COMPUTATIONAL STUDIES ON THE IDENTIFICATION AND ANALYSES OF P53 CANCER ASSOCIATED MUTATIONS

By

Nosipho Magnificat Cele

Student number: 21320624

Submitted in the fulfillment of the requirement for the Degree of Master of Applied Science in Chemistry in the Faculty of Applied Sciences at the Durban University of Technology

June 2017

DECLARATION

I, **Nosipho Magnificat Cele**, declare that the thesis submitted for the Master of Applied Science in Chemistry degree at the Durban University of Technology has not been submitted in any University or any other Institution. I, therefore, declare that this is solely my own work.

Student name: Nosipho Magnificat Cele

[redacted] 24/8/17

Signature of a student Date

Supervisor's name: Prof K. Bisetty

[redacted] 24/8/17

Signature of Supervisor Date

Co-Supervisor's name: Assoc. Prof. Md Imtaiyaz Hassan, Jamia Millia Islamia (India).

[redacted] 24/8/17

Signature of Co-Supervisor Date

ACKNOWLEDGEMENTS

The Glory belongs to **God** All Mighty, special gratitude goes to **Jesus Christ** of Nazareth my Lord and Saviour for giving me the power to start and strength to continue especially in difficult times until the end of this journey.

My outmost gratitude goes to my promoter, Professor K. Bisetty for allowing me to undertake this research, his helpful advice, guidance, and direction in the past two years have been invaluable. It is my privilege to thank my co-supervisor Professor Md. Imtaiyaz Hassan from Jamia Millia Islamia, India for his guidance and constructive suggestions throughout my studies. I would like to sincerely thank Dr. Kanchi for all his guidance and support especially during group meetings. To Myalo your patience and supervision are also appreciated. To my friends Athika, Bayu and Kwanele you were there through the discoveries, discussions and friendly conversations it was great fun to work with you.

Finally, my special thanks to family, my two mothers (E.D and Sane), my sisters Sibusisiwe you went through all the ups and downs with me, Fanelesibonge, Nontobeko and my late sister Nkosiyapha this is for you, my nephews (Kwenzakwenkosi, Bongumusa, and Owami) your prayers and support kept me through during this journey, I love you guys.

ABSTRACT

P53 is a tumour suppressor protein that is dysfunctional in most human cancer cells. Mutations in the p53 genes result in the expression of mutant proteins which accumulate to high levels in tumour cells. Several studies have shown that majority of the mutations are concentrated in the DNA-binding domain where they destabilize its conformation and eliminate the sequence-specific DNA-binding to abolish p53 transcription activities. Accordingly, this study involved an investigation of the effects of mutations associated with cancer, based on the framework of sequences and structures of p53 DNA-binding domains, analysed by SIFT, Pmut, I-mutant, MuStab, CUPSAT, EASY-MM and SDM servers. These analyses suggest that 156 mutations may be associated with cancer, and may result in protein malfunction, including the experimentally validated mutations. Thereafter, 54 mutations were further classified as disease-causing mutations and probably have a significant impact on the stability of the structure. The detailed stability analyses revealed that Val143Asp, Ala159Pro, Val197Pro, Tyr234Pro, Cys238Pro, Gly262Pro and Cys275Pro mutations caused the highest destabilization of the structure thus leading to malfunctioning of the protein. Additionally, the structural and functional consequences of the resulting highly destabilizing mutations were explored further using molecular docking and molecular dynamics simulations. Molecular docking results revealed that the p53 DNA-binding domain loses its stability and abrogates the specific DNA-binding as shown by a decrease in binding affinity characterized by the ZRANK scores. This result was confirmed by the residues Val143Asp, Ala159Pro, Val197Pro, Tyr234Pro and Cys238Pro p53-DNA mutant complexes inducing the loss of important hydrogen bonds, and introduced non-native hydrogen bonds between the two biomolecules. Furthermore, Molecular dynamics (MD) simulations of the experimental mutant forms showed that the structures of the p53 DNA-binding domains were more rigid comparing to the wild-type structure. The MD trajectories of Val134Ala, Arg213Gly

and Gly245Ser DNA-binding domain mutants clearly revealed a loss of the flexibility and stability by the structures. This might affect the structural conformation and interfere with the interaction to DNA. Understanding the effects of mutations associated with cancer at a molecular level will be helpful in designing new therapeutics for cancer diseases.

LIST OF CONTENTS

LIST OF FIGURES

ACRONYMS AND SYMBOLS

SA	South Africa
SV 40	simian virus 40
LFS	Li-Fraumeni Syndrome
Aa	Amino acid
DNA	Deoxyribonucleic acid
NES	Nuclear export signals
NLS	Nuclear localization signals
LFL	Li-Fraumeni-like
IARC	International agency for research on cancer
NMR	Nuclear magnetic resonance
DBD	DNA-binding domain
LSH	Loop-sheet-helix
MD	Molecular dynamics
SIFT	Sorting intolerant from tolerant
PSI-BLAST	Position-Specific Iterative Basic Local Alignment Search Tool
EASY-MM	Evolutionary amino acid and structural encodings with multiple models
SVM	Support vector machine
CUPSAT	Cologne university protein stability analysis tool
SDM	Site-directed mutator
FTT	Fast Fourier transform
RMSD	Root mean square deviation
PSC	Pairwise shape complementary
CAPRI	Critical assessment of predicted interactions
CHARMm	Chemistry at Harvard Macromolecular Mechanics
MM	Molecular mechanics

PME	Particle mesh Ewald
PBC	Period boundary conditions
NVT	Number of atoms, volume, and temperature
NPT	Number of atoms, pressure, and temperature
PDB	Protein data bank
Mol PDF	Molecular Probability density function
DOPE	Discrete optimized protein energy
DS	Discovery Studio
3D	3-Dimensional
1D	1-Dimensional
SASA	Solvent accessible surface area
SAS	Solvent accessible surface
GUI	Graphical user interface
CHPC	Centre for high-performance computing
NH-bond	Hydrogen bonding
Rg	Radius of gyration
GROMACS	GROningen MOlecular Simulation

RESEARCH OUTPUTS

Poster Presentation

10th CHPC Annual National Meeting held at International Convention Centre, East London, South Africa, 5 to 9 December **2016**. "A computational study of identification and analyses of p53 mutations".

Publication

Nosipho Cele, Krishna Bisetty and Md. Imtaiyaz Hassan

"Computational studies on identification and analyses of cancer associated mutations in p53 DNA-binding domain." *In preparation*.

CHAPTER 1

INTRODUCTION

This chapter provides a brief outline on the epidemiology of cancer along with the current impact on the socio-economics of cancer-related diseases in South Africa. An overview of the initiation and progression of cancer cells at a molecular level defined by molecular genes associated with cancer is included in this section. This chapter also provides a brief description of p53 tumour suppressor proteins and how missense mutations affects its stability, leading to the development of cancer. Computational chemistry methods employed in this study are briefly outlined along with a scheme of the methodology workflow. This is followed by the aims and objectives of this work.

1.1 Background and context of study

Cancer is a well-known disease which has been documented around the nineteenth century but a limited number of people able to survive this disease up to this day and age. Due to the dramatic increase in the population growth worldwide the risk of cancer in this era is great regardless of the improvement by the health sector. The increase in rates of people with cancer is expected as the population ages, since cancer is primarily known as a disease for older people. For this reason cancer research has gained popularity in the field of medicine to find the agents that will fight this human scare (Knowles and Selby 2005).

The increasing death toll is not only a health burden but an economic cost to the governments in third world countries (Busolo and Woodgate 2014) (Siegel *et al.* 2011). In 2008 an estimated

12.7 million new cases of cancer deaths were reported worldwide and it is expected that these numbers will double by 2030 (Center, Siegel and Jemal 2008). A large number of deaths pushes cancer to be the second cause of death to cardiovascular as the first (Singh *et al.* 2015). Incidences of the population suffering from cancer are increasing in the third world continents like Africa which is carrying a burden of fighting infectious diseases. However, one of the main challenges in Africa is the stigma associated with cancer which hinders the patients and families to speak about the disease in their communities.

South Africa (SA) is one of the countries that are facing a burden of chronic and infectious diseases as a result of HIV/AIDS and unhealthy lifestyle (Mayosi *et al.* 2009). Where prevention, diagnosis, and treatment come with constrained budgets due to a large population of poor people mostly affected by diseases (Mayosi *et al.* 2009). In SA 40 000 patients die of cancer with a substantial diversity of older population groups annually (Busolo and Woodgate 2014). The improved surveillance and diagnostic methods reveal that socio-economic factors such as poor living environments and limited access to basic health services are the contributors to the exponential increase of cancer deaths. Additionally, delays in screening and treatment, due to lack of proper education and unaffordability as the government is focusing on common health issues like HIV/AIDS further limits education about other diseases dominating the country. Furthermore, cancer in South Africa is largely induced by habits and unhealthy lifestyles adopted from western cultures (Sitas *et al.* 2008). However, prevention is possible for some cancers by early identification of the associated risks (Busolo and Woodgate 2014).

At a molecular level, cancer is known as a disease that is formed by uncontrolled proliferation of the damaged cells. The hallmarks of genetic instabilities stem from traumatic stresses inside the

cells as a result of DNA breakage leading to genetic alterations that form cancer cells. Genes that are mostly altered in cancer cells are classified as oncogenes or tumor suppressors. Adversely, these genes are susceptible to genetic mutations. Genetic mutations activate oncogenes or destroy the functions of tumour suppressor genes by promoting the formation of cancer cells. Genetic alterations caused by mutations might stem from gene amplifications or exchange of a single amino acid by changing residues that are important for the activity of the corresponding proteins. Examples of frequently mutated genes in cancer cells include Ras proteins, a family of small GTPases and p53 tumor suppressor proteins (Fernandez-Medarde and Santos 2011). In this spectrum substitution of a single amino acid residue that changes the sequence of the protein is called missense mutations and is the primary focus of this thesis. It is well established that missense mutations are liable for many diseases in humans as a result of protein instability, disruption of protein-protein interactions, protein-nucleic acid interactions, and characteristics of the active site (Tan *et al.* 2009).

The ability to predict whether a given missense mutation is associated with a disease would be of great importance in providing treatment that will reduce or eliminate the disease associated with the missense mutations. However, little progress has been made in drug design to provide competitive inhibitors that are capable of reversing the effects caused by a mutation associated with diseases. The present study focuses on p53 tumour suppressor protein which is a transcription factor that regulates the cell cycle upon DNA damage. But missense mutations is one of the frequent events in p53 that hinders its function and results in the formation and progression of cancer cells. A rational approach to predicting the effect of missense mutations in p53 protein using state of the art machine learning algorithms is developed in this work. The pathogenic effects of mutations were predicted at the sequence and 3D-structure levels to observe

and analyse the changes that affect the protein stability. Following this work, the impact of mutations in the binding specificity of p53 with DNA was confirmed with the molecular docking studies using ZDOCK algorithm. A robust explicit solvent molecular dynamics simulations technique concluded the structural instability of mutations in p53 protein at an atomic level as the workflow presented in **Figure 1.1**.

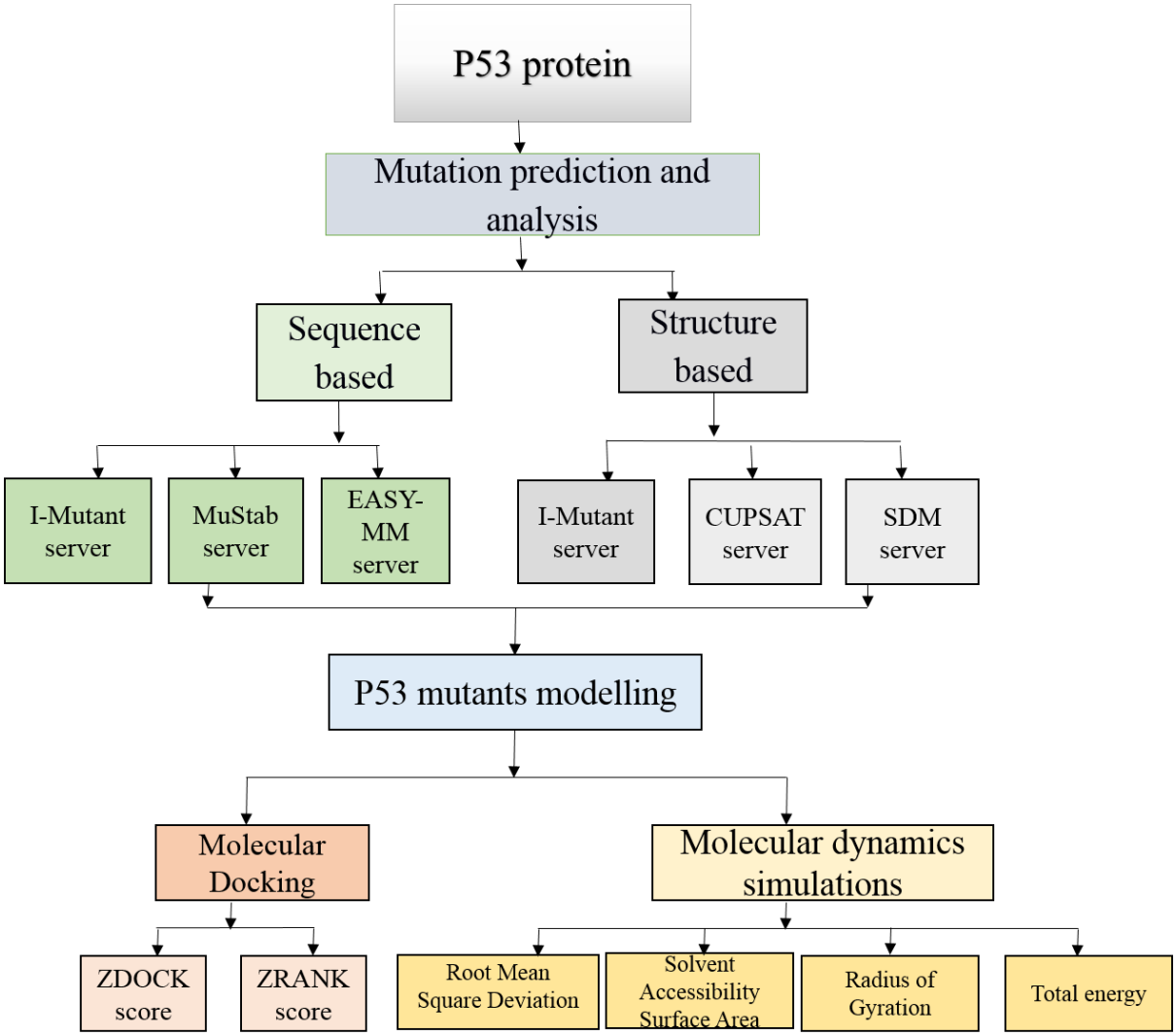


Figure 1.1 Methodology workflow scheme of the present research methodology

1.2 AIM and OBJECTIVES

Aim:

The broader goal of this study is to understand how cancer-causing mutations affect the structure and activity of p53 proteins using computational methods.

Objectives:

- Prediction of mutations that affect the stability of the p53 protein.
- Analysis of predicted mutations associated with cancer using neural network web servers.
- Modeling of selected p53 mutants.
- Determine p53-DNA binding energy through molecular docking.
- Determine specific p53-DNA contact residues that are affected by mutations in the DNA-binding domain.
- Perform molecular dynamics (MD) simulations of wild-type and selected structures of the computer generated mutants and compare the deviations from experimental mutants.

1.3 Thesis outline

After providing an overview of cancer as a disease and the role of proteins involved in cancer formation, further chapters in this thesis are as follows,:

Chapter 2

Deals with cancer as a result of DNA damage in cells and the power of genetic mutation. A review of p53 protein structure and its stability due to mutation is presented. P53 mutations in cancer are discussed. A further discussion involves mutations in p53 DNA-binding domain that changes the thermodynamic stability, p53-DNA interaction and structural arrangement of the domains and the protein tetramer is well articulated.

Chapter 3

Deals with the art of molecular modelling techniques, with the emphasis on web-server algorithms for predicting the consequences of mutation stability energies, molecular docking, and molecular dynamics simulations.

Chapter 4

Involves the protocols applied in the computational methodology of this study.

Chapter 5

Covers the results which focus on the identified destabilizing residues used to build mutants by Modeller. The developed docking approach was applied to quantify binding affinities of the p53-DNA complex. Molecular dynamics simulations focused on the DNA-binding domain inactivity brought by conformational changes of the DNA-binding domain experimental mutant structures contributing to the stability.

Chapter 6

Summarizes the achievements and the novelty of this work as well as the concluding remarks.

CHAPTER 2

LITERATURE REVIEW

This chapter provides an overview of cancer in the context of DNA damage. The remarkable breakthrough findings in the establishment of p53 genes, a leading tumor suppressor in proteins are documented here. The p53 protein structure is elucidated by prompting the transcription factor activities in the regulation of the cell cycle. A review on missense mutations in the DNA-binding domain in relation to stability and function of p53 is also presented. Moreover, a cohort of *in silico* techniques used to study the impact of mutations in the functions of p53 protein is briefly highlighted in this section.

2.1 Cancer at the molecular level

Cancer is a complex spectrum of chronic diseases caused by faults in genes that results into abnormal proliferation of damaged cells (Junttila and Evan 2009). At a molecular level, abnormal proliferation of damaged cells leads to the accumulation of masses of growth called tumours (Sudhakar 2009). Early detection of certain tumours is treatable with surgery, chemotherapy, radiation, immunotherapy and hormones. Well-developed tumours spread to far away sites causing the breakdown of the vital organs in the body. This invasive behaviour is the reason of many deaths caused by cancer (Bartkowiak, Riethdorf and Pantel 2012).

Cellular stress results into genetic mutations in cells. Genetic mutations in the gene network dictate cell growth as a necessity in the formation of cancers. These mutations in cancer cause genes to be either tumorigenic (oncogenes) or destroy functions that prevent diseases (tumour suppressor genes). Mutation of oncogenes confers cell growth by deactivating the wild-type

genes in the cells. In contrast to oncogenes, mutations of tumour suppressor genes lead to dysfunctional wild-type proteins. These alterations are caused by missense mutations in regions that are essential for protein activity (Bouafia *et al.* 2013). APC, Rb and p53 proteins are examples of expressed tumour suppressor genes with deactivated functions in cancer. Mutations of these proteins disrupt the cell cycle resulting in uncontrollable growth of cells that turn into cancer cells (Hejmad 2010).

The understanding of cancer from projects like Cancer Genome Atlas and the Human Genome Project are the main sites of identifying the mutations associated with cancer by utilizing modern experimental technologies and computational techniques. On the other hand, *in-vitro* experiments of each mutation are laboratory expensive and time-consuming, therefore, there is great dependence on computational methodologies to assess the effects of mutations associated with diseases (Ramani and Jacob 2013).

2.2 P53 is a tumour suppressor protein

P53 is a cellular protein that was discovered in 1979, where Linzer and Levine established that it complexes simian virus 40 (SV40) large T-antigen and was classified as a tumor antigen (T-antigen) (Levine, Finlay and Hinds 2004),(Bai and Zhu 2006) (Weisz, Oren and Rotter 2007). During this period, it was noted that abundant levels of this protein were expressed in many different types of transformed cells than normal cells and were responsible for driving the transformation of the cells. It was until 1984, where p53 received its first clone (Levine and Oren 2009) and was acknowledged as an oncogene that is activated by mutations with uncovered significance and mechanism of action (Levine and Oren 2009). Using the newly cloned p53 cDNA led a to series of studies in the functioning of p53 in cells and their over expression demonstrated total transformation and tumourigenesis in p53-null cells (Wolf and Rotter 1984).

Subsequent studies led to a series of important discoveries including the role of p53 proteins in DNA replication and the idea of p53 being a regulatory protein (Friedman *et al.* 1990). There were many studies to understand the functions and structure of p53 but hindered by the discovery that each clone had a different sequence from the other (Levine, Finlay and Hinds 2004). However, in 1988, molecular studies of the murine mouse cell lines revealed that the originally cloned cDNA had mutated versions of p53 that were isolated from a tumour cell (Finlay *et al.* 1988). Mutations affects conserved regions in p53 that are important for both conformation and biological activity of p53 proteins.

The turning point in p53 research occurred in 1989, where the wild-type was classified as a tumour suppressor, from the ability to suppress the transformation of E1A and Ras proteins. It was described in 1989 that tumour viruses have the ability to deactivate or lose the functions of p53 wild-type. Mutations in p53 were associated with Li-Fraumeni Syndrome (LFS) since the early 1990s. This predisposed cancer syndrome with neoplastic characteristics at multiple sites caused a variety of cancers due to the germline mutations in p53. Affected individuals will have an onset of tumours at an early age as they have inherited point mutations at highly conserved regions of the p53 genes (Hu *et al.* 2016). The results from the first p53 knockout mice firmly confirmed the function of p53 wild-type as a tumour suppressor. The mice with a mutated gene had a greater susceptibility to development of tumours by 6 months of age (Donehower *et al.* 1992). In contrast, the mice with the wild-type gene failed to develop any tumours until the age of 9 months. Therefore, p53 gene was targeted as the most common site for genetic alterations in human cancers (Wu and Yan 2003). P53 discovery timeline presented in **Figure 2.1**.

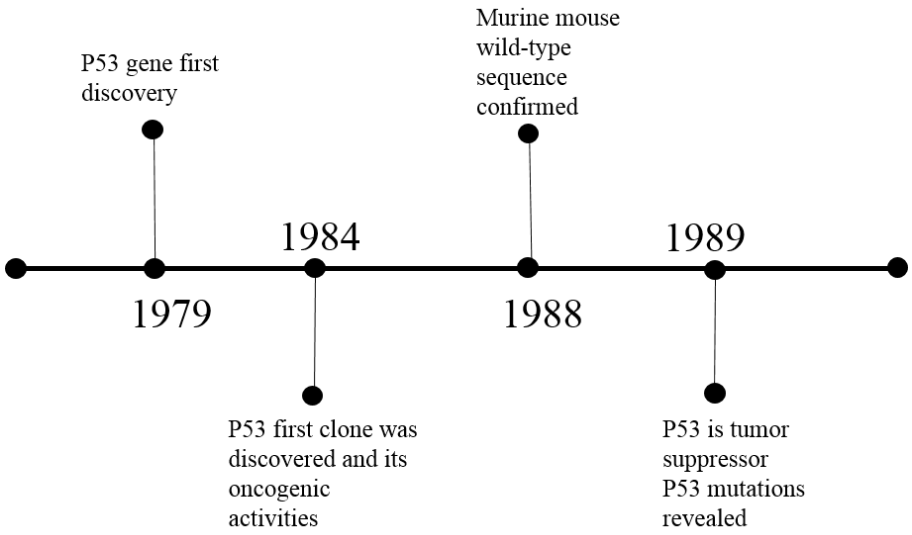


Figure 2.1 Summary of the landmark findings involved in establishing p53 as a tumour suppressor. Timeline adapted from Levine et al (Levine, Finlay and Hinds 2004)

The p53 transcriptional factor is a crucial tumour suppressor in humans and animals with the highest frequency of mutation in over 50% of all related types of cancers. This tumour suppressor protein (p53) has a vital role in the prevention of cancer and regulation of cell cycle (Vogelstein, Lane and Levine 2000). Therefore, the main function of a p53 protein is to protect the genome against genomic instabilities which are caused by cell damages and prevent genome mutations (Huang *et al.* 2011). Conversely, p53 protein acts as a transcription factor by regulating numerous genes (Mathe *et al.* 2006). When responding to stress in a cell, (for example DNA damage hypoxia, reactive oxygen species) p53 target genes will assist in cell cycle arrest, DNA repair, and apoptosis (cell death) (Kato *et al.* 2003) **Figure 2.2**.

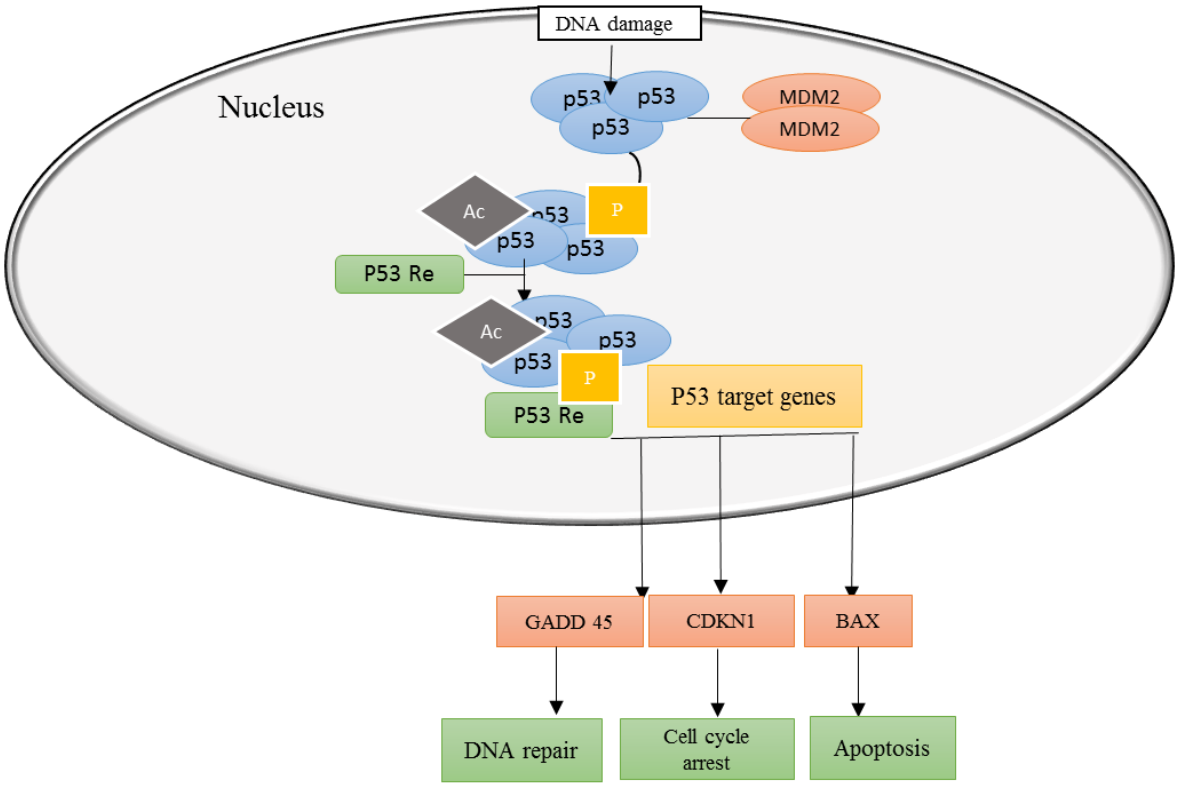


Figure 2.2 Schematic representation of p53 activation after cellular stress and transactivation of target genes that lead to tumour suppression by DNA repair, cell-cycle arrest, and apoptosis in the regulation of cell cycle

This mechanism regulates the cell cycle by preventing the propagation of genetic mismatch. The extent of cell damage is determined by p53 protein, and if the cell DNA is fixed p53 protein will thus stop the cell cycle and regulate the expression of other genes responsible for checkpoints to repair DNA before cell division (Jayaraman, Jamil and Raju 2012) or if the damage is severe and cannot be repaired p53 protein facilitate in the programming of apoptosis activity in the cell which is processed by protein-protein interactions. The p53 protein network consists of numerous proteins involved in sensing cellular stress, inducing signals that enable p53 to function. These series of signalling actions activate p53 to accumulate in the nucleus, increase its half-life, and change its conformation and its ability to bind to DNA. The activated p53 recruits its co-factors to inhibit certain genes in order to perform its transactivation. Over 120 p53-responsive genes in the human genome are regulated to p53 in the same conditions (Fu *et al.* 2012).In fact, different

stresses activate certain transcriptional programs for specific targets. Therefore, p53 tumour suppressor is a regulator of a complex protein network in the cell. The core central regulator and the activation of p53 occurring at the protein level will undergo post-translational modifications that will transform the stability and protein-protein interactions (Chehab *et al.* 1999). It is therefore not surprising that this functional complexity is mirrored in its structure as a tetramer (Boeckler *et al.* 2008).

2.3 The p53 Protein Structure

2.3.1 DNA sequence

P53 gene is present in chromosome 17 and located on the band of the short arm (17p13.1) **Figure 2.3**. P53 base pairs that measure 11 exons and 10 introns (Szymañska and Hainaut 2003). The coding sequence ranging from 2nd exon to the 11th exon (Ve´gran *et al.* 2013), **Figure 2.4**.

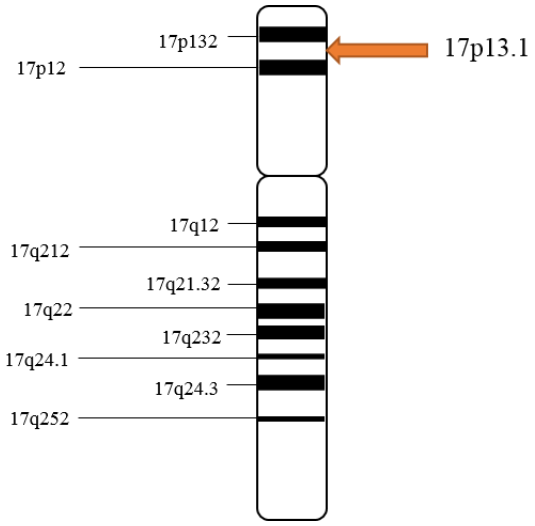


Figure 2.3 P53 gene is situated on the band of the 17p13.1 short arm of the chromosome 17

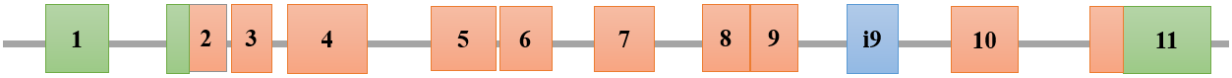


Figure 2.4 The p53 gene is made up of eleven exons. The green region symbolises untranslated region, the orange region symbolises the coding region and the blue region represents the internal exons within the introns

2.3.2 The p53 3D-structure

The human p53 protein is a 393 amino acid (aa) (Baroni *et al.* 2004) full-length protein composed of structural and functional divided domains. Each chain is comprised of four domains that form a homotetramer (Levine 1997) shown in **Figure 2.5**.

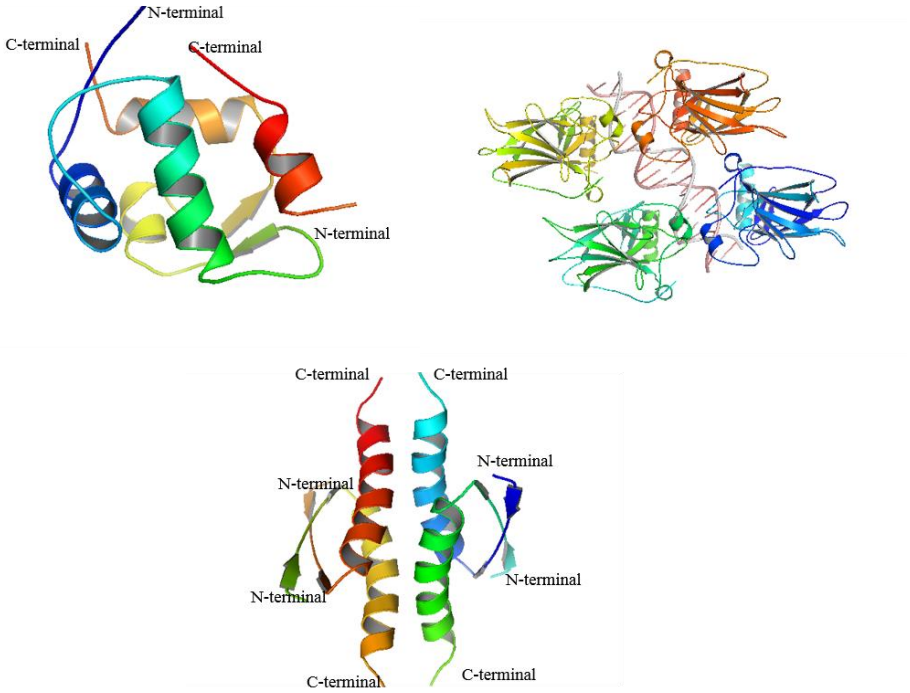


Figure 2.5 A cartoon representation of the p53 protein. **A** The acidic amino (N) terminal (PDB 1YCR) (Kussie et al. 1996) TAD1 in orange and bound MDM2 in green, yellow and blue; **B** DNA binding core domain tetramer (PDB 3KMD) (Chen, Dey and Chen 2010) with four monomers coloured in yellow, orange, green and blue; DNA is shown in stick; **C** The oligomerization region of the basic carboxy-terminal domain (PDB 1PES) (Lee et al. 1994) with four monomers coloured in red, yellow, green and cyan. Figures were modified with Pymol

The acidic amino-terminal (N-terminus) constitute of a transactivation domain (residues 1-57 aa) that is sub-divided into TAD1 (1-42 aa) and TAD2 (43-57 aa) (Ganguly and Chen 2015). The transactivation domain provides the binding sites of basal transcription in the regulation of

numerous downstream target genes (Fu *et al.* 2012). However, neighbouring sequences are needed for optimal functioning of this domain. Amino acid residues in TAD1 needed to regulate the transcription of p53 protein, including the binding of human MDM2 (a transcription initiator) protein into the amino-terminal, not only to negatively regulate the level of p53 in the cells but also prevent p53 from stimulating transcription (May and May 1999). Moreover, amino acids 13–23 in the p53 protein are similar in a wide range of species of this family (Levine 1997), (Xu and El-Gewely 2001). Amino acid residues in TAD2 interacts with MDM2, BAX and GADD45 proteins to induce apoptosis for tumour cell growth repression (El-Deiry 1998).

The proline-rich domain (58-101 aa) is harboured by the N-terminus and give rise to response into DNA damage through transcription-independent apoptosis for suppressing tumour cell growth (Toledo *et al.* 2007). This region contains five repeats of SH3 binding motifs PXXP. The multiple PXXP motifs is involved in the physical interaction with signal transduction pathways i.e. cytoplasmic signalling. This region is required for p53-mediated apoptosis and for degradation of HPV E6 targeted protein. However, the interesting issue is that p53 gene is polymorphic at codon 72 which yield either a proline or arginine at this position. The deletion of Pro with Arg terminates the PXXP motif. Subsequent studies revealed that p53 Arg is more liable than Pro to E6 mediated degradation of HPV-induced tumours in p53 Arg (Ozaki and Nakurawara 2011).

The central core domain is primarily composed of the sequence-specific DNA binding domain (102-292 aa) that is important for transcriptional activity. It includes four-five highly conserved regions which are crucial for recognizing and binding of p53 to DNA. These regions are highly conserved in p53 compared to other family members. The 3D crystal structure of p53 core domain bound to a double-stranded DNA was structurally elucidated in 1994 using X-ray

crystallography. This core domain structure formulated a basis for understanding the pathological effect of mutations associated with cancer (Cho *et al.* 1994). Under these findings, the basic structure of the core domain is a large β-sandwich of two antiparallel β-sheets that acts as a backbone of the DNA-binding interface (Fang, Simeonova and Toledo 2012). The interface anchors two loops [L2 (residues 164-194) and L3 (residues 237-250)] that are stabilized by a zinc ion and a loop–sheet–helix motif. The zinc ion is tetrahedral coordinated by Cys176, His179, Cys238 and Cys242 residues (Parks, Bolinger and Mann 1997). The Zn^{2+} ion is essential for specific sequence binding and the folding state of p53 core domain (Levine 1997).

The loop–sheet–helix motif together with coordinated zinc ion stabilizes the structure of the loops, such that the removal of the latter destabilizes the protein resulting into local structure rearrangement and cost the specific sequence DNA-binding. Six residues in the conserved regions of the loop–sheet–helix motif makes important hydrogen bond contacts with the DNA (Levine 1997). However, in this domain, more than 95% of mutations associated with cancer reside, and these mutations are differentiated into two classes (Lu, Tan and Luo 2007). Class I mutations involved Arg248Trp and Arg273His contact mutation residues that led to the loss of normal contact with DNA (Duffy *et al.* 2014). Class II mutations involved Arg175His, Gly245Ser, Arg249Ser and Arg282Trp that are classified as structural mutations as they destroy the local structure of p53 which acts as the scaffold in this domain (Szymañska and Hainaut 2003). Structural mutations alter p53 protein conformation as detected by PAb240 antibody. Analysis with monoclonal PAb240 antibody pointed out that transforming mutant p53 possess a different conformation from wild-type p53 (Rainwater *et al.* 1995).The carboxy-terminal (300-393 aa) domain consist of a flexible linker region (300-315 aa) and a tetramerization domain (316-393 aa) (Barakat *et al.* 2011). The flexible linker connects the central core domain with the C-terminal (Barakat *et al.* 2011). The oligomerization region (325-356 aa) of p53 forms a tetramer with each

monomer comprised of a dimer with two β-strands and two α-helices (Chong *et al.* 2005). The hydrophobic surfaces of two antiparallel α-helices and β-strands monomers are held together to produce a dimer that results into a tetramer which is essential for transactivation regulation of p53 *in vivo*, or DNA damage recognition (Joerger, Ang and Fersht 2006). A strongly basic carboxy-terminus domain contains a cluster of two nuclear export signals (NES) and three nuclear localization signals (NLS) (Stavridi *et al.* 2005). NLS was identified through sequence similarity and mutagenesis however, it mediates nuclear location of the protein. NLS 1 (316-324 aa) initiates the synthesis of a cytoplasmic protein of p53, however, NLS 2 (370-376 aa) and NLS 3 (380-386 aa) binds to specific receptors and permit passage of p53 through the nuclear pore complex (Liang and Clarke 1999). The highly conserved NES has been reported to be essential for nuclear retention of p53 (Stommel *et al.* 1999). As shown by the schematic representation of all the domains in **Figure 2.6**.

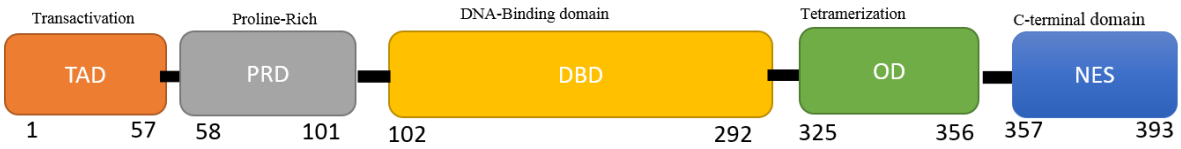


Figure 2.6 A schematic drawing of the p53 protein domains. The amino-terminus that is made up of TAD region, the proline-rich, DNA-binding domain, tetramerization forming the oligomerization domain and carboxy-terminus that is comprised of NES region

2.4 P53 Mutations in cancer

P53 tumour suppressor protein is critical in preserving the stability of the genome (Ramani and Jacob 2013), hence is given a nickname "guardian angel of the genome" (Kholmurodov 2013). P53 protein pathways inactivation by mutation is a universal feature in the development of human cancers, with mutations being the heart of progressing malignancies. The whole region of the p53 protein is vulnerable to mutations in different ways (Kim and Deppert 2007). Firstly, somatic mutations in p53 are associated with tumours in many cancers. The tumour-associated mutations

in many cancers arise from somatic cells through DNA damage. This type of mutations abolishes the function of a p53 protein that in turn negatively affects its pathway. Secondly, germline mutations in the p53 gene that are a consequence of an early onset cancer defining inherited Li–Fraumeni (LFS) and Li-Fraumeni-like (LFL) syndromes. Both these types of mutations respectively differ in their nature positions, and effects on the structure, allowing possible classification of their form in relation to the cancer type (Olivier, Hollstein and Hainaut 2010).

Li-Fraumeni syndrome (LFS) is a heterogeneous hereditary cancer condition that affects persons at early stages of life. This syndrome is characterised by autosomal dominant various cancers which include adrenocortical carcinomas, breast cancers, brain tumours ,bone and soft tissue sarcomas (Gualberto *et al.* 1998) (Giacomazzi *et al.* 2015). The distinct factor of LFS from the other cancer syndrome is not related to a single type of cancer but a broad spectrum of cancers. With, more than 70% of families with an inherited LFS syndrome have germline mutations that are only associated with p53 gene (Petitjean *et al.* 2007).

In contrast, LFL is a similar disease without all the traditional LFS characteristics but possess germline mutations of the p53 gene (Petitjean *et al.* 2007). In minor cases where the CHK2 mutation is reported, germline mutations in the p53 gene are not featured. International Agency for Research on Cancer (IARC) database which catalogues online information about p53 mutations has reported that vast majority of mutations occurring in p53 are missense mutations and least number are from deletions (Whibley, Pharoah and Hollstein 2009).

2.4.1 Somatic Mutations in p53

Mutations inactivation of p53 protein is found in more than 50% of human cancers (Parrales and Iwakuma 2015). The p53 protein is affected by missense mutations (single amino acid

substitutions) in one allele and results into a stable full-length protein that is functionally inactive. Missense mutations in p53 are associated with both, loss of p53 wild-type tumour suppressor activity and gained functions that enhance tumour progression (Silva *et al.* 2014). Missense mutations distributed along the p53 protein affects the molecule transcriptional activities to various degrees. Mutations do not possess the same deleterious effect on the protein (Oren and Rotter 2010) but exhibits a variety of different changes to the structure of the protein. The consequences of missense mutations are namely; (i) The change of the folding state of a protein (ii) The loss in the binding affinity of p53-DNA to its target genes and (iii) Dysfunction of protein-protein interactions to initiate cellular response and alteration of phosphorylation sites (Guimaraes and Hainaut 2002).

Sequencing studies revealed that overwhelming majority of 95% known tumour forming missense mutations are located in the (DNA-binding domain) DBD with over 46% of certain residues at codons Arg175, Gly245, Arg248, Arg249, Arg273 and Arg282 that correspond to the "hotspots" mutations in this core domain of the p53 protein (Brazdova *et al.* 2013). But, the six hotspot residues are frequently mutated and are detected in most human cancers (Joerger *et al.* 2005). The nature of the hotspot mutations provides insights on how p53 function is affected by mutations (Muller and Vousden 2013). A surprisingly significant note was that p53 mutations are predominantly distributed in the exon 5-8 which lead to a biased analysis of mutations only in the DBD (Soussi and Lozano 2005). The hotspot mutations are further categorized into three classes depending on how they deactivate p53 protein. Class I mutations Arg248 and Arg273 are DNA-contact residues that lead to the loss of essential contacts with target DNA which inhibits p53 protein transcription activities. Class I mutations are made up of residues from conserved regions in the loop-sheet-helix that form specific contact with the major grooves of bound DNA target (Joerger and Fersht 2007). Class II mutations Arg175, Gly245, Arg249, and Arg282 are

classified as structural residues that abolish the local structure of p53 which acts as the scaffold in this domain (Selivanova *et al.* 1999). Class III mutations, these mutations affect the globular packing causing the unfolding and structural denaturation under physiological conditions (Saha, Kar and Sa 2015).

Nuclear magnetic resonance (NMR) studies elucidated the first local and global characteristic conformational changes of p53 mutations, whereas X-ray crystallography concluded the structural annotations of the vast majority of oncogenic mutations (Cho *et al.* 1994). The above-mentioned crystal structures depicted clear features of mutations associated with cancer (Joerger and Fersht 2007). However, the global basic structure did not change while the specific local structure varied upon mutations. At a molecular level inactivation of p53 DBD is provided by either distortion or conformational changes in important regions of DNA-binding interface or removal of DNA contacts residues contributing to the DBD stability (Joerger, Allen and Fersht 2004).

2.4.2 DNA-contact mutations

DNA-contact mutations play a vital role in the substitution of residues (Arg248, Arg273, and Arg280) that are sequence specific for DBD contact with DNA leading to abolishing the rights of DNA binding (Ng *et al.* 2015). Moreover, substitution of contact residues lower the binding affinity with DNA and hence p53 protein lose its transcription function while maintaining the wild-type conformation structure (Stojnev, Golubović and Babović 2009). The importance of Arg273 residue side chain comes from its positive charge that interacts with DNA negative charge at the centre of each DNA half-site that is strongly reinforced by hydrogen bonding and electrostatic and van der Waal contacts with the sugar, phosphates, and DNA base atoms (Cho *et al.* 1994). Therefore, Arg273 is crucial for docking interaction with DNA backbone (Kamaraj

and Bogaerts 2015). Crystallographic studies gave birth to the fact that Arg273His and Arg273Cys contact mutations eliminate Arg273 DNA-contact residue without inducing structurally distortion on neighbouring residues yet, maintaining the DNA binding interface scaffold (Saha, Kar and Sa 2015). Ang *et al* urea denaturation studies revealed that kinetic and structural thermodynamic stabilities of p53 wild-type protein are similar to that of Arg273His mutant. And that Arg273His mutant possesses a greatly reduced residual DNA binding activity (Ang *et al.* 2006). Substitution of His by Cys residues results in a drastic reduction in the binding affinity (Kamaraj and Bogaerts 2015).

Contact residues Arg248 and Arg280 exhibits similar roles within local changes for cancer-associated mutations (Muller and Vousden 2013). However, crystal structure data revealed that Arg248 is openly exposed to the solvent hence is not involved in the stabilization of DBD but directly connects to target DNA. In addition, Arg248 residue directly interacts with the minor groove of target DNA (Lukman, Lane and Verma 2013). Mutants Arg248Ala, Arg248Trp and Arg248Gln have been widely considered as the mutations in this residue and they lead to the loss of p53-DNA interaction. Arg248Trp mutant reaches to communicate via loop L3 to induce structural flexibility at loop L1 (Saha, Kar and Sa 2015). Arg248Gln mutant at Arg248 residue is associated with head, neck, skin and colon cancers (Barakat *et al.* 2011). Regardless of the appearance that Arg248Gln is not negatively influencing the stability and dynamics of DBD. Wong *et al* represented the first NMR experimental evidence that this mutant experiences major conformational changes (Wong *et al.* 1999). And again urea denaturation studies using fluorescence spectroscopy presented that Arg248Gln mutant is 2 kcal/mol less stable than the wild-type (Bullock, Henckel and Fersht 2000). On the contrary, other DNA contact mutants (Ser241Phe, and Cys277Phe) does not only participate in the damage of DNA contact interface

but prevent the specificity of DNA-binding due to the introduction of their large aromatic hydrophobic side chains that are solvent expose (Saha, Kar and Sa 2015).

2.4.3 Structural mutations

Structural cancer-associated mutations (Arg175His, Gly245Ser, Arg249Ser, and Arg282Trp) are harboured in the L3 loop, which is important for the binding of DNA to the core domain. This is brought about by Arg248 residue which joins the core domain to target DNA at a significant position via direct or water directed contact. The above-mentioned mutations affect the local and global thermodynamic stability of p53 protein structure in various degrees as concluded by theoretical and experimental results (Joerger and Fersht 2010). Bioinformatics neural network studies indicated that hotspot residues are located in highly conserved regions of the p53 protein (Greenblatt *et al.* 2003).

Two structural cancer-associated mutations Gly245Ser and Arg249Ser are predominantly found in L3 loop. These mutations determine the degree of change in the local structure by mutation effects (Natan *et al.* 2011). Arg249 residue stabilizes the DBD hairpin conformation at L3 for DNA contact but is frequently targeted to structural mutations (Nikolova *et al.* 1998). Crystal structure studies showed that loss of crucial significant interactions by mutation of Arg to Ser at position 249 produces a great deal of flexibility at L3 (Joerger *et al.* 2005). This was noted by an irregular methionine switch involving Met243 and Met246 which revealed the outcome of Arg249Ser mutation. In free p53 protein, Met243 is exposed to the solvent and buried inside the core DBD when bound to DNA. However, Met246 is buried in the hydrophobic pocket in the free and complex states with DNA. The Arg249Ser mutation displaces a buried Met246 from its position in the Zn^{2+} ion hydrophobic region which weakens the α-helix position and form local

structure perturbation. This methionine switch displaces the contact of Arg248 with DNA-phosphate (Saha, Kar and Sa 2015).

The structural mutant Gly245Ser is accommodated by displacing a water molecule. Jeorger *et al* reported that the extent of this structural mutation is relatively smaller in comparison as the L3 overall conformation is conserved (Joerger, Ang and Fersht 2006). On the other hand, this led to small structural shifts that lower protein-protein interaction in G_2 checkpoint pathway upon DNA binding (Joerger and Fersht 2007). Arg282Trp mutant residue forms part of the six hotspots mutations that frequently occur in human cancers. The Arg282Trp mutation is situated at the H2-helix which is part of the loop-sheet-helix motif. Moreover, Arg282Trp mutant residue highly destabilizing effects on the DBD abrogates its function at physiological temperature (Calhoun and Daggett 2011).

2.4.4 Mutations in the DNA-binding domain

In p53 wild-type protein, residue Arg175 is located in between side chains of loop L2 and L3 which are situated in the Zn^{2+} binding region. This residue purely forms a salt bridge with Asp148 and also water mediated hydrogen bonds with Pro191 together with Met237. This interaction stabilizes the L2 and L3 loops into a correct orientation in the interaction with the DNA minor groove (Butler and Loh 2007), hence Arg175 residue is embedded near the four zinc coordinated ligands. DNA-binding studies revealed that Arg175His mutant is the most frequently mutated residue position that highly destabilizes the protein by greater than 3 kcal/mol of energy (Walker 1999). The high destabilization may contribute to the deleterious effect of this mutation. The local structure in the zinc binding region is interfered by the introducing a bulky imidazole ring from the histidine residue which results in the loss of the salt bridge. Arg175His mutation does not

interfere with structural distortion but affects the Zn^{2+} ion binding that is essential for structural integrity in the core domain and reduces the DNA-binding specificity. Substitution at codon 175 can be Arg-His or Arg-Cys, whereas Arg175His mutation is more prone than Arg175Cys as the latter is more frequently mutated than Arg175Cys. Substitution with smaller residues like cysteine at this position is less destabilizing and partially perturb the structure (Duan and Nilsson 2006; Eldar *et al.* 2013).

Mutations at the edge of DNA-binding domain are at the α-helix (H2) of the loop-sheet-helix motif C-terminal which contains the Arg282 residue. The Arg282 residue is vital in maintaining the structure of the loop-sheet-helix motif which binds the minor groove of DNA response elements. The Arg282 residue forms a centre of a network of hydrogen bonds with Gly117 from L1 and Glu286, whereas the Arg282 aliphatic side chain interacts with Phe143 through van der Waals forces. The crystal structure studies show that Arg282Trp substitution interferes with stabilization interaction by abolishing hydrogen bonding interactions of the centre network residues, which in turn induce structural distortion in loop-sheet-helix motif, but not the whole structural scaffold (Zhang *et al.* 2016). Conversely, Arg282Trp mutation displaces L1 loop across because of the bulky Trp side chain that destroys the conformation of the DNA-contact residue Lys120. The His168Arg mutation located in the β-turn region of the L2 loop adjacent to the edge of DNA-binding interface has similar characteristic effect as Arg282Trp. His168Arg mutant negatively affects the local structure but preserve the basic scaffold of the core domain. Lastly, His168Arg mutant destroys the defined conformation of residues 166-170 in the L2 loop.

L1 loop forms a portion of the loop-sheet-helix (LSH) motif region in the core domain with residues that are involved in DNA contact. The L1 loop is seldom affected by mutations, even when it contains DNA-contact residues. Structural studies have shown that loop L1 is the most

labile region in the core domain. This region is very flexible such that the core domain changes the conformation upon DNA binding. The conformational change because of DNA binding affects residues including Lys120 and Ser121 in this region. Numerous elucidated crystal structures of the p53-DNA complex and mutants' structures revealed disordered regions that explain the flexible nature of this loop. The L1 loop is significant in facilitating the binding of p53 with oxygen response elements, which specify the need for its flexibility (Merabet *et al.* 2010).

Leu114, His115, Ser116 and Gly117 residues map the flexible nature of loop L1 and hence are involved in the stabilization of loop L1. Ser116Met mutation destabilizes this region without altering the conformation of the core domain. However, His115Gln, Ser121Phe, and Thr123Pro mutant residues provide altered specific binding affinities of the p53 wild-type to target growth arrest elements such as p21 and pro-apoptotic PUMA gene at G_1 phase. Other studies showed that Thr123Ala mutant increase transactivation that lowers the binding affinity of pro-apoptotic response elements such as BAX1 and PUMA. However, this mutation does not perturb the core domain structure in its DNA-free state but changes the conformational flexibility of L1 loop (Merabet *et al.* 2010). About one-third of mutations in the p53 protein occur in the β-sandwich region but at a less frequency than the hotspots mutations. Some of the β-sandwich mutations include Leu145Gln and Val157Phe which are located at S3 and S4 β-strands respectively. The above-mentioned mutations eliminate the function of the core domain by their high instability at physiological temperatures. β-Sandwich mutations maintain sequence-specific DNA-binding interface at lower temperatures but reduce its function at physiological temperature. The destabilizing energy of the Leu145Gln and Val157Phe mutations is 3 and 3.9 kcal/mol less stable than the wild-type protein (Tan *et al.* 2009). Tyr220 residue is connected towards the end of the β-sandwich at the opening of S7 and S8 β-strands. The Tyr220 aromatic side chain stabilizes the

hydrophobic core of the β-sandwich, while the side chain OH^- group interacts with the solvent. Tyr220Cys is one of the frequently mutated positions outside the DNA-binding interface (Basse *et al.* 2010). Crystal structures revealed that the Tyr220Cys mutant lowers the hydrophobic stability of the DNA-binding domain but does not affect the whole protein backbone structure. This mutant affects the molecular surface by inducing large mutation crevices. Fortunately, the crevice is far from the functional region of DNA-binding interfaces (Accordino, Rodriguez Fris and Appignanesi 2013).

Conversely, β-sandwich mutations are not common but added together account for a considerable amount of mutations associated with cancer from the highly mutated DNA-binding domain. But, mostly the β-sandwich mutations are small to large substitutions in the hydrophobic core that are creating cavities. In p53 DNA-binding domain Val143Ala and Phe270Leu mutations contribute to small-large substitutions. Val143 and Phe270 residues are located adjacent to the strands of the β-sandwich. Further, their side chains are in close proximity to the van der Waals distance with the hydrophobic core. Other studies showed that Val143Ala and Phe270Leu mutations create cavities within the hydrophobic core that do not break the surrounding structure (Shruti *et al.* 2015). P53 studies have shown an interest in temperature-sensitive mutants which include Val143Ala that is not active in body temperature and shows reduced DNA-binding transactivation activity at sub-physiological temperature, however, the mutant residue showed high DNA-binding affinities at temperatures below 306K. Another temperature-sensitive mutant reported is Tyr220Cys that is located in the β-sandwich, where the majority of this mutations reside (Rauf *et al.* 2013).

A number of mutations that seldom occur in human cancers are situated outside of the DNA-binding domain but are situated in the tetramerization domain that interacts with DBD to maintain

the stability of the tetramer at the C-terminus. The mutations in this part of the protein are associated with LFS that is a hereditary early onset of cancer. The Arg337His mutant located in the tetramerization domain is a frequently mutated residue in the germline p53 mutations.

The tetramerization domain is made up of an intermolecular β-sandwich and a packing α-helix that is linked by a hairpin that forms a dimer of a dimer which results into a tetramer. Arg337 residue forms part of the salt bridge and is linked into hydrophobic packing with the dimers of the domain. When Arg is substituted by His at the 337 codon, the domain protonates due to histidine side chain (Kamada 2012). Other cancer-associated mutations are Leu334Pro and Phe341Leu respectively, where the introduction of the proline breaks the α-helix and the later weakens the hydrophobic core formed by the dimers to totally destroy the oligomerization region. Leu334Pro and Phe341Leu mutant residues also alter the equilibrium of the tetramerization domain to prevent the formation of the tetramer. P53 protein with Leu334Pro mutation formed a tetramer with lowered transcriptional and pro-apoptotic activities (Lomax *et al.* 1997). Another tetramerization mutant Gly334Val is associated with lung cancer. In the wild-type, p53 protein Gly at codon 334 mediates the sharp turn which is connecting the β-strand to the α-helix for a more stable conformation of this domain. Mutation at this codon by any residue besides glycine results to a destabilization of the structure. However, these mutations rarely occur in cancer and are not affecting the function of the protein (Lomax *et al.* 1997).

2.5 Computational chemistry methods employed in studying p53 mutations

A vast number of computational studies carried out to date have utilised molecular modelling methods to shed light on the structural details of p53 proteins. Together with the consequences of diverse nature of mutations that affect the interaction of p53-nucleic acids or protein-protein for the rescue of functions have been dealt with (Saha, Kar and Sa 2015). Thermodynamic

stability energies, conformational properties and dynamic behaviour upon mutations have been well documented. An increasing number of molecular docking and molecular dynamic simulations studies have been employed to bridge the structural and sequential evidence of p53 from experimental studies data (Pagano *et al.* 2013). However, the success of both approaches depends on knowing the disease-causing mutations that affect the functions of p53 proteins. This study is therefore aimed at bridging this gap by applying predictive methods that are capable of identifying mutations and their effects on the p53 protein stability.

CHAPTER 3

THEORETICAL PRINCIPLES

The theoretical principles of computational modelling and the strategies based on the sequences as well as structure-based analyses of the point mutations in the p53 proteins are outlined in this chapter. The initial approach includes the use of machine learning methods for predicting the effects of protein stability, followed by molecular docking studies of the wild-type protein with DNA and its identified mutants to better understand the binding modes between p53 wild-type and DNA. Furthermore, the molecular dynamics (MD) simulations were applied to analyse the effects of mutations on the protein flexibility that cause crucial changes in the protein conformation which influence its functions.

3.1 Principles of computational approaches

Molecular modelling is the art of learning the important relationship between a molecular structure and its function through computation by building a model (Schlick 2010a). The Oxford English Dictionary defines a "model" as 'a simplified and idealized description of a system or process, often in mathematical terms, devised to facilitate calculations and predictions' The calculations and predictions in computations include molecular dynamics simulations, molecular docking, Monte Carlo simulations, folding free energy methods, biomolecules stability measurements methods, *ab initio* quantum mechanics and other methods (Lindahl 2008a).

3.1.1 Principles of molecular modelling

Throughout the last two decades, the advent and booming access to software and supercomputers led the rapid development of biomolecular modelling. The ever growing technology and

developmental methodologies hold a great promise for biomolecular studies and their applications in computational chemistry. To name a few: advances of improved algorithms of molecular simulations, parallel processors, and improvements of fast supercomputers have been established. However, communication between theoreticians and experimentalists (multidisciplinary collaborations) is a huge contributing factor in facing the challenges of modifying models at a virtual level through biomolecular modelling (Schlick 2010b).

Molecular modelling is classified as a technique that is focused on utilizing the laws of physics and chemistry to study organic and inorganic molecules as well as biomolecules. This technique is widely used in drug discovery, biophysics and biological chemistry (Jorgensen 2004). The exponential growth of computational space allows the spread of interdisciplinary applications from biology to physics and chemistry to molecular modelling (van Gunsteren *et al.* 2006). The increasing computational processing has allowed for big data to be processed and characterized by building models. The importance of molecular modelling and simulations methods is growing because of its contribution to the study of protein functions, harboured by the structural features at an atomic level including protein flexibility, conformational changes, structural dynamics and prediction of stability changes in proteins upon mutations (Schlick 2010b).

3.1.2 Prediction of mutations and generation of stable mutants

3.1.2.1 Protein stability

Proteins are biological macromolecules that perform diverse specific functions in organisms. Their primary structure (sequence) is composed of a linear polypeptide chain of amino acids. The function of the proteins is determined by correctly folding into a unique 3D structure with a stable conformation (Kang and Kini 2009). Determination of the folding pathway of a 3D structure is

still a difficult task to date regardless of the gamut range of studies in protein folding. Despite all the challenges of protein folding the knowledge of forces that drives folding and unfolding to form a stable 3D structure is still unknown. But the folding free energy change (energy difference between two folding states) is directly related to the thermodynamic stability of proteins as represented below.

$$\Delta\Delta G = \Delta G_{unfold} - \Delta G_{fold}$$

Conversely, substitution of the key residues which alters the folding free energy difference is important for thermodynamic stability which influences proteins activation and regulation. Therefore, the most obvious reason that negatively impacts the stability of proteins is mutation a (substitution of single amino acid residue to another). And determining the degree of mutations effect on protein stability quantified by folding free energy ($\Delta\Delta G$) is of great importance. Because mutations that significantly changes the protein function through stability alteration result into diseases formation (Teng, Michonova-Alexova and Alexov 2008). And more adversely, mutations at functional regions are destabilizing the protein structure leading to its malfunction. So the interest of predicting the effect of mutations in these regions is of great importance.

The forces that are driving the small net changes in the enthalpy and entropy conformational stability are the biggest contributors in destabilizing the protein structure. Where, hydrogen bonding and hydrophobic effects are important stabilizing forces thus far. The burial of the residues side chain in the core of the protein molecule far away from the solvent builds intramolecular contacts which determine protein folding. A missense mutation with different physicochemical properties from wild-type residue generally creates a negative effect on the previous contacts in between residues and therefore cause a different folding pattern from the wild-type protein. As a result of this consequence, a stable mutant protein will form with a different conformation from wild-type protein or a correctly folded mutant protein with a less stable conformation (Stefl *et al.* 2013).

Molecular modelling techniques accurately predict the changes in protein stability upon mutations to rationalise the structural function by understanding the role of each amino acid in the protein structure. But these methods are time-consuming and require a 3D structure for a small dataset. On the other spectrum, there is a wide range of efficient web servers that utilizes both the sequence and 3D structures to predict the consequences of mutation on the large dataset at the blink of an eye.

The neural network servers mainly focus on the folding differences between wild-type and mutant proteins as a function of free energy change. They are classified according to several categories which include (i) principles approaches which used the atomic model to calculate folding or binding of free energy (ii) approaches that use statistical potential and (iii) machine learning algorithms that are applicable to protein stability by using trained experimental data while noting the temperature and pH parameters. The use of these parameters is essential for assessing free energy changes at near physiological conditions (Zhang *et al.* 2012). These methods are used in conjunction with each other to produce accurately precise results. In this thesis, numerous machine learning algorithms will be applied to predict and analyse the stability energy changes upon mutations in p53 protein.

3.1.3 Prediction of stable mutants from a protein

3.1.3.1 Pathogenic prediction (Filtering tolerant/Intolerant)

Machine learning algorithm such as Sorting Intolerant From Tolerant (SIFT) (Kumar, Henikoff and Ng 2009) and Pmut (Ferrer-Costa *et al.* 2005) are related web servers that use evolutionary geometries with an assumption that mutations at functional region are less tolerated and hence destroys protein function. SIFT is a position specific algorithm that utilises sequence alignment information to compute the aftermath of amino acid substitution in the protein function. SIFT's

underlying principle is that mutation at important positions in a protein sequence is intolerant and the protein function is affected at that position. The multi-step workflow of SIFT begins with compiling a family of protein sequences in a protein database through the PSI-BLAST algorithm. Homologous sequences are selected and aligned to look for amino acid composition at specific positions then calculate the scores. SIFT incorporate sequence conservation and physicochemical properties of each 20 amino acids neighbours in predicting whether the mutation is tolerant or intolerant. The output of SIFT is a position specific scoring matrix of 0-1 where a threshold of <0.5 is deleterious or intolerant, and >0.5 is tolerant or neutral at that position (Kumar, Henikoff and Ng 2009).

Pmut is another sequence based web server that is used in conjunction with SIFT for effective prediction of deleterious mutations. Pmut is a robust algorithm which uses neural networks of a large dataset to predict the pathogenicity index of a mutation through the application of sequence-based conservation profile (Kamaraj and Purohit 2013). This method is a neural network classifier that determines the disease associated mutations. The Pmut parameters include mutation descriptors, residue and sequence properties and solvent accessibility to calculate the pathological character of a single mutation. The success rate of Pmut is greater than 80%. The protein sequence is the Pmut input that is subjected to scanning of mutation hot spots followed by immense mutation of the whole sequence to pinpoint regions with definite pathological impact (Ferrer-Costa *et al.* 2005). The assessed genetically accessible mutations are scores with a pathogenicity scoring function of 0-1. All the mutations with indexes >0.5 are predicted to be pathological.

3.1.3.2 Stability prediction (sequence-based)

The second group of algorithms that were used to evaluate the extent of mutation on protein stability involved I-mutant 2.0 (Capriotti, Fariselli and Casadio 2005), MuStab (Teng, Srivastava and Wang 2010) and EASY-MM (Folkman *et al.* 2016) combines different physicochemical properties of amino acids that are trained from experimental mutations. I-mutant 2.0 is a unique web server that is based on support vector machine (SVM) to predict the Gibbs free energy (ΔΔG) values using both sequence and 3D structure. It correctly measures the difference in stability of the new protein from ΔΔG of the wild-type protein. A positive and a negative ΔΔG sign is said to predict the direction to where mutation will swing the stability of the protein. The accuracy of I-mutant web server reaches 80% with a Pearson correlation value of 0.71 in comparison to experimental ΔΔGs (Capriotti, Fariselli and Casadio 2005). Similarly, MuStab is another SVM sequence-based classifier web server with the same prediction accuracy as the structure based web servers. MuStab predicting properties is not only the mutation stability effect but also the function of the protein. It utilizes the amino acids biochemical, structural and other biological features in it prediction which increases its accuracy to 84.50%, sensitivity of 70.29% and specificity of 90.98% (Teng, Srivastava and Wang 2010). Lastly is Evolutionary, Amino acid, and Structural Encodings with Multiple Models (EASY-MM) web server that accurately predicts protein stability changes upon mutations. The EASY-MM is a sequence-based SVM web server that employs five specialised models in its prediction. But, EASY-MM chooses only two models in the final prediction depending on the viability of the secondary structure and available surface area of the mutation position. It accuracy reaches a Pearson correlation score of 0.59 in a 10-fold cross validation. EASY-MM prediction feature infers that mutations in conserved or functional regions result into the destabilization of the protein structure (Folkman *et al.* 2016). Mutations ΔΔG value is predicted highly accurate if supported by sequence analysis (Stefl *et al.* 2013).

3.1.3.3 Stability prediction (structure-based)

The last group of web servers relies only on the structural information. Cologne University Protein Stability Analysis Tool (CUPSAT) (Parthiban, Gromiha and Schomburg 2006) predict and analyse protein stability changes by using the difference in unfolding ΔΔG between wild-type and mutated proteins with high efficiency. It uses structural course-grained atom potentials and torsion angles potentials in its prediction. The prediction reliability and accuracy reaches over 80% when tested against trained sets. However, Site Directed Mutator (SDM) (Worth, Preissner and Blundell 2011) is fabricated on statistical potential energy function to predict the consequences of mutation on protein stability. It uses the position-specific mutation frequencies within similar protein families to calculate ΔΔG between the wild-type and mutated proteins. SDM accuracy reaches a Pearson correlation score of 0.80. SDM has an increased sensitivity in the prediction of protein stability change and its role in the disease.

3.2 Molecular docking

3.2.1 Introduction

Molecular docking is a computational method which predicts an ideal orientation of one molecule and conformation of the other molecule when they bound together to form a complex with minimum energy. This method mainly measures the binding energy of the two molecules: the receptor and the ligand (Kitchen *et al.* 2004). Docking algorithms like molecular dynamics simulations and Fast Fourier Transform (FTT) are incorporated to predict the orientation of the ligand into the active site of the receptor (Lengauer and Rarey 1996). It is known that the ligand tries to orient itself in order to fit into the receptor's cavity by a search algorithm which is a bit challenging since biomolecules contain many conformational degrees of freedom. In this way, the protein cavities become active when external ligands bound to them hence are thus called the active sites (Kitchen *et al.* 2004). Docking of small molecules i.e ligands against larger molecules

receptor like a protein is a daunting task. Since docking of these molecules is expected to be a simple *Lock and Key* mechanism, where the receptor and the ligand are treated as rigid bodies that do not change their conformations after binding. However, ligands and proteins are flexible organisms that show well defined altered conformations after binding through *Induced Fit* Mechanism especially around the active site (Meng *et al.* 2011).

In contrast to the poor representation of ligands and proteins flexibility, most docking methods employ search algorithms to maximize accuracy to produce scoring function nearer to native energy. The search algorithms provide many conformational degrees of freedom in the protein's active site and search the dihedral angles while the protein is held rigid. The scoring function evaluates the complementary binding modes of the complex and provides a ranking of the configurations where the top scoring pose has the minimum energy (Ferreira *et al.* 2015). In most cases, the success of a docking algorithm is measured by reproducing with reasonable accuracy the protein-ligand complex structures that were observed by experimental crystallography (Zhou *et al.* 2007). The accuracy of the predicted protein-ligand complexes binding pose is quantitatively measured by the Root Mean Square Deviation (RMSD) that is compared to the experimental binding pose of the complexes. The scoring function of molecular docking provides substantial degree accuracy in predicting the stable conformation of small ligands within a protein receptor hence it is therefore essential in structural biology and drug discovery design (Brooijmans and Humblet 2010). The predetermined approach of investigating the ligand-binding modes and corresponding interactions is of great importance.

3.2.2 Types of interactions

The interaction movements between the ligand and the protein is a result of forces between the two molecules (Leckband 2000). These forces are classified into electrostatic forces, steric forces, electrodynamics forces and solvent related forces (Sushma and Suresh 2012).

(i) **Electrostatic forces –** Forces with electrostatic charges that govern biomolecules. The most common charges are charges-charges, charges-dipole and dipole-dipole interactions.

(ii) **Steric forces –** Steric forces are formed when two different molecules are in contact with each other and start to impact on their reactivity.

(iii) **Electrodynamics forces-** These are non-covalent interaction forces that are known as van der Waals interactions.

(iv) **Solvent related forces-** Forces that are incorporated as a result of the chemical reactions between the solvent and protein-ligand complex. i.e hydrophilic (hydrogen bonds) interaction and hydrophobic interaction

(v) **Other physical factors-** Conformational changes of the protein and the ligand after binding is a necessity of molecular docking (Sushma and Suresh 2012).

3.2.3 Scoring functions

Scoring functions are quantitative calculations used to estimate the strengths of non-bonding interactions called binding affinity/energy of the docked protein-ligand complexes. The scoring functions are highly dependent on the positions the ligands fit to the active sites of the proteins receptor. Important physical-chemical factors such as specific non-covalent interaction forces, desolvation and entropy effects involved in the prediction of the binding energy of the protein-ligand complexes are essential for increasing the accuracy of the scoring functions. However,

increasing the number of variables in the function increases the computational time, which in turn reduces the accuracy of the docking algorithm. So preferably, a successful scoring function is key to balance between accuracy and speed (Ferreira *et al.* 2015).

Scoring functions are classified into three groups namely, force-field-based, empirical, and knowledge-based scoring functions. Force field based scoring functions prediction of the binding energy is the addition of all the covalent (bond stretching, angle bending, and dihedrals angles) and non-covalent (electrostatic and van der Waals interactions) bonding terms. However, this type of scoring function is limited in estimating the entropic effects (Meng *et al.* 2011).

Binding energy

$$\Delta G\ bind = \Delta Gvdw + \Delta Ghbond + \Delta Gelect + \Delta Gconform + \Delta Gtor + \Delta Gsol$$

Empirical scoring functions evaluate each variable such as non-bonded interactions (hydrogen bonding, ionic and non-polar bonds), desolvation and entropic effects by defining the physical events involved in the creation of the protein-ligand complex in the docking process. The limitation of this function is that it is highly dependent on the used training set when developing the model. However, empirical scoring functions are faster than force field-based scoring functions because of the simplicity in the incorporated energy variables.

Knowledge-based scoring functions are the last approach employed in the evaluating the protein-ligand complex binding energy. This method obtains its function by applying pairwise energy potentials obtained from known protein-ligand complexes. The rate at which two different atoms are positioned at particular distance cut-off in the structural dataset is considered when constructing the potentials. The essential interactions observed in the dataset are rated as the frequency of occurrence and summed up to give a final score. The benefit of knowledge-based scoring functions is that it does not depend on reproducing binding energies or employing many

variables in the function so it provides a suitable balance between speed and accuracy (Meng *et al.* 2011).

3.2.4 Docking algorithm used in the study

ZDOCK is a protein–protein docking program that incorporates Fast Fourier Transform (FTT) search algorithm (Chen, Li and Weng 2003) to predict the protein-protein interactions at an atomic level. ZDOCK treats one protein as a rigid body and allows three rotational and three translational degrees of freedom of the second protein around the first protein. The initial stage of the program is to search for different conformations by permitting a combination of geometry fit, electrostatics complementary, and desolvation factors to enable a systematic search of uniform shapes of docked protein poses on a spatial grid-base. After the initial search, the identified docked poses are ranked and scored with statistical pairwise shape complementary (PSC) scoring function. PSC comes with high accuracy because it employs desolvation, van der Waals and electrostatic energy terms to score interaction energy of the docked poses. PSC does not only score interactions on the protein active site but recognizes atomic contacts in close proximity with the protein active site at considerable distances (Chen, Li and Weng 2003). In the second stage, the selected docked poses are re-ranked and proceed to energy refining with a CHARMm based minimization (Sable and Jois 2015).

Moreover, benchmark tests have proven ZDOCK to be a successful method with high accuracy factor of greater that 70% success comparing to previous rigid-docking methods (Pierce *et al.* 2014). Benchmark tests were not the only tested success but, this method has shown notable gains in docking experiment, Critical Assessment of Predicted Interactions (CAPRI). ZDOCK program was developed to provide researchers with an efficient user-friendly interface to predict protein complexes. Even when ZDOCK was designed for protein-protein interactions it can be pushed for protein-nucleic acid interactions (Fanelli and Ferrari 2006). It can be applied to a server on its

own or can be part of a suite like Accelrys Discovery Studio. Accelrys Discovery Studio is a molecular graphic program that was developed for visualization of proteins, nucleic acids, and small organic molecules.

3.3 Molecular dynamics simulations

The basic principle of molecular modelling is that structural predictions and molecular properties are calculated from mechanical-like models subjected to the laws of classical physics. A molecule is represented by a mechanical system where atoms are treated as balls of different sizes connected together by springs that vibrates, rotate and translate in order to adopt stable conformations upon forces acting on it. The inter and intra molecular forces are portrayed as a collection of harmonic-like terms for bond-angle and bond-length deviation from ideal values, dihedral terms for rotation, non-covalent van der Waals and electrostatic potentials (Karplus and Petsko 1990).

Molecular mechanics (MM) denotes the summation of simple potential energy values (i.e harmonic oscillators or Coulombic potentials) to model biomolecular systems. MM method calculates the energy of the system as a function of atoms position only, ignoring the electronic motions as applied in quantum mechanics.

There are three principles governing the effectiveness of MM derived from Schrödinger equation.

(i) The thermodynamic hypothesis which assumes that there is a robust thermodynamic energy force as a function of position which controls the conformation of the molecule from high free energy to stable low free energy.

(ii) The principle of summation assuming that real molecular potential energy may be added to a sum of potential energy comprised of physical factors (covalent and non-covalent interactions).

(iii) Lastly, MM is applied to the belief of transferability which assumes that potential energies developed from experimental data can be utilized to build in new functional macromolecules.

MM potential energy functions are applied in energy minimization, molecular docking/scoring and subsequently molecular dynamics simulations.

3.3.1 Force Field: functional form of potential energy function

Biological processes are large systems that occur at long time scales, so it is a necessity to develop parameterization potentials (force fields) to escape the need to solve Newton equations of motion. Force fields are empirical approximations that calculate and evaluate the potential energy of the studied system. The potential energy of the system is based on the positions of point like atoms (denoted as atomic Cartesian coordinates). Force fields are composed of equations used to calculate the potential energy terms and the forces from the atom coordinates and limitations used in these equations. However, as beautiful these equations are, they cannot account for bond breaking and formation as in quantum mechanics. The potential energy equation of the force field is always composed of bonded interactions to account for covalent bond stretching, angle bending, proper and improper dihedrals and non-bonded interactions consisting of Lennard-Jones and Coulomb electrostatics terms (Ditzler *et al.* 2010) as shown in **Figure 3.1**.

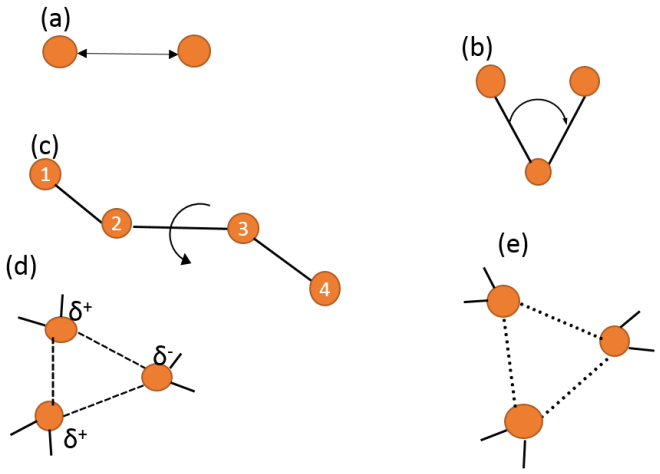


Figure 3.1 Schematic representation of the important components of molecular mechanics force field. (a) bond stretching, (b) angle bending, (c) dihedral rotation, (d) non-covalent electrostatic interactions and (e) non-covalent van der Waals interactions (Figure source (Lindahl 2008a))

$E_{total} = E_{bonded} + E_{non\text{-}bonded}$

$E\ total = \sum bonds\ Ei^{bonds}(bi) + \sum angles\ Ei^{bondangles}(\theta i) + \sum dihedrals\ Ei^{tor}(\varphi i) + \sum non\text{-}bonded\ pairs\ Ei < j^{r}(r_{ij})$

The total potential energy of a 3D structure system is defined by bonded and non-bonded interactions. The terms of 3D structure associated with the potential energy calculation are the summation of the contributing covalently bonded terms from the near equilibrium values of the bond length and bond angles (bi, θi) and internal dihedrals flexibility φi together with the contribution of non-covalently bonded interactions between the atoms positions $r_{ij.}$ A typical example of a generally used force field is CHARMm force field. Similar to any other widely used force fields CHARMm is comprised of discrete functional terms that describe the intra and intermolecular forces within the system as follows.

$$V = \sum_{\text{bonds}} k_b \left(b - b_0\right)^2 + \sum_{\text{angles}} k_\theta \left(\theta - \theta_0\right)^2 + \sum_{\text{dihedrals}} k_\phi \left[1 + \cos\left(n\phi - \delta\right)\right]$$

$$+ \sum_{\text{impropers}} k_\omega \left(\omega - \omega_0\right)^2 + \sum_{\text{Urey-Bradley}} k_u \left(u - u_0\right)^2$$

$$+ \sum_{\text{nonbonded}} \epsilon \left[\left(\frac{R_{\min_{ij}}}{r_{ij}}\right)^{12} - \left(\frac{R_{\min_{ij}}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{\epsilon r_{ij}}$$

The first two terms of the above equation are the harmonic potentials expressing atomic interactions between two (bonds stretching) and three (angles bending), which keep the bond/angles to near ideal values. These are an energy approximation of a bond/angle, as a function of position from the ideal bond length (b-b_0) or bond angle (θ-θ_0) that the atoms have moved. The bond force (k_b) and angle force (k_θ) constants determine the power of the bond/angle, respectively. The third term is a sporadic torsion angle potential function which computes the steric barriers between four atoms separated by three covalent bonds. The oscillatory nature of the terms is used to describe the associated rotational motion around a bond that leads to the changes in the position of atoms. Fourier series terms describe the dihedrals parameters that requires the ($k_ø$) force constant which is associated with the rotational barrier height, the multiplicity of the function *n* which marks the periodicity of full cycle rotation about the dihedral, the dihedral angle (ø) and (δ) phase shift dictates the location where the torsion angle passes the minimum value. The fourth term denotes the improper out of plane angle bending frequency (ω-ω_0) where (k_ω) is the force constant. The fifth term comprises of the Urey-Bradley component that denotes 1,3 distance between non-bonded interactions atoms to treat the harmonic angle bending term (U) in order to precisely model the vibrational spectra with the (k_U) force constant and between the 1,3 atoms.

The last term denotes non-bond interaction energy term that is important for computational studies of biomolecules. This term is comprised of the exchange of dispersion and repulsion terms that strike a balance between neutral atoms. These non-bond interactions are used to treat van der Waals interactions known as Lennard-Jones 6-12 potential and a Coulombic term representing the electrostatic energy. The dispersion and repulsion terms are combined in Lennard-Jones potential depending on the position *i* and *j* of the interacting atoms. The effect of repulsive/attraction terms becomes valid with decreasing inter-atomic distances ($Rmin_{ij}$) which position atoms at optimum distances to stabilize the system and giving rise to minimum energy. The Coulombic term represents the electrostatic interactions between charged groups with point-like charges q_i and q_j. Collectively all these terms and parameters are required to calculate the total energy of the system that is represented by a force field.

3.3.2 Long range interactions

As mention earlier, the sum of bonded and non-bonded interactions marks the total energy of the system. However, direct evaluation of long- range electrostatic interactions is not compensated as it is computationally expensive and time-consuming. Therefore the non-bonded intermolecular interaction forces are truncated at particular distances to account for the unavoidable slow drifts formed. These cut-off distances brought about significant non-physical effects that appear at distance cut-off. Particle Mesh Ewald (PME) process came to the rescue of solving the issue of long- range electrostatic interactions by overcoming the issue of instability in the simulation. The distance cut-off effects caused instability of the simulation that leads to errors. The use of PME in the treatment of the electrostatic interactions calculates the energy of the system coupled with periodic boundary condition to eradicate the errors (de Souza and Ornstein 1999).

Long-range electrostatic effects influence the charge and polarity of molecules which makes them significant in determining the structure, function, and motion of biomolecular system such as protein-protein interactions, protein-ligand binding, protein-nucleic acid binding and thermodynamic effects as they globally affect the whole system. Modelling methods that are based on empirical force fields are applied in molecular dynamics (MD) simulations of biomolecules which provide the motions at an atomic level and the studied free energies by conformational changes of the system. The implemented CHARMm force field has been widely used to study proteins in explicit solvents. Apart from interaction with other macromolecules, proteins interact with water solvent to form a natural reality setting. To simulate a system that matches the experiment as close as possible, water molecules and sodium chloride physiological concentrations are added to the system in order to solvate the proteins. Transferable interaction potential 3 points (TIP3P) parameters (Jorgensen, Chandrasekhar and Madura 1983) for water and Dangs' parameters for ions are used (Smith and Dang 1994).

3.3.3 Periodic Boundary Conditions

It is normal in simulations to soak a molecule in the box filled with solvent, artifacts such as preferred orientation of solvent molecules to the solute and increase in pressure by surface tension arise due to boundaries of the system. To escape from these artifacts periodic boundary conditions (PBC) are employed to avoid any surface boundary when modeling in the bulk solvent. When modelling biomolecules an infinite PBC is employed so that the system infinitely interacts with images of itself and replicate in all directions to form a lattice (Harms, Mittra and Ko 1994). Different choices on boxes are used which include cubic, octahedron and rhombic to allow the fit of all protein shapes to reduce the number of solvent molecules. The solvent models applied in the simulations of proteins include TIP3P, TIP5P and SPC water (Bader and Chandler 1992).

3.3.4 Molecular dynamics simulations principles

In order to describe any physical system, one requires the solution of time-dependent Schrodinger equation for the N-particle wave function $\hat{H} = E\Psi$, where $\hat{H}$ denotes the Hamiltonian, Ψ the wave function and E the energy of the system. Due to large systems of interacting particles, any attempts to use the Schrodinger equation to solve any system seems to be in vain. Therefore approximations which follow the dynamics of the system are needed to compensate for the expensive computing time (Derker 2001). Molecular dynamics (MD) simulations is a theoretical method that computes the motion of biomolecules under the effect of force from interacting atoms energy potentials (Derker 2001). The motion of atoms in the molecule is calculated by the rules of the applied force field. The natural evolution of atoms motion as a function of time is governed by the integration of the Newton Second Law of motion that describe the positions and velocities of atoms. This can be translated to (F_i) as the total force of each particle *i* that is defined by the slope of the potential energy U at each position of the particle (atom) r_i as a function of time (Lindahl 2008b).

$$F_i = -\frac{\partial U}{\partial r_i}$$

The acceleration a_i of the particle is the derivative of velocity over time.

$$a_i = \frac{dv}{dt}$$

The change in velocity and position is determined by the integration of acceleration and velocities as a function of time. Where the momentum p of the particle is defined by the mass m_i and velocity which are produced by the kinetic energy.

$$K(v)\ \frac{1}{2}\sum_{i=1}^{N} m_i v_i^2 \qquad \text{and} \qquad K(p) = \frac{1}{2}\sum_{i=1}^{N} \frac{p_i^2}{m_i}$$

For this system, the potential energy function is not connected to the kinetic energy function but the sum of both energy functions is termed as the classical Hamiltonian (H) equation. Thus the Newton's law of motion is represented in terms of Hamiltonian.

$$H(p,r) = K(p) + U(r)$$

Where r is the set of coordinate components, p is the velocity (momentum) of the particles and U is the potential energy of the system. At an atomic level, the system is represented by a set of r and p where the velocity v_i changes the position r_i at a particular time.

$$V_i(t) = \frac{d}{dt} r_i(t)$$

Based on the initial coordinates, the numerical equations are integrated by MD integrators algorithms such as leap frog to provide a series of solutions of velocities and positions that results in the formation of new conformations. The series are repeated for predefined cycles. The thermodynamic averages of conformations produced by this procedure are collected into a canonical ensemble. Initial coordinates are given velocities based on Maxwell-Boltzman distribution at a specified temperature (Derker 2001). As macromolecules, dynamic properties are not measured as direct observations but an ensemble of billions of molecules called systems. So quantifying equilibrium properties are not about reproducing individual atomic trajectories but to generate equilibrium of ensembles by molecular dynamics simulations. However starting coordinates are very far from equilibrium, which may distort or crash the simulation of the system, therefore, it is a vital necessity to relax the system with energy minimization of the system (Lindahl 2008b).

3.3.5 Energy-minimization processes in simulations

Energy minimization is an important concept for solving non-linear optimization problems in molecular modelling of biomolecules. Numerous energy minimization algorithms exist and are

rooted in the specified order of derivatives for locating the least potential energy surface. The widely used programs are those that include the first order derivatives namely; steepest descent and conjugate gradient. First order derivatives methods apply the gradient of energy function to locate their global minimum energy. Meanwhile, second order derivatives like Newton-Raphson apply the Hessian function to locate their minima (Adcock and McCammon 2006).

3.3.5.1 Steepest Descent method

Steepest descent method is an example of the first order derivatives that uses repetitions of the descent gradient potential-energy surface. The method uses molecular mechanics force descriptors to guide the path towards the closest minimum energy without changing the molecular structure. The applied force lowers the potential energy of the whole system by channelling atoms path. An empirical or step size iteration is randomly selected. Successive iterations are adjusted according to an oscillation of the potential energy of the system in that step. The step-size to reach the nearest local minima is adjusted according to the increase or decrease of the energy function. Meanwhile, the steepest descent method is not suitable for multidisciplinary problems it is however robust in locating the nearest minimum energy. And it can only be used to remove bad atomic contacts and quickly correcting geometries. It can also be employed to locate the closest energy of the system which is far from a minimum as a rough initial run following accurately advance algorithms like conjugate gradient method (Adcock and McCammon 2006).

3.3.5.2 Conjugate Gradient method

Conjugate gradient and steepest descent minimization methods are both first order derivatives that use gradient potential surfaces but differ from each other in the number of ways. The conjugate gradient benefit is to further employ minimization history to locate the search direction

path in and converges to the local minima faster than steepest descent. Another reason is that it uses numerous direction paths to compensate for the oscillation of the energy potential closer to the minima. Subsequent configurations gather the information about the energy function from the initial cycle to the next iteration to refine the energy potential closer to the local minimum. For each iteration, the gradient is calculated to and added a new information in determining the new direction path to locate the local minimum energy at a less computational cost time. Hence conjugate gradient is a better choice to choose for energy minimization of larger molecules.

In non-gradient energy minimizers, the initial coordinates and small increments of energy are considered when calculating the geometry of biomolecules through energy minimization. However, gradient energy minimizers utilise the atomic gradients to compute the molecular geometry. This has been proved by calculating the root mean square gradients of the forces of each atom in a molecule (Adcock and McCammon 2006).

3.3.6 Thermostats in MD

A solution of Newtonian equations of motions in the application of molecular dynamics simulations, the energy E, volume V and number of atoms N are conserved, thus MD simulations system results in micro-canonical ensembles. However, energy and volume variables are not a good choice for experimental conditions. Therefore it is important to perform MD simulations where experimental conditions are imitated. Then the interest to mimic physical properties like temperature and pressure in the MD simulations to attain a more realistic environment was developed. Thermostats that stabilize the energy fluctuation are coupled to the MD system to eliminate system boundaries and attain a more realistic setting, approximating the canonical ensemble (Berendsen *et al.* 1984).The widely used algorithms that use the thermostat in their systems allow the sampling of the isochoric-isothermal ensemble (NVT): a system where a

number of atoms N, volume V and temperature T are constant then isobaric-isothermal ensemble (NPT): where the number of atoms N, pressure P and temperature T are constant, respectively. In the instance of applying NVT to NPT ensembles, the system is set to stabilized temperature first followed by pressure. MD follows the time evolution of a system, allows atoms velocities to be rescaled to maintain a constant temperature. Hence it is a necessity to correct some degrees of freedom with constraining algorithm such as SHAKE method (Uline and Corti 2013).

3.3.7 MD simulations in practice

A workflow of standard molecular dynamics simulations is comprised of (i) obtaining the starting structural coordinates (ii) immersing the structure in the solvent box and neutralising the system (iii) relax the system through minimization (iv) heating the system to desired temperature (v) equilibrating the solvent box with a constant volume (vi) generating the trajectories in the production step (vii) Analysing the trajectories. The steps are expanded in the chapter to follow.

CHAPTER 4

COMPUTATIONAL METHODS

The computational approaches used in this study include the identification of mutations in the p53 sequence and the stability of p53 protein using sequence and structure-based methods, followed by molecular docking to determine the binding interactions between p53 with DNA. Finally, molecular dynamics (MD) simulations used to understand protein flexibility at an atomic level as articulated in Chapter 3 is presented here.

4.1 Evaluation of the prediction methods

Amino acid sequence (UniProtKB id: P04637) in a FASTA format was downloaded from UniProtKB database for the prediction and analyses of stability changes upon mutations in p53 protein. The 3D crystal structure of the p53 protein was collected from Protein Data Bank with PDB ID: 1TSR (http://www.rcsb.org) (Berman *et al.* 2000). Thereafter, the downloaded amino acid sequence was submitted to the SIFT server to screen out the intolerant amino acid positions. Amino acid positions that showed intolerance were filtered using SIFT scoring function. The collected mutations were loaded into the Pmut server to search if they are damaging, deleterious or neutral. The Pmut server pathogenic score filtered deleterious mutations. All the observed pathogenic mutations were further tested to analyse the effect of mutations in p53 protein stability using sequence and structure-based web servers. I-mutant 2.0, EASY-MM, MuStab, CUPSAT and SDM web servers were used for studying the effect of mutations in the p53 protein. In all programs default parameters were applied. Binary predictions (pathogenic/neutral) and numerical values were considered as per guidelines given by the authors of programs for output results.

4.2 P53 mutant modelling

As protein modelling requires a clean 3D protein structure. The first step involved the acquiring of structural files from Protein Data Bank with PDB ID: 1TSR (http://www.rcsb.org) (Berman *et al.* 2000). However, it was noticed that the structure was bound to ligands and contained similar chains. This was followed by the removal of unwanted ligands, repeated chains, adding hydrogen's and removing water molecules and metals.

P53 mutant structures were modelled by MODELLER (version 9.15) (Feyfant, Sali and Fiser 2007) Build mutants module present in Accelrys Discovery Studio (DS) 2016. An X-Ray crystal structure of p53 (PDB ID: 1TSR) (Cho *et al.* 1994) chain B monomer was used as the basis of modelling. The first step in build mutant module was making point mutations to the specified wild-type residue. Followed by the optimization of all atoms of the modelled mutated residue with a combination of conjugate gradient minimization and molecular dynamics simulated annealing at 4.5 Å cut-off radius within surrounding residues. Molecular mechanics energy terms that are involved in loop modelling, disulphide bridges formation, and cis-proline restraints were omitted to simplify model building.

The residue-by-residue energy profile of the predicted 3D models was evaluated by calculating the Molecular Probability Density Function Mol (Mol PDF) energy and Discrete Optimized Protein Energy (DOPE) scores (Shen and Sali 2006) to select the best fit model. DOPE energy score was calculated by comparing the relative stability of the mutant with the wild-type to measure the quality of the latter (Jaya and Saleh 2012). Five mutant models were generated for each mutation per wild-type protein. Among the generated mutants' conformations, the one with the highest Mol PDF energy score and the most negative DOPE score was selected (scores at Appendix I). The observed mutation residues modelled were Val143Asp, Ala159Pro, Val197Pro,

Try234Pro, Cys238Pro, Gly262Pro and Cys275Pro. Selected p53 mutant models were subjected to side chain refinement in order to optimize the conformation of a chain fragment based on CHARMm energy minimization using CHIRotor algorithm. The measured conformation energy is sole of the protein primary structure considering the energetic and steric constrains (Spassov, Yan and Flook 2007). Thereafter, the whole protein structure was then optimized with CHARMm 22 forcefield and subjected to energy minimization by smart energy minimizer algorithm for 4000 steps with RMS > 0.1 using Accelrys DS 2016. The 3D structures of the selected predicted and experimental p53 mutants are shown in the table in Appendix I.

4.2.1 Solvent accessible surface calculation

The surface area of an atom or residue is defined as the area where the latter is interacting with the solvent molecule (Connolly 1983). Solvent Accessible Surface Area (SASA) is defined as the area from the central radius of the solvent is measured compared to the calculated surface area the atom or residue solvent can stretch towards. The algorithm used to compute SASA calculates the surface area acquired by the propagating sphere along the surface of the molecule (a protein), obtained by merging van der Waals radii atoms in a protein. The radius of the sphere is normally set to 1.4Å, to the van der Waals of the water molecule. The area covered by the centre of the solvent while propagating along the surface of the protein is SASA as investigated by (Lee and Richards 1971). The term Solvent Accessible Surface (SAS) will be used as it is applicable to the package used to calculate SASA for this study. The majority of the identified mutations were located in the hydrophobic core area of DBD. SAS was calculated to determine whether the mutations were disrupting the hydrophobic architecture.

In this work an in-house residue solvent accessibility area module in Accelrys DS 2016 was used to calculate the percentage solvent accessibility surface (%SAS) of each residue, based on Lee and Richards work of solvent accessibility calculation procedure, a probe radius of 1.4 Å was applied. P53 wild-type and modelled mutant structures were used as the basis to calculate SAS and hydrophobicity scale was used in the prediction. P53 wild-type % SAS for each residue was compared with the mutant % SAS residue. A residue with a % SAS below 50% was classified as buried residue and vice versa. The same procedure was followed for experimental modelled mutants. Various hydrophobicity scales have been developed up to this day for different experiments. Here, Kyte and Doolittle scale will be utilised to test the extent of disruption by mutated residues (Kyte and Doolottle 1982): it combines methods of the past in a dimensionless range from -4.5 for hydrophilic to 4.5 for hydrophobic as seen in **Table 4.1**.

Table 4.1 Kyte & Doolittle hydrophobicity scale

Amino acid	Arg	Lys	Asp	Glu	Asn	Gln	His	Pro	Try	Trp
hydrophobicity	-4.5	-3.9	-3.5	-3.5	-3.5	-3.5	-3.2	-1.6	-1.3	-0.9
Amino acid	Ser	Thr	Gly	Ala	Met	Cys	Phe	Lue	Val	Iso
Hydrophobicity (cont)	-0.8	-0.7	-0.4	1.8	1.9	2.5	2.8	3.3	4.2	4.5

4.3 The p53-DNA docking

4.3.1 Computational details

After reviewing the concept of molecular docking in chapter 3, the docking protocols implemented in this work are outlined below. Accelrys Discovery Studio 2016 commercial software was used as the primary graphical user interface (GUI) for the calculations and visualization bearing in mind it is one of the few programs with built-in graphics for results viewing. The ligand was prepared using "Prepare ligand" module. The protein was prepared using "Prepare protein" module in Accelrys DS. Protein-DNA rigid docking was undertaken by ZDOCK housed at Accelrys Discovery Studio 2016. ZDOCK algorithm uses FTT in exploring docked poses and scoring them by PSC scoring function that is taking into consideration the shape, electrostatics and desolvation associations (Chen, Li and Weng 2003). ZDOCK protocol included launching ZDOCK and analysing ZDOCK results to obtain a docking ZRANK energy score (Pierce *et al.* 2014).

4.3.2 Protein preparation

An X-ray p53 DNA binding domain crystal structure bound to duplex DNA used in this study was retrieved from Protein Data Bank (PDB) repository with PDB ID: 1TSR (http://www.rcsb.org) (Berman *et al.* 2000). Specifically, chain B of the structure was selected for docking as is was gathered from the literature that it interact extensively with the consensus DNA binding site. As it is known to be bad practice to directly use raw crystal structures from the database because it might contain several unwanted details that were useful for crystallization. Therefore, the downloaded protein structure was refined by removing water molecules, bound ligands, deleting alternate conformations and added hydrogen atoms and protonating titratable residues predicted pKas by using "Prepare protein" module. The final structure atoms were typed with CHARMm 22 forcefield to prepare it prior to molecular docking. Thereafter, the whole

structure was subjected to energy minimization with smart minimizer for 4000 steps until RMS of ≤0.1 was achieved. Then the protein target was ready as an input for docking. The same experimental procedure was carried out for wild-type, predicted and experimental mutant forms of p53 modelled in their monomeric states.

4.3.3 Ligand preparation

Prepare ligand module was utilized to prepare the input coordinates of duplex DNA (with 21 base pairs 5'-TTTCCTAGACTTGCCCAATTA-3') ligand. Protonation was generated at pH 7.4 in order to reflect physiological conditions. The resultant structure was typed with CHARMm 27 force field. An *in vacuo* energy minimization was performed using smart minimizer for 2000 steps until a convergence energy of ≤0.1 Kcal/mol was obtained.

4.3.4 ZDOCK protocols

ZDOCK was designed for protein-protein interactions but it can be pushed for protein-DNA interaction (Fanelli and Ferrari 2006). The structures of p53 DBD and duplex DNA were refined and optimized with Accelrys DS 2016 tools. Initially, a systematic search of rotational and translational degrees of freedom was explored on the p53 wild-type, its mutants against duplex DNA. A rotational sampling interval of 6 Å (dense sampling) was used and kept all other parameters as default settings. The best 2000 out of 54000 docked structure poses were retained and re-ranked with ZRANK according to the best-docked structures to improve the success of docking predictions. Finally, docked protein-DNA complex poses were clustered according to their densities. Docking simulation was employed on the p53 wild-type and Val143Asp, Ala159Pro, Val197Pro, Try234Pro, Cys238Pro, Gly262Pro and Cys275Pro mutant structures of the DNA-binding domain as well as Val143Ala, Tyr163Asn, Arg213Gly, Tyr220Ser, Gly245Ser

and Gly245Val experimental mutations. The binding affinity interaction quantities were estimated with the ZRANK score, comparing the wild-type with its subsequent mutant as the workflow in **Figure 4.1**.

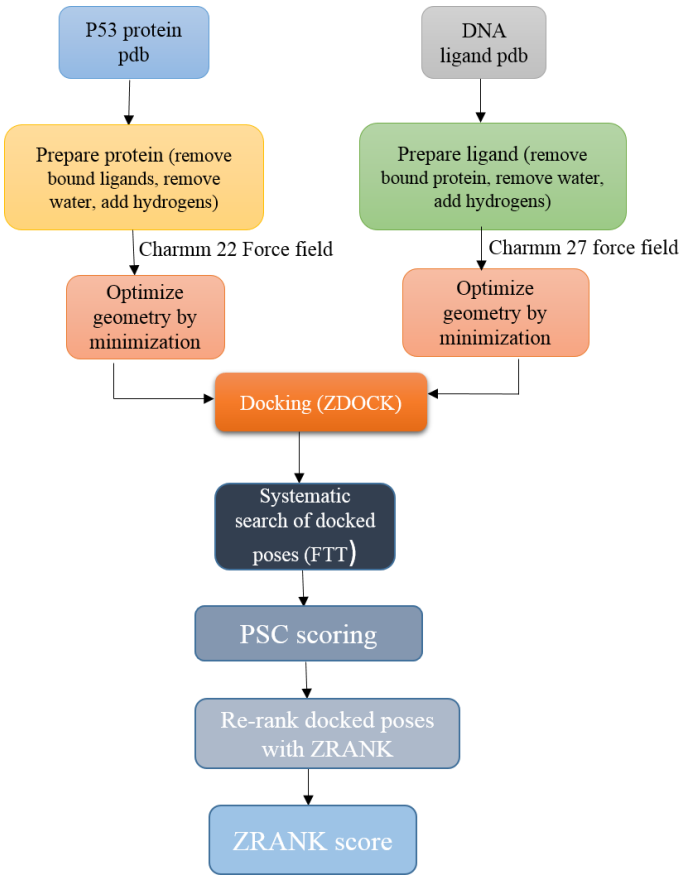


Figure 4 1 Computational workflow representing ZDOCK docking protocol

4.4 P53 and mutants MD simulation protocols

The MD simulations were performed at the molecular level using GROMACS 5.2.1 simulation package (van der Spoel *et al.* 2005). GROMACS is a user-friendly research tool with all the parameter files and topologies clearly written in the text format specially designed for protein dynamics study through molecular dynamics theory (Pronk *et al.* 2013). The atom functions of p53 protein and all the mutants were parameterized with a CHARMm27 force field. MD

simulations method was carried out on the p53 wild-type, Val143Ala, Arg213Gly and Gly245Ser experimental mutant proteins at 300K to investigate their stability profile. The p53 protein and its subsequent mutants were immersed in a cubic water box with dimensions of 10 Å separately applying periodic boundary conditions for solvation. The SPC216 (simple point charge) water molecules were added to solvate the protein and mutants. Counter ions Na^+ and Cl^- were added to neutralize the total charge of the protein and mutant systems using the genion utility in GROMACS. The solvated structures were energetically minimized using steepest descent for 50000 steps. Further, the systems were then heated from 0 K to 300 K after the required minimization process. Non-bonded interactions were counted for by PME to treat long-range interactions. The van der Waals and Coulombic cut-off were used as 14 Å and 12 Å, respectively. Hydrogen bond lengths were constrained with SHAKE algorithm to enable a 2fs time step. The equilibration was attained in two Berendsen canonical ensembles, isochoric-isothermal NVT (constant volume, temperature, and a number of atoms) for 100 ps and isobaric-isothermal ensemble NPT (constant pressure, volume, and number of atoms) for 100 ps to relax the systems. After equilibration the production run of all the MD simulations completed the length of 100ns, saving the coordinate at regular time intervals of 2ps. The resulting trajectories were analysed using g_energy, g_rmsd, g_sasa and g_rg utilities in GROMACS. The outputs were plotted using Origin software. All graphic presentation of the 3D model were prepared using Accelrys Discovery Studio 2016. The workflow of the protocol is given below **Figure 4.2**.

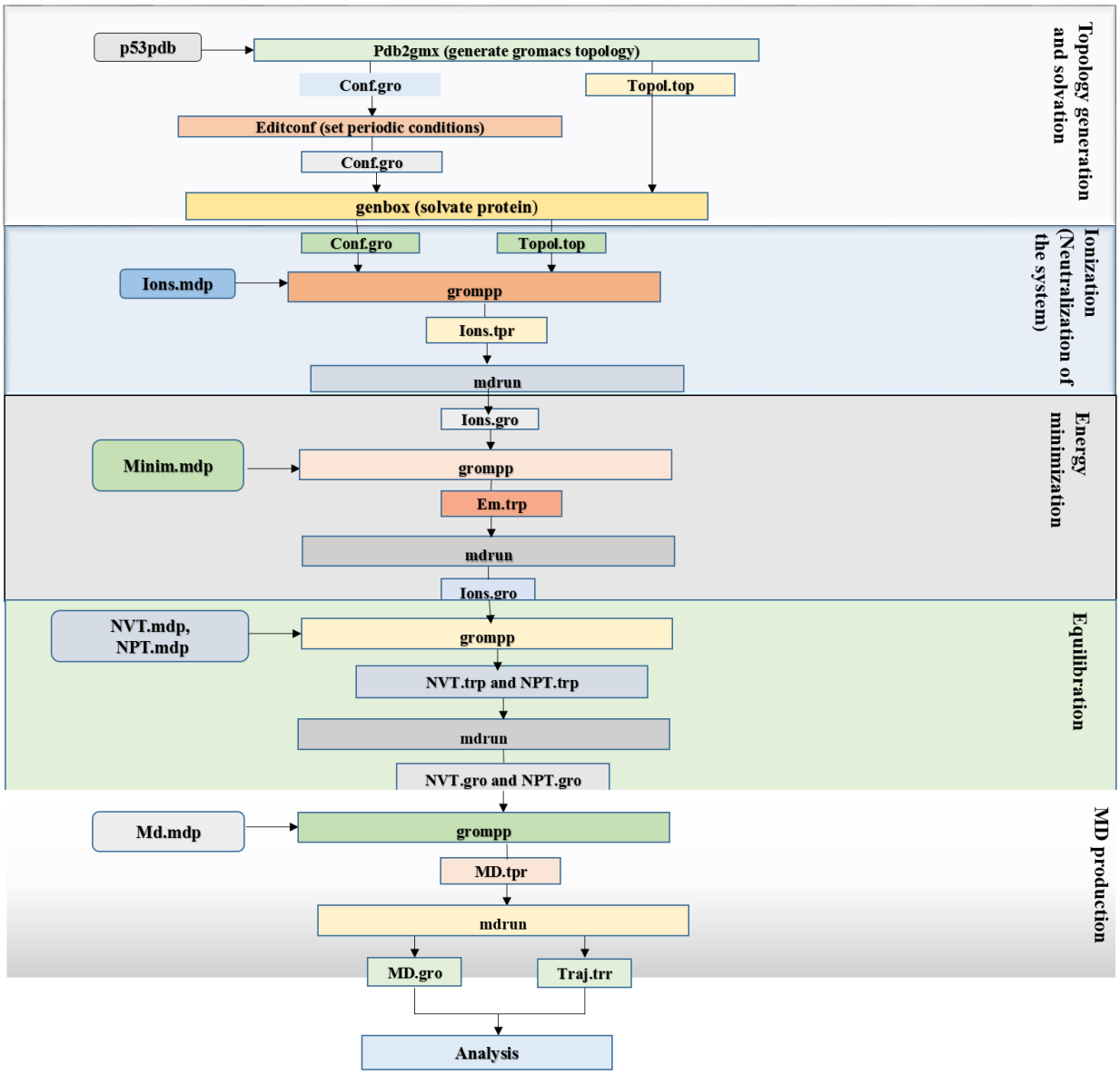


Figure 4.2 The workflow description of the molecular dynamics simulation using GROMACS package

A high resolution structure was selected and the bound ligands and water were removed while missing residues were added using Accelrys Discovery Studio 2016. The PDB file to use as the input for GROMACS must contain only the protein atoms. The topology input coordinates were created with the pdb2gmx module. Incomplete amino acid residues present in the structure were refined to prevent the fail of the pdb2gmx. Therefore, the PDB structures of p53 protein were refined with scwr14. The processed structure was parameterized using CHARMm 27 force field.

When the pdb2gmx topology file was generated, the protein and the mutant structures were solvated separately. In GROMACS the simulation box is firstly defined and poured in the solvent.

1. The editconf module was used to define the size of the box.
2. The box was filled with water using the genbox solvation module.

The cubic periodic boundary conditions were defined, imposing a minimal distance between the protein and the box walls of 1.0 nm. Thereafter the whole system was solvated using SPC/E261 water model. The charged protein in the system was neutralized by adding counter ions. Information of the pdb2gmx module was used to add Na^+ and Cl^- ions. The genion utility module was used to add Na^+ and Cl^- ions in order to neutralize the systems by the use of -mpd file parameters. Steepest descent algorithm and Velvet cut-off schemes were applied in carrying out neutralization. Thereafter, when the solvated system was neutralized minimization was carried out to remove bad atomic contacts and structure distortion. GROMACS utilize a pre-processing program called grompp to collect the coordinates, topology and all other parameters into a single run input file tpr to start running the simulation.

Energy minimization was performed with the steepest descent for a maximum of 50000 steps. This process was performed until a maximum force of <1000 kJ/mol/nm was obtained. Long range non-bonded interactions were treated with PME. Hydrogen bonds lengths restrained with SHAKE. The van der Waals and Coulombic electrostatic interactions cut-off at 14 Å and 12 Å, were specified using the –mpd file parameters. The minimized structures were heated from 0 K to 300 K to institute a proper orientation. The minimized structures were relaxed through equilibration using Berendsen weak coupling NVT and NPT ensembles. The equilibration process of the solvent and ions around the protein was rather important because the water solvent was at optimum with itself and not with the protein. The first equilibration was performed with NVT at constant volume for 100 ps at 300K. Similarly, a second equilibration at NPT to stabilize

the pressure in the system for 100 ps. The equilibrated systems with respect to temperature and pressure were produced and further applied for the production phase. Finally, MD simulations were executed at 100 ns for each run on GPU nodes using mdrun module available at CHPC (Centre for High-Performance Computing) in Cape Town. The output trajectories were analysed using statistical software tools.

CHAPTER 5

RESULTS AND DISCUSSION

This chapter provides the results obtained from the logical analysis of the effects of mutations on the stability of p53 DNA-binding domain (DBD), using the sequence and structure based methods leading to the selection and the modelling of stable p53 mutants. The predicted stable p53 mutant structures were validated using the properties of the SAVES server which included VERIFY-3D, PROCHECK and the Ramachandran plots. This was followed by Solvent Accessibility Surface (SAS) calculations to quantify the surface of the protein residues that were exposed to the solvent. Further, the predicted stable mutants were docked on to DNA using ZDOCK to better understand the p53-DNA binding interactions in relation to experimental mutant forms. Finally, Molecular Dynamics (MD) simulations were performed to better understand the structural variations of the p53 wild-type in relation to experimentally determined mutant proteins at an atomic level.

5.1 Analysis of mutation effects on p53 protein stability

5.1.1 Sorting intolerant from tolerant (SIFT)

Amino acids in the DNA-binding domain (DBD) are involved in p53 transcriptional activities however, they were found to possess a maximum degree of intolerance. Surprisingly, mutations in this domain result in an increase or decrease in the stability of the full-length protein preventing specific DNA-binding, leading to a loss of transcriptional activities. An intolerant of 19 means that these amino acids have a high susceptibility to be mutated. Important residues responsible for p53 protein binding to DNA are Lys 120, Ser241, and Arg280, but all were found to have a maximum number of intolerant equal to 19 while Ala276, Arg273, and Cys277 showed the number of intolerant equal to 18 as presented in Appendix I. All these residues are critical for the

p53-DNA interface and their mutation will prevent the protein to bind or lose specificity of interactions with DNA, however they are not identified as hotspot residues except for Arg273 (class I contact-mutation). This clearly shows the negative effect of these mutations may have on the p53-DBD binding interaction.

5.1.2 Prediction and selection of stable mutants

Effects of mutations on the structural stability of p53 protein were evaluated by sequence and structure-based web servers to predict destabilizing residues and assessed their changes in the protein structure. The DNA-binding core domain monomer was selected for this study. No specific position was targeted for point mutations in this domain. Out of 292 amino acid residues in the DNA-binding domain, 156 were predicted by online web servers employed in this work to be either stabilizing or destabilizing the structure, respectively. Amino acids with the highest frequency of substitution in their descending order of magnitude were Pro, Asp, Lys, Arg but Trp, Leu, Ile, Gln, Val, Tyr and His showed the least frequency as displayed in **Figure 5.1** showing that mutation to proline is very common compared to other amino acids but their frequency does not provide fully understanding of their individual role in development of cancer only their main biochemical activity (Martin *et al.* 2002).

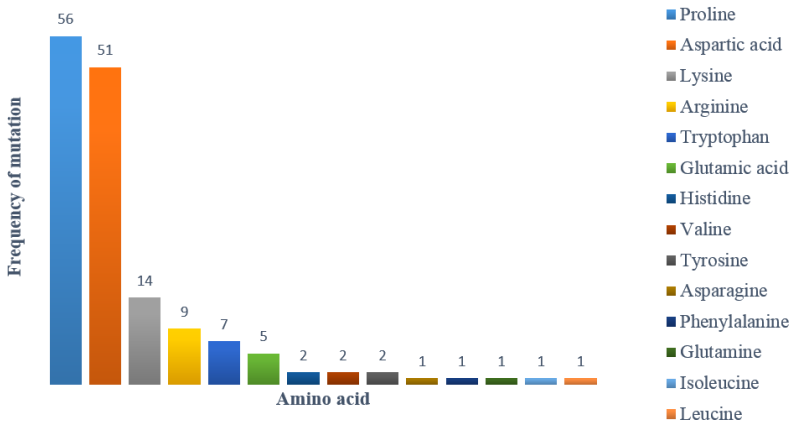


Figure 5.1 Graph of the frequency of mutation on the basis of amino acid occurrence

Free energy change (ΔΔG) scores generated from SDM server for each amino acid substitution at near physiological conditions were mostly considered for selection of mutants because this server does not only predict the effect of stability in a protein, it also provides the role that specific mutation has on the disease as presented in a table in Appendix I. The mutation positions showing the highest scores of ΔΔG for SDM server were selected for further evaluation. The highest SDM ΔΔG score was observed at position 262. Where the small wild-type glycine is replaced by a controversial cyclic proline at this position to yield a Gly262Pro mutant residue. Other residues with highest SDM ΔΔG scores selected for the study were Val143Asp, Ala159Pro, Val197Pro, Try234Pro, Cys238Pro, Gly262Pro and Cys275Pro mutant residues. The combined results of I-mutant, MuStab, EASY-MM and CUPSAT servers further suggested that the selected mutant residues were destabilizing the protein and deactivate the structure leading to a defective protein (**Table 5.1**).

Table 5.1 Computational predicted mutations combined effects of I-mutant 2.0, MuStab, EASY-MM, SDM, CUPSAT and Accelrys Discovery Studio 2016 in the prediction of mutation stability of p53 protein DBD

P53 DBD residues and mutations	**I-mutant 2.0 (ΔΔG) kcal/mol[a]**	**MuStab (ΔΔG) confidence[b]**	**CUPSAT (ΔΔG) kcal/mol[c]**	**SDM (ΔΔG) kcal/mol[d]**	**EASY-MM (ΔΔG) kcal/mol[e]**	**Discovery Studio (ΔΔG) in kcal/mol [f]**
Val143Asp	-3.65	91.07%	-10.13	-5.01	-2.80	2.53
Ala159Pro	-3.33	81.61%	-6.40	-3.35	-0.62	5.28
Val197Pro	-2.72	92.68%	-6.62	-4.14	-2.89	6.46
Tyr234Pro	-3.90	92.86%	-2.03	-3.48	-1.84	25.62
Cys238Pro	-3.32	92.86%	-8.34	-6.00	-1.57	6.79
Gly262Pro	-0.87	82.32%	-3.68	-6.62	-0.81	7.43
Cys275Pro	-2.36	83.39%	-10.06	-5.72	-1.10	20.75

[a]I-mutant stability effect = Decrease (sequence and structure-based).
[b]MuStab stability effect = Decrease (sequence-based).
[c]CUPSAT stability effect = Destabilizing/Unfavourable (structure-based).
[d]SDM stability effect = Highly destabilizing/caused protein malfunction and disease (structure-based).
[e]EASY-MM stability effect = Destabilizing (sequence-based).
[f]Discovery Studio stability effect = Destabilizing (structure-based).

The stability energy of the mutated positions were further analysed with Accelrys Discovery Studio (DS) 2016 package. The predicted magnitude of mutation stability energy for stable mutants with Accelrys DS 2016 at pH 7 and 298.15K for Val143Asp, Ala159Pro, Val197Pro, Try234Pro, Cys238Pro, Gly262Pro, and Cys275Pro were 2.53, 5.28, 6.46, 25.62, 6.79, 7.43 and 20.75 kcal/mol, respectively. The only inconsistency among the stability predictions with Accelrys DS 2016 was the Val143Asp mutation which was calculated to decrease the stability of the protein with an insignificant effect.

Prediction and analyses of the effect of Val143Asp, Ala159Pro, Val197Pro, Try234Pro, Cys238Pro, Gly262Pro and Cys275Pro p53 mutant residues using a combination of I-mutant, MuStab, EASY-MM and CUPSAT and SDM were in consensus among each other. The results of the sequence and structure-based web servers were in kcal/mol and negative energy change ($\Delta\Delta G$) indicated that the predicted mutation is decreasing the stability or destabilizing the structure and vice versa. Comparison of the mutation stability energy outputs of Accelrys DS 2016 with online web servers was established to show that they are in agreement with each other, only that the destabilizing effect in Accelrys DS 2016 has a positive energy change ($\Delta\Delta G$) sign and the destabilizing effect of other web servers have a negative energy change ($\Delta\Delta G$) sign. All programs above predicted that mutations destabilized the p53 DBD protein together with Accelrys DS 2016 as hypothesized prior.

5.1.3 Folding pattern disruption effects

Mutation stability effects can be described by structural effects that govern protein folding. Where mutation to proline may introduce a bulkier, bad-contact causing residue. Mutation to proline residues was mostly found to be destabilizing the protein structure when a wild-type residue had a torsional backbone angle which was disallowed for proline. When a protein folds to form a 3D

structure, the arrangement of the secondary structure is governed by the backbone conformation of the rotational phi and psi angles (Lovell *et al.* 2003). The deviation in the backbone arrangement of the mutant structures compared to the wild-type were studied by the phi and psi angles of the Ramachandran plots. Ramachandran plot diagram of p53 wild-type showed that 100% of residues reside in the allowed region indicating the consistency of predicted model. However, when looking at the mutant residue positions strong phi and psi angle deviations have been reported comparing to the wild-type residues (**Table 5.2**). The phi of Tyr234 was 179.23 and that of Pro234 mutant was -143.99, this shows a substitution that is disallowed for a cyclic proline and that its side chain does not fit into parent amino acid position therefore this will result into local structural distortion. Another structural distortion is observed with Gly262Pro mutation with a phi of 94.90→-1.12 and a psi of -9.34→76.95 deviation. This clearly showed that this position is clearly disallowed for proline with restricted conformation because of its cyclic side chain that limit its backbone conformation.

Table 5.2 The phi and psi degrees angle showing the arrangement of the secondary structure residues of p53 wild-type and mutants

Amino acid mutations	Native phi degrees	Mutant phi degrees	Native psi angles	Mutant psi angles
Val143Asp	-135.61	-130.26	161.05	142.32
Ala159Pro	-125.28	-99.72	158.39	167.78
Val197Pro	-92.73	-80.76	124.80	138.51
Tyr234Pro	179.23	-143.99	163.89	162.50
Cys238Pro	-152.41	-112.37	127.29	173.22
Gly262Pro	94.90	-1.12	-9.34	76.95
Cys275Pro	-147.66	-136.12	157.71	-152.54

In the Ramachandran plots, phi or psi angle represented by a square clearly depicted a position that is disallowed for that residue as presented in Appendix II. In this work, mutation residues that were highly destabilizing the protein structure were to proline. The protein structure destabilizing effects caused by proline's distinctive bulkier side chain fragment produced an

altered backbone conformation with lots of strain. In addition to web server results, mutant modelling of selected mutations positions were performed to understand whether the introduced side chain would perfectly fit the structure or completely destroy the structure and its folding state. And if, the introduced side chain destroys the structure, how the created bad steric clashes are affecting structural integrity. Below, the reasonable chemical and physical structural characteristics that are causing significant changes in the p53 DNA-binding domain structure are presented.

Val143Asp mutant residue is predicted to be destabilizing by online web-servers and DS program**.** This analysis is in part supported by the work done by Fersht and co-workers that Val143Ala destabilized the core domain with ΔΔG > 3kcal/mol (Bullock, Henckel and Fersht 2000). The Val143 residue is located in S3 β-strand at the opposite strands of the β-sandwich where its side chain forms van der Waals interactions with the important part of the hydrophobic core. As a result, substitution of Val to Asp may cause local to global structure rearrangement. Val is a non-polar hydrophobic amino acid that prefers to be buried, whereas Asp is a charged polar acidic amino acid that will pay a desolvation penalty to the buried Val and expose it to an aqueous environment. Therefore, this substitution created structural reorganization which was predicted to be severely destabilizing the p53 protein core domain by utilized web servers.

Ala159Pro mutant residue is predicted by the applied online web servers and DS program to be destabilizing the core domain. Ala159 is located in S4 β-strand connecting to L2. The Ala residue at position 159 forms part of the β-sandwich hydrophobic core. The wild-type Ala contains a β-carbon meaning its conformation is hindered to the backbone that the structure can adopt. Substitution of Ala to Pro caused structural rearrangement of the neighbouring L2 loop which in

turn disturb the beginning of upcoming H1 α-helix that is involved in the DNA binding interface. This type of reorganization induces extra strain and destabilizes the local structure.

Val197Pro mutant residue is predicted to be destabilizing the structure in agreement with DS package. Val197 resides at S5 β-strand which is connected to H1 α-helix. Val as a hydrophobic residue which is known to be buried in the core of the protein and however stabilizes the 3D core domain structure due to hydrophobic effect. The missense mutation introduced a Pro that has a different polarity and a hydrophobicity which also contain a cyclic side chain that restricts the backbone conformation. The introduction of a proline residue produces an extra strain to the structural backbone as proline is known to cause distortion of the secondary structure by introducing bad clashes into the residues in close proximity to that position. Val197 location at the S5 β-strand next to a loop that connects to S6 β-strand, the introduction of proline will reduce the flexibility of the loop and create a rigid loop. As a result, the stability of core domain 3D structure decreased upon Val197Pro mutation.

Try234Pro mutant residue is predicted to be destabilizing by the web server calculations in conjunction with DS program. Tyr234 is situated at S8 β-strand which connect to L3 loop. Tyr234 makes hydrophobic contacts with buried Val143 and Ile195 which form the heart of the hydrophobic core (Lu, Tan and Luo 2007). Therefore, Tyr234 is a core residue meaning that it is part of residues that are responsible for the protein's global stability. Substitution of Try234 to controversial Pro in the S8 β-sheet that is connected to L3 loop which is responsible for DNA binding will rearrange the structure and eradicate specific DNA binding. The introduction of proline at positions with different backbone chain torsional angles from wild-type induce a destabilizing effect on the 3D structure (MacArthur and Thornton 1991).

Cys238Pro. Cys238 is part of the ligands in an L3 loop that are responsible for Zn^{2+} metal coordination. The ligands keep the L3 loop in a critical orientation because it is required for providing a correct p53 conformation for DNA binding. Substitution of Cys238 with Pro will distort L3 loop proper orientation which will prevent crucial interaction with the DNA minor groove and eradicate specific DNA binding. Cys238Pro mutation will also interfere with the zinc metal binding which is critical for correct folding of the core domain.

Gly262Pro. Glycine plays a crucial role structurally because of its smallest size with no side chain meaning that it can adopt numerous conformations that are hindered for other amino acids providing structural flexibility (Ashworth *et al.* 2014). Substitution of glycine to proline with a restricted backbone conformation will destroy structural flexibility and destabilize protein folding. Especially missense mutation of Gly to Pro at position 262 in the turn which is connecting to S10 and S9 β-sheets that participate in the DNA binding. Again this mutation will reduce structural flexibility in this region and penalise DNA specific binding.

Cys275Pro. Cysteine's in a protein structure provides assistance in binding to DNA (Rainwater *et al.* 1995). Cys275 is located in a loop connecting S10 β-stand and H2 α helix of the loop-sheet-helix motif that is involved in DNA binding. Cys275 modulate the orientation of the loop-sheet-helix of p53 DNA binding site. Therefore, Cys275Pro produces structural reorganization of the loop-sheet-helix. This missense mutation produced the most destabilizing ΔΔG energy using DS program. However, it is within the vicinity of the DNA binding site, so the structural changes due to mutations will reduce the recognition of DNA to its binding site and lower the transactivation activities hence the function of the p53 protein.

5.1.4 Mutant model analysis and validation

The observed p53 mutant structures are displayed in Appendix II. The evaluation of the quality of the models were processed by methods like Mol PDF score and DOPE energies (Appendix II). Moreover, the selected mutant models were further assessed by VERIFY-3D (Luthy, Bowie and Eisenberg 1992) from SAVES for validation as shown in **Figure 5.2**.

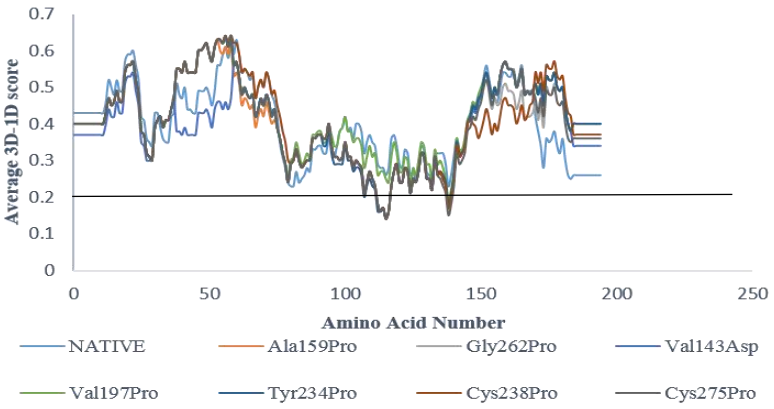


Figure 5.2 VERIFY-3D plots validating generated models

VERIFY-3D defines well-matching residue positions of the atomic model (3D) profile with its amino acid sequence (1D). A negative score of a specified residue means that an acquired conformation by that residue in the generated model is not suitable for the surroundings of the 3D profile. Models with at least 80% of the amino acids above a score of 0.2 in a 3D-1D profile are taken as good models. Figure 5.2 showed that VERIFY-3D 1D scores of the generated models are positive. The generated models were validated by the absence of negative areas in the plots. VERIFY-3D plots of all the mutant structures displayed parts of a scoring area less than 0.2. All the wild-type amino acids were located in good geometries, while more than 80% of mutant structures amino acids were above 0.2 and were taken as good models as verified by PROCHECK. PROCHECK showed that 96.91%, 96.91%, 99.48%, 96.91%, 96.39%, 91.91%, and 96.39% of Val143Asp, Ala159Pro, Val197Pro, Try234Pro, Cys238Pro, Gly262Pro and Cys275Pro mutant residues in the Ramachandran plots were in the most favoured and allowed

regions and were taken as stable models. The spotted deviations were because of the mutations to proline residue which disrupt the structure when inserted in the middle of secondary structure residues (Chou and Fasman 1974).

5.1.5 Solvent Accessibility Area (SAS)

Contributions of each amino acid mutation in the % solvent accessibility surface (SAS) and their location in the structure were computed and represented in **Table 5.3** below, revealed a clear trend of amino acid preferences.

Table 5.3 Comparison on the p53 wild-type residue % SAS and mutants residue % SAS and their location in the p53 structure of computational mutations

P53 domain residues and mutations	Location in the structure[a]	Wild-type % Solvent Accessibility Area[b]	Mutant % Solvent Accessible Area[b]
Val143Asp	S3 strand, hydrophobic core of the β-sandwich, closest to DNA-binding interface and near the LSH motif. Deeply buried in the core.	0.09	0.42
Ala159Pro	S4 strand, highly buried in hydrophobic core of the β-sandwich, connecting to L2 a central structure that connects H1, L2' with the β-sandwich.	1.55	1.67
Val197Asp	S5 strand, towards the end β-sandwich, connecting to L2 that stabilizes H1 α-helix by forming contacts with units of DNA binding region.	6.63	5.63
Tyr234Pro	S8 strand, connecting to L3 that contain DNA-binding region. Produces stabilizing contacts with the core.	2.24	14.24
Cys238Pro	L3 loop between S8 and S9 and is part of the Zn^{2+} ligands, required for the overall stability of the protein.	1.57	0.94
Gly262Pro	Located in a turn that connects S8 and S9 which forms part of DNA binding region.	50.76	52.97
Cys275Pro	Located in a loop connecting S10 and H2 α-helix of the LSH and form part of the DNA binding interface.	6.78	9.13

[a]The position according to the sequence in human p53 protein and PDB 1TSR.
[b]The Solvent accessibility area (%) defined as the solvent-accessible surface area of the amino acid residue in its parent protein.

The calculated residue % SAS and Kyte and Doolittle hydrophobicity scale predicted that almost all the amino acid mutation positions were identified as buried and hydrophobic except for Gly to Pro at position 262. The % SAS of Gly was found to be a 50.76 % and mutation Pro was

52.97% solvent accessible. This mutation was above the 50% threshold which determined the surface of amino acids side chains as buried. Although wild-type amino acids were identified to be buried, replacements to proline projected a slight increase in spaces in-between the DBD core. This might be caused by proline increased polarity that opened up spaces or its cyclic side chain. A contrary of the hydrophobicity was noted for Val at position 197 with a % SAS of 6.63% and Asp showed 5.63%. This finding is not in agreement with the Kyte and Doolittle scale, Val is more hydrophobic and Asp is more hydrophilic. Therefore, Val must have a lower % SAS than Asp. The increase in the SAS of Tyr234Pro from 2.24 %→14.24% is in agreement with distortion of the local structure predicted by the phi of the Ramachandran plot. This showed that the introduction of proline in this position did not match the backbone conformation that was adopted by tyrosine hence there was an increase in SAS and a deviation in Phi and Psi positions.

5.1.6 Comparison of computational and experimental mutants

Experimental mutations collected from the literature were comprised of Val143Ala, Tyr163Asn, Arg213Gly, Cys220Ser, Gly245Ser and Gly245Val. These mutations are known to destabilizes the p53 protein structure and are associated with cancer (Muller and Vousden 2014). When tested on the combination of I-mutant, MuStab, CUPSAT, SDM and EASY-MM and Accelrys DS 2016 their effect was destabilizing the p53 DBD as presented in **Table 5.4**.

Table 5.4 Experimental mutations combined effects of I-mutant 2.0, MuStab, EASY-MM, SDM, CUPSAT and Accelrys Discovery Studio 2016 in the prediction of mutation stability of p53 protein DBD

P53 core domain amino acids and mutations	I-mutant 2.0 (ΔΔG) kcal/mol[a]	MuStab confidence[b]	CUPSAT (ΔΔG) kcal/mol[c]	SDM (ΔΔG) kcal/mo[d]	EASY-MM (ΔΔG) kcal/mol[e]	Discovery studio (ΔΔG) in kcal/mol[f]
Val143Ala	-3.82	92.32%	-6.26	-2.73	-2.17	2.39
Tyr163Asn	-2.72	91.79%	-4.13	-4.75	-1.90	2.51
Arg213Gly	-2.57	87.86%	-6.11	-5.53	-2.92	5.06
Tyr220Ser	2.00	88.93%	-1.78	-3.04	-2.38	4.46
Gly245Val	-1.24	84.82%	-5.38	-3.25	-0.71	8.60
Gly245Ser	-1.18	88.21%	-2.49	-1.50	-0.57	3.96

[a]I-mutant stability effect = Decrease (sequence and structure-based).
[b]MuStab stability effect = Decrease (sequence-based).
[c]CUPSAT stability effect = Destabilizing/Unfavourable (structure-based).
[d]SDM stability effect = Highly destabilizing/caused protein malfunction and disease (structure-based).
[e]EASY-MM stability effect = Destabilizing (sequence-based).
[f]Discovery Studio mutation stability effect = Destabilizing (structure-based).

Similarly, to the tested computational predicted mutations. The penalty for the Val143Ala, Tyr163Asn, Arg213Gly, Cys220Ser, Gly245Ser and Gly245Val substitutions in terms of SDM ΔΔG were -2.37, -4.75, -5.53, -3.04, -3.35 and -1.50 kcal/mol respectively. The most observed deleterious mutation was Arg213Gly with an SDM ΔΔG of -5.53 kcal/mol. A combination of I-mutant, CUPSAT, EASY-MM and SDM server's system analysis of the selected experimental mutations revealed that these mutations lowered the stability of the p53 DBD and they are unfavourable. The destabilizing mutations Val143Ala, Arg213Gly, and Gly245Ser provided the highest I-mutant ΔΔG of -3.82, -2.57 and -1.18 kcal/mol and CUPAST ΔΔG of -6.26, -4.13 and -2.49 kcal/mol respectively. The Val143 side chain is not exposed to the solvent and appear to be deeply buried with a % SAS of 0.09 while the Ala143 mutant residue has % SAS of 1.9 (**Table 5.5**).

Table 5.5 Comparison of the p53 wild-type residue % SAS and mutants residue % SAS and their location in the p53 structure of experimental mutations

P53 core domain amino acids and mutations	**Location in the structure[a]**	**Wild-type % solvent accessibility[b]**	**Mutant % Solvent accessibility[b]**
Val143Ala	S3 strand, hydrophobic core far from DNA-binding interface and near the LSH motif.	0.09	1.9
Try163Asn	S4 strand, connected to L2 loop that proceeds to H1 α-helix, deeply buried, makes hydrophobic contacts with Arg249 for overall protein stability.	0.43	1.2
Arg213Gly	S7 strand, deeply buried in the core, responsible for hydrophobic contacts around L2 loop.	4.64	4.2
Tyr220Ser	Located in a loop between S7 and S8 and maintains hydrophobic contacts.	11.21	19.84
Gly245Ser	L3 loop between S8 and S9 β-strands, critical for backbone conformation (allow the formation of two NH-bonds with Cys247 a zinc ligand, and Arg249).	11.85	20.83
Gly245Val	L3 loop between S8 and S9, critical for backbone	11.85	73.95

	conformation (allow the formation of two NH-bonds with Cys247, a zinc ligand, and Arg249).

[a]The position according to the sequence in human p53 protein and PDB 1TSR.
[b]The Solvent accessibility area (%) defined as the solvent-accessible surface area of the amino acid residue in its parent protein.

In this instance, Val143 is a hydrophobic residue that is responsible for packing of the hydrophobic core during folding (Cho *et al.* 1994). And is known to be forming hydrophobic contacts that stabilize the whole protein (Lu, Tan and Luo 2007). It was predicted in the work done by Fersht and co-workers that truncation of two methyl groups from Val to Ala severely destabilize the protein by 3.50 kcal/mol (Bullock, Henckel and Fersht 2000). The Val143Ala mutation introduced structural changes in all the residues of the β-sandwich which may likely affect the packing of the hydrophobic core.

Gly245Ser is a cancer associated hotspot mutation that destabilizes p53 protein structure (Nikolova *et al.* 2000). The Gly245 is located in L3 loop of the β-sandwich and it was found out to be buried with a % SAS of 11.85. In the wild-type structure, this residue adopts a rare backbone conformation which enables the formation of hydrogen bonds between its amide and the backbone carbonyl of a zinc ligand Cys242 and with the guanidinium group of Arg249 (Cho *et al.* 1994). Substitution to serine at position 245 demolishes these interactions and abolish the structural integrity of L3 loop which provides the DNA binding site (Wong *et al.* 1999). In this work, the 245Ser mutant residue %SAS was calculated to be 20.83 accessible to the solvent. The introduction of Ser with a polar bulky side chain created lots of space in this position hence it appeared near to the surface of the structure. It must be noted that the destabilizing features predicted in this work are in agreement with the rationale provided by theory. Further, experimental mutations predicted here have the same stability effect as the computational mutations.

5.1.7 Summary

In this study, numerous web-servers were simultaneously utilised to predict and analyse the impact of missense mutations on the structure of the p53 protein. This was possible because of the abundantly available sequence and solved 3D structures of p53 DNA-binding domain. The sequence and the 3D structures subjected to energy calculations for each web-server were compared and also combined the obtained results. The distinctive mechanism established from the missense mutations studied in this work was perturbation of the 3D structure to change the wild-type phenotypic characteristics of p53, since destabilizing alone cannot conclude the formation of cancer cell tumours. (Robles and Harris 2010). There were numerous examples of missense mutations that showed a decrease in the stability of either p53 sequence or the corresponding DNA-binding domain structure which further impacted on the whole protein. The established argument was, mutations that alter the wild-type characteristics might result into pathogenic effects or natural difference

Three of the missense mutations, Val143Asp, Ala159Pro, and Val197Pro occur at the hydrophobic core of the DNA-binding domain. They were all predicted to be destabilizing the DNA-binding domain by inducing local structural rearrangements that result in conformational changes. Moreover, experimental data showed that stability of the p53 full-length is determined by the DNA-binding domain core (Ang *et al.* 2006). Kato and co-workers revealed that buried hydrophobic residues in the core domain are crucial for correctly folding of the p53 protein (Kato *et al.* 2003), their substitution affected the p53 protein function by producing an incorrectly folded protein. Missense mutations that do not affect the binding affinity of the p53 protein to DNA, express a case of missense mutations that are associated with cancer through indirect effects in the reduction of protein stability. Further, three mutations Cys238Pro, Gly262Pro and Cys275Pro located within the β-sheet sandwich which act as the basic scaffold of the DNA binding interface

and carries the active site. However, neither the wild type Cys238, Gly262 nor the mutants are in direct contact with the active site except for Cys275. (Duan and Nilsson 2006). The analysis from this work revealed a reasonable development which may involve a loss of specific sequence DNA binding by the latter mutations. More so, Cys238 residue coordinates the Zn^{2+} ion which keeps L3 loop in a proper position to allow minor groove interaction with DNA half site. So Cys238Pro mutation negatively interferes with this coordination which in turn destroys L3 loop orientation which is important for DNA minor groove interaction. Lastly, Tyr234 forms part of critical hydrophobic core domain residues that preserve the whole protein stability. It was projected that replacement of the wild-type Tyr234to Pro changed the tyrosine hydroxyl side chain which is involved in hydrogen bonding in the hydrophobic core.

Most of the amino acid mutations associated with the disease are found in buried positions in a protein comparing to amino acids on the surface. The buried amino acids control structural folding and protein stability. The changes to these amino acid positions, especially in the hydrophobic core, have tremendous disruptions to the overall hydrophobic packing. Val143, Ala159, and Val197 are known hydrophobic core residues, and mutation to Asp and Pro slightly broke the space open but remained buried as there were insignificant changes in the % SAS. Val143 maintain critical stability contacts within van der Waals distance with its neighbours in the core and this might be disrupted when this residue is altered. Gly262 was the only residue identified as exposed to the surface because Gly is usually found in turns on protein structures. So it was not buried inside the protein. Similarly, with experimental mutations, Gly245Val, where the % SAS of wild-type residue was 11.85 and introduction of Val with a bulky β-carbon in this position increased the % SAS to 73.95. It can be clearly seen that there was a local disruption of the L3 loop where Gly245Val is located. The rearrangement of this secondary structure was shown to be destabilizing as predicted by-mutant, MuStab, CUPSAT, SDM and EASY-MM and Accelrys DS 2016 servers. Changes in size, hydrophobicity, and charge of residue side chains at

buried positions might not only affect structural stability but may interfere with protein interaction with other molecules as it appeared in the p53-DNA docking analysis described below.

5.2 Analysis of ZDOCK results

Machine learning web servers have concluded that mutations affect the stability of the p53 protein and interfere with the binding actions. In order to verify this further, molecular docking was consulted to compute p53 wild-type and mutants DNA-binding domain (DBD) interactions with DNA. ZDOCK algorithm was employed to calculate the binding affinities of the two biomolecules. The Receptor-Ligand Interactions tools in Accelrys DS were used to visualize the non-bonded interactions. The binding affinities of p53-DNA and its mutant complexes were quantified with ZRANK a more accurate energetic scoring function. ZRANK scores were obtained by re-ranking highest docked poses in largest clusters to improve the ZDOCK prediction success rate. P53-DNA crystal structural studies showed that the DBD is bound to one DNA half site but adjacent consensus half-site is needed for binding. The binding specificity of the p53-DNA interaction was determined by observed interface residues. The best-docked complex pose with interface residues was chosen for evaluation from highest 10 p53-DNA complex poses in largest clusters of each run.

Wild-type p53-DNA complex produced an optimum binding affinity with the most negative ZRANK (dimensionless energy reliant (Pierce and Weng 2007; Tharakaraman *et al.* 2013),) score of -70.50 maintaining most of the interface residues. P53-DNA mutant complexes showed weaker binding affinities with less negative ZRANK scores of -68.43 to -60.13, where Try234Pro p53-DNA mutant complex showed the least negative ZRANK score of -60.13 as mentioned in **Table 5.6**.

Table 5.6 Molecular docking results obtained by the ZRANK scores and hydrogen bonding (NH-bonds) ordered by the distance between p53 wild-type, computational mutants, and DNA is indicated

P53 mutant residue	ZRANK score[a]	Amino acid[b]	Bond distance (Å)[c]	Bonding DNA part[d]	Bonding Amino acid part[e]	No NH bonds
Wild-type	-70.5	Lys120	2.8826	(PO_4^{3-})Phosphate	(NH_3^+)Side chain	11
		Ser241	3.07819	(PO_4^{3-})Phosphate	(H) Backbone	
		Arg248	1.39228	(N3)Base	(CH_2)Side chain	
			1.91434	(C=O)Base	(NH_2)Side chain	
		Arg273	2.3672	(PO_4^{3-})Phosphate	(CH_2)Side chain	
		Ala276	2.52321	(H3)Sugar	(C=O)Backbone	
		Arg280	2.47995	(PO_4^{3-})Phosphate	(NH)Side chain	
			2.76845	(O4)Sugar	(NH_2)Side chain	
		Arg283	2.71971	(PO_4^{3-})Phosphate	(H) Backbone	
Val143Pro	-67.45	Arg248	2.26369	(PO_4^{3-})Phosphate	(H) Backbone	6
			1.885513	(C=O)Base	(NH)Side chain	
		Arg273	0.892535	(O3)Sugar	(CH_2)Side chain	
		Arg280	2.13708	(N3)Base	(NH_2)Side chain	
			2.01491	(O3)Sugar	(H)Backbone	
			2.89612	(N3)Base	(CH_2)Side chain	
Ala159Pro	-68.43	Arg248	2.25791	(PO_4^{3-})Phosphate	(NH)Side chain	4
			1.84908	(C=O)Base	(NH_2)Side chain	
		Arg273	0.89857	(OH)sugar	(H) Backbone	
		Arg280	2.13375	(N3)Base	(NH_2)Side chain	
			2.89093	(N3)Base	(CH_2)Side chain	
			2.00878	(O3)Sugar	(H) Backbone	
Val197Pro	-66.57	Arg241	2.85845	(PO_4^{3-}) Phosphate	(HN)Backbone	4
			3.03162	(PO_4^{3-}) Phosphate	(HN)Backbone	
		Arg248	2.67544	(PO_4^{3-}) Phosphate	(C=O)Backbone	
		Arg273	2.89824	(PO_4^{3-}) Phosphate	(NH)Side chain	
Tyr234Pro	-60.13	Ser241	2.23656	(O3)Sugar	(CH_2 Backbone	8
			2.21966	(PO_4^{3-}) Phosphate	(CH_2)Side chain	
		Arg248	2.22979	(PO_4^{3-}) Phosphate	(H)Backbone	
		Arg273	1.28546	(O3)Sugar	(NH_2)Side chain	
			2.92736	(C=O)Base	(NH_2)Side chain	
		Arg280	2.86123	(O3)Sugar	(CH_2)Side chain	
Cys238Pro	-64.00	Lys120	2.88818	(PO_4^{3-}) Phosphate	(H)Backbone	9
		Arg248	2.25448	(PO_4^{3-}) Phosphate	(H)Backbone	
			1.85449	(C=O)Base	(NH_2)Side chain	
		Arg273	0.905118	(OH) Sugar	(NH) Side chain	
		Arg280	2.14259	(N3)Base	(NH_2)Side chain	
			2.90174	(N3) Base	(CH_2)Side chain	
			2.01803	(O3)Sugar	(H)Backbone	
			1.22367	(H)Sugar	(C=O)Backbone	
Gly262Pro	-67.92	Lys120	2.87166	(PO_4^{3-}) Phosphate	(NH_3)Side chain	8
		Ser241	2.65057	(PO_4^{3-}) Phosphate	(HN_3^+) Backbone	
		Arg248	1.44155	(N3)Base	(NH_2)Side chain	
			1.89489	(C=O)Base	(CH_2)Side chain	
		Arg273	2.35592	(PO_4^{3-}) Phosphate	(CH_2)Side chain	
		Ala276	2.51947	(H3)Sugar	(C=O)Backbone	
		Arg280	2.80285	(O4) Sugar	(NH_2)Side chain	
		Arg283	2.74143	(PO_4^{3-}) Phosphate	(H)Backbone	
Cys275Pro	-68.13	Lys120	2.8933	(PO_4^{3-}) Phosphate	(NH_3)Side chain	9
		Ser241	3.09888	(PO_4^{3-}) Phosphate	(H)Backbone	
		Arg248	1.43516	(N3)Base	(NH_2)Side chain	
		Arg273	2.34982	(PO_4^{3-}) Phosphate	(CH_2)Side chain	
		Ala276	2.52624	(O3)Sugar	(C=O)Backbone	

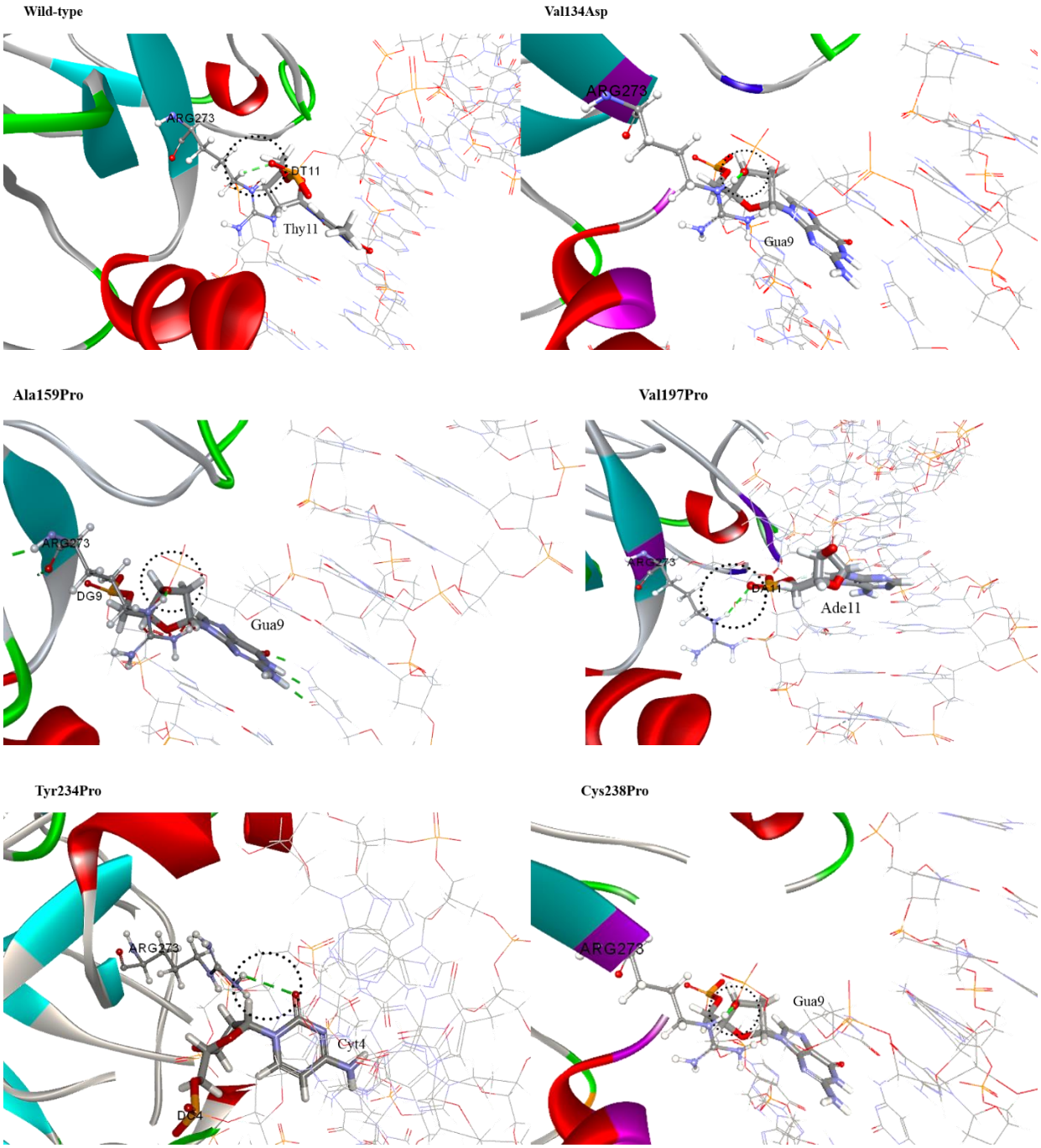
Cys277	2.22017	PO_4^{3-}) Phosphate	(SH)Side chain
Arg280	2.84032	(O4)Sugar	(NH_2)Side chain
Arg283	2.74999	(PO_4^{3-}) Phosphate	(H)Backbone

[a]ZRANK score is a measure of the binding affinity between p53 wild-type its mutant forms and DNA.
[b]Amino acids are the interface residues that are involved in forming DNA contacts.
[c]Bond distance measures the length of the hydrogen bond formed between interface residues and DNA backbone or bases.
[d]Ligand part is representing the part of the ligand responsible for forming a hydrogen bond, it can be a sugar, phosphates or DNA bases.
[e]Bonding amino acid is the interface residue part that is forming a hydrogen bond, it can be a side chain or backbone.

The p53-DNA interface is formed by the L1 loop (aa 237-250) and H2 α-helix (from loop-sheet-helix (aa 119-135 and 272-287)) which interact with the major groove, while L3 loop binding site made contacts with the minor groove stabilized by the Zn^{2+} ion (Duan and Nilsson 2006). Crystallographic references (Cho *et al.* 1994) revealed the DNA binding interface residues namely, Lys120, Ser241, Arg248, Arg273, Ala276, Cys277 Arg280 and Arg283. Structural changes of the DBD/DNA interface residues produced the basic measure of understanding the sequence-specific transcription-activation activities lost by p53 mutant forms. In particular, Arg248 and Arg273 residues have been shown to be critical for interaction with the DNA minor and major grooves, respectively (Cho *et al.* 1994).

Structural defects created by a mutation on each mutant structure produced different degrees of impairment and severity to the binding interface. Moreover, visible deviations of p53-DNA mutant complexes interface residues interacted with DNA at different bases and other interface residues did not participate in binding as it was shown in the crystal structure analysis. Earlier studies by Kamaraj *et al* noted that Arg273 residue is vital for docking because its positive guanidinium group makes major contacts with the negative phosphate DNA backbone which is supported by hydrogen bonding and electrostatic interactions (Kamaraj and Bogaerts 2015). Docking results of this work showed that wild-type p53-DNA complex, Arg273 interacted with Thy11 of DNA. While, in Val143Asp, Ala159Pro, and Cys238Pro mutant complexes, Arg273

interacted with Gua9. In the Val197Pro mutant complex, Arg273 interacted with Ade11. The Try234Pro mutant Arg273 interacted with Cyt4. In the Gly262Pro and Cys275Pro mutant complexes, Arg273 residues interacted with Thy11 separately (see **Figure 5.3**).



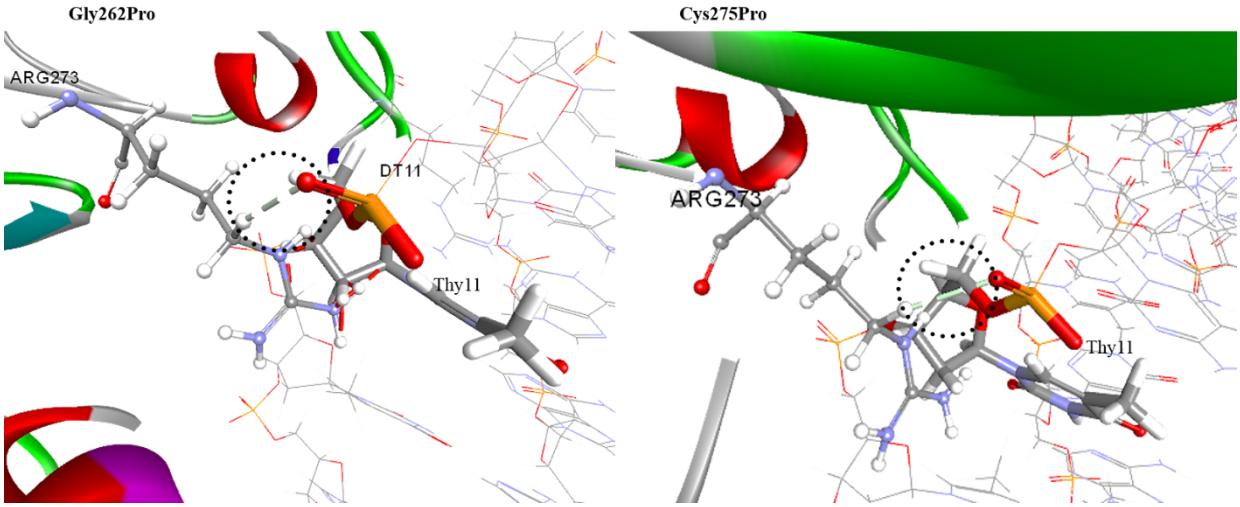


Figure 5.3 Specific Arg273-DNA interaction shown by green/cyan dotted lines (circled): Arg273 residue is shown in ball and stick, DNA base in stick form. Rest of the protein is shown in ribbon cartoon and DNA as wires

The beginning of the Val143Asp, Ala159Pro, Val197Pro, and Cys238Pro p53-DNA mutant complexes results showed that Arg273 residue was bound to the consensus site sequence 5'-PuPuPuC(A/T) DNA strand. While the wild-type p53, Try234Pro, Gly262Pro and Cys275Pro p53-DNA mutant complexes showed that, Arg273 interacted with the reverse-consensus site sequence 3'(T/A)-GPyPyPy DNA strand as investigated in crystal structure analysis by Cho and co-workers (Cho *et al.* 1994). The divergence of specific binding by the Arg273 residue to Thy11 might be the reason for the decrease in binding affinity of the mutant complexes.

5.2.1 Hydrogen bonding

Many studies of the crystal structure for p53-DNA complex have emphasized on the central role played by hydrogen bonding network in the interaction and stabilization of the two biomolecules. Some reports have shown that failure of p53 to bind DNA has been linked to the loss of hydrogen bonds by residues in the conserved regions of the DBD or the introduction of non-native hydrogen bonds (Wright and Lim 2007). Hydrogen bonds formed between p53-DNA interfaces through ZDOCK molecular docking were analysed by Receptor-Ligand Interactions tool in Accelrys DS

2016. Protein residues were mostly hydrogen bond donors and DNA bases were hydrogen bond acceptors with the exception to Asp281 and Ala276 residues which behaved as both donors and acceptors. The number of hydrogen bonds produced by the backbone and side chains interface residues of wild-type and mutants is related to the binding affinity. The wild-type p53-DNA complex produced 11 stabilizing hydrogen bonds in between the two biomolecules.

The p53-DNA mutant complexes showed a lower number of hydrogen bonds comparing to the native p53-DNA complex and produced non-native hydrogen bonds. The lower number of hydrogen bonds meant that mutation destabilizes the p53-DNA complex and decreased the binding affinity of the two biomolecules. In the stable p53-DNA complex, the primary focus was particularly the hydrogen bonds formed between interface protein residues, DNA bases and sugar-phosphate of DNA backbone. Importantly hydrogen bonds formed or lost as a result of mutations in certain residues of the DBD were noted. The DNA-binding site for p53 contains a repeat of two consensus pentamer sequences, 5'-PuPuPuC(A/T)/3'(T/A)-GPyPyPy where Pu represents a purine (Gua/Ade) and Py represents a pyrimidine (Cyt/Thy) (El-Deiry 1998).

ZDOCK results predicted that upon docking, the DNA ligand undergoes a conformational change and allowed interaction with reverse-consensus site sequence 3'(T/A)-GPyPyPy comparing to the site 5'-PuPuPuC(A/T) sequence reported in the literature (Cho *et al.* 1994). The important major groove contacts studied here appeared in the pyrimidine-rich ^{10}CTTGCC^{15} region. In the major groove, Lys120 from L1 loop produced an electrostatic hydrogen bond (called a salt bridge) to Cyt14 phosphate. Arg280 in the amino terminal of α-helix also donated a salt bridge and a hydrogen bond to Gua13 phosphate. In this context, Arg280 is compressed to form a salt bridge with an Asp281 carboxylate but this contact was not present in the p53-DNA complex but instead, Asp280 side chain made a hydrogen bond contact with Asp281. Again, Arg280 formed

another hydrogen bond with Cys277 and a carbon-hydrogen bond with Arg283. Arg283 produced a hydrogen bond with Gua8. Further, Thy12 donated a carbon hydrogen bond to Asp281. The good proximity of the two residues resulted in good hydrogen bonding that stabilized H2 α-helix and contributed to a stable loop-sheet-helix (LSH). The Cys277 to Cyt9 hydrogen bond observed in the crystal structure was not present in this docked complex (Cho *et al.* 1994). However Cys277 formed two hydrogen bonds to Arg280 and Asp281. In the minor groove, Arg248 from L3 loop produced carbon hydrogen bond contacts with Gua13 and Thy14 bases. In the L3 loop, Ser241 formed a hydrogen bond contact with Thy12 phosphate, while Ala276 backbone carbonyl accepted a hydrogen bond from Thy12 sugar. In the LSH motif, Arg273 of H2 α-helix produced a carbon hydrogen bond contact with phosphate of Thy11 at 2.3672 Å. Apart from the phosphate contact, Arg273 salt bridge contact with Asp281 carboxylate which interact with Arg280 in the major groove was not present. But instead, Arg273 formed reinforcing electrostatic interactions with Thy11 and Thy12 phosphates backbones.

For p53 tumour suppressor protein to function, it is a necessity for p53 to bind to a correct specific DNA sequence (Wright and Lim 2007). In the docking here, Arg273 was the only residue with specific binding to Thy11 in the DNA non-consensus site. However, this residue is important for docking of p53 to DNA. Zn^{2+} ions are found in DNA binding proteins and in p53 DBD structure the Zn^{2+} ion is responsible for sequence-specific binding and aggregation (Duan and Nilsson 2006). On the other note, Zn^{2+} ion was removed before docking but the metal is important for stabilizing L3 loop and is essential for specific DNA-binding to p53. The removal of Zn^{2+} ion eradicated the specificity of Lys120, Ser241 and Ala276 which in turn interfered with the orientation of S9 and S10 β-strands leading to H2 α-helix thus the whole loop-sheet-helix motif was disturbed. This whole debacle might have led to the change in the conformation of the DNA ligand that yielded to non-specific to the consensus site which resulted in the formation of

unfavourable interactions of Arg248, Arg273, Arg280 and Arg283 residues to DNA as indicated by red dotted lines. Arg248 formed unfavourable interaction with Thy14. Arg273 guanidinium group formed unfavourable interaction with Thy11. Arg280 guanidinium group formed unfavourable interactions with Thy11 and Thy12 phosphates. Arg283 formed another unfavourable interaction with the phosphate group of Thy11. The arginine's in L3 and LSH motif participates in the hydrogen bonding, electrostatics and van der Waal contacts with the sugar, phosphates, and DNA base atoms to maintain the important structural stability of the p53-DNA complex (Cho *et al.* 1994). But in this work, their guanidinium groups also took part in unfavourable bumps as donated by red dotted lines in **Figure 5.4**.

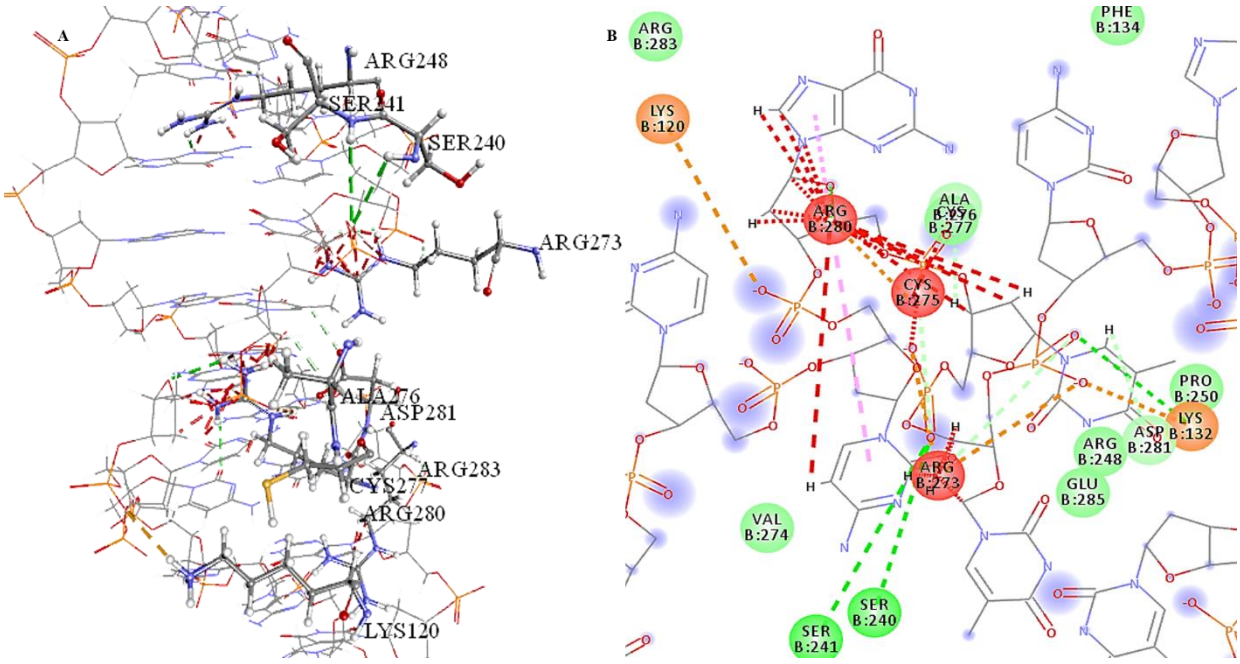


Figure 5.4 Wild-type p53-DNA binding interface. **A**) 3D interaction of DNA (wires) with residues (ball and stick). **B**) 2D binding interface residues are shown as balls coloured by type of interaction. NH-bonds shown by green dashed lines. Carbon NH-bonds denoted by cyan dashed lines. Salt bridges and electrostatic interaction orange lines. Unfavourable interactions are shown by red lines

In the Va143Asp p53-DNA mutant complex 6 stabilizing hydrogen bonds formed between the two biomolecules. Arg273 of H2 α-helix formed a hydrogen bond with Gua9 phosphate at a distance of 0.8925 Å. In the major groove, Arg280 donated a hydrogen bond to Gua7. Arg248

formed a hydrogen bond with Cyt10 in the minor groove and also with Gua13. Gua9 donated two hydrogen bonds to Asp281. The guanidinium group of Arg248 formed two electrostatic interactions with imidazole rings of Gua9 and Arg280 guanidinium group formed another two electrostatic interactions with imidazole rings of Gua8. The guanidinium group of Arg273 in the LSH formed unfavourable interactions with phosphate of Gua9 and long range unfavourable interactions with sugar oxygen of Gua8. Ala276 made an unfavourable contact with phosphate of Gua9 as denoted by red dotted lines (**Figure 5.5**).

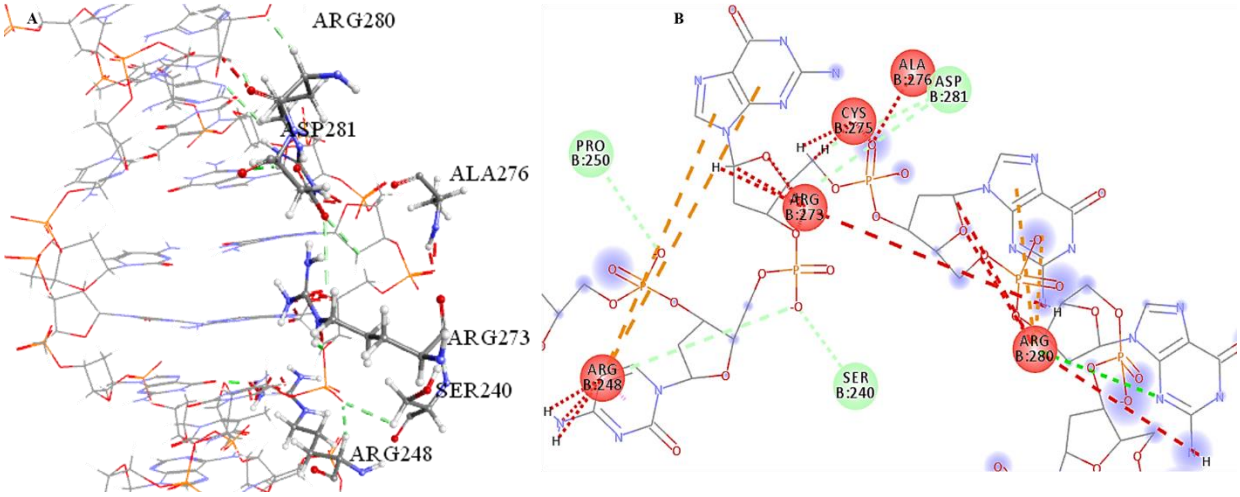


Figure 5.5 Val143Asp p53-DNA complex binding interface generated with Accelrys DS 2016 ZDOCK. **A**) 3D interaction of DNA (wires) with residues (ball and stick). **B**) 2D binding interface residues are shown as balls coloured by type of interaction. NH-bonds shown by green dashed lines. Carbon NH-bonds denoted by cyan dashed lines. Salt bridges and electrostatic interaction orange lines. Unfavourable interactions are shown by red lines

Here, L3 and other LSH residues were pushed further away from the major groove bases such that they did not make any contacts with the DNA. Important contacts of the interface residues to DNA produced unfavourable interactions as a result of DNA structure being distorted due to stacking with the reverse half site. Replacement of Val to Asp at position 143 produced masses of strain to L3 loop and LSH motif preventing other interface residues to bind to DNA.

In Ala159Pro p53-DNA mutant complex, 4 hydrogen bonds stabilized the complex. The docked structure displayed a distorted DNA binding interface as some of the contact residues were far away and did not participate in the interaction. In the major groove, Arg280 formed a hydrogen bond with Gua7 and two carbon hydrogen bonds with Ade17. The Arg248 formed two carbon hydrogen bonds with Cyt10 and another hydrogen bond with Gua13. Also, Cyt10 donated a sugar carbon hydrogen bond to Ser240. Gua9 donated two carbon hydrogen bonds to Asp281. Moreover, Asp281 shared a hydrogen bond with Arg280. Arg273 formed a hydrogen bond with Gua9 at 0.8986Å and its guanidinium group formed a salt bridge with the carboxylate of Asp281 which formed a hydrogen bond with Arg280. The guanidinium groups of Arg248 and Arg280 formed two electrostatic interactions with imidazole rings of Gua9 and Gua8 respectively. The guanidinium group of Arg273 produced many unfavourable interactions with Gua7 and Gua8. So as guanidinium group of Arg248 with Cyt10, Arg280 with Gua8 and Ala276 carbonyl with Gua9 (**Figure 5.6**).

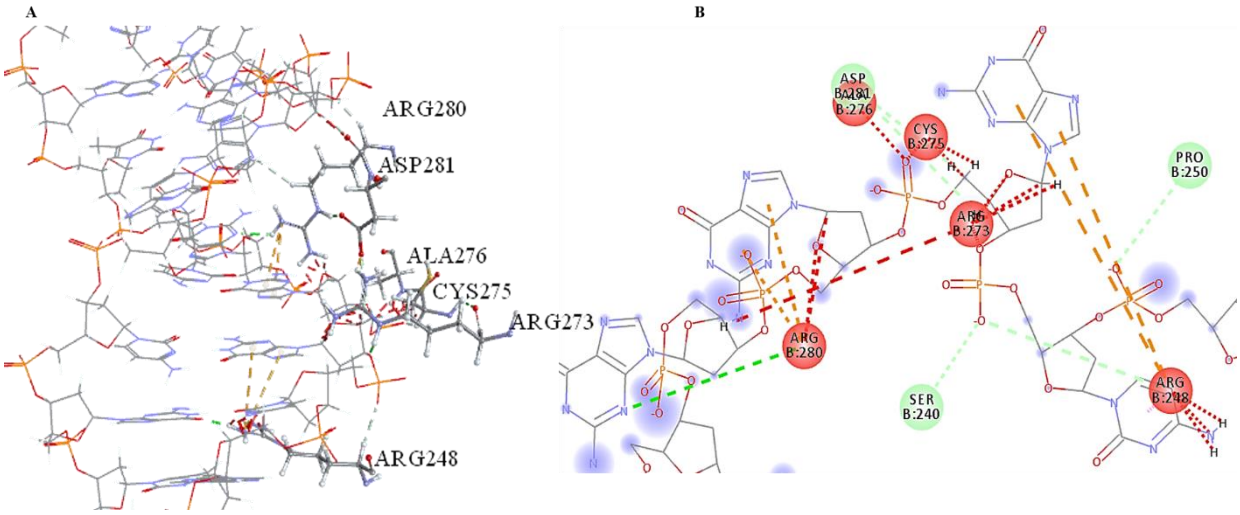


Figure 5.6 Ala159Pro p53-DNA complex binding interface **A**) 3D interaction of DNA (wires) with residues (ball and stick). **B**) 2D binding interface residues are shown as balls coloured by type of interaction. NH-bonds shown by green dashed lines. Carbon NH-bonds denoted by cyan dashed lines. Salt bridges and electrostatic interaction orange lines. Unfavourable interactions are shown by red lines

The increased in the number of unfavourable interactions is due to the close packing with symmetry-related strand that pushed further away from other interaction residues from the interface. Replacement of Ala159 to Pro at S4 which connects to L2 that indirectly interact with the binding interface by destabilizing the conformation of L3.

In Val197Pro p53-DNA mutant complex a total of 4 hydrogen bonds stabilized the two biomolecules. In the minor groove, the critical Arg248 formed a carbon hydrogen with Ade11 and an electrostatic attraction with Cty10. Two phosphate group oxygen's of Cyt10 formed carbon hydrogen bonds with Ser241. The phosphate group of Ade11 produced a hydrogen bond with Arg273 at a distance of 2.8982 Å. But the guanidinium group reinforcing electrostatic contacts were not present. Arg249 residue from the conserved region formed a carbon hydrogen with Gua13 (**Figure 5.7**).

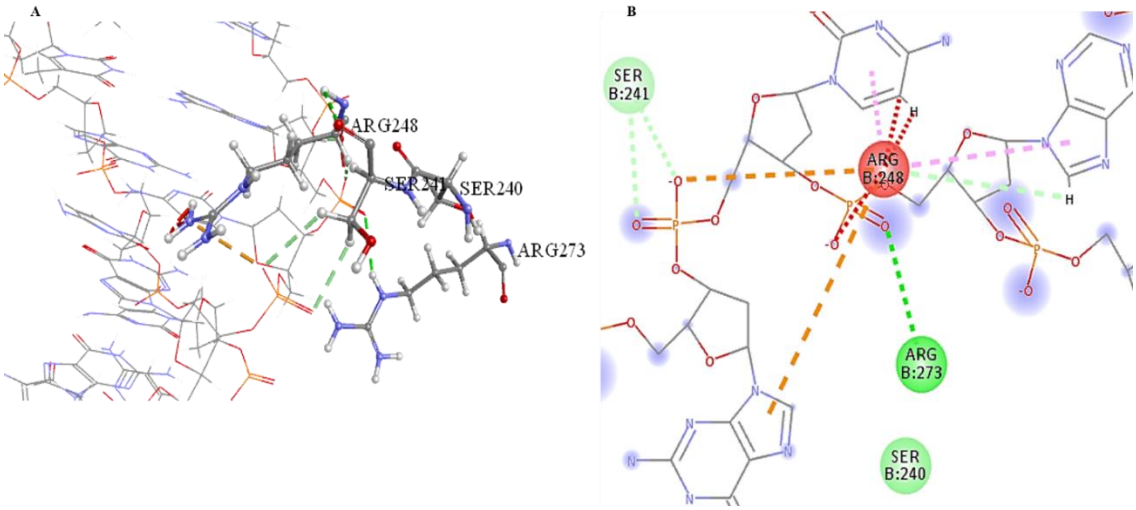


Figure 5.7 Val197Pro p53-DNA complex binding interface. **A)** 3D interaction of DNA (wires) with residues (ball and stick). **B**) 2D binding interface residues are shown as balls coloured by type of interaction. NH-bonds shown by green dashed lines. Carbon NH-bonds denoted by cyan dashed lines. Salt bridges and electrostatic interaction orange lines. Unfavourable interactions are shown by red lines

None of the interface residues made contact with the major groove of DNA. Mutation of Val to Pro at position 197 at the S5 proceeding to H1 α-helix destabilized the conformation of L1 and destroyed major groove contacts.

In the Try234Pro p53-DNA mutant complex, 8 stabilizing hydrogen bonds were established between the two biomolecules. In the minor groove, Arg248 formed a carbon hydrogen bond with the phosphate group of Cty5. The guanidinium group of Arg248 formed unfavourable interaction with Thy6. Moreover, Arg248 donated a hydrogen bond to Ser240. Adjacent to the minor groove contact (Arg248) Arg249 from L3 formed a carbon hydrogen with the phosphate group of Thy6. Arg249 from the conserved region provided a hydrogen bond to Gua13. Only one contact formed in the major groove, Arg280 from the amino terminal donated a hydrogen bond to Ade21. The DNA backbone contacts involved the Ser241 hydrogen bonds with phosphate of Cyt5 and Cyt4. A critical p53-DNA docking role player Arg273 formed two hydrogen bonds with Cyt4 and Cyt5 at magnitudes of 1.2855 Å and 2.9274 Å while, its guanidinium group was involved in the unfavourable interactions with Cyt4. Apart from the phosphate hydrogen bond, the Arg273 guanidinium group forms a salt bridge with carboxylate of Asp281 which formed a hydrogen bond with Arg280 (**Figure 5.8**).

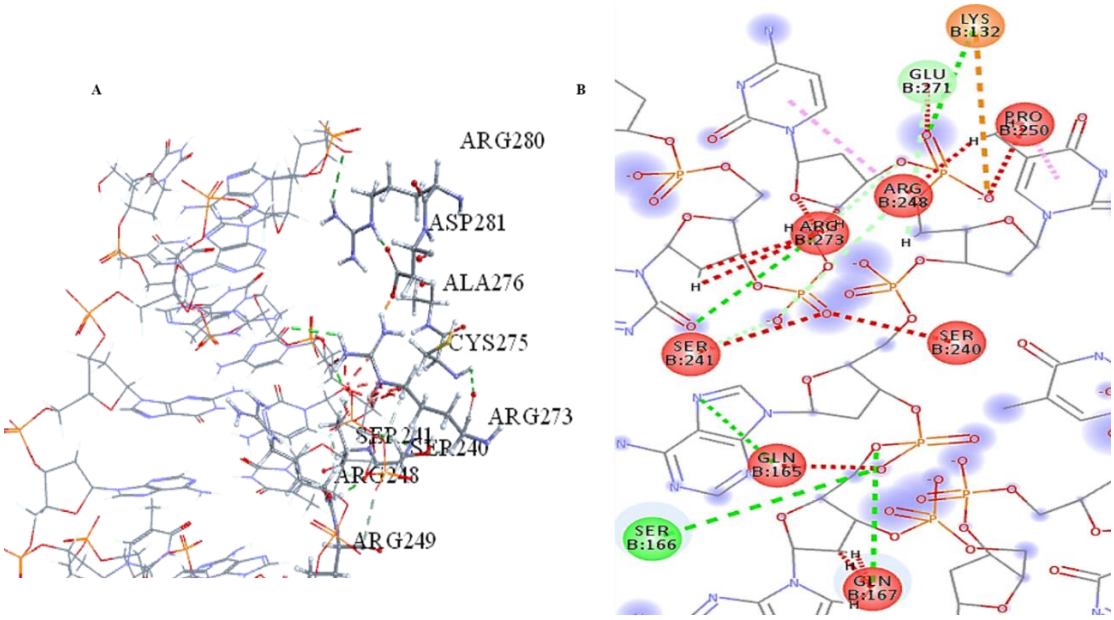


Figure 5.8 Tyr234Pro p53-DNA complex binding interface **A**) 3D interaction of DNA (wires) with residues (ball and stick). **B**) 2D binding interface residues are shown as balls coloured by type of interaction. NH-bonds shown by green dashed lines. Carbon NH-bonds denoted by cyan dashed lines. Salt bridges and electrostatic interaction orange lines. Unfavourable interactions are shown by red lines

The majority of the established hydrogen bonds were not from the interface residues. Try234Pro mutation eradicated the specific-DNA binding and further created non-native hydrogen bonds. Again, Try234Pro mutation interfered with the DBD conformation and pushed further away the interface residues to remove their interaction with the DNA molecule.

The Cys238Pro p53-DNA complex formed 9 hydrogen bonds. Major groove contacts were denoted by Arg280 forming a hydrogen bond with Gua7 and other two hydrogen bonds to Asp281 and Cys277, respectively. Ade 17 formed three long-range hydrogen bonds with Arg280. Arg273 donated a hydrogen bond with Gua9 at 0.9051 Å position and a salt bridge with Asp281. Lys120 formed a carbon hydrogen bond with Ade18. In the minor groove, Cyt10 denoted a hydrogen bond with Arg248. Again, Arg248 produced another hydrogen bond with Gua13. The guanidinium groups of Arg248 and Arg280 formed electrostatic interactions with imidazole rings of Gua8 and Gua9, respectively. Arg248 guanidinium group formed unfavourable interactions with Cyt10. Arg280 guanidinium group formed unfavourable interactions with Gua8 and Ade17. In **Figure 5.9** below the interface, residues appeared red meaning that there lots of unfavourable interactions as results of strain during the packing of atoms residues to form a stable structure.

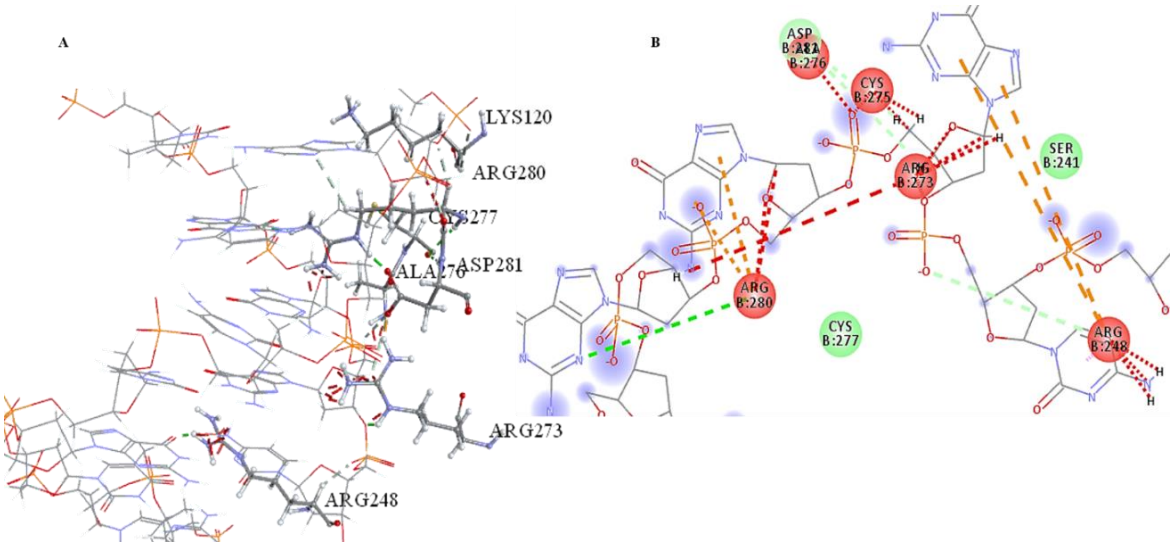


Figure 5.9 Cys238Pro p53-DNA binding interface **A**) 3D interaction of DNA (wires) with residues (ball and stick). **B**) 2D binding interface residues are shown as balls coloured by type of interaction. NH-bonds shown by green dashed lines. Carbon NH-bonds denoted by cyan dashed lines. Salt bridges and electrostatic interaction orange lines. Unfavourable interactions are shown by red lines

The Gly262Pro p53-DNA complex formed 8 intermolecular stabilizing hydrogen bonds between the two biomolecules. Major groove interactions were comprised of Lsy120 forming a salt bridge with Cyt14. Arg280 forming a hydrogen bond with Gua13 and unfavourable interactions with Thy12, Gua13, and Cyt14. Thymine 12 donated a hydrogen bond to Asp281. Arg283 formed carbon hydrogen bond with Gua8. Cys277 did not form any hydrogen bond with DNA but shared a hydrogen bond with Asp281. In the minor groove, Arg248 two short hydrogen bonds with Gua13 and Thy14 were noted. Phosphates contacts included Ser241 forming a hydrogen bond with Thy12. Arg273 formed a carbon hydrogen bond with Thy11 at a bond distance of 2.3559 Å. And the latter guanidinium group showed numerous unfavourable interactions with Thy11 and electrostatic interactions with Thy11 and Thy12. Thy12 donated a hydrogen bond to the carbonyl oxygen of Ala276. Gly262Pro mutation created many non-native residue-residue hydrogen bonds compared interacting with to with DNA (**Figure 5.10**).

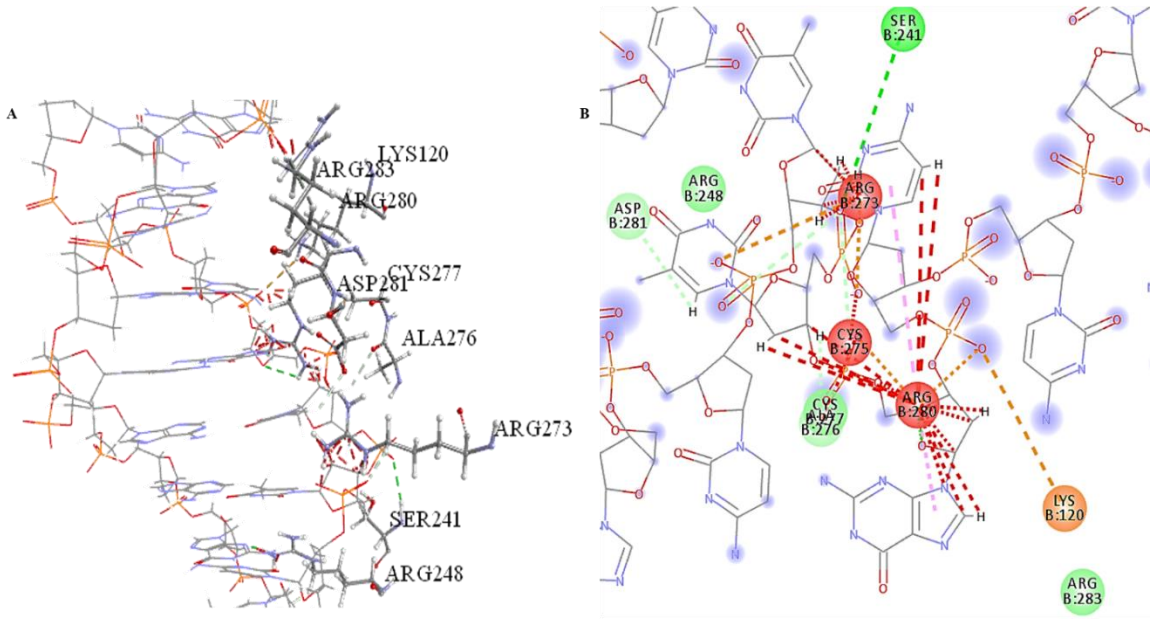


Figure 5.10 Gly262Pro p53-DNA binding interface **A**) 3D interaction of DNA (wires) with residues (ball and stick). **B**) 2D binding interface residues are shown as balls coloured by type of interaction. NH-bonds shown by green dashed lines. Carbon NH-bonds denoted by cyan dashed lines. Salt bridges and electrostatic interaction orange lines. Unfavourable interactions are shown by red lines

The Cys275Pro p53-DNA complex was stabilized by 9 intermolecular hydrogen bonds. In the major groove, Lys120 formed a salt bridge with Cyt14. Arg280 donated a hydrogen bond to Gua13. Also, Arg280 made two attractive electrostatic interactions with Gua13 and Cyt14 phosphate groups. Cys277 formed a hydrogen bond with Gua13. Thy12 donated a carbon hydrogen bond with Arg281. In the minor groove, Arg248 formed a hydrogen bond with Gua13 and another with Ser240. Ser241 formed a hydrogen bond with the phosphate group of Thy12. Arg273 formed a hydrogen bond withThy11 at 2.3498 Å while forming attractive electrostatic with Thy12. Thy12 again donated a hydrogen bond with Ala276. Pro275 formed a hydrogen bond with Thy12 (**Figure 5.11**).

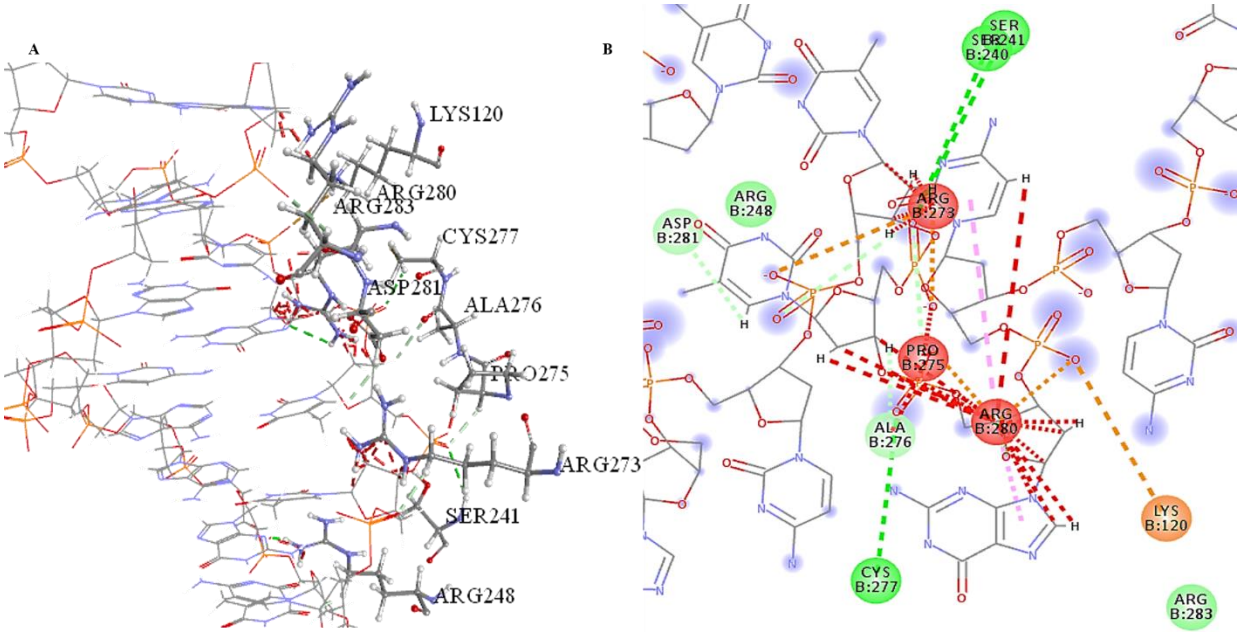


Figure 5.11 Cys275Pro p53-DNA binding interface **A**) 3D interaction of DNA (wires) with residues (ball and stick). **B**) 2D binding interface residues are shown as balls coloured by type of interaction. NH-bonds shown by green dashed lines. Carbon NH-bonds denoted by cyan dashed lines. Salt bridges and electrostatic interaction orange lines. Unfavourable interactions are shown by red lines

Cys275Pro mutation weakened the DNA binding interface with the introduction of a bulky proline side chain which prevented the specific DNA-binding. Cys275Pro mutation prevented the flexibility of the loop connecting S10 to the loop-sheet-helix and interfered with major groove contacts.

5.2.2 Comparison of ZDOCK computational results with experimental mutations

Similar work was tested on experimental mutations. The Val143Ala, Tyr163Asn, Arg213Gly, Tyr220Ser, Gly245Ser and Gly245Val p53-DNA mutant complexes ZRANK scores were as the following -67.15, -68.88, -63.20, -71.53, -70.10 and -70.39 as presented in **Table 5.7**.

Table 5.7 Molecular docking results obtained by the ZRANK scores and hydrogen bonding (NH-bonds) ordered by the distance between p53 wild-type, experimental mutants, and DNA complex

Type of complex	ZRANK Score[a]	Amino Acid[b]	Bond distance (Å)[c]	Bonding DNA part[d]	Bonding amino acid[e]	NH-bonds
Val143Ala-DNA	-67.15	Ser241	3.0983	(PO_4^{3-})Phosphate	(C=O)Backbone	8
			1.9326	(H5)Sugar	(C=O)Backbone	
		Arg248	2.7676	(NH1)Base	(NH_2)Side chain	
		Arg273	3.0499	(O3)Sugar	(NH_2)Side chain	
		Ala276	2.7488	(O5)Sugar	(HN_3^+) Backbone	
		Arg280	3.1014	(C=O)Base	NH_2)Side chain	
			2.8831	(O4)Sugar	(CH_2)Side chain	
			2.4830	(C=O)Base	(NH_2)Side chain	
Try163Asn-DNA	-68.88	Ser241	2.8972	(PO_4^{3-})Phosphate	(C=O)Backbone	10
			1.9928	(H5,1)Sugar	(C=O)Backbone	
			2.3589	(H5'2)Sugar	(C=O)Backbone	
		Arg273	2.9559	(O3)Sugar	(NH_2)Side chain	
			1.3898	(O4)Sugar	(NH_2)Side chain	
		Ala276	3.0447	(O5)Sugar	(HN_3^+) Backbone	
		Arg280	2.6128	(C=O)Base	(NH_2)Side chain	
			2.9897	(O4)Sugar	CH_2)Side chain	
			2.767	(C=O)Base	CH_2)Side chain	
			2.4512	(H4)Sugar	(C=O) Backbone	
Arg213Gly-DNA	-63.20	Ser241	1.3748	(PO_4^{3-})Phosphate	(HN_3^+) Backbone	8
			3.0159	(O4)Sugar	(CH_2)Side chain	
		Arg248	2.7783	(N3)Base	(NH_2)Side chain	
			2.6405	(C=O)Base	(CH_2)Side chain	
			2.9975	(N3)Base	(CH_2)Side chain	
		Arg273	2.6433	(PO_4^{3-})Phosphate	(NH_2)Side chain	
		Arg280	2.9687	(O4)Sugar	(NH)Side chain	
			2.2179	(N7)Base	(CH_2)Side chain	
Tyr220Ser-DNA	-71.53	Lys120	3.7982	(PO_4^{3-})Phosphate	(NH_2)Side chain	4
		Arg273	3.4358	(PO_4^{3-})Phosphate	(CH_2)Side chain	

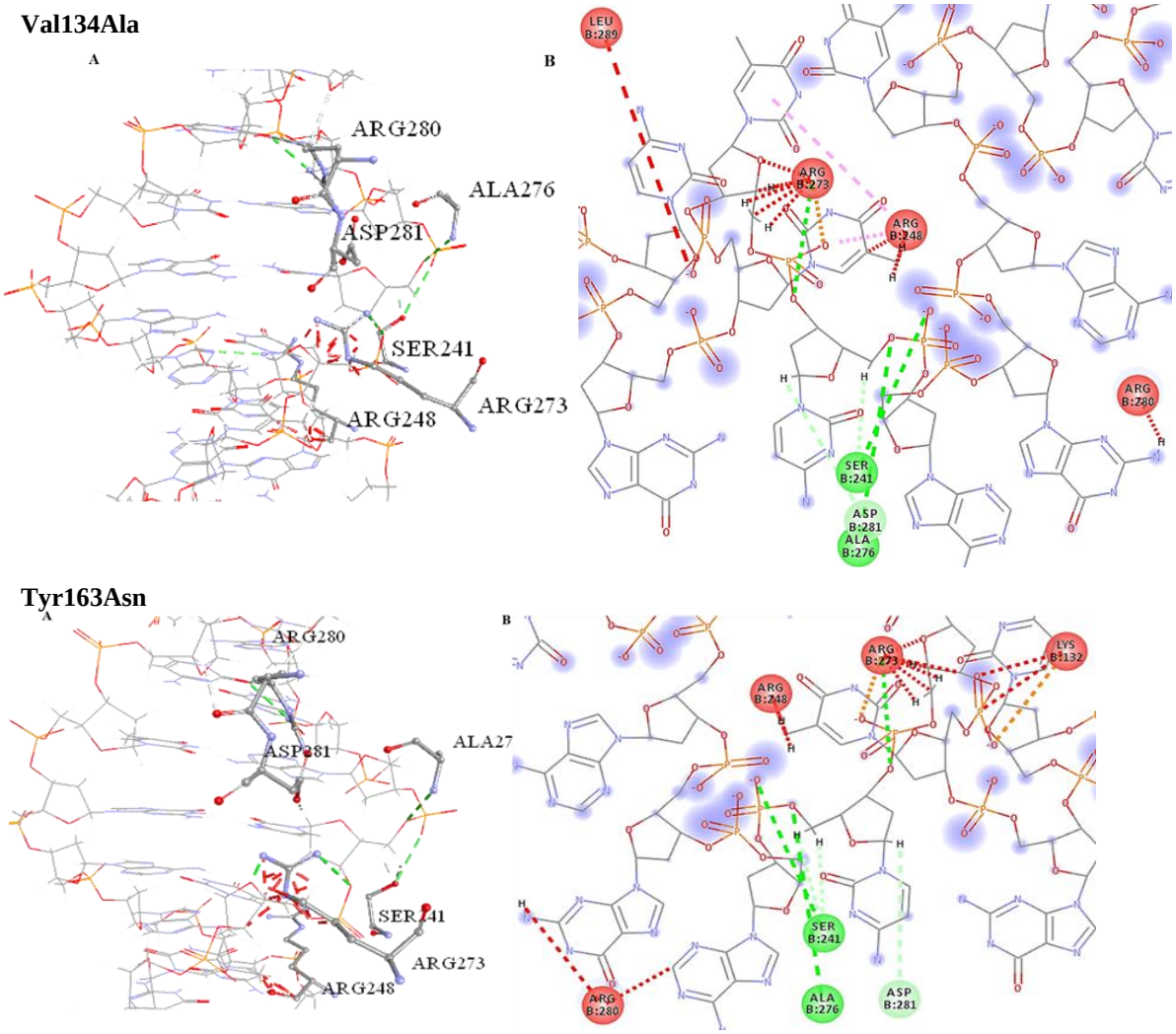
		Ala276	2.5795	(H3)Sugar	(C=O) Backbone	
		Asp281	2.5762	(H6)Base	C=O) Side chain	
Gly245Ser-DNA	-70.10	Arg248	3.0947	(H6)Base	(C=O) Backbone	4
		Arg249	3.5854	(PO_4^{3-})Phosphate	(C) Backbone	
			3.8008	(imidazole)Base	(NH_2)Side chain	
			3.1245	(N7) Base	(NH_2)Side chain	
Gly245Val-DNA	-70.39	Ser241	3.0365	(O5)Sugar	(OH)Side Chain	14
			2.7377	(PO_4^{3-})Phosphate	(H) Backbone	
			3.0408	(PO_4^{3-})Phosphate	(CH_2)Side chain	
		Arg248	2.5405	(C=O)Base	(NH_2)Side chain	
			1.6539	C=O)Base	(NH_2)Side chain	
			2.6710	(PO_4^{3-})Phosphate	(H) Backbone	
			2.2625	(C=O)Base	(CH_2)Side chain	
		Arg273	2.5167	(PO_4^{3-})Phosphate	(H) Backbone	
		Ala276	1.4188	(PO_4^{3-})Phosphate	(HN_3^{+}) Backbone	
		Arg280	2.3790	(C=O)Base	(NH_2)Side chain	
			1.1273	(C=O)Base	(NH_2)Side chain	
			2.6520	(O3)Sugar	(H) Backbone	
			2.6291	(H5)Sugar	(C=O) Backbone	
			2.0785	(H6)Sugar	(C=O) Backbone	

[a]ZRANK score is a measure of the binding affinity between the mutant and DNA.
[b]Amino acids are the interface residues that are involved in forming DNA contacts.
[c]Bond distance measures the length of the hydrogen bond formed between interface residues and DNA backbone or bases.
[d]Ligand part is representing the part of the ligand responsible for forming a hydrogen bond, it can be a sugar, phosphates or DNA bases but especially here there is also bases rings.
[e]Bonding amino acid is the interface residue part that is forming a hydrogen bond, it can be a side chain or backbone.

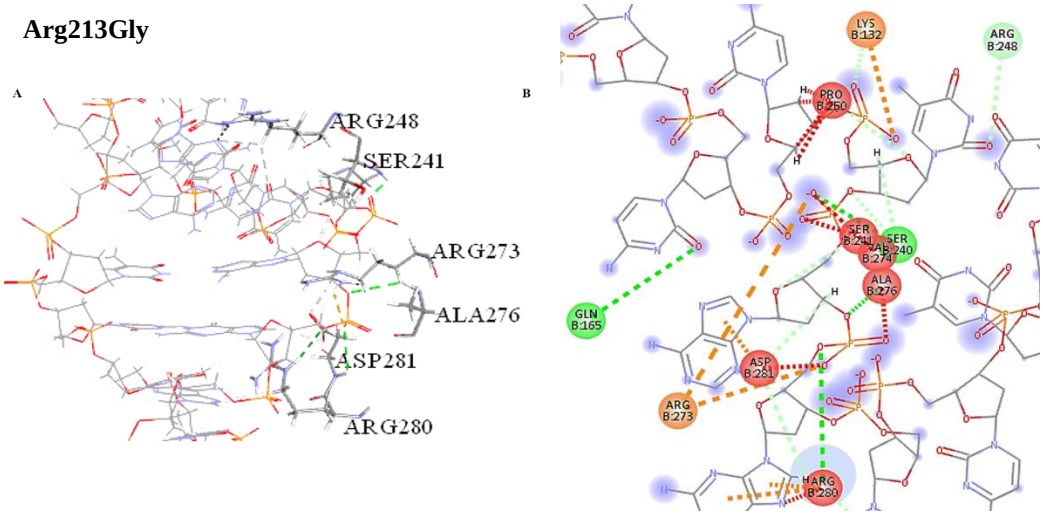
The Val143Ala, Tyr163Asn, Arg213Gly mutations decreased the binding affinity of the p53-DNA complex with considerable amounts. Wherein, Tyr220Ser, Gly245Ser, and Gly245Val mutations insignificantly increase the binding affinity compared to p53 wild-type. These ZRANK scores are in agreement with the computational mutations in the effect of p53-DNA complex.

In a similar way to the computational p53-DNA mutant complexes, experimental p53-DNA mutant complexes produced a lower number of hydrogen bonds as compared to the wild-type complex, except for Gly245Val p53-DNA complex with a higher number of hydrogen bonds. The increased number of hydrogen bonds in the Gly245Val p53-DNA mutant complex is in line with the ZRANK binding affinity score of -70.39. This meant that this mutant increased the stability of the complex. Val143Ala, Tyr163Asn, Arg213Gly, Tyr220Ser, Gly245Ser and Gly245Val p53-DNA mutant complexes generated 8, 10, 8, 4, 4 and 14 intermolecular hydrogen bonds between the two biomolecules. More interestingly, was the hydrogen bond formed by the

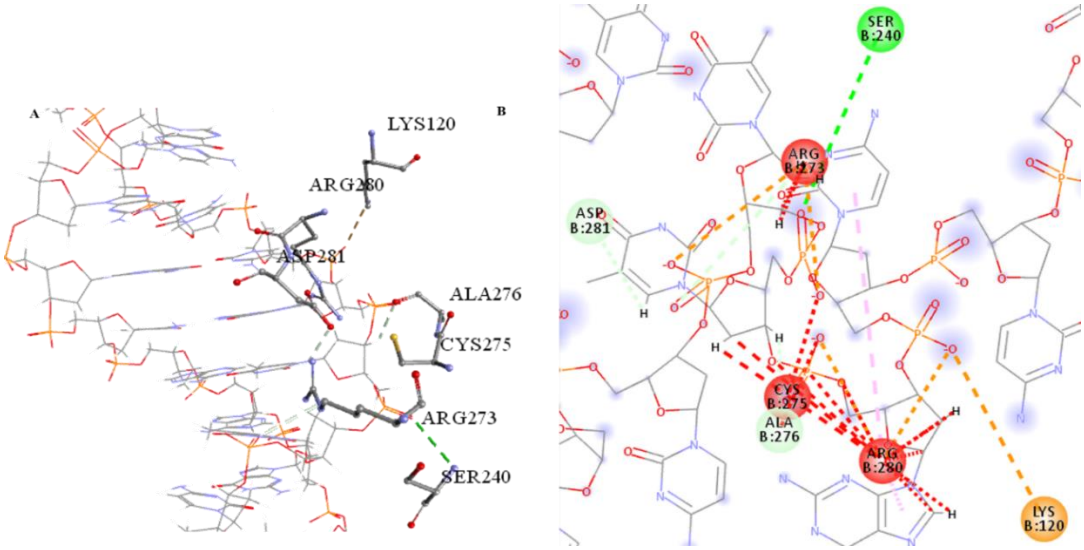
Arg273 residue with DNA. It was projected that Val143Ala p53-DNA mutant complex Arg273 produced a hydrogen bond to Cyt10 and an electrostatic interaction with Thy11. In the Tyr163Asn p53-DNA mutant complex, Arg273 donated two hydrogen bonds to Cyt10 and Thy11. Arg213Gly p53-DNA donated only a salt-bridge to the phosphate of Gua8. Tyr220Ser p53-DNA mutant complex Arg273 provided a hydrogen bond to Thy11. Gly245Ser p53-DNA Arg273 residue did not form a hydrogen bond interaction with DNA. Gly245Val p53-DNA mutant complex donated a hydrogen bond to Ade17 phosphate. It can be seen from the above argument that selected experimental p53-DNA mutant complexes either altered or abolished the critical Arg273 residue hydrogen bond interaction with DNA bases. Experimental mutations docking results are presented in **Figure 5.12**.



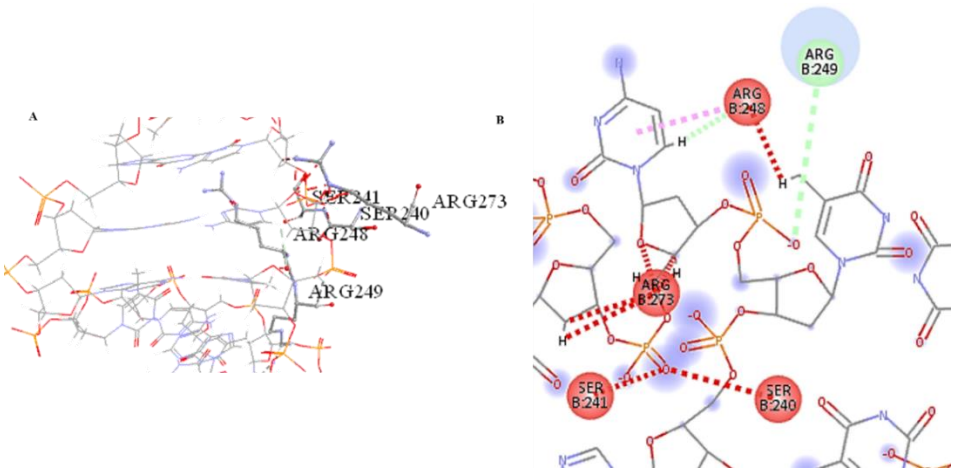
Arg213Gly



Tyr220Ser



Gly245Ser



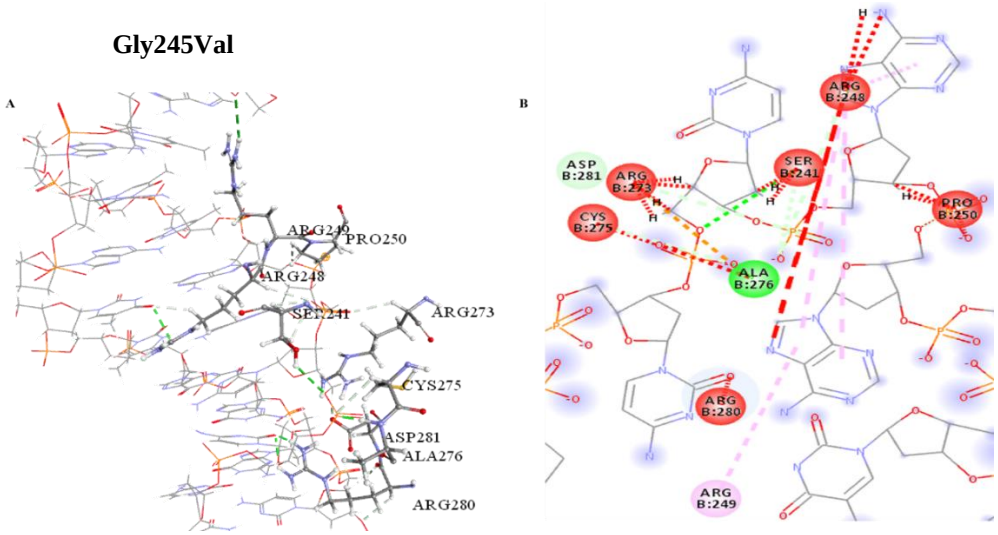


Figure 5.12 P53-DNA binding interface complexes. **I)** Val143Ala p53-DNA **II**) Tyr163Asn p53-DNA **III**) Arg213Gly p53-DNA **IV**) Tyr220Ser p53-DNA **V**) Gly245Ser p53-DNA and **VI**) Gly245Val p53-DNA **A**) 3D interaction of DNA (wires) with residues (ball and stick). **B**) 2D binding interface residues are shown as balls coloured by type of interaction. NH-bonds shown by green dashed lines. Carbon NH-bonds denoted by cyan dashed lines. Salt bridges and electrostatic interaction orange lines. Unfavourable interactions are shown by red lines

The Val143Ala p53-DNA mutant complex binding affinity was lowered to a ZRANK score of -67.15 and decreased hydrogen bonds interaction to 8 because the Val143Ala mutation resulted in the long range effects from the hydrophobic core that impaired the specificity of DNA binding. The altered binding affinity of the p53-DNA complex of Gly245Ser stems from the structural rearrangement of the L3 and L2 loops, where Gly245 is located at L3. It was reported that the secondary structure of L3 and L2 loops is deficient and is stabilized by the zinc metal coordination and their side chain interactions. From the latter, it can be seen that mutation to serine abolishes these interactions and destabilize the conformation of L2 and L3 (Freed-Pastor and Prives 2012). While L3 provides the important minor groove interactions of DNA which destroyed the hydrogen bonds interactions between the two biomolecules as a result of serine mutation.

5.2.3 Summary

Docking results inferred that p53-DNA mutant complexes produced wild-type like binding to DNA with far less stable complexes as shown by the drop in the binding affinity characterized by the ZRANK scores. This displayed the lack of sequence-specific DNA-binding because the p53 mutants failed to bind to the 5'-GGGCAAGTCTA-3'site but were seen to bind in the 3'-CTTGCC-5' site of the duplex DNA (Wright and Lim 2007) which may consequently block the transcription of genes and prevent the tumour suppression function.

An analysis of interface contacts indicated that there is a correlation between tumour associated mutations, DNA-binding and folding stability in the p53 DBD. In the following, a correlation between p53-DNA interface with the inactivation mechanisms of Val143Asp, Ala159Pro, Val197Pro, Tyr234Pro, Cys238Pro and Cys275Pro tumour associated mutations in the DNA-binding domain is discussed. It is apparent that the β-sandwich region Val143Asp, Ala159Pro and Val197Pro mutations of the hydrophobic core maybe be responsible for the DBD and the whole protein stability. It has been established in this work that tumour mutation targeting Val143, Ala159 and Val197 residues destabilize the DBD domain. Val143 is a hydrophobic core residue in the conserved region II far away from the DNA binding site. However, Val143Pro mutation indirectly interfered with p53-DNA interaction by decreasing the binding affinity and altered Arg273 specificity by attaining a different binding orientation comparing to the wild-type. This may be caused by the fact that Val143Asp mutation at S3 proceeds to the loop-sheet-helix. Therefore, the introduction of a polar Asp exposed S3 to the protein surface and changed the orientation position of the loop-sheet-helix motif which affected DNA binding. Adversely, Val143 form hydrophobic contacts with Tyr234 to stabilize the core domain. So mutation of Val143 also interferes with Tyr234 hydrophobic contact which will negatively affect the stability of DBD.

Conversely, Ala159 is situated in L2 far away from the binding interface but L2 directly interact with L3 which provides minor groove contacts and DNA binding site. So mutation of Ala to Pro at position 159 indirectly affected the DNA binding interaction by lowering the binding affinity as denoted by ZRANK score of -68.43 and deserted DNA binding of the critical Arg273. Similarly, a Val197Pro mutation in the hydrophobic core produced similar effects as Val143Pro and Ala159Pro. Another noted effect was a mutation of these hydrophobic core residues did not only weakened p53-DNA binding affinity but destroyed stabilizing hydrogen bond contacts of the complex. This proved that the hydrophobic core is important for the whole domain stability and hence the whole protein. Try234 is located at S8 which connects to L3. Mutation at this position altered the interaction of L3 with minor groove and eradicated DNA binding specificity. The cys238 residue is part of the Zn^{2+} ligands which connects L2 and maintain proper position of L3. Mutation of Cys238 to Pro changed the proper co-ordination of Zn^{2+} ligands which in turn changed the orientation of the L2 and L3 loops and interfere with specific DNA binding. In order to acquire a correct atom typing for protein minimization, the Zn^{2+} metal ion was removed before docking. The removal of Zn^{2+} ion before docking and created bad atomic clashes which changed H2 α-helix and produced side chain residues steric hindrance around Arg273 in S10. Hence the interaction of Arg273 with DNA Thy11 was not seen in most mutant complexes. Because binding of Zn^{2+} ion is compulsory for stabilization of the DBD and acknowledgment of DNA during binding (Rainwater *et al.* 1995). The gly262 residue is connected to a turn that joins S10 with S2' of the loop-sheet-helix. Gly262Pro mutation lowered the binding affinity but did not change the specific binding of the critical Arg273 to Thy11. The cys275 residue is part of loop-sheet-helix which interacts with the DNA major groove. Mutation at this site hindered the interaction of the loop-sheet-helix with DNA. The majority of p53 residues interaction to DNA produced numerous unfavourable interactions caused by unfavourable bumps. This may be caused by the removal of the Zn^{2+} ion before docking which gives a proper orientation of L3 loop and the LSH

motif. As it was shown that the removal of Zn^{2+}ion changed the site-specific DNA-binding while, non-specific DNA binding is retained (Butler and Loh 2003). Again these interactions may result into the created bad atomic clashes because of the introduction of proline in positions that does not allow its backbone conformation and thus interfered with secondary structure. The mutations of the hydrophobic core residues Val143, Ala159 and Val197 to Asp or Pro respectively, induced a sequential disruption of the hydrogen bond network that ultimately led to the loss of the DNA-binding specificity. It can be confirmed that mutation of the hydrophobic core residues reduced the stability of the DBD and altered DNA binding. Tyr234Pro, Cys238Pro, Gly262Pro and Cys275Pro mutants produced significantly fewer hydrogen bonds disruptions comparing to the hydrophobic core mutants.

Furthermore, the binding specificity of Arg273 was retained for Gly262Pro and Cys275Pro, where all the interface residues were observed. The change in binding affinity and elimination of DNA binding specificity of the experimental mutant complex Val134Ala p53-DNA can only be understood by long-range effects of the Val143Ala mutation. It was clearly visible that Gly245Ser mutation tremendously destroyed the hydrogen bond interaction in the Gly245Ser p53-DNA complex, as a result of structural changes in L2 and L3 loops due to mutation. The findings in this work from the computational and experimental mutations regarding the interaction of the p53 mutant forms with DNA are in agreement with each other. So we can precisely say that the effects of mutations on the p53 DBD structure and its interaction are similar regardless of whether the mutations were predicted theoretically or experimentally.

5.3 MD Simulation trajectory analysis

In this study, a comparative molecular dynamics (MD) simulation analysis was performed between the p53 wild-type and experimental mutants. The aim of MD was to examine the structural and functional behaviour of experimental mutations, Val143Ala, Arg213Gly and Gly245Ser comparing to p53 wild-type DNA-binding domain. During the dynamics, structural variations measured by the RMSD, Rg, SASA and total energy of the system parameters as a function of time were studied and described below.

5.3.1 Root mean square deviation (RMSD)

The MD simulation Root Mean Square Deviation (RMSD) for all the backbone atoms from the initial coordinates as the central of origin were analysed to compute the convergence of the protein system (Kuzmanic and Zagrovic 2010) and depicted in **Figure 5.13**.

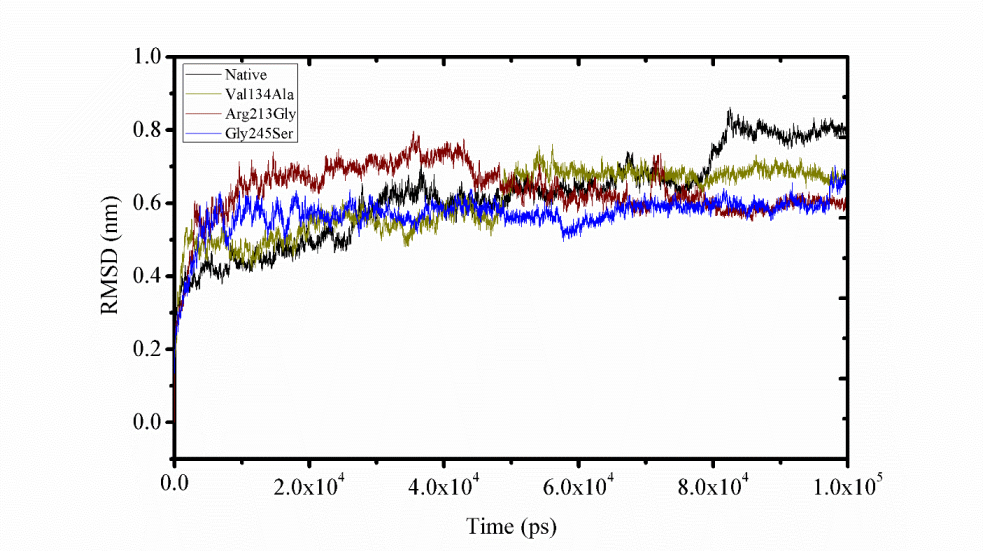


Figure 5.13 RMSD of wild-type, Val143Ala, Arg213Gly and Gly245Ser mutant of p53 proteins DBD over time at 300K presented in black, dark yellow, wine and blue colours.

In the Cα-RMSD plot, the native and Arg213Gly mutant showed a similar way of deviation from the beginning up to ~80000ps after which the mutant showed a decrease in the RMSD comparing

to the native structure till the end of the simulation. This mutant showed a higher deviation from wild-type up to ~50000ps while Va134Ala and Gly245Ser showed the high deviation up to ~28000ps. Val134Ala mutant has a similar way of deviation as the native structure up to ~38000ps, but then it decreased and showed a significant rise at ~45000ps and become similar to Gly245Ser at ~98000ps. It can be observed from the RMSD plots of experimental p53 mutant forms that an unstable profile was produced from ~3000ps as there were visible fluctuations. But starting from ~80000ps the systems stabilized for the whole duration of the simulation. The RMSD values of p53 Val143Ala, Arg213Gly, and Gly245Ser mutants are presented in Appendix II. This table indicated that mutations loses the stability of p53 DBD and affect its stable conformation. The Gly245Ser showed a rather constant nature comparing to the wild-type but there were initial fluctuations at the beginning of the simulation.

5.3.2 Radius of gyration (Rg)

The Radius of gyration (Rg) parameter provides the level of compactness in the protein structure (Lobanov, Bogatyreva and Galzitskaya 2008), as presented in **Figure 5.14**.

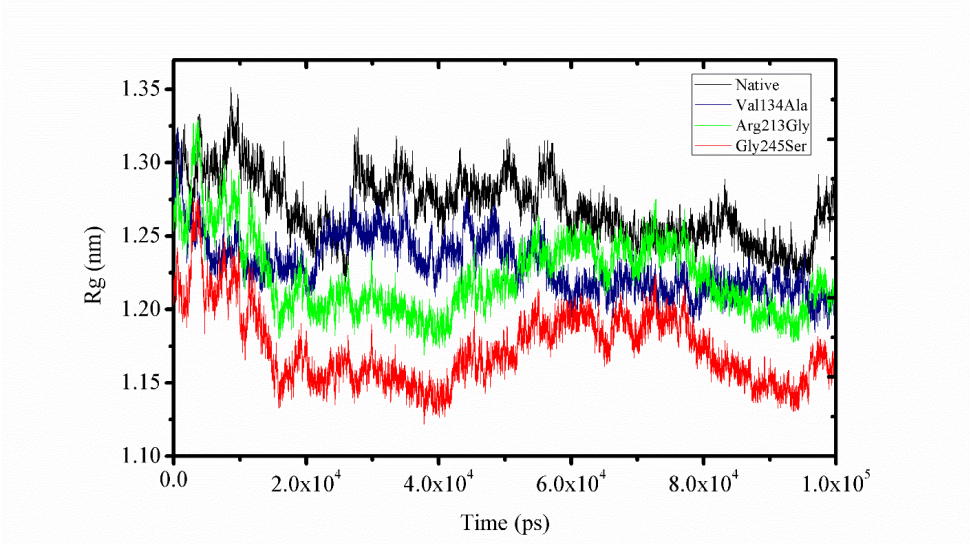


Figure 5.14 Radius of gyration (Rg) of c-alpha atoms of wild-type, Val143Ala, Arg213Gly and Gly245Ser mutant of p53 proteins DBD versus time at 300K displayed in black, blue, green and red colours.

In the Cα-Rg plot, the p53 wild-type produced the highest Rg value from the start till the end of the simulation. The Gly245Ser mutant indicated a convincing lowest value of the Rg plot from beginning to the end. Significant fluctuation similarities in the Rg values of Arg213Gly and Gly245Ser mutants from appeared from the start till the end of the simulation. The Rg values of the p53 wild-type, Val143Ala, Arg213Gly and Gly245Ser mutants are displayed in Appendix II.

5.3.2 Solvent Accessibility Surface Area (SASA)

As described in section 5.15, the Solvent Accessibility Surface Area (SASA) computed and presented in **Figure 5.15** illustrated the change of (SASA) for wild-type, Val143Ala, Arg213Gly and Gly245Ser mutant p53 proteins with time. SASA is the surface area of the biomolecules that is accessible to the solvent.

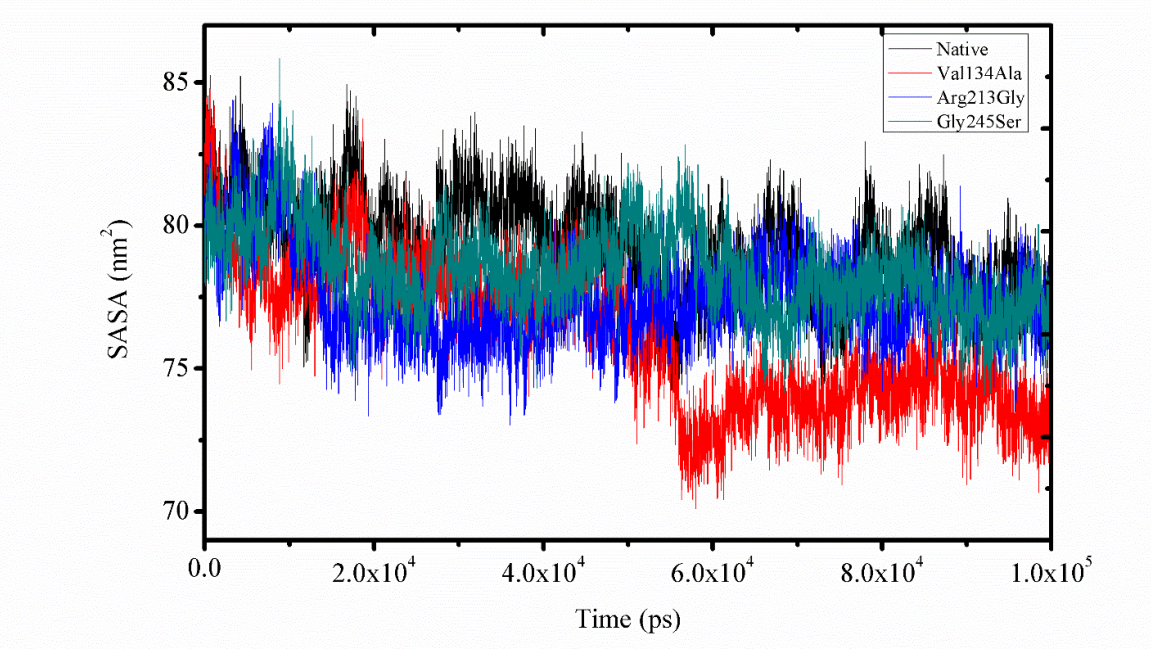


Figure 5.15 Solvent accessibility surface area of wild-type, Val143Ala, Arg213Gly and Gly245Ser mutant of p53 protein DBD over time at 300K, displayed in black, red, blue, dark cyan colours.

The SASA values of wild-type, Val143Ala, Arg213Gly and Gly245Ser mutants are again presented in Appendix II. The lower fluctuations in Rg plots of mutant structures indicated that

the protein structure might be undergoing a significant structural transition. This was further confirmed by the SASA results, where Val143Ala, Arg213Gly, and Gly245Ser mutants revealed lower SASA values comparing to the p53 wild-type protein. The lower values of SASA denoted that mutant structures have a compressed weighted-mass nature as compared to the p53 wild-type protein structure.

5.3.4 Total energy of the system

The overall energy of the wild-type and mutant proteins did not show any significant difference in magnitude as depicted in **Figure 5.16** below.

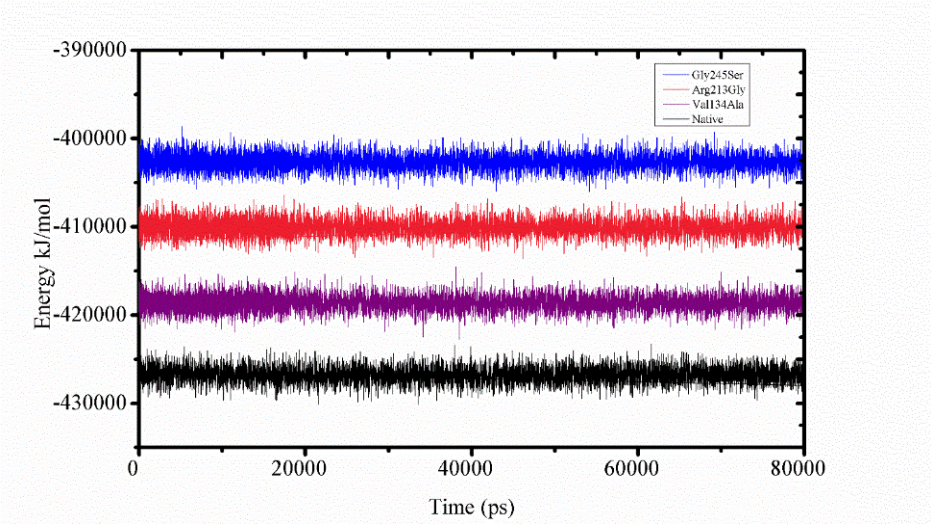


Figure 5.16 Relative stability energies of the systems over time during the simulation at 300K. The wild-type, Val143Ala, Arg213Gly and Gly245Ser mutant of p53 proteins DBD is shown in black, purple, red and royal blue colours.

The relative total energy of wild-type, Val143Ala, Arg213Gly and Gly245Ser mutants were estimated to be -426748.20 kJ/mol, -426587.63 kJ/mol, -428964.36 kJ/mol and -429175.14 kJ/mol, respectively. This finding indicated a loss of the stability by mutant protein structures. Moreover, the total energy of the whole systems produced a relatively stable nature of the systems from the start to the end of the simulations.

5.3.5 Summary

The MD simulation correlation of structural and dynamical behaviour at an atomic level of p53 DNA-binding domain (DBD) and experimental mutant forms showed that the stability of the protein structure is decreasing, due to mutations leading to a more rigid structure. The results also showed that mutations in p53 DBD decrease the stability and flexibility of the protein structure. All the mutations under study are situated in the DBD hydrophobic core. The conformational changes due to the instability of p53 DBD showed that mutations in the hydrophobic core destroyed the existing hydrophobic contact interactions with its neighbours. The increase in magnitude shown by an Arg213Gly mutant in critical positions from the Cα-RMSD results revealed that the mutant is destabilized. In this work, the zinc metal ion was removed before the simulations because it is not easy to simulate a protein structure containing metals ions as proper force field are needed for simulation but removal of zinc ion contributed into poor models as this ion is contributing to the stability of this domain. In molecular dynamics simulation of p53 DBD the removal of Zn^{2+} ion produced a high degree of fluctuations in the adjacent L2 and L3 loops (Duan and Nilsson 2006). In that note, the visible limitations shown by the various fluctuations in all simulations, indicating that the zinc metal is essential for maintaining the structural integrity of the DBD domain. The results of all mutant protein structures demonstrated a decrease in the Rg and SASA values, which showed that there was a significant structural transition due to mutations in the p53 DBD. Our analysis clearly showed a lack of flexibility by the mutant structures which have become more rigid in nature. However, the total energy of each protein structure showed a relatively stable system for the duration of the whole simulations, but the increased total energy by some of the mutant structures revealed a loss in their stability.

CHAPTER 6

CONCLUDING REMARKS

It should be pointed out that the computational analysis in this work was aimed at predicting the direction (stabilizing or destabilizing) together with the absolute value of energy changes, but more importantly to reveal the details of suggested changes. The importance of acquiring both sequence and structure-based approaches provided details causing the malfunctioning of the p53 protein. In addition, this work produced testable predictions about the effect of mutations in p53 protein DNA-binding domain (DBD) stability.

The failure of p53 mutants to bind to DNA in a specific sequence mode may be attributed to structural defects of the hydrophobic architecture that was disrupted by the replacement of hydrophobic residues due to mutations. Mutations in this region resulted in the change of stable conformations of the DBD. The observed reduction of DNA binding characterized by a decrease in the binding affinity quantified by the ZRANK scores was a key indicator illustrating the p53 DBD structural changes. The weakening of the binding affinity was due to the failure of p53 mutants to act as sequence-specific transcription factors through binding to DNA. The evidence of these penalties was the lack of the vital interface residues to form contacts with the target DNA which was demonstrated by the loss of native intermolecular hydrogen bonds and the introduction of non-native hydrogen bond contacts. The p53-DNA docking results stipulated that core domain mutations does not completely destroy the p53-DNA binding but reduce the binding affinity and alters binding specificity. The use of experimental mutations in MD simulations further confirmed their defective nature by producing structural variations in the stability lost by in the DBD and conformational changes through the application of RMSD, Rg, SASA and total energy parameters. Many fluctuations in the RMSD showed that there was a lack of crystal packing in

the mutant structures. This may be correlated to mutation inactivation of the DBD in the hydrophobic core as the selected mutations are responsible for forming hydrophobic contacts. The results showed that the long simulation time provided a clear indication of conformational changes occurring in the experimental p53 mutant forms.

Preliminary computational studies in identification and analyses of cancer associated mutations in p53 protein revealed that mutations in the DBD result in loss of the tumour suppressive function as p53 failed to bind to DNA. Collected evidence indicated that p53 DBD inactivation may be due to loss of stability that resulted in the loss of binding-DNA at the interface. Further analysis of p53 mutations provided comprehensive stabilizing mechanisms at an atomic level.

Future work

The promising future work will be to apply molecular dynamics simulation approaches of the previously docked structures from the predicted mutations using long-range simulations to further investigate the conformational changes at longer timescales. The identified tumorigenic mutations that produced a highly destabilizing effect on the structure and function of p53 may be applied in an attempt to design rescuing mutants that can be used for anti-cancer drugs. The presented biophysical data provided a more detailed structural to functional relationship as a result of mutations in p53. A diverse nature of p53 mutations studies to serve as an assisting model for predicting the effects of mutations in other related proteins.

1

2

REFERENCES

Accordino, S. R., Rodriguez Fris, J. A. and Appignanesi, G. A. 2013. Wrapping effects within a proposed function-rescue strategy for the Y220C oncogenic mutation of protein p53. *PLoS One*, 8 (1): e55123.

Adcock, S. A. and McCammon, J. A. 2006. Molecular dynamics: survey of methods for simulating the activity of proteins. *Chemical Reviews*, 106 (5): 1589–1615.

Ang, H. C., Joerger, A. C., Mayer, S. and R, F. A. 2006. Effects of common cancer mutations on stability and DNA binding of full-length p53 compared with isolated core domains. *The Journal of Biological Chemistry*, 28 (31): 21934–21941.

Ashworth, J., Bernard, B., Reynolds, S., Plaisier, C. L., Shmulevich, I. and Baliga, N. S. 2014. Structure-based predictions broadly link transcription factor mutations to gene expression changes in cancers. *Nucleic Acids Res*, 42 (21): 12973-12983.

Bader, J. S. and Chandler, D. 1992. Computer simulation study of the mean forces between ferrous and ferric ions in water. *The Journal of Physical Chemistry*, 96 (2): 6423-6427.

Bai, L. and Zhu, W.-G. 2006. p53: Structure, function and therapeutic applications. *Journal of Cancer Molecules*, 2 (4): 141-153.

Barakat, K., Issack, B. B., Stepanova, M. M. and Tuszynski, J. 2011. Effects of temperature on the p53-DNA binding interactions and their dynamical behavior: comparing the wild type to the R248Q mutant. *PLoS One*, 6 (11): e27651.

Baroni, T. E., Wang, T., Qian, H., Dearth, L. R., Truong, L. N., Zeng, J., Denes, A. E., Chen, S. W. and Brachmann, R. K. 2004. A global suppressor motif for p53 cancer mutants. *PNAS*, 101 (14): 4930-4935.

Bartkowiak, K., Riethdorf, S. and Pantel, K. 2012. The interrelating dynamics of hypoxic tumor microenvironments and cancer cell phenotypes in cancer metastasis. *Cancer Microenvironments*, 5: 59–72.

Basse, N., Kaar, J. L., Settanni, G., Joerger, A. C., Rutherford, T. J. and Fersht, A. R. 2010. Toward the rational design of p53-stabilizing drugs: probing the surface of the oncogenic Y220C mutant. *Chem Biol*, 17 (1): 46-56.

Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. and Haak, J. R. 1984. Molecular dynamics with coupling to an external bath. *Journal of Chemical Physics*
81 (8): 3684-3690.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, J., Bhat, T. N., Weissig, N., Shindyalov, I. N. and Bourne, P. E. 2000. Protein data bank. *Nucleic Acids Res*, 28 (1): 235-242.

Boeckler, F. M., Joerger, A. C., Jaggi, G., Rutherford, T. J., Veprintsev, D. B. and Fersht, A. R. 2008. Targeted rescue of a destabilized mutant of p53 by an in silico screened drug. *PNAS*, 105 (30): 10360-10365.

Bouafia, A., Corre, S., Gilot, D., Mouchet, N., Prince, S. and Galibert, M.-D. 2013. p53 requires the stress sensor USF1 to direct appropriate cell fate decision. *Plos Genetics*, 10 (5): e1004309.

References

Brazdova, M., Navratilova , L., Tichy, V., Nemcova, K., Lexa, M., Hrstka, R., Pecinka, P., Adamik, M., Vojtesek, B., Palecek, E., Deppert, W. and Fojta, M. 2013. Preferential binding of hot spot mutant p53 proteins to supercoiled DNA *In vitro* and in cells. *PLoS One*, 8 (3): e59567.

Brooijmans, N. and Humblet, C. 2010. Chemical space sampling by different scoring functions and crystal structures. *Journal of Computer Aided Molecular Design*, 24 (433–447)

Bullock, A. N., Henckel, J. and Fersht, A. R. 2000. Quantitative analysis of residual folding and DNA binding in mutant p53 core domain: definition of mutant states for rescue in cancer therapy. *Oncogene*, 19: 1245 -1256.

Busolo, D. S. and Woodgate, R. L. 2014. Cancer prevention in Africa: a review of the literature. *Global Health Promotion 1757-975*, 22 (2): 31 –39.

Butler, J. S. and Loh, S. N. 2003. Structure, function, and aggregation of the zinc-free form of the p53 DNA binding domain. *Biochemistry*, 42: 2396-2403.

Butler, J. S. and Loh, S. N. 2007. Zn^{2+} -dependent misfolding of the p53 DNA binding domain. *Biochemistry*, 46: 2630-2639.

Calhoun, S. and Daggett, V. 2011. Structural effects of the L145Q, V157F, and R282W cancer associated mutations in the p53 DNA-binding core domain. *Biochemistry*, 50 (23): 5345–5353.

Capriotti, E., Fariselli, P. and Casadio, R. 2005. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Res*, 33: 306-310.

Center, M., Siegel, R. and Jemal, A. 2008. Global cancer facts and figures. *American Cancer Society*: 1-60.

Chehab, N. H., Malikzay, A., Stavridi, E. S. and Halazonetis, T. D. 1999. Phosphorylation of Ser-20 mediates stabilization of human p53 in response to DNA damage. *PNAS*, 96 (24): 13777–13782.

Chen, R., Li, L. and Weng, Z. 2003. ZDOCK: An initial-stage protein-docking algorithm. *Proteins*, 52: 80 – 87.

Chen, Y., Dey, R. and Chen, L. 2010. Crystal structure of the p53 core domain bound to a full consensus site as a self-assembled tetramer. *Structure*, 18 (2): 246-256.

Cho, Y., Gorina, S., Jeffrey, P. D. and Pavletich, N. P. 1994. Crystal structure of a p53 tumor suppressor-DNA complex: understanding of tumorigenesis mutations. *Science*, 265 (346-355)

Chong, L. T., Snow, C. D., Rhee, Y. M. and Pande, V. S. 2005. Dimerization of the p53 oligomerization domain: identification of a folding nucleus by molecular dynamics simulations. *Journal of Molecular Biology*, 345: 869–878.

Chou, P. Y. and Fasman, G. D. 1974. Conformational parameters for amino acids in helical, β-sheet,and random coil regions calculated from proteins. *Biochemistry*, 13 (2): 211-222.

Connolly, M. L. 1983. Solvent-accessible surfaces of proteins and nucleic acids. *Science*, 221: 709-713.

de Souza, N. and Ornstein, R. L. 1999. Molecular dynamics simulations of a protein-protein dimer Particle-Mesh Ewald electrostatic model yields far superior results to standard cutoff model. *Journal of biomelecular structure and dynamics*, 16 (6): 1205-1218.

Derker, M. 2001. *Computational Biochemistry and Biophysics*. New York Marcel Dekker Inc.

Ditzler, M. A., Otyerpka, M., Spiner, J. and Walter, N. G. 2010. Molecular dynamics and quantum mechanics of RNA: conformational and chemical change we can believe in. *Accounts of chemical research*, 43 (1): 40-47.

Donehower, L. A., Harvey, M., Slagle, B. L., McArthur, M. J., Montgomery Jr, C. A., Butel, S. B. and Bradley, A. 1992. Mice deficient for p53 are developmentally normal but susceptible to spontaneous tumours. *Nature*, 356: 215-221.

Duan, J. and Nilsson, L. 2006. Effect of Zn^{2+} on DNA recognition and stability of the p53 DNA-binding domain. *Biochemistry*, 45: 7483-7492.

Duffy, M. J., Synnott, N. C., McGowan, P. M., Crown, J., D, O. C. and Gallagher, W. M. 2014. p53 as a target for cancer treatment. *Cancer Treatment Reviews*, 40: 1153–1160.

El-Deiry, W. S. 1998. Regulation of p53 downstream genes. *Cancer Biology*, 8: 345-357.

Eldar, A., Rozenberg, H., Diskin-Posner, Y., Rohs, R. and Shakked, Z. 2013. Structural studies of p53 inactivation by DNA-contact mutations and its rescue by suppressor mutations via alternative protein–DNA interactions. *Nucleic Acids Res*, 41 (18): 8748–8759.

Fanelli, F. and Ferrari, S. 2006. Prediction of MEF2A-DNA interface by rigid body docking: a tool for fast estimation of protein mutational effects on DNA binding. *J Struct Biol*, 153 (3): 278-283.

Fang, M., Simeonova, I. and Toledo, F. 2012. P53: point mutations, SNPs and cancer, point mutation. In: Logie, C. ed. *Piont mutations*. Croatia: InTech, 301-309.

Fernandez-Medarde, A. and Santos, E. 2011. Ras in cancer and developmental diseases. *Genes Cancer*, 2 (3): 344-358.

Ferreira, L. G., dos Santos, R. N., Oliva, N. and Andricopulo, A. G. 2015. Molecular Docking and Structure-Based Drug Design Strategies. *Molecules*, 20: 13384-13421.

Ferrer-Costa, C., Gelpi, J. L., Zamakola, L., Parraga, I., de la Cruz, X. and Orozco, M. 2005. PMUT: a web-based tool for the annotation of pathological mutations on proteins. *Bioinformatics*, 21 (14): 3176-3178.

Feyfant, E., Sali, A. and Fiser, A. 2007. Modeling mutations in protein structures. *Protein Science*, 16 (9): 2030-2041.

Finlay, C. A., Hinds, P. W., Tan, T. H., Eliyahu, D., Oren, M. and Levine, A. J. 1988. Activating mutations for transformation by p53 produce a gene product that forms an hsc7O-p53 complex with an altered half-life. *Molecular and cellular biology*, 8 (2): 531-539.

Folkman, L., Stantic, B., Sattar, A. and Zhou, Y. 2016. EASE-MM: Sequence-Based Prediction of Mutation-Induced Stability Changes with Feature-Based Multiple Models. *Journal of Molecular Biology*, 428 (6): 1394-1405.

Freed-Pastor, W. A. and Prives, C. 2012. Mutant p53: one name, many proteins. *Genes and development*, 26: 1268–1286.

Friedman, P. N., Kern, S. E., Vogelstein, B. and Prives, C. 1990. Wild-type, but not mutant, human p53 proteins inhibit the replication activities of simian virus 40 large tumor antigen. *PNAS*, 87: 9275-9279.

Fu, T., Min, H., Xu, Y., Chen, J. and G, L. 2012. Molecular dynamic simulation insights into the normal state and restoration of p53 function *Journal of Molecular Sciences*, 13: 9709-9740.

Ganguly, D. and Chen, J. 2015. Modulation of the disordered conformational ensembles of the p53 transactivation domain by cancer associated mutations. *PLOS Computational Biology*, 11 (4): e1004247.

Giacomazzi, C. R., Giacomazzi, J., Netto, C. B. O., Santos-Silva, P., Selistre, S. G., Maia, A. L., de Oliveira, V. Z., Camey, S. A., Goldim, J. R. and Ashton-Prolla, P. 2015. Pediatric cancer and Li-Fraumeni/Li-Fraumeni-like syndromes: a review for the pediatrician. *RAMB*, 61 (13): 282-289.

Greenblatt, M. S., Beaudet, J. G., Gump, J. R., Godin, K. S., Trombley, L., Koh, J. and Bond, J. P. 2003. Detailed computational study ofp53andp16: using evolutionary sequence analysis and disease-associated mutations to predict the functional consequences of allelic variants. *Oncogene*, 22: 1150–1163.

Gualberto, A., Aldape, K., Kozakiewicz, K. and Tlsty, T. 1998. An oncogenic form of p53 confers a dominant, gain-of-function phenotype that disrupts spindle checkpoint control. *PNAS*, 95 (5166–5171)

Guimaraes, D. P. and Hainaut, P. 2002. TP53: a key gene in human cancer. *Biochimie*, 84: 83-93.

Harms, P., Mittra, R. and Ko, W. 1994. Implementation of the periodic boundary conditions in the finite-difference time-domain algorithm for FSS structures. *IEEE*, 42 (9): 1317-1324.

Hejmad, M. 2010. Introduction to cancer biology. In. 2nd edn. 1-48.

Hu, H., Liu, J., Liao, X., Zhang, S., Li, H., Lu, R., Li, X., Lin, W., Liu, M., Xia, Z., Qing, G. and Li, J. D. 2016. Genetic and functional analysis of a Li Fraumeni syndrome family in China. *Scientific Report*, 6: 20221.

Huang, T., Niu, S., Xu, Z., Huang, Y., Kong, X., Cai, Y. D. and Chou, K. C. 2011. Predicting transcriptional activity of multiple site p53 mutants based on hybrid properties. *PLoS One*, 6 (8): e22940.

Jaya, H., S and Saleh, R. 2012. Build mutant and build homology protein structure predictions for indonesian avian influenza neuraminidase. *Journal of Biophysical Chemistry*, 3 (2): 183-190.

Jayaraman, A., Jamil, K. and Raju, S. 2012. The interaction of p53 and MDM2 genes in cancers, in silico studies and phylogenetic analysis. *Biology and medicine*, 3 (3): 1-12.

Joerger, A. C., Allen, M. D. and Fersht, A. R. 2004. Crystal structure of a superstable mutant of human p53 core domain. *Journal of Biological Chemistry*, 279 (2): 1291–1296.

Joerger, A. C., Ang, H. C. and Fersht, A. R. 2006. Structural basis for understanding oncogenic p53 mutations and designing rescue drugs. *PNAS*, 103 (41): 15056 –15061

Joerger, A. C., Ang, H. C., Veprintsev, D. B., Blair, C. M. and R, F. A. 2005. Structures of p53 cancer mutants and mechanism of rescue by second-site suppressor mutations
Journal of Biological Chemistry, 280 (16): 16030 –16037.

Joerger, A. C. and Fersht, A. R. 2007. Structure–function–rescue: the diverse nature of common p53 cancer mutants. *Oncogene*, 26: 2226–2242.

Joerger, A. C. and Fersht, A. R. 2010. The tumor suppressor p53: from structures to drug discovery. *Cold Spring Harbor Perspectives in Biology*: a000919.

Jorgensen, W. L. 2004. The many roles of computation in drug discovery. *Science*, 303 (5665): 1813-1818.

Jorgensen, W. L., Chandrasekhar, J. and Madura, J. D. 1983. Comparison of simple potential functions for simulating liquid water. *Journal of Chemical Physics*, 79 (2): 926-935.

Junttila, M. R. and Evan, G. I. 2009. p53- a jack of all trades but master of none. *Nat Rev Cancer*, 9: 821-829.

Kamada, R. 2012. Quantitative analysis for p53 tetramerization domain mutants reveals a low threshold for tumor suppressor inactivation. In: *Tetramer stability and functional regulation of tumor suppressor protein p53*. Japan Springer Theses, 13-43.

Kamaraj, B. and Bogaerts, A. 2015. Structure and function of p53-DNA complexes with inactivation and rescue mutations: a molecular dynamics simulation study. *PLoS One*, 10 (8): e0134638.

Kamaraj, B. and Purohit, R. 2013. In silico screening and molecular dynamics simulation of disease-associated nsSNP in TYRP1 gene and its structural consequences in OCA3. *BioMedical Research International*, 2013: 1-13.

Kang, T. S. and Kini, R. M. 2009. Structural determinants of protein folding. *Cellular and Molecular Life Sciences*, 66 (14): 2341-2361.

Karplus, M. and Petsko, G. A. 1990. Molecular dynamic simulation in biology. *Nature*, 347: 631-639.

Kato, S., Han, S. Y., Liu, W., Otsuka, K., Shibata, H., Kanamaru, R. and Ishioka, C. 2003. Understanding the function-structure and function-mutation relationships of p53 tumor suppressor protein by high-resolution missense mutation analysis. *PNAS*, 100 (14): 8424-8429.

Kholmurodov, K. T. 2013. Molecular dynamics study of the effect of induced mutations on the protein structures associated with diseases of a radiobiological nature. *American Journal of Bioscience and Bioengineering*, 1 (1): 7-16.

Kim, E. and Deppert, W. 2007. Interactions of mutant p53 with DNA: guilt by association. *Oncogene*, 26: 2185-2190.

Kitchen, D. B., Decornez, H., Furr, J. R. and Bajorath, J. 2004. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat Rev Drug Discov*, 3 (11): 935-949.

Knowles, M. A. and Selby, P. J. 2005. *Introduction to the cellular and molecular biology of cancer*. 4th ed. New York: Oxford University Press.

Kumar, P., Henikoff, S. and Ng, P. C. 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc*, 4 (7): 1073-1081.

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

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

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

Leckband, D. 2000. Measuring the forces that control protein interactions. *Annual Reviews Biophysical Biomolecular Structure*, 29: 1–26.

Lee, B. and Richards, F. M. 1971. The interpretation of protein structures: estimation of static accessibility. *Journal of Molecular Biology*, 55 (379-490)

Lee, W., Harvey, T. S., Yin, Y., Yau, P., Litchfield, D. and Arrowsmith, C. H. 1994. Solution structure of the tetrameric minimum transforming domain of p53. *Structural biology*, 1 (12): 877-890.

Lengauer, T. and Rarey, M. 1996. Computational methods for biomolecular docking *Current Opinion in Structural Biology*, 6 (3): 402-406.

Levine, A. J. 1997. p53 the cellular gatekeeper for growth and division. *Cell Press*, 88: 323–331.

Levine, A. J., Finlay, C. A. and Hinds, P. W. 2004. P53 is a tumor suppressor gene. *Cell Press*, 116: 67-69.

Levine, A. J. and Oren, M. 2009. The first 30 years of p53: growing ever more complex. *Nature Reviews Cancer*, 9 (10): 749–758.

Liang, S.-H. and Clarke, M. F. 1999. A bipartite nuclear localization signal is required for p53 nuclear import regulated by a carboxyl-terminal domain. *The Journal of Biological Chemistry*, 274 (46): 32699 – 32703.

Lindahl, E. R. 2008a. Molecular dynamics simulations. In: M, W. J. ed. *Molecular modeling of proteins*. United States of America: Humana Press, 3-23.

Lindahl, E. R. 2008b. *Molecular modeling of proteins*. New Jersey: Hanama Press.

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

Lomax, M. E., Barnes, D. M., Gilchrist, R., Picksley, S. M., Varley, J. M. and Camplejohn, R. S. 1997. Two functional assays employed to detect an unusual mutation in the oligomerisation domain ofp53in a Li-Fraumeni like family. *Oncogene*, 14: 1869 -1874.

Lovell, S. C., Davis, I. W., Arendall III, W. B., de Bakker, P. I. W., Word, J. M., Prisant, M. G., Richardson, J. S. and Richardson, D. C. 2003. Structure validation by Cα geometry: ϕΨ and Cβ deviation. *Proteins*, 50: 437–450

Lu, Q., Tan, Y.-H. and Luo, R. 2007. Molecular dynamics simulations of p53 DNA-binding domain. *Journal of Physical Chemistry*, 111: 11538-11545.

Lukman, S., Lane, D. P. and Verma, C. V. 2013. Mapping the structural and dynamical features of multiple p53 DNA binding domains: insights into loop 1 intrinsic dynamics. *PLoS One,* 8 (11): e80221.

Luthy, R., Bowie, J. W. and Eisenberg, W. 1992. Assesment of of protein models with three-dimensional profiles. *Nature*, 356: 83-85.

MacArthur, M. W. and Thornton, J. M. 1991. Influence of proline residues on protein conformation. *Journal of Molecular Biology*, 218 (2): 397-412.

Martin, A. C. R., Facchiano, A. M., Cuff, A. L., Hernandez-Boussard, T., Olivier, M., Hainaut, P. and Thornton, J. M. 2002. Integrating mutation data and structural analysis of the TP53 tumor-suppressor protein. *hum Mutation*, 19: 149-164.

Mathe, E., Olivier, M., Kato, S., Ishioka, C., Hainaut, P. and Tavtigian, S. V. 2006. Computational approaches for predicting the biological effect of p53 missense mutations: a comparison of three sequence analysis based methods. *Nucleic Acids Res*, 34 (5): 1317-1325.

May, P. and May, E. 1999. Twenty years of p53 research: structural and functional aspects of the p53 protein. *Oncogene*, 18: 7621-7636.

Mayosi, B. M., Flisher, A. J., Lalloo, U. G., Sitas, F., Tollman, S. M. and Bradshaw, D. 2009. The burden of non-communicable diseases in South Africa. *Lancet*, 374: 934–947.

Meng, X.-Y., Zhang, H.-X., Mezei, M. and Cui, M. 2011. Molecular docking: a powerful approach for structure-based drug discovery. *Current Computational Aided Drug Design*, 7 (2): 146–157.

Merabet, A., Houlleberghs, H., Maclagan, K., Akanho, E., Bui, T. T. T., Pagano, B., Drake, A. F., Fratenali, F. and Nikolova, P. V. 2010. Mutants of the tumour suppressor p53 L1 loop as second-site suppressors for restoring DNA binding to oncogenic p53 mutations: structural and biochemical insights. *Biochemistry*, 427: 225–236.

Muller, P. A. J. and Vousden, K. H. 2013. p53 mutations in cancer. *Nature Cell Biology*, 15 (1): 1-8.

Muller, P. A. J. and Vousden, K. H. 2014. Mutant p53 in cancer: new functions and therapeutic opportunities. *Cancer Cell*, 25: 303-317.

Natan, E., Baloglu, C., Pagel, K., Freund, S. M. V., Morgner, N., Robinson, C. V., Fersht, A. R. and Joerger, A. C. 2011. Interaction of the p53 DNA-binding domain with its N-terminal extension modulates the stability of the p53 tetramer. *Journal of Molecular Biology*, 409: 358–368.

Ng, J. W. K., Lama, D., Lukman, S., Lane, D. P., Verma, C. S. and Sim, A. Y. L. 2015. R248Q mutation—beyond p53-DNA binding. *Proteins*, 83: 2240–2250.

Nikolova, P. V., Henckel, J., Lane, D. P. and Fersht, A. R. 1998. Semirational design of active tumor suppressor p53 DNA binding domain with enhanced stability. *PNAS*, 95: 14675–14680.

Nikolova, P. V., Wong, K.-B., DeDecker, B., Henckel, J. and Fersht, A. R. 2000. Mechanisms of rescue of common p53 cancer mutations by second-site suppressor mutations. *The Embo Journal*, 19 (3): 370-378.

Olivier, M., Hollstein, M. and Hainaut, P. 2010. TP53 mutations in human cancers: origins, consequences, and clinical use. *Cold Spring Harbor Perspective Biology*, 2: a001008.

Oren, M. and Rotter, V. 2010. Mutant p53 gain of function in cancer. *Cold Spring Harbor Perspective Biology*: a001107.

Ozaki, T. and Nakurawara, A. 2011. Role of p53 in cell death and human cancers. *Cancers*, 3: 994-1013.

Pagano, B., Jama, A., Martinez, P., Akanho, E., Bui, T. T. T., Drake, A. F., Fraternali, F. and Nikolova, P. V. 2013. Structure and stability insights into tumour suppressor p53 evolutionary related proteins. *PLoS One*, 8 (10): e76014.

Parks, D., Bolinger, R. and Mann, K. 1997. Redox state regulates binding of p53 to sequence-specific DNA, but not to non-specific or mismatched DNA. *Nucleic Acids Res*, 25 (6): 1289–1295.

Parrales, A. and Iwakuma, T. 2015. Targeting oncogenic mutant p53 for cancer therapy. *Frontiers Oncology*, 5 (288): 1-13.

Parthiban, V., Gromiha, M. M. and Schomburg, D. 2006. CUPSAT: prediction of protein stability upon point mutations. *Nucleic Acids Res*, 34: 239-242.

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 tumor phenotype: lessons from recent developments in the IARC TP53 database. *Hum Mutat*, 28 (6): 622-629.

Pierce, B. and Weng, Z. 2007. ZRANK: reranking protein docking predictions with an optimized energy function. *Proteins*, 67 (4): 1078-1086.

Pierce, B., Wiehe, K., Hwang, H., Kim, H.-Y. V., T and Weng, Z. 2014. ZDOCK server: interactive docking prediction of protein–protein complexes and symmetric multimers. *Bioinformatics Applications Note*, 30 (12): 1771–1773.

Pronk, S., Pa´ll, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shirts, M. R., Smith, J. C., Kasson, P. M., van der Spoel, D., Hess, B. and Lindahl, E. R. 2013. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Structural bioinformatics*, 29 (7): 845–854.

Rainwater, R., D, P., Anderson, M. E., P, T. and K, M. 1995. Role of cysteine residues in regulation of p53 function. *Molecular and cellular biology*, 15 (7): 3892–3903.

Ramani, R. G. and Jacob, S. G. 2013. Prediction of p53 mutants (multiple sites) transcriptional activity based on structural (2D&3D) properties. *PLoS One*, 8 (2): e55401.

Rauf, S. M., Endou, A., Takaba, H. and Miyamoto, A. 2013. Effect of Y220C mutation on p53 and its rescue mechanism: a computer chemistry approach. *Protein J*, 32 (1): 68-74.

Robles, A. I. and Harris, C. C. 2010. Clinical outcomes and correlates of TP53 mutations and cancer. *Cold Spring Harbor Perspective Biology*, 2 (3): a001016.

Sable, R. and Jois, S. 2015. Surfing the protein-protein interaction surface using docking methods: application to the design of PPI inhibitors. *Molecules*, 20 (6): 11569-11603.

Saha, T., Kar, R. K. and Sa, G. 2015. Structural and sequential context of p53: a review of experimental and theoretical evidence. *Progress in Biophysics and Molecular Biology*, 117: 250-263.

Schlick, T. 2010a. Biomolecular structure and modeling: historical perspective. In: Antman, S. S., Marsden, J. E. and Sirovich, L. eds. *Molecular modeling and simulation: An interdisciplinary guide*. Springer New York: 1-40.

Schlick, T. 2010b. Theoretical and computational approaches to biomolecular structure. In: Antman, S. S., Marsden, J. E. and Sirovich, L. eds. *Molecular modeling and simulation: An interdisciplinary guide*. 2nd edn. 237-262.

Selivanova, G., Ryabchenko, L., Jansson, E., Lotsava, V. and Wiman, K. G. 1999. Reactivation of mutant p53 through interaction of a C-terminal peptide with the core domain. *Molecular and cellular biology*, 19 (5): 3395–3402.

Shen, M. Y. and Sali, A. 2006. Statistical potential for assessment and prediction of protein structures. *Protein Science*, 15 (11): 2507-2524.

Shruti, K., Archan, A., Uddhavesh, S. and Uddhavesh, J. 2015. Investigating DNA binding and conformational variation in temperature sensitive p53 cancer mutants using QM-MM simulations. *PLoS One*, 10 (11): e0143065.

Siegel, R., Ward, E., Brawley, O. and Jemal, A. 2011. The impact of eliminating socioeconomic and racial disparities on premature cancer deaths. *CA Cancer Journal for Clinicians*, 1 (61): 212–236.

Silva, J. L., De Moura Gallo, C. V., Costa, D. C. F. and Range, L. P. 2014. Prion-like aggregation of mutant p53 in cancer cell. *Trends in Biochemical Science*, 39 (6): 260-267.

Singh, E., Ruff, P., Babb, C., Sengayi, M., Beery, M., Khoali, L., Kellett, P. and Underwood, J. M. 2015. Establishment of a cancer surveillance programme: the South African experience. *Lancet Oncology*, 16: 414–421.

Sitas, F., Parkin, D. M., Chirenje, M., Stein, L., Abratt, R. and Wabinga, H. 2008. Part II: Cancer in Indigenous Africans-causes and control. *Lancet Oncology*, 9: 786–795.

Smith, D. E. and Dang, L. X. 1994. Computer simulations of NaCI association in polarizable water. *Journal of Chemical Physics*, 100 (5): 3757-3766.

Soussi, T. and Lozano, G. 2005. p53 mutation heterogeneity in cancer. *Biochem Biophys Res Commun*, 331 (3): 834-842.

Spassov, V. Z., Yan, L. and Flook, P. K. 2007. The dominant role of side-chain backbone interactions in structural realization of amino acid code. ChiRotor: a side-chain prediction algorithm based on side-chain backbone interactions. *Protein Science*, 16 (3): 494-506.

Stavridi, E. S., Huyen, Y., Sheston, E. A. and Halazonetis, T. D. 2005. The three-dimensional structure of p53. In: Zambetti, E. D. ed. *The p53 tumor suppressor pathway and cancer*. New York: Springer Science+Business Media, 25-52.

Stefl, S., Nishi, H., Petukh, M., Panchenko, A. R. and Alexov, E. 2013. Molecular mechanisms of disease-causing missense mutations. *Journal of Molecular Biology*, 425 (21): 3919-3936.

Stojnev, S., Golubović, M. and Babović, P. 2009. Tp 53 gene mutations- from gudiain of the genome to oncogene. *Acta Medica Medianae*, 48 (4): 59-63.

Stommel, J. M., Marchenko, N. D., Jimenez, G. S., Moll, U. M., Hope, T. J. and Wahl, G. M. 1999. A leucine-rich nuclear export signal in the p53 tetramerization domain: regulation of subcellular localization and p53 activity by NES masking. *The Embo Journal*, 18 (6): 1660–1672.

Sudhakar, A. 2009. History of cancer, ancient and modern treatment methods. *Journal of Cancer Science & Therapy*, 1 (2): 1-4.

Sushma, B. and Suresh, C. V. 2012. Docking-a Review. *Journal of Applicable Chemistry*, 1 (2): 167-173.

Szymañska, K. and Hainaut, P. 2003. TP53 and mutations in human cancer. *Acta Biochimica Polonica*, 50 (1): 231–238.

Tan, Y. H., Chen, Y. M., Ye, X., Lu, Q., Tretyachenko-Ladokhina, V., Yang, W., Senear, D. F. and Luo, R. 2009. Molecular mechanisms of functional rescue mediated by P53 tumor suppressor mutations. *Biophysical Chemistry*, 145: 37–44.

Teng, S., Michonova-Alexova, E. and Alexov, E. 2008. Approaches and resources for prediction of the effects of nonsynonymous single nucleotide polymorphism on protein function and interactions. *Current pharmaceutical biotechnology*, 9: 123-133.

Teng, S., Srivastava, A. K. and Wang, L. 2010. Sequence feature-based prediction of protein stability changes upon amino acid substitutions. *BioMedical Central Genomics*, 11 (2): 1-8.

Tharakaraman, K., Robinson, L. N., Hatas, A., Chen, Y. L., Siyue, L., Raguram, S., Sasisekharan, V., Wogan, G. N. and Sasisekharan, R. 2013. Redesign of a cross-reactive antibody to dengue virus with broad-spectrum activity and increased in vivo potency. *PNAS*, 110 (17): 1555-1564.

Toledo, F., Lee, C. J., Krummel, K. A., Rodewald, L.-W., Liu, C.-W. and Wahl, G. M. 2007. Mouse mutants reveal that putative protein interaction sites in the p53 proline-rich domain are dispensable for tumor suppression. *Molecular and cellular biology*, 27 (4): 1425–1432.

Uline, M. J. and Corti, D. S. 2013. Molecular dynamics at constant pressure: allowing the system to control volume fluctuations via a "shell" particle. *Entropy*, 15: 3941-3969.

van der Spoel, D., Lindahl, E. R., Hess, B., Groenhof, G., Mark, A. E. and Berendsen, H. J. C. 2005. Gromacs: Fast, Flexible and Free. *J Comp Chem*, 26: 1701-1718.

van Gunsteren, W. F., Bakowies, D., Baron, R., Chandrasekhar, I., Christen, M., Daura, X., Gee, P., Geerke, D. P., Glattli, A., Hunenberger, P. H., Kastenholz, M. A., Oostenbrink, C., Schenk, M., Trzesniak, D., van der Vegt, N. F. and Yu, H. B. 2006. Biomolecular modeling: goals, problems, perspectives. *Angewandte Chemie International Edition*, 45 (25): 4064-4092.

Ve´gran, F., Rebucci, M., Chevrier, S., Cadouot, M., Boidot, R. and S, L.-N. 2013. Only missense mutations affecting the DNA binding domain of P53 influence outcomes in patients with breast carcinoma. *PLoS One*, 8 (1): e55103.

Vogelstein, B., Lane, D. and Levine, A. J. 2000. Surfing the p53 network. *Nature*, 408: 307-310.

Walker, D. R. B., J P Tarone, R E Harris, C C Makalowski, W Boguski, M S Greenblatt, M S. 1999. Evolutionary conservation and somatic mutation hotspot maps of p53: correlation with p53 protein structural and functional features. *Oncogene*, 18 (1): 211-218.

Weisz, L., Oren, M. and Rotter, V. 2007. Transcription regulation by mutant p53. *Oncogene*, 26: 2202–2211.

Whibley, C., Pharoah, P. D. P. and Hollstein, M. 2009. p53 Polymorphisms: cancer implications. *Nature Reviews*, 9: 95-107.

Wolf, D. and Rotter, V. 1984. Inactivation of p53 gene expression by an insertion of moloney murine leukemia virus-Like DNA sequences. *Molecular and cellular biology*, 4 (7): 1402-1410.

Wong, K.-B., DeDecker, B. S., Freund, S. M. V., Proctor, M. R., M, B. and Fersht, A. R. 1999. Hot-spot mutants of p53 core domain evince characteristic local structural changes. *PNAS*, 96: 8438–8442.

Worth, C. L., Preissner, R. and Blundell, T. L. 2011. SDM-a server for predicting effects of mutations on protein stability and malfunction. *Nucleic Acids Res,* 39: 215-222.

Wright, J. D. and Lim, C. 2007. Mechanism of DNA-binding loss upon single point mutation in p53. *Journal of Biosciences*, 32: 827-839.

Wu, G. and Yan, S. 2003. Determination of amino acid pairs in human p53 protein sensitive to mutations/variants by means of a random approach. *J Mol Model*, 9: 337–341.

Xu, H. and El-Gewely, M. R. 2001. P53-responsive genes and the potential for cancer diagnostics and therapeutics development. *Biotechnology Annual Review*, 7: 131-164.

Zhang, Y., Coillie, S. V., Fang, J. Y. and Xu, J. 2016. Gain of function of mutant p53: R282W on the peak? *Oncogenesis*, 5: e196.

Zhang, Z., Miteva, M. A., Wang, L. and Alexov, M. 2012. Analyzing effects of naturally occurring missense mutations. *Computational and mathematical methods in medicine*: 1-15.

Zhou, Z., Felts, A. K., Friesner, R. A. and Levy, R. M. 2007. Comparative performance of several flexible docking programs and scoring functions: enrichment studies for a diverse set of pharmaceutically relevant targets. *Journal of Chemical Information and Modeling*, 47 (4): 1599–1608.

APPENDICES

Appendix I

```
>sp|P04637-2|P53_HUMAN Isoform 2 of Cellular tumor antigen p53 OS=Homo sapiens
GN=TP53
MEEPQSDPSVEPPLSQETFSDLWKLLPENNVLSPLPSQAMDDLMLSPDDIEQWFTEDPGP
DEAPRMPEAAPPVAPAPAAPTPAAPAPAPSWPLSSSVPSQKTYQGSYGFRLGFLHSGTAK
SVTCTYSPALNKMFCQLAKTCPVQLWVDSTPPPGTRVRAMAIYKQSQHMTEVVRRCPHHE
RCSDSDGLAPPQHLIRVEGNLRVEYLDDRNTFRHSVVVPYEPPEVGSDCTTIHYNYMCNS
SCMGGMNRRPILTIITLEDSSGNLLGRNSFEVRVCACPGRDRRTEEENLRKKGEPHHELP
PGSTKRALPNNTSSSPQPKKKPLDGEYFTLQDQTSFQKENC
```

Figure S1: P53 DNA-binding domain sequence in a FASTA format.

Table 5.1: List of all intolerant positions of the amino acid residues that are susceptible to mutations.

Number of intolerant mutations	Amino acid positions[a]
19	Phe19, Gly105, **Lys120**, Ser120, Thr125, Ser127, Leu130, Lys132, Cys132, Cys137, Lys139, Thr140, Pro152, Gly154, Ala161, Lys164, Val173, Arg175, Cys176, His178, His179, His193, Leu194, Arg196, G199, Tyr205, Ser215, Val216, Pro219, Tyr220, Gly226, Thr230, Thr231, Met237, Cys238,Asp239, Ser240, **Ser241**, Cys242, Met243, Gly244, Gly245, Met246, Arg249, Pro250, Ile251, Thr253, Ile254, Thr256, Leu265, Gly266, Phe270, Gly266, Glu271, Val272, Cys275, Arg278, **Arg280**, Asp281, Glu281, Glu286, Leu330, Gly334,Asp339, Leu348, Leu350.
18	Glu11, Leu22, Trp23, Tyr103, Phe109, Leu111, Val122, Try126, Met133, Phe134, Gln136, Cys141, Pro142, Pro151, Arg151, Arg158, Ile162, Tyr163, Val172, Arg174, Glu198, Asp208, Thr211, Arg213, His214, Val218, Glu221, Tyr236, Leu252, Gly262, Arg267, **Arg273**, Val274, **Ala276, Cys277**, Arg282, Glu326, Ile332, Arg284, Arg335, Phe338
17	Thr18, Val143, Val147, Val157, Glu157, Glu171, Val197, Asp207, Phe328, Arg337, Leu344, Ala347, Glu349
16	Leu26, Tyr107,Gly108, Met169, Ile195,Val203,Tyr234, Asn235, Leu264
15	Cys124, Gln144, Leu145, Gly187,Phe341
14	Arg110, Asp228, Arg306, Ile333
13	-
12	Glu21
11	Met160
10	Gly233, Glu336,
9	-
8	-
7	-
6	-
5	-
4	-
3	Leu383
2	-
1	Lue45, Ala83, Pro290
0	Val31, Phe72, Arg156, Arg202, Phe212, Asp259, Ser269,

[a]There exist conserved positions in the p53 structure where mutations are not tolerated and will increase/decrease stability of the protein.

Table S1: Nucleotide names and 3-letter words

Nucleotide base	3-Letter word
Guanine	Gua
Adenine	Ade
Thymine	Thy
Cytosine	Cyt

Table S2: Amino acids names and 3-letter words

Amino Acid name	3-Letter word	Amino Acid name	3-Letter word
Glycine	Gly	Threonine	Thr
Alanine	Ala	Cysteine	Cys
Valine	Val	Tyrosine	Try
Leucine	Leu	Asparagine	Asn
Isoleucine	Iso	Glutamine	Gln
Methionine	Met	Aspartic acid	Asp
Tryptophan	Trp	Glutamic Acid	Glu
Phenylalanine	Phe	Lysine	Lys
Proline	Pro	Arginine	Arg
Serine	Ser	Histidine	His

Table S3: DOPE and Mol PDF scores of predicted stable mutants

Mutant	PDF Total Energy	DOPE Score
Val143Asp	1111.1512	-17110.14257
Ala159Pro	1036.2098	-17379.6328
Val197Pro	1074.0341	-17336.36132
Try234Pro	1108.6473	-17032.0000
Cys238Pro	1080.5300	-17291.5469
Gly262Pro	1101.4941	-17413.8477
Cys275Pro	1064.3381	-17313.3925

Table S4: List of computationally predicted deleterious mutations and the stability of the resultant mutants in p53 protein

S. No.	Amino acid Mutation	SIFT[a] and Pmut[b] Predictions	Reliability Index (RI)[c]	Protein Stability											
				I-MUTANT 2.0 [d] (Sequence based)		I-MUTANT 2.0 (Structure based)		MuStab server [e]		CUPSAT server [f]		SDM server [g]		EASE-MM server	
				Predictions (Stability)	ΔΔG	Predictions (Stability)	ΔΔG	Predictions (Stability)	confidence	Predictions (Overall stability/torsions)	ΔΔG	Predictions (Stability/effect)	ΔΔG	Predictions (Stability/effect)	ΔΔG
1.	**Glu11Lys**	Pathological	5	Decrease	-0.84	-	-	Decrease	77.50%	-	-	-	-	Neutral	0.1262
2.	**Thr18Pro**	Pathological	3	Decrease	-0.84	-	-	Decrease	93.93%	-	-	-	-	Destabilizing	-1.0456
3.	**Phe19Lys**	Pathological	8	Decrease	-2.22	-	-	Decrease	94.11%	-	-	-	-	Destabilizing	-2.5054
4.	**Asp21Arg**	Pathological	4	Increase	0.26	-	-	Decrease	78.57%	-	-	-	-	Neutral	0.4110
5.	**Leu22Asn**	Pathological	7	Decrease	-2.74	-	-	Decrease	91.43%	-	-	-	-	Destabilizing	-2.4798
6.	**Trp23Pro**	Pathological	3	Decrease	-2.20	-	-	Decrease	90.17%	-	-	-	-	Destabilizing	-2.6167
7.	**Leu26His**	Pathological	7	Decrease	-1.26	-	-	Decrease	89.64%	-	-	-	-	Destabilizing	-1.6664
8.	**Val31His**	Pathological	6	Decrease	-2.50	-	-	Decrease	92.50%	-	-	-	-	Likely destabilizing	-0.5487
9.	**Leu45Asp**	Pathological	2	Decrease	-2.75	-	-	Decrease	91.96%	-	-	-	-	Destabilizing	-1.2298
10.	**Pro72Phe**	Pathological	4	Increase	-0.24	-	-	Decrease	78.57%	-	-	-	-	Destabilizing	-1.007
11.	**Ala83Trp**	Pathological	3	Decrease	0.13	-	-	Increase	22.86%	-	-	-	-	Likely destabilizing	-0.5894
12.	**Tyr103Lys**	Pathological	5	Decrease	-1.15	Decrease	-1.36	Decrease	94.11%	Stabilizing/ unfavourable	1.88	Destabilizing/ non-disease associated	-1.92	Destabilizing	-1.7140

13.	**Gly105Glu**	Pathological	8	Increase	0.80	Decrease	-0.13	Decrease	81.61%	Destabilizing/ unfavourable	-1.44	Highly destabilizing/ cause protein malfunction and disease	-3.71	Likely destabilizing	-0.6517
14.	**Tyr107Gln**	Pathological	3	Decrease	-0.72	Decrease	-1.33	Decrease	90.54%	Stabilizing/ unfavorable	-0.25	Slightly destabilizing/ non-disease associated	-0.88	Destabilizing	-1.7552
15.	**Gly108Lys**	Pathological	1	Decrease	-0.76	Decrease	-1.39	Increase	25.18%	Destabilizing/ Favourable	-0.96	Destabilizing/ non-disease associated	-1.30	Likely destabilizing	-0.7858
16.	**Phe109Lys**	Pathological	2	Decrease	-2.32	Decrease	-4.14	Decrease	94.82%	Destabilizing/ unfavourable	-3.21	Destabilizing/ cause protein malfunction and disease	-3.06	Destabilizing	-2.9295
17.	**Arg110Val**	Pathological	2	Increase	-0.04	Decrease	-0.00	Increase	26.62%	Destabilizing/ Favorable	-3.14	Slightly destabilizing/ non-disease associated	0.97	Likely destabilizing	-0.5313
18.	**Leu111Asp**	Pathological	7	Decrease	-2.46	Decrease	-2.63	Decrease	91.79%	Destabilizing/ unfavourable	-0.27	Destabilizing/ cause protein malfunction and disease	-4.17	Destabilizing	-3.1922
19.	**Lys120Asp**	Pathological	0	Decrease	-0.62	Increase	0.43	Decrease	89.64%	Stabilizing/ unfavourable	0.28	Neutral/ non-disease associated	0.01	Neutral	-0.2402
20.	**Ser121Pro**	Pathological	1	Increase	-0.70	Increase	0.06	Decrease	81.79%	Destabilizing/ unfavourable	-0.68	Neutral/ non-disease associated	0.03	Neutral	-0.2402
21.	**Val122Asp**	Pathological	9	Decrease	-2.81	Decrease	-2.20	Decrease	93.39%	Destabilizing/ unfavourable	-5.26	Neutral/ non-disease associated	0.35	Likely destabilizing	-0.5461
22.	**Cys124Pro**	Pathological	5	Decrease	-1.20	Decrease	-1.64	Decrease	91.07%	Destabilizing/ unfavourable	-8.94	Highly destabilizing/ cause protein	-4.77	Neutral	0.2852

												malfunction and disease			
23.	**Thr125Val**	Pathological	7	Decrease	-0.43	Decrease	-1.25	Decrease	81.79%	Destabilizing/ Favourable	-1.32	Highly stabilizing/ cause protein malfunction and disease	2.95	Neutral	0.0297
24.	**Tyr126Pro**	Pathological	7	Decrease	-1.69	Decrease	-2.46	Decrease	92.32%	Destabilizing/ unfavourable	-5.04	Highly destabilizing/ cause protein malfunction and disease	-2.48	Destabilizing	-1.7540
25.	**Ser127Pro**	Pathological	5	Increase	-0.92	Increase	-0.19	Decrease	89.82%	Destabilizing/ unfavourable	-2.29	Destabilizing/ non-disease associated	-1.41	Slightly destabilizing	-0.7651
26.	**Leu130Asp**	Pathological	1	Decrease	-3.68	Decrease	-1.79	Decrease	92.86%	Destabilizing/ No change	-0.56	Destabilizing/ non-disease associated	-1.27	Destabilizing	-1.2953
27.	**Lys132Asp**	Pathological	2	Decrease	-0.97	Decrease	-1.74	Decrease	92.86%	Destabilizing/ unfavourable	-5.73	Slightly stabilizing/ non-disease associated	1.00	Destabilizing	-1.2824
28.	**Met133Pro**	Pathological	9	Decrease	-0.61	Decrease	-3.07	Decrease	92.86%	Destabilizing/ unfavourable	-2.99	Highly destabilizing/ cause protein malfunction and disease	-3.76	Destabilizing	-1.5695
29.	**Phe134Pro**	Pathological	5	Decrease	-2.00	Decrease	-2.68	Decrease	94.11%	Destabilizing/ unfavourable	-6.47	Highly destabilizing/ cause protein malfunction and disease	-4.53	Destabilizing	-1.6704
30.	**Cys135Ile**	Pathological	0	Decrease	-0.30	Decrease	-2.19	Decrease	78.57%	Destabilizing/ unfavourable	-10.23	Destabilizing/ non-disease associated	-1.18	Neutral	-0.1282
31.	**Gln136Pro**	Pathological	8	Decrease	-0.89	Decrease	-1.09	Decrease	78.07%	Stabilizing/ unfavourable	0.21	Slightly stabilizing/ non-	0.70	Slightly destabilizing	-0.7451

											disease associated				
32.	**Leu137Asp**	Pathological	6	Decrease	-3.36	Decrease	-2.44	Decrease	94.11%	Destabilizing/ Favourable	-1.53	Highly destabilizing/ cause protein malfunction and disease	-2.07	Destabilizing	-1.0164
33.	**Lys139Asp**	Pathological	3	Increase	-0.80	Decrease	-2.02	Decrease	92.32%	Stabilizing/ unfavourable	0.03	Neutral/ non-disease associated	0.31	Slightly destabilizing	-0.8135
34.	**Thr140Asp**	Pathological	2	Decrease	-0.45	Decrease	-0.86	Increase	22.86%	Destabilizing/ unfavourable	-0.85	Neutral/ non-disease associated	-0.15	Likely destabilizing	-0.8419
35.	**Cys141Glu**	Pathological	2	Increase	0.40	Decrease	-1.56	Increase	25.18%	Destabilizing/ unfavourable	-8.33	Highly destabilizing/ cause protein malfunction and disease	-2.89	Destabilizing	-2.0585
36.	**Pro142Asp**	Pathological	7	Decrease	-1.49	Decrease	-1.85	Increase	23.57%	Destabilizing/ unfavourable	-1.02	Slightly stabilizing/ non-disease associated	0.71	Destabilizing	-2.0377
37.	**Val143Asp**	Pathological	1	Decrease	-2.01	Decrease	-3.65	Decrease	91.07%	Destabilizing/ unfavourable	-10.13	Highly destabilizing/ cause protein malfunction and disease	-5.07	Destabilizing	-2.8030
38.	**Gln144Pro**	Pathological	1	Decrease	-0.19	Decrease	-1.23	Decrease	88.04%	Stabilizing/ unfavourable	3.69	Highly destabilizing/ cause protein malfunction and disease	-2.28	Neutral	-0.4813
39.	**Leu145Lys**	Pathological	9	Decrease	-1.79	Decrease	-1.24	Decrease	88.39%	Destabilizing/ unfavourable	-3.44	Highly destabilizing/ cause protein malfunction and disease	-4.28	Destabilizing	-2.1784

40.	**Val147Asp**	Pathological	3	Decrease	-2.20	Decrease	-1.76	Decrease	94.11%	Destabilizing/ unfavourable	-1.73	Highly destabilizing/ cause protein malfunction and disease	-3.85	Destabilizing	-2.7449
41.	**Pro151Asp**	Pathological	2	Decrease	-1.04	Decrease	-1.22	Decrease	78.04%	Destabilizing/ unfavourable	-3.11	Destabilizing/ non-disease associated	-1.16	Destabilizing	-1.3997
42.	**Pro152Asp**	Pathological	2	Decrease	-1.28	Decrease	-1.29	Increase	28.21%	Destabilizing/ Favourable	-1.52	Slightly destabilizing/ non-disease associated	-0.95	Destabilizing	-1.3247
43.	**Gly154Trp**	Pathological	2	Decrease	-1.12	Decrease	0.67	Decrease	84.11%	Destabilizing/ unfavourable	-2.93	Destabilizing/ non-disease associated	-1.38	Destabilizing	-1.6332
44.	**Arg156Asp**	Pathological	4	Decrease	-1.16	Decrease	-1.30	Decrease	78.93%	Destabilizing/ unfavourable	-3.83	Highly destabilizing/ cause protein malfunction and disease	-2.20	Destabilizing	-1.6559
45.	**Val157Glu**	Pathological	3	Decrease	-1.44	Decrease	-3.54	Decrease	94.46%	Destabilizing/ unfavourable	-2.48	Highly destabilizing/ cause protein malfunction and disease	-3.80	Destabilizing	-2.8035
46.	**Arg158Asp**	Pathological	7	Decrease	-0.71	Decrease	-2.40	Decrease	81.79%	Destabilizing/ unfavourable	-3.58	Neutral/ non-disease associated	0.25	Destabilizing	-2.6578
47.	**Ala159Pro**	Pathological	4	Increase	-0.50	Decrease	-3.33	Decrease	81.61%	Destabilizing/ unfavourable	-6.40	Highly destabilizing/ cause protein malfunction and disease	-3.35	Likely destabilizing	-0.6225
48.	**Met160Asp**	Pathological	4	Increase	0.90	Decrease	-2.74	Decrease	90.71%	Stabilizing/ unfavourable	3.68	Slightly destabilizing/ non-disease associated	-0.70	Destabilizing	-2.6718

49.	**Ala161Trp**	Pathological	6	Decrease	-0.70	Decrease	-1.48	Increase	25.71%	Stabilizing/ favourable	0.96	Neutral/ non-disease associated	-0.24	Neutral	0.1923
50.	**Ile162Asp**	Pathological	4	Decrease	-2.47	Decrease	-2.78	Decrease	96.79%	No change/ Unfavourable	-0.00	Highly destabilizing/ cause protein malfunction and disease	-5.77	Destabilizing	-2.9070
51.	**Try163Pro**	Pathological	7	Increase	-0.75	Decrease	-2.33	Decrease	89.64%	Destabilizing/ favourable	-1.63	Highly destabilizing/ cause protein malfunction and disease	-3.48	Destabilizing	-1.8553
52.	**Lys164Asp**	Pathological	3	Decrease	-0.73	Decrease	-0.69	Decrease	82.14%	Stabilizing/ Unfavourable	1.18	Neutral/ non-disease associated	0.21	Likely destabilizing	-0.8462
53.	**His168Arg**	Pathological	5	Decrease	0.18	Decrease	-0.44	Increase	27.14%	Destabilizing/ favourable	-0.81	Neutral/ non-disease associated	0.25	Neutral	-0.0925
54.	**Met169Lys**	Pathological	6	Decrease	-1.40	Decrease	-1.69	Decrease	93.04%	Stabilizing/ Favourable	1.85	Destabilizing/ non-disease associated	-1.26	Neutral	-0.2092
55.	**Glu171Pro**	Pathological	3	Increase	-0.08	Decrease	-1.24	Decrease	89.64%	Stabilizing/ Unfavourable	1.03	Highly destabilizing/ cause protein malfunction and disease	-2.60	Neutral	-0.4593
56.	**Val172Asp**	Pathological	8	Decrease	-0.41	Decrease	-3.19	Decrease	92.32%	Destabilizing/ unfavourable	-2.23	Slightly destabilizing/ non-disease associated	-0.62	Destabilizing	-1.7633
57.	**Val173Pro**	Pathological	3	Decrease	-0.98	Decrease	-3.10	Decrease	92.86%	Destabilizing/ favourable	-2.48	Slightly destabilizing/ non-disease associated	-0.71	Destabilizing	-1.5633
58.	**Arg174Asp**	Pathological	9	Decrease	-1.48	Decrease	-1.87	Decrease	91.96%	Destabilizing/ unfavourable	-1.08	Slightly destabilizing/ non-disease associated	-0.75	Likely destabilizing	-0.7489

59.	**Arg175Asp**	Pathological	0	Decrease	-1.88	Decrease	-2.86	Decrease	96.07%	Destabilizing/ unfavourable	-3.67	Destabilizing/ non-disease associated	-1.50	Likely destabilizing	-0.7188
60.	**Cys176Asp**	Pathological	3	Decrease	-1.29	Decrease	-2.64	Decrease	94.11%	Destabilizing/ favourable	-9.06	Highly destabilizing/ cause protein malfunction and disease	-3.13	Destabilizing	-2.2183
61.	**His178Pro**	Pathological	5	Increase	0.11	Increase	0.29	Decrease	85.71%	Destabilizing/ favourable	-0.40	Slightly destabilizing/ non-disease associated	-0.87	Likely destabilizing	-0.5302
62.	**His179Lys**	Pathological	8	Decrease	-0.53	Decrease	-1.23	Decrease	81.07%	Destabilizing/ favourable	-0.10	Highly destabilizing/ cause protein malfunction and disease	-5.22	Destabilizing	-1.4244
63.	**Gly187Arg**	Pathological	7	Decrease	-0.29	Decrease	-0.78	Decrease	78.75%	Stabilizing/ Unfavourable	5.62	Destabilizing/ non-disease associated	-1.53	Neutral	-0.3348
64.	**His193Pro**	Pathological	5	Increase	-0.27	Decrease	-1.23	Decrease	89.64%	Stabilizing/ Favourable	0.48	Slightly destabilizing/ non-disease associated	-0.59	Likely destabilizing	-0.8920
65.	**Leu194Lys**	Pathological	2	Decrease	-2.96	Decrease	-2.72	Decrease	94.11%	Destabilizing/ favourable	-1.87	Highly destabilizing/ cause protein malfunction and disease	-4.02	Destabilizing	-2.0110
66.	**Ile195Asp**	Pathological	7	Decrease	-3.38	Decrease	-3.53	Decrease	92.32%	Destabilizing/ unfavourable	-4.68	Highly destabilizing/ cause protein	-5.77	Destabilizing	-3.0903

											malfunction and disease			
67.	**Arg196Asp**	Pathological	8	Decrease	-1.40	Decrease	-2.89	Decrease	79.82%	Destabilizing/ unfavourable	-12.18	Highly destabilizing/ cause protein malfunction and disease	-2.71	Destabilizing	-2.0120
68.	**Val197Pro**	Pathological	10	Decrease	-2.26	Decrease	-3.84	Decrease	92.68%	Destabilizing/ unfavourable	-6.62	Highly destabilizing/ cause protein malfunction and disease	-4.14	Destabilizing	-1.2888
69.	**Glu198Pro**	Pathological	4	Increase	0.11	Decrease	-1.14	Decrease	83.93%	Stabilizing/ Unfavourable	2.01	Destabilizing/ non-disease associated	-1.32	Neutral	-0.1746
70.	**Gly199Tyr**	Pathological	3	Decrease	-1.27	Decrease	-0.17	Increase	26.25%	Destabilizing/ unfavourable	-5.76	Neutral/ non-disease associated	0.07	Destabilizing	-1.5025
71.	**Arg202Asp**	Pathological	4	Decrease	-1.68	Decrease	-0.70	Decrease	90.00%	Stabilizing/ Favourable	1.47	Neutral/ non-disease associated	0.40	Likely destabilizing	-0.5911
72.	**Val203Pro**	Pathological	3	Decrease	-3.23	Decrease	-3.18	Decrease	92.68%	Destabilizing/ favourable	-1.24	Slightly destabilizing/ non-disease associated	-0.71	Likely destabilizing	-0.8578
73.	**Tyr205Arg**	Pathological	6	Decrease	-0.91	Decrease	-2.07	Decrease	85.89%	Stabilizing/ Unfavourable	2.71	Highly destabilizing/ cause protein malfunction and disease	-2.76	Destabilizing	-2.0620

74.	**Asp207Pro**	Pathological	1	Decrease	-2.44	Decrease	-1.33	Decrease	89.64%	Destabilizing/ unfavourable	-1.31	Neutral/ non-disease associated	-0.17	Neutral	0.0913
75.	**Asp208Pro**	Pathological	6	Decrease	-2.58	Decrease	-2.09	Increase	27.14%	Stabilizing/ Favourable	1.52	Neutral/ non-disease associated	-0.04	Neutral	-0.4333
76.	**Thr211Pro**	Pathological	5	Decrease	-1.50	Decrease	-1.63	Decrease	92.68%	Stabilizing/ Unfavourable	0.87	Slightly destabilizing/ non-disease associated	-0.60	Likely destabilizing	-0.7816
77.	**Phe212Leu**	Pathological	7	Decrease	-2.74	Decrease	-2.29	Decrease	87.86%	Destabilizing/ unfavourable	-1.89	Neutral/ non-disease associated	0.39	Likely destabilizing	-0.8563
78.	**Arg213Glu**	Pathological	2	Decrease	-1.80	Decrease	-1.54	Decrease	83.57%	Destabilizing/ unfavourable	-1.39	Highly destabilizing/ cause protein malfunction and disease	-2.90	Destabilizing	-2.3822
79.	**His214Pro**	Pathological	4	Increase	0.17	Decrease	-0.41	Decrease	91.25%	Destabilizing/ unfavourable	-1.89	Highly destabilizing/ cause protein malfunction and disease	-2.55	Destabilizing	-1.5180
80.	**Ser215Trp**	Pathological	2	Increase	-0.40	Decrease	-0.78	Decrease	81.79%	Stabilizing/ Favourable	3.69	Neutral/ non-disease associated	0.31	Neutral	-0.1427
81.	**Val216Asp**	Pathological	4	Decrease	-2.03	Decrease	-3.09	Decrease	91.96%	Destabilizing/ unfavourable	-4.30	Highly destabilizing/ cause protein malfunction and disease	-5.07	Destabilizing	-2.8681

82.	**Val218Arg**	Pathological	9	Decrease	-2.11	Decrease	-3.68	Decrease	90.71%	Stabilizing/ Favourable	3.51	Highly destabilizing/ cause protein malfunction and disease	-3.34	Destabilizing	-2.2677
83.	**Pro219Asp**	Pathological	6	Decrease	-1.93	Decrease	-2.02	Decrease	86.43%	Destabilizing/ unfavourable	-1.40	Slightly stabilizing/ non-disease associated	0.92	Destabilizing	-1.9152
84.	**Tyr220Lys**	Pathological	5	Increase	-0.18	Decrease	-1.77	Decrease	89.64%	Destabilizing/ favourable	-2.08	Highly destabilizing/ cause protein malfunction and disease	-3.21	Destabilizing	-1.4987
85.	**Glu221Pro**	Pathological	4	Decrease	-0.31	Decrease	-0.40	Decrease	84.82%	Destabilizing/ unfavourable	-1.23	Highly destabilizing/ cause protein malfunction and disease	-2.60	Neutral	-0.1389
86.	**Gly226Lys**	Pathological	6	Decrease	-0.55	Decrease	-0.56	Decrease	79.82%	Destabilizing/ unfavourable	-0.37	Destabilizing/ non-disease associated	-1.27	Neutral	-0.3388
87.	**Asp228Pro**	Pathological	1	Decrease	-0.13	Increase	-0.33	Decrease	86.61%	Stabilizing/ Unfavourable	1.21	Neutral/ non-disease associated	-0.38	Neutral	-0.0417
88.	**Thr230Asp**	Pathological	6	Decrease	-0.15	Decrease	0.17	Decrease	86.96%	Destabilizing/ unfavourable	-4.33	Destabilizing/ non-disease associated	-1.17	Destabilizing	-1.6739
89.	**Thr231Arg**	Pathological	6	Decrease	0.59	Decrease	0.83	Decrease	81.43%	Stabilizing/ Unfavourable	3.00	Slightly stabilizing/ non-disease associated	0.58	Neutral	-0.4285

90.	**Ile232Pro**	Pathological	4	Decrease	-2.21	Decrease	-2.92	Decrease	93.39%	Destabilizing/ unfavourable	-3.1	Highly destabilizing/ cause protein malfunction and disease	-4.88	Destabilizing	-1.6938
91.	**Tyr234Pro**	Pathological	7	Decrease	-1.52	Decrease	-3.09	Decrease	92.86%	Destabilizing/ unfavourable	-2.03	Highly destabilizing/ cause protein malfunction and disease	-3.48	Destabilizing	-1.8384
92.	**Asn235Pro**	Pathological	8	Increase	-0.43	Decrease	-1.87	Decrease	83.75%	Stabilizing/ Unfavourable	3.45	Sightly stabilizing/ non-disease associated	0.90	Neutral	-0.1330
93.	**Tyr236Pro**	Pathological	4	Decrease	-1.53	Decrease	-2.44	Decrease	92.32%	Destabilizing/ unfavourable	-7.04	Highly destabilizing/ cause protein malfunction and disease	-3.48	Destabilizing	-1.3574
94.	**Met237Pro**	Pathological	5	Decrease	-1.64	Decrease	-2.32	Decrease	81.61%	Stabilizing/ Unfavourable	2.76	Destabilizing/ non-disease associated	-1.44	Likely destabilizing	-0.7546
95.	**Cys238Pro**	Pathological	5	Decrease	-1.20	Decrease	-3.32	Decrease	92.86%	Destabilizing/ unfavourable	-8.34	Highly destabilizing/ cause protein malfunction and disease	-6.00	Destabilizing	-1.5721
96.	**Asn239Pro**	Pathological	7	Increase	-0.42	Decrease	-1.43	Decrease	78.21%	Stabilizing/ Favourable	1.94	Stabilizing/ non-disease associated	1.84	Neutral	-0.3435
97.	**Ser240Pro**	Pathological	5	Increase	-0.59	Decrease	-1.63	Decrease	85.54%	Destabilizing/ favourable	-2.20	Slightly stabilizing/ non-disease associated	0.67	Likely destabilizing	-0.5194

98.	**Ser241Pro**	Pathological	6	Increase	-0.41	Decrease	-1.28	Decrease	92.86%	Destabilizing/ unfavourable	-0.69	Stabilizing/ non-disease associated	1.09	Neutral	-0.3662
99.	**Cys242Lys**	Pathological	7	Decrease	-1.60	Decrease	-2.36	Decrease	94.64%	Destabilizing/ Favourable	-13.11	Slightly destabilizing/ non-disease associated	-0.75	Destabilizing	-1.7499
100.	**Met243Asp**	Pathological	8	Decrease	-0.20	Decrease	0.03	Decrease	79.64%	Destabilizing/ unfavourable	12.99	Destabilizing/ non-disease associated	-1.23	Neutral	-0.4069
101.	**Gly244Lys**	Pathological	6	Decrease	-0.19	Decrease	-0.47	Decrease	86.43%	Destabilizing/ unfavourable	-3.14	Highly destabilizing/ cause protein malfunction and disease	-3.32	Likely destabilizing	-0.6232
102.	**Gly245Pro**	Pathological	5	Decrease	-0.02	Decrease	-1.20	Decrease	87.68%	Stabilizing/ Unfavourable	2.59	Highly destabilizing/ cause protein malfunction and disease	-2.52	Neutral	-0.4083
103.	**Met246Asp**	Pathological	8	Decrease	-0.35	Decrease	-0.31	Decrease	92.32%	Destabilizing/ No change	-3.06	Highly destabilizing/ cause protein malfunction and disease	-2.49	Destabilizing	-1.1998
104.	**Arg249Asp**	Pathological	7	Decrease	-0.62	Decrease	-1.96	Decrease	94.46%	Destabilizing/ unfavourable	-1.68	Destabilizing/ non-disease associated	-1.22	Likely destabilizing	-0.9671
105.	**Ile251Glu**	Pathological	1	Decrease	-1.50	Decrease	-3.41	Decrease	92.86%	Destabilizing/ unfavourable	-0.81	Highly destabilizing/ cause protein	-4.28	Destabilizing	-2.7358

											malfunction and disease			
106.	**Leu252Pro**	Pathological	8	Decrease	-1.24	Decrease	-2.31	Decrease	92.32%	Destabilizing/ unfavourable	-6.17	Highly destabilizing/ cause protein malfunction and disease	-3.53	Destabilizing	-1.4938
107.	**Thr253Pro**	Pathological	6	Decrease	-1.78	Decrease	-3.25	Decrease	81.25%	Destabilizing/ unfavourable	-5.64	Destabilizing/ non-disease associated	-1.57	Likely destabilizing	-0.6124
108.	**Ile254Arg**	Pathological	4	Decrease	-1.64	Decrease	-3.83	Decrease	90.71%	Stabilizing/ Unfavourable	1.69	Highly destabilizing/ cause protein malfunction and disease	-2.86	Destabilizing	-2.3146
109.	**Ile255Pro**	Pathological	4	Decrease	-2.20	Decrease	-4.45	Decrease	84.11%	Destabilizing/ unfavourable	-6.64	Highly destabilizing/ cause protein malfunction and disease	-5.35	Destabilizing	-1.9279
110.	**Thr256Lys**	Pathological	7	Decrease	-1.35	Decrease	-2.17	Decrease	83.57%	Destabilizing/ unfavourable	-0.28	Neutral/ non-disease associated	-0.39	Destabilizing	-1.8774
111.	**Leu257Asp**	Pathological	8	Decrease	-2.29	Decrease	-2.85	Decrease	92.86%	Destabilizing/ Favourable	-4.24	Highly destabilizing/ cause protein malfunction and disease	-4.17	Destabilizing	-3.2793
112.	**Glu258Pro**	Pathological	6	Increase	0.22	Decrease	-1.82	Decrease	91.79%	Stabilizing/ unfavourable	4.11	Neutral/ non-disease associated	-0.49	Neutral	0.0348

113.	**Asp259Trp**	Pathological	1	Decrease	-1.28	Decrease	-1.11	Increase	27.14%	Destabilizing/ unfavourable	-0.27	Highly stabilizing/ cause protein malfunction and disease	3.67	Likely stabilizing	0.7070
114.	**Gly262Pro**	Pathological	7	Decrease	-1.94	Decrease	-0.87	Decrease	82.32%	Destabilizing/ unfavourable	-3.68	Highly destabilizing/ cause protein malfunction and disease	-6.46	Likely destabilizing	-0.8149
115.	**Leu264Asp**	Pathological	3	Decrease	-2.76	Decrease	-1.67	Decrease	85.71%	Stabilizing/ unfavourable	0.27	Highly destabilizing/ cause protein malfunction and disease	-2.57	Destabilizing	-2.1642
116.	**Leu265Asp**	Pathological	8	Decrease	-3.03	Decrease	-1.68	Decrease	92.68%	Destabilizing/ unfavourable	-0.25	Highly destabilizing/ cause protein malfunction and disease	-2.07	Destabilizing	-2.6571
117.	**Gly266Asp**	Pathological	7	Decrease	-0.75	Decrease	-1.74	Decrease	79.29%	Destabilizing/ unfavourable	-4.38	Destabilizing/ non-disease associated	-1.57	Destabilizing	-1.1571
118.	**Arg267Asp**	Pathological	8	Decrease	-1.39	Decrease	-1.12	Decrease	85.54%	Destabilizing/ unfavourable	-0.89	Destabilizing/ non-disease associated	-1.30	Destabilizing	-1.9756
119.	**Ser269Trp**	Pathological	8	Decrease	-0.81	Decrease	-0.05	Decrease	84.46%	Stabilizing/ Favourable	1.93	Stabilizing/ non-disease associated	1.06	Neutral	-0.0397
120.	**Phe270Asp**	Pathological	8	Decrease	-2.10	Decrease	-3.51	Decrease	89.64%	Destabilizing/ unfavourable	-5.97	Highly destabilizing/ cause protein	-4.50	Destabilizing	-3.3540

											malfunction and disease			
121.	**Glu271Arg**	Pathological	1	Decrease	-0.48	Decrease	-1.17	Decrease	79.64%	Destabilizing/ unfavourable	-0.78	Neutral/ non-disease associated	-0.30	Neutral	-0.3740
122.	**Val272Pro**	Pathological	2	Decrease	-2.32	Decrease	-3.22	Decrease	92.68%	Destabilizing/ unfavourable	-11.16	Highly destabilizing/ cause protein malfunction and disease	-4.14	Destabilizing	-1.8374
123.	**Arg273Tyr**	Pathological	2	Decrease	-0.37	Decrease	-0.57	Decrease	94.11%	Destabilizing/ Favourable	-0.21	Slightly stabilizing/ non-disease associated	0.61	Likely destabilizing	-0.5083
124.	**Val274Pro**	Pathological	3	Decrease	-1.76	Decrease	-2.15	Decrease	91.79%	Destabilizing/ unfavourable	-1.15	Highly destabilizing/ cause protein malfunction and disease	-4.14	Destabilizing	-1.7899
125.	**Cys275Pro**	Pathological	3	Decrease	-1.48	Decrease	-2.36	Decrease	83.39%	Destabilizing/ unfavourable	-10.06	Highly destabilizing/ cause protein malfunction and disease	-5.72	Destabilizing	-1.1013
126.	**Ala276Asp**	Pathological	9	Decrease	-0.39	Decrease	-0.40	Decrease	84.11%	Stabilizing/ Unfavourable	4.53	Neutral/ non-disease associated	-0.18	Likely destabilizing	-0.8474
127.	**Cys277Pro**	Pathological	1	Decrease	-1.46	Decrease	-1.01	Decrease	91.79%	Destabilizing/ unfavourable	-3.95	Highly destabilizing/ cause protein malfunction and disease	-2.92	Destabilizing	-1.7457

128.	**Pro278Asp**	Pathological	6	Decrease	-1.71	Decrease	-3.35	Decrease	87.86%	Stabilizing/ Unfavourable	1.93	Destabilizing/ non-disease associated	-1.23	Destabilizing	-1.1266
129.	**Arg280Asp**	Pathological	6	Decrease	-1.61	Decrease	-0.29	Decrease	85.71%	Destabilizing/ unfavourable	-2.01	Slightly stabilizing/ non-disease associated	-0.88	Likely destabilizing	-0.6532
130.	**Asp281Trp**	Pathological	6	Decrease	-1.06	Decrease	-2.81	Decrease	77.68%	Destabilizing/ Favourable	-1.17	Highly stabilizing/ cause protein malfunction and disease	2.09	Neutral	0.1607
131.	**Arg282Asp**	Pathological	6	Decrease	-1.43	Decrease	-1.86	Decrease	88.75%	Destabilizing/ unfavourable	-0.45	Neutral/ non-disease associated	-0.40	Likely destabilizing	-0.7705
132.	**Glu286Pro**	Pathological	3	Increase	-1.03	Increase	-0.80	Increase	23.21%	Destabilizing/ unfavourable	-0.86	Highly destabilizing/ cause protein malfunction and disease	-2.31	Destabilizing	-1.3264
133.	**Arg290Trp**	Pathological		Decrease	-0.71	-	-	Decrease	81.07%	-	-	-	-	Likely destabilizing	-0.8970
134.	**Arg306Asp**	Pathological		Decrease	-1.12	-	-	Decrease	86.07%	-	-	-	-	Neutral	-0.2038
135.	**Glu326Pro**	Pathological		Increase	-1.36	-	-	Decrease	90.71%	-	-	-	-	Neutral	-0.4447
136.	**Phe328Pro**	Pathological		Decrease	-2.65	-	-	Decrease	92.86%	-	-	-	-	Destabilizing	-1.9358
137.	**Leu330Pro**	Pathological		Decrease	-1.69	-	-	Decrease	83.57%	-	-	-	-	Destabilizing	-2.4097
138.	**Ile332Pro**	Pathological		Decrease	-2.64	-	-	Decrease	94.64%	-	-	-	-	Destabilizing	-2.3669

139.	**Arg333Asp**	Pathological		Decrease	-0.87	-	-	Decrease	88.04%	-	-	-	-	Destabilizing	-1.7324
140.	**Gly334Arg**	Pathological		Decrease	-0.26	-	-	Decrease	82.32%	-	-	-	-	Likely destabilizing	-0.7521
141.	**Arg335Asp**	Pathological		Decrease	-0.76	-	--	Decrease	85.89%	-	-	-	-	Likely destabilizing	-0.6098
142.	**Glu336Pro**	Pathological		Increase	-0.33	-	-	Decrease	80.54%	-	-	-	-	Destabilizing	-1.3619
143.	**Arg337Asp**	Pathological		Decrease	-0.88	-	-	Decrease	93.04%	-	-	-	-	Destabilizing	-1.6760
144.	**Phe338Pro**	Pathological		Decrease	-2.05	-	-	Decrease	91.96%	-	-	-	-	Destabilizing	-1.7103
145.	**Glu339Pro**	Pathological		Increase	-0.01	-	-	Decrease	88.75%	-	-	-	-	Destabilizing	-1.4218
146.	**Phe341Asp**	Pathological		Decrease	-2.40	-	-	Decrease	91.79%	-	-	-	-	Destabilizing	-3.1838
147.	**Leu344Asp**	Pathological		Decrease	-4.43	-	-	Decrease	94.11%	-	-	-	-	Destabilizing	-2.4357
148.	**Glu346Pro**	Pathological		Increase	-0.39	-	-	Decrease	78.04%	-	-	-	-	Destabilizing	-1.4705
149.	**Ala347Pro**	Pathological		Decrease	-2.59	-	-	Decrease	85.89%	-	-	-	-	Likely destabilizing	-0.8952
150.	**Leu348Pro**	Pathological		Decrease	-2.39	-	-	Decrease	93.93%	-	-	-	-	Destabilizing	-1.8585
151.	**Glu349Pro**	Pathological		Increase	-0.47	-	-	Decrease	77.68%	-	-	-	-	Destabilizing	-1.4719
152.	**Leu350Arg**	Pathological		Decrease	-1.82	-	-	Decrease	91.79%	-	-	-	-	Neutral	0.1400
153.	**Lys381Asp**	Pathological		Increase	0.57	-	-	Decrease	89.64%	-	-	-	-	Likely destabilizing	-0.5053
154.	**Lys382Asp**	Pathological		Increase	0.17	-	-	Decrease	82.32%	-	-	-	-	Likely destabilizing	-0.5299
155.	**Leu383Asp**	Pathological		Decrease	-2.02	-	-	Decrease	92.86%	-	-	-	-	Destabilizing	-1.2648
156.	**Lys386Asp**	Pathological		Decrease	-0.54	-	-	Decrease	86.07%	-	-	-	-	Likely destabilizing	-0.6634

Table S5: Experimentally predicted deleterious mutations and the stability of the resultant mutants in p53 protein*

S. No.	Amino acid Mutation	PMut [a] Predictions	Reliability Index (RI)[c]	Protein Stability											
				I-MUTANT 2.0 [d] (Sequence based)		I-MUTANT 2.0 (Structure based)		MuStab server [e]		CUPSAT server [f]		SDM server [g]		EASY-MM	
				Predictions (Stability)	ΔΔG	Predictions (Stability)	ΔΔG	Predictions (Stability)	confidence	Predictions (Overall stability/torsions)	ΔΔG	Predictions (Stability/effect)	ΔΔG		ΔΔG
1.	**Tyr126Cys**	Pathological	3	Increase	0.43	Decrease	-1.82	Decrease	90.36%	Stabilizing/ Favourable	5.69	Highly Stabilizing/ cause protein malfunction and disease	2.87	Destabilizing	-1.7847
2.	**Cys135Tyr**	Pathological	3	Decrease	-0.78	Decrease	-1.87	Decrease	78.04%	Destabilizing/ Favourable	-5.01	Highly destabilizing/ cause protein malfunction and disease	-3.91	Destabilizing	-1.1986
3.	**Val143Ala**	Pathological	7	Decrease	-0.88	Decrease	-3.82	Decrease	92.32%	Destabilizing/ Unfavourable	-6.26	Highly destabilizing/ cause protein malfunction and disease	-2.73	Destabilizing	-2.1709
4.	**Pro151Ser**	Pathological	4	Decrease	-1.26	Decrease	-1.02	Decrease	78.75%	Destabilizing/ Favourable	-3.06	Destabilizing/ non-disease associated	-1.92	Destabilizing	-1.5070

5.	**Arg158His**	Pathological	8	Decrease	-0.98	Decrease	-2.95	Decrease	78.75%	Destabilizing/ Unfavourable	-1.55	Stabilizing/ non-disease associated	1.10	Destabilizing	-1.5071
6.	**Tyr163Asn**	Pathological	3	Decrease	-0.86	Decrease	-2.72	Decrease	91.79%	Destabilizing/ Unfavourable	-4.13	Highly destabilizing/ cause protein malfunction and disease	-4.75	Destabilizing	-1.8971
7.	**Try163Cys**	Pathological	2	Decrease	-0.95	Decrease	-1.29	Decrease	83.57%	Destabilizing/ Unfavourable	-0.79	Stabilizing/ non-disease associated	1.65	Destabilizing	-1.5028
8.	**His168Tyr**	Pathological	2	Decrease	0.30	Increase	0.58	Increase	30.36%	Destabilizing/ Favourable	-0.17	Neutral/non-disease associated	0.44	Neutral	-0.3005
9.	**Val173Leu**	Pathological	5	Decrease	-0.10	Decrease	-0.86	Decrease	91.79%	Destabilizing/ Favourable	-3.26	Neutral/non-disease associated	0.35	Neutral	-0.0807
10.	**Arg175His**	Pathological	7	Decrease	-1.35	Decrease	-2.78	Decrease	99.11%	Destabilizing/ Unfavourable	-3.45	Neutral/non-disease associated	0.26	Neutral	-0.2842
11.	**Arg175Leu**	Pathological	5	Increase	1.01	Decrease	1.03	Decrease	96.43%	Destabilizing/ Unfavourable	-4.20	Neutral/non-disease associated	0.17	Likely destabilizing	-0.6803
12.	**Arg175Cys**	Pathological	2	Increase	-0.65	Decrease	-1.49	Decrease	96.43%	Stabilizing/ Unfavourable	6.14	Highly destabilizing/ cause protein malfunction and disease	-3.66	Neutral	0.3709
13.	**Arg175Pro**	Pathological	4	Increase	-0.57	Decrease	-2.13	Decrease	94.11%	Stabilizing/ Favourable	0.18	Destabilizing/ non-disease associated	-1.01	Neutral	0.1250
14.	**Cys176Phe**	Pathological	4	Increase	1.20	Decrease	-1.60	Decrease	78.75%	Destabilizing/ Unfavourable	-9.12	Highly destabilizing/ cause protein	-2.14	Destabilizing	-1.4097

											malfunction and disease			
15.	**His179Leu**	Pathological	3	Decrease	-0.99	Decrease	-0.47	Increase	25.00%	Stabilizing/ Favourable	0.93	Stabilizing/ non-disease associated	1.64	Likely destabilizing	-0.9441
16.	**Leu194Arg**	Pathological	0	Increase	0.47	Decrease	-2.73	Decrease	91.79%	Destabilizing/ Favourable	-4.47	Highly destabilizing/ cause protein malfunction and disease	-2.32	Destabilizing	-1.1871
17.	**Arg213Gly**	Pathological	9	Decrease	-3.19	Decrease	-2.57	Decrease	87.86%	Destabilizing/ Unfavourable	-6.11	Highly destabilizing/ cause protein malfunction and disease	-5.53	Destabilizing	-2.9207
18.	**His214Arg**	Pathological	0	Decrease	0.04	Decrease	-0.88	Decrease	81.61%	Stabilizing/ No change	2.12	Destabilizing/ non-disease associated	-1.09	Destabilizing	-1.1914
19.	**Tyr220Ser**	Pathological	4	Decrease	-1.13	Decrease	-2.00	Decrease	88.93%	Destabilizing/ Unfavourable	-1.78	Highly destabilizing/ cause protein malfunction and disease	-3.04	Destabilizing	-2.3839
20.	**Pro223Ser**	Pathological	2	Decrease	0.04	Decrease	-1.40	Decrease	88.39%	Destabilizing/ Unfavourable	-0.52	Destabilizing/ non-disease associated	-1.02	Destabilizing	-1.0309
21.	**Try234Cys**	Pathological	8	Decrease	-1.31	Decrease	-2.60	Decrease	92.32%	Destabilizing/ Favourable	-2.25	Slightly stabilizing/ non-disease associated	0.98	Likely destabilizing	-1.6749
22.	**Asn239Ser**	Pathological	9	Decrease	-3.63	Decrease	-1.75	Increase	26.96%	Stabilizing/ Favourable	1.34	Neutral/non-disease associated	0.00	Likely destabilizing	-0.6646

Appendices

23.	**Gly245Ser**	Pathological	9	Decrease	-3.83	Decrease	-1.18	Decrease	88.21%	Stabilizing/ unfavourable	2.49	Stabilizing/ non-disease associated	1.50	Likely destabilizing	-0.5741
24.	**Gly245Asp**	Pathological	8	Decrease	-4.12	Decrease	-1.75	Decrease	88.93%	Destabilizing/ Unfavourable	-1.95	Slightly stabilizing/ non-disease associated	0.70	Likely destabilizing	-0.5775
25.	**Gly245Cys**	Pathological	8	Decrease	-3.63	Decrease	-1.00	Decrease	88.57%	Stabilizing/ Favourable	7.13	Highly stabilizing/ cause protein malfunction and disease.	3.73	Likely destabilizing	-0.5102
26.	**Gly245Val**	Pathological	8	Decrease	-1.79	Decrease	-1.24	Decrease	84.82%	Destabilizing/ Unfavourable	-5.38	Highly stabilizing/ cause protein malfunction and disease	3.25	Likely destabilizing	-0.7145
27.	**Met246Ile**	Pathological	8	Decrease	-0.71	Decrease	-1.86	Decrease	90.36%	Destabilizing/ Unfavourable	-5.14	Destabilizing/ non-disease associated	-1.33	Neutral	-0.1573
28.	**Arg248Gln**	Pathological	5	Decrease	-1.34	Decrease	-1.68	Decrease	97.50%	Destabilizing/ Unfavourable	-3.17	Destabilizing/ non-disease associated	-1.59	Neutral	0.2885
29.	**Arg248Trp**	Pathological	6	Decrease	-0.37	Decrease	-0.63	Decrease	90.71%	Destabilizing/ Unfavourable	-31.1	Slightly stabilizing/ non-disease associated	0.82	Likely stabilizing	0.8963
30.	**Arg248Gly**	Pathological	5	Decrease	-1.15	Decrease	-1.28	Decrease	91.96%	Destabilizing/ Favourable	-0.73	Stabilizing/ non-disease associated	1.54	Neutral	-0.2703

31.	**Arg248Pro**	Pathological	1	Increase	-0.80	Decrease	-0.97	Decrease	92.50%	Destabilizing/ Unfavourable	-5.91	Highly destabilizing/ cause protein malfunction and disease	-9.22	Neutral	0.0260
32.	**Arg249Ser**	Pathological	9	Decrease	-1.96	Decrease	-1.85	Decrease	98.21%	Destabilizing/ Unfavourable	-3.46	Destabilizing/ non-disease associated	-1.06	Neutral	-0.4084
33.	**Arg249Thr**	Pathological	8	Decrease	-1.27	Decrease	-1.72	Decrease	96.43%	Destabilizing/ Unfavourable	-2.49	Slightly stabilizing/ non-disease associated	0.83	Neutral	-0.2405
34.	**Gly266Glu**	Pathological	2	Decrease	-1.16	Decrease	-0.96	Decrease	81.25%	Destabilizing/ Unfavourable	-3.52	Destabilizing/ non-disease associated	-1.26	Destabilizing	1.3588
35.	**Arg273His**	Pathological	7	Decrease	-1.05	Decrease	-0.59	Decrease	97.68%	Stabilizing/ Favourable	0.10	Neutral/non-disease associated	-0.47	Likely destabilizing	-0.9030
36.	**Arg273Leu**	Pathological	4	Decrease	-0.83	Decrease	-0.51	Decrease	92.50%	Stabilizing/ Favourable	0.37	Slightly stabilizing/ non-disease associated	0.63	Likely destabilizing	-0.6260
37.	**Val274Phe**	Pathological	1	Decrease	1.25	Decrease	-1.60	Decrease	88.57%	Destabilizing/ Favourable	-3.76	Destabilizing/ non-disease associated	-1.05	Destabilizing	-1.2751
38.	**Pro278Ser**	Pathological	7	Decrease	-2.00	Decrease	-2.76	Decrease	83.57%	Destabilizing/ Unfavourable	-1.28	Slightly destabilizing/ non-disease associated	-0.63	Destabilizing	1.1614
39.	**Arg280Lys**	Pathological	4	Decrease	-0.84	Decrease	0.13	Decrease	83.57%	Stabilizing/ Favourable	0.38	Neutral/non-disease associated	-0.23	Neutral	-0.2030

40.	**Arg280Thr**	Pathological	6	Decrease	-0.86	Decrease	0.73	Decrease	88.75%	Destabilizing/ Unfavourable	-0.73	Destabilizing/ non-disease associated	-1.82	Neutral	-0.0679
41.	**Asp281Gly**	Pathological	9	Decrease	-3.22	Decrease	-3.12	Decrease	85.89%	Destabilizing/ Unfavourable	-1.28	Destabilizing/ non-disease associated	-1.92	Neutral	-0.0893
42.	**Arg282Trp**	Pathological	5	Decrease	-0.73	Decrease	-0.32	Increase	29.11%	Destabilizing/ Favourable	-0.36	Highly stabilizing/ cause protein malfunction and disease.	3.50	Likely destabilizing	-0.7958
43.	**Arg282Gln**	Pathological	7	Decrease	-1.19	Decrease	-1.60	Decrease	82.86%	Destabilizing/ Favourable	-1.24	Slightly stabilizing/ non-disease associated	0.73	Neutral	-0.2567
44.	**Pro309Ser**	Pathological	-	-	-	-	-	-	-	-	-	-	-	-	-
45.	**Arg337His**	Pathological	-	--	-	-	-	-		-	-	-	-		

***Experimentally validated mutations**

a-http://sift.jcvi.org/

b-http://mmb2.pcb.ub.es:8080/PMut/

RI value (≥5) is considered as significant hit

http://folding.biofold.org/i-mutant/i-mutant2.0.html

http://bioinfo.ggc.org/mustab/

http://cupsat.tu-bs.de/

http://mordred.bioc.cam.ac.uk/~sdm/sdm.php

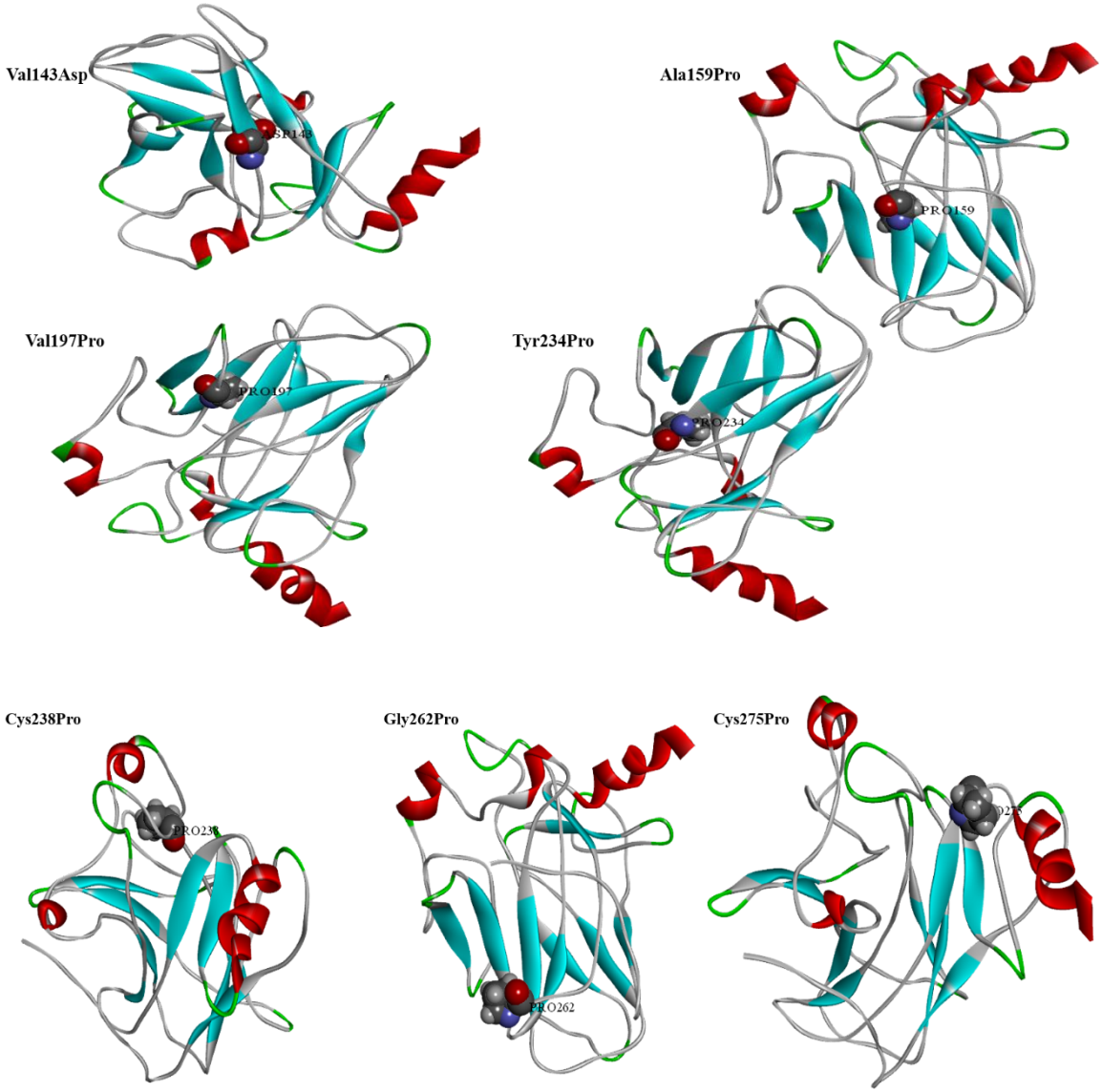
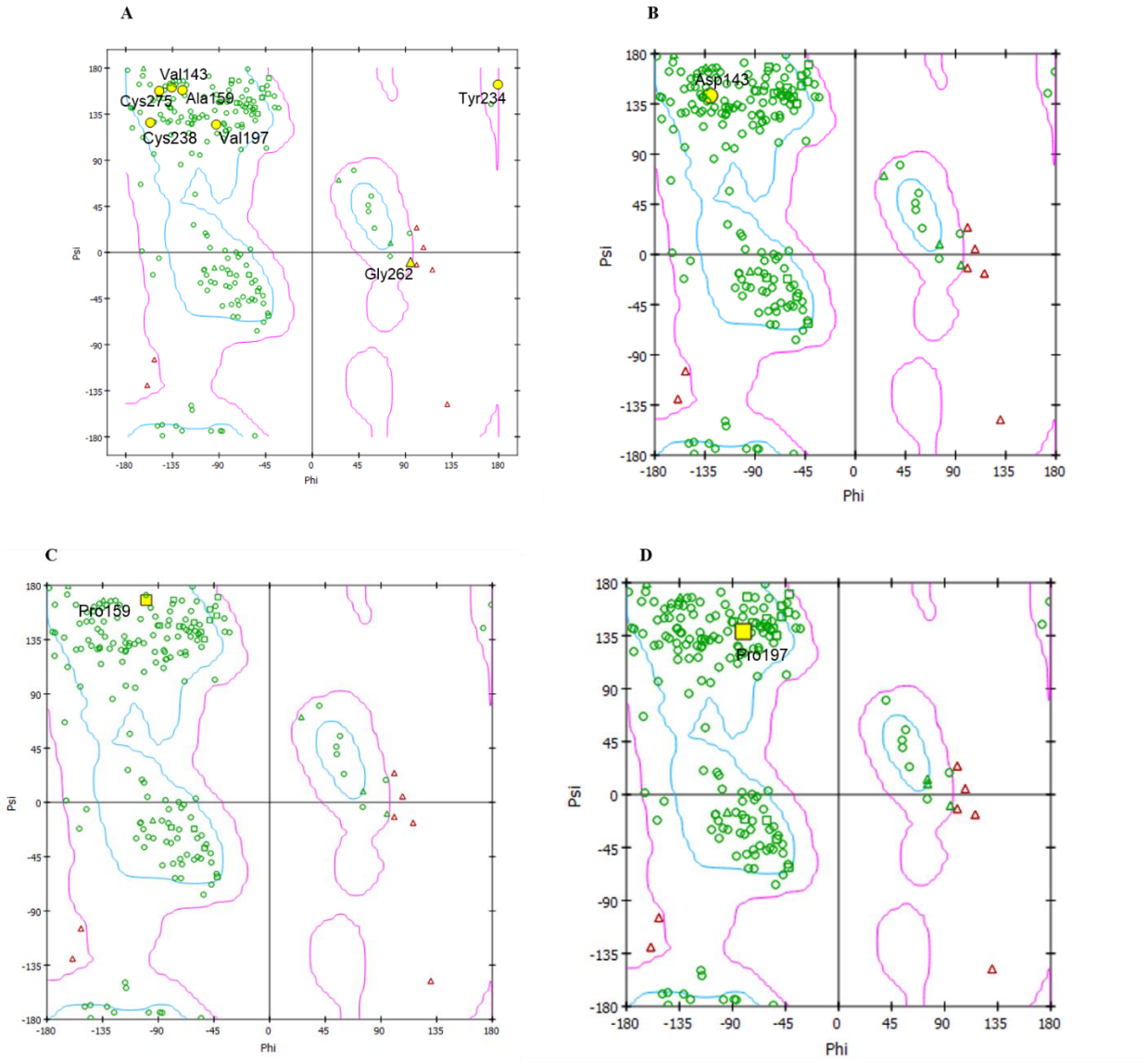


Figure S2: Predicted stable p53 mutants modelled in Accelrys DS 2016.

Appendix II


A
Val143
Cys275
Ala159
Cys238
Val197
Tyr234
Gly262
B
Asp143
C
Pro159
D
Pro197
Psi
Phi

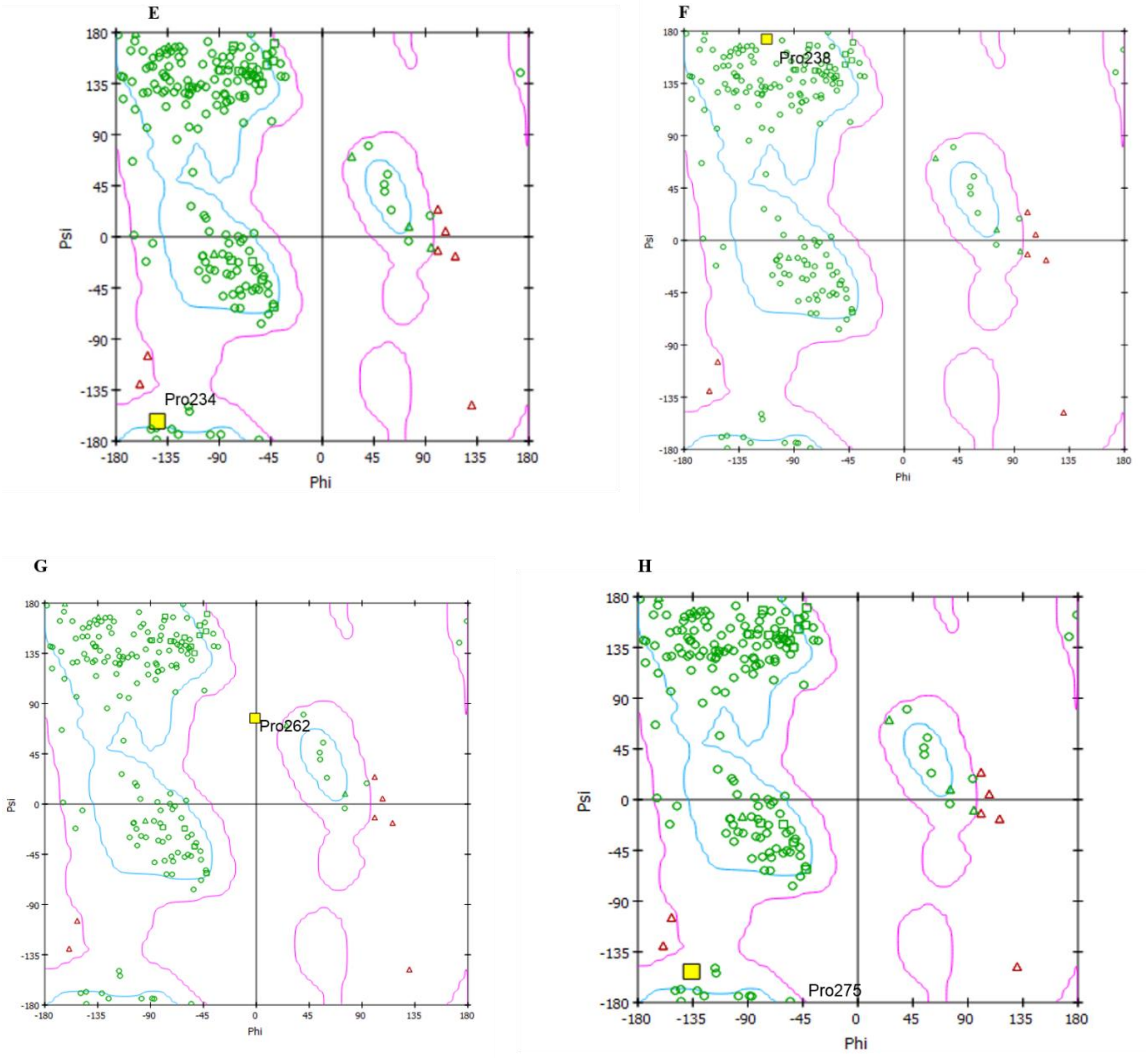


Figure S3: Ramachandran plots of the wild-type and predicted p53 mutants modelled structures.

Table S6: Average values of RMSD, Rg, SASA and total energy of the system for p53 wild-type and its experimental mutant forms (Discussed in chapter 5)

Parameters	**Native**	**P53 Mutations**		
		Val134Ala	**Arg213Gly**	**Gly245Ser**
RMSD (nm)	0.618	0.617	0.617	0.617
Rg (nm)	1.27	1.23	1.22	1.17
SASA (nm^2)	79.29	76.14	77.53	78.42
Total energy (kJ/mol)	-426748	-426587.63	-428964.36	-429175.14