

# **Reactivation of mutant p53: constraints on mechanism highlighted by principal component analysis of the DNA binding domain**

Zahra Ouaray<sup>1,2</sup>, Karim M. ElSawy<sup>3</sup>, David P. Lane<sup>4</sup>, Jonathan W. Essex<sup>1,\*</sup>, Chandra Verma<sup>2,5,6,\*</sup>

<sup>1</sup> School of Chemistry, University of Southampton, Southampton, SO17 1BJ, United Kingdom

<sup>2</sup> Bioinformatics Institute, Agency for Science, Technology and Research, Singapore, 138671, Singapore

<sup>3</sup> Department of Chemistry, Qassim University, Qassim, 51452, Saudi Arabia

<sup>4</sup> p53 laboratory, Agency for Science, Technology and Research, Singapore, 138648 Singapore

<sup>5</sup> School of Biological Sciences, Nanyang Technological University, 60 Nanyang Drive, Singapore 637551

<sup>6</sup> Department of Biological Sciences, National University of Singapore, 14 Science Drive 4, Singapore 117543

\* Corresponding Authors: [J.W.Essex@soton.ac.uk](mailto:J.W.Essex@soton.ac.uk), [chandra@bii.a-star.edu.sg](mailto:chandra@bii.a-star.edu.sg)

Short Title: Principal component analysis of p53 DNA binding domain and its DNA binding site

Keywords: Principal component analysis, p53 DNA binding domain, DNA, induced fit, flexibility, specificity

## 1 ABSTRACT

Most p53 mutations associated with cancer are located in its DNA binding domain (DBD). Many structures (Xray and NMR) of this domain are available in the protein data bank (PDB) and a vast conformational heterogeneity characterizes the various free and complexed states. The major difference between the apo and the holo-complexed states appears to lie in the L1 loop. In particular, the conformations of this loop appear to depend intimately on the sequence of DNA to which it binds. This conclusion builds upon recent observations that implicate the tetramerisation and the C-terminal domains (respectively TD and Cter) in DNA binding specificity. Detailed PCA analysis of the most recent collection of DBD structures from the PDB have been carried out. In contrast to recommendations that small molecules/drugs stabilise the flexible L1 loop to rescue mutant p53, our study highlights a need to retain the flexibility of the p53 DNA binding surface (DBS). It is the adaptability of this region that enables p53 to engage in the diverse interactions responsible for its functionality.

## 2 INTRODUCTION

p53 is a tumour suppressor protein that regulates the cell cycle in multicellular organisms and guards the cells against oncogenic transformations; this has earned it the name of the “guardian of the genome”<sup>1</sup>. This function is brought about by cell cycle arrest, senescence, apoptosis, inhibition of angiogenesis etc <sup>2</sup>. Human p53 is 393 amino acids long and is made up of multiple domains: an N-terminal transactivation domain from residues 1-63; a proline-rich domain from residues 64-92; a DNA binding domain (DBD) from residues 94-312, a tetramerization domain (TD) from residues 320-355 and a C-terminal regulatory domain (Cter) from residues 356-393. p53 is active in its tetrameric form. If p53 becomes damaged, for example by mutagens (chemicals, radiation or viruses), tumour suppression is hampered, resulting in an uncontrolled division of the cells. More than half of human tumours contain a mutation in the p53 protein and 95% of these mutations, commonly referred to as ‘hot spot’ mutations, are located in the DBD. These mutations are classified into three groups depending on their phenotypes <sup>3</sup>: DNA-contact mutants that do not have significant effects on folding and stability; mutations resulting in localized conformational changes, and located near the DNA binding surface (DBS), and are destabilizing by less than 2 kcal.mol<sup>-1</sup> relative to the wild-type; mutations leading to global unfolding of the  $\beta$ -sandwich, and these are destabilizing by more than 3 kcal.mol<sup>-1</sup>. There are 6 hot spot mutations located on or close to the DBS. These residues have been classified as “contact” (Arg248 and Arg273: inhibit DNA interactions) or “structural” mutants (Arg175, Gly245, Arg249 and Arg282: destabilizing mutants) <sup>4</sup> and are frequently mutated as follows: R175H, G245S, R248Q, R249S, R273H, and R282W. 30% of the clinical cancer cases involve p53 “structural” mutants while 20% are “contact” mutants. Increasing the amount of wild type

p53, to treat tumours or prevent them from spreading turns out to be unsuitable as it can result in premature ageing <sup>5</sup>. However, an increasing number of studies <sup>6-9</sup> are beginning to suggest that restoring endogenous p53 function holds potential therapeutic benefits. As a result, many experimental structures of p53 can now be found in the protein data bank (PDB). These are mainly of the TD and the DBD, the two other domains being intrinsically disordered. Monomeric DBD also binds to DNA although its affinity is much lower than that of tetrameric p53 <sup>10</sup>. Biophysical analysis of the wild-type monomeric p53 DBD shows that it is stable in vitro <sup>11, 12</sup>, which makes it a good model for structural studies.

The DBD consists of a central immunoglobulin-like  $\beta$  sandwich (2 anti-parallel  $\beta$ -sheets) which provides the basic scaffold for the DBS. The DBS is characterised by a loop-sheet-helix motif (L1, S2/S2' and H2) and two large loops, L2 (residues 164-194) and L3 (residues 240-250). These latter two loops are stabilized by a zinc ion coordinated to Cys176, Cys238, Cys242 and His179 (Fig. 1 A). p53 binds to DNA in a sequence-specific manner via this domain. The targeted consensus DNA-binding sites are two decameric motifs (half-sites) with a general form RRRCWWGYYY (where: R = A, G; W = A, T; Y = C, T) separated by 0-13 base pairs. The spacer length is correlated with p53 affinity and transactivation; the smaller the spacer, the stronger is the affinity for p53. Two DBDs bind to a half site DNA, forming a symmetrical dimer with a relatively small DBD-DBD interface (Pro177, His178, Arg181, Met243 and Gly244, see figure 1 B). Conserved residues from the loop-sheet-helix motif make 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 engage in several interactions with base pairs in the major groove. To date there are 79 experimental structures of human p53 DBD available in the PDB,

of which 77 have been resolved by Xray diffraction spectroscopy and 2 have been resolved by NMR spectroscopy. The experimental structures of DBD exist in both wild type and mutant states and in various forms including polymeric, DNA-bound, ligand-bound, and protein-bound (Table 1).

In order to understand the links between structural variability and function of the DBD we used Principal Component Analysis (PCA). This is a powerful technique used to reduce the dimensionality of high dimensional systems and distil out the important features of complex data, which in this case are the coordinates of the multiple conformations of the DBD (such as are available from MD simulations or Xray/NMR) <sup>13</sup>. One of the powerful features of PCA, is the identification in conformational phase space of the large spread in coordinates which is associated with high amplitude motions of proteins (which often are also the functional motions). Multivariate PCA performed on NMR structures of the native globular structure of prion protein revealed subdomain differences between the different pathogenic mutants <sup>14</sup>. PCA of the crystallographic and MD generated conformations of Ras isoforms <sup>15</sup>, separated the conformations that were functionally associated with binding to different ligands (GDP or GTP) and also separated the different Ras mutants. A recent study utilizing PCA to examine the conformations of human and mouse experimental structures (NMR and Xray) and molecular dynamics (MD) simulations of the human p53 DBD 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 <sup>16</sup>. The observation that this recessed L1 conformation is not commonly observed in the experimental structures was explained to result from the crystallization procedures used and the associated effects of crystal packing. Half of

the conformations explored by the L1 loop in the first NMR structure solved (apo p53 monomer - 2FEJ) <sup>17</sup> were also thought to be similar to most of the DNA-bound p53. These results stimulated several groups to consider L1 loop mutations and drugs to specifically stabilise L1 as a possible way to rescue mutant p53 <sup>8, 18–20</sup>.

We now build upon this earlier study, but focus only on the human p53 DBD experimental structures, thus allowing us to distinguish between apo, DNA bound, and antigen bound DBDs. The differences between these structures are mainly located in the loops: L1 while p53 is bound to DNA and L3 when it is bound to the LTag antigen. The flexibility of L1 has been highlighted in recent studies (not available at the time the earlier study <sup>16</sup> was carried out) using an extended model of p53 (DBD and TD), that we will refer to as engineered DBD (Table 1) <sup>21–23</sup>. Experimental studies showed that the full p53 protein has a higher DNA binding affinity than DBD (micromolar affinity for the DBD and nanomolar for the structure with both DBD and TD) <sup>24</sup>. To investigate the mechanism underlying the higher affinity of the full-length p53, Halazonetis *et al.* constructed and crystallized an engineered structure of the DBD structure (Xray). In their structural studies, L1 adopts a recessed conformation while binding to the DNA. The recessed L1 loop observed in the engineered structures (and not in the other experimental structures) is thought to be due to the presence of the TD of p53, which explains its absence from the remaining experimental structures (as they do not contain the TD) found in the PDB. According to Halazonetis *et al.* p53 undergoes an induced fit mechanism while binding to DNA <sup>21–23</sup>, which is responsible for the binding specificity of p53 to the DNA. In their studies they also show that the L1 loop does not adopt the recessed state if the DBD is not bound to DNA; however the mutants S121F and V122G in the L1 loop appear to stabilize L1

in its recessed state even in its apo form <sup>23</sup>. In addition to this, the Cter domain has also been implicated in DNA binding <sup>25</sup>. Recently, Prives *et al.* <sup>25</sup> examined the binding of p53 WT, p53 without Cter and p53 with mutated Cter lysines (mutated to arginine or glutamine) to DNA. They discovered four groups of DNA sequences that bind to p53: from those that bind to all p53 variants (more similar to the p53 targeted DNA-binding sites) to those that bind only to the WT (less similar to the p53 targeted DNA-binding sites). The authors concluded that the Cter of p53 appears to participate in DNA binding by stabilising the full-length protein through non-specific DNA interactions and induces conformational changes within the DBD, allowing p53 to bind to diverse DNA sequences.

Together these studies and the findings of the current and earlier PCA analyses <sup>16</sup> are beginning to establish the intimate role of the p53 domains and the flexibility of the DBS (mainly the L1 loop) in specific DNA binding. The availability of the new structures of DBD (NMR, engineered and DNA bound) and new experimental observations on p53 (in particular the finding that the Cter is linked to DNA binding specificity) encourages us to carry out a new PCA-based analysis of the most updated collection in the PDB of structures of the DBD, and PCA of the DNA complexed to DBD (interestingly the DNA component has not been considered in earlier modelling studies). Our findings highlight the flexibility of the p53 DBS and its necessity for specific binding to DNA. This suggests that attempts at stabilizing mutant DBD should ensure the retention of the flexibility of the DBS.

### 3 MATERIAL AND METHODS

#### 3.1 PDB structures

To perform PCA on the experimental structures (Xray and NMR) of p53 DBD, it is necessary to ensure that the structures have the same number of atoms. There are 79 human p53 DBD structures available in the PDB (as of January 2016); however some of them have missing residues, and hence were excluded from our analysis. This resulted in 63 structures. This set of proteins contains wild-type and mutant structures which are either free or bound to DNA/protein/small molecules (Table 1). Two NMR structures exist for the DBD (PDB codes 2FEJ and 2MEJ) containing 36 and 20 conformers respectively. Among the 79 p53 DBD structures, 12 are bound to DNA and used to perform the PCA of DNA.

Among the 61 Xray structures selected (Table 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 TD linked together via a short linker sequence. These four groups were then each divided into three sub-groups depending on whether they were uncomplexed i.e. apo structures, bound to a DNA consensus sequence or bound to a protein/small molecule. The label of the Xray structures in the different PCA analyses will be coloured depending on the subgroup to which the structure belongs (colours indicated in Table 1). In addition, two NMR structures were also 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

structure conformers will be labelled by numbers (2FEJ → red numbers and 2MEJ → blue numbers).

### 3.2 Structural preparation and PCA

The focus of this study is on the monomeric form of p53 DBD. In this context, a PDB file for each chain of the selected structure was created (totalling 219 chains in the 63 structures of which 61 are Xray structures and 2 are NMR structures). As several structures have missing residues at their termini, only residues from Val97 to Glu287 of the DBD (normally defined as extending from residue 94 to 312) 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 and complete DBD structures) was chosen as the reference structure for superimposition; in the rest of this manuscript we will refer to this as the reference structure. Each structure was superimposed on to 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 and these were subject to PCA analysis. The structure alignment was performed using the Visual Molecular Dynamic software (VMD) <sup>26</sup>.

In the case of DNA, the torsional angles of their sugar-phosphate backbone ( $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ ,  $\epsilon$ ,  $\zeta$  and  $\chi$ ), extracted using the program 3DNA <sup>27</sup>, were considered for PCA; for each structure the torsional angles of all the possible dinucleotides were extracted (Fig. 1) and adjusted such that differences from their mean values lie in the range  $-180^0$  to  $+180^0$ , as described elsewhere <sup>28</sup>, which results in 421 dinucleotides obtained from the 12 p53 DNA sequences. PCA has been performed on the dinucleotides using 14 torsional angles (7 per nucleotide).

The PCA technique involves the calculation of the covariance between all dimensions and for an  $n$  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 an  $n$  dimensional 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. The PCA was performed using the package Bio3D via the R statistical programming language <sup>29</sup>, which allows the analysis of protein structure, sequence and trajectories.

## 4 RESULTS

### 4.1 Overview of the possible conformations of p53 DBD monomers

#### 4.1.1 Free DBD conformations observed in an NMR structure: 2FEJ.

PCA was initially carried out on the 219 chains of the 61 Xray and the 2 NMR (36 and 20 conformers) structures chosen. The graph of the variance captured by each principal component (PC) (Fig. 2 A) shows that the first three PCs represent ~60% of the major motions in the DBD while it takes up to 8 PCs to cover 80% of the variance in motions. The loadings of these three PCs (Fig. 2 B) show the regions of the DBD that are the locations of these motions. It is apparent that PC1 captures motions that are globally spread over the DBD structure, with loops L1, L2 and L6 showing high amplitude motion. In contrast, PC2 and PC3 show motions mainly within the L1 and L3 loops, with PC2 displaying a similar variance in the two loops while for

PC3, the L3 loop shows more variance than L1. To examine the contributions of individual structures to these variances, the 219 p53 DBD structures were projected onto the three PCs (Fig. 3 A to C, Fig. S1 displays the PDB codes and chain names of the structures within Figure 3 B). The projection along PC1 and PC2 (Fig. 3 A) shows 4 clusters. The NMR structures 2FEJ (red circle) cluster separately from both the crystallographic structures and from the NMR structure 2MEJ, along PC1. Along PC2, there are three clusters, the central cluster (light blue circle) includes most of the Xray structures and 2MEJ, a top cluster (green circle) formed by all the 2H1L Xray structures (which bind to the L-Tag antigen) and a bottom cluster (orange circle) containing several chains of the engineered structures. The projection along PC3 and PC2 (Fig. 3 C) displays 3 main clusters, most of the Xray structures and the two NMR structures cluster together (light blue circle) while 2H1L (green circle) and some chains of the engineered structures (orange circle) cluster separately from the central cluster along PC2 and PC3.

In a previous PCA study <sup>16</sup> the clustering difference between Xray structures and the NMR structure 2FEJ was thought to originate from differences in compactness between the tighter Xray and the more “loose” NMR structures. However, we find that the more recent p53 DBD NMR structure 2MEJ co-localises with most of the other Xray structures in the central cluster (zoom Fig. 3 B). It is clear that the Xray structures and 2FEJ separate along PC1 (Fig. 3 B) which captures differences in motions mainly in the loops L1, L2 and L6. Two separate PCAs were then performed, one containing only Xray structures and the other only 2FEJ. The three first PCs of these two sets, 2FEJ and Xray capture 32.9 % and 71.4% of the variance respectively (Fig. 4). The first 2 PCs of 2FEJ capture largely motions in loops L1 and L2 while

for Xrays, they capture motions mainly in loops L1 and L3 (Fig. 5 A and B). Although both sets of structures show motion in L1, these motions appear to correspond to different possible conformational minima. Fersht *et al.* termed the two conformations of L1 as the open and closed states of the loop. It has been reported <sup>17</sup> that the closed conformation of 2FEJ is similar to DBD bound to DNA and the open is similar to one of the mutant DBD structures (however the mutant was not specified in the paper). The open conformation is thought to arise from a lack of hydrogen bonds involving the L1 loop <sup>17</sup>. However, hydrogen bond analysis using the program HBonanza (default settings) <sup>30</sup> was performed on the open and closed conformations of 2FEJ (Fig. 6). Both groups of conformations have three hydrogen bonds in L1, two within the loop (Lys120-Ala119 and Thr123-Cys124) and one between L1 and a residue located in the turn between S2'-S3 (Val122-Gln136). However, these three hydrogen bonds are found in only 7% of the ensemble in 2FEJ and it is clear from Figure 6 that they are unlikely to affect the secondary structure of loop L1. A visualisation of the two conformations shows differences in the orientation of Ser121; the sidechain points towards the H2 helix in the open conformation and away from it in the closed conformation (Fig. 7).

The first 3 PCs of 2FEJ also show that L2 (residues 164 to 194) undergoes significant loop motions (Fig. 5 A). The L2 loop 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 coordinated zinc atom <sup>31</sup>. These two loops and a loop-sheet-helix motif form the DBS <sup>32</sup>. When comparing 2FEJ with the recent NMR structure, 2MEJ, which binds the Bcl-2-like protein 1 (Fig. 8), it can be seen that the 20 conformers of 2MEJ (Fig. 8 B) are clustered tightly compared to the 36 conformers

of 2FEJ (Fig. 8 A). A PCA performed on the 20 conformers of 2MEJ shows little flexibility of the structure (data not shown). The flexibility of the second part of loop L2 seems to be seen in all apo DBD monomers (as seen in the NMR; 2FEJ and the xray structure; 3Q01) while these motions are clearly dampened when the region interacts with other proteins (*e.g.* Bcl-2-like protein 1).

While the earlier study <sup>16</sup>, combining the then available Xray and NMR structures had found that apo-p53 appears to explore the same conformational landscape as p53 complexed with other proteins/DNA <sup>16</sup>, our PCA analysis carried out on the new NMR structure (2MEJ) highlights conformational differences between apo and holo p53 within the DBS loops. Although the 2FEJ clustering away from the Xray structures was thought to be due to the technique used, the availability of another NMR structure 2MEJ (this does cluster in the same region of space as the Xray structures) suggests that the flexibility of the DBS seen in 2FEJ (apo-DBD) leads to the separation. We complement the earlier study <sup>16</sup> with an observation that loop L2 adopts multiple conformations in the apo state of DBD and yet is conformationally restricted to a common conformation upon complexation (as seen in Fig. 8).

#### **4.1.2 L3 loop distortion.**

It is clear from Figures 2 and 3 that the differences between the Xray structures are captured mainly along PC2 and PC3 (Fig. 3 C). Projections of the structures along these two PCs indicates that 2H1L (the L-tag antigen bound structure) and some chains of the engineered structures cluster away from the rest of the DBD monomers (Xray and NMR, see Fig. 3C). These two principal components highlight variances in loops L1 (from residue 112 to 124) and

L3 (from residue 240 to 250) (Fig. 2 B). Comparisons (Fig. 9) of these structures after aligning onto the  $\alpha$ -carbons of 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 2H1L and the reference structure is  $\approx 1.5$  Å, and  $\approx 5.6$  Å over loop L3 (RMSD  $\approx 0.9$  Å for the L1 loop). The second cluster contains the engineered structures 3Q05, 3Q06, 3TS8 and 4MZR (chains A and D of all the 4 structures). Their RMSD relative to the reference structure ranges from  $\sim 1.4$ - $2.1$  Å (over all Ca atoms); the RMSD over only loop L1 and ranges from  $5$ - $7.6$  Å while over loop L3 ranges from  $0.7$ -  $0.9$  Å. These results are in agreement with the PCA results showing L1 and L3 conformational differences for 2H1L and some chains of the engineered structures compared to the rest of the Xray and NMR structures.

The L3 loop motion was highlighted in the study of the SV40 large T-antigen (LTag) binding to DBD (PDB code: 2H1L, see Fig. 10 A) <sup>33</sup>. The 2H1L structure contains two LTag hexamers in which each monomer binds to a monomer of the DBD of p53 (Fig. 10 A); all 12 DBD structures were used in the first PCA analysis. The Xray structure of the complex LTag-p53 DBD shows that LTag disturbs the interaction of p53 with DNA by binding at the p53 DBS. The binding in this complex has been considered to be a two-region interaction <sup>33</sup>; the first region contains residues Phe177, His178 and Arg181 from the p53  $\alpha$ -helix 1, which makes direct contact with LTag  $\alpha$ -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 present in region 2, the strong hydrophobic interactions 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 p53 DBD conformation of L3 (Fig. 9 A). This conformational change in L3 is the result of a methionine switch between Met246 (originally buried) and Met243 (Fig. 10 B). In the wild type p53 structure, Met246 is buried and Met243 is exposed (Fig. 10 B) (Met243 is implicated in the formation of stable p53 dimer-DNA complexes<sup>32, 34, 35</sup> and in p53-protein interactions<sup>36–38</sup>). The position of these two methionines is important in partner interactions<sup>33</sup>: while one participates in hydrophobic packing with the DBD of p53 to stabilise the L3 loop, the other is exposed to bind with partners (DNA or protein). The packing also has to be weak 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<sup>32</sup>; in 2H1L the methionine switch induces the positioning of Arg248 in order to achieve optimal interactions with LTag (Fig. 10 A and B) (33).

Two oncogenic mutations in p53 DBD are also known to disturb L3 stability. The hot spot mutation R249S, a structural mutation (destabilising by almost 2 kcal.mol<sup>-1</sup>), induces distortions in the p53 DBS 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  $\alpha$ -helix conformation by residues 239-244. This is similar to the LTag-p53 DBD complex, where the L3 loop movement is associated with a methionine switch. In this mutant, Met243 is buried in a hydrophobic pocket (formed

by Val173 and Leu194) originally occupied by Met246 <sup>39</sup>. Fersht *et al.* solved three Xray structures containing the R249S mutant <sup>39</sup>: 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 p53 DBD). However, 2BIO, which shows enhanced DNA binding compared to the p53 DBD R249S mutant, displays L3 distortion thought to be the deleterious result of the R249S mutation <sup>39</sup>. In the 2BIP and 2BIQ structures, the residues Met243 and Gly244, implicated in the methionine switch, are not resolved by the Xray. Although the Met246 of these two structures is still buried, the lack of a complete loop L3 does not allow us to make any definitive conclusion about the L3 loop in the R249S mutant. These structures were not included in the PCA because of missing atoms or residues, as indicated in the Materials and Methods. The previous p53 PCA study <sup>16</sup> was able to highlight an intermediate of the methionine switch. This L3 conformation was highlighted in the mouse p53 DBD structure without zinc <sup>40</sup>. The present analysis focusing solely on human p53 DBD with zinc (essential to p53 DBS stability <sup>31</sup>).

The second mutation, R282Q (pdb code: 2PCX), defined as a low frequency hot spot mutation, is located in the H2 helix. R282Q destabilises p53 DBD by 8.8 kJ/mol (2.1 kcal/mol) at 283 K <sup>41</sup>. This destabilisation is thought to originate mainly from 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 induction of flexibility in the L3 loop through long-range effects <sup>42</sup>; this loop L3

can exist in 2 different conformations in the mutant. The rigidity of L1 in this mutant R282Q is induced by a rotation of the peptide bond (between Leu114 and His115) stabilised by three hydrogen bonds, including one between Gln282 and Ser116. This latter induces a distortion in Ser116, leading to a shift in L1 towards the hydrophobic core of the  $\beta$  sandwich. The first L3 loop conformation ( $\beta$ -conformer) observed is similar to the L3 loop conformation seen in the wild type structure and is unlike the second conformation ( $\alpha$ -conformer) which exhibits a shift of the  $\alpha$ -carbon atom of Ser241 of about 2.8 Å (Fig. 10 C and D). In the wild-type L3 loop conformation the Ser241 oxygen makes a hydrogen bond with the phosphodiester backbone of the DNA minor groove<sup>42</sup>. Interestingly this last mutant has been crystallised without any rescue mutations included in the structure (PDB code: 2PCX), unlike the R249S mutant Xray structure that contains rescue mutations.

The  $\alpha$ -carbon alignment of the two molecules of the R282Q mutant structure in the unit cell, i.e. conformer  $\alpha$  (L3 loop destabilised, red in Fig. 10 C) and conformer  $\beta$  (wild-type like structure, purple in Fig. 10 C), 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. 10 C). This L3 loop motion arises from a shift of the  $\alpha$ -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. 10 D). 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 (on the  $\alpha$ -carbons) and the RMSD with 2XWR as reference was calculated (Fig. 10 E). 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 – from residue 117 to 120 and from residue 182 to 185 - the complete structures were not used in this RMSD measure). The methionine switched L3 loop form might be a characteristic preventing p53 from binding to DNA, as has been 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 the DBD displays a conformational change in L3 which was highlighted by PCA.

#### 4.1.3 L1 and L2 loop motions.

The structural and functional resolution of a set of engineered structures of the DBD by the group of Halazonetis<sup>42</sup> demonstrated for the first time, that the L1 loop can exist in a ‘recessed’ conformation when bound to DNA, in contrast to all prior structures showing this loop to exist in an ‘extended’ state. They resolved 6 structures with a construct that included the DBD together with the TD, linked by a shortened linker region and containing several stabilizing mutations (Fig. 11 A). Each structure contains the following stabilizing mutations: C135V, C141V, W146Y, C182S, V203A, R209P, C229Y, H233Y, Y234F, N235K, Y236F, T253V, N268D and these mutations were in regions that were not part of the DBS. All the mutations resulted in the same DNA binding specificity and cell functionality as the wild type. Of the 6 structures resolved, 4 are bound as dimers to various DNA sequences (these are 3Q05, 3Q06, 3TS8 and 4MZR) while 2 structures (3Q01 and 4MZI) are in a monomeric apo state. We see the L1 loop motion highlighted in the second cluster in figure 3 (orange circle) containing the 4 engineered DNA-bound structures (3Q05, 3Q06, 3TS8 and 4MZR chains A and D) and this loop motion arises from an induced fit mechanism which has been hypothesized to enable the

specific binding of DBD to the DNA via a dimeric arrangement of p53 (21–23). The DNA sequence targeted in this study consisted of 4 contiguous pentamer repeats (Fig. 11 B). Each engineered dimer binds to DNA with the monomers bound to the inner DNA pentamers adopting an extended L1 loop (WT like conformation) and the monomers bound to the outer DNA pentamers adopting a recessed conformation (Fig. 11 B and C). The structure 3TS8 consists of this engineered construct complexed with the promoter of the CDKN1A (p21) gene, this being the first structure of the p53 DBD bound to a natural p53-response element <sup>22</sup>. 3TS8 shows a more extreme recessed change in the loop L1 due to its specific binding compared to 3Q05 and 3Q06 structures which suggests that the 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 enhance the affinity of p53 for specific DNA sequences but decreases the residence time of p53 on DNA <sup>23</sup>. In contrast to the other engineered structures, 4MZI (the apo p53 engineered model) contains a recessed L1 loop, but upon binding to DNA, L1 is extended (4MZR). The authors proposed that the high affinity and the low residency of 4MZR is due to the high flexibility of the L1 loop confirmed by the crystallographic B-factor of the structures.

The engineered structures 3Q01 (apo state and with L1 loop in ‘extended’ state), 3Q05 and 3TS8, 4MZI (apo state and with L1 loop in ‘recessed’ state, caused by the S121F and V122G mutations) and 4MZR (chain A) were superimposed onto the  $\alpha$ -carbons of the wild type reference structure 2XWR (chain A; Fig. 11 C) and the RMSD of the L1 loop is ~ 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<sup>17</sup>. Indeed, in 2FEJ, the L1 loop changes conformation while DBD is not bound and shows a twist of the L1 loop (Fig. 8). In the engineered structures (3Q05, 3Q06 and 3TS8) the L1 loop is pulled towards the central immunoglobulin-like  $\beta$  sandwich to bind to the outer sequence of the DNA and in the L1 loop mutated structure (4MZR), it is pulled back towards the H2 helix to bind to the inner DNA sequence. Indeed the engineered structures binding to the inner DNA, and the other Xray structures containing only the DBD, show a similar L1 conformation, as shown by the co-localisation of all these structures in the main cluster (Fig. 3 B 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.

This first PCA that we performed on all the p53 DBD structures available highlighted conformational changes localized mainly in the loops. To determine if these loop changes also occur, albeit with low amplitude, in the Xray structures containing only the DBD, separate PCA was performed on these structures. An interesting result was found while performing PCA on wild-type structures that included 26 wild-type structures bound to DNA and 11 in apo states. The first three PCs capture 66.4% of variance (Fig. 12 A) with the first PC highlighting motions along loops L1 and L6 and PCs 2 and 3 capturing motions mainly in L2. The L6 loop is known to be highly flexible<sup>12</sup>, and hence will not be discussed here. The structural projections along PC1 and PC2 (Fig. 13 and Fig. S2, which displays the PDB codes and chain names of each structure) exhibit two clusters along PC1: one containing most of the DBD

structures bound to DNA (red circle) and the other comprising all the free DBDs (blue circle). A visual comparison between the two sets of structures (Fig. 14) shows that the main difference between these two groups is located in the L1 loop. Superimposition of the free DBD on to the reference structure (2XWR chain A) shows a perfect alignment of their L1 loop (Fig. 14 A). In contrast, when 2XWR (blue structure) is superimposed on to the DNA bound structures (Fig. 14 B) the L1 loop is seen to adopt multiple conformations. In the engineered structures with recessed loop 1 (bound to outer DNA) neither Lys120 nor Ser121 contact the DNA compared to the extended L1 conformation (bound to inner DNA) in which both residues bind to DNA<sup>21,22</sup>. The L1 loop contains three residues that contact DNA bases directly: Lys120, Ser121 and Thr123. It has been also shown that mutation of these residues changes the DNA binding specificity<sup>18,43</sup>. Visual comparisons of the structures bound to DNA and those unbound, and between structures bound to inner and outer DNA sequences show that Lys120 and Ser121 in L1 do play a role in DNA recognition (Fig. 15). The three structures compared, 2XWR chain A, 2OCJ chain A and 3KMD chain C, all show different orientations of Ser121, which seems to pull out Lys120 and the L1 loop in the case of the free p53DBD or when bound to an outer DNA sequence. Ser121 adopts the same orientation in all the unbound 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 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.

This PCA also brings out the case of 1TSR and 1TUP structures. As seen in the structure projection (Fig. 13 and Fig. S2 for structure labels) none of the chains constituting these

structures cluster where expected. 1TSR and 1TUP contain three DBD bound to DNA (chain A to C). Only the chain B and C are bound to DNA, chain B binds specifically to DNA while chain C non-specifically <sup>44</sup> which explains the localisation of 1TSR and 1TUP chain A with the free DBD structures (Fig. 13). Also p53 DBD is known to bind to DNA as a dimer of dimers <sup>45</sup>, which is not the case in 1TSR and 1TUP structures. The visualisation of 1TSR reveals that each chain (B and C) bound independently to the DNA. This structures are thought to represent p53 binding to a Holliday junction DNA when p53 is implicated in homologous recombination (DNA repair) <sup>46</sup> and not in genes transcription as the other experimental structures.

A PCA was performed on only the structures bound to DNA as a dimer of dimers (2AHI, 2AC0, 2ADY, 2ATA, 3KMD and 4HJE; these comprise 22 p53DBD monomers; Fig. S3 and 4) did not reveal any differences. This might be due to the fact that in 2AHI, 2ATA, 2ADY and 2AC0 structures there is no distinction between monomers binding to inner/outer DNA (Fig. S3 and S4). In these structures, four DBDs self-assemble onto 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 (Fig. 11 B) in this group of PDBs (Fig. 16). 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.

In contrast to previous findings <sup>16</sup>, we now show that the L1 loop in p53 DBD adopts different conformations depending on whether or not the DBD binds to DNA, and this is captured nicely by the use of PCA. In the particular case of the engineered structures, PCA has also been able

to distinguish between the L1 loop binding to the inner sequence of the DNA (extended L1 loop) and the loop binding the outer part (recessed L1).

## 4.2 DNA PCA

The torsion angles of the DNA sugar-phosphate backbone show different distributions depending on the conformation adopted (defined as A, B, Z forms) in its various apo/complexed states. Thus, it is possible to distinguish between B and A DNA and their sub-forms which exhibit differences in the two torsion angles  $\epsilon$  (C4'-C3'-P-O5) and  $\zeta$  (C3'-O3'-P-O5') for the BI and BII-forms, and differences between  $\alpha$  (O3'-P-O5'-C5') and  $\gamma$  (O5'-C5'-C4'-C3') torsion angles for the A and Crankshaft-forms<sup>28</sup>. Changes in torsion angles have also been observed upon the binding of DNA to proteins when the uncomplexed DNA, which usually exists in the B-form, undergoes transitions in its  $\alpha$  and  $\gamma$  torsion angles which are characteristic of the A-form of DNA<sup>47</sup>. PCA performed on the sugar phosphate backbone of single-strand dinucleotide DNA structures not bound to proteins was shown to be able to discriminate between A and B DNA as well as their respective sub-forms, A and Crankshaft A, BI and BII and the Crankshaft B specific to the B-DNA forms bound to proteins<sup>28</sup>. PCA of the DNA sequences bound by p53 therefore appears to be a useful tool to explore the binding characteristics of p53 DBD.

Several studies have attempted to address this question. Some of the first investigations showed DNA bending when p53 binds to linear DNA as a tetramer and find that the magnitude of the bending depends on the 4-base sequence at the centre of the DNA consensus half sites<sup>48-50</sup>. A later study investigating the different potential modes of binding of the DBD as a tetramer to

DNA, finds contradictory results for linear DNA to which p53 binds and concludes that the bending seen in earlier studies might be due to the DNA sequence or due to a special simulation setup <sup>46</sup>. Those studies were mainly based on the two p53-DNA crystallographic structures available at the time, 1TSR and 1TUP. We currently have access to 12 structures consisting of p53 DBD bound to double helical sequences of DNA; these are used for the current analysis. Among these, seven (3KMD, 4HJE, 3Q05, 3Q06, 3TS8, 4IBU and 4MZR) bind to the DNA as dimers and in the 5 structures (2AHI, 2ATA, 2ADY, 2AC0 and 3D0A), each dimer binds to a different helix and the two DNA helices are parts of separate molecules. The DNA sequences of 1TSR and 1TUP structures were not used in this analysis as they are thought to represent p53-DNA binding during DNA repair <sup>46</sup>.

The 12 DNA helices extracted from the p53 DBD Xray structures are all in the B-form. The sugar phosphate backbone torsion angles were extracted from each dinucleotide (Fig. 1), which results in 421 dinucleotides studied using 14 torsion angles (7 per nucleotide, see Material and Methods). PCA was performed on the 421 dinucleotides using the PCA method applied previously for characterising the Crankshaft B-form of DNA <sup>28</sup>. 60% of the variance is captured by the first 3 PCs (Fig. 17 A), with PC1 and PC2 capturing variance mainly of 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  $\alpha$  and  $\gamma$  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) <sup>28, 47</sup>. 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 a change from B-form DNA. The projection of the structures in the

PC1-PC2 plane (Fig. 17 B) shows a main central cluster surrounded by four clusters. The dinucleotides located in these four clusters (Fig. S5) strikingly originate in the 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 represent p53-DNA binding closer than previous structures suggested<sup>21-23</sup>. The other structures (3KMD, 2AHI, 2ADY, 2AC0, 2ATA and 3DOA) show fewer dinucleotides with large variations in  $\alpha$  and  $\gamma$  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. One symmetry is due to the dinucleotides arrangement when preparing the DNA for the PCA. The division of the DNA helices into dinucleotides resulted in nucleotides being expressed two times in two different dinucleotides (as an example, in structure 3Q06, chain K dinucleotide 7-8 appears in cluster 1 while dinucleotide 8-9 appears in cluster 2, this is due to the common nucleotide 8, see Table 2). As a result cluster 1 highlights the same nucleotide as cluster 2 (the same is true in cluster 3 and 4), which helped us clarify which specific nucleotide is involved by the  $\alpha/\gamma$  transition characteristic to complexed DNA. The nucleotides in clusters 1 and 2 and in clusters 3 and 4 for the three engineered structures are shown within the corresponding DNA structures in Figure 19. Two of the nucleotides in clusters 1 and 2 (Table 2), are common in the three engineered structures (nucleotides 8 and 17 chain K, in 3Q05, 3TS8 and 4MZR). These two nucleotides are located close to (within 5 Å) Arg280 of the different p53 chains (Fig. S5 A and B). In clusters 3 and 4, three nucleotides are common to the four engineered structures (3Q05, 3Q06, 3TS8 and 4MZR), nucleotides 12 and 22 of chain K and nucleotide 34 of chain L. These are also found to be close or in direct contact with

Arg280. 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  $\alpha$  and  $\gamma$  angles of the nucleotides 6, 24 and 33 of the 3Q06 structure (blue coloured Fig. S6 A). Nucleotides 17 and 8 that are not directly bound to a DBD residue are the only two nucleotides of the engineered structures that appear in clusters 1 and 2. These nucleotides although not directly bound to a DBD residue, are located close to Arg280 of chains B and D. Their DNA backbone superimposition shows a good alignment of their  $\alpha/\gamma$  torsional angles (Fig. S6 B).

The other symmetry of cluster 1 and 2 to cluster 3 and 4 results from the position of the nucleotides within the DNA helix. The nucleotides presenting the  $\alpha/\gamma$  transition common to all the engineered structures within the DNA sequence shows that nucleotides from cluster 1 and 2 are located opposite to the ones from cluster 3 and 4 (Fig. 18). In fact, superimposition of the average structures obtained from the nucleotides in clusters 1 and 2 with those from nucleotides in clusters 3 and 4 (Fig. S6 C) show that, as expected, their  $\alpha$  and  $\gamma$  angles are oriented in opposite directions. Concerning the other structures found in the four main clusters, especially 3KMD, these dinucleotides show contact with the p53 DBD (data not shown), but none in the same way as in the engineered structures, except for nucleotide 4 in chain F of 3KMD which is found in both clusters 3 and 4. Nucleotide 4 is, however, not bound to any residue from DBD but rather is adjacent to Arg280 of chain B that binds nucleotide 3 and Arg248 of chain B (NH<sub>2</sub> at 6.85 Å from nucleotide 4 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. 17 B); as these structures are considered as binding as dimers

complexed with two distinct DNA helices, no DNA bending was expected, and their location in the structure projection (Fig. 17 B) confirms our results.

In summary, DNA PCA appears to offer a good tool for highlighting p53 DBD experimental structures that bind properly to p53. Indeed this analysis performed on all the p53 structures known to bind double helix DNA shows  $\alpha/\gamma$  transition (specific to complexed DNA <sup>28, 47</sup>) on mainly nucleotides from the engineered structures, described as being the best model of p53 DBD-DNA binding <sup>21, 22</sup>. Those structures are thought to be better DNA binders because of their extra TD. The small number of nucleotides affected by this bending might be explained by the fact that the engineered structures bind to DNA as dimers and not tetramers.

## 5 CONCLUSIONS

PCA of the experimentally resolved atomic structures of the p53 DBD structures highlights the importance of the flexibility of the DBS as judged from the distribution of the conformational phase space. The availability of new structural information helps us resolve an earlier conclusion that the 2FEJ NMR structure was unique in its separation in the phase space of coordinates because of the NMR techniques used <sup>16</sup>. However with the availability of a new NMR structure, 2MEJ, which is complexed to another protein, and is located in the same region of phase space together with the Xray structures of DBD in its complexed states, it is clear that 2FEJ separates out because it represents an apo state of DBD.

Most of the flexibility observed in the DBS of the NMR structure of the apo-p53 (2FEJ) is located in the loops (L1, L2 and L3). This flexibility was also present in some Xray structures. The specific analysis of these structures allowed us to draw some interesting conclusions about

the necessity to keep p53 DBD flexible enough to allow a plethora of interactions important for its function. The p53 DBD apo state is thought to have a flexible L2 loop (residues 181 to 194). Although most of the free Xray structures of the DBD present an L2 loop conformation similar to that of the DBD bound to DNA, there exist apo structures (such as 3Q01, 4MZR and 2FEJ) that show L2 to adopt different conformations. Unlike L1, no specific conformation of L2 has been associated with the apo state. In the case of the loop L3, two distinct conformations were highlighted by PCA: a commonly observed conformation present in most DBD structures (both apo and complexed) in which Met246 is buried and Met243 exposed and a methionine switched conformation which is mainly observed in DBD linked to oncogenic situations (mutant or binding to an oncogenic partner). These results remind us that Xray spectroscopy is limited in its ability to capture flexibility. It is possible that the loop L3 conformation highlighted in 2H1L and in the mutant R249S may actually represent a conformation accessible to wild type p53 and is not just the result of an oncogenic transformation.

Little information can be found on the L2 and L3 loop in the literature compare to the third DBS loop, L1. Its high flexibility and reputation as a cancer mutation “cold spot” made it the ideal target against p53 related cancers. As 30% of cancers are induced by mutations that destabilize p53 (the structural mutants), several studies suggested increasing p53 stability by reducing the flexibility of some of its regions. A study showed that *C. elegans* p53 is more stable than human p53<sup>51</sup>, and MD simulations of p53 from both species showed that the loop L1, the second part of L2 (after the H1 helix) and L6 were more stable in the worm p53 than in the human p53. The authors proposed mutants stabilising these loops, and highlighted that shortening L6 loop (as L6 is shorter in p53 worm than in the human) of human p53 reduced

the flexibility of p53. Strikingly the shortened L6 loop in human p53 resulted in significant reduction in the flexibility of L1 loop. This idea of inducing a more stable L1 to rescue p53 mutants has been further strengthened by the predominance of similar L1 loop conformations (extended) in most experimental structures. Another suggestion was to make stabilizing mutations within the L1 loop<sup>8</sup>. While several cancer mutations were tested in the context of two second site mutations (H115N and S116M) that stabilized the L1 loop, only R248Q showed reactivation when combined with the H115N mutation (induced transcription of *gadd45*). In a separate study that employed MD simulations to investigate the druggability of R175H mutant p53, a transiently open L1/S3 pocket (appears when L1 adopts an extended conformation, and occurs with similar frequencies in the simulations of WT and other mutants) was seen and screened against small molecules<sup>52</sup>. The authors identified stictic acid to be able to bind within the L1/S3 pocket and lock the position of L1 in its extended conformation (thought to be the most favourable conformation for DNA binding). Stictic acid tested on the R175H mutant reactivated p21 expression<sup>52</sup>.

However, the flexibility of the p53 DBS, and of the L1 loop in particular, has been shown to be essential for DNA binding specificity. Indeed a recent comparison of p53 stability and function in different species showed that the more stable p53 is, the less it interacts (DNA and protein-protein)<sup>53</sup>, and p53-DNA binding studies showed that p53 scans the DNA before binding specifically to it<sup>54,55</sup>. First, p53 binds non-specifically to the DNA, then it drifts along the DNA to find an optimal position on the promoter to which it binds specifically<sup>54,55</sup>. This motion along the DNA is driven by K120 within the L1 loop. These computational studies of p53 DNA binding<sup>54,55</sup> were the first to highlight the importance of an induced fit mechanism

operating, and was later confirmed by the construction and activity of the engineered p53 structures<sup>21–23</sup>. This then suggests that stabilising L1 might limit p53 function (it is likely that this may result in better binding to some partners - DNA and/or proteins, and inhibiting the binding to others).

The PCA of the DBD in its various complexes with sequences of DNA demonstrates that one of the most flexible regions of the p53 DBS is the loop L1. This loop has been shown to adopt different conformations depending on the complexed state of the DBD and on the DNA sequence and can effectively be separated in conformational phase space by PCA. It is now clear that the L1 loop conformations observed in the 2FEJ structures are not DNA binding-like (as stated in the NMR study<sup>17</sup>) but possible L1 conformations of the DBD apo state, and this flexibility was shown by Halazonetis *et al.* to be important for DNA binding specificity<sup>21–23</sup>. In parallel, the PCA of the torsional angles of DNA sequences bound to p53 showed that only dinucleotides bound to the engineered structures present conformations known to be associated to protein binding ( $\alpha/\gamma$  transition). Most of these dinucleotides were found to be directly bound to DBD residues known to be implicated in DNA binding. The engineered structures studied are thought to be the first to show a proper p53 DBD-DNA binding process suggesting an induced fit mechanism (characteristic of most protein binding to DNA)<sup>21–23</sup>. Another interesting result comes from a molecular dynamics study, investigating the conformational changes in tetrameric p53 DBD bound to DNA, highlighting a transition of the DNA form (from B to A -  $\alpha/\gamma$  transition) during the simulation<sup>56</sup>. While this result was not further elaborated in the study, their simulations, based on the Xray structure 2AC0, confirms that p53

within the crystal was not specifically bound and that specific binding only took place during the simulation.

The analysis of the experimentally resolved (Xray and NMR) structures of the DBD of p53 points out a major limitation of the models arrived at to date: it was concluded that the loop L1 in all the free DBD structures adopted a DNA binding conformation. However the current PCA study allowed us to delve deeper into the mechanism, as exemplified by the differences between the DNA bound structures and the apo structures, exposing a new level of complexity in the interactions between p53 and DNA. Our work further stresses that the flexibility of the L1 loop is essential for the multiple functions of p53. This flexibility facilitates the binding of p53 to various DNA consensus sequences; the p53 promoters of cell cycle arrest and DNA repair genes have a specific form (two decameric motifs) but the promoters of apoptosis contain mismatches and are longer. Indeed, a recent study showed that the apoptotic promoter was efficiently bound by p53 only when the Cterminal domain was present <sup>25</sup>. Our PCA analysis further highlights the importance of domains other than the DBD for binding to DNA by showing that only the engineered structure (DBD + TD) displays the  $\alpha/\gamma$  transition specific to DNA bound to proteins. Together our computational results and the observations of Halazonetis *et al* suggest that the Xray structures of p53-DNA complexes exhibit only a non-specific binding and it is the presence of other domains of p53 domain <sup>21-23, 25, 57</sup> and dynamics <sup>56</sup> that combine to enable p53 to bind specifically to DNA.

## **6 ACKNOWLEDGEMENT**

We acknowledge support from the University of Southampton and the Agency for Science, Technology and Research (A\*STAR) under the ARAP program. The authors would like to thank Christopher Bull for his help on PCA. We would also like to thank the research computing platforms Iridis and Emerald.

## 7 REFERENCES

1. Lane DP. p53, guardian of the genome. *Nature* 1992;358.
2. Naga Deepthi C, VVL Pavan Kumar A, Rameshbadu, Indirapriyadarshini U. Role of tumor suppressor protein p53 in apoptosis and cancer therapy. *J Cancer Sci Ther* 2011;1–6.
3. Joerger AC, Allen MD, Fersht AR. Crystal structure of a superstable mutant of human p53 core domain. Insights into the mechanism of rescuing oncogenic mutations. *J Biol Chem* 2004;279(2):1291–1296.
4. Joerger AC, Ang HC, Fersht AR. Structural basis for understanding oncogenic p53 mutations and designing rescue drugs. *Proc Natl Acad Sci U S A* 2006;103(41):15056–15061.
5. Rodier F, Campisi J, Bhaumik D. Two faces of p53: aging and tumor suppression. *Nucleic Acids Res* 2007;35(22):7475–7484.
6. Nikolova P V, Wong KB, DeDecker B, Henckel J, Fersht a R. Mechanism of rescue of common p53 cancer mutations by second-site suppressor mutations. *EMBO J* 2000;19(3):370–378.
7. Bykov VJN, Issaeva N, Shilov A, Hultcrantz M, Pugacheva E, Chumakov P, Bergman J, Wiman KG, Selivanova G. Restoration of the tumor suppressor function to mutant p53 by a low-molecular-weight compound. *Nat Med* 2002;8(3):282–288.
8. Merabet A, Houlleberghs H, Maclagan K, Akanho E, Bui TTT, Pagano B, Drake AF, Fraternali F, Nikolova P V. 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* 2010;427(2):225–236.
9. Lambert JMR, Gorzov P, Veprintsev DB, Söderqvist M, Segerbäck D, Bergman J, Fersht AR, Hainaut P, Wiman KG, Bykov VJN. PRIMA-1 reactivates mutant p53 by

- covalent binding to the core domain. *Cancer Cell* 2009;15(5):376–388.
10. Pavletich NP, Chambers KA, Pabo CO. The DNA-binding domain of p53 contains the four conserved regions and the major mutation hot spots. *Genes Dev* 1993;7(12b):2556–2564.
  11. Bullock AN, Fersht AR. Rescuing the function of mutant p53. *Nat Rev Cancer* 2001;1(1):68–76.
  12. Lu Q, Tan Y, Luo R. Molecular dynamics simulations of p53 DNA-binding domain. 2008;111(39):11538–11545.
  13. Amadei A, Linssen ABM, Berendsen HJC. Essential dynamics of proteins. *Proteins* 1993;17:412–425.
  14. Gendoo DMA, Harrison PM. The landscape of the prion protein's structural response to mutation revealed by principal component analysis of multiple NMR ensembles. *PLoS Comput Biol* 2012;8(8).
  15. Lukman S, Grant BJ, Gorfe AA, Grant GH, McCammon JA. The distinct conformational dynamics of K-Ras and H-Ras A59G. *PLoS Comput Biol* 2010;6(9):4–12.
  16. Lukman S, Lane DP, Verma CS. Mapping the structural and dynamical features of multiple p53 DNA binding domains: insights into loop 1 intrinsic dynamics. *PLoS One* 2013;8(11).
  17. Cañadillas JMP, Tidow H, Freund SM V, Rutherford TJ, Ang HC, Fersht AR. Solution structure of p53 core domain: structural basis for its instability. *Proc Natl Acad Sci U S A* 2006;103(7):2109–2114.
  18. Zupnick A, Prives C. Mutational analysis of the p53 core domain L1 loop. *J Biol Chem* 2006;281(29):20464–20473.
  19. Pan Y, Ma B, Venkataraghavan RB, Levine AJ, Nussinov R. In the quest for stable

- rescuing mutants of p53: computational mutagenesis of flexible loop L1. *Biochemistry* 2005;44(5):1423–1432.
20. Demir Ö, Baronio R, Salehi F, Wassman CD, Hall L, Hatfield GW, Chamberlin R, Kaiser P, Lathrop RH, Amaro RE. Ensemble-based computational approach discriminates functional activity of p53 cancer and rescue mutants. *PLoS Comput Biol* 2011;7(10).
  21. Petty TJ, Emamzadah S, Costantino L, Petkova I, Stavridi ES, Saven JG, Vauthey E, Halazonetis TD. An induced fit mechanism regulates p53 DNA binding kinetics to confer sequence specificity. *EMBO J* 2011;30(11):2167–2176.
  22. Emamzadah S, Tropia L, Halazonetis TD. Crystal structure of a multidomain human p53 tetramer bound to the natural CDKN1A (p21) p53-response element. *Mol Cancer Res* 2011;9(11):1493–1499.
  23. Emamzadah S, Tropia L, Vincenti I, Falquet B, Halazonetis TD. Reversal of the DNA-binding-induced loop L1 conformational switch in an engineered human p53 protein. *J Mol Biol* 2014;426(4):936–944.
  24. Weinberg RL, Veprintsev DB, Bycroft M, Fersht AR. Comparative binding of p53 to its promoter and DNA recognition elements. *J Mol Biol* 2005;348:589–596.
  25. Laptenko O, Shiff I, Freed-Pastor W, Zupnick A, Mattia M, Freulich E, Shamir I, Kadouri N, Kahan T, Manfredi J, Simon I, Prives C. The p53 C terminus controls site-specific DNA binding and promotes structural changes within the central DNA binding domain. *Mol Cell* 2015;57(6):1034–1046.
  26. Humphrey W, Dalke A, Schulten K. VMD: Visual Molecular Dynamics. *J Mol Graph* 1996;14:33–38.
  27. Lu X-J. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res* 2003;31(17):5108–5121.

28. Elsawy KM, Hodgson MK, Caves LSD. 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* 2005;33(18):5749–5762.
29. Grant BJ, Rodrigues APC, ElSawy KM, McCammon JA, Caves LSD. Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics* 2006;22(21):2695–2696.
30. Durrant JD, McCammon JA. HBonanza: A computer algorithm for molecular-dynamics-trajectory hydrogen-bond analysis. *J Mol Graph Model* 2011;31:5–9.
31. Duan J, Nilsson L. Effect of Zn<sup>2+</sup> on DNA recognition and stability of the p53 DNA-binding domain. *Biochemistry* 2006;45(24):7483–7492.
32. Joerger AC, Fersht AR. The tumor suppressor p53: from structures to drug discovery. *Cold Spring Harb Perspect Biol* 2010;2(6).
33. Lilyestrom W, Klein MG, Zhang R, Joachimiak A, Chen XS. Crystal structure of SV40 large T-antigen bound to p53 : interplay between a viral oncoprotein and a cellular tumor suppressor. *Genes Dev* 2006;20:2373–2382.
34. Bista M, Freund SM, Fersht AR. Domain-domain interactions in full-length p53 and a specific DNA complex probed by methyl NMR spectroscopy. *Proc Natl Acad Sci U S A* 2012;109(39):15752–15756.
35. Ma B, Pan Y, Gunasekaran K, Venkataraghavan RB, Levine AJ, Nussinov R. 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* 2005;102(11):3988–3993.
36. Derbyshire DJ, Basu BP, Serpell LC, Joo WS, Date T, Iwabuchi K, Doherty AJ. Crystal structure of human 53BP1 BRCT domains bound to p53 tumour suppressor. *EMBO J* 2002;21(14):3863–3872.

37. Joo WS, Jeffrey PD, Cantor SB, Finnin MS, Livingston DM, Pavletich NP. Structure of the 53BP1 BRCT region bound to p53 and its comparison to the Brca1 BRCT structure. *Genes Dev* 2002;16(5):583–593.
38. Gorina S, Pavletich NP. Structure of the p53 tumor suppressor bound to the ankyrin and SH3 domains of 53BP2. *Am Assoc Adv Sci* 2012;274(5289):1001–1005.
39. Joerger AC, Ang HC, Veprintsev DB, Blair CM, Fersht AR. Structures of p53 cancer mutants and mechanism of rescue by second-site suppressor mutations. *J Biol Chem* 2005;280(16):16030–16037.
40. Kwon E, Kim DY, Suh SW, Kim KK. Crystal structure of the mouse p53 core domain in zinc-free state. *Proteins* 2007;70(1):280–283.
41. Butler JS, Loh SN. Structure, function, and aggregation of the zinc-free form of the p53 DNA binding domain. *Biochemistry* 2003;42(8):2396–2403.
42. Tu C, Tan YH, Shaw G, Zhou Z, Bai Y, Luo R, Ji X. 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* 2008;64(Pt 5):471–477.
43. Saller E, Tom E, Brunori M, Otter M, Estreicher a, Mack DH, Iggo R. Increased apoptosis induction by 121F mutant p53. *EMBO J* 1999;18(16):4424–4437.
44. Joerger AC, Allen MD, Fersht AR. Crystal structure of a superstable mutant of human p53 core domain. Insights into the mechanism of rescuing oncogenic mutations. *J Biol Chem* 2004;279(2):1291–1296.
45. Kitayner M, Rozenberg H, Kessler N, Rabinovich D, Shaulov L, Haran TE, Shakked Z. Structural basis of DNA recognition by p53 tetramers. *Mol Cell* 2006;22(6):741–753.
46. Ma B, Levine AJ. Probing potential binding modes of the p53 tetramer to DNA based

- on the symmetries encoded in p53 response elements. *Nucleic Acids Res* 2007;35(22):7733–7747.
47. Djuranovic D, Lavery R, Hartmann B. a / g Transitions in the B-DNA backbone. *Nucleic Acids Res* 2002;30(24):5398–5406.
  48. Nagaich a K, Zhurkin VB, Durell SR, Jernigan RL, Appella E, Harrington RE. 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* 1999;96(5):1875–1880.
  49. Pan Y, Nussinov R. Structural basis for p53 binding-induced DNA bending. *J Biol Chem* 2007;282(1):691–699.
  50. Nussinov YP and R. p53-induced DNA bending: the interplay between p53-DNA and p53-p53 interactions. *J Phys Chem B* 2009;112(21):6716–6724.
  51. Pan Y, Ma B, Levine AJ, Nussinov R. Comparison of the human and worm p53 structures suggests a way for enhancing stability. *Biochemistry* 2006;45(12):3925–3933.
  52. Wassman CD, Baronio R, Demir Ö, Wallentine BD, Chen C-K, Hall L V, Salehi F, Lin D-W, Chung BP, Hatfield GW, Richard Chamberlin a, Luecke H, Lathrop RH, Kaiser P, Amaro RE. Computational identification of a transiently open L1/S3 pocket for reactivation of mutant p53. *Nat Commun* 2013;4:1407.
  53. Pagano B, Jama A, Martinez P, Akanho E, Bui TTT, Drake AF, Fraternali F, Nikolova P V. Structure and stability insights into tumour suppressor p53 evolutionary related proteins. *PLoS One* 2013;8(10):1–17.
  54. Pan Y, Nussinov R. Preferred drifting along the DNA major groove and cooperative anchoring of the p53 core domain: mechanisms and scenarios. *J Mol Recognit* 2010;23(2):232–240.
  55. Pan Y, Nussinov R. Lysine120 interactions with p53 response elements can

- allosterically direct p53 organization. PLoS Comput Biol 2010;6(8).
56. Pan Y, Nussinov R. Cooperativity dominates the genomic organization of p53-response elements: a mechanistic view. PLoS Comput Biol 2009;5(7).
  57. Abramo MD, Be N, Desideri A, Levine AJ, Melino G, Chillemi G. The p53 tetramer shows an induced- fit interaction of the C-terminal domain with the DNA-binding domain. Oncogene 2015:1–10.

## 8 TABLE LEGENDS

Table 1. List of the Xray structures of human p53 used to perform principal component analysis (PCA). The colours in parentheses are used in the figures of structure projections along the different PCs.

Table 2. List of the dinucleotides found in the clusters. The first 4 alpha-numeric code correspond to the PDB id, 'Ch' designates the chain containing the dinucleotide and 'Nuc' the dinucleotide number.

## 9 FIGURE LEGENDS

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

Figure 2. All monomers' PCA. **A)** Proportion of variance (in %) captured by each principal component (PC) from a data set of 219 monomeric p53DBD Xray and NMR structures. The first 3 PCs contribute 59.4% of the variance of the dataset. **B)** Tube representation of the motion projections along the three first PCs (PC1, PC2 and PC3) of the 219 structures using VMD. The DBS loops are coloured in blue. **The varying thickness along the tubes represents the variance (fluctuation) of each residue along that particular PC.**

Figure 3. Structure projections along the three first PCs. **A)** Projections along PC1 and PC2. **B)** zoom of the PC1-PC2 main cluster (light blue circle in panel A). **C)** projections along PC3 and PC2. Xray structures are coloured according to table 1. NMR structures are represented by stars (2 FEJ containing 36 conformers → red circled and 2MEJ containing 20 conformers → dark blue circled).

Figure 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 Xray structures, 163 DBD monomers. The three first PCs capture 71.4% of variance.

Figure 5. Tube representation of the motion projections along the first 3 PCs (PC1, PC2 and PC3) using VMD. **A)** 2FEJ, 36 conformers. **B)** All Xray monomers, 163 monomers. The DBS loops are coloured in blue. **The varying thickness along the tubes represents the variance (fluctuation) of each residue along that particular PC.**

Figure 6. Hydrogen bonds in the L1 loop. p53DBD is shown in cartoon and coloured by secondary structure and the residues implicated in hydrogen bonds are shown in liquorice with the hydrogen bonding atoms shown as spheres.

Figure 7. 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 8. Comparison of the NMR structures. **A)** 2FEJ, 36 conformers. **B)** 2MEJ, 20 conformers. The structures were aligned on the reference structure (2XWR chain A) and are shown in cartoon and coloured by secondary structure.

Figure 9. L1 and L3 loop conformational comparison. **A)** Chain M (green) of PDB 2H1L is aligned on to chain A of PDB 2XWR (reference structure, blue). **B)** Engineered structures (3Q05 → red, 3TS8 → orange and 4MZR → black) are aligned on to chain A of PDB 2XWR (reference structure in blue).

Figure 10. L3 loop conformations. **A)** 2H1L (PDB id), LTag hexamer (light blue) bound to six p53 DBD monomers (dark blue). **B)** Methionines 243 and 246 in PDB id 2XWR, chain A (blue) and in PDB id 2H1L, chain M (green). 2XWR and 2H1L were aligned on to 2AHI (in trace, grey) **C)** Structural comparison of 2PCX  $\alpha$ -conformation (red) and  $\beta$ -conformation (purple) of the L3 loop when superposed on to 2XWR chain A (blue). **D)** Met243 adopts

different conformations in 2PCX  $\alpha$ -conformation (red),  $\beta$ -conformation (purple) and 2XWR (blue). **E)** Structural comparison of 2H1L (green), 2BIO (red) aligned on 2XWR (blue).

Figure 11. Engineered p53. **A)** 3Q05 monomer coloured by secondary structure. **B)** DNA sequence to which 3Q05 and 3Q06 bind containing the four contiguous pentamer repeats (a to d). In the manuscript the outer DNA sequences refer to a and d and the inner sequences refer to b and c. **C)** Superposition of the DBD of the engineered p53 structures on to 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).

Figure 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 Xray structures. The 3 first PCs contribute to 66.4% of the variance of the dataset. **B)** Tube representation of the motion projections along the 3 first PCs (PC1, PC2 and PC3) of the set of structures using VMD. The DBS loops are coloured in blue. **The varying thickness along the tubes represents the variance (fluctuation) of each residue along that particular PC.**

Figure 13. Projection of the set of 37 structures on to principal components PC1 and PC2. The structure were coloured as follow: the monomers bound to DNA (purple) and the free monomers (blue).

Figure 14. Comparison of apo and DNA bound p53DBD structures. **A)** Apo p53DBD monomers aligned on to the  $\alpha$ -carbons of the 2XWR chain A (blue): 2OCJ all chains (red) 2YBG all chains (green) **B)** DNA bound p53DBD monomers aligned on to 2XWR chain A

(blue): 2AHI chain A (purple), 3KMD chain A (red) and 4HJE chain A (green) aligned on carbon alpha of 2XWR chain A.

Figure 15. The orientations of L1 loop residues Lys120 and Ser121 in the wild-type p53DBD structures: 2XWR (blue), 2OCJ (green), 3KMD (red). 2OCJ and 3KMD were aligned on to 2XWR.

Figure 16. 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.

Figure 17. 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)** Dinucleotide projection onto PC1 and PC2.

Figure 18. Nucleotides with  $\alpha/\gamma$  transition in the engineered structures. Common nucleotides for all engineered structures found in clusters 1 and 2 (liquorice representation, blue coloured) in cluster 3 and 4 (liquorice representation red coloured).

Figure S1. Structure projections along the two first PCs of the 219 p53DBD (main structures), along PC1 and PC2. Xray structures are labelled by PDB name and chains and coloured according to table 1. The NMR structure is numbered by conformers and the numbering are coloured in blue.

Figure S2. Projection of the set of 37 structures of p53 DBD structure apo and DNA bound 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 S3. 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.

Figure S4. 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 sequence are coloured in purple while the monomers bound to the outer part are coloured in green.

Figure S5. Dinucleotide location on the structures. **A)** Cluster 1 **B)** Cluster 2 **C)** Cluster 3 **D)** Cluster 4. VMD visualisation of p53 DBD monomer in cartoon coloured by structure, dinucleotides in bonds coloured by name, and the rest of the DNA helix in line representation and coloured by name.

Figure S6.  $\alpha$  and  $\gamma$  torsional angle comparison of the common nucleotides found in clusters 1-2 and in clusters 3-4 from the engineered structures. **A)** DNA backbone alignment of the common 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 the three nucleotides from 3Q06 structure that does not aligned (Nuc. 3, 24 and 33) are coloured in blue. **B)** DNA backbone alignment of the redundant nucleotides in cluster 1 and 2 (3TS8-nuc8-chainK chosen as reference). **C)** the average structures from the nucleotides presented in A and B were calculated using ptraj and were aligned. The average structure calculated from

the common nucleotide in cluster1-2 is coloured in blue. The visualisation was performed in VMD.