UNIVERSITY

OF ZULULAND

Biophysical Investigations of the Interaction Between Substrate Binding Domain of Hsp40 and the RING finger domain of Retinoblastoma Binding Protein 6

By

SANELE MAUD NONDUMISO MDUNGE

(200964011)

BSc (Biological Sciences); BSc Hons (Biochemistry)

Dissertation submitted in partial fulfilment of the requirement for the degree **MASTER OF SCIENCE (MSc) IN BIOCHEMISTRY**

Department of Biochemistry and Microbiology, Faculty of Science and Agriculture, University of Zululand, KwaDlangezwa, KwaZulu Natal

Supervisor: Prof Abidemi Paul Kappo

Co-Supervisor: Mr Mathula Lancelot Ngwenya

December 2021

ABSTRACT

Irrespective of the tireless efforts and resources that have been pumped into cancer elimination and management, the disease remains one of the foremost causes of death worldwide, with associated mortality and morbidity cases on the rise every year. The disease has therefore become a major public health concern as both the young and aged are greatly affected. For years, efforts have been directed towards finding permanent solutions for cancer, not only because of the risk of reoccurrence but also due to the non-specificity and side effects of current chemotherapy treatments. Protein-protein interactions have recently come onto the scene as potential targets for novel treatment and management options for the disease. Through a yeast-two hybrid study, the RING finger domain of the Retinoblastoma binding protein 6 (RBBP6) and the substrate binding domain of heat shock protein 40 (SBDHsp40) suggested a putative interaction between these two proteins. In this study molecular dynamic simulations were used to confirm the interaction between the proteins and associated binding parameters. This was followed by the recombinant expression of the proteins within two difference expression vectors and *E. coli* bacterial strains. The SBDHsp40 protein was purified using affinity chromatography while the RING finger domain protein was purified to homogeneity by affinity and size exclusion chromatography. Biophysical characterization in the form of Fourier Transform–infrared spectroscopy and Raman spectroscopy were then used to determine the functional groups within the proteins that made up their secondary structural elements. Due to time constraints, the experimental interactive studies could not be performed. However, the *in-silico* studies done confirm this putative interaction and form a basis for consequent application in the discovery of new anti-cancer drugs.

Keywords: Cancer, heat shock protein 40 (Hsp40), protein-protein interaction (PPI), retinoblastoma-binding protein-6 (RBBP6), RING finger domain, substrate binding domain, ubiquitination

DECLARATION

I, Sanele Mdunge (Student No: 200964011), solemnly declare that the "***Biophysical Investigations of the Interaction between Substrate Binding Domain of Heat Shock Protein 40 and the RING finger domain of Retinoblastoma Binding Protein 6***" is my own, unaided work. It is being submitted for an MSc Masters Degree in Biochemistry at the University of Zululand and it has not been submitted before for any degree or examination in any other University.

Masters Candidate: Sanele Maud Nondumiso Mdunge

Signature:……..

Date:……..

Supervisor: Prof Abidemi Paul Kappo

Signature:…….

Date:…….

Co-Supervisor: Mr Mathula Lancelot Ngwenya

Signature:…….

Date:…….

DECLARATION ON PLAGIARISM

I acknowledge that I have read and understood the University's policies and rules applicable to postgraduate research, and I certify that I have, to the best of my knowledge and belief, complied with their requirements.

I declare that the content of this dissertation, save for the supervisory guidance received, is the product of my own work and effort. I have, to the best of my knowledge and belief, acknowledged all sources of information in line with normal academic conventions.

I further certify that the content of this dissertation is original, and that the material to be submitted for examination has not been submitted, either in whole or on part, for a degree at this or any other university.

I have subjected this document to the University's text-matching and/or similarity-checking procedures and I consider it to be free of any form of plagiarism.

Signature: ……...

Date:……...

DEDICATION

This dissertation is dedicated to my late parents

Adv. CM & Mrs BS Mdunge

May your souls continue to rest in eternal peace

ACKNOWLEDGEMENTS

To God be the Glory, great things he has done…

If it was not for His grace, I surely would not have made it till this far. When I look back to where it all started – I can never thank you enough dear Lord. To my family, my brothers: thank you for the love, support and not forgetting the pressure and criticism whenever I started slacking. Thank you for being there for me even with your wallets whenever I fell short. My niece: Nosipho, or rather my right-hand MAN…yes MAN! Thank you for being my keeper and making sure you drag me to bed, fed and looked after me even during the most hectic of days, thank you for staying up during my late nights to make sure I was well tucked in.

My late parents, as sad as I am that I get to acknowledge you in your absence, I am eternally grateful for how you have brought me up, molded me to becoming the woman I am and for always being my number one cheerleader… I am very confident that I have the best guardian angel squad in you guys, if only heaven had a phone.

Dr Priscila Masamba, words fail me in expressing how thankful I am for literally EVERYTHING you have done, since the very first day I walked into the Department of Biochemistry and Microbiology – this includes giving up your bed for me to crash during my Joburg visits. Thank you for your assistance with my work and for going over and beyond in ensuring that all was above board. We have not known each other for long, but watching you transition from the tiny girl I met a couple of years ago to this beautiful, intelligent young woman (with a heart of gold) has been a marvel to watch. Dynamites do come in small packages; the world is not ready for you. God bless your heart.

My partner in crime Siphosethu Buthelezi, thank you for being my punching bag, ear when I needed to vent, shoulder when I needed to cry and for always being my greatest supporter. Your presence in my life brings out the best of me, even when the going gets tough. In you I found a best friend. To Zinhle Ntuli, thank you so much Sisi for being there throughout - on a personal and professional level. You may not know what an impact your support had on lessening the load, but it surely does not go unnoticed.

To the rest of my family, friends and BSB Lab: thank you so much for assisting whenever/wherever you could. Good to know there are always people who are a phone call away whenever you require assistance.

Last but not least… Professor Abidemi Paul Kappo, HA!!!.. where do I even start! Do I mention your tea and rusks that I used to finish when working in the lab on weekends? Do I mention you forking out money to satisfy Chicken Licken cravings? Thank you for being such a dedicated and hardworking supervisor, I am yet to meet another supervisor who is this involved and driven in ensuring all student needs are met. Thank you to your wife Dr Christiana Kappo-Abidemi and family for allowing me to use and abuse you throughout all my academic frustrations. Thank you for being a mentor, father and friend and for all your 'serious talks' when I was quarter to burn out and not taking care of myself. Thank you for preaching the gospel of working hard and for playing such a tremendous role in the development of my career. You're a blessing to all those around you, people like yourself are not usually recognized and acknowledged but I would like to take this opportunity once again to let you know how much all your efforts are appreciated. May the Lord richly bless you.

LIST OF ABBREVIATIONS

ATP	Adenosine triphosphate
CD	Circular Dichroism
DWNN	Domain With No Name
HB	Hydrogen bond
HECT	Homologous to E6-associated protein carboxyl terminus
Hsp40	Heat Shock Protein 40
IMAC	Immobilized metal-affinity chromatography
RBBP6	Retinoblastoma Binding Protein 6
RFD	RING finger Domain
RING	**R**eally **I**nteresting **N**ew **G**ene
ITC	Isothermal Titration Calorimetry
RMSD	Root Mean Square Deviation
RMSF	Root Mean Square Fluctuation
RoG	Radius of Gyration
SASA	Solvent Accessible Surface Area
SBD	Substrate Binding Domain
SPR	Surface Plasmon Resonance
WHO	World Health Organization

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE: INTRODUCTION

1.1Introduction

In the year 1971, the then US President Richard Nixon foretold victory over the "war on cancer" in the next 5 years. Fast forward to 5 decades later, this goal is unfortunately far from the word attainable. In fact, cancer has now become the first and second leading cause of death in both economically developed and developing countries respectively (Kaiser, 2021). According to a Global Burden of Disease (GBD) study, an estimated 1 335 100 new cases and 397 583 deaths were attributed to cancer in the year 2021 amongst adolescents and young adults, while 2.3 million new cases were reported in adults older than the age of 80 in 2013, making 13% of all cases worldwide (Pilleron *et al.*, 2021; You *et al.*, 2021). Adding more salt to the injury, the latter is expected to triple from 14.3 million in 2019 to 426 million by the year 2050. Most recently, GLOBACAN 2020 estimated 19.3 and 10 million new cancer cases and deaths respectively in 2020 globally. The study also showed that the cumulative risk of developing the disease in African women was comparable to that experienced by North American women and those from European high income countries (Sung *et al.*, 2021). The growing incidence of cancer in Africa is mainly driven by poor and inadequate resources and facilities for the treatment, prevention, and diagnosis of the disease, which have in turn contributed to the high number of morbidity and mortality cases in the region (Ferlay *et al.*, 2015). Female breast cancer (2.26 million cases), lung (2.21) and prostate (1.41) make up the most frequently diagnosed cancers globally, while lung (1.79 million deaths), liver (830000) and stomach cancers (769000) are responsible for the most common cancer-causing deaths (Ferlay *et al.*, 2021).

Cancer has been described as the uncontrolled growth of abnormal cells that may occur anywhere in the body (Anand *et al.*, 2008). This may be caused by both genetic and epigenetic factors within somatic cells forming a subset of neoplasm or tumour, which in basic terms is a lump or mass of cells. The causes of cancer ranges from tobacco consumption, which is responsible for 22% of all cancer-related deaths, alcohol ingestion, poor diet, inadequate physical exercise, obesity, exposure to occupational and environmental carcinogens or pollutants as well as ionizing radiation, which all account for 10% of all deaths (Vineis and Wild, 2014). At least 15% of cancers are attributed to infections caused by biological pathogens such as *Helicobacter pylori*, Schistosoma spp, human papillomavirus (HPV), hepatitis B and C, as well as HIV. Other factors that account for 5 – 10% of cancers caused are inherited genetic factors, which is the most common. Under normal circumstances, cells can detect and repair DNA damage. If a cell undergoes severe damage and is unable to repair itself, it under-goes a program called cell death or apoptosis (Roberts *et al.*, 2002). Cancer occurs when damaged cells grow, divide, and spread abnormally in lieu of self-destructing as they most commonly should. Using biochemical and physical mechanisms, this uncontrolled proliferation disrupts both the structure and function of tissues surrounding the tumor (Nia *et al.*, 2020). Normal body tissues are infiltrated and destroyed and the physical abnormalities that result affect both the cancer cells and their microenvironment, causing tumorigenesis and resistance to treatment. With all the studies invested in understanding cancer biology and major efforts towards the elimination of the disease since the 1970s, current treatment options have sadly fallen short of fulfilling all expectations. Till date, developing anti-cancer treatment remains an important health problem, as more astute strategies to combat the disease are being sought and has encouraged the pharmaceutical industry to explore protein-protein interactions (PPIs) as an alternative drug target class.

Many functions in a cell are carried out through the specific interactions of individual proteins with each other. Proteins in a cell therefore form complex interaction networks known as the interactome (Cusick *et al.*, 2005). Thus, knowing which other protein interacts with it and where it fits in the interactome allows the prediction of the function of that protein. Over the last 10 years, protein-protein interactions have received immense recognition for their potential as drug targets. Within the human interactome, there are over 650 000 PPIs with only 2% having been explored as drugs by the year 2011 (Sawyer, 2020). Although they are complex due to their unconventional physicochemical features which have made them challenging or almost "undruggable targets", they are the basis of several important cellular processes, many of which are disease relevant. Regrettably, only a small number of PPIs have been approved or have been integrated into clinical trials, largely because they are difficult to study, and their mechanisms of action are largely elusive (Sawyer, 2020). Nevertheless, they hold immense attraction as molecular targets in modern day drug discovery that could easily see them being used to address specific disease pathways.

Accumulated evidence has shown how heat shock proteins (HSPs) and the Retinoblastoma Binding Protein 6 (RBBP6) play a central role in buffering the cancer proteome against stress and also maintaining key oncoproteins through its protein folding and degrading activity (Esser *et al.*, 2004). Moreover, these proteins have been found to be essential in maintaining the activity of key oncoproteins that endow cancer cells with the qualities required for tumorigenesis, for example limitless cell division, tissue invasion and metastasis, avoidance of apoptosis, resistance to chemotherapy and immune destruction (Xu *et al.*, 2002). This standpoint has thus become a prominent target in cancer drug discovery. Based on the implications and their central role in the maintenance of oncogenes and resistance to cell stress required for tumorigenicity, more evidence can be obtained that the RING finger domain of RBBP6 interacts with the substrate binding

domain of Hsp40. Exploring and understanding the relationship and function of these proteins may potentially overcome the bottleneck of producing new inexpensive, accessible, and less invasive anti-cancer treatment. In this study, the purpose is to biophysically investigate their relationship by exploring their interaction.

1.2 Problem Statement

A significant number of deaths caused by cancer have generated a critical need for the development of new intervention strategies to populate the anticancer drug pipeline. A crippling impact of the disease on the fragile socio-economic and public health structure of some of the world's poorest nations has alarmed researchers to move towards overcoming the salient obstacles to effective cancer therapy such as less invasive and more affordable treatment. Current strategies to treat cancer are centred around invasive therapy such as chemotherapy, radiation, hormonal therapy, and immunotherapy, which has shown severe side effects on individuals and lack of affordability to a huge proportion of the poor, especially people living in sub-Saharan Africa. Several biochemical reactions in cells are mediated by protein-protein interactions, thus the decrease in numbers of new chemical entities from these interactions injected into the drug market has encouraged the pharmaceutical industry to focus and explore these interactions as an alternative target class in new cancer drug discovery, design, and development.

1.3Aim and Objectives of the Study

The aim of this study is to biophysically investigate the putative interaction between the substrate binding domain between Hsp40 (SBDHsp40) and the RING finger domain of Retinoblastoma binding protein 6 (RBBP6). This aim was achieved through the following objectives:

- Homology modelling of the RING finger domain and substrate binding domain of Hsp40 using MODELLER,
- Validation of predicted RING and SBDHsp40 models by PROCHECK tools,
- Molecular docking of RING finger domain protein onto SBDHsp40 using AutoDock tools,
- Investigation of dynamic evolution of the RING protein and RING~SBDHsp40 complex over time using molecular dynamic simulations,
- Determination of the stability and conformational flexibility of the complexes through post dynamic analyses
- Heterologous recombinant expression of RING finger domain and SBDHsp40 proteins in *E. coli* bacterial strains,
- Purification of the RING finger domain and SBDHsp40 proteins to homogeneity using affinity chromatography and gel filtration techniques
- Identification of protein functional groups using Fourier Transform Infrared (FTIR) Spectroscopy,
- Determination of protein molecular interactions using Raman Spectroscopy.

CHAPTER TWO: LITERATURE REVIEW

2.1. The Incidence of Cancer

In the shortest of definitions, cancer is a group of diseases caused by an uncontrollable division of abnormal cells with the potential of spreading to different parts of the body (Anand *et al.*, 2008). It is one of the most widespread and diagnosed diseases all over the world. To date, over 200 types of cancer have been reported and the signs and symptoms vary depending on the specific type and grade. It has been referred to as the leading cause of death in our modern society, responsible for severe causes of morbidity and mortality world-wide with an estimated 14 million new cases in 2012 and the second leading cause of death globally, accounting for about 8.8 million deaths in 2015 (Ferlay *et al.*, 2015; Parkin *et al.,* 2005). According to the Cancer Association of South Africa (CANSA), recent statistics show that 100 000 South Africans are diagnosed with cancer on an annual basis, with a survival rate of six out of every ten individuals affected, which calls for serious interventions in the country. Although this number may even be understated as cancer cases in South Africa are poorly documented, reports have shown that prostrate and breast cancer are most common among men and women respectively and are also the leading causes of death within the two sexes respectively (Singh *et al.*, 2015).

Cancer occurs due to a series of mutations within genes such that these mutations interrupt the functioning of cells. An increase in predisposing factors that include smoking, unhealthy lifestyle and poor dietary choices that lead to obesity, as well as alcohol and lack of exercise have been attributed to the causes of the disease. Special emphasis on smoking and environmental chemical substances play vital roles in gene mutation, genetic disorders and cancer cells by possessing carcinogenic chemical compounds that may influence cell cytoplasm or the nucleus either directly or indirectly (Hassanpour and Dehghani, 2017). The disruption of cellular relations leads to the

dysfunction of vital genes that in turn affect the cell cycle and result in abnormal proliferation. Cell division and growth under normal circumstances are regulated by proto-oncogenes but these become oncogenes during genetic mutation, which is harmful for cell existence. Despite the widespread and currently the most common methods pertaining to cancer management, which include surgery, laser therapy, endocrine therapy, radiation, cytotoxic chemotherapy, targeted therapy, immunotherapy or combination therapy, the incidence of cancer has still not slowed down but has actually increased (Mansoori *et al.*, 2017; Wang *et al.*, 2019). A significant reason for cancer relapses, which is one of the major causes of cancer deaths, has been attributed to the occurrence of resistance, which may be generated either before (intrinsic) or after treatment (acquired) to classical chemotherapy and/or novel targeted drugs. This coupled with the physical, emotional, psychological, and financial burden associated with the disease highlights the ongoing need to mitigate the effects of cancer.

The function of a particular protein is highly dependent on its three-dimensional shape. Abnormalities in proteins often occur within cells and usually the accumulation of these abnormal proteins can be a distinctive feature of different pathological states. The incorrect folding of proteins into their functional shape may result in exposure of the hydrophilic core, causing the formation of aggregates that play a role in several diseases or disorders that include cystic fibrosis, Alzheimer's, or Parkinson's disease as well as cancer. Several years have been dedicated to the study of cancer and its hallmarks with one of them being the structural and functional proteins that play a biological role for example in the formation of tumors. Heat shock proteins and RBBP6 are two completely different proteins that have been hypothesized to interact through their domains to play a role in the workings of cancer. This chapter lays a background and foundation for the current

study and how their possible interaction may potentially result in novel and cheaper forms of treatment for the disease.

2.2 Protein-Protein Interactions

Proteins are versatile biomolecules that are literally referred to as the building blocks of life. Uncountable biological functions within the human body are heavily reliant on protein activity, which in turn carry out their functions in combination with other proteins, or biomolecules such as DNA or RNA, as well as inorganic groups or organic co-factors that include Zn^{2+}, Mg^{2+} or O_2 (Ali and Bagdi, 2015). These are known as protein-protein interactions (PPIs), and they play a vital role in biological functions such as immunological function, enzyme-substrate interactions, and hormone-receptor binding. Therefore, understanding their structural basis is relatively important. Within the human interactome, there are at least 40 000–200 000 PPIs that have been predicted (Garner and Janda, 2011). The interactome plays a vital role in pathological and physiological processes that include signal transduction, apoptosis, cell growth, differentiation, and proliferation (Lu *et al.*, 2020). PPI networks lay a foundation for understanding the workings of many cellular and biological processes; thus, they provide vital insight on the molecular mechanisms of diseases such as cancer, neurodegenerative and infectious diseases (Lu *et al.*, 2020; Kar *et al.*, 2009). Since these binding occurrences form a significant part of several important cellular functions and their poor regulation causes disease, the modulation of these interactions has become a popular potential yet an extremely tasking area of therapeutics.

Drug development strategies or PPI modulation attempts mostly focus on the discovery of small molecule inhibitors that can bind to PPI interfaces, but this is generally considered a very difficult concept, with this are of therapeutics considered almost ‘undruggable’. This is because the PPI

area interface where the interactors meet is relatively large and highly hydrophobic and is also a flat surface that contains pockets and grooves that make the design of small molecules difficult (Lu *et al.,* 2020). In addition, inhibition of the amino acid residues involved in the interaction is also difficult because of their high affinity and lastly, PPIs lack reference small molecule ligands in comparison to traditional drug target receptors or enzymes. However, rapid development in the field of structural biology and available high-throughput screening techniques has made this challenge less tasking. Moreover, some drugs have already been placed on the market while other drug-like candidates are currently being developed and explored in clinical trials such as Maraviroc for the treatment of HIV infection, Navarixinin (Phase II) chronic obstructive pulmonary disease, BCL6/SMRT PPI in B cell lymphoma, bortezomib, Cullin RING E3 ligases and MLN4924 in the ubiquitination pathway and navitoclax, obatoclax and ABT-199 (Venetoclax) for apoptosis (Mabonga and Kappo, 2019). This shows that the PPI modulator field is an exciting one that bears great prospective in the development of drugs that possess great specificity and less side effects.

Over the past few years, various techniques for identifying different sets of interacting proteins have emerged such as mass spectrometry, affinity purification and yeast two-hybrid (Y2H) systems (Ali and Bagdi, 2015). The latter has, however, established itself as a well-known and powerful routine tool for investigating PPIs. The assay is a genetic screen that detects PPIs when reconstituted transcription factors respond to activated reporter genes (Mehla *et al.,* 2015). cDNA expression libraries are screened to isolate interacting proteins, which in turn reveal the cloned cDNAs encoding the proteins, thus simplifying downstream processes. Due to its high sensitivity, weak interactions can be picked up and mutational analyses are also simplified, thus allowing for the identification of residues or motifs involved in the binding (Causier and Davis, 2002). In as

much as the system offers several advantages over other biochemical techniques that include cost effectiveness, sensitivity, and convenience, it also has its own challenges, the largest which would probably be the generation of false positives. This arises from two-hybrid library screens where transcription of the reporter gene is activated in the absence of any bait-prey interaction, which occurs in at least 10-20% of all baits (Mehla *et al.,* 2015). Taken together, this project seeks to validate a study by Kappo and co-workers who in 2012 performed a Y2H and showed possible interaction between the substrate binding domain of Hsp40 and the RING finger domain of RBBP6 (Kappo *et al.*, 2012).

2.3 Heat Shock Proteins

Heat shock proteins (HSPs) were first discovered in overheated larvae from *Drosophila melanogaster* by Ritossa and colleagues in 1962, first isolated by Lindquist and co-workers and first published by Pelham in 1986 (Kaul and Thippeswamy, 2011; Van Drie, 2011). With subsequent research, it was then clear that HSPs are widely distributed throughout the three domains of life i.e., eukaryotes, archaea and bacteria and demonstrate a wide range of functions. Although termed heat shock proteins, their functions are not solely induced by heat shock. Elevated levels of HSP expression are triggered by various forms of cellular insults that include but are not limited to of course heat stress, chemical exposure such as alcohol or heavy metal treatment, magnetic fields, ischemia, shearing and tearing, nutrient deficiencies, hypoxia, compression, oxidative stress, pH shift and ultraviolet irradiation (Kaul and Thippeswamy, 2011; Tóth *et al.*, 2015). These in addition include biological stress such as cold, inflammation, fever, and infection. Consequently, their presence has been implicated in various diseases such as cancer, neurodegenerative disorders and cardiac diseases (Seigneuric *et al.*, 2011).

The HSP family is one that is quite large and evolutionarily conserved that contains various proteins classified under three main groups according to their molecular weight. Group I constitute the most notable which are Hsp60, Hsp70, Hsp90 and Hsp110, while Group II is made up of the "minor HSPs" such as glucose-regulated proteins (GRPs) and these are mostly activated by conditions caused by inadequate glucose. Lastly Group III HSPs are the low molecular weight HSPs containing about 10 members (HSPB1–B10) of 12 to 30 kDa proteins (Patni *et al.*, 2021). This ubiquitous class of proteins are intracellular molecules that are usually referred to as molecular chaperones due to their ability to assist in the covalent unfolding and folding of proteins, as well as the disassembly and assembly of macromolecules (Silva and Madusanka, 2019). They present themselves as the first line of defence against cell stress and therefore function in various protective and physiological processes to sustain cellular homoeostasis (Yun *et al.*, 2020).

2.4 Role of HSPs in Cancer

Cancer cells are subjected to several abnormal cell stresses and tumorigenicity requires active oncoproteins. Abnormalities in proteins often occur within cells and usually the accumulation in these abnormal proteins can be a distinctive feature of different kinds of pathological state. HSPs have recently been the focus of both intense and extensive study in various types of cancer because they are actively involved in metastases, invasion, cell proliferation, differentiation, and death (Lianos *et al.*, 2015). In addition, they are highly expressed in several tumours. For example, Hsp27 regulates transcriptional activity, stability of androgen receptors and nuclear shuttling, and plays a role in modulating the β-catenin/Slug signalling pathway for mediating epidermal growth from epithelial to mesenchymal transition. The interaction of WASF3 with Hsp70 and Hsp90 is implicated in the invasion and metastasis of prostate cancer, while the involvement of Hsp40 in

cancer has been suggested due to its ability to inhibit apoptosis via the Hsp70 dependent protein kinase (PKR) and Fanconi anaemia group C protein complex (Lianos *et al.*, 2015).

Hsp40 (DnaJ in *E. coli*) is a molecular chaperone which was first discovered in the early 1960s, with its main function being the regulation of ATPase activity by binding and transferring substrate proteins to Hsp70, required for the stability and the function of numerous conditionally activated and expressed signalling proteins (Neckers, 2007). The ATPase activity of heat shock proteins is important for controlling the chaperone cycle and directing binding, induction of the active conformation and the release of its client proteins. Hsp40 binds to unfolded or non-native polypeptides through its C-terminal substrate binding domain. Through its highly conserved ATPase domain, the Hsp40-bound unfolded substrate then interacts with Hsp70 via the N-terminal ATPase domain, which regulates and stimulates the ATPase activity of Hsp70 thus allowing the stable Hsp70/protein complex and passing down of substrate to Hsp70. Upon ATP hydrolysis, Hsp70 adopts a closed conformation and becomes a mature complex that is essential for it to carry out its function of client protein folding and stabilization (Pearl *et al.*, 2008). Hsp40 target Hsp70 activity to clients at different locations in cells and others bind client proteins directly, thereby delivering specific clients to Hsp70 and directly determining their fate (Figure 2.1: Kampinga *et al.*, 2010). Table 1 depicts the role of Hsp40 either as a tumour suppressor or as an oncogene.

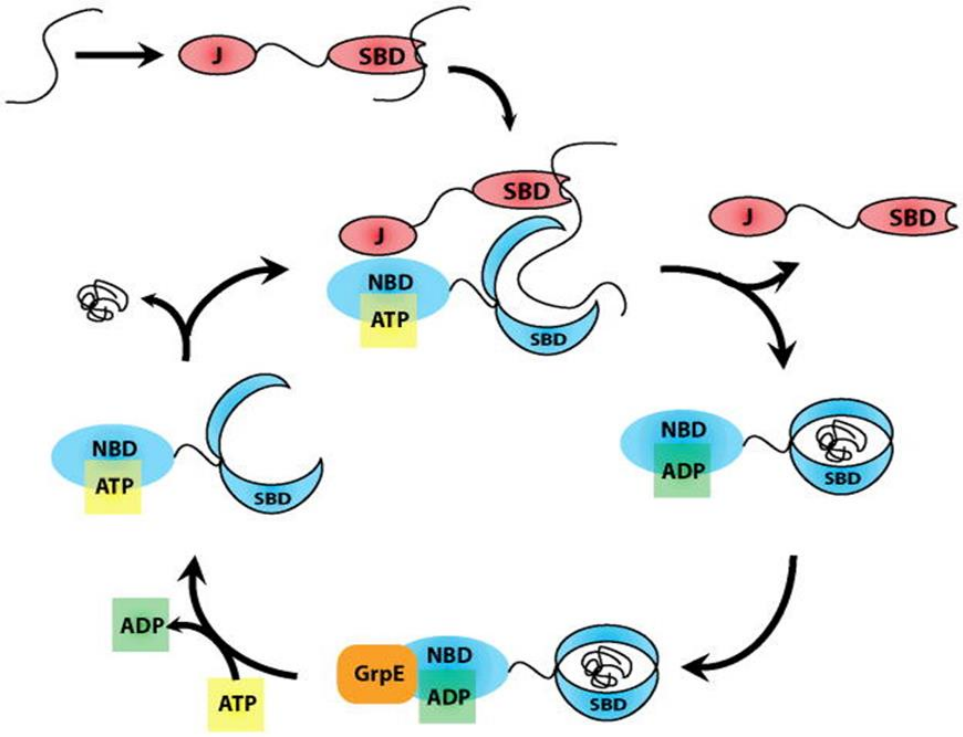


Figure 2.1: The Catalytic Cycle of Hsp70/Hsp40 upon substrate recognition. Upon binding of client protein to the substrate binding domain (SBD) of DnaJ, the J domain recognizes and binds to the ATP bound N terminal ATPase domain (NBD) of DnaK (through its HPD motif) and delivers the client protein. DnaJ stimulates the hydrolysis of ATP to ADP and then gets released. The exchange of ATP and ADP leads to conformational change, where the SBD of DnaK adopts a closed conformation through its groove and lid. The release of ADP is triggered by Nuclear Elongation Factors (eg. GrpE), which allows the binding of ATP thus completing the cycle by releasing the client protein (Srinivasan *et al.*, 2012).

Table 2.1: The implications of Hsp40 in various types of cancers. This information is extracted from Mitra *et al.*, 2009.

Name	HUGO nomenclature	Cellular location	No. of isoforms	Chromosomal location	Cancer type	Functionally tested	Role
HDJ2	DNAJA1	N/Cyt	2	9p13-p12	Glioblastoma	Tested in vitro in glioblastoma patient study in metastatic pancreatic adenocarcinoma	Radio-resistance, two splice variants farnesylated
Htid1	DNAJA3	Cyt/Mito	2	16p13.3	Breast	Tested in animals using MDA-MB-231 cells	Tumor suppressor, two splice variants negative regulator of breast cancer metastasis by increased NF-Kβ activity on IL8 promoter
HLJ1	DNAJB4			Ip31.1	Lung	In Vitro human lung adenocarcinoma cells; NSCLC patient study	Tumor suppressor in NSCLC; Suppressed invasion by upregulating E cadherin in lung cancer cells
MRJ	DNAJB6	N/Cyt	2	11q24.3	Breast	Tested in animals using MDA-MB-231 and MDAMB435 cells	Transcriptional expression via NFAT, reduces tumorigenicity and metastasis
MDGI	DNAJB9	ER		7q31		Tested in primary hamster fibroblasts	Present in cells with low metastatic potential
HDJC9	DNAJC9	N		10q22.2	Lung, Breast	Tested in human lung carcinoma A549 and MCF7 cells	Role not investigated
JDP1	DNAJC12		2	10q22.1	Breast	Studied in human breast tumor samples and MCF7	Found in ER+ breast cancer, two isoforms

MCJ	DNAJC15	Cyt		13q14.1	Ovary, brain	Studied in childhood brain tumor of humans; human ovarian tumors	Transmembrane CD chaperone, methylation controlled; expressed in drug sensitive breast cancer cells; MW 16-17 kDa
PPID	CYP40	Cyt		4q31.3	Breast	Studied in breast cancer patients	Allelic loss in advanced breast cancer modulation of apoptosis and mitosis in peroxisome proliferator-dependent heptacellular neoplasms. Associates with HSP90

2.6 Retinoblastoma binding protein 6 (RBBP6)

Retinoblastoma binding protein 6 (RBBP6) is a multi-faceted protein that has been identified as an E3 ubiquitin ligase due to its implication in promoting protein degradation via ubiquitination through the presence of a RING finger domain. It consists of an N-terminal ubiquitin-like domain and a cysteine-rich RING finger like domain (Pugh *et al.*, 2006)**.** It is characterised by a p53 domain – responsible for binding tumour suppressors by supressing binding of p53 to DNA, Rb domain – which interacts with retinoblastoma (a tumour suppressor) and the RING finger domain – responsible for the ubiquitination of proteins destined for degradation. Other domains found in the protein include the domain with no name (DWNN), zinc knuckle, the proline-rich domain, the serine-rich (SR) domain, and the nuclear localization signal (Figure 2.2: Dlamini *et al.*, 2019; Pugh *et al.*, 2006; Simons *et al.*, 1997).

The domain with no name (DWNN) is a 118–150 residue protein located at the N-terminus of RBBP6 that is highly conserved and exhibits a ubiquitin-like fold (Pugh *et al.*, 2006). The protein shares 18% similarity with ubiquitin and includes the presence of an anti-parallel β-sheet stationed in front of the α-helix, which is absent in ubiquitin. It has been suggested that the DWNN functions in cell cycle regulation, protein turnover and camptothecin-induced apoptosis, while also possibly playing a role in splice regulation and ubiquitin modification (Mbita *et al.*, 2011; Pugh *et al.*, 2006). The zinc knuckle domain, also known as the zinc finger domain, is suggested to possibly serve in protein, RNA, and DNA binding. It does this by binding histidine and cysteine to zinc ions as it comprises the C2HC (Cys-Cys-His-Cys) motif that contains a β-hairpin and a one-turn α-helix (Krishna *et al.*, 2003).

.

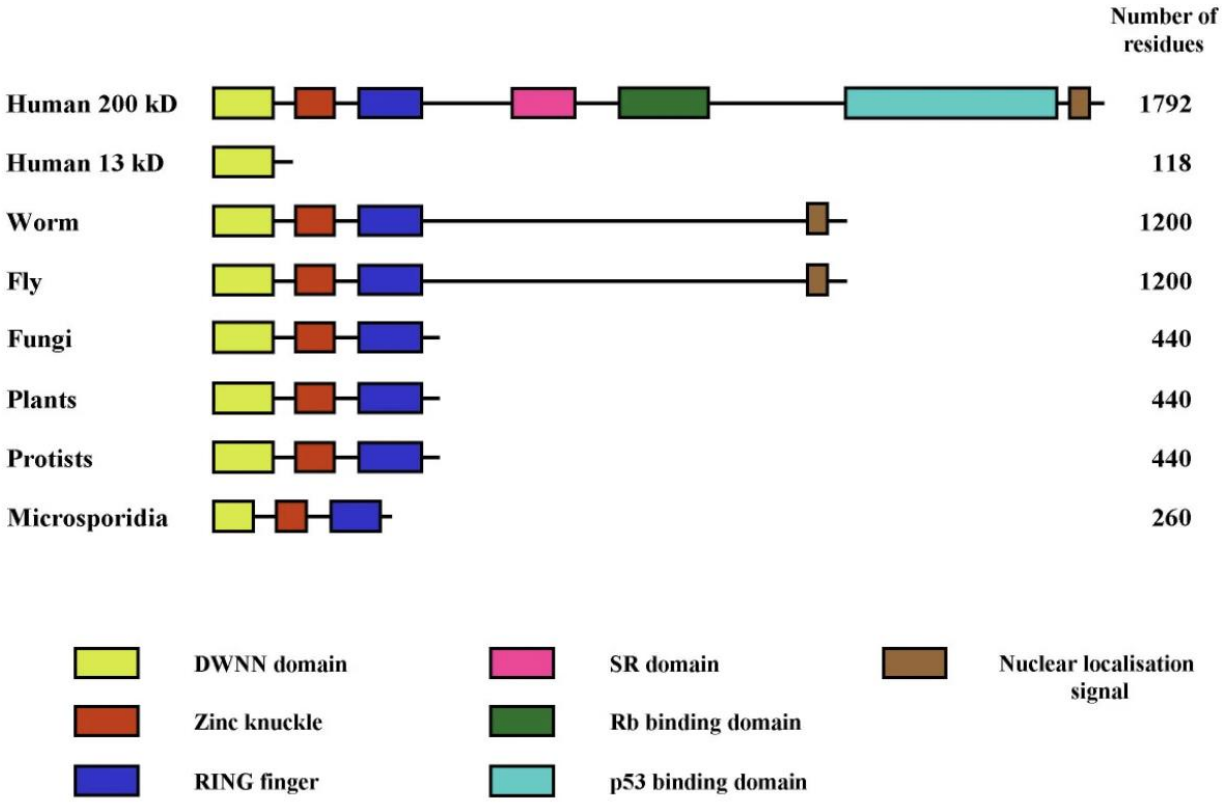


Figure 2.2: Pictorial representation of the domains within the Retinoblastoma binding protein 6 (RBBP6) protein. The multifaceted macromolecule contains at least 7 domains that all involved directly or indirectly in functions pertaining to cell cycle regulation and apoptosis and are thus involved in the incidence of cancer (Pugh *et al.*, 2006).

Human zinc knuckles are also known to play a role in protein-protein interactions, DNA interactions and protein-RNA interactions (Krishna *et al.*, 2003; Matthews and Sunde, 2002).

Found on the C-terminus of RBBP6 on residues 1142–1727 is the p53 binding domain that interacts with the tumour suppressor p53-binding protein (Pugh *et al.*, 2006; Simons *et al.*, 1997). Interaction of p53 with the p53 binding domain of RBBP6 has an involvement in ubiquitination (Moela *et al.*, 2014). Regulation of p53 function takes place by tetramerization, assisted by the p53 DNA binding site (Balagurumoorthy *et al.*, 1995). Activation of p53 plays a role in apoptosis, cell cycle arrest, and DNA repair. Another RBBP6 C-terminal domain protein located on residues 964-1120 is the Rb domain which also interacts with another tumour suppressor protein called Rb (Pugh *et al.*, 2006). This interaction is extremely vital as Rb is involved in gene transcription within the regulation of the cell cycle and DNA synthesis (Sellers and Kaelin, 1997). It plays important roles in protein differentiation, apoptosis, and cell proliferation regulation through inhibition of the cell cycle S-phase (Du and Searle, 2009). Mutations within these networks lead to various diseases including cancer.

2.7 The Really Interesting New Gene (RING) finger domain

The Really Interesting New Gene (RING) is often found in the E3-ubiquitin ligases and is located on the N-terminus of RBBP6 (Pugh *et al.*, 2006). Its location assists in the movement of the E2 molecule during the ubiquitination process without the formation of thioester bonds (Kandias *et al.*, 2009). It is highly conserved, rich in cysteines and contains two zinc ions to support protein folding. The motifs are differentiated based on the presence of a metal-chelating residue either in the form of a cysteine or histidine and these allow for the domain's cross brace topology, where the first and second Zn^{2+} ions are attached to Pairs 1 and 3 and Pairs 2 and 4, respectively (Figure

2.3: Kandias *et al.*, 2009; Kappo *et al.*, 2012; Kosarev *et al.*, 2002). The fold contains two large loops stabilized by one zinc ion that is situated next to three β-sheets and an α-helix. RING motifs are mostly presented in the Cys-X_2-Cys-X_{9-39}-Cys-X_{1-3}-His-X_{2-3}-Cys or His-X_2-Cys-X_{4-48}-Cys-X_2-Cys formations. RING domains come in different patterns enabled by the Zn^{2+} ions coordination in the form of C_3HC_4, C_3HHC_3, $C_2H_2C_4$ and C_4C_4 with C_3HC_4 as shown in Figure 2.3 being the most common. The ability of RING finger domains to dimerize and form homodimers allow them to play an important role in the ubiquitination process (Kappo *et al.*, 2012).

2.8 Ubiquitination

The RING finger domain function in a considerable number of significant processes that include negatively regulating tumour suppressors Rb and p53, regulation of translation and transcription, cell cycle regulation, apoptosis, and proteolysis (Chibi *et al.*, 2008; Kappo *et al.*, 2012; Kosarev *et al.*, 2002; Mbita *et al.*, 2011; Pugh *et al.*, 2006; Simons *et al.*, 1995; Xiao *et al.*, 2019). All these pathways form part of the ubiquitination process. The default fate of soluble proteins to be secreted from the cell along the secretory pathway is to be gradually taken up by endocytosis and digested in lysosomes (Lodish *et al.*, 2000). Some proteins are diverted from the default secretory pathway to regulated secretory vesicles, where they possess important functions such as degradation.

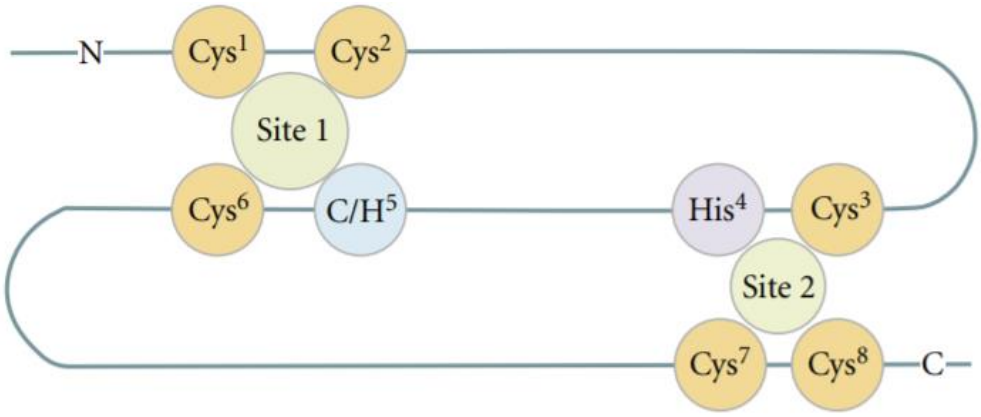


Figure 2.3: Schematic diagram of the RING finger domain topology. It is a cysteine rich domain that is structured in a "cross brace" formation containing two zinc binding sites (Chasapis *et al.*, 2010)

Proteins that are destined for degradation become ubiquitinylated on their cytoplasmic domains (tagged with a ubiquitin group) and play a crucial part in most signalling pathways that include cancerous cells. In the mid-1970s, Goldstein and colleagues identified a small protein of 76 amino acids called ubiquitin, found in almost every living cell in some organisms, with its primary function being unknown. Over the years it was then discovered that ubiquitin served as a tag, passed down by a cascade of enzymes known as ubiquitin-activating (E1), ubiquitin-conjugating (E2) and ubiquitin-ligating (E3) enzymes that direct proteins to their designated degradation destination. Damaged and toxic proteins are removed for regulation of the cell cycle and in turn, cell death and signalling. E1 first begins the process by activating ubiquitin through the action of adenosine triphosphate (ATP). Ubiquitin is transferred to E2, which is accompanied by the formation of a thioester bond between E1 and ubiquitin. Ubiquitin is then transferred to E3 through a catalytic reaction involving a lysine residue of the substrate bound to E3. The process is continuous to the point of a ubiquitin chain forming on the substrate, followed by a propelled signal for the ubiquitin chain to be sent for proteasomal degradation (Figure 2.4).

2.9 Ubiquitination and Cancer

The understanding of ubiquitin mediated degradation has increased the knowledge of their role in protein modification and cancer, which has enabled the development of several targeted therapeutics, E3 ligases being the most common therapeutic target due to their specificity in the ubiquitination system and their participation in cancer development (Diehl *et al*, 2010). E3 ligases are classified into three different groups, and each of these group of classes differs based on their structure and mechanism used to transfer ubiquitin to the protein substrate.

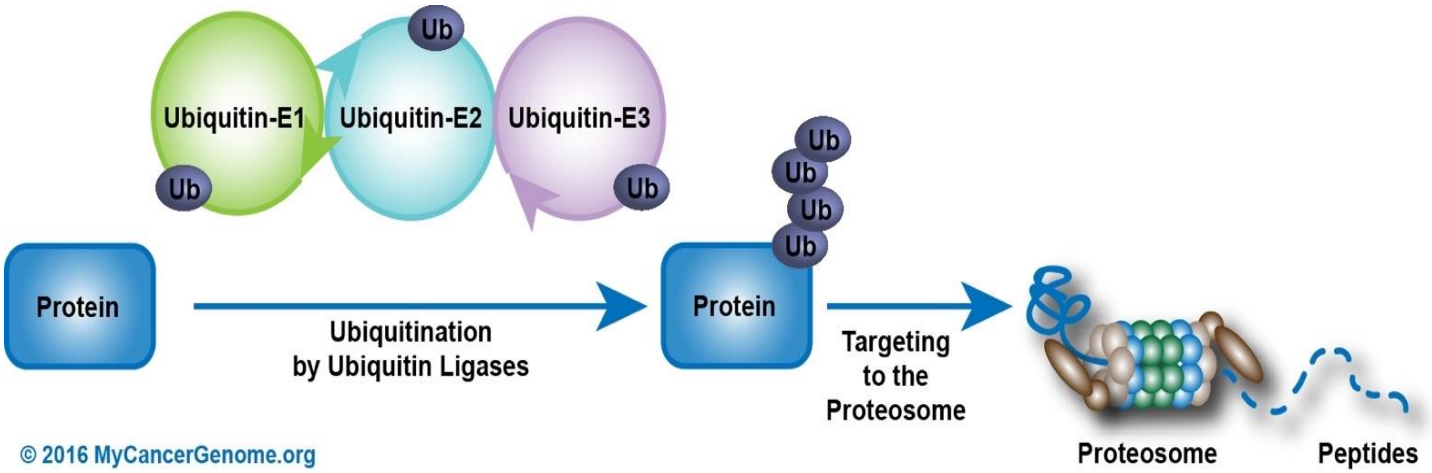


Figure 2.4: Pathway illustrating ubiquitination of proteins destined for proteasomal degradation. Toxic and damaged proteins are removed. Defects in the ubiquitination pathway lead to diseases such as cancer. Figure taken from https://www.mycancergenome.org/content/pathways/protein-degradation-ubiquitination/ [Accessed 12 June 2018].

The three classes are named the RING, HECT (homologous to E6-associated protein carboxyl terminus) and U-box. As mentioned earlier, RING domains negatively regulate p53 and this is driven by the E3 ligases. This is extremely important as p53 is involved in apoptosis, DNA repair and cell signalling. Regulation of p53 however is also performed by ubiquitination (Deng *et al.*, 2020; Sane and Rezvani, 2017). The activity, stability, and expression of E2 and E3 ligases are affected when there is an interaction with p53. The expression of RBBP6 has been associated in both the proliferation and apoptosis of numerous types of cancers and has thus enabled researchers to unravel their function and association with other proteins to deliver new intervention strategies.

CHAPTER THREE: MATERIALS AND METHODS

3.1 Computational Analysis

3.1.1 Homology Modeling

The National Centre for Biotechnology Information (NCBI) (https://blast.ncbi.nlm.nih.gov/Blast) was probed to retrieve the gene sequences for the RING finger domain and Hsp40 substrate binding domain proteins. The Translate tool on the ExPASY web server (https://web.expasy.org/translate/) was then consulted to translate the gene sequences into amino acid sequences. These were then used to select templates using the BLAST tool (Basic Local Alignment Search) (https://blast.ncbi.nlm.nih.gov/Blast.cgi) and those with either the highest sequence identities or percentages higher than 30% were selected. These included: 3ZTG_A (Solution structure of the RING finger-like domain of Retinoblastoma Binding Protein-6 (RBBP6) – 35.85%), 2YUR_A (Solution structure of the RING finger of human Retinoblastoma-binding protein 6 – 47.37%) and 5H7R_D (Structural basis of the flanking zinc-finger motifs crucial for the E3 ligase activity of the LNX1 RING domain – 54.84%) for the RING finger domain protein; 3AGZ_A (Crystal structure of human Hsp40 Hdj1 peptide-binding domain complexed with a C-terminal peptide of Hsp70 [Homo sapiens] – 97.3%), 2QLD_A (Human Hsp40 Hdj1 [Homo sapiens] – 97.3%), 3AGX_A (Crystal structure of human Hsp40 Hdj1 peptide-binding domain [Homo sapiens] – 100%) for Hsp40 substrate binding domain, respectively. The PDB files of these proteins were then retrieved from the Protein Data Bank database (https://www.rcsb.org) and used to model the proteins of interest using the built-in version of MODELLER in Chimera (https://www.cgl.ucsf.edu/chimera/). Validation of the models were then done by mapping Ramachandran plots using the PROCHECK server (https://saves.mbi.ucla.edu/) and further analysis was employed on the resultant models.

3.1.2 Molecular Docking

Molecular docking is a technique that investigates how small molecules behave within the binding site of a target protein (Pagadala *et al.*, 2017). Based on the estimated molecular weights of the two proteins, the RING finger domain protein was considered ligand and Hsp40 the receptor. UCSF Chimera was used to prepare the proteins for docking by adding hydrogens to the receptor and deleting them from the ligand. The files were then saved in .pdb and .mol2 formats respectively. Before docking was performed, the binding pockets available on Hsp40 were first identified using MetaPocket 2.0 and thereafter, AutoDock tools were used to open and convert the previously prepared formats into rec.pdb and .pdbqt files. The residues were then enclosed by a grid box with specific X, Y and Z centers, after which Raccoon and Autodock Graphical user interface, which are software provided by MGL tools, were then used to dock the RING finger domain protein onto Hsp40SBD using default docking parameters. The complex (RING~Hsp40SBD) that exhibited the lowest Z-score or most negative binding score, which represents the least energy, was then selected, and subjected to molecular dynamics simulations.

3.1.3 Molecular Dynamics Simulations

For some time now, molecular dynamics (MD) simulations has developed into a robust tool that not only studies the dynamic properties of a biological system, but is also a powerful technique that assists in understanding the structure-to-function association of macromolecules (Hospital *et al.*, 2015). In this project, the GPU version of the PMEMD engine available within the AMBER 18 package was used to perform MD simulations (Nair and Miners, 2014). Atomic partial charges were generated by ANTECHAMBER using General Amber Force Field (GAFF) processes. All atoms were subjected to a 10Å distance of any box edge by suspending each system within an

orthorhombic box of TIP3P molecules through the LEAP module of AMBER 18. The systems were also neutralized by adding Na^+ or Cl^- counter ions. An initial minimization of 20000 steps was carried out with an applied restraint potential of 500kcal/mol. Without any restraints, a conjugate gradient algorithm was used to perform a full minimization of 1000steps. To make sure that all systems retained the same volume and number of atoms, a 0K to 300K gradual heating MD simulation was applied for 50ps, while the system's solutes were subjected to a potential harmonic restraint and collision frequency of 10kcal/mol and 1ps, respectively. This was followed by a 500ps equilibration step with the temperature kept at a constant 300K and all atom and pressure numbers also kept constant to imitate an isobaric-isothermal ensemble. The Berendsen barostat was used to maintain pressure of the systems at 1bar and the simulation was run for 130ns. The SHAKE algorithm was used to limit the hydrogen atom bonds during each simulation. The step size of each simulation was 2fs and an SPFP precision model was used. This coincided with other parameters that were previously mentioned such as 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.

3.1.4 Post-Molecular Dynamics Simulation Analysis

Post dynamic analyses were done by investigating parameters such as the Root Mean Square Deviation (RMSD), Root Mean Square Fluctuation (RMSF), Radius of Gyration (RoG), Solvent Accessible Surface Area (SASA) and hydrogen bonding. This was done by saving the coordinates of the free and bound complexes after every 1ps and analyzing the trajectories using the CPPTRAJ module within the AMBER 18 suite.

3.2 Wet-lab Analysis

3.2.1 General Stock Solutions, Buffers and Media

4X Sample Buffer: 10mL Stacking Buffer, 8mL Glycerol, 0.8g Sodium Dodecyl Sulfate (SDS), 0.8mL 2-mercapto-ethanol/Dithiothreitol (DTT), 0.2mg Bromophenol Blue and 1.2mL dH_2O

5X Running Buffer: 15.1g Tris, 72g Glycine, 5g SDS dissolved in 1L dH_2O

10X Phosphate Buffered Saline (PBS): 80g NaCl, 2g KCl, 14.4g Na_2HPO_4, 2.4g KH_2PO_4 dissolved in 1L dH_2O. The pH of the solution was adjusted to 7.4. The final working concentration used was 1X

Acrylamide/Bis: 30g Acrylamide, 0.8g *bis*/Acrylamide dissolved in 100mL dH_2O

Ammonium Persulphate (APS): 1g APS dissolved in 10mL dH_2O to make 10% stock solution

Ampicillin: 1g Ampicillin in 10mL dH_2O to make 100mg/ml stock solution. The final working concentration used was 100µg/mL

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

Cleavage Buffer: 50mM Tris-HCL pH7.0, 150mM NaCl, 1mM DTT and 0.01% Triton X-100

Coomasie Blue Stain: 0.25g Coomasie Blue R-250, 100mL acetic acid, 450mL Methanol/Ethanol and 450mL dH_2O

Destaining Solution: 15% Acetic Acid OR 100mL Acetic Acid, 200mL Methanol and 700mL dH_2O

Dithiothreitol (DTT): 1.5g DTT dissolved in 10mL dH_2O to make a 1M stock solution

Elution Buffer (His-tagged proteins): 50mM Tris (pH 8.0), 500mM NaCl and 500mM Imidazole. The pH of the solution was adjusted to 7.4–8.0

Elution Buffer (GST-tagged proteins): 50mM Tris (pH8.0), 15mM reduced Glutathione and 0.1% Triton X-100

Equilibration/Wash Buffer (His-tagged proteins): 50mM Tris (pH 8.0), 500mM NaCl, 40mM Imidazole. The pH of the solution was adjusted to 7.4–8.0

Equilibration/Wash Buffer (GST-tagged proteins): 1X PBS (pH7.4), 1mM DTT and 1mM PMSF

Isopropyl β-d-1-thiogalactopyranoside (IPTG): 2.38g IPTG dissolved in 10mL dH_2O to make a 1M stock solution. The final working concentration of the solution was 0.5mM

Luria Broth: 50g Luria Broth dissolved in 2L dH_2O, autoclaved for 1hr at 120°C

Lysis Buffer (His-tagged proteins): 50mM Tris-HCL, 100mM NaCl and 10mM Imidazole

Lysis Buffer (GST-tagged proteins): 1X Phosphate Buffered Saline (PBS), 100µg/mL Lysozyme, 1mM DTT and 1mM PMSF

Lysozyme: 0.5g Lysozyme dissolved in 20mL dH_2O water to make 25mg/mL stock solution. The solution was aliquoted and stored at -20°C. The final working concentration used was 100µg/mL

NMR Buffer: 100mM NaH_2PO_4 pH6.0, 50mM NaCl, 5mM DTT and 0.02% NaN_3.

Nutrient Agar: 31g Nutrient agar dissolved in 1L dH_2O and autoclaved at 120°C for 1hr

Phenylmethylsulfonyl fluoride (PMSF): 1.74g PMSF dissolved in 100mL isopropanol, methanol, or ethanol to make a 1M stock solution. Final working concentration used was 1mM

5X Running Buffer: 15.1g Tris, 72g Glycine and 5g SDS dissolved in 1L dH_2O. The pH was approximately 8.3 – 8.5 but was not adjusted

Separating/Resolving Gel Buffer: 18.2g Tris and 0.4g SDS dissolved in 100mL dH_2O and pH was adjusted to 8.8

Stacking Gel Buffer: 6.05g Tris and 0.4g SDS dissolved in 100mL dH_2O and pH was adjusted to 6.8

Sodium Azide: 10g NaN_3 dissolved in 100mL dH_2O to make a 10% stock solution. The final working concentration used was 0.02%

Sodium Phosphate Buffer: 138g $NaH_2PO_4.H_2O$ in 1L dH_2O to make 1M stock solution. The final working concentration used was 100mM

Table 3.1: 15% SDS PAGE Gel Constituents

	Separating Gel	**Stacking Gel**
dH_20	3.20mL	3.17mL
Separating/Resolving Gel	2.63mL	_
4X Stacking Buffer	_	1.25mL
40% BisAcrylamide 37:5:1	4.00ml	500µL
10% SDS	150µL	50µL
10% APS	20µL	25µL
TEMED	7.5µL	5µL

3.2.2 Bacterial Strains Used

1. *E. coli* BL21 (DE3) (Promega) pLysS strain: F–, ompT, hsdSB (rB–, mB–), dcm, gal, λ(DE3), pLysS.

2. *E. coli* JM109 (Sigma) strain: endA1, recA1, gyrA96, thi, hsdR17 (rk–, mk+), relA1, supE44, Δ(lac-proAB), [F´ traD36, proAB, laqIqZΔM15].

3.2.3 Competent Cell Preparation

Glycerol stocks of *E. coli* BL21 and JM109 cells were streaked on agar plates and grown at 37˚C overnight. Single colonies were picked from the overnight plates and incubated in separate test tubes containing 5mL YT broth with shaking at 37˚C overnight. The next morning, these were scaled up, inoculated into 50ml broth, and further incubated at 37˚C until an optical density (OD_{600}) of 0.4–0.6 was reached. The cultures were then transferred to sterile 50mL tubes and centrifuged for 10minutes. The supernatants were discarded, 10mL of cold $MgCl_2$ added to the pellets and incubated on ice for 30minutes. The mixtures were centrifuged at 5000rpm for 10minutes followed by the discarding of the supernatants and re-suspension of the cells in 10mL ice-cold $CaCl_2$. The mixtures were once again incubated on ice for 4–5hours followed by centrifugation and discarding of the supernatants. Equal amounts of $CaCl_2$ and 30% glycerol were added to the pellets and the resulting cell suspensions, now considered competent cells, were aliquoted and stored at -80˚C for further use.

3.2.4 Transformation of Expression Constructs into *E. coli* BL21 cells

Both the pGEX-6P-2~RING and pQE30~SBDHsp40 DNA constructs were purchased from GeneScript (USA) (Figure 3.1. and 3.2) and the DNA material re-suspended in 30µL dH_2O. BL21

and JM109 competent cells were allowed to thaw on ice and 2µL of each construct and 100µL of competent cells was combined in sterile microcentrifuge tubes, respectively. The vials were incubated on ice for 30minutes. A negative control was also set up by following the same procedure but without the inclusion of any construct. The vials were transferred to a water bath and the mixtures were heat shocked at 37˚C for 5minutes. These were afterwards immediately returned on ice for 5minutes.

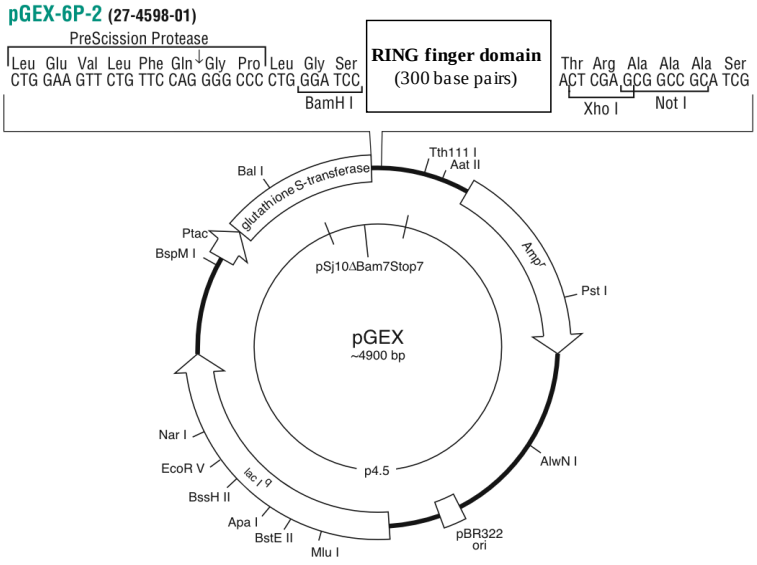


Figure 3.1: Illustration of pGEX-6P-2~RING DNA construct. The gene encoding the RING protein was cloned into the pGEX-6P-2 expression vector using BamHI and Xho I restriction enzymes.

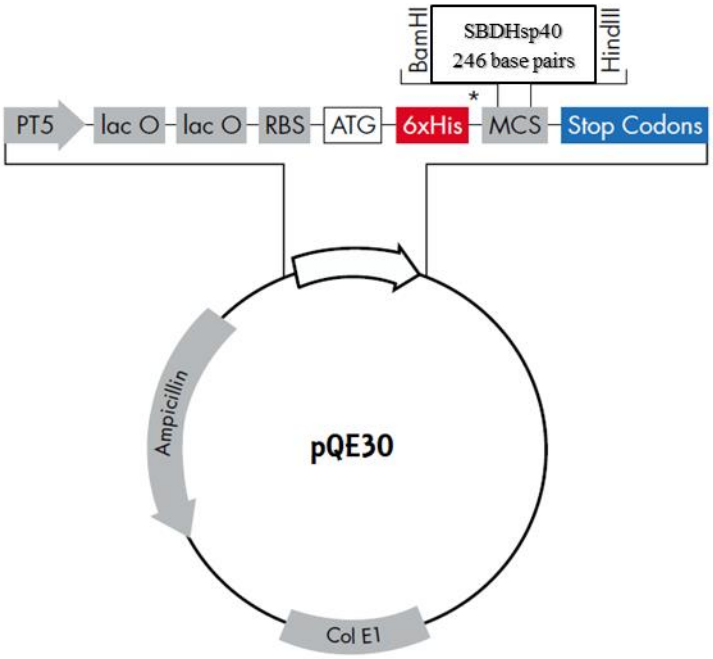


Figure 3.2: Illustration of pQE30~SBDHsp40 DNA construct. The gene encoding the RING protein was cloned into the pQE-30 expression vector using BamHI and HindIII restriction enzymes.

Following this, 900µL of pre-warm YT broth without antibiotics was added and mixed gently by inverting the tubes a few times. The mixtures were then incubated at 37˚C for 2hrs in a shaking incubator. In the meantime, agar plates containing ampicillin were prepared and the cell suspensions were spun down at the highest speed for 5minutes. The supernatant was discarded, and the pellets re-suspended in 100µL YT broth, which was spread onto the agar plates and incubated overnight at 37˚C. The appearance of colonies the following morning and clean control plates indicated successful transformation.

3.2.5 Expression Screening

Four colonies from the transformation plates were randomly picked, inoculated in 5mL Luria broth containing ampicillin and incubated for 4hrs at 37˚C with vigorous shaking. Thereafter, 1mL

samples were removed from each culture and transferred to sterile microcentrifuge tubes with the addition of 0.5mM Isopropyl beta-D-thiogalactoside (IPTG); this served as the induced culture. Un-induced samples were also set up without the addition of IPTG, serving as negative controls. Both the samples were incubated for a further 2hrs at 37˚C with vigorous shaking alongside the remaining 3mL cultures, after which the cells were harvested by centrifugation at 5000Xg for 10minutes. This was followed by re-suspension of the pellets in 50µL of 2X sample buffer and heating of the samples at 95˚C for 10minutes. The mixtures were then centrifuged at the highest speed for 1minute and 20µL immediately electrophoresed on a 15% SDS-PAGE gel. Staining and destaining of the gel the revealed the best expressing colony to be used for subsequent expression.

3.2.6 Large Scale Protein Expression and Extraction

A 100µL of the best expressing culture was inoculated in 100mL of LB broth containing ampicillin and incubated at 37˚C with shaking overnight. The next day, the culture was scaled up to 2L and left incubating at 37˚C with shaking until an OD_{600} of 0.4–0.6 was reached. Protein expression was then induced through the addition of 0.5mM IPTG. The expression temperature was then decreased to 25˚C and the cultures was left to grow overnight with shaking. The following day, the cultures were transferred to 250mL centrifuge tubes, and the cells harvested at 10 000Xg for 20mins. The supernatant was then discarded, and the pellets were re-suspended in their appropriate lysis buffer. For the GST~RING protein, the pellet was re-suspended in 1X PBS (pH7.4) containing 1 mM DTT, 1 mM PMSF, 100 µg/mL Lysozyme, 0.02% of 10% NaN_3 and Triton X-100. For the His~Hsp40SBD protein, the cells were re-suspended in lysis buffer (pH7-8) constituting 50mM Tris-HCl, 100mM NaCl and 10mM Imidazole. The cells were then stored at -80˚C until further use.

To release the proteins of interest, protein extraction was done. The re-suspended total bacterial cell pellets stored in -80°C were allowed to thaw on ice. To each 50mL sample, a tablet of cOmplete EDTA-free Protease inhibitor cocktail tablets (Merck, Darmstadt, Germany) was added to prevent protein lysis. The samples were subjected to 3–5 cycles of freeze-thawing followed by sonication at 4˚C for 5minutes using 30seconds pulse intervals. Once the cells are broken down, the proteins of interest were released from the cells and clear protein lysate was obtaining by centrifugation at 20 000Xg for 20minutes. This was then stored at 4˚C until required for purification. To determine the expression profile of the proteins, 50µL samples were collected from the clear lysates and prepared for electrophoresis by adding 50µL of 2X sample buffer and heating the mixtures for 10minutes at 95˚C. A 15% SDS-PAGE gel was used to analyse 20µL from each sample after which protein expression cultures of up to 10L were then set up for over-expression of the proteins.

3.2.7 Purification of GST~RING Protein by Affinity Chromatography and Gel Filtration

The RING finger domain of RBBP6 was cloned into the pGEX-6P-2 vector which encompasses a GST tag, thus requiring the use of Glutathione-agarose linked beads (Merck, Darmstadt, Germany). Affinity chromatography by gravity was performed, first by packing the beads into a column and allowing the beads to swell overnight. The beads were then washed with 5CV of water and equilibrated with 5CV of 1X PBS (pH7.4) containing 1mM DTT and 1mM PMSF. The lysate was then poured into the column and the flowthrough collected after which the column was washed with 5CV equilibration buffer to eliminate any non-specific binding. The GST-fusion protein was then eluted using a buffer containing 50mM Tris (pH8.0), 15 mM reduced Glutathione and 0.1%

Triton X-100. To cleave off the GST-tag from the protein of interest, the GST-fusion protein was subjected to dialysis overnight at 4˚C using SnakeSkin Dialysis Tubing (ThermoFisher Scientific) with a molecular weight cut-off (MWCO) of 3500 Da and 5µL of human rhinovirus (HRV) 3C protease. This was done in ice cold cleavage buffer constituting 50mM Tris-HCl pH7.0, 150mM NaCl, 1mM DTT and 0.01% Triton X-100 with continuous stirring. The sample was then subjected once again to the glutathione agarose column to retrieve the cleaved RING finger domain protein via the flowthrough and the fractions were then analyzed on a15% SDS PAGE gel. Following this, the protein was purified to homogeneity using size exclusion chromatography/gel filtration. Samples corresponding to the eluted and cleaved semi-purified RING finger domain protein were pooled together, concentrated down, and subjected to a HiLoad Superdex75 prep grade (pg) column constituting a 16mm and 600mm diameter and bed height 16/600 respectively. The column was equilibrated overnight with 1CV NMR buffer (100 mM sodium phosphate buffer pH 6.0 containing 50 mM NaCl, 5 mM DTT and 0.02 % sodium azide) as the storage buffer. The column was running at a flow rate of 0.5mL/min and the final purified protein was eluted in fractions of 375µL. These were stored at 4˚C until required for biophysical characterization.

3.2.8 Affinity Purification of His~SBDHsp40 Protein

The pQE30~Hsp40SBD incorporates six moieties of polyhistidine (6X His) because of its tag which have an affinity for nickel. Thus Nickel-NTA (Merck, Darmstadt, Germany) beads were used during affinity chromatography, which was performed by gravity. The beads were packed onto a column and allowed to swell and settle at 4°C overnight in its solution. Once packed, the beads were thoroughly washed with 5CV water the next morning and thereafter equilibrated with 5CV of equilibration buffer (pH7.4–8.0) containing 50mM Tris (pH8.0), 40mM Imidazole and

150mM NaCl. The lysate was then poured down the column and the flow-through collected. The column was then washed with 5CV of equilibration buffer and the protein eluted for up to 5 fractions with elution buffer (pH7.4–8.0) containing 50mM Tris (pH8.0), 400mM NaCl and 400mM Imidazole. After purification the beads were washed thoroughly with equilibration buffer followed by the addition of 30% ethanol for storage at 20˚C.

3.2.9 Protein Characterization

3.2.9.1 Fourier Transform- Infrared Spectroscopy (FTIR)

Fourier-transform infrared spectroscopy (FTIR) is a spectroscopic technique used to determine the secondary structure and conformational changes that occur within a protein (Yang *et al.*, 2015). A Fourier Transform – Infrared Spectrometer (SHIMADZU) was used to analyze the functional groups present in the proteins at a range of 500–4000cm^{-1} wavelength. The technique was performed using 3ml samples in cuvettes. NMR and the elution buffer were used as blanks for the RING finger domain and SBDHsp40 protein respectively. To determine the vibrational modes of the protein, infrared light was passed through the samples at different wavenumbers to disclose the functional groups.

3.2.9.2 Raman Spectroscopy

Raman spectral data was ascertained using a WITec Alpha300R Confocal Raman spectrometer using 3µl of each protein sample. Spectral data was measured by using an excitation (laser) wavelength of 532.110 nm, a 50X objective and a 100µm optical fiber. A grating 600 grooves per mm (g/mm) for T1 and 500nm for BLZ was used, which resulted in a 50–3800cm^{-1} range, a spectral center of 2103.106 rel. 1/cm and wavelength of 599.161nm. The 600g/mm grating was

used instead of a 400g/mm grating because it yielded a better dispersion of light resulting in a better resolution of spectra, hence the greater the groove density the better the spectra resolution (Gyakwaa *et al.*, 2020) A laser power of 2mW was used on both protein samples. The chemical structure and molecular interactions within the protein were identified through excitation of its photons due to the vibrations produced by the bonds of the molecule when interacting with light (Butler *et al.*, 2016).

CHAPTER FOUR: RESULTS AND DISCUSSION

4.1 Introduction

The main aim of this study is to analyze and ascertain the putative interaction between the RING finger domain of RBBP6 with the substrate binding domain of Hsp40. This chapter first introduces the computational analysis of the two proteins and its interaction. It then goes on to describe the expression and purification of both proteins in preparation for biophysical characterization in a bid to confirm the computational work done earlier. This was a significant study as both proteins have been implicated in the incidence of cancer. The aim of this study was to therefore investigate and affirm whether this interaction exists, mainly because yeast-two hybrid systems are largely known for producing false positives. In addition, a positive interaction between the proteins establishes a basis for the discovery of alternative treatment regimen against cancer.

4.2 Computational Analysis

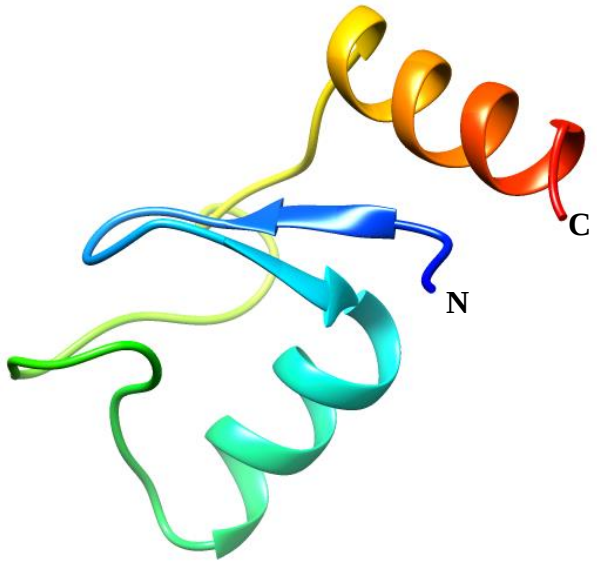
4.2.1 Homology Modeling

The three-dimensional structures and complexes of proteins have formed an integral part in providing information on the many cellular processes that drive biological systems. Through such knowledge, avenues are created on how these operations within such networks can be modulated for downstream processes such as drug design and development (Waterhouse *et al.*, 2018). For this reason, there has been significant growth in the number of proteins structures and complexes that have been deposited in the Protein Data Bank. However, in comparison to the high throughput methods that have been developed for characterizing proteins and their associated interactions, the structure determination rate is still considered relatively slow. Homology modeling aims at

bridging the gap between known protein sequences and structures that have been determined experimentally (Dalton and Jackson, 2007). The basics of this is done by computing the structure of the protein based on the alignment of its amino acid sequence with that of a homologous protein (referred to as a template) whose structure has been solved (Venselaar *et al.*, 2010). In the absence of any experimentally solved structures for the Hsp40 substrate binding domain (SBDHsp40), the protein was modeled using 3 templates of very high sequence identity (all above 90%) and the model revealed that the protein constitutes 2 anti-parallel β-strands and a short α-helix, also referred to as a 3_{10} helix (Figure 4.2). Although the structure of the RING finger domain has been determined experimentally, modeling re-confirmed that the protein contains 2 α-helices and 2 anti-parallel β-sheets (Figure 4.1).

Since homology models are simply predictions and may therefore cast doubt on their confidence, structure validation is therefore necessary. The Ramachandran plot is a 'quality' indicator for protein models based on residue φ–ψ torsion angles (Laskowski *et al.*, 2006). The conceptual framework is that residues apart from glycine and proline i.e. alanine-like residues occupy the three core "allowed regions"; two large ones known as the alpha- and beta-regions that give rise to α-helices and β-strands respectively, and a small region called the alpha1-regions, which represents the backbone conformations of those present in the alpha region (Hollingsworth and Karplus, 2010). Only under the circumstances that peptide atoms do not collide when given standard radii are these regions 'allowed'.

A



B

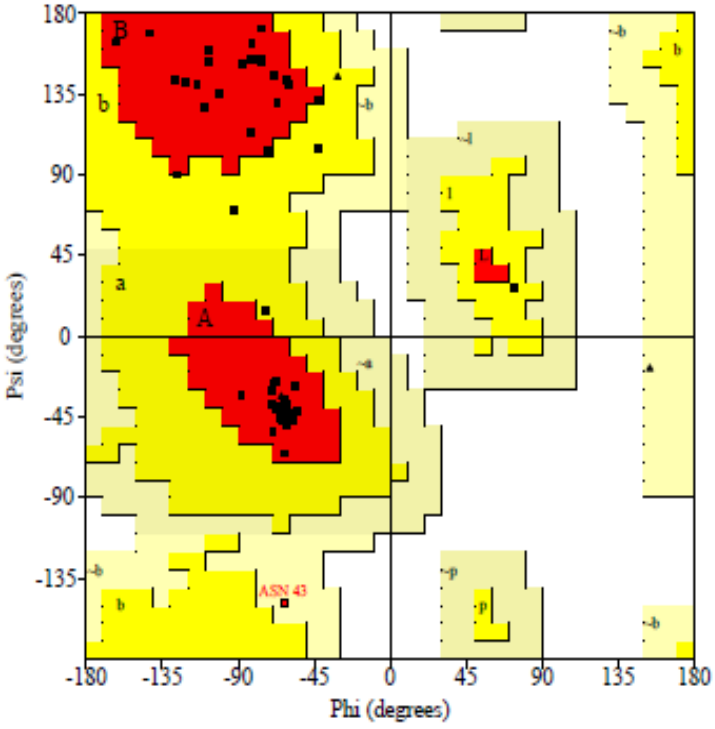


Figure 4.1: Homology model (A) and Ramachandran plot (B) of RING finger domain. The RING finger domain consists of two α-helices and two anti-parallel β-strands. The C-terminal and the N-terminal are both indicated on the model. The Ramachandran plot shows 92.9% of the residues are found in the favored region.

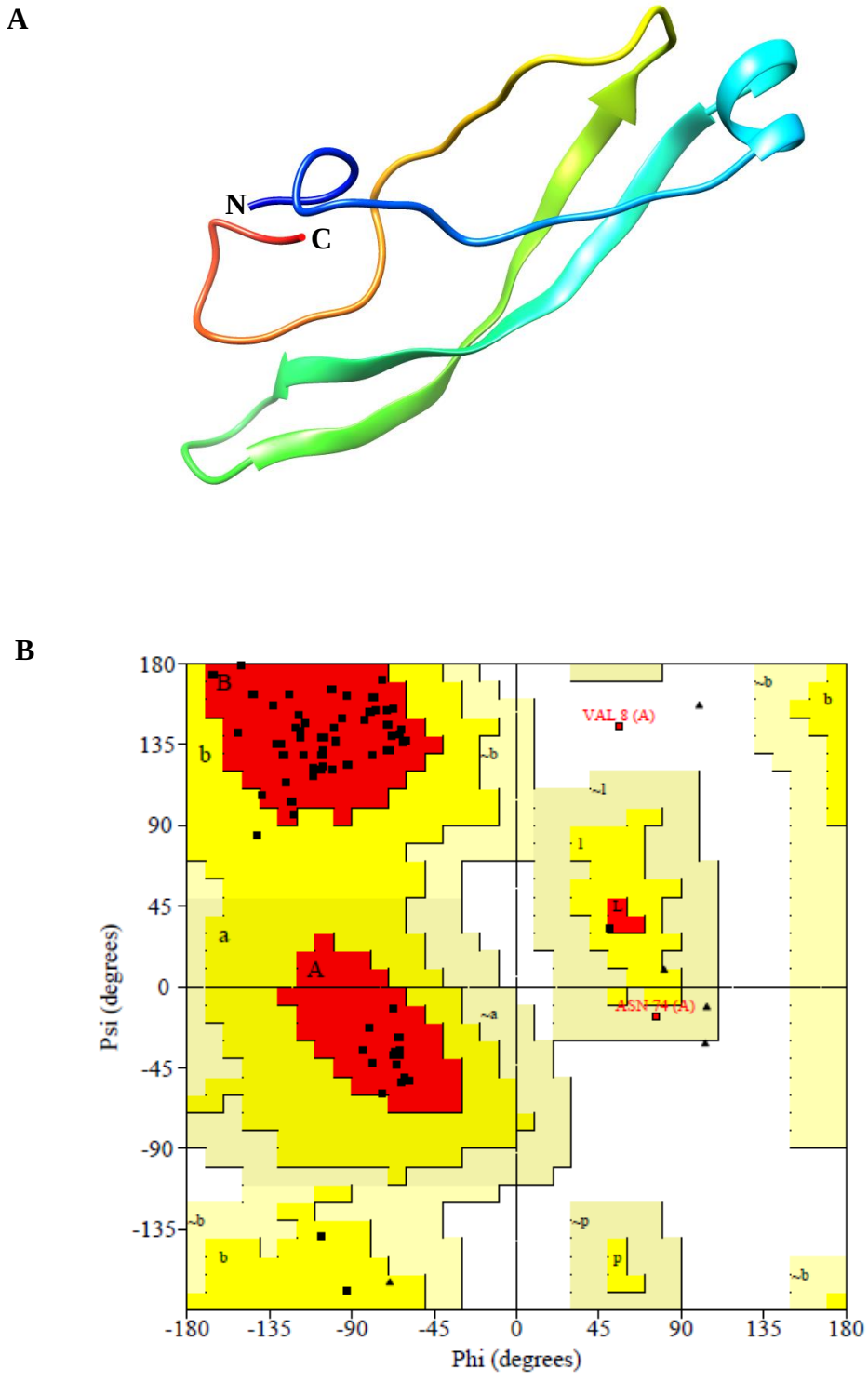


Figure 4.2: Homology model (A) and Ramachandran plot (B) of substrate binding domain of Hsp40. The protein consists of one α-helix and two anti-parallel β-strands. The C-terminal and the N-terminal are both indicated on the model. The Ramachandran plot shows 92.3% of the residues are found in the favored region.

Both the RING and Hsp40 protein models ranked an overall confidence score of 92.9% and 92.3% respectively, indicating that these were trustworthy models and no stereochemical parameters were violated during model building. As a normal rule of thumb, over 90% of all residues within the most favoured region (A, B, L), are a requirement, which both models possessed (Adekiya *et al.*, 2018). For the other regions, 5.4% of the residues were present in the additional allowed regions (a, b, l, p), 1.8% in the generously allowed region (~a, ~b, ~l, ~p) and 0.0% in the disallowed region for the RING protein, while for SBDHsp40 4.6% and 1.5% of the residues alike were found in the additionally favored, generously allowed and disallowed regions respectively. These regions are defined based on how densely populated or tight the clustering of points within the regions is (Laskowski *et al.*, 2006).

4.2.2 Molecular Docking

Modern day drug discovery relies heavily on molecular docking to predict and quantify the binding orientation and affinity between ligands and receptors, either in the form of proteins or enzymes, through the use of electrostatic interactions and shape (Fan *et al.*, 2019; Pagadala *et al.*, 2017). This method has gained immense popularity as it greatly reduces research costs while improving efficiency for further lead optimization processes. The conformation of the complex is evaluated until convergence to the minimum energy has been obtained and an affinity score has been calculated by the addition of Van der Waals and electrostatic forces. The RING finger domain and substrate binding domain of Hsp40 were docked (RING~SBDHsp40) using Autodock tools and five complexes (-7.1, -6.5, -5.4, -7.1 and -8.4) obtained from which the one exhibiting the least binding score or energy of -8.4 kcal/mol (Figure 4.3) was selected.

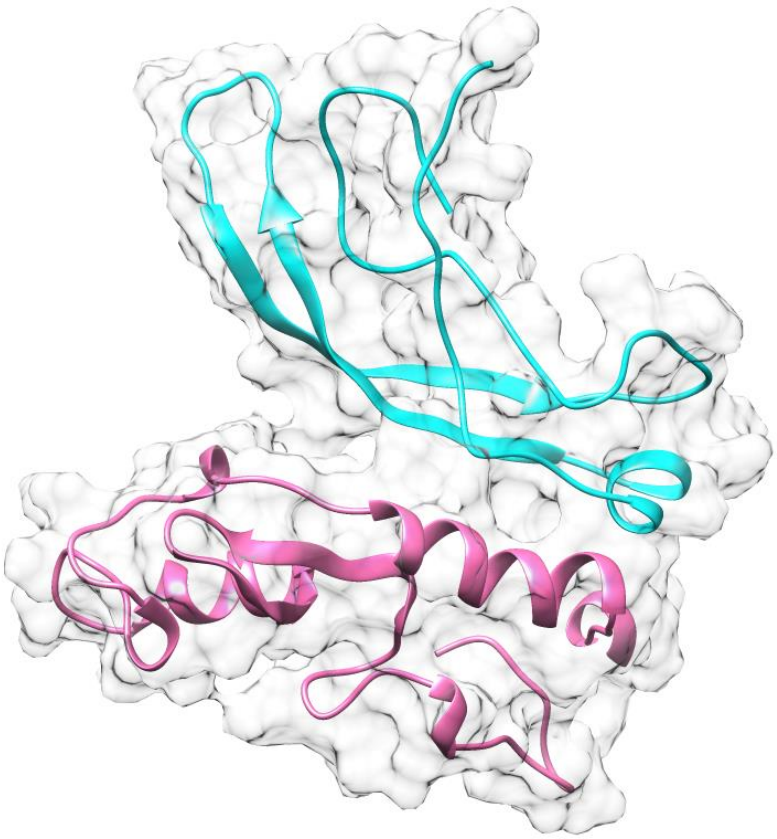


Figure 4.3: Docked complex of the RING finger domain protein and SBDHsp40. The proteins were docked using Autodock tools and the complex with the lowest score was selected for molecular dynamic simulations.

4.2.3 Molecular Dynamics Simulation

Protein 3D structures only provide a partial outlook mainly because proteins and nucleic acids are flexible systems that are driven by dynamics (Hospital *et al.*, 2015). Molecular dynamics (MD) simulation bring biomolecular structures to life by providing insights into the dynamics or movement of biomolecules in solution during different time frames, thereby explaining the physicality of the structure and function of biological macromolecules and therefore systems (Hansson *et al.*, 2002; Karplus and Mccammon, 2002). In addition, the technique calculates thermal averages and which conformations of the complex are thermally accessible. MD simulation of the RING finger domain protein against the RING~SBDHsp40 complex was done and compared by analyzing parameters such as the root mean square deviation, root mean square fluctuation, radius of gyration, solvent accessible surface area and hydrogen bonding.

4.2.3.1 Analysis of System Stability

The RING~SBDHsp40 complex was prepared and using the AMBER18 program, the docked pose was subjected to MD simulations, which were performed for a total of 130ns. Analysis of trajectories for interaction blueprints of the ligands within the active site of the protein is shown over time in Figure 4.4. The root mean square deviation (RMSD) refers to structure variation of a protein-ligand complex and therefore measures the stability or stereochemical variance of a structure, similar to the crystallographic B parameter (Masamba *et al.*, 2020; Peele *et al.*, 2020). It also indicates the movement of atoms where the value is directly proportional to the mobility and contrariwise.

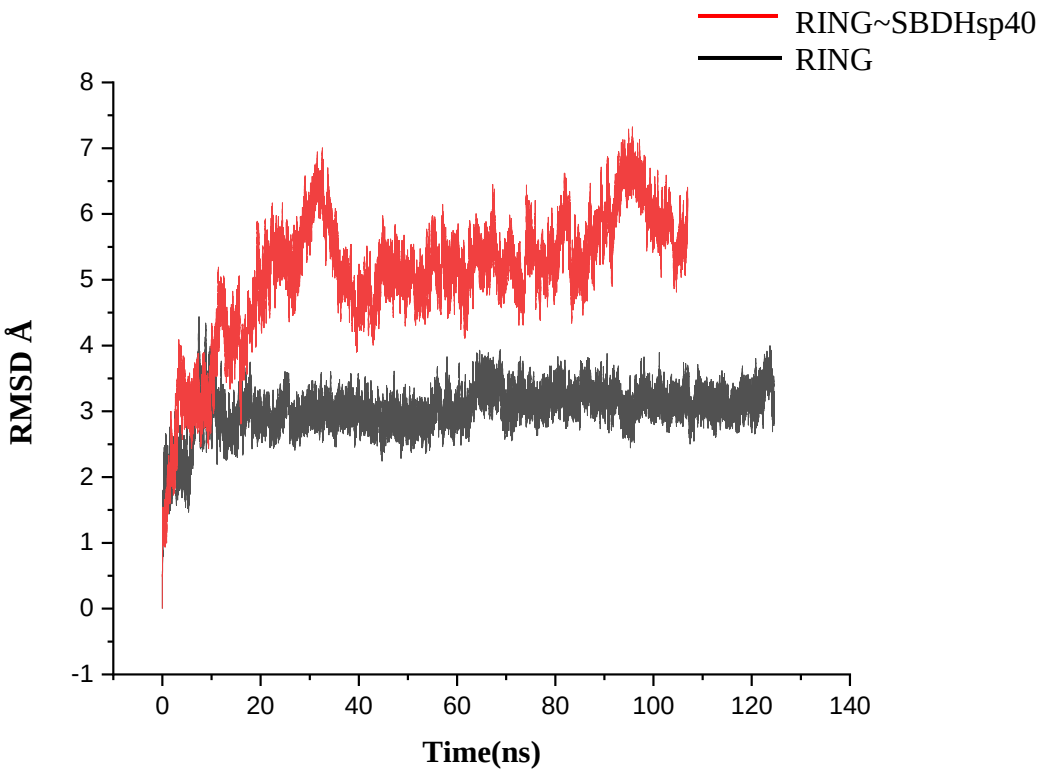


Figure 4.4: Graph showing root mean square fluctuation of the Cα backbone of the apo protein (RING) and complex with RING~SBDHsp40, where the black profile represents RING and red represents the complex.

A high fluctuation was observed throughout the simulation of the RING~SBDHsp40 complex in comparison to the apo (RING finger domain), which has a lower RMSD profile (Figure 4.4). In the apo, fluctuation was observed for the first ~10ns and stabilized for the rest of the simulation indicating a state of protein stability. According to Olotu and co-workers (2019), a low RMSD is representative of a structurally stable entity, while a high value coincides with an increase in instability and structural disparity (Olotu *et al.*, 2019). Although a high level of fluctuation for the RING~SBDHsp40 complex was seen throughout the simulation, most notable was ~30ns and ~100ns. The fluctuation observed may be indicative of the positioning or adjustment of ligand residues (RING) within the active site or hydrophobic pocket of the protein (Hsp40).

4.2.3.2 Analysis of Structural Flexibility

In comparison to RMSD, which provides information on how similar members of an ensemble are based on their global structure, root mean square fluctuation (RMSF) is often studied to determine the thermal stability, variability and most importantly the local structural flexibility of macromolecules (Kuzmanic and Zagrovic, 2010). The Cα backbone of a protein may undergo conformational changes whether in a bound or free state and therefore a protein may adopt a different conformation once bound to another biological molecule (Masamba *et al.*, 2020). The result in Figure 4.5 shows the areas or regions of the apo and protein complex undergoing fluctuation or the restriction of intrinsic flexibility as well as the rearrangement of amino acid residues. The simulation shows a high level of fluctuation that eventually decreases and becomes stabilized after the formation of the complex.

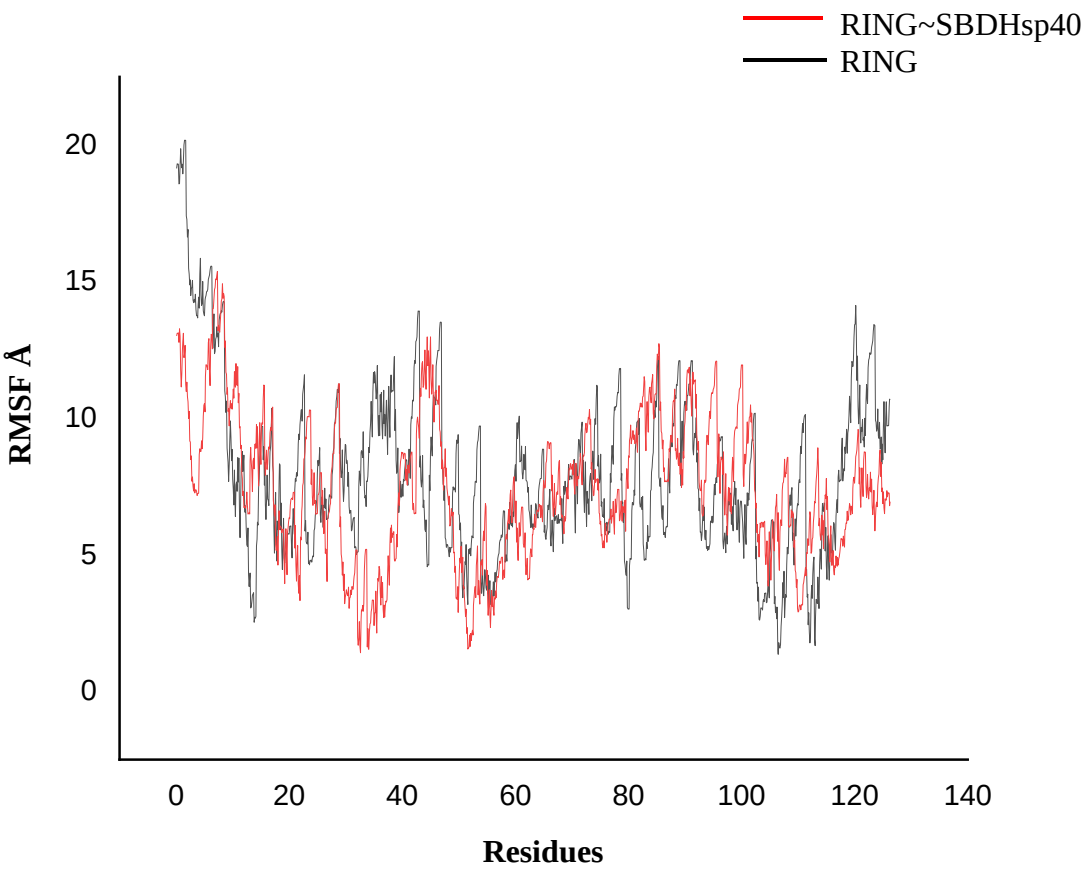


Figure 4.5: Root mean square fluctuation (RMSF) plot for the apo RING protein and RING~SBDHsp40 complex. This was obtained so as to determine the average mobility of atoms within the backbone of the proteins during MD simulation.

The region between ~30 – ~40ns shows a sharp increase and decrease in fluctuation in the RING and RING~SBDHsp40 profiles respectively, which coincides with the zinc-coordination site within the RING finger domain protein. Within the RING protein, one of the zinc coordination sites, which is one of the most flexible regions in the protein, contains Asn34 whose side chains binds Zn ions (Kappo, 2009). As the complex is formed, the area does not coordinate to any ion but remains flexible, which is accounted for in the RMSF profile. As similar profile was observed in the RMSD results which showed an increase in fluctuation at ~30ns.

4.2.3.3 Analysis of Structural Compactness

A protein can refold under normal physiological conditions from a denatured to its native state. Although proteins have an endless number of potential conformations, they somehow are able to form their unique stable 3D structure over time (Lobanov *et al.,* 2008). One of the challenges in molecular biology and biophysics is understanding this phenomenon. The radius of gyration (RoG) is an important parameter that describes this action as well as the static compactness of a protein (Hong and Lei, 2009). It calculates the root mean square distance from the centre of gravity to the edges of the protein. A very significant difference is seen between the profiles of the protein and complex in Figure 4.6. The results show that the Cα backbone of the RING protein is not completely rigid and undergoes distortion once the Hsp40 protein binds and a complex is formed. This flexibility is seen during the first ~90ns of the RING~SBDHsp40 profile and eventually stabilizes for the rest of the simulation. This result coincides to the RMSD profile of the RING~SBDHsp40 complex and as well as the RMSF results.

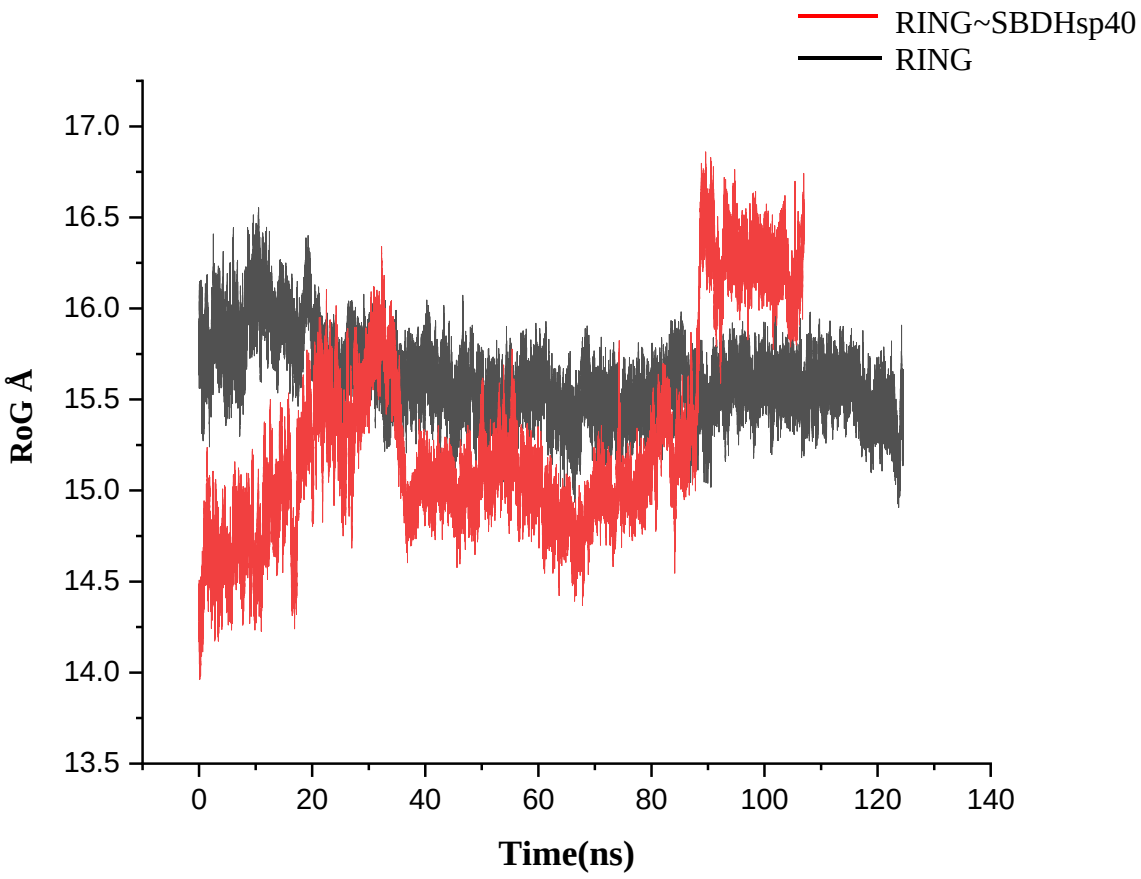


Figure 4.6: Radius of gyration (RoG) profile of the RING protein and RING~SBDHsp40 complex during the MD simulations. Results show how rigid or flexible the fluctuation within the Cα backbone of the systems are.

4.2.3.4 Hydrogen Bond Profile

The binding of ligand to protein results in an increase in the formation of electrostatic interactions and hydrogen bonds, which accompanies protein stability. The profile in Figure 4 represents the hydrogen bonds present in the apo and those formed once ligand binding occurs. As is clear from the profile, a larger number of hydrogen bonds are formed over time within the RING~SBDHsp40 complex. From the beginning of the simulation a large catalytic space is open, which reaches a state of stability once the ligand binds, caused by the formation of hydrogen bonds.

4.2.3.5 Solvent Accessible Surface Area Analysis

An important driving force in the folding of proteins is the burial of hydrophobic amino acids within the core of the protein (Durham *et al.*, 2009). This force takes into cognizance the burial of hydrophobic side chains, density of sidechain packing and minimization of surface area. Solvent accessible surface area (SASA) is a geometric measure of how amino acids interact with the protein solvent and core, which is proportional to the exposed surface area. The results in Figure 4.8 show that the RING~SBDHsp40 complex possesses more solvent accessible area than the apo structure. This may have taken place due to a loss of hydrophobicity. To confirm and further understand this result, a Kyte and Dolittle plot was generated via the Expasy Protscale tool (https://web.expasy.org/protscale/) for both the apo structure and the protein complex.

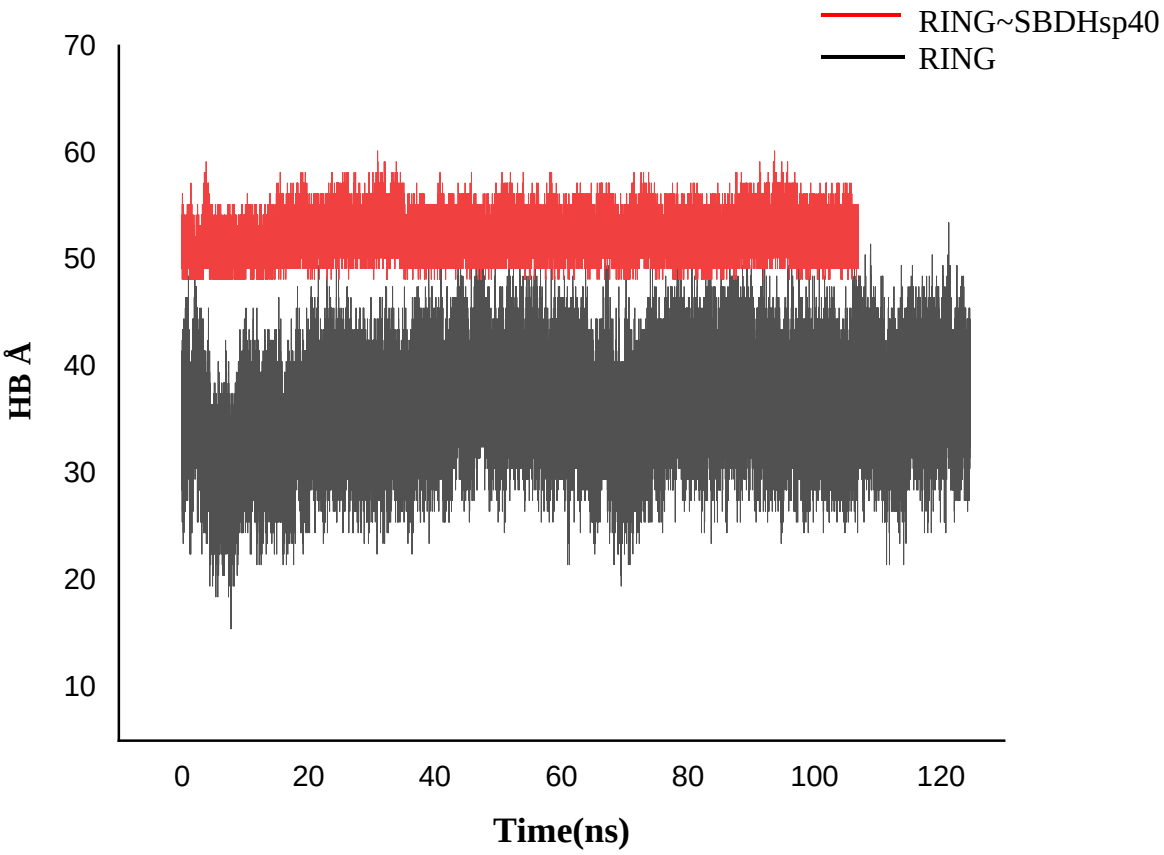


Figure 4.7: Representation of Hydrogen bonding profile of the RING and RING~SBDHsp40 profiles. The graph represents the formation of hydrogen bonds and electrostatic forces within the systems during the MD simulations.

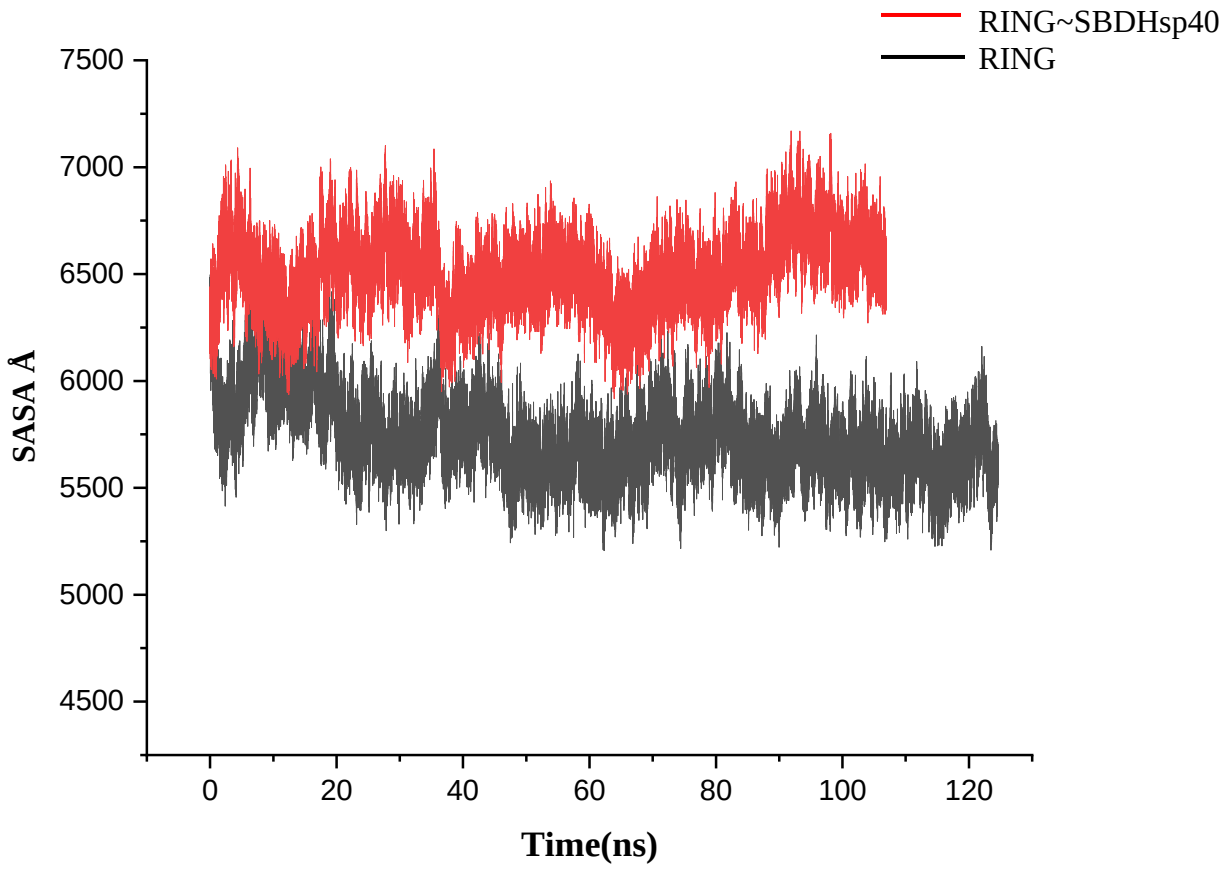


Figure 4.8: Solvent accessible surface area (SASA) profile of RING and RING~SBDHsp40 systems. Analysis of how amino acids within the proteins become exposed or accessible to surface was determined over time and confirmed by using Kyte and Dolittle plots shown below.

The Kyte and Dolittle plot is a measure of the hydrophobic and hydrophilic regions, which are determined based on the properties of the amino acid side chains (Kyte and Doolittle, 1982). Accordingly, the presence of several peaks below zero indicates the overall hydrophobic nature of the protein, while the bulk of peaks above zero represent the hydrophilic regions of the protein (Adekiya *et al.*, 2018). The results displayed in Figures 4.9 and 4.10 show that the RING protein is hydrophobic in nature but loses its hydrophobicity once in complex with Hsp40. This correlates with the SASA results, which indicate that immediately the RING protein binds to Hsp40, the complex becomes more accessible to solvent.

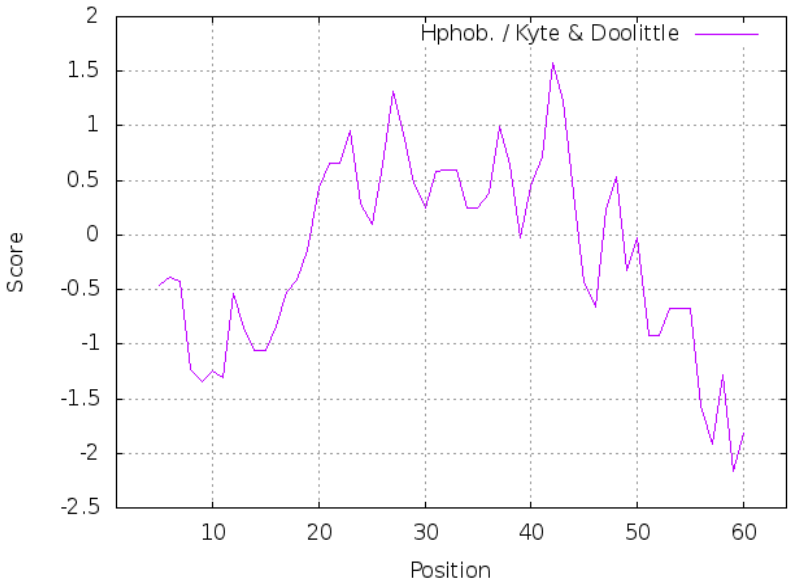


Figure 4.9: Kyte and Dolittle plot of the RING finger domain showing its hydrophobic nature. The presence of peaks above zero indicates its overall hydrophobicity.

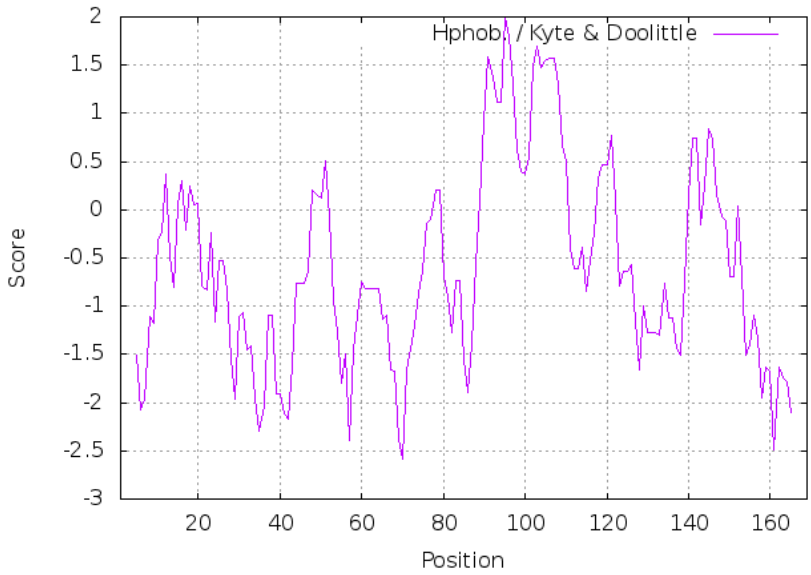


Figure 4.10: Kyte and Dolittle plot of the RING~SBDHsp40 complex showing its hydrophilic nature. The presence of peaks below zero indicate that the RING protein lost its hydrophobicity once the complex was formed, thereby becomes more accessible to solvent.

4.3 Wet Lab Analysis

4.3.1Transformation and Expression Screening (Small scale expression)

The genes encoding the RING finger domain and substrate binding domain of Hsp40 (SBDHsp40) proteins were cloned into the pGEX-6P-2 and pQE30 protein expression vectors respectively as earlier shown in Figures 3.1 and 3.2. Within the natural environment, bacteria rarely if ever engulf DNA organically, although this does happen in a few cases spontaneously, due to the hydrophilic nature of DNA; hence bacterial transformation is an extremely important practice in cloning, protein expression and other molecular biology practices (Asif *et al.*, 2017; Shutz, 2014). Cells are therefore made 'competent', which manipulates the permeability of the cell wall or membrane to uptake foreign genetic matter. Various methods have been adopted for making cells using polyethylene glycol (PEG), dimethyl sulfoxide (DMSO) or divalent cations that include $CaCl_2$ and $MgCl_2$. In this project, BL21 and JM109 cells were used to express the pGEX-6P-2~RING finger domain protein and pQE30~SBDHsp40 respectively, which were prepared using the $CaCl_2$ chemical method. It is hypothesized that calcium ions act as a cation bridge between the phosphate backbone of DNA and the negative charges of phosphorylated lipid A in lipopolysaccharide (LPS); hence treating the cells with ice cold $CaCl_2$ solution facilitates DNA binding to the cell surface (Chang *et al.*, 2017). Exposure to temperature imbalances during transformation first with heat shocking causes porosity, allowing voluntary uptake of DNA, which is immediately followed by a cold shock treatment for closing of the pores. Both methods of chemical and physical manipulation work in synergy to produce the act of "artificial DNA internalization" (Asif *et al.*, 2017).

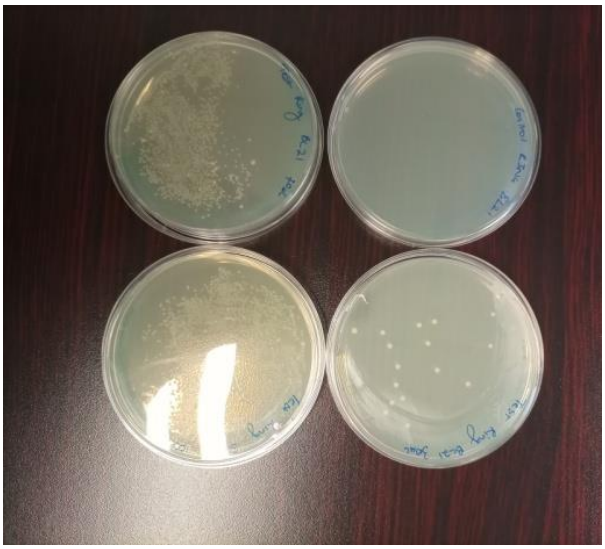


Figure 4.11: Transformation of BL21 bacterial cells with pGEX-6P-2~RING construct. BL21 competent cells were used and the above image depicts successful transformation. A – Test (70µL); B – Control; C – Test (100µL); D – Test (30µL).

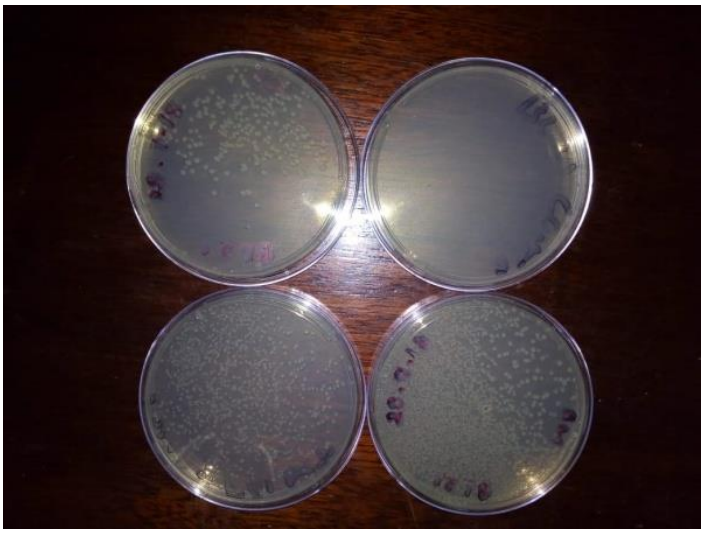


Figure 4.12: Transformation of JM109 bacterial cells with pQE30~SBDHsp40 construct. The image above shows successful transformation of pQE30-SBDHsp40 construct in JM109 competent cells. A – Test (30µL); B – Control; C – Test (70µL); D – Test (100µL).

The results in Figures 4.11 and 4.12 show successful transformation of the constructs observed by the presence of single colonies and clean control plates. This was achieved by incubation of the cells at 37˚C overnight on agar plates containing ampicillin, which was used as a selective or antibiotic resistance marker. From these plates, single colonies were screened for expression. The use of *E. coli* as a host is the most reliable, economical, rapid, and relatively easiest means of producing proteins in comparison to eukaryotic expression system that are generally more complex to use. In spite of this however, it is also highly unpredictable how much expression yield and solubility of protein the host can produce, hence it is imperative to screen multiple constructs of the target protein (Bird, 2011). This was also done to influence the decision on which clone would be used for large scale expression and purification. The expression profiles of the colonies selected were analyzed on a 15% SDS-PAGE gel indicated in Figures 4.13 and 4.14. Both RING and SBDHsp40 proteins showed expression identified by the presence of protein bands corresponding to the correct estimated sizes of 37kDa and 40kDa, respectively. However, leaky expression was observed in the SDS-PAGE profile for SBDHsp40. *Escherichia coli* cells contain the lactose (*lac*) operon which controls genetic transcription in bacteria. The operon is considered "turned off" or is down-regulated in the presence or absence of glucose and lactose respectively (Nielsen *et al.*, 2007). However, the expression of high protein levels may occur even in the absence of an inducer and presence of glucose, in other words due to the negative control of the *lac* promoter resulting in 'leakiness', also termed basal expression. This type of expression may eventually lead to plasmid instability, later causing poor target protein yields. This can however, be rectified by various means, one of the simplest being the addition of 0.2 – 1% glucose in enrich media containing peptone or tryptone for tighter control of the promoter (Rosano and Ceccarelli, 2014).

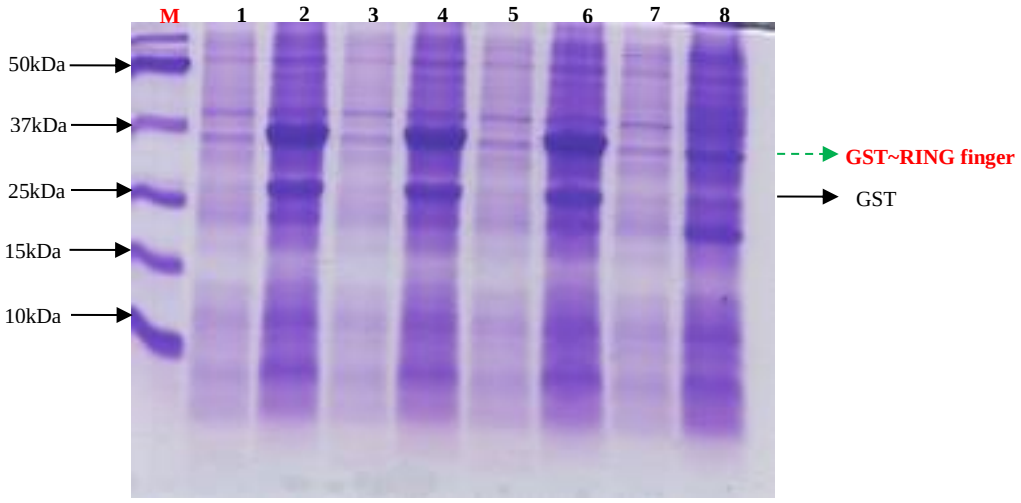


Figure 4.13: Expression screening of RING finger domains construct in BL21 cells. M – Molecular Marker; Lanes 1, 3, 5, 7 – Un-induced samples; Lanes 2, 4, 6, 8 – Induced samples

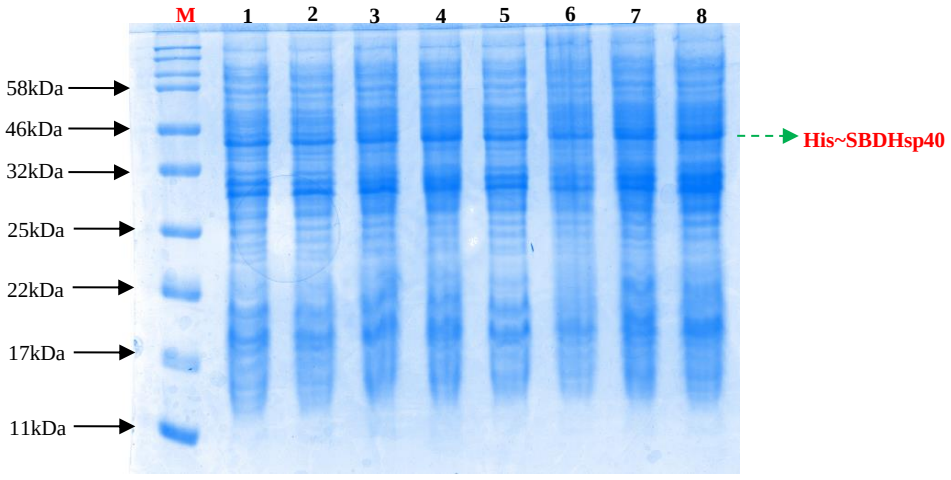


Figure 4.14: Expression screening of SBDHsp40 construct in JM109 cells. M – Molecular Marker; Lanes 1, 3, 5, 7 – Induced samples; Lanes 2, 4, 6, 8 – Un-induced samples

Using the expression screening gels as a point of reference, clones 1 for RING and 3 for SBDHsp40 were chosen for large scale protein expression.

4.3.2 Large Scale Protein Expression

The over-expression of recombinant proteins is an imperative step in modern day biochemistry, biotechnology or structural biology (Price *et al.*, 2011). The study of processes such as enzyme activity, ligand binding, as well as protein interactions, which are especially important for pharmaceuticals, are highly dependent on the production of recombinant proteins. Unfortunately, this experimental procedure is accompanied by several challenges. Despite the many advantages associated with the use of *E. coli* as an expression host, over-expression within this host may also face several disadvantages. These may include failure to express or poor yield, expression of proteins in inclusion bodies, or the production of insoluble, unfolded, partially folded or misfolded proteins (Francis and Page, 2010). These may be attributed to the lack of post-translational modifications that *E. coli* cannot perform or the highly reductive environment of the bacterial cytosol.

Overnight cultures of 100mL LB broth containing 100µg/mL ampicillin and 100µL of the best selected glycerol stocks were set up and scaled to 2L cultures the next morning. Figures 4.15 and 4.16 show successful expression of the RING finger domain and SBDHsp40 proteins, which was induced at an induction temperature of 25˚C using 0.5mM IPTG. Protein expression was induced once the OD_{600} 0.4–0.6 was reached as it is at this level (exponential growth phase) that most of the cells are most healthy, viable and actively dividing.

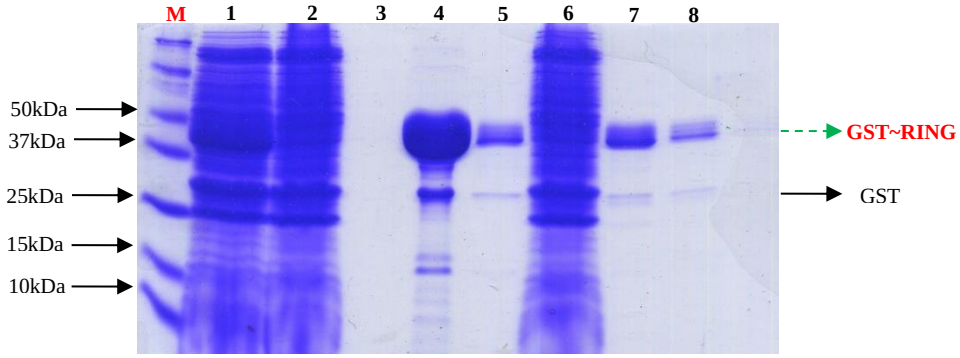


Figure 4.15: Recombinant large scale expression and affinity purification of GST~RING. M – Molecular Marker; Lane 1 – Total cell lysate; Lane 2 – Flowthrough; Lane 3 – Clean Wash; Lane 4 – Elute 1; Lane 5 – Elute 2; Lane 6 – Second Flowthrough; Lane 7 – Elute 1; Lane 8 – Elute 2

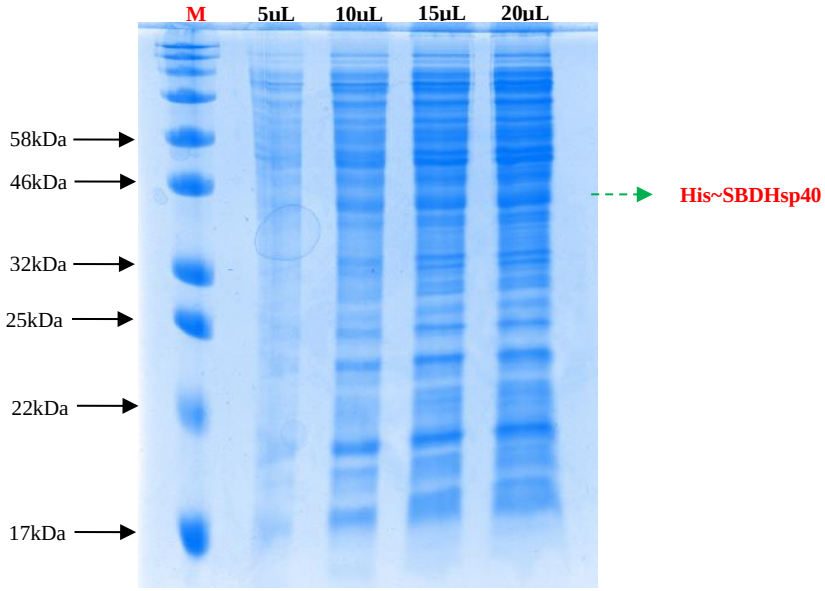


Figure 4.16: Recombinant large scale expression of His~SBDHsp40. M – Molecular Marker; Lanes represent amount of sample loaded in 5µL increments.

A succession of cultures was then set up to attain enough of the RING and SBDHsp40 proteins for purification and downstream processes.

4.3.3 Protein Purification

In the world of biopharmaceuticals, proteins are one of the most valuable products since they are essentially the key elements of life and perform pivotal roles in various cellular functions. It is therefore no surprise that the therapeutic protein industry is valued at over USD140 billion per year (Gomari *et al.*, 2020). It is extremely crucial that these proteins and other biomedical applications associated with their products are of a high level of purity, which may require several steps and complicated procedures that are not necessarily affordable. Protein affinity tags have now since been introduced as more straightforward and cheaper alternatives for purification purposes.

4.3.3.1. Purification of GST-tagged RING finger domain protein

The pGEX-6P-2~RING finger protein was purified to homogeneity using a three-step purification process that included affinity chromatography, cleavage and gel filtration. Affinity chromatography was performed by gravity using glutathione agarose beads. Clear lysate was poured down the column and the flowthrough collected. This is reflected in Figure 4.15 Lanes 1 and 2 respectively. The column was then washed with equilibration buffer to prevent the occurrence of non-specific binding and the last two drops were collected which is reflected in Lane 3. The GST-fusion protein was then eluted (Lane 4and 5) using 15mM reduced glutathione.

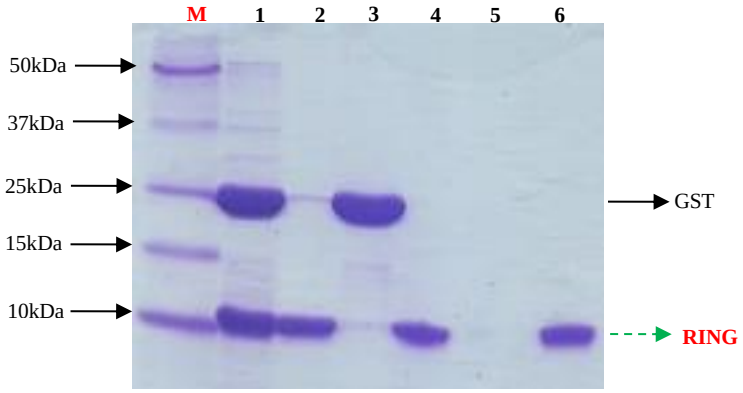


Figure 4.17: Cleavage and gel filtration of RING finger domain protein. M – Molecular Marker; Lane 1 – Cleaved protein; Lane 2 – RING finger protein; Lane 3 – Eluted GST; Lane 4 – Remaining RING finger protein; Lane 5 – Clean wash; Lane 6 – Gel filtrated RING finger domain protein

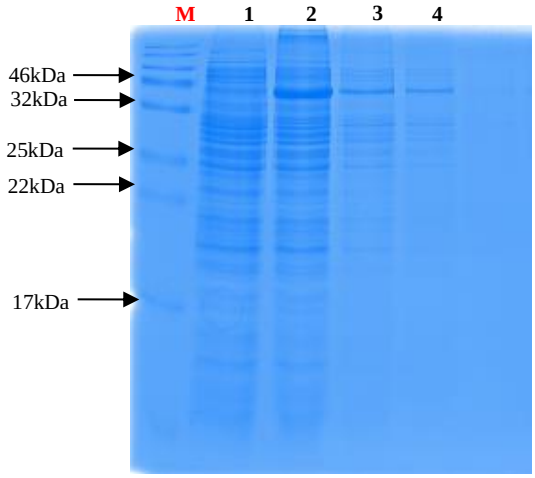


Figure 4.18: Affinity purification of substrate binding domain of Hsp40. M – Molecular Marker; Lane 1 – Flowthrough; Lane 2 – Lysate; Lane 3 – Elute 1; Lane 4 – Elute 2

The flowthrough from Lane 3 was poured down the column again (Lane 6) and the purification procedure repeated to recover as much protein as possible and these elutes are represented by Lanes 7 and 8.

Glutathione (GST) is a 26kDa 211 amino acid-containing protein that is expressed in pGEX vectors at the N-terminal and acts as a chaperone to promote protein folding and higher yields of recombinant protein expression and solubility (Harper and Speicher, 2011). GST possesses high affinity for glutathione and therefore the fusion protein is drawn to the immobilized glutathione substrate while impurities are washed out. To septate the GST tag from the protein of interest, an overnight dialysis was set up. This was done with the inclusion of PreScission protease, a fusion protein of human rhinovirus (HRV) 3C protease that recognises the sequence LeuGluValLeuPheGln/GluPro and performs specific cleavage between the Gln and Glu residues. Lane 1 in Figure 4.17 represents the cleaved protein. Following dialysis where cleavage had taken place, separation from the tag was achieved by pouring the sample down the glutathione agarose column again to collect the flowthrough containing the free RING finger protein (Lane 2). Lane 3 represents the GST tag, which was eluted from the affinity column, while Lane 4 represents all other collected remnants of the RING finger domain protein. Lastly, the column was cleansed using 2M NaCl (Lane 5).

The purified samples were then pooled together and subjected to gel filtration as the final polishing step for protein purification. Also known as size exclusion, this technique separated impurities based on the flow speed of the contaminants due to their different molecular weights or hydrodynamic radiuses (Mojsiewicz-Pieńkowska, 2014; Striegel, 2008). Lane 6 shows the protein purified to homogeneity by gel filtration to a final concentration of 5.7mg/mL, which was determined using a Nanodrop ND2000 spectrophotometer.

4.3.3.2 Purification of His-tagged SBDHsp40

Among all the protein tags that are currently available, histidine is still the most popular. His-tagged proteins can be purified in a one-step process that is bind-wash-elute procedure that can be performed under native or denaturing conditions (Spriestersbach *et al.*, 2015). Immobilized metal-affinity chromatography (IMAC) is the most prevalent method for polyhistidine-tagged recombinant proteins as it relies on the interactions between transition metal ions, in this case Ni^{2+}, immobilized on a solid matrix. Histidine, due to its electron donor groups, possesses a strong affinity for immobilized metal ion matrices by forming coordination bonds with the immobilized metal. Therefore, proteins containing repeating histidine residues are drawn to IMAC column matrices and easily be eluted by the addition of free imidazole to elution buffer (Bornhorst and Falke, 2000). The SBDHsp40 protein was purified using a Nickel-NTA column by gravity and the elutes, which were performed using 400mM Imidazole are shown in Lanes 3 and 4 of Figure 4.18. Unfortunately, the protein was not entirely purified to homogeneity despite several attempts. This was due to the significant presence of non-specific contaminants within the eluted samples.

4.3.4 Protein Characterisation

Protein quality assurance is a key aspect in several structural, proteomic, pharmaceutical, biotechnology, molecular biology, and medicine as well as other related science downstream processes. Proteins differ from each other in terms of their different characteristics that include size, physiochemical properties, and molecular structure. Hence, analysis and characterization of recombinant macromolecules is essential, mostly done by identification and separation (Oyinloye, 2017). Assessment of purity and biological or enzymatic activity are the most important analyses

determined. Various experimental methods are employed for the determination of protein structure and function. In this project, protein characterization included Fourier Transform Infrared (FTIR) spectroscopy and Raman spectroscopy.

4.3.4.1. Fourier Transform Infrared Spectroscopy

One of the most popular techniques for protein, protein-membrane and peptide structural characterization is Fourier Transform Infrared (FTIR) Spectroscopy. The method is based on the wide spectral range within a single spectrum that considers frequencies of vibrations of the chemical groups that exist in the polypeptide chain and all organic or inorganic compounds (Tatulian, 2013). Information of individual functional groups of molecules can be obtained from FTIR spectra because of the unique chemical compositions of every structural or functional biomolecule, which are able to absorb at unique or fingerprint vibrational frequencies (Tatulian, 2019). The FTIR spectrometer is sensitive enough to sense the conformational changes that occur when intermolecular interactions or functional transitions take place.

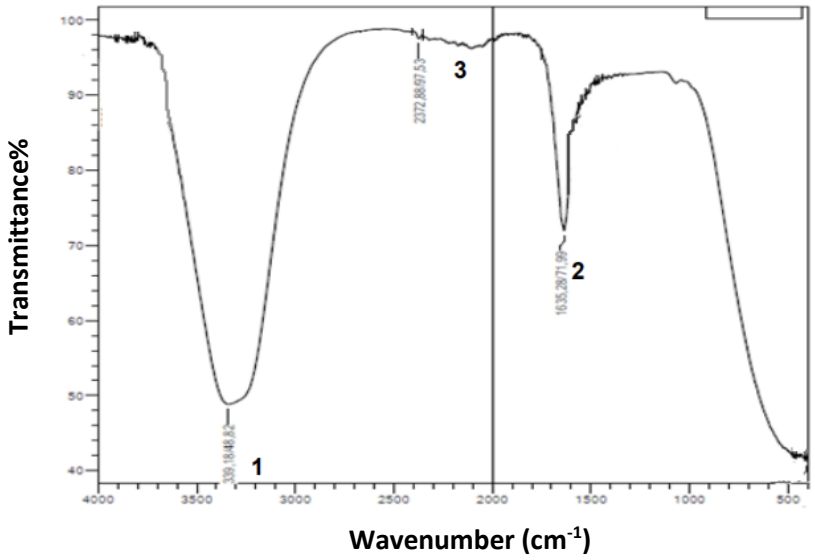


Figure 4.19: Fourier Transform Infrared (FTIR) Spectroscopy of RING finger domain protein. The functional groups were identified using FTIR where 1- 3 indicates the most relevant peaks for analyses.

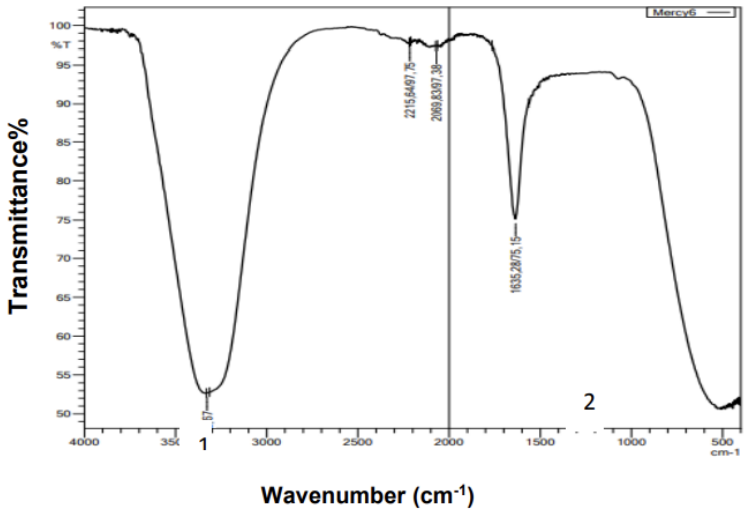


Figure 4.20: Fourier Transform Infrared (FTIR) Spectroscopy of SBDHsp40 protein. Two prominent peaks were identified and analyzed within the protein.

The functional groups determined for the RING finger domain protein are shown in Figure 4.19 where the prominent peak that falls in the 3339–3348cm^{-1} (Peak 1) shows the presence of an N-H stretch vibration (Merck, 2021). This may be indicative of an amide group of carboxyl side chains. Peak 2 within the 1635–1671cm^{-1} range represents C=O stretching vibration, also referred to as the Amide I band. This peak is also a representation of the secondary structures of the protein. Once peaks are present between the 1647–1658cm^{-1} and 1632- 1638cm^{-1} ranges, this shows the presence of an α–helix and antiparallel β–sheet as seen in the spectrum (Tatulian, 2019). The presence of both these secondary structural elements were seen during the homology model prediction of the protein (Figure 4.1A). Lastly the region between 2000 – 2397 cm^{-1} (labelled 3) is an area that contains alkynes, C≡C or C≡N groups (Merck, 2021).

Figure 4.20 shows the FTIR spectrum generated for the SBDHsp40 protein. The spectrum obtained is a Mid-IR infrared spectrum as the wavelengths of all the peaks fall between 400–4000cm^{-1}. The peak at 3300cm^{-1} (Peak 1) shows an alkyne with carbon hydrogen bonds C-H, O-H and N-H stretch functional groups because they are found in the single bond region between 2500-4000 cm^{-1}. This is considered an amide II band because it results from N-H bending and C-N stretching vibrations. The 1635.28 cm^{-1} peak (Peak 2) falls in the double bond region as it is between the 1500–2000 cm^{-1} range; this indicates the presence of C=C and C=O functional groups. In this region, the peak is considered an Amide I major band of the protein infrared spectrum because of the C=O stretching vibrations (Nandiyanto *et al.*, 2019). The region measured at 2215.54 cm^{-1} and 2069.83 cm^{-1} fall in the triple bond region between 2000cm^{-1} and 2500cm^{-1} and signify the presence of C≡C bond or C≡N bonds. There was no fingerprint region under the 600cm^{-1} wavenumber. In overall, the protein is made up of Amide I bands that show that the analyte is structurally made up

of carbon-hydrogen bonds, hydroxyl and amine functional groups in the single bond regions and carbonyl groups shown by the presence of aldehydes in the double bond region (Silva *et al.*, 2020).

4.3.4.2. Raman Spectroscopy

Raman Spectroscopy is a technique that combines the Raman spectrometer and a microscope, whose principle is based on the inelastic scattering of light (Mitsutake *et al.*, 2019). Raman spectroscopy can analyze different matrices and provide information on the unique secondary structure fingerprints of proteins such as α-helices, β-sheets, random coils and turns as well as amino acids such as phenylalanine, tyrosine and tryptophan (Kuhar *et al.*, 2021). Since misfolded proteins are associated with the occurrence of various life-threatening diseases, it is therefore essential to gather information on the shape and structure of proteins. Raman spectra, although insufficient to determine this as a single technique, is a good additional feature point for understanding the characteristics of protein structure and shape. Raman imaging is used to generate detailed chemical images that identify various chemical species and their localization. The spectroscopic images allow for the study of compound distribution on the spectra and the characteristics of the different compounds that are related to the sample being analyzed (Mitsutake *et al.*, 2019). These images provide both spatial and spectra information.

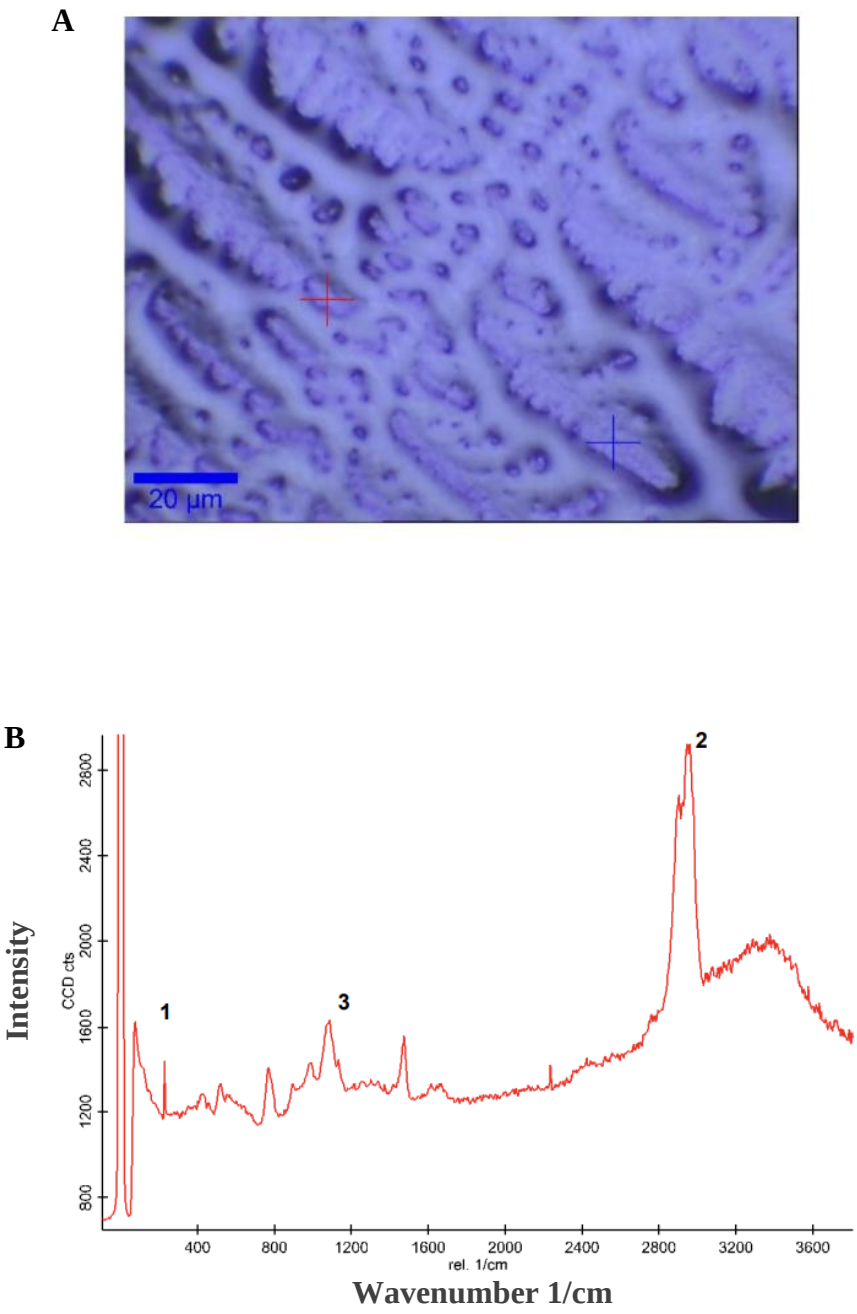


Figure 4.21: Raman imaging (A) and Raman spectroscopy (B) of the RING finger domain protein. The red and blue cross indicate spectra generated reveals the chemical structure and molecular interactions of the protein. Peaks labelled 1–3 indicate those relevant for the RING finger domain protein.

Raman imaging for the RING finger domain protein was generated using two laser powers indicated by two crosses; 1mW for the red cross and 2mW for the blue cross (Figure 4.21A). The image corresponded to two spectra which had similar peaks, indicating that the protein was homogenous. The spectra generated for the blue cross was therefore used for further analysis (Figure 4.21B). On the other hand, Figure 4.22A shows a Raman image with three marked regions (red, blue, and green) determined for the SBDHsp40 protein. Integration of the regions gave rise to three identical peaks, also demonstrating the homogenous nature of the sample. Figure 4.22B shows the spectrum generated for the SBDHsp40 protein.

The prominent peak at the beginning of the RING finger domain protein spectrum (Figure 4.21B) represents the baseline, while Peak 1 indicates possible lattice vibrations or C-C aliphatic chains (Adamo, 2021). The next prominent peak (Peak 2), which is found between 2800 and 3000cm^{-1} shows the presence of CH=CH stretching. The final peaks indicated by 3 range from 400 to 1600 cm^{-1} and show that there are α–helices and β–sheets present within the protein (Rygula *et al.*, 2013). This result coincided with the FTIR (Figure 4.19) and homology modelling (Figure 4.1) results, substantiating the two α-helices and β-sheets present within the protein. For the SBDHsp40 protein, the first peak at 600cm^{-1} reveals the presence of a C=S bond with strong intensity, while Peak 2 was observed at 1000cm^{-1} showing that the protein possesses aromatic rings with high intensity. Peak 3 positioned at 2400cm^{-1} shows that the protein contains thiol groups, while the last peak indicates the presence of Amide groups with moderate intensity (Gyakwaa *et al.*, 2020). Overall, this technique identified a total of 4 functional groups within the SBDHsp40 protein sample.

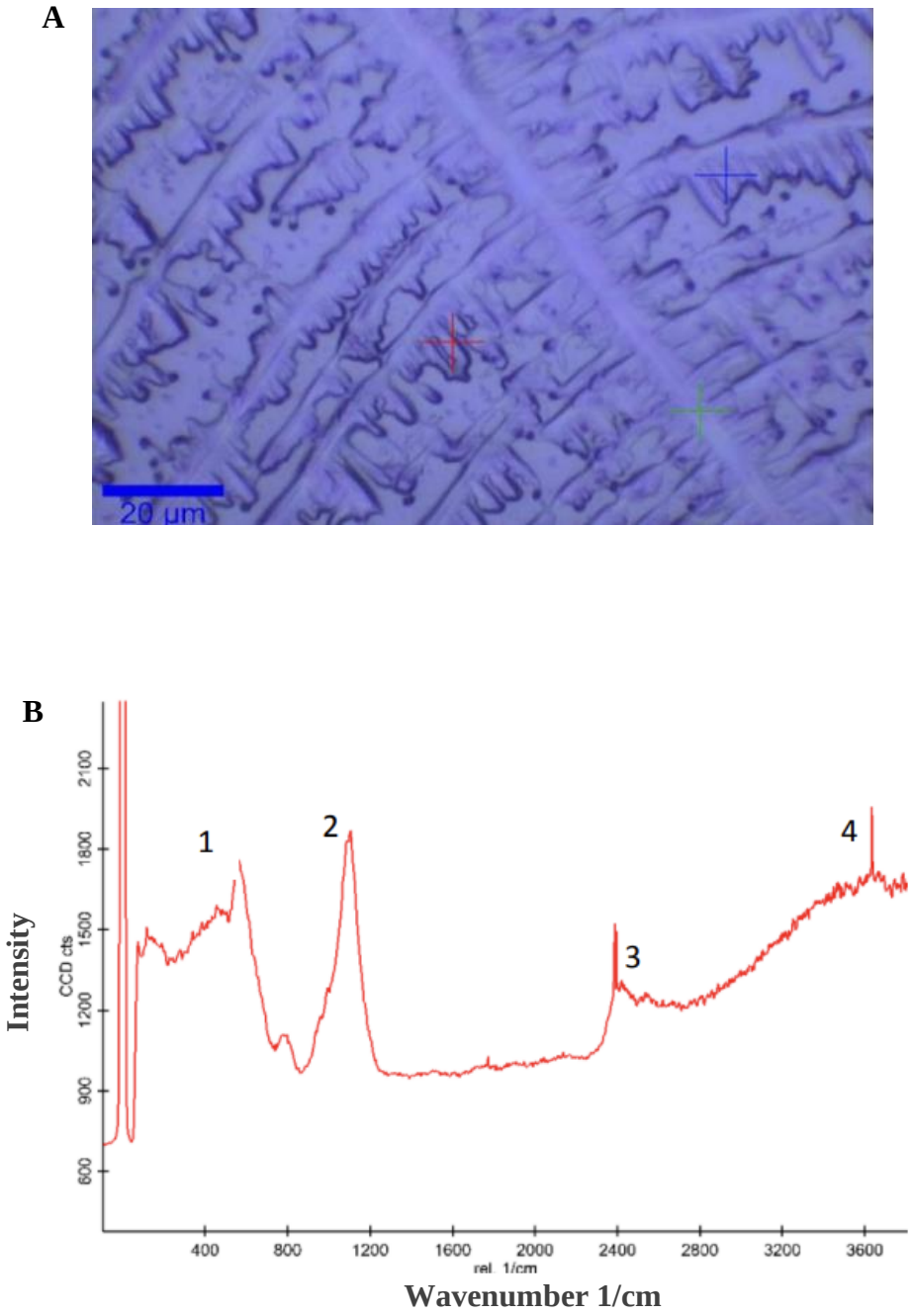


Figure 4.22: Raman imaging (A) and Raman spectroscopy (B) of the SBDHsp40 protein. The Raman image shows the distribution of compounds that are reflected on the subsequent spectrum.

CHAPTER FIVE: GENERAL DISCUSSION, CONCLUSION AND FUTURE RECOMMENDATIONS

Cells contain a group of chaperones that are responsible for the folding, maturation and maintenance of the native 3-dimensional conformations of cellular proteins under normal and in particular stress conditions. Consequently, it has been found that chaperones are implicated in ubiquitin-mediated proteasomal degradation. The activities surrounding chaperone-mediated ubiquitination constitutes a pervasive protein-protein interaction network and is highly prominent in events that may activate signalling pathways for diseases such as cancer. Cancer cells are subjected to a number of abnormal cell stresses and tumorigenicity requires active oncoproteins (Banerji, 2009). Cancer is a condition propelled by the irrepressible growth of cells that eventually infiltrate proximate tissues to become tumorous, thereby causing physical abnormalities that lead to high causes of morbidity and death globally.

Various treatment options for the disease have evolved over the years but has still been an area of concern till date as there are no current treatment options that come without side effects. Protein-protein interactions therefore come onto the scene as a novel solution directed towards these challenges. Although their mechanisms of action remain largely unexplained and under-explored and are regarded as generally ‘undruggable’, it is widely agreed that PPIs hold tremendous potential as more specific and better drug candidates simply because most drug targets are in the form of enzymes, receptors, and ion channels (Mabonga and Kappo, 2019).

In cancer, PPIs form indication hubs and nodes for the transmission of physiological signals along various molecular networks for a biological response, which promotes the occurrence of metastasis, invasion, tumour progression or tumorigenesis (Ivanov *et al.*, 2013). Hence, the

disruption of PPIs presents as an effective and innovative approach to inhibit oncogenic signal transmission. This study aimed at substantiating the proposed interaction between RING finger domain of RBBP6 and SBDHsp40, whose binding has been hypothesized to play a crucial role in cancer.

Using *in silico* techniques, the current study validated this claim. The proteins were first modelled using the built-in Chimera function for MODELLER and validated by a Ramachandran plot generated via PROCHECK. This revealed that both protein models had not violated any stereochemical parameters as both models exhibited over 90% of their residues in the most favoured and favoured regions. Thereafter, molecular docking ensued of the two proteins where the complex with the lowest binding score of -8.4kcal/mol was selected for further analysis. The lower the binding score, the more it is assumed that this is the best conformation the complex demonstrates in nature (Masamba *et al.*, 2020). MD simulations analyzed system stability, flexibility and compactness, the formation of hydrogen bonds and how accessible the surface area of the systems is to the solvent. In comparison to the apo RING finger domain, the RING~SBDHsp40 complex showed a system with high levels of fluctuation, which in literature coincides with the zinc coordination site within the RING finger domain (Kappo, 2009). The fluctuation observed reached eventual stability (RoG) and demonstrated the formation of hydrogen bonds, which was mirrored by the RMSD and RMSF results. Most notable among the analyses done were the SASA results that showed that the RING~SBDHsp40 complex displays more area accessible to solvent than the RING finger domain due to a loss of hydrophobicity. These results were corroborated by the generation of Kyte and Dolittle plots that revealed that the RING protein is hydrophobic in nature and loses it hydrophobicity once in complex with SBDHsp40. According to Lu and co-workers, the hydrophobic nature of PPIs makes it difficult to target them as

therapeutic drugs (Lu *et al.*, 2020). This result therefore reveals the 'druggability' of this PPI as a potential druggable target in the eventual development of novel treatment for the management of cancer.

In vitro analyses were performed next, which began with the overexpression of the RING and SBDHsp40 proteins in BL21 and JM109 competent bacterial cells respectively. The RING protein was then purified to homogeneity using affinity chromatography by gravity using a GST agarose column, cleavage by overnight dialysis and incorporation of 3C protease enzyme and finally gel filtration also known as size exclusion chromatography. The SBDHsp40 protein, on the other hand, was only semi-purified by affinity chromatography using a His-select Nickel NTA column. This is therefore the first recommendation for future studies. Despite various efforts to purify the protein to homogeneity using affinity chromatography, the protein was only purified to a certain limited extent. This may have been caused by poor binding of the His-tagged protein to the column, high and tight occurrence of non-specific interactions or a pH or imidazole concentration problem. Future studies should therefore invest in optimizing purification conditions or using other purification methods such as ion exchange chromatography before the use of gel filtration to allow for proper separation of the contaminants.

Structural characterization of the proteins was performed by Fourier Transform Infrared (FTIR) Spectroscopy and Raman Spectroscopy that analysed the functional groups and structural fingerprints of the proteins. Both results confirmed the secondary structural elements of both the RING and SBDHsp40 proteins as shown during homology modelling. However, these methods could have been substantiated with the use of other molecular techniques such Circular Dichroism (CD), which is an excellent spectroscopic tool that considers differential absorption of circular polarized light. Therefore, another recommendation for future studies concerning this study is the

use of CD spectroscopy to not only determine the folding state, conformation, or stability of the expressed and purified proteins but also confirm the secondary structure components of the proteins, such which include α-helices, β-turns and sheets as well as 'other' disordered structures (Greenfield, 2006; Micsonai *et al.*, 2018).

The main aim of this study would have finalized by performing biophysical interaction studies of the protein complex using techniques such as Isothermal Titration Calorimetry (ITC) that measures and characterizes PPI thermodynamic properties that reveal entropy, enthalpy, association constant, stoichiometry of interaction (n) and free energy, Surface Plasmon Resonance (SPR), which determines interactions using refractive index changes that give rise to dissociation constant (K_D) or Microscale Thermophoresis (MST) analysis that quantifies binding affinities through fluorescence variations occurring from infrared laser-induced temperature gradients (Mabonga *et al.*, 2021; Masamba 2017). However, due to extremely stringent time constraints and the heavy impacts of the COVID-19 pandemic that made working conditions in the laboratory extremely arduous, this part of the work was not completed, and it is therefore recommended that efforts directed towards this aspect should be performed in future, not only because yeast two-hybrid system possess the ability to produce false positives but also because *in silico* characterizations are, at best, purely predictions.

Lastly, one of the main purposes for identifying PPIs, particularly in disease pathways is for the purpose of novel drug discovery, which in most cases, is in the form of small molecules. Therefore, this study recommends the generation of a pharmacophore model to identify potential inhibitors and the repetition of both *in silico* and *in vitro* experiments to analyse and validate the identified inhibitors. It is imperative as well that the strength of the interactions, residues involved in the binding, and subsequent mechanisms of action be determined. Performance of these assays may

then form a substantive basis for the application of the RING-SBDHsp40 PPI in the development of novel treatment against cancer.

REFERENCES

Adamo T. (2021). Understanding Raman Spectroscopy Principles and Theory Basic Raman Instrumentation. [online] Traces Centre, pp.1–5. Available at: https://www.utsc.utoronto.ca/~traceslab/PDFs/raman_understanding.pdf [Accessed Nov 2021].

Adekiya T.A., Aruleba R.T., Khanyile S., Masamba P., Oyinloye B.E. and Kappo A.P. (2018). Structural analysis and epitope prediction of MHC class-1-chain related protein-a for cancer vaccine development. *Vaccines* **6**(1): 1.

Ali A. and Bagdi. (2015). An Overview of Protein-Protein Interaction. *Current Chemical Biology* **9**: 53-65.

Anand P., Kunnumakara A.B., Sundaram C., Harikumar K.B., Tharakan S.T., Lai O.S., Sung B. and Aggarwal B.B. (2008). Cancer is a preventable disease that requires major lifestyle changes. *Pharmaceutical Research* **25**(9): 2097-2116.

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.

Balagurumoorthy P., Sakamoto H., Lewis M.S., Zambrano N., Clore G.M., Gronenborn A.M., Appella E. and Harrington R.E. (1995). Four p53 DNA-binding domain peptides bind natural p53-response elements and bend the DNA. *Proceedings of the National Academy of Sciences* **92**(19): 8591–8595.

Banerji U. (2009). Heat shock protein 90 as a drug target: some like it hot. *Clinical Cancer Research* **15**(1): 9-14.

Bird L.E. (2011). High throughput construction and small scale expression screening of multi-tag vectors in Escherichia coli. *Methods* **55**(1): 29-37.

Bornhorst J.A. and Falke J.J. (2000). Purification of Proteins Using Polyhistidine Affinity Tags. *Methods in Enzymology* **326**: 245–254.

Butler H.J., Ashton L., Bird B., Cinque G., Curtis K., Dorney J., Esmonde-White K., Fullwood N.J., Gardner B., Martin-Hirsch P.L., Walsh M.J., McAinsh M.R., Stone N. and Martin F.L. (2016). Using Raman spectroscopy to characterize biological materials. *Nature Protocols* **11**(4): 664–687.

Causier B. and Davies B. (2002). Analysing protein-protein interactions with the yeast two-hybrid system. *Plant Molecular Biology* **50**(6): 855-870.

Chang A.Y., Chau V., Landas J.A. and Pang Y. (2017). Preparation of calcium competent Escherichia coli and heat-shock transformation. *JEMI Methods* **1**: 22-25.

Chasapis C.T., Loutsidou A.K., Orkoula M.G. and Spyroulias G.A. (2010). Zinc binding properties of engineered RING finger domain of Arkadia E3 ubiquitin ligase. *Bioinorganic Chemistry and Applications* **2010**.

Cusick M.E., Klitgord N., Vidal M. and Hill D.E. (2005). Interactome: gateway into systems biology. *Human Molecular Genetics* **14**(suppl_2): R171-R181.

Dalton J.A. and Jackson R.M. (2007). An evaluation of automated homology modelling methods at low target–template sequence similarity. *Bioinformatics* **23**(15): 1901-1908.

Deng L., Meng T., Chen L., Wei W. and Wang P. (2020). The role of ubiquitination in tumorigenesis and targeted drug discovery. *Signal Transduction and Targeted Therapy* **5**(1): 1–28.

Dlamini Z., Ledwaba T., Hull R., Naicker S. and Mbita Z. (2019). RBBP6 Is Abundantly Expressed in Human Cervical Carcinoma and May Be Implicated in Its Malignant Progression. *Biomarkers in Cancer* **11**: 1–10.

Du W. and Searle J. (2009). The Rb Pathway and Cancer Therapeutics. *Current Drug Targets* **10**(7): 581–589.

Durham E., Dorr B., Woetzel N., Staritzbichler R. and Jens M. (2009). Solvent accessible surface area approximations for rapid and accurate protein structure prediction. *Journal of Molecular Modeling* **15**: 1093–1108.

Esser C., Alberti S. and Höhfeld J. (2004). Cooperation of molecular chaperones with the ubiquitin/proteasome system. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research* **1695**(1-3): 171-188.

Fan J., Fu A. and Zhang L. (2019). Progress in molecular docking. *Quantitative Biology* 1-7.

Ferlay J., Soerjomataram I., Dikshit R., Eser S., Mathers C., Rebelo M., Parkin D.M., Forman D. and Bray F. (2015). Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *International Journal of Cancer* **136**(5): E359-E386.

Ferlay J., Colombet M., Soerjomataram I., Parkin D.M., Piñeros M., Znaor A. and Bray F. (2021). Cancer statistics for the year 2020: An overview. *International Journal of Cancer.*

Francis D.M. and Page R. (2010). Strategies to optimize protein expression in *E. coli*. *Current Protocols in Protein Science* **61**(1): 5-24.

Garner A.L. and Janda K.D. (2011). Protein-protein interactions and cancer targeting the central dogma. *Current Topics in Medicinal Chemstry* **11**(3): 258-280.

Gomari M.M., Saraygord-Afshari N., Farsimadan M., Rostami N., Aghamiri S. and Farajollahi M.M. (2020). Opportunities and challenges of the tag-assisted protein purification techniques: Applications in the pharmaceutical industry. *Biotechnology Advances* 107653.

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

Gyakwaa F., Aula M., Alatarvas T., Vuolio T., Shu Q., Huttula M. and Fabritius T. (2020). Application of Raman Spectroscopy for Characterizing Synthetic Non-Metallic Inclusions Consisting of Calcium Sulphide and Oxides. *Applied Sciences* **10**(6): 2113.

Hansson T., Oostenbrink C. and Van Gunsteren W. (2002). Molecular dynamics simulations. *Current Opinion in Structural Biology* **12**(2): 190-196.

Harper S. and Speicher D.W. (2011). Purification of proteins fused to glutathione S-transferase. *Methods in Molecular Biology* **681**: 259–280.

Hassanpour S.H. and Dehghani M. (2017). Review of cancer from perspective of molecular. *Journal of Cancer Research and Practice* **4**(4): 127-129.

Hollingsworth S.A. and Karplus P.A. (2010). A fresh look at the Ramachandran plot and the occurrence of standard structures in proteins 271-283.

Hong L. and Lei J. (2009). Scaling law for the radius of gyration of proteins and its dependence on hydrophobicity. *Journal of Polymer Science Part B: Polymer Physics* **47**(2): 207-214.

Hospital A., Goñi J.R., Orozco M. and Gelpí J.L. (2015). Molecular dynamics simulations: advances and applications. *Advances and applications in bioinformatics and chemistry: AABC* **8**: 37.

Ivanov A.A., Khuri F.R. and Fu H. (2013). Targeting protein–protein interactions as an anticancer strategy. *Trends in Pharmacological Sciences* **34**(7): 393-400.

Kaiser J. (2021). Cancer. *Encyclopedia of Evolutionary Psychological Science* 871-874.

Kandias N.G., Chasapis C.T., Bentrop D., Episkopou V. and Spyroulias G.A. (2009). High yield expression and NMR characterization of Arkadia E3 ubiquitin ligase RING-H2 finger domain. *Biochemical and Biophysical Research Communications* **378**(3): 498–502.

Kappo A.M. (2009). *Solution structure of the RING finger domain from the human splicing-associated protein RBBP6 using heteronuclear Nuclear Magnetic Resonance (NMR) Spectroscopy.* University of the Western Cape.

Kappo M.A., Eiso A.B., Hassem F., Atkinson R.A., Faro A., Muleya V., Mulaudzi T., Poole J.O., McKenzie J.M., Chibi M. and Moolman-Smook J.C. (2012). Solution structure of RING finger-like domain of retinoblastoma-binding protein-6 (RBBP6) suggests it functions as a U-box. *Journal of Biological Chemistry* **287**(10): 7146-7158.

Kar G., Gursoy A. and Keskin O. (2009). Human Cancer Protein-Protein Interaction Network: A Structural Perspective. *PLoS Computational Biology* **5**(12): e1000601.

Karplus M. and Mccammon J.A. (2002). Molecular dynamics simulations of biomolecules. *Nature Structural Biology* **9**(9): 646-652.

Kaul G. and Thippeswamy H. (2011). Role of heat shock proteins in diseases and their therapeutic potential. *Indian Journal of Microbiology* **51**(2): 124-131.

Kosarev P., Mayer K.F. and Hardtke C.S. (2002). Evaluation and classification of RINGfinger domains encoded by the Arabidopsis genome. *Genome Biology* **3**(4): 1-12.

Krishna S.S., Majumdar I. and Grishin N.V. (2003). Structural classification of zinc fingers: Survey and summary. *Nucleic Acids Research* **31**(2): 532–550.

Kuhar N., Sil S. and Umapathy S. (2021). Potential of Raman spectroscopic techniques to study proteins. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* **258:** 119712.

Kuzmanic A. and Zagrovic B. (2010). Determination of ensemble-average pairwise root mean-square deviation from experimental B-factors. *Biophysical Journal* **98**(5): 861-871.

Kyte J and Doolittle R F. (1982). A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology* **157**(1): 105-132.

Laskowski R., Macarthur M. and Thornton J. (2006). PROCHECK: validation of protein-structure coordinates.

Lianos G.D., Alexiou G.A., Mangano A., Mangano A., Rausei S., Boni L., Dionigi G. and Roukos D.H. (2015). The role of heat shock proteins in cancer. *Cancer Letters* **360**(2): 114-118.

Lobanov M., Bogatyreva N. and Galzitskaya O. (2008). Radius of gyration as an indicator of protein structure compactness. *Molecular Biology* **42**(4): 623-628.

Lu H., Zhou Q., He J., Jiang Z., Peng C., Tong R. and Shi J. (2020). Recent advances in the development of protein–proteininteractions modulators: mechanisms and clinical trials. *Signal Transduction and Targeted Therapy* **5**: 213.

Mabonga L. and Kappo A.P. (2019). Protein-protein interaction modulators: advances, successes and remaining challenges. *Biophysical Reviews* **11**: 559–581.

Mabonga L., Masamba P., Basson A.K., Kappo A.P. (2021). Microscale thermophoresis analysis of the molecular interaction between small nuclear ribonucleoprotein polypeptide G and the RING finger domain of RBBP6 towards anti-cancer drug discovery. *American Journal of Translational Research* **13**(11): 12775-12785.

Matthews J.M. and Sunde M. (2002). Zinc Fingers--Folds for Many Occasions. *IUBMB Life (International Union of Biochemistry and Molecular Biology: Life)* **54**(6): 351–355.

Mehla J., Caufield J.H., Uetz P. (2015). The yeast two-hybrid system: a tool for mapping protein–protein interactions. *Cold Spring Harbor Protocols* **2015**(5): pdb-top083345.

Mansoori B., Mohammadi A., Davudian S., Shirjang S. and Baradaran B. (2017). The different mechanisms of cancer drug resistance: a brief review. *Advanced Pharmaceutical Bulletin* **7**(3): 339.

Masamba P. (2017). Structural Characterization of the interaction between divalent cations and the novel Schistosoma mansoni Universal Stress Protein G4LZI3. (Masters dissertation, University of Zululand).

Masamba P., Munsamy G. and Kappo A.P. (2020). Closing the gap: an atomistic structural and functional perspective of S. mansoni universal stress G4LZI3 protein in complex with phenolic compounds. *Current Drug Discovery Technologies.*

Mbita Z., Meyer M., Skepu A., Hosie M., Rees J. and Dlamini Z. (2011). De-regulation of the RBBP6 isoform 3/DWNN in human cancers. *Molecular and Cellular Biochemistry* **362**(1- 2): 249–262.

Merck (2021). IR Spectrum Table & Chart. [online] SigmaAldrich. Available at: https://www.sigmaaldrich.com/ZA/en/technical-documents/technical-article/analyticalchemistry/photometry-and-reflectometry/ir-spectrum-table [Accessed 29 Oct. 2021].

Micsonai A., Wien F., Bulyáki É., Kun J., Moussong É., Lee Y.H., Goto Y., Réfrégiers M. and Kardos J. (2018). BeStSel: a web server for accurate protein secondary structure prediction and fold recognition from the circular dichroism spectra. *Nucleic Acids Research* **46**(W1): W315-W322.

Mitra A., Shevde L.A. and Samant R.S. (2009). Multi-faceted role of HSP40 in cancer. *Clinical and Experimental Metastasis* **26**(6): 559-567.

Mitsutake H., Poppi R.J. and Breitkreitz M.C. (2019). Raman Imaging Spectroscopy: History, Fundamentals and Current Scenario of the Technique. *Journal of the Brazilian Chemical Society* **30**(11); 2.

Moela P., Choene M.M.S. and Motadi L.R. (2014). Silencing RBBP6 (Retinoblastoma Binding Protein 6) sensitises breast cancer cells MCF7 to staurosporine and camptothecin-induced cell death. *Immunobiology* **219**(8): 593–601.

Mojsiewicz-Pieńkowska K. (2014). Size Exclusion Chromatography a Useful Technique For Speciation Analysis of Polydimethylsiloxanes. In *Green Chromatographic Techniques* (pp. 181-202). Springer, Dordrecht.

Nair P.C. and Miners J.O. (2014). Molecular dynamics simulations: from structure function relationships to drug discovery. *In silico Pharmacology* **2**(1): 1-4.

Nandiyonto A.B.D., Oktiani R. and Ragadhita R. (2019). How to Read and Interpret FTIR Spectroscope of Organic Material. *Indonesian Journal of Science and Technology* **4**(1): 97-118.

Nia H.T., Munn L.L. and Jain R.K. (2020). Physical traits of cancer. *Science* **370**(6516).

Nielsen B.L., Willis V.C. and Lin C.Y. (2007). Western blot analysis to illustrate relative control levels of the lac and ara promoters in Escherichia coli. *Biochemistry and Molecular Biology Education* **35**(2): 133-137.

Oyinloye B.E. (2017). Structural Characterization of the Switch domain regions of Schistosoma mansoni druggable Adenylate cyclase-stimulating Gα-protein (Doctoral dissertation, University of Zululand).

Olotu F.A., Agoni C., Adeniji E., Abdullahi M. and Soliman M.E. (2019). Probing gallate-mediated selectivity and high-affinity binding of epigallocatechin gallate: A way-forward in the

design of selective inhibitors for anti-apoptotic bcl-2 proteins. *Applied Biochemistry and Biotechnology* **187**(3): 1061-1080.

Pagadala N.S., Syed K. and Tuszynski J. (2017). Software for molecular docking: a review. *Biophysical reviews* **9**(2): 91-102.

Patni A.P., Joseph J.P., Macrin D. and Devi A. 2021. Progress in human embryonic stem cell research and aging. *Stem Cells and Aging.* Elsevier.

Peele K.A., Durthi C.P., Srihansa T., Krupanidhi S., Ayyagari V.S., Babu D.J., Indira M., Reddy A.R. and Venkateswarulu T. (2020). Molecular docking and dynamic simulations for antiviral compounds against SARS-CoV-2: A computational study. *Informatics in Medicine Unlocked* **19:**100345.

Pilleron S., Soto-Perez-De-Celis E., Vignat J., Ferlay J., Soerjomataram I., Bray F. and Sarfati D. (2021). Estimated global cancer incidence in the oldest adults in 2018 and projections to 2050. *International Journal of Cancer* **148**(3): 601-608.

Price W.N., Handelman S.K., Everett J.K., Tong S.N., Bracic A., Luff J.D., Naumov V., Acton T., Manor P. and Xiao R. (2011). Large-scale experimental studies show unexpected amino acid effects on protein expression and solubility in vivo in E. coli. *Microbial Informatics and Experimentation* **1**(1): 1-20.

Pugh D.J., Eiso A.B., Faro A., Lutya P.T., Hoffmann E. and Rees D.J.G. (2006). DWNN, a novel ubiquitin-like domain, implicates RBBP6 in mRNA processing and ubiquitin-like pathways. *BMC Structural Biology* **6**(1): 1-12.

Roberts K., Alberts B., Johnson A., Walter P. and Hunt T. (2002). Molecular biology of the cell. *New York: Garland Science* **32**(2).

Rosano G.L. and Ceccarelli E.A. (2014). Recombinant protein expression in Escherichia coli: advances and challenges. *Frontiers in Microbiology* **5:** 172.

Rygula A., Majzner K., Marzec K.M., Kaczor A., Pilarczyk M. and Baranska M. (2013). Raman spectroscopy of proteins: a review. *Journal of Raman Spectroscopy* **44**(8): 1061–1076.

Sane S. and Rezvani K. (2017). Essential Roles of E3 Ubiquitin Ligases in p53 Regulation. *International Journal of Molecular Sciences* **18**(2): 1–20.

Sawyer A. (2020). Developing drugs for the 'undruggable'. *Biotechniques* **69**(4).

Seigneuric R., Mjahed H., Gobbo J., Joly A.-L., Berthenet K., Shirley S. and Garrido C. (2011). Heat shock proteins as danger signals for cancer detection. *Frontiers in Oncology* **1**: 37.

Sellers W.R. and Kaelin W.G. (1997). Role of the retinoblastoma protein in the pathogenesis of human cancer. *Journal of Clinical Oncology* **15**(11): 3301–3312.

Shutz G. (2014). Functional characterization of a novel bacterial peroxidockerin with homology to human heme peroxidases. *Publikationsserver der Universitätsbibliothek Bodenkultur Wien.*

Silva P. and Madusanka T. (2019). The Role of Heat Shock Proteins in Cancer. *International Research Journal of Natural and Applied Sciences* **6**(9): 10-37.

Simons A., Melamed-Bessudo C., Wolkowicz R., Sperling J., Sperling R., Eisenbach L. and Rotter V. (1997). PACT: cloning and characterization of a cellular p53 binding protein that interacts with Rb. *Oncogene* **14**: 145–155

Simons M., Ariyoshi H., Salzman E.W. and Rosenberg R.D. (1995). c-myb affects intracellular calcium handling in vascular smooth muscle cells. *American Journal of Physiology-Cell Physiology* **268**(4): C856–C868

Singh E., Ruff P., Babb C., Sengayi M., Beery M., Khoali L., Kellett P. and Underwood J.M. (2015). Establishment of a cancer surveillance programme: the South African experience. *The Lancet Oncology* **16**(8): e414-e421.

Spriestersbach A, Kubicek J, Schäfer F, Block H and Maertens B (2015). Purification of His-tagged proteins *Methods in Enzymology* **559**: 1-15.

Srinivasan S.R., Gillies A.T., Chang L., Thompson A.D. and Gestwicki J.E. (2012). Molecular chaperones DnaK and DnaJ share predicted binding sites on most proteins in the *E. coli* proteome. *Molecular BioSystems* **8**(9): 2323-2333.

Striegel AM (2008). Size-exclusion chromatography: smaller, faster, multi-detection, and multi-dimensions. *Analytical and Bioanalytical Chemistry* **390**(1): 303-305.

Sung H., Ferlay J., Siegel R.L., Laversanne M., Soerjomataram I., Jemal A. and Bray F. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians* **71**(3): 209-249.

Tatulian S.A. (2013). Structural characterization of membrane proteins and peptides by FTIR and ATR-FTIR spectroscopy. In *Lipid-Protein Interactions* (pp. 177-218). Humana Press, Totowa, NJ.

Tatulian S.A. (2019). FTIR analysis of proteins and protein–membrane interactions. In *Lipid-Protein Interactions* (pp. 281-325). Humana, New York, NY.

Tóth M.E., Gombos I. and Sántha M. (2015). Heat shock proteins and their role in human diseases. *Acta Biologica Szegediensis* **59**(suppl. 1.): 121-141.

Van Drie J.H. (2011). Protein folding, protein homeostasis, and cancer. *Chinese Journal of Cancer* **30**(2): 124.

Venselaar H., Joosten R.P., Vroling B., Baakman C.A., Hekkelman M.L., Krieger E. and Vriend G. (2010). Homology modelling and spectroscopy, a never-ending love story. *European Biophysics Journal* **39**(4): 551-563.

Vineis P. and Wild C.P. (2014). Global cancer patterns: causes and prevention. *The Lancet* **383**(9916): 549-557.

Wang X., Zhang H. and Chen X. (2019). Drug resistance and combating drug resistance in cancer. *Cancer Drug Resistance* **2**(2): 141-160.

Waterhouse A., Bertoni M., Bienert S., Studer G., Tauriello G., Gumienny R., Heer F.T., De Beer T.a.P., Rempfer C. and Bordoli L. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research* **46**(W1): W296-W303.

Xiao C., Wu G., Zhou Z., Zhang X., Wang Y., Song G., Ding E., Sun X., Zhong L., Li S., Weng J., Zhu Z., Chen J. and Wang X. (2019). RBBP6, a RING finger-domain E3 ubiquitin ligase, induces epithelial–mesenchymal transition and promotes metastasis of colorectal cancer. *Cell Death and Disease* **10**(11): 1–17.

Xu W., Marcu M., Yuan X., Mimnaugh E., Patterson C. and Neckers L. (2002). Chaperone-dependent E3 ubiquitin ligase CHIP mediates a degradative pathway for c-ErbB2/Neu. *Proceedings of the National Academy of Sciences* **99**(20): 12847-12852.

Yang H., Yang S., Kong J., Dong A. and Yu S. (2015). Obtaining information about protein secondary structures in aqueous solution using Fourier transform IR spectroscopy. *Nature Protocols* 10(3): 382-396.

You L., Lv Z., Li C., Ye W., Zhou Y., Jin J. and Han Q. (2021). Worldwide cancer statistics of adolescents and young adults in 2019: a systematic analysis of the Global Burden of Disease Study 2019. *ESMO Open* **6**(5): 100255.

Yun C.W., Kim H.J., Lim J.H. and Lee S.H. (2020). Heat shock proteins: agents of cancer development and therapeutic targets in anti-cancer therapy. *Cells* **9**(1): 60.