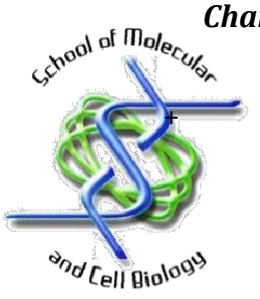


SCHOOL OF MOLECULAR AND CELL BIOLOGY

DISSERTATION: MASTER OF SCIENCE IN BIOCHEMISTRY & BIOINFORMATICS

MSC CANDIDATE: MR CHARMY STARNOD TWALA

SUPERVISOR: PROF. M. NTWASA

PROJECT ADVISOR: DR NIKKI GENTLE

POSTGRADUATE CO-ORDINATOR: DR J. MOLLETT

DRUGS TARGETING THE RETINOBLASTOMA BINDING PROTEIN 6 (RBBP6)

"The collision of computers and biochemistry"

A dissertation submitted to the Faculty of Science, University of the Witwatersrand, Johannesburg, in fulfilment of the requirements for the Master of Science degree.

Declaration

I, Charmy Starnod Twala (560944), am a student registered for the degree of Master of Science (MSc) in the academic year 2017.

I hereby declare the following:

- I am aware that plagiarism (the use of someone else's work without their permission and/or without acknowledging the original source) is wrong.
- I confirm that this dissertation submitted for assessment of the Master of Science degree is my own unaided work except where explicitly indicated otherwise and acknowledged.
- I have not submitted this work before for any other degree or examination at this or any other University.
- I have followed the required conventions in referencing the thoughts and ideas of others.
- I understand that the University of the Witwatersrand may take disciplinary action against me if there is a belief that this is not my own unaided work or that I have failed to acknowledge the source of the ideas or words in my writing.

Charmy S. Twala

03-November-2017

Date

Quotation

"As we let our own light shine, we unconsciously give others permission to do the same…"

- *Marianne Williamson*

Dedication

This work is dedicated to God Almighty who has shown me grace that surpasses human intellect. Lord, your guidance, comfort and provision have been priceless to me.

Thank you so much for your unfailing love.

Research Output

- Conference Proceedings

Oral Presentation

1. Twala CS, Ntwasa MM. ***Drugs targeting the Retinoblastoma binding protein 6 (RBBP6).*** Molecular Bioscience Research Thrust (MBRT) 2016 Research day, University of the Witwatersrand, 8th December 2016.
2. Twala CS, Ntwasa MM. ***Drugs targeting the Retinoblastoma binding protein 6 (RBBP6).*** School of Molecular and Cell Biology (MCB) 2017, MSc Exit Seminar, University of the Witwatersrand, 11th October 2017.

Poster Presentation

1. Twala CS, Ntwasa MM. ***Drugs targeting the Retinoblastoma binding protein 6 (RBBP6).*** Wits Cancer Research Symposium 2017. Medical School Adler Museum, University of the Witwatersrand, 9th February 2017.

Abstract

Screening methodologies have identified specific targets that could serve as potential therapeutic markers in cancer drug design, and the Retinoblastoma binding protein 6 (RBBP6) which is predominately expressed in lung and breast cancers is one critical protein identified. This study seeks to understand the 3D structure of RBBP6 domains, with emphasis on cancer. Three of these domains have been studied in this project, *i.e.* the *Domain With No Name* (DWNN), RING Finger, and the p53-binding domain. The ubiquitin-like structure of the DWNN has implicated this domain as a ubiquitin-like modifier of other proteins such as p53, whilst the RING Finger domain has intrinsic E3 Ligase activity like MDM2 the prototypical negative regulator of p53. The DWNN and RING Finger domains have resolved solution NMR structures, whilst the p53-binding domain has none. Thus, the first initiative undertaken was to model the RBBP6 p53-binding domain using I-TASSER and *e*Thread-Modeller web-severs. Our results demonstrated that this domain mainly constitutes of alpha-helices and loop structures. Structural quality validations of both I-TASSER and *e*Thread-Modeller models were further assessed using Swiss-Model and ProSA (Protein structure analysis) web-servers. Analyses were focussed on specific statistical parameters (Anolea, DFire, QMEAN, ProCheck and the ProSA Z-score). Results from these analyses show that the first I-TASSER model is the best possible representation of the RBBP6 p53-binding domain depicting minimal deviation from native state. Furthermore, screening and docking studies were performed using Schrödinger-Maestro v10.7: Glide SP and drug-like molecules that would potentially serve as agonist or antagonist of RBBP6 were identified.

Keywords: Retinoblastoma binding protein-6, MDM2, p53, DWNN, RING Finger, Virtual screening, Molecular modelling, Structural quality assessment, Docking, Agonist, Antagonist.

Acknowledgements

I would like to express my profound gratitude to my supervisor, Professor Monde Ntwasa for his guidance, research support, motivation and most importantly believing in me. I am thankful for his constructive criticism, brilliant ideas in experimental designs, as well as the write-up of this manuscript.

My sincere appreciation is extended to my advisor, Dr Nikki Gentle for encouraging and advising me throughout the project.

Special thanks to my NRF referees: Dr De Assis Rosa, the late Dr B. Letsolo (Molecular & Cell biology) and Mr Mike Mnchunu (Computer science & Applied mathematics) their motivation granted me a Masters Scholarship for this project.

My special thanks is also extended to the Schrödinger-Team (New York City, California) for providing me with their latest Molecular docking platform (Maestro v10.7).

My sincere and special gratitude goes to my mother, Delsile Mericah Nkosi for her moral and financial support as well as the ceaseless love and care she has shown. Her assistance has helped me tremendously.

My special appreciation also goes to my uncle, Mr David Solomon Nkosi for his sincere moral and financial support.

Special thanks to my older sister, Mrs Charmaine Fakude and my girlfriend Miss Lerato Mononyane for their encouragement and love, I'm highly grateful.

I am also very grateful of the assistance and support I got from my colleagues: Dr E. Nweke, Mr P. Ubanako and Ms B. Monchusi as well as the entire Flylab community; I thank you all.

Lastly, I would like to appreciate the National Research Foundation (NRF) for funding.

TABLE OF CONTENTS **PAGE**

List of Figures **Page**

List of Tables **Page**

List of Abbreviations

Abbreviation	***Full-Name***
ADME	Absorption, Distribution, Metabolism, Excretion
ATP	Adenosine Triphosphate
Cd	Cadium ion
CXCR4	Chemokine Receptor type 4
DWNN	Domain With No Name
E1	Ubiquitin-activating enzyme
E2	Ubiquitin-conjugating enzyme
E3	Substrate specific ubiquitin-protein ligase
E4	Conjugation Factor
EMEA	European Medicine Agency
FDA	Food and Drug Administration
FEP+ Tech	Free Energy Perturbation Technology
Glide SP	Glide Standard Precision
Glide XP	Glide Extra Precision
GUI	Graphic User Interface
HIV	Human Immunodefiency Virus
HMQC	Heteronuclear Multiple Quantum Coherence
HTVS	High Throughput Virtual Screening
HSC	Hematopoietic Stem Cell
HSQC	Heteronuclear Single Quantum Coherence
HDM2	Human homologue MDM2

HDX-MS	Hydrogen deuterium exchange-mass spectrometry
I-TASSER	Iterative-Threading ASSEmbly Refinement
MM-GBSA	Molecular Mechanics, the Generalised Born model & Solvent Accessibility
MDM2	Mouse Double Minute 2 – homologue
NADH	Nicotinamide adenine dinucleotide
NMR	Nuclear Magnetic Resonance
NPC1L1	Niemann-Pick C1-Like 1
OXPHOS	Oxidative Phosphorylation
p53BD	p53 Binding domain
PACT	P53-Associated cellular-testes derived
PDB	Protein Data Bank
ProSA	Protein Structure Assessment
QMEAN	Qualitative Model Energy ANalysis
QSAR	Quantity Structure-Activity Relationship
QSPR	Quantity Structure-Property Relationship
RBBP6	Retinoblastoma Binding protein–6
Rb	Retinoblastoma protein
ResQ	Residue Quality
SDF-1	Stromal cell-Derived Factor-1
TCA	Tricarboxylic Acid cycle
YB-1	Y-box binding protein 1
Zdd	Zinc Drug Database
Zn	Zinc

CHAPTER ONE: LITERATURE REVIEW

1.1 Insight into the Retinoblastoma Binding Protein 6 (RBBP6)

RBBP6 is a 250-kD multi-functional nuclear protein with several domains, found in all eukaryotes. It is implicated in a variety of cellular functions such as cell cycle regulation (Scott *et al.*, 2003 and Kappo *et al.*, 2012), mRNA metabolism (Witte *et al.*, 1997 and Shi *et al.*, 2009) , and development (Yoshitake *et al.*, 2004 and Li *et al.*, 2007). In humans, RBBP6 forms part of the pre-mRNA 3'-end processing complex (Vo *et al.*, 2001) and is hypothesised to interact with tumour suppressors such as p53 and pRb (Witte *et al.*, 1997). However, in mice PACT has been shown to exhibit a direct interaction with these tumour suppressors (Li *et al.*, 2007).

RBBP6 is composed of an N-terminal ubiquitin-like domain known as the DWNN (Domain With No Name) (Pugh *et al.*, 2006), through which it facilitates ubiquitination. It also constitutes a cysteine-rich RING Finger-like domain which further promotes p53 ubiquitination by Hdm2 in an E4-like manner (Li *et al.*,2007). The RBBP6 RING Finger domain also interacts directly with the pro-proliferative transcription factor Y-box binding protein-1 (YB-1) (Eliseeva *et al.*, 2011). In this instance RBBP6 acts as a negative regulator of YB-1, as expression studies demonstrated that its over-activation in cultured mammalian cells leads to the suppression of YB-1 in a proteasome-dependent manner (Chibi *et al.*, 2008). It remains unclear whether this illustrates an example of E4 (additional conjugation factor) or E3 (substrate specific ubiquitin-protein ligase)-like behaviour. Significantly, this over-activation triggers a down regulation of both the pro-apoptotic p53 and anti-apoptotic YB-1 which hints on the different roles of RBBP6 during tumorigenesis.

The Retinoblastoma binding protein 6 gene is located on chromosome 16p12.2 in humans and encodes the RBBP6 isoforms. The RBBP6 gene is transcribed into two mRNA sequences of different sizes (1.1kb and 6.1kb), which then encodes four RBBP6 protein isoforms. Isoform 1 is encoded by the

6.1kb transcript and upon alternative splicing; Isoform 2 and 4 emerges from the same transcript. Isoform 3 also known as the Domain With No Name (DWNN) as it is comprised of only this domain, is encoded by the 1.1kb transcript (Pugh *et al.*, 2006).

1.2 Isoforms and 3-dimensional structures of RBBP6 domains

The RBBP6 gene is composed of 18 exons within its coding region. Its isoforms: isoforms 1, 2, 3 and 4 share a common N-terminal domain and are encoded by 18, 17, 3 and 11 exons respectively. These exons encode different domains within a single isoform. Consequently, the isoforms constitute different domain structures which implicate them in performing different cellular functions (Figure 1.1). Thus isoforms 1 and 3 are significantly implicated in carcinogenesis whilst the functions of isoforms 2 and 4 have not been elucidated (Ntwasa, 2016).

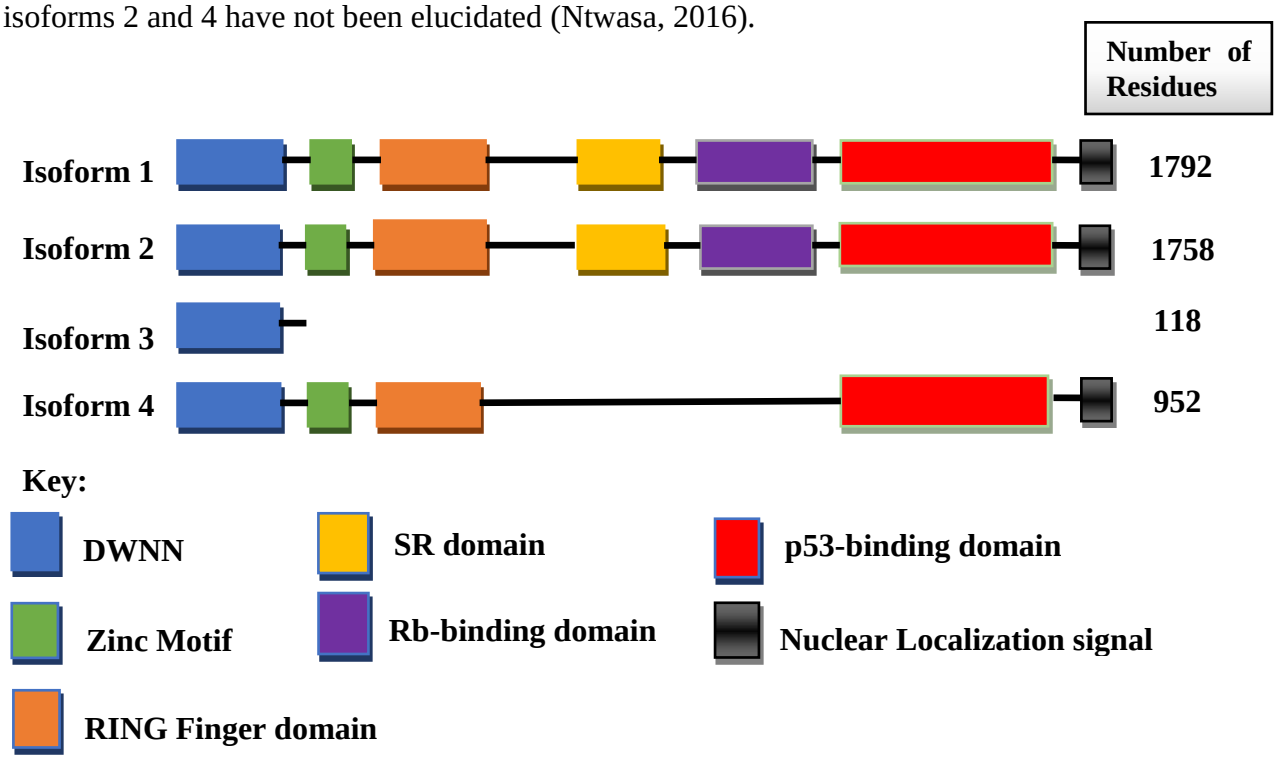


Figure 1.1: Illustrated domains encodes across different isoforms of RBBP6: The number of amino acid residues that constitutes each isoform is also represented.

In this study, only isoform 1 was considered and three of its domains (DWNN, RING Finger, and p53-binding domains) were analysed as potential drug targets. This is mainly because the zinc motif, serine rich domain (SR domain) as well as the nuclear localisation signal have no significant implication in carcinogenesis. However, the retinoblastoma binding domain (Rb-binding domain) targets the pro-apoptotic retinoblastoma protein (pRb) which has critical cell-cycle regulating functions. This tumour suppressor protein (pRb) is mostly critical for retinoblastoma carcinomas in children, thus despite its' tumour inhibitory effects the p53-binding domain (which binds the p53 tumour suppressor protein) was examined for potential drug-like inhibitors since the p53 protein is widely expressed in numerous cancers. This was done to provide a broader perspective of the potential drug-like inhibitors in carcinogenesis.

Currently, both the DWNN and RING Finger domains have resolved solution NMR (Nuclear magnetic resonance) structures (Figure 1.2&1.3), whilst the p53-binding domain has none. The N-terminal ubiquitin-like domain (DWNN) structure was firstly resolved using the Modeller software in the presence of a ubiquitin protein template. This application predicted an exact 3-dimensional structure like that resolved by solution NMR spectroscopy, despite the 22% sequence identity between them (Mather *et al.,* 2005). The lysine residues on DWNN (mostly Lys48 and Lys63) are very critical in the ubiquitin template during substrate specific ubiquitin-protein ligase, they specifically facilitate the formation of peptide bonds between the ubiquitin moieties and various protein substrates (Figure 1.2). Thus, their presence in DWNN suggests that RBBP6 might be functioning with some form of ubiquitin-like modification (Pugh *et al.,* 2006).

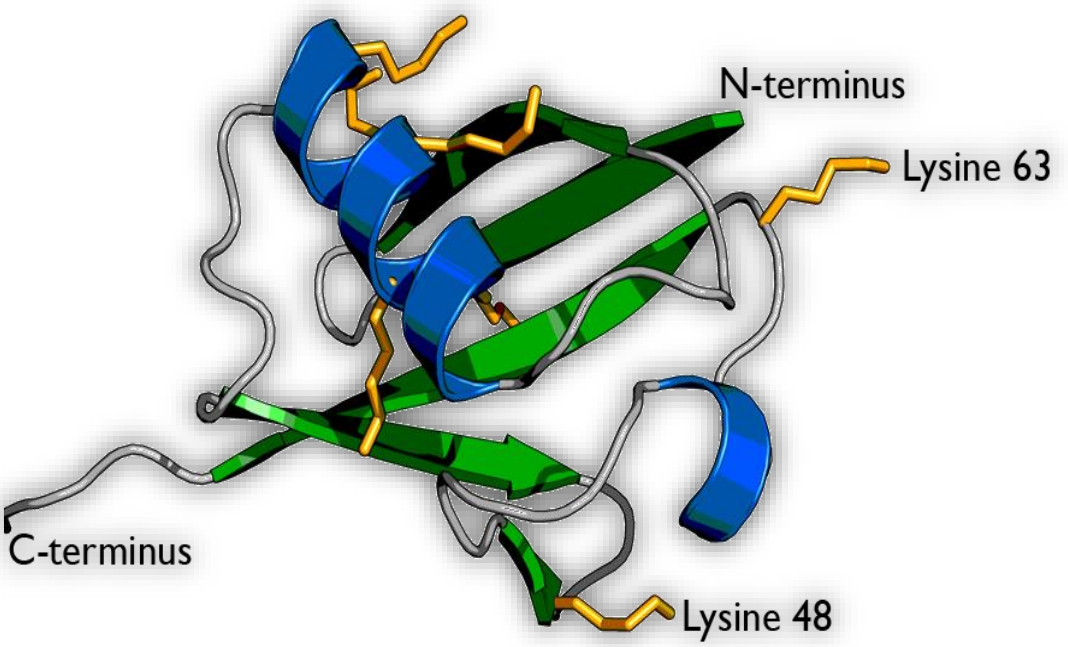


Figure 1.2: Solution structure of the DWNN domain from human RBBP6: The structure was resolved using Nuclear magnetic resonance spectroscopy (NMR), and is composed of both alpha helices (blue) and beta sheets (green), as well as seven lysine residues (yellow). It was viewed using PyMOL and constitutes a PDB_ID: 1UBI, (Adapted from: Pugh *et al.*, 2006).

The DWNN domain is expressed in higher vertebrates as a small independent protein (*i.e.* isoform 3), which further suggests that it may function as a novel ubiquitin-like modifier of other proteins (Pugh *et al.*, 2006). Furthermore, isoform 3 has been shown to be down-regulated in most human cancers which hints its role in inducing cell cycle arrest or apoptosis (Mbita *el al.*, 2012). The DWWN has also been shown to form a crucial part of the multi-subunit protein complex, which mediates polyadenylation of mRNA precursors. This suggests that RBBP6 plays a role in regulation and covalent modification (addition of ubiquitin-like moieties) of pre-mRNA processing proteins (Pugh *et al.*, 2006).

The RBBP6 RING Finger domain is made up of approximately 40 residues (including 8 conserved cysteines), and folds independently with the help of two zinc ions. These ions are coordinated by four pairs of cysteine residues in a “cross-braced” fashion. This means that one ion is co-ordinated by the

first and third Cys pairs while the other ion by the second and forth Cys pairs (Kappo *et al.*, 2012). RING Fingers are therefore generally classified based on this zinc coordination pattern.

There are several known RING Fingers such as C4C4, C3HHC3, C2H2C4 and C3HC4 which is the most common. The presence of the eight conserved cysteine residues suggests that RBBP6 contains a C4C4 RING Finger domain binding two zinc ions. Based on current literature this was confirmed using a directly detected ^{113}Cd spectrum and the ^{1}H – ^{113}Cd HMQC spectroscopy (Kappo *et al.*, 2012). In solution the RBBP6 RING Finger domain adopts a homodimeric structure, which forms a deep hydrophobic cleft at the interface of the two monomers (Kappo *et al.*, 2012) (Figure 1.3).

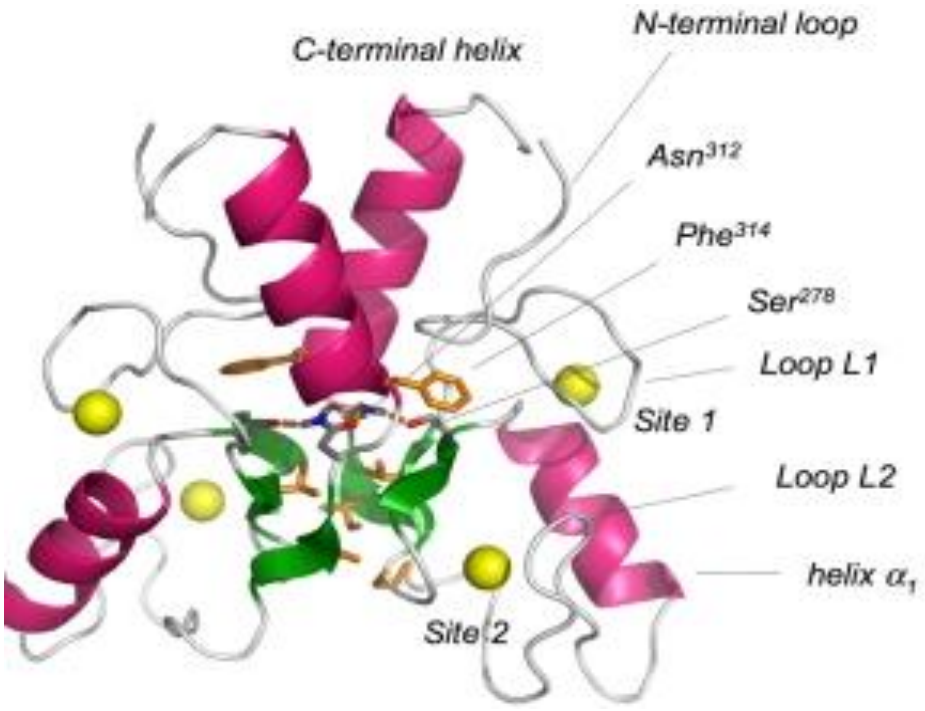


Figure 1.3: **Solution structure of the RING Finger-like domain from human RBBP6:** The schematic depicts a secondary structure of the homodimer, with a hydrophobic interface consisting of a collection of hydrophobic residues (brown), including PHE314 which deeply projects into the opposite monomer. This interface is further stabilised by the *triple-stranded β-sheets* (green), C-terminal helices (pink) and asparagine (ASN312) which is positioned at the centre and forms a pair of reciprocal hydrogen bonds to its counterpart on the opposing monomer. The zinc ions are shown in yellow spheres, the structure was resolved using solution NMR and has been uploaded on the protein databank with PDB_ID: 3ZTG, (Adapted from: Kappo *et al.*, 2012).

Dimerization occurs within *triple stranded β-sheets* (green) forming a deep hydrophobic core interface, which is then stabilised by hydrophobic site chain interactions between the C-terminal helix and the N-terminal loop (including Loop L1 and L2 as shown in Figure 1.3) of the two monomers. Furthermore, positioned at the centre/core of the interface close to the 2-fold axis is asparagine (ASN312) which co-ordinates a pair of reciprocal hydrogen bonds to its counterpart on the opposite monomer (Kappo *et al.,* 2012). This co-ordination is like that observed in many U-boxes facilitated by ASN314 which occupies the same position in murine CHIP (Zhang *et al.*, 2005).

The structure also depicts a hydrogen bond co-ordination between SER278 and ASN312, which is supported by broadening of all resonances associated with SER278 (Kappo *et al.*, 2012). Altogether, this network of polar interactions is likely to exhibit a strong stabilising effect on the homodimer and thus maintain the hydrophobic interface of the RING Finger domain. Site-directed mutagenic studies involving the change of ASN312 to aspartic acid (N312D) further abolished the homodimer completely (Kappo *et al.*, 2012). Which hints on the significance of this residue (ASN312) in forming polar interactions that stabilises the homodimeric nature of this domain.

Therefore, this hydrophobic interface is likely the catalytic site of the RING Finger domain, and thus gives RBBP6 its E3-ligase activity which then implicates it in p53 ubiquitination and other protein substrates such as pRb. Therefore, targeting this region with small molecular drug-like inhibitors might significantly re-activate p53 and pRb levels in cancer cells and thus prove useful in carcinogenesis and cancer therapy.

1.3 Retinoblastoma Binding protein 6 as a potential drug target

1.3.1 P53 protein as a tumour suppressor

P53 is a widely-studied protein implicated in tumorigenesis. It is mutated or silenced in more than half of human cancers (Petitjean *et al.*, 2007 and Danilova *et al.*, 2010). The tumour suppressor protein

(p53) functions by modulating apoptosis and the cell cycle through regulating expression levels of its transcriptional targets (Ko *et al*., 2004; Ko *et al.*, 2005 and Lee *et al.*, 2008). Thus, the absence of p53 in cancer cells enable them to either evade cell cycle control, or fail to activate p53-dependent cell cycle arrest or apoptosis.

Under normal conditions p53 responds to different forms of cellular stresses, such as hypoxia and DNA damage where it activates genes involved in growth arrest at G0 or G2/M phase of the cell cycle (Cheng *et al.*, 2008). Furthermore, it induces the expression of pro-apoptotic proteins within the bcl-2 family, resulting in the induction of apoptosis in cells that cannot be repaired like those with damaged DNA. The p53 responses are therefore carried out by transcription factors of various p53 response genes (bax and p21) in normal cells.

1.3.2 The MDM2-p53 complex interaction

The 3D structure of MDM2-p53 complex has been explored, and its stability relies on the role played by RBBP6 as a scaffold protein (Li *et al.*, 2007). HDM2 (Human homologue of MDM2) is a phosphoprotein with 491 amino acids, and interacts with the α-helix NH_2 terminal transactivation domain of p53 through its NH_2 terminal domain (Kussie *et al.*, 1996). This interaction induces the blockage of p53 transcriptional activity and subsequent ubiquitination by MDM2 for proteasomal degradation (Moll *et al.*, 2003). The biochemistry underlying MDM2 mediated p53 function inhibition have been explored using crystallography. This demonstrated that the N-terminal domain of MDM2 forms a deep hydrophobic pocket where binding of the p53 transactivation domain occurs, implying that it is not involved in the interaction with the transcriptional machinery (Kussie *et al.*, 1996).

The interaction between MDM2 and p53 is restricted to a small hydrophobic cleft (25-109 amino acids residue) and an amphipathic peptide containing 15 amino acid residues at the NH_2 terminus of MDM2 and p53 respectively (Kussie *et al.*, 1996 and Moll *et al.*, 2003).

Consequently, the minimal binding site of MDM2 on p53 was mapped within 18-26 residues, and site-directed mutagenesis within this region demonstrated the significance of certain p53 residues such as LEU14, PHE19, LEU22, TRP23, and LEU26.

Amongst these; PHE19, TRP23, and LEU26 are vital in binding the MDM2-hydrophobic cleft (Moll and Petrenko, 2003). Therefore, mutated p53 proteins are resistant to the MDM2-mediated proteasomal degradation, and similarly the mutation of MDM2 residues (GLY58, GLU68, VAL75, or CYS77) causes an inability to bind wildtype p53 (Moll *et al.*, 2003). This hints on the significance of these residues from both proteins in the formation of stable MDM2-p53 complexes.

The interacting domains depict a lock and key configuration of the MDM2-p53 complex, forming a dual-site interaction model (Figure 1.5). The p53 NH_2-terminus constitutes amphipathic peptides, and therefore uses the alpha helical hydrophobic side which is formed by PHE19, TRP23, and LEU26 to bind deeply into the MDM2 hydrophobic pocket (Kussie *et al.*, 1996 and Patil *et al.*, 2014) (Figure 1.4).

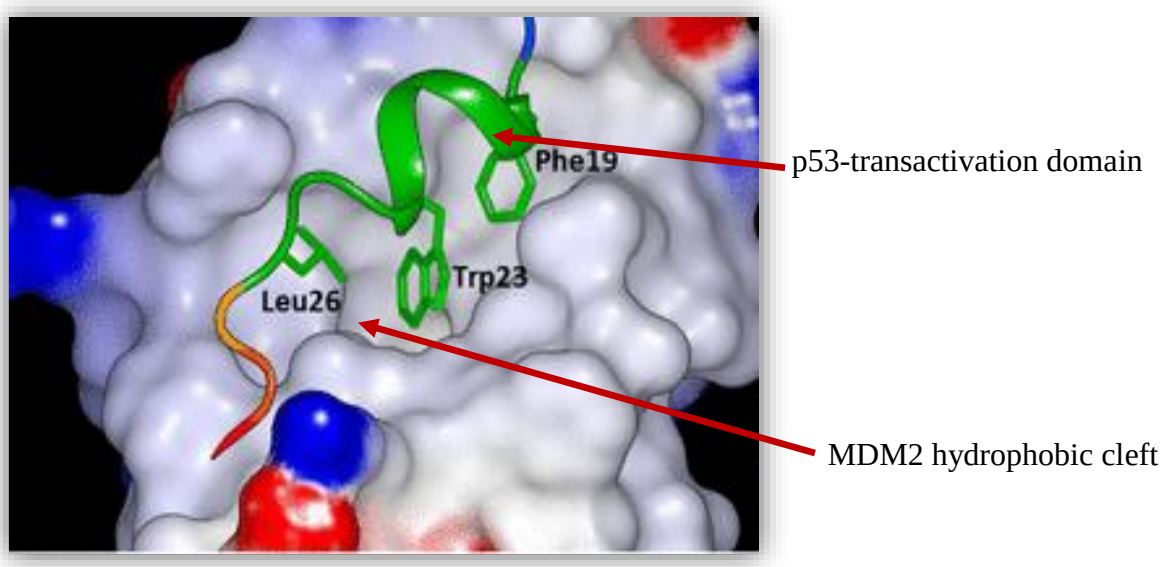


Figure 1.4: The Crystal structure of MDM2-p53 interaction (pdb_id: 1ycq): Shows the p53-transactivation domain (green) with 3 main residues (LEU26, TRP23 and PHE19) bound into a deep hydrophobic pocket (grey) of MDM2, which then promotes p53 inhibition and MDM2-mediated proteasomal degradation. The surface representation of grey concave regions in the MDM2 hydrophobic cleft illustrates its pocket-like nature of the domain (Adapted from: Patil *et al.*, 2014).

The MDM2 cleft is formed by the 26-108 amino acid region and contains two structurally similar portions that folds into a deep groove of aromatic and 14 hydrophobic residues (Kussie *et al.*, 1996, Moll and Petrenko, 2003). Hence, a tight binding mode is maintained between the p53-transactivation domain and MDM2.

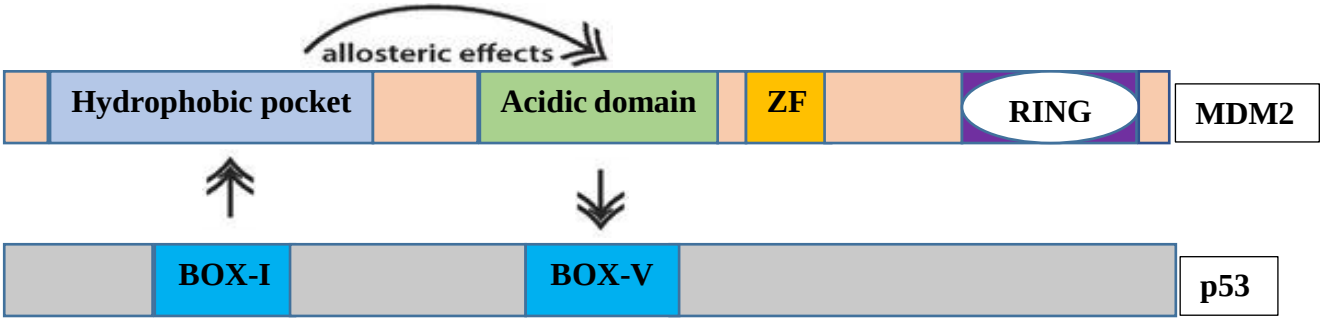


Figure 1.5: Domain structures of MDM2 and p53 illustrating positive effect binding: The p53 BOX-I domain binds the N-terminal hydrophobic pocket of MDM2, whilst there is also binding of the MDM2 central acidic domain to the p53 BOX-V domain. This forms the "dual-site" model of p53 MDM2-mediated ubiquitination (Goh *et al.*, 2014).

2.3.3 Scaffold activity of RBBP6

The E3 ubiquitin ligase activity of p53-associated cellular protein-testes derived (PACT also known as RBBP6) implicates it in acting as a scaffold protein promoting the assembly of HDM2 (Human double minute 2)-p53 complexes, resulting in an increased HDM2-mediated ubiquination and proteasomal degradation of p53 (Li *et al.*, 2007). This suggests that RBBP6 might act as a negative regulator of p53, inducing cell growth or cell proliferation when upregulated and promoting a p53-dependent cell cycle arrest or apoptosis when down regulated. Thus, RBBP6 indirectly affects the cell cycle through interacting with HDM2 and together regulate cellular levels of p53 (Li *et al.*, 2007).

Therefore, the involvement of RBBP6 in facilitating HDM2-p53 assemble is crucial in hypothesizing its role as a possible anti-cancer drug target. Designing of small drug-like molecules that can disrupt the RBBP6-HDM2 interaction (scaffolding activity), could serve as an alternative way of inhibiting the formation of MDM2-p53 complexes to reactivate p53 levels in cancer cells.

1.4 Rational drug design

Drug discovery has increased quite significantly based on the availability of X-ray Crystallographic or NMR (Nuclear magnetic resonance) protein structures, development of structure guided computer-aided methodologies, as well as docking tools (Mandal *et al.,* 2009*)*. Despite this increase, developing lead molecules (small molecules of desired properties) and effective drug targets remains challenging even for known targets. The protein data bank (PDB) is currently estimated to hold over 100,000 3D structures, but this number is still relatively insignificant as some of these biomolecules have multiple structures binding to different molecules. Most important target 3D structures are still unknown based on the limitations on the throughput and applicability of protein size presented by both NMR and X-ray crystallography. Therefore, this delays the drug-discovery process and have prompted a quest for complementary analytical tools that can detect local conformation changes with high sensitivity and experimental throughput (Yu *et al.,* 2013).

Recent advances in elucidating high-order structures (HOS) is based on the underlying principles of HDX-MS, which stands for Hydrogen-deuterium exchange mass spectrometry. This technique seeks to counter act the limitations posed by conventional methods, by increasing its sensitivity upon flexibility and molecular dynamics analysis of protein structures (Wales *et al.,* 2006; Fang *et al.,* 2010 and Hebling *et al.,* 2010). HDX-MS is also capable of monitoring domain interaction, localised protein breathing and folding or unfolding in solution. The principle behind HDX-MS is mainly based on a chemical phenomenon in which the amide hydrogen atoms (NH) on the protein backbone, exchange with deuterium atoms on the surrounding solvent or gas (Yu *et al.,* 2013). The coupled mass spectrometer then detects where the exchange took place in the molecule (i.e. at the protein, peptide or single amino acid level) (Engen *et al.,* 2009).

In addition to these advances, newly improved approaches of rational drug design are also necessary to overcome problems in structure guided computer aided-drug design.

The rational drug design methodology generally encompasses two critical approaches:

- Development of small drug-like molecules for targeting cellular proteins or enzymes whose functional roles in cellular processes and 3D structural information are well-defined (Mandal *et al.*, 2009*)*.
- Development of small drug-like molecules with desired chemical properties for targets, whose cellular functions and 3D structural information may or may not be known. Advanced computational tools can be utilised to retrieve knowledge of unknown targets by performing global gene expression data analysis of drug treated and untreated samples (Mandal *et al.*, 2009).

The final process in rational drug design focuses on the balance between hydrophilicity and pharmacokinetics (ADME) as well as the electrophilicity, nucleophilicity and biodegradation (radical attack). During this process toxicity is also thoroughly examined due to the biotransformation at different phases of metabolism. This examination is done using QSAR (Quantitative structure-activity relationship) and QSPR (Quantitative structure property relationship). Gene expression profiling and bioinformatics analysis are further applied to gain insight of the gene expression patterns immediately after the initial evaluation and identification of lead molecules (Mandal *et al.*, 2009).

This insight can then be utilised to improve drug potency which measures Eradication of disease, elimination or minimization of toxic side effects, reduction of undesirable biotransformation, improvement in distribution (bioavailability), overcoming of drug resistance, and improvement of the immune response (Mandal *et al.*, 2009).

1.4.1 Structure guided computer-aided drug design

Structure guided methodologies remains a major part of drug discovery and development for known 3D structures of potential drug binding site. In structure guided-drug design various *state-of-the-art tools* (X-ray Crystallography, Nuclear magnetic resonance (NMR) and Hydrogen Deuterium Exchange – Mass spectrometry (HDX-MS)) are used to determine the 3D structures of targets complexed with their ligands or drug molecules, thus providing hints on binding site location. This serves as the first step towards structure guided drug design for known targets in lead discovery. After the ligand bound 3D structure is known, virtual screening of a large collection of chemical compounds such as those of the ZINC database can be performed (Irwin *et al.*, 2005). This screening approach enables identification of new potential drugs by performing docking experiments of these chemical compounds.

The newly discovered compounds (lead compounds) are further tested for biological potency to improve their binding affinity and specificity, then statistical analyses and relationships (QSAR, QSPR) can be performed to ascertain the potential molecules for lead drug discovery. When working with an FDA approved database, lead molecules can be used to determine their current pharmacokinetics behaviour (ADME) and reactivity (Mandal *et al.*, 2009). To prevent undesirable metabolic fate, further modifications of the lead molecule should be performed through examination of molecular reactivity. This is done by considering the following properties: electrophilicity, nucleophilicity, or radical attacks.

Furthermore, the optimised pre-clinical drug candidates should be examined to ensure that they possess desirable drug-like properties (pharmacokinetics). These properties include, Log P, molar refractivity, number of acidic & basic like molecules, and molecular weights (Mandal *et al.*, 2009), and they were formulated by Lipinski et al (Lipinski *et al.*, 1997) and later developed by Ghose et al (Ghose *et al.*, 1999). Computational power is then used to assess the toxicity of the optimised drug-candidates and

their respective metabolites. However, this is thoroughly further investigated and confirmed using *in vitro* assays.

Small molecules must possess specific chemical properties (mentioned above), and exert the desired pharmacological effect to be classified as a drug-like molecule. Thus, the principle underlying rational drug design encompasses the detailed evaluation of all such properties (Mandal *et al.*, 2009).

1.5 Protein Modelling

Generally, molecular modelling accommodates all kinds of theoretical methods and computational techniques used to model and mimic the behaviour of molecules. It is a widely-used technique to study molecular systems ranging from small chemical systems to large bio-molecular systems.

One of the ultimate goals in protein modelling is the prediction of protein 3D structures from their amino acid sequences. This prediction, however, remains intricately complex as multiple different amino acid sequences have the potential to form similar functional protein structures. Protein structural alignment is therefore important in protein modelling and uses computational graphics to super-position two or more structures. This can reveal conserved features in two related proteins or differences in the conformational changes during allosteric transition or upon complexation (Tsai, 2002).

There are two ways in which structures can be superimposed. First, by displaying structures using interactive graphics to provide the facility to rotate and translate one with respect to the other. This involves superimposing the two structures by “simple visualisation”. The second method is numerical and involves selecting corresponding sets of atoms from two structures, a program can then calculate the best ‘‘least-square fit’’ of one set of atoms to another. These two methods are equally important; eye fitting or visualisation permits rapid experimentation to assess the goodness of fit in different

portions of the molecules, while numerical fitting permits quantitative comparisons of the relative goodness of fit of different structures and sub-structures (Tsai, 2002).

When superposing two protein structures such that the corresponding atoms in the well-fitting regions are optimally matched, two aspects needs to be considered. They include; choosing the corresponding sets of atoms from the two structures, and finding the best match of these corresponding atoms (Tsai, 2002).

However, the principle underlying homology modelling comprises major limitations when modelling proteins or domains that has no conserved sequence or structural regions to currently resolved structures (in PDB). Thus, fragment assembly modelling has been one major approach utilised to counter-act such limitations. This technique, works by using approximately 10 different structural templates on which all have some degree of homology to the query sequence. These templates are then fragmented based on their aligned regions, and then later ligated by Replica-exchange monte Carlo simulations, creating structural decoys. The published softwares using such simulations include I-TASSER, 3D-Jigsaw etc.

Recent advances on this field also led to the development of template-free modelling tools such as QUARK (Xu *et al.*, 2012). This works by incorporating *ab intio* algorithms with Replica-exchange monte Carlo simulations and both work to construct protein's 3D models from only the amino acid sequence. These techniques attempt to mimic *in vivo* molecular dynamics and incorporates protein folding mechanics to achieve high quality models with supreme accuracy (Xu *et al.*, 2012). Thus, they are currently major drivers of protein modelling.

CHAPTER TWO: AIMS AND OBJECTIVES

2.1 Main aim

To find drugs that would target the retinoblastoma binding protein 6 (RBBP6) domains for cancer therapy

2.2 Objectives

1. Modelling the RBBP6 p53-binding domain
    - Protein domain Modelling using I-TASSER and *e*Thread-Modeller web-applications
    - Structural Quality Assessment using Swiss-Model and ProSA web-applications
2. Binding site prediction of the DWNN, RING Finger, and p53-binding domains
3. Screening of chemical compounds and Molecular docking using Schrödinger-Maestro v10.7: Glide SP (Standard precision)
    - The virtual screening was done to determine potential drug-like compounds that could serve as agonist and antagonist of RBBP6
    - Molecular docking was performed to 3D structures having the lowest possible Gibbs free energy (ΔG)
    - The Zinc drug database (Zdd) containing 2924 structures was utilised in all docking studies

2.3 Hypothesis

Drugs can bind the retinoblastoma binding protein 6 (RBBP6) domains (DWNN, RING Finger, and p53-binding domains) with high affinity.

CHAPTER THREE: MATERIALS & METHODS

3.1 Materials

3.1.1 Computational tools

Table 3.1: Software tools used during molecular modelling and docking studies.

Software Name	**Web-site**	**Publisher**
I-TASSER (Iterative-Threading ASSEmbly Refinement)	http://zhanglab.ccmb.med.umich.edu/I-TASSER/	Yang Zhang Lab, University of Michigan
*e*Thread-Modeller	http://brylinski.cct.lsu.edu/content/webservices	Computational Systems Biology Group, Louisiana State University
Swiss-Model	http://swissmodel.expasy.org/workspace/	Swiss-Institute of Bioinformatics, Biozentrum University of Basel
ProSA (Protein Structure Assessment)	https://prosa.services.came.sbg.ac.at/prosa.php	Markus Wiederstein, Centre of Applied Molecular Engineering, University of Salzburg
FTSite: Binding site prediction server	http://ftsite.bu.edu/	Structural Bioinformatics Lab, Boston University
Zinc Drug Database (Zdd)	http://zinc.docking.org/browse/subsets/special	Shoichet Lab, University of California, San Francisco
Schrödinger-Maestro v10.7	https://www.schrodinger.com/ (Licence Needed)	Schrödinger, New York City, California
UniprotKB	http://www.uniprot.org/uniprot/Q7Z6E9	Uniprot Consortium

3.2 Methods

3.2.1 Modelling the RBBP6 p53-binding domain

The RBBP6 p53-binding domain (p53BD) has not been modelled or neither does it have an NMR or X-ray crystallographic structure. Therefore, in this study the p53BD was modelled, using I-TASSER (*Iterative-Threading ASSEmbly Refinement*) and *e*Thread-Modeller web-servers.

I-TASSER constitutes an inherent structure fragment assembly approach, which consistently drives the threading alignments closer to native state (Zhang, 2008). This application identifies structural templates from the PDB library using a multiple threading approach called LOMETS. These structural templates are excised into continuous fragments based on the threading alignments, and full-length atomic models are constructed by replica-exchange Monte Carlo simulations (iterative template fragment assembly simulations) with the threading unaligned regions built by ab initio modelling (Roy *et al.*, 2010). Using spatial restraints from both LOMETS templates and PDB structures by TM-align, a further fragment assembly calculation is performed starting from the SPICKER cluster centroids (Figure 3.1). This is done to remove steric clashes and refine the global topology of the cluster centroids. The computed decoys of this second simulation are clustered in which the lowest energy structures are selected (Roy *et al.*, 2010). Ultimately, using the selected I-TASSER decoys REMO builds the atomic details of the final full-atomic models through hydrogen-bonding network optimisation. Lastly, functional insights of the target are derived by threading the 3D models through a protein function database (BioLiP) (Yang *et al.*, 2015) (Figure 3.1).

Furthermore, *e*Thread-Modeller is a template-based protein structure prediction tool that uses a machine learning approach to Meta-threading and modelling of protein tertiary structures (Brylinski *et al.*, 2012). Templates are also selected from the PDB library using Meta-threading that combines ten individual threading or fold recognition algorithms.

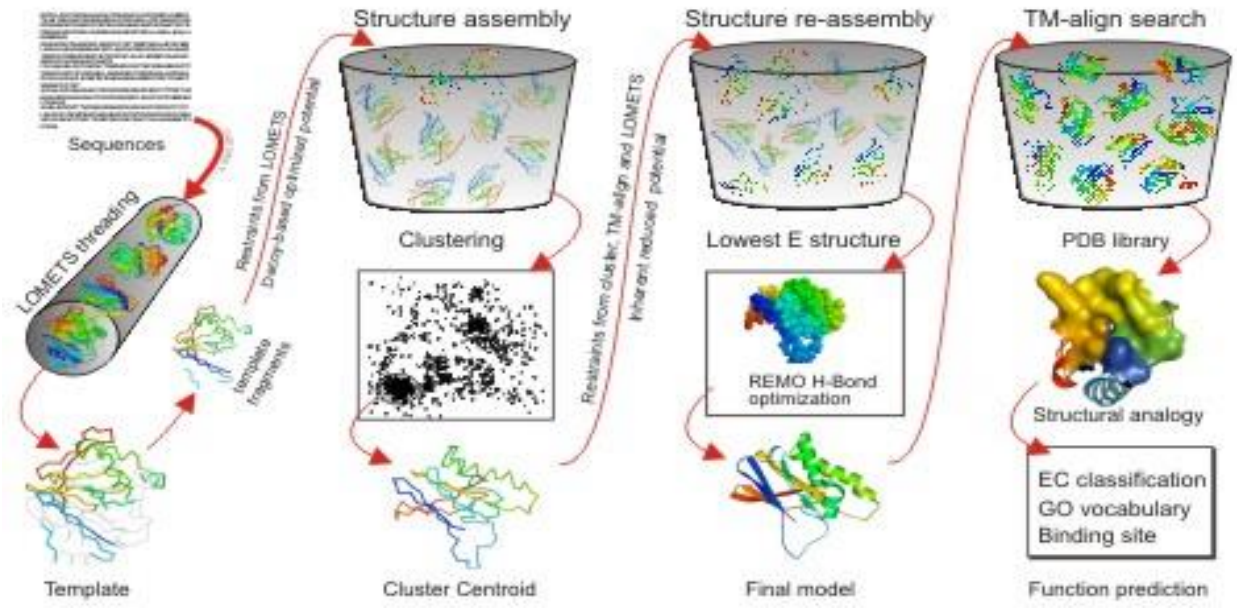


Figure 3.1: I-TASSER protocol for protein structure and function prediction: The server firstly retrieves templates from PDB using LOMETS, and further performs multiple iterative-structure fragment assembly simulations to determine the final models. Functional insights are finally derived by threading the final 3D models through a protein function database (BioLiP) (Adapted from: Zhang, 2008).

The RBBP6 p53BD constitutes 111 amino acids sequence in length and was retrieved from the UniprotKB database (Unique entry identifier: Q7Z6E9), with domain positions ranging from 1433-1544 amino acids. This region was taken up by both servers as inputs. Default settings were used in both I-TASSER and *e*Thread-Modeller such that no algorithmic scripts were written to assist the servers. The generated outputs (predicted p53BD models) were then assessed for quality using quality assessment web-servers such as Swiss-Model and ProSA (*Protein Structure Assessment*). These servers considered a wide spectrum of statistical parameters, ranging from both local to global accuracy predictions to precisely evaluate the models' native state. Thus, the most plausible RBBP6 p53BD model was selected.

3.2.2 *Structural Quality Assessment using I-TASSER, eThread-Modeller, Swiss-Model and ProSA web-servers*

The Iterative-Threading ASSEmbly Refinement and *e*Thread-Modeller applications used various statistics to rank obtained models. These statistical parameters ranged from the Confidence score (c-score), Root Mean Square Deviation of atomic positions (RMSD), Template Modelling score (TM score), and the Global Distance Test-total score (GDT_TS). The RMSD is the measure of the average distance between atoms (backbone atoms) of two superimposed protein structures, and is represented by the equation below:

$$rmsd = \sqrt{\frac{\sum_i d_i^2}{n}}$$ ***Equation 3.1***

Where d represents the square of the distance between two given atoms, and n represents the total number of atoms present. In instances where the two superimposed structures are identical, the RMSD value would be 0Å. Therefore, the larger the RMSD the more dissimilar the structures are. TM-score measures the degree of similarity between two proteins possessing different tertiary structures. The TM-score measures this similarity using a score ranging from (0 – 1), where 1 indicates a perfect match. The GDT_TS is the measure of similarity between two proteins with identical amino acid sequences but different tertiary structures. This is a preferred parameter compared to RMSD, since it's sensitive to outliers generated by poor modelling of individual loops in a reasonably accurate structure.

Both the Swiss-Model and ProSA *(Protein Structure Assessment)* web-servers stipulated further statistical predictions on the generated models. These predictions were based on five parameters namely; ANOLEA, Dfire, QMEAN, ProCheck, and the ProSA z-score. The ANOLEA (Atomic Non-Local Environment Assessment) is the atomic empirical mean force potential, and is used to assess

packing quality of the models (Melo *et al.*, 1998). The program performs energy calculations on a protein chain, evaluating "Non-Local Environment" (NLE) of each heavy atom in the molecule.

The Dfire is an all atom statistical potential based on a distance-scaled finite ideal-gas reference state. This parameter is used to assess non-bonded atomic interactions in a protein model. A pseudo energy for the entire model reflecting its quality was provided, a lower energy indicates that the model is closer to native state (Zhou *et al.*, 2002). The QMEAN is a composite scoring function for both the estimation of global quality of the entire model as well as for local per-residue analysis of different regions within the models. For this study, the local per-residue analysis was performed. The ProCheck focuses on the stereochemistry assessment of the models, and it assesses the geometry of residues in each protein structure, as compared with stereochemical parameters derived from well-refined, high-resolution structures. ProSA *z*-score indicates overall model quality and its value is displayed in a plot that contains the *z*-scores of all experimentally determined protein chains in the current PDB (Wiederstein *et al.*, 2007). Therefore, it was used in this study to check whether the *z*-score of the input structure (modelled RBBP6 p53-binding domain) is within the range of scores typically found for native proteins of similar size.

3.2.3 Binding site prediction

The modelled RBBP6 p53-binding domain, DWNN, and the RING Finger domain structures were then subjected to FTSite Web-server. This was to identify potential binding pockets in all three domains. Binding site prediction is an important technique and constitutes a variety of applications such as structure-based function prediction, protein functional relationship elucidation, protein engineering and drug design. In this study, the technique was applied in the context of small-molecular drug design. The FTSite web-server was used as a binding site identification tool and is based on NMR (Hajduk *et al.*, 2005) and X-ray crystallographic (Mattos *et al.*, 1996) experimental evidence, that ligand binding sites possess a tendency to bind small organic molecules of various shapes, size and polarity. FTSite

constitutes a solvent mapping algorithm (Brenke *et al.*, 2009) which is an exact computational analogue of NMR and X-ray Crystallography.

FTSite identifies regions that have the highest number of non-bonded interactions with overlapping low energy clusters of several small molecular probes (Ngan *et al.*, 2012). This is achieved by using computational mapping that works by placing 1 of 16 different small molecular probes on a dense grid around a given protein, and finds favourable positions using empirical free energy functions. Individual probes are then clustered and ranked based on average free energy (Ngan *et al.*, 2012). Following this, consensus clusters are identified as binding sites where different probe clusters overlap. The consensus clusters are then ranked based on the total number of non-bonded interactions between the protein and all probes in that cluster (Ngan *et al.*, 2012). Consensus clusters are ranked from the highest number of contacts to the lowest. The amino acid residues in contact with the probes of the newly defined cluster constitute the top ranked predicted ligand binding site. Thus, clusters with fewer contacts are defined as lower ranked predictions (Ngan *et al.*, 2012). The use of multiple probes sets FTSite apart from other energy-based methods, and the use of NMR and X-ray Crystallographic evidence improves its robustness on conformational chances. Thus, it is currently one of the leading protein binding site prediction severs with supreme accuracy on prediction irrespective of evolution or statistical data. Results obtained from this application were then used as basis of Grid generation prior the ligand docking process.

3.2.4 Screening of chemical compounds and Molecular docking

The three RBBP6 domains (DWNN, RING Finger, and p53-binding domain) were used as templates on the Schrödinger-Maestro v10.7: Glide SP (Standard Precision) application to screen the Zinc drug database (Zdd) for chemical compounds that can target characterised binding sites. The Zdd constitutes commercially FDA approved drugs, available worldwide as pure compounds. In this study, the entire Zdd database which had 2924 structures was screened. The use of Glide enabled both virtual screening

and molecular docking studies to be simultaneously performed. This application took-in respective domain structures and the Zdd as inputs, and generated a collection of chemical compounds docked into specified pockets with different binding affinities.

The Schrödinger's Receptor-based ligand docking protocol employs a multi-step procedure, which in this study involved the preparation and manipulation of protein domains as well as ligands prior to ligand-docking. These steps were sequentially performed as follows: Protein domain preparation, Ligand Preparation, Grid generation and Receptor-based ligand docking.

3.2.4.1 Protein domain preparation

The Protein Databank (PDB) structures are not suitable for immediate use in molecular modelling calculations, as they usually consist of only heavy atoms. They may also include: core crystallised ligand, water molecules, metal ions and co-factors. Additionally, some structures are multi-meric and need to be reduced to a single unit, and because of the limited resolution of X-ray experiments it can be difficult to distinguish between the carbonyl oxygen and the secondary amine nitrogen of the amides in crystal structures thus the placement of the groups must be checked. PDB structures may also be missing atoms or connectivity information which must be assigned along with bond-orders and formal charges.

Therefore, in this study the Schrödinger-Maestro v10.7 protein preparation wizard was used. This took all the three RBBP6 domains from their raw state (having missing atoms or incorrect bond-order assignments, incorrect charge states and orientation of various groups) and brought them to a suitable state of being utilised by Glide. The wizard contains a Graphical User Interphase (GUI) with a systematic functional procedure that can be briefly described as follows;

- The "import and process" tab imports structures and perform basic tasks fixing the structure.

- The "Review and Modify" enabled the deletion of unwanted chains and waters and fix or delete head groups.
- The "Refine" tab allowed for the optimising orientations of Hydrogen bonded groups and minimise the structure.

These functionalities provided a very convenient way in the successful preparation of all three RBBP6 domains.

3.2.4.2 Ligand Preparation

The Zinc drug database was downloaded (http://zinc.docking.org/browse/subsets/special) in SMILE and SDF formats which contained 2D structures of chemical compounds. This configuration state is not suitable when performing molecular docking calculations, or to simulate using computational docking algorithms. Proteins exist in 3-dimensional space, thus drugs that would successfully target them should also exhibit such configuration. Therefore, the Schrödinger-Maestro v10 .7 ligand preparation wizard was used to convert 2924 2D structures into lowest energy possible 4909 3D structures in a maestro format. This program allows for an expansion of each input structure by generating variations on the ionisation state, tautomer's, stereochemistry, and ring conformations.

The possible ionisation states of the ligands were generated at a target pH range; 7.0 +/- 2, using Epik which is a built-in application within Glide that predicts both the ionisation states and their associated penalties. Epik also predicted different tautomeric forms and calculated energetic penalties for every ligand state it predicts. In Glide, the Epik state penalty is also used to differentiate active from inactive compounds during docking, in fact the use of Epik is known to improve virtual screening enrichment.

3.2.4.3 Grid Generation

The outer scoring grids were generated with different dimensions ranging from 20x20x20Å to 50x50x50Å in x, y, z - axis respectively. Generally, it is important to make the outer grid consistent

with the shape of the protein's active site, thus this was done to only cover-up the active site volume in each of the three RBBP6 domains (DWNN, RING Finger and p53-binding domain) plus the MDM2 p53-binding hydrophobic cleft. The results obtained from FTSite (Binding site characterisation) were used as a premise to accurately map-out critical residues facilitating the binding co-ordination in all the above-mentioned domains.

Furthermore, a ligand centre box (inner grid) was generated. This defines the acceptable ligand centre positions during the side point search, and gives a true measure of the effective search space size. The ligand centre box is useful for ligands to find usual or asymmetric binding modes in the active site or to confine their midpoints into a smaller box to save calculation time. The "centroid of selected residues" option, which specifies the residues that best define the active site was also used, and the inner grid was then centred on the centroid of these selected residues.

3.2.4.4 Receptor-based Ligand docking

The final docking algorithm utilised in this study is the Glide SP-algorithm, better known as the standard precision. The nature of docking simulations employed by this algorithm are the same to that of the High Throughput Virtual Screening (HTVS). Thus, both methods use the same scoring functions, except that HTVS reduces the number of intermediate conformations throughout the docking funnel, the thoroughness of the final torsional refinement and sampling. During the docking process, the domain-structures were kept rigid (not even the hydroxyl and thiol groups could rotate), and flexibility was induced to all docking ligands. This was achieved through the ligand preparation wizard, which had generated a collection of multiple possess of the ligand database (Zdd). The entire work was done using a Core i7 with 4 cores, 8 processors, 8GB of RAM, and only six docking jobs were performed.

Summary: Methodology underlying structure guided computer-aided drug design

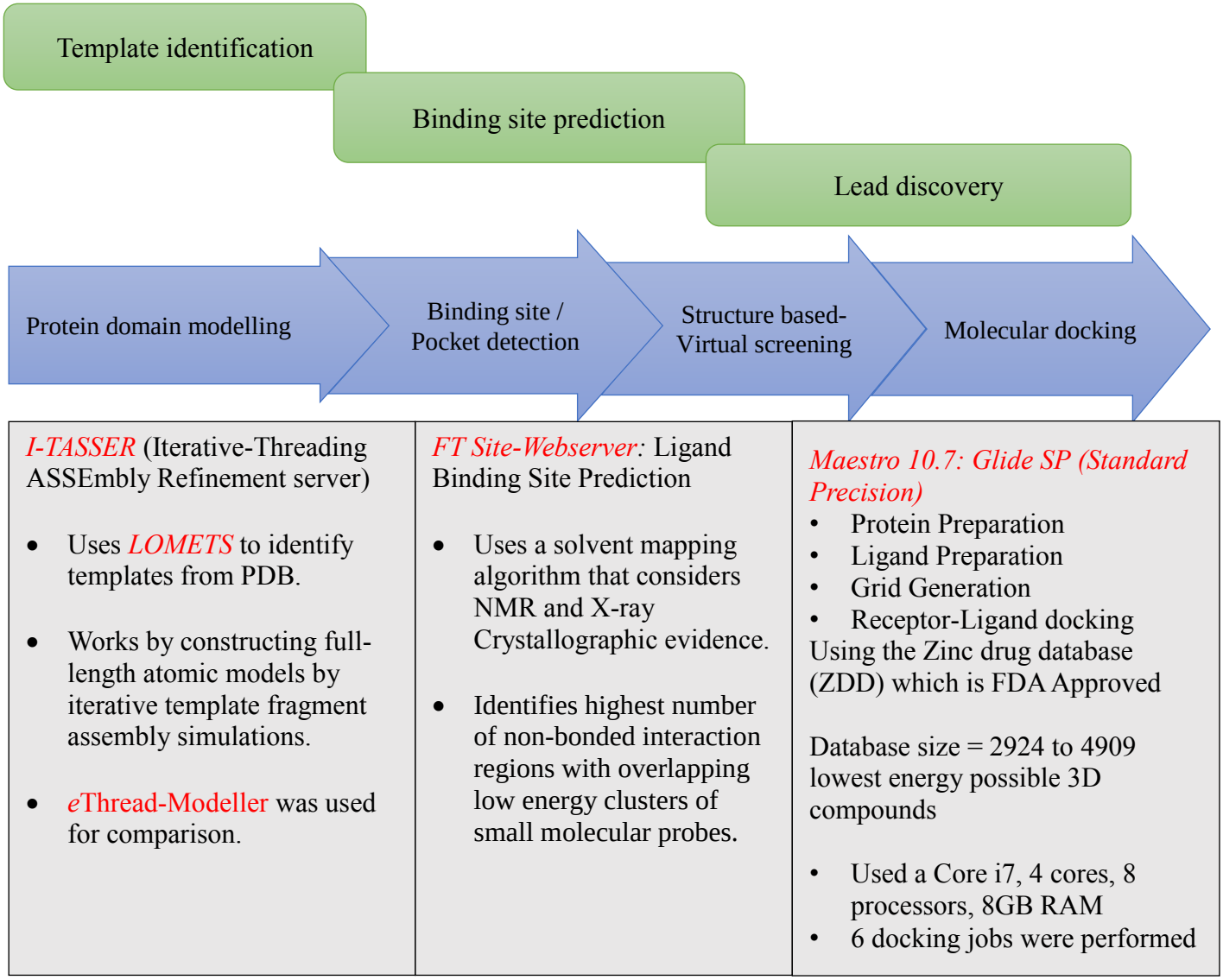


Figure 3.2: **Flow diagram of the adopted methodology:** Representing the adopted methodology underlying structure guided computer-aided drug design. The green region indicates a broader concept of the method performed, whilst the blue area specifies the processes performed in each method. Applications and web-servers used in each process are highlighted in red on the bottom light-grey boxes.

CHAPTER FOUR: RESULTS

4.1 Six plausible RBBP6 p53-binding domain models

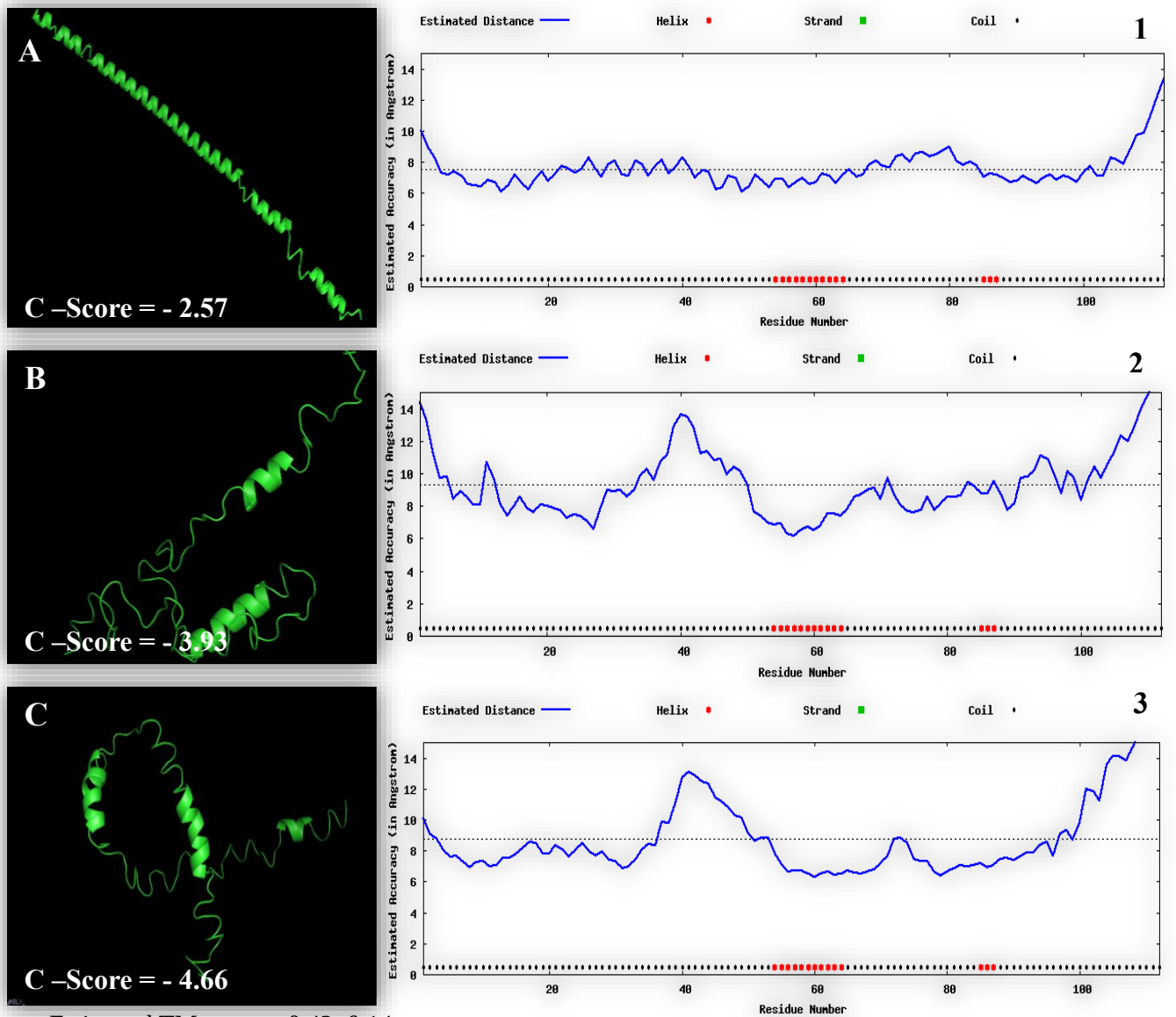
Since the RBBP6 p53-binding domain (p53BD) has not been modelled nor crystallised, this section seeks to close such gaps. Six top models (*i.e*. Model-A, Model-B, Model-C, Model-D, Model-E, and Model-F) demonstrating a tertiary structural fold of the p53BD sequence were chosen and are reported. These models were predicted using I-TASSER (Model-A, Model-B and Model-C) and for comparison, *e*Thread-Modeller (Model-D, Model-E, and Model-F) was also utilised.

Both I-TASSER and *e*Thread-Modeller used various statistical parameters composed of both local and global accuracy predictions to rank the predicted models. I-TASSER used the c-score, TM-Score, and the RMSD for global accuracy, whilst the built-in Residue Quality program (ResQ) was utilised for local accuracy predictions (Figure 4.1). Similarly, *e*Thread-Modeller calculated the global distance test total scores (GDT_TS) for each generated model (Figure 4.2). Analyses based on these parameters show that Model-A and Model-D, generated from both severs, are the best models which may resemble the native state. Other web-severs; Swiss-Model and ProSA were used to validate the structural quality of these models and to select one model that best represent the native state of the RBBP6 p53BD.

4.1.1 Modelling by Iterative-Threading ASSEmbly Refinement (I-TASSER)

I-TASSER simulations generated a large ensemble of structural conformations, called decoys for the RBBP6 p53BD query sequence. When selecting the final models, I-TASSER used the SPICKER program to cluster all decoys based on their pair-wise structure similarity. Here, the top 3 of 5 models representing the RBBP6 p53BD produced by LOMETS are reported (Figure 4.1). The confidence of each model is quantitatively measured by the c-score that is calculated based on the significance of threading template alignments and the convergence parameters of the structure assembly simulations.

The acceptable c-score ranges from -5 to 2, where a c-score of a higher value signifies a model with higher confidence and vice-versa. The TM-align and RMSD scores are estimated based on the c-score and protein length. Thus, of the three models, Model-A with the highest c-score is the most probable model representing the RBBP6 p53BD.



- Estimated TM-score = 0.42±0.14,
- Estimated RMSD = 9.8±4.6Å

Figure 4.1: Three possible RBBP6 p53-binding domain models resolved by I-TASSER: [A–C] Represent the predicted tertiary structures of Model-A, Model-B and Model-C along with their confidence scores (C-scores). The corresponding graphs **[1–3]** depict local accuracy predictions of all three models resolved by the ResQ (Residue Quality and B-factor) algorithm, black and red dots below the graphs represent coils and helices respectively which corresponds to the secondary structure prediction of the query sequence and has been resolved using the built-in *Psipred* algorithm within I-TASSER.

Global accuracy prediction of the models was also assessed using both the TM-align and RMSD scores, as shown in bullet form below the models.

Since the top 5 models are ranked by the decoy cluster size, it is possible that the lower-ranked models have a higher c-score in rare cases. High c-scores of lower ranked models can result in better quality compared to highly ranked models with lower c-scores. If I-TASSER simulations converge (reaches a perfect match), it is possible to have less than 5 clusters or models generated. This would mean that the query sequence precisely aligned with one template from the PDB library, thus indicating good quality of the generated model because of the converged simulations.

The local accuracy in figure 4 (1-3) reflects the distance of deviation (in Angstrom) between residue positions of the models and the native state (Figure 4.1(1-3)). Since the native structure is unknown, the distance errors in the plots above are estimated by the ResQ (Estimation of Residue Quality and B-factor) algorithm using support vector regression that makes use of the coverage of threading alignment, divergence of I-TASSER simulation decoys, sequence-based secondary structure as well as solvent accessibility predictions.

All these models form a mixture of helices and loops structure. However, Model-A seems to have very little deviation (approximately 1-1.5Å) from native state in terms of its local accuracy prediction (Figure 4.1 (1)). Therefore, these local accuracy prediction scores imply that Model-A is the best possible representation of the native state compared to Model-B and Model-C, since there is greater local accuracy deviation (above 3.5Å) in graph 2 and 3 respectively. The Global accuracy of the models has also been predicted by I-TASSER using usual statistical parameters (*i.e.* c-score, TM-score and RMSD score). Reported here is the quality prediction (TM and RMSD) of Model-A, since the correlation between c-score and TM-score is weaker for lower rank models. However, the c-score values of the models decrease from Model-A to Model-C as shown, suggesting that the RBBP6 p53BD is more likely to fold in a pattern resembling that of Model-A.

4.1.2 Modelling by eThread-Modeller

The *e*Thread-Modeller web-application also employs a Meta-threading approach that combines ten individual threading or fold recognition algorithms. Here, *e*Thread-Modeller was used as a comparison tool to observe the relationship between its predicted models and those generated by I-TASSER. Interestingly, a similar folding pattern of the models predicted by these two applications was observed (*i.e.* the helix-loop pattern nature) (Figure 4.2). Since these web-applications consider different sets of parameters and use unique computational calculations for structure prediction, these findings increased confidence in the models and importantly the structural geometry of the RBBP6 p53BD. Notably, the GDT_ TS below decreases from Model-D to Model-F implicating that the Global accuracy of Model-D surpasses that of Model-E and Model-F and thus represents the best model of the RBBP6 p53BD generated by *e*Thread-Modeller.

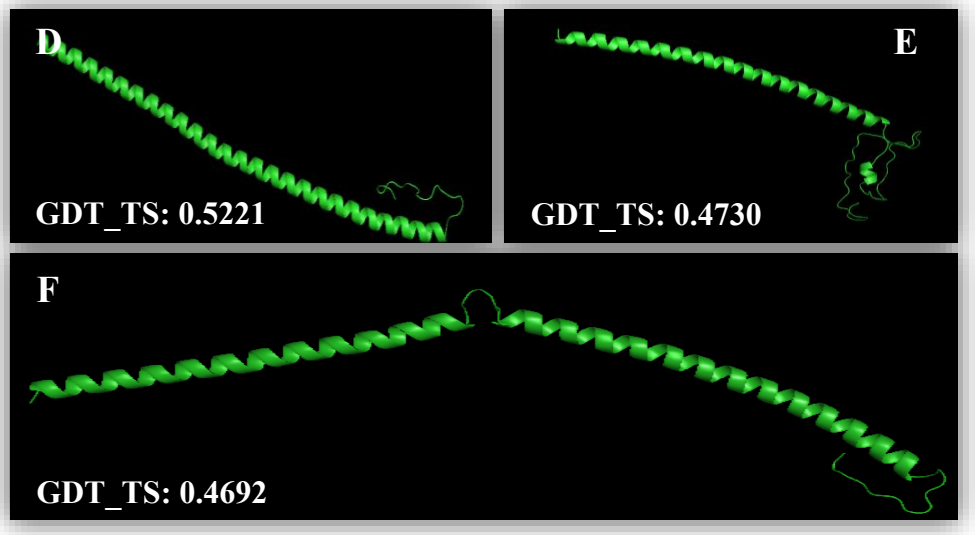


Figure 4.2: Three possible RBBP6 p53-binding domain models resolved by *e*Thread-Modeller: [D–F] Represents the tertiary structures of Model-D, Model-E and Model-F respectively, the structures are ranked based on their Global distance test total scores (GDT_TS) and they only exhibit alpha-helices & coils/loops as shown in each model. The NH_2 to C-terminal positions of all models spans across from left to right.

4.2 Sequence analysis and modelling quality assessment

The RBBP6 p53-binding domain (p53BD) amino acid stretch was used as a query sequence in I-TASSER for the prediction of structural properties (Figures 4.3-4.5). Both I-TASSER and *e*Thread-

Modeller models were then analysed and assessed for structural quality (Figures 4.7&4.8). In this section, structural properties (secondary structure prediction, solvent accessibility and B-Factor) generated by I-TASSER prior to the molecular modelling process are reported. Furthermore, the quality of Model-A and Model-D was assessed using Swiss-Model and ProSA (Protein Structure Assessment) web-applications. These applications added value to the statistical base to make confident statements regarding the quality of the models and enabled the selection of one that could better represent the RBBP6 p53BD.

4.2.1 Sequence analysis and properties prediction using I-TASSER

I-TASSER predicted various properties of the submitted query sequence (RBBP6 p53BD sequence). The secondary structure corresponding to the query sequence was then determined together with the respective confidence scores (Figure 4.3). A helix-loop pattern feature which is like that illustrated by the 3D-models above was observed. However, this domain seems to be conserved at a tertiary structural level given the observed secondary and tertiary structural prediction differences. Solvent accessibility was also explored at a residue-per-residue level, to determine hydrophilic and hydrophobic regions. Noticeably, the helical regions including the 3_{10}-helix were shown to have solvent exposed residues meaning they are more likely to serve as potential hydration pockets (Figure 4.4).

```
                    20        40        60        80        100
                    |         |         |         |         |
Sequence    RLNEQGNFKSLSQSSKEARTSDKHDSTRASSNKDFTPNRDKKTDYDTREYSSSKRRDEKNELTRRKDSPSRNKDSASGQKNKPREERDLPKKGTGDSKKSNSSPSRDRKPHD
Prediction  CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHCCCCCCCCCCCCCCCCCCCCHHHCCCCCCCCCCCCCCCCCCCCCCC
Conf.Score  954346754221013333333323444554445668886544555532123100012454665513688755333344447740010156567643333678986556888
            H:Helix; S:Strand; C:Coil
```

Figure 4.3: Secondary structure prediction by I-TASSER using the built-in *Psipred* algorithm: A high score (Conf.Score) implies higher confidence of secondary structural feature in that region, the key of the prediction is given below the sequences, *i.e.* Helices (Red), Strand (Blue), and Coil (Black).

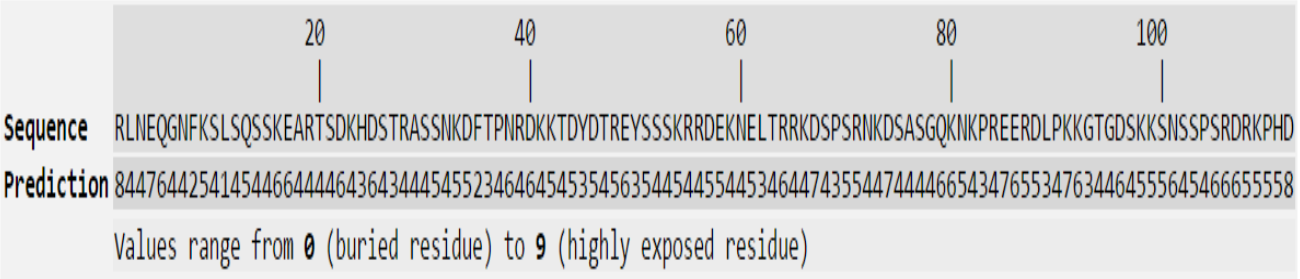


Figure 4.4: Solvent accessibility prediction of the RBBP6 p53BD protein sequence: Regions with solvent exposed residues (hydrophilic) are potential hydration or functional sites, the top numbers represents a residue count in the sequence and the key is shown below the prediction scores.

I-TASSER further predicted B-Factor values to indicate inherent thermal mobility of residues / atoms in the protein domain. This value is deduced from the threading template proteins retrieved from PDB in combination with the sequence profiles derived from sequence databases. Thus, the reported B-factor corresponds to the normalised B-factor of the target protein (Figure 4.5), defined by the following equation;

$$\mathrm{B} = \frac{(B' - u)}{s} \qquad \textit{Equation 4.1}$$

Where B' is the raw B-factor value, u and s are the respective mean and standard deviation of the raw B-factors along the sequence. The observed trend-line in figure 4.5 shows that there isn't much fluctuation of atoms about their average positions. This indicates that the molecular dynamics of the RBBP6 p53BD is unlikely to affect modelling and docking calculations.

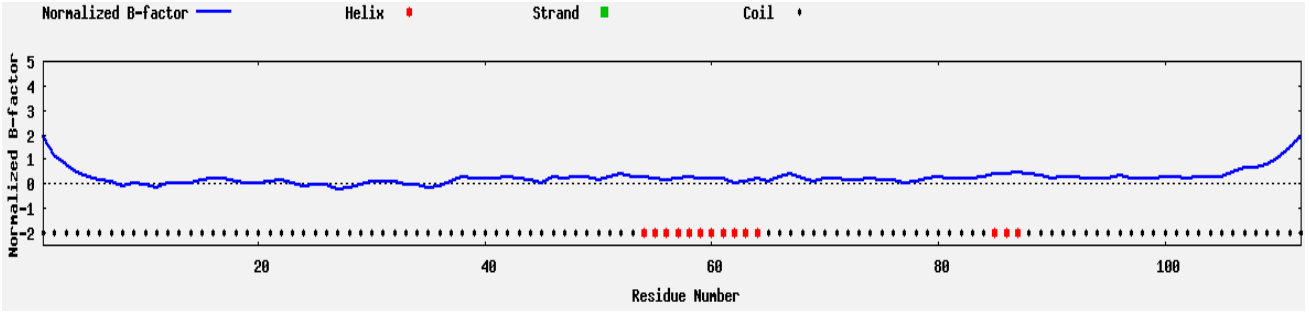


Figure 4.5: Normalised B-factor predicted by the built-in ResQ (Residue Quality) algorithm: The blue trend-line indicate normalised B-factor values on a residue basis, the black and red dots below represents coils and helices respectively which is secondary structure prediction of the p53BD sequence, the key for the prediction is also given on top of the graph.

I-TASSER modelling starts from the structure templates identified by LOMETS, which is a Meta-sever threading approach containing multiple threading programs. Each threading program can generate tens of thousands of templates alignments. I-TASSER then selects ten best temples from the LOMETS threading programs (Figure 4.6), these templates are of highest significance from each threading program as measured by the z-score (*i.e*. the difference between the raw and average scores in a unit of standard deviation). The top ten reported alignments or templates (in order of their ranking) in this section are from the following programs; 1. MUSTER 2. SPARKS-X 3. Neff-PPAS 4. MUSTER 5. SPARKS-X 6. Neff-PPAS 7. MUSTER 8. SPARKS-X 9. Neff-PPAS 10. SPARKS-X.

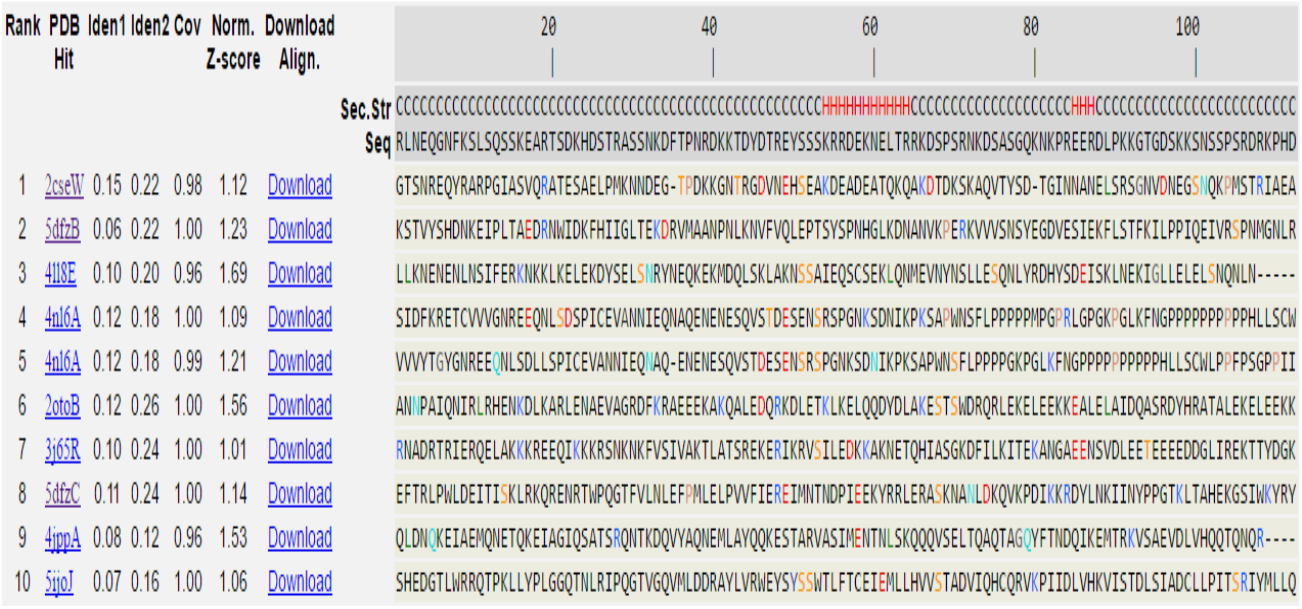


Figure 4.6: **The top 10 threading templates utilised by I-TASSER**, represented along with their corresponding sequence alignment data. [**Rank**] Represents the threading templates used by I-TASSER, [**Iden1**] is the percentage sequence identity of the templates in the threading aligned region with the query sequence, [**Iden2**] is the percentage sequence identity of the whole template chains with query sequence, [**Cov**] represents the coverage of the threading alignment and is equal to the number of aligned residues divided by the length of query protein, [**Norm.Z-score**] is the normalized z-score of the threading alignments. Alignment with a Normalized z-score > 1 means a good alignment and vice versa, [**Download Align**] provides the 3D structure of the aligned regions of the threading templates. All residues are coloured in black; however, those residues in the template which are identical to the residue in the query sequence (Seq) are highlighted in colour. Colouring scheme is based on the property of amino acids, where polar are brightly coloured while non-polar residues are coloured in dark shade.

The *e*Thread-Modeller program also retrieved templates from PDB prior the molecular modelling process (Appendix 5). In this section, top five templates retrieved by both servers are reported (Table 4.1). When comparing, it is noted that there is no identity of the retrieved templates between these web-severs. However, most of these protein templates are alpha-helical with segments of loop structures, which resembles the helix-loop nature of the above reported models. These findings are critical in confirming that the RBBP6 p53BD would fold to exhibit a helix-loop pattern configuration, and despite observed differences on retrieved templates the tertiary structural prediction remains consistent as observed from the reported models (Figure 4.1&4.2). This indicates that despite the use of any method in protein modelling the p53BD would more likely adopt this conformation. Therefore, this suggests that the helix-loop nature plays a critical role in stabilising the native state configuration of this domain (p53BD).

Table 4.1 The top 5 RBBP6 p53BD templates retrieved by I-TASSER and *e*Thread-Modeller from PDB (Protein Databank)

I-TASSER (Iterative-Threading ASSEmbly Refinement)		***e*Thread-Modeller**	
PDB ID	**Protein Name**	**PDB ID**	**Protein Name**
2cse	Major outer-capsid protein-mu1	4fmu	Histone-lysine N-methyltransferase SETD2
5df2	C.elegans FOG-3 BTG/Tob domain	1e5w	MOESIN (Helix), Isolated FERM domain
4118	Sensor protein kdpD	5fm9	Neurogenic Locus Notch Homolog protein 1
4n16	Carbonic anhydrase 2	2mbf	Fork head domain containing protein
2oto	Streptococcus pyogenes M1 protein	4yyc	Putative trimethylamine methyltransferase

4.2.2 Model quality assessment using Swiss-Model and ProSA

Based on the six models generated by both web-applications (I-TASSER and *e*Thread-Modeller), Model-A and Model-D seem to be the best representations of the RBBP6 p53BD. In this sub-section, further analyses of the two models using Swiss-Model and ProSA were conducted to evaluate their quality. These severs use four statistical parameters (*i.e.* Anolea, QMEAN, Dfire and ProSA z-score) that cover both local and global accuracy predictions, and one model which best represent the RBBP6 p53BD was then selected for further docking studies.

A comparative view of Model-A and Model-D in terms of both local (Anolea and QMEAN) and global (Dfire and ProSA) accuracy predictions is depicted in Figure 4.7 and 4.8. The local accuracy analyses in terms of favourable energy environments (Anolea) indicates the presence of negative energies in approximately all residues of Model-A (Figure 4.7A). This means that all these residues are in fovarable local environments with only one exception (SER53 (red arrow)) having a positive energy score. However, the Anolea graph of Model-D has significant differences especially for the first 40 residues (Figure 4.8A). This graph indicates that about 9 residues are not in favourable energy environments, as shown by the red bars indicating positive energies (Figure 4.8A). The QMEAN (Qualitative Model Energy Analysis) is at an average energy range of 3-6 units throughout the query sequence for both models (Model-A & Model-D). This implies that there are no significant differences in terms of energy analysis between the models. Furthermore, the observed positive energy range on the QMEAN graphs indicates the absence of bad contacts between the residues.

The Global accuracy analyses focused on two parameters, *i.e.* Dfire_energy and ProSA z-score. These statistics have critical differences when compared between Model-A and Model-D. The Dfire_energy score for Model-A is more negative (-85.28) than that of Model-D (-75.30), meaning that the RBBP6 p53BD is likely to fold in a pattern resembling that of Model-A, as it's closer to native state than Model-D. As previously mentioned, the ProSA-web measures the overall quality of the model and

produces a z-score in comparison to all experimentally determined crystal structures in current PDB. Here, both models seem to fall within the NMR region (blue). However, models of similar size seem to contain more negative z-scores in a range of -8 to 0. Implying that the more negative is the model's z-score at this size (111 amino acids) the more it is likely to resemble native state. The above results further indicate that Model-A (ProSA z-score = -1.92) is closer to native state than Model-D (ProSA z-score = -1.08), hence it is the best possible representation of the RBBP6 p53BD to be utilised for further docking studies.

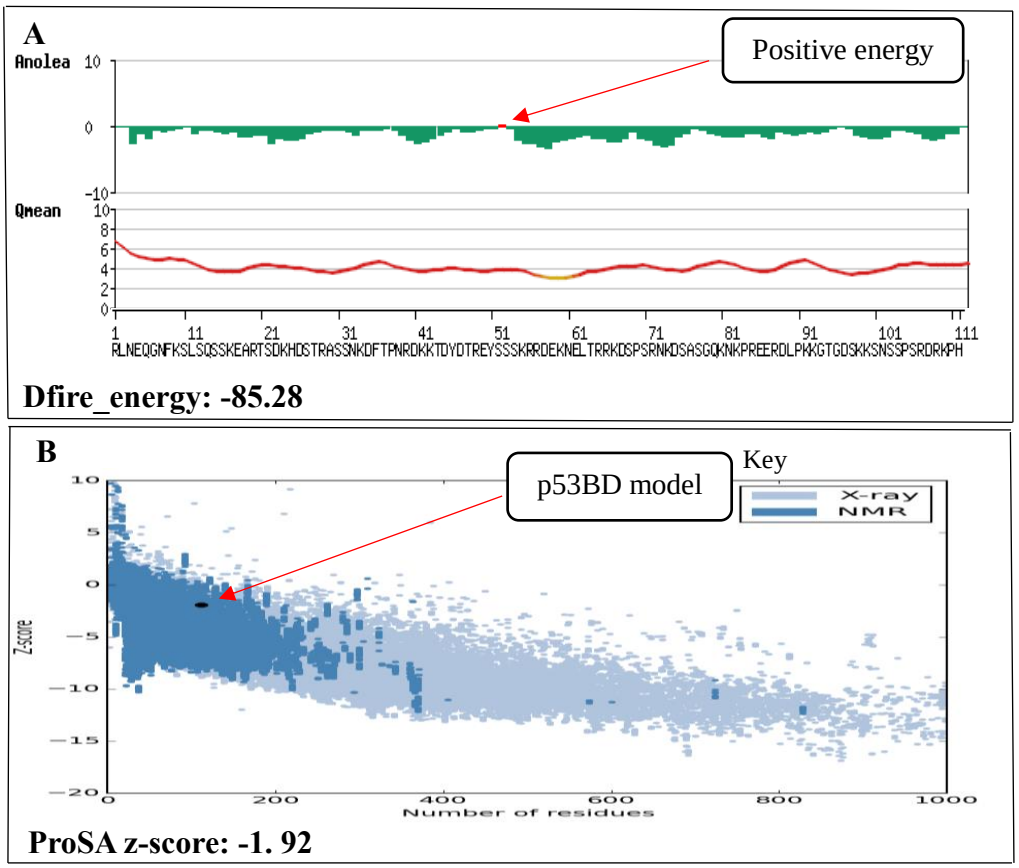


Figure 4.7: Structural quality assessment of Model-A: Resolved using Swiss-model & ProSA web-severs, **[A]** Depicts two local statistics i.e. ANOLEA and QMEAN represented by the top bars (green & red) and the red trend-line respectively. The Dfire_energy is associated with global accuracy of the model. **[B]** ProSA z-score statistics examines the global accuracy by determining where the model's z-score sits in comparison to all structures in current PDB (Protein Data bank), thus the black dot on the graph represents the region at which the model is located relative to all crystallised proteins. The key to the graph is provided on the top-right corner.

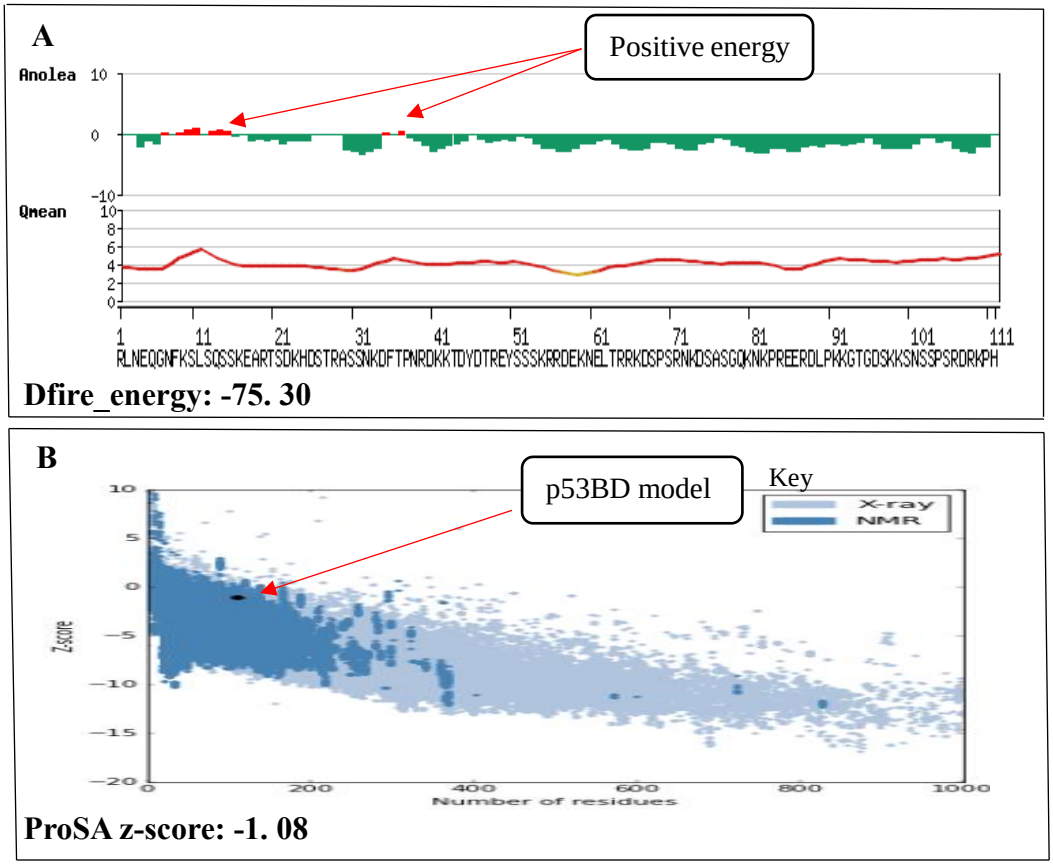


Figure 4.8: Structural quality assessment of Model-D: Resolved using Swiss-model & ProSA web-severs, **[A]** Depicts two local statistics *i.e.* ANOLEA and QMEAN represented by the top bars (green & red) and the red trend-line respectively. The Dfire_energy is associated with global accuracy of the model. **[B]** ProSA z-score statistics examines the global accuracy by determining where the model's z-score sits in comparison to all structures in current PDB (Protein Database), thus the black dot on the graph represents the region at which the model is located relative to all crystallised proteins. The key to the graph is provided on the top-right corner.

4.3 Binding site prediction using FTSite: Binding site prediction server

Prior to receptor-based molecular docking binding sites of all three RBBP6 domains (Figure 4.9-4.11) were determined. This is critical during Grid generation as it specifies residues facilitating the binding co-ordination between the pocket and a given potential drug-candidate. Thus, for successful docking studies binding site prediction is necessary. FTSite binding site prediction server was used, and it predicted three potential binding pockets in all domains. These binding pockets are represented with putative ligand binding sites (net-like structures) and their respective residues.

4.3.1 Binding sites for the Domain With No Name (DWNN)

Three potential binding sites of the DWNN domain were identified (Figure 4.9). These findings would help in discovering potential drugs (agonist / antagonist) through docking experiments. Hydrophilic domains are potential binding or functional sites, meaning that more hydrophobic regions are difficult to target with small organic compounds. Thus, site 1 of the DWNN below would be more efficiently targeted based on the high hydrophilic state it possesses in comparison to the other two sites (site 2 and 3). The DWNN domain has been implicated in mitotic apoptosis, however, there has not been any records of structural work reported. Thus, three docking studies for all sites were performed in search of potential lead compounds from the zinc drug database (Zdd).

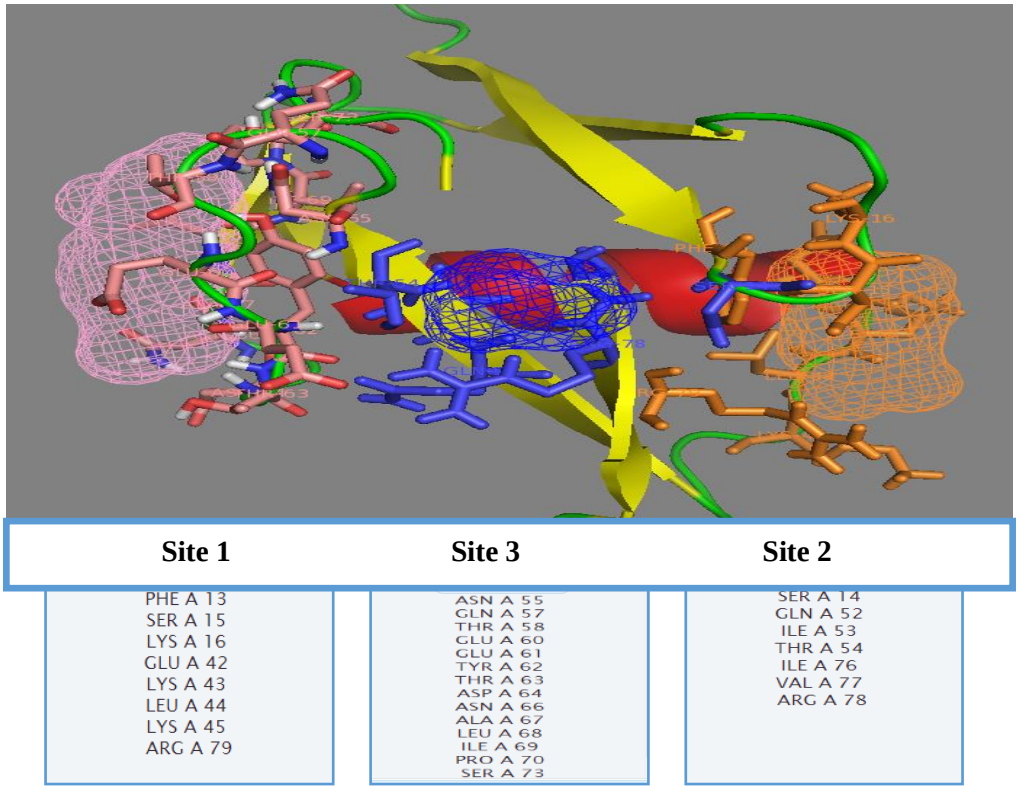


Figure 4.9: Binding site prediction for the Domain With No name (DWNN): Showing three potential drug-binding pockets in pink, blue and golden brown, these were resolved by FTSite and viewed using PyMol. The number of non-bonded interactions with low energy clusters increases from site 3, site 2 and site 1. The residues below also indicate that hydrophilicity increases from site 3, site 2 and site 1. Stick models represent the residues involved in binding.

4.3.2 Binding sites for the RING Finger domain

The RING Finger domain adopts a homodimeric structure in which the interface contains hydrophobic residues similar to U-boxes (Kappo *et al.*, 2012). This hydrophobic interface has intrinsic E3 Ligase activity like MDM2 the prototypical negative regulator of p53. Three binding sites on the RING Finger were determined. Interestingly, site 3 (blue) is located at the exact interface between the monomers and is characterised by hydrophobic residues (Figure 4.10). Since RBBP6 can bind both p53 and pRB, this site is likely to promote ubiquitination and subsequent proteasomal degradation of these pro-apoptotic proteins. Therefore, only one docking job was performed in search for potential drugs that can inhibit the E3 ligase catalytic site (site 3) and re-activate p53 and pRB levels in cancer cells.

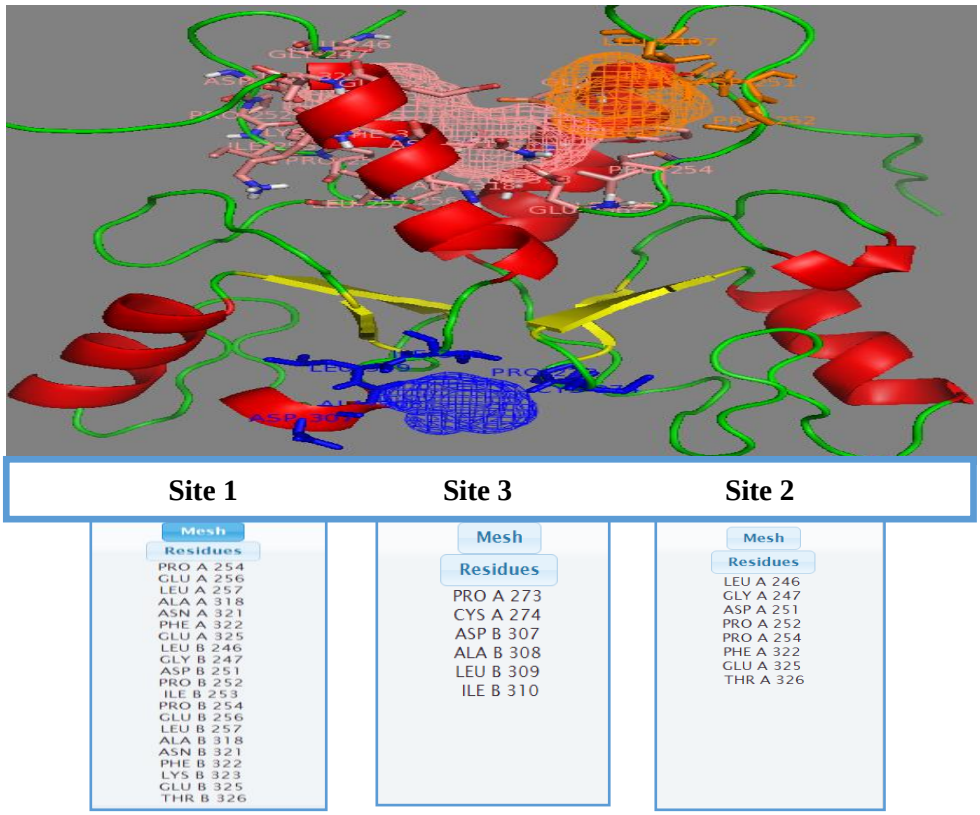


Figure 4.10: Binding site prediction for the RING Finger domain: Showing three critical binding pockets (Site 1, Site 3, Site 2) in pink, blue and golden brown respectively. These were resolved by FTSite and viewed using PyMol. The number of non-bonded interactions with low energy clusters increases from site 3, site 2 and site 1. The residues below also indicate that hydrophilicity increases from site 3, site 2 and site 1 and the A & B represent the two monomers on which the residues are located. Stick models represent the residues involved in binding.

4.3.3 Binding sites for the p53-binding domain (Model-A)

Mode-A was selected as the best representation of the RBBP6 p53BD. Therefore, it was used in binding site prediction and further docking experiments. Three potential binding sites were identified in this domain and site 1 was selected for docking (Figure 4.11). Due to its high hydrophilicity, site 1 is likely the most potential binding region for p53. The p53BD adopts a helix-loop pattern structural fold and the helix regions are more likely to serve as potential binding sites (*i.e*. site 1 and site 3), with only site 2 (yellow) being an exception as it is located within the loop region (Figure 4.11). This confirms the results obtained in section 4.2.1 (Figure 4.3&4.4).

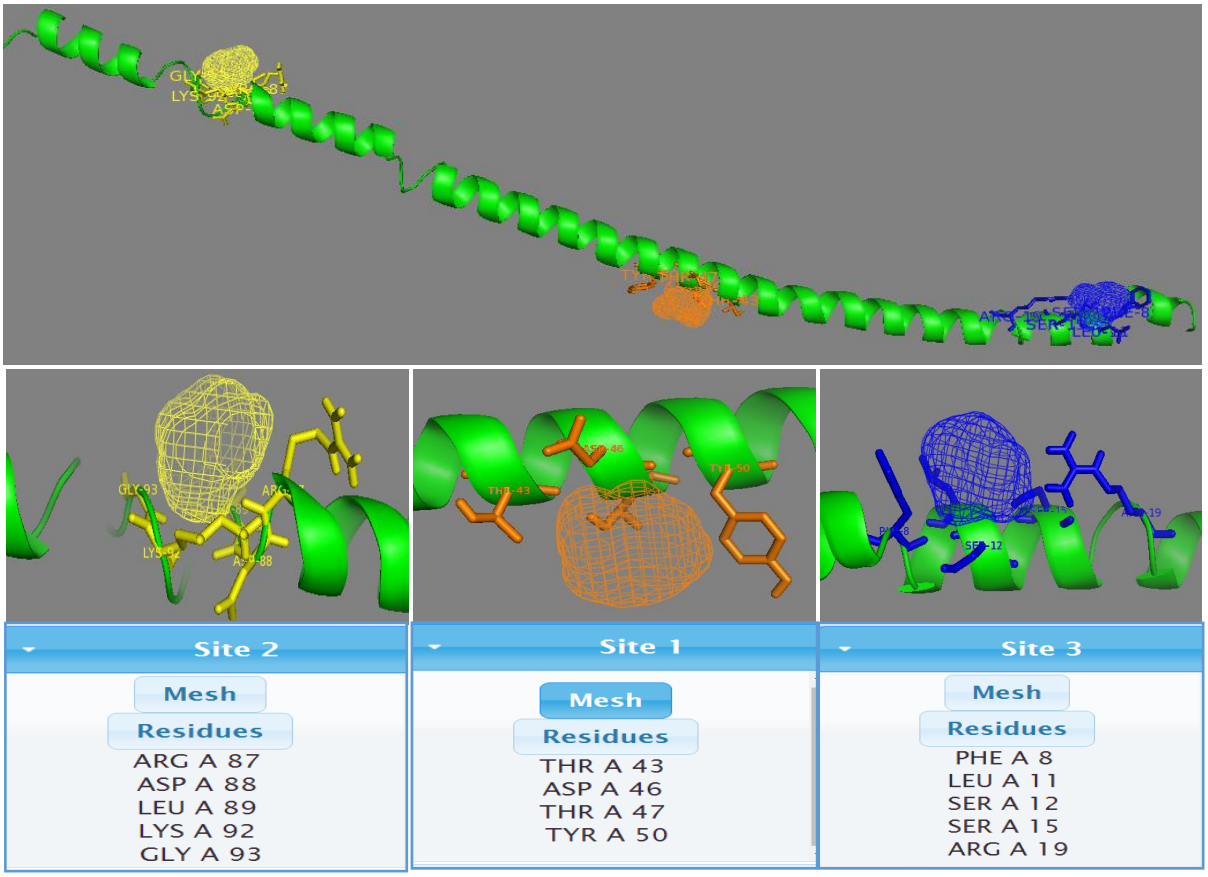


Figure 4.11: Binding site prediction for the RBBP6 p53-binding domain (Model-A): Showing three critical binding pockets in yellow, golden brown and blue respectively, resolved by FTSite and viewed using PyMol. The number of non-bonded interactions with low energy clusters increases from site 3, site 2 and site 1. The residues below also indicate that hydrophilicity increases from site 3, site 2 and site 1. Stick models represent the residues involved in binding.

4.4 Virtual screening & Molecular docking using Schrödinger-Maestro v10.7: Glide SP

Molecular docking studies were performed to determine potential lead compounds for RBBP6. The screened Zinc drug database (Zdd) contains 2924 readily available FDA approved compounds. In the present study, docking calculations were performed onto all three domains of RBBP6 (Figure 4.12-4.17). This was achieved using the Schrödinger-Maestro v10.7: Glide SP, targeting selected binding pockets such as those known to bind cell-cycle regulatory proteins. For comparison purposes, a further docking calculation onto the MDM2 p53-binding domain was performed (Figure 4.18). Additionally, figure 4.19 representing binding posses of the p53 peptide and the most potent nutlins (*Nutlin-2* and MI-219) is shown and compared to the first MDM2 binding lead compound (*Ezetimibe*).

During the docking process, the Glide standard precision algorithm generated over two thousand different lead compounds with multiple posses per each docking run. Therefore, the generated outputs contained approximately over two thousand lead compounds in each site. Analyses focusing on the docking scores, Glide scores and Glide emodel scores of the lead compounds enabled selection of the top 3 leads in each targeted site (Table 4.2).

4.4.1 (+)-Catechin binds into the Domain With No Name (DWNN) binding pocket

The DWNN was docked in all three sites. Among these binding pockets site 1 is the most hydrophilic and thus the most potential binding site. Therefore, here only the first binding pocket (site 1) bound to *(+)-catechin* is reported (Figure 4.12). Interestingly, a phosphate molecule was observed to tightly bind on site 2, suggesting that this may be a phosphorylation site on DWNN. The site 1 lead compound, *(+)-catechin* is an antioxidant, a type of natural phenol that serve as a plant secondary metabolite and belong to a group of flavanols. Catechins are the major components of green tea extracts and possess several properties that tend to have wide implications in disease and lifestyle conditions, such as

cardiovascular pathology, cancer, arthritis as well as obesity (Reygaert, 2017). Here, a very good binding mode consisting of four hydrogen bonds (yellow dotted-lines) and possible hydrophobic interactions is observed between *(+)-catechin* and DWNN (Figure 4.12A). There is also a collection of good steric contacts (green dotted-lines) suggesting a stereo-chemically feasible binding configuration of *(+)-catechin* to DWNN (Figure 4.12B). This finding suggests that the catechins' tumour inhibitory effect is likely via RBBP6 DWNN, indicating that RBBP6 may be involved in their pharmacologic activity.

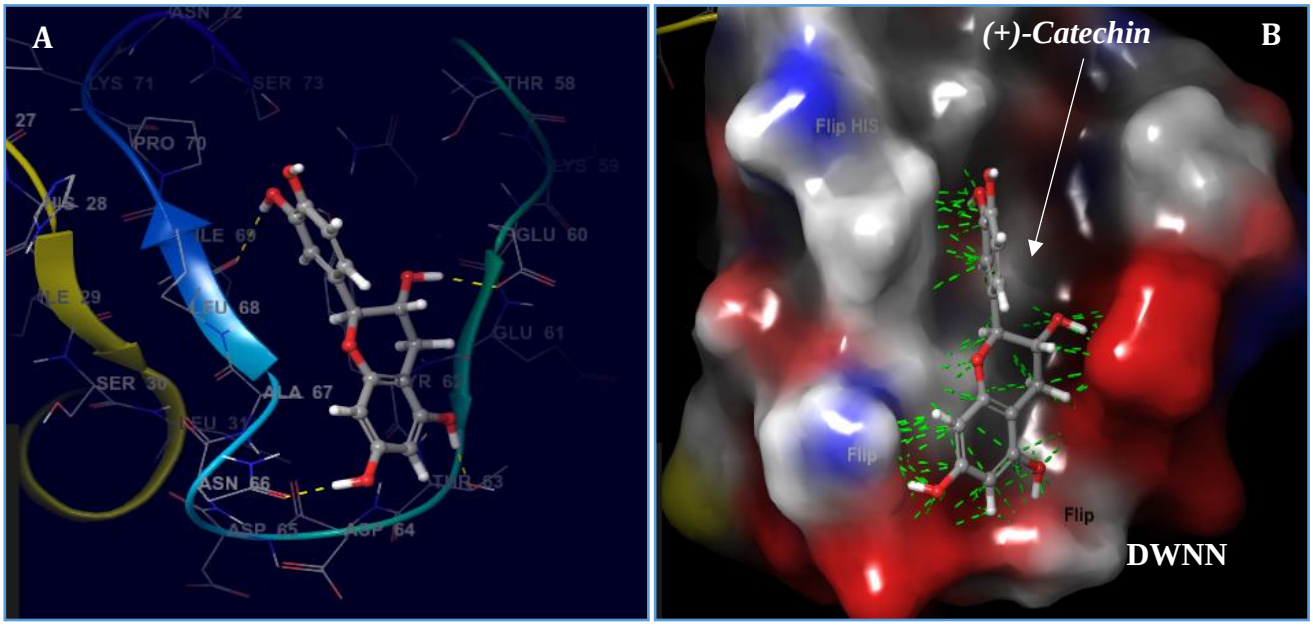


Figure 4.12: The first lead compound *((+)-Catechin)* docked into site 1 of the DWNN: [A] Shows *(+)-Catechin* (ZINC00119985) and DWNN interaction, hydrogen bonds are coloured in yellow dotted-lines and the presence of hydrophobic residues (ILE69, LEU68, ALA67, LEU31, PRO70) within the drug vicinity indicates the possibility of hydrophobic interactions. The binding site is largely composed of *β-sheets and loop* structures (yellow, blue & cyan) **[B]** Represents a residue property surface together with a network of good steric contacts between the *(+)-Catechin-DWNN* complex.

The chemical structure of *(+)-catechin* is composed of several water-soluble regions such as the hydroxyl groups ($^{-}$OH) which makes it a good lead compound (Figure 4.13). This compound possesses two benzene rings and a hydroxyl group attached to carbon 3 of the dihydropyran heterocycle. There are two chiral centres on carbon 2 and 3 of the molecule which makes four diastereoisomers, two in trans-configuration (catechin) and two in cis-configuration (epicatechin) (Figure 4.13A).

Among the catechins, *(+)-catechin* is the most common, but it has a stereoisomer *(-)-catechin.* Glide SP (Standard precision) also computed a ligand interaction diagram of the *(+)-Catechin-DWNN* complex showing all residues at 3Å axis within the binding pocket (site 1) (Figure 4.13B). The presence of ionic groups also suggests the possibility of electrostatic and van der Waals interactions, which would greatly contribute towards the binding affinity of the complex.

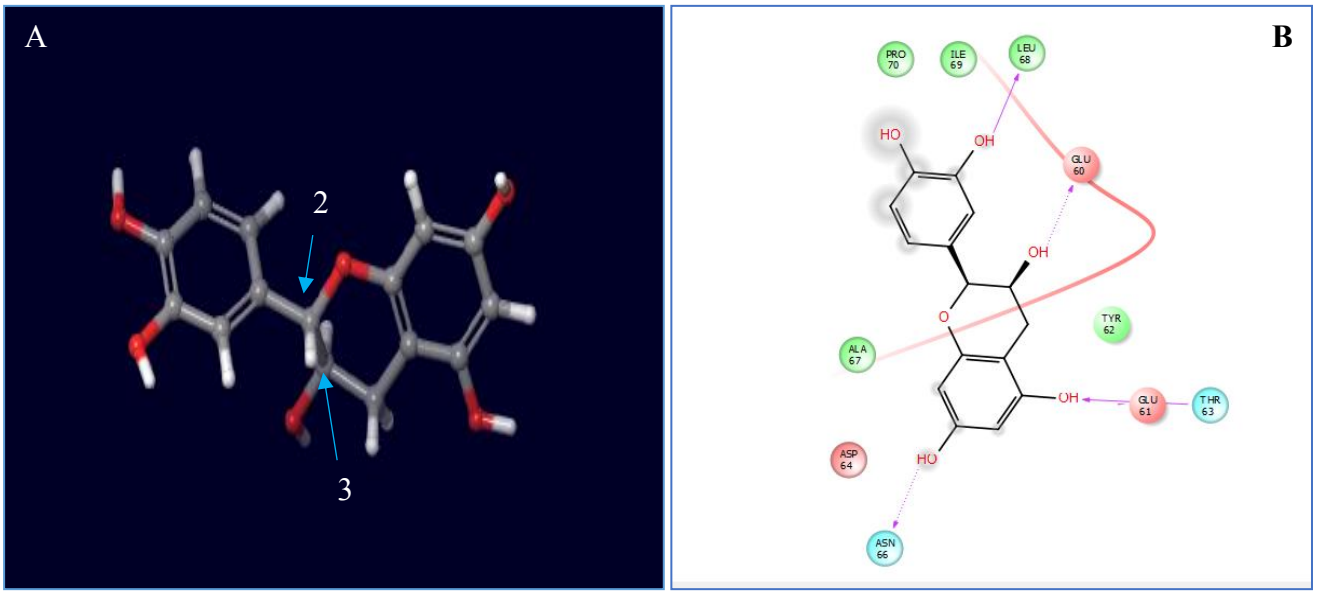


Figure 4.13: Structural composition and interaction analysis of *(+)-Catechin* with DWNN: [A] Shows a ball and stick model of *(+)-Catechin* which is a potential DWNN domain agonist or antagonist, the red, white and grey spheres represents Oxygen (O_2), hydrogens (H), and Carbons (C) respectively, chiral centres are labelled 2 & 3 **[B]** Demonstrates a 2D ligand interaction diagram with a collection of nearby hydrophobic residues as well as four hydrogen bond coordination to LEU68, GLU60, ASN66 and GLU61**.**

4.4.2 Plerixafor binds into the RBBP6 RING Finger E3 Ligase site (site 3)

Plerixafor is an immunostimulant used to mobilise hematopoietic stem cells in cancer patients. This drug is currently utilised synergistically with granulocyte colony-stimulating factor (G-CSF) to release hematopoietic stem cells from the born marrow into the peripheral blood for collection and subsequent autologous transplantation in Non-Hodgkin's lymphoma (NHL) and multiple myeloma (MM) patients (Uy *et al.*, 2008). *Plerixafor* exhibits an antagonistic effect to the CXCR4 chemokine receptor binding

to three acidic residues (ASP171, ASP262 and GLU288) within the pocket and thus blocking binding of its cognate ligand, stromal cell derived factor-1 (SDF-1α).

In the current study, this drug binds tightly (docking score -7.756) to the E3 ligase site of the RING Finger domain implicating RBBP6 as a potential off-target from CXCR4 (Figure 4.14). This finding suggests that *plerixafor* can also block the E3 ligase activity of RBBP6. The figure 4.14 below depicts a range of bonding networks such as hydrogen bonds (yellow dotted-lines) and possible hydrophobic interactions, stipulated by the presence of hydrophobic residues (VAL304, PRO273, LEU309, and ALA308) near the drug within the binding pocket of the RING Finger domain (site 3) (Figure 4.14A). The collection of good steric contacts (green dotted-lines) below also affirms the stereochemical feasibility of the complex, indicating no steric clashes or bad contacts between the drug and neighbouring residues (Figure 4.14B).

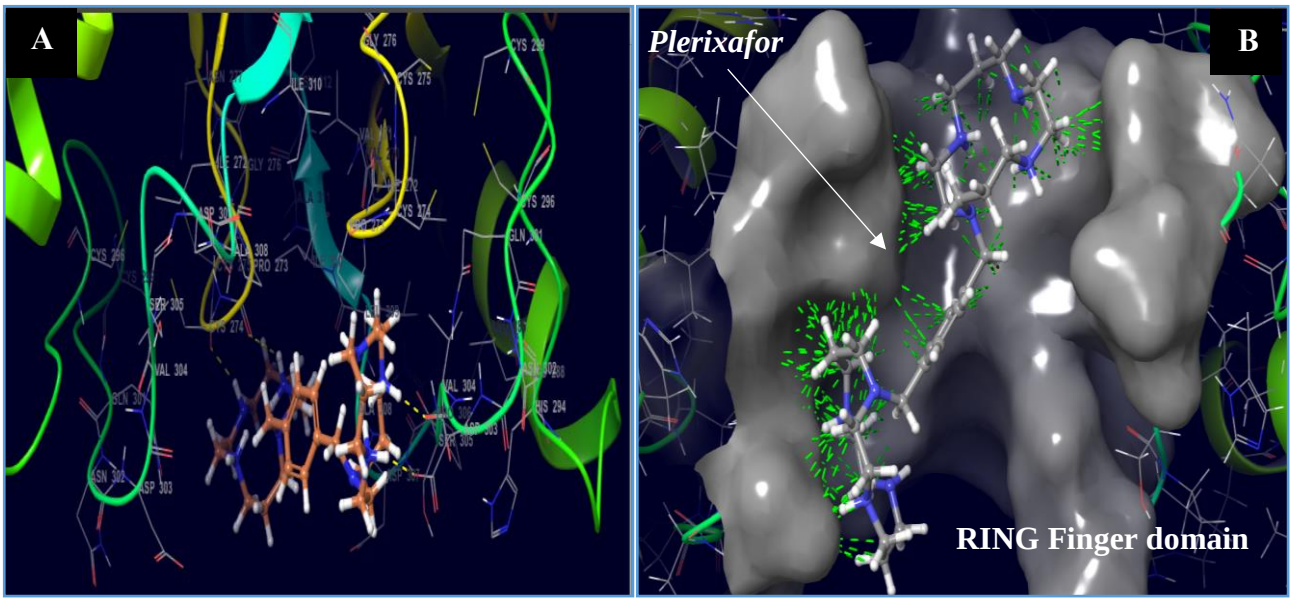


Figure 4.14: The first lead compound (plerixafor) docked into site 3 of the RING Finger domain (E3-Ligase site): [A] Shows the *plerixafor-RING Finger's* hydrogen bonds and hydrophobic interactions, this drug is bound within the interface just below the triple-stranded *β-sheets (light blue & yellow).* Helices and loops are shown in green within the drug vicinity **[B]** Depicts a constant surface along with a network of good steric contacts (green dotted-lines) for the *Plerixafor-RING Finger* domain complex.

Plerixafor adopts a bicyclam structure with two symmetrical rings separated by a *para*-substituted benzene ring (Figure 4.15). The bulkiness of this lead compound enables it to fill-up the entire active site volume of the hydrophobic interface on the catalytic site, suggesting that it exerts complete inhibition of the E3 ligase activity. The ligand interaction diagram below provides a picture of how nearby residues (at 3Å axis) interact with *plerixafor* (Figure 4.15B). These results indicate that four hydrogen bonds (purple dotted-lines) are formed and there is a high density of hydrophobic residues within the drug vicinity. Furthermore, the tight binding of *Plerixafor-RING Finger* domain complex is facilitated by the fact that *plerixafor* is also highly hydrophobic in its chemical nature. The positively charged nitrogen ions (N^+) on both rings indicate the possibility of ionic or electrostatic interactions, perhaps contributing to high binding affinity of the complex.

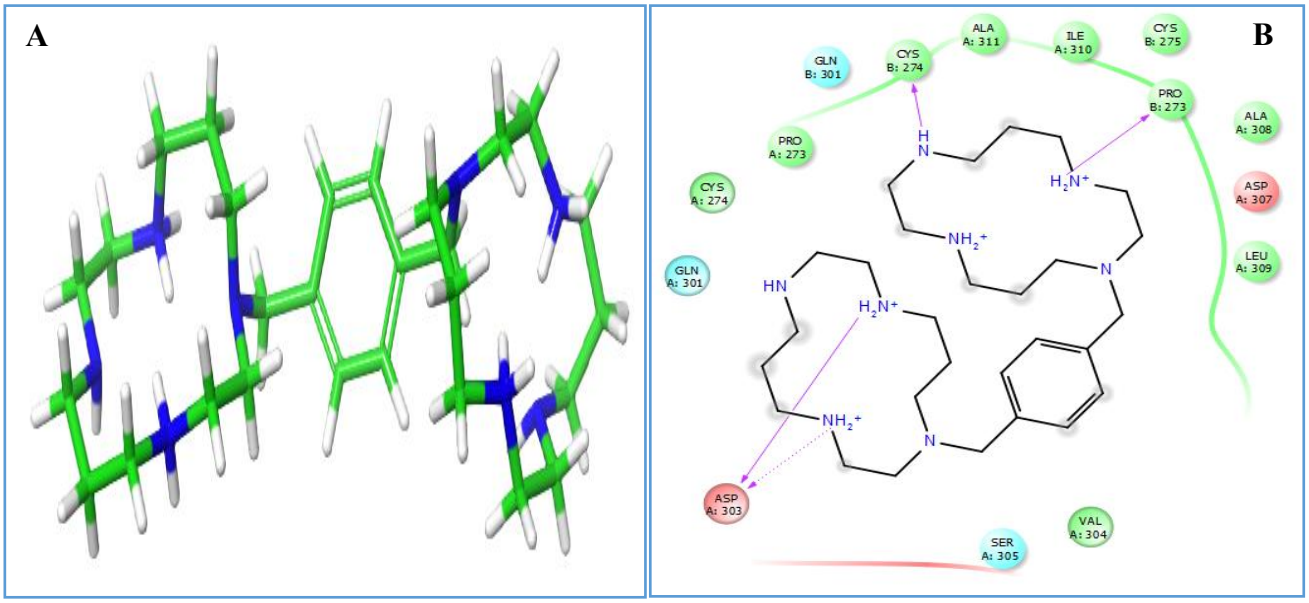


Figure 4.15: Structural composition and interaction analysis of *plerixafor* with the RING Finger domain: [A] Shows a ball and stick model of *plerixafor* (ZINC22443609) a potential RBBP6 RING Finger domain inhibitor, the blue, green and white sticks represents nitrogen (N), carbons (C), and hydrogens (H) respectively **[B]** Demonstrates a 2D ligand interaction diagram with a collection of nearby hydrophobic residues as well as four hydrogen bond coordination (purple arrows) to CYS274, PRO273, and ASP303.

4.4.3 *Nicotinamide adenine dinucleotide (NADH) docks into the RBBP6 p53-binding domain*

NADH is a well-studied co-enzyme produced in many cellular contexts, especially during glucose metabolism and is found in all living cells. Since it exhibits both oxidising and reducing properties it is involved in redox reactions during metabolism. This means that NAD^+ (oxidising agent) accept electrons from other molecules and become reduced to form NADH (reducing agent), which can then be used as a reducing agent to donate electrons to the electron transport chain occurring in the mitochondria. NADH is in turn oxidised by the electron transport chain which pumps protons across the membrane to generate ATP through oxidative phosphorylation (OXPHOS).

In the present study, NADH is observed to bind the RBBP6 p53-binding domain (p53BD) forming four hydrogen bonds and electrostatic interactions with the p53BD hydrophilic pocket (Figure 4.16). The tight binding of NADH to the p53BD is further depicted by the collection of good steric contacts (green dotted-lines) with no bad clashes between the co-ordinating residues (Figure 4.16B). These findings implicate NADH as a good lead compound that can be utilised as a scaffold to develop drugs which could disrupt the RBBP6-p53 interaction.

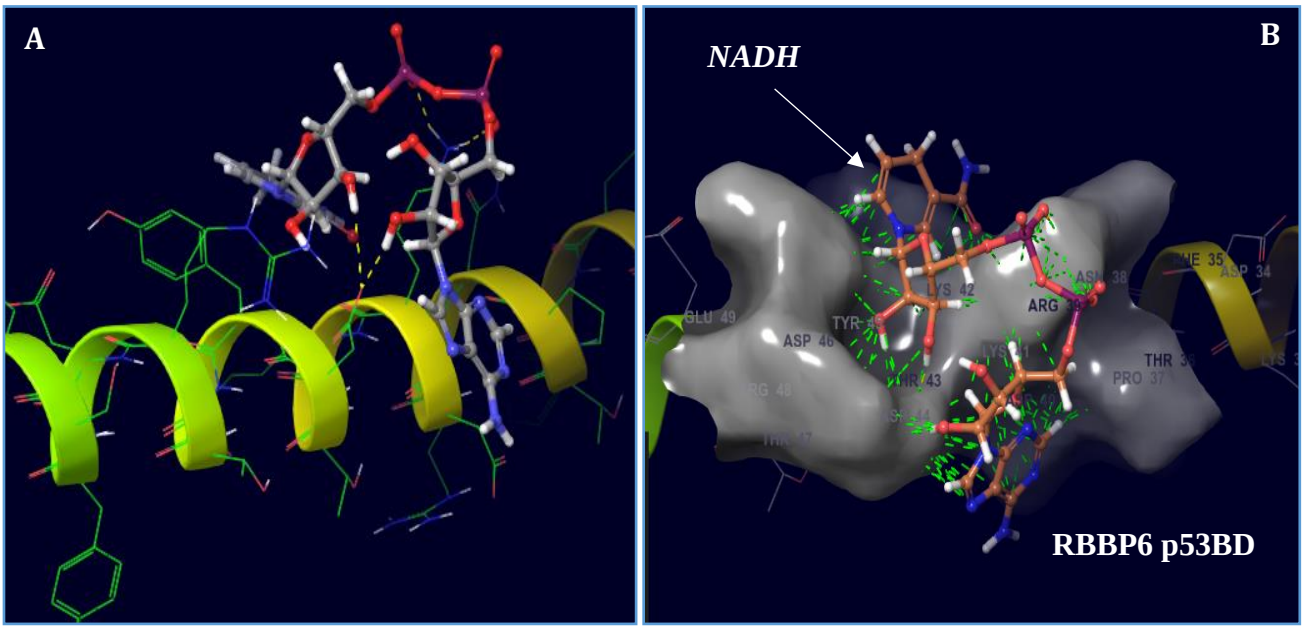


Figure 4.16: **The first lead compound (NADH) docked into site 1 of the RBBP6 p53BD: [A]** Shows the NADH (ZINC53682927) and p53BD interaction, hydrogen bonds are coloured in yellow dotted-lines and hydrophilic interactions are expected since the first binding site (hydrophilic) was targeted in this case. The binding site is within the helix region of the domain (greenish-yellow) **[B]** Depicts a constant surface along with a network of good steric contacts (green dotted-lines) for the drug-receptor complex.

The NADH structure is composed of two nucleotides (nicotinamide and adenine) joined together through their phosphate groups, thus it's referred to as a dinucleotide (Figure 4.17A). The presence of these phosphate groups provides negative charges and thus determine the overall net-charge of NADH. This further indicates the possibility of ionic interactions which greatly contributes to the docking score and thus binding affinity of the interacting complex. The ligand interaction diagram was measured at 3Å axis around the drug and shows four hydrogen bonds co-ordinated by nearby residues (Figure 4.17B).

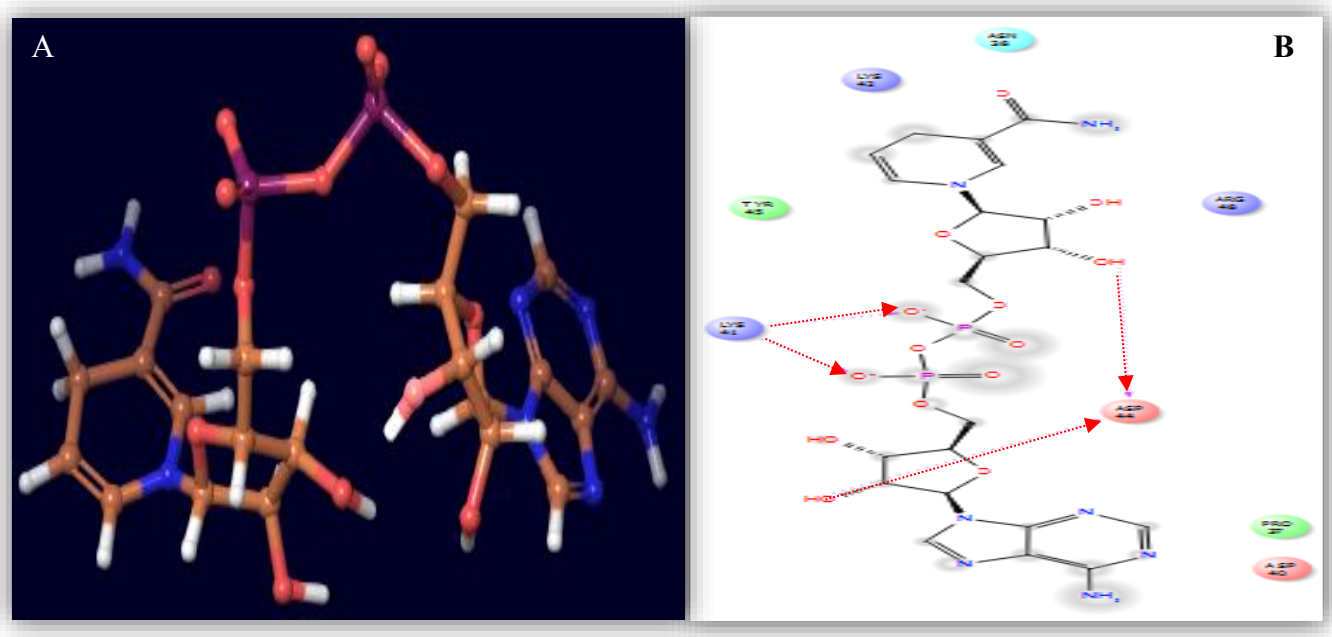


Figure 4.17: Structural composition and interaction analysis of *Nicotinamide adenine dinucleotide* (NADH) with the RBBP6 p53BD: [A] Shows a ball and stick model of NADH (ZINC53682927) a potential RBBP6 p53-binding domain inhibitor, the maroon, pick, blue, brown and white sticks represents phosphates (P), oxygen (O_2), nitrogen (N), carbon (C) and hydrogen (H) respectively **[B]** Demonstrates a 2D ligand interaction diagram with a collection of nearby residues as well as four hydrogen bond (red arrows) coordination to ASP44 (red), and LYS41(blue).

4.4.4 Ezetimibe (Zetia) docks into the MDM2 p53-binding domain (hydrophobic pocket)

Multiple studies have shown that the MDM2 hydrophobic cleft binds the p53-transactivation domain and thus targeting p53 for proteasomal degradation. For comparison purposes, the Zinc drug database was docked into the MDM2 p53-binding domain to determine if there is any match among the lead compounds picked-up by the RBBP6 p53-binding domain, and thus whether there are some expected off-target effects. Here, *ezetimibe* was observed to bind the MDM2 hydrophobic cleft with high affinity (with a docking score of -7.885) and no lead compound matches were noted between these two proteins (MDM2 and RBBP6) (Figure 4.18).

Ezetimibe is a drug that falls within a class of lipid-lowering compounds. This drug is currently used to selectivity inhibit intestinal absorption of cholesterol and related phytosterols from dietary and biliary sources. *Ezetimibe* has been shown to have a direct impact on the reduction of plasma total

cholesterol concentrations in several pre-clinical models (Catapano, 2001). This is because the drug inhibits cholesterol absorption from the small intestines, thus reduces the amount of cholesterol available to liver cells. This forces the cells to absorb cholesterol from the circulatory system, hence lowering total plasma cholesterol concentration.

Ezetimibe undergoes glucuronidation (addition of glucuronic acid) to a single metabolite, and then it targets and blocks the Niemann-Pick C1-like 1 (NPC1L1) protein which keeps cholesterol in the intestinal lumen for excretion. The glucuronidation metabolite form of *ezetimibe* is more water-soluble and binds NPC1L1 protein with higher affinity than *ezetimibe* to prevent cholesterol absorption from the gastrointestinal tracts. Interestingly, through docking experiments, the present study has determined that *ezetimibe* forms a single hydrogen bond with VAL93 (just like the p53 peptide does with TRP23) and mimics all three key residues of p53 (*i.e.* TRP23, LEU26, and PHE19) upon binding the MDM2 hydrophobic pocket (Figure 4.18A and 4.19A).

The collection of good steric-contacts (green dotted-lines) upon binding conforms the stereochemical feasibility and high affinity of the complex (Figure 4.18B). The structural composition of *ezetimibe* is known as monobactam and comprises the beta-lactam ring alone or fused to another ring (Figure 4.18C). It belongs to a class of organic compounds under the organoheterocyclic compounds superclass. This drug is currently used in patients suffering from hyperlipidaemia or hypercholesterolemia, but the ligand interaction diagram below shows that *ezetimibe* tightly binds the MDM2-p53BD with hydrophobic and ionic interactions at 3Å axis around the drug (Figure 4.18D). This finding is likely to have significant impact in cancers with high MDM2 expression.

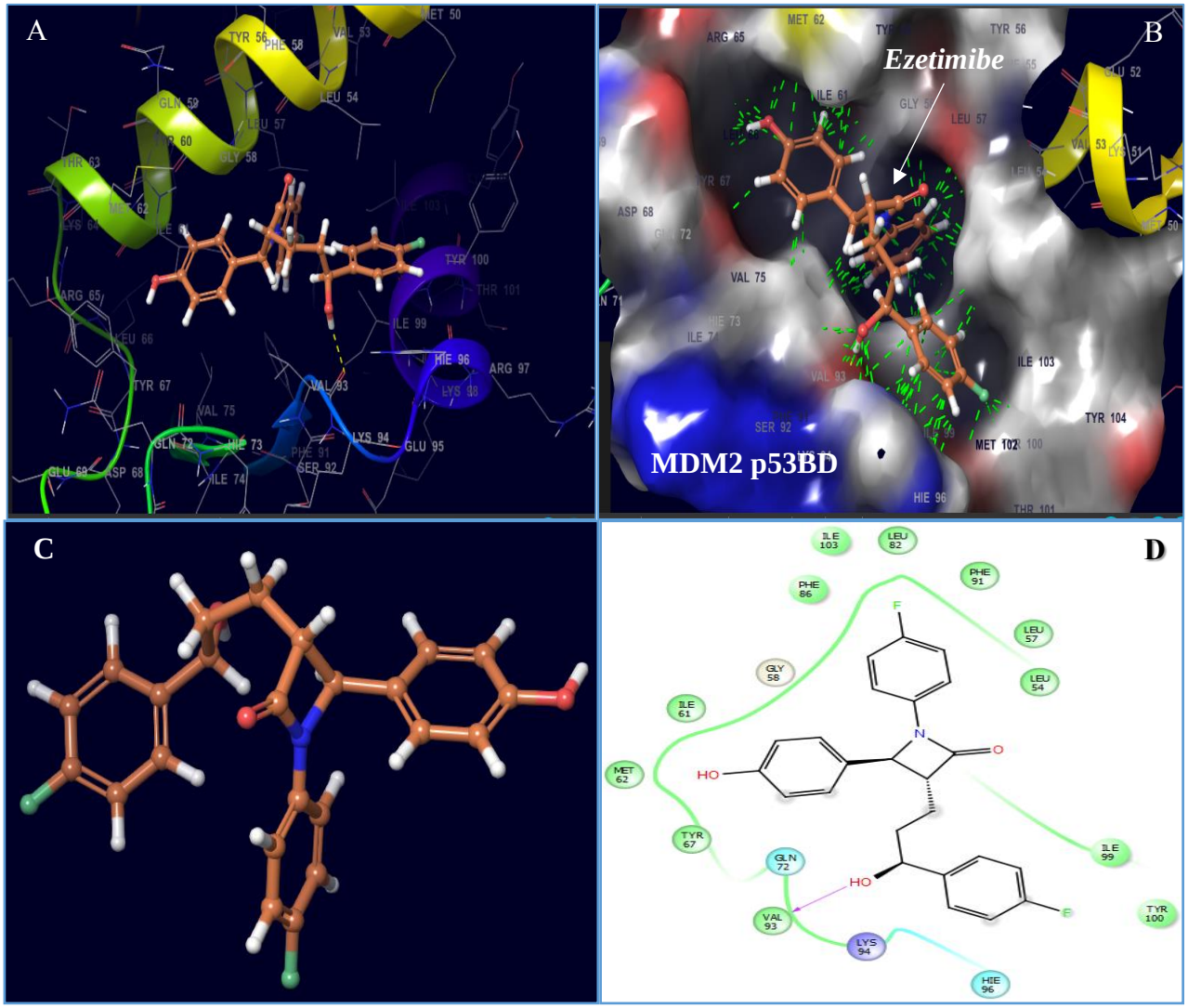


Figure 4.18: The first compound (*Ezetimibe*) docked into the p53-binding domain of MDM2 (hydrophobic pocket): [A] Shows *Ezetimibe* and MDM2 interaction, a single hydrogen bond with VAL93 (yellow dotted-line) and possible hydrophobic interactions are observed. **[B]** Depicts a residue property surface along with a network of good steric contacts (green dotted-lines) of the *Ezetimibe-MDM2* complex. This surface representation shows the pocket-like nature of the binding site and how *Ezetimibe* accurately mimics the three critical p53 binding residues (PHE19, TRP23, and LEU26). **[C]** Shows a ball and stick model of *Ezetimibe* (ZINC03810860) a potential MDM2 p53-binding domain inhibitor, where the white, green, blue and red spheres represent hydrogen (H), fluorine (F), nitrogen (N) and oxygen (O_2) respectively. **[D]** Demonstrates a 2D ligand interaction diagram with a collection of nearby residues and a single hydrogen bond (purple arrow) co-ordination to VAL93 (green).

The observed binding mode affirms an accurate inhibition of p53 binding, thus should prevent the formation of MDM2-p53 complexes. A comparative view of both figure 4.18 and 4.19 shows that *Ezetimibe* is not structurally like the well-known nutlins (Nutlin-2, Nutlin-3a and MI-219), but seem

to mimic the p53 peptide upon binding the MDM2 hydrophobic pocket more accurately than them (Figure 4.18 and 4.19). However, this does not negate the fact that nutlins are very good inhibitors of the MDM2-p53 interaction that have been thoroughly studied thus far, but rather presents another promising class of compounds which interacts with MDM2 in a similar function.

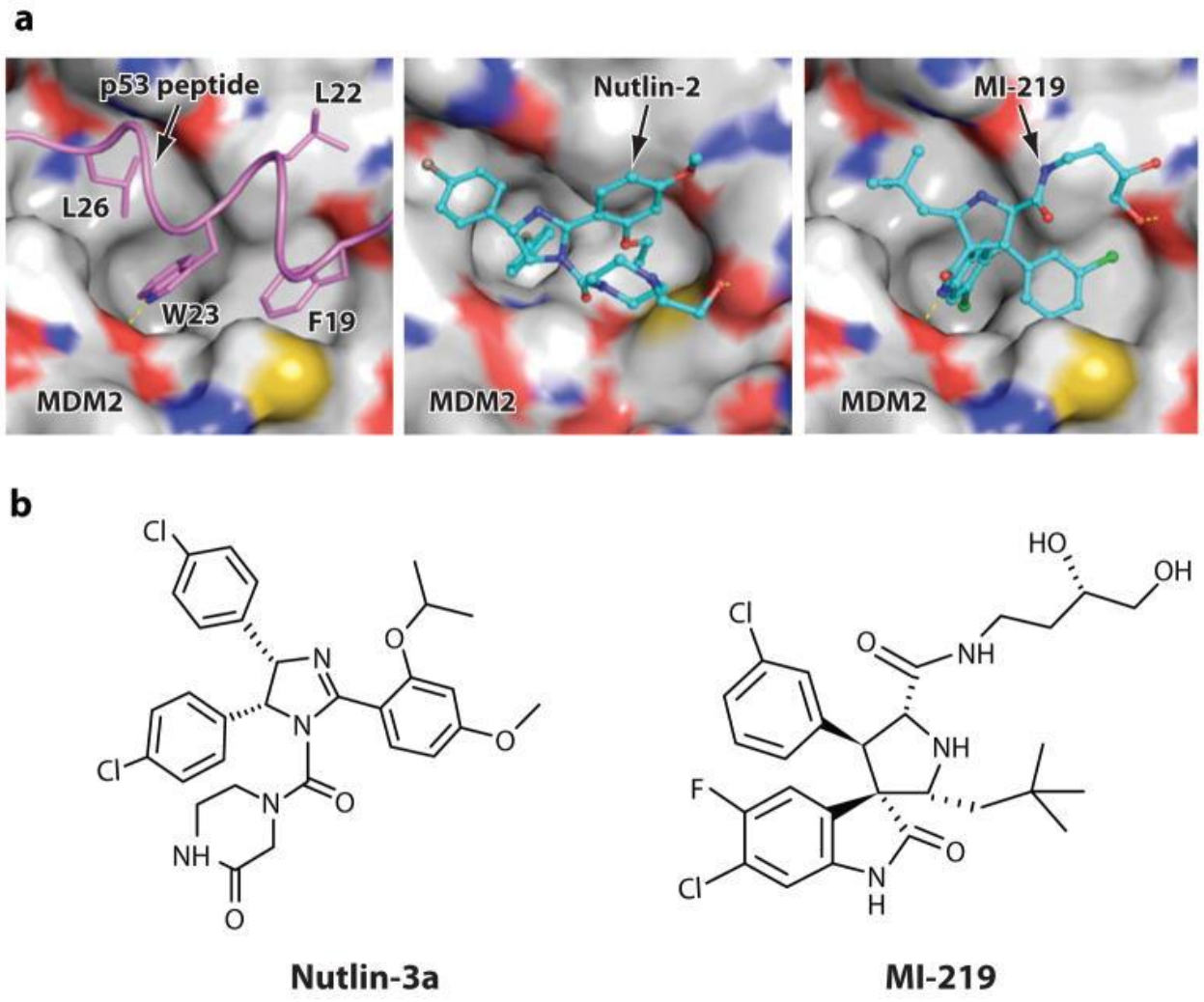
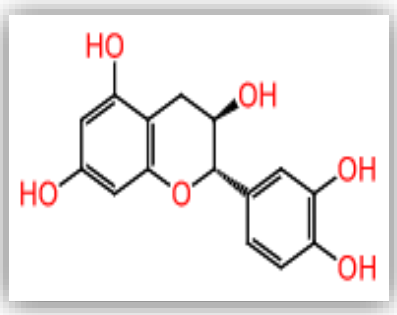
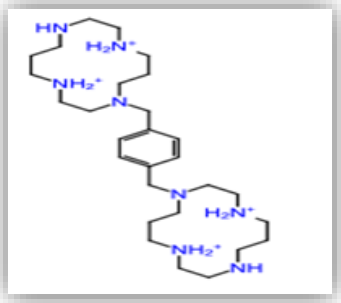
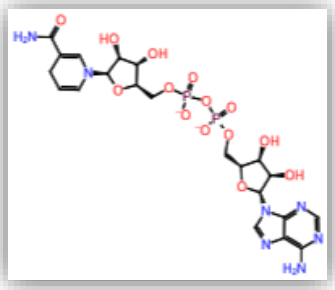
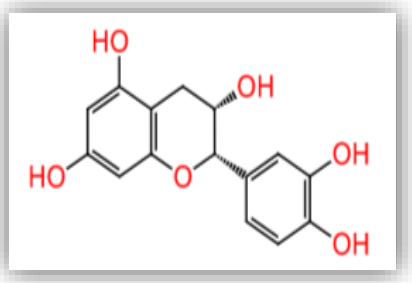
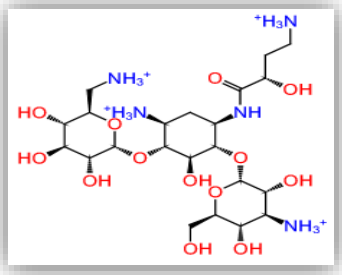
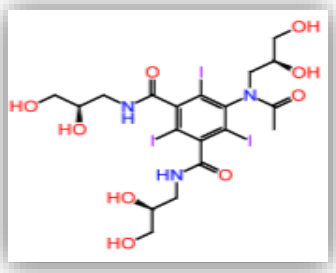
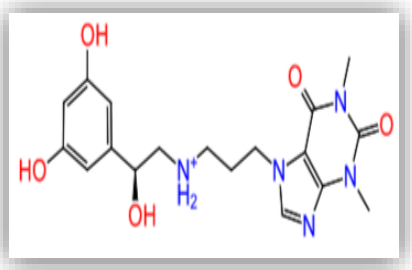
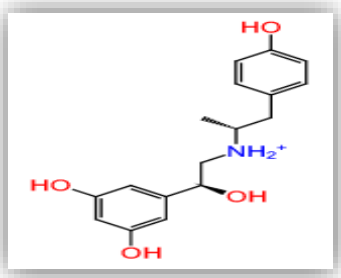
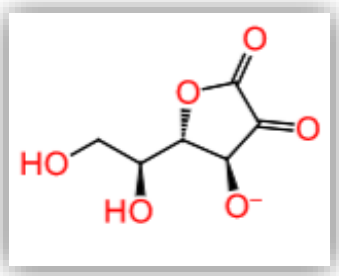


Figure 4.19: Structural binding modes of the p53 peptide (pdb:1YCR), the cis-imidazoline analog, Nutlin-2 (pdb:1rv1), and the spiro-oxindole analog MI-219 to MDM2: [A] Shows the p53 transactivation domain residues that are very critical for the MDM2-p53 interaction in stick representation, Nutlin-2 is shown in ball and stick models with carbons in cyan, nitrogen in blue, oxygen in red, and bromine in brown. The predicted binding of MI-219 to MDM2 is shown with carbons in cyan, nitrogen in blue, oxygen in red, fluorine in light blue, and chlorine in green. The surface representation of MDM2 in each case is shown with carbons in gray, nitrogen in blue, oxygen in red, and sulfur in yellow. Hydrogen bonds are depicted with yellow lines. **[B]** Shows chemical structures of Nutlin-3a and MI-219, the two potent and specific small-molecule inhibitors of the MDM2-p53 interaction **(Adapted from: Shangary *et al.*, 2009).**

The RBBP6 domains have shown to bind multiple potential drugs upon virtual screening of the zinc drug database (Zdd). This database was chosen primarily based on two critical parameters, namely; database size and FDA approval. Therefore, the small zinc drug database size of 2924 compounds enabled that all six docking studies are performed with high quality, high processing power, and maximum storage of outputs. Hence the entire process was also not financially expensive. Furthermore, the used of an FDA approved database would speed-up the optimisation process since the obtained lead compounds are already "drugs" themselves (*i.e.* they already possess drug-like properties) targeting different proteins.

The targeted dataset has not been previously screened for RBBP6, thus these drugs could serve as activators or inhibitors depending on the domain and site at which they bind. In Glide, the Zdd was firstly prepared and generated compounds about double its original size (from 2924 to 4909). The newly generated compounds contained different possesses of the original dataset to account for possible stereoisomers and had adopted a lowest energy possible 3-dimensional configuration suitable for docking. This was to further permit ligand flexibility during docking whilst the receptor is kept rigid, and a top 3 of over thousand compounds were selected in each domain (Table 4.2). Shown in this chapter is raw A, which are the first lead compounds in each of the RBBP6 domains. However, all top three are good lead compounds of RBBP6 that could serve as scaffolds for further optimisation depending on the domains at which they bind and its context in cancer therapy.

Table 4.2: Top three lead compounds of RBBP6 shown along with their binding energies.

Drugs	Domain With No Name (DWNN)	RING Finger domain	p53-binding domain
A	 glide energy: -34.418 docking score: -6.784 glide emodel: -47.466	 glide energy: -60.179 glide emodel: -121.038 docking score: -7.756	 glide emodel: -61.698 docking score: -6.136 glide energy: -47.387
B	 glide energy: -35.265 docking score: -6.588 glide emodel: -48.361	 glide energy: -68.137 glide emodel: -128.478 docking score: -7.566	 glide emodel: -56.8 docking score: -5.677 glide energy: -42.891
C	 glide energy: -41.212 docking score: -6.246 glide emodel: -57.786	 glide energy: -45.334 glide emodel: -70.443 docking score: -7.494	 glide emodel: -32.215 docking score: -5.317 glide energy: -22.14

CHAPTER FIVE: DISCUSSION

RBBP6 is a eukaryotic multifunctional protein and is implicated in diverse cellular functions such as nucleic acid metabolism, cell cycle regulation (Scott *et al.,* 2003 and Kappo *et al.,* 2012), as well as carcinogenesis as it associates with both p53 and pRB which are key cell cycle regulators largely linked to carcinogenesis (Witte and Scott, 1997). Overexpression of RBBP6 in cervical cancer also indicated that it positively influences the S-phase of the cell cycle, and ultimately promotes cell proliferation. Conversely, the down regulation of RBBP6 is associated with apoptosis (Moela *et al.,* 2016). These findings suggest that RBBP6 may be a plausible target for cancer therapy (Ntwasa, 2008).

In this study, a computational approach was taken to determine whether RBBP6 is druggable. The 3-dimensional structures of three RBBP6 domains (DWNN, RING Finger and p53-binding domain) were targeted as potential drug binding sites. Molecular docking studies were conducted using the zinc drug database. Three-dimensional structures of the DWNN and the RING Finger domain are already known but the p53-binding domain tertiary structure has not been reported. In this study, the p53-binding domain was modelled using I-TASSER and *e*Thread-Modeller prior the ligand docking process (Figure 4.1&4.2). Three binding sites were predicted using FTSite, and docking studies produced 3 potential drugs, namely; (+)-catechin, plerixafor and nicotinamide adenine dinucleotide binding to DWNN, RING Finger and the p53BD respectively. Two of these are known drugs but the one binding to the p53BD is a novel drug considered as lead compound worthy for further development. These findings suggest that RBBP6 can bind small molecular agonist or antagonist, implying that it is druggable.

For comparison, a docking study was conducted targeting the MDM2 p53-binding domain. This was to determine if there are any similar molecules binding to the p53-binding domains of RBBP6 and MDM2. The MDM2 p53BD binds to Nutlins which are in clinical trials for treatment of certain cancers (Burgess *et. al.*, 2016).

The Zdd used for docking does not contain Nutlins, however it has a collection of piperazine compounds (class of nutlins) which did not show any binding with MDM2. The lack of matches on the retrieved lead compounds between these proteins confirms that there is no link or similarity between them.

5.1 The I-TASSER model (Model-A) seem to be the best possible representation of the RBBP6 p53-binding domain.

The RBBP6 p53-binding domain was modelled using I-TASSER, which attempts to efficiently mimic *in vivo* molecular dynamics, by using a hierarchal approach that consistently drives the threading alignments closer to native state. The models generated by I-TASSER are represented together with their local accuracy prediction graphs which clearly stipulates structural differences between the predicted models and native state, at a residue-per residue basis (Figure 4.1(1-3)).

In this study, it has been observed that Model-A possess minimal deviation from native state as compared to all models generated by I-TASSER. Furthermore, this model constitutes a *c-score* of -2.57, which is a good score higher than that of the other I-TASSER models. However, this *c-score* is less than -1.5 indicating that the global topology of Model-A still needs further refinements. This might be because, firstly, the domain was modelled separately from the rest of the protein chain, thus losing the interaction influence and molecular dynamics of the neighbouring domains. Secondly, the modelling process assumed that the p53-binding domain functions as a monomer, whilst homo or hetero-dimerization is also possible, and this could change the binding interface appreciably.

These results were corroborated by the models generated by *e*Thread-Modeller because there is consistency of the helix-loop pattern generated, indicating that the native configuration of the domain is likely composed of helices and coil / loop structures with no beta sheets. Altogether, using molecular modelling, these results suggests that the RBBP6 p53BD adopts a native helix-loop pattern structural fold.

Furthermore, both I-TASSER and *e*Thread-Modeller models (Model-A and Model-D) were assessed for structural quality using Swiss-Model and ProSA (Protein Structure Assessment) web-servers. Quality assessment was based on various statistical parameters namely; Anolea, Dfire, QMEAN, ProCheck, and ProSA Z-score. These parameters cover both local and global quality assessments. The Anolea graph evaluates the "Non-Local Environment" (NLE) of each heavy atom in the molecule, and in this instance Model-A constitutes a range of green bars (negative energy values) in the entire Anolea graph despite a single red bar (SER53) (Figure 4.7A). This implies that 99.5% of residues in this model are in favourable energy environments. The Dfire_energy score of the model is also quite negative (-85.28), implicating that the model is very close to native state (Figure 4.7A). The ProSA Z-score is -1.95, which places the model at an NMR region of previously crystallised protein structures. Therefore, these results confirm that the Model-A is quite closer to native state, and thus depicts minimal deviation to known crystallographic structures as compared to others (the four and five models generated by I-TASSER and *e*Thread-Modeller respectively).

Structural quality assessment was also performed to the first model (Model-D) generated by *e*Thread-Modeller. The Anolea graph results have further demonstrated the presence of unfavourable energy residues, thus implying that there are 91.9% residues in the model occupying favourable energy environments. This percentage has been significantly reduced when compared to the 99.5% in Model-A, the assumption is that the other four models generated by *e*Thread-Modeller would even constitute a lesser percentage of residues with favourable energy environments. The DFire energy score is also less negative when compared to that of Model-A, meaning that Model-D is not that closer to native state in comparison. Therefore, it appears statistically feasible to propose that Model-A is the best possible representation of the RBBP6 p53-binding domain. However, its global topology would still need further refinements as previously stated.

5.2 The binding of catechins to the RBBP6 DWNN may have positive implications in cancer

Catechins have numerous health benefits although its mechanism of action is not currently well-understood. It is known, however, that catechins ((2R, 3S)-2-(3, 4-dihydroxyphenyl)-3,4-dihydro-2H-chromene-3,5,7-triol) are the major components of green tea, and several studies (*in vitro*, *in vivo* and epidemiological studies) have shown that consumption of green tea may significantly reduce risk to cancer (Khan *et al.*, 2007). These green tea extracts (epigallocatechin-3-gallate (EGCG)), have been further shown to inhibit tumour invasion and angiogenesis which are crucial for tumour growth and metastasis (Lecumberri *et al.*, 2013).

Camellia sinensis is a plant derived beverage that is largely consumed as green tea in Asia, parts of North Africa, the United states and Europe (Khan *et al.*, 2007; Chacko *et al.*, 2010). Therefore, a very limited number of individuals with cancer (especially skin, lung, prostate, breast, Liver and Gastrointestinal cancers) is reported in these parts of the world compared to other continents and countries. Consequently, beneficial effects on human health are attributed to green tea (Khan *et al.*, 2007).

In this study, using molecular docking, the two trans-stereoisomers (*(+)-catechins and (-)-catechins*) of catechin are observed to tightly bind the RBBP6 DWNN. The *(+)-catechin* binds more tightly (Glide energy: -34.418, Glide emodel: -47.466, docking score: -6.784) to the domain than its corresponding stereo-isomer, *(-)-catechins* (Glide energy: -35.265, Glide emodel: -48.361, docking score: -6.588) (Figure 4.12). Thus, making it the top lead compound (within the Zinc drug database) upon binding the RBBP6 DWNN. These are interesting findings since *catechins* are known to have tumour inhibitory effects.

Knock down studies in human cancers have revealed that tumour progression is related to increased levels of DWNN (Mbita *et al.*, 2011). Indicating that this domain is involved in mitotic apoptosis, as

its over-expression led to an increased G2/M cell cycle arrest (Mbita *et al.*, 2011). Therefore, if the tumour inhibitory effects of *catechins* are via the RBBP6 DWNN, they could serve as activators (agonists) of this domain.

5.3 Plerixafor binds to the catalytic site (hydrophobic interface) of RBBP6

Before its current use as an immunostimulant to increase white blood cell count by significantly increasing the number of hematopoietic stem cells (HSC) circulating in the peripheral blood, *plerixafor* (1, 4-Bis ((1, 4, 8, 11-tetraazacyclotetradecan-1-yl) methyl) benzene) was primarily designed for HIV-infection therapy. This is because it targets the CXCR4 chemokine receptor, which is a co-receptor for Human Immunodeficiency virus (HIV-1) entry into the cell. HIV-1 enters the cell via CD4 which is the main receptor on the cell surface, however CXCR4 and CCR5 serve as co-receptors that take-up T-lymphotropic (T-tropic) and M-lymphotropic (M-tropic) HIV-1 strains respectively (Feng *et al.*, 1996; Wu *el at.*, 1996).

During asymptomatic stages of the disease, the M-tropic HIV-1 strains enters the cell via CCR5 co-receptors (Alkhatib *et al.*, 1996), and the T-tropic HIV-1 strains targeting the CXCR4 eventually replaces them later as they are associated with rapid disease progression (Cheng-Mayer *et al.*, 1988). Therefore, *plerixafor* would target the CXCR4 and block the HIV-1 co-receptor function. Thus, it was the first drug with antagonistic effect to CXCR4 to enter clinical trials for the treatment of HIV infection. Unfortunately, it was found to exhibit cardiac toxicity before further clinical development could be performed.

However, this led to the identification of an unexpected side effect and was later hypothesised that *plerixafor* may find its application on hematopoietic stem cell (HSC) mobilisation for the purposes of peripheral blood stem cell transplantation (PBSCT) (De Clercq, 2009). Thus, *plerixafor* has been

approved recently (2009) by the European Medicines Agency (EMEA) as a drug to help patience with lymphoma and multiple myeloma since their cells poorly mobilise.

In the present study, molecular docking studies show that *plerixafor* binds tightly to the RBBP6 RING Finger domain (Figure 4.14). This observed binding mode possesses good energies (Glide energy: -60.179, Glide emodel: -121.038, docking score: -7.756), suggesting that RBBP6 might be involved in hematopoietic stem cell mobilisation. Altogether, these results suggest that the cancer application of *plerixafor* may be via the RING Finger domain and can have positive implication in cancer patients with high RBBP6 expression (such as cervical, lung, breast, colorectal, prostate, head & neck as well as pencreatic cancers). This binding mode suggests a complete inhibition of the RBBP6's E3 ligase activity, thus likely to reactive intracellular levels of pro-apoptotic proteins, such p53 and pRB. RBBP6 contains p53 and pRB binding domains, hence these proteins are likely substrates ubiquitinated by the RING Finger domain.

5.4 Nicotinamide adenine dinucleotide (NADH) binds the RBBP6 p53-binding domain possibly inhibiting p53

Using Model-A, potential p53 inhibitors were investigated through docking experiments. Interestingly, the co-enzyme Nicotinamide adenine dinucleotide (NADH) was determined and shown to bind this domain with high affinity (Figure 4.16). This binding affinity was estimated using the docking score to give an indication of the bond-strength between NADH and the RBBP6 p53BD, and a docking score of -6.136 implying high affinity was observed (Table 4.2).

In metabolism, NADH is involved in redox reaction and exhibits both oxidizing and reducing agent species in solution. Although this is a well-studied molecule its binding to RBBP6 is yet to be reported. This study demonstrated that NADH binds RBBP6 with good binding energies (Glide energy: -47.387, Glide emodel: -61.698, docking score: -6.136) making it the best molecule (within the zinc drug database) that binds the p53-binding domain of RBBP6 (Table 4.2). Therefore, this finding sets-up an

excellent premise on which pre-clinical drug inhibitors could be developed, thus implicating NADH as a good lead compound that could be optimised to block RBBP6-p53 interaction leading to the re-activation of p53 levels in cancer cells. The inhibition of RBBP6-p53 binding leads to the stabilisation of p53 in cancer cells consequently promoting p53-dependent cell cycle arrest or apoptosis. This study has identified a possible inhibitor for the RBBP6 p53BD, and future studies optimising this lead compound (in the context of blocking the formation of RBBP6-p53 complexes) may prove useful in cancer therapy.

5.5 RBBP6 probably binds independently to p53 on a different site compared to the transactivation domain targeted by MDM2

Both RBBP6 and MDM2 are well-known cell cycle regulators as they bind and promote proteasomal degradation of p53. In the present study, sequence alignment data of the p53BDs revealed very poor similarity and molecular docking studies further demonstrated major differences of the p53-binding pockets together with their retrieved lead compounds (Figure 4.16&4.18). This suggests that RBBP6 and MDM2 bind p53 at different sites.

It is important to note that these findings do not negate those of other studies that RBBP6 (also known as PACT) may act as a scaffold promoting the assembly of p53 and MDM2 (also known as HDM2) proteins (Li *et al.*, 2007). These results only imply that the binding is different and independent, but RBBP6 can still bind both proteins and facilitate their interaction independent of MDM2. Altogether, this finding suggests that the formation of MDM2-p53 complexes relies on the availability of RBBP6, whilst RBBP6-p53 complexes does not need MDM2. Interestingly, this finding enables for an increased specificity upon optimisation of drugs targeting the RBBP6 p53BD.

5.6 Ezetimibe (Zetia) accurately mimics p53 upon binding the MDM2 hydrophobic pocket

The tumour suppressor protein (p53) is a major cell cycle regulator that tightly binds the MDM2 hydrophobic cleft. The pocket-like nature of the MDM2 p53-binding domain targets the N-terminal transactivation domain of p53 allowing three critical residues (PHE19, TRP23, and LEU26) to be deeply buried (Patil *et al.*, 2014). In the present study, molecular docking studies show that *ezetimibe* binds the MDM2 hydrophobic pocket in a manner that accurately resemble p53 (Figure 4.18 and 4.19). This binding mode guarantees a complete inhibition of p53, and thus blocking the formation of MDM2-p53 complexes.

Current studies, show that nutlins (categorised within the piperazine class) are the major p53 inhibitors binding the MDM2 hydrophobic cleft. In this study, the screened database (Zdd) does not include nutlins, however there is a collection of piperazine compounds which did not exhibit any binding to the MDM2 p53-binding domain. Therefore, these findings do not negate those of previous studies, but rather reports another class of compounds (Organoheterocyclic) more likely to target the MDM2 hydrophobic pocket and thus reactive intracellular levels of p53. Therefore, this finding suggests that *ezetimibe* might be a very good lead compound possessing have significant implications in cancer therapy.

CHAPTER SIX: CONCLUSIONS

- Model-A represents the best possible RBBP6 p53-binding domain with minimal deviation from native state.
- The RBBP6 p53-binding domain constitutes a mixture of helices and loop structures, however its global topology still needs further refinements.
- RBBP6 interact with p53 in a different site compared to the transactivation domain targeted by MDM2, thus confirming that it acts as a scaffold promoting p53 ubiquination and its subsequent degradation.
- RBBP6 is draggable and might be a good therapeutic target in cancer.

CHAPTER SEVEN: FUTURE PROSPECTS

Modelling improvement

In future studies, the global topology of the p53-binding domain would be improved by incorporating a detailed statistical assessment on structure prediction. This should be done in a setting that would take most factors into account; such as molecular dynamics of neighbouring domains and the possibility of homo or hetero-dimerisation. Furthermore, the fundamental principle underlying conventional methods in protein modelling relies on "homology'' and that brings major limitations when modelling proteins with no homologous protein template from PDB (Protein databank). This limitation has significantly delayed the drug discovery process as multiple critical protein structures remains unknown.

Therefore, I-TASSER, 3D-Jigsaw and other softwares were developed to work based on fragment assembly approach rather than searching for a specific protein template. These techniques have shown great advancement and significantly improved drug discovery processes. However, for poor aligning proteins the modelling is still severely challenged. Hence the next adventure in modelling the RBBP6 p53BD would be to use QUARK, which is a template-free modelling software that uses *ab initio* algorithms and Replica Exchange Monte Carlo Simulations. Finally, after these *in sillico* attempts HDX-MS and X-ray Crystallography would be used to predict the domain structure in an *in vitro* setup.

Lead Optimisation

The optimisation of predicted lead compounds would also be performed using Glide XP (Extra Precision), Prime MM-GBSA and FEP+ Technology to generate preclinical drug candidates. For example; the optimisation of *catechins* with an aim of eliminating binding to anticancer drugs whilst maintaining their binding efficacy with RBBP6, might shed new light of their use in cancer therapy. However, *ezetimibe* has been shown to possess vulnerable sites on which glucuronidation takes place

(appendix 6). These sites would be of greater significance during optimisation to prevent metabolic fate and degradation of *ezetimibe*.

Toxicity prediction and final testing

This would then be followed up by *in silico* toxicity prediction of all the most promising lead compounds using Schrödinger-Maestro v10.7 and discovery studio. Druggable properties including pharmacokinetics would then be tested to all optimised compounds (pre-clinical drug candidates), and several *in vitro* assays (toxicity, specificity assays etc) would be done to validate these *in silico* predicted drug-like properties. Modern technologies such as Xcelligence and flow cytometry would be performed to assess the impact of the optimised drug candidates on cell-viability using various normal and cancer cell-lines (MCF-7, MRC-5, HEK293 etc). Finally, these drugs would be further tested in mince (*in vivo studies*) with varying concentrations. The data retrieved from all these tests would then be used as basis for designing new cancer therapies.

CHAPTER EIGHT: REFERENCES

Alkhatib, G., Combadiere, C., Broder, C.C., Feng, Y., Kennedy, P.E., Murphy, P.M., 1996. CC CKR5: a RANTES, MIP-1alpha, MIP-1beta receptor as a fusion cofactor for macrophage-tropic HIV-1. Science. 272:1955-1958.

Brenke, R., Kozakov, D., Chuang, G.-Y., Beglov, D., Hall, D., Landon, M. R., Vajda, S. 2009. Fragment-based identification of druggable "hot spots" of proteins using Fourier domain correlation techniques. Bioinformatics, 25(5), 621–627. doi.org/10.1093/ bioinformatics/btp036

Brylinski, M., Lingam, D., 2012. *e*Thread: A Highly Optimized Machine Learning-Based Approach to Meta-Threading and the Modelling of Protein Tertiary Structures. PLoS ONE, 7(11), e50200. doi.org/10.1371/journal.pone.0050200

Burgess, A., Chia K.M., Haupt, S., Thomas, D., Haupt, Y., Lim, E.,2016. Clinical Overview of MDM2/X-Targeted Therapies. Front. Oncol.6:7. doi: 10.3389/fonc.2016.00007

Catapano, A.L., 2009. Perspectives on low-density lipoprotein cholesterol goal achievement. Curr Med Res Opin. 25(2), 431-447. doi.org/10.1185/03007990802631438

Chacko, S. M., Thambi, P. T., Kuttan, R., Nishigaki, I., 2010. Beneficial effects of green tea: A literature review. Chin Med. 5, 13. doi.org/10.1186/1749-8546-5-13

Cheng, Y., Patel, N.H., Ting, C., 2008. Functional Studies of the Hybrid Male Sterility Gene, Odysseus (OdsH), in Drosophila melanogaster. A. Dros. Res. Conf. 49: 707B.

Cheng-Mayer, C., Seto, D., Tateno, M., Levy, J.A., 1988. Biologic features of HIV-1 that correlate with virulence in the host. Science. 240:80–82

Chibi, M., Meyer, M., Skepu, A., G. Rees, D.J., Moolman-Smook, J.C., Pugh, D.J.R., 2008. RBBP6 Interacts with Multifunctional Protein YB-1 through Its RING Finger Domain, Leading to Ubiquitination and Proteosomal Degradation of YB-1. J. Mol. Biol. 384, 908–916. doi: 10.1016/j.jmb.2008.09.060

Danilova, N., Kumagai, A., Lin J., 2010.p53 Upregulation Is a Frequent Response to Deficiency of Cell-Essential Genes. PLoS ONE. 5(12): e15938.

De Clercq, E., 2009. The AMD3100 story: the path to the discovery of a stem cell mobilizer (Mozobil). Biochem Pharmacol. 77: 1655-1664.

Eliseeva, I.A., Kim, E.R., Guryanov, S.G., Ovchinnikov, L.P., Lyabin, D.N., 2011. Y-box-binding protein 1 (YB-1) and its functions. Biochem. Mosc. 76, 1402–1433. doi:10.1134/S0006297911130049

Engen, J.R., 2009. Analysis of Protein Conformation and Dynamics by Hydrogen Deuterium Exchange MS. Analytical chemistry.81, 7870-7875.

Fang, X., Lu, Q., Emoto, K., Adler, P.N., 2010. The Drosophila Fry protein interacts with Trc and is highly mobile in vivo. BMC Developmental Biology.10,40.

Feng, Y., Broder, C.C., Kennedy, P.E, Berger, E.A., 1996. HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. Sci. 272:872–877.

Ghose, A.K., Viswanadhan, V.N., Wendoloski, J.J., 1999. A knowledge-based approach in designing combinatorial and medicinal chemistry libraries for drug discovery. 1. A qualitative and quantitative characterisation of drug databases. Journal of combinatorial chemistry. 1, 55-68.

Goh, W.L., Lee, M.Y., Joseph, T.L., Quah, S.T., Brown, C.J., Verma, C., Brenner, S., Ghadessy, F.J., Teo, Y.N., 2014. Molecular rotors as conditionally fluorescent labels for rapid detection of biomolecular interactions. J. Am. Chem. Soc. 136, 6159–6162.

Hajduk, P.J., Huth, J.R., Fesik, S.W., 2005. Druggability Indices for Protein Targets Derived from NMR-Based Screening Data. J. Med. Chem. 48 (7), 2518-2525 doi:10.1021/jm049131r

Hebling, C.M., Morgan, C.R., Stafford, D.W., Jorgenson, J.W., Rand, K.D., Engen, J.R., 2010. Conformational analysis of membrane proteins in phospholipid bilayer nanodiscs by hydrogen exchange mass spectrometry. Analytical chemistry. 82, 5415-5419.

Irwin, J.J., Shoichet, B.K., 2005. ZINC-a free database of commercially available compounds for virtual screening. J. Chem. Inf. Model. 45, 177–182.

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., others, 2012. Solution structure of RING finger-like domain of retinoblastoma-binding protein-6 (RBBP6) suggests it functions as a U-box. J. Biol. Chem. 287, 7146–7158.

Khan, N., Mukhtar, H., 2007. Tea polyphenols for health promotion. Life Sci. 81(7), 519-533. doi.org/10.1016/j.lfs.2007.06.011

Kosoglou, T., Shatkevich, P., Johnson-Levonous, A.O., Paolini, J.F., Bergman, A.J., Alton, K.B., 2005. Ezetimibe: a review of its metabolism, pharmacokinetics and drug interactions. Clin. Pharmacokenetics. 44(5): 467-494.

Kussie, P.H., Gorina, S., Marechal, V., Elenbaas, B., Moreau, J., Levine, A.J., Pavletich, N.P., 1996. Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. Science 274, 948–953.

Lee, H.G., Kim, Y.C., Dunning, J.S., Han, K.A, 2008. Recurring ethanol exposure induces disinhibited courtship in Drosophila. PLoS ONE. 3(1): e1391.

Lecumberri, E., Dupertuis, Y.M., Miralbell, R., Pichard, C., 2013. Green tea polyphenol epigallocatechin-3-gallate (EGCG) as adjuvant in cancer therapy. Clinical Nutr.32, 894-903.

Li, L., Deng, B., Xing, G., Teng, Y., Tian, C., Cheng, X., Yin, X., Yang, J., Gao, X., Zhu, Y., others, 2007. PACT is a negative regulator of p53 and essential for cell growth and embryonic development. Proc. Natl. Acad. Sci. 104, 7951–7956.

Lipinski, C.A., Lombardo, F., Dominy, B.W., Feeney, P.J., 1997. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Advanced drug discovery reviews. 23, 3-25.

Mather, A., Rakgotho, M., Ntwasa, M., 2005. SNAMA, a novel protein with a DWNN domain and a RING finger-like motif: A possible role in apoptosis. Biochim. Biophys. Acta BBA - Gene Struct. Expr. 1727, 169–176.

Mandal, S., Moudgil, M. ’nal, Mandal, S.K., 2009. Rational drug design. Eur. J. Pharmacol. 625, 90–100. doi: 10.1016/j.ejphar.2009.06.065

Mattos, C., Ringe, D., 1996. Locating and characterizing binding sites on proteins. Nat. Biotechnol. 14:595–599.

Mbita, Z., 2012. Molecular analysis of the domain with no name (DWNN)/RBBP6 in human cancers. University of the Witwatersrand.

Melo, F., Feytmans, E., 1998. "Assessing protein structures with a non-local atomic interaction energy." J Mol Biol. 277(5): 1141-1152.

Moela, P., Motadi, L. R., 2016. RBBP6: a potential biomarker of apoptosis induction in human cervical cancer cell lines. OncoTargets and Ther. 9, 4721–4735. doi.org/10.2147/OTT.S100964

Moll, U.M., Petrenko, O., 2003. The MDM2-p53 interaction. Mol. Cancer Res. 1, 1001–1008.

Ngan, C.-H., Hall, D. R., Zerbe, B., Grove, L. E., Kozakov, D., Vajda, S., 2012. FTSite: high accuracy detection of ligand binding sites on unbound protein structures. Bioinformatics, 28(2), 286–287. doi.org/10.1093/bioinformatics/btr651

Ntwasa, M., 2016. Retinoblastoma Binding Protein 6, Another p53 Monitor. Trends in Cancer. 2, 635-637.

Ntwasa, M., 2008. The retinoblastoma binding protein 6 is a potential target for therapeutic drugs. Biotechnol Mol Biol Rev. 3, 24–31.

Patil, S.P., Ballester, P.J., Kerezsi, C.R., 2014. Prospective virtual screening for novel p53–MDM2 inhibitors using ultrafast shape recognition. J. Comput. Aided Mol. Des. 28, 89–97.

Petitjean, A., Mathe, E., Kato, S., Ishioka, C., Tavtigian, S. V., Hainaut, P. and Olivier, M., 2007. Impact of mutant p53 functional properties on TP53 mutation patterns and tumour phenotype: lessons from recent developments in the IARC TP53 database. Hum. Mutat., 28: 622–629. doi:10.1002/humu.20495

Pugh, D.J., Eiso, A.B., Faro, A., Lutya, P.T., Hoffmann, E., Rees, D.J.G., 2006. DWNN, a novel ubiquitin-like domain, implicates RBBP6 in mRNA processing and ubiquitin-like pathways. BMC Struct. Biol. 6, 1.

Reygaert, W.C., 2017. An Update on the Health Benefits of Green Tea. MDPI Bev. 3, 6.

Roy, A., Kucukural, A., & Zhang, Y., 2010. I-TASSER: a unified platform for automated protein structure and function prediction. Nat. Prot. 5(4),725–738. doi.org/10.1038/nprot.2010.5

Scott, R.E., Giannakouros, T., Gao, S., Peidis, P., 2003. Functional potential of P2P-R: A role in the cell cycle and cell differentiation related to its interactions with proteins that bind to matrix associated regions of DNA? J. Cell. Biochem. 90, 6–12. doi:10.1002/jcb.10618

Shangary, S., Wang, S., 2009. Small-Molecule Inhibitors of the MDM2-p53 Protein-Protein Interaction to Reactivate p53 Function: A Novel Approach for Cancer Therapy. Annual Review of Pharmacology and Toxicology, 49, 223–241. http://doi.org/10.1146/annurev. pharmtox. 48.113006.094723

Shi, Y., Di Giammartino, D.C., Taylor, D., Sarkeshik, A., Rice, W.J., Yates, J.R., Frank, J., Manley, J.L., 2009. Molecular Architecture of the Human Pre-mRNA 3′ Processing Complex. Mol. Cell 33, 365–376. doi: 10.1016/j.molcel.2008.12.028

Tsai, C. S., 2002. An Introduction to Computational Biochemistry, J. Wiley & Sons, Inc., Hoboken, NJ, USA. doi: 10.1002/0471223840.ch1

Uy, G.L., Rettig, M.P., Cashen, A.F., 2008. Plerixafor, a CXCR4 antagonist for the mobilization of hematopoietic stem cells. Expert Opin Biol. Ther. 8, 1797-1804. doi.org/101517/14712598. 8.11.1797

Vo, L.T.A., Minet, M., Schmitter, J.-M., Lacroute, F., Wyers, F., 2001. Mpe1, a Zinc Knuckle Protein, Is an Essential Component of Yeast Cleavage and Polyadenylation Factor Required for the Cleavage and Polyadenylation of mRNA. Mol. Cell. Biol. 21, 8346–8356. doi:10.1128/MCB.21.24.8346-8356.2001

Wales, T. E. and Engen, J. R. (2006), Hydrogen exchange mass spectrometry for the analysis of protein dynamics. Mass Spectrometry. Rev., 25: 158–170. doi:10.1002/mas.20064

Wiederstein, M., Sippl, M. J., 2007. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Research, 35. W407–W410. doi.org/10.1093/nar/gkm290

Witte, M.M., Scott, R.E., 1997. The proliferation potential protein-related (P2P-R) gene with domains encoding heterogeneous nuclear ribonucleoprotein association and Rb1 binding shows repressed expression during terminal differentiation. Proc. Natl. Acad. Sci. 94, 1212–1217. doi:10.1073/pnas.94.4.1212

Wu, L., Gerard, N.P., Wyatt, R., Choe, H., Parolin, C., Ruffing, N., Borsetti, A., Cardoso, A.A., Desjardin, E., Newman, W., Gerard, C., Sodroski, J.G., 1996. CD4-induced interaction of primary HIV-1 gp120 glycoproteins with the chemokine receptor CCR-5. Nature. 6605,179–183.

Xu, D., Zhang, Y., 2012. *Ab initio* protein structure assembly using continuous structure fragments and optimised knowledge-based force field. Proteins. 80, 1715-1735.

Yang, J., Zhang, Y., 2015. I-TASSER server: new development for protein structure and function predictions. Nucleic Acids Research, 43, W174–W181. doi.org/10.1093/nar/gkv342

Yoshitake, Y., Nakatsura, T., Monji, M., Senju, S., Matsuyoshi, H., Tsukamoto, H., Hosaka, S., Komori, H., Fukuma, D., Ikuta, Y., 2004. Proliferation potential-related protein, an ideal oesophageal cancer antigen for immunotherapy, identified using complementary DNA microarray analysis. Clin. Cancer Res. 10, 6437–6448.

Yu, J.Y., Kanai, M.I., Demir, E., Jefferis, G.S., Dickson, B.J., 2010. Cellular Organization of the Neural Circuit that Drives Drosophila Courtship Behaviour. Curr. Biol. 20(18): 1602-1614.

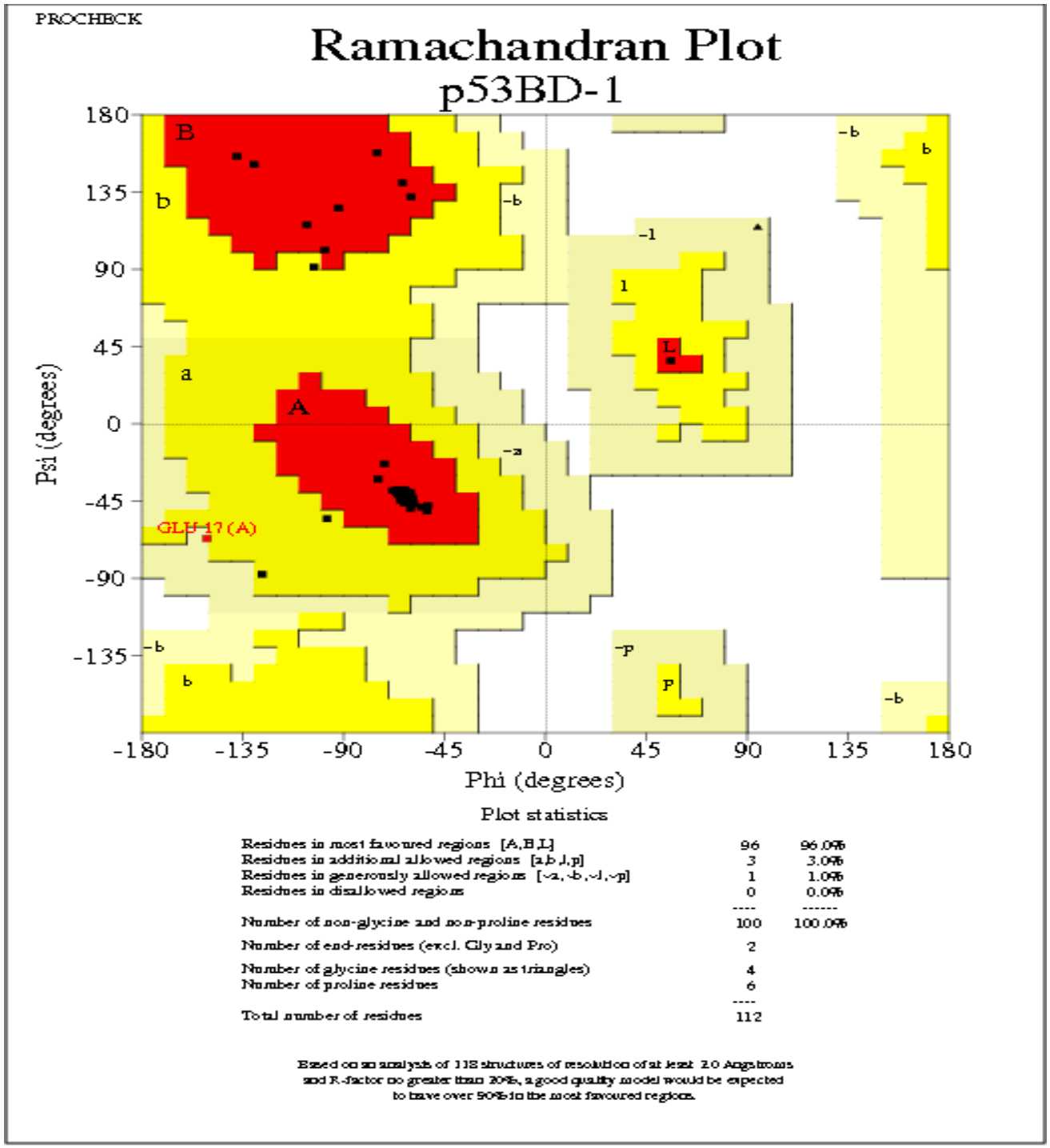
Zhang, Y., 2008. I-TASSER server for protein 3D structure prediction. BMC Bioinformatics, 9, 40. doi.org/10.1186/1471-2105-9-40

Zhou, H., Zhou, Y., 2002. Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction. Protein Sci. 11(11), 2714–2726.

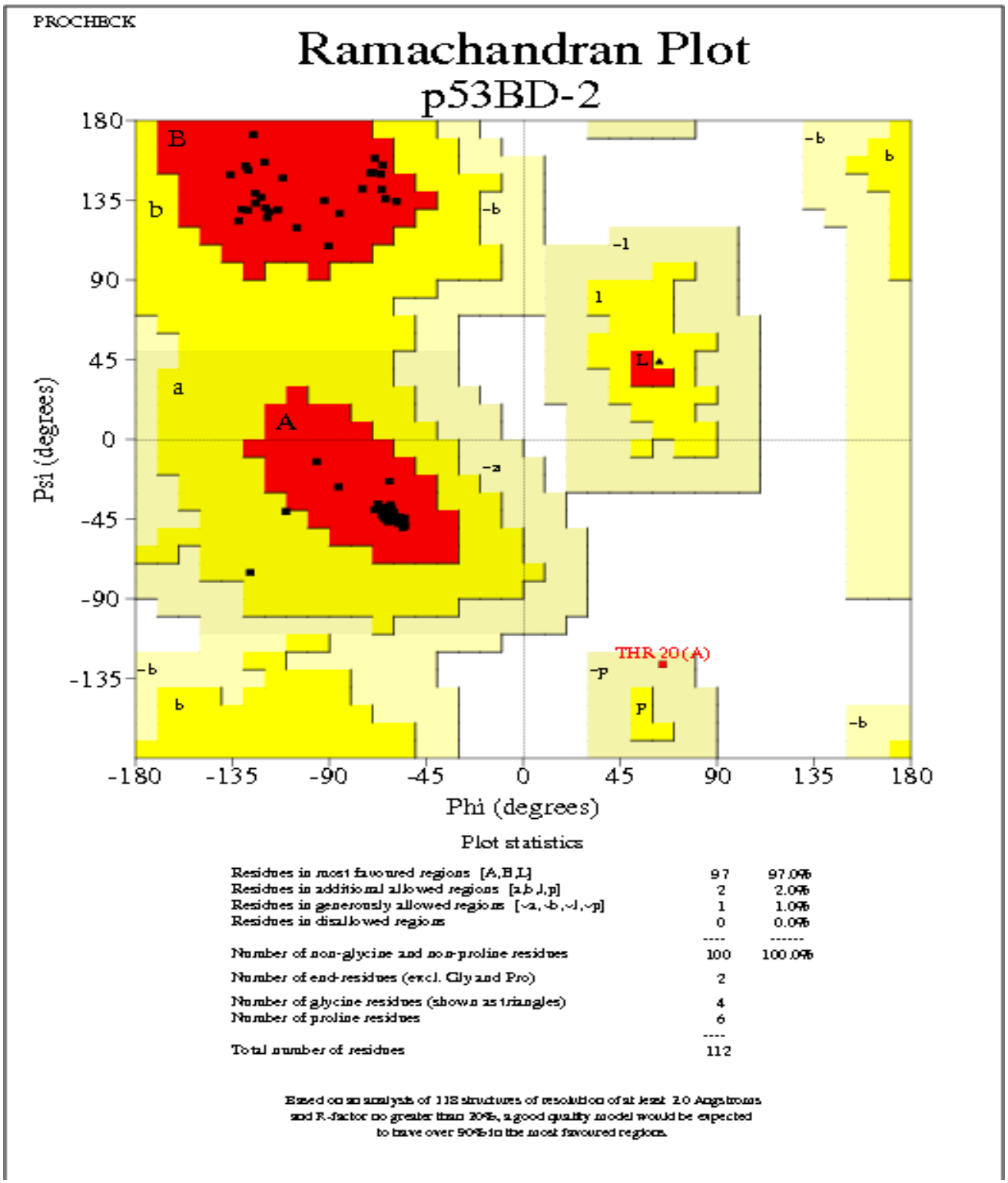
8.2 Appendices

Ramachandran plots of the first three models generated by *e*Thread–Modeller: Stereochemistry Assessment

Appendix 1: Model-D plot showing sterically allowed and disallowed residues based on Psi and Phi angles



Appendix 2: Model-E plot showing sterically allowed and disallowed residues based on Psi and Phi angles

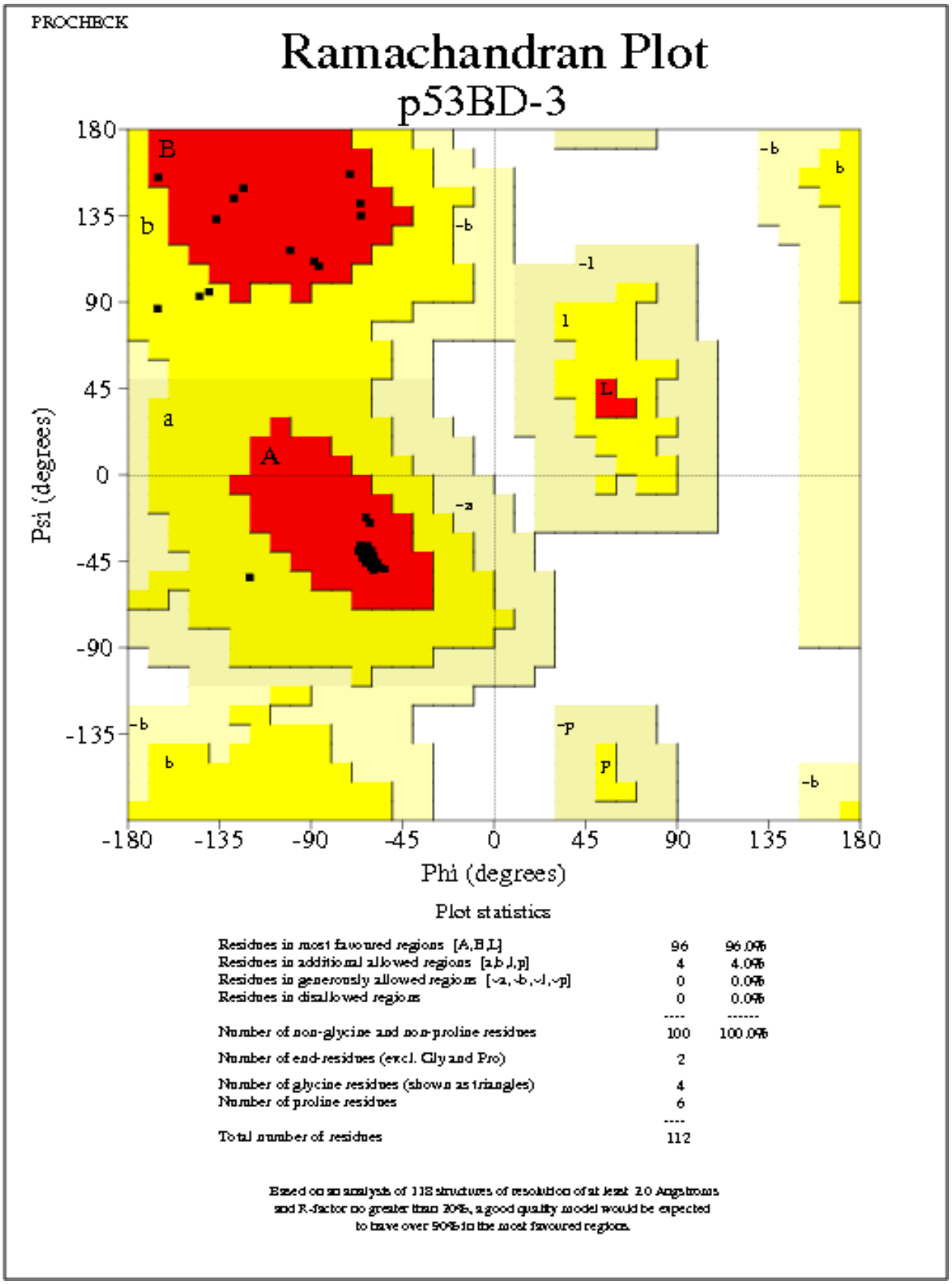


Plot statistics

Residues in most favoured regions [A,B,L]	97	97.0%
Residues in additional allowed regions [a,b,l,p]	2	2.0%
Residues in generously allowed regions [~a,~b,~l,~p]	1	1.0%
Residues in disallowed regions	0	0.0%
Number of non-glycine and non-proline residues	100	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	4	
Number of proline residues	6	
Total number of residues	112	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

Appendix 3: Model-F plot showing sterically allowed and disallowed residues based on Psi and Phi angles



Plot statistics

Residues in most favoured regions [A,B,L]	96	96.0%
Residues in additional allowed regions [a,b,l,p]	4	4.0%
Residues in generously allowed regions [~a,~b,~l,~p]	0	0.0%
Residues in disallowed regions	0	0.0%
Number of non-glycine and non-proline residues	100	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	4	
Number of proline residues	6	
Total number of residues	112	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

Appendix 4: *e*Thread-Modeller: Main chain parameters of Model-D, E and F respectively

D

Plot statistics

Stereochemical parameter	No. of data pts	Parameter value	Comparison values Typical value	Band width	No. of band widths from mean	
a. %-tage residues in A, B, L	100	96.0	56.4	10.0	4.0	BETTER
b. Omega angle st dev	111	8.4	6.0	3.0	0.8	Inside
c. Bad contacts / 100 residues	2	1.8	29.8	10.0	-2.8	BETTER
d. Zeta angle st dev	108	1.0	3.1	1.6	-1.3	BETTER
e. H-bond energy st dev	88	0.4	1.1	0.2	-3.8	BETTER
f. Overall G-factor	112	-0.5	-1.0	0.3	1.5	BETTER

E

Plot statistics

Stereochemical parameter	No. of data pts	Parameter value	Comparison values Typical value	Band width	No. of band widths from mean	
a. %-tage residues in A, B, L	100	97.0	50.6	10.0	4.6	BETTER
b. Omega angle st dev	111	9.6	6.0	3.0	1.2	WORSE
c. Bad contacts / 100 residues	3	2.7	35.5	10.0	-3.3	BETTER
d. Zeta angle st dev	108	1.1	3.1	1.6	-1.3	BETTER
e. H-bond energy st dev	63	0.6	1.2	0.2	-2.9	BETTER
f. Overall G-factor	112	-0.6	-1.0	0.3	1.5	BETTER

F

Plot statistics

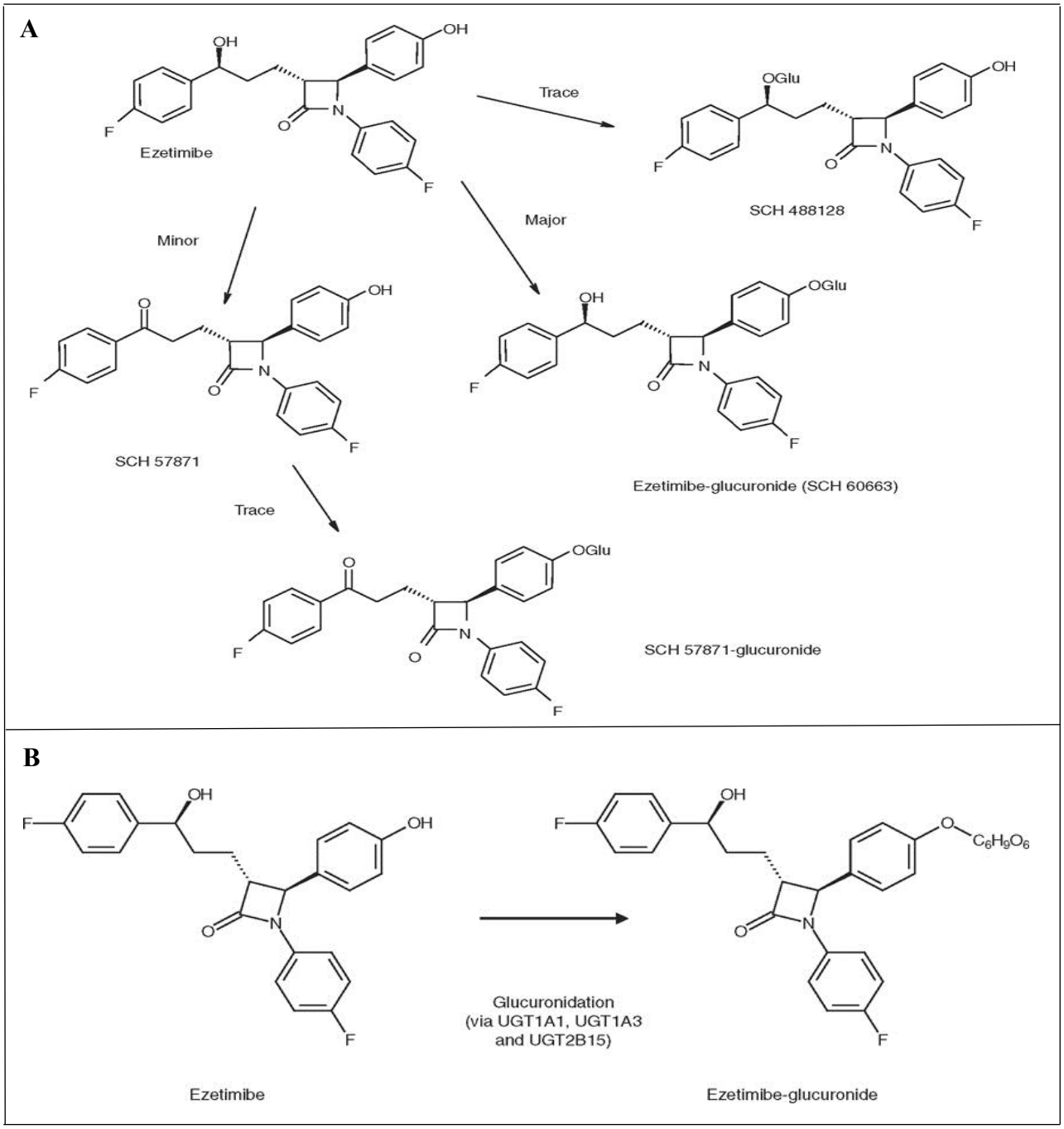
Stereochemical parameter	No. of data pts	Parameter value	Comparison values Typical value	Band width	No. of band widths from mean	
a. %-tage residues in A, B, L	100	96.0	47.6	10.0	4.8	BETTER
b. Omega angle st dev	111	12.4	6.0	3.0	2.1	WORSE
c. Bad contacts / 100 residues	0	0.0	38.6	10.0	-3.9	BETTER
d. Zeta angle st dev	108	1.1	3.1	1.6	-1.3	BETTER
e. H-bond energy st dev	88	0.3	1.2	0.2	-4.4	BETTER
f. Overall G-factor	112	-0.3	-1.1	0.3	2.5	BETTER

Appendix 5: PDB retrieved templates by *e*Thread-Modeller during the p53BD domain modelling

Template	Identity	Confidence	Coverage	Length
4fmuA	0.315	T/0.00	0.330	37
1e5wA	0.286	T/0.23	1.000	112
5fm9A	0.243	T/0.78	1.000	112
2mbfA	0.192	L/0.49	0.562	63
4yycA	0.378	L/0.69	0.625	70
1lt7A	0.342	L/0.52	0.339	38
4kbgB	0.339	L/0.47	0.339	38
1vtnC	0.162	L/0.53	0.562	63
3soaA	0.423	L/0.57	0.161	18
2k7nA	0.234	L/0.59	0.116	13
4p4kC	0.270	L/0.48	0.330	37
4om8B	0.355	L/0.51	1.000	112
4onhB	0.315	L/0.47	1.000	112

Identity: Is the global sequence identity to the target.
Confidence: Estimates the chances that the template is structurally similar to the target (***H*** - high, ***L*** - low, ***T*** - top).
Coverage: Gives the coverage of the target sequence by the template.
Length: Is the alignment length

Appendix 6: Biotransformation pathway of *ezetimibe*



Appendix 6: Biotransformation pathway of *ezetimibe*: [A] The metabolic pathway for *ezetimibe* in humans and pre-clinical species. **[B]** shows the major metabolite (Ezetimibe-glucuronide) formation together with the enzymes responsible for the reaction (Adapted from: Kosoglou *et al.*, 2005).