crystal Diffraction and Data Collection**

Crystals were mounted remotely using the BART sample changer onto the I03 MX beamline (Appendix IV) using the Generic Data Acquisition (GDA) software. Once samples were aligned (Fig.), x-ray centring was done guided by a two-dimensional grid scan to ensure the best positions for diffraction. Data collection automatically followed was accessed via the ISPyB (<https://ispyb.diamond.ac.uk>) database that now operates via a modern interface known as *SynchWeb* (<https://diamondlightsource.github.io/SynchWeb/>) (Fisher *et al.*, 2015). For each dataset, crystal snapshots (Fig. 6.6A), a diffraction pattern (Fig 6.6C), a DISTL plot (Fig 6.6B) and experimental parameters (Table 6.1) were displayed. Two datasets were eventually chosen and merged for structure determination.

**A****B****C**

**Fig. 6.6: Crystal Diffraction of Dataset 1** (A) Crystal mounted in cryoloop and centered for the acquisition of diffraction data. (B) DISTL Plot displaying the identification of Bragg spots for autoindexing. (C) Diffraction pattern of the crystal. Data obtained possessed a resolution of 3.36Å.

**A****B****C**

**Fig. 6.7: Crystal Diffraction of Dataset 2** (A) Crystal mounted in cryoloop and centered for the acquisition of diffraction data. (B) DISTL Plot displaying the identification of Bragg spots for autoindexing. (C) Diffraction pattern of the crystal. Data obtained possessed a resolution of 3.05Å.

**Table 6.2: Obtained Crystallographic Data Statistics**

|                        |   |   |       | Dataset 1  |        |       | Dataset 2  |        |       |
|------------------------|---|---|-------|------------|--------|-------|------------|--------|-------|
| Wavelength (Å)         |   |   |       | 0.9736     |        |       | 0.9736     |        |       |
| Exposure               |   |   |       | 0.010 s    |        |       | 0.020 s    |        |       |
| Beamsize               |   |   |       | 50 x 20 μm |        |       | 80 x 20 μm |        |       |
| Resolution (Å)         |   |   |       | 3.36       |        |       | 3.05       |        |       |
| Spacegroup             |   |   |       | P21 21 21  |        |       | P21 21 21  |        |       |
| Unit Cell Statistics   |   |   |       |            |        |       |            |        |       |
| A                      | B | C | 67.98 | 72.33      | 175.94 | 70.38 | 74.76      | 174.01 |       |
| A                      | β | γ | 90.00 | 90.00      | 90.00  | 90.00 | 90.00      | 90.00  | 90.00 |
| Mn < 1 /sig (i) >      |   |   |       | 9.0        |        |       | 9.2        |        |       |
| Rmeas Inner            |   |   |       | 0.045      |        |       | 0.055      |        |       |
| Rmeas Outer            |   |   |       | 1.388      |        |       | 6.174      |        |       |
| Rmeas Overall          |   |   |       | 0.144      |        |       | 0.191      |        |       |
| CC Half                |   |   |       | 1.0        |        |       | 1.0        |        |       |
| Multiplicity           |   |   |       | 13.2       |        |       | 13.1       |        |       |
| Anomolous Completeness |   |   |       | 99.9       |        |       | 99.5       |        |       |
| CC Anom                |   |   |       | -0.2       |        |       | 0.0        |        |       |
| Completeness           |   |   |       | 100.0      |        |       | 99.8       |        |       |

Several datasets were acquired but displayed poor diffraction. The best diffraction dataset is presented in Fig 6.6 whose resolution was at  $\sim 3.4\text{\AA}$ . Typically, a crystallographic resolution from  $1\text{\AA}$  to less than  $3\text{\AA}$  is ideal for trustable and excellent structures. Despite various collections from various crystals, the data statistics remained consistent until a dataset of  $3.05\text{\AA}$  was obtained. This data was used to improve the data that was already processed for structure determination.

The quantity and quality of data largely contributes to the quality of determined crystal structures. Diffraction data is most often assessed by the R value, which includes factors such as the  $R_{\text{work}}$ ,  $R_{\text{free}}$ ,  $R_{\text{sym}}$ ,  $R_{\text{merge}}$  and the  $R_{\text{meas}}$  factors (Karplus and Diederichs, 2012). These set of values are especially useful in equating observed and calculated data. The  $R_{\text{meas}}$  factor is a numerically realistic indicator of redundancy independence defined by:

$$R_{\text{meas}} = \sum_h (n_h/n_h - 1)^{1/2} \sum_i | \langle I_h \rangle - I_{h,i} | / \sum_h \sum_i I_{h,i}$$

The  $R_{\text{meas}}$  value observed in the diffraction sets ranged from 14–19%; values of less than 5% are indicators of high quality data. Initially,  $R_{\text{merge}}$  (alternatively known as  $R_{\text{sym}}$ ) was the main indicator for data quality but was eventually discarded as an indicator due to the fact that it depends on multiplicity while  $R_{\text{meas}}$  does not and is considered the corrected multiplicity factor (Karplus and Diederichs, 2012; Diederichs and Karplus, 1997). Although this seemed daunting,  $R_{\text{merge}}$  or  $R_{\text{meas}}$  factors are on the other hand, inappropriate determinants for resolution cut-off, value employed to dispose of data considered too ‘noisy’ and downgrading factors of model quality (Karplus and Diederichs, 2012; Evans and Murshudov, 2013). Therefore,  $R_{\text{merge}}/R_{\text{meas}}$  factors may not necessarily indicate weak data. In general, values for experimental systems that are highly optimized and comprise good-quality crystals tend to possess values of less than 4–5%, while sub-optimal data may possess values of up to 20% (Wlodawer *et al.*, 2013). Such

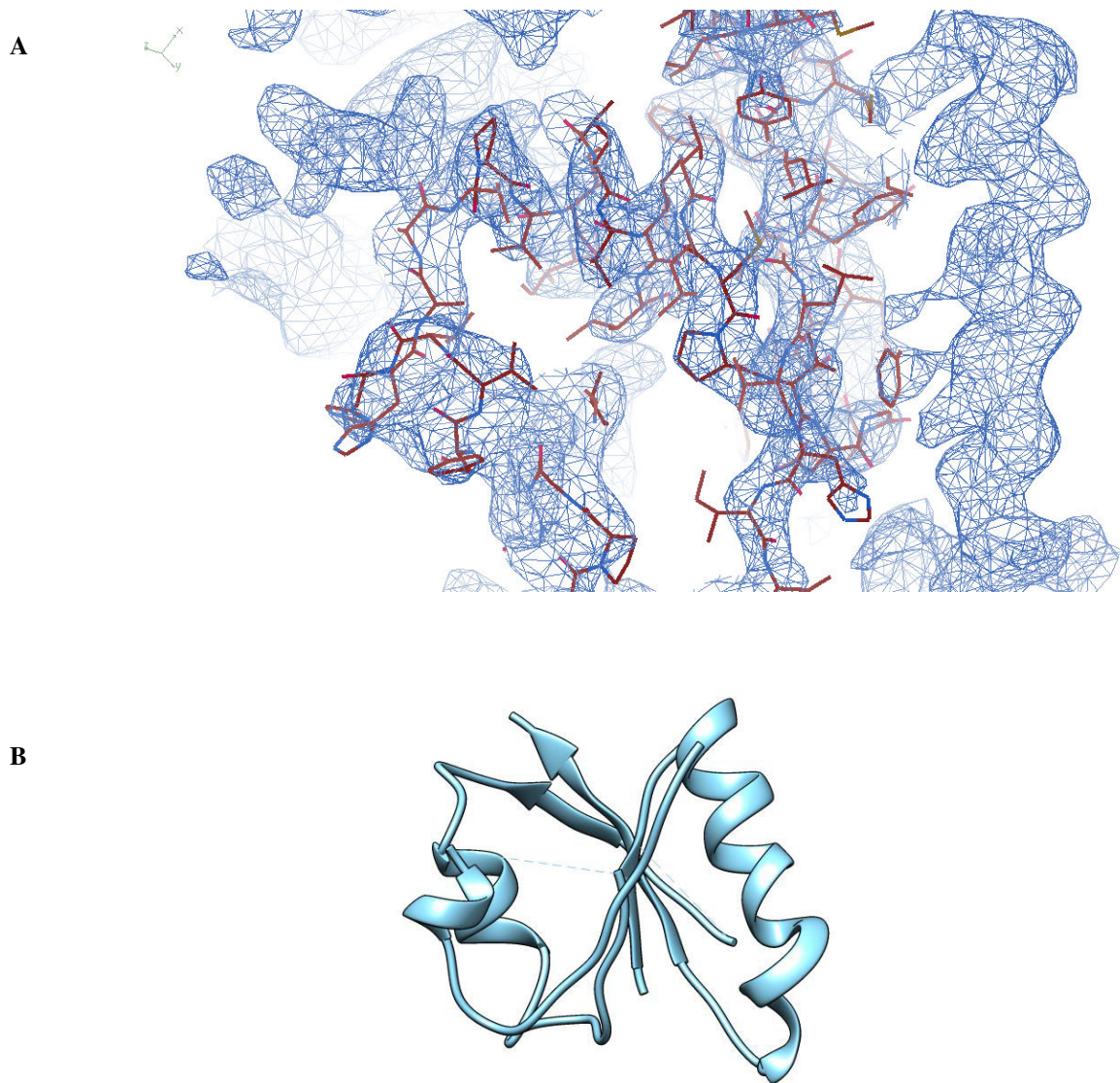
values however, signify difficult or important structures. Hence, the chosen datasets were still considered for structure determination with  $R_{\text{meas}}$  factors of 14% and 19%.

The symmetry of a structure is defined by its spacegroup, which when applied in combination with the asymmetric unit, the complete protein structure is built (Moyo, 2017). The spacegroup of  $P2_1 2_1 2_1$  (which is the most common) indicated that the structure is of orthorhombic crystal system with an estimated presence of 4 molecules in the unit cell. This was taken into consideration when solving the structure by Molecular Replacement.

### ***6.3.3 Structure Determination and Refinement***

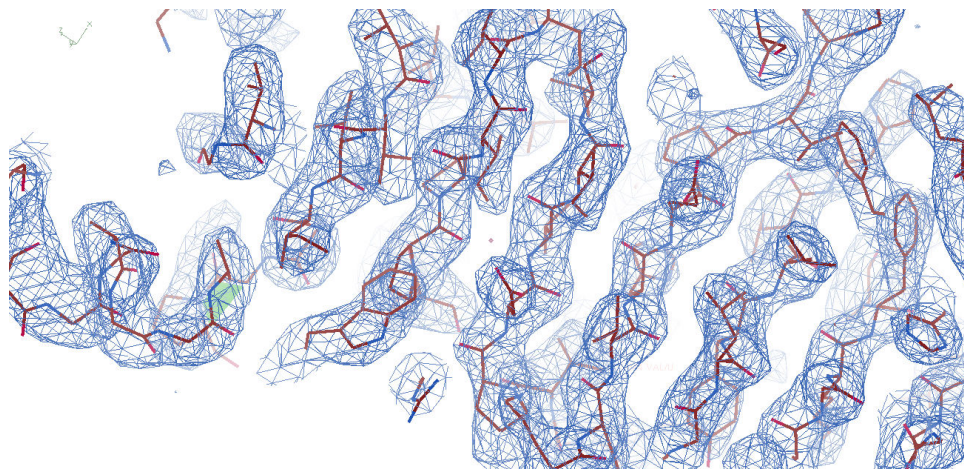
As its name implies, x-ray crystallography makes use of x-rays for the visualization of molecular structures down to their atomic level by x-ray diffraction. The intensities of the diffracted rays, which can be described as  $\text{intensities} = \text{amplitude}^2$ , can be measured effortlessly but during the experiment, the relative phases of the waves are unfortunately lost, which is a critical requirement for the elucidation of crystal structures (Prendergast, 2010). Molecular Replacement (MR) presents itself as one of the most popular methods for solving the phase problem using a homologue of the crystallized protein, which is oriented in agreement with the obtained crystallographic data for the calculation of phases and generation of an electron density map (DiMaio, 2017). Crystallographic data with high resolutions and extreme likeness to their homologues are easier to deal with as the generated map is easily interpreted. Low-resolution data are somewhat more difficult to tackle but are not impossible. Several software, mostly automated, are used for the MR process and for this project, MR\_Rosetta was used as it has been described in literature as a powerful tool in solving difficult molecular replacement problems due to the low resolution data acquired (DiMaio, 2017). Fig. 6.8 below shows the generated electron density map and partially solved structure visualized in Coot and Chimera respectively.

The homologues used for MR were of 31.06% (PDB ID: 2DUM) and 28.93% (PDB ID: 1MJH) identity to the G4LZI3 sequence. Using the generated electron density map, model building ensued and this was first done using the *PHENIX AutoBuild* wizard. The *AutoBuild* wizard forms part of the *PHENIX* package and is a highly automated system that uses experimental phase data, MR starting models when available and any other provided data files and specified parameters for the rebuilding and completion of model structures (Terwilliger *et al.*, 2008). The software produces R factors that are independent of resolution but however, reliant on the electron density map provided. The results were visualized in Coot and Chimera as shown in Fig. 6.9A and B. The resulting  $R_{\text{work}}$  and  $R_{\text{free}}$  factors were 0.28 and 0.36 respectively with 4 molecules present in the asymmetric unit. Despite the technique, the chains were labeled wrongly, the structure contained a number of significant gaps and many residues in accordance with the protein sequence were replaced with other suggestive amino acids that did not necessarily correlate to the electron density map.

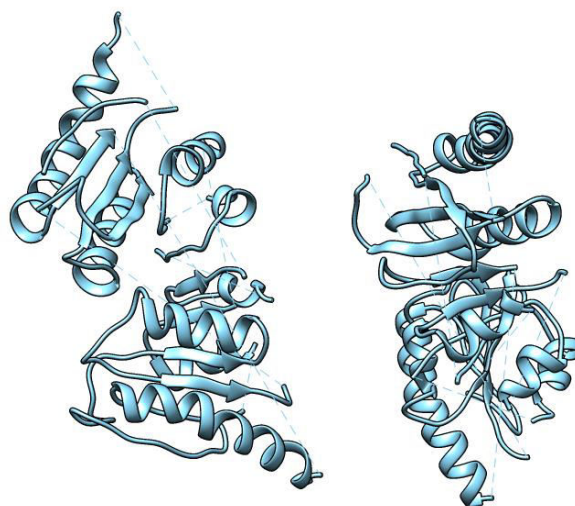


**Fig. 6.8: Starting Solution obtained by Molecular Replacement.** (A) Density map of the crystal data as solved by MR\_ROSETTA. (B) Structure of the starting model viewed in Chimera displaying an obviously unfinished model.

**A**



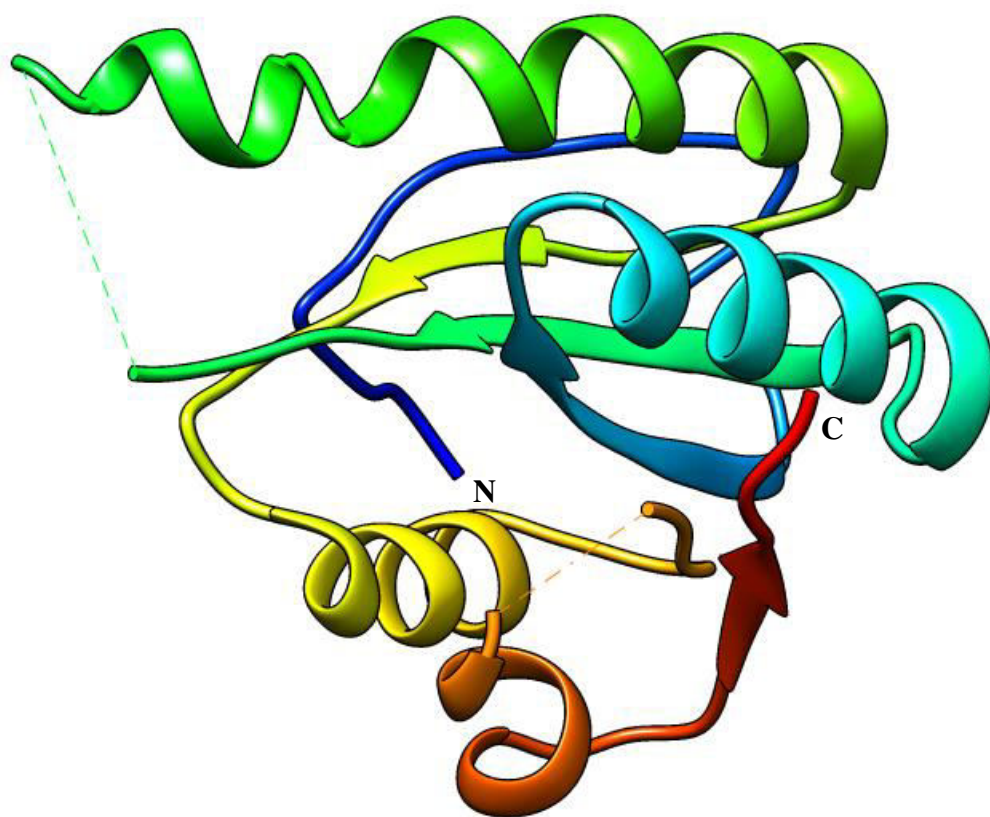
**B**



**Fig. 6.9: Model obtained from AutoBuild.refine.** (A) Clearly improved density map of the crystal data having undergone autobuild. (B) Sturcture of the model viewed in Chimera showing four molecules present in the unit cell.

Manual building and correction was then continued in Coot with by the addition, deletion or proper positioning of amino acid residues within the map using the protein sequence as a guide as well as refinement of the corrections in-house by Step-Refinement. With every correction, the model was put through series of more accurate and vigorous refinements through PHENIX.refine. 'Refinement' in simple terms refers to the modifying of model parameters or coordinates to equate calculated structure factors as precisely as possible. Fig 6.10 shows results of the current structure after 22 cycles of refinement, while Table 6.2 displays statistics comparisons between initial and final R factors.

The solved structure shows that the G4LZI3 protein consists of four alpha-helices and four beta-sheets. The  $\beta$ -sheets are of a parallel arrangement as a  $\beta\alpha\beta\alpha\beta\alpha$  motif composed of  $\beta$ -sheet and  $\alpha$ -helices alternating with each other. One out of the four  $\alpha$ -helices is a very short one known as a  $3_{10}$ -helice at the C-terminal end of the protein consisting of six amino acid residues. Subsequent to  $\alpha$ -helices,  $\beta$ -sheets and reverse turn secondary structures, the  $3_{10}$  helix is the fourth most common and is thought to be an extension of and an  $\alpha$ -helical folding/unfolding intermediate (Armen *et al.*, 2003).



**Fig. 6.10: Final Model of the G4LZI3 Structure.** Model was obtained after 22 cycles of refinement. Model reveals that the structure is composed of 4 alpha-helices and beta-sheets respectively with a long N-terminal loop and short C-terminal loop.



As much as protein structure prediction and determination is one of the most important objectives XRD, NMR and all other determination tools seek to accomplish, protein function determination is just as critical. This may however, be quite difficult to achieve especially in cases where structure has been solved with little or no initial functional characterization (Redfern *et al.*, 2009). The Dali server (<http://ekhidna2.biocenter.helsinki.fi/dali/>) was contacted for the purpose of predicting possible functions of the G4LZI3 protein based on sequence-structure comparisons in 3D (Holm and Laakso, 2016). A total 3232 models were compared to the G4LZI3 protein and scored according to the Z-score. The table in Appendix V lists the best 15 models with the highest Z-scores while Fig. 6.11 shows superimposition of the solved structure with the best model. This model for comparison because it possessed the highest Z-score; the higher the score the more similar the structures are. Z-scores of less than 2 are considered spurious and since all 15 structures exhibited very high z-scores, Multiple Sequence Alignment (MSA) was done to determine structural conservation. MSA was conducted using CLUSTALW (<https://www.genome.jp/tools-bin/clustalw>) while shading was done using the Boxshade program ([https://embnet.vital-it.ch/software/BOX\\_form.html](https://embnet.vital-it.ch/software/BOX_form.html)).

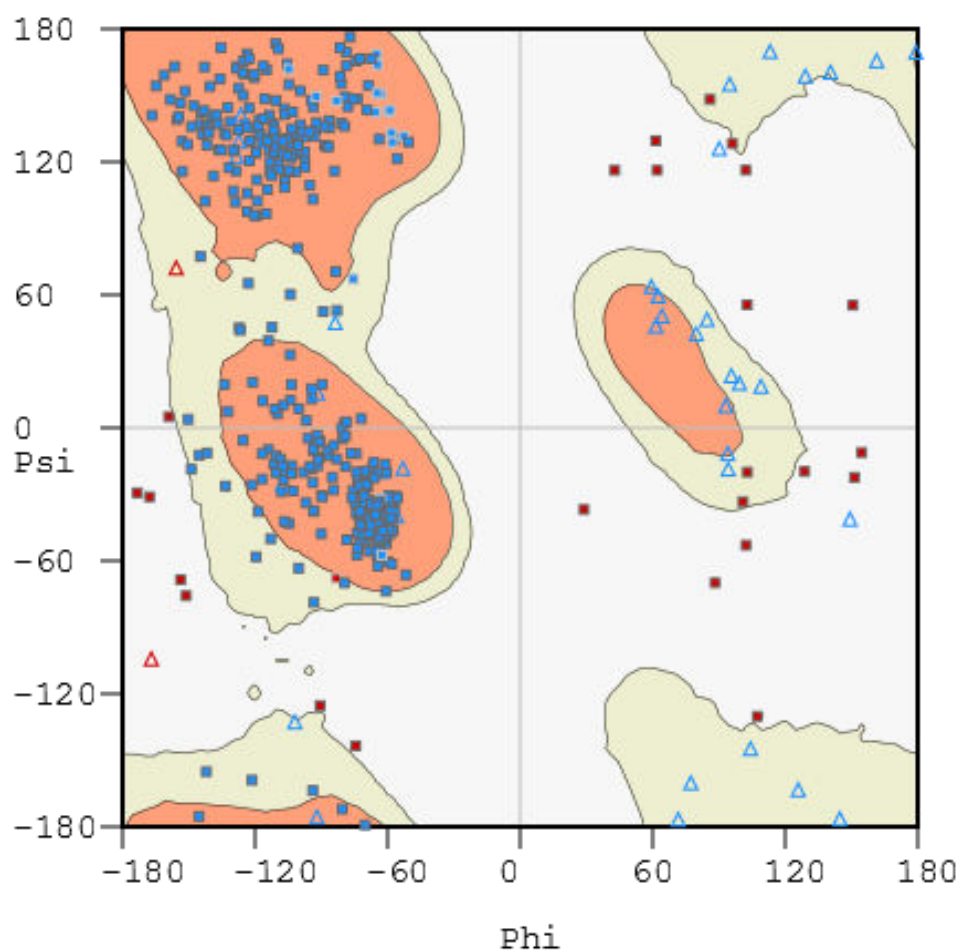
At3g01520 (PDB ID: 2GM3) is a USP-like protein from *Arabidopsis thaliana* that possesses six AMP-bound promoters and is classified as a 1MJH-like plant protein due its strong sequence similarity to the MJ0577 Usp from *Methanococcus janaschi*(Kim *et al.*, 2015). USPs are classified according to their ATP binding abilities defined by the G-2X-G-9X-(S/T) motif where x represents any amino acid (Sousa and McKay, 2001; Tkaczuk *et al.*, 2013). The electron density modeled naturally acquired AMP onto the model, supporting evidence that the protein may also belong to the nucleotide-binding USP subfamily. The G4LZI3 protein has been predicted to bind biologically relevant ligands which include adenosine triphosphate (ATP),

adenosine monophosphate (AMP), guanosine triphosphate (GTP) and three metallic ion ligands (calcium, zinc and magnesium) (Mbah *et al.*, 2013). Currently, the ATP-binding region of the G4LZI3 protein has not been modeled and thus the binding of ligands calls for more investigation. Looking at the similarities between the G4LZI3 protein and At3g01520, it may be a possibility that the G4LZI3 protein may also form a part of the nucleotide-binding USP subfamily. Loops constituting residues 56–65 and 137–145 in the At3g01520 are missing in all chains of the asymmetric unit, a trait that was also observed in the G4LZI3 model from residues 144–161 and 67–89; it is suggested that these regions may be of very high flexibility. Key residues were however, observed on one flexible loop involved in the binding of the AMP molecule onto the protein, which also corresponded to the ATP binding site of other USP homologues. It was proposed that the binding of a proper ligand to the protein may assist in the re-positioning and ordering of flexible loops.

A study by Mbah and co-workers analyzed, integrated and visualized 13 USP sequences from the *S. mansoni* and *S. japonicum spp.* Sequence alignment, including the G4LZI3 amino acid sequence, revealed the proteins to possess conserved binding sites for amino acids for five key residues: Asp, Leu, Gly, His and Pro at positions 57, 101, 127, 166 and 176 respectively (Mbah *et al.*, 2013). These residues were still present and were not replaced even after model building and refinement. It has been suggested that these amino acids may be important regulators for protein functioning. Functionally characterizing these residues may therefore be an interesting future step in discovering how Usp functions for accomplishing schistosomiasis interventions.

**Table 6.3: Obtained Refinement Statistics**

|                         |                 |
|-------------------------|-----------------|
| $R_{\text{work}}$       | 0.2707 (0.3701) |
| $R_{\text{free}}$       | 0.3293 (0.4421) |
| Molprobity Score        | 2.46            |
| RMS (angles)            | 1.15            |
| RMS (bonds)             | 0.0065          |
| Clashscore              | 18.03           |
| Ramachandran Statistics |                 |
| Favoured (%)            | 89%             |
| Allowed (%)             | 5.60%           |
| Outliers (%)            | 5.40%           |



**In Preferred Regions: 445 (89.00%)**

**In Allowed Regions: 28 ( 5.60%)**

**Outliers: 27 ( 5.40%)**

**Fig 6.12: Ramachandran Plot of the final G4LZI3 model.** The plot was generated from Coot and shows an overall 94.6% of the residues in the preferred and allowed regions while a meagre 5.4% are present in the outliers

Table 6.3 highlights the statistics acquired after refinement of the structure. Final  $R_{\text{work}}$  and  $R_{\text{free}}$  values of 0.27 and 0.33 were acquired respectively, seeing a decrease from initial 0.37 and 0.44 values. The lowered values as well as a decrease in the gap between the  $R_{\text{work}}$  and  $R_{\text{free}}$  range were a positive indication of structure improvement (Karplus and Diederichs, 2012; Headd *et al.*, 2012). A Ramachandran plot was generated to validate model quality by viewing the energy levels of residues based on their phi-psi torsion angles to determine the violation of any stereochemical parameters (Laskowski and Swaminathan, 2013). Results presented in Table 6.2 show 89% of the residues present in the favoured region, 5.6% in the allowed and a meager 5.4% in the outlier region. An improvement in the plot was actually observed after it underwent refinement from values ranging from 83.6%, 8.40% and 8% of the residues present in the favoured, allowed and outlier regions of the plot. Although more refinement may improve the model even further, current results show that the model is a relatively good one.

## References

- Armen R, Alonso DO and Daggett V (2003). The role of  $\alpha$ -,  $\beta$ -, and  $\pi$ -helix in helix $\rightarrow$  coil transitions. *Protein science* **12**(6): 1145-1157.
- Asif A, Mohsin H, Tanvir R and Rehman Y (2017). Revisiting the mechanisms involved in calcium chloride induced bacterial transformation. *Frontiers in Microbiology* **8**: 2169.
- Department of Chemistry, Division of Biochemistry, Protein Biochemistry Group: 1 - 80*
- Diederichs K and Karplus PA (1997). Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nature Structural Biology* **4**(4): 269.
- Dimaio F (2017). Rosetta structure prediction as a tool for solving difficult molecular replacement problems *Protein Crystallography*: Springer pp. 455-466.
- Evans PR and Murshudov GN (2013). How good are my data and what is the resolution? *Acta Crystallographica Section D: Biological Crystallography* **69**(7): 1204-1214.
- Fisher S, Levik K, Williams M, Ashton A and McAuley K (2015). SynchWeb: a modern interface for ISPyB. *Journal of applied crystallography* **48**(3): 927-932.
- Headd JJ, Echols N, Afonine PV, Grosse-Kunstleve RW, Chen VB, Moriarty NW, Richardson DC, Richardson JS and Adams PD (2012). Use of knowledge-based restraints in phenix.refine to improve macromolecular refinement at low resolution. *Acta Crystallographica Section D: Biological Crystallography* **68**(4): 381-390.
- Holm L and Laakso LM (2016). Dali server update. *Nucleic Acids Research* **44**(W1): W351-W355.

Karplus PA and Diederichs K (2012). Linking crystallographic model and data quality. *Science* **336**(6084): 1030-1033.

Kim DJ, Bitto E, Bingman CA, Kim HJ, Han BW and Phillips Jr GN (2015). Crystal structure of the protein A t3g01520, a eukaryotic universal stress protein-like protein from *arabidopsis thaliana* in complex with AMP. *Proteins: Structure, Function, and Bioinformatics* **83**(7): 1368-1373.

Laskowski R and Swaminathan G (2013). Problems of protein three-dimensional structures.

Luft JR, Wolfley JR and Snell EH (2011). What's in a drop? Correlating observations and outcomes to guide macromolecular crystallization experiments. *Crystal growth & design* **11**(3): 651-663.

Mbah AN, Mahmud O, Awofolu OR and Isokpehi RD (2013). Inferences on the biochemical and environmental regulation of universal stress proteins from *Schistosomiasis* parasites. *Advances and applications in bioinformatics and chemistry: AABC* **6**: 15.

Moyo T. (2017). *Role of envelope compactness and glycosylation in HIV-1 resistance to neutralising antibody responses*. University of Cape Town.

Prendergast R. (2010). *Structure determination of small and large molecules using single crystal X-ray crystallography*. The University of Manchester (United Kingdom).

Redfern OC, Dessailly BH, Dallman TJ, Sillitoe I and Orengo CA (2009). FLORA: a novel method to predict protein function from structure in diverse superfamilies. *PLoS Computational Biology* **5**(8): e1000485.

Schutz G (2014). Functional characterization of a novel bacterial peroxidockerin with homology to human heme peroxidases. *University of Natural Resources and Life Sciences, Vienna*

- Sousa MC and McKay DB (2001). Structure of the universal stress protein of *Haemophilus influenzae*. *Structure* **9**(12): 1135-1141.
- Terwilliger TC, Grosse-Kunstleve RW, Afonine PV, Moriarty NW, Zwart PH, Hung L-W, Read RJ and Adams PD (2008). Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallographica Section D: Biological Crystallography* **64**(1): 61-69.
- Thorne RE, Berejnov V and Kalinin Y. (2010). Screening and crystallization plates for manual and high-throughput protein crystal growth: Google Patents.
- Tkaczuk KL, A. 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.
- Wlodawer A, Minor W, Dauter Z and Jaskolski M (2013). Protein crystallography for aspiring crystallographers or how to avoid pitfalls and traps in macromolecular structure determination. *The FEBS journal* **280**(22): 5705-5736.

# Chapter 7: Identification of Possible Inhibitors against G4LZI3 Protein Using Pharmacophore Modeling towards the Design of New Schistosomicides

<sup>1</sup>Priscilla Masamba, <sup>1</sup>Geraldene Munsamy and <sup>1,2\*</sup>Abidemi Paul Kappo

<sup>1</sup>Biotechnology and Structural Biochemistry Group, Department of Biochemistry and Microbiology, University of Zululand, KwaDlangezwa 3886, South Africa; [presh4u@rocketmail.com](mailto:presh4u@rocketmail.com) (PM); [geraldenemunsamy@gmail.com](mailto:geraldenemunsamy@gmail.com) (GM); <sup>2</sup>Molecular Biophysics and Structural Biology (MBSB) Group, Department of Biochemistry, Faculty of Science, University of Johannesburg, Kingsway Campus, Auckland Park 2006, South Africa; [crystallopaul@gmail.com](mailto:crystallopaul@gmail.com) (APK)

## 7.1 Abstract

Pharmacophore modeling is a relatively new yet successful and varied section within the field of computer-aided drug design that has become a useful platform for the rationale design of novel drugs. Among the various applications pharmacophore modeling is employed, virtual screening is often the most common for the identification of molecules that can stimulate a desired biological response. In this manuscript, we report the construction of a pharmacophore model for the Universal Stress G4LZI3 Protein, which has been identified as a promising target in schistosomiasis, a waterborne disease linked to extreme cases of morbidity and mortality. Recent *in silico* studies demonstrated the potential inhibitory effect of a select group polyphenols that showed. Curcumin was used to generate the pharmacophore, which was in turn employed to screen for potential inhibitors. Molecular dynamic simulations were ran on the best three systems and revealed potential inhibitory activity. The best results were of Inhibitor III which portrayed ligand binding and strong interaction. This study forms a basis for the synthesis of compounds that can be tested both *in vivo* and *in vitro* in a bid to design alternative treatment for schistosomiasis.

**Keywords:** Curcumin, G4LZI3, inhibitor, pharmacophore, schistosomiasis.

## 7.2 Introduction

In terms of socioeconomic significance, schistosomiasis ranks second after malaria as a parasitic disease of poverty that showcases severe morbidity and mortality, mostly in tropical and sub-tropical rural areas satiated with poverty. Up to 280 000 deaths are recorded annually by direct cause of the disease with a global prevalence of over 207 million people in 74 countries and a hefty 90% of all schistosome infections occurring in sub-Saharan Africa (Adenowo *et al.*, 2015; Hajissa *et al.*, 2018; Molyneux *et al.*, 2017). The disease displays diverse and severe chronic infection, ranging from haematuria (bloody stool) to anaemia, visual impairment, muscle weakness, abdominal pain down to severe complications such as hepatosplenic periportal fibrosis (Barsoum *et al.*, 2013; Colley *et al.*, 2014; Gryseels *et al.*, 2006; Jauréguiberry *et al.*, 2010). Various pharmacological drugs with anti-schistosomal properties have been in use over the years for its treatment but have been discontinued for various reasons that span from specie specificity, manifestation of toxic side effects and most of all, drug resistance. Only one drug, Praziquantel (PZQ), has been in consistent use for over 30 years, renowned for its activity against all schistosome species (Huang *et al.*, 2012). The drug is highly tolerable with very mild side effects and impressive cure rates of up to 90% (Woldegerima *et al.*, 2019). In light of this however, major and legitimate concerns have been raised on the emergence of drug resistance due to the sole dependence on the drug, with reduced susceptibility already being exhibited among various strains (Melman *et al.*, 2009; Vale *et al.*, 2017). Therefore novel, reputable and effective anti-schistosomal agents with specific and selective activity are critical for the control and eradication of schistosomiasis.

The incorporation of computational approaches to complement the experimental has been of immense value in the discovery and development of novel potential compounds in the research-based pharmaceutical industry. Within this spectroscopic framework includes methods such as nuclear magnetic resonance (NMR) spectroscopy and X-ray crystallography (XRD), which have greatly contributed in providing valuable structural information on key macromolecular targets. Structure-based drug design (SBDD) opens up as an additional and significant avenue in understanding the intricacies surrounding structural and molecular biology (Ferreira *et al.*, 2015). This technique in conjunction with virtual ligand screening (VLS) provides information on the specificities of protein-ligand and protein-protein interactions while eliminating the generation of dubious and unnecessary data in an otherwise expensive and time-consuming process (Lounnas *et al.*, 2013). Target-based pharmacophore modeling is a ligand-based virtual screening (LBVS) strategy that uses compounds to generate structural patterns or spatially distributed chemical features essential for the specific binding of ligands to targets for the exertion of a biological effect (Ferreira *et al.*, 2015; Qing *et al.*, 2014; Yu and MacKerell, 2017).

For this study, the solved structure of the G4LZI3 protein (as described in Chapter 6) was employed for the generation of a pharmacophore for the discovery of small molecule inhibitors. The protein falls under the multi-gene Universal Stress Protein (USP) family, which makes up a highly evolutionarily conserved group of proteins whose members are spread across bacterial, plant, archaea, some invertebrate, yeast, protist and fungal genomes (Vollmer and Bark, 2018; Wang *et al.*, 2017). Genomics analyses using various bioinformatics tools by Isokpehi and colleagues identified the up-regulation of eight genes within the schistosome parasite that encode the USP domain (Isokpehi *et al.*, 2011). Mbah and co-workers eventually identified the G4LZI3 protein, labeling it a potential druggable target in the design of novel therapeutics targeted

against schistosomiasis due to its up-regulation throughout the schistosome multifaceted developmental cycle. Chapter 5 saw molecular dynamic simulations run on docked G4LZI3-polyphenol complexes, selected based on potential inhibitory activity against universal stress proteins. A study by Aanandhi and co-workers demonstrated the hypothetical inhibitory effects of a selective range of polyphenols of the Universal Stress Rv1636 Protein from *Mycobacterium tuberculosis* (Aanandhi *et al.*, 2014). Due to the absence of any schistosome USP structures in the Protein Data Bank database, the structure of the G4LZI3 was predicted computationally and used to conduct post-MD analyses and binding free energy calculations with the polyphenols. Results revealed high binding energy as well as hydrogen and ionic bond interactions between the curcumin and catechin polyphenols with the G4LZI3 protein. Chapter 6 demonstrated the determination of the three dimensional structure of the G4LZI3 protein by X-ray crystallography. Hence, this chapter employed the use of the modeled structure and curcumin to generate a pharmacophore for the identification of novel inhibitors of the protein.

## 7.3 Materials and Methods

### 7.3.1 Molecular Docking

Based on the study conducted in Chapter 5, curcumin and catechin, which exhibited the highest binding energies, were docked onto the G4LZI3 crystal structure. Raccoon, Autodock Graphical user interface supplied by MGL tools and AutoDockVina were used for docking and were set to default docking parameters (Ramírez and Caballero, 2016; Sanner, 2008; Trott and Olson, 2010). Gasteiger charges were added to the compounds prior to docking and the non-polar hydrogen atoms were merged to carbon atoms. The compounds were then docked into the binding pocket of G4LZI3 by defining the grid box with 1 Å spacing and 46 × 60 × 56 size pointing in x, y and z directions. Due to the high number of hydroxyl groups attached to the benzene ring of

polyphenols, Curcumin was deemed fit for further analysis and was selected as the drug template because of the large structure attributed to catechin thus allowing the presence of more interacting residues as opposed to curcumin.

### **7.3.2 Pharmacophore Model Generation and Screening**

The G4LZI3-curcumin complex was subjected to a 20 ns MD simulation to generate the bound conformation and highest contributing residues. Per residue energy decomposition was achieved using the CPPTRAJ module incorporated in the AMBER14 package (Case *et al.*, 2014). The highest contributing pharmacophoric moieties were then used to construct the pharmacophore model which was in turn employed to screen for additional inhibitors. This was achieved using the ZINC PHARMER online platform (<http://zincpharmer.csb.pitt.edu>) (Koes and Camacho, 2012).

### **7.3.3 Molecular Dynamic Simulations**

Two sets of hits were generated from the screen, each containing 31 and 26 hits. Molecular docking of the hits was conducted using UCSF Chimera software and Auto Dock vina (Morris *et al.*, 2009). The best three complexes with the most negative binding energies (kcal.mol<sup>-1</sup>) were then selected for further analysis using molecular dynamic simulations achieved using the GPU version of the PMEMD engine integrated within the AMBER14 suit (Case *et al.*, 2014).

Partial charges for the compounds were first generated by ANTECHAMBER using General Amber Force Field (GAFF) procedures (Wang *et al.*, 2005). The leap module of the AMBER14 package ensured all atoms were within 10Å of any box edge while allowing for the addition of Na<sup>+</sup> or Cl<sup>-</sup> counter ions for system neutralization (Galindo-Murillo *et al.*, 2016). An initial minimization of 20000 steps followed by a 10000steps full minimization was carried out with an

applied restraint potential of 500 kcal/mol and the use of conjugate gradient algorithms in the absence of restraints respectively. A gradual heating MD simulation from 0K to 300 K was run for 50 ps, to ensure that all systems maintained a fixed number of atoms and volume. The solutes within the system were exposed to a potential harmonic restraint of 10 kcal/mol and collision frequency of 1 ps. Following heating, an equilibration step estimating 500 ps for each of the systems was conducted with the operating temperature being kept constant at 300K. Additional features such as the number of atoms and pressure were also kept constant, mimicking an isobaric-isothermal ensemble. The systems pressure was maintained at 1 bar using the Berendsen barostat (Berendsen *et al.*, 1984). The total time for the MD simulation conducted was 100ns. In each simulation, the SHAKE algorithm was employed to constrict the bonds of hydrogen atoms. The step size of each simulation was 2 fs and an SPFP precision model was used. The simulations coincided with isobaric-isothermal ensemble, with randomized seeding, constant pressure of 1 bar, a pressure-coupling constant of 2 ps, a temperature of 300 K and Langevin thermostat with collision frequency of 1 ps.

## 7.4 Results and Discussion

### 7.4.1 Pharmacophore Generation

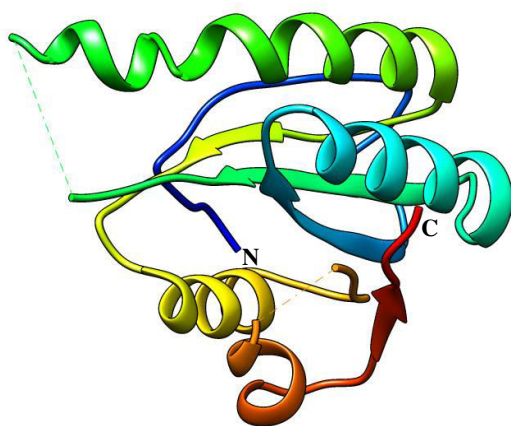
In the field of drug discovery, the pharmacophore technique represents a bridge between medicinal and computational chemistry for the hit discovery and optimization of compounds considered as lead targets for the development of drug candidates (Roy *et al.*, 2015). The pharmacophore can be defined as an ensemble of features and chemical properties required for the screening of potential inhibitors. Pharmacophores are generally constructed in one of two ways; by identifying and selecting common features for their bioactivity through the superimposition of active molecules or by identifying the key interacting points from the ligand-

binding site or active site using the known 3D structure of a target macromolecule. Based on previous studies (Chapter 5), a wide polyphenols were identified as potential inhibitors against the G4LZI3 USP, with curcumin and catechin exhibiting the greatest promise. Curcumin, however, showed the most positive result and was therefore selected for the generation of the pharmacophore (Fig 7.1), which was done using the ZINC PHARMER online platform (<http://zincpharmer.csb.pitt.edu>) (Koes and Camacho, 2012). The pharmacophore model was built using three pharmacophore elements as highlighted with spheres in Fig 7.1. These included the aromatic group (purple), hydrogen bond donating (orange) and hydrogen bond acceptor sites (green). These features were employed as a query into the ZINC database, which enables the search of 18.3 million compounds in over 176 million conformational states in a short period of time (~1min) (Koes and Camacho, 2012). Every one of these conformations is aligned with the features of the pharmacophore and hit molecules are identified by their fit into the spheres (Qing *et al.*, 2014). Two sets of hits were generated from the screen, each containing 31 and 26 hits and the best three were selected and docked onto the G4LZI3 protein (Fig. 7.2-7.5). The docked complexes with Inhibitors I, II and III had Z-scores of 5.6, 5.9 and 5.7 respectively. All the residues involved in the binding were conserved across all complexes.

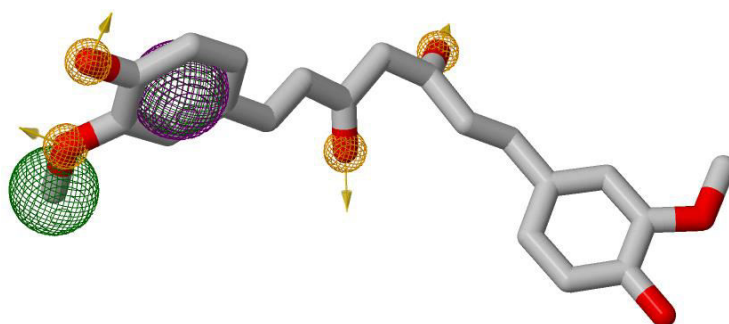
### G4LZI3 Chain A Sequence

STSDGLDIGETKGTISMTRKVLMPVDGSEHSERAFNWMNDVMKITDGLYLVIHIV  
EAKHLNSLVESGRALRAKFFTRCEDSGLSARFTIHVGTKPGENIVRIAHEHGVDLVIISDY  
VLHHANVPVVIIPPP

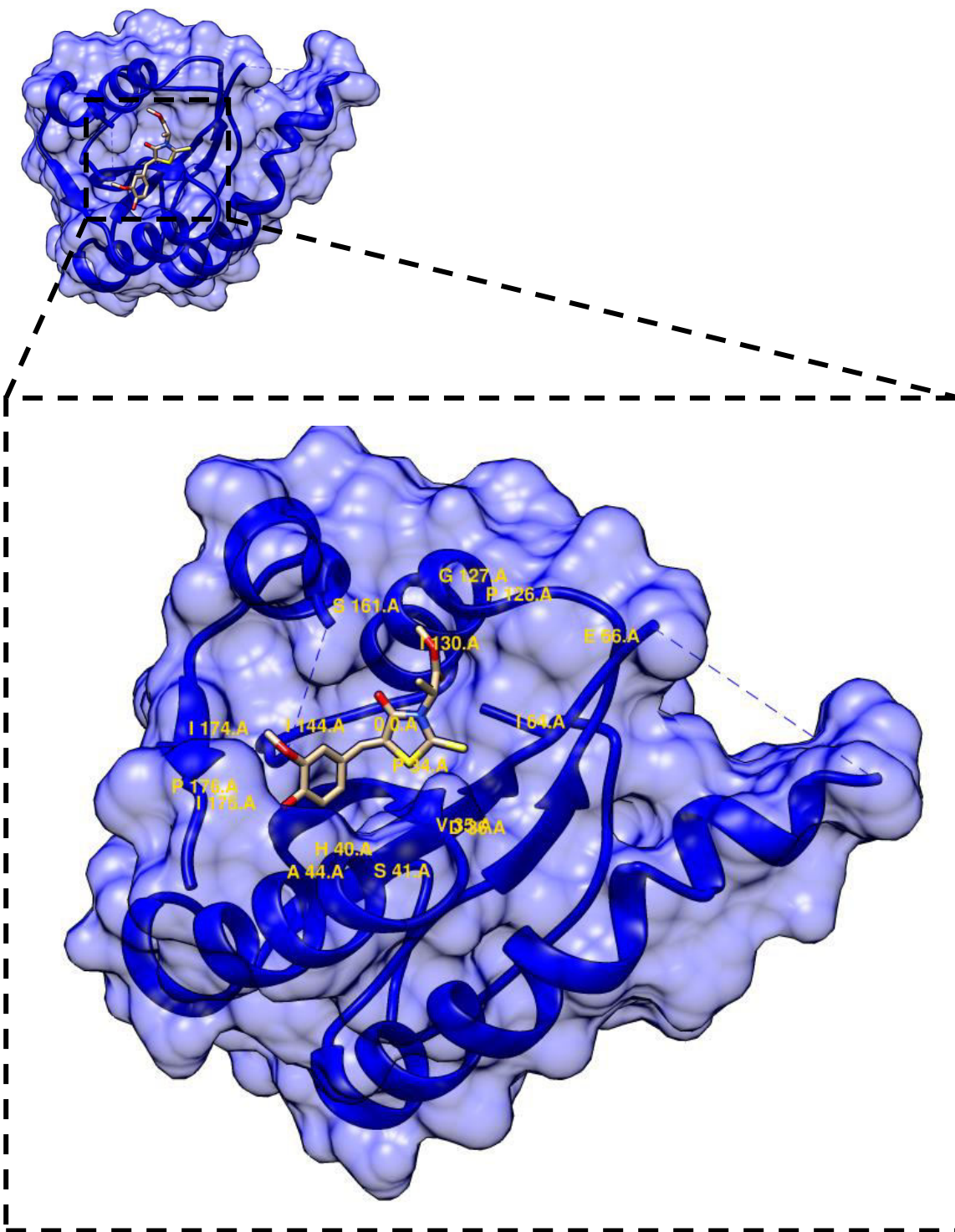
Structure Determination by X-ray  
Crystallography



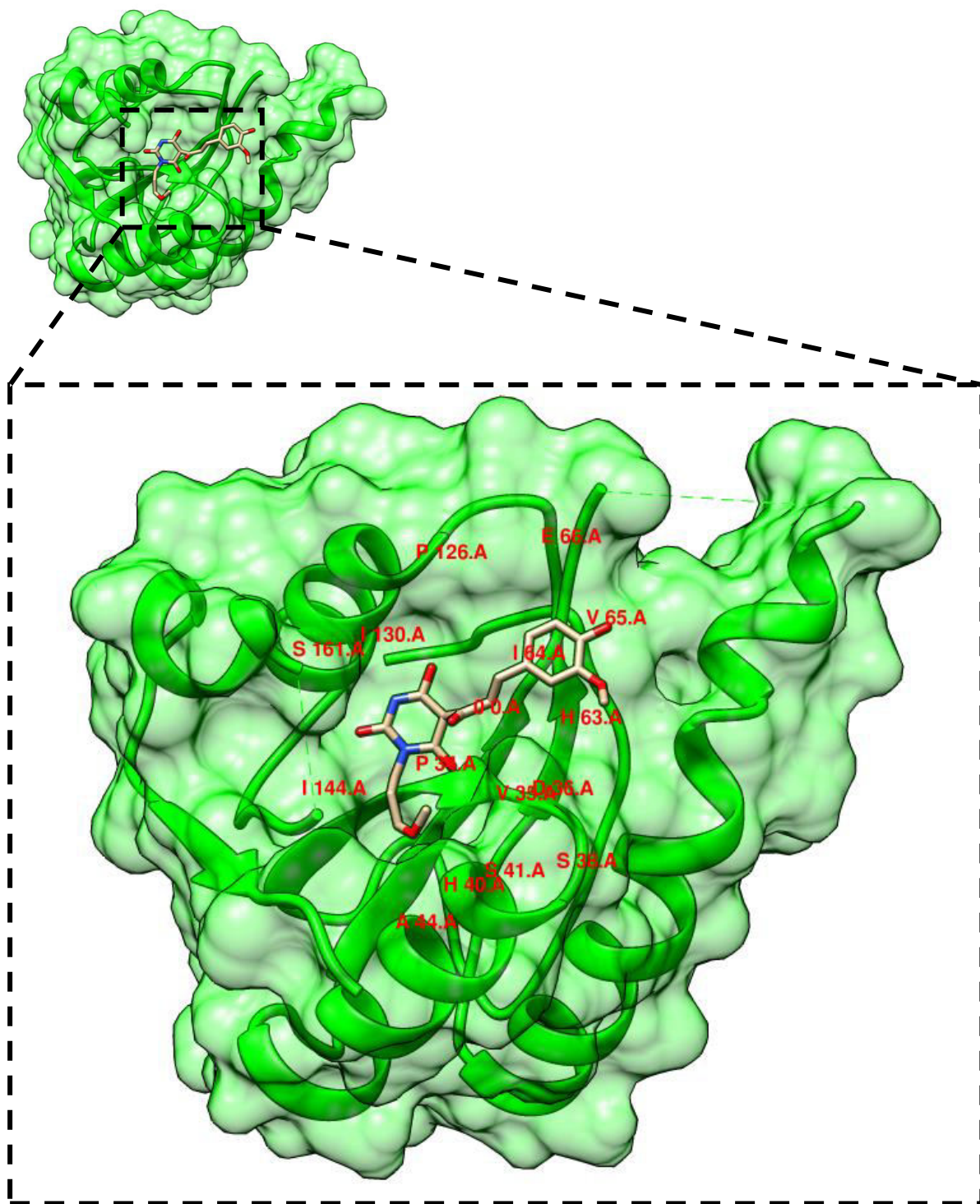
Curcumin Identified as a Potential  
Inhibitor



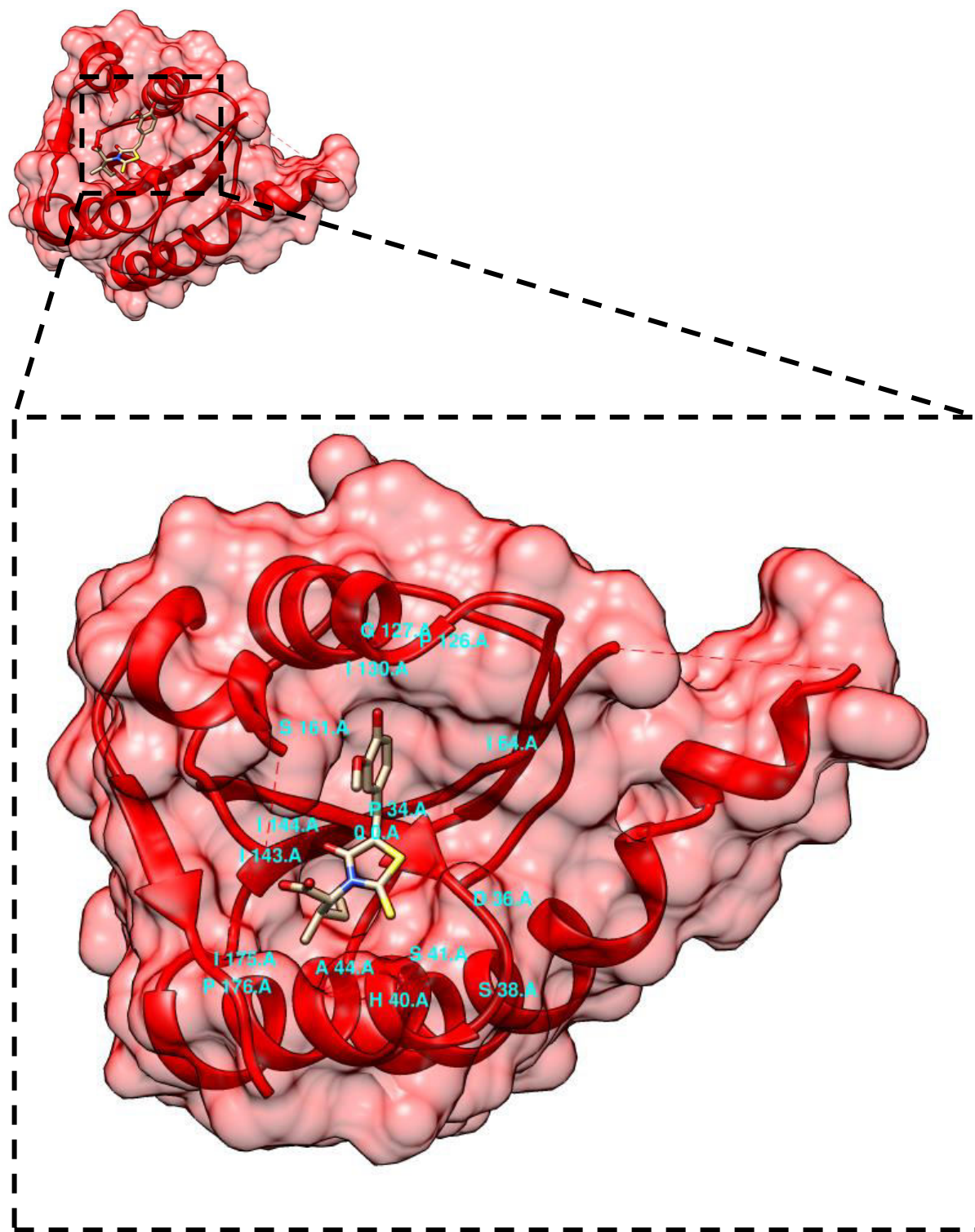
**Fig 7.1: Generated Pharmacophore model from ZINC PHARMER.** The 3D structure of the G4LZI3 protein was constructed by XRD and curcumin was identified as a potential inhibitor. Features used for screening are highlighted with spheres: purple (aromatic ring), orange (hydrogen bond donating sites) and green (hydrogen bond acceptor sites).



**Fig. 7.2: Model of G4LZI3 in complex with Inhibitor I.** The structure of the G4LZI3 structure docked with inhibitor with ID: ZINC00927367. The docked complex had a Z-score of 5.6.



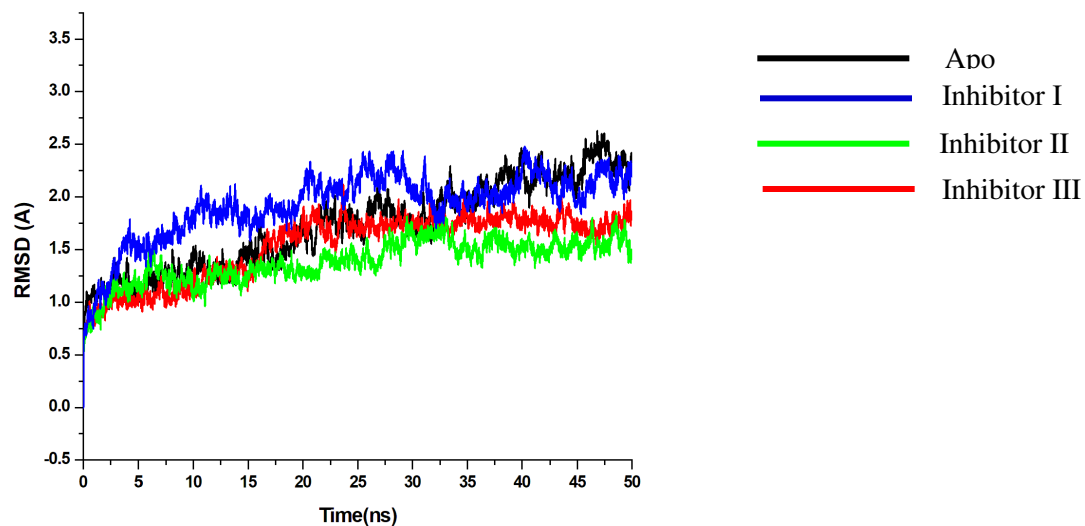
**Fig. 7.3: Model of G4LZI3 in complex with Inhibitor II.** The structure of the G4LZI3 structure docked with inhibitor with ID: ZINC13576641. The docked complex had a Z-score of 5.9.



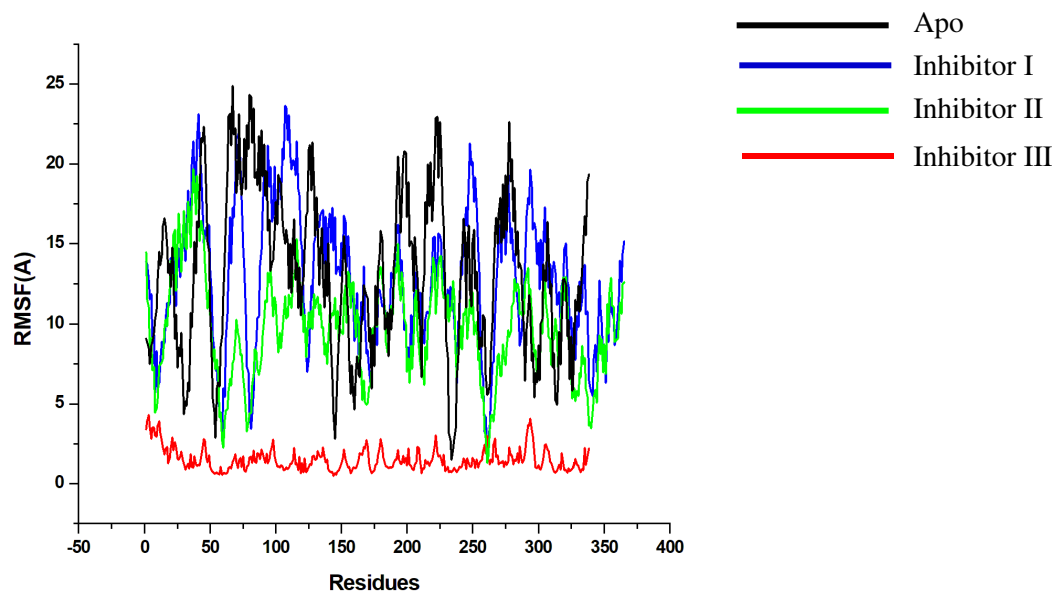
**Fig. 7.4: Model of G4LZI3 in complex with Inhibitor III.** The structure of the G4LZI3 structure docked with inhibitor with ID: ZINC01276959. The docked complex had a Z-score of 5.7.

#### 7.4.2 Molecular Dynamics and Post Molecular Dynamic Simulations

Conformational analyses of the docked complexes were performed for partial insight into the biological effect of the ligands with the protein. Root mean square deviation (RMSD) portrays the stability of a system and its associated stereochemical variability (Pitera, 2014; Schreiner *et al.*, 2012). It depicts the atom mobility and changes a system undergoes during MD simulations (Fu *et al.*, 2018). Therefore, RMSD values are directly proportional to mobility. The results in Fig. 7.5 indicate that the systems remained stable with minimal fluctuations or deviations from the unbound system with the binding of each ligand to the active site of the protein. Stability of a system is also correlated by the amount of fluctuations not going beyond 2Å (Fernando *et al.*, 2019). The average RMSD fluctuation values of the apoprotein were 2.25Å, Complex 1 (3.78Å), Complex 2 (2.83Å), Complex 3 (2.79Å). These values were slightly above 2 and higher than that of the apoprotein. However, in comparison to the apo system, the RMSD profile of Inhibitor II and III decreased at around 20 ns after ligand binding. A decrease in fluctuation is expected indicating restricted motion which the complexes displayed. Inhibitor I on the other hand increased and became stable around 45 ns but the profile was the same as the apo system. This indicated weak inhibitory effect of the ligand.



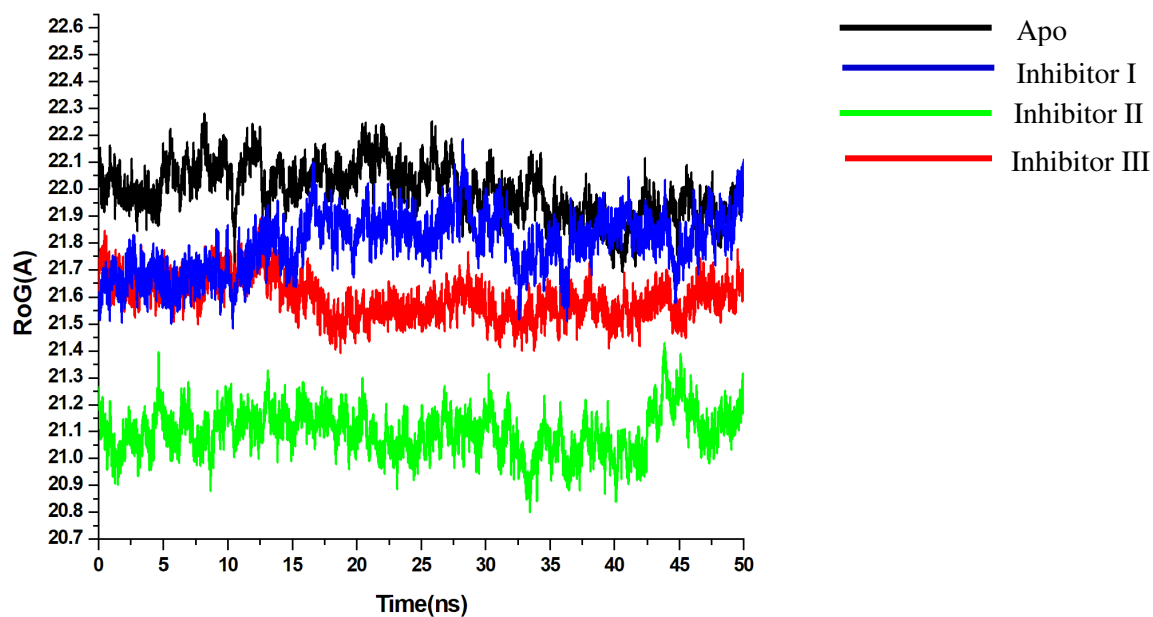
**Fig. 7.5:** Graph showing Root mean square deviation of G4LZI3 in complex with Inhibitor I, II and III. Apo structure is represented by Black; in complex with Inhibitor I (Blue), Inhibitor II (Green) and Inhibitor III (Red).



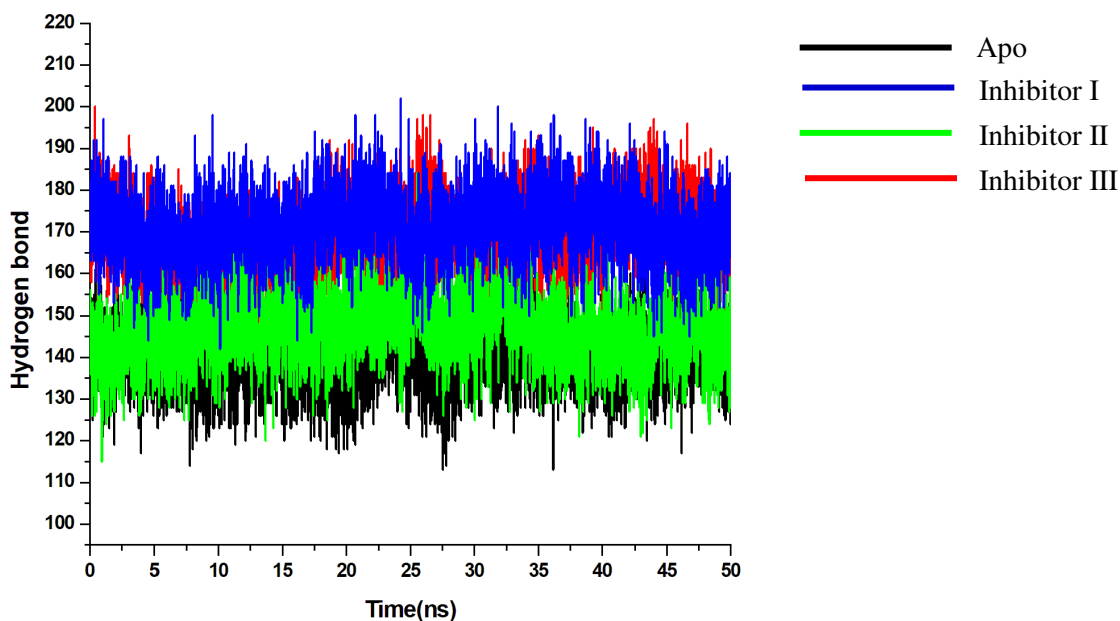
**Fig. 7.6:** Graph showing Root mean square fluctuation of G4LZI3 in complex with Inhibitor I, II and III. Apo structure is represented by Black; in complex with Inhibitor I (Blue), Inhibitor II (Green) and Inhibitor III (Red).

The conformational changes a protein undergoes upon interaction with another biological molecule may indicate function of that particular protein (Condic-Jurkic *et al.*, 2018). Root mean square fluctuation quantifies the flexibility of the C $\alpha$  backbone of a protein. The instability experienced by the system caused by conformational changes that are induced by ligand binding, can restrict the intrinsic flexibility of a protein or the region to which the ligand bound to. Inhibitor 3 did not experience any significant fluctuation and may indicate instantaneous and tight binding of the inhibitor to the protein thereby restricting motion. Significant fluctuations were observed in all other systems indicating weak binding (Fig. 7.6).

The compactness of a system is defined by its radius of gyration (RoG), which is denoted by the root mean square distance between the centre of gravity and the edges of an object (Lobanov *et al.*, 2008; Verma *et al.*, 2016). It is also indicator of protein mass and the folding state of a complex. In addition it complements information pertaining to the stability of biological macromolecules during the molecular dynamic simulation process. The RoG results displayed in Fig. 7.7 saw a decrease in the complex bound to Inhibitor III in comparison to the unbound complex. However, Inhibitor II exhibited the most significant decrease, revealing a more compact structure in comparison to the other complexes. The presence of hydrogen bonds is key for the interactions between various ligands to a macromolecule and is crucial for protein stability (Fu *et al.*, 2018). Fig. 7.8 shows the formation of hydrogen bonds and electrostatic interactions which play a critical role in the binding of each ligand to the G4LZI3 protein. As the catalytic space in the protein increased, the flexibility of the protein increased but it was quickly stabilized by the binding of the ligand. Stability of the complexes was facilitated by the formation of hydrogen bond interactions.



**Fig. 7.7: Radius of gyration profile of G4LZI3 G4LZI3 in complex with Inhibitor I, II and III.** Apo structure is represented by Black; in complex with Inhibitor I (Blue), Inhibitor II (Green) and Inhibitor III (Red).



**Fig. 7.8: Hydrogen Bond Profile of G4LZI3 in complex with Inhibitor I, II and III.** Apo structure is represented by Black; in complex with Inhibitor I (Blue), Inhibitor II (Green) and Inhibitor III (Red).

## 7.5 Conclusion

This chapter was able to show the inhibitory potential of compounds screened via the generation of a pharmacophore for structure based drug design. Inhibitor I and II complexes signified very weak binding based on observations of the RMSF and RoG profiles. Significant fluctuations without the systems reaching stability were observed and the RoG profiles also did not observe any decrease in values, which is suggestive on ligand binding and reduced flexibility. Inhibiter III on the other hand, displayed instant ligand binding and consistent fluctuation in the RMSF profile. This was confirmed by the RoG profile, which displayed a significant decrease at around 20ns. It is advised therefore that *in vivo* and *in vitro* studies should concentrate on the synthesizing of this compound for experimental studies.

## References

- Aanandhi MV, Bhattacharjee D, George PSG and Ray A (2014). Natural polyphenols down-regulate universal stress protein in Mycobacterium tuberculosis: An in-silico approach. *Journal of advanced pharmaceutical technology & research* **5**(4): 171.
- Adenowo AF, Oyinloye BE, Oggunyinka BI and Kappo AP (2015). Impact of human schistosomiasis in sub-Saharan Africa. *Brazilian Journal of Infectious Diseases* **19**(2): 196-205.
- Barsoum RS, Esmat G and El-Baz T (2013). Human schistosomiasis: clinical perspective: review. *J Adv Res* **4**(5): 433-444.
- Berendsen HJ, Postma JV, Van Gunsteren WF, Dinola A and Haak JR (1984). Molecular dynamics with coupling to an external bath. *The Journal of chemical physics* **81**(8): 3684-3690.
- Case D, Babin V, Berryman J, Betz R, Cai Q, Cerutti D, Cheatham Iii T, Darden T, Duke R and Gohlke H (2014). AMBER 14 Reference Manual. *Table of Contents graphic* **1**(2): 3.
- Colley DG, Bustinduy AL, Secor WE and King CH (2014). Human schistosomiasis. *The Lancet* **383**(9936): 2253-2264.
- Condic-Jurkic K, Subramanian N, Mark AE and O'mara ML (2018). The reliability of molecular dynamics simulations of the multidrug transporter P-glycoprotein in a membrane environment. *PloS one* **13**(1): e0191882.
- Fernando F, Irawan MI and Fadlan A. (2019). *Bat Algorithm for Solving Molecular Docking of Alkaloid Compound SA2014 Towards Cyclin D1 Protein in Cancer*. Paper presented at the Journal of Physics: Conference Series.

- Ferreira LG, Dos Santos RN, Oliva G and Andricopulo AD (2015). Molecular docking and structure-based drug design strategies. *Molecules* **20**(7): 13384-13421.
- Fu Y, Zhao J and Chen Z (2018). Insights into the Molecular Mechanisms of Protein-Ligand Interactions by Molecular Docking and Molecular Dynamics Simulation: A Case of Oligopeptide Binding Protein. *Computational and mathematical methods in medicine* **2018**.
- Galindo-Murillo R, Robertson JC, Zgarbova M, Sponer J, Otyepka M, Jurečka P and Cheatham Iii TE (2016). Assessing the current state of amber force field modifications for DNA. *Journal of Chemical Theory and Computation* **12**(8): 4114-4127.
- Gryseels B, Polman K, Clerinx J and Kestens L (2006). Human schistosomiasis. *The Lancet* **368**(9541): 1106-1118.
- Hajissa K, Muhajir AEM, Eshag HA, Alfadel A, Nahied E, Dahab R, Ali SM, Mohammed M, Gaafar M and Mohamed Z (2018). Prevalence of schistosomiasis and associated risk factors among school children in Um-Asher Area, Khartoum, Sudan. *BMC research notes* **11**(1): 779.
- Huang H-H, Rigouin C and Williams DL (2012). The redox biology of schistosome parasites and applications for drug development. *Current pharmaceutical design* **18**(24): 3595-3611.
- Isokpehi RD, Mahmud O, Mbah AN, Simmons SS, Avelar L, Rajnarayanan RV, Udensi UK, Ayensu WK, Cohly HH and Brown SD (2011). Developmental regulation of genes encoding universal stress proteins in *Schistosoma mansoni*. *Gene regulation and systems biology* **5**: GRSB. S7491.
- Jauréguiberry S, Paris L and Caumes E (2010). Acute schistosomiasis, a diagnostic and therapeutic challenge. *Clinical Microbiology and Infection* **16**(3): 225-231.

- Koes DR and Camacho CJ (2012). ZINCPharmer: pharmacophore search of the ZINC database. *Nucleic Acids Research* **40**(W1): W409-W414.
- Lobanov MY, Bogatyreva N and Galzitskaya O (2008). Radius of gyration as an indicator of protein structure compactness. *Molecular Biology* **42**(4): 623-628.
- Lounnas V, Ritschel T, Kelder J, Mcguire R, Bywater RP and Foloppe N (2013). Current progress in structure-based rational drug design marks a new mindset in drug discovery. *Computational and structural biotechnology journal* **5**(6): e201302011.
- Melman SD, Steinauer ML, Cunningham C, Kubatko LS, Mwangi IN, Wynn NB, Mutuku MW, Karanja DM, Colley DG and Black CL (2009). Reduced susceptibility to praziquantel among naturally occurring Kenyan isolates of *Schistosoma mansoni*. *PLoS neglected tropical diseases* **3**(8): e504.
- Molyneux DH, Savioli L and Engels D (2017). Neglected tropical diseases: progress towards addressing the chronic pandemic. *The Lancet* **389**(10066): 312-325.
- Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS and Olson AJ (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry* **30**(16): 2785-2791.
- Pitera JW (2014). Expected distributions of root-mean-square positional deviations in proteins. *The Journal of Physical Chemistry B* **118**(24): 6526-6530.
- Qing X, Lee XY, De Raeymaecker J, Tame JR, Zhang KY, De Maeyer M and Voet A (2014). Pharmacophore modeling: advances, limitations, and current utility in drug discovery. *Journal of Receptor, Ligand and Channel Research* **7**: 81-92.

- Ramírez D and Caballero J (2016). Is it reliable to use common molecular docking methods for comparing the binding affinities of enantiomer pairs for their protein target? *International journal of molecular sciences* **17**(4): 525.
- Roy K, Kar S and Das RN (2015). *Understanding the basics of QSAR for applications in pharmaceutical sciences and risk assessment*: Academic press.
- Sanner MF (2008). The Python interpreter as a framework for integrating scientific computing software-components. *Scripps Res. Inst.* **26**(1): 1-12.
- Schreiner W, Karch R, Knapp B and Ilieva N (2012). Relaxation estimation of RMSD in molecular dynamics immunosimulations. *Computational and mathematical methods in medicine* **2012**.
- Trott O and Olson AJ (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry* **31**(2): 455-461.
- Vale N, Gouveia MJ, Rinaldi G, Brindley PJ, Gärtner F and Da Costa JMC (2017). Praziquantel for schistosomiasis: single-drug metabolism revisited, mode of action, and resistance. *Antimicrobial agents and chemotherapy* **61**(5): e02582-02516.
- Verma S, Grover S, Tyagi C, Goyal S, Jamal S, Singh A and Grover A (2016). Hydrophobic interactions are a key to MDM2 inhibition by polyphenols as revealed by molecular dynamics simulations and MM/PBSA free energy calculations. *PloS one* **11**(2): e0149014.
- Vollmer AC and Bark SJ (2018). Twenty-five years of investigating the universal stress protein: function, structure, and applications *Advances in Applied Microbiology* (Vol. 102.): Elsevier pp. 1-36.

Wang J, Wolf RM, Caldwell JW, Kollman PA and Case D (2005). Erratum: Development and testing of a general amber force field (Journal of Computational Chemistry (2004) 25 (1157)). *Journal of Computational Chemistry* **26**(1).

Wang X-F, Su J, Yang N, Zhang H, Cao X-Y and Kang J-F (2017). Functional characterization of selected universal stress protein from *Salvia miltiorrhiza* (SmUSP) in *Escherichia coli*. *Genes* **8**(9): 224.

Woldegerima E, Bayih AG, Tegegne Y, Aemero M and Jejawi Zeleke A (2019). Prevalence and Reinfection Rates of *Schistosoma mansoni* and Praziquantel Efficacy against the Parasite among Primary School Children in Sanja Town, Northwest Ethiopia. *Journal of parasitology research* **2019**.

Yu W and Mackerell AD (2017). Computer-aided drug design methods *Antibiotics*: Springer pp. 85-106.

## Chapter 8: General Discussion and Conclusion

The schistosome parasite causes the deaths of over 200 000 people annually worldwide, while leaving most survivors with chronic infections and severe cases of morbidity. Despite the plethora of environmental insults the parasite is plagued with throughout its developmental cycle, the pathogen still thrived and even has the ability to co-exist within the human host for up to three years, producing millions of eggs per day. The release of antigens from tissue-trapped eggs is the main underlying factors for the pathologies associated with the disease. Although reactive oxygen species are produced by the host along with strong immunological responses to rid the host of these parasites, inflammation and tissue damage manifest causing harm to the host instead. It has been proposed that through the action of Usps, schistosomes possess the ability to avert and survive the diverse stress conditions they are exposed to during their lifecycle. Therefore, studying this class of proteins opens up avenues for the design of alternative treatment in light of the setbacks associated with Praziquantel, only chemotherapeutic drug for the treatment of the disease, the biggest of which is drug resistance.

Based on previous studies, this project began by taking an atomistic structural and functional look into the binding of a select range of polyphenols onto the G4LZI3 protein. These were considered, not only because they are natural compounds but also because of their free radical scavenging and inflammation-reducing activities. Hence they were investigated as potential inhibitors of the G4LZI3 protein. Amongst the array of compounds, distinct structural and conformational alterations were observed in the interaction between the G4LZI3 Usp and catechin and curcumin respectively. The high binding energies exhibited by the docked complexes suggested inhibitory activity given that high values indicate increased non-covalent

interactions, in turn restricting protein mobility. Docking and molecular dynamics studies were conducted by Aanandi and co-workers between the polyphenols mentioned in this current project and the Rv1636 Universal Stress Protein from *Mycobacterium tuberculosis*. Results pointed towards curcumin as the polyphenol exhibiting the highest binding energy, followed by catechin and quercetin (Aanandhi *et al.*, 2014). Hydrogen and ionic bond interactions were observed between the ligands and the protein within the docked complexes, which were suggested to take place via carboxyl groups of the ligands. The G4LZI3 protein in complex with Catechin, Dalbergin and Quercetin systems exhibited the lowest average residue fluctuations as a result of the binding of the ligands to the protein thereby reducing the fluctuation of the C $\alpha$  backbone of the protein. Amid the polyphenols chosen in this study, of particular interest was curcumin, a turmeric-derived micronutrient shown to regulate various enzymes, cytokines and transcription factors implemented in inflammatory conditions (Allam, 2009). Hepatic enzyme activities were restored back to normal while catalase activity was up-regulated in the liver of acute *S. mansoni* infected mice following treatment with curcumin. IL-12 and TNF- $\alpha$ , IL-10, IgM and IgG2a serum levels as well as soluble worm antigens (SWAP) and soluble egg antigens (SEA) saw significant reduction, maintenance and increase in their levels respectively, proving curcumin's role in the modulation and reduction of liver pathology and parasitic burden in schistosome infected mice (Allam, 2009). These results were further confirmed by Al-Agamy *et al* who showed the effective anti-fibrotic activity of curcumin in blocking and reversing liver fibrosis developed in *S. mansoni* infected mice (El-Agamy *et al.*, 2011). Other studies have demonstrated curcumin's ability to detach schistosome worm pairs, disrupt egg fertility and decrease parasite viability and oviposition resulting on the death of worms (Morais *et al.*, 2013). All these highlight curcumin as a potent molecule that should be investigated further for alternative anti-schistosomes.

Key interacting residues (Asp11; Ile39; Gly120; Arg122 and Gly123) were discovered within the best five docked complexes, which all play an active role in the binding and stabilization of each ligand to the G4LZI3 protein. However, Asp11 displayed the highest energy contribution, due to the formation of hydrogen bonds with each ligand. The discovery of these residues provides a route map for structure-based drug design of novel compounds, which could possess minimal toxicity against the G4LZI3 protein and enhanced inhibitory potency, even Praziquantel (PZQ). A study by Hussein and colleagues showed significant reductions in worm burden, exhibiting total clearance of adult worms and decrease in granuloma size following treatment of *S. mansoni* infected mice with combinatory turmeric and PZQ treatment (Hussein *et al.*, 2017). The ability of polyphenols to scavenge oxidative stress associates these compounds with the prevention of diseases or infections that are linked to the production of ROS. Therefore one future research aim may be the investigation of polyphenols as potential nutraceuticals for schistosomiasis.

The main aim of this project aimed at the determination of the three dimensional structure of the G4LZI3 by X-ray crystallography. This was achieved using the vapor diffusion sitting drop method. This technique proved efficient as crystals appeared in a matter of days at room temperature. Heterologous over-expression of the G4LZI3 protein was done in M15 and JM109 cells at 25°C using 0.5 mM IPTG. Purification of the protein to homogeneity was performed using affinity chromatography on a Nickel-NTA column by size exclusion. SDS-PAGE analysis revealed pure protein, which was subjected to crystallization. Yielded crystals were diffracted and two of the best datasets with resolutions of 3.36Å and 3.05Å were used for structure determination. Molecular replacement was done using MR\_rosetta to generate a starting model and a density electron map. Model building and refinement were achieved using *PHENIX Autobuild*, Coot and *PHENIX.refine*. After 22 cycles of refinement, the model achieved  $R_{work}$  and

$R_{\text{free}}$  values of 0.27 and 0.33, a significant improvement from 0.37 and 0.44 initial values. The model revealed that the structure consists of 4 parallel  $\beta$ -sheets and 3  $\alpha$ -helices with the presence of a short  $3_{10}$  helix positioned close to the C-terminal of the protein. A generated Ramachandran plot showed that though there is still room for improvement, the solved structure is a relatively good model with 92% of the residues present in the favored and allowed regions. Comparison of the model with the At3g01520 USP-like protein from *Arabidopsis thaliana* suggests that the G4LZI3 protein may potentially be a member of the nucleotide-binding USP subfamily due to the possibility of the protein binding to either ATP or AMP. The missing residues in the G4LZI3 structure indicate the presence of flexible loops, which may be stabilized by the binding of ATP or AMP molecules as was suggested for the At3g01520 USP (Kim *et al.*, 2015). Despite structure determination of the At3g01520 protein by XRD, the biological function of the protein is still unclear but has been suggested to play a role as a molecular switch through its ATP binding and hydrolyzing properties. Therefore, designing experiments to identify the relationship between the ATP hydrolyzing activities and the G4LZI3 protein, after determining its binding potential to ATP, would elucidate on the function of this protein in various stress-related signaling pathways within the schistosome parasite.

Lastly, a pharmacophore model was generated using ZINC PHARMER for the G4LZI3 USP. Curcumin was used as the template due to its high binding affinity observed in Chapter 5. The best results were of Inhibitor III which portrayed consistent and decreased fluctuations in the RMSF and RoG post MD simulations. RMSD profiles displayed reduced flexibility of the protein once the inhibitor was bound. This study forms a basis for the synthesis of compounds that can be tested both *in vivo* and *in vitro* in a bid to design alternative treatment for schistosomiasis.

The biggest shortfall of this study is the poor resolution of the structure. Due to the fact that missing residues are still present within the solved structure, the first port of call regarding any future work would first revolve around additional structure building and refinement. This would bring insight into the supposed interaction of the protein with either ATP or AMP. Other future perspectives would include conducting interaction studies of the G4LZI3 protein with the above-mentioned ligands by Isothermal Titration Calorimetry or Surface Plasmon Resonance. This would served as a complementary and confirmatory tool as regarding the first suggestion. This would then be followed by the co-crystallization of the G4LZI3 protein with ATP so as to solve the complex with x-ray crystallography. Finally, it would be ideal to design and synthesize the screened small molecule inhibitors for *in vivo* and *in vitro* experiments to ultimately determine and confirm the inhibitors as alternative and novel anti-schistosomids.

## References

- Aanandhi MV, Bhattacharjee D, George PSG and Ray A (2014). Natural polyphenols down-regulate universal stress protein in Mycobacterium tuberculosis: An in-silico approach. *Journal of advanced pharmaceutical technology & research* **5**(4): 171.
- Allam G (2009). Immunomodulatory effects of curcumin treatment on murine schistosomiasis mansoni. *Immunobiology* **214**(8): 712-727.
- El-Agamy DS, Shebl AM and Said SA (2011). Prevention and treatment of Schistosoma mansoni-induced liver fibrosis in mice. *Inflammopharmacology* **19**(6): 307-316.
- Hussein A, Rashed S, El Hayawan I, El-Sayed R and Ali H (2017). Evaluation of the Anti-schistosomal Effects of Turmeric (Curcuma longa) Versus Praziquantel in Schistosoma mansoni Infected Mice. *Iranian journal of parasitology* **12**(4): 587.
- Kim DJ, Bitto E, Bingman CA, Kim HJ, Han BW and Phillips Jr GN (2015). Crystal structure of the protein A t3g01520, a eukaryotic universal stress protein-like protein from arabidopsis thaliana in complex with AMP. *Proteins: Structure, Function, and Bioinformatics* **83**(7): 1368-1373.
- Morais ER, Oliveira KC, Magalhães LG, Moreira ÉB, Verjovski-Almeida S and Rodrigues V (2013). Effects of curcumin on the parasite Schistosoma mansoni: a transcriptomic approach. *Molecular and biochemical parasitology* **187**(2): 91-97.

# APPENDIX I: General Stock Solutions, Buffers and Media

**4X Sample Buffer:** 10 mL Stacking Buffer, 8 mL Glycerol, 0.8 g Sodium Dodecyl Sulfate (SDS), 0.8 mL 2-mercapto-ethanol/DTT, 0.2 mg Bromophenol Blue and 1.2 mL deionized water

**5X Running Buffer:** 15.1 g Tris, 72 g Glycine and 5 g SDS dissolved in 1 L deionized water

**Acrylamide/Bis:** 30 g Acrylamide and 0.8 g *bis*/Acrylamide dissolved in 100 mL deionized water

**Ammonium Persulphate (APS):** A 10% stock solution was made by dissolving 1 g APS in deionized water and storing at -20°C

**Ampicillin:** A 100 mg/ml stock solution was made by dissolving 1 g Ampicillin in 10 mL deionized water, was aliquoted and stored at -20°C. The final working concentration of the solution was 100 µg/mL

**Cell Lysis Buffer:** 50 mM Tris-HCL, 100 mM NaCl and 10 mM Imidazole

**Coomasie Blue Stain:** 0.25 g Coomasie Blue R-250, 100 mL acetic acid, 450 mL Methanol/Ethanol and 450 mL deionized water

**Destaining Solution:** 15% Acetic Acid OR 100 mL Acetic Acid, 200 mL Methanol and 700 mL deionized water

**Dithiothreitol (DTT):** A 1 M stock solution was made by dissolving 1.5 g DTT in 10 mL deionized water and stored in -20°C

**Elution Buffer:** 50 mM Tris (pH 8.0), 500 mM NaCl, 500 mM Imidazole. This solution was adjusted to pH7.4–8.0

**Equilibration/Wash Buffer:** 50 mM Tris (pH 8.0), 500 mM NaCl, 40 mM Imidazole. This solution was adjusted to pH7.4–8.0

**IPTG:** A 1 M stock solution was made by dissolving 2.38 g IPTG in 10 mL deionized water and stored at -20°C. The final working concentration of the solution was 0.5 mM

**Kanamycin:** A stock solution of 50 mg/mL was made by dissolving 0.5 g Kanamycin monosulphate in 10 mL distilled water. The final working concentration of the solution was 25 µg/mL

**Luria Broth:** 50 g Luria Broth (10.0 g/L Tryptone, 5.0 g/L Yeast extract and 10.0 g/L Sodium chloride) was dissolved in 2 L deionized water and autoclaved for 1 hour at 120°C

**Lysozyme:** A stock solution of 25 mg/mL was made by dissolving 0.5 g Lysozyme in 20 mL distilled water. The solution was aliquoted and stored at -20°C. The final working concentration of the solution was 100 µg/mL

**Nutrient Agar:** 31 g Nutrient agar 1.0 g/L Meat extract, 5.0 g/L Peptone, 2.0 g/L Yeast extract, 8.0 g/L Sodium chloride, 15 g/L Agar was dissolved in 1 L distilled water and autoclaved at 120°C for 1 hr

**PMSF:** A 1 M stock solution was made by dissolving 1.74 g PMSF in 100 mL isopropanol, methanol or ethanol. Final working concentration used was 1 mM

**Running Buffer (5X):** 15.1 g Tris, 72 g Glycine and 5 g SDS dissolved in 1 L deionized water. The pH was approximately 8.3–8.5 but was not adjusted

**Size Exclusion Storage Buffer:** 50 mM Tris and 150 mM NaCl adjusted to pH8.0

**Separating/Resolving Gel Buffer:** 18.2 g Tris and 0.4 g SDS dissolved in 100 mL deionized water and adjusted to pH8.8

**Stacking Gel Buffer:** 6.05 g Tris and 0.4 g SDS dissolved in 100 mL deionized water and adjusted to pH6.8

**Sodium Azide:** A 10% stock solution was made dissolving 10 g  $\text{NaN}_3$  in 100 mL distilled water.

The final working concentration used was 0.02%

**TFB1 Transformation Buffer:** 12 g/L 100 mM  $\text{RbCl}$ , 9.89 g/L 50 mM  $\text{MnCl}_2 \cdot 4\text{H}_2\text{O}$ , 2.94 g/L 30 mM Potassium acetate, 1.47 g/L 10 mM  $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$  and 150 ml/L 15% (v/v) Glycerol. This solution was adjusted to pH5.8 with dilute acetic acid, filter sterilized and stored in a brown bottle at 4°C.

**TFB2 Transformation Buffer:** 2.09 g/L 10 mM MOPS, 1.2 g 10 mM  $\text{RbCl}$ , 11 g/L 75 mM  $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$  and 150 ml/L 15% (v/v) glycerol. This solution was adjusted to pH6.5 with dilute acetic acid, filter sterilized and stored in a brown bottle at 4°C.

**Transformation Storage Buffer (TSB):** 1.6 g Peptone, 1.0 g Yeast Extract, 0.5 g  $\text{NaCl}$ , 10.0 g PEG (MW3350), 5.0 mL DMSO, 1.0 mL 1 M  $\text{MgCl}_2$  and 1.0 mL 1 M  $\text{MgSO}_4$ . This was made up to 100 mL with deionized water and filter-sterilized

**Transformation Storage Buffer Glucose (TSBG):** 1.6 g Peptone, 1.0 g Yeast Extract, 0.5 g  $\text{NaCl}$ , 10.0 g PEG (MW3350), 5.0 mL DMSO, 1.0 mL 1 M  $\text{MgCl}_2$ , 1.0 mL 1 M  $\text{MgSO}_4$  and 2 mL 20% glucose. This was made up to 100 mL with deionized water and filter-sterilized

## APPENDIX II: Bacterial Strains Used

1. *E. coli* DH5 $\alpha$  (Invitrogen): F<sup>-</sup> endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG  $\Phi$ 80 $\Delta$ lacZ $\Delta$ M15  $\Delta$ (lacZYA-argF)U169, hsdR17(r<sub>K</sub><sup>-</sup> m<sub>K</sub><sup>+</sup>),  $\lambda$ -
2. *E. coli* JM109 (DE3) (Promega): endA1 glnV44 thi-1 relA1 gyrA96 recA1 mcrB<sup>+</sup>  $\Delta$ (lac-proAB) e14- [F' traD36 proAB<sup>+</sup> lacIq lacZ $\Delta$ M15] hsdR17(rK-mK<sup>+</sup>).
3. *E. coli* M15 (pREP4) (Qiagen): F<sup>-</sup>,  $\Phi$ 80 $\Delta$ lacM15, thi, lac<sup>-</sup>, mtl<sup>-</sup>, recA<sup>+</sup> , KmR

# APPENDIX III: Morpheus Screen Constituents

Table 1: List of PDB ligands in Morpheus®

| PDB Ligand name(s)                                         | Class            | PDB ID(s)     | Structure Hits* |
|------------------------------------------------------------|------------------|---------------|-----------------|
| 1,2-Ethanediol (ethylene glycol)                           | Precipitant      | EDO, EGL      | 1081            |
| 1,2-Propanediol (enantiomers R and S)                      | Alcohols         | PGO, PGR      | 41              |
| 1,3-Propanediol                                            | Alcohols         | PDO           | 7               |
| 1,4-Butanediol                                             | Alcohols         | BU1           | 11              |
| 1,6-Hexanediol                                             | Alcohols         | HEZ           | 19              |
| 1-Butanol                                                  | Alcohols         | 1BO           | 7               |
| 2-(N-Morpholino)-ethane sulfonic acid (MES)                | Buffer           | MES           | 315             |
| 2-Amino-2-hydroxymethyl-propane-1,3-diol (Tris)            | Buffer           | TRS           | 334             |
| 2-Methyl-2,4-pentanediol (MPD, enantiomers R and S)        | Precipitant      | MPD, MRD      | 504             |
| 3-Morpholinopropane-1-sulfonic acid (MOPS)                 | Buffer           | MPO           | 21              |
| 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) | Buffer           | EPE           | 201             |
| Acetic acid, acetate, acetyl                               | Carboxylic acids | ACY, ACT, ACE | 1890            |
| (S)-2-Aminopropanoic acid (Alanine, (enantiomers L and D)  | Amino acids      | ALA, DAL      | 35              |
| Amino, Ammonia, Ammonium                                   | multiple         | NH2, NH3, NH4 | 582             |
| N,N-bis(2-hydroxyethyl)glycine (Bicine)                    | Buffer           | BCN           | 13              |
| Bromide                                                    | Halogens         | BR            | 120             |
| Calcium                                                    | Divalents        | CA            | 3959            |
| Chloride                                                   | Multiple         | CL            | 2842            |
| Citric acid, citrate                                       | Carboxylic acids | CIT, FLC      | 384             |
| D-Galactose (anomers $\alpha$ and $\beta$ )                | Monosaccharides  | GAL, GLA      | 86              |
| D-Glucose (anomers $\alpha$ and $\beta$ )                  | Monosaccharides  | GLC, BGC      | 206             |
| Glutamic acid (enantiomers L and D)                        | Precipitant      | GLU, DGL      | 75              |
| Di(Hydroxyethyl)ether (Di-Ethyleneglycol)                  | Ethylene glycols | PEG           | 209             |
| D-Mannose (anomers $\alpha$ and $\beta$ )                  | Monosaccharides  | MAN, BMA      | 178             |
| D-Xylopyranose (anomers $\alpha$ and $\beta$ )             | Monosaccharides  | XYL, XYP      | 41              |
| Fluoride                                                   | Halogens         | F             | 16              |
| Formic acid                                                | Carboxylic acids | FMT           | 267             |
| Glycerol                                                   | Amino acids      | GOL           | 2884            |
| Glycine                                                    | Buffer           | GLY           | 50              |
| Imidazole                                                  | Halogens         | IMD           | 154             |
| Iodide                                                     | Alcohols         | IOD           | 178             |
| Isopropyl alcohol (iso-propanol, 2-Propanol)               | Monosaccharides  | IPA, IOH      | 174             |
| L-Fucose (anomers $\alpha$ and $\beta$ )                   | Amino acids      | FUC, FUL      | 62              |
| Lysine (enantiomers L and D)                               | Amino acids      | LYS, DLY      | 36              |
| Magnesium                                                  | Divalents        | MG            | 3991            |
| N-Acetyl-d-glucosamine (anomers $\alpha$ and $\beta$ )     | Monosaccharides  | NAG,NBG       | 1150            |
| Nitrate                                                    | NPS              | NO3           | 156             |
| Oxamic acid                                                | Carboxylic acids | OXM           | 17              |
| Penta(hydroxyethyl)ether (Penta-Ethyleneglycol)            | Ethylene glycols | 1PE           | 91              |
| Phosphates                                                 | NPS              | PO4, PI, 2HP  | 1687            |
| Potassium                                                  | Carboxylic acids | K             | 720             |
| Serine (enantiomers L and D)                               | Amino acids      | SER, DSN      | 38              |
| Sodium                                                     | multiple         | NA            | 1926            |
| Sulfate                                                    | NPS              | SO4           | 5793            |
| Tartaric acid (enantiomers R and S)                        | Carboxylic acids | TAR, TLA      | 113             |
| Tetra(hydroxyethyl)ether (Tetra-Ethyleneglycol)            | Ethylene glycols | PG4           | 194             |
| Tri(Hydroxyethyl)ether (Tri-Ethyleneglycol)                | Ethylene glycols | PGE           | 107             |
| SUM                                                        |                  |               | 32956           |

\* as of July 2008.

**Table 2: Mixes of additives used in Morpheus®**

| Mix name         | Composition                                                                                                                                             | Catalogue Number (100 mL) | Catalogue Number (250 mL) |
|------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------|---------------------------|
| Divalents        | 0.3M Magnesium chloride hexahydrate; 0.3M Calcium chloride dihydrate                                                                                    | MD2-100-70                | MD2-250-70                |
| Halogens         | 0.3M Sodium fluoride; 0.3M Sodium bromide; 0.3M Sodium iodide                                                                                           | MD2-100-71                | MD2-250-71                |
| NPS <sup>†</sup> | 0.3M Sodium nitrate, 0.3 Sodium phosphate dibasic, 0.3M Ammonium sulfate                                                                                | MD2-100-72                | MD2-250-72                |
| Alcohols         | 0.2M 1,6-Hexanediol; 0.2M 1-Butanol<br>0.2M 1,2-Propanediol; 0.2M 2-Propanol; 0.2M 1,4-Butanediol; 0.2M 1,3-Propanediol                                 | MD2-100-73                | MD2-250-73                |
| Ethylene glycols | 0.3M Diethylene glycol; 0.3M Triethylene glycol;<br>0.3M Tetraethylene glycol; 0.3M Pentaethylene glycol                                                | MD2-100-74                | MD2-250-74                |
| Monosaccharides  | 0.2M D-Glucose; 0.2M D-Mannose; 0.2M D-Galactose; 0.2M L-Fucose; 0.2M D-Xylose; 0.2M N-Acetyl-D-Glucosamine                                             | MD2-100-75                | MD2-250-75                |
| Carboxylic acids | 0.2M Sodium formate; 0.2M Ammonium acetate;<br>0.2M Sodium citrate tribasic dihydrate; 0.2M Sodium potassium tartrate tetrahydrate; 0.2M Sodium oxamate | MD2-100-76                | MD2-250-76                |
| Amino acids      | 0.2M DL-Glutamic acid monohydrate; 0.2M DL-Alanine; 0.2M Glycine; 0.2M DL-Lysine monohydrochloride; 0.2M DL-Serine                                      | MD2-100-77                | MD2-250-77                |

<sup>†</sup>NPS; Nitrate Phosphate Sulfate

**Table 3: Buffer systems used in Morpheus®**

| Mix name        | Conc. | pH @ 20°C | Composition                       | Catalogue Number (100 mL) | Catalogue Number (250 mL) |
|-----------------|-------|-----------|-----------------------------------|---------------------------|---------------------------|
| Buffer System 1 | 1.0M  | 6.5       | Imidazole; MES monohydrate (acid) | MD2-100-100               | MD2-250-100               |
| Buffer System 2 | 1.0M  | 7.5       | Sodium HEPES; MOPS (acid)         | MD2-100-101               | MD2-250-101               |
| Buffer System 3 | 1.0M  | 8.5       | Tris (base); BICINE               | MD2-100-102               | MD2-250-102               |

**Table 4: Mixes of Precipitants<sup>1</sup> used in Morpheus®**

| Mix name          | Old Mix Name  | Composition                                 | Catalogue Number (100 mL) | Catalogue Number (250 mL) |
|-------------------|---------------|---------------------------------------------|---------------------------|---------------------------|
| Precipitant Mix 1 | P500MME_P20K  | 40% v/v PEG 500* MME; 20 % w/v PEG 20000    | MD2-100-81                | MD2-250-81                |
| Precipitant Mix 2 | EDO_P8K       | 40% v/v Ethylene glycol; 20 % w/v PEG 8000  | MD2-100-82                | MD2-250-82                |
| Precipitant Mix 3 | GOL_P4K       | 40% v/v Glycerol; 20% w/v PEG 4000          | MD2-100-83                | MD2-250-83                |
| Precipitant Mix 4 | MPD_P1K_P3350 | 25% v/v MPD; 25% PEG 1000; 25% w/v PEG 3350 | MD2-100-84                | MD2-250-84                |

\*N.B. The PEG 550 MME that was originally used in this screen has been discontinued and replaced with PEG 500 MME.

<sup>1</sup> The precipitants are made up as stock mixes and these are used to make up the final conditions at the concentrations show on the datasheet. For recipes of the above mixes please contact us.

**Morpheus®****Conditions 1-48 (Box 1)****MD1-46**

| Tube | Conc   | Ligands   | Conc  | Buffer          | pH  | Conc     | Precipitant       |
|------|--------|-----------|-------|-----------------|-----|----------|-------------------|
| 1-1  | 0.06 M | Divalents | 0.1 M | Buffer System 1 | 6.5 | 50 % v/v | Precipitant Mix 1 |
| 1-2  | 0.06 M | Divalents | 0.1 M | Buffer System 1 | 6.5 | 50 % v/v | Precipitant Mix 2 |
| 1-3  | 0.06 M | Divalents | 0.1 M | Buffer System 1 | 6.5 | 50 % v/v | Precipitant Mix 3 |
| 1-4  | 0.06 M | Divalents | 0.1 M | Buffer System 1 | 6.5 | 50 % v/v | Precipitant Mix 4 |
| 1-5  | 0.06 M | Divalents | 0.1 M | Buffer System 2 | 7.5 | 50 % v/v | Precipitant Mix 1 |
| 1-6  | 0.06 M | Divalents | 0.1 M | Buffer System 2 | 7.5 | 50 % v/v | Precipitant Mix 2 |
| 1-7  | 0.06 M | Divalents | 0.1 M | Buffer System 2 | 7.5 | 50 % v/v | Precipitant Mix 3 |
| 1-8  | 0.06 M | Divalents | 0.1 M | Buffer System 2 | 7.5 | 50 % v/v | Precipitant Mix 4 |
| 1-9  | 0.06 M | Divalents | 0.1 M | Buffer System 3 | 8.5 | 50 % v/v | Precipitant Mix 1 |
| 1-10 | 0.06 M | Divalents | 0.1 M | Buffer System 3 | 8.5 | 50 % v/v | Precipitant Mix 2 |
| 1-11 | 0.06 M | Divalents | 0.1 M | Buffer System 3 | 8.5 | 50 % v/v | Precipitant Mix 3 |
| 1-12 | 0.06 M | Divalents | 0.1 M | Buffer System 3 | 8.5 | 50 % v/v | Precipitant Mix 4 |
| 1-13 | 0.09 M | Halogens  | 0.1 M | Buffer System 1 | 6.5 | 50 % v/v | Precipitant Mix 1 |
| 1-14 | 0.09 M | Halogens  | 0.1 M | Buffer System 1 | 6.5 | 50 % v/v | Precipitant Mix 2 |
| 1-15 | 0.09 M | Halogens  | 0.1 M | Buffer System 1 | 6.5 | 50 % v/v | Precipitant Mix 3 |
| 1-16 | 0.09 M | Halogens  | 0.1 M | Buffer System 1 | 6.5 | 50 % v/v | Precipitant Mix 4 |
| 1-17 | 0.09 M | Halogens  | 0.1 M | Buffer System 2 | 7.5 | 50 % v/v | Precipitant Mix 1 |
| 1-18 | 0.09 M | Halogens  | 0.1 M | Buffer System 2 | 7.5 | 50 % v/v | Precipitant Mix 2 |
| 1-19 | 0.09 M | Halogens  | 0.1 M | Buffer System 2 | 7.5 | 50 % v/v | Precipitant Mix 3 |
| 1-20 | 0.09 M | Halogens  | 0.1 M | Buffer System 2 | 7.5 | 50 % v/v | Precipitant Mix 4 |
| 1-21 | 0.09 M | Halogens  | 0.1 M | Buffer System 3 | 8.5 | 50 % v/v | Precipitant Mix 1 |
| 1-22 | 0.09 M | Halogens  | 0.1 M | Buffer System 3 | 8.5 | 50 % v/v | Precipitant Mix 2 |
| 1-23 | 0.09 M | Halogens  | 0.1 M | Buffer System 3 | 8.5 | 50 % v/v | Precipitant Mix 3 |
| 1-24 | 0.09 M | Halogens  | 0.1 M | Buffer System 3 | 8.5 | 50 % v/v | Precipitant Mix 4 |
| 1-25 | 0.09 M | NPS       | 0.1 M | Buffer System 1 | 6.5 | 50 % v/v | Precipitant Mix 1 |
| 1-26 | 0.09 M | NPS       | 0.1 M | Buffer System 1 | 6.5 | 50 % v/v | Precipitant Mix 2 |
| 1-27 | 0.09 M | NPS       | 0.1 M | Buffer System 1 | 6.5 | 50 % v/v | Precipitant Mix 3 |
| 1-28 | 0.09 M | NPS       | 0.1 M | Buffer System 1 | 6.5 | 50 % v/v | Precipitant Mix 4 |
| 1-29 | 0.09 M | NPS       | 0.1 M | Buffer System 2 | 7.5 | 50 % v/v | Precipitant Mix 1 |
| 1-30 | 0.09 M | NPS       | 0.1 M | Buffer System 2 | 7.5 | 50 % v/v | Precipitant Mix 2 |
| 1-31 | 0.09 M | NPS       | 0.1 M | Buffer System 2 | 7.5 | 50 % v/v | Precipitant Mix 3 |
| 1-32 | 0.09 M | NPS       | 0.1 M | Buffer System 2 | 7.5 | 50 % v/v | Precipitant Mix 4 |
| 1-33 | 0.09 M | NPS       | 0.1 M | Buffer System 3 | 8.5 | 50 % v/v | Precipitant Mix 1 |
| 1-34 | 0.09 M | NPS       | 0.1 M | Buffer System 3 | 8.5 | 50 % v/v | Precipitant Mix 2 |
| 1-35 | 0.09 M | NPS       | 0.1 M | Buffer System 3 | 8.5 | 50 % v/v | Precipitant Mix 3 |
| 1-36 | 0.09 M | NPS       | 0.1 M | Buffer System 3 | 8.5 | 50 % v/v | Precipitant Mix 4 |
| 1-37 | 0.12 M | Alcohols  | 0.1 M | Buffer System 1 | 6.5 | 50 % v/v | Precipitant Mix 1 |
| 1-38 | 0.12 M | Alcohols  | 0.1 M | Buffer System 1 | 6.5 | 50 % v/v | Precipitant Mix 2 |
| 1-39 | 0.12 M | Alcohols  | 0.1 M | Buffer System 1 | 6.5 | 50 % v/v | Precipitant Mix 3 |
| 1-40 | 0.12 M | Alcohols  | 0.1 M | Buffer System 1 | 6.5 | 50 % v/v | Precipitant Mix 4 |
| 1-41 | 0.12 M | Alcohols  | 0.1 M | Buffer System 2 | 7.5 | 50 % v/v | Precipitant Mix 1 |
| 1-42 | 0.12 M | Alcohols  | 0.1 M | Buffer System 2 | 7.5 | 50 % v/v | Precipitant Mix 2 |
| 1-43 | 0.12 M | Alcohols  | 0.1 M | Buffer System 2 | 7.5 | 50 % v/v | Precipitant Mix 3 |
| 1-44 | 0.12 M | Alcohols  | 0.1 M | Buffer System 2 | 7.5 | 50 % v/v | Precipitant Mix 4 |
| 1-45 | 0.12 M | Alcohols  | 0.1 M | Buffer System 3 | 8.5 | 50 % v/v | Precipitant Mix 1 |
| 1-46 | 0.12 M | Alcohols  | 0.1 M | Buffer System 3 | 8.5 | 50 % v/v | Precipitant Mix 2 |
| 1-47 | 0.12 M | Alcohols  | 0.1 M | Buffer System 3 | 8.5 | 50 % v/v | Precipitant Mix 3 |
| 1-48 | 0.12 M | Alcohols  | 0.1 M | Buffer System 3 | 8.5 | 50 % v/v | Precipitant Mix 4 |

**Morpheus®****Conditions 1-48 (Box 2)****MD1-46**

| Tube | Conc   | Ligands          | Conc  | Buffer          | pH  | Conc     | Precipitant       |
|------|--------|------------------|-------|-----------------|-----|----------|-------------------|
| 2-1  | 0.12 M | Ethylene glycols | 0.1 M | Buffer System 1 | 6.5 | 50 % v/v | Precipitant Mix 1 |
| 2-2  | 0.12 M | Ethylene glycols | 0.1 M | Buffer System 1 | 6.5 | 50 % v/v | Precipitant Mix 2 |
| 2-3  | 0.12 M | Ethylene glycols | 0.1 M | Buffer System 1 | 6.5 | 50 % v/v | Precipitant Mix 3 |
| 2-4  | 0.12 M | Ethylene glycols | 0.1 M | Buffer System 1 | 6.5 | 50 % v/v | Precipitant Mix 4 |
| 2-5  | 0.12 M | Ethylene glycols | 0.1 M | Buffer System 2 | 7.5 | 50 % v/v | Precipitant Mix 1 |
| 2-6  | 0.12 M | Ethylene glycols | 0.1 M | Buffer System 2 | 7.5 | 50 % v/v | Precipitant Mix 2 |
| 2-7  | 0.12 M | Ethylene glycols | 0.1 M | Buffer System 2 | 7.5 | 50 % v/v | Precipitant Mix 3 |
| 2-8  | 0.12 M | Ethylene glycols | 0.1 M | Buffer System 2 | 7.5 | 50 % v/v | Precipitant Mix 4 |
| 2-9  | 0.12 M | Ethylene glycols | 0.1 M | Buffer System 3 | 8.5 | 50 % v/v | Precipitant Mix 1 |
| 2-10 | 0.12 M | Ethylene glycols | 0.1 M | Buffer System 3 | 8.5 | 50 % v/v | Precipitant Mix 2 |
| 2-11 | 0.12 M | Ethylene glycols | 0.1 M | Buffer System 3 | 8.5 | 50 % v/v | Precipitant Mix 3 |
| 2-12 | 0.12 M | Ethylene glycols | 0.1 M | Buffer System 3 | 8.5 | 50 % v/v | Precipitant Mix 4 |
| 2-13 | 0.12 M | Monosaccharides  | 0.1 M | Buffer System 1 | 6.5 | 50 % v/v | Precipitant Mix 1 |
| 2-14 | 0.12 M | Monosaccharides  | 0.1 M | Buffer System 1 | 6.5 | 50 % v/v | Precipitant Mix 2 |
| 2-15 | 0.12 M | Monosaccharides  | 0.1 M | Buffer System 1 | 6.5 | 50 % v/v | Precipitant Mix 3 |
| 2-16 | 0.12 M | Monosaccharides  | 0.1 M | Buffer System 1 | 6.5 | 50 % v/v | Precipitant Mix 4 |
| 2-17 | 0.12 M | Monosaccharides  | 0.1 M | Buffer System 2 | 7.5 | 50 % v/v | Precipitant Mix 1 |
| 2-18 | 0.12 M | Monosaccharides  | 0.1 M | Buffer System 2 | 7.5 | 50 % v/v | Precipitant Mix 2 |
| 2-19 | 0.12 M | Monosaccharides  | 0.1 M | Buffer System 2 | 7.5 | 50 % v/v | Precipitant Mix 3 |
| 2-20 | 0.12 M | Monosaccharides  | 0.1 M | Buffer System 2 | 7.5 | 50 % v/v | Precipitant Mix 4 |
| 2-21 | 0.12 M | Monosaccharides  | 0.1 M | Buffer System 3 | 8.5 | 50 % v/v | Precipitant Mix 1 |
| 2-22 | 0.12 M | Monosaccharides  | 0.1 M | Buffer System 3 | 8.5 | 50 % v/v | Precipitant Mix 2 |
| 2-23 | 0.12 M | Monosaccharides  | 0.1 M | Buffer System 3 | 8.5 | 50 % v/v | Precipitant Mix 3 |
| 2-24 | 0.12 M | Monosaccharides  | 0.1 M | Buffer System 3 | 8.5 | 50 % v/v | Precipitant Mix 4 |
| 2-25 | 0.1 M  | Carboxylic acids | 0.1 M | Buffer System 1 | 6.5 | 50 % v/v | Precipitant Mix 1 |
| 2-26 | 0.1 M  | Carboxylic acids | 0.1 M | Buffer System 1 | 6.5 | 50 % v/v | Precipitant Mix 2 |
| 2-27 | 0.1 M  | Carboxylic acids | 0.1 M | Buffer System 1 | 6.5 | 50 % v/v | Precipitant Mix 3 |
| 2-28 | 0.1 M  | Carboxylic acids | 0.1 M | Buffer System 1 | 6.5 | 50 % v/v | Precipitant Mix 4 |
| 2-29 | 0.1 M  | Carboxylic acids | 0.1 M | Buffer System 2 | 7.5 | 50 % v/v | Precipitant Mix 1 |
| 2-30 | 0.1 M  | Carboxylic acids | 0.1 M | Buffer System 2 | 7.5 | 50 % v/v | Precipitant Mix 2 |
| 2-31 | 0.1 M  | Carboxylic acids | 0.1 M | Buffer System 2 | 7.5 | 50 % v/v | Precipitant Mix 3 |
| 2-32 | 0.1 M  | Carboxylic acids | 0.1 M | Buffer System 2 | 7.5 | 50 % v/v | Precipitant Mix 4 |
| 2-33 | 0.1 M  | Carboxylic acids | 0.1 M | Buffer System 3 | 8.5 | 50 % v/v | Precipitant Mix 1 |
| 2-34 | 0.1 M  | Carboxylic acids | 0.1 M | Buffer System 3 | 8.5 | 50 % v/v | Precipitant Mix 2 |
| 2-35 | 0.1 M  | Carboxylic acids | 0.1 M | Buffer System 3 | 8.5 | 50 % v/v | Precipitant Mix 3 |
| 2-36 | 0.1 M  | Carboxylic acids | 0.1 M | Buffer System 3 | 8.5 | 50 % v/v | Precipitant Mix 4 |
| 2-37 | 0.1 M  | Amino acids      | 0.1 M | Buffer System 1 | 6.5 | 50 % v/v | Precipitant Mix 1 |
| 2-38 | 0.1 M  | Amino acids      | 0.1 M | Buffer System 1 | 6.5 | 50 % v/v | Precipitant Mix 2 |
| 2-39 | 0.1 M  | Amino acids      | 0.1 M | Buffer System 1 | 6.5 | 50 % v/v | Precipitant Mix 3 |
| 2-40 | 0.1 M  | Amino acids      | 0.1 M | Buffer System 1 | 6.5 | 50 % v/v | Precipitant Mix 4 |
| 2-41 | 0.1 M  | Amino acids      | 0.1 M | Buffer System 2 | 7.5 | 50 % v/v | Precipitant Mix 1 |
| 2-42 | 0.1 M  | Amino acids      | 0.1 M | Buffer System 2 | 7.5 | 50 % v/v | Precipitant Mix 2 |
| 2-43 | 0.1 M  | Amino acids      | 0.1 M | Buffer System 2 | 7.5 | 50 % v/v | Precipitant Mix 3 |
| 2-44 | 0.1 M  | Amino acids      | 0.1 M | Buffer System 2 | 7.5 | 50 % v/v | Precipitant Mix 4 |
| 2-45 | 0.1 M  | Amino acids      | 0.1 M | Buffer System 3 | 8.5 | 50 % v/v | Precipitant Mix 1 |
| 2-46 | 0.1 M  | Amino acids      | 0.1 M | Buffer System 3 | 8.5 | 50 % v/v | Precipitant Mix 2 |
| 2-47 | 0.1 M  | Amino acids      | 0.1 M | Buffer System 3 | 8.5 | 50 % v/v | Precipitant Mix 3 |
| 2-48 | 0.1 M  | Amino acids      | 0.1 M | Buffer System 3 | 8.5 | 50 % v/v | Precipitant Mix 4 |

## APPENDIX IV: I03 MX Beamline Specifications

|                                                   |                   |
|---------------------------------------------------|-------------------|
| Standard Energy (Å / keV)                         | 0.98 / 12.7       |
| Energy Range (keV)                                | 5.2 – 21.0        |
| Flux at 12.7 keV, focus at sample                 |                   |
| Beam Size Options:                                |                   |
| Focus at sample                                   | 90 x 20 µm        |
| Focus on detector                                 | 90 x 30 µm        |
| Defocussed                                        | 100 x 100 µm      |
| Focus at sample with CRLs (12.7keV)               | 55 x 10 µm        |
| Wavelength range (Å)                              | 0.6 – 2.3         |
| Working wavelength                                | 0.6 – 2.48        |
| Standard working wavelength (Å)                   | 0.976Å (12.7 keV) |
| Detector                                          | Eiger2 XE 16 M    |
| Time to collect 180° data set (seconds)           | 3.2*              |
| Sample changer                                    | BART              |
| No pins in dewar                                  | 592               |
| Sample Exchange Time (seconds)                    | 18                |
| Guide to samples screened / hour (X-ray centring) | 15                |

\*1800 frames, 0.1° oscillation at 560Hz

Information taken from <https://www.diamond.ac.uk/Instruments/Mx/I03.html>. Accessed 25/11/2019

## APPENDIX V: Best 15 matches against G4LZI3 model derived from Dali

### Server

| No | PDB ID/Chain | Z-score | rmsd | %id | Description                                     |
|----|--------------|---------|------|-----|-------------------------------------------------|
| 1  | 2gm3-A       | 14.3    | 2.1  | 24  | Unknown Protein                                 |
| 2  | 1mjh-B       | 13.7    | 1.5  | 30  | ATP-Binding Domain of MJ0577                    |
| 3  | 4wny-A       | 13.5    | 2.0  | 23  | Universal Stress Protein                        |
| 4  | 3s3t-A       | 13.0    | 2.0  | 23  | Nucleotide-Binding Protein, USP                 |
| 5  | 3idf-A       | 12.5    | 2.2  | 28  | Universal Stress Protein-like Protein           |
| 6  | 3dlo-D       | 12.4    | 2.7  | 21  | Universal Stress Protein                        |
| 7  | 4r3l-A       | 12.2    | 2.1  | 21  | Universal Stress Protein F                      |
| 8  | 3cis-G       | 12.0    | 2.5  | 21  | Uncharacterized Protein                         |
| 9  | 3hgm-A       | 11.8    | 2.1  | 25  | Universal Stress Protein TEAD                   |
| 10 | 3tnj-A       | 11.8    | 2.2  | 22  | Universal Stress Protein (USP)                  |
| 11 | 5ahw-C       | 11.7    | 2.4  | 20  | Universal Stress Protein                        |
| 12 | 2z3v-A       | 11.3    | 2.0  | 25  | Universal Stress Protein Family                 |
| 13 | 2dum-C       | 11.3    | 2.3  | 29  | Hypothetical Protein PH0823                     |
| 14 | 3ab8-A       | 11.1    | 2.0  | 21  | Putative Uncharacterized Protein TTHA0350       |
| 15 | 3fg9-A       | 11.0    | 2.6  | 19  | Protein of Universal Stress Protein USPA Family |

## APPENDIX IV: General Chemicals and Reagents

| Chemical                                             | Company              |
|------------------------------------------------------|----------------------|
| 40% 37.5:1 Acrylamide: bis-acrylamide                | Lasec                |
| Acetic Acid                                          | Capital Lab Supplies |
| Ammonium Persulphate (APS)                           | Inqaba Biotechnical  |
| Ampicillin                                           | Capital Lab Supplies |
| Bromophenol Blue                                     | Inqaba Biotechnical  |
| Calcium Chloride dehydrate                           | Sigma-Aldrich        |
| Coomasie Blue R-250                                  | Inqaba Biotechnical  |
| Dimethyl sulfoxide (DMSO)                            | Capital Lab Supplies |
| Dithiothreitol (DTT)                                 | Inqaba Biotechnical  |
| Glucose                                              | BioRad/Lasec         |
| Glycerol                                             | Capital Lab Supplies |
| Glycine                                              | Inqaba Biotechnical  |
| Hydrochloric Acid                                    | Capital Lab Supplies |
| Imidazole                                            | Inqaba Biotechnical  |
| Isopropyl $\beta$ - d-1-thiogalactopyranoside (IPTG) | Capital Lab Supplies |
| Kanamycin                                            | Inqaba Biotechnical  |
| Luria Bertanit (LB) Broth                            | Biorad/Lasec         |
| Lysozyme                                             | Inqaba Biotechnical  |
| Magnesium Chloride                                   | BioRad/Lasec         |

|                                             |                      |
|---------------------------------------------|----------------------|
| Magnesium Sulphate                          | Sigma-Aldrich        |
| Manganese Chloride tetrahydrate             | Sigma-Aldrich        |
| MOPS                                        | Sigma-Aldrich        |
| Nutrient Agar                               | Biorad/Lasec         |
| Peptone                                     | Sigma-Aldrich        |
| Phenylmethylsulfonyl fluoride (PMSF)        | Capital Lab Supplies |
| Polyethylene glycol (PEG 3350)              | Sigma-Aldrich        |
| Potassium Acetate                           | Sigma-Aldrich        |
| Precision Plus Protein Dual Color Standards | BioRad/Lasec         |
| Rubidium Chloride                           | Sigma-Aldrich        |
| Sodium Azide                                | Sigma-Aldrich        |
| Sodium Chloride                             | BioRad/Lasec         |
| Sodium Dodecyl Sulfate (SDS)                | BioRad/Lasec         |
| Tris                                        | BioRad/Lasec         |
| Yeast                                       | BioRad/Lasec         |