UNIVERSITY OF SOUTHAMPTON

FACULTY OF NATURAL AND ENVIRONMENTAL SCIENCES

Chemistry

&

BIOINFORMATICS INSTITUTE

A*STAR Singapore

Simulating the role of conformational stabilization in p53 and its effect on oncogenesis

by

Zahra Ouaray

Thesis for the degree of Doctor of Philosophy

March 2015

Supervisor: Prof. Jonathan Essex

A*Star Supervisor: Prof. Chandra Verma

Advisor: Dr. Chris-Kriton Skylaris

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF NATURAL & ENVIRONMENTAL SCIENCES

SCHOOL OF CHEMISTRY

&

BIOINFORMATICS INSTITUTE

A*STAR Singapore

Thesis for the degree of Doctor of Philosophy

SIMULATING THE ROLE OF CONFORMATIONAL STABILIZATION IN p53 AND ITS EFFECT ON ONCOGENESIS

By Zahra Ouaray

The p53 tumour suppressor protein functions as a transcription factor to prevent cancer by inducing cell cycle arrest, DNA repair and apoptosis. 50% of cancers are associated with a mutation on this protein; indeed mutated p53 either loses its capacity to bind DNA or is destabilised which induces the proliferation of DNA-damaged cells. 95% of these mutations occur on the p53 DNA-binding domain (DBD). In the present study, the use of computational methods and structural analysis allowed us to investigate p53DBD dynamics at an atomistic level. Principal component analysis of p53DBD experimental structures (NMR and Xray), as well as of the DNA sequences to which these were bound, confirmed the role of the tetramerisation domain in p53 DNA binding specificity and contradicted the belief that the DNA-bound p53DBD and the apo p53DBD can present similar conformations. Conventional and accelerated molecular dynamics (cMD and aMD) were used to investigate the role of p53DBD cysteine alkylation by Michael acceptors (MAs) and the destabilisation of a p53DBD temperature sensitive mutant (V143A). Both techniques highlighted a stable p53DBD dimerization interface (H1 α-helix) as being implicated in the mutant rescue process. While further investigations are needed to confirm our conclusions, these early results might help to design more specific anti-cancer drugs.

Table of Contents

List of tables

List of figures

DECLARATION OF AUTHORSHIP

I, Zahra Ouaray .....

declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

Simulating the role of conformational stabilization in p53 and its effect on oncogenesis.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;

6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;

7. [Delete as appropriate] None of this work has been published before submission [or] Parts of this work have been published as: [please list references below]:

Signed: .....

Date: .....

Acknowledgements

I would like to thank my supervisors Prof. Jonathan Essex and Prof. Chandra Verma for their help, useful insights and patience throughout this project.

I would also like to express my gratitude to the members of Essex's and Verma's groups, with a special mention to Chris Bull for his help with the dimensional reduction methods, to Thomas Joseph for his guidance throughout Amber parameterization of NEM, and to Ragav Kannan for his advice on NAMD. I am also thankful to Dr. Karim M. ElSawy, for his help on DNA PCA.

Special thanks to my colleagues and friends, Michael Carter, Barbara Sander, Ioannis Haldoupis, Alvaro Ruiz-Serrano and Chris Sampson for their support and the enjoyable work environment.

Finally I am thankful to my family and loved ones, to my parents, my source of inspiration and motivation and to Francesco for his loving support and encouragement that helped me during this memorable experience.

Chapter 1. Introduction

Proteins are essential parts of organisms and are involved in virtually every process within cells. Each protein within the body has a specific function. They function as catalysts, transport and store other molecules, provide mechanical support and immune protection, generate movement, transmit nerve impulses, and control growth and differentiation (1). A single protein can be also engage in different functions. Proteins vary in structure as well as function. Their construction is based on a set of 20 amino acids (primary structure) which are organised in secondary structures which then adopt distinct three dimensional shapes (tertiary or quaternary structures). Proteins are not rigid; they usually shift between several related three dimensional shapes or conformations while they perform their functions. The study of the tertiary structure of a protein, or the quaternary structure of its complexes, can provide important clues about how the protein performs its function. The activities and structures of proteins are examined both *in vitro* and *in vivo*. The *in vitro* studies (X-ray crystallography, NMR, mass spectroscopy, molecular dynamics) of proteins in controlled environments are useful for learning how a protein carries out its function. In contrast, *in vivo* experiments (for example, site directed mutagenesis or immune-electron microscopy, which use antibodies labelled with heavy metal particles, directly visualised using transmission electron microscopy) on proteins within cells or the whole organisms can provide complementary information about where in a cell a protein carries out its functions and how it is regulated.

There are several diseases in which certain proteins become structurally abnormal, and thereby lead to the disruption of the functions of cells, tissues and organs of the body. Frequently the proteins fail to fold into their normal configuration; in this misfolded state, the proteins can become toxic or they can lose normal function. These diseases include prion diseases, type 2 diabetes, Alzheimer's disease, and a wide range of other disorders. In most of these diseases a change in the 3-dimensional fold increases the tendency of a specific protein to aggregate. In other cases this induces a loss of protein's

usual functions or enhances their functions that can lead to uncontrollable signalling such as is frequently seen in cancers. Formation of abnormal structures can originate in different factors, including destabilizing changes in the primary amino acid sequence of the protein, post-translational modifications, changes in temperature or pH, an increase in the production of the protein, or a decrease in its clearance (1). It is now well known that the structural changes, chemical pathways and the thermodynamic character of cellular and physiological reactions cannot be fully characterized at the atomic level by experimentation alone (2).

The increase in computational power and the development of new theoretical methods play an important role in the analysis of protein function. Computational calculations attempt to associate motion with mechanism and characterize detailed kinetic and thermodynamic parameters of biomolecular interactions by using site-directed mutagenesis, biochemical, biophysical and bioinformatic techniques. Molecular simulations of protein motions are usually performed by molecular dynamics (MD) in which atoms and molecules are allowed to interact over time at a given temperature following the laws of classical mechanics and which provide detailed description of atomic motions. Technical advances in computational and simulation methods have been matched by the accurate measurement of kinetic and thermodynamic parameters. Calculated data compared to experimental results have given a confidence in the validity of simulations (2).

In this study we focus on the conformational characteristics of the p53 protein. This protein is a tumor suppressor; its role is crucial in multicellular organisms where it regulates the cell cycle and functions in preventing cancer. p53 has been described as the “guardian of the genome” referring to its role in conserving stability by preventing genome mutation (3). This protein has many mechanisms of anticancer function and plays a role in apoptosis, genomic stability and inhibition of angiogenesis (4). Human p53 is 393 amino-acids long and is activated in its tetrameric form. If p53 becomes damaged, tumour suppression is severely reduced. The p53 gene in cells can be damaged by mutagens (chemicals, radiation or viruses) which will induce an uncontrolled division of the cells. More than 50 % of human tumours contain a mutation in the p53 protein. These mutations can cause either p53 destabilization or the

loss of its DNA-binding or both. Increasing the amount of p53, which may initially seem a good way to treat tumours or prevent them from spreading, is actually not a suitable method of treatment, since it can cause premature aging (5). However restoring endogenous p53 function holds a lot of promise.

We used computational models to investigate a set of properties of p53. The physics behind the models are based on molecular mechanics (classical/Newtonian mechanics) that were described in chapter 2 along with the techniques used to run the molecular dynamics (MD) simulations (conventional and accelerated MD) and the structural analysis methods. In chapter 3, a non-exhaustive summary of p53 molecular biology was proposed to help us understand and interpret the results. In a first part of the thesis, a statistical analysis, principal component analysis (PCA), was used to clarify some p53DBD experimental results (Chapter 4). The second part of the thesis, presents the results of the investigation of p53DBD stability using MD techniques (cMD and aMD). Two cases were investigated in these simulations:

- The role of the alkylation of p53DBD cysteines by Michael acceptors (MAs, p53 mutant rescue molecules), in chapter 5.
- The destabilisation of the p53DBD by a temperature sensitive 'structural' mutant: V143A, in chapter 6.

Taken together this study allowed us to highlight a common mechanism that might explain the rescue of the p53 mutants by MAs and the V143A mutation at low temperature.

Chapter 2. Computer simulation methodology

2.1 Introduction

Molecular Dynamics (MD) simulations provide valuable insights into the dynamics of protein structure at an atomistic level. They have been used to describe many protein mechanisms, such as their folding and unfolding and the effect of ligand binding on their structure (6–10). This theoretical method is based on mathematical models and physical laws to produce time dependent conformations of the proteins that will be analysed. In this chapter the molecular modelling theory and the analysis techniques used in the thesis will be described, based on description found in the book by A.R Leach (11).

2.2 Force fields

Molecular mechanics refers to the use of Classical/Newtonian mechanics to describe the physical basis behind the model of molecular systems. This model allows the molecular structure and properties to be described in as accurate manner as possible by ignoring the electronic motions (unlike in the quantum mechanical method) and calculates the energy of a system in its ground electronic state as a function of the nuclear positions only (Born-Oppenheimer approximation). This allow less time consuming calculations on large systems. For this purpose, atoms are described as spheres with a radius, a mass and an atomic charge and interactions are based on springs and classical potentials which will determine the spatial distribution of atoms and their energies. The springs and atom names contain the chemical information in the system; their parameters are determined by fitting to experimental results (X-ray crystallography and NMR spectroscopy) or *ab initio* calculations (12). A collection of constants defining the spheres and interaction potentials is referred to as a force field. The general form of a force field is described by four components that capture intra- and inter-molecular forces within the

system. The three first components describe the bonded interactions as bond stretching, angle bending and bond torsion and the last component take into account the non-bonded interactions, electrostatic and van der Waals interactions. One functional form of such a force field used to model single molecules or assemblies of atoms and/or molecules is the AMBER force field (13):

$$\mathcal{V}(\boldsymbol{r}^N) = \sum_{bonds} \frac{k_i}{2} \left(l_i - l_{i,0}\right)^2 + \sum_{angles} \frac{k_i}{2} \left(\theta_i - \theta_{i,0}\right)^2 + \sum_{torsions} \frac{V_n}{2} \left(1 + \cos(n\omega - \gamma)\right) + \sum_{i=1}^{N} \sum_{j=i+1}^{N} \left(4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6} \right] + \frac{q_i\, q_j}{4\pi\varepsilon_0 r_{ij}} \right)$$ Equation 2-1

Where $\mathcal{V}(\boldsymbol{r}^N)$ is the total potential energy, which is a function of the positions $(\boldsymbol{r})$ of N particles (atoms). The two first terms in equation (2-1) use a Hooke's law formula or harmonic potential in which the energy varies with the square of the displacement of the bond length l from the reference bond length l_0 or the angle θ from the reference angle θ_0 and k_i refers to the force constant of bond stretching and angle distortion respectively.

The third term in the equation is the torsional potential with a contribution from each bonded quartet of atoms A-B-C-D in the system. In this term, ω is the torsion angle, V_n gives a quantitative indication of the relative barriers to rotation, n is the multiplicity; its value gives the number of minimum points in the function as the bond is rotated through 360°. The phase factor γ determines where the torsion angle passes through its minimum value.

The last term in equation (2-1) corresponding to non-bonded interactions is usually considered in two groups, one comprising van der Waals interactions (first part) modelled by using Lennard-Jones (Fig. 2-1) potential and electrostatic interactions (second part) modelled by using the coulomb potential. The non-bonded energy represents the pair-wise sum of the energies of all possible interacting non-bonded atoms i and j.

The van der Waals attraction refers to an all-atom long-range attraction which is proportional to r^{-6} (r = distance between 2 atoms). At sufficiently short distances the interactions of all atoms are repulsive. In this function, ε_{ij} represent the depth of Lennard-Jones potential well, σ_{ij} is the distance at which the interatomic potential is zero, $\boldsymbol{r}_{ij}$ is the distance between the atoms.

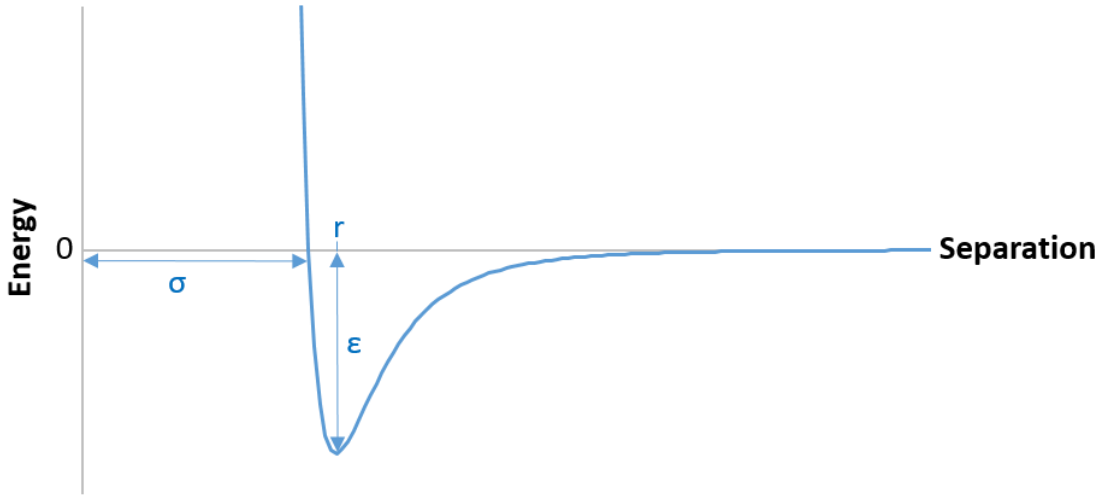


Figure 2-1: The Lennard-Jones potential.

Electrostatic interactions play a critical role in determining the structure and packing of molecules. In this function, $\boldsymbol{r}_{ij}$ is the interatomic distance, q_i and q_j are the corresponding charges on the atoms and ε_0 is the relative permittivity of free space.

Most of the variation in the structure and relative energies is due to the complex interplay between the torsional and non-bonded contributions

Force fields also contain a set of parameters for each atom type. These parameter sets include values for atomic mass, van der Waals radius, and partial charge for individual atoms, and equilibrium values of bond lengths, bond angles, and dihedrals, and values corresponding to the effective spring constant (Hooke's law) for each potential. Usually these parameters are taken from experimental data (crystallographic, NMR structure) or obtained by *ab initio* calculations.

Several force fields have been developed for studying biomolecules based on this definition, such as AMBER (13), CHARMM (14) and OPLS (15).

2.3 Solvent model

Several explicit-atom solvent models exist; they can be classified by the number of points used to define the model (atoms plus dummy sites). Their structures can be rigid or flexible and some models can include polarization effects (16).

The use of explicit solvent is essential to reproduce properties of a solute molecule. This is especially important to reproduce kinetics (7). The simple water model TIP3P in which each water molecule is maintained in a rigid geometry and the interaction between molecules is described using pairwise coulombic and Lennard-Jones expressions, is used in this study (17).

Implicit solvation models (continuum models) can also be used. There are two basic types of this model; the accessible surface area-based models (18, 19) and continuum electrostatics models (20, 21).

2.4 Energy minimisation

Molecular mechanics allows minimisation of the calculated energy to obtain low energy configurations of a molecular system and reduce high initial forces which induce unstable simulation trajectories. The minimization algorithm finds the minimum value of a function f for which its first derivative will be equal to zero and its second derivative > 0. The methods of 'steepest descents' and 'conjugate gradient' method, frequently used algorithms in molecular modelling for minimisation, are 1st order minimisation methods (use the 1st derivative of the energy with respect to coordinates). They gradually change the atomic coordinates as they move the system closer and closer to the minimum energy location. In this way, the initial configuration of the system is moved almost exclusively downward on the energy surface until it reaches the closest local minimum.

2.4.1 Steepest descents method

This method moves the atoms in the direction parallel to the net force (walking straight downhill). In a $3N$ dimension system, this direction is represented by a $3N$ dimensional unit vector $\boldsymbol{s}_k$.

$$\text{With} \quad \boldsymbol{s}_k = \frac{-\boldsymbol{g}_k}{|\boldsymbol{g}_k|} \quad \text{Equation 2-2}$$

Once the direction is determined, how far to move along the gradient $\boldsymbol{g}$ can be specified by two methods:

- Line search that finds three points along the gradient such that the energy in the middle point is lower than the energy of the 2 outer points.
- or take a step with an arbitrary size along the gradient. The new set of coordinates $\boldsymbol{x}_{k+1}$ will be form the current coordinates $\boldsymbol{x}_k$ by:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \lambda_k \boldsymbol{s}_k \quad \text{Equation 2-3}$$

  In this method if the step results in reduced energy, the step size λ_k will be increased by a multiplicative factor of 1.2 at the second iteration and inversely reduced by a factor of 0.5 if the energy increases (11).

2.4.2 Conjugate gradients method

In this method the gradient direction $\boldsymbol{g}_k$ depend on the previous direction $\boldsymbol{v}_{k-1}$ (method also called conjugate direction method). The method moves in a direction $\boldsymbol{v}_k$ from the current position $\boldsymbol{x}_k$ where $\boldsymbol{v}_k$ is computed from the gradient at the point and the previous direction vector $\boldsymbol{v}_{k-1}$.

$$\boldsymbol{v}_k = -\boldsymbol{g}_k + \gamma_k \boldsymbol{v}_{k-1} \quad \text{Equation 2-4}$$

Where γ_k is a constant scalar:

$$\gamma_k = \frac{\boldsymbol{g}_k \cdot \boldsymbol{g}_k}{\boldsymbol{g}_{k-1} \cdot \boldsymbol{g}_{k-1}} \quad \text{Equation 2-5}$$

In the current study, both steepest descent and conjugate gradient algorithm were used. The steepest descent removes the larges strain but converge slowly when close to a minima, in this case the conjugate gradient is more efficient.

2.5 Periodic boundary conditions

To eliminate surface effects in a finite system, periodic boundary conditions are used. The simulated system is introduced in a cubic box surrounded by identical translated images of this box. In this situation the force calculation on the central system (box) include the contributions of the other neighbouring systems. When a molecule leaves the box, one of its images enters through the opposite side. The use of periodic boundary conditions allows interactions of each particle the particles within the box and with their image. This time consuming process can be controlled by using a potential with a finite range to consider only the closest images of a particle and neglect the rest.

2.6 Particle mesh Ewald (11)

This method is used to model long range forces, applied in conjunction with periodic boundary conditions, it sums the long-range interactions (charges). This allows the particle to interact with all other particles in the simulation box and with all their images in an infinite array of periodic cells. The Ewald summation method general expression:

$$\mathcal{V} = \frac{1}{2} \sum'_{|\boldsymbol{n}|=0} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{q_i \, q_j}{4\pi\varepsilon_0 \, |\boldsymbol{r}_{ij}+\boldsymbol{n}|} \qquad \text{Equation 2-6}$$

Where $\boldsymbol{n}$ is a cubic lattice point and $'$ excludes $i = j$ for $\boldsymbol{n} = 0$.

To correct the slow convergence, the summation is divided into two series and a function $f(r)$ is applied dealing with the rapid variation of $1/r$ at small r (by using real space) and the slow decay at long r (by using reciprocal space).

In the particle mesh Ewald method the charges are distributed in a grid, speeding up the computational reciprocal space part of the simulation.

2.7 Ensemble average

Thermodynamic properties can be calculated using the partition function derived from statistical mechanical formulae. This formula depends on the positions $\boldsymbol{r}^N$ and the momenta $\boldsymbol{p}^N$ of the N particles of the system that will be averaged over time, time average A:

$$A_{ave} = \lim_{\mathrm{T}\to\infty} \frac{1}{T} \int_{t=0}^{T} A(\boldsymbol{p}^N(t), \boldsymbol{r}^N(t))dt \quad \text{Equation 2-7}$$

For large systems, the full exploration of the energy surface is impossible. For these systems, Boltzmann and Gibbs developed statistical mechanics in which a single system evolving in time is replaced by a large number of replica and for which the ensemble average over these replicas is equal to the time average, the ergodic hypothesis. In this case the ensemble average $\langle A \rangle$ is as follow:

$$\langle A \rangle = \iint d\boldsymbol{p}^N d\boldsymbol{r}^N A(\boldsymbol{p}^N, \boldsymbol{r}^N) \rho(\boldsymbol{p}^N, \boldsymbol{r}^N) \quad \text{Equation 2-8}$$

Where $\rho(\boldsymbol{p}^N, \boldsymbol{r}^N)$ is the probability density of the ensemble.

There are different ensembles with different characteristics:

- The microcanonical ensemble (NVE), in which the thermodynamic state is characterized by fixed number of atoms (N), a fixed volume (V) and a fixed energy (E).
- The canonical ensemble (NVT), characterized by a fixed number of atoms (N), volume (V) and temperature (T).
- The isobaric isothermal ensemble (NPT), defined by a fixed number of atoms (N), pressure (P) and temperature (T).
- The grand canonical ensemble (μVT), defined by a fixed chemical potential (μ), volume (V) and temperature (T).

Under the canonical ensemble, the probability of finding a configuration with momenta $\boldsymbol{p}^N$ and position $\boldsymbol{r}^N$ is equal to the Boltzmann distribution:

$$\rho(\boldsymbol{p}^N, \boldsymbol{r}^N) = e^{\left(-E(\boldsymbol{p}^N, \boldsymbol{r}^N)/k_B T\right)/Q} \quad \text{Equation 2-9}$$

Where $E(\boldsymbol{p}^N, \boldsymbol{r}^N)$ is the energy, Q is the partition function (sum of all possible energy stated), k_B is the Boltzmann's constant and T is the temperature.

The partition function Q is generally written in terms of the Hamiltonian $\mathcal{H}$. In a canonical ensemble:

$$Q = \frac{1}{N!}\frac{1}{h^{3N}}\iint d\boldsymbol{p}^N d\boldsymbol{r}^N e^{\left[-\mathcal{H}(\boldsymbol{p}^N,\boldsymbol{r}^N)/k_BT\right]} \quad \text{Equation 2-10}$$

The factor $N!$ is used as the particles are indistinguishable and the factor $\frac{1}{h^{3N}}$ is required to ensure that the partition function is equal to quantum mechanical results for a particle in a box.

The thermodynamics properties of a system are then derived from the partition function Q. As an example the energy of a system will be written as follow:

$$\langle E \rangle = U = kT^2\left(\frac{d\ln q}{dT}\right) \quad \text{Equation 2-11}$$

Which in the canonical ensemble corresponds to the Helmholtz free energy A:

$$A = -kT\ln Q \quad \text{Equation 2-12}$$

2.8 Molecular dynamics

Computer simulations allow the understanding of molecular systems in terms of their structure and microscopic interactions.

2.8.1 Equation of motion

Newton's second law is used to describe the time dependent evolution of atomic coordinates in phase space, which for a simple atomic system may be written as:

$$m_i\boldsymbol{a}_i = \boldsymbol{f}_i \quad \text{Equation 2-13}$$

$$\boldsymbol{f}_i = -\frac{\partial}{\partial \boldsymbol{r}_i}\mathcal{V} \quad \text{Equation 2-14}$$

Where m_i is the mass of atom i, $\boldsymbol{f}_i$ is the force acting on the atom i, and $\mathcal{V}$ is the potential energy.

The acceleration $\boldsymbol{a}$ is the first derivative of velocity with respect to time and the second derivative of position with respect to time:

$$\boldsymbol{a}_i = \frac{d\boldsymbol{v}_i}{dt} = \frac{d^2\boldsymbol{r}_i}{dt^2} \qquad \text{Equation 2-15}$$

The combination of equation 2-13 to 2-15 shows how Newton's second law relates the derivative of the potential energy to the changes of position as a function of time.

$$-\frac{\partial}{\partial \boldsymbol{r}_i}\,\mathcal{V} = m_i \frac{d^2\boldsymbol{r}_i}{dt^2} \qquad \text{Equation 2-16}$$

Furthermore the forces $\boldsymbol{f}_i$ acting on the atoms have to be calculated, and these are usually derived from a potential energy $\mathcal{V}(\boldsymbol{r}^N)$, where $\boldsymbol{r}^N = (r_1, r_2, \ldots, r_N)$ represents the complete set of $3N$ atomic coordinates.

2.8.2 Molecular dynamic algorithm

In molecular dynamics simulations, successive configurations of the system are generated by integrating Newton's laws of motion in time, which induces a force change on each particle whenever the particle changes its position, or whenever any of the other particles with which it interacts changes position, so the equations of motion must be calculated simultaneously. This cannot be solved analytically for more than two interacting bodies, but can be solvated numerically by using a finite difference method. Several algorithms exist for integrating the equations of motion using this method. All algorithms assume that the positions and dynamic properties (velocities, accelerations, etc.) can be approximated as Taylor series expansions.

The velocity Verlet algorithm (11) is one of the most commonly used methods for this integration. This method gives positions, velocities and accelerations at the same time and does not compromise precision, unlike the standard Verlet algorithm in which the positions $\boldsymbol{r}(t + \delta t)$ are obtained by adding a small

term $\left(\delta t^2 \boldsymbol{a}(t)\right)$ to the difference of two much larger terms, $2\boldsymbol{r}(t)$and $\boldsymbol{r}\,(t - \delta(t))$.

The velocity Verlet algorithm is implemented in three-stages. As shown in equation 2-17; to calculate the new velocities requires the accelerations at t and $t + \delta t$ where δt is a fixed time step. Thus in the first step the position at $t + \delta t$ are calculated based on equation 2-12 using velocities and accelerations at time t. The velocities at time $t + \frac{1}{2}\delta t$ are then determined by using equation 2-18. New forces are computed from the current positions, thus giving $\boldsymbol{a}(t + \delta t)$. In the final step, the velocities at time $t + \delta t$ are determined using equation 2-19.

$$\boldsymbol{r}(t + \delta t) = \boldsymbol{r}(t) + \delta t\boldsymbol{v}(t) + \frac{1}{2}\delta t^2 \boldsymbol{a}(t) \quad \text{Equation 2-17}$$

$$\boldsymbol{v}(t + \delta t) = \boldsymbol{v}(t) + \frac{1}{2}\delta t[\boldsymbol{a}(t) + \boldsymbol{a}(t + \delta t)] \quad \text{Equation 2-18}$$

$$\boldsymbol{v}\left(t + \frac{1}{2}\delta t\right) = \boldsymbol{v}(t) + \frac{1}{2}\delta t\boldsymbol{a}(t) \quad \text{Equation 2-19}$$

$$\boldsymbol{v}(t + \delta t) = \boldsymbol{v}\left(t + \frac{1}{2}\delta t\right) + \frac{1}{2}\delta t\boldsymbol{a}(t + \delta t) \quad \text{Equation 2-20}$$

Another variation of Verlet is the leap-frog algorithm (11). In this method the velocities 'leap frog' over the positions to give their values at $t + \frac{1}{2}\delta t$:

$$\boldsymbol{v}\left(t + \frac{1}{2}\delta t\right) = \boldsymbol{v}\left(t - \frac{1}{2}\delta t\right) + \delta t\boldsymbol{a}(t) \quad \text{Equation 2-21}$$

Then the positions leap over the velocities to give their new values at $t + \delta t$:

$$\boldsymbol{r}(t + \delta t) = \boldsymbol{r}(t) + \delta t\boldsymbol{v}\left(t + \frac{1}{2}\delta t\right) \quad \text{Equation 2-22}$$

This is ready for the velocities at $t + \frac{3}{2}\delta t$, and so on.

2.8.3 Isothermal isobaric (NTP) ensemble

Molecular dynamics is usually performed under conditions of constant number of particles (N), volume (V) and energy (E), the microcanonical ensemble NVE. But this can be modified to sample from other ensembles. For this study the Isothermal isobaric (NTP) has been used in which number of particles (N), temperature (T) and pressure (P) are conserved. This ensemble corresponds most closely to experimental conditions with a flask open to ambient temperature and pressure.

2.8.3.1 Constant temperature

The temperature T of a system is related to the time average kinetic energy $\langle \mathcal{K} \rangle_t$.

$$\langle \mathcal{K} \rangle_{NVT} = \frac{k_B T}{2}(3N - N_c) \quad \text{Equation 2-23}$$

$$\text{And also } \langle \mathcal{K} \rangle_{NVT} = \frac{1}{2} m v^2 \quad \text{Equation 2-24}$$

In this equation (2-18 and 2-19), k_B is the Boltzmann's constant, N is the number of particles, each with three degrees of freedom, N_c is the number of constraints on the system and m is the masse of a particle.

An obvious way to alter the temperature is thus to scale the velocities $\boldsymbol{v}$. If the temperature at time t is $T(t)$ and the velocities are multiplied by a factor λ then the associated temperature change can be calculated as follow:

$$\Delta T = \frac{1}{2}\sum_{i=1}^{N} \frac{2}{3}\frac{m_i(\lambda v_i)^2}{N k_B} - \frac{1}{2}\sum_{i=1}^{N} \frac{2}{3}\frac{m_i v_i^2}{N k_B} \quad \text{Equation 2-25}$$

$$\leftrightarrow \Delta T = (\lambda^2 - 1)T(t) \quad \text{Equation 2-26}$$

Which allow the control of the temperature by multiplying the velocities at each time step by the factor $\sqrt{T_{new}/T(t)}$.

An alternative way to control the temperature is by using the Berendsen thermostat (11). This technique couples the system to an external heat bath fixed at the desired temperature. The bath supplies and removes heat from the system as appropriate. In this case the velocities will be scaled using the following scaling factor:

$$\lambda^2 = 1 + \frac{\delta t}{\mathfrak{T}}\left(\frac{T_{bath}}{T(t)} - 1\right)$$ Equation 2-27

Where $\mathfrak{T}$, is a coupling parameter; its value determines how tightly the system and the bath are coupled.

2.8.3.2 Constant pressure

A macroscopic system maintains a constant pressure P by changing its volume V, and the volume fluctuation is related to the isothermal compressibility $\mathcal{K}$, as follow:

$$\kappa = \frac{-1}{V}\left(\frac{\partial V}{\partial P}\right)_T$$ Equation 2-28

The methods used to control pressure are analogous to those used to control temperature: volume scaling and pressure bath. In this case the volume scaling factor λ is equal to:

$$\lambda = 1 - \kappa\frac{\delta t}{\mathfrak{T}_P}(P - P_{bath})$$ Equation 2-29

And the new position $\boldsymbol{r}_i'$ is calculated as follow:

$$\boldsymbol{r}_i' = \lambda^{1/3}\boldsymbol{r}_i$$ Equation 2-30

With $\boldsymbol{r}_i$ been the current position.

In molecular dynamics, the pressure is calculated using the virial theorem of Clausius, in which the total virial for a real system equals the sum of an ideal gas part $(-3PV)$ and a contribution due to interactions between the particles. Also the virial theorem states that the virial is equal to $-3Nk_BT$. The result obtained is:

$$W = -3PV + \sum_{i=1}^{N}\sum_{j=i+1}^{N} \boldsymbol{r}_{ij}\boldsymbol{f}_{ij} = -3Nk_BT$$ Equation 2-31

From which the pressure expression is derived as:

$$P = \frac{1}{V}\left[Nk_BT - \frac{1}{3}\sum_{i=1}^{N}\sum_{j=i+1}^{N} \boldsymbol{r}_{ij}\boldsymbol{f}_{ij}\right]$$ Equation 2-32

Where V is the volume, N is the Avogadro number, T the temperature, $\boldsymbol{r}_{ij}$ is the distance and $\boldsymbol{f}_{ij}$ is the force acting between atoms i and j (11).

2.8.4 Constraints

As high frequency motions (bond vibrations) are not interesting features to focus on to observe major conformational changes of a system, and as they dictate the time step of molecular dynamics, these motions will be constrained to their equilibrium values. Constraint dynamics enables individual internal coordinates or combinations of specified coordinates to be constrained during the simulation without affecting the other internal degree of freedom. The most commonly used method for this purpose is the SHAKE procedure (22), where desired coordinates will be fixed by using the method of Lagrange multipliers. SHAKE will keep the desired coordinates at their equilibrium value and allow the exploration of more phase space by allowing a large timestep to be used. During MD simulations, these constraints are usually applied on bonds involving hydrogen. In constraint dynamics the equations of motion are solved while simultaneously satisfying the imposed constraints.

2.9 Enhanced sampling method: accelerated MD

For most systems their energy landscape cannot be explored adequately by conventional Molecular Dynamics (cMD), due to their multiple minima and high free energy barriers. Several methods have been proposed to overcome this problem while being able to simulate rare transitions between potential energy minima (23, 24).

In this thesis accelerated molecular dynamics (aMD), a biased potential method was tested. This technique was first presented by McCammon *et al.* to simulate the transition of high energy barrier of a system without previous knowledge

of its energy landscape (25). To do so, this method proposes a 'boost energy' E_b that will be used as a threshold during the potential energy $V(\boldsymbol{r})$ calculation:

If $V(\boldsymbol{r}) \geq E_b$ no bias potential is applied and the potential energy remains equal to $V(\boldsymbol{r})$.

$$V^*(\boldsymbol{r}) = V(\boldsymbol{r})$$

On the other hand if $V(\boldsymbol{r}) < E_b$ a bias potential $\Delta V(\boldsymbol{r})$ is added as such:

$$V^*(\boldsymbol{r}) = V(\boldsymbol{r}) + \Delta V(\boldsymbol{r})$$

The bias potential:

$$\Delta V(\boldsymbol{r}) = \frac{(E_b - V(\boldsymbol{r}))^2}{E_b - V(\boldsymbol{r}) + \alpha} \qquad \text{Equation 2-33}$$

Depends on the choice of the boost energy, E_b and the acceleration parameter α.

When this bias potential is applied, the forces on the atoms are recalculated for the modified potential as follow:

$$F_i^* = -\frac{d}{d\boldsymbol{r}}[V(\boldsymbol{r}) + \Delta V(\boldsymbol{r})] = F_i \times \left[\frac{\alpha^2}{(\alpha + E_b - V(\boldsymbol{r}))^2}\right]$$

In aMD the use of the bias potential allows more frequent jumps between the different systems states and *per se* explore more phase space that the cMD (although do not reproduce the exact dynamics of the system) (26).

In this study a dual potential boost was used: one to bias the overall potential energy and the second to bias the torsional terms. While the use of two bias potentials enhances the conformation sampling it also allows to produce similar configurations distributions to the cMD (27).

2.10 Analysis methods

Several analysis techniques were used to monitor the conformational changes of our system.

2.10.1 Root-Mean-Square Deviation (RMSD)

The root-mean-square deviation is the measure of the average distance between the atoms (usually backbone or C-α atom) of superimposed protein structures. It quantifies how different the MD conformations are from a reference structure. RMSD is calculated following the equation:

$$RMSD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\delta_i^2} \qquad \text{Equation 2-34}$$

Where δ is the distance between N pairs of equivalent atoms.

2.10.2 Root-Mean-Square Fluctuation (RMSF)

The root-mean-square fluctuation calculates the average fluctuation of each residue along the simulations. It measures the deviation between the position x_i of a particle and some reference position $\tilde{x}_i$ at time t_j.

$$RMSF = \frac{1}{N}\sum_{t_j=1}^{T}\left(x_i\left(t_j\right) - \tilde{x}_i\right)^2 \qquad \text{Equation 2-35}$$

Where T is the time over which one wants to average.

RMSD and RMSF have been performed using PTRAJ module in Amber 11 (28) on alpha-carbons of the superimposed structures and the structure used as reference for the RMSD and RMSF.

2.10.3 Secondary structure analysis: STRIDE

Secondary structure analyses have been performed via Timeline, a VMD (29) plugin for trajectory analysis using the STRIDE software (30) for secondary structure determination. STRIDE is a knowledge based approach which uses a carefully verified set of secondary structure elements found in the protein data base. It determines secondary structure by considering hydrogen bonding pattern and backbone geometries. This techniques allows to monitor the secondary structure changes within a system.

2.10.4 Hydrogen bond analysis

This analysis has been performed using HBonanza (31), an open source, python implemented computer program. This program allows analysis and visualization of hydrogen bond networks (via VMD) of both trajectory and single-structure PDB files. HBonanza defines a hydrogen bond through several characteristics; heavy atoms must be oxygen, nitrogen, fluorine or sulphur; distances between heavy atoms must be no greater than a certain cut-off distance (3.5 Å by default), the angle θ (Fig. 2.2) formed by the hydrogen-bond must be less than user-specified angle (30° by default) and finally for a MD trajectory analysis the hydrogen bonds must be well represented (percentage of occurence) among the frames of the simulation: if a hydrogen bond appears only infrequently, it is less likely to play an important role in protein structure.

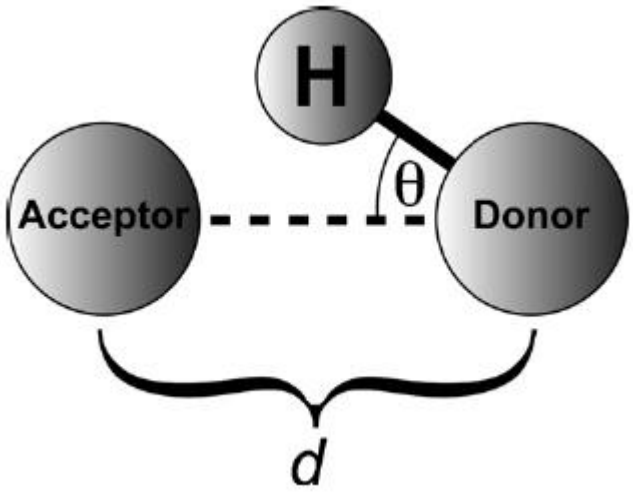


Figure 2-2: Schematic representation of a hydrogen bond.
d is the distance between heavy atoms and θ is the angle formed by the donor-acceptor distance and the hydrogen linked to the donor.

2.10.5 Dynamic cross-correlation map (DCCM) (32)

The correlation of the atomic motions in proteins is essential to understand their biological functions. This analysis allows low frequency, collective motions to be determined and to determine the influence of the neighbouring atoms on the dynamic of active site residues (32).

For the displacement (with respect of a reference structure) vectors $\Delta \boldsymbol{r}_i$ and $\Delta \boldsymbol{r}_j$ for atoms i and j, the covariance $c(i,j)$ is given by:

$$c(i,j) = < \Delta \boldsymbol{r}_i \cdot \Delta \boldsymbol{r}_j > \quad \text{Equation 2-36}$$

Where <>, denote an ensemble average. The diagonal elements of the covariance matrix ($i = j$) are the mean-square (MS) atomic fluctuation, $< \boldsymbol{r}_i^2 >$. The cross-correlation is given by:

$$C(i,j) = \frac{c(i,j)}{[c(i,i)c(j,j)]^{1/2}} = \frac{<\Delta \boldsymbol{r}_i \cdot \Delta \boldsymbol{r}_j>}{<\Delta \boldsymbol{r}_i^2>^{1/2} <\Delta \boldsymbol{r}_j^2>^{1/2}} \quad \text{Equation 2-37}$$

Where $c(i,i)$ is the variance of atom i and $c(j,j)$ is the variance of atom j.

When $C(i,j) = 1$, the motions are correlated, they have the same phase as well as the same period. $C(i,j) = -1$ means that the motions are anticorrelated and when $C(i,j) = 0$, the two atoms have fluctuations of the same period and phase but the displacements are oriented at an angle of 90° (32).

In molecular dynamics, covariances are calculated by estimating the ensemble average over a set of discrete time points (or frames) t_n $(n = 1, 2, \ldots, N)$. The correlation between two atoms i and j will be written as:

$$c(i,j) = \frac{1}{N} \sum_{n=1}^{N} \Delta \boldsymbol{r}_i(t_n) \cdot \Delta \boldsymbol{r}_j(t_n) \quad \text{Equation 2-38}$$

Where $\Delta \boldsymbol{r}_i(t_n)$, is the displacement of atom i from its average position at (t_n).

To examine the covariances of residues i and j, the atomic displacements $\Delta \boldsymbol{r}$ are replaced by the average vector of the appropriate atoms for a given residue, based on backbone average or carbon-alpha for example. Bio3D, an R package (33) containing utilities for the analysis of protein structure, sequence and trajectory data has been used to perform the DCCM based on the alpha-carbons.

2.10.6 Principal components analysis (PCA)

This technique allows dominant patterns and representative distributions from noisy data to be highlighted. This is performed by mapping the investigated complex system from a multidimensional space to a reduced space spanned by a few principal components (PCs) (34). PCA has the potential to identify converged structures automatically by providing a visualization of the detected motion (unlike RMSD analysis, for which a manual checking is needed) (35). In this technique, the mean from each of the data dimensions is first subtracted. This produces a data set whose mean is zero. Then the covariance between 2 dimensions (X and Y coordinates of a specific atom for example) will be calculated as follows:

$$cov(X,Y) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{(n-1)} \qquad \text{Equation 2-39}$$

Where n is the number of dimensions, $\bar{X}$ and $\bar{Y}$ are the average structures of dimensions X and Y.

As in a molecular dynamic simulation, $3N$ (N = number of atoms in a system) dimensions are studied, the covariance between all these dimensions will be calculated by a covariance matrix, and for a n dimensional system ($n = 3N$), $\frac{n!}{(n-2)!*2}$ different covariances will be calculated. From this covariance matrix, the eigenvectors (corresponding to the principal components) and their eigenvalues are calculated. For a n dimension system n eigenvectors exist. These eigenvectors are then ordered according to their eigenvalues. Only the first few eigenvectors (direction of a common motion in a system) with high eigenvalues (intensity of the motion) are studied, as the majority are of little importance.

In this study the principal component analysis of the different molecular dynamic simulations have been performed via bio3D only on the alpha-carbons of the structures.

2.10.7 Trj-cavity

Trj-cavity allows identification of cavities in molecular dynamics simulation trajectories (36). This technique uses grid voxels to represent the protein structure and a new algorithm for a fast neighbour search. This method transforms the molecular coordinates of the system in a 3D grid of user-defined voxel size. The grid contains each atom with its corresponding van der Waals radius. One can vary the voxel size in order to modulate the precision of the cavity detection and an index file can be used to restrict the grid calculation to a group of atoms to limit the computational time needed.

The cavity search can be performed by scanning the entire atom-mapped grid or can start from a seed point provided by the user. This process is incremental with potential cavity members searched along the immediate six positive/negative neighbours until no other empty voxels are found. This algorithm is based on three characteristics to identify a cavity member: (i) the voxel must be empty; (ii) it must be connected to at least one cavity point and (iii) be surrounded by protein atoms in a certain number of directions.

A linear search will be performed along the positive and negative directions of the three grid dimensions until an atom or the end of the grid is found to define the “burial state” of the voxel. The “burial state” is linked to the cavity type determined by the user. Optionally a cut-off can also be used (in order to consider atoms from only one domain for example).

As the output for the cavity information is coordinate-based; (i) one cavity from a frame to another is considered to be the same when sharing some Cartesian coordinates; (ii) the output PDB for the cavity contains coordinates and mean occupancies of all detected free voxels, it is recommended to perform a sensible trajectory fit based on the cavity of interest. Generating a static cavity maps using trj-cavity might help with the trajectory fitting.

Several cavity characteristics are obtained using this method: the cavity structure and its trajectory, the cavity voxel volume versus time and an optional PDB file listing all the voxels corresponding to a possible cavity and their occupancy.

In addition to detect cavity trj-cavity can also monitor the presence of solvent or ligand in the cavity.

2.11 Summary

In this chapter, the simulation methods applied to the models of wild-type p53, its mutant and p53 within alkylated cysteines have been described. In subsequent chapters, the biochemical context of p53 will be described in more detail, along with the results of the computer simulations performed.

Chapter 3. The molecular biology of p53

3.1 Introduction

Cancer is one of the most studied diseases of our time; however, our comprehension of the cellular and molecular pathology of malignant transformation is incomplete. One of the criteria that characterize all malignant growths is the disregulation of growth-controlling genes commonly called oncogenes and tumour suppressor genes. This involves uncontrolled cell division and tissue invasion (metastasis). Proto-oncogenes are normal genes that promote cell growth and mitosis; once mutated by carcinogenic agents they become oncogenes producing excessive levels of growth promoting proteins. Tumour suppressor genes discourage cell growth; they are frequently transcription factors that suppress mitosis and cell growth to allow DNA repair. Suppressor genes are recessive (unlike oncogenes); they contain loss of function mutations (while gaining oncogenic functions). Consequently, to cause the loss of suppressor function, both copies of a suppressor gene need to mutate. But there are exceptions to this rule, such as certain mutations in the p53 gene product. p53 mutations can function as a “dominant negative” meaning that it prevents the function of the normal protein. Nearly half of all cancers involve altered p53 tumour suppressor genes.

3.2 p53 function

Unlike the other tumour suppressors, which are active in both normal and tumour cells, p53 is activated only by stress signals that have the potential to damage cells. It orchestrates the cellular response to DNA damage in concert with the DNA repair machinery in order to select among multiple effector pathways toward DNA repair, survival or apoptosis (37). Thus it is commonly named the “guardian of the genome”. In normal cells, p53 activity is kept at low levels through protein-protein interactions such as with MDM2 which is well known to promote p53 degradation via the ubiquitin/proteasome pathway (38). In response to DNA damage, such as ionizing irradiation, UV light, and

ribonucleotide depletion, p53 undergoes a complex and hierarchical series of post-translational modifications (phosphorylation-acetylation cascade) (39), which result in its activation (see Fig. 3.2). The exact mechanisms through which p53 functions are not clear. However, it is well known that activated p53 functions as a transcription factor to activate genes that modulate cell cycle arrest, DNA repair and apoptosis.

3.2.1 Cell cycle arrest and p53 implication in the checkpoint

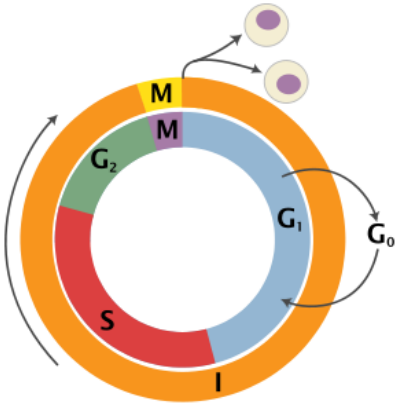


Figure 3-1: Schematic of the cell cycle.
M = Mitosis, when cells divide. G1 = G1 phase, in this phase the cell increases in size. S = Synthesis, this is the phase of DNA replication. G2 = G2 phase, this is the second phase of cell growth. G0 = G0 phase/resting, in this phase the cell has left the cycle and stopped dividing. I = Interphase (includes G1, S, G2)

At key transitions during cell cycle progression, temporary halt (arrest) occurs as checking processes take place to make sure there are no errors prior to proceeding to the next phase. These regulatory pathways are called cell cycle checkpoints. This enables: (i) cellular damage to be repaired; (ii) the dissipation of cellular stress signals (iii) availability of essential growth factors, hormones, or nutrients (37). Two main checkpoints exist: the G1/S and the G2/M checkpoint. The G1/S checkpoint prevents replication of damaged DNA, whereas the G2/M transition makes sure there is no damaged and/or incompletely replicated DNA. The G1/S transition is known as the restriction point and is a rate limiting step in the cell cycle. The implication of p53 in the G1/S arrest mechanism is much better documented than its ability to trigger apoptosis.

The accumulation of cells in the G1 phase after exposure to DNA damage is dependent on p53. p53 activation will induce the transcription of $p21^{Waf/Cip1}$,

which can inhibit Cdk activity during all phases of the cell cycle. Elevated levels of p21 bind and inactivate cyclin D/Cdk4,6 and cyclin E/Cdk2 complexes which in turn induce pRB hypophosphorylation, leading to cell cycle arrest.

It is also thought that p53 plays a role in the G2/M transition. However p53 knock-out and p53 mutated cells show a DNA damage-induced G2 arrest. In this case, p21 (p53 transcript) will inhibit cyclin B1/Cdc2 activity and a reduction of cyclin B1 and Cdc2 levels will result. At the G2 checkpoint, p53 also modulates additional downstream target genes: 14-3-3σ which can modulate the subcellular localization of cyclin B1/Cdc2 and GADD45 which induces a G2 arrest on cells (40). p53 senses DNA damage at several stages of the cell cycle and accordingly determines whether the cell needs to arrest at one of the checkpoints to undergo DNA repair or proceed through it (41).

3.2.2 p53 implication in DNA repair

DNA repair after DNA damage is critical to maintaining healthy genomes. For example, UV induced formation of pyrimidine dimers are removed by Nucleotide excision repair (NER), nucleotide modifications resulting from DNA hydrolysis and alkylation are removed by base excision repair (BER), mismatch repair corrects errors of DNA replication and DNA recombination repairs strand breaks (37). It has been shown that p53 is implicated in NER and BER. The intervention of p53 in these mechanisms is not totally clear yet, however some studies give some indications. In NER, p53 regulates the TFIIH repair complex by direct interactions its components, with XPB and XPD proteins, and by modulating Gadd45 and p48XPE genes (42). There is some evidence that p53 also localizes to sites of DNA damage (43). In BER, some evidence points towards the modulation of the DNA polymerase β (β-pol) by p53 and direct interactions with the APE endonuclease required in the BER machinery (42, 44).

3.2.3 p53 implication in apoptosis

p53 is also known to induce apoptosis through transcriptional regulation of genes as well as through transcription-independent mechanisms. It is thought that cell death is only observed when several of the genes controlled by p53 are expressed in concert (37). In the p53-independent pathway, p53 elicits the

release of cytochrome C to promote caspase activation (Apaf1/caspase 9 pathway) (45).

The exact mechanisms that govern the induction of cell cycle arrest and DNA repair or apoptosis are little understood. Understandably, these are affected by p53 expression levels, the cell type or the type of stress signal (46). Further complexities were revealed by the observation that lower levels of genotoxic agents enhanced BER activity but when these levels were increased, immediate inhibition of BER activity was followed by apoptosis (47).

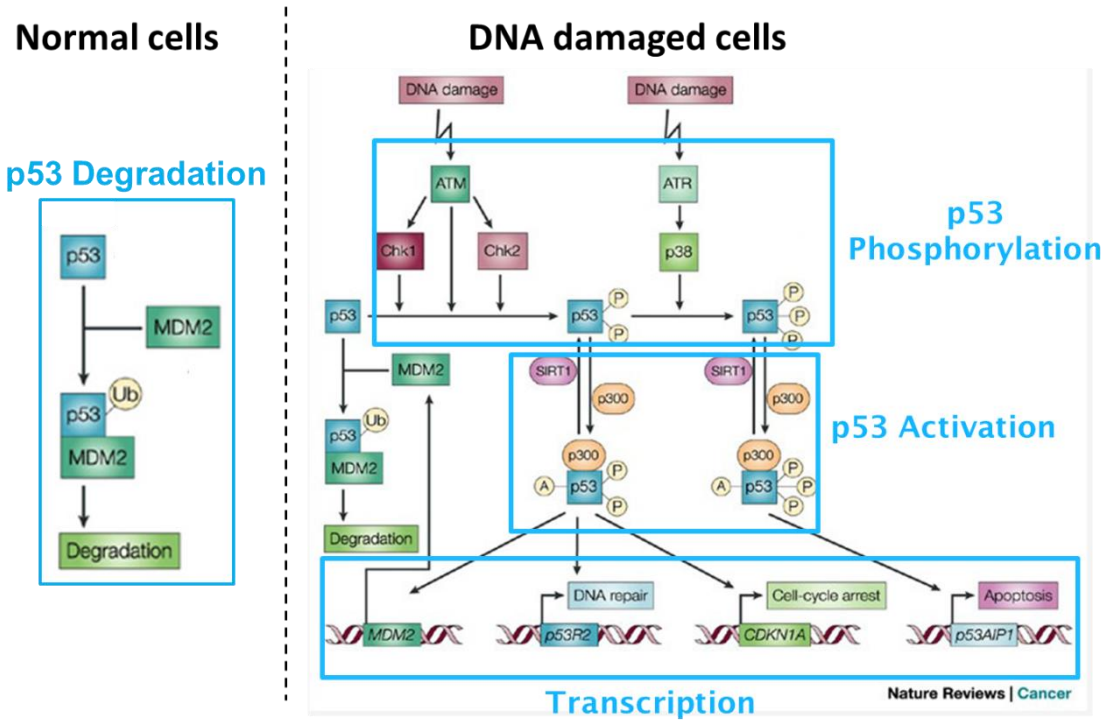


Figure 3-2: Illustration of p53 function.
In normal cells (left), p53 is kept at low leves by MDM2 which promotes p53 degradation via the ubiquitin/proteasome pathway. In DNA damaged cells (right), p53 is at first phosphorylated by DNA damage sensors (kinases), followed by p53 activation resulting in cell cycle arrest and DNA-repair or apoptosis (48).

3.3 Structure (49) (50)

p53 is biologically active as a homotetramer comprising 4*393 amino-acid residues and has a modular domain structure. There are two domains - a folded DNA-binding domain and a tetramerisation/oligomerization domain, and the protein is flanked by intrinsically disordered regions at both amino- and carboxy-termini (Figure 3.3). The DNA-binding core domain and oligomerisation domain are connected through a flexible linker region (residues 313-325) which contains the nuclear localisation signal (NLS); the oligomerization domain contains a nuclear export signal (NES, from residue 340 to 351).

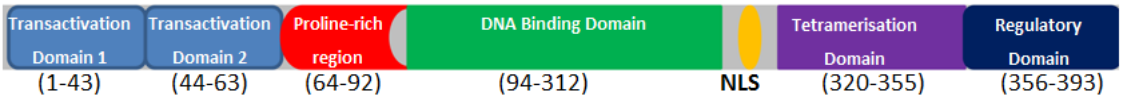


Figure 3-3: p53 domain distribution.
NLS standing for Nuclear Localisation Signal.

3.3.1 DNA-Binding Domain (DBD) structural motif

The DBD is constituted by a central immunoglobulin-like β sandwich (2 anti-parallel β-sheets) which provides the basic scaffold for the DNA-binding surface. The DNA-binding surface is characterised by a loop-sheet-helix motif and two large loops L2 (residues 164-194) and L3 (residues 237-250). These two loops are stabilized by a zinc ion coordinated to Cys176, 238, 242 and to His179 (see Figure 3.4).

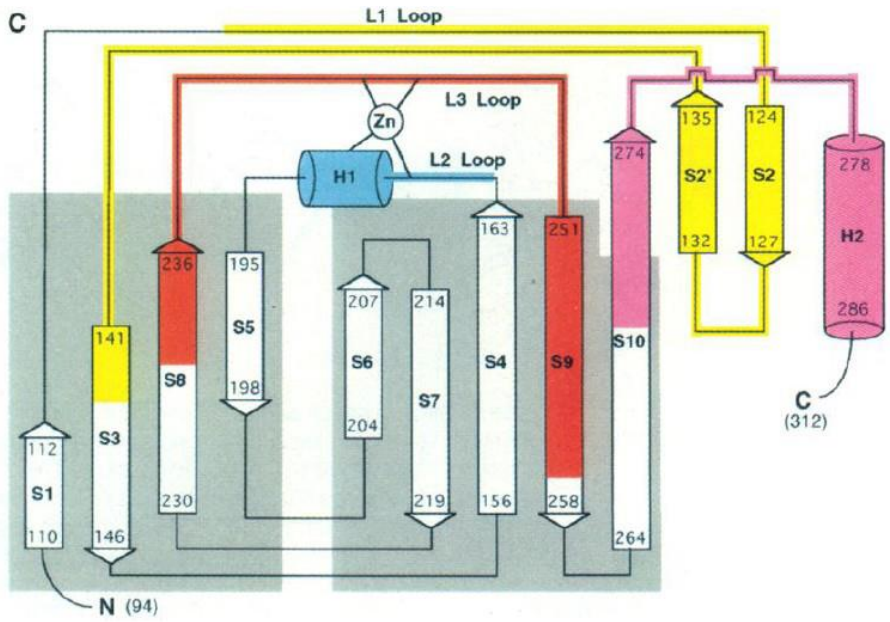


Figure 3-4: p53 DNA-binding domain (51).

p53 binds to DNA in a specific manner via this domain. The targeted DNA-binding sites are two decameric motifs (half-sites) with a general form as RRRCWWGYYY (where: R = A, G; W = A, T; Y = C, T) separated by 0-13 base pairs (52). The spacer length is correlated to p53 affinity and transactivation. Two core domains bind to a half site DNA, forming a symmetrical dimer with a relatively small-complementary helical DBD-DBD interface (Pro177, His178, Arg181, Met243 and Gly244). Conserved residues from the loop-sheet-helix motif have specific contacts with the major groove of the DNA and the L3 loop is anchored to the minor groove via Arg248. The other residues make varying interactions with base pairs in the major groove (Figure 3.5).

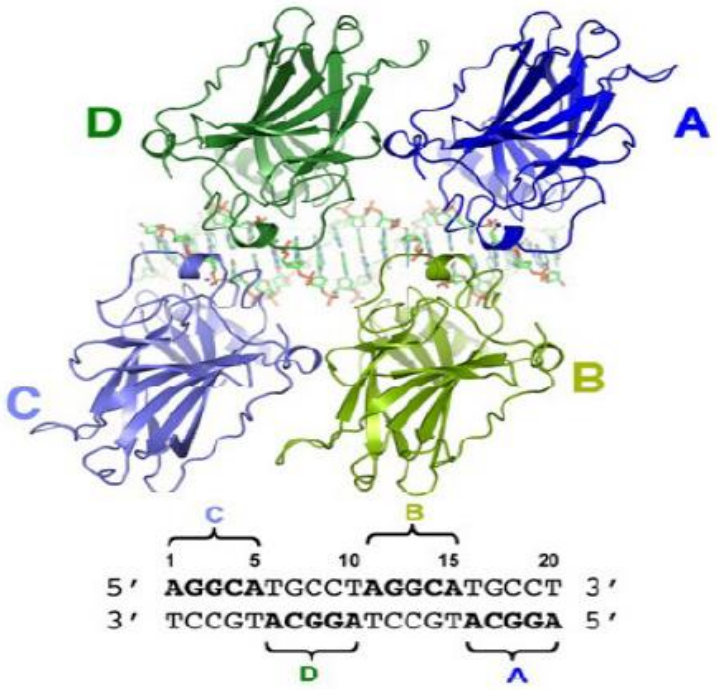


Figure 3-5: Specific DNA Binds to p53 DBD (53)

3.3.2 Oligomerization/Tetramerisation Domain (TD)

The tetramerisation domain is a dimer of dimers. Each monomer is constituted by a short β-strand followed by an alpha-helix, both connected by a sharp turn induced by a conserved glycine residue, Gly334. It forms a V-shape and each element of the secondary structure is a branch of the V. Three amino-acids (Ile332, Phe338 and Phe341) form a small hydrophobic cluster at the hinge region (see figure 3.6 A). Two monomers associate via their β-strands to form an antiparallel double-stranded sheet and via the antiparallel association of their helices to create a double-helical bundle. The formation of the antiparallel β-sheet leads to the creation of eight backbone hydrogen bonds. A hydrophobic core, made up of Phe328, Leu330, Ile332 from the β-strands and Phe338, Phe341 and Asn345 from the α-helices, is also created at the interface between the monomers. A salt bridge possibly exists between Arg337 from one monomer and Asp352 from the other monomer (see figure 3.6 B). Two dimers interact via their α-helices. The β-strands are on the outside of the tetramer, and their residues are therefore not directly involved in the association between the two dimers. The interface between the helices is mainly hydrophobic and involves Met340, Leu344, Ala347, Leu348 and Leu350 (see figure 3.6 C). The residues located in the hydrophobic core are crucial for

the formation of the tetramers and the hydrophobic effect is therefore the major stabilizing force. Therefore a highly conserved intermolecular salt bridge between Arg337 and Asp352 is also implicated in the tetramer stabilization.

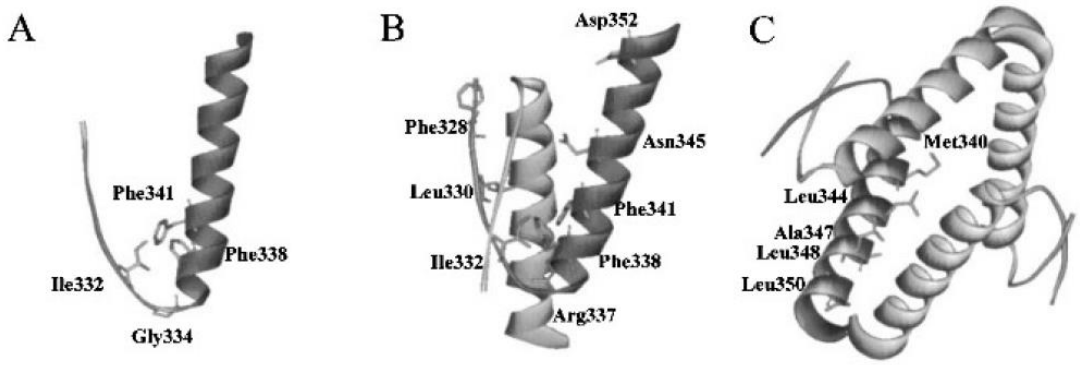


Figure 3-6: Tetramerization domain.
With **A)** representing the monomer, **B)** the dimer and **C)** the tetramer (54).

3.3.3 N-terminal transactivation domain and C-terminal regulatory domain

These domains correspond to the disordered regions of the p53 protein. But they will adopt a folded state upon binding to other proteins or non-specific DNA. The amino terminal transactivation domain (TAD) is essential for the p53 transcriptional activity as TAD will directly bind to transcriptional co-activators (p300/CBP) and to components of the basal transcription machinery (55). This domain is often represented as two subdomains: TAD1 which binds strongly to MDM2 and MDMX and TAD2, important for the apoptotic activity of p53. These intrinsically disordered regions bind to multiple partner proteins in several processes (56). Further control is exerted by post-translational modifications (9 phosphorylation sites in the TAD). In this way, TAD provides specificity in response to a particular cellular signal. The TAD is flanked by a Proline-rich Region from residue 64 to residue 91, which contains five repeats of the sequence PXXP (P → Proline and X → any other amino acid). Deletion of this region was first thought to mainly compromise growth suppression (57), but was later found that together with its neighbouring domain, TAD2, is implicated in apoptosis.

The carboxyl terminal regulatory domain (CTD) undergoes various post-translational modifications such as acetylation at the 6 carboxy-terminal

lysines and phosphorylation of serine and threonine residues (58, 59). These modifications play a role in the modulation of p53 function and regulation of cellular protein levels. Also, CTD presents a chameleon sequence (residues 376-387) which can adopt different secondary structure (α-helix or β-sheet) to bind to different proteins and to DNA in a non-specific manner.

3.3.4 The full p53 structure

Several techniques were used to study the full structure of p53. The first cryoelectron microscopy (Cryo-EM) study showed a tetramerisation of p53 matching the first human p53DBD Xray structure solved (60). The p53 full structure organisation proposed in this Cryo-EM study is thought to allow p53 to bind to long and mismatched DNA sequence. A later electron microscopy study displayed a relaxed p53 structure compared to the DNA bound (61). On the other hand the SAXS studies of p53 full structure proposed an organisation matching most of the p53DBD experimental structure (62, 63). In this studies, the authors focused on the organisation of the N and Cter domains. While the prolin rich region was found to be rigid, the TAD was shown to be highly flexible and to present a nascent helix (that becomes fully helical when binding to MDM2). Recently, mass spectrometry with chemical cross link has been used to solve p53 (64). This study was in agreement with the SAXS structures and highlighted a more compact Cter domain.

3.4 p53 post-translational modifications

The multiple-functionality of p53 is thought to be due to its regulation at several levels. The TP53 gene regulation generates twelve p53 isoforms (65, 66) produced by alternative slicing, promoters and translation. The main differences between these isoforms are located in the N and C-terminal regions that make the protein shorter. Diverse functions are allocated to these isoforms including the observation that shorter isoforms or those lacking the N-terminus act as antagonists. There is also evidence that while most of them will oligomerize with p53, some may act autonomously (65, 66).

The second level of regulation concerns p53 expression, as its transcription needs multiple positive and negative regulations from several promoters (67). The majority of the p53 regulatory elements are found upstream of the

transcription initiation site. To date several proteins binding sites have been identified: for example, promoters for the PAX family repress p53 transcription, while for the c-Myc protein enhances p53 expression.

Once the p53 protein is expressed, a next level of regulation is through post-translational modifications that are responsible for p53 differential function (58, 59).

Several post-translational modifications have been reported for p53, with the majority located outside the DBD, mainly on the N- and C-terminal domains. These modifications are important in p53 regulation and activation, however complexity arises because these modifications will have different functions depending on the tissue and cell type as well as on the genotype and on the kind of stimuli causing the stress. In the following, we will go through some of these modifications to develop an understanding of their effects on the conformational dynamics of p53.

3.4.1 Ubiquitination

p53 lysines undergo mono and poly-ubiquitination, orchestrated by Mdm2. At low concentration of Mdm2, p53 will be mono-ubiquitinated and exported form the nucleus while a high concentration of Mdm2 will promote poly-ubiquitination of p53 and its degradation (68). Human p53 contains 20 lysines, mainly located in the C-terminus and in the DBD and only one lysine in the N-terminal domain. Studies have shown that the lysines located in the DBD are essential for p53 degradation via ubiquitination (69). Indeed the DBD truncated p53 shows not only reduced ubiquitination but is also more stable (not degraded) compared to the full structure.

3.4.2 Acetylation

p53 acetylation is a reversible process that occurs in response to DNA damage and genotoxic stress and is crucial for transcription (70). Although it occurs preferentially on the C-terminal domain and the linker region containing the Nuclear Localisation Signal (NLS), acetylations also occur on the DNA binding domain on lysines 120 and 164. While the acetylation of K164 (as most of the other acetylation sites) was shown to be important for the docking of

transcription cofactors to the promoter regions (55, 71), the acetylation of K120 (located within the DNA binding surface, sometimes implicated in direct DNA binding) is thought to be involved in the differential activation of p53 (choice between apoptosis or DNA damage repair) (72). K120 mutation still binds to DNA, however it abrogates apoptosis, as it is unable to activate the proapoptotic genes PUMA and BAX. p53 acetylation regulates not only its activation and transcription activity, but also to its stability. It is clear that there must be some complex inter-relationship between ubiquitination and acetylation as both take place on the same lysines; so for example, once p53 is, acetylated it will clearly inhibit its ubiquitination by Mdm2 (73).

3.4.3 Cysteines and redox modulation of p53

Human p53DBD contains 10 cysteines (Fig. 3-7 A), nine of which are conserved in murine and human p53DBD (Fig. 3-7), while only seven are conserved in the p53 homologues p63 and p73 (Fig. 3-8): C141 (C164 in p63 and C153 in p73), C135 (C170 in p63 and C159 in p73), the three cysteines coordinated to the zinc C176 (C205 in p63 and C194 in p73), C238 (C269 in p63 and C258 in p73) and C242 (C273 in p63 and C262 in p73) and the two cysteines in the turn region prior to the H2 α-helix: C275 (C306 in p63 and C295 in p73) and C277 (C308 in p63 and C297 in p73). The oxidation-reduction of the cysteines has been shown to modulate p53-DNA binding, with oxidation inhibiting the binding (74, 75). Hainau and Milner (74) showed that p53 exposed to metal chelators adopts an oncogenic mutant like conformation (reactive with antibody PAb240, specific to mutant p53) and does not bind to DNA. The DNA binding loss at 0.625 mM of chelator can be partially recovered by addition of low concentration of Zinc (125 μM). The metal ion removal has been associated with an oxidation of the thiols in p53 (74). Wild type p53 exposed to DTT (sulfhydryl reducing agent) showed a WT conformation (p53 reactive with antibodies PAb246 and PAb1620, specific to WT p53) and keeps its ability to bind to the DNA. On the other hand, addition of diamide (sulfhydryl oxidizing agent) disturbed p53 tertiary conformation, induced the formation of inter- and intramolecular dissulfide bridges and inhibited p53-DNA binding. The authors suggest that redox modulation plays a crucial role in the ability of p53 to suppress/promote cell proliferation.

In a study from Mann et *al.* (76), enquiring the specific cysteines implicated in the activation/inhibition of DNA binding (in mouse p53), the authors mutate the cysteines to serines and examine the DNA binding ability of each mutant. The mutation of the cysteines coordinating the zinc ion; C173, 235 and 239 (C176, 238 and 242 in human p53) induce a nearly complete loss of function and the mutations of C121, 132, 138 and 272 (C124, 135, 141 and 275 for human p53) induce an intermediate loss of function (DNA binding). The authors conclude that cysteine oxidation-reduction influences p53 function in multiple ways (76).

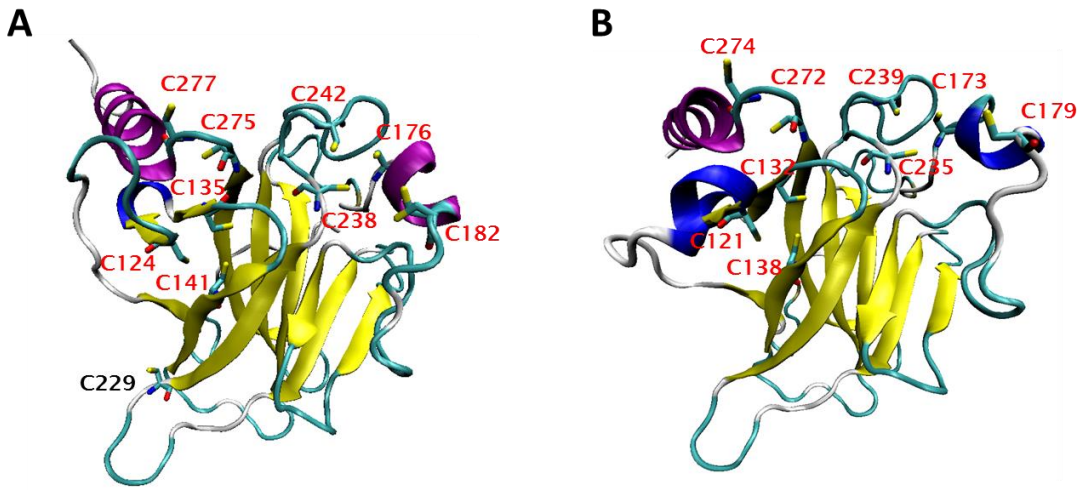


Figure 3-7: Location of the cysteines in p53.
A) Human p53DBD, conserved cysteines are in red. B) Mouse p53DBD.

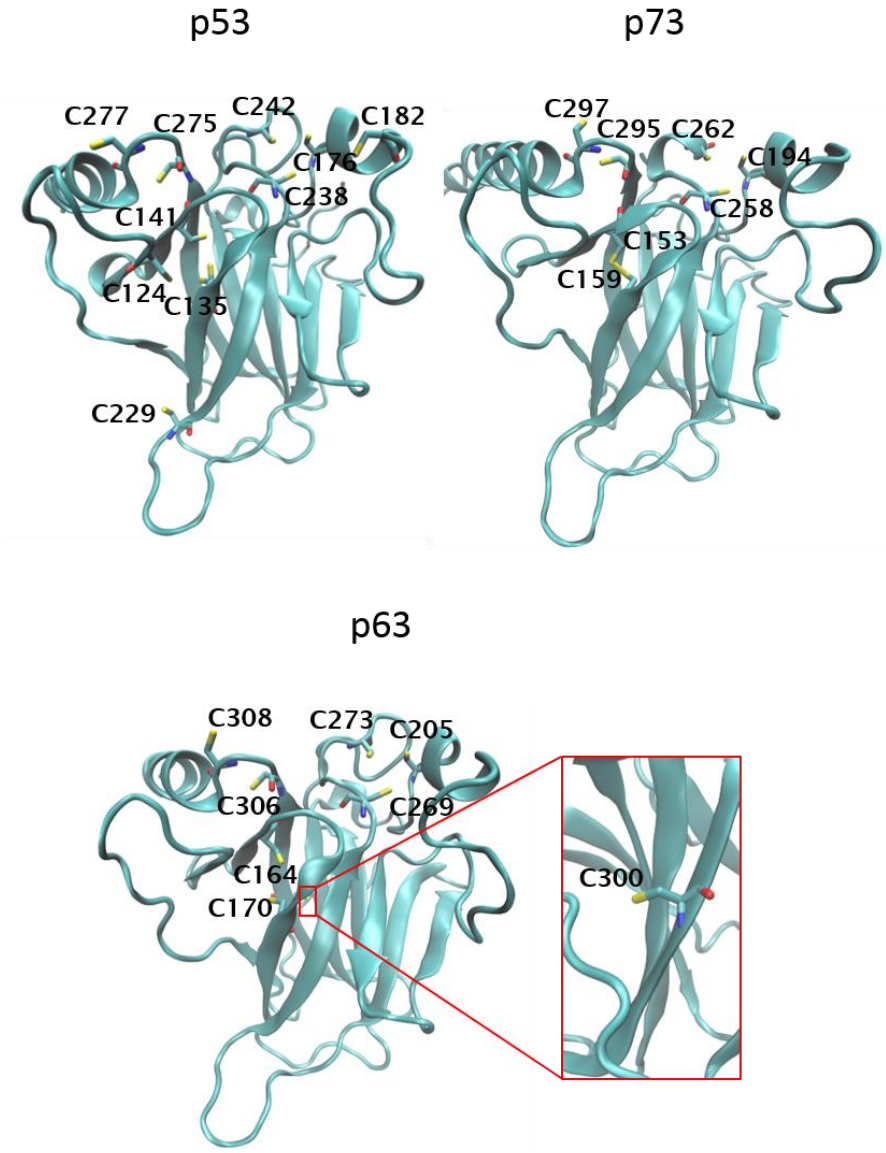


Figure 3-8: Cysteines in the p53, p73 and p63.
The structures are represented in cartoon and the cysteines in licorice.

3.5 p53 mutations

As outlined earlier, the p53 protein prevents cancer by inhibiting cell division, inducing DNA damage repair or promoting apoptosis mainly via its function as a transcriptional factor. This tumour suppressor is mutated in ~50% of reported human cancers, making it a major target for anticancer therapy. It has also been reported that over 95% of these mutations occur in the DNA-Binding Domain. These mutations are commonly called hot spot mutations.

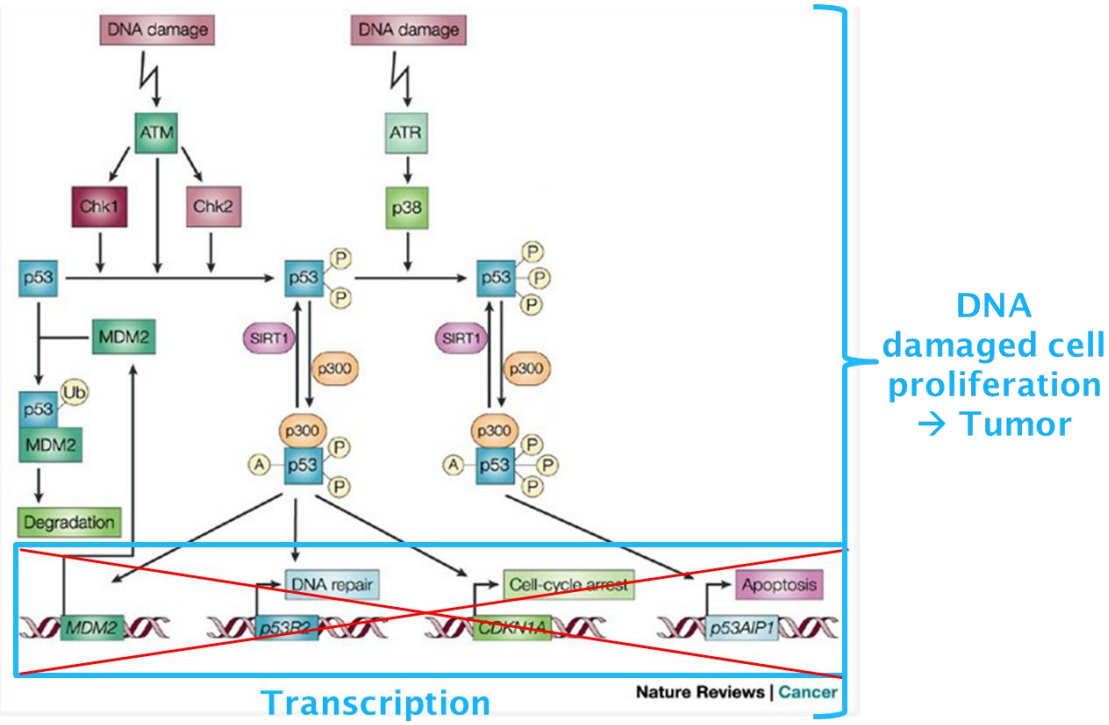


Figure 3-9: p53 mutation inducing cancer (48).

3.5.1 p53 DBD mutations

Mutations in the DBD of p53 are classified into three groups depending on their phenotypes (77):

- DNA-contact mutants with only minor effects on folding and stability.
- Substitutions resulting in local changes, mainly in proximity to the DNA-binding surface, which are destabilized by < 2 kcal.mol^{-1} relative to the wild-type.
- Mutations causing global unfolding of the β-sandwich, which are destabilized by > 3 kcal.mol^{-1}.

Six amino-acid residues are most frequently mutated in human cancer (about 40% of all missense mutations). They are located in or close to the DNA-binding surface. These residues have been classified as "contact" (Arg248 and Arg273 → inhibit DNA interaction) or "structural" (Arg175, Gly245, Arg249 and Arg282 → destabilizing mutants) (78). The mutations frequently found are: R175H, G245S, R248Q, R249S, R273H, and R282W.

The “structural” mutants make up ~30% of the clinical cancer cases reported and “contact” mutants account for ~20% of the cases.

3.5.2 Mutant p53 gain of function

p53 mutations not only abolish normal p53 functions but also contribute to the mechanism of maintenance and growth of the tumour by inducing resistance to cancer treatments, genomic instability and increasing metastasis. This gain of function of p53 mutants can be explained by different mechanisms (79).

3.5.2.1 p53 mutants bind transcriptional regulators and other proteins

Wild-type p53 interacts with several transcription factors, allowing the transcription/inhibition of genes whose promoters lack p53 binding sites. This interaction is also seen for some p53 mutants, but the transcriptional outcome is different from that of the wild-type (80–85).

NF-Y, a heterotrimeric transcription factor binding the CCAAT consensus motif, is implicated in the modulation (transcription/inhibition) of several genes controlling different phases of the cell cycle in case of DNA damage. In a normal cell with DNA damage, NF-Y will recruit p53 WT which in turn will bind to HDACs and inhibit the transcription of the genes. In contrast, p53 mutants R175H will preferentially bind to the p300 protein and induce the transcription of the genes implicated in the cell cycle. The recruitment of p300, triggering acetylation, is thought to be the key process involved (84). The NF-Y study highlighted the opposing functions as elicited from WT and mutant forms of p53, while binding to the same partners.

In breast cancer, mutant forms of p53 (R273H, R280K, V143A, R175H etc) play a role in the mevalonate pathway (implicated in breast cell architecture) and appears to work with the SERBP family of transcription factors (83).

When p53 mutants are introduced into cells latently infected by the HIV virus, an active HIV replication results. This is thought to result from the cooperation of mutant p53 (V143A, R175H, R249S and R273H) with Sp1 transcription factor; p53 mutants are able to activate the transcription of genes containing the Sp1 promoter (82).

Several of these transcription factors were found to bind p53 mutants towards the C-terminal domain. This recruitment might be facilitated by proteins or induce the binding of p53 mutants to proteins (such as p300 and Pin1) which facilitate the progression and aggressiveness of the tumours (86).

3.5.2.2 DNA structure selectivity of p53 mutants (87)

The binding of p53 WT to a sequence specific DNA is fundamental for its activity as a transcription factor. This characteristic, which confer p53 its pluripotency, is conserved in the p53 mutants. DNA consensus sequence specific to p53 mutants remain under investigation. On the other hand, p53 mutants were shown to preferentially bind to non-B DNA in a conformation specific manner (88). This was called the DNA structure-specific binding (DSSB) and in vivo was observed to involve MAR/SAR sequences and trinucleotides (CTG·CAG). This binding is thought to be driven by the DBD and the C-terminal domains.

It is thus clear that there exist a variety of different mechanisms by which p53 mutants gain more functions and induce progression and resistance/aggressiveness of cancer cells, thus highlighting the need to restore p53 wild type function and more specifically p53DBD stability.

3.5.3 Destabilized mutant p53 rescue

The restabilization of mutant p53 can be achieved in: either by introducing stabilizing mutations into the p53 core domain or by the addition of stabilizing agents (89):

- The “chaperone strategy” to rescue oncogenic mutants is based on the theory that conformationally destabilized mutants can be rescued by generic small molecules binding the native but not the denatured state.
    - CDB3: a nine-residue peptide derived from the crystal structure of p53 core domain in complex with 53BP2 stabilizes various oncogenic mutants (90).
    - PRIMA-1: increases p53 levels in vivo (91).
- The stabilizing mutations act as second-site suppressors for various cancer-associated mutations. The most stable substitutions have been combined in a quadruple mutant (M133L, V203A, N239Y and

N268D) (77). This combination of mutations induces a super-stable p53 DBD mutant which is stabilized by 2.65 kcal.mol^{-1} but retains wild-type DNA-binding function. The main contributions come from the N239Y and N268D mutations.

3.5.4 p53 mutant V143A

In this thesis, we focus on the computational studies of one such mutant of the p53 DBD: V143A. The valine 143 is located in the strand S3 of the β sandwich. The side chain of V143 is deeply buried and participates in the packing of the hydrophobic core. Mutation of V143 to alanine dramatically destabilizes the fold (by 3.34 kcal.mol^{-1}) and is accompanied by the creation of a cavity (Fig.3.9). This mutation is temperature sensitive [43], at 25 and 32.5^{0}C it still binds to DNA and presents a transcriptional activity stronger than the wild-type p53, but at 37.5^{0}C both DNA binding and transcription are compromised. This change in activity is proven to be due to a change of conformation, as at 25-32.5^{0}C V143A still binds to PAb1620 (antibody specific to wild-type p53 conformation) and at 37.5^{0}C it immunoprecipitates with PAb240 (antibody specific to mutant p53).

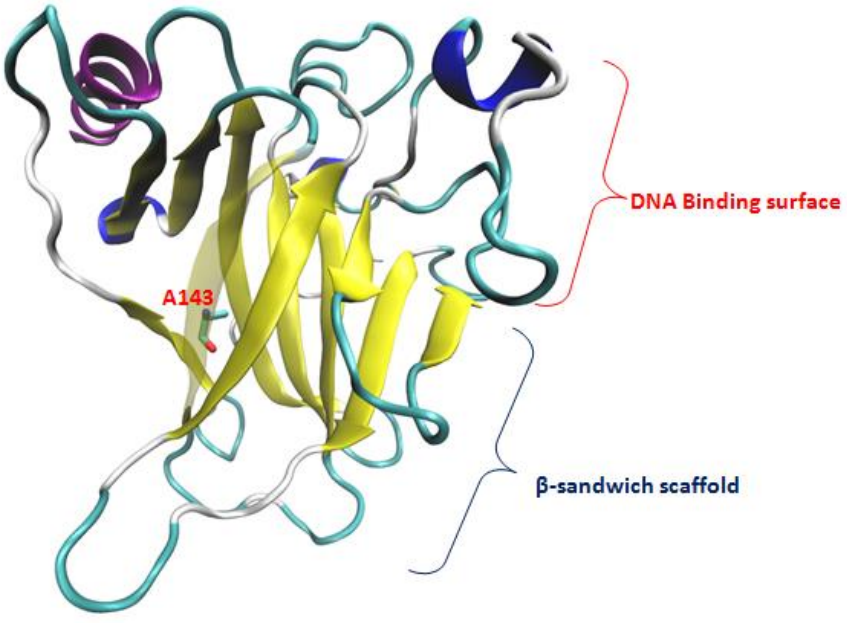


Figure 3-10: Valine 143 mutated to Alanine.
VMD visualisation of the V143A mutant structure (PDB code 2J1W). Chain A in cartoon coloured by secondary structure; the strand S3 is transparent in order to highlight Alanine 143 which is shown as licorice.

At 37^{0}C, V143A is estimated to be about 80% denatured. But even at low temperature the DNA binding specificity is different from that of WT p53; for example V143A induces transcription of p21 protein but not of Bax protein at 24 and 30^{0}C (89, 92) (WT p53 itself presents a lower binding affinity to Bax promoter compared to other response elements (RE) (93) and no binding at all to Bax at 24^{0}C while still binding to other oligonucleotides (92)). Bax RE does not contain a typical consensus sequence (92–94), as can be seen in Fig. 3-11. Bax contains only one well defined pentamer (red), the three other pentamers surrounding it (underlined and overlined) are described as being imperfect as not all the base pairs match the DNA consensus sequence.

p21 promoter

5′ CTCAACATGTTGGGACATGTTCCTTT 3′
3′ AAGAGTTGTACAACCCTGTACAAGGA 5′

Bax promoter

5′ TCACAAGTTAGAGACAAGCCTGGGCGTGGGCTATATT 3′
3′ AGTGTTCAATCTCTGTTCGGACCCGCACCCGATATAA 5′

Figure 3-11: Promoters on which p53 binds.
p21 promoter (92, 95, 96) and Bax promoter (92, 94).

The differential binding observed between mutant and WT p53 is thought to be due to a long range effect (LRE). Residues trying to fill the V143A cavity cause chemical shift changes in almost all residues in the β sandwich, inducing changes in residues in helix H2 and loop L3 which affect the DNA-binding region and so the specificity of DNA binding (97, 98) . V143A has a dominant negative effect. Indeed in several studies the transfection of V143A mutant in cells expressing WT p53 exhibits a decrease of the normal functions of p53 (apoptosis, G1 arrest) (99–101). V143A mutant is rescued by the suppressor mutant N268D. This second site suppressor stabilises V143A mutant by 1.1 kcal.mol-1 and retains the ability of mutant p53 to bind to DNA (89).

The V143A mutant structure was determine experimentally (78) after including four extra stabilising mutations (M133L, V203A, N239Y and N268D), which stabilise the mutant by 2.6 Kcal.mol^{-1}. In this study, the conformation of several DBD destabilising mutants is studied while combined with the four stabilizing mutations cited previously. It has been shown that V143A creates a cavity with

a volume of 47 $Å^3$ in the hydrophobic core of the β-sandwich but conserves its overall structure (78).

3.6 p53 simulation studies

There are several studies of p53DBD using atomistic computer simulations. The importance of Zinc in p53DBD has been investigated by Nilsson et *al.* (102). In a study of four systems: p53DBD, p53DBD-apo (no zinc), complex (p53DBD-DNA) and the apo-complex (p53DBD-DNA no zinc), L. Nilsson showed the importance of the zinc not only for the DNA binding domain stability but also in DNA binding specificity. Indeed, it has been shown that the apo-DBD is prone to aggregation mainly due to the loss of H1 helix, which will drive the L2 loop to become an aggregation site. It is thought that ZN^{2+}, by being attracted to the DNA phosphate backbone, facilitates the insertion of Arg248 into the minor groove.

In another study, the stability of p53 was examined through simulations. p53 DBD is thought to be unstable (melting temperature ~ 42-44^{0}C) compared to its homologues p63 and p73 (~ 61^{0}C). Fersht et *al.* (103) highlighted the presence of two polar residues in p53 (Tyr236 and Thr253) that are apolar (Phe238 and Ile255) in its homologues (p63/73) which are more stable than p53. They mutated the Tyr-Thr pair in p53 to Phe-Ile and found that this lead to an increase in the stability of the new p53 by about ~1.6 kcal.mol^{-1}. Verma *et al.* carried out MD simulations of the three structures (p53, 63 and 73) and their double mutants (Tyr236Phe and Thr235Ile for p53 and Phe238Tyr and Ile255Thr for p63/73). The computational results showed an increase of p53 double mutant structure fluctuations as well as an increase of the phase space explored compared to p53WT and an inverse behaviour for p73 (decrease of structure fluctuation and phase space explored). They concluded that the gain in flexibility is likely linked to the gain of stability.

In a separate simulation study, Lim *et al.* proposed a mechanism by which p53 R273H loses its capacity to bind DNA and how it is rescued by the T284R mutation (104). They first examined the binding of wild type p53DBD to DNA and found that it involved hydrogen bonds between side chains of K120, C277 and R280 with base oxygen or nitrogen of the DNA in the major groove, and non-specific binding interactions arose from interactions made by salt

bridges/charge interactions between K120, S241, R273, A273 and R283 and the DNA backbone phosphate oxygens. Binding free energy calculations confirmed a weaker binding of p53R273H to DNA compared to wild type p53, and a binding affinity comparable to the wild type for the double mutant p53R273H/T284R. The DNA affinity loss is thought to be triggered by the loss of the salt bridge between R273 and D281 which disrupts a hydrogen-bonding network involving R273, D281 and R280. And the rescue by T284R is shown to be due to its ability to bind DNA (which squeezes some water out of the DNA binding surface) and enable R280 and K120 to regain native major groove contacts that are thought to be essential for sequence specific binding. Later, the Lim group examined the mechanism underlying the observed loss of DNA binding in the p53R273C mutant (105). Experimental studies had concluded that the DNA binding loss was due to the loss of charge-charge interactions between the arginine side chain and the DNA phosphate backbone. Lim et al found that along with the loss of the Arg273-Asp281 salt bridge that is responsible for the p53R273H/C DNA binding loss, there is also reduced flexibility in the mutant structures.

The crystallisation and molecular dynamics study of another hot spot p53 mutant, R282Q, highlighted an inversion of the flexibility in the L1 and L3 loop of the DNA binding surface (106). R282Q is located in the helix H2 of the DNA binding surface, and is the only mutant crystallised without rescue mutations or stabilising molecules. The mutation induces a salt bridge loss between Arg282 and Glu286 which induces p53DBD structure instability. The hydrogen bond between Glu282 and the L1 loop appears to lead to loss of flexibility of the loop and the L3 loop presents two conformations, one rigid and identic to wild-type p53DBD (presenting Ser241 to the DNA) and a second one in which the orientation of Ser241 is not compatible with DNA binding.

3.7 Summary

To date, no long MD simulations were performed on p53 protein and most of the mutants studied *in silico* were 'contact' mutants (104, 106). In this study, p53 DNA binding domain (p53DBD) dynamics was investigated. Experimental data (p53DBD NMR and Xray structure) were collected from the protein data bank (PDB) and inspected using principal component analysis (PCA). The insights gathered from these studies and our analysis allowed us to propose a method to distinguish the apo p53DBD from the DNA bound conformation and to highlight the limitation of the p53DBD model to study its DNA binding. A computational model of p53DBD was then proposed, long range (from 500 ns to 1 μs) molecular dynamics (MD) and accelerated molecular dynamics (aMD) simulations were performed, and p53DBD stability was analysed. Two cases were explored:

(i) The effect of Michael acceptors (MAs).

(ii) The effect of a temperature sensitive p53DBD mutant, V143A.

MAs were shown to enhance DNA binding of the p53DBD WT at low concentration and to inhibit it at high concentration (107). These molecules tested on p53DBD mutants (mainly hot-spot mutants: R175H, Y220C, G245S, R249S and R282W) showed the same DNA binding enhancement effect and an increase of their stability (108). While the implication of the p53DBD cysteines in the mechanism of action of the MAs has been demonstrated (91, 108, 109), which specific cysteines are implicated and how does these cysteines alkylation stabilise p53 and enhance its DNA binding is still under investigation. Based on Langridge-Smith et *al.* experimental study, we built several p53DBD alkylated model, on which we applied MD and aMD. The structural analysis of the resulting p53DBD conformations showed that the alkylation of C182 might be implicated in the DNA binding enhancement at low concentration of MAs (by stabilising the H1 α-helix) and the alkylation of the C277 might inhibit DNA binding at high concentration (by preventing two direct DNA contacts).

p53DBD V143A still binds to some promoters and presents a transcriptional activity stronger than the wild-type at low temperature (89, 92), while at body temperature it has been reported to be 80 % denatured. The MD and aMD simulations (500 ns) at low temperature and the structural analysis of the

resulting p53DBD WT and V143A conformations showed that V143A had a more stable H1 α-helix compare to the WT. This was thought to allow a more stable p53 V143A dimerization while binding to the DNA and explains its strong transcriptions compare to the WT at low temperature. On the other hand, the simulations at high temperature (310K and 323K) of p53DBD WT and V143A did not highlight p53 destabilisation. This is thought to be due to the time scale of our simulation (ns), which is too short to allow large scale motions (folding/unfolding) to occur.

Chapter 4. Principal Component Analysis (PCA) of the p53DBD experimental structures

4.1 Introduction

To date there are 70 experimental structures of human p53DBD available in the Protein Data Bank (PDB), 68 by X-Ray diffraction and two by NMR. The experimental structures of p53DBD exist in both wild type and mutant states and in various forms including polymeric, DNA-bound, ligand-bound, and protein-bound (see Table 4.1). As stated in section 3.2, wild type and mutant p53 have several functions, and most of these functions implicate the DNA binding domain. PCA of monomeric p53DBDs and of the DNA sequences to which they bind will allow us to highlight the conformational mechanisms involved in these functions.

PCA is routinely used for analysing multiple conformations of a protein (such as are available from MD simulations or X-ray/NMR) to extract high amplitude motions. Multivariable PCA performed on NMR structures of the native globular structure of prion protein (PrP) highlighted subdomain differences between the different pathogenic mutants (110). PCA was used on Ras isoforms' crystallographic structures and MD simulations to examine if the conformational differences between these structures is related to their functional difference (111). In that study, PCA applied to Ras X-ray structures was shown to be able to distinguish between structures binding to different ligands (GDP or GTP) as well as between different Ras mutants.

A recent study combining PCA of human and mouse experimental structures (RMN and X-rays) and molecular dynamics (MD) simulations of the human p53DBD in its monomeric form, showed that the recessed L1 conformation explored in some experimental structures with binding to DNA is accessed easily in the MD simulations (112). The observation that this recessed L1

conformation is not commonly observed in the experimental structures was explained to result from the crystallization procedures and the associated effects of crystal packing. To rescue certain destabilised mutants, the author suggested some stabilising mutations in the L1 loop.

In contrast, in this study, a more selective PCA applied to only human p53DBD experimental structures, allows us to distinguish between apo, DNA bound, and antigen bound p53DBDs. The difference between these structures is mainly located on the loops: L1 while p53 is bound to DNA and L3 when it is bound to the LTag antigen. The L1 flexibility has been highlighted in recent studies using an extended model of p53 (DBD and tetramerisation domain), that we will call engineered p53DBD (see Table 4-1) (96, 113, 114). Experimental studies showed that the full p53 protein has a higher DNA binding affinity than p53DBD (micromolar affinity for the DBD and nanomolar for the structure with both DBD and TD) (115). To investigate the mechanism allowing a better affinity of the full p53 protein, Halazonetis et *al.* proposed the engineered p53DBD structure. In these studies the L1 loop adopts a recessed conformation while binding to the DNA. The recessed L1 loop observed on the engineered structures (and not on the remaining experimental structures) is thought to be due to the presence of the tetramerisation domain of p53, which explains its absence from the remaining experimental structures found in the PDB. According to Halazonetis et *al.* p53 undergoes an induced fit mechanism while binding to DNA (96, 113, 114), which is responsible for the binding specificity of p53 to the DNA. In their studies they also show that the recessed L1 loop is never present in the p53DBD not bound to DNA, only the mutations of S121F and V122G in the L1 loop allow them to detect the recessed L1 conformation in the apo-p53DBD (114). To test the binding specificity of the experimental structures of p53DBD we performed a DNA PCA. This analysis shows DNA bending, specific to protein-DNA bound, only on the engineered structures.

4.2 Materials and methods

4.2.1 PDB structures

To perform PCA on the experimental structures (X-rays and NMR) of p53DBD, it is necessary to ensure that the structures have the same number of atoms. There are 70 human p53DBD structures available in the PDB (as of April 2014); however some of them have missing residues, and hence were excluded from our analysis. This resulted in 54 structures. This set of proteins contains wild-type and mutant structures which are either free or bound to DNA/protein/small molecule (Table 4-1). Two NMR structure exist for the p53DBD (PDB codes 2FEJ and 2MEJ) containing 36 and 20 conformers respectively. Among the 70 p53DBD structures 14 are bound to DNA and used to perform the PCA of DNA.

Among the 52 X-ray structures selected (see Table 4-1), we can distinguish four main groups: WT, mutant, engineered structures and mutated engineered structures (containing two mutations on the L1 loop). The WT and the mutant structures are composed of the DBD while the engineered structures (3Q01, 3Q05, 3Q06, 3TS8) and the mutated engineered structures (4ZMI and 4ZMR) contain the DBD as well as the tetramerisation domain linked together via a shortened linker sequence. These four groups were then divided in three sub-groups depending on if they were unbound i.e. apo structures, bound to a DNA consensus sequence or bound to a protein/small molecule. The label of the X-ray structures in the different PCA analyses will be coloured depending on the subgroup to which the structure belongs (colours indicated in Table 4-1). Also two NMR structures were used in this study, both containing only one monomer of p53. 2FEJ is apo while 2MEJ is bound to the Bcl-2-like protein 1. These two structures will be labelled by numbers indicating their source (2FEJ → red coloured numbers and 2MEJ → blue coloured numbers).

Wild type structures					
Apo (Blue)		Bound to DNA (Purple)		Bound to proteins (Green)	
PDB code	Resolution (Å)	PDB code	Resolution (Å)	PDB code	Resolution (Å)
2OCJ	2.05	1TSR	1.91	2H1L	3.16
2XWR	1.68	1TUP	2.20	1GZH	2.60
2YBG	1.90	3KMD	2.13	1KZY	2.50
		2AC0	1.80	1YCS	2.20
		2ADY	2.50		
		2AHI	1.85		
		2ATA	2.20		
		4HJE	1.91		
Mutant structures					
Apo (Red)		Bound to DNA (Pink)		Bound to proteins or small molecules (Brown)	
2WGX	1.75	3D0A	1.80	2VUK	1.50
2PCX	1.54	4IBU	1.70	2X0U	1.60
2J1W	1.80			2X0V	1.80
2J1X	1.65			2X0W	2.1
2J1Z	1.80			3ZME	1.35
2J20	1.80			4AGL	1.70
2BIM	1.98			4AGM	1.52
1UOL	1.90			4AGN	1.60
4IBQ	1.80			4AGO	1.45
4IBS	1.78			4AGP	1.50
4IBT	1.70			4AGQ	1.42
4IBY	1.45				
4IBZ	1.92				
4IJT	1.78				
4KVP	1.50				
4L09	2.50				
4L0E	1.85				
4L0F	2.00				
Engineered structures					
Apo (Grey)		Bound to DNA (Orange)			
3Q01	2.10	3Q05	2.40		
		3Q06	3.20		
		3TS8	2.80		
Engineered structures containing L1 loop mutations					
Apo (Yellow)		Bound to DNA (Black)			
4MZI	1.25	4MZR	2.90		

Table 4-1: List of the X-ray structures used to perform principal component analysis (PCA).
The colours in parentheses are used in the structure projections along the different PCs.

4.2.2 Structural preparation and PCA

The focus of this study is the monomeric form of p53DBD. In this context, a PDB file for each chain of the selected structure was created (totalling 202 chains in the 54 structures, 52 X-ray and 2 NMR). As several structures have missing residues at their termini, only residues from Val97 to Glu287 of the p53DBD (normally defined as extending from residue 94 to 293) were considered. The chain A of the PDB structure 2XWR, a wild-type structure resolved at 1.68 Å (one of the most highly resolved DBD structures) was chosen as the reference structure for superimposition. Each structure was superimposed on the alpha carbons of the reference structure and a general PDB file containing all the structures was produced. Only the alpha-carbons were kept in this set of PDB files to perform PCA. The structure alignment was performed using the Visual Molecular Dynamic software (VMD) (29).

In the case of DNA, the torsional angles of their sugar-phosphate backbone (α, β, ϒ, δ, ε, ζ and χ), extracted using 3DNA (116), were considered for PCA; for each structure the torsional angles of all the possible dinucleotides were extracted (Fig. 4-1) and adjusted such that differences from their mean values lie in the range -180º to +180º, as described in (117) which results in 421 dinucleotides obtained from the 14 p53 DNA sequences. PCA has been performed on the dinucleotides using 14 torsional angles (7 per nucleotide).

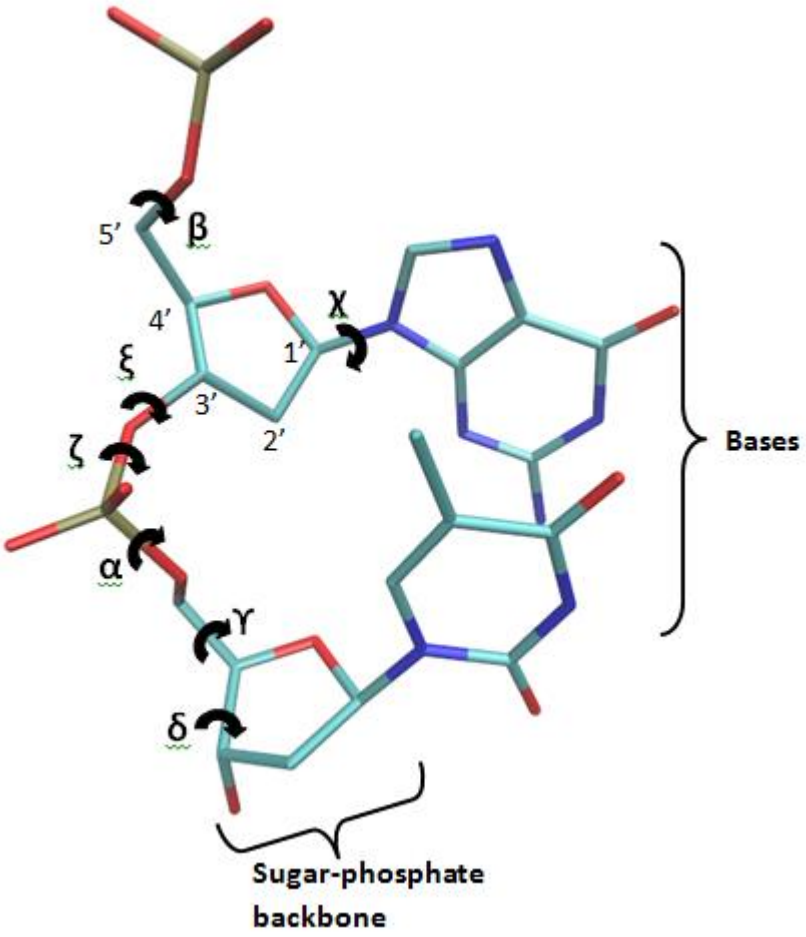


Figure 4-1: Schematic representation of DNA.
Torsional angles of the DNA sugar-phosphate backbone used to perform PCA.

The PCA technique involves the calculation of the covariance between all dimensions and for a n dimensional system, $\frac{n!}{(n-2)!*2}$ different covariances will be calculated. From this covariance matrix, the eigenvectors (corresponding to the principal components) and their eigenvalues are calculated. For a n dimension system, n eigenvectors exist. These eigenvectors are then ordered according to their eigenvalues. Only the first few eigenvectors (direction of a common motion in a system) with high eigenvalues (intensity of the motion) are studied, as the majority are of little importance. Within our system, the atom projection within each eigenvalues showed that the most relevant motions were captured within the three first eigenvalues.

The PCA was performed using the package Bio3D via the R statistical programming language (118), which allows the analysis of protein structure, sequence and trajectories.

4.3 Results

4.3.1 Overview of the possible conformations of p53DBD monomers

4.3.1.1 Free p53DBD conformations observed in an NMR structure: 2FEJ

PCA was initially carried out on the 202 chains of the 52 X-ray and the 2 NMR structures chosen. The graph of the variance captured by each principal component (PC) (Fig. 4-2A) shows that the first three PCs represent 59.8 % of the major motions in the DBD while it takes up to 8 PCs to cover 80% of the variance in motions. The structures projected along these three PCs (Fig. 4-2B) show the regions of p53DBD that are the sites of these motions. It is apparent that PC1 captures motions that are globally spread over the p53DBD structure, with loops L1, L2 and L6 showing higher amplitudes. In contrast, PC2 and PC3 show motions mainly along L1 and L3 loops, with PC2 displaying a comparable variance on the two loops while along PC3, the L3 loop shows more variance than L1. To examine the contributions of individual structures to these variances, the 202 p53DBD structures were projected onto the three PCs (Fig. 4-3A and B). The projection along PC1 and PC2 (Fig. 4-3A) shows 4 clusters. The NMR structures 2FEJ (red circled) cluster separately from the crystallographic structures and 2MEJ along PC1. Along PC2, there are three clusters, the central cluster (blue circled) includes most of the X-ray structures and 2MEJ, a top cluster (green circled) formed by all the 2H1L X-ray structures (which bind to the L-Tag antigen) and a bottom cluster (orange circled) containing several chains of the engineered structures. The projection along PC3 and PC2 (Fig 4-3B) displays 3 main clusters, most of the X-ray structures and the two NMR structures cluster together (blue circled) while 2H1L (green circled) and some chains of the engineered structures (orange circled) cluster separately from the central cluster along PC2 and PC3.

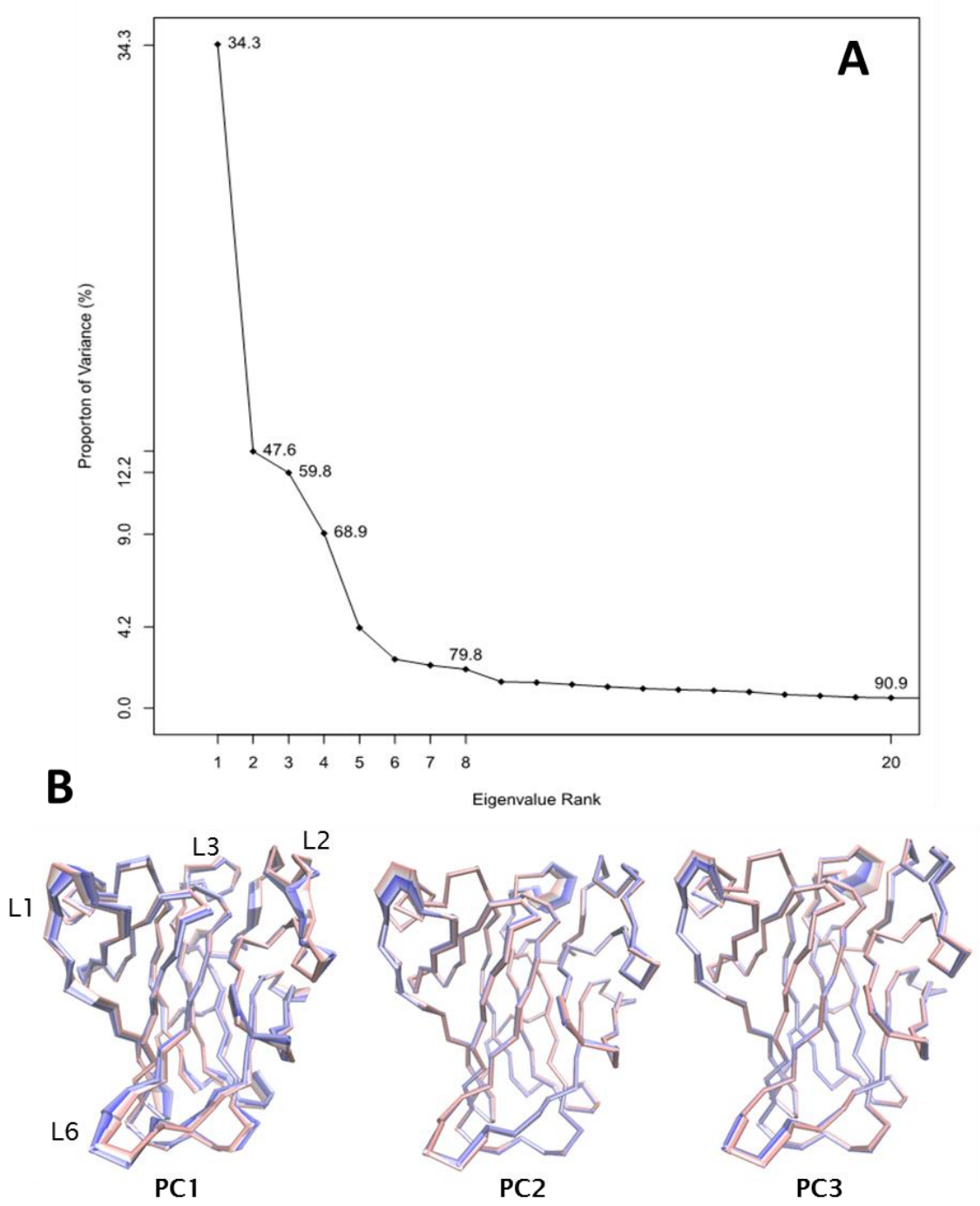
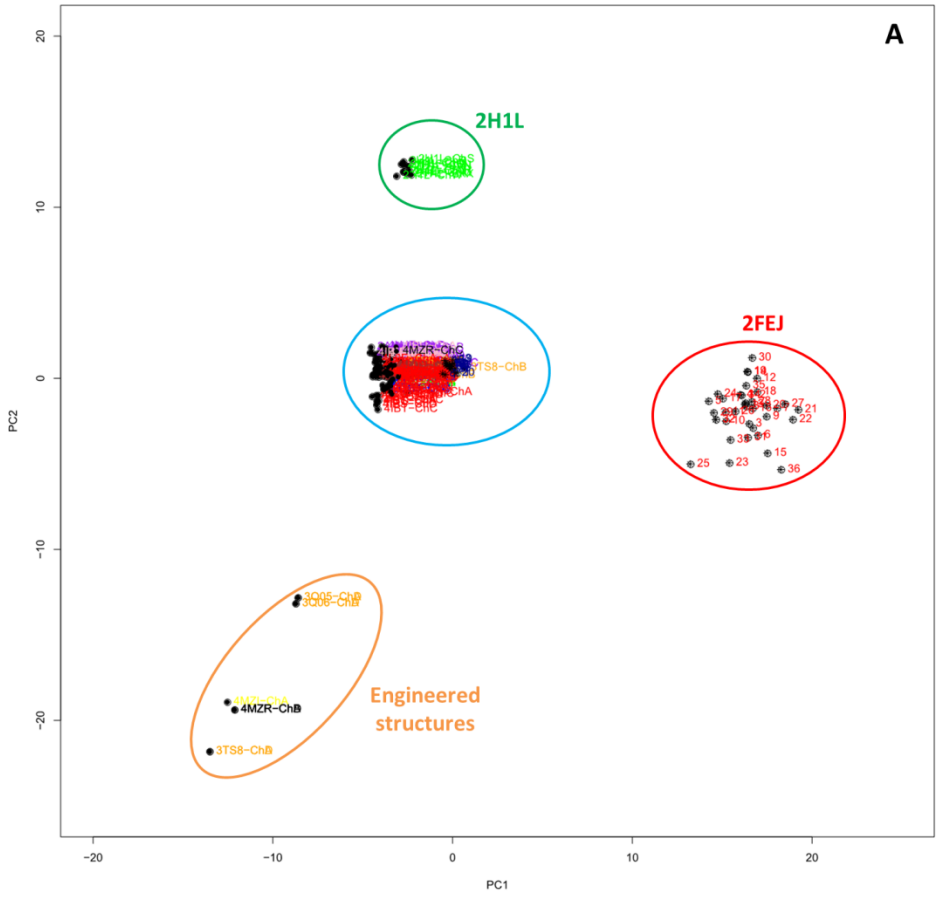
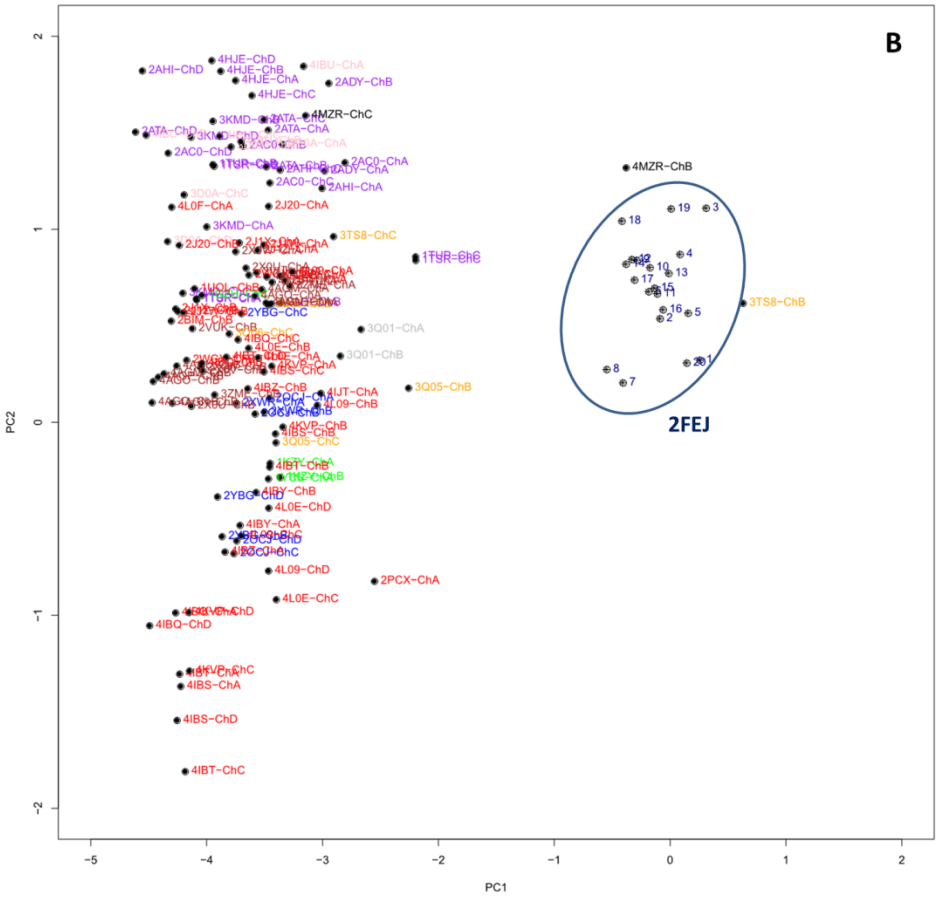


Figure 4-2: All monomers PCA.
A) Proportion of variance (in %) captured by each principal component (PC) from a data set of 202 monomeric p53 X-ray and NMR structures. The 3 first PCs contribute to 59.8% of the variance of the dataset. **B)** Trace representation of the motion projections along the three first PCs (PC1, PC2 and PC3) of the 202 structures using VMD. The colouring range of the motions goes from red, to white, to blue as function of time (timestep option on VMD), red being the early motion.


A
2H1L
2FEJ
Engineered
structures
4MZR-ChB
3TS8-ChB
PC1
PC2


B
4MZR-ChB
3TS8-ChB
2FEJ
PC1
PC2

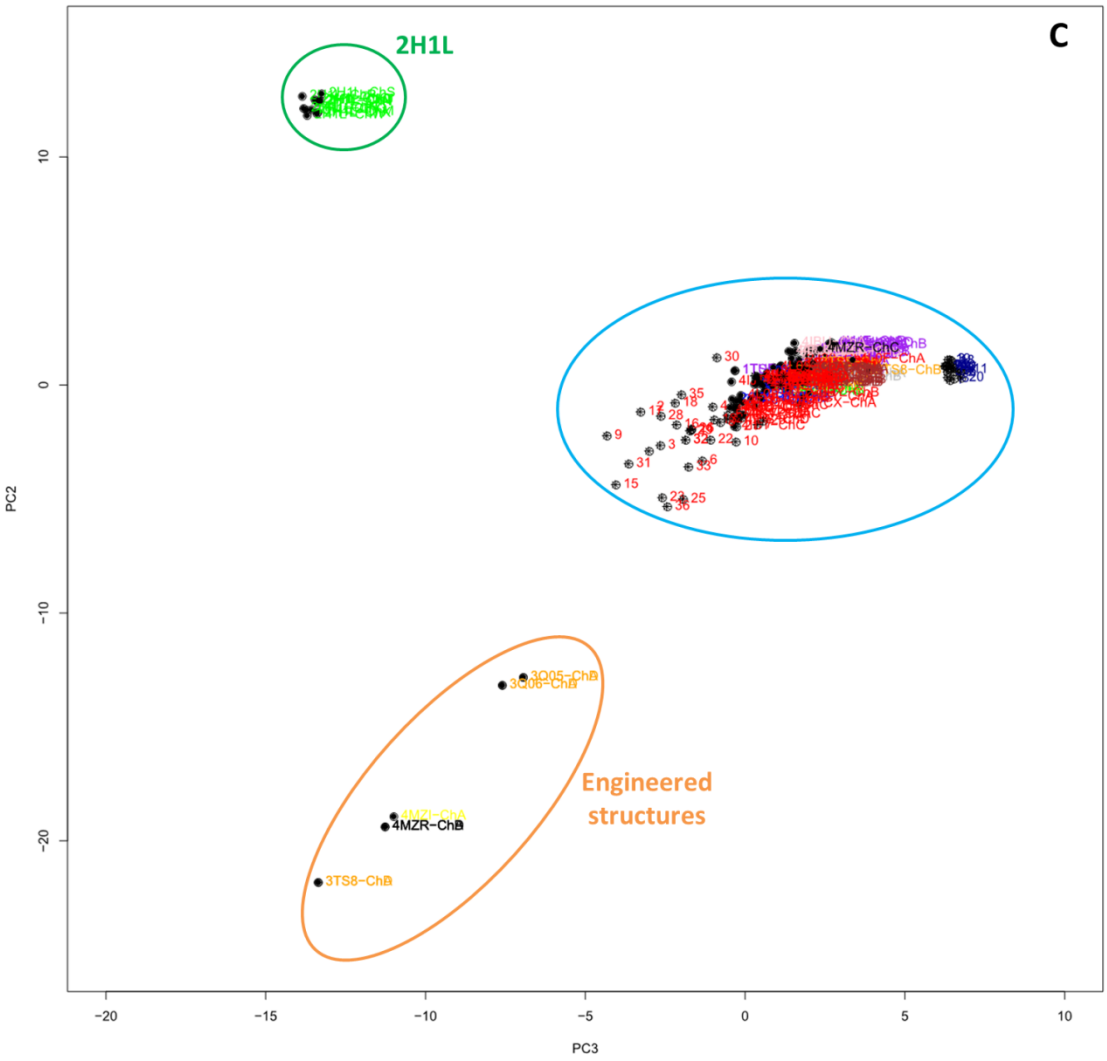
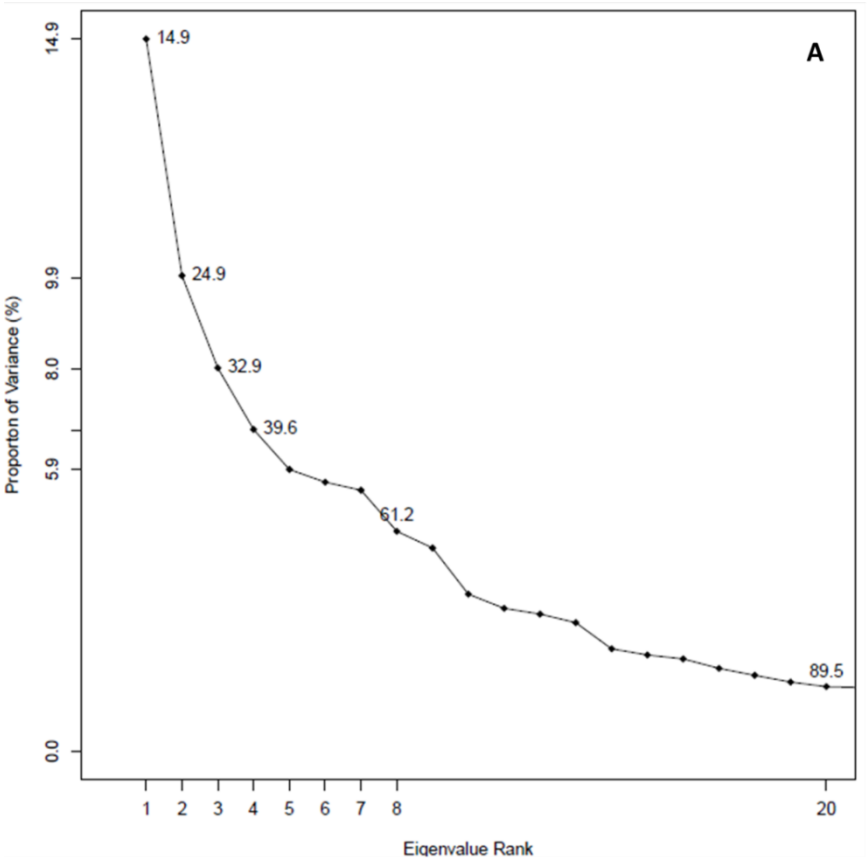


Figure 4-3: Structure projections along the three first PCs.
A) Along PC1 and PC2. **B)** Zoom of the PC1-PC2 main cluster (blue circle in A). **C)** Along PC3 and PC2. X-ray structures are labelled by PDB name and chains and coloured according to table 1. NMR structures are numbered by conformers and coloured (2 FEJ containing 36 conformers → red numbering/circled and 2MEJ containing 20 conformers → blue numbering).

In a previous PCA study (112) the clustering difference between X-ray structures and the NMR structure 2FEJ was thought to originate in differences in compactness between the tighter X-ray and the more "loose" NMR structures. However, we find that the recent p53DBD NMR structure 2MEJ co-localises with most of the other X-ray structures in the central cluster (zoom Fig. 4-3 B). It is clear that the X-ray structures and 2FEJ separate along PC1 (Fig. 4-2B) which captures differences in motions mainly in the loops L1, L2 and L6. Two separate PCAs were then performed, one containing only X-ray structures and the other only 2FEJ. The three first PCs of these two sets, 2FEJ and X-ray

capture 32.9 % and 72.1% of the variance respectively (Fig. 4-4). The first 2 PCs of 2FEJ capture largely motions in L1 and L2 while for X-ray, they capture motions mainly in L1 and L3 loop (Fig. 4-5A and B). Although both sets of structures show motion of L1, these motions appear to correspond to different possible conformations. The open/closed conformations of the L1 loop observed in the 2FEJ structure (correspond to two different geometry of L1, see Fig. 4-6), which are possible conformations of an apo p53DBD monomer (103) and the recessed/extended L1 loop conformations seen in the engineered structures thought to be due to DNA binding (will be developed later on the paper). The L1 motion captured along PC1 of the 2FEJ structures shows clusters of the open and the closed conformations (Fig. 4-6). It has been reported (103) that the closed conformation of 2FEJ is similar to p53DBD bound to DNA and the open is similar to one of their mutant p53DBD structures. The open conformation is thought to be possible due to the lack of hydrogen bonds involving the L1 loop. A hydrogen bond analysis using HBonanza (31) was performed on the open and closed conformations of 2FEJ (Fig. 4-7). Both group of conformations present three hydrogen bonds in the loop L1, two within the loop (Lys120-Ala119 and Thr123-Cys124) and one between L1 loop residue and a residue located in the turn between S2'-S3 (Val122-Gln136). The occurrence of the three hydrogen bonds is 7% in both groups, open and closed conformations. A visualisation of the different conformations show that these two conformations seem to be modulated by Ser121 which points towards the H2 helix in the open conformation and away from it in the closed conformation (Fig. 4-6).


A
14.9
24.9
32.9
39.6
61.2
89.5
Proporton of Variance (%)
0.0
5.9
8.0
9.9
14.9
1 2 3 4 5 6 7 8 20
Eigenvalue Rank

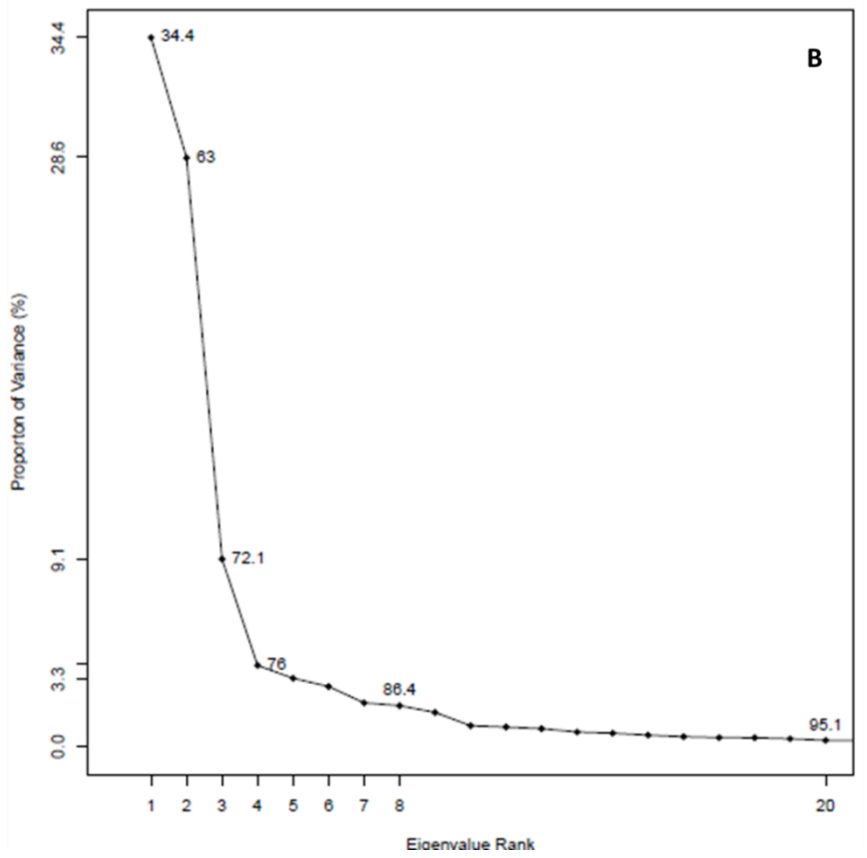

B
34.4
63
72.1
76
86.4
95.1
Proporton of Variance (%)
0.0
3.3
9.1
28.6
34.4
1 2 3 4 5 6 7 8 20
Eigenvalue Rank

Figure 4-4: Proportion of variance (in %) captured by each principal component (PC).
A) 2FEJ NMR structures consisting of 36 conformers. The first three PCs capture 32.9% of variance **B)** all X-ray structures, 146 DBD monomers. The three first PCs capture 72.1% of variance.

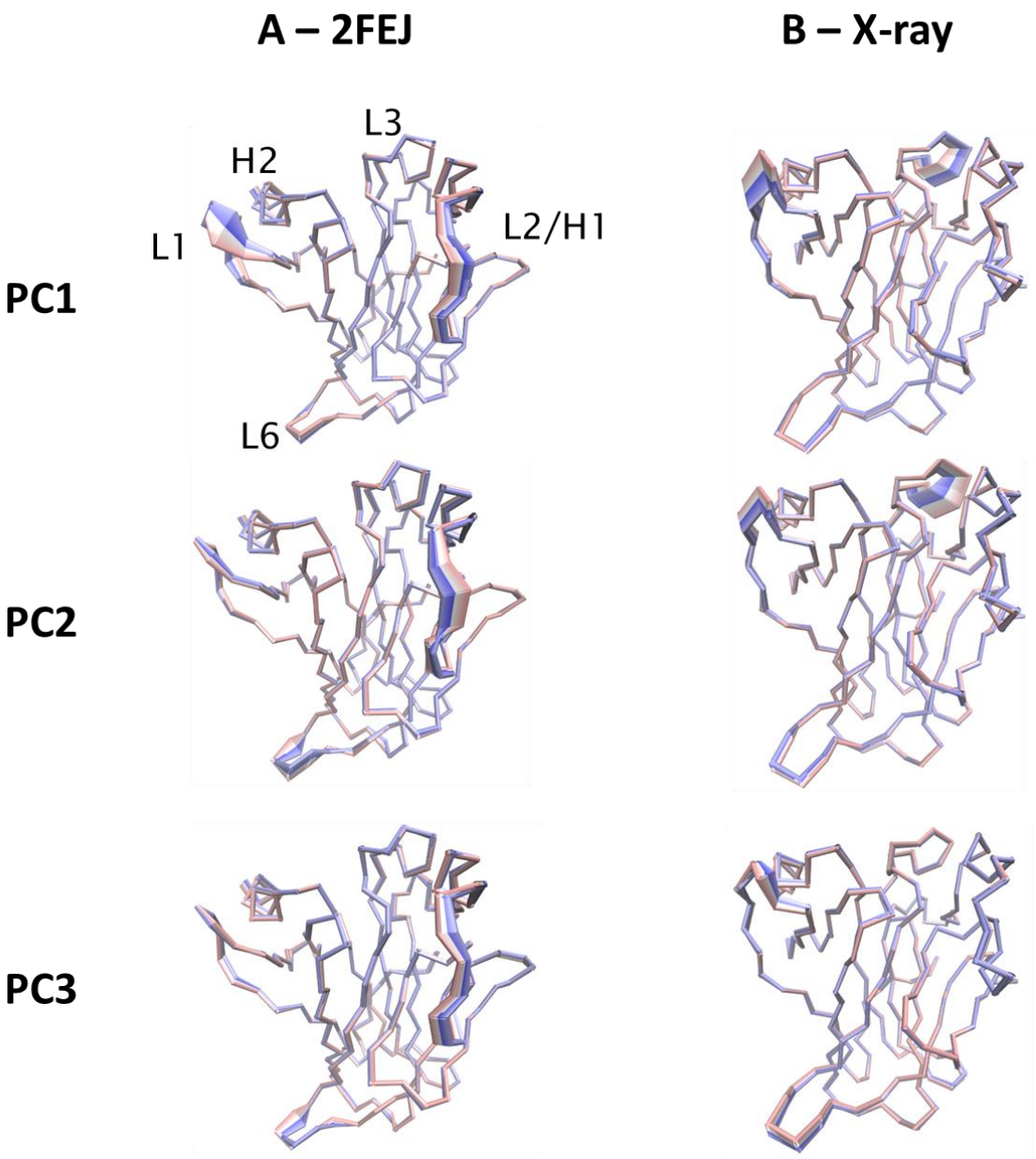


Figure 4-5: Trace representation of the motion projections along the 3 first PCs (PC1, PC2 and PC3) using VMD.
A) 2FEJ, 36 conformers. **B)** All X-ray monomers, 146 monomers. The colouring range of the motions goes from red, to white, to blue as function of time (timestep option on VMD), red being the early motion.

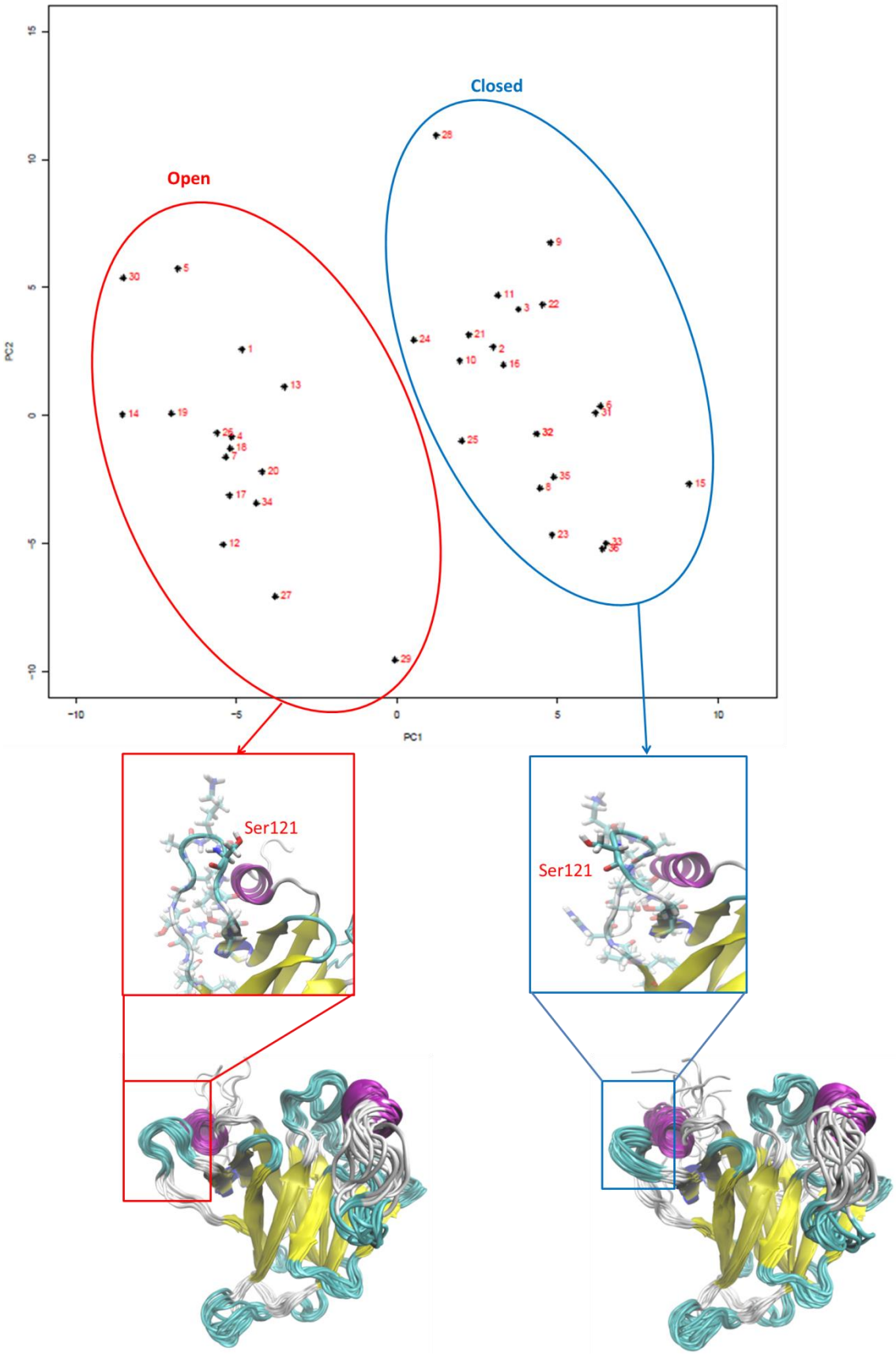


Figure 4-6: 2FEJ, structure projections along the two first PCs

The 36 2FEJ conformers projected along PC1 and PC2 cluster according to L1 loop conformation: open against closed. Open conformations (structure number: 1, 4, 5, 7, 12, 13, 14, 17, 18, 19, 26, 27, 29, 30, 34) and closed conformations (structure number: 2, 3, 6, 8, 9, 10, 11, 15, 16, 21, 22, 23, 24, 25, 28, 31, 32, 33, 35, 36). VMD visualisation.

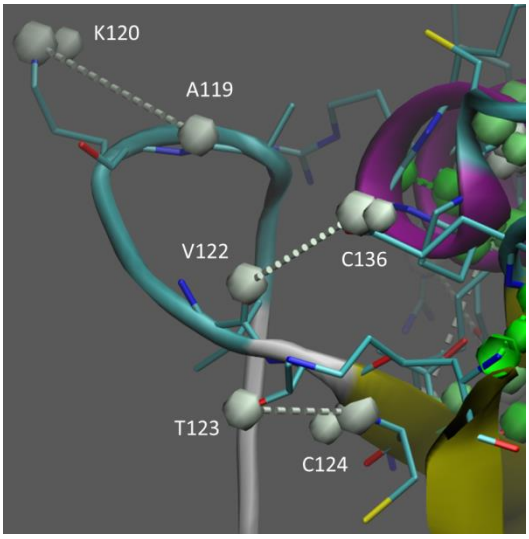


Figure 4-7: Hydrogen bonds in the L1 loop.
VMD visualisation after HBonanza process. p53DBD structure in new cartoon and coloured by secondary structure, the residues implicated in hydrogen bonds are in licorice coloured by name and the hydrogen bonding atoms in van der Waals. Hydrogen bond under 25% of occurrence are coloured in white, hydrogen bonds with an occurrence between 25-50% are coloured in pale green and the ones with an occurrence above 50% are coloured in dark green. the hydrogen bonds labelled within the L1 loop have an occurrence of 7%.

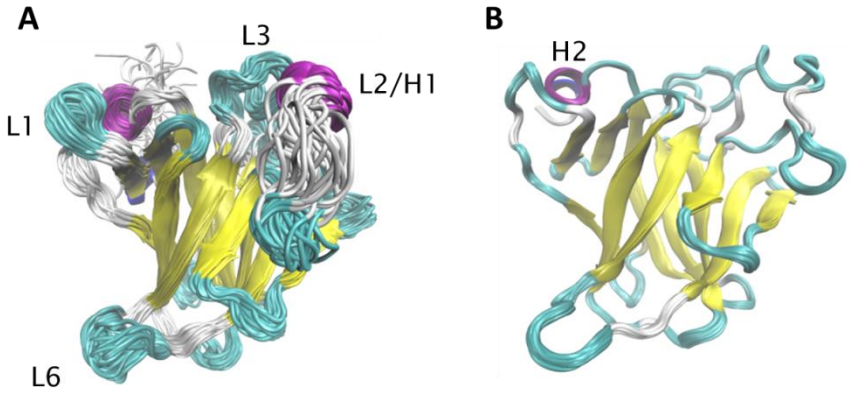


Figure 4-8: NMR structures comparison.
A) 2FEJ, 36 conformers. B) 2MEJ, 20 conformers. The structures were aligned on the reference structure (2XWR chain A) and were presented in cartoon and coloured by structure.

The first 3 PCs of 2FEJ also show that L2 undergoes significant loop motion (Fig. 4-5A). The L2 loop (residues 164 to 194) has a small helical insert H1 (residues 177 to 180) and most of the motion in this loop is observed in the

second part of the loop (after the H1 helix) from residues 181 to 194. Loops L2 and L3 are thought to be kept stable by a zinc atom (102). These two loops and a loop-sheet-helix motif form the DNA binding surface (119). When comparing 2FEJ with the recent NMR structure, 2MEJ, which binds to the Bcl-2-like protein 1 (Fig. 4-8), it can be seen that the 20 conformers of 2MEJ (Fig. 4-8B) are clustered tightly compared to the 36 conformers of 2FEJ (Fig. 4-8A). A PCA performed on the 20 conformers of 2MEJ show little flexibility of the structure (data not shown).The flexibility of the second part of loop L2 seems to be seen in all apo p53DBD monomers (as seen in 2FEJ and 3Q01) while these motions are clearly dampened when the region interacts with other proteins (*e.g.* Bcl-2-like protein 1).

4.3.1.2 L3 loop distortion

It is clear from Fig. 4-2 and Fig. 4-3 that the differences between X-ray structures are captured mainly along PC2 and PC3 (Fig. 4-3B). Projections of the structures along these two PCs indicates that 2H1L and some chains of the engineered structures cluster away from the rest of the p53DBD monomers (X-rays and NMR, see Fig. 4-3 C). These two principal components highlight variances in loops L1 (from residue 112 to 124) and L3 (from residue 240 to 250) (Fig. 4-2B). Comparisons (Fig. 4-9) of these structures after aligning onto the reference structure, 2XWR chain A, show that 2H1L has a different L3 loop conformation while the engineered structure shows differences in the L1 loop. The Root Mean Square Deviation (RMSD) of the whole protein (residues 97 to 287 alpha-carbon) between the 2H1L and the reference structure is ≈ 1.5 Å, and ≈ 5.6 Å over loop L3 (RMSD ≈ 0.9 Å for the L1 loop). The second cluster containing the engineered structures (3Q05, 3Q06, 3TS8 and 4MZR chain A and D) yield an RMSD of the full structure between 1.4 and 2.15 Å and an L1 RMSD of respectively 5 Å (3Q05 and 3Q06), 7.6 Å, and 6.7Å (RMSD between 0.7 and 0.9 Å for the L3 loop). These results are in agreement with the PCA results showing L1 and L3 conformation differences for 2H1L and some chains of the engineered structures compared to the rest of the X-rays and NMR structures.

The L3 loop motion of p53DBD was highlighted in the study of the SV40 large T-antigen (LTag) binding to p53DBD (PDB code: 2H1L) (120). The 2H1L

structure contains two LTag hexamers in which each monomer binds to a monomer of the DBD of p53 (Fig. 4-9A); the 12 p53DBD structures were used in the first PCA analysis. The X-ray structure of the complex LTag-p53DBD shows that LTag disturbs the interaction of p53 with DNA by binding at the p53 DNA-binding surface. This complex binding has been considered as a two-region binding (120); the first region contains residues Phe177, His178 and Arg181 from p53 α-helix 1, which makes direct contact with LTag α-helix 15, and the second region (mostly H2 and L3) in which 16 p53 residues make contacts with residues on the surface of LTag. Among the large number of interactions induced in region 2, the strong hydrophobic interaction of p53-Met246 within the LTag hydrophobic pocket (residues L609, Y612, W581 and Y582) induces a conformational change in the L3 loop compared to the free p53DBD conformation of L3 (Fig. 4-3B). This conformational change in L3 is the result of a methionine switch between Met246 (originally buried) and Met243. In the wild type p53 structure, Met246 is buried and Met243 is exposed instead (Met243 is implicated in the formation of stable p53 dimer-DNA complexes (121–123) and in p53-protein interactions (124–126)). The position of these two methionines is important in the partner interactions (120): while one participates in hydrophobic packing with the core of p53 in order to stabilise the L3 loop, the other is exposed to bind with partners (DNA or protein). The packing also has to be weak in order to allow the methionine switch, which is helped by the presence of the two consecutive glycines that provide the necessary flexibility within this region. This mechanism of adaptive binding of p53 depending on the partner allows a flexible and robust interaction. In the wild type, the stable conformation adopted by L3 also appears to depend on the guanidinium group of Arg249. This residue is implicated in a network of hydrogen bonds and a salt bridge with Glu171, and modulates the positioning of Arg248 for DNA binding (122); in 2H1L the methionine switch induces the positioning of Arg248 in order to achieve optimal interactions with LTag (120).

Two oncogenic mutations in p53 DBD are also known to disturb L3 stability. The hot spot mutation R249S, a structural mutation (destabilised by almost 2 $kcal.mol^{-1}$), induces distortions in the p53 DNA-binding surface and hence loss of interactions with DNA due to the loss of the stabilising interactions mediated by Arg249. This distortion is associated also with a reorientation of

Met243, resulting in the adoption of an α-helix conformation by the residues 239-244. This is similar to the LTag-p53DBD complex, where the L3 loop movement is associated with a methionine switch; Met243 is buried in a hydrophobic pocket (formed by Val173 and Leu194) originally occupied by Met246 (127). Three models were built and studied for the R249S mutant by Fersht et *al.* (127): 2BIO (PDB code), a super stable mutant M133L, V203A, N239Y and N268D associated with the R249S mutation which shows the L3 loop distortion and 2BIP and 2BIQ also R249S super stable mutants with rescue mutations specific to R249S (H168R and T123A respectively). Both 2BIP and 2BIQ have their DNA binding restored by the rescue mutations (Kd equivalent to WT p53DBD) unlike 2BIO, which although shows enhanced DNA binding compared to P53DBD R249S mutant, displays L3 distortion thought to be the deleterious result of the R249S mutation (127). In the 2BIP and 2BIQ structures, the residues Met243 and Gly244, implicated in the methionine switch, are missing. Although the Met246 of these two structures is still buried, the lack of a complete loop L3 does not allow us to state any definitive conclusion about L3 loop in the R249S mutant. These structures were not included in PCA because of missing atoms or residues, as cited in Materials and Methods.

The second mutation, R282Q (pdb code: 2PCX), defined as a low frequency hot spot mutation is located in the H2 helix. R282Q destabilises p53DBD by 8.8 kJ/mol (2.1 Kcal/mol) at 283K (128). This is thought to originate mainly in the elimination of an intramolecular salt bridge between Arg282 and Glu286, which also has a significant impact on the DNA binding loops (L1 and L3). Indeed the R282Q mutation is associated with a stiffening of the L1 loop and via a long-range effect, induction of flexibility in the L3 loop; this loop can present 2 different conformations in the mutant. The rigidity of L1 is induced by: a peptide bond flip (Leu114 and His115) stabilised by two hydrogen bonds and the formation of 3 more hydrogen bonds between Gln282 and its neighbouring residues including one with Ser116. This latter induces a distortion in the Serine, leads to a shift in L1 towards the hydrophobic core of the β sandwich. The second L3 loop conformation (β-conformer) observed is similar to the wild type structure unlike the second conformation (α-conformer) which exhibits a shift of the α-carbon atom of Ser241 of about 2.8 Å (Fig. 4-10 C and D). In the wild-type L3 loop conformation Ser241 ϒ oxygen makes a

hydrogen bond with the phosphodiester backbone of the DNA minor groove (106). Interestingly this last mutant has been crystallised without any rescue mutations included in the structure (PDB code: 2PCX), unlike the R249S mutant X-ray structures which contains rescue mutations.

The α-carbon alignment of the two molecules of the R282Q mutant structure in the unit cell, i.e. conformer A (L3 loop destabilised, red, Fig. 10C) and conformer B (wild-type like structure, purple, Fig. 4-10C), on to the reference structure (2XWR) shows a small shift of the L3 loop of conformer A (L3 loop RMSD = 1.2 and 0.6 Å, respectively) (Fig. 4-10C). This L3 loop motion arises from a shift of the α-carbon atom of Ser241 and is also associated with a slight movement of the Met243 side chain in the B conformer which is wild-type like (Fig. 4-10D). It is possible that these two conformers might represent conformations of the L3 loop that lie along the pathway of the Methionine switch. 2XWR chain A (WT), 2H1L (LTag-p53 complex) and 2BIO (R249S mutant with rescue mutations) were aligned and the RMSD with 2XWR as reference was calculated (Fig. 4-10E). As expected 2H1L and 2BIO show L3 loop movement due to the methionine switch, with respective RMSD of 5.6 Å and 5.3 Å compared to the wild-type (as 2BIO has residues missing, the complete structures were not used in this RMSD measure). This L3 loop form is a characteristic preventing p53 from binding to DNA, as found in a mutant (2BIO) and in a p53 structure for which the function has been inhibited by virus antigen binding (2H1L). Thus an oncogenic form of p53DBD displays a L3 conformational change which can be highlighted by PCA.

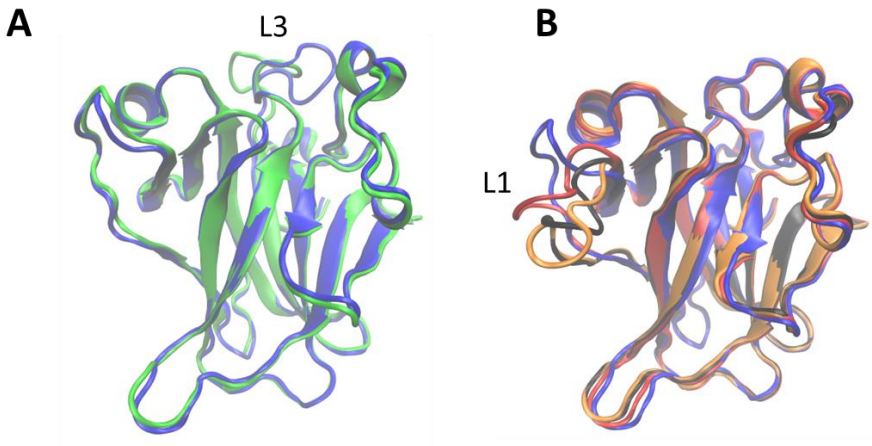


Figure 4-9: L1 and L3 loop conformational comparison.
A) 2H1L chain M (green) is aligned on the carbon-alpha of 2XWR chain A (reference structure, blue). **B)** Engineered structure (3Q05 → blue, 3TS8 → orange and 4MZR → black) are aligned on the carbon alpha of the reference structure (blue).

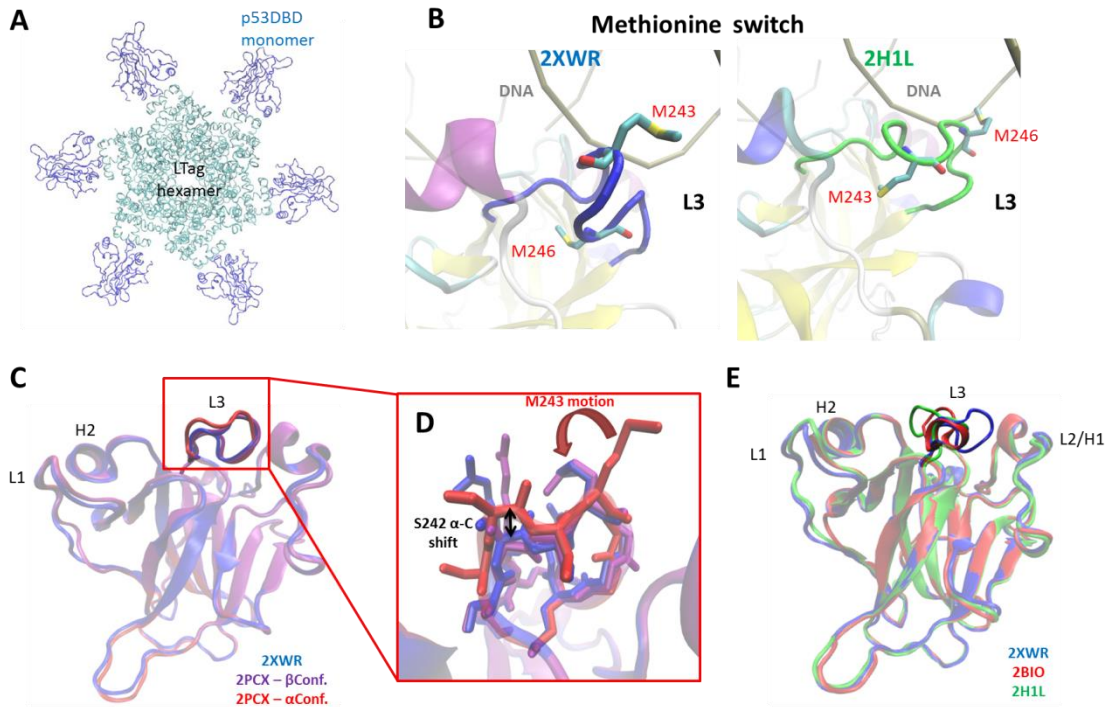


Figure 4-10: L3 loop conformations.
A) LTag hexamer binding (light blue) to six p53 DBD monomers dark blue). New cartoon representation. **B)** Methionines 243 and 246 orientation in 2XWR chain A (blue) and in 2H1L chain M (green). 2XWR and 2H1L were aligned on 2AHI and the 2AHI DNA structure was kept for orientation (in trace, grey) **C)** Structure comparison of 2PCX α-conformation (red) and β-conformation (purple) of the L3 loop to 2XWR chain A (blue). **D)** Met243 different conformations in 2PCX α-conformation (red), β-conformation (purple) and 2XWR (blue). **E)** Structure comparison of 2H1L (green), 2BIO (red) aligned on 2XWR (blue). VMD visualisation.

4.3.1.3 L1 and L2 loop motions

The L1 loop motion highlighted in the second cluster containing the engineered structures (3Q05, 3Q06, 3TS8, 4MZI and 4MZR chain A and D) is due to an induced fit mechanism event allowing the specific binding of p53DBD to the DNA via a dimeric arrangement of p53. The new L1 conformation observed in these structures is called 'recessed' while the L1 conformation common to most experimental structures is termed 'extended'. These engineered p53 structures are composed of the DNA binding domain, the tetramerisation domain (TD) and some residues from the linker sequence between these two domains (Fig. 4-11A). Some mutations performed in the DBD in order to stabilise the domain (C135V, C141V, W146Y, C182S, V203A, R209P, C229Y, H233Y, Y234F, N235K, Y236F, T253V, N268D), allow the same DNA specificity and cell functionality as the wild type, and two amino-acid substitutions were introduced in the tetramerisation domain to convert it to a dimerization domain. A total of 18 mutations were applied in non-conserved residues, away from the DNA binding surface, using residues present at the same position as in p53 from other species. The three first structures studied, two binding to an artificial consensus DNA site as a dimer (3Q05 and 3Q06, two dimers binding to DNA in each X-ray structure) and one dimeric structure not bound to DNA (3Q01) (113), highlight this L1 loop motion when p53 binds to DNA in 3Q05 and 3Q06 structures but not in 3Q01. The DNA sequence targeted in this study, is constituted by 4 contiguous pentamer repeats (Fig. 4-11B), each p53 engineered dimer will bind to DNA in a way that the monomer bound to the inner DNA pentamer will adopt an extended L1 loop (WT like conformation) and the monomer bound to the outer sequence of the DNA will present a recessed conformation. In a second study of this engineered p53 structure, 3TS8 binds to the promoter of the CDKN1A (p21) gene, this was the first structure of the p53DBD bound to a natural p53-response element (96). 3TS8 shows a more extreme change in the loop L1 due to its specific binding compared to 3Q05 and 3Q06 structures which suggests that the first L1 loop conformation observed (in 3Q05 and 3Q06) is a conformation intermediate between p53 DNA bound and DNA free states. In their third study, the group proposed two new engineered structures, one bound to DNA (4MZR) and the other free (4ZMI). These two structures contain two mutations in the L1 loop (S121F and V122G) that enhances the affinity of p53 for specific DNA but

decreases the residency time of p53 on DNA (114). In contrast to the other engineered structure, 4MZI (the free p53 engineered model) contains a recessed L1 loop. When binding to DNA the two chains switch to the extended conformation (4MZR).

The α-carbon of the L1 engineered structures 3Q01(with no recessed L1 loop, as not bound to DNA) , 3Q05 and 3TS8, 4MZI and 4MZR chain A were superimposed onto the wild type reference structure 2XWR chain A (Fig. 4-11C) and the RMSD of the L1 loop is about 1.2, 4.6, 7.6, 7.6 and 6.7 Å respectively. This L1 loop conformational change is different from the one described in the NMR structure 2FEJ (103). Indeed 2FEJ L1 loop changes conformation while p53DBD is free and shows a twist of the L1 loop driven by Ser121 (Fig. 4-6). In the engineered structures (3Q05, 3Q06 and 3TS8) the L1 loop is pulled toward the central immunoglobulin-like β sandwich to bind to the outer sequence of the DNA and pulled back towards the H2 helix to bind to the inner DNA sequence in the L1 loop mutated structure (4MZR). Indeed the chain of the engineered structures binding to the inner DNA, and the other X-ray structures containing only the DBD, show a similar L1 conformation, as shown by the co-localisation of all these structures in the main cluster (Figure 4-3A Zoom). The two apo engineered structures 3Q01 and 4MZI display a different conformation of a part of their L2 loop (from residue 181 to 194). This conformational change as mentioned earlier is thought to be due to the apo state of the proteins.

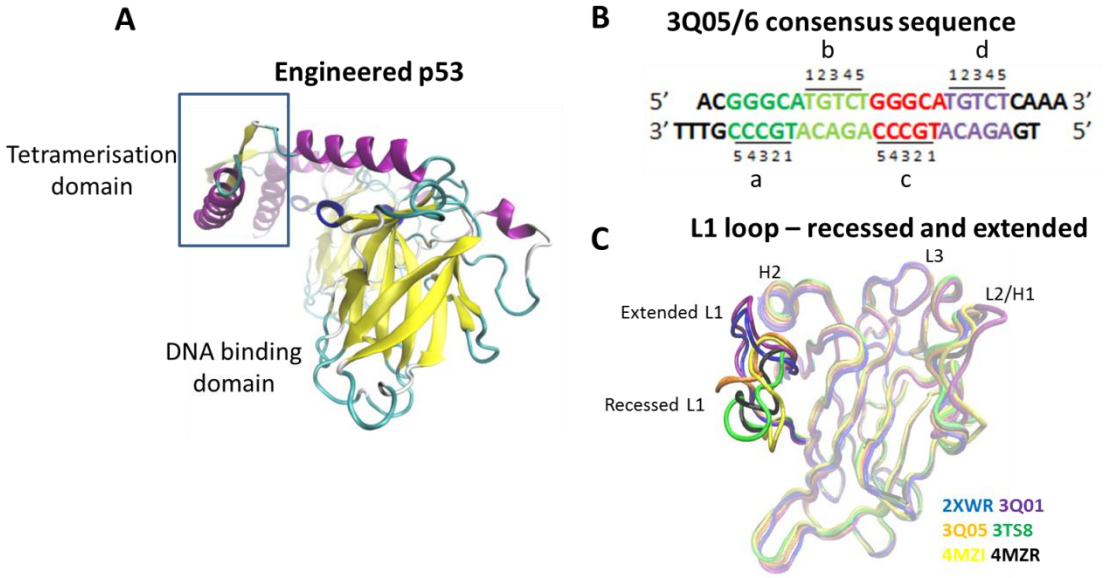


Figure 4-11: Engineered p53.
A) 3Q05 dimer new cartoon representation and coloured by secondary structure. **B)** DNA sequence on which 3Q05 and 3Q06 binds containing the four contiguous pentamer repeats (a to d). In the manuscript outer DNA sequences refers to a and d and the inner sequences refers to b and c. **C)** Superposition of the 5 DBD of the engineered p53 on C-α of 2XWR (the reference WT structure, blue). 3Q01 chain A (non-bound structure, in purple), 4MZI chain A (non-bound structure, yellow), 3Q05 chain A (in orange), 3TS8 chain A (in green) and 4MZR chain A (black). VMD visualisation.

This first PCA that we performed on all the p53DBD structures available highlighted conformational changes localized mainly in the loops. In order to determine if these loop changes also occur, albeit with low amplitude, in the X-ray structures containing only the DBD, separate PCA was performed on these structures. An interesting result was found while performing PCA on wild-type structures: 26 wild-type structures bound to DNA and 11 in apo states. The first three PCs capture 66.4% of variance (Fig. 4-12A), the first PC highlights motions along loops L1 and L6 and PCs2 and 3 capture motions mainly in L2. The L6 loop is known to be highly flexible (129), we will ignore this loop motion here. The structures projection along PC1 and PC2 (Fig. 4-13) exhibits two clusters along PC1; one containing most of the p53DBD structures bound to DNA (red circled) and the other comprising all the free p53DBD (blue circled). A visual comparison between the two set of structure (Fig. 4-14) shows that the main difference between these two groups is located in the L1 loop. The C-α superimposition of the free p53DBD on the reference structure (2XWR chain A) show a perfect alignment of their L1 loop (Fig. 4-14A). Whereas when

2XWR (blue structure) is superimposed on the DNA bound structures (Fig. 4-14B) a difference in the L1 loop conformation is observed. In the engineered recessed loop 1 conformation (binding to outer DNA) neither Lys120 nor Ser121 contact the DNA compared to the extended L1 conformation (binding to the inner DNA) in which both residues bind to DNA (96, 113). In these structures, the L1 loop of the chains binding to the outer DNA sequences are not in contact with DNA, the monomers binding to the inner DNA sequences seem to bind more specifically. The L1 loop contains three residues that contact DNA bases directly, Lys120, Ser121 and T123. It has been also shown that mutation of these residues changes the DNA binding specificity (130, 131). Visual comparisons of structures bound to DNA against those not bound and between structures bound to inner against those bound to outer DNA consensus confirm a role of Lys120 and Ser121 in loop 1 conformation (Fig. 4-15). The three structures compared, 2XWR chain A, 2OCJ chain A and 3KMD chain C show different orientations of Ser121, which seems to pull out Lys120 and L1 loop in the case of the free p53 structure and of the structure bound to outer DNA sequence. Ser121 adopts the same orientation in all the free wild type monomers (data not shown) whereas in the DNA bound monomers its orientation is different for each monomer even between two monomers of the same structure both bound to the outer sequence of the DNA. The location/organisation of the DNA consensus sequences in the DNA helix seems to dictate the L1 loop conformation.

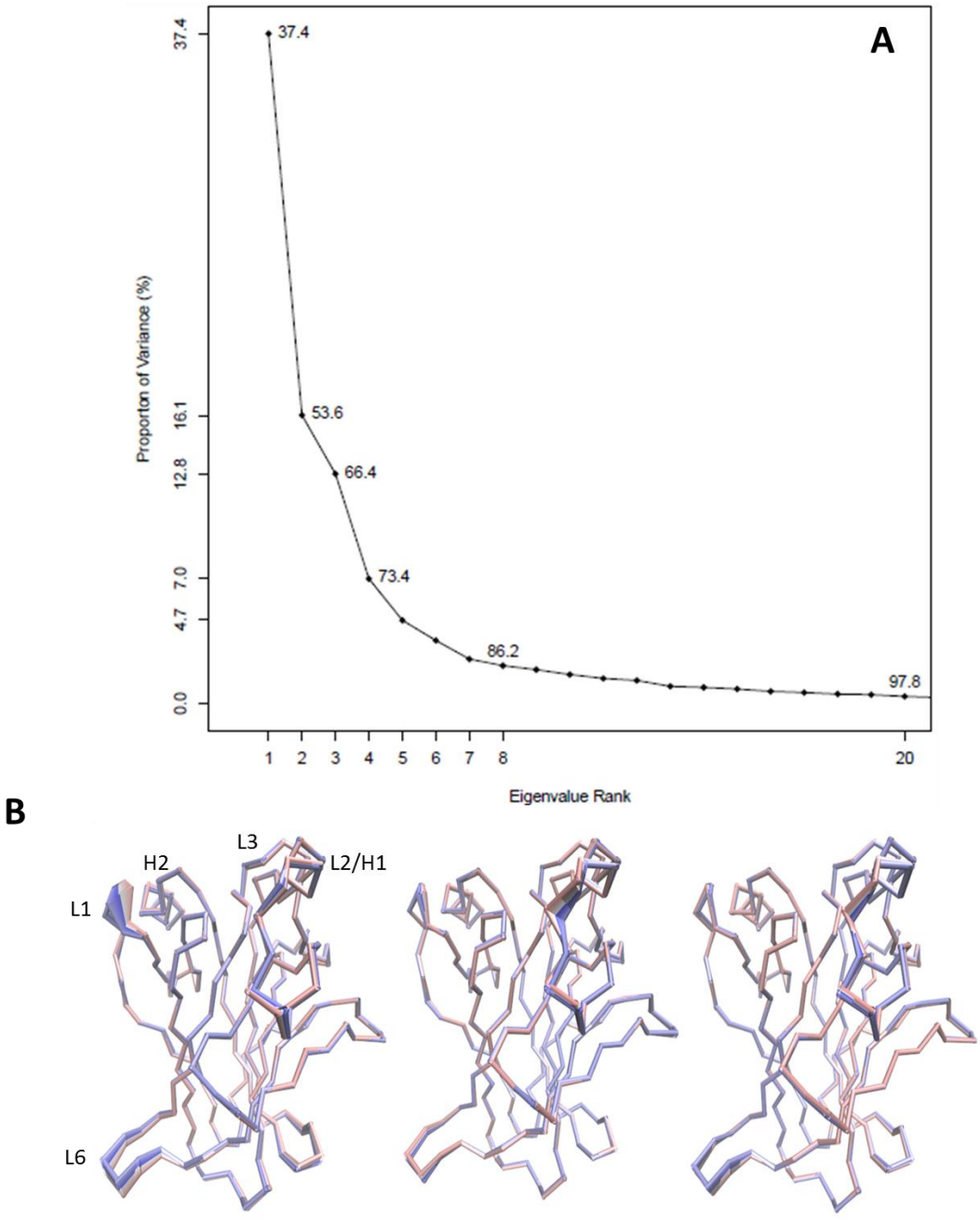


Figure 4-12: PCA of wild type p53DBD.
Proportion of variance (in %) captured by each principal component (PC) from a data set of the 37 wild-type X-ray structures. The 3 first PCs contribute to 66.4% of the variance of the dataset. **B)** Trace representation of the motion projections along the 3 first PCs (PC1, PC2 and PC3) of the set of structures using VMD. The colouring range of the motions goes from red, to white, to blue as function of time (timestep option on VMD), red being the early motion.

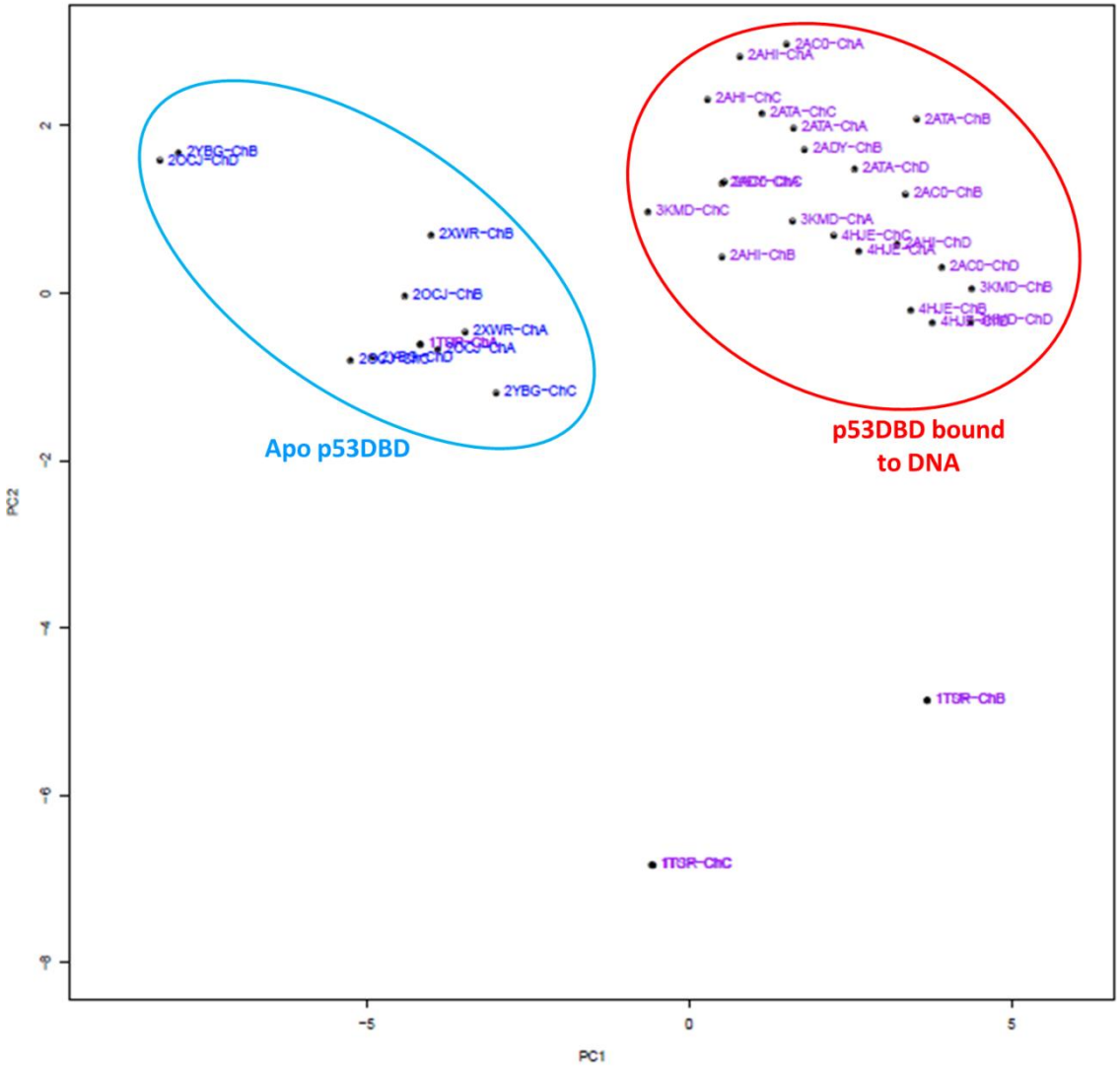


Figure 4-13: Projection of the set of 37 structures on PC1 and PC2.
The structure labels were colored as follow: the monomers binding to DNA sequence (purple) and the monomer not bound to DNA (blue).

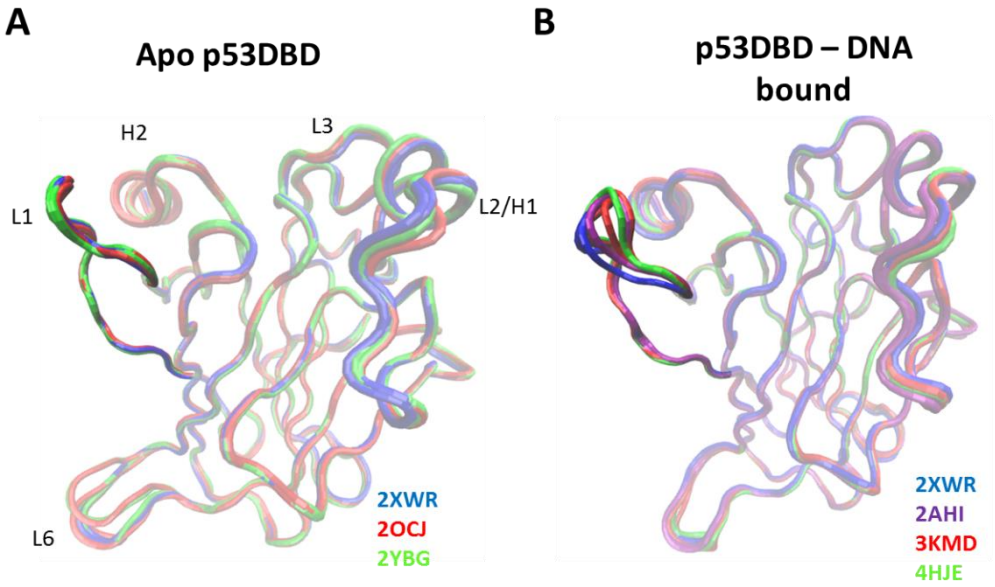


Figure 4-14: Apo against DNA bound p53DBD comparison.
A) Apo p53DBD monomers aligned on carbon alpha of 2XWR chain A (blue), 2OCJ all chains (red) 2YBG all chains (green) **B)** DNA bound p53; 2AHI chain A (purple), 3KMD chain A (red) and 4HJE chain A (green) aligned on carbon alpha of 2XWR chain A. VMD visualisation using tube representation.

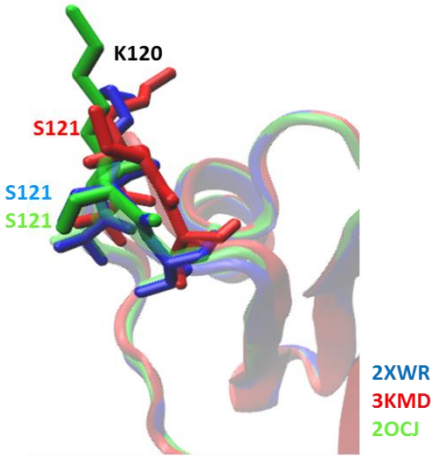


Figure 4-15: Lys120 and Ser121 orientation in L1 loop wild-type structures. 2XWR (blue), 2OCJ (green), 3KMD (red). 2OCJ and 3KMD were aligned on carbon alpha of 2XWR.

This PCA also brings out the case of 1TSR and 1TUP structures. As seen in the structure projection (Fig. 4-14) none of the chains constituting these structures cluster where expected. 1TSR and 1TUP contain three p53DBD monomers bound to DNA. According to the study only the chain B and C are bound to DNA, chain B binds specifically to DNA while chain C non-specifically (123) which explains the localisation of 1TSR/TUP chain A with the free p53DBD

structures (Fig. 4-14). Also p53DBD is known to bind to DNA as a dimer of dimers (132), which is not the case in 1TSR/TUP structures. The visualisation of 1TSR (Fig. 4-16A) reveals that each of the DNA binding chain (B and C) bound independently to DNA. On the contrary, in 3KMD (Fig. 4-16B) each monomer binds to DNA head to head with another monomer, involving protein-DNA and protein-protein interactions, to form a dimer (Chain A and B in Fig. 4-16B). In 1TSR/TUP no dimer is formed while binding to DNA which can explain the location of the chain B and C further away from the DNA binding structures. This can explain why 1TSR/TUP chain B and C are located further away along PC2. Indeed PC2 captures L2 loop motion (Fig. 4-12B and Fig. 4-17), that was shown previously to be a characteristic of free p53DBD structures.

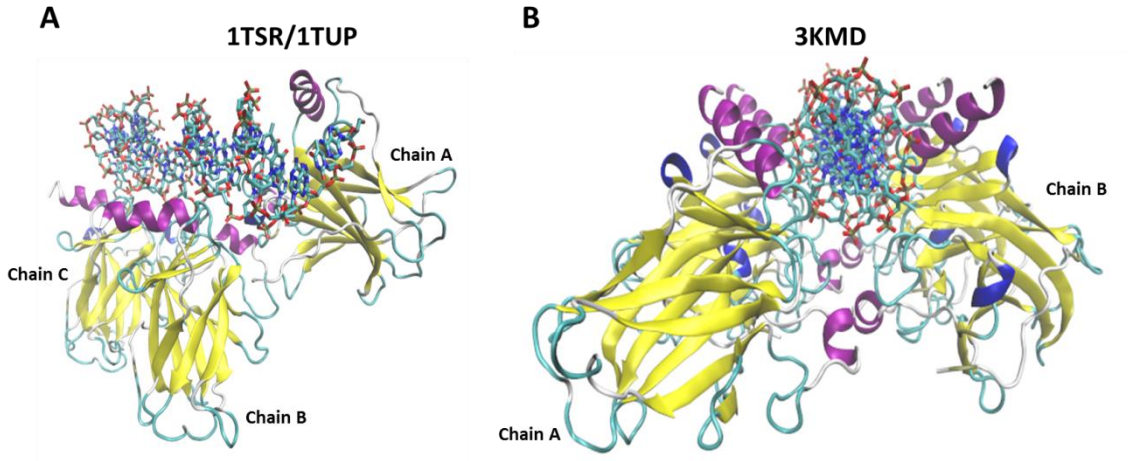


Figure 4-16: DNA binding mode.
A) 1TSR/TUP chain A, B and C binding to DNA**. B)** 3KMD. The p53 DBDs are represented in new-cartoon and coloured by structure and the DNA is represented in licorice coloured by name.

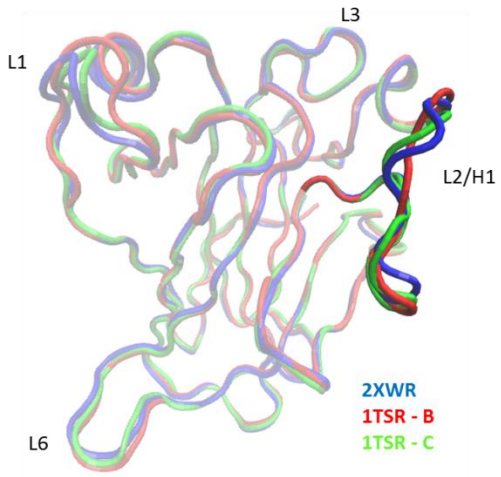


Figure 4-17: L2 loop comparison.
1TSR chain B (red), chain C (green) aligned on carbon alpha of 2XWR chain A (blue). The structures are represented in trace.

To further investigate the influence of DNA binding on p53 DBD conformation, a PCA was performed on only the structures bound to DNA as a dimer of dimers (2AHI, 2AC0, 2ADY, 2ATA, 3KMD and 4HJE); these comprised 22 p53DBD monomers (data not shown). We explored if the p53 binding to DNA as a dimer, monomer binding to the inner and the outer DNA, has an influence in the L1 loop of these structures. Although PCA on bound/not bound structures gives relevant result based on structural visualisation, it does not highlight the differences between structures binding to inner/outer DNA. This might be due to the Loop 6 motion and to the fact that in 2AHI, 2ATA, 2ADY and 2AC0 structures there is no distinction between monomers binding to inner/outer DNA (see Fig. App. A 1). In these structures, four p53DBD self-assemble on two B-DNA half sites to form a tetramer. The DNA consensus sequences, stacked end to end mimicking a continuous double helix separated by two base pairs, are at the same position in the two strands. There are no inner or outer DNA consensus sequences in this group of PDBs (see Tab. 4-2). These structures although being tetrameric, present few interactions between the two dimers due to the base pair separation and off-set axes of the DNA molecules. A better result can be expected using a larger data set and structures naturally bound to DNA.

L1 loop adopts different conformations depending on whether p53DBD binds to DNA or not, and this result is captured by the use of the PCA technique. In the particular case of the engineered structures, PCA has also been able to

distinguish between the L1 loop binding the inner sequence of the DNA (extended L1 loop) and the loop binding the outer part (recessed L1). Although the technique has not been able to differentiate the inner/outer p53DBD binders among the rest of the X-ray structures, mainly due to the lack of proper p53DBD-DNA tetramers (only two structures: 3KMD and 4HJE).

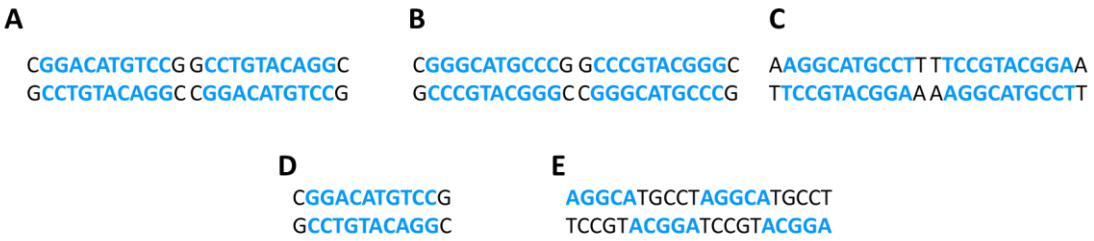


Table 4-2: DNA consensus sequences.
DNA sequences to which p53DBD binds. The consensus sequences were coloured in blue. **A)** 2AHI, **B)** 2AC0, **C)** 2ATA, **D)** 2ADY and **E)** 3KMD. Only 3KMD presents inner and outer consensus sequence.

4.3.2 DNA PCA

We now explore the influence of the binding of p53DBD on the conformations of DNA. Several studies have attempted to address this question. Some of the first investigations show DNA bending when p53 binds to linear DNA as a tetramer and that the magnitude of bending depends on the 4-bases sequence at the centre of the DNA consensus half sites (133–135). A later study investigating the different potential mode of binding of p53DBD as a tetramer to DNA, finds contradictory results for linear DNA on which p53 binds in a Q1234 mode and concludes that the bending seen in those studies might be due to the DNA sequence or to a special simulation setup (136). Those studies are mainly based on the p53-DNA crystallographic structures available at the time, 1TSR and 1TUP. Up to date 19 structures binding to a double helix DNA are available and are used for the current analysis. Among these, seven (3KMD, 4HJE, 3Q05, 3Q06, 3TS8, 4IBU and 4MZR) show Q1234 binding mode (Fig. 4-18), the 1TSR and 1TUP structures show a binding mode similar to T14 and T13 (136) and the 5 other structures (2AHI, 2ATA, 2ADY, 2AC0 and 3D0A) although existing as tetramers seem to form dimer complexes with DNA (as each dimer binds to a different helix and the two DNA helices are not aligned).

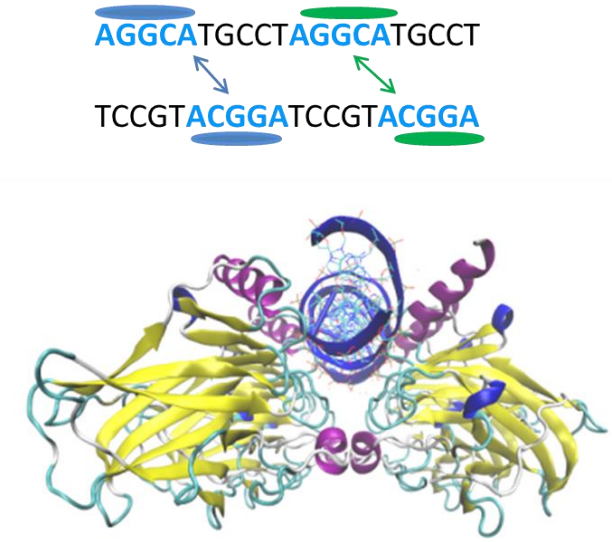
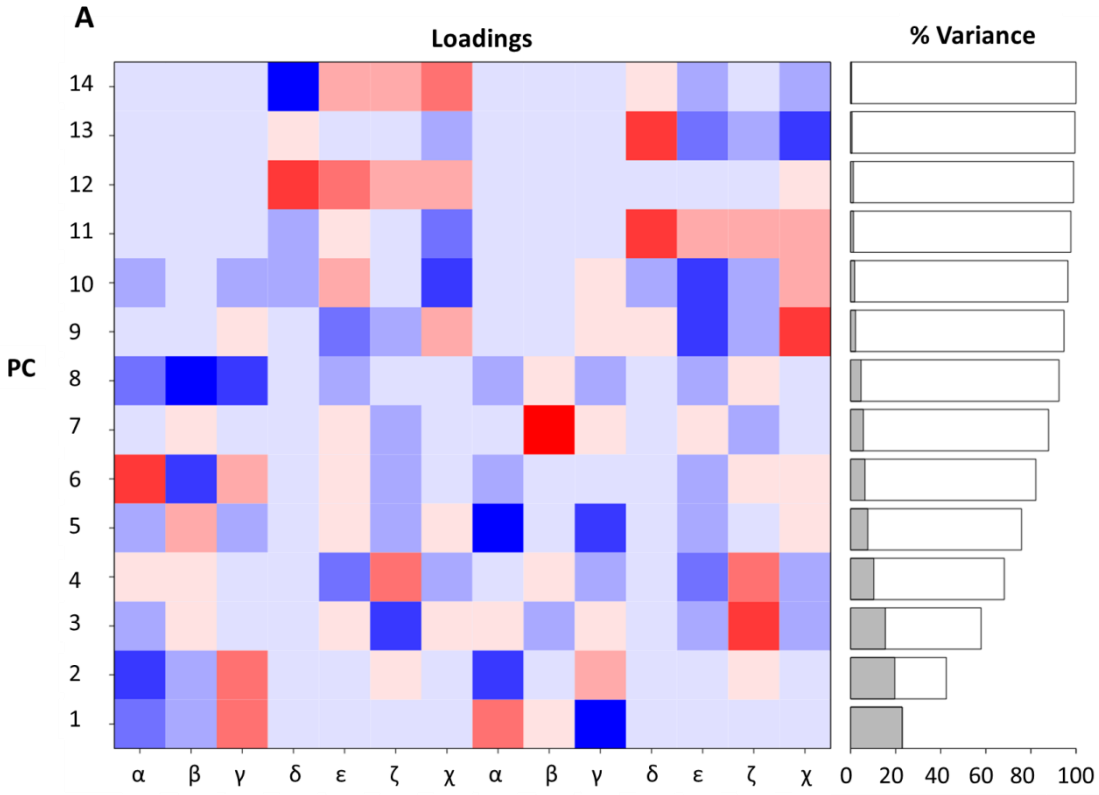


Figure 4-18: Q1234 p53-DNA binding mode of 3KMD structure.
DNA sequence on which 3KMD binds, consensus sequence highlighted in blue. Schematic representation of p53 binding (blue dimer1 and green dimer2). 3KMD representation in new cartoon. VMD visualisation.

The torsion angles of the DNA sugar-phosphate backbone show different features depending on its form and whether it is bound to a protein or not. Thus, it is possible to distinguish between B and A DNA and their sub-forms which exhibit differences in the two torsion angles ε (C4'-C3'-P-O5) and ζ (C3'-O3'-P-O5') for the BI and BII-forms and differences between α (O3'-P-O5'-C5') and ϒ (O5'-C5'-C4'-C3') torsion angles for the A and Crankshaft-forms (117). These torsion angles differences have also been observed on DNA binding to proteins, usually the B-form, which exhibits α and ϒ transitions generally found in the A-form of DNA (137). PCA performed on the sugar phosphate backbone of single-strand dinucleotides DNA structures not bound to proteins was shown to be able to discriminate between A and B DNA forms as well as their respective sub-forms, A-form and Crankshaft A-form, BI and BII-forms and the Crankshaft B-form specific to the B-DNA form bound to proteins (117).

The 14 DNA helices extracted from the p53 DBD X-ray structures are all in the B-from. The sugar phosphate backbone torsion angles were extracted from each dinucleotide (Fig. 4-1), which results in 421 dinucleotides studied using 14 torsion angles (7 per nucleotide). PCA was performed on the 421 dinucleotides using the PCA method used previously for characterising the Crankshaft B-form (117). 60% of the variance is captured by the first 3 PCs (Fig. 4-19A), PC1 and PC2 capture variance mainly on the alpha and gamma torsion

angles of both nucleotides (as indicated by the magnitude of their loadings), with more variance captured in the second nucleotide. The α and ϒ loadings along PC1 and PC2 are of opposite signs in each nucleotide which fits the description of the canonical DNA conformation bound to protein (Crankshaft B-form) (117, 137). It is noted that the loadings of the delta and chi angles along PC1 and PC2 are negligible which indicates that moving in the PC1-PC2 plane does not incur change from B-form DNA.



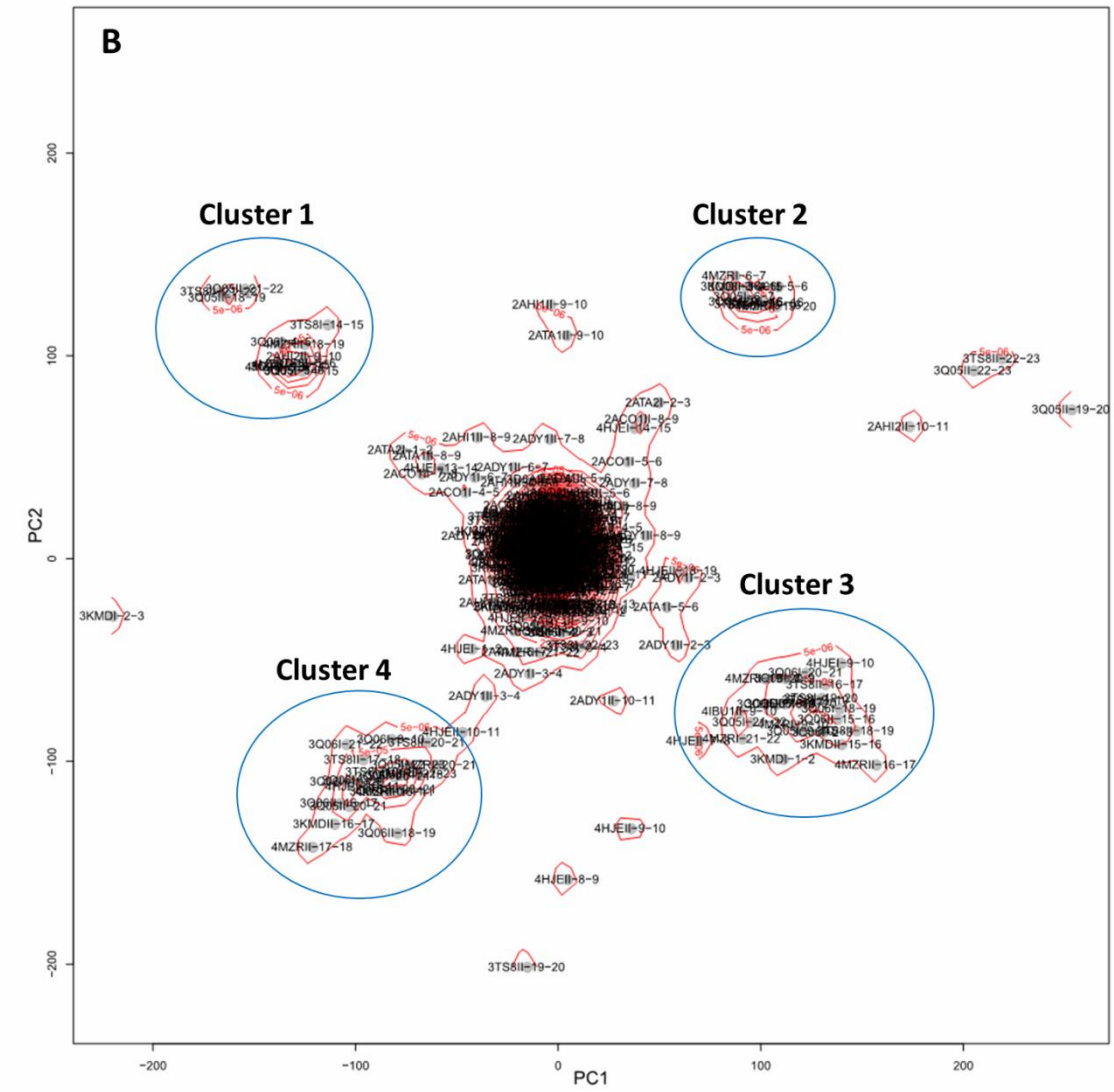
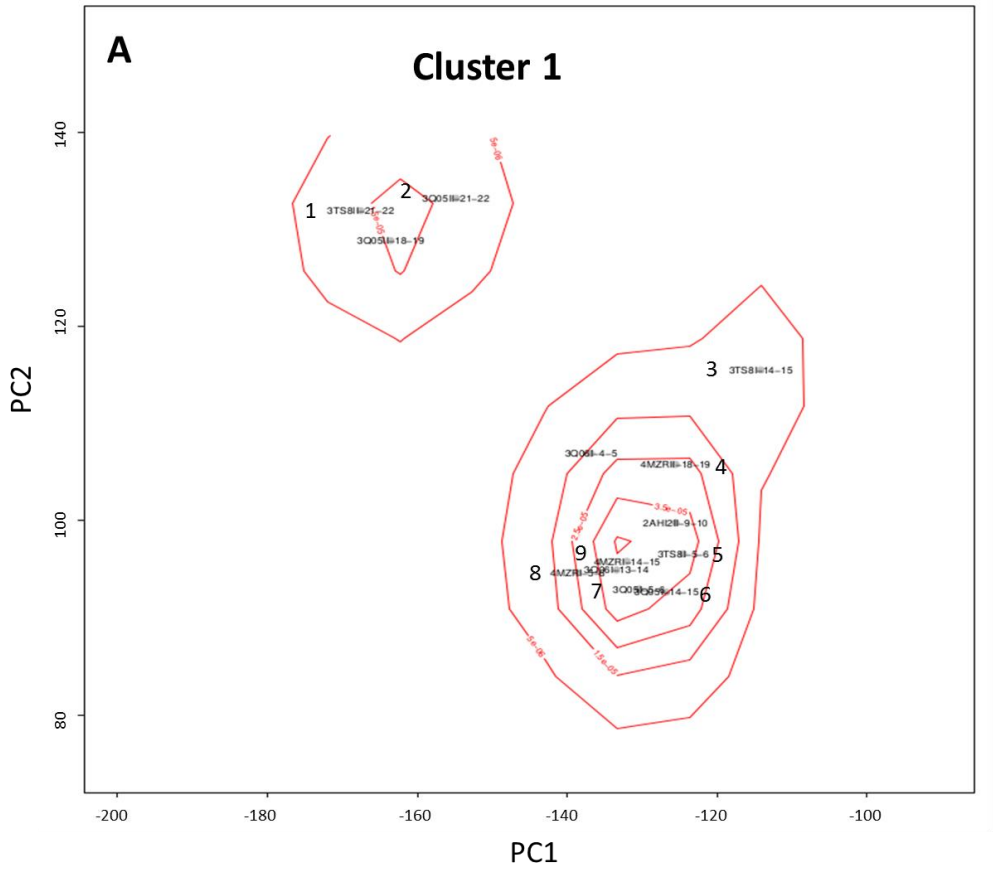


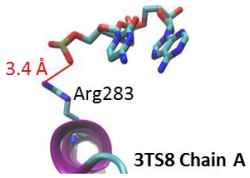
Figure 4-19: DNA PCA.
A) Proportion of variance captured by each PC; grey boxes and cumulative variance; grey + white boxes (right plot) and torsional angles implicated in the variance, red: positive motion, blue: negative and white: neutral. **B)** Dinucleotides projection onto PC1 and PC2.

The projection of the structures in the PC1-PC2 plane (Fig. 4-19B) shows a main central cluster surrounded by four clusters. The dinucleotides located in these four clusters were visualised in the corresponding DNA structures (Fig. 4-20). It is striking that most of the dinucleotides in these clusters come from engineered structures where the nucleotides are bound or close to the p53 residues that are known to bind DNA. These structures (3Q05, 3Q06, 3TS8 and 4MZR) are thought to be the perfect p53-DNA binding model. This is confirmed by the observation that most of the dinucleotides in the four clusters come from those structures and are directly implicated in the p53 binding.

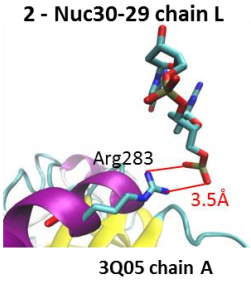
Interestingly, the 1TSR and 1TUP dinucleotides are not located in the clusters and show large variation in their alpha and gamma torsion angles compared to the dinucleotides in the central cluster (dinucleotides scattered around the clusters) (Fig. 4-19). In fact, these structures are thought to show a different DNA binding mode of p53 (136), even though p53 is not bound as a tetramer it has an influence on the DNA backbone. The other structures (3KMD, 2AHI, 2ADY, 2AC0, 2ATA and 3DOA) show fewer dinucleotides with large variations in α and ϒ torsion angles and most of them cluster in the main central cluster. The second striking result of the projection along the PC1-PC2 plane is the symmetry of the 4 clusters. Interestingly, nucleotides common to different dinucleotides appear in different clusters (example in structure 3Q06, chain K dinucleotide 7-8 appears in cluster 1 while dinucleotide 8-9 appears in cluster 2, see Table 4-3). The nucleotides in clusters 1 and 2 and the ones in clusters 3 and 4 for the three engineered structures are shown within the corresponding DNA structures in figure 4-20, two of the nucleotides in cluster 1 and 2 (Table 4-3), are common in the three engineered structures (nucleotides 8 and 17 chain K). These three nucleotides located close to (within 5 Å) Arg280 of the different p53 chains (Fig. 4-20A and B). In cluster 3 and 4, three nucleotides are common to the four engineered structures, nucleotide 12 and 22 chain K and nucleotide 34 chain L which also are found to be close or in direct contact with Arg280.



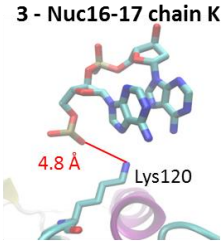
1 - Nuc29-30 chain L



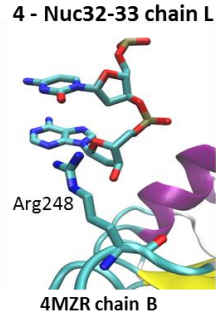
2 - Nuc30-29 chain L



3 - Nuc16-17 chain K

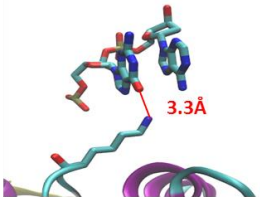


4 - Nuc32-33 chain L



5 - Nuc7-8 chain K: close to Arg280 chain D - 3TS8

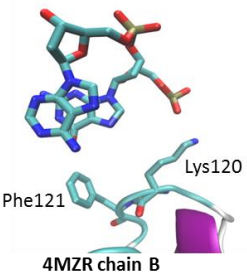
6 - Nuc16-17 chain K

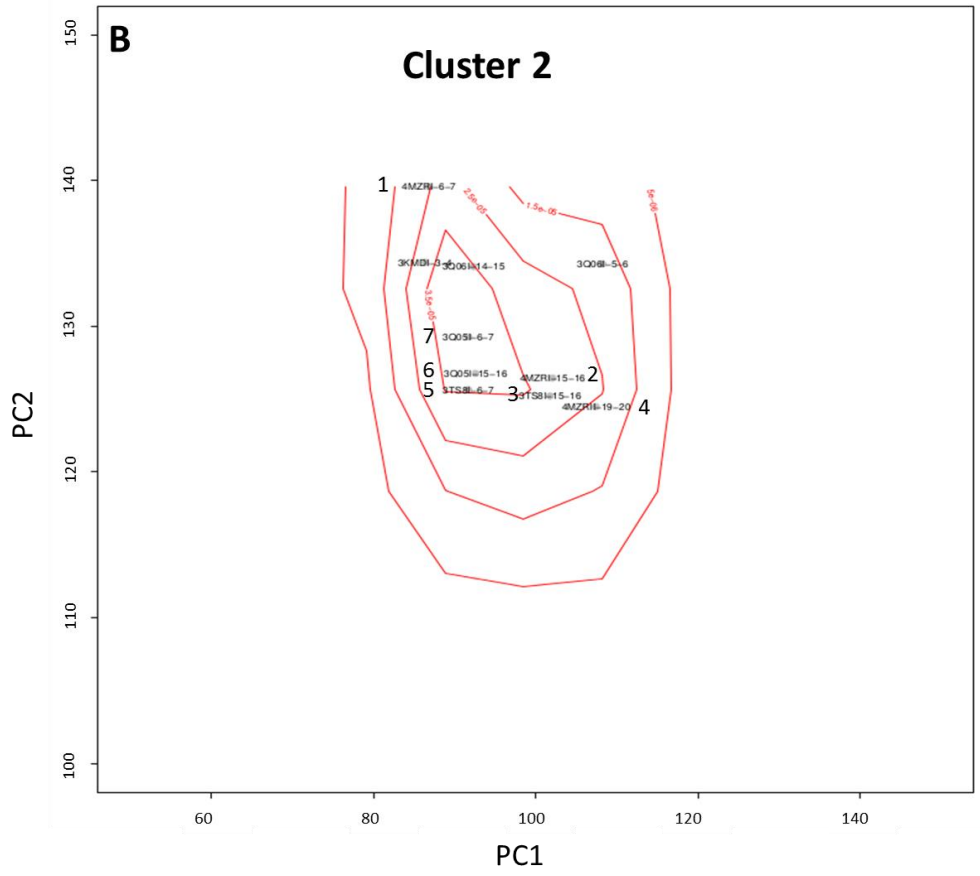


7 - Nuc7-8 chain K: close to Arg280 chain D - 3Q05

8 - Nuc7-8 chain K close to Arg280 chain D - 4MZR

9 - Chain K Nuc17-16

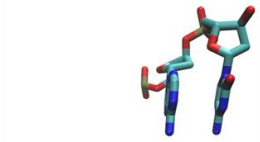


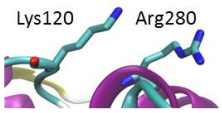


1 - Nuc8-9 chain K: close Arg280 chain D - 4MZR

2 - Nuc17-18 chain K: close Arg280 chain B - 4MZR

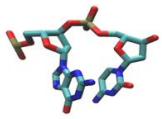
3 - Nuc17-18 chain K

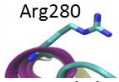




3TS8 Chain B

4 - Nuc31-32 chain L

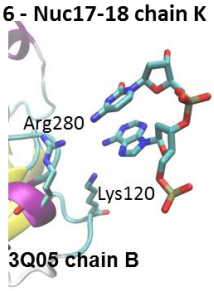




4MZR Chain A

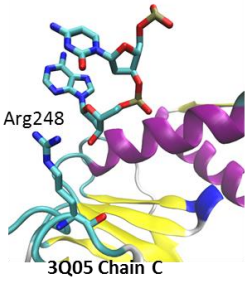
5 - Nuc8-9 chain K close: Arg248 and Arg273 - 3TS8

6 - Nuc17-18 chain K

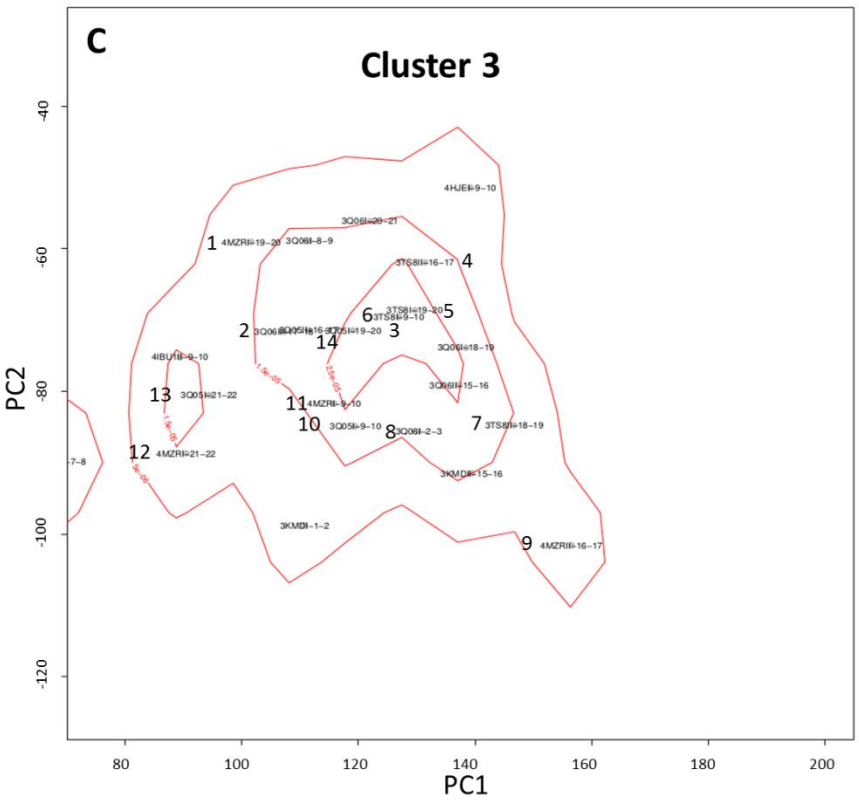


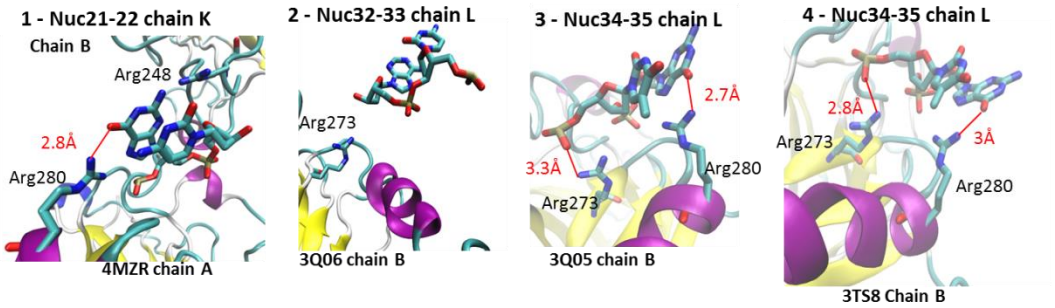
3Q05 chain B

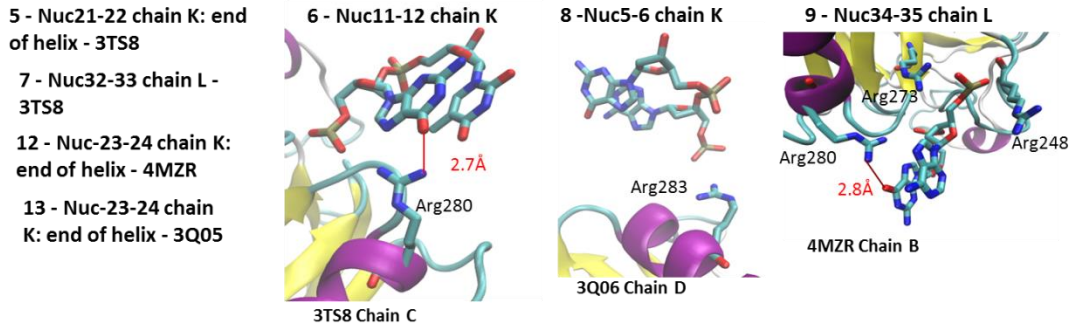
7 - Nuc9-8 chain K

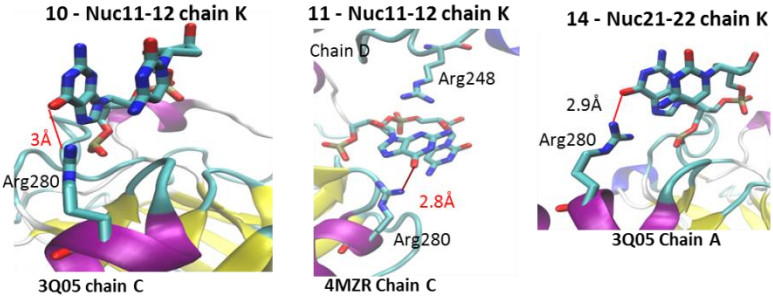


3Q05 Chain C


C
Cluster 3
PC2
PC1
-40
-60
-80
-100
-120
80
100
120
140
160
180
200
1
2
3
4
5
6
7
8
9
10
11
12
13
14


1 - Nuc21-22 chain K
Chain B
Arg248
2.8Å
Arg280
4MZR chain A
2 - Nuc32-33 chain L
Arg273
3Q06 chain B
3 - Nuc34-35 chain L
2.7Å
3.3Å
Arg280
Arg273
3Q05 chain B
4 - Nuc34-35 chain L
2.8Å
3Å
Arg273
Arg280
3TS8 Chain B


5 - Nuc21-22 chain K: end of helix - 3TS8
7 - Nuc32-33 chain L - 3TS8
12 - Nuc-23-24 chain K: end of helix - 4MZR
13 - Nuc-23-24 chain K: end of helix - 3Q05
6 - Nuc11-12 chain K
2.7Å
Arg280
3TS8 Chain C
8 -Nuc5-6 chain K
Arg283
3Q06 Chain D
9 - Nuc34-35 chain L
Arg273
Arg280
Arg248
2.8Å
4MZR Chain B


10 - Nuc11-12 chain K
3Å
Arg280
3Q05 chain C
11 - Nuc11-12 chain K
Chain D
Arg248
2.8Å
Arg280
4MZR Chain C
14 - Nuc21-22 chain K
2.9Å
Arg280
3Q05 Chain A

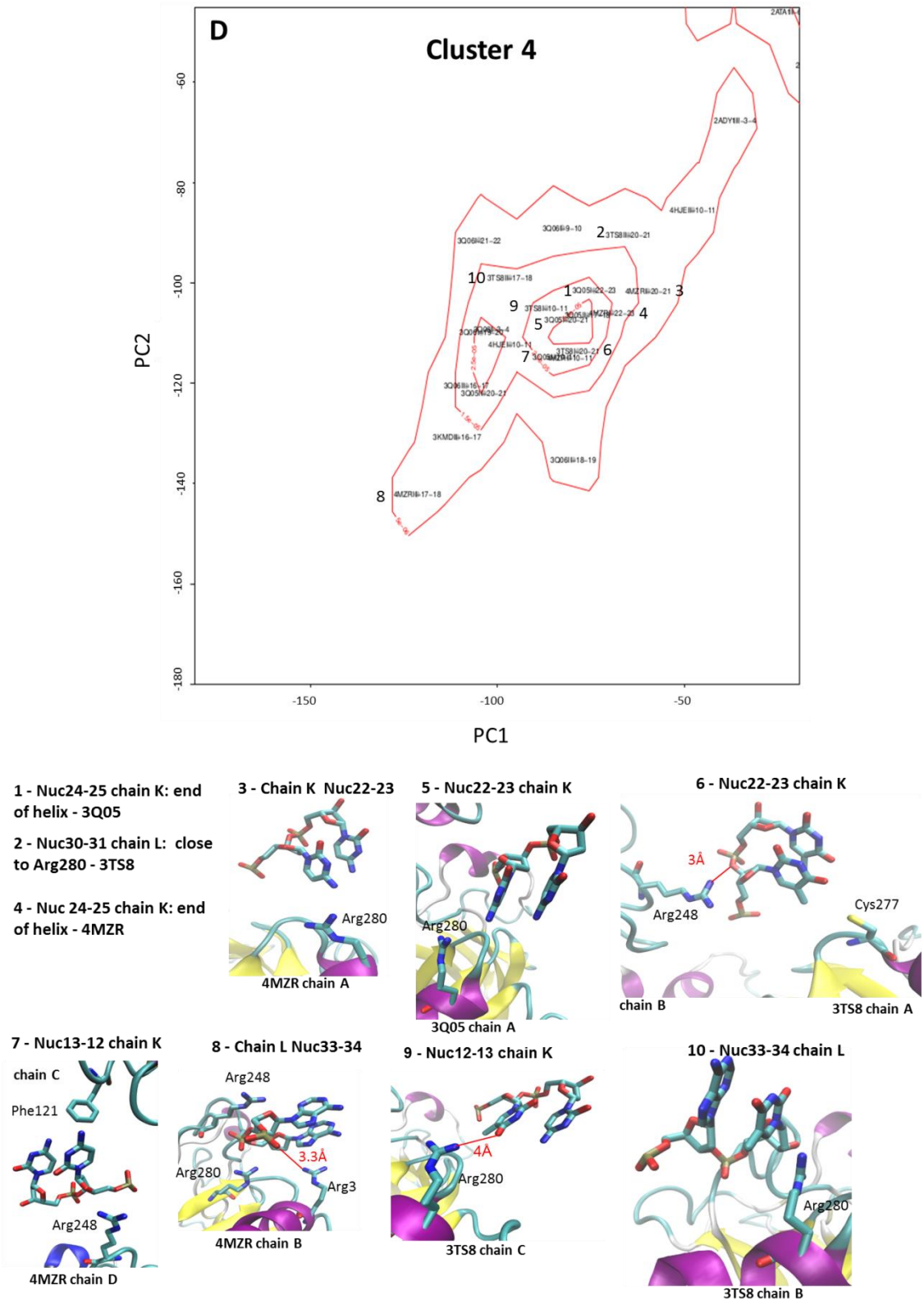


Figure 4-20: Dinucleotide location on the structures.
A) Cluster 1 **B)** Cluster 2 **C)** Cluster 3 **D)** Cluster 4. VMD visualisation p53 DBD monomer in new-cartoon coloured by structure, dinucleotides in bonds coloured by name and the rest of the DNA helix is kept in line representation coloured by name.

Cluster 1	Cluster 2	Cluster 3	Cluster 4
3Q05-ChL-Nuc29-30	3Q05-ChK-Nuc8-9	3Q05-ChL-Nuc34-35	3Q05-ChK-Nuc24-25
3Q05-ChK-Nuc7-8	3Q05-ChK-Nuc17-18	3Q05-ChK-Nuc21-22	3Q05-ChL-Nuc33-34
3Q05-ChK-Nuc16-17	3Q06-ChK-Nuc17-18	3Q05-ChK-Nuc23-24	3Q05-ChK-Nuc22-23
3Q06-ChK-Nuc7-8	3Q06-ChK-Nuc8-9	3Q05-ChK-Nuc11-12	3Q05-ChK-Nuc12-13
3Q06-ChK-Nuc16-17	3TS8-ChK-Nuc8-9	3Q06-ChK-Nuc23-24	3Q05-ChL-Nuc30-31
3TS8-ChL-Nuc29-30	3TS8-ChK-Nuc17-18	3Q06-ChK-Nuc11-12	3Q06-ChK-Nuc12-13
3TS8-ChK-Nuc7-8	4MZR-ChK-Nuc8-9	3Q06-ChK-Nuc21-22	3Q06-ChK-Nuc24-25
3TS8-ChK-Nuc16-17	4MZR-ChK-Nuc17-18	3Q06-ChK-Nuc5-6	3Q06-CK-Nuc6-7
4MZR-ChL-Nuc32-33	4MZR-ChL-Nuc31-32	3Q06-ChL-Nuc32-33	3Q06-ChK-Nuc22-23
4MZR-ChK-Nuc16-17	3KMD-ChE-Nuc4-5	3Q06-ChL-Nuc34-35	3Q06-ChL-Nuc33-34
4MZR-ChK-Nuc7-8		3TS8-ChL-Nuc34-35	3Q06-ChL-Nuc31-32
2AHI-ChH-Nuc3-4		3TS8-ChK-Nuc21-22	3TS8-ChL-Nuc30-31
		3TS8-ChK-Nuc11-12	3TS8-ChK-Nuc12-13
		3TS8-ChL-Nuc32-33	3TS8-ChK-Nuc22-23
		4MZR-ChK-Nuc21-22	4MZR-ChK-22-23
		4MZR-ChK-Nuc11-12	4MZR-ChK-24-25
		4MZR-ChK-Nuc23-24	4MZR-ChK-Nuc12-13
		4MZR-ChL-Nuc34-35	4MZR-ChL-Nuc33-34
		3KMD-ChF-Nuc4-5	2ADY-ChF-Nuc9-10
		3KMD-ChE-Nuc18-19	3KMD-ChF-Nuc2-3
		4HJE-ChE-Nuc10-11	4HJE-ChF-Nuc10-11
		4IBU-ChF-Nuc3-4	4HJE-ChE-Nuc11-12

Table 4-3: List of the dinucleotides found in clusters.

The DNA backbone superimposition of the nucleotides, that are bound to Arg280 (cluster 3-4) show a perfect alignment yet with a different conformation for the α and ϒ angles of the nucleotides 6, 24 and 33 of the 3Q06 structure (Blue coloured Fig. 4-21A). Nucleotides 17 and 8 that are not directly bound to a p53DBD residue are the only two nucleotides of the engineered structures that appear in cluster 1 and 2. These nucleotides although not directly bound to a p53DBD residue, are located close to Arg280 of the chain B and D. Their DNA backbone superimposition shows a perfect alignment of their α/ϒ torsional angles (Fig 4-21B).

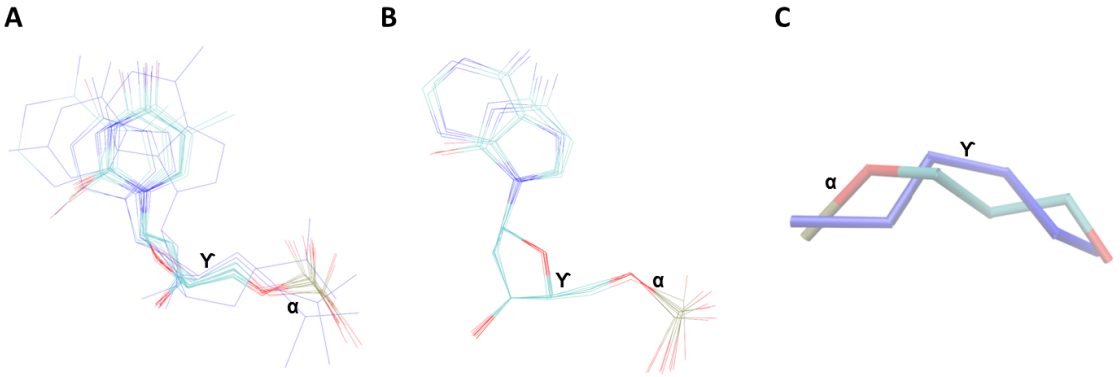


Figure 4-21: α and ϒ torsional angles comparison of the common nucleotides found in cluster 1-2 and in cluster 3-4 from the engineered structures.
A) DNA backbone alignment of the redundant nucleotides in cluster 3 and 4 (3TS8-nuc12-chainK chosen as reference). The common nucleotides found in the three structures are coloured by name and 3Q05-nuc24-chainK is coloured in red, 3Q06-nuc33-chainL coloured in green and 3Q06-nuc6-chainK coloured in purple. **B)** DNA backbone alignment of the redundant nucleotides in cluster 1 and 2 (3TS8-nuc8-chainK chosen as reference). **C)** the average strutures from the nucleotides presented in A and B were calculated using ptraj and were aligned. The average structure calculated from the redondant nucleotide in cluster1-2 is coloured in blue. The visualisation was performed in VMD.

Cluster 1 and 3 are related by reflection through the origin as well as cluster 2 and 4 which implies that α and ϒ torsional angles in cluster 1 and 2 are counterpoised in cluster 3 and 4. In fact, superimposition of the average structures obtained from the nucleotides in cluster 1 and 2 and those from nucleotides in cluster 3 and 4 (Fig. 4-21C) shows an opposite orientation of their α and ϒ angles as expected. Moreover, the dinucleotides in cluster 1 and 2 are located at one side of the DNA helix while the dinucleotides in clusters 3 and 4 are located on the opposite side of the helix (Fig. 4-22). This cluster organisation although highlighting nucleotides in direct contact with the Arg280 or close to residues known to bind DNA, in the engineered structures, is mainly due to the location of the nucleotide in which there is α and ϒ variance in the DNA helix. Concerning the other structures found in the four main clusters, especially 1TSR/1TUP and 3KMD, these dinucleotides show contact with p53DBD structure (data not shown), but none of them in the same way as the engineered structure except for nucleotide 4 in chain F of 3KMD structure which is found in both cluster 3 and 4. Nucleotide 4 is, however, not bound to any residue from p53DBD rather it is surrounded by the Arg280 of chain B that binds nucleotide 3 and Arg248 of chain B (NH_2 at 6.85 Å from nuc4 O2). Concerning the structures binding to two DNA helices (2AHI, 2ATA,

2AC0, 2ADY and 3D0A) most of the dinucleotides extracted from these structures are located in the central cluster (Fig. 4-19B), as these structures are considered binding as dimers and not as tetramers, no DNA bending was expected, their location in the structure projection (Fig. 4-19B) comfirms our results.

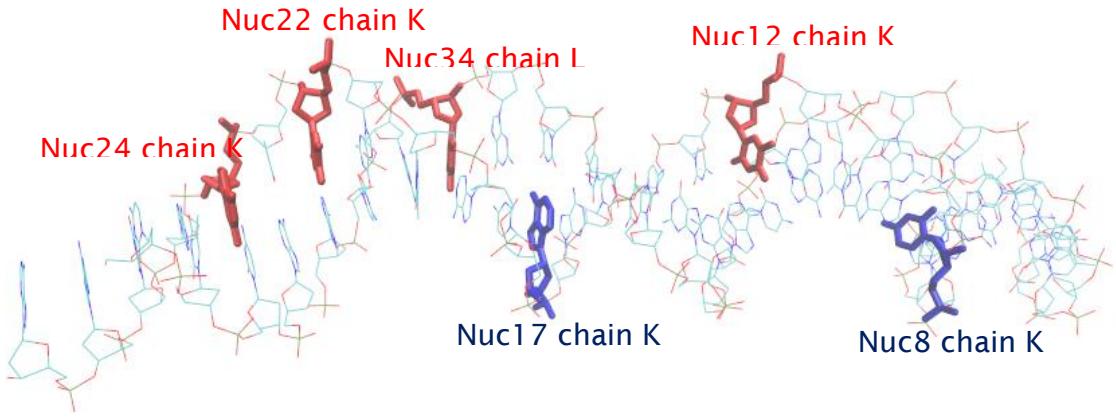


Figure 4-22: Nucleotides with α/ϒ variance in the engineered structures. Common nucleotides for all engineered structures found in cluster 1 and 2 (licorice representation, blue coloured) in cluster 3 and 4 (licorice representation red coloured).

In the case of 1TSR and 1TUP, containing three p53DBD monomers and a long DNA helix, only the chain B and C are bound to DNA, chain B binds specifically to DNA while chain C non-specifically. Chain A interacts with DNA but not as expected via the DNA interface (loop-sheet-helix motif), for this reason this structure has been described as representing another mode of interaction compared to the p53DBD-DNA complexes experimentally solved (123). However visualisation of 1TSR/1TUP dinucleotides located in the four clusters and the ones scattered around the clusters, shows that the α/ϒ variance involves any type of dinucleotide – in direct contact with p53 chain B and C, some in between two half-sites and end of helix nucleotides. In the case of 1TSR/1TUP dinucleotides directly bound to p53 there is no common α/ϒ variance as observed in the engineered structures. Indeed, dinucleotides found in cluster 1 are not represented in cluster 2 and the alignment of 1TSR nucleotides in cluster 1 and scattered around the clusters which directly bind Arg280 does not align properly and does not align on the average structure calculated from common nucleotides found in cluster 1 and 2 for the engineered structures. Also no DNA bending was expected for these structures

as only one monomer of p53DBD is specifically bound to the DNA in these structures.

DNA PCA appears to offer a good tool for highlighting p53DBD experimental structures that bind properly to p53. Indeed this analysis performed on all the p53 structures known to bind a double helix DNA shows DNA bending in engineered structures, described as been the best model of p53DBD-DNA binding (96, 113). Those structures are thought to be better DNA binders because of their extra tetramer domain. The little amount of nucleotides concerned by this bending might be explained by the fact that engineered structures bind to DNA as dimers and not as tetramers.

4.4 Conclusions

The PCA of the experimental structures highlights the importance of the flexibility of p53 DNA binding surface. The clustering of 2FEJ NMR structures which was originally thought to be only caused by the NMR technique (112) is now proven to be due to its apo states (as 2MEJ NMR structures cluster with the other X-ray structures). This result taken together with the PCA of the wild type structures (apo states versus p53DBD bound to DNA), show that one of the most flexible region of p53 DNA binding surface is the loop L1. This loop has been shown to present different conformation depending on the p53DBD state (bound or apo) and on which part of the DNA p53DBD binds; PCA has been shown to be able to distinguish between these different conformations. We can now state without doubt that the L1 loop conformations observed in the 2FEJ structure are not DNA binding-like (as stated in the NMR study (103)) but possible L1 conformation of p53DBD apo state and this flexibility was shown by Halazonetis et *al.* to be important for the DNA binding specificity (96, 113, 114).

p53DBD apo state is also thought to have a flexible L2 loop (from residue 181 to 194), even though most of the free X-ray structures present a L2 loop conformation similar to the one of the p53DBD bound to DNA, some structures as 3Q01, 4MZR and 2FEJ show different conformations of L2. Unlike L1 no specific conformation of L2 was highlighted for the p53DBD apo state. In the case of the loop L3, two distinct conformations were highlighted by PCA: a

common loop L3 conformation present in most p53DBD structures in which Met246 is buried and Met243 exposed and a methionine switched conformation of L3 mainly observed on p53DBD linked to oncogenic situations (mutant or binding to an oncogenic partner). These results exhibit a limitation of X-ray technique to capture the flexibility of p53 free state, which can be misleading as this flexibility (i.e. loop L1) can be important for p53DBD function. Knowing that, one can be inclined to think that the loop L3 conformation highlighted in 2H1L and in the mutant R249S instead of only been a oncogenic conformation can be a normal p53DBD conformation not yet highlighted by experimental techniques.

The PCA applied to torsional angles of DNA sequences bound to p53DBD showed that only dinucleotides bound to engineered structures present angle conformations known to be due to protein binding. Most of these dinucleotides were found to be directly bound to p53DBD residues known to be implicated in the DNA binding. The engineered structures studies are thought to be the first to show a proper p53DBD-DNA binding process implicating an induce fit mechanism (characteristic of most protein binding to DNA) (11, 30, 31). Also, in these studies a natural or natural-like DNA consensus sequence is used. The results of the PCA applied to the DNA torsional angle support their findings that a specific p53 DNA binding needs the cooperation of several domains, pinpointing another limitation of the P53DBD crystallographic structures.

The PCA technique has been successfully able to distinguish between the different p53DBD states based on the loop conformations and to highlight which dinucleotides bind to p53DBD residues. The use of this technique on p53DBD experimental structures (X-rays and NMR) brings out limitations of crystallographic method to study p53DBD due to its flexibility. Indeed, to date all free p53DBD loop L1s were describe as having a DNA binding conformation which is contradict by the PCA technique that is able to distinguish p53DBD structures bound to DNA from the free structures based on the loop L1 differences. A second limitations pinpointed by PCA is on the DNA binding of p53DBD. Although these structures are bound to DNA, one can legitimately question the binding specificity as it is commonly accepted that the p53DBD DNA binding specificity is driven by the C-ter domain and none of the experimental structures contain this domain.

Chapter 5. Structural dynamics of the DNA binding domain of p53: effects of Cysteine alkylation

5.1 Introduction

The conformation of the p53DBD required for DNA binding has been found to be influenced by redox conditions. DNA binding assay experiments have suggested that a reduced environment is essential for the formation of p53-DNA complexes, as the oxidation of cysteine residues inhibits the DNA-binding capacity of p53 and induces the cysteine sulfhydryl groups to form inter and intramolecular disulfide bonds (74). There are 10 cysteine residues located in the DBD (Cys124, 135, 141, 176, 182, 229, 238, 242, 275 and 277, see Fig. 5-1), and 3 of them are coordinated to a zinc ion (Cys176, 238 and 242). All except Cys229 are highly conserved; this is a common structural feature of many redox sensitive transcription factors (138, 139) (see section 3.4.3).

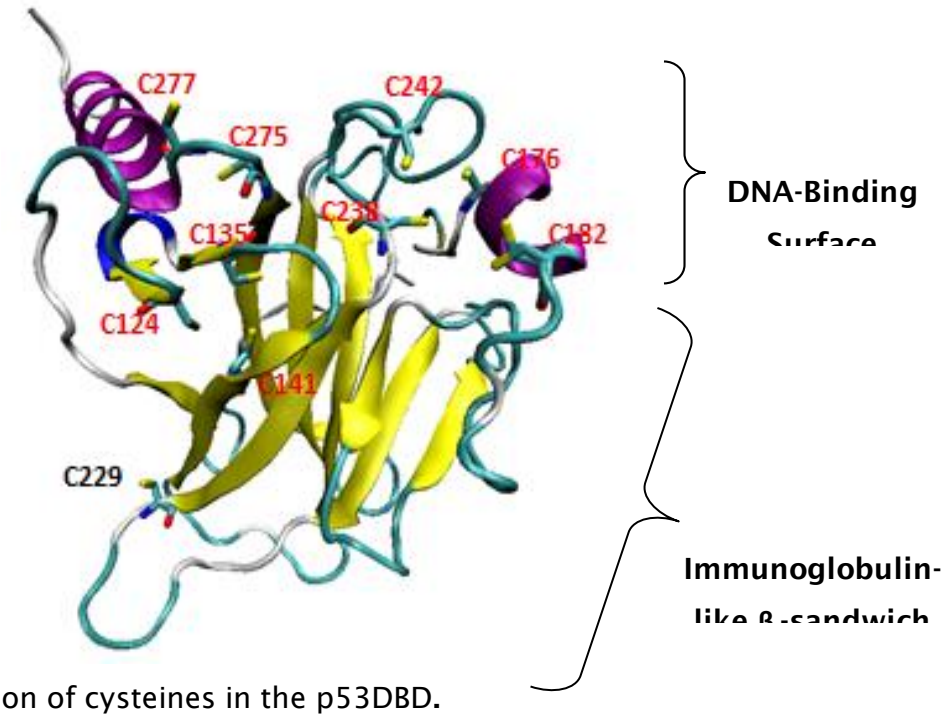


Figure 5-1: Location of cysteines in the p53DBD.
The 2XWR structure is shown in cartoon and coloured by secondary-structure, with the cysteines represented in liquorice and labelled. The conserved cysteines are labelled in red and the non-conserved cysteine (C229) is labelled in black (image generated using VMD.

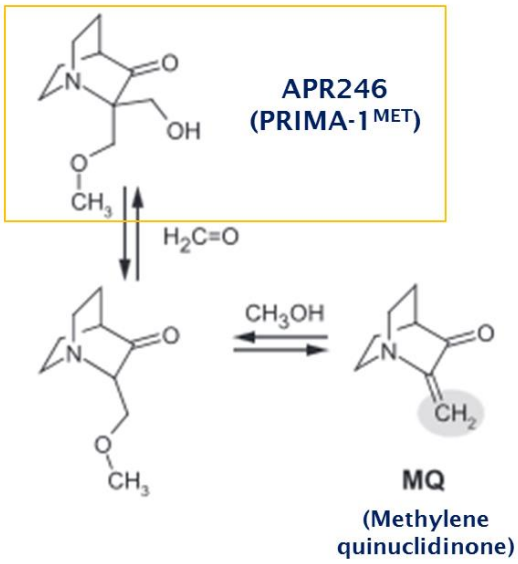
Investigations of the effects of oxidation on cysteine residues (in full p53), particularly the effect of cysteine alkylation using Michael acceptors such as N-ethylmaleimide (NEM) and other Maleimide analogues (MIRA1), have shown that NEM alkylation of wild type and mutant p53 enhances the formation of wild-type p53-DNA complexes when NEM is added in low concentrations (10 to 20 μM) (107, 140), while it disturbs wild-type p53 binding to DNA binding at high concentrations (2-4 mM) (76). Mass spectrometry experiments reported by Langridge-Smith *et al.* were used to hypothesize that such concentration-dependent effects of NEM alkylation on p53 DNA binding could be due to a stepwise mechanism involving the initial alkylation of 2 cysteines, followed by a third alkylation which in turn leads to global unfolding of p53, thus enabling the alkylation of the remaining seven cysteines (109). The first two cysteines to be alkylated were reported to be C182 and C277, while the identity of the third remained ambiguous, although two possible candidates were identified as C229 and C275.

A study by *Fersht et al.* (108) on the reactivity of cysteines towards Michael acceptors reported results that varied depending on the technique used

(inspection of electron density in X-ray data, mass spectrometry) and the structure used (p53DBD; p53-full-length; p53DBD-Y220C, which is a stable variant with 4 additional stabilising mutations other than Y220C) (108). Inspection of the electron density in p53DBD-Y220C highlighted density changes on residues C182, C277 and C229 when the crystal was soaked with 3-benzoylacrylic acid (an alkylating agent, MA), while mass spectrometric analysis of p53DBD-Y220C found C124 and C141 to be involved in the first two alkylation steps and C182 and C277 in the 4th and 5th steps, respectively. From these results they concluded that a good strategy to rescue p53 mutants would be to target the cysteine cluster of C124, C135 and C141. An interesting result from this work is that by studying the alkylation of the whole p53 protein the authors were able to highlight a mono-alkylation step (at the beginning of the alkylation process), consistent with the findings of Langridge-Smith *et al*. (109). Indeed, neither groups could identify the cysteine that was alkylated or its role in p53 function.

Langridge-Smith *et al*. (109) correlated the stepwise alkylation of the cysteines in p53DBD to the concentration dependent effects that several other Michael acceptors (PRIMA-1, MIRA-1 and PRIMA1-MET) have on p53 DNA binding. However, in contrast to the findings of *Fersht et al.,* they ascribed the inhibition of p53 DNA binding to the 3rd alkylation, which induces p53DBD unfolding. On the other hand, *Fersht et al.* suggested that inhibition of the binding is driven by the alkylation of C277, while alkylation of C182 in the DBD-DBD interface might enhance dimerization during the DNA binding process and in turn facilitate the formation of the dimer-DNA complex.

Such keen interest in MAs started in the early 2000s when APR246, a pro-drug that is metabolised into a Michael acceptor (Scheme 5-1) and re-activates p53 mutant and unfolded wild-type (91), was discovered by a research group at the Karolinska Institute, Sweden (141). This pro-drug, now in clinical trials, showed anticancer effects in several *in vivo* models. Its minimal secondary effects on test patients even following extended exposure made it an ideal candidate to treat cancers characterized by p53 mutants. Recent studies (91, 107–109) using MAs have tried to understand how thiol alkylation in mutant p53 could restore the native conformation in the DBD. Answering this key question will enable the design of potent and p53 mutant specific anticancer drugs.



Scheme 5-1: PRIMA-1^{MET} metabolism (91).

In this study, conventional and accelerated molecular dynamics (cMD and aMD) simulations have been used to model the stepwise alkylation mechanism in p53DBD. The results obtained allowed us to:

(i) confirm the identity of the 2 cysteines that are the target of the first alkylation step by monitoring the solvent accessibility of the thiol groups

(ii) suggest the identity of the cysteine which is alkylated in the second step and how the third alkylation enhances the reactivity of the 7 remaining cysteines towards MAs

(iii) monitor the changes caused in the p53DBD by mono and di-alkylation, and how these changes help explain the experimental results

(iv) propose a mechanism by which cysteine alkylation regulates DNA binding in p53DBD.

5.2 Material and methods

In these simulations, we modelled the alkylation of p53 cysteines (especially C182, C277, C229 and C275) with N-ethylmaleimide (NEM) (see Fig. 5-2). NEM, a derivative of maleic acid, contains an imide functional group and an alkene bond. This is reactive towards thiols and is commonly used to modify ('cap') cysteine residues in proteins and peptides. The resulting thioether features a strong C-S bond and the reaction is virtually irreversible (see Fig. 5-2, red circled). This reaction occurs in a pH range of 6.5-7.5.

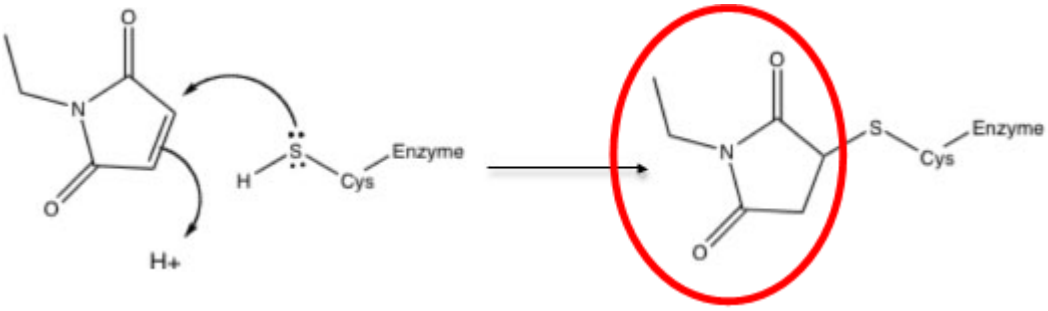


Figure 5-2: Mechanism of NEM alkylation of a representative cysteine residue.

5.2.1 Alkene group setting

ff99SB and GAFF force fields (142, 143) were used in our simulations, but they do not contain parameters for the NEM group. The first step was to develop the parameters for the alkene in accordance with the ff99SB and GAFF force fields. The alkene group was drawn using DS Visualizer (144), a chemical editor for drawing chemical structures and reactions. The alkene part of NEM interacting with the cysteine thiol was drawn, hydrogen atoms added (in order to satisfy the valence) and the molecular geometry corrected via minimization. We used the Red Server (145) to assign charges to the molecule. The charge assignment allowed us to create a library of parameters for NEM. The molecule was loaded in XLEAP and edited in order to draw the bonds (as XLEAP does not show them). The 'edit selected atoms' option in XLEAP opens a table in which the atom types and the charges calculated previously for the NEM group can be loaded and saved. The missing parameters for NEM were taken from the supplementary table S4 of (146).

5.2.2 The zinc model

The p53DBD contains a single Zinc coordinated by Cys176, Cys238, Cys242 and His179 connecting the L2 and L3 loops. Simulation studies of the Zinc influence on p53 stability and behaviour show that it is involved in the sequence specific DNA-binding and also coordinates movements of different structural elements in the protein (102). There are several approaches to incorporate metal ions into force fields. All are based on three different Zinc models: the bonded, non-bonded and semi-bonded models (147, 148).

Considering the advantages and disadvantages of each model, the bonded model was chosen for the Zinc parameterisation. Indeed p53 is a Zinc containing system in which the zinc is coordinated only to p53 residues, not to the DNA or other ligands. Furthermore Zinc is not implicated in a catalytic activity. In this way, we can suppose that Zinc keeps a tetrahedral conformation which in turn allows the use of the bonded model.

For this purpose some manual changes were made to the pdbfile:

- Hydrogen atoms of the cysteine residues coordinating the Zinc have been removed.
- The cysteine names have been changed in CYM from compatibility with the force field (negatively charged).
- The Zinc name has been changed as “ZN” and its atom name as “Zn”.
- The pseudo bond parameters and the charge for the zinc coordinates have been taken from other Zinc parameterization studies (149).

5.2.3 Protein model

The crystallographic structure 2XWR (PDB code) (150) was chosen as the model for this study. This structure includes residues W91 to L289; W91 is located at the N-ter region and is important for p53 tetramer stabilization (150). Only chain A of this structure was kept, without crystallographic water or zinc (these are removed by the MODLOOP and WHATIF (151, 152)processes and added again later). Several residues, S106, M133, C141 and S166, can adopt two possible conformations (α and β). For these residues only the conformation with the highest occupancy was kept. The coordinates for the missing

residues: R290, K291 and K292 were taken from the X-ray structure 1GZH (PDB code) chain A, and G293 from 3IGL (PDB code), wild type structures for which these residues are complete (i.e. there are no missing atoms). The missing residues at the C-ter of the p53DBD (R290, K291, K292, G293) were added using MODLOOP, to adjust the coordinates to the new structure. The system was protonated using WHATIF, which adds the polar hydrogens (other hydrogens are added later by LEAP). After protonation, the model was checked via MOLPROBITY for clashes, and the residue Q165 was flipped. WHATIF does not correctly protonate the histidines (resulting in bad naming and wrong protonation on the histidine binding to the zinc). The hydrogens added on these histidines by WHATIF were removed and their naming changed as follows: Histidines 178, 214 and 233 were labelled HID and Histidines 179, 115, 168 and 193 were labelled HIE. The hydrogen bonding potential around each of the sidechain nitrogen atoms was examined and the one with maximum donor potential was protonated (in HIE, the hydrogen atom is attached to the ε nitrogen atom and in HID, the hydrogen atom is attached to the δ nitrogen atoms). This was followed by adding the remaining hydrogens using LEAP. Finally the coordinates for the crystallographic water and the zinc were added and the file was converted in order to be readable by AMBER.

5.2.4 p53 cysteine alkylation

The NEM groups were added to the model prepared above using PYMOL, in order to position them close to the cysteines of interest (C182, C277, C229 and C275). The cysteines were deprotonated and bonded to NEM using leap. The new pdb file obtained was terminated at the N-terminus and C-terminus by adding capping moities (ACE and NME), the zinc bonded to atom SG of the cysteines C176, 238 and 242 and the ND1 atom of H179, the C5 atom of the NEM molecules were bound to the corresponding deprotonated cysteines (to the SG atom). The system obtained was then minimised step by step: the caps, followed by the flipped residue (Q165), followed by the protein and finally the whole system.

5.2.5 Systems

Based on Langridge-Smith *et al.* study (109), the following systems were modelled: apo-p53DBD; 1NEM, in which alkylation was modelled only on one cysteine, either at C182 or C277; 2NEM in which C182 and C277 were both alkylated; 3NEM, in which C229 or C275 were separately alkylated, after 100 ns, 500 ns and 1µs of the 2NEM simulations, resulting in 3 Cysteines being alkylated; 4NEM systems were modelled after 100 ns of 3NEM simulations by adding a fourth alkylation to the system at either C229 or C275. A total of 12 systems were built and for each model, 3 repeat simulations were run for lengths as shown in Figure 5-3.

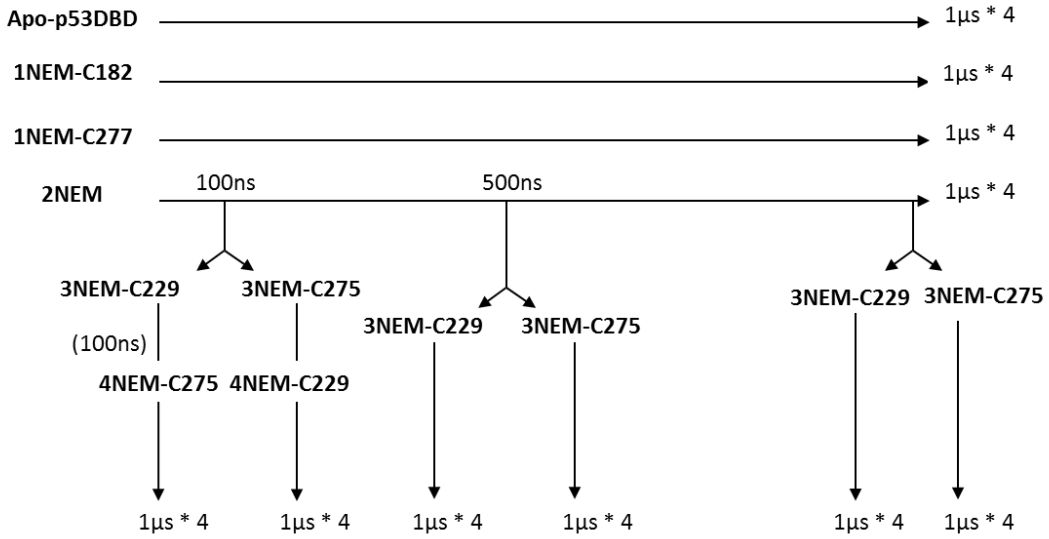


Figure 5-3: Schematic representation of the systems produced.
*4 refers to 4 simulations of the MD simulations, with each repeat started with a different random number seed for the MD simulations.

5.2.6 Water box and ion concentration

The systems were modelled at physiological conditions with an ionic strength of 0.05 M (to accelerate the equilibration) and a cubic box of TIP3P waters with a distance of 10 Å. Details for the different systems are reported in the table below.

	Apo-p53DBD	1NEM	2NEM	3NEM	4NEM
Number of ions	14 Cl^- (5 to neutralize) and 9 K^+	14 Cl^- (4 to neutralize) and 10 K^+	14 Cl^- (3 to neutralize) and 11 K^+	13 Cl^- (2 to neutralize) and 11 K^+	12 Cl^- (1 to neutralize) and 11 K^+
Water box dimension, Å (Volume, Å³)	79.7, 88.1, 71.4 (501315.6)	**C182** 81.9, 86.8, 73.2 (520473.5) **C277** 79.1, 90.6, 75.2 (538676.3)	78.7, 94.1, 78 (577688.6)	**C229** 83.5, 92.1, 77.7 (597519.6) **C275** 84.1, 90.6, 73.3 (557917.8)	**C229** 82.2, 94.2, 72.2 (559203) **C275** 87.7, 88.6, 76.8 (597155.6)
Number of atoms	39695	**C182** 41701 **C277** 43714	46758	**C229** 48471 **C275** 44373	**C229** 44829 **C275** 48222

Table 5-1: Details of the different systems modelled.
Apo-p53DBD, mono-alkylated **1NEM** (p53-C182 and p53-C277), **2NEM** with alkylations on C182 and C277 (p53-2NEM), **3NEM** with alkylations on C182, C277 and a third alkylation on C229 or C275 (only the systems modelled after 100 ns of p53-2NEM are reported, details for the 500 ns and 1 µs are not shown) and **4NEM** with 4 alkylations (in the order C182-C277-C229-C275 and C182-C277-C275-C229), modelled after 100 ns of 3NEM.

The minimization and equilibration processes were carried out using the SANDER module of Amber 12. The minimization process was carried out gradually using a restraint of 1000 $kcal.mol^{-1}.Å^{-2}$ (minimization order: caps first, NEM, water, water + ions, then whole protein and finally the whole system) to remove bad contacts, using 1000 steepest descent steps and 5000 conjugate gradient steps. The equilibration was also performed gradually. During the first step, a restraint was applied to the system to induce ion and hydrogen atom relaxation, by applying a positional restraint on protein heavy atoms (1000 $kcal.mol^{-1}.Å^{-2}$). Then the temperature was raised from 100 K to 300 K (using Langevin thermostat) over 200 ps at a constant volume and protein restraints in order to equilibrate the solvent. To adjust the size of the water box, a second equilibration step was then performed on the solvent (protein restrained) at constant temperature (300 K) and pressure (1 bar) over 200 ps using the Langevin piston coupling algorithm. The third equilibration step was applied to decrease the temperature of the system from 300 K to 100

K in 100 ps (protein restrained and at constant volume). Once the solvent relaxed, six minimization processes were applied using decreasing restraint weights on the protein (from 1000 to 0 kcal.mol^{-1}.Å^2). The temperature was raised gradually (60 K increments over 10 ps) from 0 to 300 K at constant volume. As a final step, the system was maintained at constant temperature (300 K) and constant pressure (1 bar) over 200 ps. The SHAKE algorithm (22) was used to constrain bonds involving hydrogen atoms, allowing for an integration time step of 2.0 fs. The Particle Mesh Ewald method (153) was employed to treat long range electrostatics with a 12 Å cut-off. The whole process was carried out to enable water molecules and ions to find optimal positions in the system and to obtain an equilibrated state suitable to perform MD simulations.

5.2.7 MD production

48 (12 systems * 4 repeats each) constant NPT simulations were run for 1 μs each using the PMEMD module of Amber 12 (12, 28) and ff99SB and GAFF force fields (142, 143). These simulations were performed on the equilibrated structures kept at 310 K and a pressure of 1 bar using the Langevin piston coupling algorithm. The SHAKE algorithm was used to freeze bonds involving hydrogen atoms, allowing for an integration time step of 2.0 ps. The Particle Mesh Ewald method (PME) was employed to treat long range electrostatics with a 12 Å cut-off. Conformations were collected every 50 ps, resulting in 20000 snapshots per simulation.

5.2.8 aMD setup

The parameters for each aMD simulation run were calculated as indicated in the AMBER manual from the last 10 ns (990 ns to 1μs) of each previous cMD simulation. The values used in the aMD script are reported in table 5-2.

	Alpha D	Ethresh D	Alpha P	Ethresh P
Apo-p53DBD	142.1	2840.5	7939.0	-115175.0
2NEM-1st	142.1	2892.0	9351.6	-136023.4
2NEM-RP3	142.1	2883.5	9351.6	-136012.4
3NEM-C229	142.1	2906.5	8629.2	-125315.3
3NEM-C275	142.1	2909.5	8024.4	-116397.0

Table 5-2: aMD parameters.

5.2.9 Dimer models

The apo-p53DBD and 1NEM-C182 dimers were built by superimposing two monomers taken from the corresponding simulations (at 0, 500 and 1000 ns) on to the crystal structure of a p53DBD bound to DNA (PDB id 2AHI, resolution of 1.85 Å) (132). One monomer was aligned on to chain A and the other on to chain B of 2AHI. Three dimers were produced for each system. These dimers were then minimised using 1000 steepest descent steps and 5000 conjugate gradient steps.

5.3 Results

5.3.1 The 'reactive' cysteines (C182 and C277)

Langridge-Smith *et al.* (109) point out that the two cysteines that were alkylated first, the so-called 'reactive' cysteines, were C182 and C277 and were also the most exposed to the solvent (Table 5-3) based on the solvent accessibility surface area (SASA) calculated for the X-ray structure PDB code 2OCJ. However, the SASA of the X-ray structure (2XWR chain A) used in our study to build the different systems showed differing results depending on whether the accessibility of the whole residue was considered or only that of the thiol terminal groups. When the whole residue was accounted for, C182 was the most exposed (44.4 $Å^2$), while C277 displayed a lower solvent accessibility (15 $Å^2$), and was even lower than that of C229 (21.6 $Å^2$) and C124 (19.7 $Å^2$), making it less susceptible to alkylation during the early stage of the

process (Table 5-3). However, when the accessibility of only the thiol groups was examined, C229 was found to be the most exposed (15 Å^2), followed by C277 (8.5 Å^2), and C182 (7.5 Å^2). Such discrepancies can be explained by differences in crystals obtained by different methods and the rigidity of the protein once packed into a crystal. These results highlight the importance of considering the dynamics of p53DBD when evaluating the reactivity (i.e. the SASA) of the cysteines towards MAs. In our study we focus on the solvent accessibility of the thiol groups only as they are the sites of alkylation by MAs.

Cysteine	Whole cysteine solvent accessibility (Å^2)	Thiol group solvent accessibility (Å^2)	2OCJ solvent accessibility (Å^2) [113]
124	**19.7**	**6.2**	**6.3**
135	0	0	0
141	0.3	0.3	0
176	4.4	4.4	2.9
182	**44.4**	**7.5**	**39.3**
229	**21.6**	**15**	**14.8**
238	0.5	0	1.1
242	**11.6**	**8**	**12**
275	7.5	0	0
277	**15**	**8.5**	**43**

Table 5-3: The solvent accessibility (in Å^2) of p53DBD (PDB code: 2XWR chain A) cysteines and thiol groups calculated using VMD.

Four 1 μs molecular dynamics (MD) simulations were run for the apo structure, resulting in 20000 frames for each simulation and 80000 frames in total. The solvent accessibility over time of all the thiol groups in the four apo simulations was monitored (an example of this analysis for the 1st 1μs simulation is reported in the Appendix B Figure 1). In all simulations, three cysteines, C182, C229 and C277, (and C242 too, but this is coordinated to the Zinc atom and hence not considered in our analysis) were found to be the most exposed (solvent accessibility ranging from 0 to 80 Å^2). The average SASA calculations for each repeat and for the overall simulations (see Fig. 5-4 and app. B Tab. 1) confirm the results by Langridge-Smith *et al.* (109). In fact we

found that C182 (average SASA over four simulations = 48.1 Å^2) and C277 (average SASA = 54 Å^2) were the two most exposed cysteines, which is also in agreement with the p53DBD electron density data reported by Fersht *et al.* (108). The SASA distribution analysis of all thiol groups for the overall simulations (see Fig. 5-5) showed that C277 (majority of SASA lies between 50 and 75 Å^2) is more exposed than C182 (majority of SASA lies between 40 and 60 Å^2). Two other thiol groups presented moderate solvent accessibility, C229 with an average SASA of 23.8 Å^2 (distributed between 10 and 40 Å^2), and C242 with an average SASA of 20.7 Å^2 (distributed between 10 and 30 Å^2).

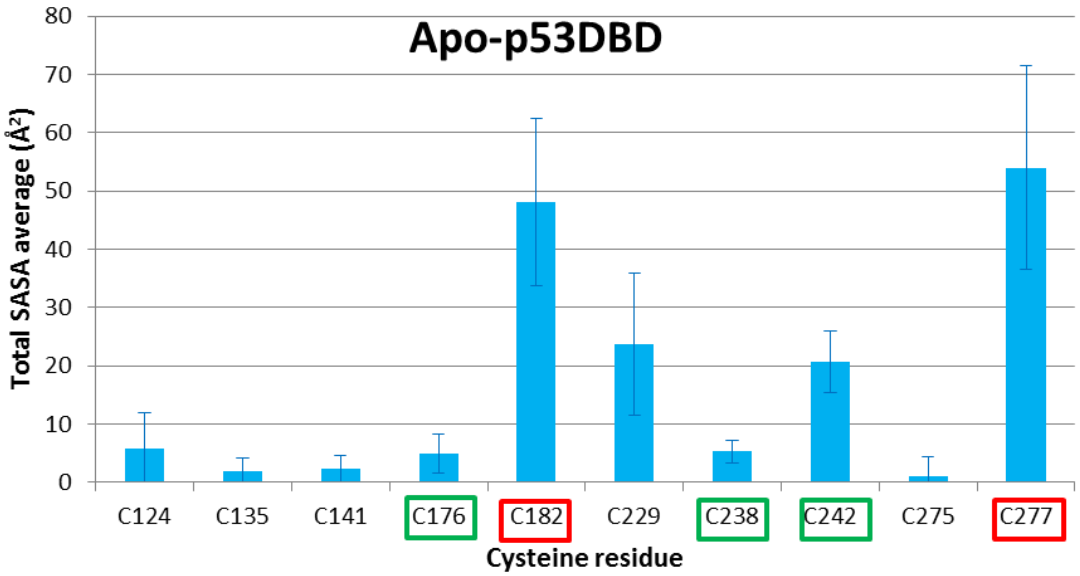


Figure 5-4: Average SASA of cysteines in the apo-p53DBD simulations. C182 and C277 (red boxes) contain the most exposed thiols. The cysteines coordinated to the zinc are in green boxes. The SASA was calculated using VMD and the average SASA (histogram) and the standard deviation (error bars) were calculated using R.

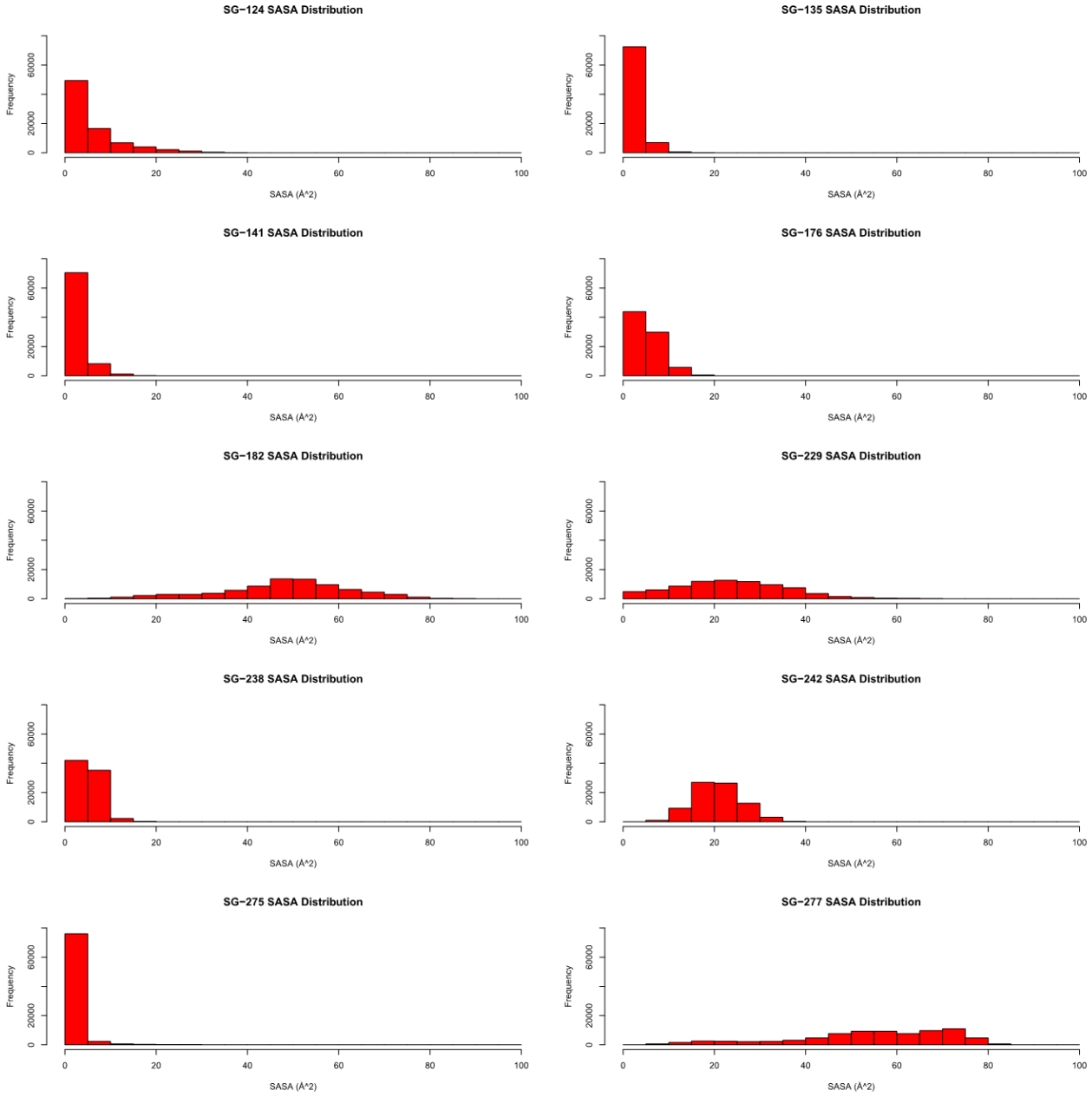


Figure 5-5: Solvent accessibility distribution of thiol groups in the apo-p53DBD simulation.

Both experimental studies (from Langridge-Smith and Fersht) highlighted the presence of a mono-alkylated protein prior to the di-alkylated protein, but they were not able to elucidate which cysteine (C182 or C277) is alkylated first. In our study, two mono-alkylated systems, 1NEM-C182 and 1NEM-C277, were modelled to confirm initially that C182 and C277 are indeed the two reactive cysteines. This was done by checking if any of these mono-alkylations changes the solvent accessibility of the other thiol groups. The solvent accessibility of the thiol groups was determined by simulating the two mono-alkylated systems (1NEM-C182 and 1NEM-C277; see Fig. 5-6, and App. B Fig. 2, 3 and Tables 2, 3).

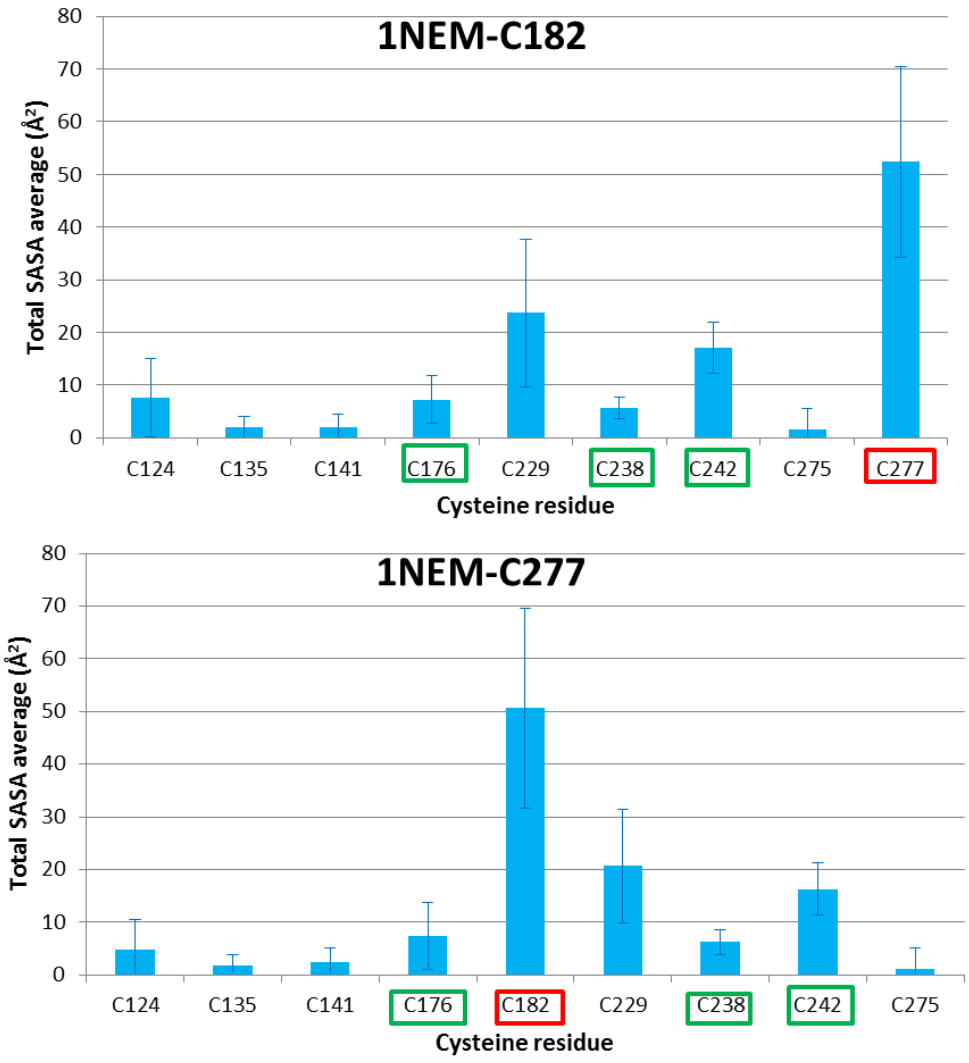


Figure 5-6: Average SASA of the 1NEM simulations.
The SASA was calculated using VMD and the average SASA (histogram) and the standard deviation (error bars) were calculated using R.

The solvent accessibility for the alkylated cysteine is not considered in the SASA calculation. The most reactive cysteine is red boxed and the cysteines coordinated to the zinc are green boxed.

No significant change was observed in the solvent accessibility of the thiols when C182 or C277 are alkylated (Fig. 5-6). C182 and C277 have the most exposed thiols in both the apo and mono-alkylated systems, which makes them the most susceptible to react with NEM. Based on the idea that Michael acceptors will react only with the exposed thiols at low concentration, our results are in agreement with those of Langridge-Smith *et al.* (109) according to which C182 and C277 are the reactive cysteines.

5.3.2 Localization of the third alkylation and its effect on p53DBD stability

Langridge-Smith *et al.* showed that the alkylation of C182 and C277 is followed by a third alkylation, which then results in the alkylation of all the remaining cysteines, and conclude that this can only be accounted for by induction of unfolding of the p53DBD (109). In this experimental study two cysteines were identified as the possible candidates for the third alkylation, C229 and C275. Since these two cysteines are located at each end of the p53 immunoglobulin-like β-sandwich scaffold (named scaffold throughout the thesis) (Fig. 5-1), the authors suggested C229 and C275 could be implicated in p53DBD unfolding that accompanies their alkylation. To investigate the likely cysteine that could be the 3rd site for alkylation, we performed a SASA analysis on the 2NEM simulations. The average SASA (of thiol groups) among our four 1 μs simulations (Fig. 5-7) showed that C229 and C242 are the two cysteines most exposed to the solvent, with an average SASA of 22.7 Å^2 (± 15.4) and 22.6 Å^2 (± 5.9), respectively. Since C242 is coordinated to zinc, we will assume that it is unlikely that the third alkylation occurs on this cysteine. On the other hand C275 has the lowest average SASA (0.9 ± 2.7 Å^2) which makes it less likely to be alkylated by NEM. These results point C229 to be the most susceptible to the third alkylation. Langridge-Smith *et al.* reported the third alkylation as occurring after 5 min of NEM incubation. Of course our simulations last only 1 μs and hence we cannot rule out the fact that C275 may eventually be exposed to solvent. The SASA analysis also does not highlight other thiol groups that

can potentially interact with NEM (C124, C135, C141all with SASA < 10 Å^2; see Fig. 5-7).

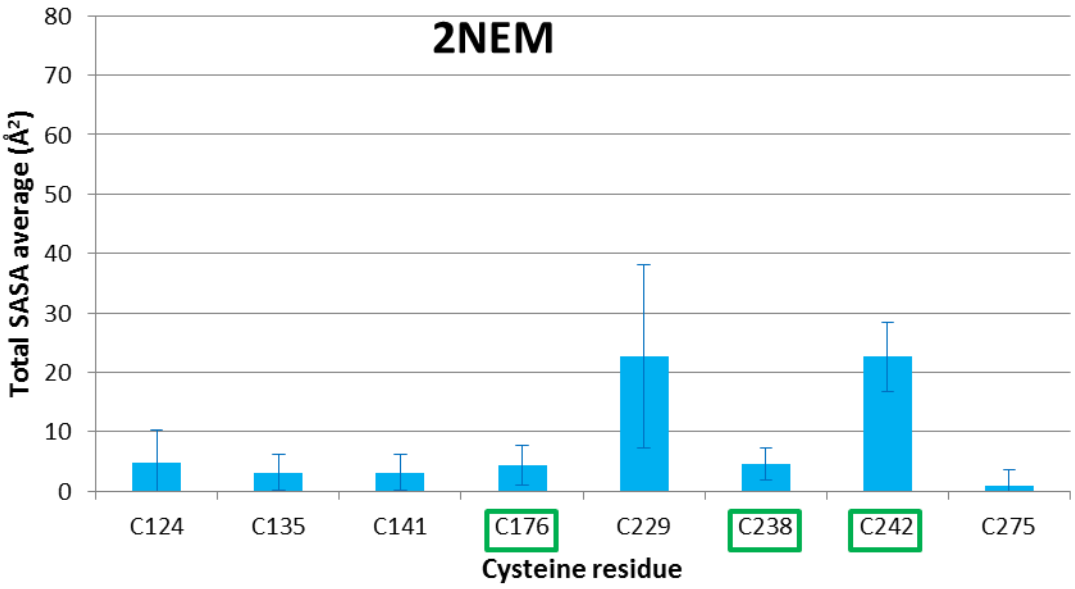


Figure 5-7: Average SASA among p53DBD cysteines di-alkylated on C182 and C277.
The cysteines coordinated to the zinc are green boxed. The SASA was calculated using VMD and the average SASA (histogram) and the standard deviation (error bars) were calculated using R.

To further test the hypothesis of Langridge-Smith et *al.,* we modelled several tri-alkylated systems from the simulations of the di-alkylated systems (see paragraph 5.2.4). Each 3NEM system was derived from a 2NEM conformation extracted at 100 ns (3NEM-C229/275-Aft100ns), 500 ns (3NEM-C229/275-Aft500ns) and 1 µs (3NEM-C229/275-Aft1µs) of the p53-2NEM simulations, on which a third alkylation was modelled, either on C229 or on C275 and a 1 µs simulation performed.

When visually comparing the tri-alkylated systems with the reference structure 2XWR chain A (Fig. 5-8 and 5.9), it can be seen that the overall structure of p53DBD is conserved, even after 2 µs of simulation (in the case of the 3NEM-Aft1µs structures). Most of the fluctuations are located on the N-ter and C-ter of the protein and on the DNA binding surface (DBS), as evidenced in the RMSF analysis (Fig 5-10). Comparison of the RMSFs of the 3NEM-C229 and 3NEM-C275 systems shows that the DNA binding surface in the 3NEM-C229 structure (DBS containing L1, S2, S2', H2, L2/H1 and L3) is more flexible than in the 3NEM-C275 system. This difference is even more obvious in the systems

alkylated on the third cysteine after 1 µs of 2NEM simulation (3NEM-C229/275-Aft1µs). In 3NEM-C229-Aft1µs the RMSFs of the L1 and L3 loops reach 4.5 Å and 3.8 Å respectively, while in 3NEM-C275-Aft1µs these two loops have much lower RMSFs of 2.5 and 1 Å respectively, even lower than in the control system (apo-p53DBD, 3.8 and 1.9 Å respectively; Fig. 5-11). C275 is located on the DBS which explains its effect on this region (see Fig. 5-12). The combination of the three alkylations on the DBS cysteines C182, C275 and C277 seems to have a rigidifying effect on this region.

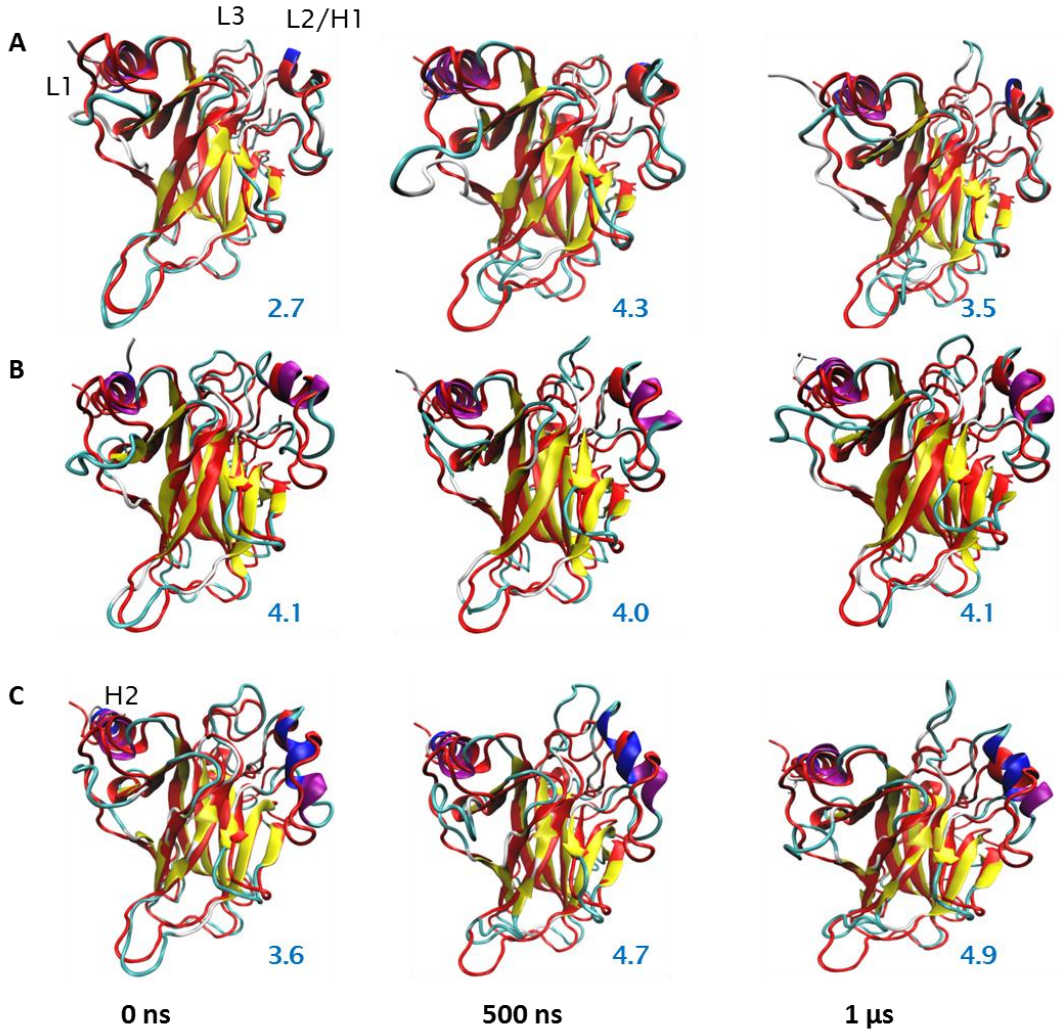


Figure 5-8: VMD visualisation of the 3NEM-C229 system.
The tri-alkylated systems (third alkylation on C229) were aligned on 2XWR chain A (reference structure) alpha-carbon; the coordinates were taken at different times of the simulations. Here we show only the first simulation of the tri-alkylated systems at **A)** 100 ns, **B)** 500 ns, and **C)** 1 µs of 2NEM. The structures are represented in cartoon, the 3NEM-C229 systems are coloured by secondary structure and the reference structure is coloured in red. RMSD of the frame with respect to the reference structure is indicated in blue.

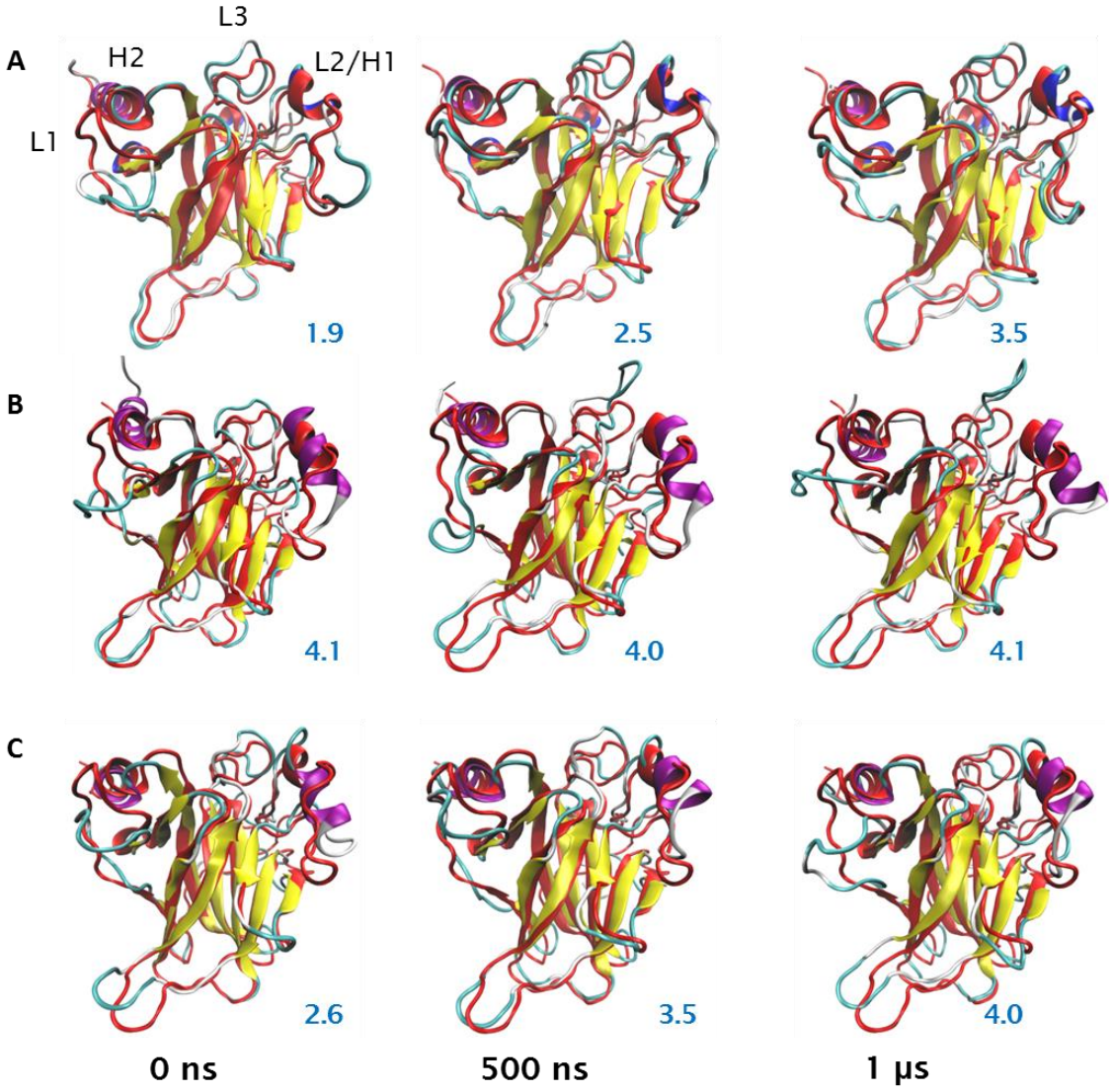


Figure 5-9: VMD visualisation of the 3NEM-C275 system.
The tri-alkylated systems (third alkylation on C275) were aligned on 2XWR chain A (reference structure) alpha-carbon; the coordinates were taken at different times of the simulations. Here we show only the first simulation of the tri-alkylated systems at **A)** 100 ns, **B)** 500 ns, and **C)** 1 µs of 2NEM. The structures are represented in cartoon, the 3NEM-C275 systems are coloured by secondary structure and the reference structure is coloured in red. RMSD of the frame with respect to the reference structure is indicated in blue.

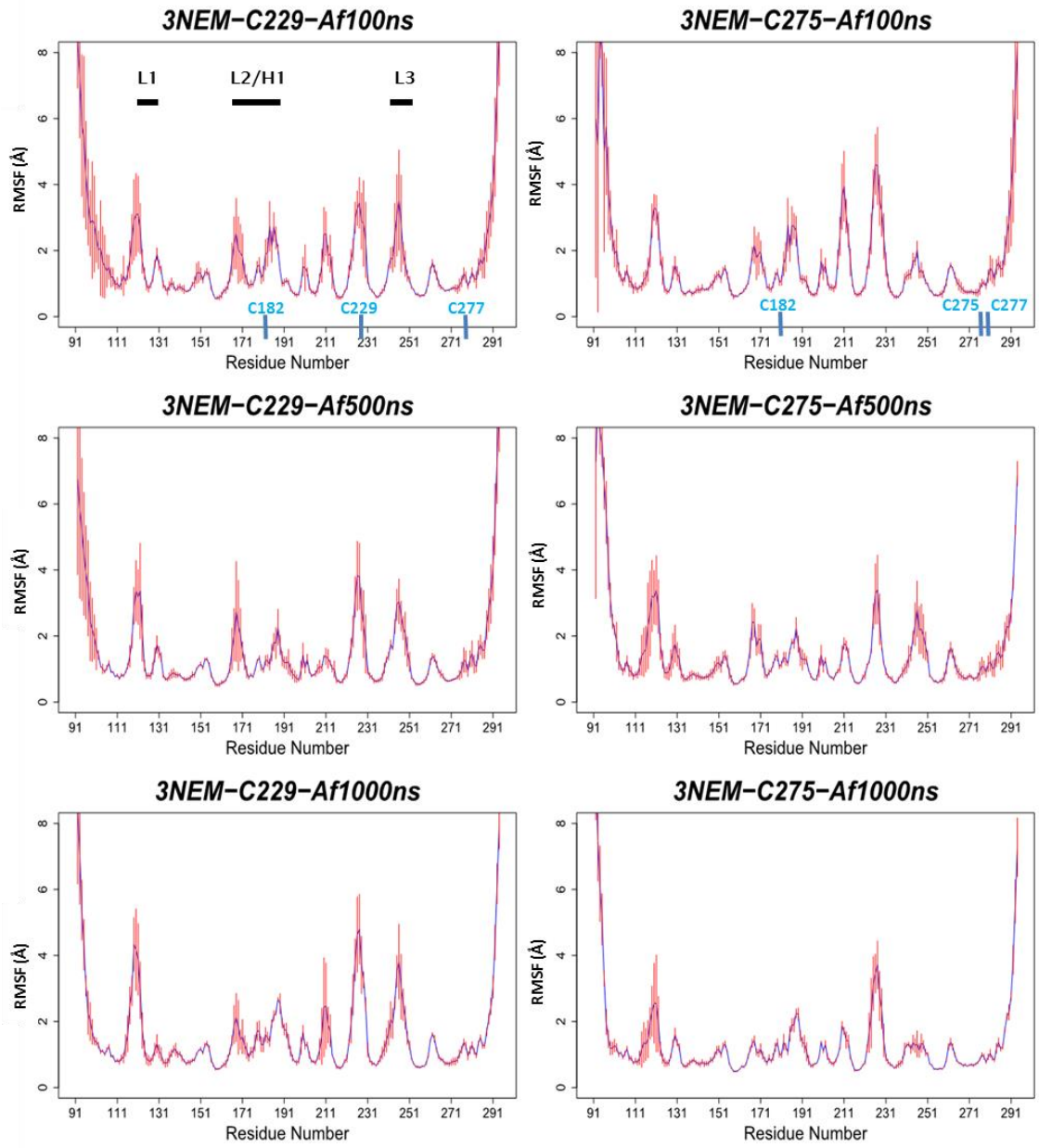


Figure 5-10: RMSF of the 3NEM systems.
The RMSF for each 1 μs simulation was determined using cpptraj in Amber. The average RMSF (blue) among the four cMD and the standard deviation (red) for each system were calculated using R. The alkylated cysteines are labelled in blue.

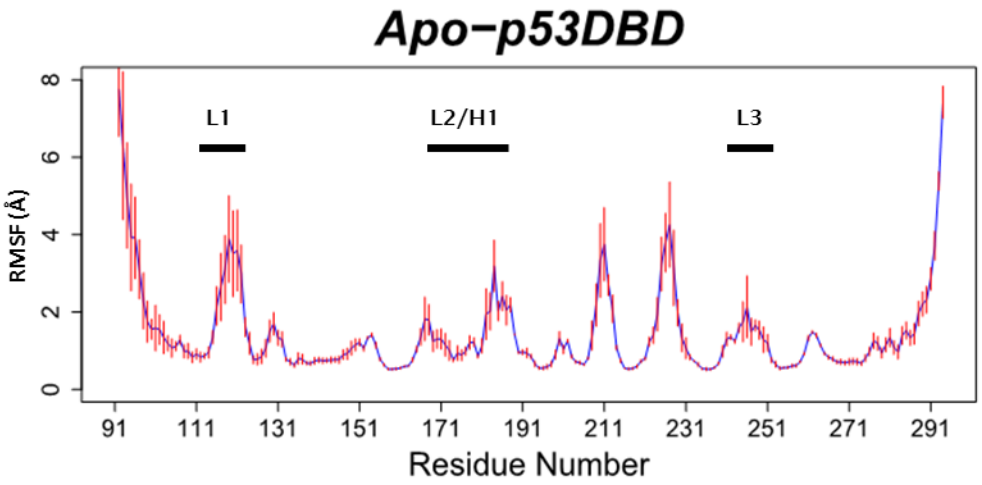


Figure 5-11: RMSF of apo-p53DBD.
The RMSF for each 1 μs simulation was determined using cpptraj in Amber. The average RMSF (blue) among the four cMD and the standard deviation (red) for each system were calculated using R.

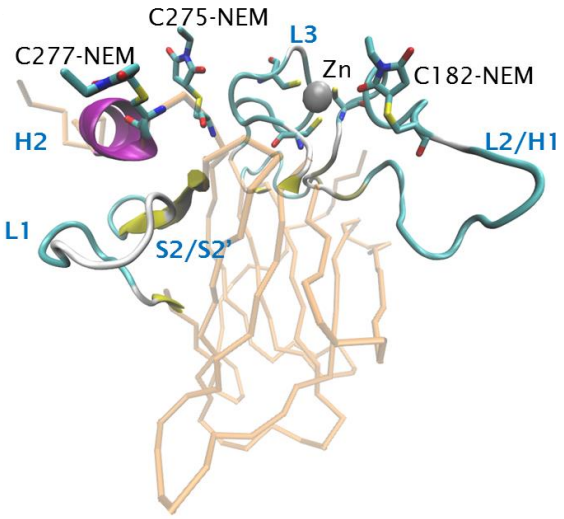


Figure 5-12: The 3NEM-C275 system with the sites of the 3 alkylations (from 3NEM-C275-Aft100ns).
The DBS is presented in cartoon and coloured by secondary structure. The p53 scaffold is shown in trace and coloured in orange. The alkylated cysteines are shown in licorice and labelled.

To compare the fluctuations of the DBS and the immunoglobulin-like β-sandwich scaffold of the DBD, we performed a separated RMSD analysis of these two regions within our simulations (see Fig. 5-13). For simplicity only the RMSD of the 1st simulation of each system was considered (as this was similar in the repeats). The simulations were aligned on the α-carbon of the corresponding region (DBD or scaffold) of the reference structure (2XWR chain A). For comparison the DBS (black line) and the scaffold (red line) RMSD of the same simulation were plotted together. These plots confirmed that most of the fluctuations indeed occur in the DBS region, both in the 3NEM and the apo p53DBD structures. The scaffold region does not fluctuate much during any simulation with RMSD of ~1 Å consistently found for all the systems. On the other hand, the DBS fluctuates during each simulation and varying RMSD values were obtained depending on the system considered. An average DBS RMSD below 3 Å was found for the apo p53DBD, the 3NEM-C275-Aft100ns and 3NEM-C275-Aft1μs, while the RMSD was above 3 Å in all the 3NEM-C229 systems and the 3NEM-C275-Aft500ns. These results confirm that alkylation of C275 affects the rigidity of the DBS region. The fact that the solvent accessibility of the remaining thiol groups in the 3NEM systems (C124, C135, C141 and C229/C275 depending on the system; Fig. 5-14) was found to be close to that of the apo p53DBD structure (Fig. 5-4) confirmed that the p53DBD maintains a high stability throughout the 3rd alkylation.

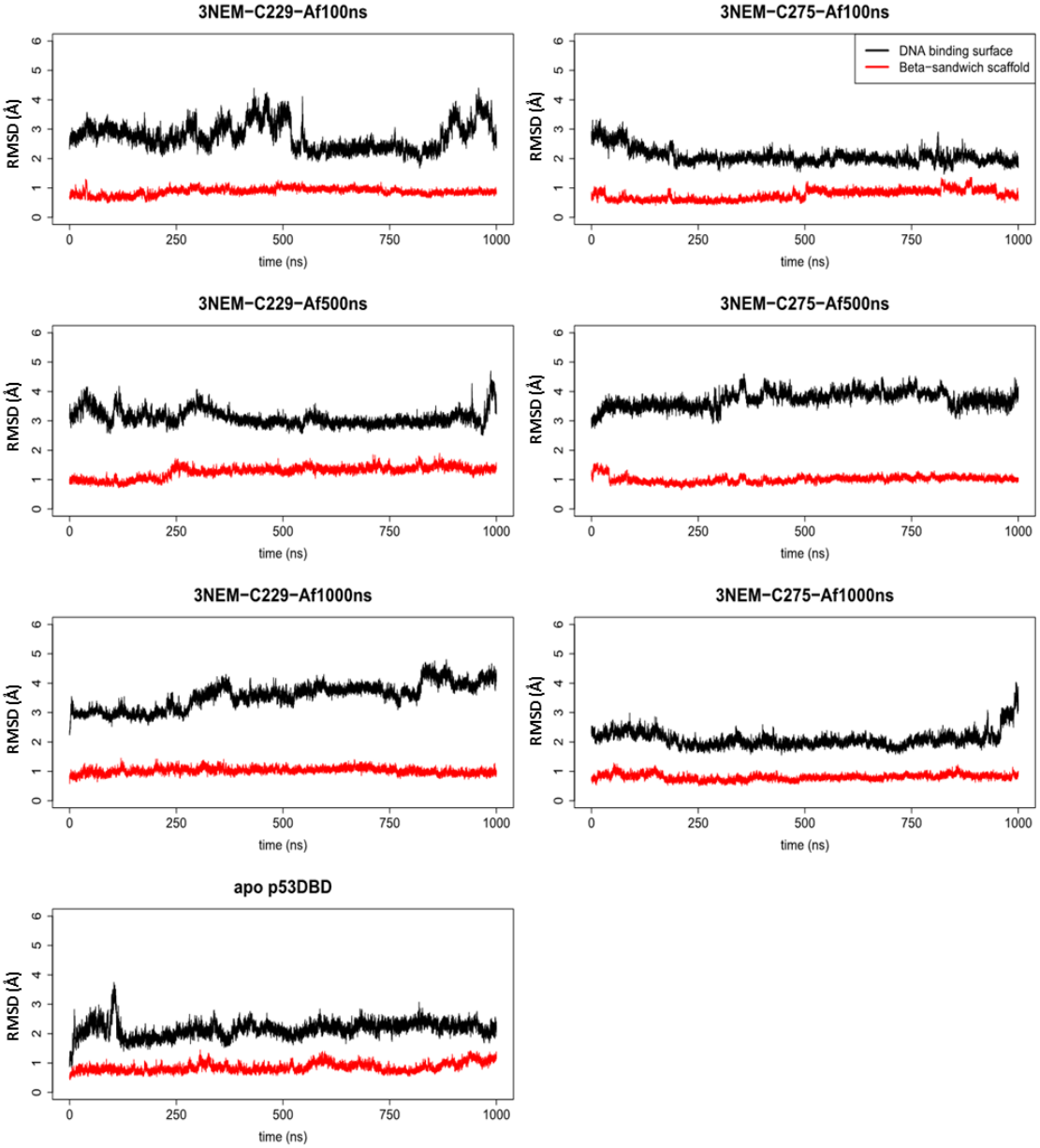


Figure 5-13: RMSD of the 1st simulation of the 3NEM and apo p53 systems. RMSD of the DNA binding surface (DBS, black) against the immunoglobulin-like β-sandwich (scaffold, red) was measured for each 1st simulation of the 3NEM and the p53-apo systems after alignment on the p53 DBS or scaffold.

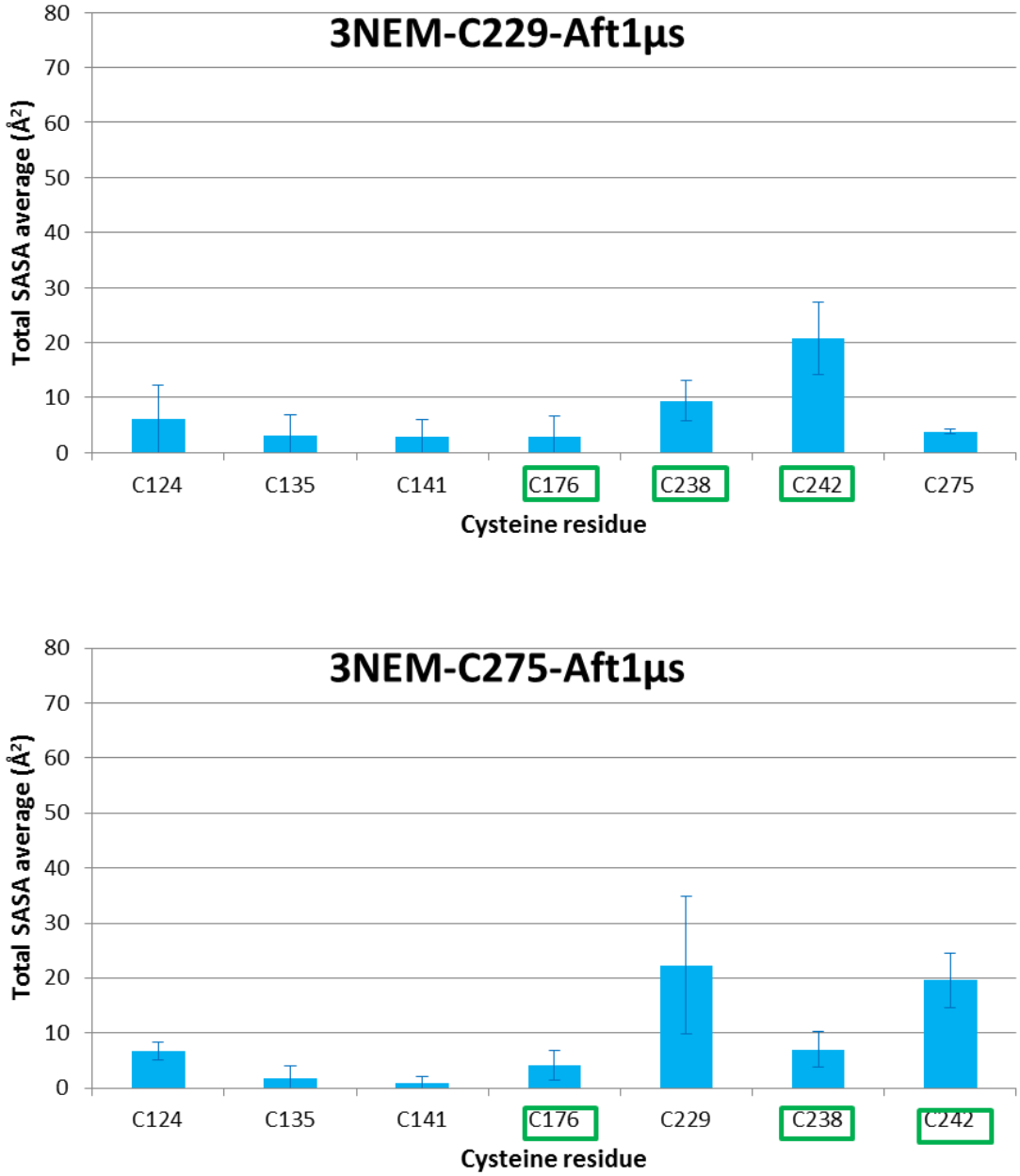


Figure 5-14: Average SASA of cysteines in the 3NEM systems after 1µs.
The cysteines coordinated to the zinc are green boxed.

These results confirmed the stability of the 3NEM systems despite the third alkylation being expected to unfold p53DBD. Although such unfolding and alkylation of the remaining 7 cysteines were described by Langridge-Smith *et al.* as happening instantly after the third alkylation, this event is not observed in our 1 µs simulations. Since long range motions such as protein unfolding typically require times in the order of 10^{-7} to 10^{4} second, both the 1st simulations of the 3NEM-C229/275-Aft1µs systems and the apo-p53DBD control simulations were extended for a further 300 ns using aMD (one repeat

for each simulation was also run). Differently from the cMD simulations, the RMSF (see Fig. 5-15) and visualisation (see Fig. 5-16) of our aMD runs showed more fluctuation in the apo-p53DBD and 3NEM-C275 systems. The unusual flexibility of the N-ter in the apo-p53DBD and 3NEM-C275 systems (residues 94-109, RMSF above (3 ± 2 Å) is due to the fact that this unstructured region adopts either an α-helix or a β-strand conformation (see Fig. 5-17). While no trend towards one or the other conformation was observed for the 3NEM-C275 systems, the apo-structure was found to adopt preferentially an α-helix conformation (visible mainly in the 1st simulation).

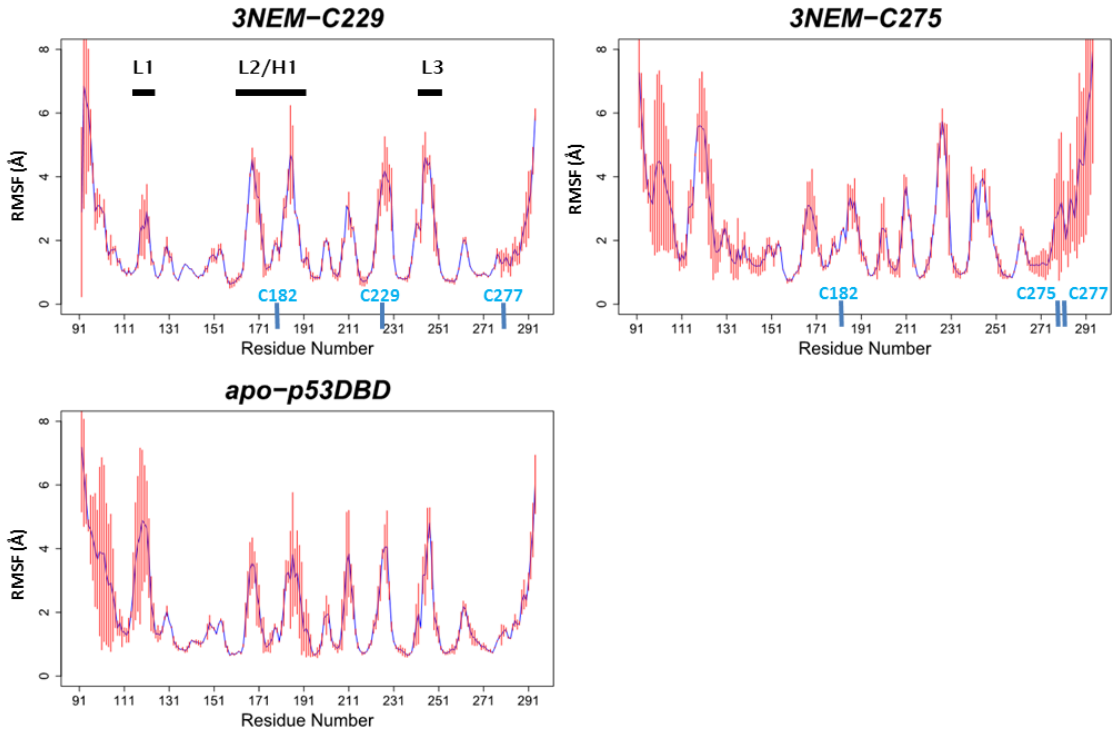


Figure 5-15: RMSF of the p53-3NEM and the apo-p53DBD aMD simulations. The RMSF for each 300 ns simulation was determined using cpptraj in Amber. The average RMSF (blue) among the two aMD and the standard deviation (red) for each system were calculated using R. The alkylated cysteines are labelled in blue.

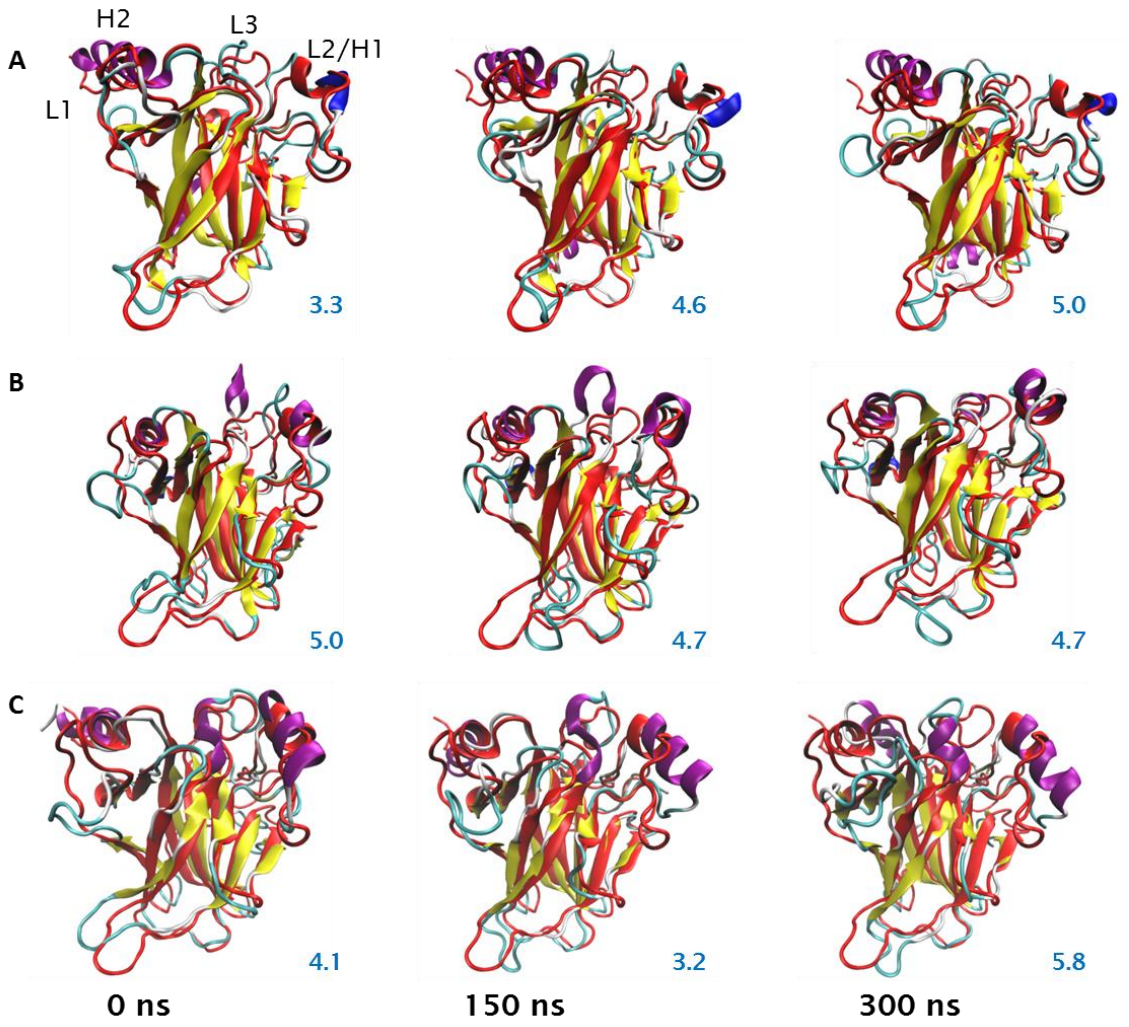


Figure 5-16: VMD visualisation of the aMD simulation of the 3NEM and apo-p53DBD systems.
The simulations were aligned on the reference structure 2XWR chain A (in red and cartoon representation) alpha-carbon at different time of the simulation. Here is shown only the first simulation for each system: **A)** apo-p53DBD, **B)** 3NEM-C229, and **C)** 3NEM-C275. The structures are represented in cartoon, the aMD frames were coloured by secondary structure and the reference structure was coloured in red. The RMSD (in Å) of each frame with respect to the reference structure is indicated in blue.

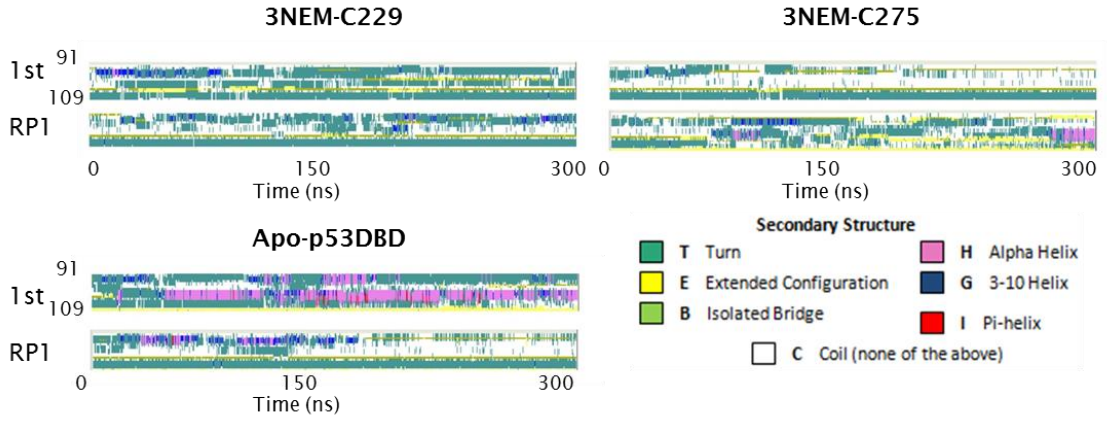


Figure 5-17: Secondary structure determination of the N-ter region in 3NEM and apo-p53DBD aMD simulations.
The secondary structure analysis was performed using timeline analysis in VMD and only the N-ter region (from residue 91 to 109) is shown.

A second unusual fluctuation in the 3NEM-C275 aMD simulation was found to concern the H2 helix. Although this region keeps an intact α-helix secondary structure over the two aMD simulations (Fig. 5-18 A), in the 1st 3NEM-C275 run the H2 helix flips away from its usual location (Fig. 5-18 B). This motion away from the L1 loop exposes the buried cysteine cluster (C124, C135, C141) to the solvent (Fig. 5-18 C), as confirmed by the SASA analysis on the three aMD simulations (Fig. 5-19). Comparison of the SASA averages of the 3NEM and the apo-p53DBD systems showed that C124, C135, C141 in the 3NEM-C275 simulation are more exposed (9.6 Å ± 18.4, 8.2 Å ± 15.1 and 4 Å ± 7.8, respectively) than those in the apo-p53DBD (C124 = 3.6 Å ± 4.9, C135 = 1.7 Å ± 2.1 and C141 = 1.9 Å ± 2.9) and 3NEM-C229 structures (C124 = 2.2 Å ± 2.7, C135 = 1.9 Å ± 2.2 and C141 = 2.5 Å ± 2.5). Indeed alkylation of C275, thought to rigidify p53 DBS when cMD is used, might actually be the one responsible for the unfolding and consequent alkylation of the remaining p53DBD cysteines. Most of the fluctuation was observed on the DBS, in both the aMD and cMD simulations (see Fig. 5.20). Despite the buried cysteines being more exposed, 3NEM-C275 maintained a scaffold RMSD of ± 1Å. In this system, it is the disruption of the H2 α-helix in the DBS, and not the scaffold destabilisation, that leads C124, C135, C141 to become more exposed to the solvent. Alkylation of C275 (also buried in the 2NEM system) was found to increase the stability of DBS secondary structure in cMD, while on the other

hand it was shown to reorganise the DBS in aMD (as H2 helix, while still stable, changes orientation). Such reorganisation has the effect of increasing the accessibility to the solvent of the cluster of buried cysteines. These results point to C275 as being the most likely target for the third alkylation.

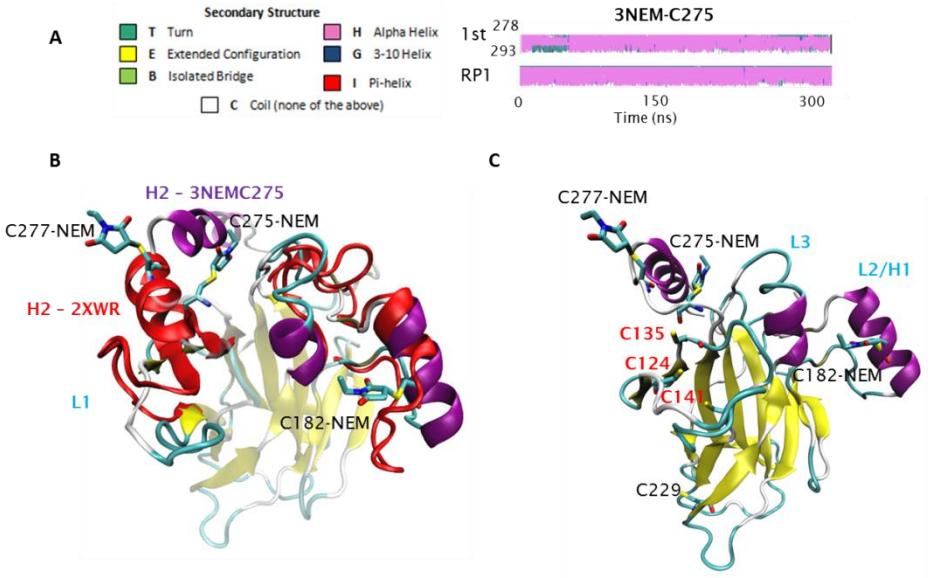


Figure 5-18: H2 α-helix within the 3NEM-C275 aMD simulations.
A) H2 helix secondary structure determination from the 1st simulation and the repeat. **B)** 3NEM-C275 aMD last frame superposed on the carbon alpha of the full reference structure (2XWR chain A, in cartoon coloured in red). **C)** Exposed cysteine cluster: C124, C135, C141. The alkylated systems are presented in cartoon coloured by structure, the alkylated cysteines and the exposed cysteines are shown in licorice.

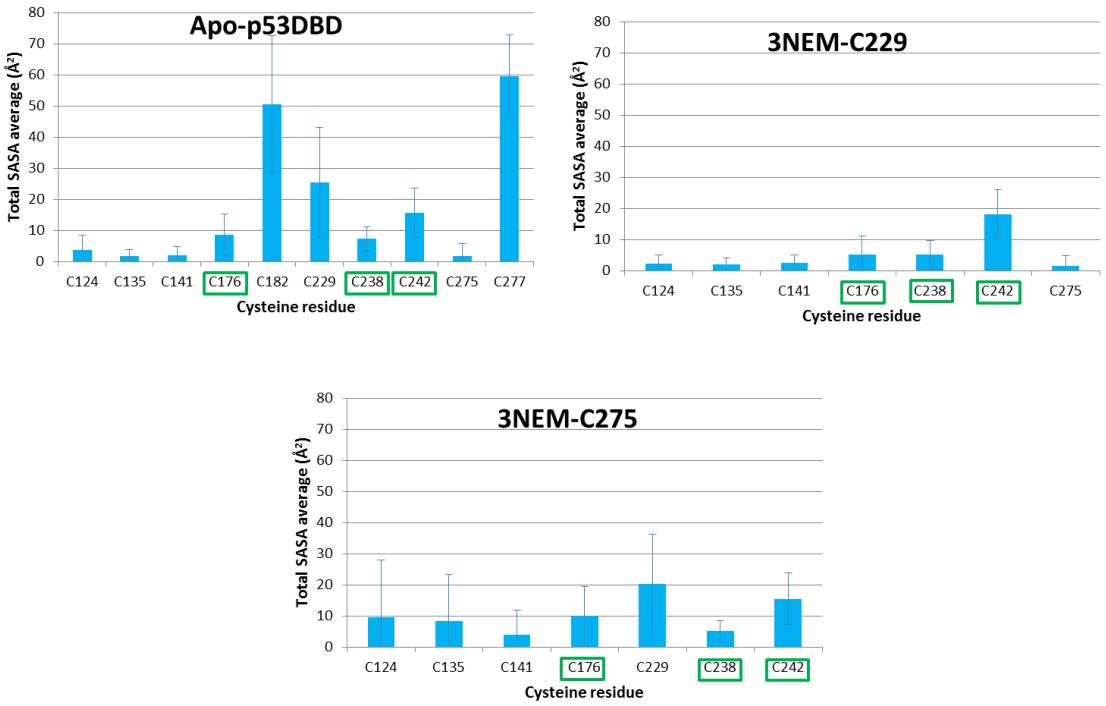


Figure 5-19: Average SASA of cysteines in the p53DBD apo-structure and the tri-alkylated systems as calculated from the aMD simulations.
The cysteines coordinated to the zinc are green boxed.

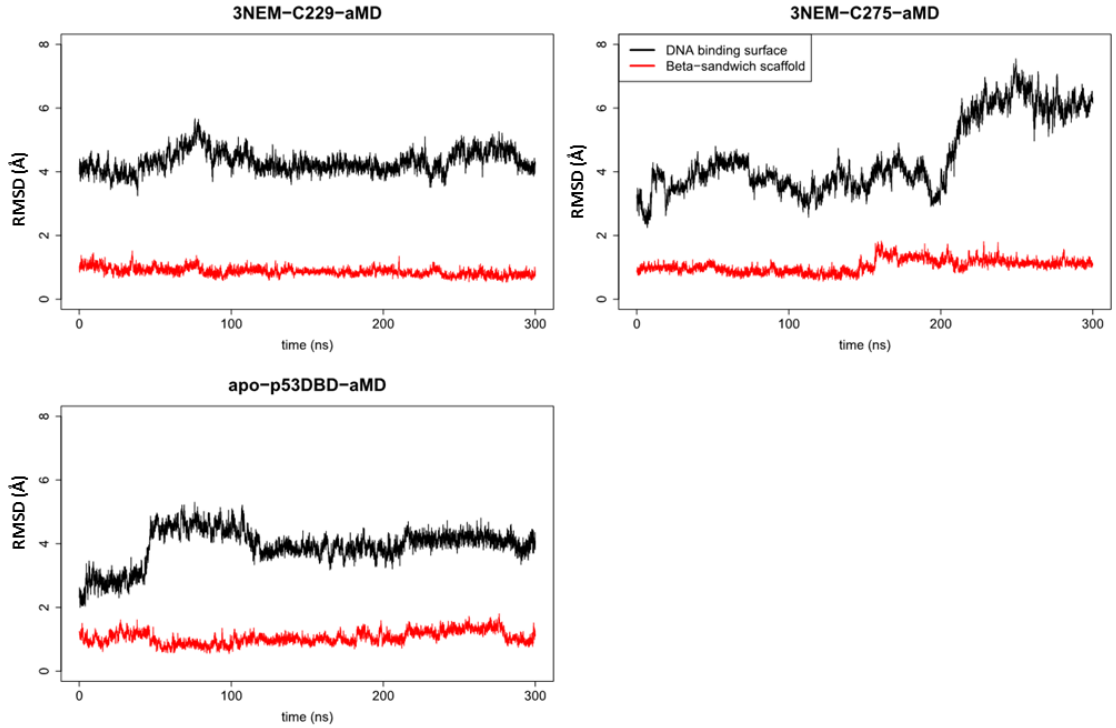


Figure 5-20: RMSD of the 3NEM and apo-p53DBD systems obtained from the 1st aMD simulations.
RMSD of the DNA binding surface (DBS, black) against the immunoglobulin-like β-sandwich (scaffold, red) was determined for each 1st simulation of the 3NEM and the apo-p53DBD systems after alignment on the reference structure DBS or scaffold.

Another mechanism that could explain why C275 is not exposed to the solvent after 1μs of 2NEM simulation is that its alkylation might occur after or concurrently with the alkylation of C229. This hypothesis could explain why the authors found two cysteines implicated in the unfolding of p53DBD and the remaining cysteines alkylated simultaneously. To investigate the possibility of the involvement of a fourth alkylation in p53DBD unfolding, we also proposed two four-alkylated systems. These were modelled from the 3NEM-C229/C275 frames at 100 ns of simulation, and were run for 1 μs each. RMSF analysis of the two 4NEM systems (4NEM-C229 and 4NEM-C275, see Fig. 5-21) showed fluctuations mainly in the C-ter region of 4NEM-C229. As in the 3NEM-C275 aMD simulation, 4NEM-C229 is characterized by high flexibility in its H2 helix. Alignment of the 4NEM-C229 simulation on the reference structure confirmed that in 4NEM-C229 the H2 helix is highly flexible (although only noticeable in repeat 1; see Fig. 5.21). The snapshots taken from the simulation of 4NEM-C229 (aligned on 2XWR) showed that the H2 helix is first destabilised (0-50 ns) and then the unstructured region becomes more flexible (100 ns-1 μs). While 3NEM-C275 also showed a similar flexibility in its C-ter region, this could not be ascribed to the same mechanism. In 3NEM-C275, the H2 helix is still conformationally stable but flipped away from its initial position (Fig. 5-18). On the other hand in 4NEM-C229 it is the destabilisation of the H2 helix that determines its flexibility (Fig. 5-22 and 5-23). Additionally, the fourth alkylation does not induce a higher solvent accessibility in the non-alkylated thiols (Fig. 5-24), unlike in the case of 3NEM-C275 (Fig. 5-7). Despite this, the scaffold of 4NEM-C229 still keeps an RMSD of 1 Å (same as in the reference structure; see Fig. 5-25). The fourth alkylation seems unable to induce the destabilisation of the p53DBD scaffold or enhance the solvent accessibility of the remaining thiols. Taken together, these results suggest that alkylation of C229 might be inconsequential in terms of stability. Indeed, its solvent accessibility above 20 Å both in the apo-p53DBD and 2NEM makes it the most likely target for the 3rd alkylation. However its alkylation (both in 3NEM and 4NEM) does not expose other thiols to the solvent, unlike the alkylation of C275 in 3NEM.

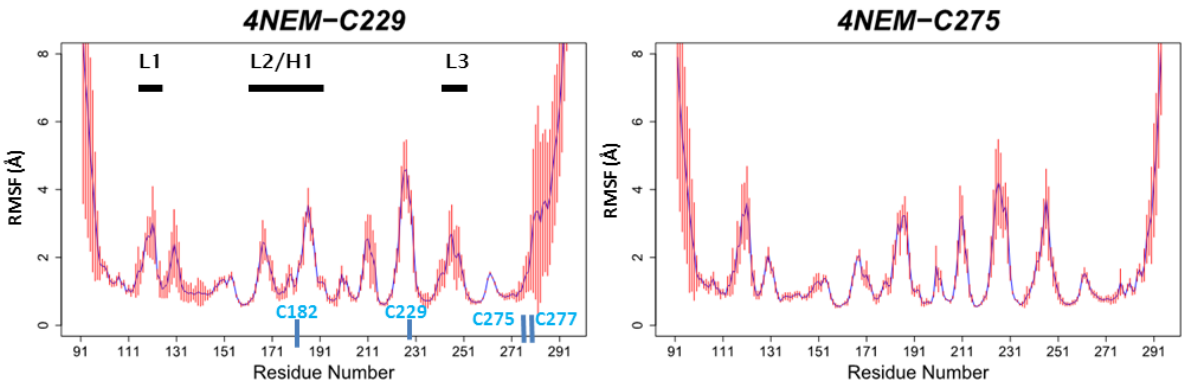


Figure 5-21: RMSF of the p53-4NEM systems.
The RMSF for each 1 μs simulation was determined using cpptraj in Amber. The average RMSF (blue) among the four cMD and the standard deviation (red) for each system were calculated using R. The alkylated cysteines are labelled in blue.

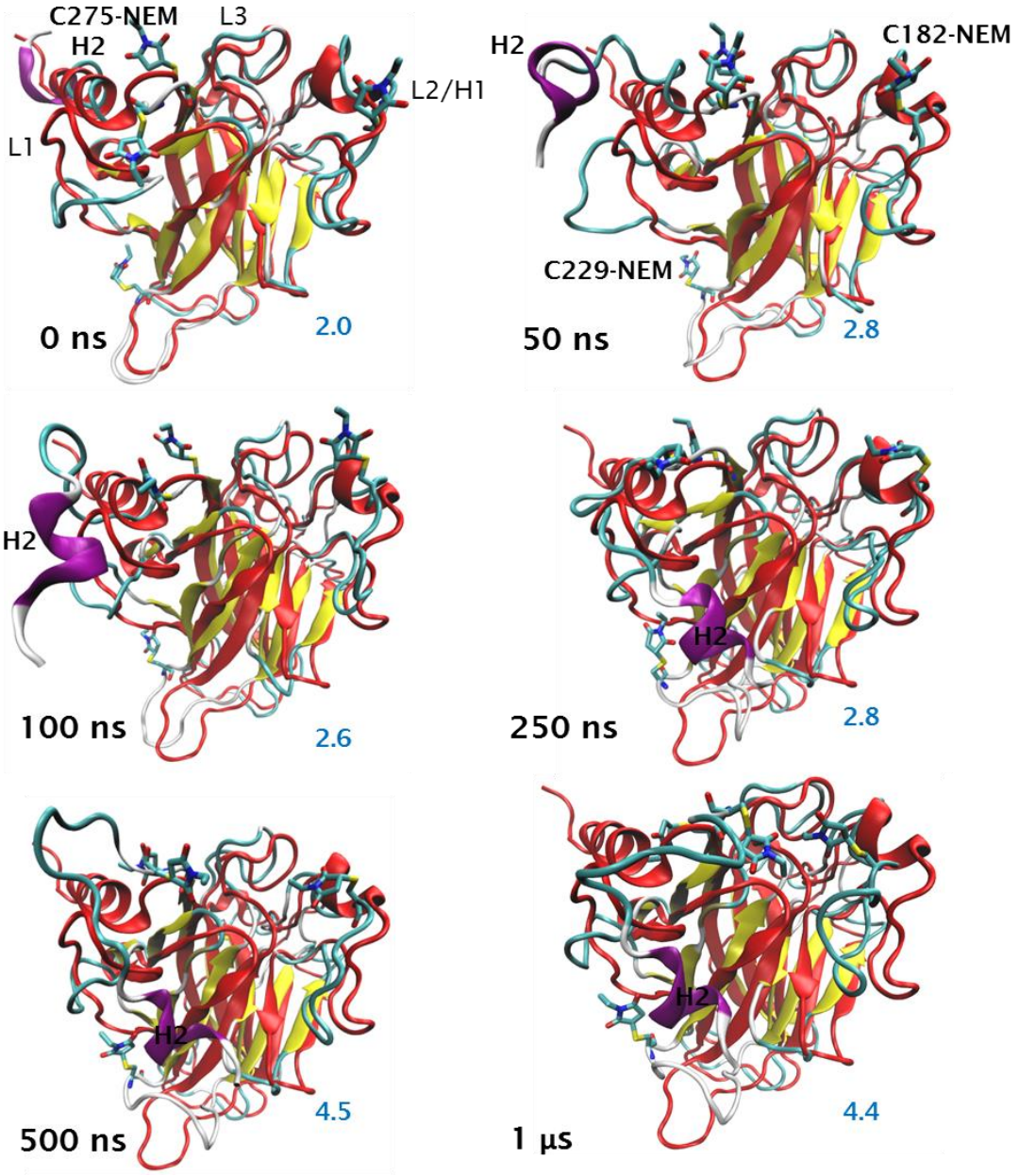


Figure 5-22: VMD visualisation of the 4NEM-C229 system, repeat 1.
The simulations were aligned on 2XWR chain A (reference structure) alpha-carbon at different time of the simulation (indicated in black). The structures are represented in cartoon, the cMD frames were coloured by secondary structure and the reference structure was coloured in red. The RMSD of the frame with respect to the reference structure is indicated in blue.

Figure 5-23: Secondary structures of H2 helix obtained from 3NEM-C275 aMD and 4NEM-C279 cMD simulations.

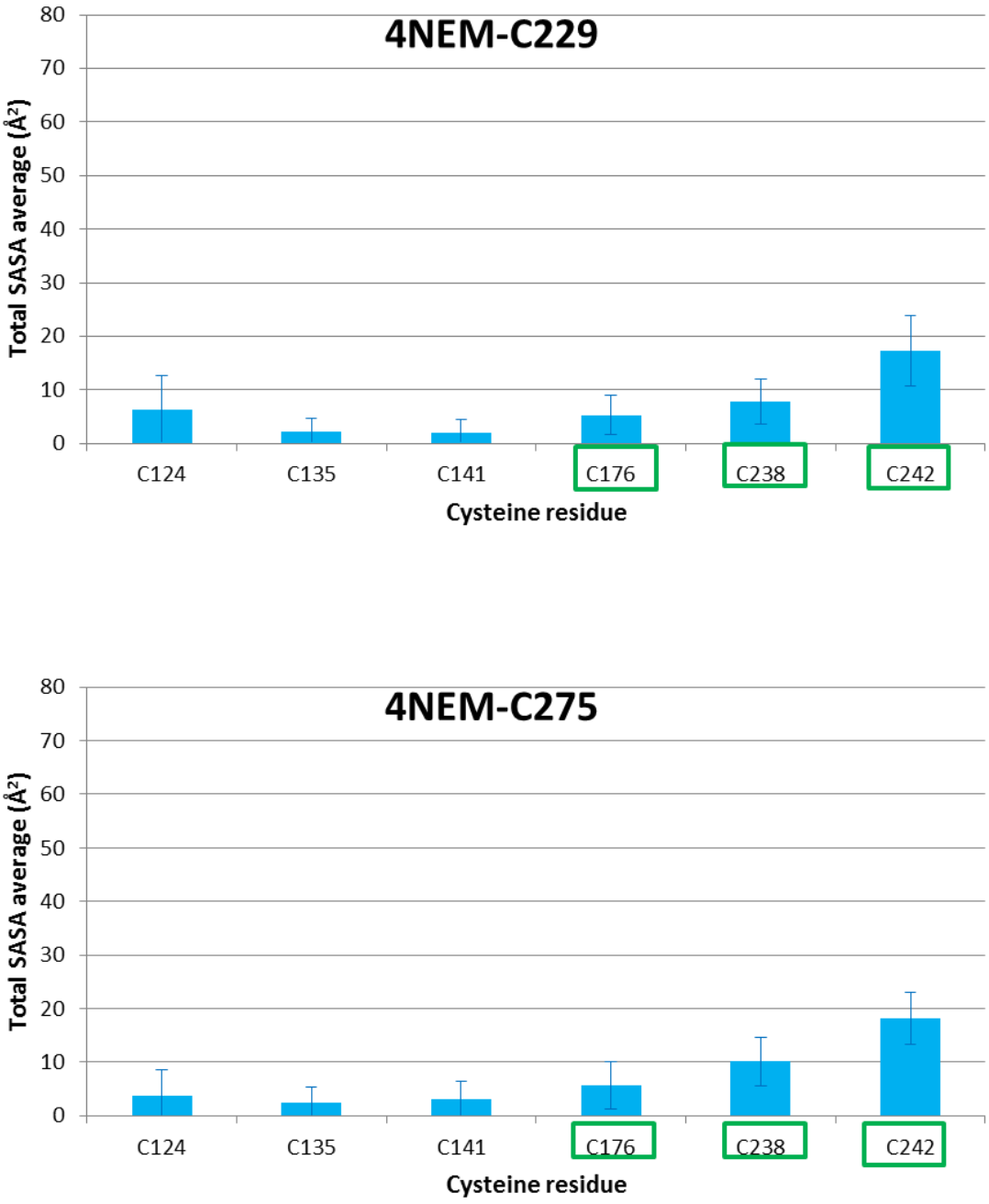


Figure 5-24: Average SASA among cysteines of p53DBD 4NEM systems.

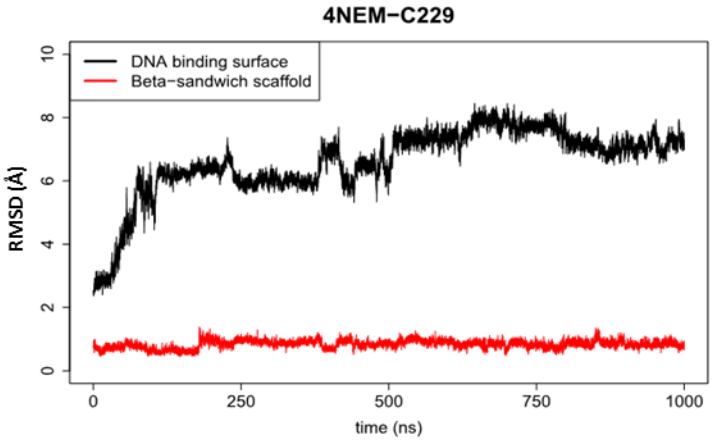


Figure 5-25: RMSD of the 4NEM-C229 system, repeat 1.
RMSD (in Å) of the DNA binding surface (DBS, black) against the immunoglobulin-like β-sandwich (scaffold, red) was calculated from the 4NEM-C229-RP1 simulation after alignment on the reference DBS or scaffold.Role of alkylation of C182 and C277

In section 5.3.1 we concluded that in our system the two cysteines most likely to be alkylated in the first step are C182 and C277, which is in agreement with the findings of Langridge-Smith *et al.* (109). The authors attributed the rescue of mutant p53 and the DNA binding enhancement of its wild type to this double alkylation. The involvement of these two cysteines was confirmed by our simulations. Since the particular mechanism responsible for the changes in p53DBD conformation is still a matter of debate in the literature, we analysed in more detail both the mono- (1NEM-C182 and 1NEM-C277) and the di-alkylated (2NEM) systems. This allowed us to propose a mechanism for how these alkylations influence p53 function.

Visualisation of the apo-p53DBD, 1NEM and 2NEM systems aligned on the reference structure (2XWR) showed stability in the structures through the whole 1 μs cMD simulation (Fig. 5-26). Most of the flexibility was observed in the L1 and L6 loops (see chapter 4), which was confirmed by RMSF analysis (Fig. 5-27). As for the previous models, the RMSF showed flexibility mainly in the DBS and the N-ter (unstructured) regions. Visualisation of the snapshots also highlighted a more structured H1 helix in the C182-alkylated systems (1NEM-C182 and 2NEM), while this helix was not stable in the apo-p53DBD and 1NEM-C277 structures. Analysis of the secondary structure of the L2/H1 region in the different systems and their repeats (Fig. 5-28) showed the H1 helix to be more conserved in the C182-alkylated systems, particularly in the 1st simulation

of 1NEM-C182 and 2NEM. Analysis of the 2NEM 1st simulation, in which the H1 helix is stabilised and extended (Fig. 5-29), showed that the NEM bound to C182 is pulled toward the cysteine cluster binding to the zinc. This makes H1 stable and even extended in the case of the 2NEM system. The H1 helix was shown to be implicated in p53 dimerization while binding to DNA (154). This structure forms the dimer interface by interacting with the H1 helix from the second monomer via hydrogen bonds, van der Waals and electrostatic interactions. Some mutations within this region (mainly those affecting H177, H178, E180 and R181) have been found in human cancers. In particular substitution of R181 was shown to inhibit DNA binding. In the wild-type, R181 from one p53DBD chain forms a salt-bridge with E180 from the other monomer, which contributes to the dimer stability (Fig. 5-30). To investigate the dimer stability, 7 dimers were modelled: 6 based on monomers extracted at 100, 500 and 1μs of the apo-p53DBD and 1NEM-C182 simulations (1st simulation, see Fig 5-31), and one modelled by combining frames from the 500 ns and 1 μs time points of the 1st 2NEM simulation (Fig 5-32). Comparison of the apo-p53DBD and the 1NEM-C182 dimers showed that the salt bridge between R181 and E180 is more likely to form in 1NEM-C182 than in the apo-p53DBD dimer, due to the higher stability of the H1 helix in the former. Since a stable and extended H1 helix is also present in the 2NEM system, we modelled the 2NEM dimer to further investigate its dimer interface. This 2NEM dimer (Fig. 5-32) was found to adopt a conformation allowing two salt bridges, instead of only one (as observed in the experimental structure and in our apo-p53DBD and 1NEM-C182 dimers). The presence of an additional salt bridge confers better stability to the dimer, which may account for the higher binding affinity to DNA. This partly explains the enhancement of DNA binding following exposure of the protein to low concentrations of MA, as these are sufficient for the alkylation of C182.

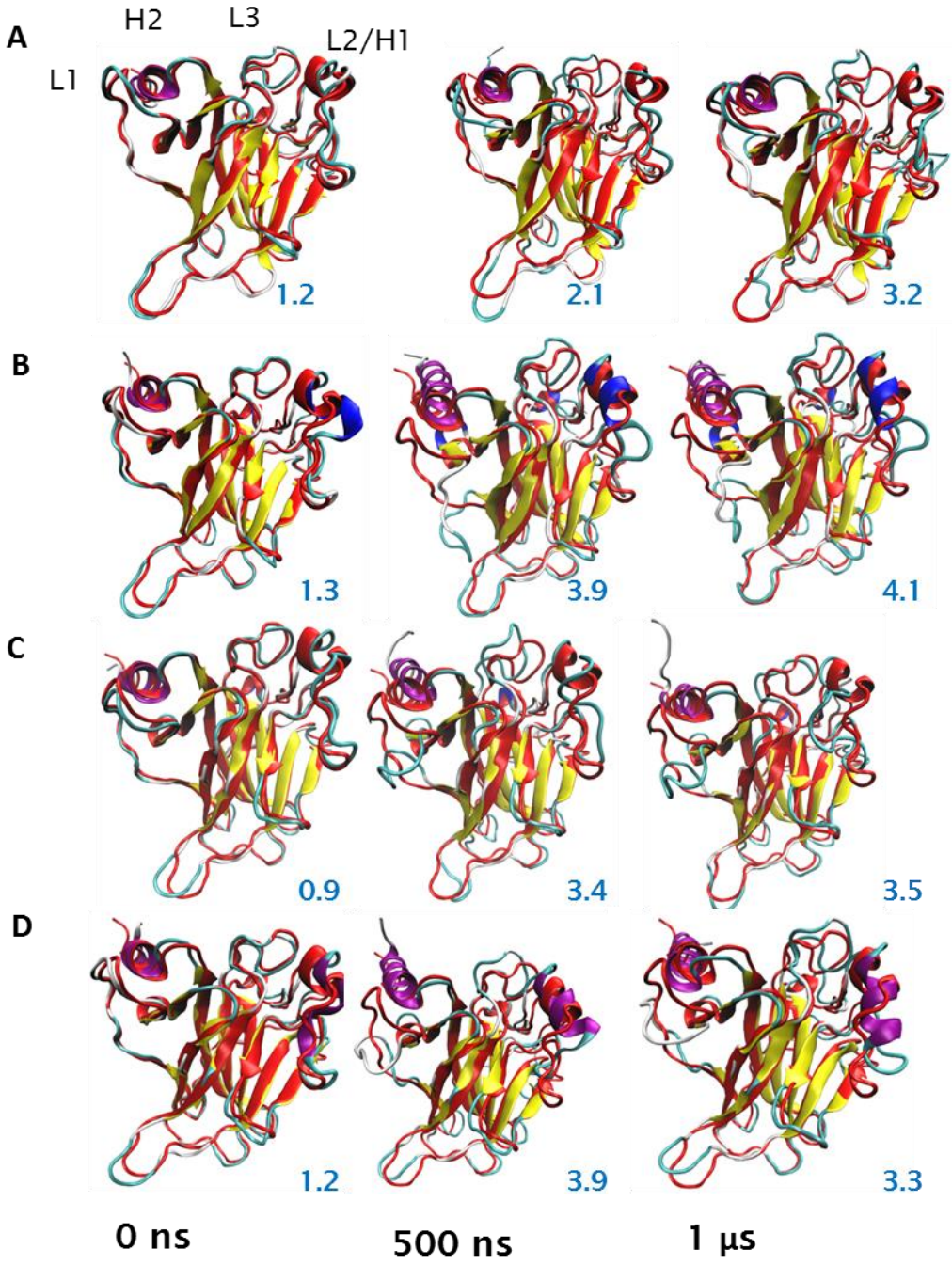


Figure 5-26: VMD visualisation of the apo-p53DBD, 1NEM-C182, 1NEM-C277 and 2NEM systems.
The systems were aligned on 2XWR chain A (reference structure) alpha-carbon at different time of the simulations. Here is shown only the first simulation. **A)** apo-p53DBD, **B)** 1NEM-C182, **C)** 1NEM-C277 and **D)** 2NEM. The structures are represented in cartoon, the simulation frames were coloured by structure and the reference structure was coloured in red. The RMSD (in Å) of the frame with respect to the reference structure is indicated in blue.

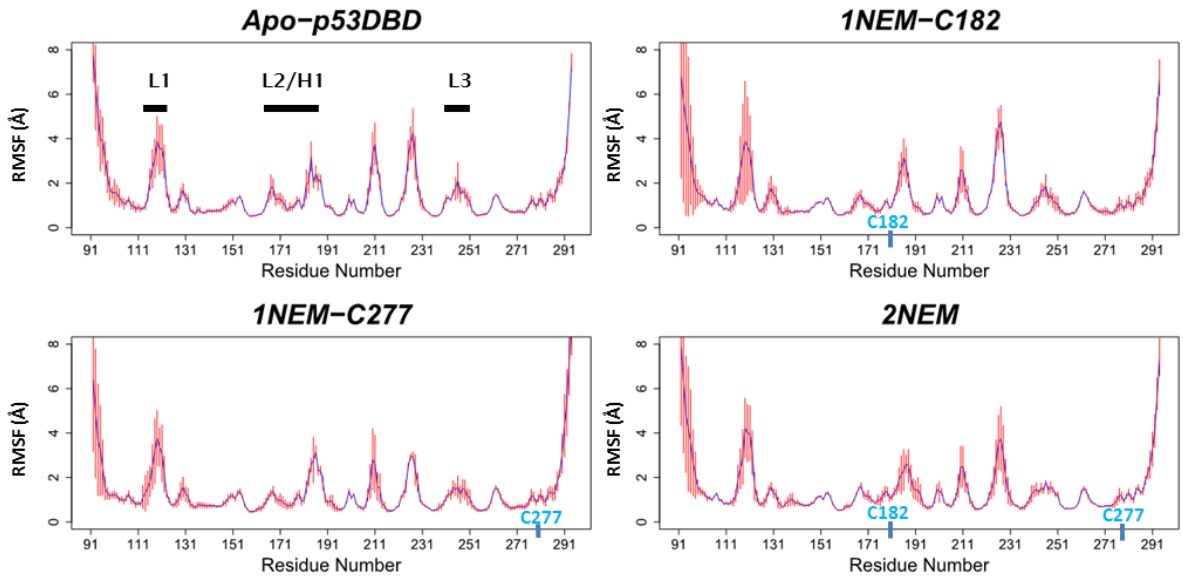


Figure 5-27: RMSF of the apo-p53DBD, 1NEM and 2NEM systems.
The RMSF for each 1μs simulation was determined using cpptraj in Amber package. The average RMSF (blue) among the four cMD and the standard deviation (red) for each system were calculated using R.

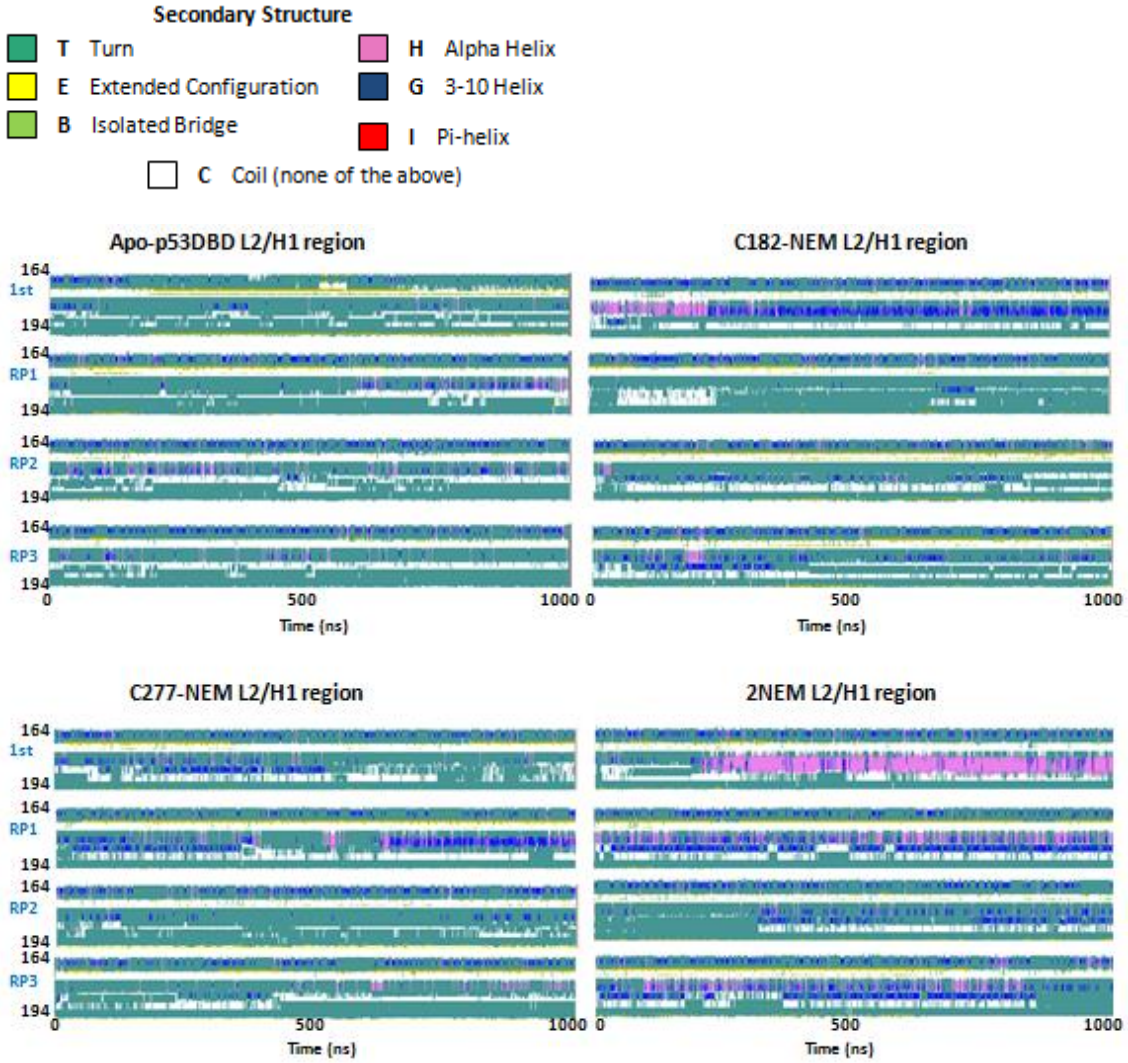


Figure 5-28: Determination of the L2/H1 region secondary structure in apo-p53DBD, 1NEM and 2NEM simulations.

The secondary structure analysis was performed using timeline analysis in VMD and only the L2/H1 region (from residue 164 to 194) is shown.

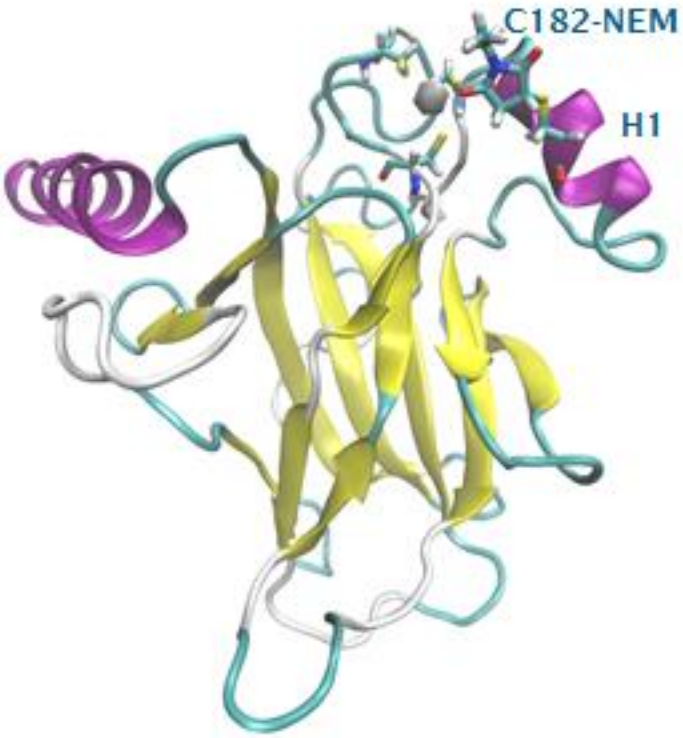


Figure 5-29: Alkylated C182 shifting towards the cluster of zinc-bound cysteines in the 1NEM-C182 simulation.
The structure is represented in new cartoon and coloured by secondary structure. The C182 alkylated cysteine and the cluster of cysteines binding to the zinc are shown in licorice and coloured by name.

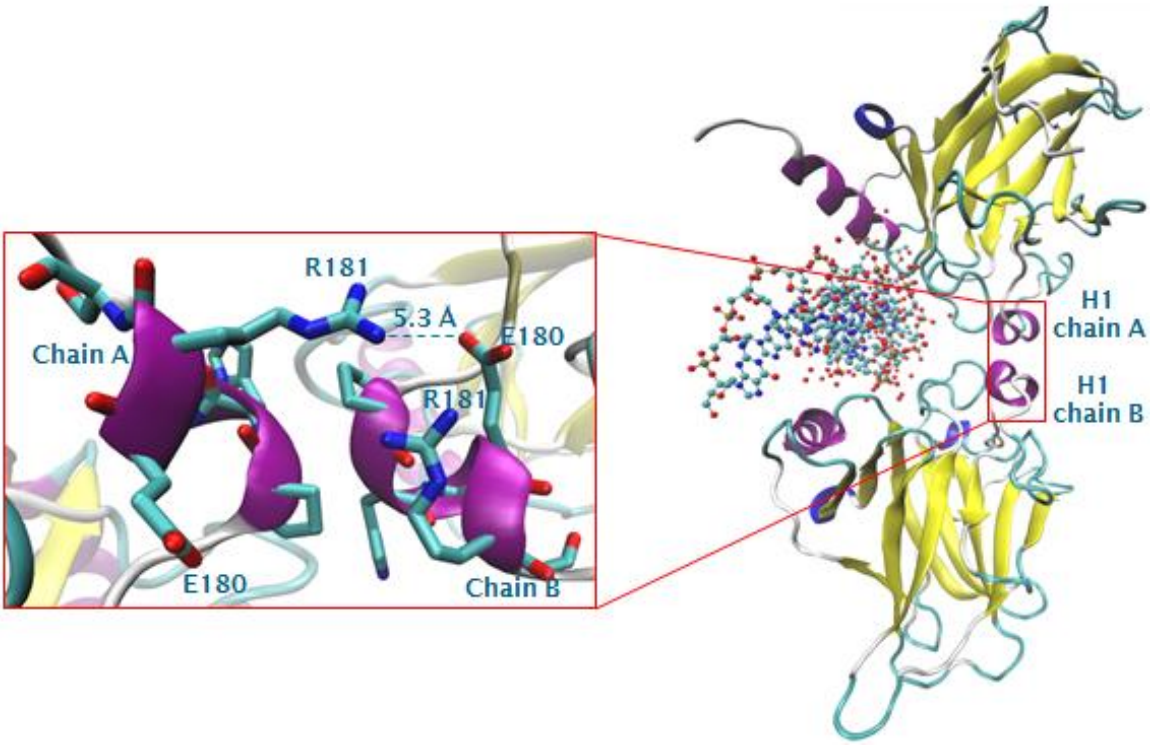


Figure 5-30: VMD visualisation of the p53 dimer interface in a crystal structure (pdb code: 3TS8).
The structure is represented in cartoon and coloured by structure, the DNA is shown by bond and coloured by name, the residues within H1 helix are represented in licorice and coloured by name. The salt bridge is represented by a dotted line.

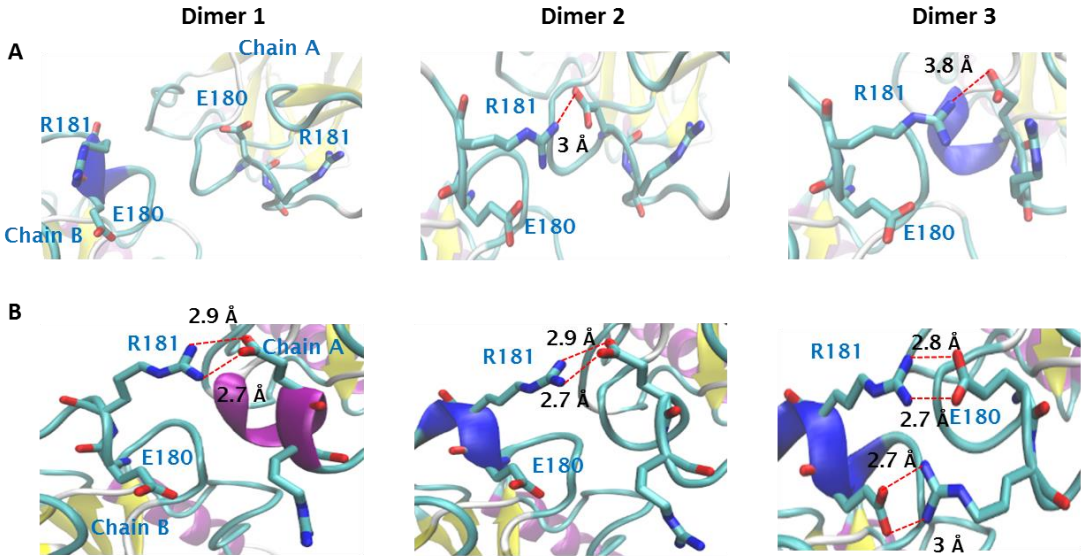


Figure 5-31: VMD visualisation of the apo-p53DBD and 1NEM-C182 dimer interfaces.
A) Apo-p53DBD **B)** 1NEM-C182. The structures are represented in new cartoon and coloured by structure, the residues responsible for the salt bridges are shown in licorice and coloured by name. The salt bridges are indicated by a dotted line.

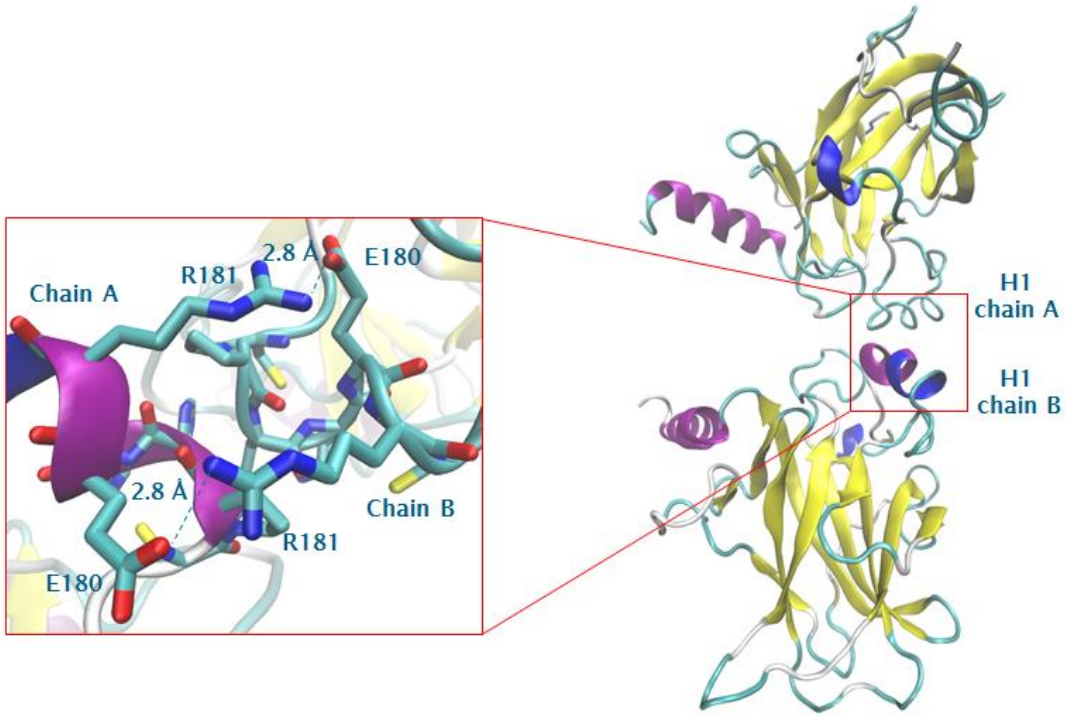


Figure 5-32: VMD visualisation of the p53 dimer interface in a 2NEM system. The structure is represented in new cartoon and coloured by structure, the DNA is shown by bond and coloured by name, the residues within H1 helix are represented in licorice and coloured by name. The salt bridges are represented by a dotted line.

Since C277 was also shown by Langridge-Smith *et al.* to be susceptible to alkylation by NEM, and it was reported to interact with some promoters by other authors (132) (Fig. 5-33), we modelled 1NEM-C277 in order to assess the influence of the alkylation of this residue on p53DBD conformation. Our analysis of the 1NEM-C277 simulations showed that alkylation of C277 indeed could result in the inhibition of DNA binding, and this effect could be attributed to conformational changes induced in the L1 loop by the alkylation.

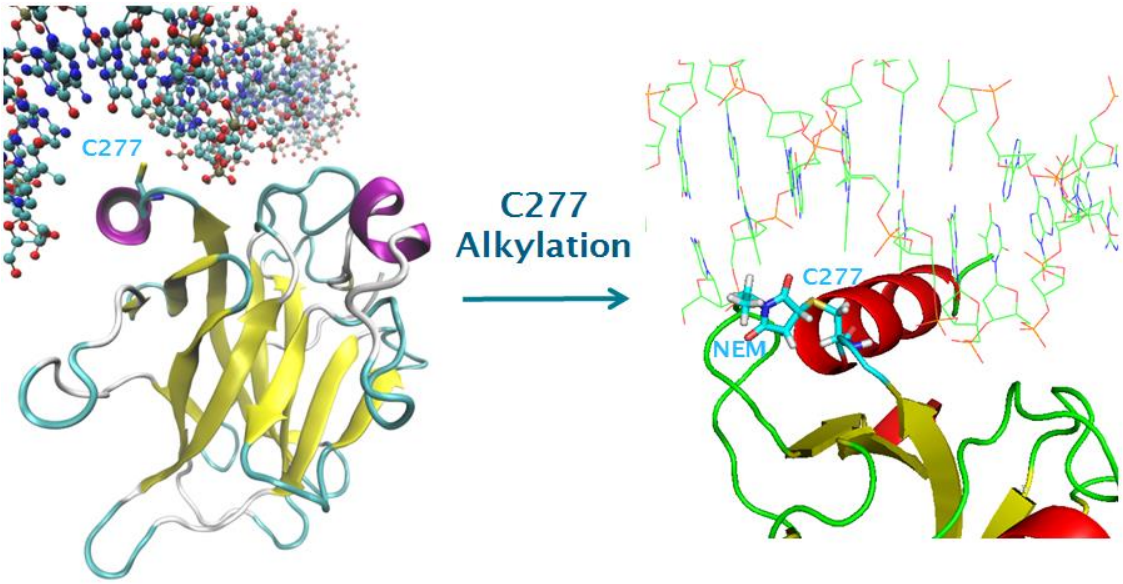


Figure 5-33: C277 binding directly to DNA of some promoters.
The right structure is 3TS8 (pdb code) and the left structure is the 1NEM-C277 model.

Analysis of the secondary structure of the L1 loop in our apo-p53DBD, 1NEM (1NEM-C182 and 1NEM-C277) and 2NEM systems showed that this loop adopts in 1NEM-C277 a conformation different than in the other systems (Fig. 5-34). In the 1NEM-C277 system the L1 loop was found to rigidify and adopt either an α-helix or a β-strand configuration (Fig. 5-35), while in the other systems this loop is still flexible and represented as a turn. Conformational changes of the L1 loop were also observed in the 3NEM-C275 cMD (Fig. 5-36) and 3NEM-C275 aMD simulations (Fig. 5-37), but not in the 3NEM-C229 system (cMD and aMD). Since C275 and C277 are both located within the turn which links the H2 helix and the S10 sheet, the similar effects of their alkylation on the L1 loop are not surprising. This result reinforces the hypothesis that C275 might be the target for the third alkylation and be responsible for p53DBD unfolding. The L1 loop of the apo-p53DBD in the aMD repeat showed that some residues form an α-

helix. Unlike in the case of 1NEM-C277 and 3NEM-C275, the changes observed in L1 of the apo-structure (aMD) occur only in one of the repeat simulations. Such changes are not persistent along the simulation and engage different residues. In this case the changes observed in the L1 loop were not considered relevant to this model.

As highlighted in chapter 4, flexibility of the L1 loop is important for p53 to bind specifically to DNA, thus its conformational change into a more rigid structure (α-helix or β-strand) in the 1NEM-C277 system might inhibit DNA binding or at least restrict the kinds of promoter sequences to which it will bind (155). The L1 loop contains the lysine residue K120 which, like C277, is also in direct contact with the DNA. Visualisation of the L1 conformational changes in the 1NEM-C277 structure (Fig. 5-38) showed that, in the system in which L1 forms an α-helix, K120 is too far from the DBS and it is therefore unable to interact with DNA. According to our results, alkylation of C277 by NEM might inhibit p53 DNA binding by inhibiting direct contacts of C277 with DNA and by rigidifying the L1 loop.

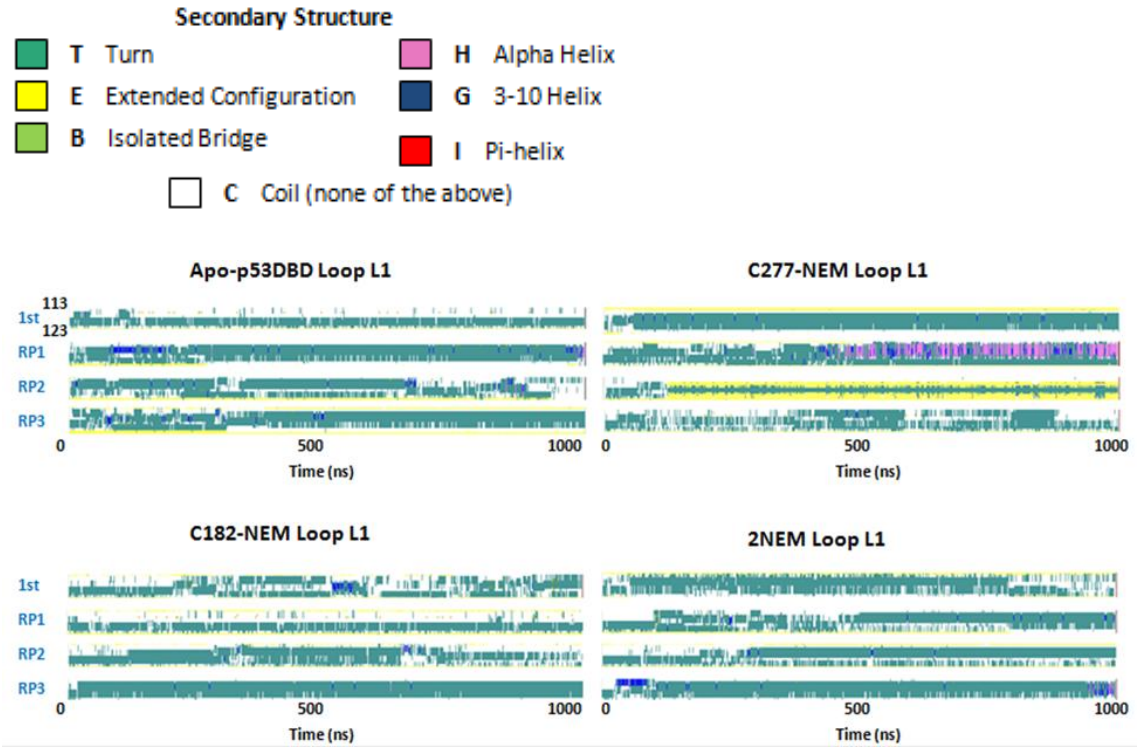


Figure 5-34: Determination of secondary structure of L1 loop from the apo-p53DBD, 1NEM and 2NEM systems.
The secondary structure analysis was performed using timeline analysis in VMD and only the L1 loop (from residue 113 to 123) is shown.

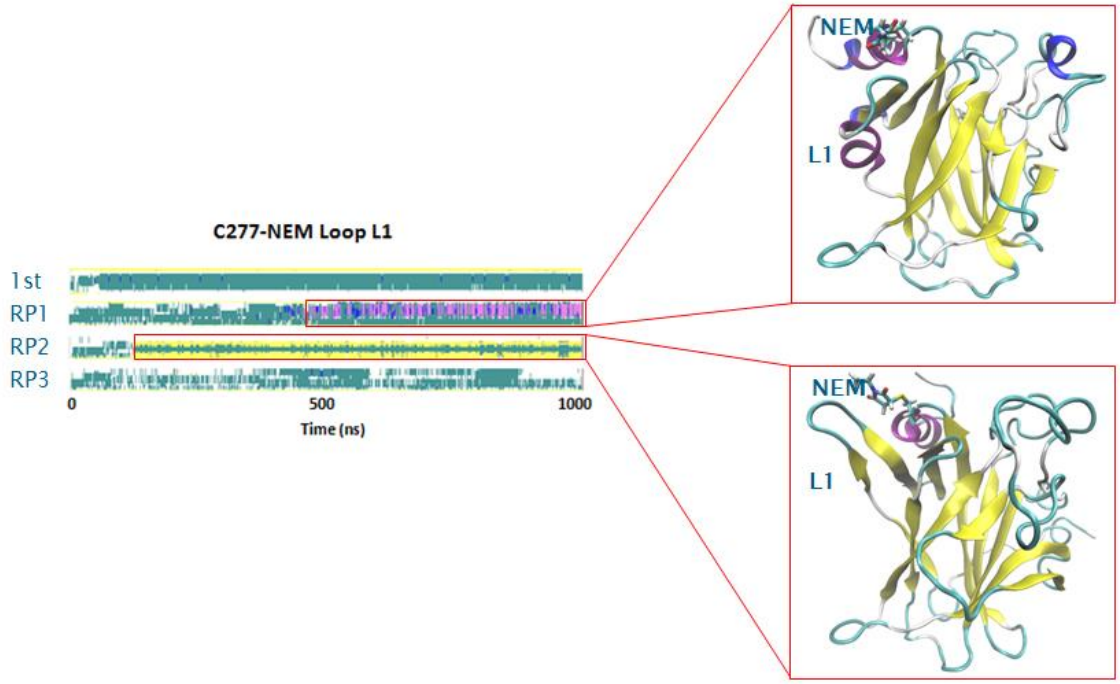


Figure 5-35: Conformational changes in the L1 loop of 1NEM-C277.

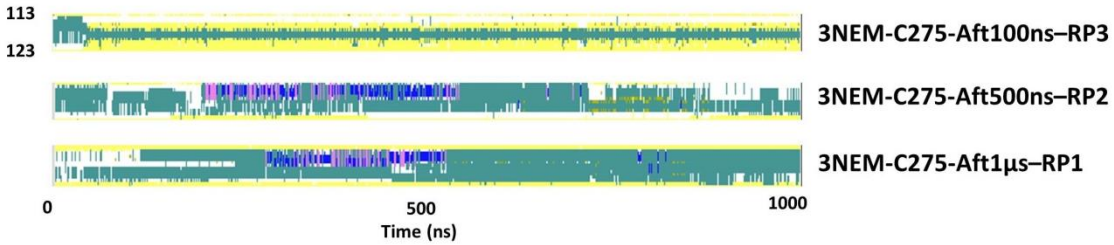


Figure 5-36: Determination of the secondary structure of L1 loop from the 3NEM-C275 system.

The secondary structure analysis was performed using timeline analysis in VMD and only the L1 loop (from residue 113 to 123) is shown.

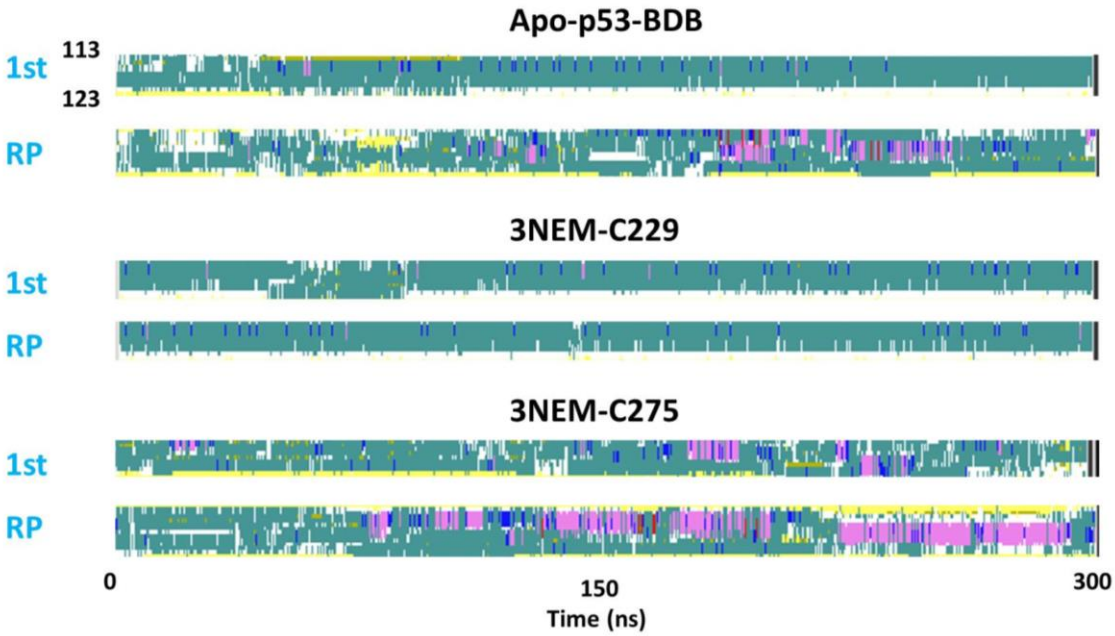


Figure 5-37: Determination of the secondary structure of L1 loop from apo-p53DBD and 3NEM aMD simulations.
The secondary structure analysis was performed using timeline analysis in VMD and only the L1 loop (from residue 113 to 123) is shown.

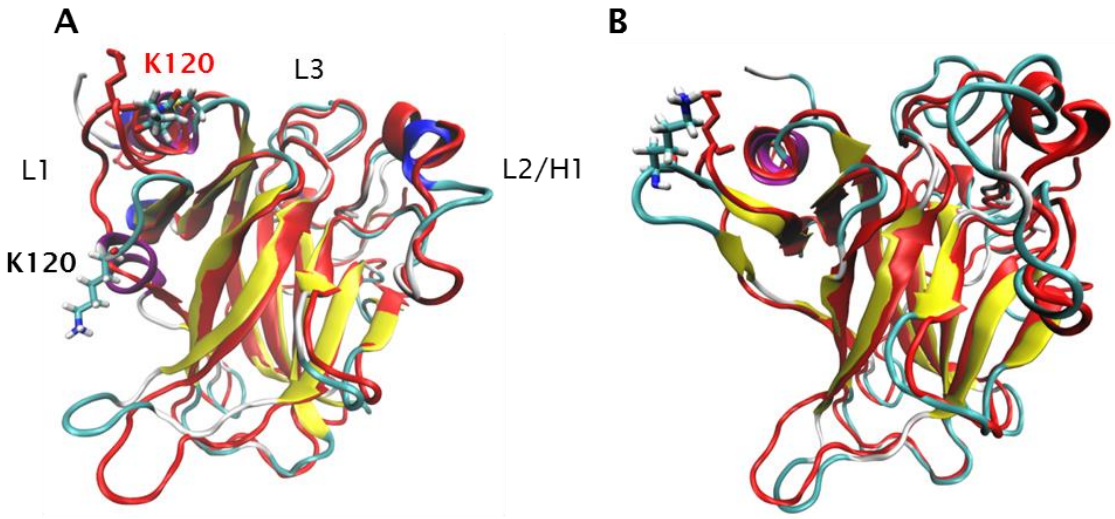


Figure 5-38: Position of 1NEM-C277 residue K120 in the rigidified L1 loop. The last frames of 1NEM-C277 **A)** repeat 1 and **B)** repeat 2 were aligned on the reference structure and visualised via VMD. The structures are presented in new cartoon. The 1NEM-C277 structures are coloured by structure and the reference is coloured in red.

In contrast to the suggestions of Langridge-Smith *et al.* (109), our results suggest that NEM alkylation of C182 might alone be responsible for enhancing DNA binding in p53, while alkylation of C277 is more likely to inhibit DNA binding. Indeed, in both the experimental study and our simulations, only the DBD of p53 is considered. In a full p53 structure, C277, located in the C-ter of the DBD, might be more buried due to its close proximity with the tetramerisation domain (Fig 5-39). In this case only C182 can be alkylated at low concentration of Michael acceptors and be responsible for the enhanced DNA binding in wild-type p53, and the rescue of the mutants. Alkylation of C277 might instead only occur at higher concentrations of Michael acceptors, which induce inhibition of DNA-binding in p53 wild type.

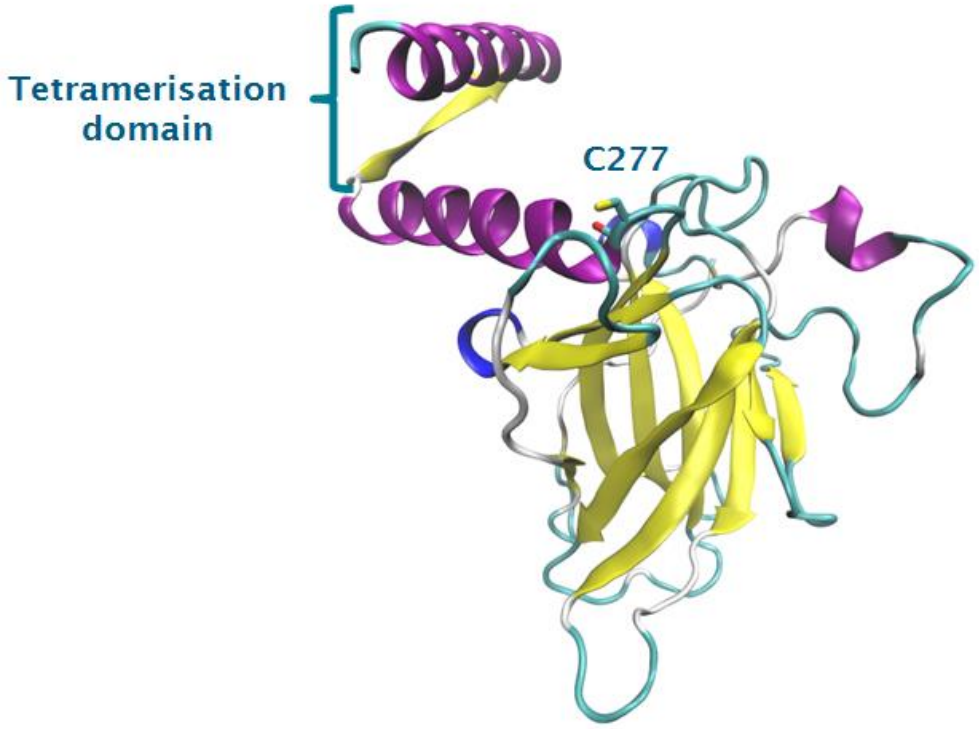


Figure 5-39: VMD visualisation of the C277 residue in the engineered p53 structure (pdb code: 3TS8).

On the other hand, the 2NEM systems showed changes (stable H1 α-helix) only due to the C182 alkylation (see Fig. 5-28). Indeed in these systems the L1 loop maintains a turn like conformation throughout the simulations, similarly to the case of the apo-structure.

To confirm our hypothesis that the alkylation of C277 is the one inhibiting DNA binding, we decided to prolong the 2NEM simulations using aMD. Two starting frames were chosen for this, the last frame of the 1st 2NEM simulation in which the H1 helix is stabilised and extended (Fig. 5-28 and 5-40), and the last frame of the 2NEM simulation (RP3) in which no H1 is formed but an α-helix starts forming in the L1 loop (Fig. 5-34 and 40). 200 ns aMD and one repeat simulation were run from each frame. The 2NEM aMD run from the 1st 2NEM simulation has been called 2NEM-aMD1 and its repeat 2NEM-aMD1-RP; the aMD run from RP3 has been named 2NEM-aMD2 and its repeat 2NEM-aMD2-RP.

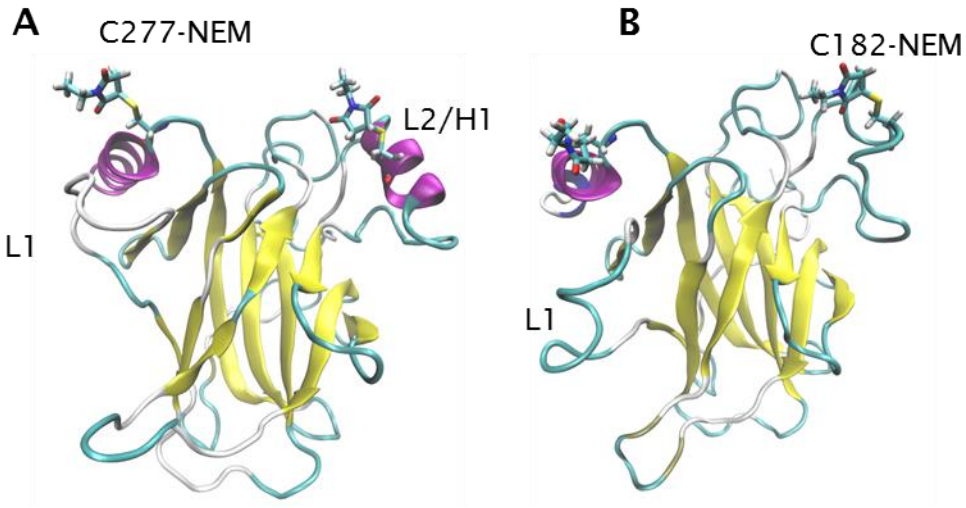


Figure 5-40: VMD visualisation of the frames from which extended 2NEM aMD simulations were started.
A) Last frame of the 1st 2NEM simulation, **B)** last frame of the 2NEM simulation repeat 3 (RP3). The structures are represented in new cartoon and coloured by structure, the alkylated cysteines (C182 and C277) are shown in licorice and coloured by name.

VMD visualisation of the 2NEM-aMD simulations aligned on the reference structure shows conservation of the p53DBD overall structure (Fig. 5-41). RMSD of the scaffold (≈ 1Å) confirmed stability in the system (Fig. 5-42), with fluctuations located mainly in the DBS (Fig. 5-42 and 5-43). The RMSF analysis highlighted fluctuations in the N-ter region (residues 91-109), as seen in the 3NEM and apo-p53DBD aMD simulations. This region, typically unstructured, tends to adopt an α-helix or a β-strand conformation in some aMD simulations, as can be seen in the secondary structure analysis (Fig. 5-44). The occurrence of more conformational changes in the apo-p53DBD (1st) than in the alkylated systems allowed us to conclude that these changes are not due to the cysteine

alkylation but are characteristic of the aMD technique applied to unstructured protein sequences.

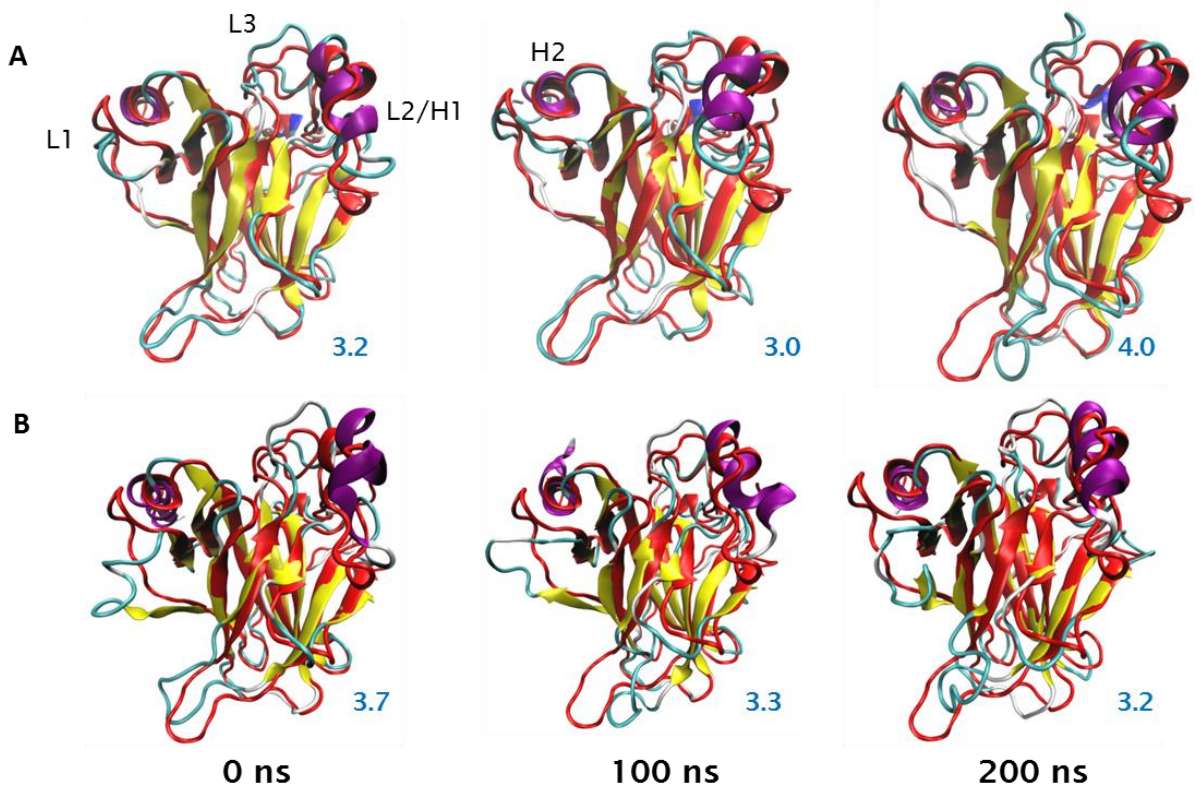


Figure 5-41: VMD visualisation of the 2NEM systems from the aMD simulation. The systems were aligned on 2XWR chain A (reference structure) alpha-carbon at different time of the simulations. Here is shown only the first simulation. **A)** 2NEM-aMD1 **B)** 2NEM-aMD2. The structures are represented in new-cartoon, the simulation frames are coloured by secondary structure and the reference structure is coloured in red.

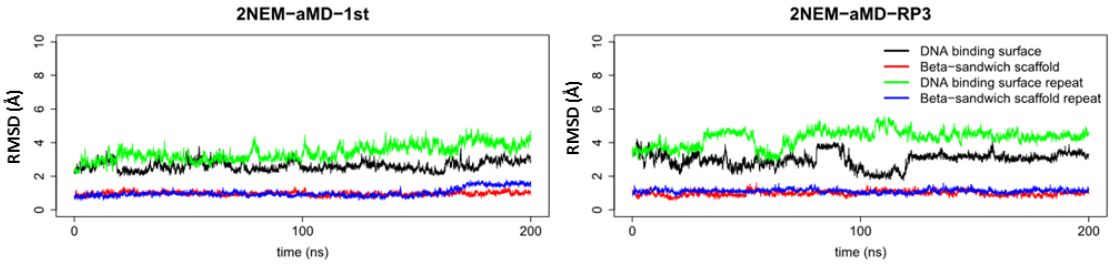


Figure 5-42: RMSD (in Å) of the 2NEM aMD simulation.
RMSD of the DNA binding surface (DBS, black) against the immunoglobulin-like β-sandwich (scaffold, red) was calculated from the 2NEM aMD simulations after alignment on the reference DBS or scaffold.

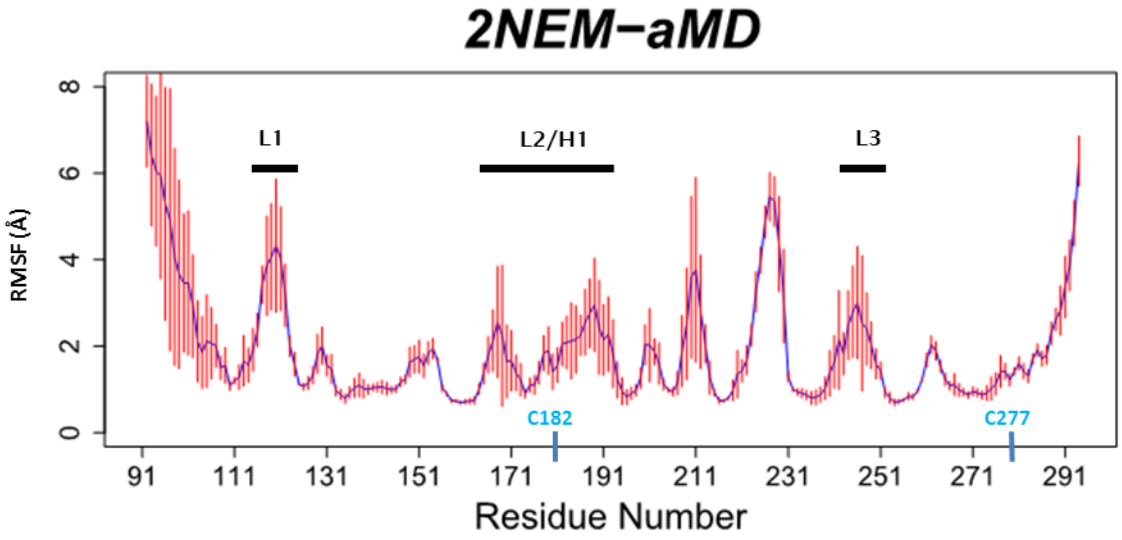


Figure 5-43: RMSF of the 2NEM aMD simulations.
The RMSF for each 200 ns simulation was determined using cpptraj in Amber package, the average RMSF (blue) among the four cMD and the standard deviation (red) for each system were calculated using R.

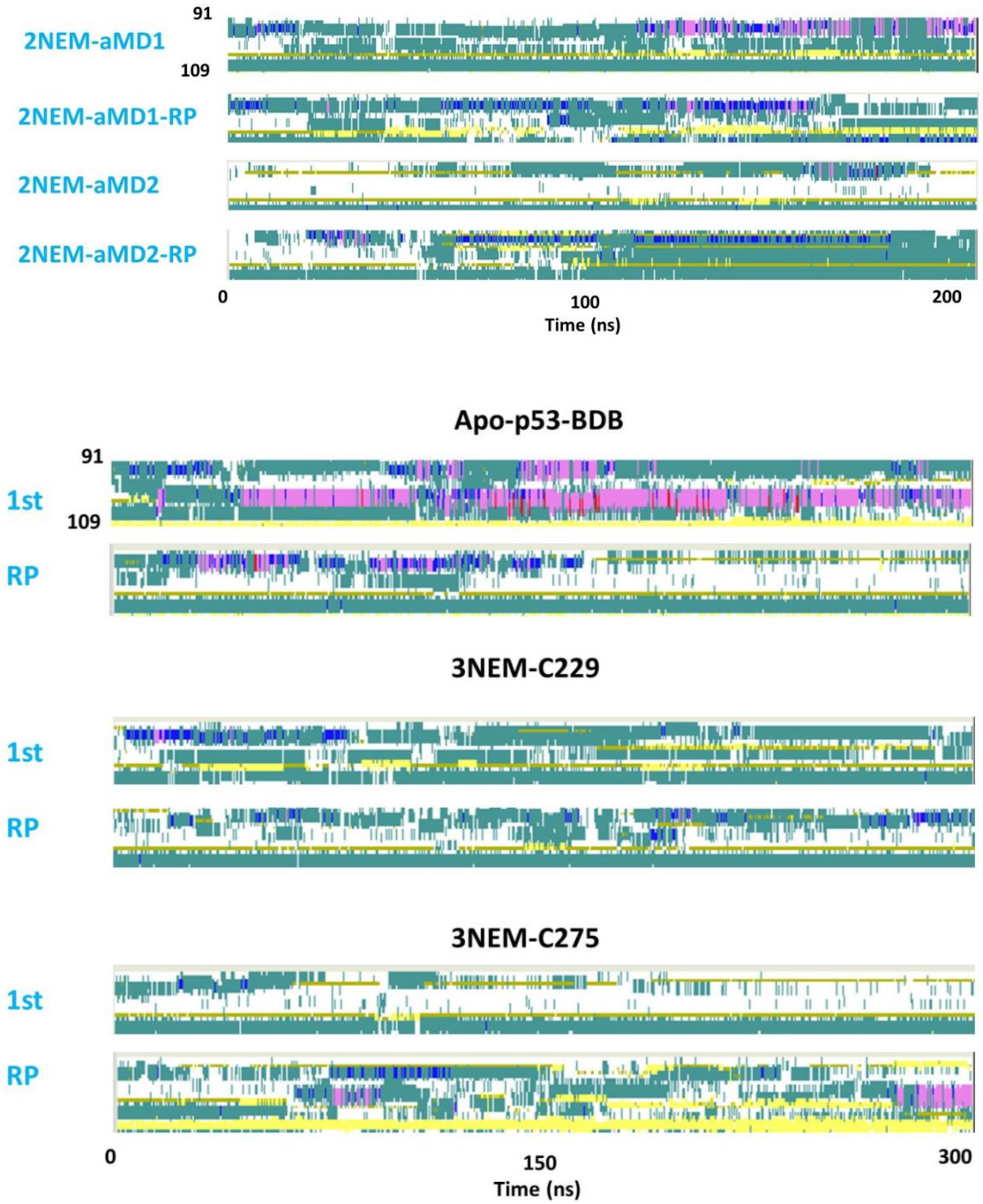


Figure 5-44: Determination of N-ter region secondary structure from aMD simulations.
Analysis of the secondary structure was performed using timeline analysis in VMD and only the N-ter region (from residue 91 to 109) is shown.

To examine the influence of C182 and C277 on the 2NEM-aMD systems, analysis of the secondary structure of the L2/H1 region and the L1 loop of the 2NEM-aMD systems was performed. The 2NEM-aMD simulation based on the last frame of the 2NEM-1st showed conservation of the H1 loop in the L2/H1 region (2NEM-aMD1 and 2NEM-aMD1-RP; see Fig. 4-45 A), while aMD based on

the 2NEM-RP3 frame resulted in a structured H1 α-helix in only one of the simulations (2NEM-aMD2). In 2NEM-aMD2 the H1 α-helix forms after 50 ns and is conserved throughout the simulation, while in 2NEM-aMD2-RP the α-helix structure appears within the first 50 ns of simulation but is not conserved.

In the apo-p53DBD-aMD simulations an α-helix like structure was also observed (Fig. 5-46). This structure was however not stable over the 1st simulation (apo-p53-aMD-1st) or engaged only some residues (apo-p53-aMD-RP), unlike the C182 alkylated systems where an extended H1 α-helix was observed. As explained in chapter 4, the H1 helix is present in the majority of the experimental structures (monomers, dimers and tetramers). This helix was found to be relatively unstable in our apo-p53DBD simulations (not present at all in cMD, see Fig. 5-28; unstable in aMD), while in the 1NEM-C182 and 2NEM cMD simulations it was often found to be stabilised and extended by the alkylation of C182.

Alkylation of C277 in the 2NEM-aMD systems was found to affect the conformation of the L1 loop (Fig. 5-45 B) and, as seen in 1NEM-C277, this may be responsible for the inhibition of DNA binding. In the 2NEM system the alkylation induced preferentially the formation of an α-helix, which does not only rigidify the L1 loop but also prevents K120 from binding to DNA. Under these conditions p53DBD cannot bind specifically to DNA. Hence it appears that the di-alkylation cannot be the reason for the enhancement of DNA binding observed in wild-type p53 exposed to low concentrations of MAs, but might instead be responsible for the inhibition of DNA binding observed in p53DBD WT incubated with high concentration of MAs.

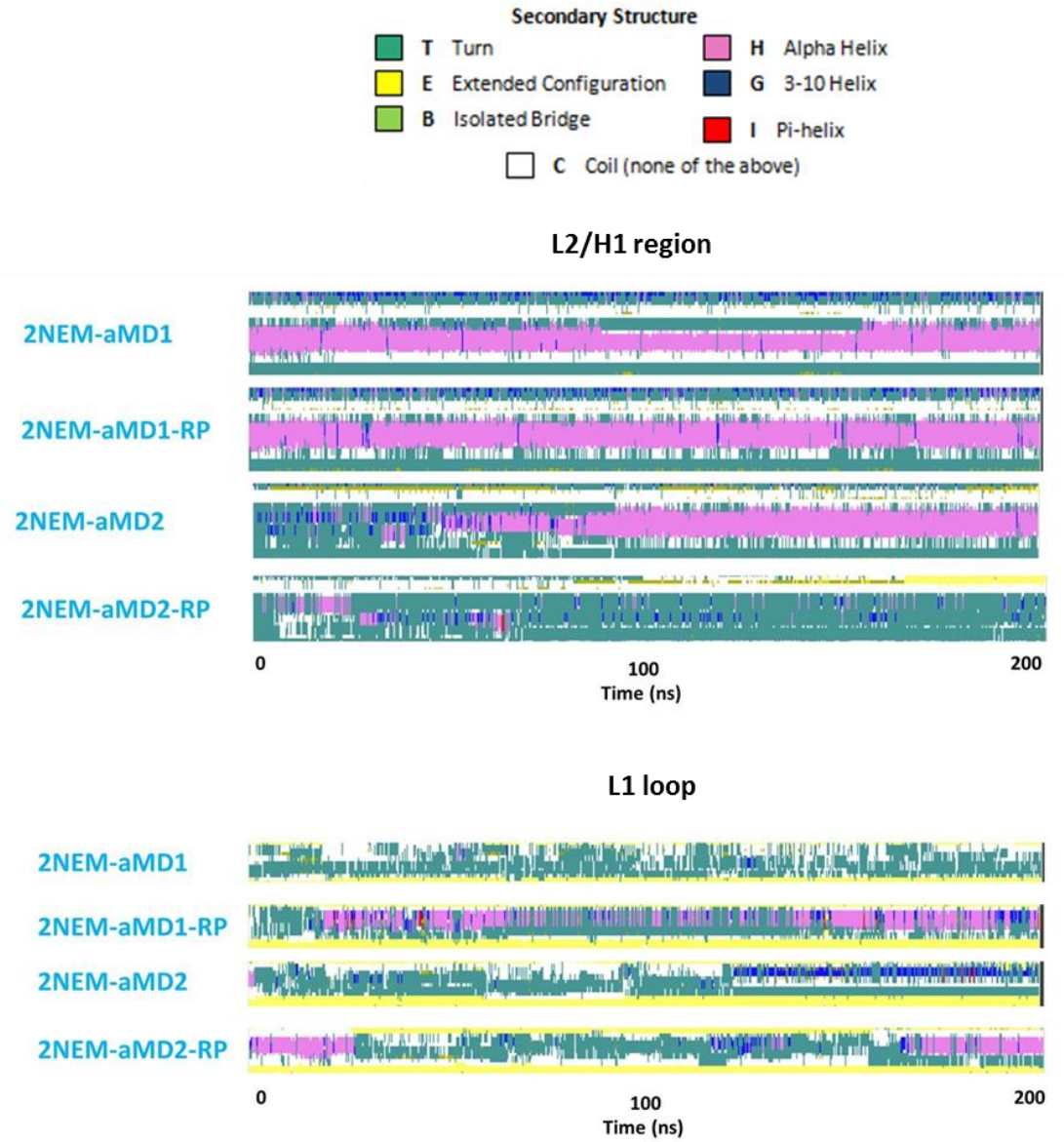


Figure 5-45: Determination of secondary structure of the L2/H1 region and L1 loop from 2NEM aMD simulations.
Analysis of the secondary structure was performed using timeline analysis in VMD.

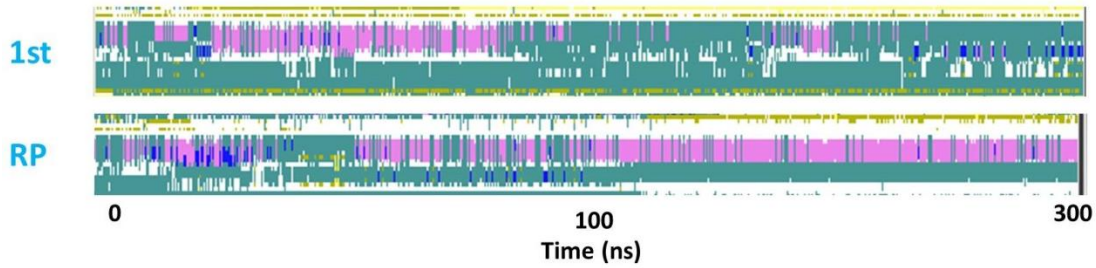


Figure 5-46: Determination of secondary structure of the L2/H1 region from apo-p53DBD aMD simulations.
Analysis of the secondary structure was performed using timeline analysis in VMD.

5.4 Conclusions

Michael acceptors are promising molecules for the stabilisation of mutant and unfolded wild type p53. Several authors showed that their effect is concentration dependent (enhancing p53 DNA binding at low concentration and inhibiting it at high concentration). In particular Langridge-Smith *et al.* were able to suggest two cysteines (C182 and C277) as likely targets for the alkylation by MAs at low concentration. To investigate this, we performed SASA analysis on our apo-p53DBD and 1NEM simulations which confirmed the solvent accessibility of these two cysteines in our systems and *per se* their reactivity.

Langridge-Smith *et al.* explained the high MA concentration effect on p53 by introducing a third alkylation responsible for p53DBD unfolding and alkylation of the remaining buried cysteines. In our simulations, C229 and C275 were alkylated after alkylating C182 and C277. Although our 2NEM model showed that C229 was more accessible to the solvent that C275 (buried), the 3NEM system simulations showed that alkylation of C275 is likely responsible for the solvent accessibility of the remaining p53DBD cysteines.

C275 is located two residues away from C277 (reactive cysteine), upon alkylation, rigidified the DBS (cMD) prior to destabilising it. This proximity of C275 (site of a possible destabilising alkylation) with one of the 'reactive cysteine' and the knowledge that C277 can bind directly to some DNA promoters, made us question the involvement of C277 in the enhancement of p53 DNA binding. The two 1NEM systems modelled (C182 and C277) were investigated further to explain the implication of each cysteine alkylation in the

effect of MAs on p53. 1NEM-C182 alkylation was found to stabilise and extend the H1 α-helix. This stabilised helix was suggested to allow a better p53 dimerization, stabilised with an extra salt bridge, which is thought to be responsible for the enhancement of the DNA binding. Recent studies of the other members of the p53 family, p63 and p73, showed that the affinity of p73 and p63 for DNA is lower than that of p53 and this has been shown to be due to their DBD oligomerisation interfaces that are different from p53 (156, 157). Both p63 and p73 lack R181(stabilises p53DBD dimer via a salt bridge with E180 from the opposite p53DBD monomer) in their dimer interfaces and additionally, p73 presents a smaller interface than p53 (157). These studies provide compelling evidence that C182 alkylation (stabilising and extending p53DBD dimer interface) alone is responsible for the enhancement of p53 function.

On the other hand, 1NEM-C277 was shown to rigidify the L1 loop by changing its conformation into a β-strand or an α-helix. This last conformation was shown to result in the motion away of K120 from the DBS. These results were confirmed by extending 2NEM simulations using aMD. Indeed, in these simulations the L1 loop was shown to adopt preferentially an α-helix than a β-strand.

These results confirmed our hypothesis that alkylation of C182 might be responsible for the enhancement of DNA binding in p53DBD wild-type. At low concentration, Michael acceptors react preferentially with the most solvent accessible cysteine in the full p53 structure, C182. This alkylation stabilises and extends the H1 helix, allowing more salt bridges and higher stability in the p53 dimer while binding to DNA.

At higher concentration, the Michael acceptors might interact with C277, which is thought to be more buried in the full p53 structure than in the p53DBD, the object of this study. Alkylation of C277 was found to induce inhibition of p53 DNA binding by rigidifying the L1 loop and preventing two direct contacts with DNA (mediated by DBS residues C277 and K120). Alkylation of C275 in the third step was found to expose a cluster of buried cysteines (C124, C135, C141) making them susceptible to alkylation by Michael acceptors. One can imagine both these processes contributing to the inhibition of p53DBD.

These results emphasize the importance of choosing an appropriate dosage of Michael acceptors in the rescue of mutant p53 and to specifically target C182. It has to be noted that the rescue of p53 by MAs does not reactivate full p53 function but only apoptosis via the implication of $p21^{waf1/cip1}$ (107). A study (158) showing the rescue of temperature sensitive mutants p63 and p73 by PRIMA-1^{MET} (APR-246) involves the implication of another cysteine alkylation. PRIMA-1^{MET} was shown to increase the transcription activity of the mutant TAp63γ and TAp73β (less in mutant Tap73α) and to induce cell cycle arrest and apoptosis in H1299 cells (158). The authors link this common rescue of mutant p53, p63 and p73 by PRIMA-1^{MET} to a common mechanism (cysteine alkylation), yet our simulations implicated C182 alkylation in the rescue of mutant p53DBD, a cysteine that does not exist in the p63 and p73 proteins (see section 3.4.3). Within the seven conserved cysteines in p63 and p73, the most probable candidate for the rescue alkylation in these structures is the C308 (in p63) and C297 (in p73) equivalent to C277 in p53DBD shown to be the most solvent accessible in p53DBD. This study sowed a doubt in our hypothesis that C277 alkylation inhibits p53DBD DNA binding. Indeed its alkylation alone might be responsible for p63 and p73 mutant rescue. Also, no evidence was found in the literature about the implication of p63 and p73 cysteines in the rescue by PRIMA-1^{MET}. The implication of C277 in the inhibition of the DNA binding needs more investigations to be able to conclude in the alkylation role

Chapter 6. Study of a destabilising p53DBD mutant: the V143A mutation

6.1 Introduction

In this project computer simulations are used to study the stability of p53DBD and the effect of a "structural" mutant, V143A (section 3.4.5). This mutant is temperature sensitive and is partially functional at low temperature and 80 % denatured at body temperature (92). The p53DBD WT on the other hand was found to be marginally stable at body temperature (half-life of 9 min) (90) and to still bind to some DNA promoters at low temperature. Both V143A and WT were shown to aggregate at certain temperatures (37 ^{0}C for V143A and 42 ^{o}C for the WT) (159). These changes of functionality according to the temperature are due to reorganisations of the DBD. To investigate the modifications responsible for p53DBD activity and loss of function in the WT and the V143A, models of these proteins were proposed on which molecular dynamics simulations were performed at different temperature. The use of molecular dynamics and other enhanced sampling methods enables the generation of as much data as needed to adequately study the conformational distribution of p53 and its mutants. The conformational data obtained via these simulations and their structural analysis allowed

(i) To propose an adequate model to study the proteins (WT and mutant) stability.

(ii) To highlight the conformational changes enabling p53DBD V143A functionality and inducing p53DBD WT's limited DNA-binding at 298K.

(iii) To introduce an alternate mechanism that can explain WT and V143A aggregation at 42 ^{0}C and 37 ^{0}C respectively.

6.2 Materials and methods

6.2.1 Protein set up

Among the 70 human p53DBD available in the protein data bank, a crystallographic structure of the p53 DBD tetramer binding to DNA (pdb code: 2AHI) was chosen based on its high resolution (1.85 Å) and the fact that the structure of its A chain is more complete (only one residue missing; Ser94 and on Ser95, Lys292 and Gly293 atoms are missing). A simulation of the complete tetramer of p53DBD in explicit water would be extremely difficult considering the number of water molecules needed to solvate the protein. Fortunately, it was shown that the monomeric p53DBD still binds to DNA, though weaker (160) and further, all existing biophysical analysis of the wild-type and mutants of monomeric p53DBD show that it is stable in vitro. This indicates that the monomer p53DBD domain can be used for stability analysis (161, 162). Knowing this, only chain A of the p53DBD tetramer, its coordinated Zinc and crystallographic water were kept from the PDB file. The missing residues Ser94, Ser95, Lys292 and Gly293 were replaced by the corresponding residues taken from other p53DBD crystallographic structures (pdbcode structure 3IGL for residues Ser94, Ser95 and Gly293 and structure 1GZH for Lys292). MODLOOP (151, 163) was used to adjust the coordinates and the conformation (via a short minimization) of the new residues added in the system.

To model the V143A p53 mutant, the alanine 143 coordinates was taken from the mutant crystallographic structure pdbcode 2J1W. This p53DBD structure also contains four suppressor mutations (M133L, V203A, N239Y and N268D). The alanine coordinates was introduced in the pdbfile of the model structure (2AHI chain A). As was the case for the addition of the missing residues, MODLOOP was used to adjust the V143A mutation coordinates in the structure.

Explicit hydrogen atoms are required for all-atom molecular mechanics, dynamics, docking and electrostatic calculations, and the initial state of the hydrogen bond network and ionization state of titratable groups can have a dramatic effect on simulation results. Unfortunately, most macromolecular crystal structures contain little or no hydrogen coordinates data due to limited resolution or disorder in the crystal. In this protocol only polar hydrogens were

added by WHATIF (152). The resulting structure has then been visually checked and some manual changes have been performed:

- His179 cannot be protonated on its delta nitrogen which is supposed to be coordinated to the Zinc. For this histidine the hydrogen on the δ nitrogen has been removed and the naming of the residue has been change on HIE, ie the hydrogen atom has been moved to the ε nitrogen atom (tleap was used to add the hydrogen on the right position).
- For His178, 214 and 233, hydrogens were added on the ε nitrogen by WHATIF (152). However a visualisation of these histidines shows that the protonation of their δ nitrogen generated hydrogen bonds or avoid clashes. These histidines were renamed in HID and the hydrogens on the ε nitrogens were moved to the δ nitrogen atoms.
- On the other histidines (H115, 168 and 193) the hydrogen was added on the right position but the residue naming is still HIS. So the naming was manually change to HIE.

Also some residues were manually flipped, as the glutamines 165 and 167 and the asparagine 288:

- The proposal to flip the Asn288 was made by the software Molprobity (164), so for this residue the new coordinates for the flipped version was taken from the protonation result of Molprobity.
- Considering Gln165 and 167, the new coordinates were manually calculated based on the fact that the atom OE1 will exchange his position with the atom NE2 on these residues.

The protein structure file obtained was then been converted to be understandable in AMBER.

C-termini N-methyl amide (NME) and a (ACE) capping groups were added to the model. These caps create a neutral terminus for the truncated structure chosen (only p53DBD considered). The cap addition was performed following the R. Walker tutorial (165).

Using the LEAP module of AMBER 10 (28), (12), with the ff99SB and gaff parameter set, two tleap processes in vacuum were performed to adjust the caps (NME and ACE) and include the Zinc parameterization (see section 5-2-2).

Finally two short minimizations (1000 steepest descent steps and 5000 conjugate gradient steps using SANDER module of AMBER 10) were applied on the caps and the flipped residues.

The systems were modelled in three different ionic conditions (using tleap); 150 mM, 75 mM and a condition in which only counter ions are added, named without extra ion (WEI). The structures were solvated in a cubic box of TIP3P waters with a buffering distance of 10 Å. The 6 different systems were constructed following this protocol. The details of these systems are reported in Table 6-1.

	Wild type			V143A mutant		
NaCl concentration (mM)	**WEI**	**75**	**150**	**WEI**	**75**	**150**
Number of ions	5 Cl^- and 0 Na^+	21 Cl^- (5 to neutralize) and 16 Na^+	37 Cl^- (5 to neutralize) and 32 Na^+	5 Cl^- and 0 Na^+	21 Cl^- (5 to neutralize) and 16 Na^+	37 Cl^- (5 to neutralize) and 32 Na^+
Water box dimension, Å (Volume, $Å^3$)	77.1, 74.2, 77.7 (444151.7)	92.9, 92.9, 85.0 (734275.9)	101.9, 97.9, 105.3 (1051003.8)	77.1, 74.2, 77.7 (444151.7)	95.9, 84.8, 83.3 (677036.5)	104.0, 97.6, 103.2 (1051444.2)
Number of atoms	35828	60616	88233	35753	55549	88347

Table 6-1: Information on the different systems modelled at different ionic concentrations.

The minimisation and equilibration have been performed using SANDER module of Amber 10. The minimization process was performed gradually (caps first, water, water + ions, the all protein and finally the full system) on the systems to remove bad contacts, using 1000 steepest descent steps and 5000 conjugate gradient steps. The equilibration was also performed gradually. During the first step, a restraint is applied to the system to induce ion and hydrogen atoms relaxation by applying potential position restraint on protein heavy atoms (1000 $kcal.mol^{-1}.Å^{-2}$). Then the temperature was raised from 100 K to 300 K (using Langevin thermostat (166)) over 200 ps at a constant volume and protein restraint in order to equilibrate the solvent. To adjust the size of the water box, a second equilibration step then performed on the solvent (protein restrained) at constant temperature (300 K) and pressure (1 bar) over 200 ps using the Langevin piston coupling algorithm (167). The third equilibration step was applied to decrease the temperature of the system from 300 K to 100 K in 100 ps (protein restraint and constant volume). Once the

solvent relaxed, six minimization processes were applied using decreasing restraint weights (from 1000 $kcal.mol^{-1}.Å^2$ to 0) on the protein. The temperature was raised gradually (60 K over 10 ps) from 0 to 300 K at constant volume. As a final step, the system was maintained at constant temperature (300 K) and constant pressure (1 bar) over 200 ps. The SHAKE algorithm (22) was used to constrain bonds involving hydrogen atoms, allowing for an integration time step of 2.0 fs. The Particle Mesh Ewald method was employed to treat long range electrostatics with a 12 Å cut-off.

All this process allowed water molecules and ions to find more natural positions in the system to obtain its equilibrated state in order to perform molecular dynamic simulations.

50 (10 systems * 5 repeats) constant NPT simulations were run for 500 ns each using the PMEMD module of Amber 12 (12, 28) and ff99SB and GAFF force fields. These simulations were performed on the equilibrated structures kept at 298 K, 310K and 323K and a pressure of 1 bar using the Langevin piston coupling algorithm. The SHAKE algorithm was used to freeze bonds involving hydrogen atoms, allowing for an integration time step of 2.0 fs. The Particle Mesh Ewald method (PME) was employed to treat long range electrostatics with a 12 Å cut-off. Conformations were collected every 50 ps, resulting on 10000 snapshots per simulation.

The parameters for the aMD (see section 2-9) were calculated from the cMD simulations from 50 to 60 ns. The parameters used are reported in the table below.

	Alpha D	Ethresh D	Alpha P	Ethresh P
WT at 298K	139.3	2791.5	12123.2	-191091.1
WT at 310K	139.3	2793.5	12123.2	-178254.8
WT at 323K	139.3	2809.5	12123.2	-174891.3
V143A at 298K	139.3	2789.5	11109.8	-167109.1
V143 at 310K	139.3	2909.5	11109.8	-163378.9
V143A at 323K	139.3	2804.1	11109.8	-160307.6

Table 6-2: aMD parameters.

6.3 Results

6.3.1 The ion concentration effect on p53DBD

A preoccupation while studying cell or protein behaviour is to provide a physiological environment to the system. The common definition of physiological conditions for most earth organisms is a temperature range of 20-40 °C, an atmospheric pressure of 1, a pH of 6-8, glucose concentration of 1-20 mM and earth gravity (168). The ionic strength applied to a system is based on the known concentration in the organism serum and cells. In almost all cells, the cytosol pH is near 7.2 and the cytosolic concentration of K^+ is much higher than that of Na^+ and Ca^{2+} (169). This trend is maintained even if the cells are cultured in very dilute salt solutions. A typical ion concentration for vertebrates is shown in table 1 (170). Knowing this, in biological buffers, the physiological ionic strength is between 100 to 200 mM of KCl or NaCl (171) in order to be as close as possible to cell or blood concentration of ions.

Ion	Cell (mM)	Blood (mM)
K^+	139	4
Na^+	12	145
Cl^-	4	116
$HCO3^-$	12	29
X^-	138	9
$Mg2^+$	0.8	1.5
$Ca2^+$	<0.0002	1.8

Table 6-3: Typical ion concentration in mammalian cells (170).

In experimental procedures, protein stabilization or destabilisation caused by ionic compounds binding to specific residues should increase as the difference between pI (isoelectric point, pH at which a particular molecule or surface carries no net electrical charge) and pH becomes greater. Owing to the various protein stabilization mechanisms, the way salts may cause protein stabilization (or destabilization) are complex and not well understood (172). This question has been addressed in protein simulation with a similar result. Indeed, the

effect of explicit salt ions on protein stability seems to depend on the system studied. When performing MD simulation on a small protein marginally stable (Yap-WW domain, about 40 residues), the use of a high ionic strength in addition of the charge-balancing counterions (a necessity while using the PME method for highly charged proteins) allowed a stable protein structure to be miantained (173). The computational study of ion concentration effects on βARK1 PH domain (119 residues) shows alteration of the average structure in the loop regions and of the fluctuation amplitudes of dihedral angles when increasing the number of ions, but in the absence of ions, unnatural strong interactions between side chains were formed. They conclude that a moderate ion concentration, 27 mM of NaCl shortened the time needed to equilibrate the ionic environment around the protein (174). A later study investigating human Lymphotactin stability highlights structural changes at C-terminal and N-terminal regions at different temperature and ion concentration (175), showing that even within the same structure each region might be sensitive to different environmental conditions.

Salt concentration impact on full length p53 has been investigated experimentally (176), indicating that low salt concentration (50 mM to 150 mM) promotes p53-DNA interactions and affects the folding of the C-terminal domain of p53 while high ion concentration (beyond 150 mM to 300 mM) induces the loss of p53 DNA binding by disrupting the p53 quaternary structure. The influence of ion concentration in p53DBD was investigated, to propose an optimal simulation condition for our system. Three different conditions were built: one in which only ions to neutralise p53DBD were added, one with 75 mM of NaCl and a third with 150 mM of NaCl (Table 6-1). Five 500 ns simulation were performed for each of these systems at 298K. A visualisation of the systems at different conditions showed stability of the systems (Fig 6-1) at different concentrations. The average α-carbon RMSD with respect to the reference structure (2AHI chain A) were of 2.8 Å (SD = 0.5) for the WEI system, 2.6 Å (SD = 0.5) for the system at 75 mM, and 2.7 Å (SD = 0.5) for the system at 150 mM of NaCl. Most of the fluctuation is located in the DNA binding surface (DBS) of the protein (Fig. 6-2) while the p53 immunoglobulin like β-sandwich is stable along the simulations and the different systems with a constant RMSD of about 1 Å. The DBS (loops L1, L2

and L3, the sheet S2/S2' and the helices H1 H2) showed mainly fluctuation on the L1 loop with a RMSF greater than 4 Å while the rest of the DBS had a RMSF ≤ 2 Å (Fig. 6-3). A VMD visualisation of the simulation confirmed the implication of the L1 loop alone in most of the DBS fluctuation (Fig 6-4). Indeed while the L1 loop is located close to the H2 α-helix (like in the reference structure, see Fig. 6-1) as seen in the repeat 3 of the WT WEI simulation (Fig 6-4, from 0 to 400 ns), the DBS RMSD is ≤ 2 Å. When the L1 loop moves away from the H2 α-helix (Fig. 6-4, frame 10000) the DBS RMSD is ≥ 2 Å. The fluctuation of the L1 loop has been observed previously in our PCA of the experimental p53DBD structures. This fluctuation has been described to be important for the DNA binding specificity (96, 113, 114).

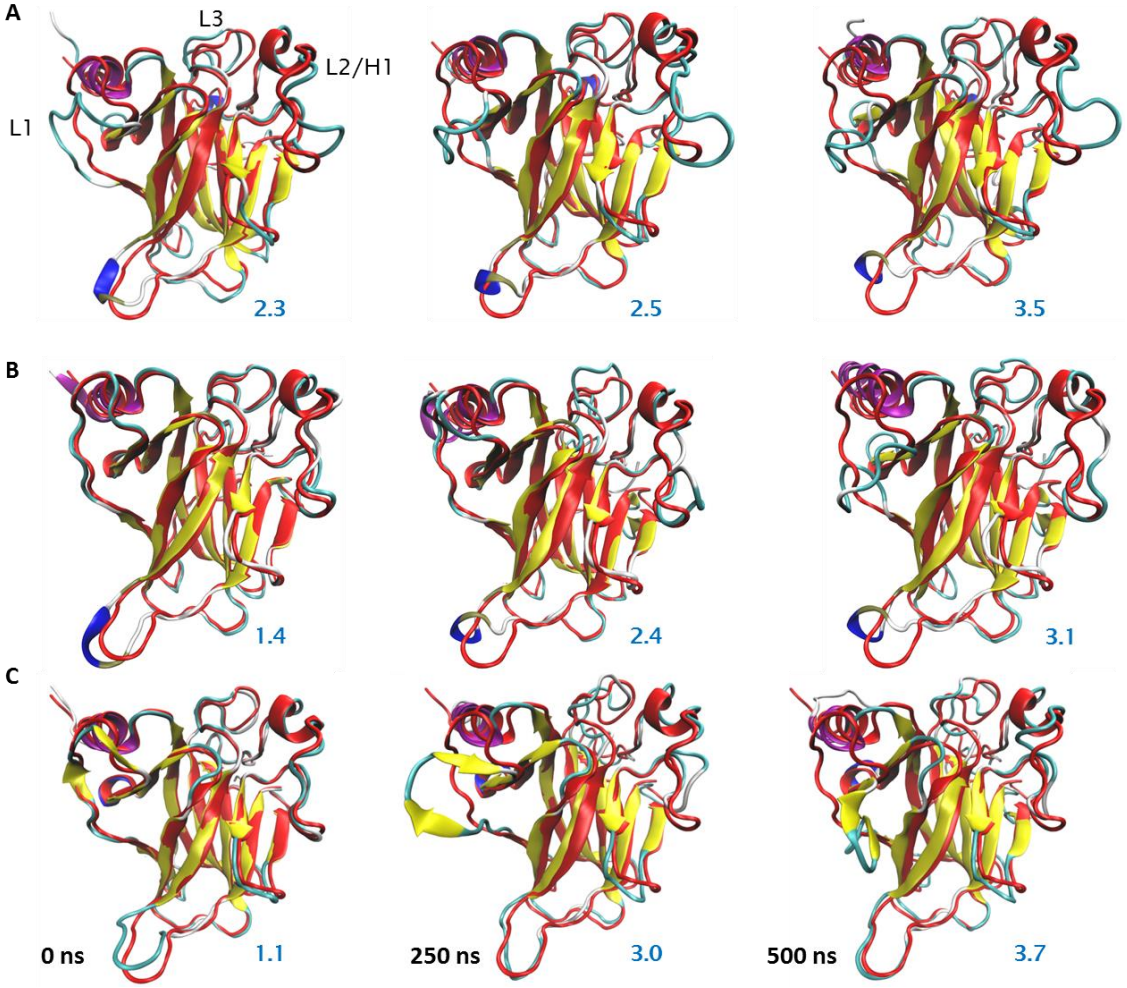


Figure 6-1: VMD visualisation of the p53DBD wild-type systems at different ion concentrations.
The systems were aligned on 2AHI chain A (reference structure) alpha-carbon at different times of the simulations. Here is shown only the first simulation. **A)** WEI, **B)** 75 mM, **C)** 150 mM of NaCl. The structures are represented in cartoon, the simulation frames were coloured by structure and the reference structure was coloured in red. The RMSD of the frame with respect to the reference structure (in Å) is indicated in blue.

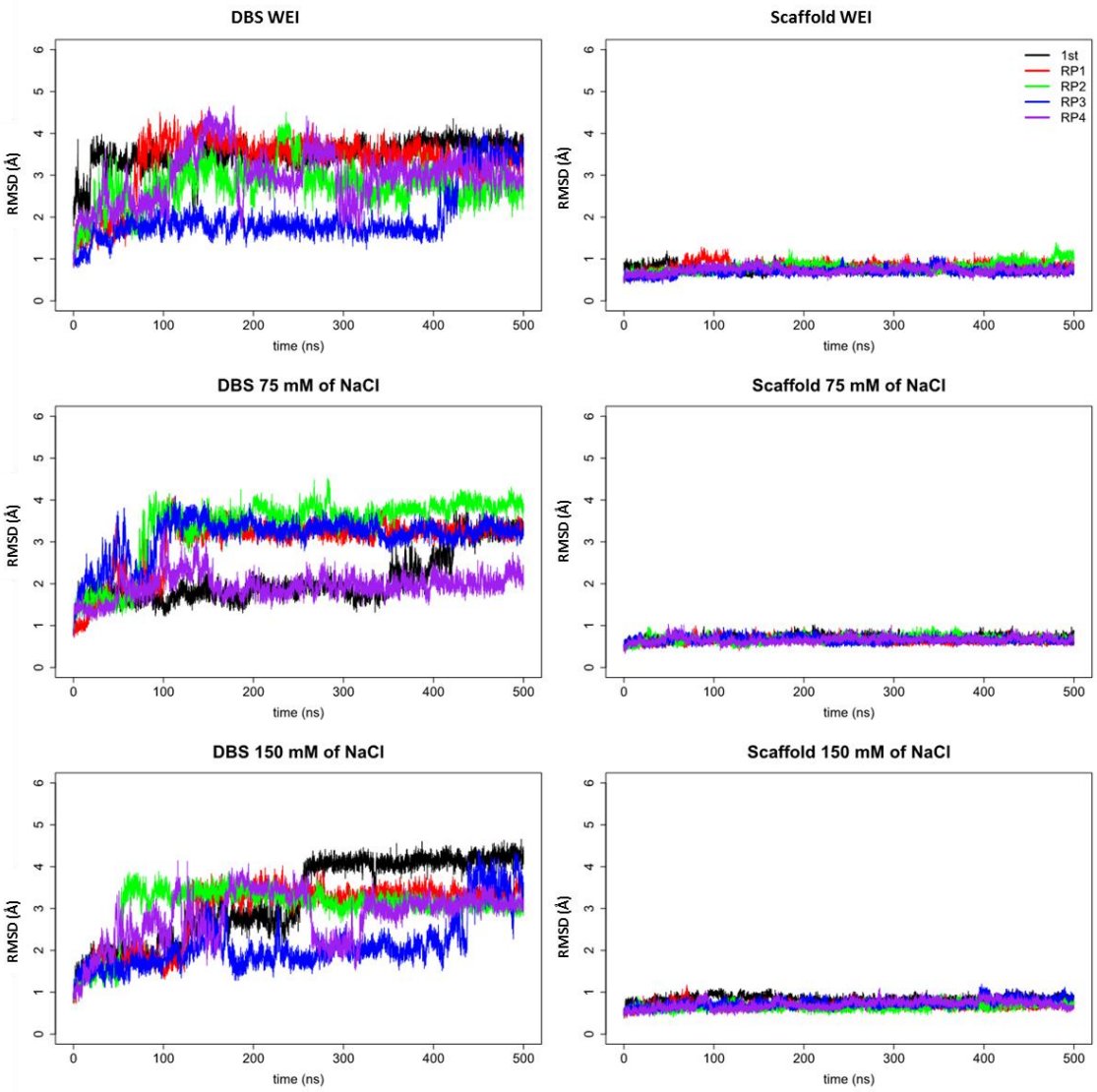


Figure 6-2: RMSD of the p53DBD wild-type systems at different ion concentrations.

RMSD of the DNA binding surface (DBS) against the immunoglobulin-like β-sandwich (scaffold) was calculated from the simulations and each repeat after alignment on the α-C of the reference DBS or scaffold (2AHI chain A).

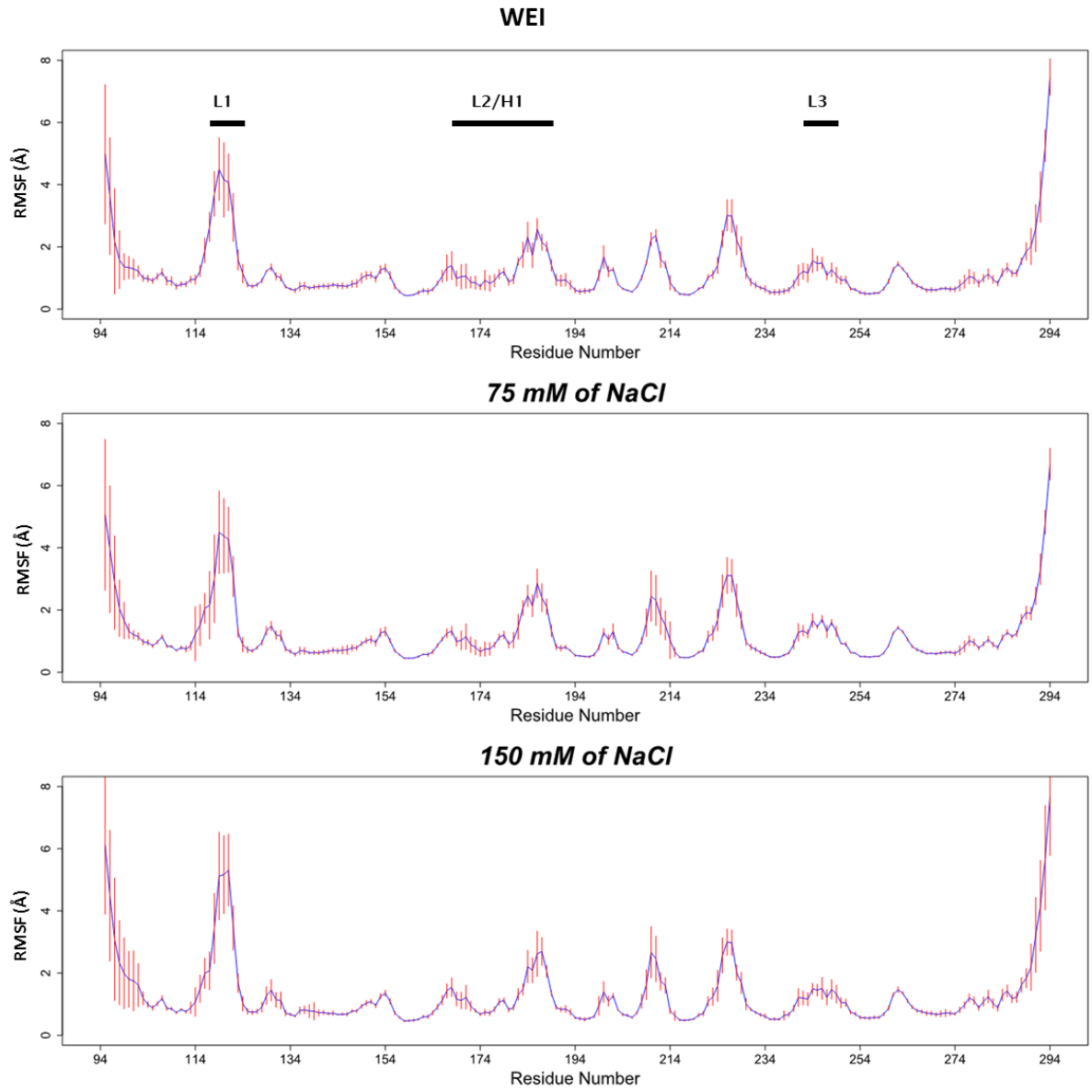


Figure 6-3: RMSF of the p53DBD wild-type systems at different ion concentrations.
The RMSF for each 500 ns simulation was determined using cpptraj in Amber. The average RMSF (blue) among the five conventional MD (cMD) and the standard deviation (red) for each system were calculated using R.

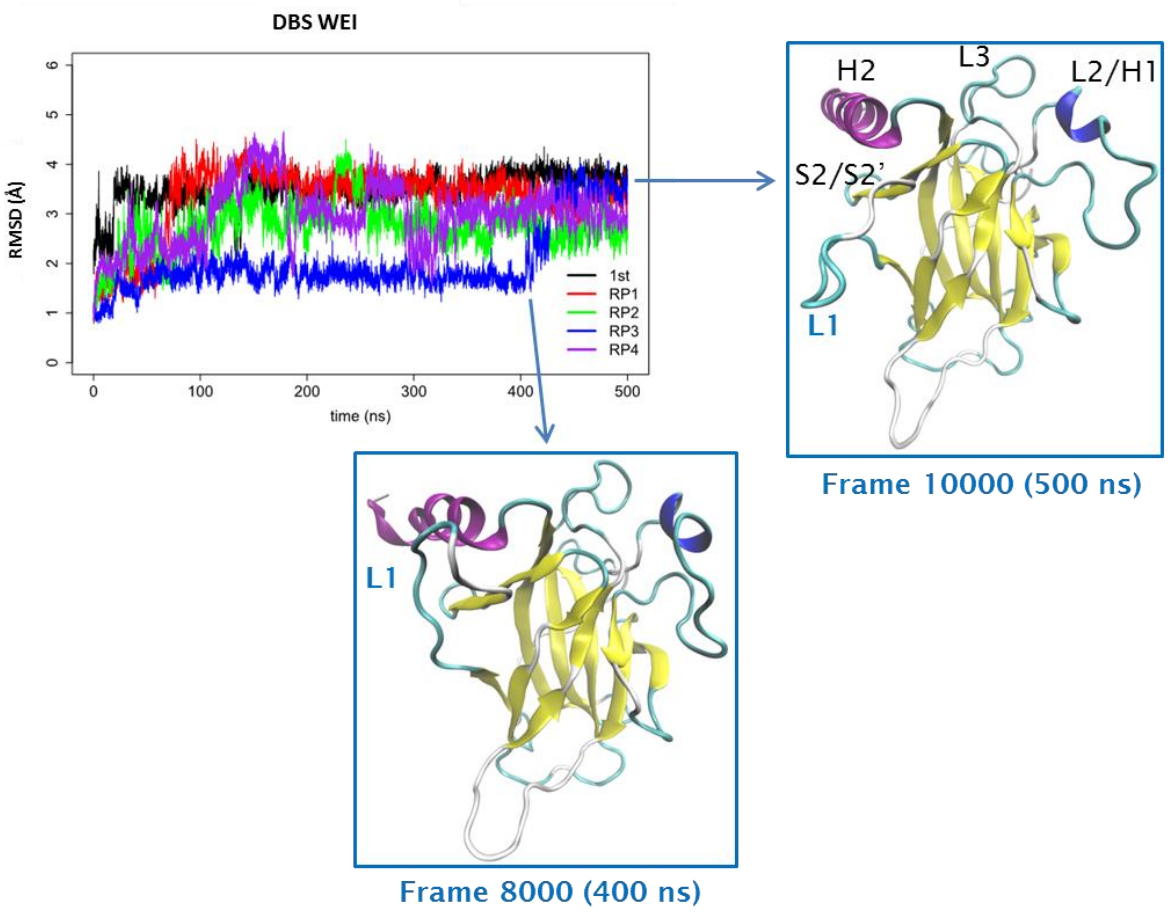


Figure 6-4: The DBS fluctuation involves mainly the L1 loop.
The RMSD of the α-carbon with respect of the reference structure was performed using Amber and plotted using R. The frames corresponding to the 400 ns and 500 ns of the WT WEI simulation (repeat 3, RP3) are shown in cartoon and coloured by secondary structure.

The secondary structure analysis of the simulations and their repeats showed conservation of the secondary structure. The main difference between the simulations at different ion concentrations was located in the L1 loop (Fig. 6-5). In the WEI, the L1 loop tends to form an α-helix while in the 150 mM of NaCl system it forms a β-sheet (both present in only one repeat). As it can be seen in the VMD visualisation (Fig. 6-5), these conformational changes drive the L1 loop away from the rest of the DBS, making it impossible to interact with the DNA. Several residues within the DBS are known to contact directly the DNA (177), among them K120 is situated in the L1 loop. This contact is likely to be abolished in the p53DBD structure having an α-helix or a β-sheet instead of the L1 loop. These results highlights another way by which ion concentration might influence DNA binding specificity other than the inhibition of the p53 tetramerisation found in the experimental study (176). As the L1 loop changes

concerned only one repeat (not reproducible), the actual results did not allow us to conclude on the effect of ion concentration on the p53DBD.

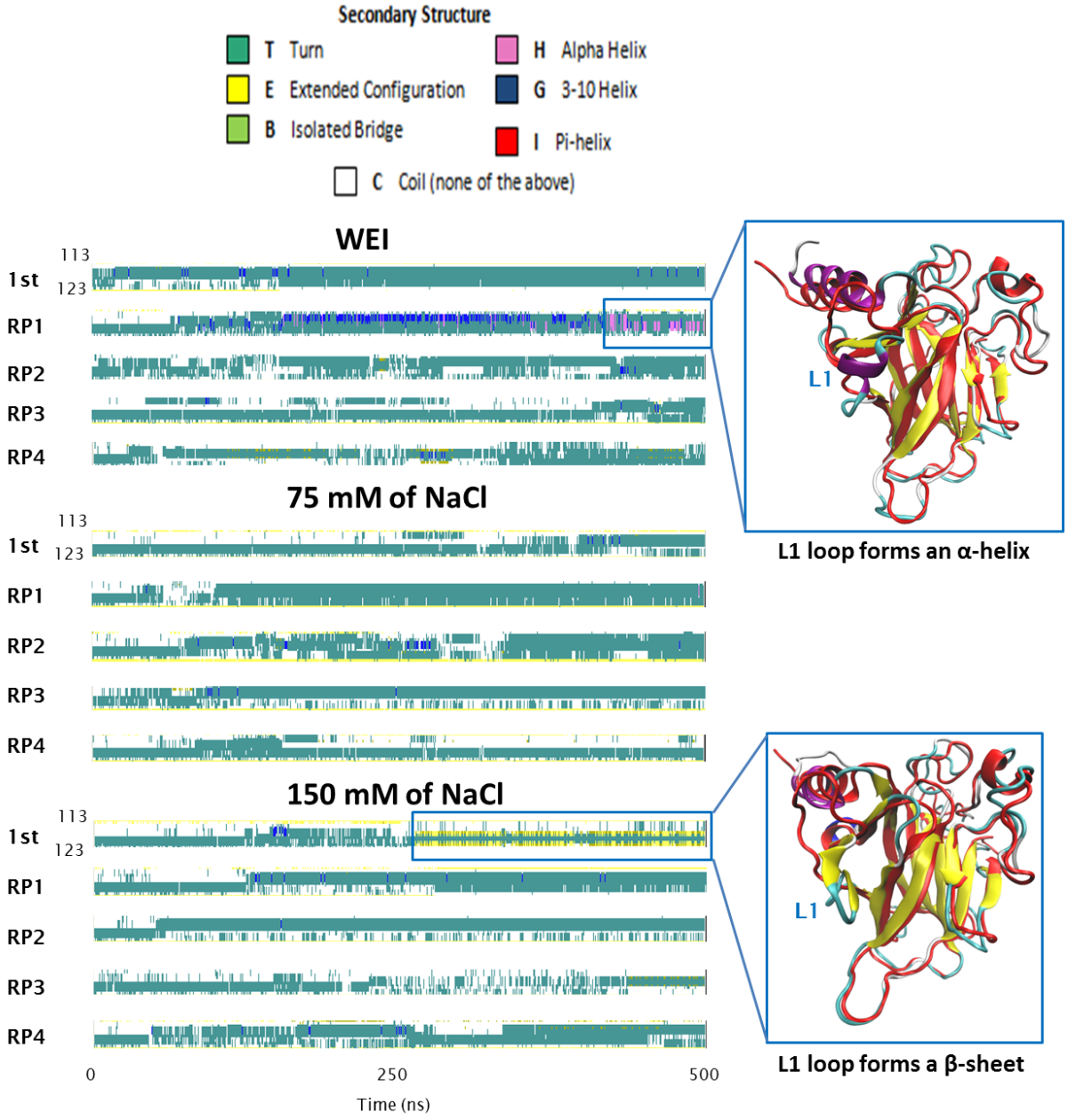


Figure 6-5: Determination of secondary structure of L1 loop from the wild type systems at different ion concentrations.
The secondary structure analysis was performed using timeline analysis in VMD and only the L1 loop (from residue 113 to 123) is shown.

A principal component analysis (PCA) of the α- carbons was performed (Fig. 6-6), on a trajectory obtained by combining the wild type systems at different concentrations (aligned on the reference structure, 2XWR chain A), to highlight the main variance in the simulations and estimate the phase space explored by

each ion concentration system. While three PCs captured 36.3 % of the systems variance 20 PCs are needed to display 80.6 % of the variance (Fig. 6-6 A). The motion projection along the three first PCs (Fig. 6-6 B) showed motion mainly on the L1 loop which persist along the 3 PCs. This confirms that the main fluctuation along these systems is located on the L1 loop. The projection of the frames along the three first PCs (Fig. 6-7) showed a similar distribution of the systems from the three different ion concentration conditions along the three first PCs. While the systems explored the same phase space, it can be noticed that the lower ion concentration systems explored more phase space. Given these results, the ionic strength of 75 mM of NaCl was chosen to investigate the V143A mutation on p53DBD.

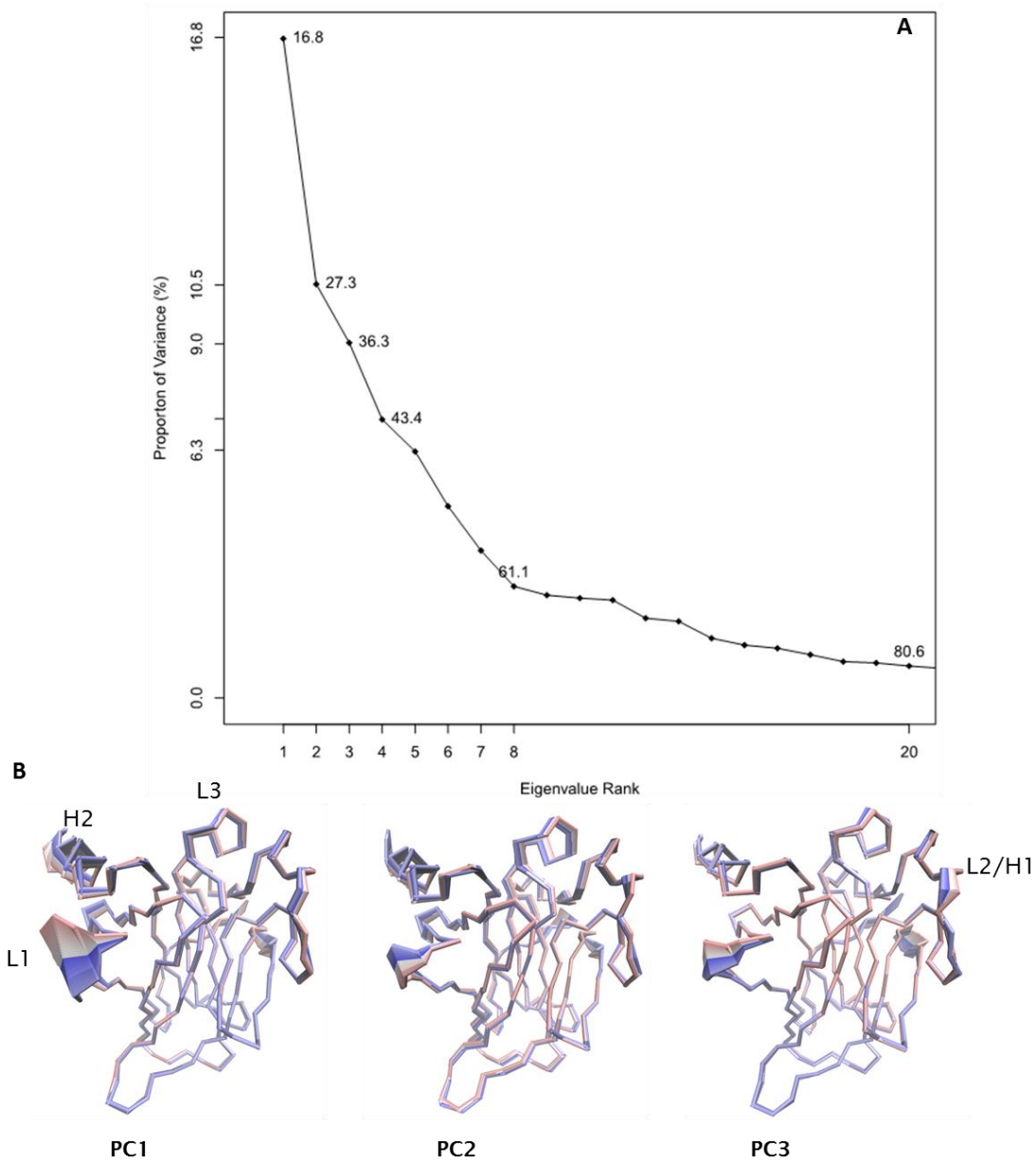


Figure 6-6: PCA of the p53DBD wild type at different ion concentrations.
A) Proportion of variance (in %) captured by each principal component (PC) from a data set of 150000 p53DBD frames (WEI, 75 mM and 150 mM of NaCl + repeats). The 3 first PCs contribute to 36.3 % of the variance of the dataset. **B)** Trace representation of the motion projections along the 3 first PCs (PC1, PC2 and PC3) of the 150000 frames using VMD. The colouring range of the motions goes from red, to white, to blue as function of time (timestep option on VMD), red being the early motion.

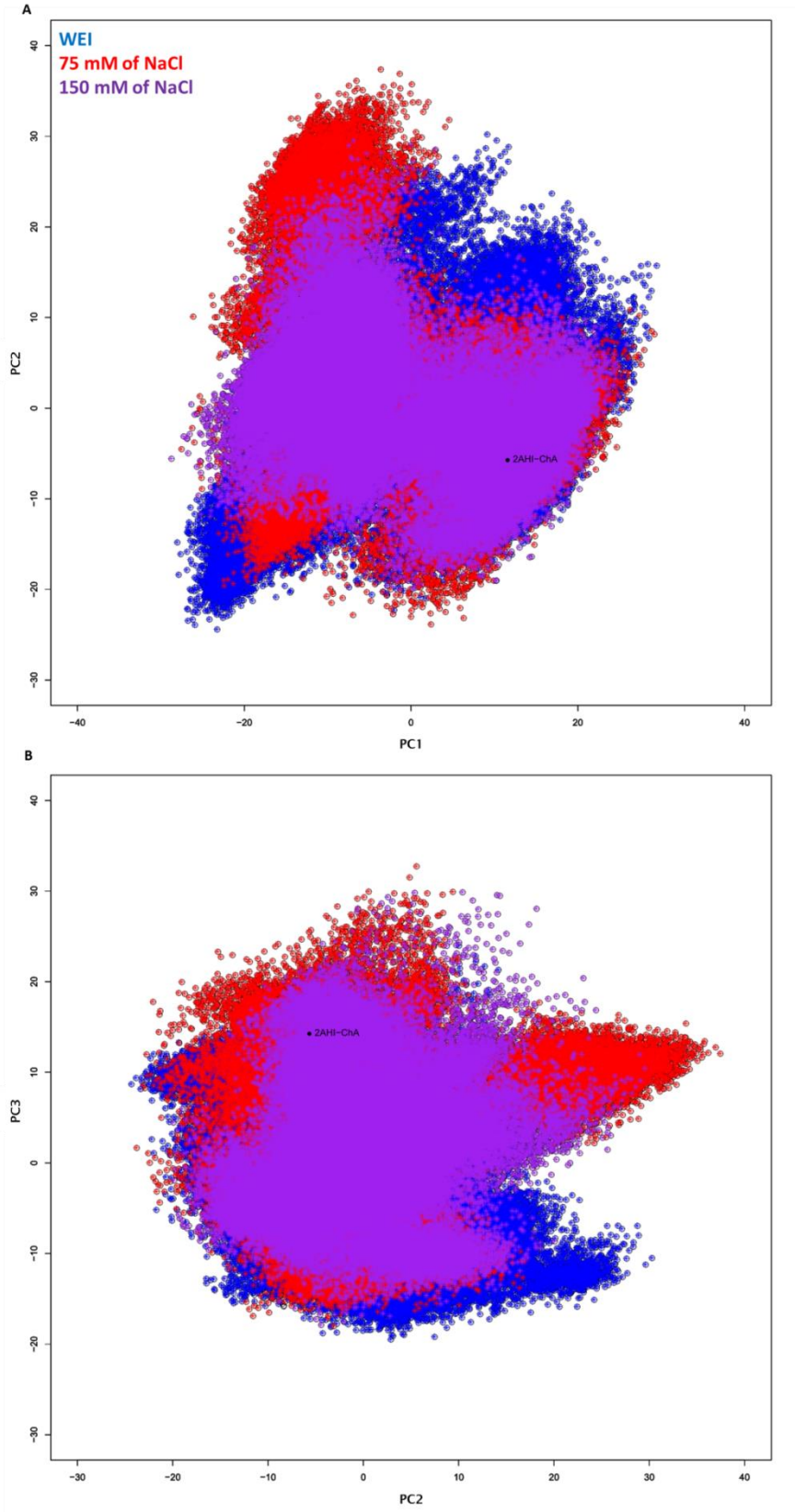


Figure 6-7: Structures projections of the p53DBD WT along the three first PCs. **A)** Along PC1 and PC2. **B)** Along PC2 and PC3. The 150000 frames (50000 for each ion concentration condition) were aligned on the reference structure (2AHI chain A). The projected reference structure is labelled in black.

The influence of ion concentration was also investigated for the p53DBD mutant V143A model. While ion concentration might influence the p53DBD WT DNA binding, no major changes were observed in the mutant. As for the WT the stability of p53DBD V143A mutant was investigated. The RMSD of the full structure and a VMD visualisation of the simulations (Fig. 6-8) showed a stability of the overall structure along the simulation. Most of the fluctuations were shown to be located on the DBS with a maximum RMSD reaching 4.5 Å for all simulations whatever ion concentration considered (Fig. 6-9) while the p53DBD scaffold remained stable with a RMSD ≤ 1 Å along the simulations. The fluctuation in the DBS is mainly located in the L1 loop as confirmed by the RMSF (Fig. 6-10), which increased as the ion concentration was raised (as in the p53DBD WT). Unlike the WT, the secondary structure of the p53DBD mutant was conserved along the simulations, the L1 loop did not change conformation (data not shown).

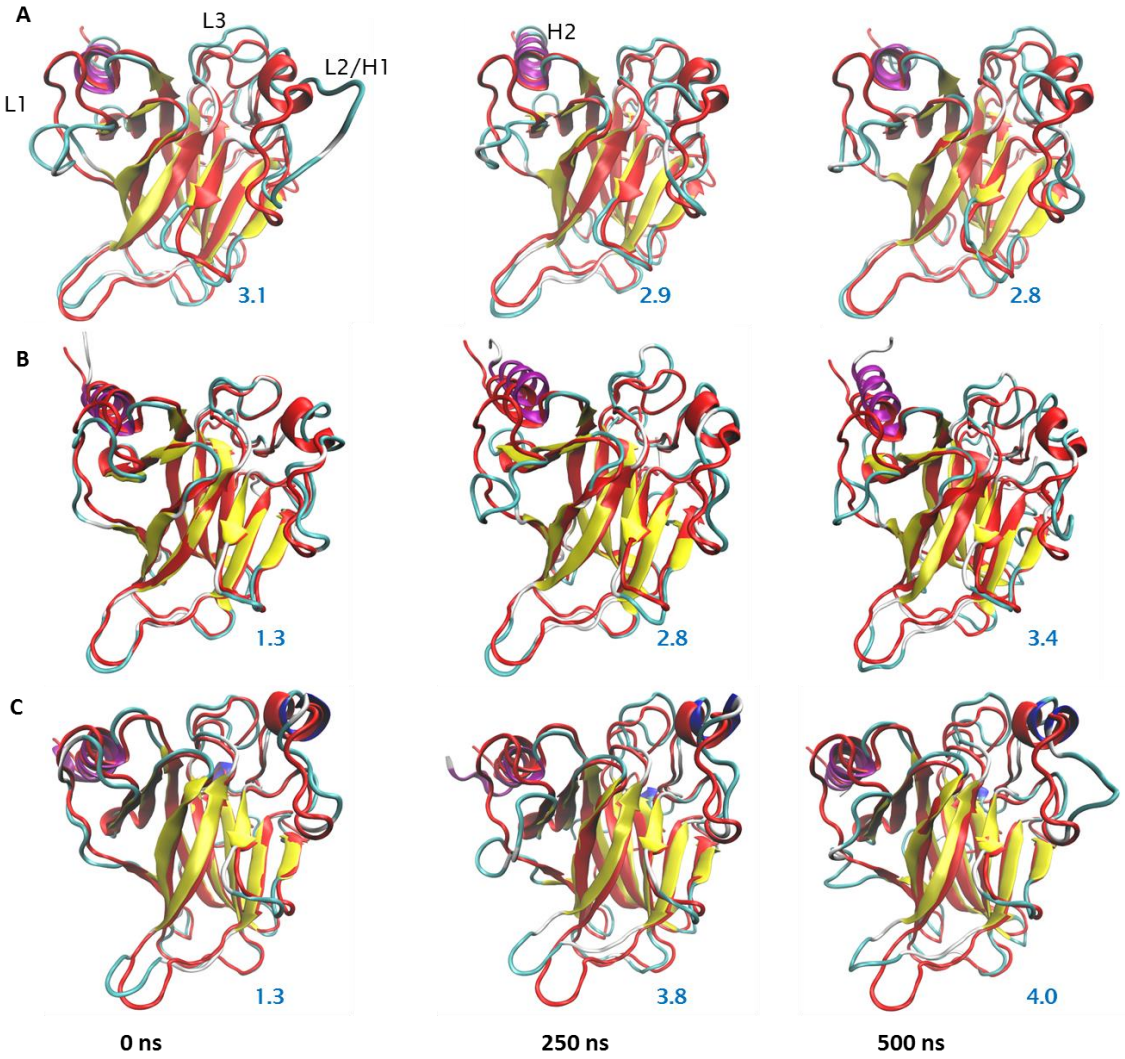


Figure 6-8: VMD visualisation of the p53DBD V143A systems at different ion concentrations.
The systems were aligned on 2AHI chain A (reference structure) alpha-carbon at different time of the simulations. Here is shown only the first simulation. **A)** WEI, **B)** 75 mM, **C)** 150 mM of NaCl. The structures are represented in cartoon, the simulation frames were coloured by structure and the reference structure was coloured in red. The RMSD of the frame with respect to the reference structure (in Å) is indicated in blue.

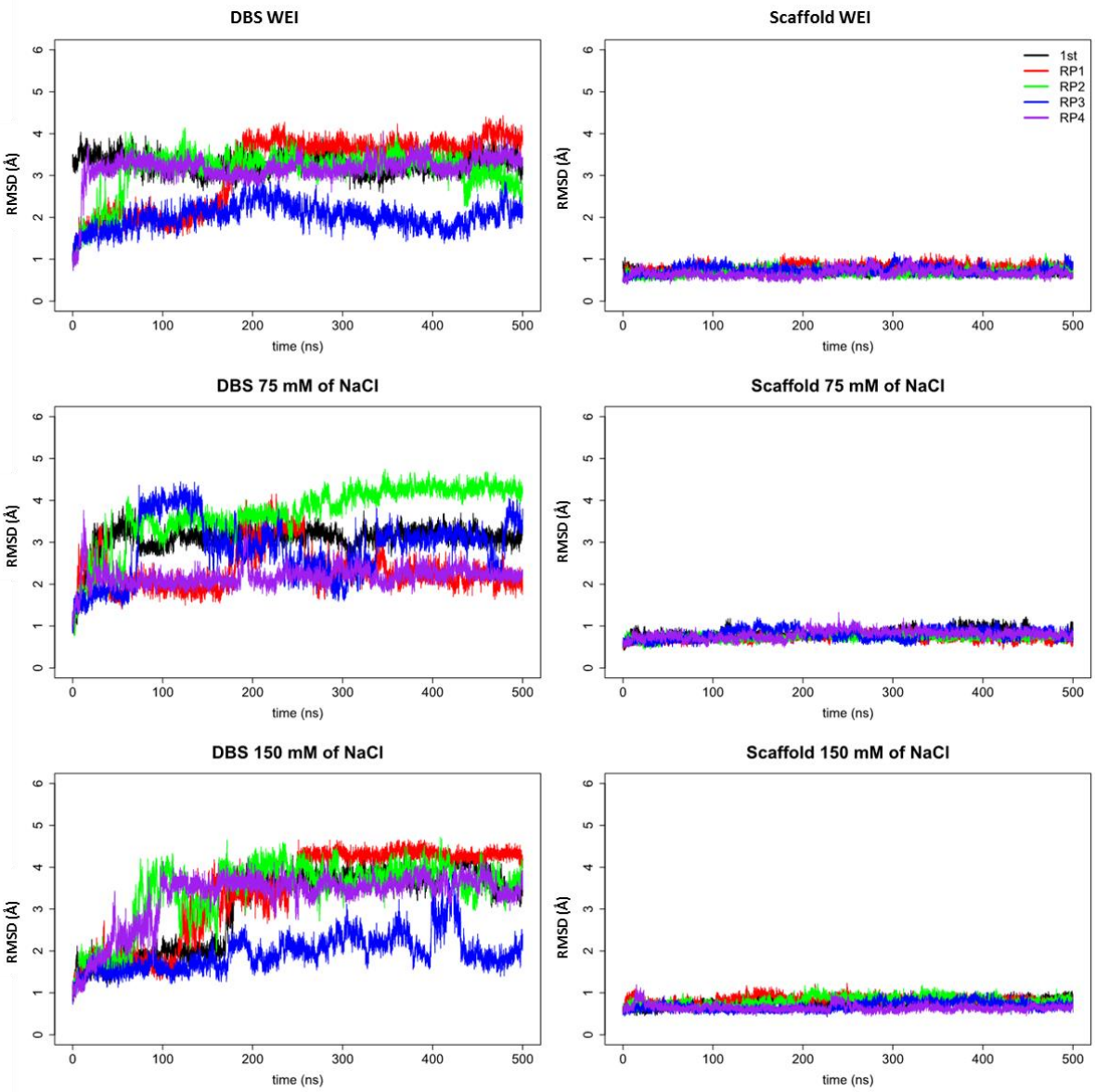


Figure 6-9: RMSD of the p53DBD V143A systems at different ion concentrations.

RMSD of the DNA binding surface (DBS) against the immunoglobulin-like β-sandwich (scaffold) was calculated from the simulations and each repeat after alignment on the reference DBS or scaffold (2AHI chain A).

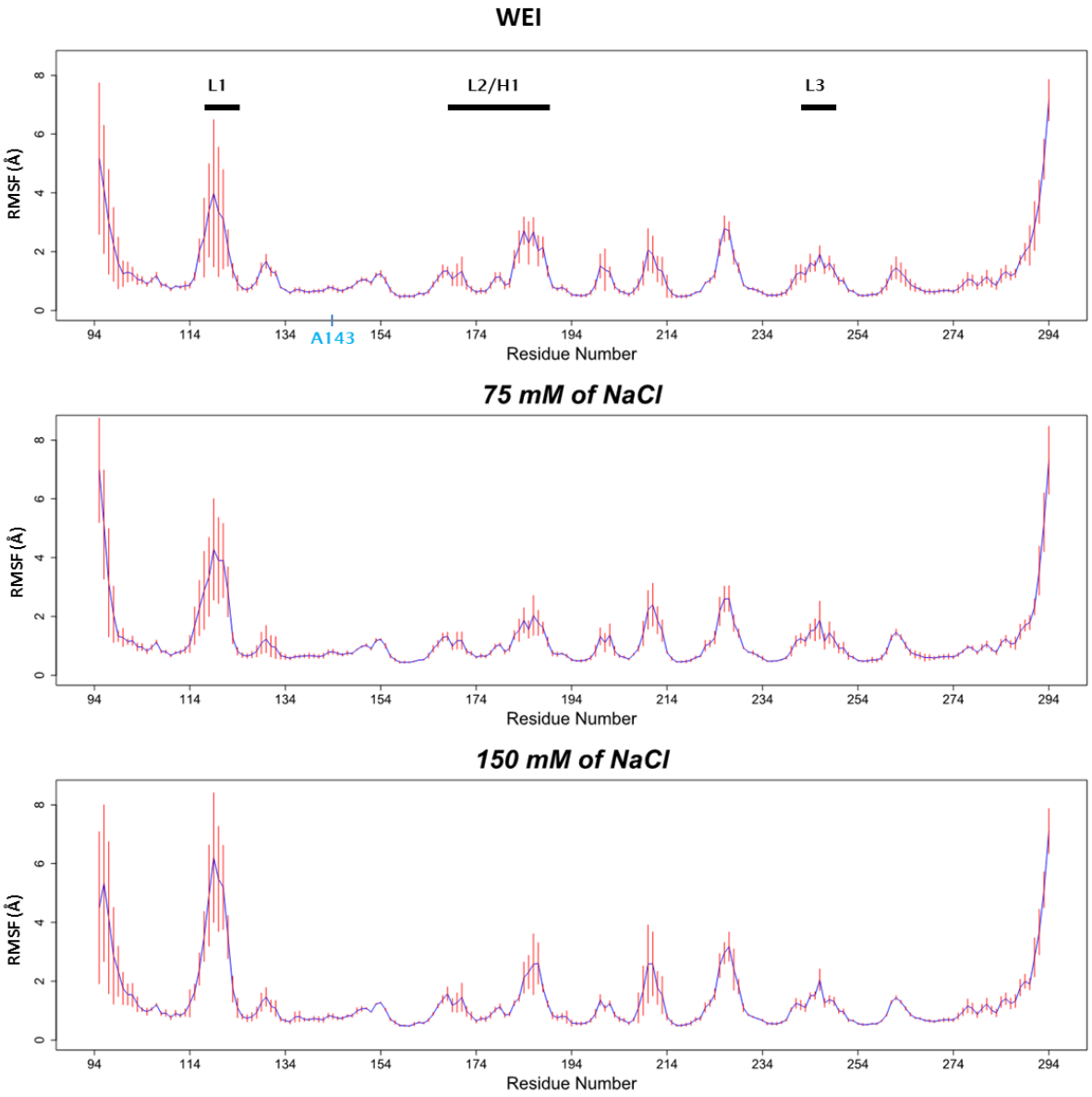


Figure 6-10: RMSF of the p53DBD V143A systems at different ion concentrations.
The RMSF for each 500 ns simulation was determined using cpptraj in Amber. The average RMSF (blue) among the five cMD and the standard deviation (red) for each system were calculated using R. The mutated residue is labelled in blue.

On the other hand, the PCA of the p53DBD V143A showed a different tendency than the WT. The analysis performed on all the mutant simulations at different concentration captured 40.2 % of variance when only the first three PCs are considered (20 PCs are needed to highlight 82.7 % of the variance, see Fig. 6-11 A). All three PCs presented fluctuation in the L1 loop (Fig. 6-11 B), which confirmed our previous results. In addition PC2 highlighted fluctuation on the L6 loop and PC3 in the L2/H1 region. The projection of the structures along these three PCs (Fig. 6-12) showed that the p53DBD V143A simulation at 150 mM of NaCl explored more phase space than the simulation at lower ion concentration. This difference is only observed along PC2 which captures the L6 loop motion. A second structure projection highlighting only the 150 mM simulations (coloured per simulations, see Fig. 6.13) showed that this difference is only due to the 1st simulation. In this simulation the L6 loop adopts a conformation different from the reference structure (Fig. 6-8 C). The fluctuation of this loop, as discussed previously in chapter 4, is a normal behaviour in p53DBD. At first, the L1 loop was thought to stabilise the new positon of L6 loop via interactions. In the 1st simulation of p53DBD V143A (150 mM of NaCl), the dynamic cross correlation matrix (DCCM, see Fig. 6-14) and VMD visualisation highlighted neither a specific correlation between the L1 and L6 loop, nor interactions that could explain the stability of the L6 loop in this particular position. As a result, the L6 loop deviation (consequently PC2) was ignored in considering which ionic concentration is most appropriate. The projection of the p53DBD V143A frames along PC1 and PC3 (Fig. 6-15), showed the same tendency as the p53DBD WT PCA (Fig. 6.7). The mutant systems at different concentration explore the same phase space, and this, decreases when the ion concentration is raised.

From these findings, the ion concentration of 75 mM was believed to be the best condition to study the stability of the p53DBD V143A mutant. This concentration was shown to conserve the L1 loop conformation in the p53DBD WT simulations and allow a wider phase space exploration than the 150 mM condition, and enabled us to propose a model closer to the physiological condition than the WEI.

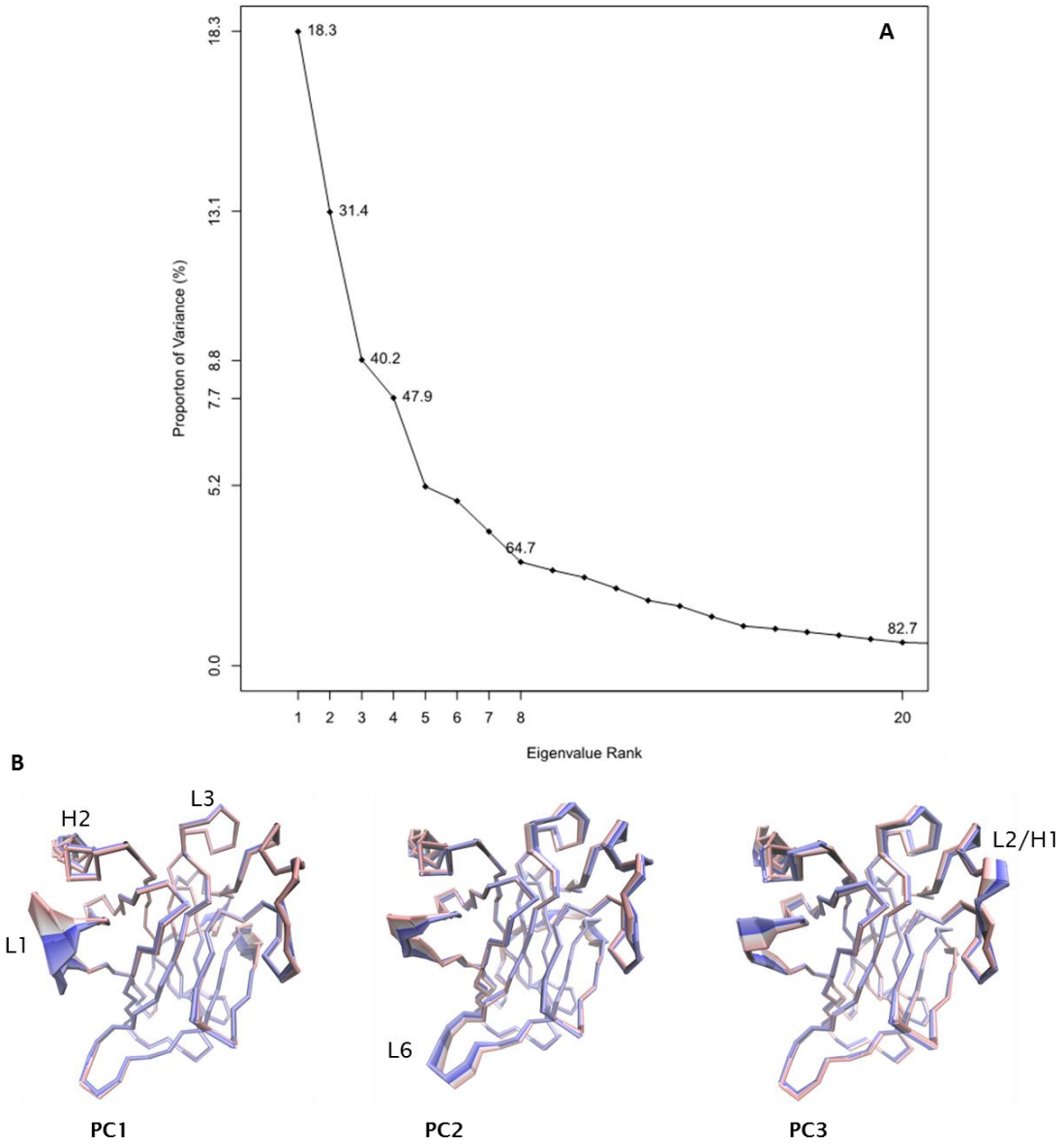


Figure 6-11: PCA of the p53DBD V143A at different ion concentrations.
A) Proportion of variance (in %) captured by each principal component (PC) from a data set of 150000 p53DBD frames (WEI, 75 mM and 150 mM of NaCl + repeats). The 3 first PCs contribute to 40.2 % of the variance of the dataset. **B)** Trace representation of the motion projections along the 3 first PCs (PC1, PC2 and PC3) of the 150000 frames using VMD. The colouring range of the motions goes from red, to white, to blue as function of time (timestep option on VMD), red being the early motion.

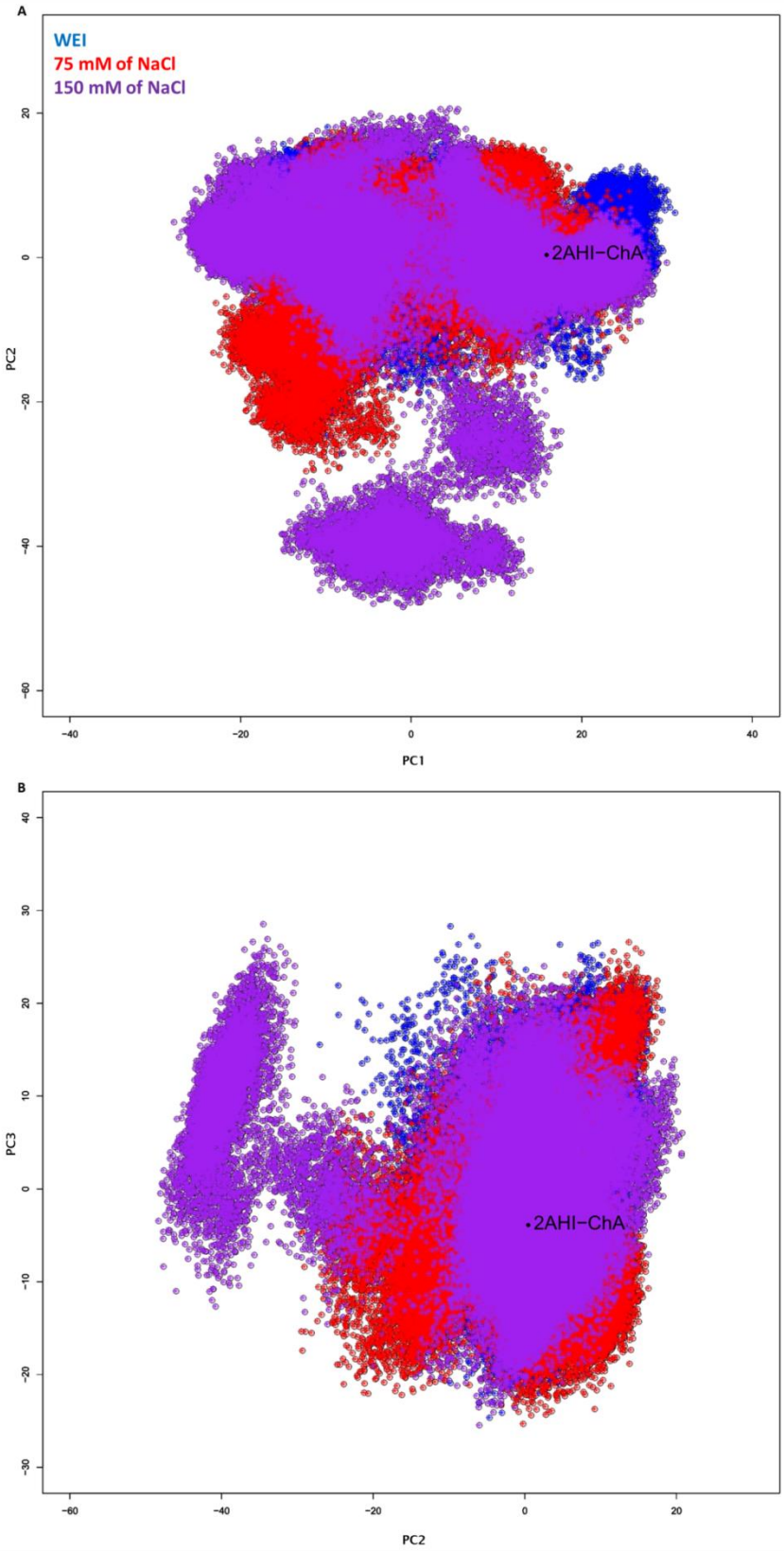


Figure 6-12: Structures projections along the three first PCs of the p53DBD V143A simulations.

A) Along PC1 and PC2. **B)** Along PC2 and PC3. The 150000 frames (50000 for each ion concentration condition) were aligned on the reference structure (2AHI chain A). The projected reference structure is labelled in black.

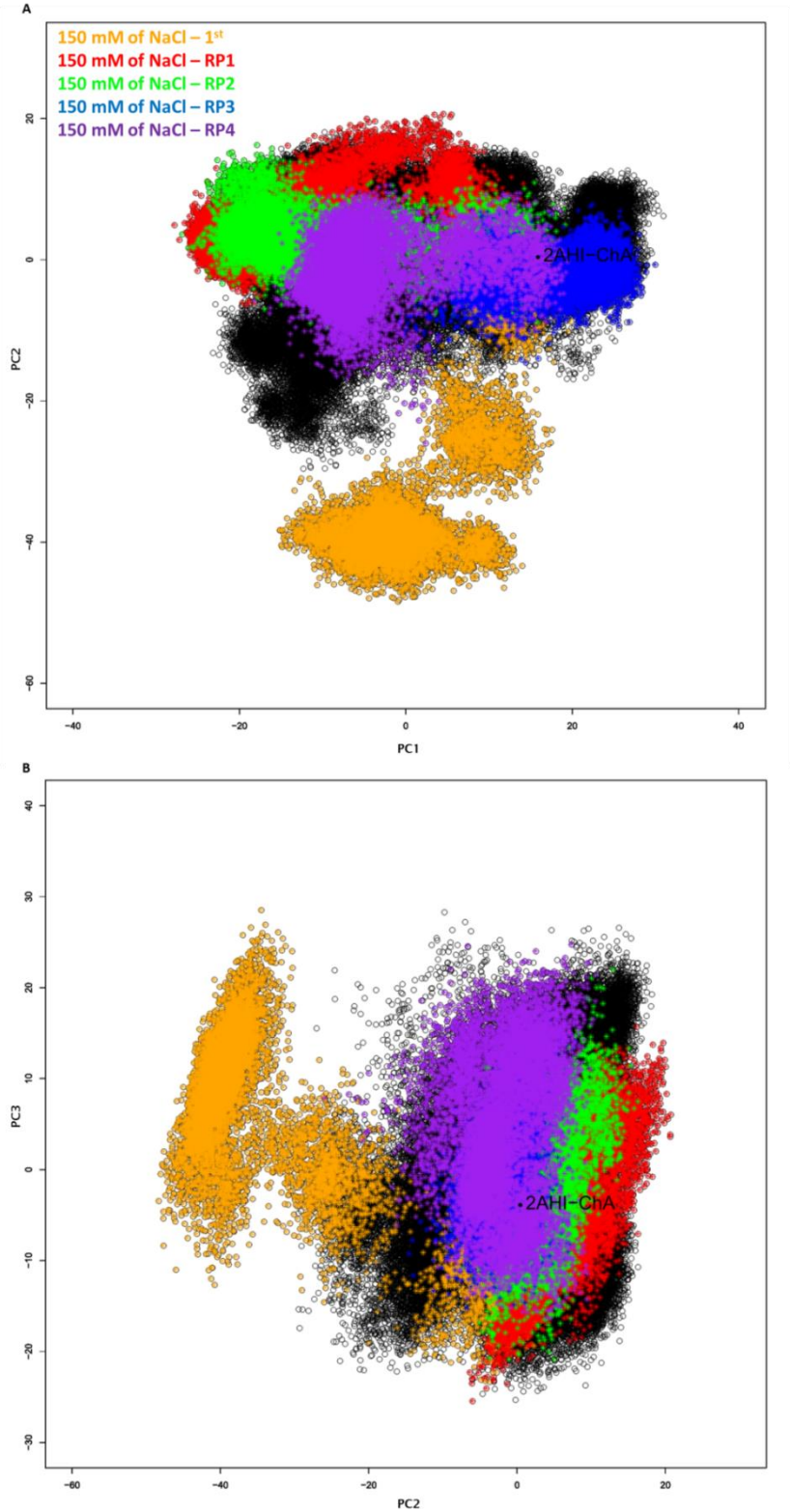


Figure 6-13: Structures projections along the three first PCs of the p53DBD V143A frames (only 150 mM of NaCl highlighted).
A) Along PC1 and PC2. **B)** Along PC2 and PC3. In this graph only p53DBD V143A at 150 mM of NaCl are coloured per simulation. The 150000 frames (50000 for each ion concentration condition) were aligned on the reference structure (2AHI chain A). The projected reference structure is labelled in black.

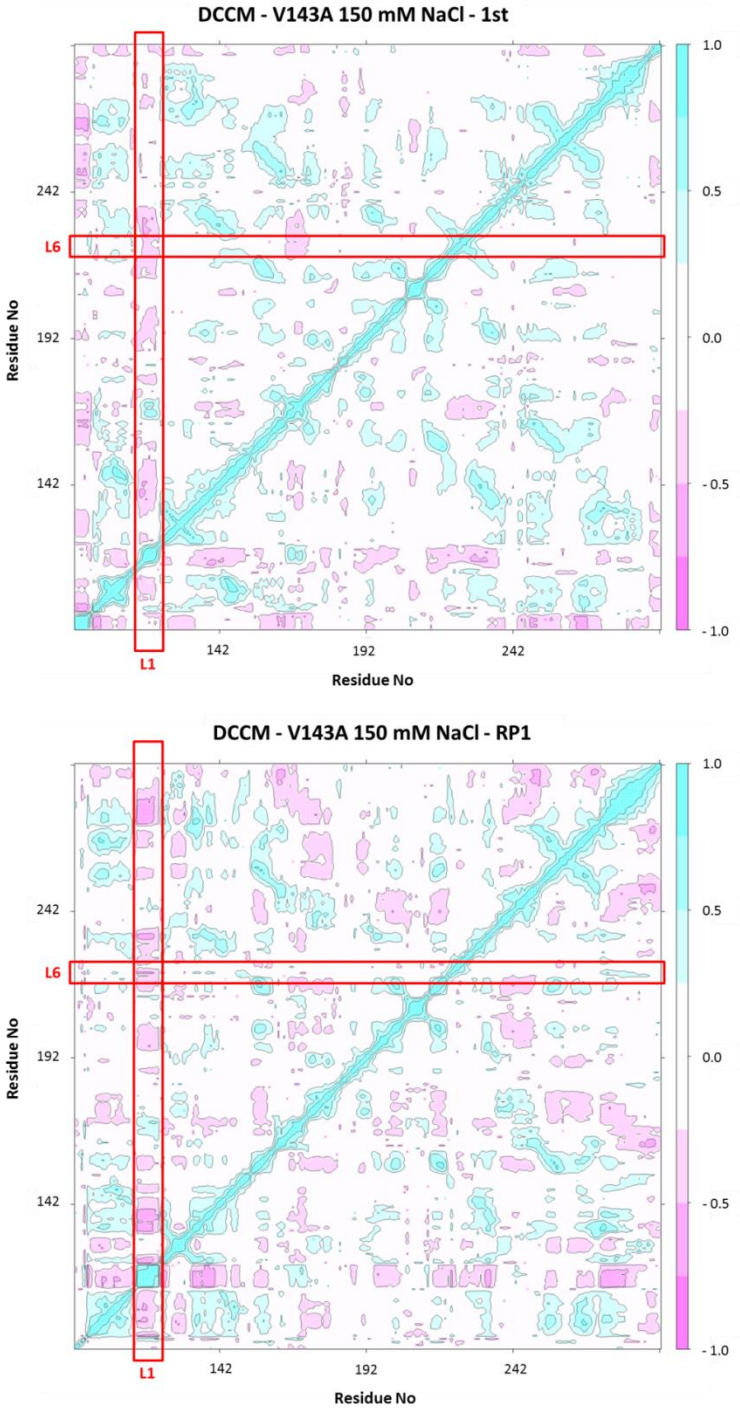


Figure 6-14: DCCM of the p53DBD V143A simulation at 150 mM of NaCl (1st and RP1).

The DCCM was performed on the 500 ns cMD simulations (10000 frames), in which only the α-C were kept. The frames were aligned on the reference structure (2AHI chain A), which also served as a reference for the DCCM. The correlated dynamic between atoms is indicated in blue and the anti-correlated motions are coloured in pink, see legend. The regions corresponding to L1 and L6 loops are red boxed.

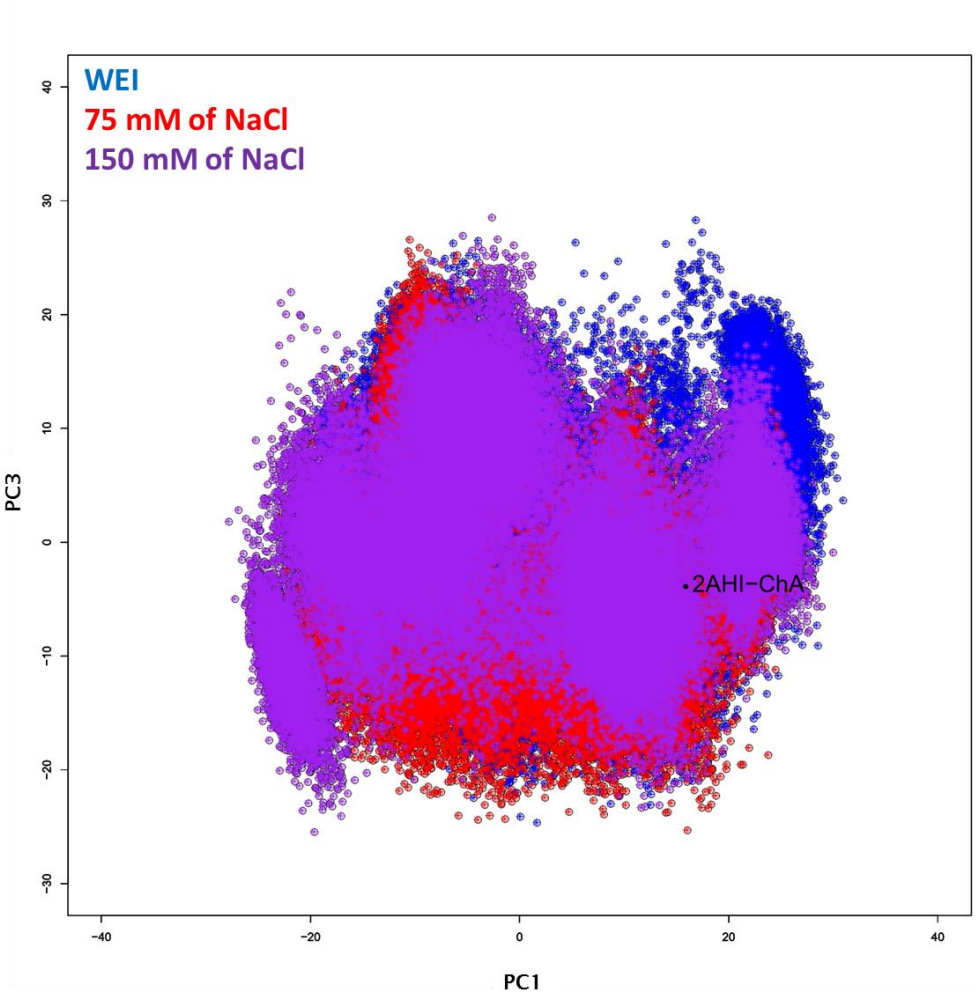


Figure 6-15: p53DBD V143A structure projection along PC1 and PC3.
The 150000 frames (50000 for each ion concentration condition) were aligned on the reference structure (2AHI chain A). The projected reference structure is labelled in black.

6.3.2 V143A mutation on the p53DBD: a temperature dependent effect

The influence of the temperature on the p53DBD mutant V143A was investigated. The WT and the mutant models (75 mM of NaCl) were simulated at different temperature: 298 K, 310 K and 323 K. At room temperature (298 K) both WT and mutant bind to DNA. The WT was shown to have a limited binding at 298 K (do not bind to the Bax promoter (93)) than at 310 K and V143A binds only to some promoters. At body temperature (310 K), the wild type is fully functional, while the mutant structure is destabilised and induces aggregation, and at 323 K, both WT and mutant are destabilised.

6.3.2.1 At permissive temperature: 298 K

As seen in section 6.3.1, at 298 K both WT and V143A mutant are stable (Fig. 6-1 B and 6-8 B). Most of the fluctuation observed in these two systems was located in the DBS and more specifically in the L1 loop (Fig 6-2, 6-3, 6-9 and 6-10) while their scaffold remained stable with a RMSD ≤ 1 Å.

The exploration of the systems' secondary structure (Fig. 6-16) showed that the H1 α-helix in the V143A is more stable and more frequent than in the WT (H1 α-helix present at 4 % in the WT and at 18 % in the V143A from residue P177 to R181). This helix is implicated in p53 dimerization while binding to the DNA (154). The stability of this helix in p53DBD V143A might be the cause of its functionality at low temperature. As seen in the cysteine alkylation study (Chapter 5), this H1 α-helix allows a stable p53 dimerization while binding to the DNA.

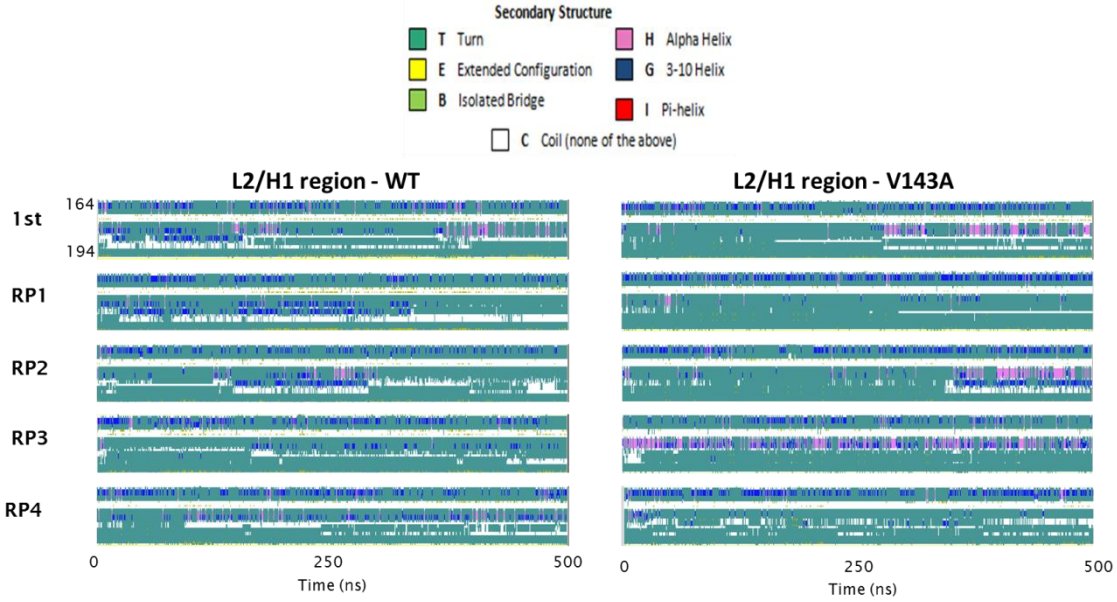


Figure 6-16: Determination of the L2/H1 region secondary structure in WT and V143A simulations.
The secondary structure analysis was performed using timeline analysis in VMD and only the L2/H1 region (from residue 164 to 194) is shown.

While the secondary structure determination highlighted a conformational difference between p53DBD WT and V143A, their PCA showed a similar phase space exploration at 298 K. The PCA performed on the trajectory combining the WT and mutant simulations captured 43.1 % of the variance within the 3 first PCs (20 PCs are necessary to present 82.6 % of the variance, see Fig. 6-

17). Most of the variance was displayed in the loops and the DBS: PC1and PC3 captured L1 loop deviation and PC2 showed variance in the overall DBS. The structure projections along these three first PCs showed that both the WT and V143A explore the same phase space and that p53DBD WT explore more phase space than the mutant. This PCA revealed the high flexibility of the WT compare to the mutant within the DBS, which might be due to the propensity of the V143A L2 loop to form a stable H1 α-helix.

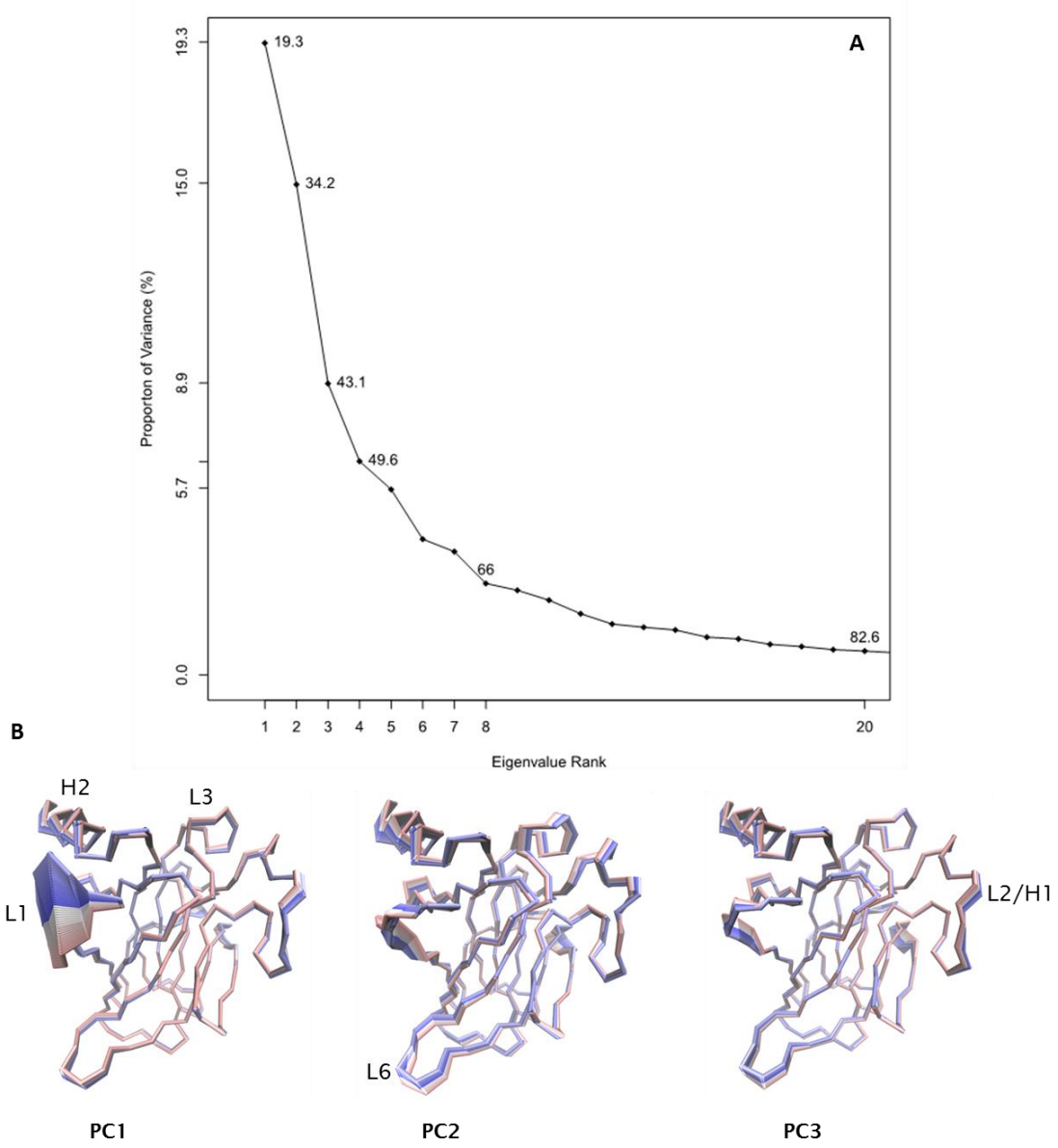


Figure 6-17: PCA of the p53DBD WT and V143A at 298K.
A) Proportion of variance (in %) captured by each principal component (PC) from a data set of 100000 p53DBD frames. The 3 first PCs contribute to 43.1 % of the variance of the dataset. **B)** Trace representation of the motion projections along the 3 first PCs (PC1, PC2 and PC3) of the 100000 frames using VMD. The colouring range of the motions goes from red, to white, to blue as function of time (timestep option on VMD), red being the early motion.

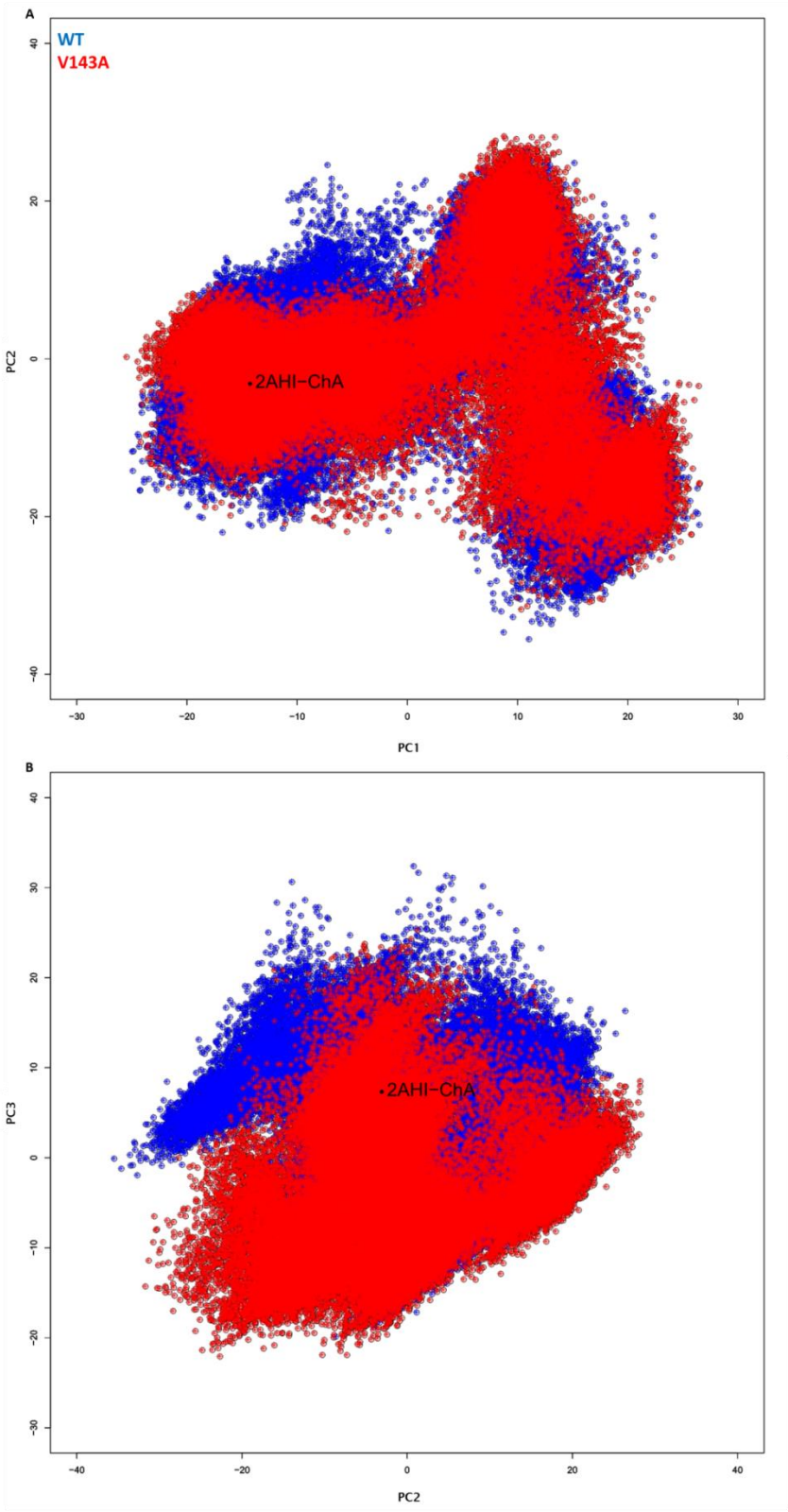


Figure 6-18: Structures projections along the three first PCs of the p53DBD WT and V143A simulations.
A) Along PC1 and PC2. **B)** Along PC2 and PC3. The 100000 frames (50000 per system) were aligned on the reference structure (2AHI chain A). The projected reference structure is labelled in black.

The effect of V143A on the DBS is thought to be caused by the cavity that the mutation creates in the p53DBD scaffold (78). This cavity was monitored in our simulations using *trj_cavity* (36). The reference structure (2AHI chain A) and the V143A mutant X-ray structure (protein data bank – PDB code: 2J1W) were analysed in the first instance to confirm the presence of this cavity in the V143A structure. 2J1W chain A and B were aligned on the reference structure and the cavity was calculated by giving *trj_cavity* seed coordinates (atom CG1 of V143, see Fig. 6-19) from which the software started searching for possible cavities. As expected no cavity was found in the WT structure (2AHI) and a cavity of 120.7 $Å^3$ was highlighted in the 2J1W structures (Fig. 6-20). *trj_cavity* was also run in our WT and V143A simulations at 298K. The analysis was performed in 500 frames (every 100 frames extracted from the simulations) of each system and the volume distribution is shown in figure 6-21. While the X-ray structure of the WT did not have a cavity, its simulation showed the presence of a cavity in the V143 region. This result was expected due to the fluctuation of the residues side chains that might from time to time be oriented in such a way that a cavity is located in this region. The volume distribution of the cavity showed more frames without cavity in the WT than in the mutant (~220 for the WT and ~140 for V143A) which explains the difference in average volume between the two systems (98.1 $Å^3$, SD = 130 $Å^3$ for the WT and 131.2 $Å^3$, SD = 143.4 $Å^3$). In the experimental study (89), it was hypothesised that the long-range effect of V143A on the DBS is due to the A143 surrounding residues filling the cavity. To verify this we calculated the average volume of the cavity over time (Fig. 6-22). The graphs of both WT and V143A simulation showed a constant volume of the cavity over time with a value of ~150 $Å^3$. These results contradict the experimental study; the V143A influence on the DBS was not due to the collapse of the surrounding residues in the cavity created by A143.

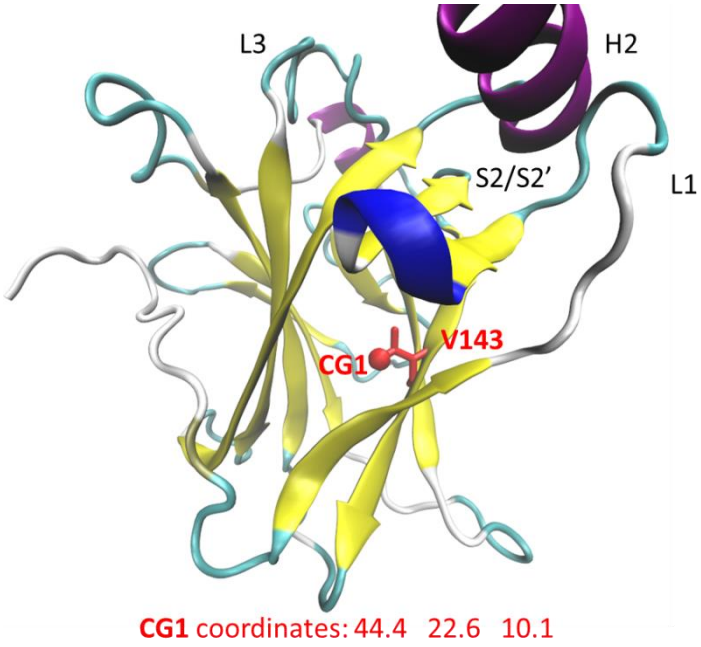


Figure 6-19: Seed coordinates used for the cavity search.
The reference structure (2AHI chain A) is represented in cartoon and coloured by structure and V143 is in liquorice and coloured in red. The seed coordinates (in Å) correspond to the atom CG1 of the V143.

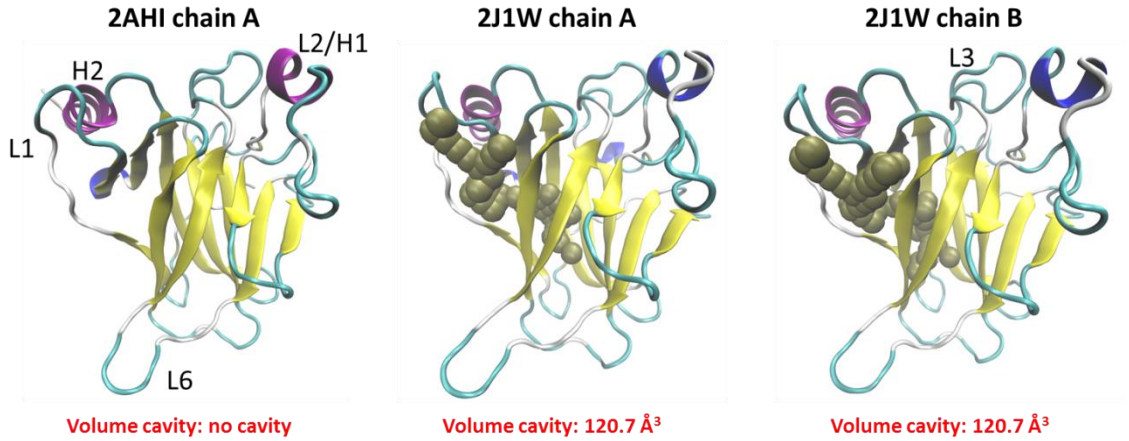


Figure 6-20: VMD visualisation of the cavities in the WT and V143A X-ray structures.
The structures are represented in cartoon and coloured by structure and the cavity is represented by VDW and coloured by name. The volume of the cavities is indicated in red.

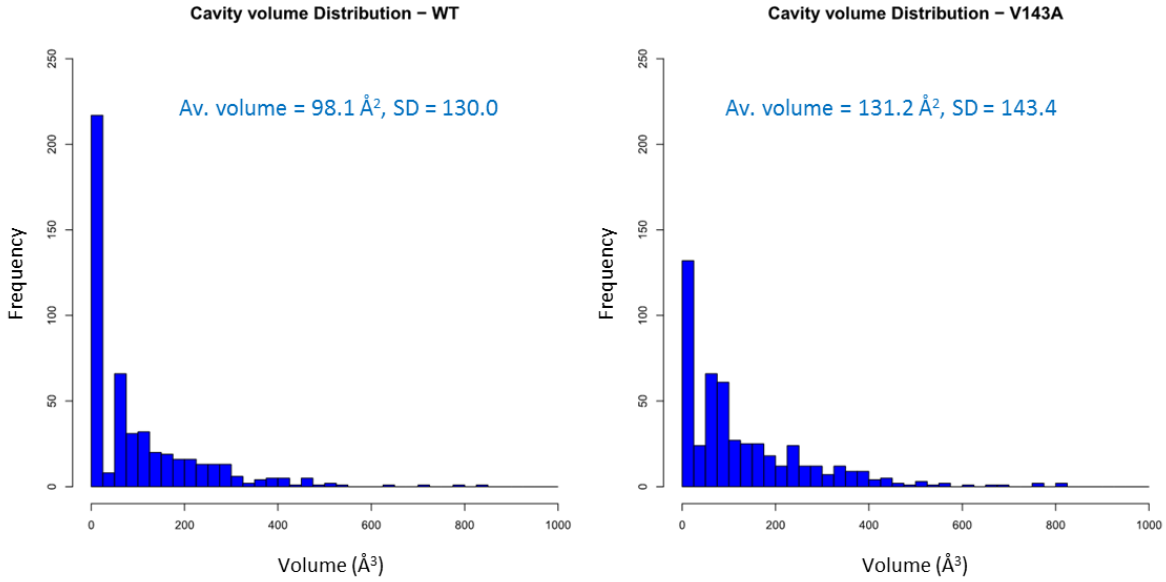


Figure 6-21: *trj_cavity* analysis, volume distribution for the WT and V143A at 298 K.

The simulations were aligned on the reference structure (2AHI chain A) and the volume of the cavity is calculated using *trj_cavity* from seed coordinates (atom CG1 of the V143 in the reference structure). The average volume and the standard deviation (SD) over all the simulation are shown in blue.

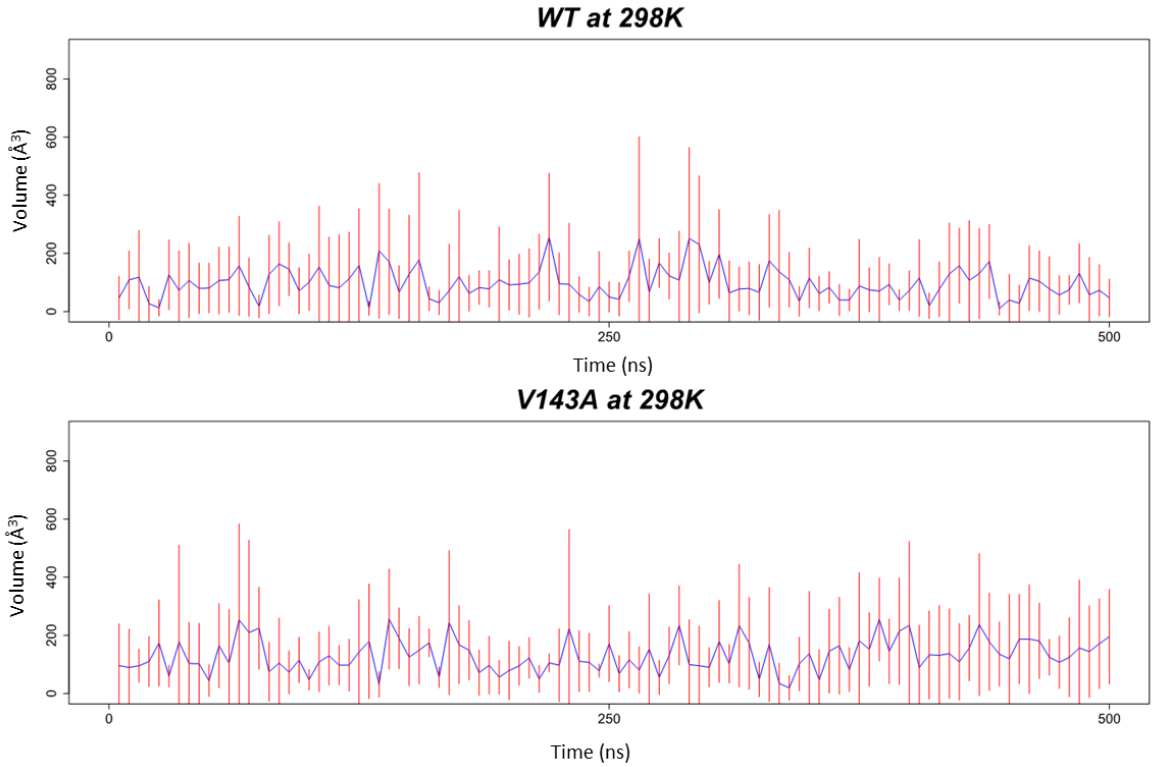


Figure 6-22: Average volume of the cavity created at the CG1 atom of V143A over time.
The simulations were aligned on the reference structure (2AHI chain A) and the volume of the cavity is calculated using *trj_cavity* from seed coordinates (atom CG1 of the V143 in the reference structure). The average volume (blue) and standard deviation (red) were calculated using R.

6.3.2.2 At denaturing temperature: 310 K and 323 K

At 310 K, V143A is estimated to be about 80% denatured and p53DBD WT is marginally stable (half-life of 9 min) (89, 90, 178), while at 323 K both systems are expected to be destabilised. The five cMD simulations run for each systems showed the stability of both systems with V143A been more stable than the WT.

The full structure RMSD, although presenting some fluctuations on the DBS and the L6 loop, showed that the DBD structure is conserved along the WT and V143A simulations at both temperature (310 K and 323 K, see Fig. 6-23 and 6-24). RMSD of the DBS/scaffold and the RMSF confirmed the p53DBD conservation. Most of the fluctuation was located in the DBS and the scaffold kept a RMSD ≤ 1.5 Å (Fig. 6-25 and 6-26). Unlike the simulations at 298 K (Fig 6-3 and 6-10), in which most of the DBS fluctuation was due to the L1 loop, at 310 K and 323 K, the flexibility increased in the three DBS loops (L1, L2 and

L3). In the V143A system, while the flexibility of its L2 loop increased its ability to for the H1 α-helix was conserved as can be seen in the VMD visualisation (Fig. 6-23 B and 6-24 B).

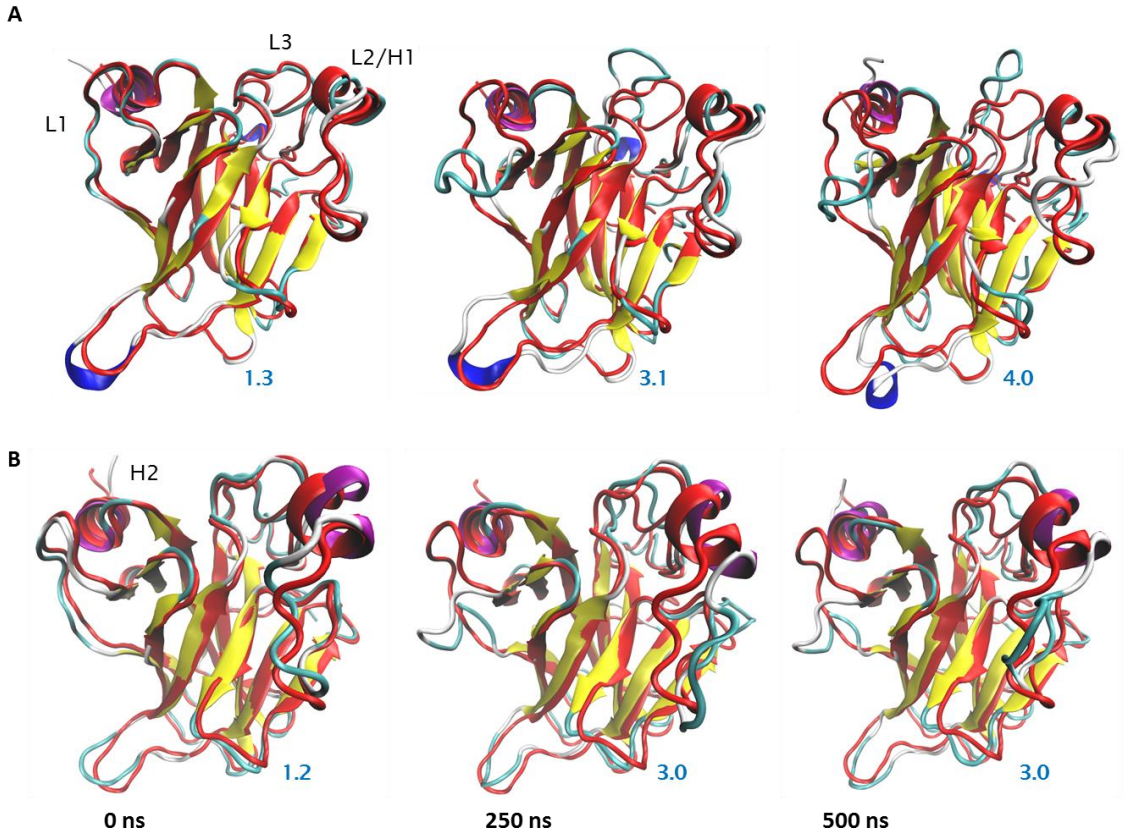


Figure 6-23: VMD visualisation of the p53DBD wild-type and V143A systems at 310 K.
The systems were aligned on 2AHI chain A (reference structure) alpha-carbon at different time of the simulations. Here is shown only the first simulation. **A)** WT, **B)** V143A. The structures are represented in cartoon, the simulation frames were coloured by structure and the reference structure was coloured in red. The RMSD (in Å) of the frame with respect to the reference structure is indicated in blue.

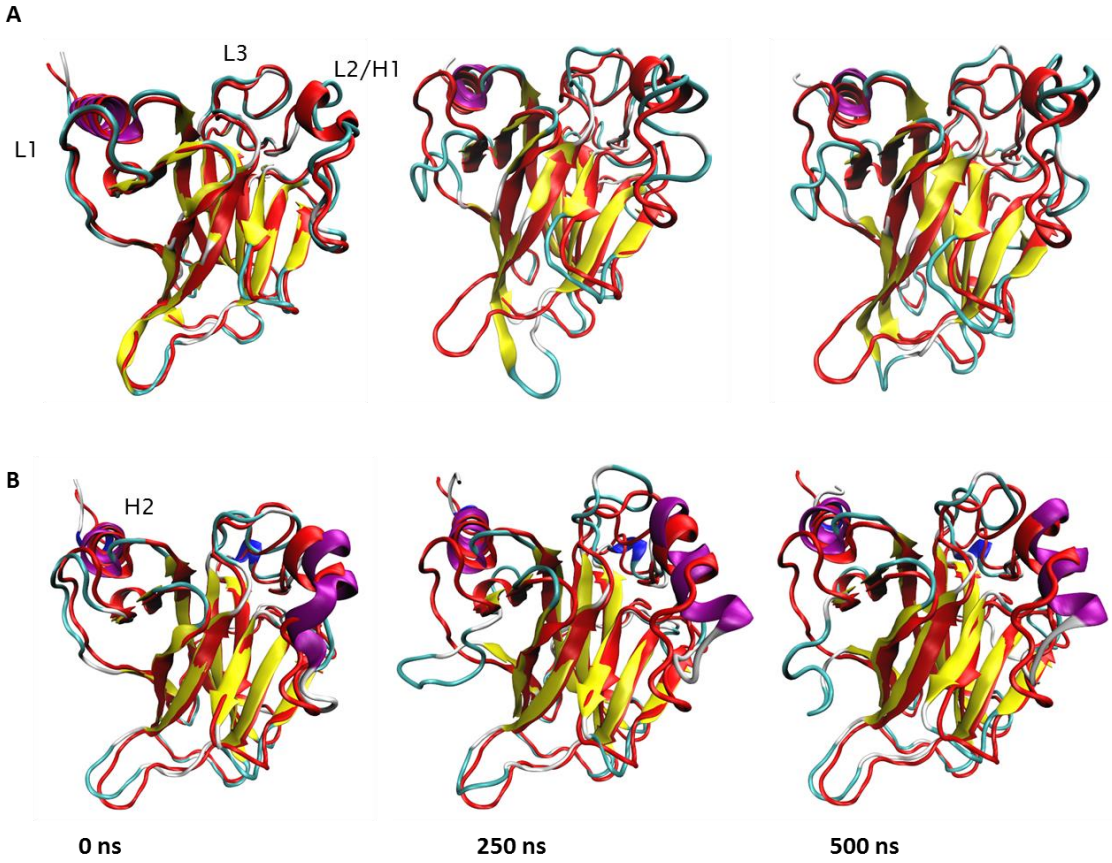


Figure 6-24: VMD visualisation of the p53DBD wild-type and V143A systems at 323 K.
The systems were aligned on 2AHI chain A (reference structure) alpha-carbon at different time of the simulations. Here is shown only the first simulation. **A)** WT, **B)** V143A. The structures are represented in cartoon, the simulation frames were coloured by structure and the reference structure was coloured in red. The RMSD (in Å) of the frame with respect to the reference structure is indicated in blue.

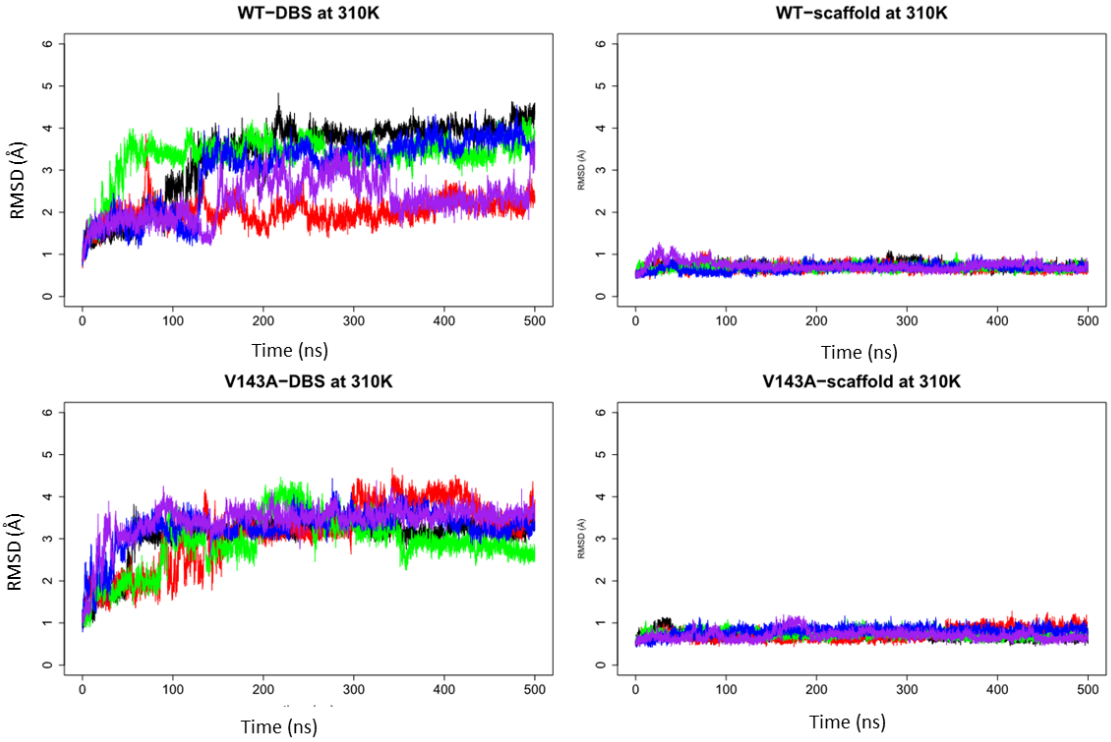


Figure 6-25: RMSD of the p53DBD wild-type and V143A systems at 310 K. RMSD of the DNA binding surface (DBS) against the immunoglobulin-like β-sandwich (scaffold) was calculated from the simulations and each repeat after alignment on the reference DBS or scaffold (2AHI chain A).

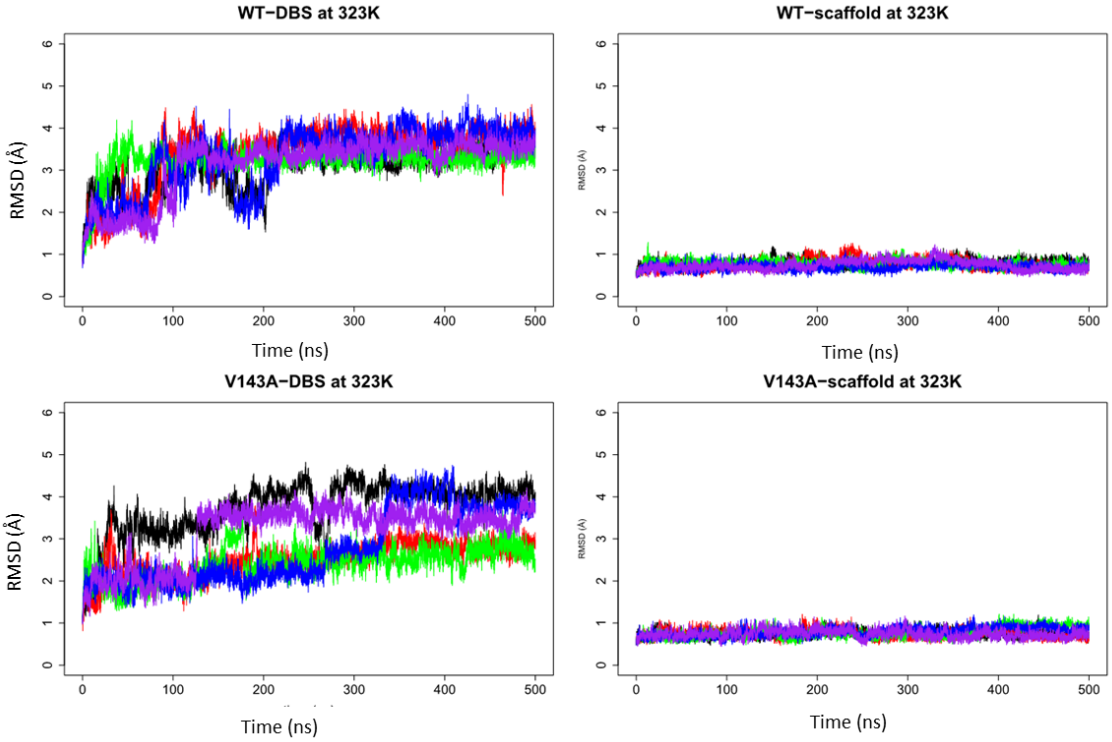


Figure 6-26: RMSD of the p53DBD wild-type and V143A systems at 323 K. RMSD of the DNA binding surface (DBS) against the immunoglobulin-like β-sandwich (scaffold) was calculated from the simulations and each repeat after alignment on the reference DBS or scaffold (2AHI chain A).

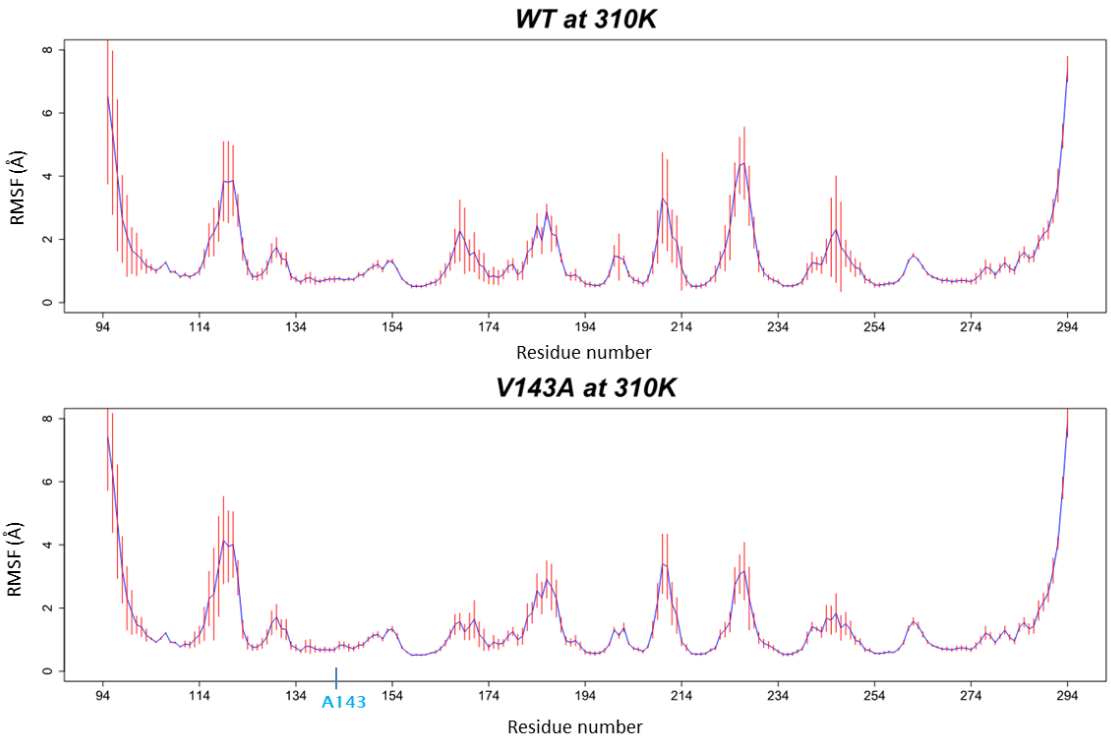


Figure 6-27: RMSF of the p53DBD wild-type and V143A systems at 310 K. The RMSF for each 500 ns simulation was determined using cpptraj in Amber. The average RMSF (blue) among the five cMD and the standard deviation (red) for each system were calculated using R. The mutated residue is labelled in blue.

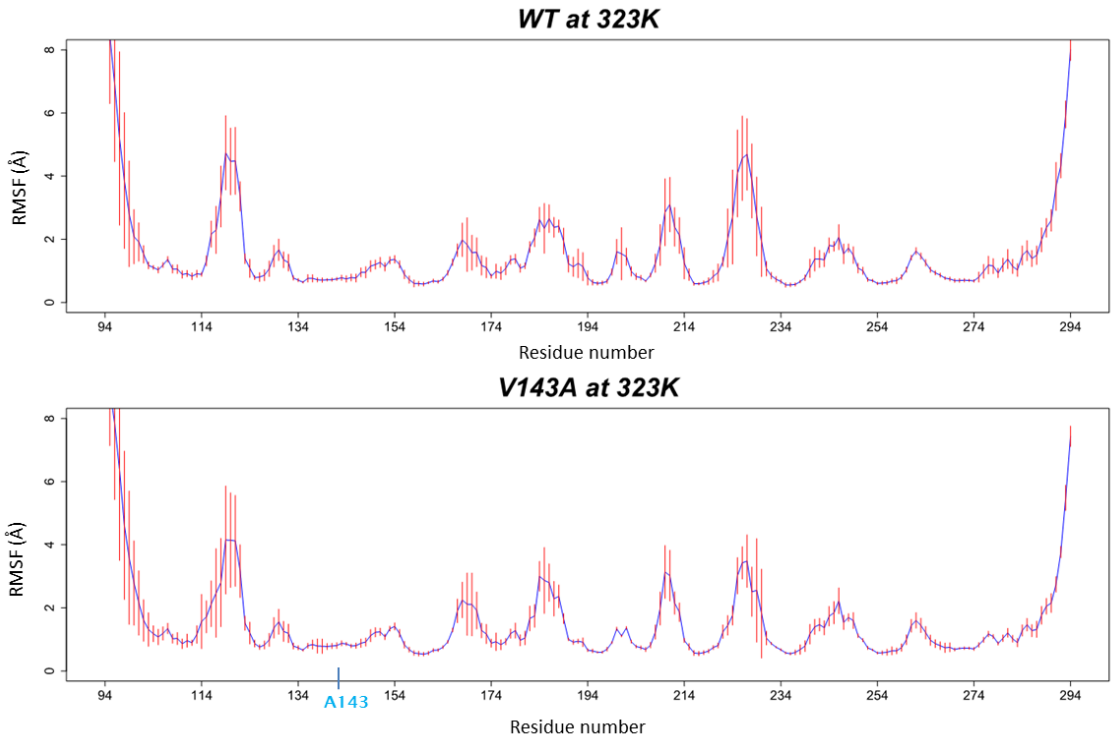


Figure 6-28: RMSF of the p53DBD wild-type and V143A systems at 323 K. The RMSF for each 500 ns simulation was determined using cpptraj in Amber. The average RMSF (blue) among the five cMD and the standard deviation (red) for each system were calculated using R. The mutated residue is labelled in blue.

This finding was confirmed by the secondary structure determination of the models (Fig. 6-29 and 6-30). While the L1 loop kept a loop conformation along the simulations in the WT and V143A, the L2/H1 region of the V143A showed that the H1 α-helix is restored in most of the simulation at both temperatures. On the other hand the H1 α-helix in the WT, although present in a couple of simulations, was not stable along the cMD trajectory. These results are similar to the 298 K simulations. Instead of been denatured (at 310 K for the V143A and 323 K for the WT), both simulations were found to be stable, and V143A was shown to restore and stabilise H1 α-helix.

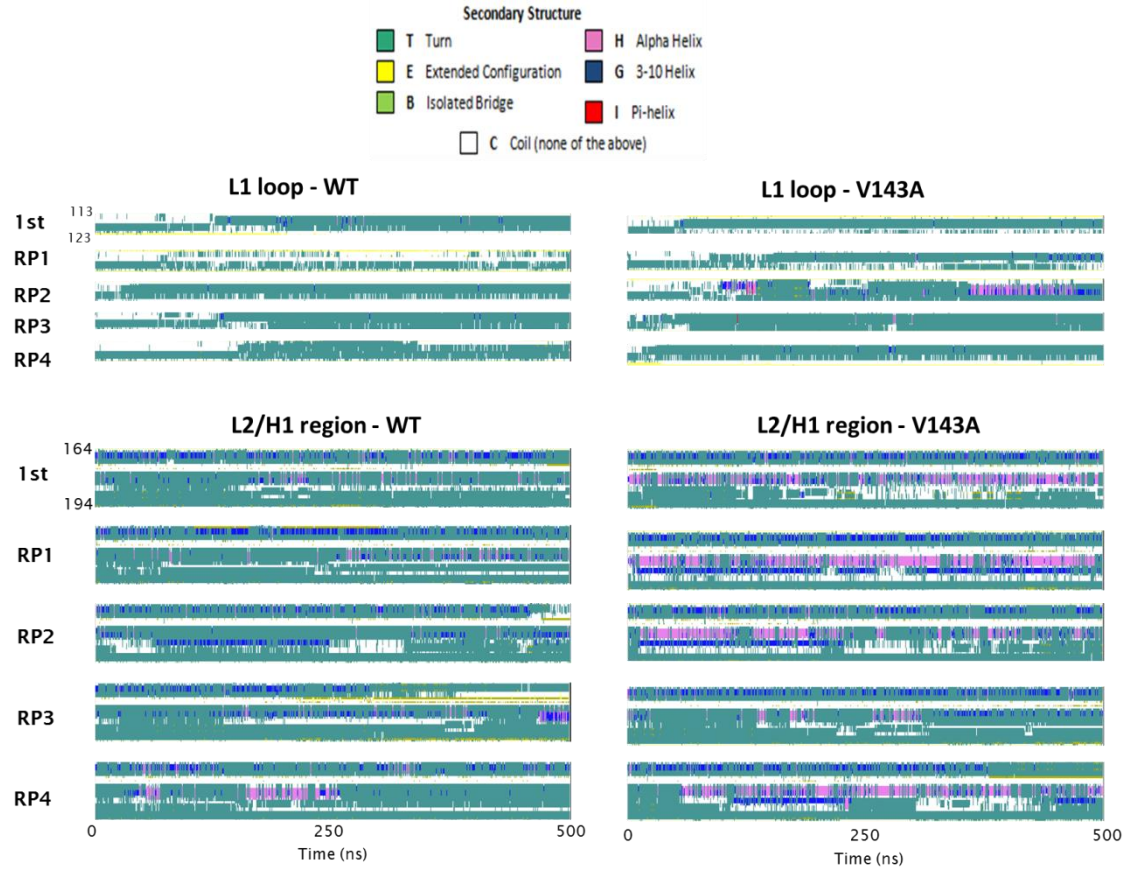


Figure 6-29: Determination of the L1 loop and the L2/H1 region secondary structure in WT and V143A simulations at 310 K.

The secondary structure analysis was performed using timeline analysis in VMD and only the L1 loop (from residue 113 to 123) and L2/H1 region (from residue 164 to 194) is shown.

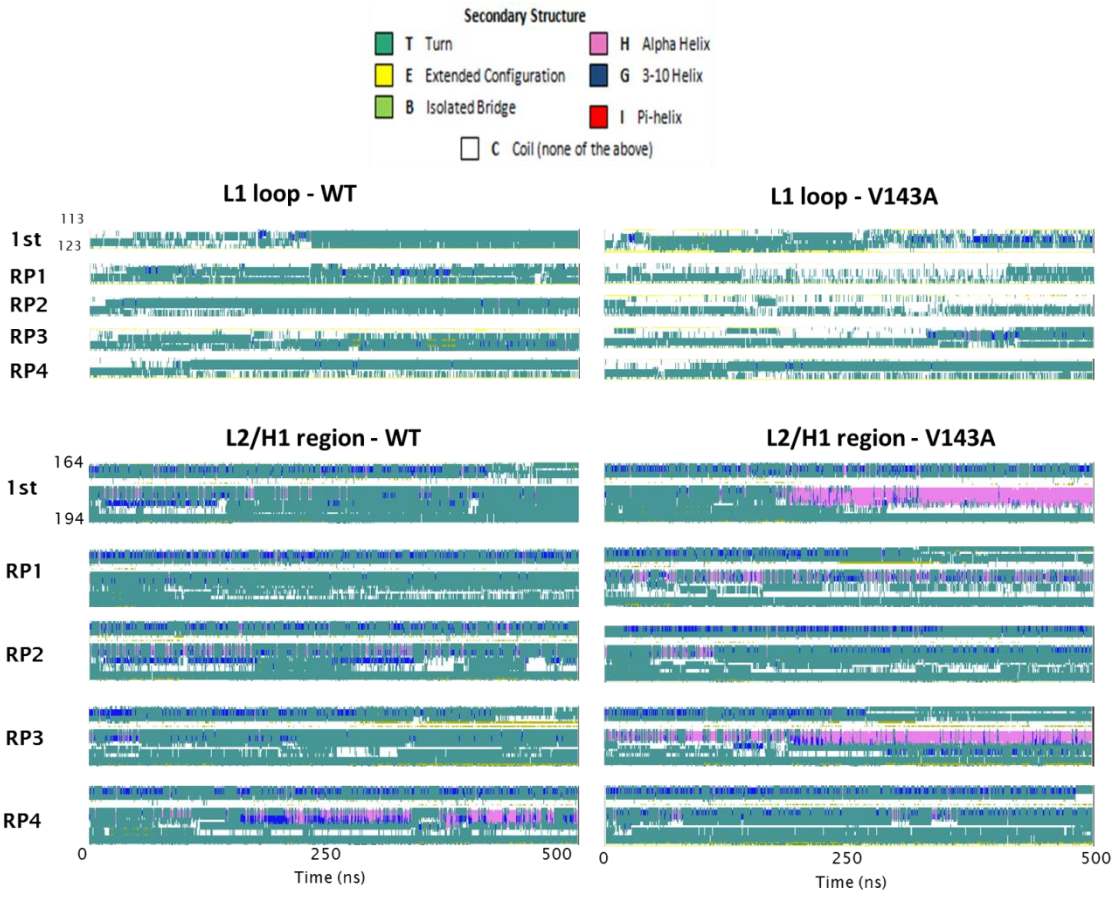


Figure 6-30: Determination of the L1 loop and the L2/H1 region secondary structure in WT and V143A simulations at 323 K.
The secondary structure analysis was performed using timeline analysis in VMD and only the L1 loop (from residue 113 to 123) and L2/H1 region (from residue 164 to 194) is shown.

The PCA analysis of the two systems was performed to check the phase space explored by each simulation at different temperature (at 298 K, 310 K and 323 K). The simulations from different temperature were merged and aligned on the α-carbon atoms of the reference structure. The PCA was performed on the α-carbon of each system, WT and V143A. The three first PCs of the WT PCA captured 40.1 % of variance (20 PCs are needed to capture 81.6 % of the variance, see Fig 6-31 A). PC1 and PC2 highlighted fluctuation in the loops and the turns and on the H2 α-helix for PC2 and PC3 showed mainly fluctuation of the loops (L1, L2 and L3) (Fig. 6-31 B). The projection of the structure along PC1 and PC2 (Fig. 6-32 A) showed that the WT at 298 K, 310 K and 323 K explore the same phase space. The simulation at 323 K resides less in the central region compare to the two other simulations which is in agreement with our previous results showing the increase of temperature increases the

p53DBD fluctuation. On the other hand, the structure projection along PC2 and PC3 (Fig. 6-32 B) showed that the p53DBD WT system at 310 K explores a region not visited by the two other systems. This difference sits along PC3 which captures mainly fluctuation in the L3 loop. This loop fluctuation, although observed in the other systems, has been shown to be more flexible only in the p53DBD WT at 310 K. This is mainly due to the first simulation, in which the M246 buried in p53DBD WT, flipped out of the structure (Fig. 6-33). M246 is thought to participate to the hydrophobic packing with the scaffold of p53, to stabilise the L3 loop and allow p53 dimerization (121–123) and to bind to other proteins (124–126). In the literature this L3 loop motion has been associated to the LTag antigen binding (cancer related) and to a hot spot p53 mutant (section 4.3.1.2) (106, 120, 127). While this L3 loop fluctuation might highlight a beginning of p53DBD destabilisation, its presence in only one repeat, did not allow us to draw any conclusion.

As at 298K, the PCAs comparing WT and V143A at 310 and 323 K (capturing mainly fluctuation in the DBS loops) also showed that V143A explores less phase space than the WT (App. C Fig. 1 to 4), due to the H1 α-helix in V143A. In the VMD visualisation of the p53DBD V143A at 323 K (Fig. 6-24 B) it was shown that the H1 α-helix is larger than the reference structure. The H1 α-helix in V143A DBS tends to be more stable as the temperature was increased.

Although meaningful structure changes are observed on the p53DBD V143A and the WT, the systems did not show denaturation. Such large scale motions (folding and unfolding) on biological molecules occur over 10^{-7} to 10^{4} s, and p53DBD WT was shown to denature after 9 min (178). The time scale used in our study being too short to observe p53DBD unfolding, aMD, an enhanced sampling method was used. This method applies a function to bias the simulation sampling to induce a better exploration of the free energy surface and allowed us to gather more conformations (26).

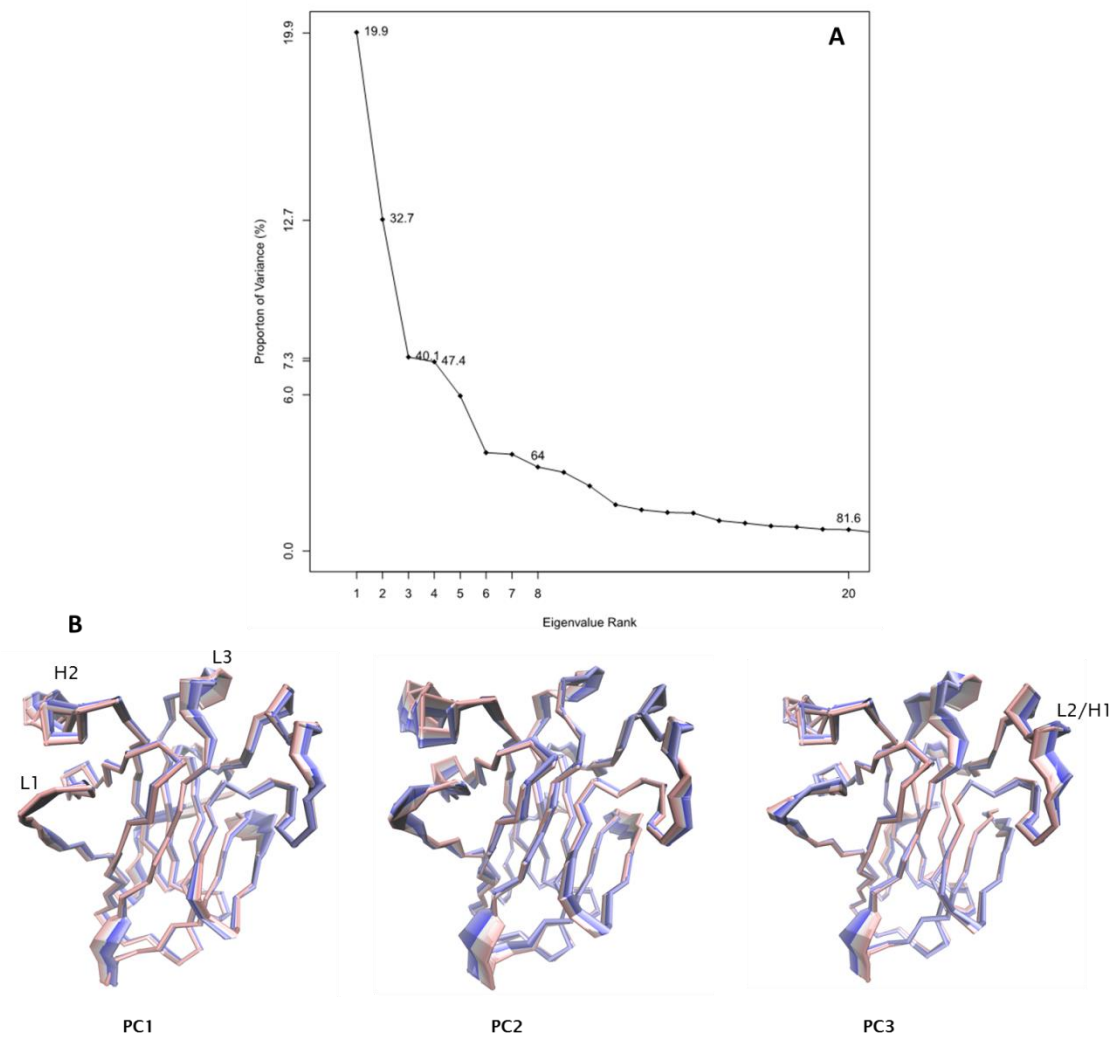


Figure 6-31: PCA of the p53DBD WT at different temperature.
A) Proportion of variance (in %) captured by each principal component (PC) from a data set of 150000 p53DBD frames. The 3 first PCs contribute to 40.1 % of the variance of the dataset. **B)** Trace representation of the motion projections along the 3 first PCs (PC1, PC2 and PC3) of the 150000 frames using VMD. The colouring range of the motions goes from red, to white, to blue as function of time (timestep option on VMD), red being the early motion.

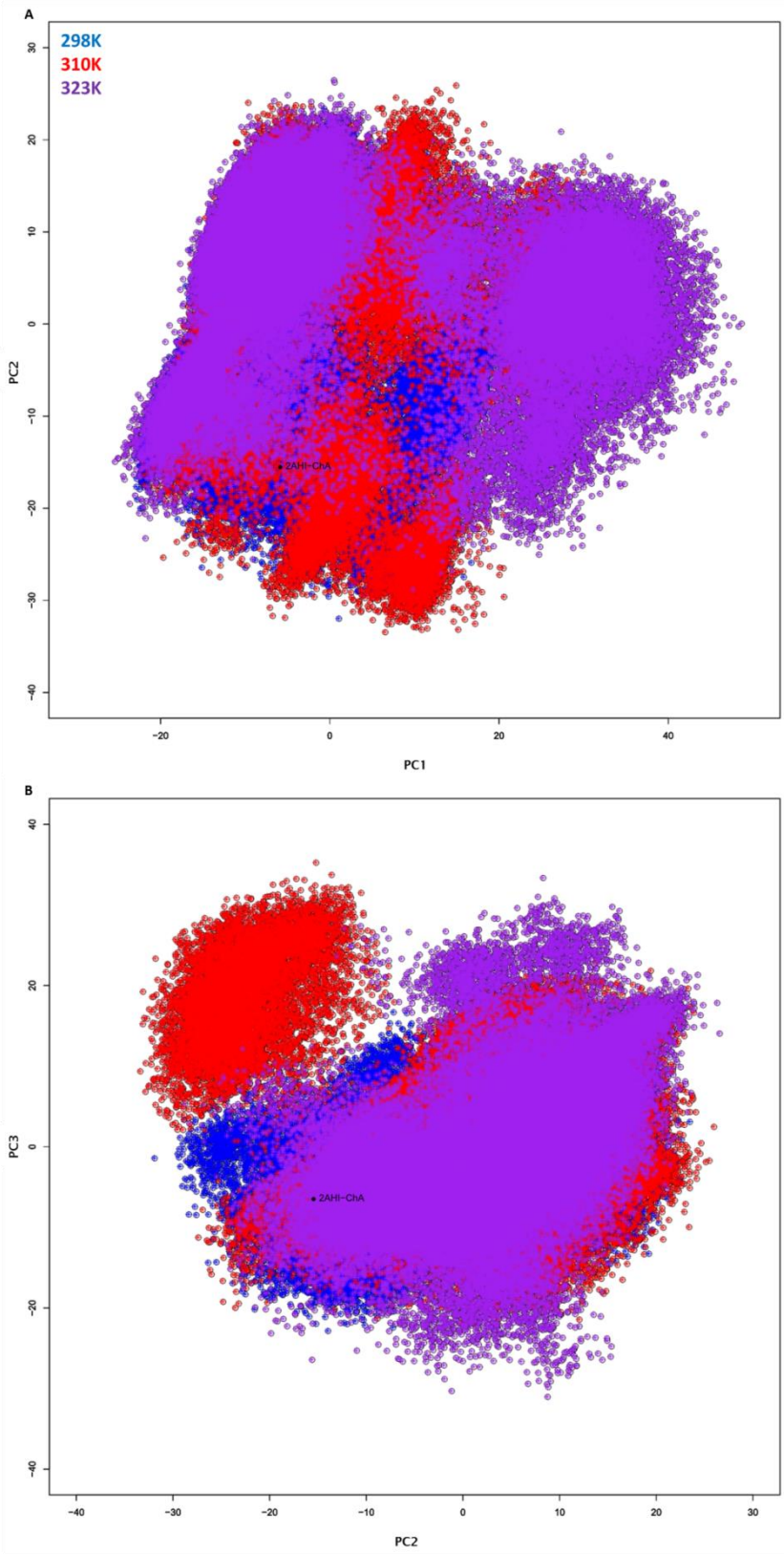


Figure 6-32: Structures projections along the three first PCs of the p53DBD WT at different temperature.
A) Along PC1 and PC2. **B)** Along PC2 and PC3. The 150000 frames (50000 per system) were aligned on the reference structure (2AHI chain A). The projected reference structure is labelled in black.

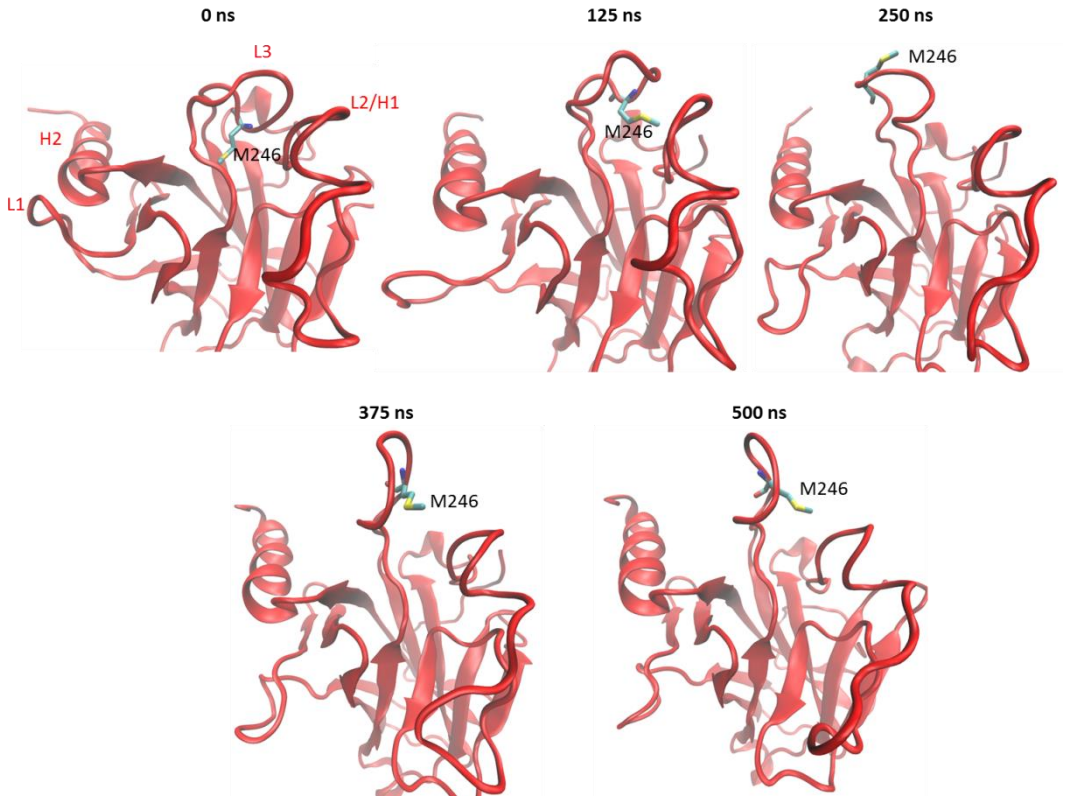


Figure 6-33: L3 loop flexibility due to M246 motion.
p53DBD WT at 310 K (1st simulation) is represented in cartoon and coloured in red and the M243 is in liquorice end coloured by name.

6.3.2.3 Enhanced sampling: aMD

aMD has been shown to explore more phase space in 500 ns than 1 μs of cMD simulation of the bovine pancreatic trypsin inhibitor (BPTI) (26). The use of aMD on the p53DBD WT and V143A showed the same trend. The WT simulation at 298 K cMD and aMD were merged, aligned on the reference structure and a PCA was performed on the α-carbon coordinates (Fig. 6-35). The first three PCs captured 42.4 % of the variance (20 PCs are needed to capture 81.8 %, See Fig.6-35 A). PC1 captured mainly fluctuation L1 and L6 loop, PC2 highlighted L1, L3 and L6 loops variance while PC3 displayed variance on L2 and L6 loops (Fig. 6-35 B). The structure projection of p53DBD WT at 298 K confirmed that the aMD simulations explored more phase space than the WT cMD (Fig. 6-36) and different region are explored along PC2 (capturing mainly fluctuation of the L1, L2 and L3 loops within the DBS).

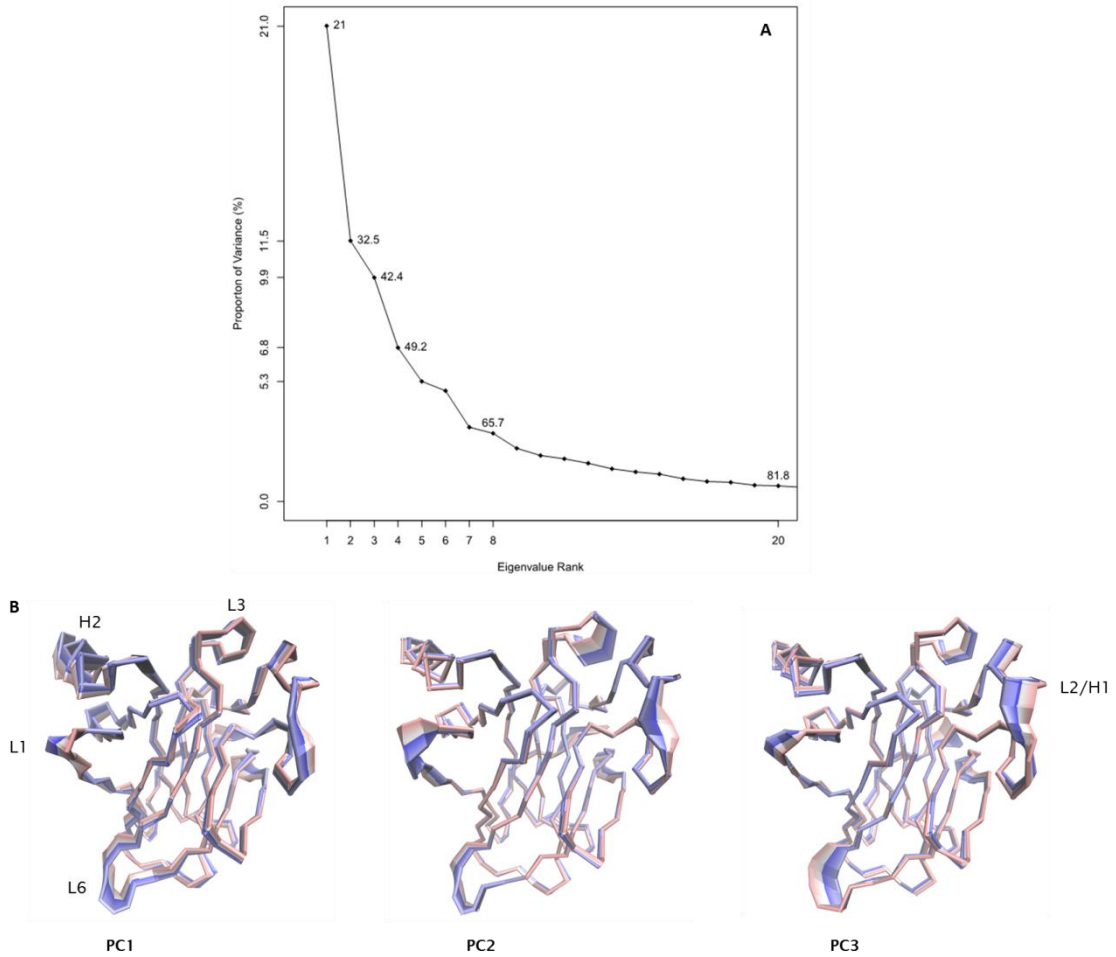


Figure 6-34: PCA of the p53DBD WT at 298K, cMD and aMD simulations. **A)** Proportion of variance (in %) captured by each principal component (PC) from a data set of 80000 p53DBD frames. The 3 first PCs contribute to 42.4% of the variance of the dataset. **B)** Trace representation of the motion projections along the 3 first PCs (PC1, PC2 and PC3) of the 80000 frames using VMD. The colouring range of the motions goes from red, to white, to blue as function of time (timestep option on VMD), red being the early motion.

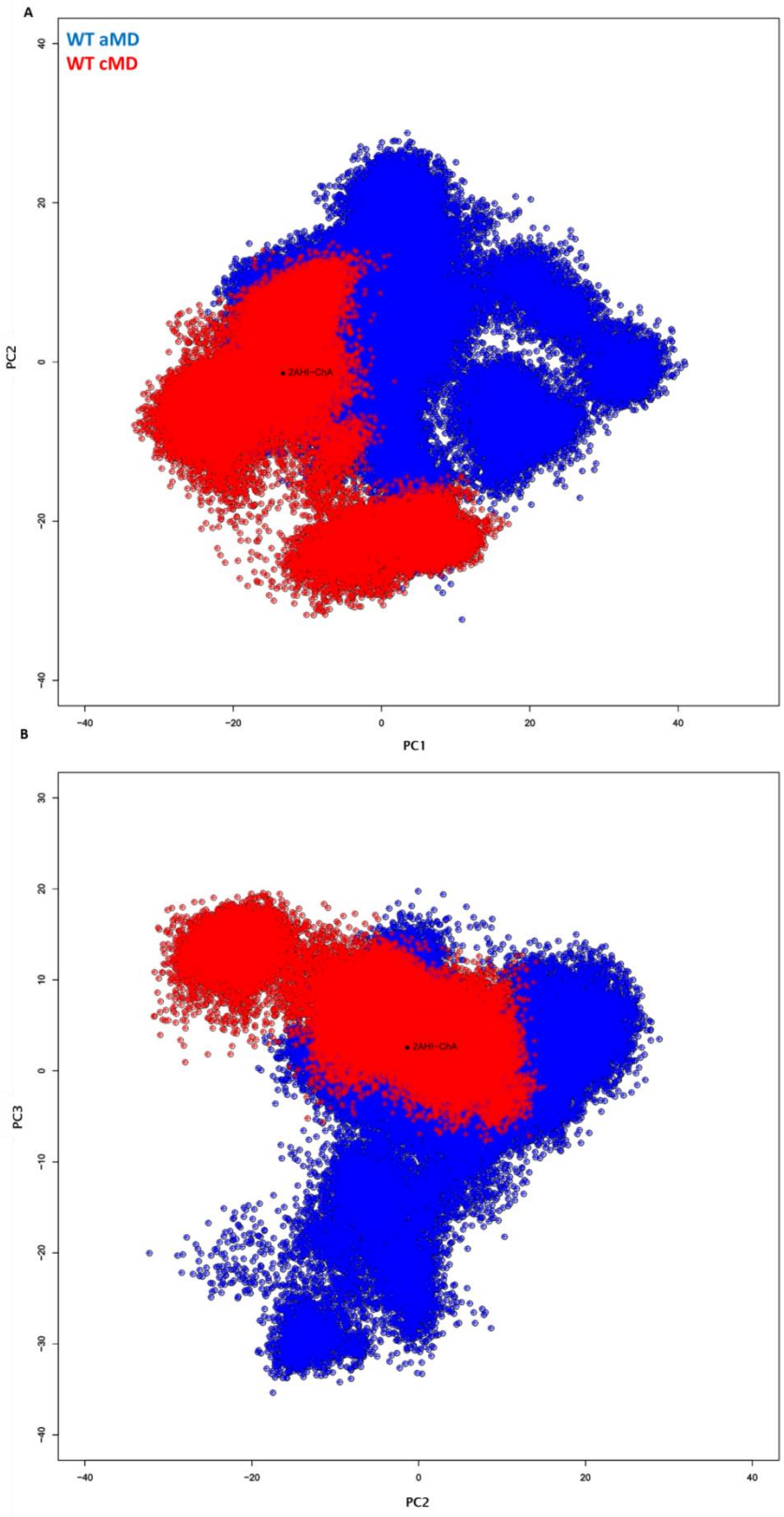


Figure 6-35: Structure projections along the three first PCs of the p53DBD WT at 298 K, cMD and aMD simulations.
A) Along PC1 and PC2. **B)** Along PC2 and PC3. The 80000 frames (40000 per system) were aligned on the reference structure (2AHI chain A). The projected reference structure is labelled in black.

A secondary structure determination of the p53DBD WT (aMD simulation at 298K, see Fig. 6-37) showed conformational changes in the L1, L2 and L3 loops of the aMD simulations compared to the cMD. While aMD tended to restaure and stabilise the H1 α-helix (p53 dimer interface) within the L2 loop compared to the cMD (Fig. 6-16), it also induced α-helices to form within L1 and L3 loops (responsible for p53 DNA binding inhibition - L1, see chapter 5 and in L3 due to the R249S mutant (127)). This propensy of the loops (mainly L1, L2 and L3) to for α-helix was obersed in all the aMD simulations (WT and V143A at different temperature, see App. C Fig. 5 and 6).

Amber force fields (mainly ff94 and ff99) were shown to induce helices within sequences that does not present one experimentally. Although ff99SB (used in this study) does not suffer of helical bias (143), the use of this force field combined to aMD seemed to reintroduce the helical errors. Knowing this, the helix formation within the aMD simulation will be ignored.

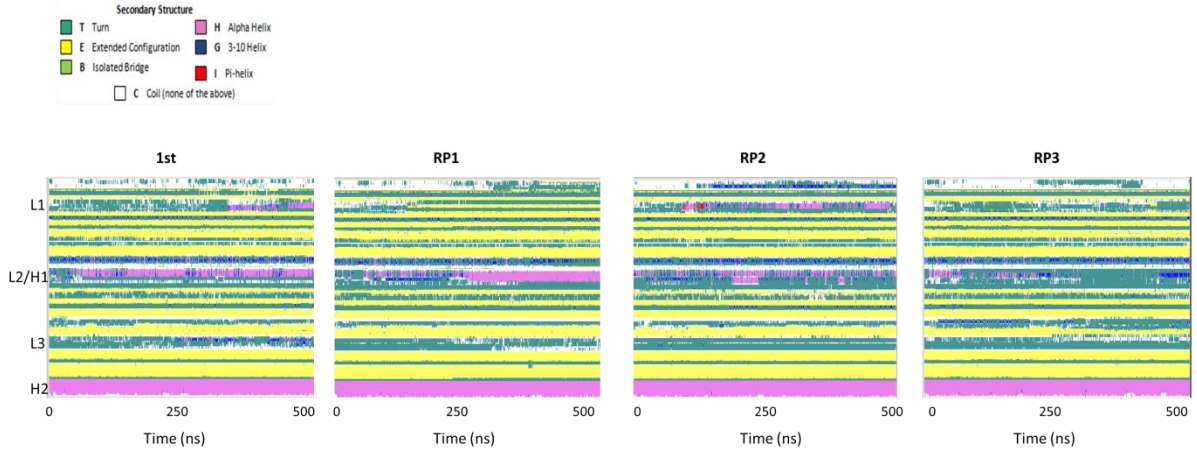


Figure 6-36: Determination of the secondary structure in p53DBD WT at 298 K (aMD simulations).
The secondary structure analysis was performed using timeline analysis in VMD

The RMSD analysis of the DBS/scaffold carried out on the aMD simulations exhibited stability of the WT scaffold along the simulation and at different temperatures (~ 1.5 Å Fig. 6-38, as cMD simulations), while the V143A showed slightly more fluctuations within its scaffold (the repeat 2 at 323 K reached ~ 2 Å, Fig. 6-38). V143A scaffold displayed more fluctuation as the temperature increased. Unlike cMD, aMD of the WT and V143A at different temperature highlighted difference in stability between the WT and the mutant and that the stability of the mutant decrease while the temperature increases.

The aMD simulations presented fluctuation of the loops, turns and the Nter region. The RMSF analysis of the simulations exhibited more fluctuations in the loops than for the cMD simulations in both WT and V143A. This is mainly observed in the L2/H1 region and the L3 loop (Fig. 6-39 and 40), which confirms the cMD/aMD PCA results (Fig. 6-35 and 36) showing that most of the differences between these two types of simulations were located in the L2 and L3 loops. While the RMSF of the L2 and L3 loops were of ~ 3 Å in the cMD for WT and mutant in the aMD simulations reached ~ 5 Å in the aMD simulations at 323 K for both WT and V143A.

The aMD RMSF comparison of the WT and V143A (Fig. 6-39 and 40), emphasised mainly differences at 298K between the two systems. At this temperature, the WT displayed less DBS loop fluctuations (L1, L2 and L3) than the mutant. Indeed in the mutant all the loops and turns as well as the Nter region, presented a high flexibility at all temperature (Fig. 6-40), while the WT aMD simulation at 298 K (Fig. 6-39) showed a RMSF similar to the WT cMD simulations at different temperatures (Fig. 6-10, at 75 mM, 6-27 and 6-28). The difference in flexibility observed within the WT DBS at 298K aMD (compared to the other temperature and the V143A) might be the cause of its limited function at low temperature reported in literature (no binding to Bax promoter at 298 K) (93). This promoter was shown in section 3.5.4, to contain only one well-defined pentamer surrounded by three imperfect pentamers. One of the imperfect pentamer is located at 5 bp from the well-defined one (overlined in Fig. 3-11, section 3.5.4). The restricted flexibility of the WT DBS at 298K in the aMD simulations (compared to the WT at 310K) might explain the WT inability to bind the Bax promoter at this temperature.

Within the V143A aMD simulations, while all the DBS loops (L1, L2 and L3), turns and the Nter region showed more flexibility than the cMD simulations a main difference was noticed within the L2/H1 region at different temperature. Indeed this region showed less fluctuation at 298 K (V143A functional temperature) than at high temperatures (310 and 323 K, V143A denatured temperatures). These results supported the cMD findings that V143A tend to have a more stable H1 α-helix at low temperature. The stability of this helix (dimerization interface (154)) might suffice to restore V143A function at low temperature.

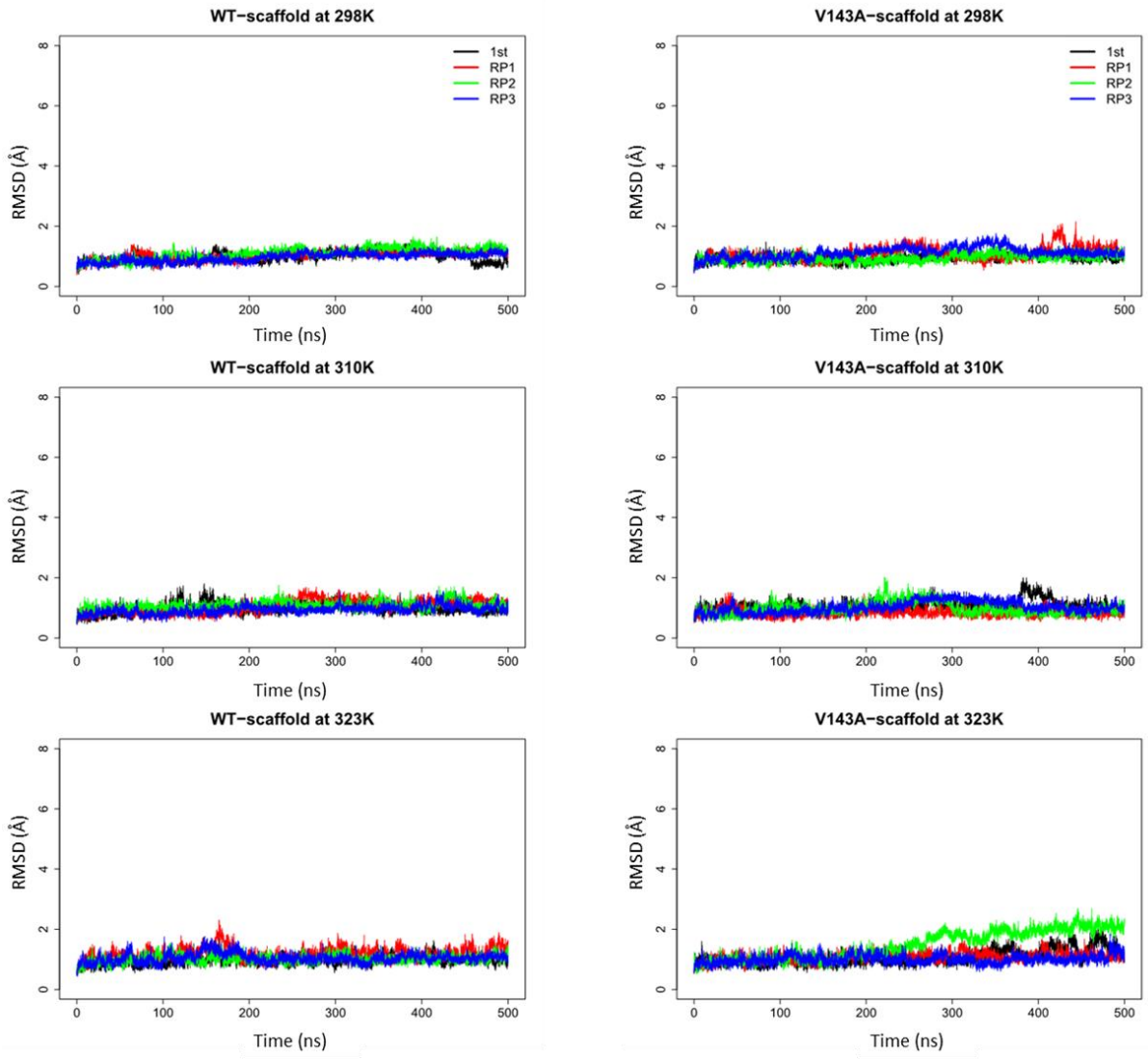


Figure 6-37: RMSD of the p53DBD WT and V143A aMD at different temperatures.

RMSD of the immunoglobulin-like β-sandwich (scaffold) was calculated from the simulations and each repeat after alignment on the reference scaffold (2AHI chain A).

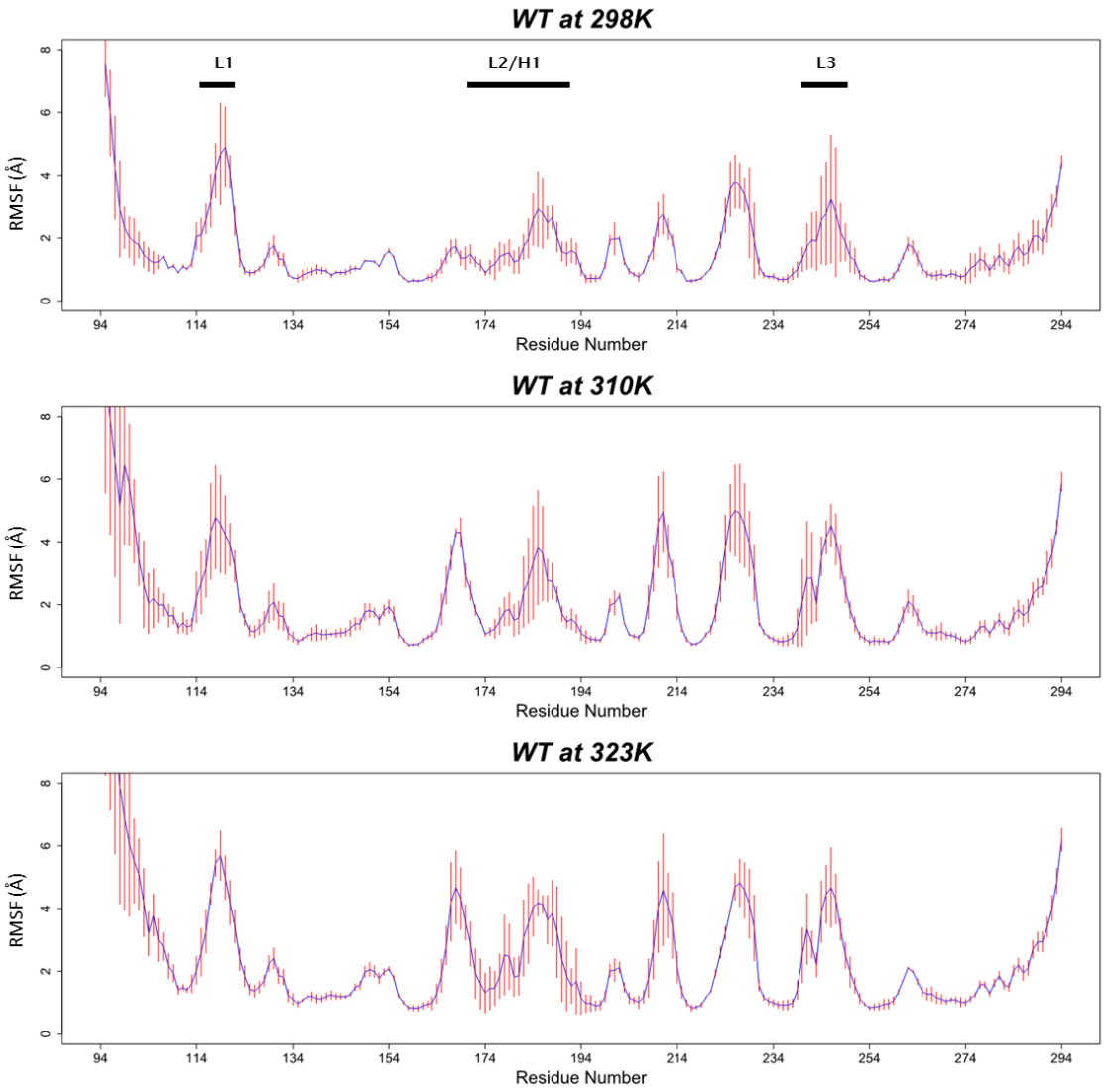


Figure 6-38: RMSF of the p53DBD WT aMD at different temperature.
The RMSF for each 500 ns simulation was determined using cpptraj in Amber. The average RMSF (blue) among the five cMD and the standard deviation (red) for each system were calculated using R.

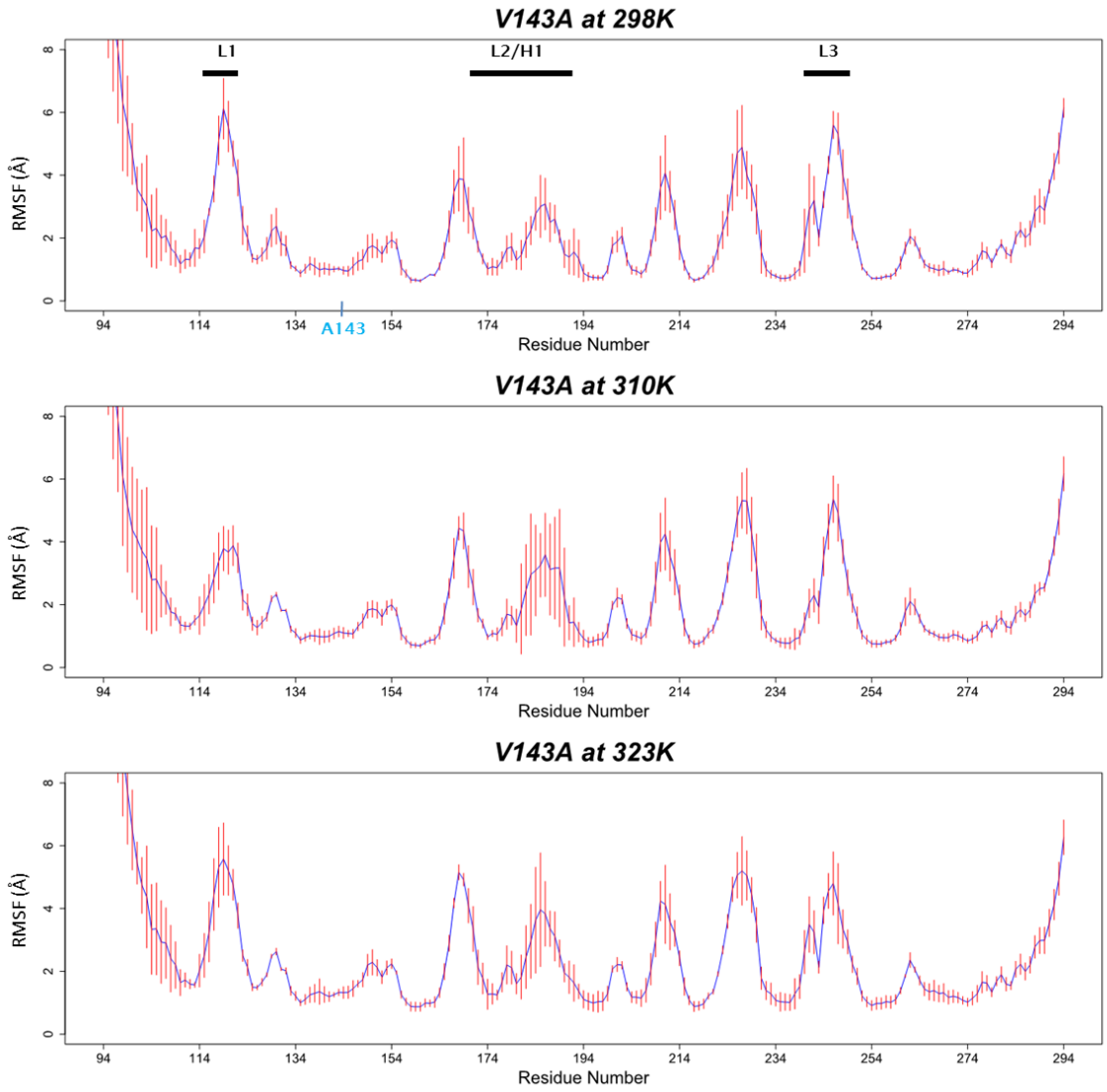


Figure 6-39: RMSF of the p53DBD V143A aMD at different temperature. The RMSF for each 500 ns simulation was determined using cpptraj in Amber. The average RMSF (blue) among the five cMD and the standard deviation (red) for each system were calculated using R. The mutated residue is labelled in blue.

While aMD simulations highlighted a slight destabilisation of the V143A scaffold (RMSD ≥ 2 Å, in the repeat 2 at 323 K) compared to the WT, the time scale used in the simulations (500 ns) did not allowed us to observe the denaturation of the system. Denatured V143A is known to bind to the PAb240 antibody (From residue 213 to 217, within the β-sheet S7, see Fig. 6-41), but not to the p53 WT. A SASA analysis of the PAb240 sequence over the cMD and aMD simulations of the WT and the mutant (data not shown), displayed a constant solvent accessibility between the systems and at different temperatures. The accessibility of the PAb240 binding sequence did not

increase in the V143A simulations, indicating that while V143A showed some stability differences compared to the WT, it still kept a WT-like conformation.

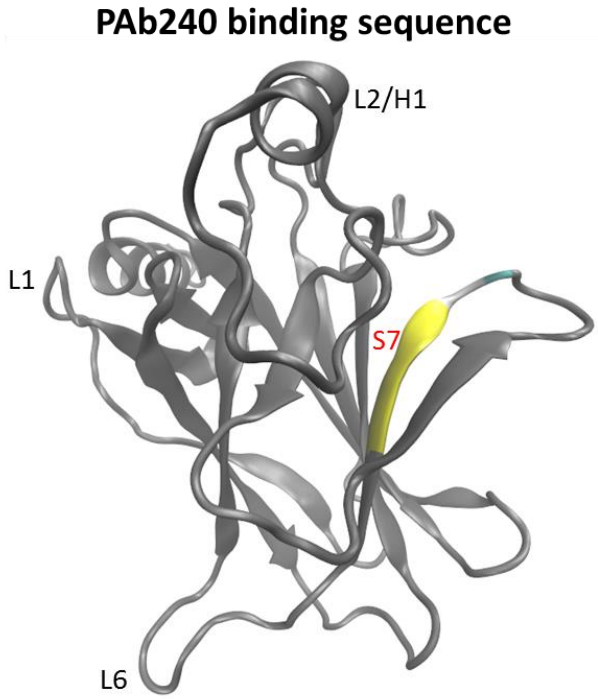


Figure 6-40: PAb240 antibody binding sequence within p53DBD.
P53DBD (pdb code: 2XWR) is represented in cartoon and coloured in grey and the PAb240 antibody binding sequence (β-sheet S7) is coloured by structure.

6.4 Conclusions

To propose a good model for p53DBD simulations, the influence of ion concentration was investigated. A concentration of 75 mM of NaCl was shown to be a good compromise, allowing to perform simulations of p53DBD close to natural condition (with the presence of both Na^+ and Cl^-, not only counter ions) and to explore a larger phase space that the simulation with 150 mM of NaCl. While the ion concentration seemed to have an influence on the L1 loop conformation of the p53DBD WT (changed in an α-helix in the system WEI and in a β-sheet at 150 mM, section 1.3.1), the occurrence of these changes (only one simulation at each concentration) did not allow us to draw a conclusion about ionic effect on the p53DBD.

The cMD showed that V143A restores and stabilises the H1 α-helix, known to be implicated in the p53 dimerization while binding to DNA promoter. This V143A H1 α-helix tended to be more stable and larger as the temperature was

raised. While these results confirmed the experimental study stating that the V143A mutation induces DBS changes (that explains V143A ability to bind to DNA at 298K (78)), no cavity was shown to be responsible for these changes.

The aMD simulations highlighted a limit of the ff99SB force field. Indeed while it was known that previous Amber force fields (ff94 and ff99) tended to form helices within non-helical sequence, ff99SB was thought to have overcome the problem (143). The use of ff99SB and aMD on p53DBD showed a high propensity of α-helix formation within the loops (not supposed to form helices). As a result, these helices (in L1 and L3 loops) within the aMD simulations were ignored.

The p53DBD WT is known to be unable to bind to the Bax promoter at 298 K (93). This promoter organisation suggests that for p53DBD to be able to bind to it, it has to be flexible, while p53DBD WT aMD showed less flexibility of its DBS at 298K (compare to the aMD at 310 and 323K). The decrease of DBS flexibility of p53DBD WT at 298K is thought to be the cause of its limited DNA binding.

The scaffold RMSD of both WT and V143A aMD simulations showed that V143A scaffold is less stable than the WT. Although V143A scaffold displayed a start of destabilization at 323K, The SASA of the sequence bound by PAb240 antibody (specific to V143A but not WT) showed no sign of V143A denaturation (SASA similar and constant for WT and V143A). The length of our cMD and aMD simulations (500 ns) were too short to observe V143A destabilisation, indeed protein folding and unfolding are known to generally occur within a time scale between 10^{-7} to 10^{4} s

Chapter 7. Conclusions

Due to its implication in cancer, p53 has been under scrutiny these last 3 decades. Computational methods were used in this study to interpret some of the important findings at an atomistic level.

Experimental structures (NMR and Xray, taken from the protein data bank - pdb) of the p53 DNA binding domain (p53DBD) were first gathered and their conformational differences studied using principal component analysis (PCA). This analysis highlighted that most of p53DBD conformational differences occur within its DNA binding surface (DBS) and more specifically in the DBS loops (L1, L2 and L3). The PCA of the p53DBD experimental structures clustered depending on the varying loop conformations. This clustering allowed us to distinguish changes in the L3 loop conformation, inhibiting p53 function (179, 180) and in the L1 loop rendering the DNA binding state of p53DBD. While the L1 loop conformational changes (when p53 binds to the DNA) within the engineered structures (p53DBD + tetramerisation domain) (96, 113, 114) highlighted differences depending on which part of the DNA (inner or outer sequence) the p53 monomer binds, this feature was absent in the p53DBD Xray structures (bound to DNA). The tetramerisation domain is thought to enable a specific binding of p53DBD to the DNA promoters which was confirmed by the DNA PCA. Indeed in this PCA analysis only the dinucleotides of the engineered structures were found to show torsional angles characteristics of DNA bound to a protein (α and γ transitions) and most of these dinucleotides were shown to be in direct contact with (or close to) a p53DBD residue.

An interesting result of the PCA performed on the p53DBD structures was the distinction between apo p53DBD and p53DBD bound to DNA, based on the L1 loop conformation differences. Indeed the L1 conformation of p53DBD bound to DNA was thought to be explored by the apo p53DBD (103, 112), while their PCA (without the engineered structures) showed a clear difference between those structures based on the L1 loop conformation they adopted.

In the second part of this project, the effects of a promising anti-cancer drug and a 'structural' mutant on p53DBD were investigated using conventional and accelerated molecular dynamics (cMD and aMD).

PRIMA-1^{MET} , a pro-drug metabolised in a Michael acceptor (MA) was shown to restore p53DBD mutant function and to enhance DNA binding of the p53 wild type (WT) (91, 141), highlighting a common mechanism. MAs displayed a concentration dependent effect on the WT (enhanced its DNA binding at low concentration and inhibit it at a higher concentration). While it has been demonstrated that the effect of MAs on p53 was due to their covalent binding with the p53DBD cysteines (91), doubts remained on the identities of the cysteines that were implicated and how this alkylation might restore mutant function and enhance the WT DNA binding. Based on the mass spectrometry study of Langridge-Smith *et al*. (109) investigating the alkylation of the p53DBD cysteine by N-ethylmaleimide (NEM), we proposed several models of p53DBD alkylation; mono-alkylated (1NEM, C182 or C277), di-alkylated (2NEM, C182 and C277), tri-alkylated (3NEM, 2NEM + C229 or C275) and tetra-alkylated (4NEM, 3NEM + C229 or C275). The analysis of the p53DBD conformations produced by cMD and aMD showed that the alkylation of C182 alone might be responsible for the beneficial effect of MAs on p53 WT at low concentration. Indeed C182 (together with C277) being one of the most solvent accessible cysteines, its alkylation induced the formation of a more stable H1 α-helix (dimer interface). The stabilisation of the H1 α-helix created a second salt bridge within the p53 dimer when binding to the DNA. On the other hand, the alkylation of C277 (also highly solvent accessible, and known to directly contact some DNA promoters) was shown to be responsible for the DNA binding inhibition effect observed at high concentration of MAs. The 1NEM-C277 model displayed conformational changes within the L1 loop (α-helix or β-sheet) that might abolish direct binding of K120 with the DNA, in addition to abolishing the C277 DNA contact.

While this study gave us good insights into the mechanism of action of MAs, the ability of the alkylated systems (1NEM-C182, 1NEM-C277 and 2NEM) to bind to DNA should be investigated to confirm our results. Therefore future work will focus on building and investigating the stability of the p53DBD dimer models (apo and alkylated systems) bound to DNA.

In chapter 6, a temperature sensitive mutant of p53, V143A, was investigated. This p53 mutant was shown to bind to some DNA promoters at low temperature (298K) while being 80 % destabilised at body temperature (310K) (92). The crystal structure of the V143A mutant (in the presence of 3 stabilising mutations) highlighted the presence of a cavity created by the mutation (180). The long range effect on the DBS of this mutation located in the p53 scaffold (in the β-strand S3) is thought to be due to the perturbation caused by the reduction in sidechain size (from Valine to Alanine) on the residues surrounding A143 (180). While cMD and aMD simulations of the WT and V143A at different temperatures (298 K, 310 K and 323 K) did not highlight p53 mutant unfolding, interesting results allowed to propose a mechanism by which p53DBD V143A binds to DNA at low temperatures and to explain why p53 WT presents a limited DNA binding at 298K. The cMD simulations highlighted a higher propensity of V143A to stabilise the H1 α-helix (the dimer interface) compared to the WT. This stable H1 α-helix, might be sufficient to allow the mutant to recover partial transcription function at low temperatures. While these results confirmed the effect of V143A on the DBS, no cavity was found to be responsible for this long-range effect. The exact mechanism by which V143A forms a more stable H1 α-helix than the WT remained elusive. No unfolding of the systems was observed, likely due to the small time scale used in this study (500 ns). On the other hand, V143A showed some signs of scaffold destabilisation at 323 K while the WT remained stable.

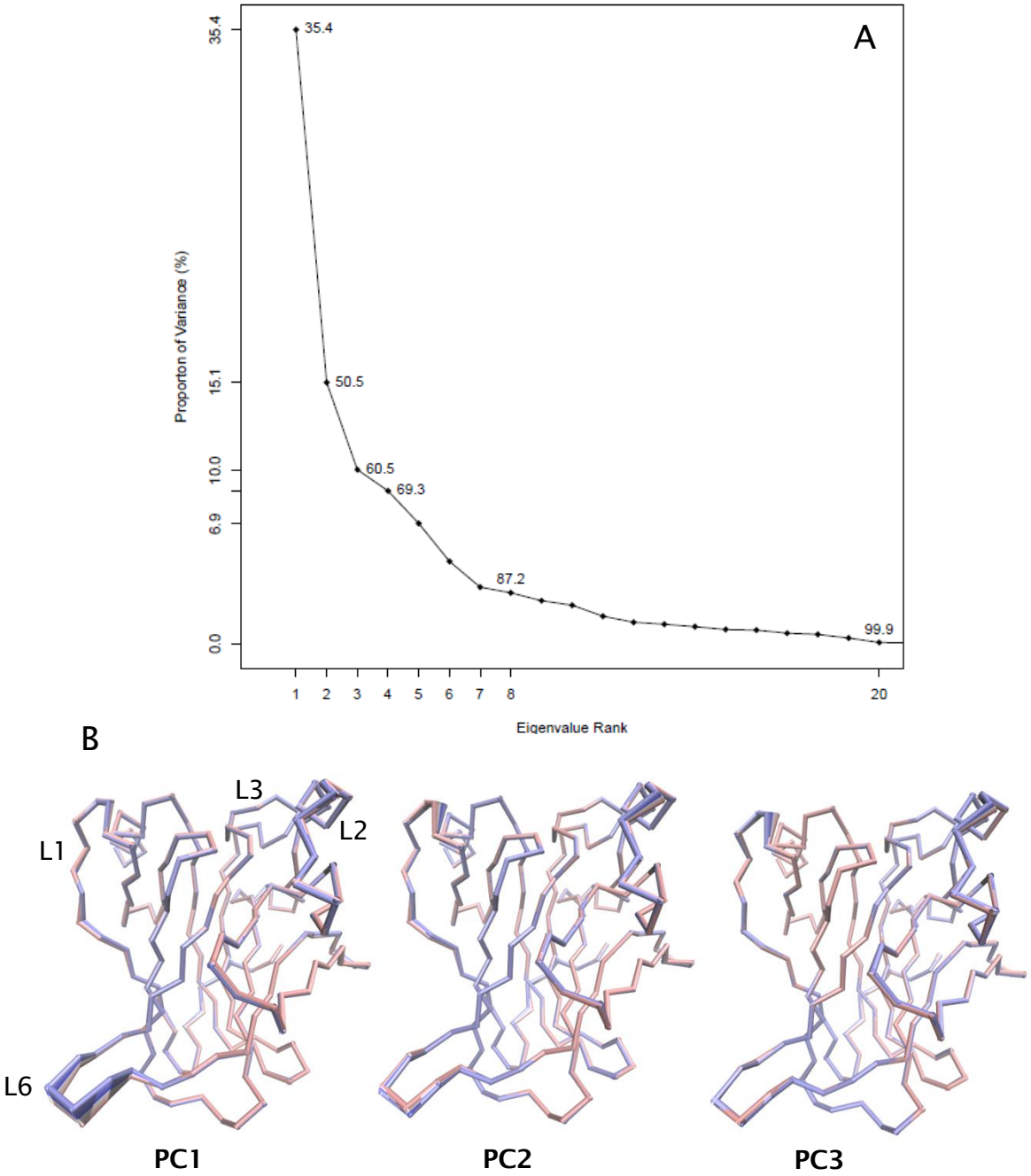
At 298K, p53 WT is known to not bind to the Bax promoter (92). The organisation of this promoter (section 3.5.4, Fig. 3-11), highlight the necessity for the p53 DBS to be flexible to be able to bind to Bax. The aMD simulations of p53DBD WT at 298K displayed reduced flexibility in the DBS that might be responsible for its inability to bind to the Bax promoter.

Due to the large time scale of p53DBD unfolding (half-life of 9 min at 310K), the denaturation pathway was not observed for these systems (WT and V143A). The use of longer enhanced sampling simulations (≥ 500 ns) might allow p53DBD V143A denaturation (scaffold started destabilising at 323K in aMD) to be observed.

Both cysteine alkylation and mutant V143A studies emphasised the implication of a stable H1 α-helix to enhance the DNA binding in the p53 WT or to restore p53 V143A function at 298K. A good strategy to rescue p53 mutants might then be to design a drug (MA) specific to C182.

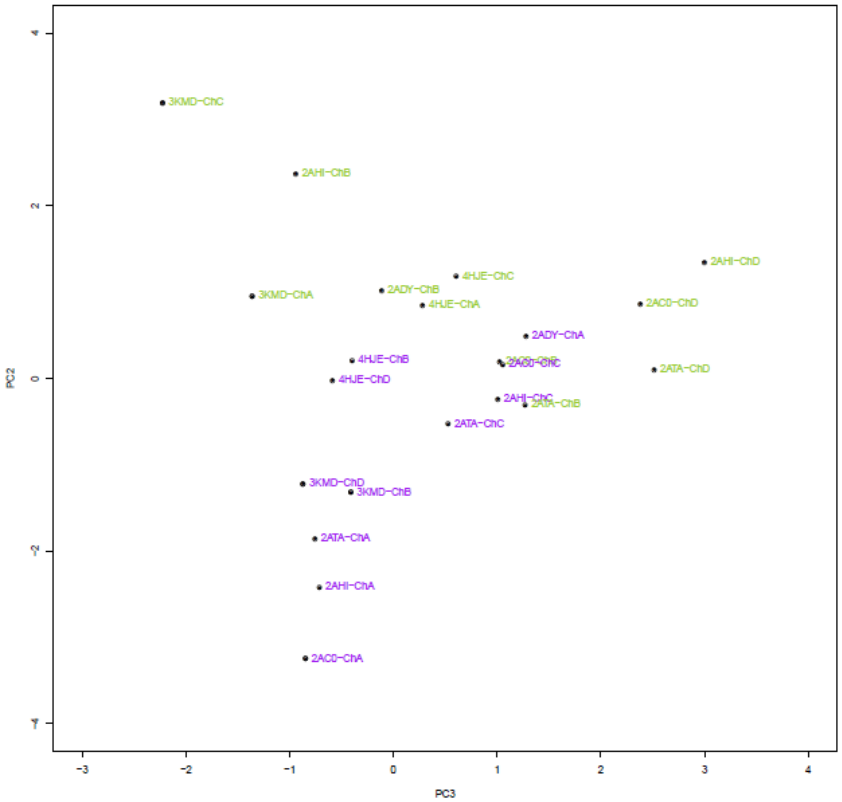
Appendices

Appendix A p53DBD experimental structure PCA



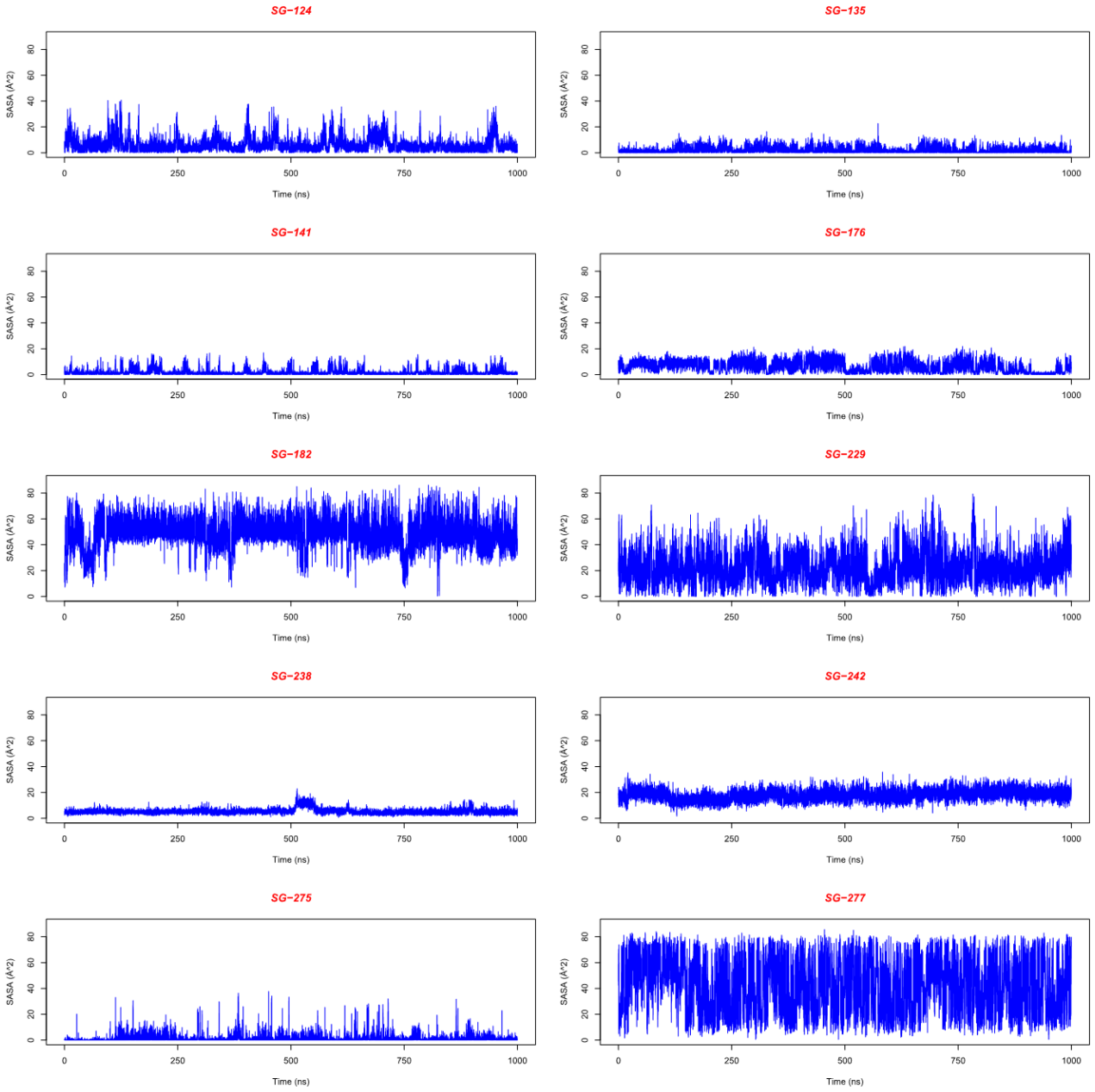
Appendix A Figure 1: p53DBD wild type bound to DNA, proportion of variance. Proportion of variance (in %) captured by each principal component from a data set of 22 monomers. The three first PCs contribute to 60.5% of the variance of the dataset.

Appendix A



Appendix A Figure 2: Structures projections along the PC2 and PC3 for the wild-type monomers bound to DNA.
Monomers bound to the inner part of the DNA consensus are coloured in purple while the monomers bound to the outer part are coloured in green.

Appendix B The alkylation of p53 by NEM



Appendix B Figure 1: p53DBD thiol groups solvent accessibility of the apo-p53DBD simulation (1st simulation) over time. Calculated using VMD.

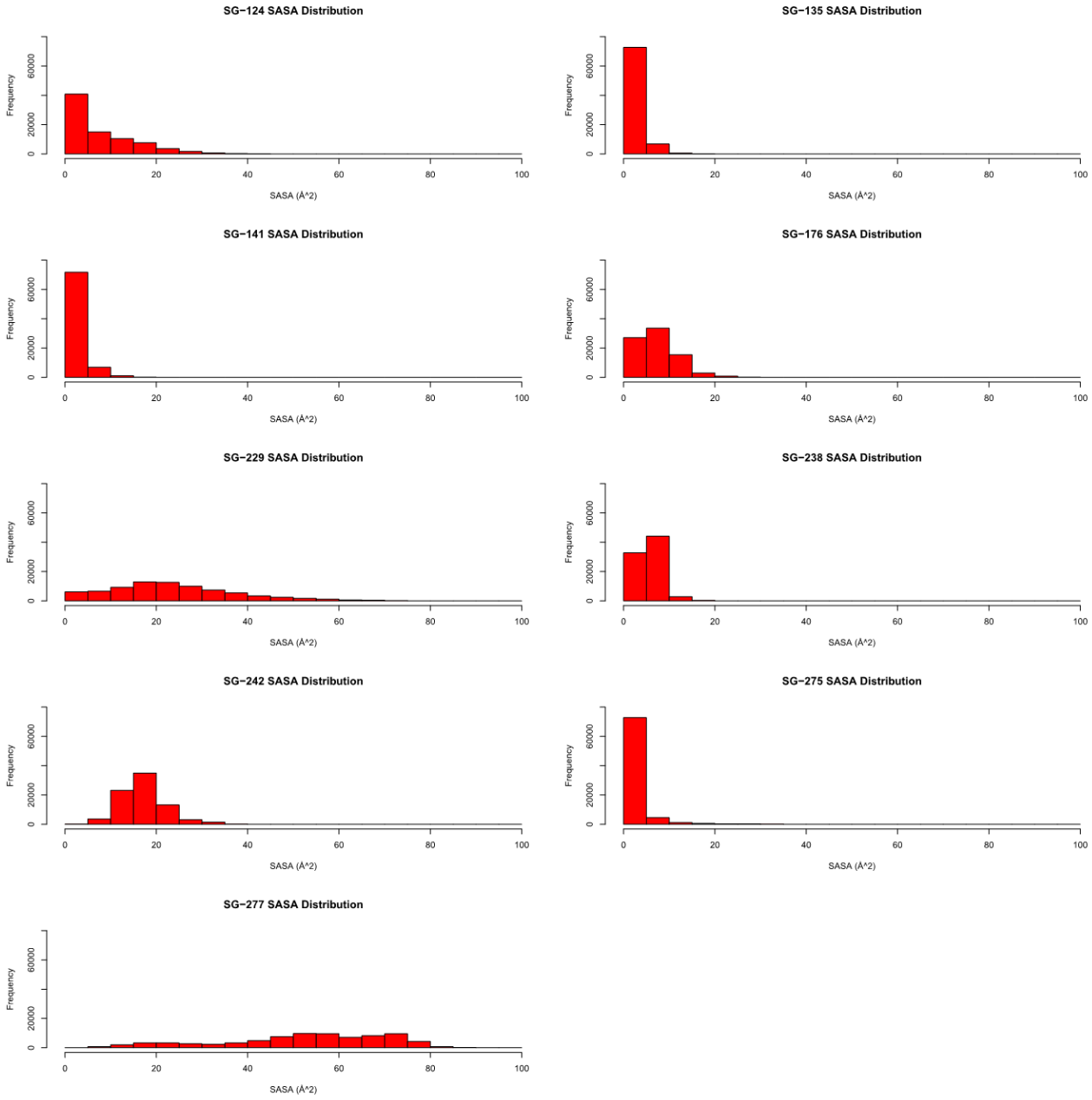
Appendix B

Thiol groups (SG)	SASA average per repeat (Å^2)				Total average (Å^2)
	1st simulation	Repeat 1	Repeat 2	Repeat 3	
124	**6.1** (± 5.4)	**5** (± 5.8)	**7.9** (± 7.9)	**3.9** (± 4.5)	**5.7** (± 6.2)
135	**1.9** (± 2.2)	**1.9** (± 2)	**1.9** (± 2.1)	**2.2** (± 2.3)	**2** (± 2.1)
141	**1.6** (± 2.2)	**2.6** (± 2.5)	**1.8** (± 2)	**2.8** (± 2.7)	**2.3** (± 2.4)
176	**6.3** (± 4.3)	**5.3** (± 3)	**3.7** (± 2.7)	**4.6** (± 2.6)	**5** (± 3.3)
182	**48.6** (± 11)	**51.4** (± 12.2)	**45.6** (± 17)	**47.8** (± 16)	**48.1** (± 14.4)
229	**22.1** (± 11.5)	**21** (± 13.3)	**28.2** (± 10.3)	**23.9** (± 12.2)	**23.8** (± 12.2)
238	**5.3** (± 2)	**5.5** (± 2.5)	**4.7** (± 1.8)	**5.5** (± 1.7)	**5.3** (± 2)
242	**18** (± 4)	**20** (± 4.9)	**22.9** (± 5)	**22** (± 5.1)	**20.7** (± 5.2)
275	**1.1** (± 3)	**1.3** (± 4.2)	**0.7** (± 2.7)	**1** (± 3.4)	**1** (± 3.4)
277	**41.8** (± 20.7)	**59.2** (± 12.6)	**56.6** (± 16)	**58.6** (± 12.7)	**54** (± 17.4)

Appendix B Table 1: Average solvent accessibility (Å2) of the thiol groups for each apo-p53DBD simulation.
The numbers in brackets refer to the standard deviation.

Thiol groups (SG)	SASA average per repeat (Å^2)				Total average (Å^2)
	1st simulation	Repeat 1	Repeat 2	Repeat 3	
124	**6.8 (± 7.6)**	**7.0 (± 5.8)**	**14.1 (± 7.0)**	**2.3 (± 2.9)**	**7.6 (± 7.4)**
135	**2.3 (± 2.4)**	**2.1 (± 2.1)**	**1.4 (± 1.9)**	**2.0 (± 2.0)**	**2.0 (±2.1)**
141	**1.7 (± 2.1)**	**1.4 (± 1.7)**	**1.4 (± 2.6)**	**3.5 (± 2.9)**	**2.0 (± 2.4)**
176	**7.1 (± 3.8)**	**9.8 (± 5.5)**	**6.0 (± 3.5)**	**5.8 (± 3.6)**	**7.2 (± 4.5)**
229	**20.0 (± 15.2)**	**25.8 (± 13.6)**	**26.6 (± 14.1)**	**22.4 (± 11.7)**	**23.7 (± 14.0)**
238	**5.0 (± 1.2)**	**7.0 (± 2.8)**	**5.7 (± 1.3)**	**5.4 (± 1.5)**	**5.7 (±2.0)**
242	**18.0 (± 6.5)**	**16.0 (± 4.1)**	**17.7 (± 4.1)**	**16.7 (± 3.9)**	**17.1 (± 4.8)**
275	**1.5 (± 2.8)**	**2.1 (± 4.6)**	**1.2 (± 3.4)**	**1.6 (± 4.3)**	**1.6 (± 4.0)**
277	**53.6 (± 17.3)**	**40.4 (± 20.2)**	**56.0 (± 15.2)**	**59.7 (± 12.8)**	**52.4 (± 18.1)**

Appendix B Table 2: Average solvent accessibility (Å^2) of the thiol groups for each C182-NEM simulation.
The numbers in brackets refer to the standard deviation.

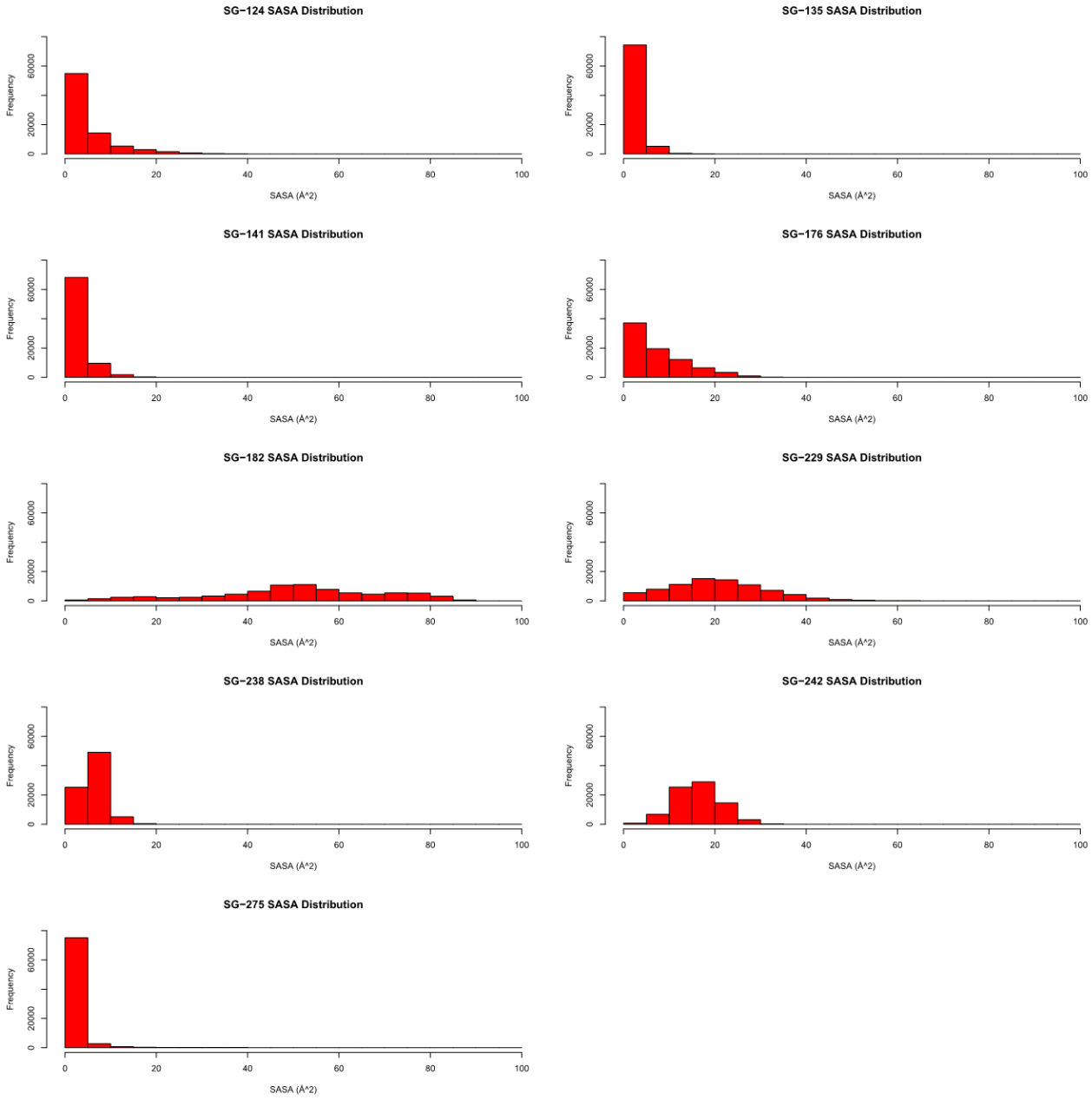


Appendix B Figure 2: Thiol group solvent accessibility distribution of the C182-NEM simulation.

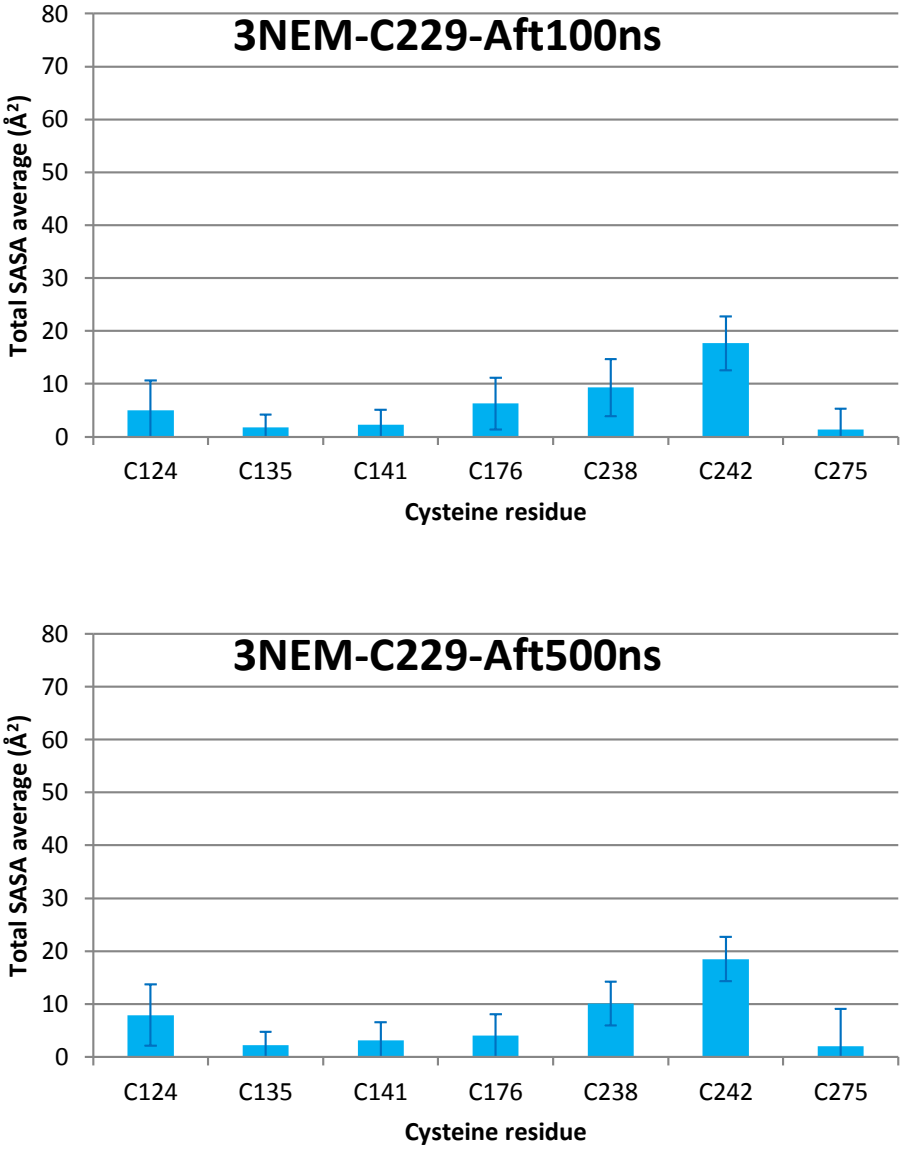
Thiol groups (SG)	SASA average per repeat (Å^2)				Total average (Å^2)
	1st simulation	Repeat 1	Repeat 2	Repeat 3	
124	**2.3 (± 2.7)**	**3.5 (± 5.3)**	**4.1 (± 4.2)**	**9.0 (± 7.3)**	**4.7 (± 5.7)**
135	**2.0 (± 2.0)**	**2.1 (± 2.3)**	**1.1 (± 1.4)**	**2.0 (± 2.0)**	**1.8 (± 2.0)**
141	**4.4 (± 3.5)**	**1.1 (± 1.5)**	**1.8 (± 2.0)**	**2.3 (± 2.5)**	**2.4 (± 2.8)**
176	**4.8 (± 3.7)**	**6.5 (± 3.6)**	**15.8 (± 5.9)**	**2.6 (± 2.1)**	**7.4 (± 6.4)**
182	**39.3 (± 17.7)**	**41.5 (± 16.7)**	**51.4 (± 9.3)**	**70.4 (± 12.2)**	**50.6 (± 18.9)**
229	**17.7 (± 10.0)**	**19.8 (± 10.3)**	**22.6 (± 11.5)**	**22.6 (± 10.5)**	**20.7 (± 10.8)**
238	**7.1 (± 2.7)**	**6.1 (± 1.6)**	**7.0 (± 2.4)**	**4.7 (± 1.5)**	**6.2 (± 2.3)**
242	**15.4 (± 4.5)**	**14.4 (± 3.8)**	**14.4 (± 4.3)**	**21.0 (± 4.0)**	**16.3 (± 5.0)**
275	**1.8 (± 5.7)**	**1.6 (± 3.6)**	**0.8 (± 3.1)**	**0.3 (± 1.8)**	**1.1 (± 3.9)**

Appendix B Table 3: Average solvent accessibility (Å^2) of the thiol groups for each C277-NEM simulation.
The numbers in brackets refer to the standard deviation.

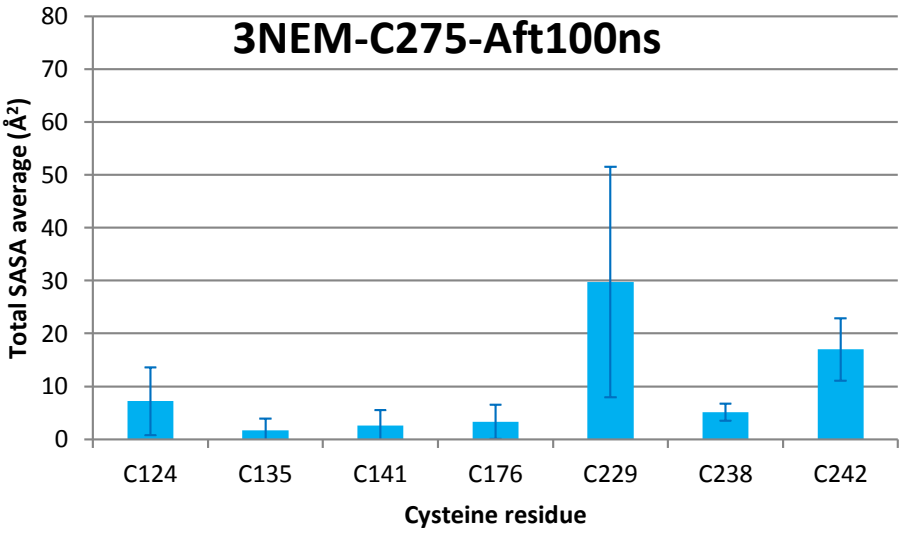
Appendix B

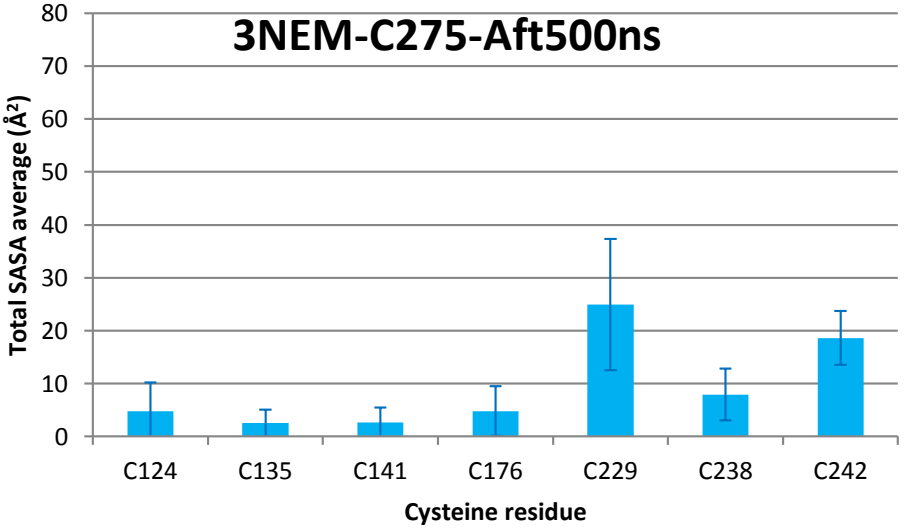


Appendix B Figure 3: Thiol group solvent accessibility distribution of the C277-NEM simulation.



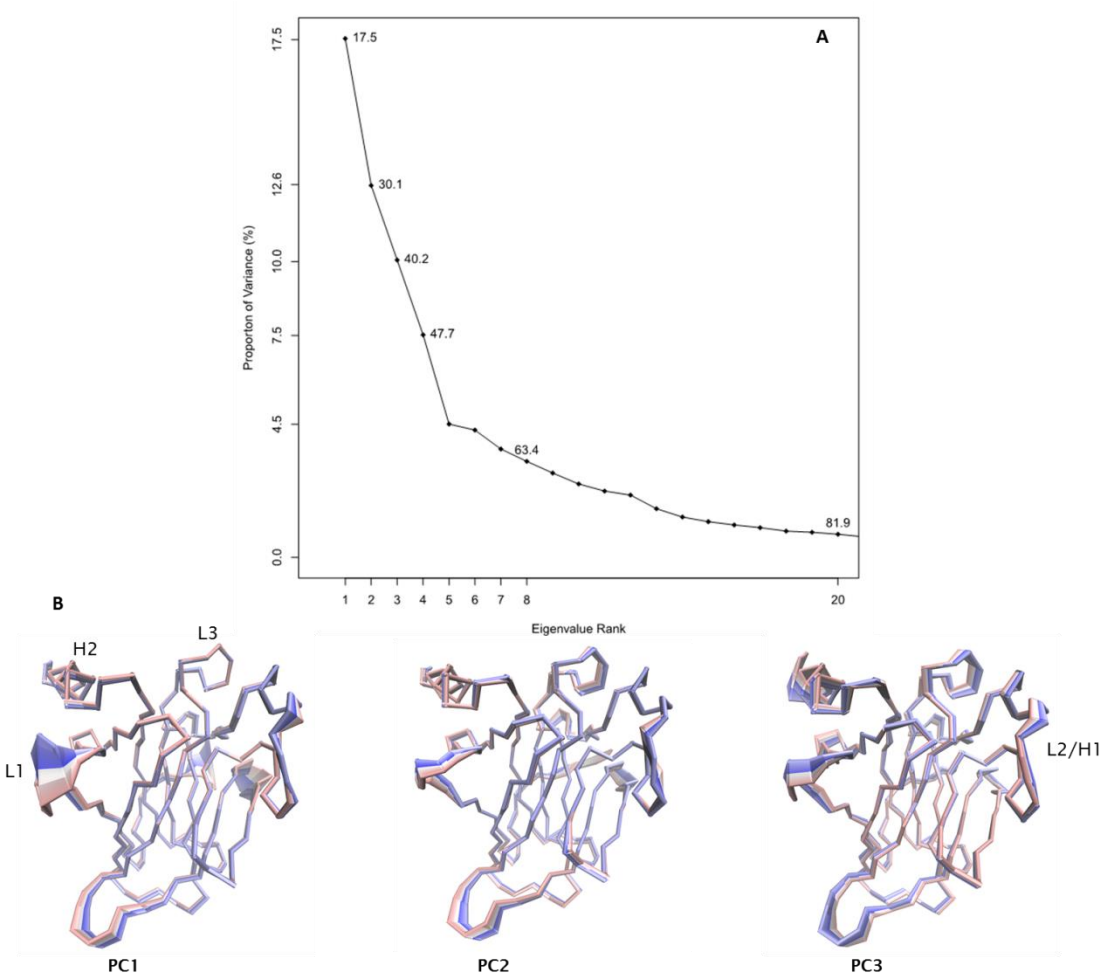
Appendix B Figure 4: Average SASA among cysteines of p53DBD tri-alkylated on C182 and C277 and C229.





Appendix B Figure 5: Average SASA among cysteines of p53DBD tri-alkylated on C182 and C277 and C275.

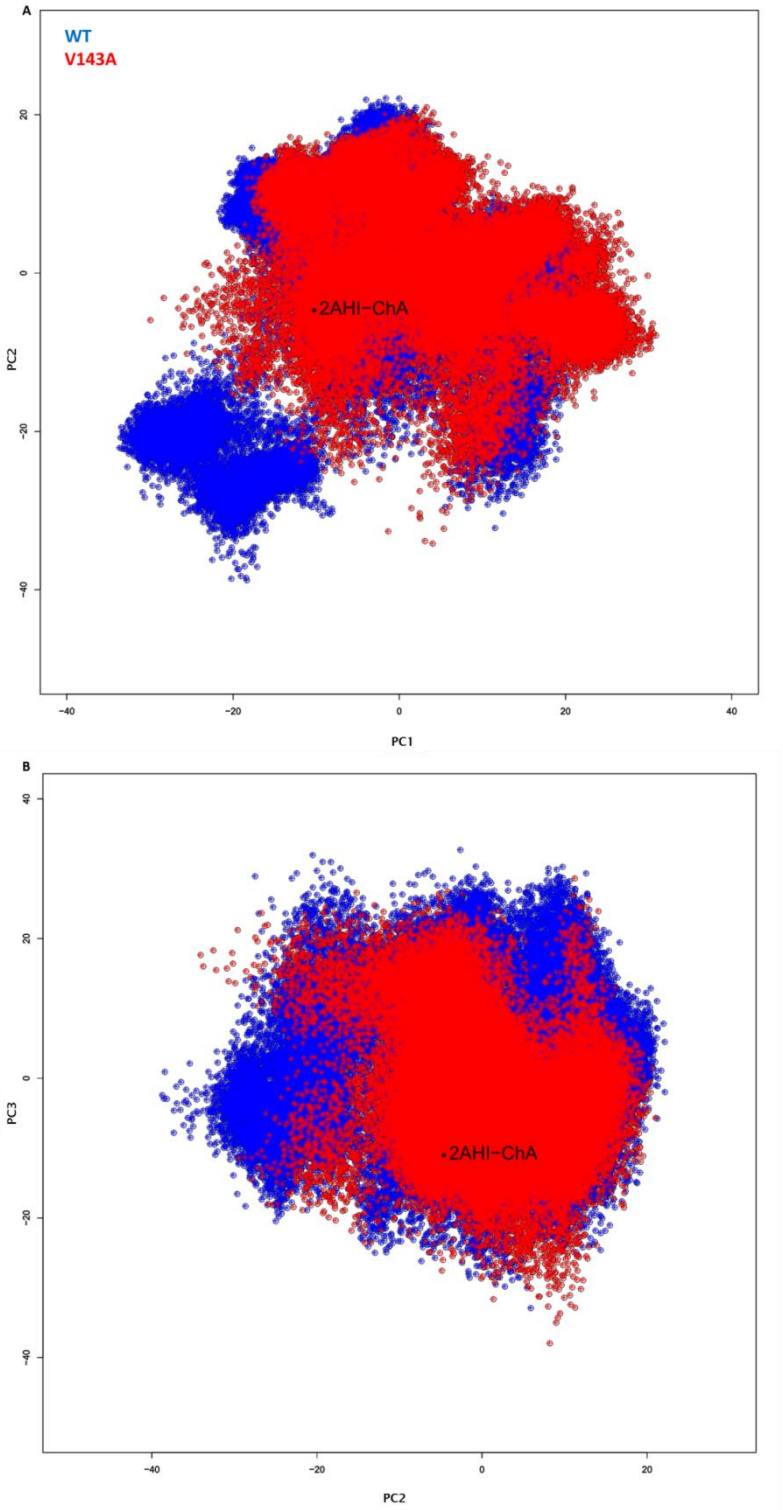
Appendix C p53DBD mutant: V143A



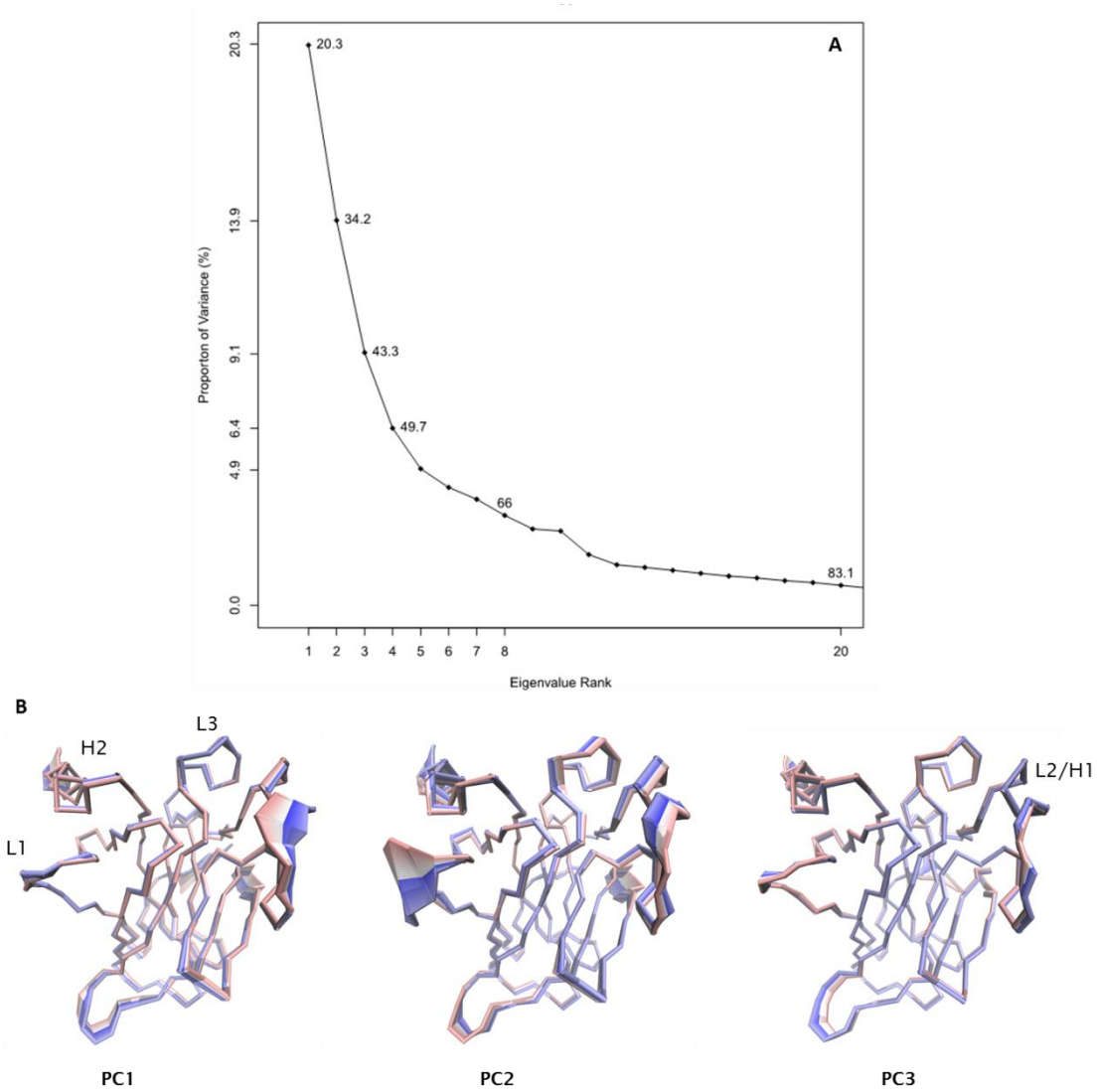
Appendix C Figure 1: PCA of the p53DBD WT and V143A at 310K, cMD simulation.

A) Proportion of variance (in %) captured by each principal component (PC) from a data set of 100000 p53DBD frames. The 3 first PCs contribute to 40.2 % of the variance of the dataset. **B)** Trace representation of the motion projections along the 3 first PCs (PC1, PC2 and PC3) of the 100000 frames using VMD. The colouring range of the motions goes from red, to white, to blue as function of time (timestep option on VMD), red being the early motion.

Appendix C



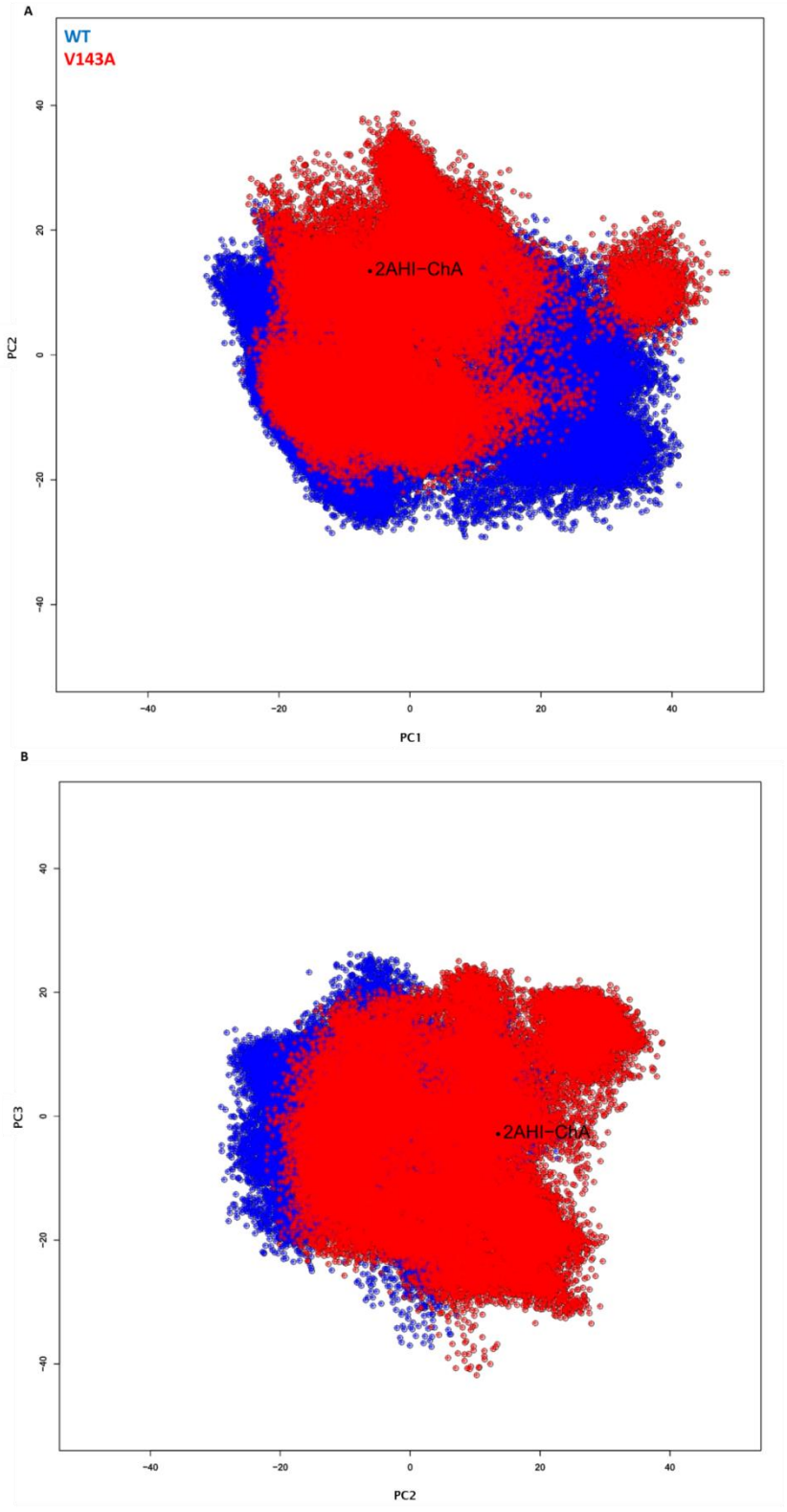
Appendix C Figure 2: Structure projections along the three first PCs of the p53DBD WT and mutant at 310 K, cMD simulations.
A) Along PC1 and PC2. **B)** Along PC2 and PC3. The 100000 frames (50000 per system) were aligned on the reference structure (2AHI chain A). The projected reference structure is labelled in black.



Appendix C Figure 3: PCA of the p53DBD WT and V143A at 323K, cMD simulations.

A) Proportion of variance (in %) captured by each principal component (PC) from a data set of 100000 p53DBD frames. The 3 first PCs contribute to 43.1 % of the variance of the dataset. **B)** Trace representation of the motion projections along the 3 first PCs (PC1, PC2 and PC3) of the 100000 frames using VMD. The colouring range of the motions goes from red, to white, to blue as function of time (timestep option on VMD), red being the early motion.

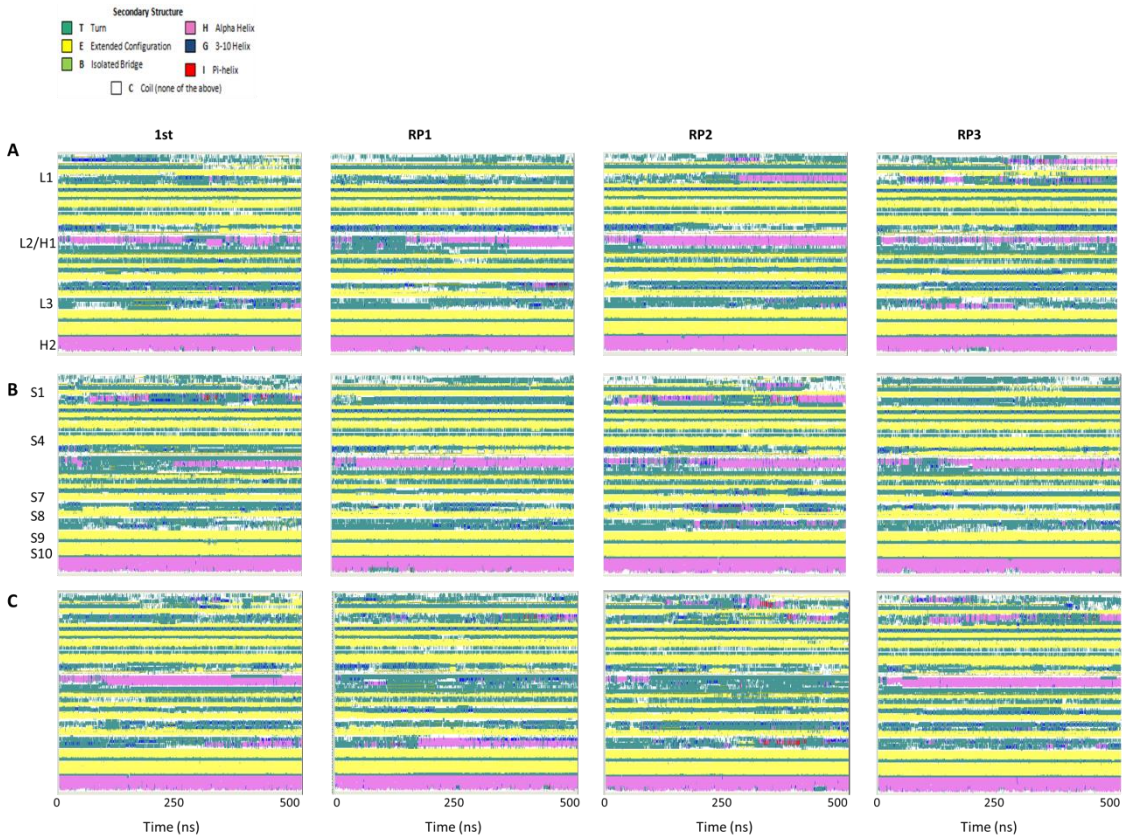
Appendix C



Appendix C Figure 4: Structure projections along the three first PCs of the p53DBD WT and mutant at 323 K, cMD simulations.
A) Along PC1 and PC2. **B)** Along PC2 and PC3. The 100000 frames (50000 per system) were aligned on the reference structure (2AHI chain A). The projected reference structure is labelled in black.

Appendix C Figure 5: Determination of the secondary structure in p53DBD WT at different temperature, aMD simulations.
A) 298 K, **B**) 310 K and **C**) 323 K. The secondary structure analysis was performed using timeline analysis in VMD.

Appendix C



Appendix C Figure 6: Determination of the secondary structure in p53DBD V143A at different temperatures, aMD simulations.
A) 298K, **B**) 310K and **C**) 323K. The secondary structure analysis was performed using timeline analysis in VMD.

List of References

1. Berg,J.M., Tymoczko,J.L. and Stryer,L. (2002) Biochemistry 5th editio. Freeman,W.H. (ed).

2. Dodson,G.G., Lane,D.P. and Verma,C.S. (2008) Molecular simulations of protein dynamics: new windows on mechanisms in biology. *EMBO Rep.*, **9**, 144–50.

3. Lane,D.P. (1992) p53, guardian of the genome. *Nature*, **358**.

4. Naga Deepthi,C., VVL Pavan Kumar,A., Rameshbadu and Indirapriyadarshini,U. (2011) Role of Tumor Suppressor Protein p53 in Apoptosis and Cancer Therapy. *J. Cancer Sci Ther*.

5. Rodier,F., Campisi,J. and Bhaumik,D. (2007) Two faces of p53: aging and tumor suppression. *Nucleic Acids Res.*, **35**, 7475–84.

6. Levitt,M. and Warshel, a (1975) Computer simulation of protein folding. *Nature*, **253**, 694–698.

7. Levitt,M. and Sharon,R. (1988) Accurate simulation of protein dynamics in solution. *Proc. Natl. Acad. Sci. U. S. A.*, **85**, 7557–7561.

8. American,O.F.T.H.E. and Society,C. (1976) Of t h e american chemical society. **101**, 3067–3071.

9. Li, a and Daggett,V. (1994) Characterization of the transition state of protein unfolding by use of molecular dynamics: chymotrypsin inhibitor 2. *Proc. Natl. Acad. Sci. U. S. A.*, **91**, 10430–10434.

10. Fersht, a R. and Daggett,V. (2002) Protein folding and unfolding at atomic resolution. *Cell*, **108**, 573–582.

11. Leach,A.R. (2001) Molecular modelling: principles and applications 2nd editio.

12. Ponder,J.W. and Case,D. a (2003) Force fields for protein simulations. *Adv. Protein Chem.*, **66**, 27–85.

13. Cornell,W.D., Cieplak,P., Bayly,C.I., Gould,I.R., Merz,K.M., Ferguson,D.M., Spellmeyer,D.C., Fox,T., Caldwell,J.W. and Kollman,P. a. (1995) A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.*, **117**, 5179–5197.

14. Brooks,B.R., Iii,C.L.B., Mackerell,A.D., Nilsson,L., Petrella,R.J., Roux,B., Won,Y., Archontis,G., Bartels,C., Boresch,S., *et al.* (2009) CHARMM : The Biomolecular Simulation Program. *J. Comput. Chem.*

References

15. Jorgensen,W.L. and Tirado-rives,J. (1988) The OPLS potential functions for proteins. Energy minimizations for crystals of cyclic peptides and Crambin. *Proteins*, **110**.

16. Mashayak,S.Y. and Tanner,D.E. (2011) Comparing Solvent Models for Molecular Dynamics of Protein. *Time*.

17. Jorgensen,W.L., Chandrasekhar,J., Madura,J.D., Impey,R.W. and Klein,M.L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, **79**, 926.

18. Still,W.C., Tempczyk,A., Hawley,R.C. and Hendrickson,T. (1990) Semianalytical treatment of solvation for molecular mechanics and dynamics. *Am. Chem. Soc.*, **112**, 6127–6129.

19. Onufriev,A., Case,D.A. and Bashford,D. (2002) Effective Born radii in the generalized Born approximation: The importance of being perfect. *J. Comput. Chem.*, **23**, 1297–1304.

20. Rashin,A.A. and Namboodiri,K. (1987) A simple method for the calculation of hydration enthalpies of polar molecules with arbitrary shapes. *J. Phys. Chem.*, **91**, 6003–6012.

21. Rashin,A.A. (1990) Hydration phenomena, classical electrostatics, and the boundary element method. *J. Phys. Chem.*, **94**, 1725–1733.

22. Forester,T.R. and Smith,W. (1998) SHAKE, rattle, and roll: Efficient constraint algorithms for linked rigid bodies. *J. Comput. Chem.*, **19**, 102–111.

23. Laio,A. and Gervasio,F.L. (2008) Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Reports Prog. Phys.*, **71**, 126601.

24. Voter,A.F. (1997) A method for accelerating the molecular dynamics simulation of infrequent events. *J. Chem. Phys.*, **106**, 4665–4677.

25. Hamelberg,D., Mongan,J. and McCammon,J.A. (2004) Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. *J. Chem. Phys.*, **120**, 11919–29.

26. Pierce,L.C.T., Salomon-Ferrer,R., Augusto F de Oliveira,C., McCammon,J.A. and Walker,R.C. (2012) Routine Access to Millisecond Time Scale Events with Accelerated Molecular Dynamics. *J. Chem. Theory Comput.*, **8**, 2997–3002.

27. Hamelberg,D., De Oliveira,C.A.F. and McCammon,J.A. (2007) Sampling of slow diffusive conformational transitions with accelerated molecular dynamics. *J. Chem. Phys.*, **127**.

28. Case,D. a, Cheatham,T.E., Darden,T., Gohlke,H., Luo,R., Merz,K.M., Onufriev,A., Simmerling,C., Wang,B. and Woods,R.J. (2005) The Amber biomolecular simulation programs. *J. Comput. Chem.*, **26**, 1668–88.

29. Humphrey,W., Dalke,A. and Schulten,K. (1996) VMD: Visual Molecular Dynamics. *J. Mol. Graph.*, **14**, 33–38.

30. Heinig,M. and Frishman,D. (2004) STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins. *Nucleic Acids Res.*, **32**, W500–2.

31. Durrant,J.D. and McCammon,J.A. (2011) HBonanza: A computer algorithm for molecular-dynamics-trajectory hydrogen-bond analysis. *J. Mol. Graph. Model.*, **31**, 5–9.

32. Ichiye,T. and Karplus,M. (1991) Collective motions in proteins: a covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. *Proteins*, **11**, 205–17.

33. An introduction to the R package mechanism (2002) *Alternatives*.

34. Yang,L.-W., Eyal,E., Bahar,I. and Kitao,A. (2009) Principal component analysis of native ensembles of biomolecular structures (PCA_NEST): insights into functional dynamics. *Bioinformatics*, **25**, 606–14.

35. Howe,P.W. (2001) Principal components analysis of protein structure ensembles calculated using NMR data. *J. Biomol. NMR*, **20**, 61–70.

36. Paramo,T., East,A., Garzo,D., Ulmschneider,M.B. and Bond,P.J. (2014) Efficient Characterization of Protein Cavities within Molecular Simulation Trajectories: trj_cavity. *J. Chem. Theory Comput.*, **10**, 2151–2164.

37. Liu,Y. and Kulesz-Martin,M. (2001) p53 protein at the hub of cellular DNA damage response pathways through sequence-specific and non-sequence-specific DNA binding. *Carcinogenesis*, **22**, 851–60.

38. Freedman,D. a and Levine, a J. (1999) Regulation of the p53 protein by the MDM2 oncoprotein--thirty-eighth G.H.A. Clowes Memorial Award Lecture. *Cancer Res.*, **59**, 1–7.

39. Sakaguchi,K., Herrera,J.E., Saito,S. 'i., Miki,T., Bustin,M., Vassilev, a., Anderson,C.W. and Appella,E. (1998) DNA damage activates p53 through a phosphorylation-acetylation cascade. *Genes Dev.*, **12**, 2831–2841.

40. Pietenpol,J. a and Stewart,Z. a (2002) Cell cycle checkpoint signaling: cell cycle arrest versus apoptosis. *Toxicology*, **181-182**, 475–81.

41. Taylor,W.R. and Stark,G.R. (2001) Regulation of the G2/M transition by p53. *Oncogene*, **20**, 1803–15.

References

42. Smith,M.L. and Seo,Y.R. (2002) p53 regulation of DNA excision repair pathways. *Mutagenesis*, **17**, 149–56.

43. Yamane,K., Katayama,E. and Tsuruo,T. (2001) p53 contains a DNA break-binding motif similar to the functional part of BRCT-related region of Rb. *Oncogene*, **20**, 2859–67.

44. Smith,M.L. and Kumar,M.A.S. (2010) The ' Two faces ' of Tumor Suppressor p53-revisited. *Mol. Cell*, **2**, 117–119.

45. Benchimol,S. (2001) P53-Dependent Pathways of Apoptosis. *Cell Death Differ.*, **8**, 1049–51.

46. Haupt,S., Berger,M., Goldberg,Z. and Haupt,Y. (2003) Apoptosis - the p53 network. *J. Cell Sci.*, **116**, 4077–85.

47. Offer,H., Erez,N., Zurer,I., Tang,X., Milyavsky,M., Goldfinger,N. and Rotter,V. (2002) The onset of p53-dependent DNA repair or apoptosis is determined by the level of accumulated damaged DNA. *Carcinogenesis*, **23**, 1025–32.

48. Kitano,H. (2004) Cancer as a robust system: implications for anticancer therapy. *Nat. Rev. Cancer*, **4**, 227–235.

49. Joerger,A.C. and Fersht,A.R. (2010) The tumor suppressor p53: from structures to drug discovery. *Cold Spring Harb. Perspect. Biol.*, **2**, a000919.

50. Joerger, a C. and Fersht, a R. (2007) Structure-function-rescue: the diverse nature of common p53 cancer mutants. *Oncogene*, **26**, 2226–42.

51. Oijen,M.G.C.T. Van (2000) Gain-of-Function Mutations in the Tumor Suppressor Gene Gain-of-Function Mutations in the Tumor Suppressor Gene p53. *Clin. Cancer Res.*

52. el-Deiry WS, Kern SE, Pietenpol JA, Kinzler KW,V.B. (1992) Definition of a consensus binding site for p53. *Nat. Genet.*, **1**, 45–9.

53. 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**, 246–56.

54. Che,P. (2001) The role of tetramerization in p53 function. *Oncogene*, **393**.

55. Goodman,R.H. and Smolik,S. (2000) CBP / p300 in cell growth , transformation , and development. *Genes Dev.*, **14**, 1553–1577.

56. Brady,C. a, Jiang,D., Mello,S.S., Johnson,T.M., Lesley, a, Kozak,M.M., Broz,D.K., Basak,S., Eunice,J., Mclaughlin,M.E., *et al.* (2012) Damage Responses and Tumor Suppression. **145**, 571–583.

57. Sciences,M. (1996) Identification of a novel p53 functional domain that is necessary. **93**, 15335–15340.

58. Meek,D.W. (1994) Post-translational modification of p53. *Semin. Cancer Biol.*, **5**, 203–210.

59. Gu,B. and Zhu,W.G. (2012) Surf the Post-translational Modification Network of p53 Regulation. *Int. J. Biol. Sci.*, **8**, 672–684.

60. Okorokov,A.L., Sherman,M.B., Plisson,C., Grinkevich,V., Sigmundsson,K., Selivanova,G., Milner,J. and Orlova,E. V (2006) The structure of p53 tumour suppressor protein reveals the basis for its functional plasticity. *EMBO J.*, **25**, 5191–5200.

61. Pham,N., Lucumi,A., Cheung,N. and Viadiu,H. (2012) The tetramer of p53 in the absence of DNA forms a relaxed quaternary state. *Biochemistry*, **51**, 8053–8055.

62. Tidow,H., Melero,R., Mylonas,E., Freund,S.M. V, Grossmann,J.G., Carazo,J.M., Svergun,D.I., Valle,M. and Fersht,A.R. (2007) Quaternary structures of tumor suppressor p53 and a specific p53 DNA complex. *Proc. Natl. Acad. Sci. U. S. A.*, **104**, 12324–12329.

63. Wells,M., Tidow,H., Rutherford,T.J., Markwick,P., Jensen,M.R., Mylonas,E., Svergun,D.I., Blackledge,M. and Fersht,A.R. (2008) Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation domain. *Proc. Natl. Acad. Sci. U. S. A.*, **105**, 5762–5767.

64. Arlt,C., Ihling,C.H. and Sinz,A. (2015) Structure of full-length p53 tumor suppressor probed by chemical cross-linking and mass spectrometry. *Proteomics*, 10.1002/pmic.201400549.

65. Khoury,M.P. and Bourdon,J.-C. (2010) The isoforms of the p53 protein. *Cold Spring Harb. Perspect. Biol.*, **2**, a000927.

66. Hollstein,M. and Hainaut,P. (2010) Massively regulated genes : the example of TP53. *J. Pathol.*

67. Reisman,D. and Loging,W.T. (1998) Transcriptional regulation of the p53 tumor suppressor gene. *Semin. Cancer Biol.*, **8**, 317–324.

68. Li,M., Brooks,C.L., Wu-Baer,F., Chen,D., Baer,R. and Gu,W. (2003) Mono- versus polyubiquitination: differential control of p53 fate by Mdm2. *Science*, **302**, 1972–1975.

69. Chan,W.M., Mak,M.C., Fung,T.K., Lau,A., Siu,W.Y. and Poon,R.Y.C. (2006) Ubiquitination of p53 at multiple sites in the DNA-binding domain. *Mol. Cancer Res.*, **4**, 15–25.

70. Brooks,C.L. and Gu,W. (2011) The impact of acetylation and deacetylation on the p53 pathway. *Protein Cell*, **2**, 456–462.

References

71. Iyer,N.G., Chin,S., Ozdag,H., Daigo,Y., Hu,D., Cariati,M., Brindle,K., Aparicio,S. and Caldas,C. (2004) p300 regulates p53-dependent apoptosis after DNA damage in colorectal cancer cells by modulation of PUMA ¯p21 levels. *PNAS*, **101**, 4–6.

72. Tang,Y., Luo,J., Zhang,W. and Gu,W. (2006) Tip60-dependent acetylation of p53 modulates the decision between cell-cycle arrest and apoptosis. *Mol. Cell*, **24**, 827–39.

73. Li,M., Luo,J., Brooks,C.L. and Gu,W. (2002) Acetylation of p53 inhibits its ubiquitination by Mdm2. *J. Biol. Chem.*, **277**, 50607–11.

74. Hainaut,P. and Milner,J. (1993) Redox Modulation of p53 Conformation and Sequence-specific DNA Binding in Vitro Advances in Brief Redox Modulation of p53 Conformation and Sequence-specific DNA Binding in Vitro. *Cancer Res.*, **53**, 4469–4473.

75. Hupp,T.R., Meek,D.W., Midgley,C.A. and Lane,D.P. (1993) Activation of the cryptic DNA binding function forms of p53. **21**, 3167–3174.

76. Rainwater,R., Parks,D., Anderson,M.E., Tegtmeyer,P. and Mann,K. (1995) Role of cysteine residues in regulation of p53 function. *Mol. Cell. Biol.*, **15**, 3892–903.

77. Joerger,A.C., Allen,M.D. and Fersht,A.R. (2004) Crystal structure of a superstable mutant of human p53 core domain. Insights into the mechanism of rescuing oncogenic mutations. *J. Biol. Chem.*, **279**, 1291–6.

78. Joerger,A.C., Ang,H.C. and Fersht,A.R. (2006) Structural basis for understanding oncogenic p53 mutations and designing rescue drugs. *Proc. Natl. Acad. Sci. U. S. A.*, **103**, 15056–61.

79. Freed-Pastor,W. a and Prives,C. (2012) Mutant p53: one name, many proteins. *Genes Dev.*, **26**, 1268–86.

80. Weisz,L., Oren,M. and Rotter,V. (2007) Transcription regulation by mutant p53. *Oncogene*, **26**, 2202–11.

81. Vaughan,C.A., Deb,S.P., Deb,S. and Windle,B. (2014) Preferred binding of gain-of-function mutant p53 to bidirectional promoters with coordinated binding of ETS1 and GABPA to multiple binding sites ABSTRACT : **5**.

82. Chicas, a, Molina,P. and Bargonetti,J. (2000) Mutant p53 forms a complex with Sp1 on HIV-LTR DNA. *Biochem. Biophys. Res. Commun.*, **279**, 383–90.

83. Freed-Pastor,W.A., Mizuno,H., Xi Zhao,X., Langerod,A., Moon,S., Rodriguez-Barrueco,R., Barsotti,A., Chicas,A., Li,W., Polotskaia,A., *et al.* (2012) Mutant p53 disrupts mammary acinar morphogenesis via the mevalonate pathway. *Cell*, **148**, 244–258.

84. Di Agostino,S., Strano,S., Emiliozzi,V., Zerbini,V., Mottolese,M., Sacchi,A., Blandino,G. and Piaggio,G. (2006) Gain of function of mutant p53: the mutant p53/NF-Y protein complex reveals an aberrant transcriptional mechanism of cell cycle regulation. *Cancer Cell*, **10**, 191–202.

85. Stambolsky,P., Tabach,Y., Fontemaggi,G., Weisz,L., Maor-,R., Shiff,I., Kogan,I., Shay,M., Kalo,E., Blandino,G., *et al.* (2010) Modulation of the vitamin D3 response by cancer-associated mutant p53. *Cancer Cell*, **17**, 273–285.

86. Girardini,J.E., Napoli,M., Piazza,S., Rustighi,A., Marotta,C., Radaelli,E., Capaci,V., Jordan,L., Quinlan,P., Thompson,A., *et al.* (2011) A Pin1/mutant p53 axis promotes aggressiveness in breast cancer. *Cancer Cell*, **20**, 79–91.

87. Kim,E. and Deppert,W. (2007) Interactions of mutant p53 with DNA: guilt by association. *Oncogene*, **26**, 2185–90.

88. Göhler,T., Jäger,S., Warnecke,G., Yasuda,H., Kim,E. and Deppert,W. (2005) Mutant p53 proteins bind DNA in a DNA structure-selective mode. *Nucleic Acids Res.*, **33**, 1087–100.

89. Nikolova,P. V, Wong,K.B., DeDecker,B., Henckel,J. and Fersht, a R. (2000) Mechanism of rescue of common p53 cancer mutations by second-site suppressor mutations. *EMBO J.*, **19**, 370–8.

90. Friedler,A., Veprintsev,D.B., Hansson,L.O. and Fersht,A.R. (2003) Kinetic instability of p53 core domain mutants. Implications for rescue by small molecules. *J. Biol. Chem.*, **278**, 24108–24112.

91. Lambert,J.M.R., Gorzov,P., Veprintsev,D.B., Söderqvist,M., Segerbäck,D., Bergman,J., Fersht,A.R., Hainaut,P., Wiman,K.G. and Bykov,V.J.N. (2009) PRIMA-1 reactivates mutant p53 by covalent binding to the core domain. *Cancer Cell*, **15**, 376–88.

92. Di Como,C.J. and Prives,C. (1998) Human tumor-derived p53 proteins exhibit binding site selectivity and temperature sensitivity for transactivation in a yeast-based assay. *Oncogene*, **16**, 2527–39.

93. Friedlander,P., Haupt,Y., Prives,C. and Oren,M. (1996) A mutant p53 that discriminates between p53-responsive genes cannot induce apoptosis. *Mol. Cell. Biol.*, **16**, 4961–71.

94. Miyashita,T. and Reed,J.C. (1995) Tumor suppressor p53 is a direct transcriptional activator of the human bax gene. *Cell*, **80**, 293–9.

95. el-Deiry,W.S., Tokino,T., Velculescu,V.E., Levy,D.B., Parsons,R., Trent,J.M., Lin,D., Mercer,W.E., Kinzler,K.W. and Vogelstein,B. (1993) WAF1, a potential mediator of p53 tumor suppression. *Cell*, **75**, 817–25.

96. Emamzadah,S., Tropia,L. and Halazonetis,T.D. (2011) Crystal structure of a multidomain human p53 tetramer bound to the natural CDKN1A (p21) p53-response element. *Mol. Cancer Res.*, **9**, 1493–9.

97. Wong,K.B., DeDecker,B.S., Freund,S.M., Proctor,M.R., Bycroft,M. and Fersht, a R. (1999) Hot-spot mutants of p53 core domain evince characteristic local structural changes. *Proc. Natl. Acad. Sci. U. S. A.*, **96**, 8438–42.

98. Merabet,A., Houlleberghs,H., Maclagan,K., Akanho,E., Bui,T.T.T., Pagano,B., Drake,A.F., Fraternali,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. *Biochem. J.*, **427**, 225–36.

99. Bartussek,C., Naumann,U. and Weller,M. (1999) Accumulation of mutant p53(V143A) modulates the growth, clonogenicity, and radiochemosensitivity of malignant glioma cells independent of endogenous p53 status. *Exp. Cell Res.*, **253**, 432–9.

100. Mcilwrath,A.J., Vasey,P.A., Ross,G.M., Mcllwrath,A.J. and Brown,R. (1994) Cell Cycle Arrests and Radiosensitivity of Human Tumor Cell Lines: Dependence on Wild-Type p53 for Radiosensitivity. *Cancer Res.*, **54**, 3718–3722.

101. McDonald, A.C. and Brown,R. (1998) Induction of p53-dependent and p53-independent cellular responses by topoisomerase 1 inhibitors. *Br. J. Cancer*, **78**, 745–51.

102. Duan,J. and Nilsson,L. (2006) Effect of Zn2+ on DNA recognition and stability of the p53 DNA-binding domain. *Biochemistry*, **45**, 7483–92.

103. Cañadillas,J.M.P., Tidow,H., Freund,S.M. V, Rutherford,T.J., Ang,H.C. and Fersht,A.R. (2006) Solution structure of p53 core domain: structural basis for its instability. *Proc. Natl. Acad. Sci. U. S. A.*, **103**, 2109–14.

104. Wright,J.D., Noskov,S.Y. and Lim,C. (2002) Factors governing loss and rescue of DNA binding upon single and double mutations in the p53 core domain. *Nucleic Acids Res.*, **30**, 1563–74.

105. Wright,J.D. and Lim,C. (2007) Mechanism of DNA-binding loss upon single-point mutation in p53. *J. Biosci.*, **32**, 827–39.

106. Tu,C., Tan,Y.H., Shaw,G., Zhou,Z., Bai,Y., Luo,R. and Ji,X. (2008) Impact of low-frequency hotspot mutation R282Q on the structure of p53 DNA-binding domain as revealed by crystallography at 1.54 angstroms resolution. *Acta Crystallogr. D. Biol. Crystallogr.*, **64**, 471–7.

107. Bhanoori,M., Yellaturu,C.R., Ghosh,S.K., Hassid,A., Jennings,L.K. and Rao,G.N. (2003) Thiol alkylation inhibits the mitogenic effects of platelet-derived growth factor and renders it proapoptotic via activation of STATs and p53 and induction of expression of caspase1 and p21(waf1/cip1). *Oncogene*, **22**, 117–130.

108. Kaar,J.L., Basse,N., Joerger,A.C., Stephens,E., Rutherford,T.J. and Fersht,A.R. (2010) Stabilization of mutant p53 via alkylation of cysteines and effects on DNA binding. *Protein Sci.*, **19**, 2267–78.

109. Scotcher,J., Clarke,D.J., Weidt,S.K., Mackay,C.L., Hupp,T.R., Sadler,P.J. and Langridge-Smith,P.R.R. (2011) Identification of two reactive cysteine residues in the tumor suppressor protein p53 using top-down FTICR mass spectrometry. *J. Am. Soc. Mass Spectrom.*, **22**, 888–97.

110. Gendoo,D.M.A. and Harrison,P.M. (2012) The landscape of the prion protein's structural response to mutation revealed by principal component analysis of multiple NMR ensembles. *PLoS Comput. Biol.*, **8**.

111. Lukman,S., Grant,B.J., Gorfe,A.A., Grant,G.H. and McCammon,J.A. (2010) The distinct conformational dynamics of K-Ras and H-Ras A59G. *PLoS Comput. Biol.*, **6**, 4–12.

112. Lukman,S., Lane,D.P. and Verma,C.S. (2013) Mapping the Structural and Dynamical Features of Multiple p53 DNA Binding Domains: Insights into Loop 1 Intrinsic Dynamics. *PLoS One*, **8**.

113. Petty,T.J., Emamzadah,S., Costantino,L., Petkova,I., Stavridi,E.S., Saven,J.G., Vauthey,E. and Halazonetis,T.D. (2011) An induced fit mechanism regulates p53 DNA binding kinetics to confer sequence specificity. *EMBO J.*, **30**, 2167–76.

114. Emamzadah,S., Tropia,L., Vincenti,I., Falquet,B. and Halazonetis,T.D. (2014) Reversal of the DNA-Binding-Induced Loop L1 Conformational Switch in an Engineered Human p53 Protein. *J. Mol. Biol.*, **426**, 936–44.

115. Weinberg,R.L., Veprintsev,D.B., Bycroft,M. and Fersht,A.R. (2005) Comparative binding of p53 to its promoter and DNA recognition elements. *J. Mol. Biol.*, **348**, 589–596.

116. Lu,X.-J. (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.*, **31**, 5108–5121.

117. Elsawy,K.M., Hodgson,M.K. and Caves,L.S.D. (2005) The physical determinants of the DNA conformational landscape: an analysis of the potential energy surface of single-strand dinucleotides in the conformational space of duplex DNA. *Nucleic Acids Res.*, **33**, 5749–62.

118. Grant,B.J., Rodrigues,A.P.C., ElSawy,K.M., McCammon,J.A. and Caves,L.S.D. (2006) Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics*, **22**, 2695–6.

119. Joerger,A.C. and Fersht,A.R. (2010) The tumor suppressor p53: from structures to drug discovery. *Cold Spring Harb. Perspect. Biol.*, **2**, a000919.

120. Lilyestrom,W., Klein,M.G., Zhang,R., Joachimiak,A. and Chen,X.S. (2006) Crystal structure of SV40 large T-antigen bound to p53 : interplay between a viral oncoprotein and a cellular tumor suppressor. *Genes Dev.*, **20**, 2373–2382.

121. Bista,M., Freund,S.M. and Fersht,A.R. (2012) Domain-domain interactions in full-length p53 and a specific DNA complex probed by methyl NMR spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.*, **109**, 15752–6.

122. Joerger,A.C. and Fersht,A.R. (2010) The tumor suppressor p53: from structures to drug discovery. *Cold Spring Harb. Perspect. Biol.*, **2**, a000919.

123. Ma,B., Pan,Y., Gunasekaran,K., Venkataraghavan,R.B., Levine,A.J. and Nussinov,R. (2005) Comparison of the protein-protein interfaces in the p53-DNA crystal structures: towards elucidation of the biological interface. *Proc. Natl. Acad. Sci. U. S. A.*, **102**, 3988–93.

124. Derbyshire,D.J., Basu,B.P., Serpell,L.C., Joo,W.S., Date,T., Iwabuchi,K. and Doherty,A.J. (2002) Crystal structure of human 53BP1 BRCT domains bound to p53 tumour suppressor. *EMBO J.*, **21**, 3863–72.

125. Joo,W.S., Jeffrey,P.D., Cantor,S.B., Finnin,M.S., Livingston,D.M. and Pavletich,N.P. (2002) Structure of the 53BP1 BRCT region bound to p53 and its comparison to the Brca1 BRCT structure. *Genes Dev.*, **16**, 583–93.

126. Structure of the p53 Tumor Suppressor Bound to the Ankyrin and SH3 Domains of 53BP2 (2012) **274**, 1001–1005.

127. Joerger,A.C., Ang,H.C., Veprintsev,D.B., Blair,C.M. and Fersht,A.R. (2005) Structures of p53 cancer mutants and mechanism of rescue by second-site suppressor mutations. *J. Biol. Chem.*, **280**, 16030–7.

128. 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**, 2396–403.

129. Lu,Q., Tan,Y. and Luo,R. (2008) Molecular Dynamics Simulations of p53 DNA-Binding Domain. **111**, 11538–11545.

130. Zupnick,A. and Prives,C. (2006) Mutational analysis of the p53 core domain L1 loop. *J. Biol. Chem.*, **281**, 20464–73.

131. Saller,E., Tom,E., Brunori,M., Otter,M., Estreicher, a, Mack,D.H. and Iggo,R. (1999) Increased apoptosis induction by 121F mutant p53. *EMBO J.*, **18**, 4424–37.

132. Kitayner,M., Rozenberg,H., Kessler,N., Rabinovich,D., Shaulov,L., Haran,T.E. and Shakked,Z. (2006) Structural basis of DNA recognition by p53 tetramers. *Mol. Cell*, **22**, 741–53.

133. Nagaich, A.K., Zhurkin,V.B., Durell,S.R., Jernigan,R.L., Appella,E. and Harrington,R.E. (1999) p53-induced DNA bending and twisting: p53 tetramer binds on the outer side of a DNA loop and increases DNA twisting. *Proc. Natl. Acad. Sci. U. S. A.*, **96**, 1875–80.

134. Pan,Y. and Nussinov,R. (2007) Structural basis for p53 binding-induced DNA bending. *J. Biol. Chem.*, **282**, 691–9.

135. Nussinov,Y.P. and R. (2009) p53-Induced DNA Bending: the Interplay between p53-DNA and p53-p53 interactions. **112**, 6716–6724.

136. Ma,B. and Levine,A.J. (2007) Probing potential binding modes of the p53 tetramer to DNA based on the symmetries encoded in p53 response elements. *Nucleic Acids Res.*, **35**, 7733–47.

137. Djuranovic,D., Lavery,R. and Hartmann,B. (2002) a / g Transitions in the B-DNA backbone. **30**, 5398–5406.

138. Soussi,T. and May,P. (1996) Structural aspects of the p53 protein in relation to gene evolution: a second look. *J. Mol. Biol.*, **260**, 623–37.

139. Kim,E.-H. and Surh,Y.-J. (2006) 15-deoxy-Delta12,14-prostaglandin J2 as a potential endogenous regulator of redox-sensitive transcription factors. *Biochem. Pharmacol.*, **72**, 1516–28.

140. Bykov,V.J.N., Issaeva,N., Zache,N., Shilov,A., Hultcrantz,M., Bergman,J., Selivanova,G. and Wiman,K.G. (2005) Reactivation of mutant p53 and induction of apoptosis in human tumor cells by maleimide analogs. *J. Biol. Chem.*, **280**, 30384–91.

141. Bykov,V.J.N., Issaeva,N., Shilov,A., Hultcrantz,M., Pugacheva,E., Chumakov,P., Bergman,J., Wiman,K.G. and Selivanova,G. (2002) Restoration of the tumor suppressor function to mutant p53 by a low-molecular-weight compound. *Nat. Med.*, **8**, 282–288.

142. Wang,J.M., Wolf,R.M., Caldwell,J.W., Kollman,P. a and Case,D. a (2004) Development and testing of a general amber force field. *J. Comput. Chem.*, **25**, 1157–1174.

143. Wickstrom,L., Okur,A. and Simmerling,C. (2009) Evaluating the performance of the FF99SB force field based on NMR scalar coupling data. *Biophys. J.*, **97**, 853–856.

144. accelrys DISCOVERY STUDIO VISUALIZER.

145. Vanquelef,E., Simon,S., Marquant,G., Garcia,E., Klimerak,G., Delepine,J.C., Cieplak,P. and Dupradeau,F.Y. (2011) R.E.D. Server: A web service for deriving RESP and ESP charges and building force field libraries for new molecules and molecular fragments. *Nucleic Acids Res.*, **39**, 511–517.

References

146. Wang,Q. and Pang,Y.-P. (2007) Preference of small molecules for local minimum conformations when binding to proteins. *PLoS One*, **2**, e820.

147. Peters,M.B., Yang,Y., Wang,B., Füsti-molnár,L., Weaver,M.N. and Jr,K.M.M. (2011) NIH Public Access. **6**, 2935–2947.

148. Lin,F. and Wang,R. (2010) Systematic Derivation of AMBER Force Field Parameters Applicable to Zinc-Containing Systems. *Methods*.

149. Madhumalar,A., Smith,D.J. and Verma,C. (2008) Stability of the core domain of p53: insights from computer simulations. *BMC Bioinformatics*, **9**, S17.

150. 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. *J. Mol. Biol.*, **409**, 358–68.

151. Fiser, A. and Sali, A. (2003) ModLoop: automated modeling of loops in protein structures. *Bioinformatics*, **19**, 2500–2501.

152. Vriend,G. (1990) WHAT IF : A molecular modeling and drug design program. *Molecules*, **8**, 52–56.

153. Steffen,C., Thomas,K., Huniar,U., Hellweg,A., Rubner,O. and Schroer,A. (2010) The implementation of a fast and accurate QM/MM potential method in Amber. *J. Comput. Chem.*, **31**, 2967–2970.

154. Ho,W.C., Fitzgerald,M.X. and Marmorstein,R. (2006) Structure of the p53 core domain dimer bound to DNA. *J. Biol. Chem.*, **281**, 20494–502.

155. Pan,Y. and Nussinov,R. (2010) Lysine120 interactions with p53 response elements can allosterically direct p53 organization. *PLoS Comput. Biol.*, **6**.

156. Chen,C., Gorlatova,N., Kelman,Z. and Herzberg,O. (2011) Structures of p63 DNA binding domain in complexes with half-site and with spacer-containing full response elements. *Proc. Natl. Acad. Sci. U. S. A.*, **108**, 6456–6461.

157. Ethayathulla,A.S., Nguyen,H.T. and Viadiu,H. (2013) Crystal structures of the DNA-binding domain tetramer of the p53 tumor suppressor family member p73 bound to different full-site response elements. *J. Biol. Chem.*, **288**, 4744–4754.

158. Rökaeus,N., Shen,J., Eckhardt,I., Bykov,V.J.N., Wiman,K.G. and Wilhelm,M.T. (2010) PRIMA-1(MET)/APR-246 targets mutant forms of p53 family members p63 and p73. *Oncogene*, **29**, 6442–6451.

159. Wiech,M., Olszewski,M.B., Tracz-Gaszewska,Z., Wawrzynow,B., Zylicz,M. and Zylicz,A. (2012) Molecular mechanism of mutant p53 stabilization: the role of HSP70 and MDM2. *PLoS One*, **7**, e51426.

160. Pavletich,N.P., Chambers,K.A. and Pabo,C.O. (1993) The DNA-binding domain of p53 contains the four conserved regions and the major mutation hot spots. *Genes Dev.*, **7**, 2556–2564.

161. Lu,Q., Tan,Y. and Luo,R. (2008) Molecular dynamics simulations of p53 DNA-binding domain. *J. Phys. Chem.*, **111**, 11538–11545.

162. Bullock,A.N. and Fersht,A.R. (2001) Rescuing the function of mutant p53. *Nat. Rev. Cancer*, **1**, 68–76.

163. Fiser, a, Do,R.K. and Sali, a (2000) Modeling of loops in protein structures. *Protein Sci.*, **9**, 1753–73.

164. Davis,I.W., Leaver-Fay,A., Chen,V.B., Block,J.N., Kapral,G.J., Wang,X., Murray,L.W., Arendall,W.B., Snoeyink,J., Richardson,J.S., *et al.* (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.*, **35**, W375–83.

165. http://www.rosswalker.co.uk/tutorials/amber_worksh.

166. Loncharich,R.J., Brooks,B.R. and Pastor,R.W. (1992) Langevin dynamics of peptides: The frictional dependence of isomerization rates of N-acetylalanyl-N′ -methylamide. *Biopolymers*, **32**, 523–535.

167. Martyna,G.J., Tobias,D.J. and Klein,M.L. (1994) Constant pressure molecular dynamics algorithms. *J.Chem.Phys*, **101**, 4177–4189.

168. http://en.wikipedia.org/wiki/Physiological_condition.

169. Palmer,L.G. and Civan,M.M. (1977) Distribution of Na+, K+ and Cl- between nucleus and cytoplasm in Chironomus salivary gland cells. *J. Membr. Biol.*, **33**, 41–61.

170. Lodish H, Berk A, Zipursky SL, et al. (2000) Molecular Cell Biology. 4th edition.

171. Link,P. (2008) Biological Buffers.

172. Ugwu,S.O. and Apte,S.P. (2004) The Effect of Buffers on Protein Conformational Stability. *Pharm. Technol.*

173. Ibragimova,G.T. and Wade,R.C. (1998) Importance of explicit salt ions for protein stability in molecular dynamics simulation. *Biophys. J.*, **74**, 2906–11.

174. Pfeiffer,S., Fushman,D. and Cowburn,D. (1999) Impact of Cl- and Na+ ions on simulated structure and dynamics of betaARK1 PH domain. *Proteins*, **35**, 206–17.

References

175. Mark S. Formaneck,L.M. and Q.C. (2008) Effects of temperature and Salt Concentration on the Structural Stability of Human Lymphotactin: INsights from Molecular Simulations. **128**, 9506–9517.

176. Butcher,S., Hainaut,P. and Milner,J. (1994) Increased salt concentration reversibly destabilizes p53 quaternary structure and sequence-specific DNA binding. *Biochem. J.*, **298 Pt 3**, 513–6.

177. Joerger,A.C. and Fersht,A.R. (2008) Structural biology of the tumor suppressor p53. *Annu. Rev. Biochem.*, **77**, 557–82.

178. Bullock, a N., Henckel,J., DeDecker,B.S., Johnson,C.M., Nikolova,P. V, Proctor,M.R., Lane,D.P. and Fersht, a R. (1997) Thermodynamic stability of wild-type and mutant p53 core domain. *Proc. Natl. Acad. Sci. U. S. A.*, **94**, 14338–14342.

179. Fu,T., Min,H., Xu,Y., Chen,J. and Li,G. (2012) Molecular Dynamic Simulation Insights into the Normal State and Restoration of p53 Function. *Int. J. Mol. Sci*, **13**, 9709–40.

180. Wang,Y. (2007) Crystal Structures of Oncogenic , Suppressor and Rescued p53 Core Domain Mutants Suggest the Rescue Mechanisms of Global Suppressor Mutations.