Cancer-related Mutations with Local or Long-range Effects on an Allosteric Loop of p53

Degn, Kristine; Beltrame, Ludovica; Dahl Hede, Freja; Sora, Valentina; Nicolaci, Vincenzo; Vabistsevits, Marina; Schmiegelow, Kjeld; Wadt, Karin; Tiberti, Matteo; Lambrughi, Matteo

Total number of authors: 11

Published in: Journal of molecular biology

Link to article, DOI: 10.1016/j.jmb.2022.167663

Publication date: 2022

Document Version
Publisher's PDF, also known as Version of record

Citation (APA):
Degn, K., Beltrame, L., Dahl Hede, F., Sora, V., Nicolaci, V., Vabistsevits, M., Schmiegelow, K., Wadt, K., Tiberti, M., Lambrughi, M., & Papaleo, E. (2022). Cancer-related Mutations with Local or Long-range Effects on an Allosteric Loop of p53. Journal of molecular biology, 434(17), [167663]. https://doi.org/10.1016/j.jmb.2022.167663
Cancer-related Mutations with Local or Long-range Effects on an Allosteric Loop of p53

Kristine Degn¹, Ludovica Beltrame¹,², Freja Dahl Hede¹, Valentina Sora¹,², Vincenzo Nicolaci², Marina Vabistsevs²†, Kjeld Schmiegelow³, Karin Wadt⁴, Matteo Tiberti², Matteo Lambrughi² and Elena Papaleo¹,²

¹ - Cancer Systems Biology, Section for Bioinformatics, Department of Health and Technology, Technical University of Denmark, 2800 Lyngby, Denmark
² - Cancer Structural Biology, Danish Cancer Society Research Center, 2100 Copenhagen, Denmark
³ - Department of Pediatrics and Adolescent Medicine, University Hospital Rigshospitalet, Copenhagen, Denmark
⁴ - Department of Clinical Genetics, Copenhagen University Hospital Rigshospitalet, Denmark

Correspondence to Elena Papaleo: Cancer Systems Biology, Section for Bioinformatics, Department of Health and Technology, Technical University of Denmark, 2800 Lyngby, Denmark. elpap@dtu.dk, elenap@cancer.dk (E. Papaleo) @KristineDegn (K. Degn), @LudoBeltrame (L. Beltrame), @teoKusaMatteo (M. Tiberti), @LambrighMatteo (M. Lambrughi), @ElePapaleo (E. Papaleo)

https://doi.org/10.1016/j.jmb.2022.167663

Edited by Igor Berezovsky

Abstract

The tumor protein 53 (p53) is involved in transcription-dependent and independent processes. Several p53 variants related to cancer have been found to impact protein stability. Other variants, on the contrary, might have little impact on structural stability and have local or long-range effects on the p53 interactome. Our group previously identified a loop in the DNA binding domain (DBD) of p53 (residues 207–213) which can recruit different interactors. Experimental structures of p53 in complex with other proteins strengthen the importance of this interface for protein–protein interactions. We here characterized with structure-based approaches somatic and germline variants of p53 which could have a marginal effect in terms of stability and act locally or allosterically on the region 207–213 with consequences on the cytosolic functions of this protein. To this goal, we studied 1132 variants in the p53 DBD with structure-based approaches, accounting also for protein dynamics. We focused on variants predicted with marginal effects on structural stability. We then investigated each of these variants for their impact on DNA binding, dimerization of the p53 DBD, and intramolecular contacts with the 207–213 region. Furthermore, we identified variants that could modulate long-range conformation of the region 207–213 using a coarse-grain model for allostery and all-atom molecular dynamics simulations. Our predictions have been further validated using enhanced sampling methods for 15 variants. The methodologies used in this study could be more broadly applied to other p53 variants or cases where conformational changes of loop regions are essential in the function of disease-related proteins.

© 2022 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).
Introduction

The TP53 gene, and its protein product p53, is best known as the guardian of the human genome but recently also as the guardian of cellular homeostasis. However, while p53 is predominantly known as one of the most important tumor suppressors, it can support cell survival at the benefit of cancer cells in certain scenarios, implying that p53 has a more complex role at the cellular level than previously assumed. This complexity may be investigated by looking at p53 as a dual-role gene, i.e., a gene that acts in a context-dependent manner as a tumor suppressor or tumor promoter. The p53 protein has remarkably diverse functions in the cell, ranging from transcription-dependent (mostly associated with its nuclear localization) and transcription-independent (mostly cytosolic) functions. For example, p53 transcription-independent functions that depend on its cytosolic pool are related to mitochondrial apoptosis, autophagy or accumulation of Ca²⁺ ions within the endoplasmic reticulum. In normal cells, the p53 protein is generally expressed at a low level, and it is often post-translationally modified and stabilized upon stimuli such as DNA damage, cellular stresses, and other alterations.

The complexity of p53 functions and behavior suggests that mutations can have heterogeneous consequences depending on the context and cellular compartment. TP53 is among the most frequently mutated genes in human cancers, with a high occurrence of missense mutations, which can result in loss of its transcriptional activity or gain-of-function (GOF).

Recent research has highlighted that cancer-predisposing syndromes are a major cause of malignancies development, especially in pediatric patients. Germline variants in the TP53 gene cause the so-called Li-Fraumeni Syndrome (LFS), which is often underdiagnosed, and approximately one-third of the affected people develop cancer during their childhood. Mutant variants of p53 have been claimed to be undruggable for a long time, but recent studies suggest the opposite and pave new directions for the possibility of rescuing structural mutants of p53.

p53 is a multi-domain protein, and the vast majority of mutations found in human cancers are located in the DBD (residues 101–292 as reported in UniProt). P53 DBD features an immunoglobulin-like β-sandwich fold with several disordered loops, including those for DNA binding. The L1 loop (residues 113–123) can adopt different conformations that are more (extended conformations) or less (recessed states) favorable for interactions with DNA through K120. p53 DBD evolved to be characterized by marginal stability. The effects exerted by p53 variants at the structural and functional levels can be different and context dependent. Some alterations might impact interaction with DNA or other biological partners, post-translational modifications (PTMs), or alter structural stability. Other variants can have more elusive and long-range effects, acting as allosteric variants. It cannot be ruled out that some variants can affect different properties at the same time, influencing the structural stability and exerting local effects on binding or more complex long-range mechanisms. In previous work, we illustrated the importance of the dynamics and conformations of the S6-S7 loop (i.e., the loop connecting the beta-strands 6 and 7 of the DBD) to recruit different binding partners and for long-range communication with the DNA binding loops. We also highlighted the importance of the disordered region upstream to the DBD, i.e., residues 91–100, in modulating the conformation of the S6-S7 loop. We thus focus our investigation on a construct including the DBD and the disordered region (namely p5391-289). Moreover, molecular dynamics (MD) simulations of destabilizing mutants of p53 DBD suggested that open states of the S6-S7 loop can form a pocket that is likely to be druggable.

Here, we studied 1132 variants in p5391-289 for their effect on the allosteric communication regulating the conformation of the S6-S7 loop of p53. As a consequence, these variants are likely to reshape the interaction preferences of the protein or the long-range communication to the DNA binding site. A schematic workflow of the study is reported in Figure 1.

Results

Overview of the variants investigated in this study

We aggregated all the variants that have been found in the p5391-289 and reported in cancer genomic databases, including CBioPortal and COSMIC (see Materials and Methods), for a total of 1091 different variants (Table S1, Figure 2(A)). All the positions of the DBD have been found to be mutated in the aggregated dataset. We compared these mutations to 1098 somatic and 237 germline missense variants reported in the TP53 database and we observed that the majority overlap with somatic variants (Figure 2(A)). Nevertheless, more than 150 variants have been found both as somatic and germline variants, depending on the cohorts under investigation (Figure 2(A)). 215 variants in our aggregated dataset from cBioPortal and COSMIC were not reported in the TP53 database. We also noticed that the majority of the germline variants overlapped with the somatic ones and seven of them were not found in our aggregated dataset (i.e., N210Y, R290H/L/C, K292I, F134Y, and K291R).

In addition, we collected 62 p53 variants found in germline samples of patients who developed cancer either through curation from the literature or found in Danish patients with child-
Our work can be divided into two parts: i) the identification of cancer-related p53 mutations to study for local and long-range effects, ii) and their respective study. The cancer mutations originate from several sources such as COSMIC, cBioPortal, and gnomAD and cover both somatic and germline mutations. The mutations were selected based on their position, as the assessment requires coverage by the protein model, and either (i) a neutral effect on structural stability, or (ii) the possibility of allosteric or local effects on the conformation of the S6-S7 loop. We retained, upon free energy calculations, 395 variants covering 143 wild-type residues. Thus, the variants were analyzed based on local and long-range effects on S6-S7. We predicted local effects from the different variants on DNA interaction, dimerization of the p53 DBD, and intramolecular contacts. Furthermore, we used methods based on protein structure networks, dimensionality reduction, and a coarse-grain model for allostery to predict the effects of the variants on the conformation of the S6-S7 loop. Finally, we used metadynamics to verify if the variants can exert an effect on the populations of the conformation of the S6-S7 loop.
hood cancer (i.e., R273H and R213G),\textsuperscript{14} or affected by LFS (OMIM entry 151623, i.e., R273H, R248W, R282Q, and R175H),\textsuperscript{20} as summarized in Table S1. In addition, we identified a previously reported germline variant (R181H) in two Danish families (see Materials and Methods). The variant has previously been identified in individuals with early-onset breast cancer and other LFS-related cancer but does not clinically appear to result in classic LFS syndrome.\textsuperscript{43–50} Most variants from our literature curation overlapped with the variants analyzed in Figure 2(A), and only V110G\textsuperscript{36} and P295S\textsuperscript{38} emerged as additional p53 germline variants.

Comparing these data with 118 variants reported in gnomAD\textsuperscript{51} for p53, we also identified three additional variants occurring with low frequency in the population that have not been reported in any of the datasets mentioned above (i.e., S116A, L114S, and G112A) which we also included in this study.

Selection of structures to use in the study

We scrutinized 31 available structures for the p53 DNA binding domain (DBD) and selected high-resolution ones in free or DNA-bound states for structure-based calculations (Table S2). We refer to p53 in the different assemblies as p53\textsuperscript{91-289}_monomer and p53\textsuperscript{91-289}_dimer. In particular, we used the monomeric state as a convenient minimalistic model to understand the effect of the mutations on the DBD, as we did in our previous study.\textsuperscript{25} Additionally, we studied the monomeric state in complex with DNA (DNA:p53\textsuperscript{91-289})

Figure 2. Overview of the datasets used in the study. (a) The Venn diagram illustrates the source of the cancer variants. The majority of mutations come from COSMIC, cBioPortal and the TP53 database. Ten variants are not covered by these, including N120Y, R290H, R290L, R290C, K292I, F134Y, S116A, L114S, and G112A. Furthermore, the Venn diagram depicts how the same variants have been found either as somatic or germline, depending on the study. The encircling bar chart represents the sequence of the p53 DBD starting with tryptophan at position 91 and ending with leucine at position 289. The height of the bar indicates the number of variants at that site in the datasets. G245, the position with the highest number of alterations, has 13 observed amino acid substitutions. Furthermore, sites of interest for this study such as the L1 loop, S6-S7 loop, and zinc-binding sites are marked on the chart and colored in grey. (b) The flow diagram in panel b illustrates the workflow for the selection of the experimentally determined structures to use for the calculations carried out in this study. We aimed to select: i) a structure to use to represent the monomeric state of p53 DBD, ii) a structure to use in complex with DNA, iii) and a structure representative of the dimeric assembly observed when p53 is localized in the cytosol. We selected 29 out of 221 structures for further inspection and we selected the PDB entries 2XWR (chain A), 3KZ8 (chain B), and 2OCJ (chain B and D).
Figure 3. Comparison of methods used to predict effects on structural stability. We evaluated the performances of the methods and influence of the initial structure on the prediction of changes in folding free energy upon mutation of p53 DBD. To this goal, we compared the predicted values to experimentally determined ones, selecting variants of p53 that are reported in the ThermoMutDB database. (a) Illustrates how nine of the mutations have multiple estimated experimental values, depicting the variability in the experimental results. (b) The panel shows the average $\Delta\Delta G$ experimental values for each of the ThermoMutDB database mutations. It is these average values used for the comparison. (c) The panel illustrates the linear relationship as Pearson correlation between the experimental values and the predicted values of the five different combinations: FoldX calculations based on a molecular dynamics (MD) ensemble, the X-ray structure, the X-ray structure after refinement with PDB-REDO, along with Rosetta calculations based on an X-ray structure and the structure after PDB-REDO refinement. The calculations based on FoldX and Rosetta using the MD ensemble and the X-ray structure respectively featured the highest Pearson coefficients. (d) We evaluated the performance of the different combinations. In this case, we divided in classes of effect, i.e., stabilizing ($\Delta G < -1.2$ kcal/mol), neutral ($-1.2 \leq \Delta G \leq 1.2$ kcal/mol), destabilizing ($1.2 < \Delta G < 3$ kcal/mol), and highly destabilizing (>3 kcal/mol). The evaluation is not solely based on correlation and introduces flexibility. FoldX has the best accuracy for this case.
to assess the impact of the mutations on DNA binding. We used the dimeric state to simulate the conformation in the cytosol, where the favored state for p53 appears to be a dimer, according to cellular studies in which the two monomers interact together in a conformation different from the ones observed in the DNA-bound tetramer of p53. The process of selection is illustrated in Figure 2(B).

Evaluation of the approaches used for free energy calculations to predict the effects of mutations on structural stability

Here, we aim to identify mutations that could alter the conformation of the S6-S7 loop either through local or distal (allosteric) effects and, in turn, affect the function of the loop in engaging in protein–protein interactions to modulate the transcriptional activity of p53. In the first part of the work, we aimed to identify mutations that are expected to have a marginal effect on the protein structural stability of p5391-289_monomer and, as such, could be more meaningful for the characterization of their functional impact. We selected two different methods based on different energy functions and sampling approaches to estimate changes in folding free energy (i.e., FoldX and Rosetta, see Materials and Methods).

Before performing the free energy calculations to select the variants to retain in the study, we evaluated the accuracy of the predictions against a pool of p5391-289 variants extracted from the ThermoMutDB database, which reports experimentally determined unfolding ΔG values. We identified 43 annotated variants with single mutations. Of these variants, 35 correspond to mutations that we also identified in the cancer datasets described above (Table S1). We identified nine mutations with at least two different experimental measurements for the same mutation (Table S1), which allows us to appreciate the differences in the estimates at the experimental level (Figure 3(A)). We could observe that some discrepancies in the experimentally determined values for these substitutions are likely due to differences in the experimental protocols or in the biophysical methods. Overall, the standard deviations for these experimental measurements were in the range of 0.01–1.21 kcal/mol, which suggests a large spread in the values. To avoid potential over-reliance on the singular experimental values, we used the average value for each amino acidic substitution (Figure 3(B)).

Furthermore, we aimed to assess the accuracy of different approaches to predict changes in folding ΔG-values, along with the influence of using different initial structures for the calculations. Thus, we compared the predicted and experimentally determined ΔG values for the 43 mutations reported in ThermoMutDB. The initial structures were an X-ray structure covering p5391-289, the same structure upon refinement with PDB-REDO, and an ensemble of 20 structures from a Molecular Dynamics (MD) simulation (see Materials and Methods). We collected the following calculations: i) scans with the Rosetta-based free energy function for the X-ray and X-rayPDB-REDO structures, ii) scans with the FoldX free energy function for the X-ray, X-rayPDB-REDO, and the MD ensemble. The MD ensemble was used exclusively for FoldX scans to evaluate if it was possible to compensate for limitations with backbone remodeling of FoldX using an average over an ensemble of structures. The accuracy of the predictive methods was evaluated by the linear relationship between experimental measurements and predictive estimation (Figure 3(C)). According to the values reported in Figure 3(C), we observed that: i) there is no clear advantage of applying the PDB-REDO refinement to repair the X-ray structure before the mutational scans of the p5391-289_monomer structure used in this study, ii) an MD ensemble yields better results than a single structure with FoldX and, finally (iii) while the Rosetta algorithm on the X-ray structure generates a better correlation than any other method, the algorithm cannot be deemed superior under these conditions due to the low correlation found on the X-rayPDB-REDO structure. In light of the variability observed among the experimentally derived values (Figure 3(A)) and the fact that the empirical energies functions used in the study generally properly recapitulate differences observed for classes of effects rather than exact ΔG values, we classified the mutations found in ThermoMutDB in four different classes: i) stabilizing (ΔG < −1.2 kcal/mol), ii) neutral (−1.2 ≤ ΔG < 1.2 kcal/mol), iii) destabilizing (1.2 < ΔG ≤ 3 kcal/mol), and iv) highly destabilizing (>3 kcal/mol) (Table S1). We then applied the same criteria to the predicted values from Rosetta or FoldX. Introducing classes based on threshold values offers some flexibility and may help overcome the limitations due to variability in the predicted and experimental values. We then evaluated the prediction accuracy according to this classification (Figure 3(D)) and observed that FoldX performed slightly better than Rosetta on p53 in this application.

Hence, the validity of the predictive values has been evaluated based on multiple methods and structural starting points. While regression correlations and accuracy measures of a confusion matrix are not comparable, both approaches illustrate reasonable predictive powers indicating that both FoldX and Rosetta can be applied for this study. Based on these analyses, we excluded the combination of the X-rayPDB-REDO structure and the Rosetta energy function due to the poor agreement with the experimental data from further analyses.

Lastly, we assessed the extent to which each of the remaining methods classified the mutations in the same way. Accordingly, each amino acid

K. Degn, L. Beltrame, F. Dahl Hede, et al. Journal of Molecular Biology 434 (2022) 167663
substitution was classified according to its associated change in free energy using the aforementioned criteria. A majority vote then determined a single class for the mutation. Based on these two pieces of information, we calculated a fraction indicating the level of agreement leading to the majority vote classification. Given the four methods, this fraction could be either 0.5 for a 2/4 majority, 0.75 for a 3/4 majority, or 1.0 for a complete agreement. We also calculated the percentage of cases in each class fraction to evaluate if method agreement is higher in one class than another. According to this analysis, the methods agree when it comes to the distinction between neutral and highly destabilizing mutations, while the stabilizing and destabilizing mutations are more difficult to discriminate from the neutral mutations (Figure S1). This result is also in line with the fact that FoldX energy function suffers from limitations in predicting stabilizing mutations.57–59

Identification of p53 variants found in cancer studies with impact on structural stability

In light of the comparison with the experimental data, we used the saturation scans to extract information on all the mutations reported in Figure 4(A). For this part of the work, we relied on the results from the MD ensemble with FoldX which is in better agreement with experimental data than using single conformations (Figure 3 (C)), and the results with Rosetta on the X-ray structure. We thus classified each of the p53 mutations collected in the datasets reported in Figure 2(A) in these four categories: stabilizing, neutral, destabilizing, and highly destabilizing for each of the two approaches mentioned above (Table S1). As mentioned above, our goal was to identify mutations with the potential to alter the transcription-independent functions of p53 through local or long-range effects on the conformation of the S6-S7 loop (Figure 4(B)). In this context, for mutations that caused an increase in folding ΔG values, the destabilization of the protein structure is expected to be the predominant signature at the cellular level and hide other functional effects due to changes in protein–protein interactions. In consideration of this, we applied strict selection criteria and retained only those mutations predicted to be neutral (or stabilizing) from both FoldX and Rosetta. Accordingly, we identified 395 likely neutral variants for the protein structural stability (Table S1). These mutations are distributed over 143 residues of the protein (Figure 4(B)).

Furthermore, we used the MD ensemble to derive a Protein Structure Network based on atomic contacts (acPSN) representation and identified residues with hub behavior,60–61 as detailed in the Materials and Methods. Hub residues in a PSN are known to be important hotspots for structural stability or key residues in long-range structural communication. We thus evaluated if each of the selected 143 mutation sites has a hub behavior (Figure 4(C)). 45 sites are predicted as hub residues but the majority with a degree of 3 or 4. Some of the mutations at these sites might not affect this property as they could retain similar capabilities to engage in atomic contacts (for example mutations to phenylalanine at Y163, Y205, Y234, and Y236) and should be re-assessed upon analyses of MD simulations of each mutant variant (see below).

Local effect of the mutations on the DNA binding interface

In the nucleus, p53 binds DNA to activate or inhibit transcription. In this compartment, the protein is mostly found in its tetrameric state, where each monomer binds its DNA target.52–57 We used a minimalistic model for the interaction between a monomer of p53 and DNA (Figure 2 (B)) that resembles the interaction of a p53 mono-

---

4. Selection of neutral variants for structural stability. (a) On the left, the panel displays the structure of the DNA binding domain where the L1 and S6-S7 loops are shown in blue and green. On the right, the saturation mutagenesis scans are illustrated as a heatmap for the two loops as an example. The full set of data is reported in the GitHub repository associated with the publication. Highly destabilizing amino acid substitutions for structural stability are highlighted in yellow, while dark blue indicates mutations that are predicted to be stabilizing the folded structure of the protein. These saturation scans are used to extract predicted ΔΔG values for all the possible variants of p53 DBD, including the ones of interest for the study (Figure 1). (b) The panel shows the predicted ΔΔG values with Rosetta or FoldX along the protein sequence of the p53 DBD. The goal was to identify and retain for further analyses alterations with marginal effects on structural stability. The red line indicates the threshold defined to discriminate between possible neutral and destabilizing amino acidic substitutions (i.e., 1.2 kcal/mol). We selected variants from the aggregated datasets of Figure 1 that were classified as neutral in both the FoldX and Rosetta datasets, resulting in 395 variants overall. The 395 mutations are distributed over 143 mutational sites. (c) We obtained a protein structure network based on atomic contacts (acPSN) from the MD ensemble of p53$_{91-289}$ monomer. We then estimated the hub residues (node degree equal to 3 or higher) that are mapped on the 3D structure of p53$_{91-289}$ monomer. The color code and ribbon thickness of the structure highlight the node degree of each hub using a color gradient from dark green (node degree = 3) to yellow (node degree = 7).
mer with DNA in the tetrameric assembly (DNA: p53\textsubscript{191-289_monomer}). Before assessing the effect of the 395 selected mutations on the long-range communication to S6-S7, we evaluated if any of them also has local effects on DNA binding. To this goal, we estimated the effects of the mutations found in our datasets on the interaction with DNA using FoldX-based calculations of changes in binding Gibbs free energy upon mutation (Figure 5(A)).

In addition, to gain an overview of potential hotspots, we performed the same calculation considering the 1132 variants collected in this study. We identified five main areas that had a change in the interaction energy. The hotspots were respectively in positions 119, 120 and 121, which are known positions to mediate DNA binding, position 132, positions 239–249, positions 273–281, and positions 285, all of which directly face the DNA. Within these hotspots, we identified 36 substitutions with a destabilizing effect on the binding between p53 and DNA (i.e., $\Delta \Delta G$ values higher than 0.8 kcal/mol) and three mutations with the opposite effect (i.e., $\Delta \Delta G$ values lower than −0.8 kcal/mol). When reducing the pool of mutations to the 395 selected variants, we identified 21 variants that are likely to have a local effect and alter the interaction with DNA (Figure 5(A)). Of these, S121F, N247Y, A276F were predicted to have a stabilizing effect on binding, whereas K120N/Q/E,

Figure 5. Local effects of mutations on binding interfaces. (a) The panel shows p53 monomer in complex with DNA, where 10 residues account for a total of 21 mutations that are all neutral for the stability of the protein but impact the binding with DNA. The structure illustrates the positional placement of the residues, and the bar chart shows the predicted changes in DNA binding free energy upon mutation. The black error bars represent the standard deviations of the performed runs. Mutations S121F, N247Y, and A276F are all predicted as stabilizing the binding of p53 DBD to the DNA, while the remaining 19 mutations are predicted as destabilizing the interaction, possibly resulting in decreased transcription. (b) The panel illustrates the cartoon representation of the p53 DBD dimer (which is representative of the assembly of p53 DBD in the cytosol). In particular, we focused on the residues at the binding interface for the sake of clarity. Each position is colored according to the sum of mutations that have a destabilizing effect at each position upon a saturation scan with the FoldX energy function. The residues for which neutral mutations in terms of structural stability were identified are shown as sticks. The bar chart shows on the x-axis the list of mutations for which the $\Delta \Delta G$ of binding is higher than 0.8 kcal/mol or lower than −0.8 kcal/mol and on the y-axis the average $\Delta \Delta G$ of binding in kcal/mol. Positive values indicate a destabilizing effect, while negative values a stabilizing one. The black bars represent the standard deviation of the predicted values.
S141V, R273H, A276P/S/T, C277W/Y/R, R280I/T/K, D281V/Y, and E285A/Q had destabilizing effects. These alterations may result in a reduction of DNA-p53 binding in terms of time and affinity which subsequently could lead to alteration of p53 activity as a transcription inhibitor or activator. Hence, these 21 mutations exemplify non-destabilizing mutations with functional impact. The 21 mutations are found in Table S3.

Local effect of the mutations on the binding interface in the dimeric state of p53 in the cytosol

In the cytosol, where the transcription-independent functions of p53 are carried out, the preferred state is as a dimer.52 Hence, we evaluated if any of the 395 selected variants had an impact on the local binding between the two monomers of p53 in the p5391-289_dimer structure using the FoldX energy function to estimate the changes in $\Delta\Delta G$ of binding (Figure 5(B)). Most of the mutations did not have any direct effect on the binding according to our calculation with the exception of 23 positions for which the $\Delta\Delta G$ of binding upon certain mutations was $\Delta\Delta G < -0.8$ kcal/mol or $\Delta\Delta G > 0.8$ kcal/mol (see Materials and Methods), thus indicating a stabilizing or destabilizing effect on the dimer formation respectively. According to all the performed scans (see Materials and Methods), C277Y, H179Y, and N263I mutations seem to stabilize the interaction. C277Y, which is predicted to destabilize the binding to the DNA (see above), might favor the binding of two p53 monomers through the formation of a hydrogen bond with D148. On the other hand, A276D, H178D/L/N/Q/Y, M243I/K/R/T/V, R181C/G/H/L/S/Y, S106R, and S261C/G/N were classified as destabilizing the binding, with H178 predicted to be a hotspot of highly destabilizing mutations (i.e., $\Delta\Delta G > 3$ kcal/mol). Interestingly, the newly identified germline mutation R181H appears to be deleterious for dimer formation, and, in turn, for interaction with other binding partners. Indeed, despite its neutral classification in terms of protein stability, we observed an average binding $\Delta\Delta G$ of 1.249 ± 0.093 kcal/mol for p5391-289_dimer. Analysis of the structures generated by FoldX upon the R181H mutation indicates the tendency of the H181 sidechain to form a hydrogen bond with the carbonyl oxygen of P177, located on the same chain, instead of pointing towards the other monomer. However, it is difficult to predict the behavior of the histidine sidechain without experimental data available that could suggest the correct protonation and tautomeric state of the imidazole ring (see Discussion Section).

Identification of mutations with local effects on conformation and dynamics of S6-S7 loop

In a previous study,25 we showed that the conformation of the S6-S7 loop (residues 207–213) can...
Identification of long-range allosteric mutations in p53 DBD with marginal effect on protein stability

In a previous work, we also illustrated that the S6-S7 loop (residues 207–213) can be allosterically modulated by conformational changes occurring in the DNA binding loop L1 (residues 114–124). Here, we carried out similar PSN-based analyses to estimate the communication between the two loops. We then evaluated which of the 143 mutation sites is part of these communication routes (Table S1 and Figure 7(A)). Apart from mutation sites located in the N-terminal disordered region (P92, L93, S95, S96, and V97), which are also of interest since movements of this region can regulate the conformation of the S6-S7 loop according to our previous study, and NMR experiments. We also identified M160 and M169, whose conformation is sensitive to conformational changes in the area where the disordered region of the N-terminal and the S6-S7 loop are located according to NMR experiments.

**Figure 7. The S6-S7 loop communicates with L1 loop and with some of the distal mutational sites.** We estimated the shortest paths of communication between the residues of S6-S7 loop and the L1 loop (a) or the 143 mutation sites (b) by calculating a Protein Structure Network based on atomic contacts (acPSN) for the MD ensemble of p53	extsubscript{391-289}_monomer. We represent the atomic contacts as sticks connecting the Cα atoms of the pair of residues involved in the interaction. We show as spheres the Cα atoms of the residues in the paths of communication. (b) The panels from the top left to bottom right show on the structure of p53	extsubscript{391-289} the shortest paths of communication calculated for different sets of residue types corresponding to mutation sites: i) R,K,H, ii) D,E, iii) S,T,N,Q, iv) A,V,I,L, v) F,Y,W, vi) C,M,P mutation sites. We identify 17 mutation sites involved in the paths of communication between the L1 loop and S6-S7 loop and several mutation sites with paths of communication to the S6-S7 loop. (c) Using the approach implemented by AlloSigMa, we predicted the potential effects of mutations on the allosteric regulation of the S6-S7 loop. This approach evaluates the effect of mutations that decrease (DOWN) or increase (UP) the steric hindrance at a certain site (e.g., our selected 143 mutation sites) and estimate the changes in allosteric free energy at distal sites (e.g., the S6-S7 loop) (Figure 7(C)). We noticed that the majority of the DOWN mutations resulted in a negative allosteric free energy, Δg, for the residues of the S6-S7 loop with only two sites interested by destabilizing
effects (i.e., positive $\Delta g$ values), as in the case of R174T/S/M and E204D/A. In the case of UP mutations, the effects are more variable with cases of long-range destabilization or stabilization of the S6-S7 loop dynamics.

We then analyzed the mutation sites that occur in the long-range communication paths between L1 and S6-S7 loops or that communicate long-range to S6-S7 (Figure 7(C)) for their changes in allosteric $\Delta g$ values. This analysis allowed us to
pinpoint some variants that do not seem to alter the allosteric free energy substantially despite being in intermediate or terminal residues for long-range communication (such as T118 and V122 of the L1 loop). In addition, the analysis allowed us to confirm variants that are involved in the shortest communication paths to the S6-S7 loop that can have a destabilizing (in 12 mutation sites) or stabilizing effect (in 22 mutation sites) on the dynamics and conformation of the loop. Intriguingly, the effect can be dual in some sites, with mutations destabilizing the loop conformation (E198Q, R273Y, T140I, T230I, and T231I) or stabilizing it, reducing its flexibility (E198D, R273H/L/C/Q, T140S, T230S, and T231A). The latter is interesting to investigate further with molecular simulations (see below) since the reduced flexibility of the S6-S7 loop could correlate with its entrapment in one of the occluded or open states that could decrease or enhance its interactions with biological partners, respectively. Furthermore, we also noticed other sites that seem to exert an allosteric effect on the S6-S7 loop if mutated, such as R181, R280, and R282.

**Molecular dynamics investigation of 15 variants for their long-range effects on S6-S7 loop dynamics and conformation**

To overcome the limitations of the AlloSigMA 2 approach, which is a convenient high-throughput approach for assessing the effects of mutations but does not allow the study of the protein structure at the atom level, we turned our attention to explicit solvent MD simulations.

We thus selected 15 interesting variants (K120E, K120N, R158L, R181H, R248Q, R248W, R273C, R273H, R280K, R280T, R282Q, S215G, S215I, S215R, and V217M) to carry out unbiased MD and/or enhanced sampling metadynamics simulations. We selected the 15 variants to cover different cases: i) four mutations at sites that could have local effects on the conformation of the S6-S7 loop (S215G/I/R); ii) in regions in proximity of the L1 loop that can allosterically communicate to S6-S7 loop (K120E/N); iii) in allosteric communication according to the PSN and AlloSigMA 2 analyses with S6-S7 (R181H, R273C/H, R282Q, R248W, and R280K/T, V217M) or iv) as a control since are expected not to have major effects (R158L, R248Q). Some of these variants have been also investigated in other studies and seem to be connected to changes in the loop of interest for this study.

At first, we evaluated if any of the mutations caused marked changes in the local atomic contacts (Figure 8). In particular, we focused, as we did for the wild type, on the effects of the mutations on the atomic contacts involving residues of the L1 loop or the S6-S7 loop in comparison with the wild-type variant. None of the variants showed a significant change in the pattern of intramolecular contacts during the simulation time, except for R181H, R273C, and S215G substitutions. In the first two cases, the residues of the L1 loop establish contacts with higher occurrence with the surrounding residues, while the S215G variant has a marked effect in the inter-residue contacts of the S6-S7 loop. The results are summarized in Table S4.

Next, we used the PSN-MD approach also for the 15 variants with a focus on changes in hub behavior and shortest paths for long-range communication. In terms of hubs, 80 of the 143 sites did not change their hub or non-hub behavior upon any of the mutations of interest. We noticed that the mutations caused hub fluctuations (i.e., changes in the number of edges connected to the hub, that kept the hub behavior) at 18 sites. We identified 21 sites that are hubs in the wild type but not in some variants, and 24 sites that become hubs upon mutation. Interestingly, the mutations R273C and R282Q cause the loss of the corresponding hub at that position, while S215I and S215R acquired the hub behavior at position 215. The results are summarized in Table S5.

Finally, we evaluated if any of the mutations could alter the long-range communication paths that we identified for the wild-type variant (Table 1). We observed that the communication paths between the loops L1 and S6-S7 were not completely abolished upon most of the amino acid substitutions. Many variants only interfered with a small subset of intermediate nodes for long-range communication preserving the endpoints for it. Only R181H and mutations at the 215 sites to glycine or arginine seem to have pronounced effects. In addition, R158L, R273C/H, and R282Q abolished additional routes for possible allosteric regulation of the S6-S7 loop that is not connected to the L1 loop (Table 1, third column).

MD simulations can also be analyzed using dimensionality reduction techniques, such as Principal Component Analysis (PCA). PCA allows to identify the most important motions during the protein dynamics and can also be used to compare different simulations in a reduced subspace. We here applied PCA to compare each of the variants to the wild-type in the same essential subspace with a focus on the conformation of the S6-S7 loop (Figure 9) and the L1 loop that is in communication with it (Figure S2). We then analyzed the projection along the first and the second principal components in detail, which gives an indication of the conformation space covered by the S6-S7 loop in the different simulations, and, as such, provides a measurement of which mutations are likely to retain the native dynamics (and conformation) of the loop and which ones can alter it. We observed that S215G, R181H, and R280T seem to have the most pronounced effects on the conformation of the S6-S7 loop.
Validation of allosteric perturbation using all-atom simulations with enhanced sampling

Unbiased classical MD simulations suffer from limitations when it comes to the analyses of effects induced by mutations and they encounter the risk of giving a limited view of the conformational space accessible to the protein, especially in the case of loop dynamics. We thus relied on a method to enhance the sampling and accurate estimate of the free energy landscape associated with the conformation of the S6-S7 loop based on metadynamics (see Materials and Methods). We obtained converged free energy profiles for ten of the mentioned variants (K120E, R158L, R181H, R248Q, R248W, R273C, R273H, S215G, S215I, and S215R), which allowed us to more accurately quantify if the mutations influence the conformation of the S6-S7 loop and the related population of each state (open or occluded) in solution (Figure 10). In Figure 10, we showed as an example the results for one of the collective variables from the metadynamics simulations (i.e., the distance between the C\textalpha\ atoms of R156 and D207). This is also one of the most representative collective variables used for the previous study on the S6-S7 loop. According to our previous work, occluded states of the loop correspond to values of this collective variable lower than 1.5 nm, whereas the open and accessible states for biological interactions are generally at values higher than 2 nm. It has to be taken into consideration that this distance refers to the C\textalpha\ atoms only and is used here to pro-

Figure 8. Mutations in the p53 DBD can alter the intramolecular contacts of the S6-S7 loop. The panels illustrate an example of how some of the variants can reshape the contact map of the S6-S7 loop. We used the CONAN software to perform the analyses. The heatmaps elucidate how the inter-residue contacts of R213, belonging to the S6-S7 loop, changed between the wild-type and the S215G variant. The contacts are colored with a gradient from yellow (low occurrence) to dark purple (high occurrence). On the right, we report the contact observed for the S215G variant with a focus on the residue R213 of the S6-S7 loop. The panel shows only the interactions with occurrence higher than 0.1 (i.e., 10%). Each contact is represented as a stick that connects the C\textalpha\ atoms of the interacting residues. The radius of the stick is proportional to the occurrence in the simulation frames (i.e., 100% occurrence corresponds to a radius of 0.4 Å) and the colors of the sticks correspond to the one of the corresponding heatmaps on the left. The C\textalpha\ atoms are represented as white (R213) and gray (all the other residues not belonging to the S6-S7 loop) spheres.
provide a suitable first screening of the effect of many variants. Nonetheless, to identify the structural details of the conformational changes of the S6-S7 loop other properties should be measured as well, such as solvent accessibility and the location of specific side chains important for intermolecular interactions. These properties will be the subject of future studies in which we will investigate the mechanistic details of each variant.

The K120E substitution does not feature a marked impact on the sampling of the two states (Figure 10(A)) but it seems to decrease the energy barrier for the interconversion among the two states. The R248W substitution causes a long-range perturbation that promotes a higher conformational heterogeneity in the S6-S7 loop, where also intermediate states between occluded and open states are observed (Figure 10(B)). The R273C mutation disfavors the sampling of occluded states and seems to stabilize more open conformations of the S6-S7 loop (Figure 10(C)). In our previous study, open states of this loop correlated with conformations of L1 less prone to DNA binding, thus, suggesting that the R273C mutation could result in a p53 variant more prone to protein-protein interactions for transcription-independent functions of the protein and with compromised transcriptional activity. R273H, instead, does not seem to compromise the capability to explore states of the loop characterized by different degrees of accessibility but it decreases the energy barrier associated with the conformational changes (Figure 10(D)).

The S215G mutation (Figure 10(E)), which alters a phosphorylation site that was known to entrap the protein with an open state of the S6-S7 loop, has the opposite effect of stabilizing more occluded states of the loop. This mutation is thus expected to inhibit the protein-protein interactions mediated by the S6-S7 loop and remove the possibility of inhibiting DNA transcription through post-translational modification in position 215. S215I caused a shift in the loop conformation towards more intermediate states between the occluded and the fully open ones (Figure 10(F)). The S215R substitution (Figure 10(G)) introduces a pos-
A positively charged residue that promotes open states of the loop, which points in the direction that including a charged residue at this position can regulate the loop conformation. Indeed, a similar effect was previously observed upon phosphorylation at S215, which introduces a negatively charged residue. This means that both electrostatic repulsive and attractive mechanisms could be in play in this region to achieve the same conformational effect of activating the loop-related functions.

R158L does not have a marked effect on the loop conformation apart from disfavoring the more extreme open states (Figure 10(H)). R248Q (Figure 10(I)), which was not predicted by the previous analyses using unbiased classical MD simulations to have an effect, causes a stabilization of the loop in its occluded state, suggesting that this mutation will affect the recruitment of p53 interactors through the S6-S7 loop. This result also emphasizes the importance of employing enhanced sampling approaches in molecular simulations to accurately understand the effects of mutations. The germline variant of unknown significance R181H tends to disfavor conformations of the S6-S7 loop that allow for interactions with other biomolecules with a potential effect on the transcription-independent functions of p53 (Figure 10(J)).
For the 15 variants that we studied using molecular simulations, we summarized the results from the different properties estimated in this study. The results are reported in Table 2 and we also compared our results with the scores for pathogenic or driver effects, as estimated by REVEL\textsuperscript{77} and CSscape-somatic\textsuperscript{78} respectively. According to the predictors for pathogenicity or driver mutations, each of these mutations features very high scores despite their marginal effect on the conformation of the loop.

### Summary of the effects of the selected 15 variants and comparison with scores for pathogenicity or driver impact

For the 15 variants that we studied using molecular simulations, we summarized the results from the different properties estimated in this study. The results are reported in Table 2 and we also compared our results with the scores for pathogenic or driver effects, as estimated by REVEL\textsuperscript{77} and CSscape-somatic\textsuperscript{78} respectively. According to the predictors for pathogenicity or driver mutations, each of these mutations features very high scores despite their marginal effect on the conformation of the loop.

![Figure 10: Mutations in the proximity of S6-S7 loop or at distal sites can alter the conformation of the loop.](image)
Table 2 Summary of the results for each of the 15 variants studied with all-atom molecular dynamics simulations. ‘*’ indicates mutation sites that are at a distance higher than 6 Å from the DNA binding site or the dimerization interface. ‘INT’ indicates that the residue is an intermediate of a communication path between the S6-S7 and the L1 loops. We indicated the specific residues of the S6-S7 loop in the cases in which the mutation site was involved in long-range communication with the loop through alternative routes not depending on L1. N.A. indicates data that are not available in terms of metadynamics results. In the case of S215R, the amino acid substitution was associated with distinct genomic variants from the data sources. In this case, we thus reported the range of scores obtained from REVEL and CScape-somatic.

| Variant | DNA binding $\Delta G$ (kcal/mol) | Dimer binding $\Delta G$, (kcal/mol) | Local vs distal effects (MD) | AllostERIC $\Delta g$ (kcal/mol) | Conformational changes (S6-S7 loop) | Revel score | CScape score |
|---------|----------------------------------|-----------------------------------|-----------------------------|-------------------------------|---------------------------------|-------------|-------------|
| K120E   | 1.627                            | 0.272*                            | NONE                        | $-1.967 < \Delta g < -1.385$ | -                               | N.A.        | 0.909       |
| K120N   | 1.627                            | 0.161*                            | NONE                        | $-1.967 < \Delta g < -1.385$ | N.A.                           | N.A.        | 0.879       |
| R158L   | 0*                               | -0.049*                           | PATHS(INT.)                 | $-0.319 < \Delta g < 0.179$  | -                               | 0.891       | 0.960       |
| R181H   | 0*                               | 1.256                             |                              | $-1.375 < \Delta g < -0.773$ | Favored occluded states         | 0.797       | 0.966       |
| R248Q   | -0.506                           | -0.581                            | NONE                        | $-0.973 < \Delta g < -0.509$ | Favored occluded states         | 0.934       | 0.957       |
| R248W   | -0.480                           | -0.857                            | NONE                        | $0.740 < \Delta g < 1.274$   | Intermediate states             | 0.927       | 0.950       |
| R273C   | 0.437                            | -0.138*                           | PATHS (208,211,213) (INT.)  | $-1.038 < \Delta g < 0.710$  | Favored open states             | 0.901       | 0.964       |
| R273H   | 0.986                            | -0.039*                           | PATHS (208,211,213) (INT.)  | $-1.038 < \Delta g < 0.710$  | Both states, smaller energy barriers for transitions | 0.868       | 0.970       |
| R280K   | 3.084                            | 0.040                             | NONE                        | N.A.                          | N.A.                            | 0.935       | 0.910       |
| R280T   | 3.061                            | 0.078                             | NONE                        | $-1.152 < \Delta g < -0.910$ | N.A.                            | 0.940       | 0.941       |
| R282Q   | -0.025*                          | 0.004*                            | PATHS (208,211,213)         | $-0.749 < \Delta g < -1.133$ | N.A.                            | 0.887       | 0.954       |
| S215G   | 0*                               | 0.003*                            | LOCAL                       | $-0.114 < \Delta g < 1.654$  | Favored occluded states         | 0.896       | 0.958       |
| S215I   | 0*                               | -0.018*                           | LOCAL                       | $-1.521 < \Delta g < -0.916$ | Intermediate states             | 0.913       | 0.944       |
| S215R   | 0*                               | -0.027*                           | LOCAL                       | $-1.521 < \Delta g < -0.916$ | Intermediate states             | 0.948–0.973 | 0.812–0.91 |
| V217M   | 0*                               | -0.006*                           | PATHS (208,211,213)         | $-0.913 < \Delta g < -0.169$ | N.A.                            | N.A.        | 0.943       |
protein stability. Our functional assessment allows us to provide complementary information on the more likely effect of each of them and divide the mutations into the different classes: i) effects on direct interaction with DNA (K120E/N, R280T/K); ii) allosteric effect on the S6-S7 interface (R248Q/W, R273C, R273H, R282Q), iii) local effects on S6-S7 (S215G/I/R), iv) combined local effects and allosteric (R181H); v) minor or neutral effects on all the properties under investigation (R158L).

Discussion

Limitations and strengths of the approach

The process of analyzing cancer variants from a structural viewpoint with computational methods may shed light on biological processes yet to be researched comprehensively. While this is a great strength, the validity of our findings must be considered in light of the methodological framework and its known limitations. The methodological framework used here can be divided into two main areas: an assessment of either local or long-range effects of the mutations at the structural level. Both areas of this research depend on the choices of the starting structure, the employed algorithms and analysis strategies, and inherent challenges with molecular simulations or free energy calculations.

In this study, we focused on the DNA binding domain of p53 since it has a key role in the function of the S6-S7 loop which is at the center of the work. It was also a convenient case study to establish a multi-layer framework for analyzing mutations, that could be reused in the future to study other structural domains or full-length proteins. The framework was designed also according to our previous study in which we discussed how the conformational properties of the loop are not substantially affected by the tetrameric state of p53. A future extension to this framework could include the possibility to account for the interactions of the DBD with the other p53 domains such as the transactivation domain, the proline-rich region, the nuclear localization signal, and the tetramerization domain.

In the first part of the work, we applied two different methods to estimate changes in folding free energy based on different algorithms to achieve more robust results. We also used a structural ensemble based on MD simulations to overcome the limits of FoldX in sampling the flexibility of the backbone of a protein. However, the two energy functions used here have both limitations in relation to the conformational sampling accessible during the calculations. We found a reasonable correlation with experimental data for p53, in the range of 0.7–0.8, using both approaches which support their usage for high-throughput screening of many variants. The usage of more accurate methodologies such as the ones based on alchemical free energy protocols could in the future be used to better appreciate and understand the effects of amino acid substitutions that are challenging for FoldX or Rosetta to predict.

We also used FoldX to calculate changes in binding free energy upon mutation for the binding to DNA or for the monomer–monomer interaction of the p53 DBD in the dimeric structure, which represents its cytosolic assembly. These calculations also suffer from limitations due to backbone rigidity when considering FoldX. This implies that we have been able to identify only very local effects with these methodologies. Other variants could also impact the binding to DNA or the dimerization which will require further modeling of the conformational changes of these complexes. A challenge in this context is posed by mutations to histidine, such as R181H and R273H. Indeed, the different protonation and tautomeric state of the histidine imidazole ring can dramatically affect both the interaction pattern that the residue can establish and, in turn, the dynamics of the whole system. Unfortunately, the lack of experimental data, especially from NMR spectroscopy, hampers the correct prediction of the preferred protonation and tautomeric state of the histidine side chain. This issue could be mitigated by the prediction of pKa values of ionizable residues, like PROPKA3 on an experimental structure of the mutant variants of interest or the availability of chemical shifts from NMR experiments that can allow to also define the tautomeric state in solution.

In light of the above observation, approaches based on energy decomposition could be an additional tool to introduce in the structure-based assessment of the effects of variants as applied to other studies. The advantage of this approach is that it is applied to MD simulations and all-atom force fields. It can then be useful to allow a better description of mutations to residues as histidines or better explore effects depending on backbone changes. It calculates an interaction matrix for the protein structure by considering the non-bonded interaction energies between residue pairs. The principal eigenvector resulting from the diagonalization of the matrix estimates the non-bonded interactions and residues with the most important contributions to the structural stability. The energy decomposition approach has been employed to estimate the effects of mutations on structural stability, showing performance comparable to FoldX-based calculations. Integrating this method into our framework could contribute to the characterization of variants that are challenging to estimate by FoldX or Rosetta.

Overall, despite the limitations described above, the possibility to carry out high-throughput free energy calculations with empirical energy functions has allowed us to conveniently identify a subset of mutations to further explore in terms of
elusive allosteric effects. In parallel, the approach has been important to disclose cases in which multiple mechanisms elicited by the mutations are at play, which could have different implications at the cellular level.

We here used PCA in a non-quantitative manner to compare the effects of the mutated variants on the dynamics and conformations of the loop of interest. This analysis has been mostly applied as a first exploration before engaging in enhanced sampling with metadynamics. Methods based on the analysis of the covariance matrix of atomic fluctuations and PCA, such as the one employed here, are widely used to investigate MD ensembles. However, they have limitations in describing collective motions and identifying conformational states.87

Approaches for structure-based assessments of variants as the ones presented here, if carried out systematically and linked to a biological readout, as we recently proposed for another protein,88 have the potential to generate a large amount of data on 3D protein features that could be exploited in machine-learning setting for predictive or classification purposes, as recently done in other fields of molecular modeling and simulations.89

Of note, we also noticed that the usage of PSN-based methods alone to predict allosteric mutations is not sufficient since very often the effects on the communication paths that we can measure are limited due to the existence of multiple paths and the robustness of the nodes of a network towards mutations.90–91 It is thus important to integrate this conventional formalism with other approaches to better predict whether the amino acidic substitutions of interest influence the conformations of the endpoints of communication or if, on the contrary, their effect is negligible. In this context, we here illustrated that the PSN approach together with enhanced sampling simulations based on collective variables that capture the conformational states of interest are a remarkable combination to disentangle the effects of the amino acid substitutions. These methods also offer the possibility of understanding the effects of the mutations in terms of changes in the population of the different conformational states of the protein during dynamics.

Comparison of our data with available data on p53 variants

Our investigation provides a detailed understanding of the impact of 15 cancer-related variants of the p53 DBD, covering changes in folding free energies, interactions with DNA, dimerization occurring in the cytosol, and local or allosteric effects on the conformations of the functional interface provided by the S6-S7 loop (Table 2). In addition, we also provided pre-calculated scores for the changes in folding or binding free energies for more than 1000 germline or somatic variants of p53.

Of note, R248Q is predicted in our study with a low impact on structural stability. At the same time, it is one of those mutations for which more than an experimental measurement is available92–94 in ThermoMutDB and they can vary from neutral to destabilizing effects (i.e., 1.87, 3.07, or −0.74 kcal/mol). Thus, a destabilizing effect due to this variant cannot be ruled out. Many studies have been carried out with cellular or mouse models which point to a destabilizing effect for the R248Q variant resulting in misfolding and aggregation of p53.95–96 The empirical free energy functions used in this study predict a marginal effect of this mutation on the stability of the monomer, representing one of those cases, as discussed above, in which more sophisticated methods accounting for protein dynamics and large-scale conformational changes are needed to appreciate the effects of this substitution. It is also important to highlight that p53 is characterized by low cellular stability already in its wild-type form and thus even mutations that destabilize it by a few kcal/mol might be enough to elicit a marked effect at the cellular level. This also suggests that, when using predictions from methods such as FoldX and Rosetta, variants with ΔΔGs values around 1 kcal/mol, i.e., close to the thresholds used to classify the variants, could be worthwhile investigating. On the other side, the effect of R248Q might be highly context-dependent and other studies suggest that it is a variant associated with higher levels of expression of p53 and with an oncogenic effect.97–98

We observed a long-range effect triggered by the R248Q variant on the S6-S7 loop. In another unbiased 500-ns molecular dynamics study,32 R248Q was observed to perturb allosterically the DBD and to favor open states of the loop. These contrast with the free energy profile from the metadynamics used in this study, where we observe the opposite behavior (i.e., R248Q stabilizes more occluded states of the loop), according to the selected collective variables (Figure 10(I), Table 2). These discrepancies could be due to multiple factors, such as different coverage of the conformational space accessible to the loop during the simulations or different force fields used in the two studies or different temperature conditions. Enhanced sampling approaches, like the ones used here, should in principle allow for a better estimate of the population of different conformational states for a highly dynamic protein such as p53, however their accuracy also depends on how well-converged they are. More calculations applying different approaches for enhanced sampling or different starting structures could help to shed light on these discrepancies. An interesting direction to pursue in the future using the R248Q variant as a case study could be to test the influence of using different force fields and different sampling approaches on...
the conformational dynamics of the S6-S7 loop and in parallel collect NMR data on its dynamics to better understand the mechanism in play. In addition, occluded states of the loop, according to our mechanistic model should imply a decreased binding to the p53 interactors recruited at the S6-S7 interface (see below) and experimental assays for protein–protein interactions on the wild-type and mutant variant could also provide a suitable strategy to validate the results.

We investigated variants at residues of p53 DBD that make direct contact with DNA and are frequently mutated in human cancers (R248Q/W, R273C/H). For R248Q/W, R273C, and R273H we observed allosteric effects on the conformations of the S6-S7 loop (Table 2), suggesting that these variants likely impact interaction preferences of the protein at this interface more than just acting locally by affecting the binding with DNA. This result suggests that multiple mechanisms are at play and can be triggered simultaneously by these substitutions. Experimental and computational data support our results, indicating that different variants of p53 DBD at R273 (R273H, R273C, and R273L) not only alter the DNA binding and transcriptional activity of p53, but also affect other properties and this may have a role in pathogenesis. Furthermore, experimental studies show that variants in the p53 DBD, such as R273C/H and R248Q, act via diverse context–dependent mechanisms, such as by affecting physiological functions, have dominant-negative effects, or could gain oncogenic activities. Our assessment provides additional information to profile the effect of these variants in the p53 DBD. In addition, we characterized alterations at site 215 of p53 DBD that not only abolish a regulatory post-translational modification but also act locally on the conformation of the S6-S7 loop and can have both inhibitory or activatory effects depending on the substitution (Figure 1), also pointing to multiple mechanisms that can have different relevance depending on the cellular context and conditions.

In this study, we systematically investigated the mutational impact one residue at a time. Further work may consider the mutational burden and effect of hypermutability of tumors to understand the global alterations of the protein and the context of cancer. Indeed, multiple occurring mutations may lead to either synergistic effects, where the collection of mutations is more damaging than the sum of the individual alterations (negative epistasis) or neutralize each other due to the compensatory nature of the mutations (positive epistasis). This is also true for p53, where some compensatory mutations can rescue the protein function in the face of a pathogenic mutation. Although the calculation of the mutational burden is common practice, using a mutational rate evaluating the number of mutations in a particular gene compared to genomic data, complex epistatic interactions are often omitted since assessment and prediction of compensatory mutations is not a straightforward task. Approaches to assessing these epistatic interactions often rely heavily on evolutionary mechanisms, fitness landscapes, or as presented using double force scanning assessing rescue sites without any evolutionary assumptions such as distance between rescue mutation and rescued site. Most recently, a new and faster implementation of the double force scanning has been proposed which could be included in an effective manner within our structure-based framework. Other methods derive epistatic information more indirectly using evolutive approaches, e.g., based on coevolution or using epistasis as a basis to understand the fitness or pathogenicity of a given mutation. In conclusion, to comprehensively understand the impact of p53 variants or other similar cancer targets, tools to unveil the intricate combinations and their effects on conformational alterations are a field of active development. Combining these tools with a firm understanding of multimerization, cellular compartment, and thus preferred conformational state could provide invaluable information in the future. Prominently, this research may play a part to bridge the gap between mutational understanding in a structural space and hypermutability of tumors found in the clinic. Future research could lead to effectively explaining some hallmarks of cancer from a structural perspective unveiling sites of interest that could be targeted by therapeutic agents.

**Biological relevance of the assessment of cancer-related mutations on the allosteric regulation of the S6-S7 loop**

One of the main results of this work is that we identified a group of mutations of p53 that are likely to have a marginal impact on structural stability but still with a possible pathogenic impact due to alteration of an allosteric mechanism to regulate an interface for recruitment of protein–protein interactions, i.e., the S6-S7 loop. These results are of broader relevance if we consider that the p53 DBD is central in the cellular hub activity of p53. It interacts with DNA, transcription factors, and other protein partners to elicit p53 transcriptional-dependent and independent functions. Different interaction partners for the p53 DBD, which binds in the proximity of the S6-S7 interface, have been predicted or experimentally identified. In addition, a recent X-ray structure shows the complex of p53 DBD with a short peptide of the p53 upregulated modulator of apoptosis (PUMA) protein. In this complex, PUMA directly interacts with the p53 DBD via residues of the N-terminal region and the S6-S7 loop.

In addition, molecular simulations of p53 DBD in complex either with the DNA or the repressor protein iASPP suggest a communication path from...
K120 in loop L1 to R213 of the S6-S7 loop, like the ones observed in our previous work. The authors propose that in complex with DNA, this communication path favors the binding with partners. On the other hand, the interaction with iASPP rewires the communication path making the interaction with DNA.

As we stated above, alterations in the coding region of the TP53 gene can elicit different outcomes, such as loss-of-function, gain-of-function, and dominant-negative effects. It becomes thus important to understand how cancer mutations can alter the conformation of the S6-S7 loop toward occluded (i.e., not favorable to intermolecular interactions) or open (i.e., accessible to intermolecular interactions) conformations. Substitutions such as R181H, R248Q, and S215G, which favor occluded states are likely to have an inhibitory effect on the intermolecular interactions either through a local effect (S215G) or an allosteric effect elicited at a long distance from the mutation site (R248Q). Other substitutions, such as R273C or S215R, could promote the opposite effect and enhance these interactions, making them constitutively active. The availability of a nanobody binding at this interface and of the interaction with PUMA could be exploited to design experiments (and simulations) to test these hypotheses.

In addition, in a recent study of the p53 DBD, it was shown that mutations that affect the motions of loop S6-S7 could favor its open states and the formation of a putative pocket with druggable properties, which makes the efforts of cancer-related variants of p53 on this interface even more interesting for applicative purposes.

**Towards a more generalized framework for structure-based assessments of disease-related mutations**

In this and previous works, we showed how using a multi-tier approach based on computational structural methods can help to shed light on the complexity behind the effects of mutations found in cancer or other diseases. Many of the techniques applied in this study will work more broadly for other proteins and can help to fill our gap of knowledge in the field of cancer structural biology. In addition, as shown in a previous study on a ubiquitin-like protein, we can use these analyses to prioritize variants for experimental validation and help to identify the best biological readouts for the experiments depending on the functional effect that we predict.

A supplementary applied endpoint is the targeting of alterations therapeutically. Indeed, well-founded knowledge of structural cancer alterations may give rise to the development of therapeutics such as antibodies or CAR T-cells, where the alteration is the epitope. Additionally, such an approach could even identify isoform alterations purely on the sequence, suggesting a progression towards a personalized treatment approach.

**Methods**

**Identification of variants to use in the study**

We collected and aggregated cancer-related missense variants for the DBD region of p53 from cBioPortal and COSMIC version 94. The results are reported in Table S1. The data have been collected in November 2021 and updated/verified in January 2022.

We have used the gnomAD API to gather all the variants associated with TP53 in gnomAD version 2.1. We have filtered them by keeping variants with a defined protein change that were defined on the canonical Uniprot isoform. We have then calculated the allele frequency as the ratio between the mutation allele count and the total number of high-quality genotypes for whole-genome sequencing data. Finally, we have filtered the dataset by keeping only the mutations with low frequency (<0.0001).

We retrieved curated somatic and germline variants from the TP53 database and filtered the data to retain only the missense variants in the region 91–289 of interest for this study. In addition, we have added in-house missense variants in the region of interest identified with 30x whole genome sequencing (method: clinical NGS gene panels).

The Venn Diagram was produced with DeepVenn.

**Selection of PDB structures for p53 DBD**

The selection of experimentally solved three-dimensional (3D) structures for the modeling of p53 is not trivial due to the uncommonly large number of options within the Protein Data Bank (PDB). Consequently, the choice was based on a set of criteria implemented in a Python script. We identified 226 PDB entries associated with UniProt P04637 and we retained information regarding the experimental method used, the resolution where applicable, and the deposition date of the structure. Each structure was chosen applying the following criteria: X-ray structures were preferred to electron microscopy, which was preferred to NMR spectroscopy. We selected the X-ray structure characterized by the best resolution. In cases of more structures with the same resolution, we selected the most recent structure.

For each structure the sequence was retrieved and compared to the sequence of the DNA binding domain (residues 102–292), discarding any structures that did not contain an exact match (n = 31). 16 of these were bound to DNA, 11 were in complex with other peptides, one had undergone a post-translational modification and three were free chains. We selected three
structures for further analyses: one free single-chain structure (2XWR, $R = 1.68\ \AA$), a single chain structure in complex with DNA (3KZ8, $R = 1.91\ \AA$, selected chain B). In addition, we used the structure 2OCJ (chains B and D) as a reference for the dimeric free state of p53 in the cytosol. See Table S2 for the full list of the 31 structures, each annotated for potential utility.

**Modeling of the dimeric state and monomeric state in complex with DNA of p5391-289**

We used MODELLER version 10.1 to generate models of the dimeric state and monomeric state in complex with DNA of p5391-289 (p5391-289_dimer and p5391-289_monomer, respectively). For p5391-289_dimer, we used as templates the X-ray structure of the dimer of p5396-289 (PDB entry 2OCJ) and the structure after PDB-REDO refinement. We used the orientation of the two monomers defined by chains B and D in the PDB files. We selected this orientation since it closely resembles the orientation of the monomers in the X-ray structure of the dimer of p5396-289 in complex with BCL-xL (PDB entry 6LHD), which is the dimeric structure expected for p53 in the cytosol. We modeled only the N-terminal region (residues 91–96) for which the coordinates are not available in the PDB file using as a template the X-ray structure of p5391-289 (PDB entry 2XWR chain A). For DNA: p5391-289_monomer, we used as a template the X-ray structure of the tetramer of p5396-289 (PDB entry 3KZ8). We used the monomer of p5391-289 in chain B and the single DNA strand in chain C. We generated the second DNA strand (chain D) via symmetry operations in PyMOL (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC). We modeled only the regions for which the coordinates are not available in the PDB file (the N-terminal region (residues 91–96), residues S185-G187, and part of the side chain of residues K101, Q104, R110, D148, L188, L201, R202, E204, E221, E224, V225, D228, E287) using as a template the X-ray structure of p5391-289 (PDB entry 2XWR chain A).

We generated 100 models from each template using six distance restraints between five pairs of residues forming non-covalent contacts in the N-terminal region of the p53 domain (i.e., W91-R174, S94-T211, S94-S96, S94-T170, P92-F212). We used harmonic potential restraints, restraining each distance around its value in the template X-ray structure of p5391-289 (PDB entry 2XWR, chain A). We then calculated in the 100 models the six sidechains that could be improved with respect to the residues in the surrounding (self-scan with the lowest RMSD as the final one for each MODELLER run.

**Free energy calculations of stability and binding**

To calculate changes in Gibbs free energy, we applied the FoldX energy function performing in-silico saturation mutagenesis using MutateX. We followed the protocol described in, applying the BuildModel module from the FoldX suite with five independent runs and subsequently finding the average $\Delta G$ of the difference in $G$ between mutant and wild-type variants. We used different initial structures for the FoldX scans: i) the chain A from the X-ray structure with PDB entry 2XWR, ii) a structure obtained from the PDB-REDO refinement of 2XWR, and iii) an ensemble of 20 conformations extracted from a MD simulation (see below) to account for flexibility in the protein. We have previously illustrated that the performance of the FoldX algorithm increases with the use of MD ensembles as opposed to a single crystallographic structure. We also showed that the proportional addition of frames only increases performance until a certain point as the application of 50 frames yielded a higher correlation with experimental values than 100 frames.

Additionally, stability prediction upon mutation was calculated using the protocol devised by Park and coworkers to predict the change in folding free energy in a protein monomer upon mutation Rosetta. As recommended by the original publication, we used the ref2015 Rosetta energy function. This protocol takes a single structure as input, relaxes it using the Rosetta relax application, and further optimizes it in Cartesian space. Three rounds of Cartesian space optimization provide three pairs of wild type and mutant structures for each mutation. The change in folding free energy is calculated for each pair, and the final change in folding free energy associated with each mutation corresponds to the average over the three values obtained.

We also used the self-scan option of MutateX using the BuildModel module from the FoldX suite to mutate only the original amino acids to themselves, which was used as a control experiment allowing us to assess the local quality of the initial structures. A similar approach was applied to Rosetta. In the case of MutateX, we carried out the self-scans with and without the repair step of FoldX. In this regard, we expect that the self mutations should be very close to 0 kcal/mol since we are replacing a side chain with itself. In cases of deviations from this behavior, it could mean that upon repair that position or the surroundings are entrapped into some unfavorable conformations (self-scan with repair function on) or that the initial structure used for the calculation had some local placement of sidechains that could be improved with respect to the residues in the surrounding (self-scan with...
The FoldX energy function. The only exceptions are the positions H193 and H214 in FoldX-based calculations on the MD ensemble. The deviations for H193 and H214 were only observed in two or three models, and we verified that the final average results used for the selection of variants of interest at these positions were not affected by the local inaccuracy highlighted by the self-scan results. In the case of Rosetta, we also noticed 78 self-mutations with values above the cutoffs when we used the structure refined by PDB-REDO as the initial structure for the calculations. This result might be also due to sampling issues with this specific calculation more than to be related to the quality of the initial structure. For the mutations introduced with Rosetta on the X-ray structure, four self-mutations exceed the thresholds, at positions 131, 132, 165, and 247. The results of the predicted changes upon mutations at these positions thus need to be taken with caution.

Furthermore, we evaluated the accuracy of the predictions using a pool of variants of p5391-289 from the ThermoMutDB database42 with experimentally determined unfolding ΔΔG values. We collected the values available in the database and sign-inverted the values so that they are consistent with our computational predictions (i.e., negative values represent stabilizing mutations while positive values represent destabilizing mutations). For mutations with multiple experimental measurements for the same mutation, we compared these and calculated the standard deviations to illustrate the error margin within the experimental findings. Furthermore, we calculated the average ΔΔG value, resulting in one single value for comparison for each mutation.

We verified that the p53 mutations from ThermoMutDB used in this study were not part of the original training dataset for the development of the FoldX energy function.128 The dataset contained 339 experimentally characterized mutant variants in nine different proteins (barnase, CI-2, spectric, Src SH3, Sso7d, Tenascin, FKBP, Ada2h and CheY).

We then compared the predicted and experimentally determined ΔΔGs values for the mutations reported in ThermoMutDB to evaluate: i) the accuracy of the approaches to predict changes in unfolding ΔΔGs, ii) the influence of different initial structures for the calculations. As stated above, we used, as starting structures: i) the X-ray crystallographic structure as taken from the Protein Data Bank (X-ray), ii) a version of the X-ray structure upon refinement with PDB-REDO (X-rayPDB-REDO), iii) and an ensemble of 20 structures from a Molecular Dynamics (MD) simulation with the CHARMM22* force field.

We used the X-ray and X-rayPDB-REDO to perform scans with both the FoldX and Rosetta free energy functions. The MD ensemble was used only for FoldX scans to evaluate if it can rescue some limitations due to lack of backbone remodeling during the calculations. To assess the accuracy of the methods used for prediction or the influence of the starting structure, we evaluated both the correlation with experimental data and the classification accuracy, which describes the performance of a classification model. The correlation was assessed by calculating the Pearson correlation coefficient by comparing the experimental values to the predicted ones. The classification accuracy was assessed based on the following classification model: stabilizing (ΔΔG < −1.2 kcal/mol), neutral (−1.2 ≤ ΔΔG ≤ 1.2 kcal/mol), destabilizing (1.2 < ΔΔG ≤ 3 kcal/mol), and highly destabilizing (>3 kcal/mol) (Table S1), applied to both the experimental and predicted values for each mutation.

We also used MutateX to calculate binding ΔΔGs upon mutation using the structures of the dimer of p53. Both saturation and self-scans were run on the X-ray structure of the homodimer (PDB entry 2OCJ),123 on the X-rayPDB-REDO structure, or their corresponding models designed with MODELLER (see above). We selected chains B and D for the MutateX scans since their orientation resembles the one observed in the cytosolic p53 fusion complex with BCL-xL (PDB entry 6LHD).53 The self scans provided values within the cutoff to classify the mutations. We used as thresholds binding ΔΔGs values higher than 0.8 kcal/mol or lower than −0.8 kcal/mol for destabilizing or stabilizing mutations, respectively (i.e., the error of the predictor suggested by the developers of the FoldX energy function). We chose the same threshold values to investigate the effects of the mutations on the binding ΔΔGs upon saturation scan at the interface between the two p53 monomers.

Changes in DNA binding free energies upon mutation

We used FoldX 5 to evaluate the change in binding energy for the binding between p53 and DNA upon mutations. This evaluation comprises a four-step process, encompassing a preparation step, a repair step, a step to build the model with the relevant mutation, and a step to analyze the DNA binding energy and calculate the difference in energy between the mutated and wild-type models. This investigation relies on a crystallographic structure of p53 bound to DNA (PDB entry 3KZB, chain B) upon preparatory and remodeling steps with PyMOL and MODELLER version 10.1.124 PyMOL was used to model the missing parts of the double-stranded DNA, whereas MODELLER was used to model the missing residues of the disordered N-terminal region (residues 91–96) and the residues 185–187 using the X-ray structure upon refinement with PDB-REDO as the starting structure, we evaluated both the methods used for prediction or the influence of the starting structure, we evaluated both the correlation with experimental data and the classification accuracy, which describes the performance of a classification model. The correlation was assessed by calculating the Pearson correlation coefficient by comparing the experimental values to the predicted ones. The classification accuracy was assessed based on the following classification model: stabilizing (ΔΔG < −1.2 kcal/mol), neutral (−1.2 ≤ ΔΔG ≤ 1.2 kcal/mol), destabilizing (1.2 < ΔΔG ≤ 3 kcal/mol), and highly destabilizing (>3 kcal/mol) (Table S1), applied to both the experimental and predicted values for each mutation.

We also used MutateX to calculate binding ΔΔGs upon mutation using the structures of the dimer of p53. Both saturation and self-scans were run on the X-ray structure of the homodimer (PDB entry 2OCJ),123 on the X-rayPDB-REDO structure, or their corresponding models designed with MODELLER (see above). We selected chains B and D for the MutateX scans since their orientation resembles the one observed in the cytosolic p53 fusion complex with BCL-xL (PDB entry 6LHD).53 The self scans provided values within the cutoff to classify the mutations. We used as thresholds binding ΔΔGs values higher than 0.8 kcal/mol or lower than −0.8 kcal/mol for destabilizing or stabilizing mutations, respectively (i.e., the error of the predictor suggested by the developers of the FoldX energy function). We chose the same threshold values to investigate the effects of the mutations on the binding ΔΔGs upon saturation scan at the interface between the two p53 monomers.

Changes in DNA binding free energies upon mutation

We used FoldX 5 to evaluate the change in binding energy for the binding between p53 and DNA upon mutations. This evaluation comprises a four-step process, encompassing a preparation step, a repair step, a step to build the model with the relevant mutation, and a step to analyze the DNA binding energy and calculate the difference in energy between the mutated and wild-type models. This investigation relies on a crystallographic structure of p53 bound to DNA (PDB entry 3KZB, chain B) upon preparatory and remodeling steps with PyMOL and MODELLER version 10.1.124 PyMOL was used to model the missing parts of the double-stranded DNA, whereas MODELLER was used to model the missing residues of the disordered N-terminal region (residues 91–96) and the residues 185–187 using the X-ray
structure of p53,91–289 (PDB entry 2XWR chain A) as a template. We generated 100 models from each template using six distance restraints between pairs of residues forming non-covalent interactions in the N-terminal region of the p53 domain, as done for the dimer model described above. We then ranked the models on the root mean square deviation (RMSD) of atomic positions in the N-terminal region to the template structure of p53,91–289 (PDB entry 2XWR chain A) and finally we selected the model with the lowest RMSD. The resulting structure is labelled 3KZb-DNA. The additional steps were completed using a Snakemake pipeline incorporating the prepared PDB file, FoldX, and a mutational input file, which included the previously described aggregated mutations.

The pipeline includes a repair step, which aims to identify and repair residues containing suboptimal torsion angles, Van der Waals clashes, or total energy. After repair, the mutation step was carried out using the FoldX BuildModel command, where each mutation is modeled five times to avoid reliance on the deviation of a singular run. Concurrently, five self scans were carried out as a control, as we did for the monomer and the dimer above. For all ten models for each mutation, the analysis of Gibbs’s interaction free energy between DNA and protein was performed using the FoldX command AnalyzeComplex. We estimated the average ΔΔG values from the five runs.

### Molecular dynamics and metadynamics simulations

We carried out all-atom explicit-solvent MD and well-tempered metadynamics (WT-MetaD) simulations of wild-type and a selection of mutant variants of p53,91–289 in its free state (i.e., DNA-unbound) using GROMACS and the CHARMM22* force field. We selected the X-ray template structure of p53,91–289 (PDB entry 2XWR chain A) as a starting structure for the simulations. The structure of each mutant variant was obtained from the original orientation of the side chain. We used a dodecahedral box of TIP3P water molecules with a minimum distance between the protein and the box edges of 15 Å applying periodic boundary conditions. We used a concentration of NaCl of 150 mM, neutralizing the net charge of the system. We modeled all the histidine residues in the N2-H tautomeric state and the system was equilibrated in multiple steps. We used the Particle-mesh Ewald summation scheme and we truncated Van der Waals and Coulomb interactions at 9 Å. We collected 500-ns productive unbiased MD simulations and 300-ns WT-MetaD. The simulations were carried out at 298 K in the canonical ensemble using the velocity rescale thermostat. As a control, WT-MetaD simulations for the wild-type and some selected variants have been extended to 700 ns to ensure that our estimate of the populations of the conformational states of the S6-S7 loop was reasonably converged.

The WT-MetaD simulations were carried out with GROMACS and PLUMED. We used the same four CVs that we applied in a previous study of p53 DBD and which were discriminating between the solvent-exposed and occluded states of the S6-S7 loop and that were sensitive enough to understand the effects of residue modifications on the populations of the loop.

### Atomic contact analyses with CONAN

We employed the software CONtact Analysis (CONAN) to perform statistical analyses of intramolecular contacts along MD trajectories. The tool computes the inter-residue contacts using different user-defined cutoff distances, with a protocol similar to the one used in a previous study. We processed the data to estimate the occurrence of each contact, calculated as the ratio between the number of frames in which the contact is detected and the total number of frames of the trajectory, and to estimate the number of encounters during the simulation. For each pair of residues, the number of frames in which the contact is detected and the total number of frames of the trajectory, and to estimate the number of encounters during the simulation (i.e., a number indicating how many times the contacts is formed and broken).

### Protein Structure Networks (PSN)

We used a PSN method based on atomic contacts (acPSN) as originally formulated by Vishveshwara’s group and implemented in PyInteraph. We applied this method to MD simulations (PSN-MD). We retained only pairs of residues whose sequence distance is higher than 4.5 Å. The calculation, normalization factors for each residue are applied as previously discussed. The normalization factors are used as proxies for the propensity of the residues to form contacts. For each pair of residues, the number of pairs of atoms lying within 4.5 Å is divided by the square root of the product of the normalization factors of the two residues, and the result is multiplied by 100. This value represents the interaction strength between the two residues. If the interaction strength exceeds a pre-defined cut-off (I<sub>min</sub> estimated as 2.5 for our simulations), an edge is drawn between the two residues with weight equal to the interaction strength. In addition, the acPSN will only include as edges those that have an occurrence higher than 50% of the frames in the ensemble. The interaction strength associated with each edge is in the final acPSN the average interaction strength. The interaction strength associated with each edge is in the final acPSN the average interaction strength between the edge over all the structures of the ensemble. The final acPSN is weighted according to the average interaction strength. We estimated the paths of communications between selected starting and end residues using PyInteraph, which integrates the algorithm implemented in NetworkX.
for the calculation of all shortest paths between pairs of nodes. We estimated as hubs in the acPSN residues with an edge degree higher than 3, as previously suggested.136

Estimate of allosteric free energy changes upon mutation

We used the structure-based statistical mechanic model of allostery implemented in AlloSigMA 272 to predict the changes in allosteric free energy upon mutations. This coarse grained model can model mutations as ‘UP’ or ‘DOWN’ mutations depending on the changes in steric hindrance after the substitution. A UP mutation will mimic a residue replacement with a bulky residue, whereas the DOWN mutation will resemble the effect of a replacement with a smaller side chain. To match each of the selected 395 mutations with the corresponding AlloSigMA mode, we estimated the changes in the volume of the amino acid elicited by the mutation with an in-house Python script (allosigma-classify), starting from an available list of residue volumes.137 If the absolute value of the change was ≤ 5 Å³ we discarded the mutation from this analysis since we do not expect that the prediction could be reliable for small changes of a steric hindrance (examples are substitutions from tyrosine to phenylalanine or serine to alanine). This analysis led to discarding 66 variants from the allosteric energy estimates. In addition, we removed variants that are in residues in direct contact with the S6-S7 loop according to the atomic contacts or that are located in the S6-S7 loop (residues 207–213). As a result, 334 variants were retained for analyses with AlloSigMA 2. We then performed the full allosteric signaling map analysis as available on the AlloSigMA 2 server on the X-ray p53 DBD structure (PDB ID 2XWR, chain A) and downloaded the corresponding session file (February 2022). Of note, in the discussion of the results from AlloSigMA 2 we refer to the changes in allosteric free energy upon mutations as Δg, in alignment with the original AlloSigMA 2 terminology.

We used an in-house developed Python script (allosigma-heatmap) to extract the response information on the full protein as ΔgUP or ΔgDOWN for each position on which a mutation was found and depending on the mutation class (UP or DOWN). We further restricted the response sites to specific regions of interest. The resulting allosteric free energy changes (ΔgUP or ΔgDOWN) can have positive or negative signs. They predict the regions of the protein sensitive to destabilization (i.e., increased flexibility) upon mutation and the ones that are stabilized (i.e., decreased dynamics), respectively. The effect of a mutation is modeled through the modification of the strength of the intramolecular contacts among the protein residues.

Principal Component Analysis (PCA) for L1 and S6-S7 loop dynamics

PCA can be used to analyze MD trajectories with reference to the eigenvectors (i.e., principal components, PCs) of the mass-weighted covariance matrix of the atomic positional fluctuations.76 We here applied PCA, upon fitting on the Cα atoms, to calculate the eigenvectors associated with the Cα of the residues of the L1 and S6-S7 loops. This provides a description in a reduced conformational space of the dynamics and conformation of these loops during the simulations. We used for the comparison a common subspace derived from a PCA of a concatenated MD trajectory including all the 15 variants and the wild type protein. The first three PCs are sufficient to account for more than 80% of the variance for the loops.

CScape-somatic

CScape-somatic is a tool that uses a single kernel method to combine several features to predict somatic mutation driver potential.78 It is based on a forward selection approach where sequential learning identified an optimal combination of feature groups based on 30 possible data sources. The final model feature groups included conservation, GC content, sequence uniqueness, local mutation frequency, proximity to gene features, spectrum, and functional elements. All used data were based on the GRCh37/hg19 version of the human genome. The prediction scores are given as probability estimates in the range [0, 1]. We here applied the cutoff of 0.89 for high confidence predictions of variants in the coding region.

To apply CScape-somatic, we used an R script to convert the data available from the aggregated metatable of variants. From this, we extracted index, residue position, wild type residue, mutated residue, and genomic mutation. The potential list of several genomic mutations for each residue was converted from HGVS format to separate columns. Furthermore, we converted all coordinates to hg19 with liftOver to match the reference genome used by CScape-somatic.

REVEL

We used MyVariant.info, Web resource138 to retrieve the REVEL score77 for each mutation of the p53 DBD. REVEL is an ensemble method to predict pathogenic mutations. The score can range from 0 to 1, with higher values indicating more detrimental mutations. We used as a cutoff a REVEL score of 0.4 to classify the mutations, as suggested in a recent benchmarking study.139 This value represents a good trade-off between sensitivity and specificity.
Data availability

The scripts, input, and main outputs from this study will be made available in a GitHub repository associated with the publication https://github.com/ELELAB/p53_S6-S7. The trajectories and other large files are deposited in the OSF repository https://osf.io/38khf/.

CRediT authorship contribution statement

Kristine Degn: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. Ludovica Beltrame: Conceptualization, Formal analysis, Investigation, Validation, Visualization, Writing – original draft, Writing – review & editing. Freja Dahl Hede: Data curation, Formal analysis, Writing – review & editing. Valentina Sora: Methodology, Software, Writing – review & editing. Vincenzo Nicolaci: Formal analysis, Writing – review & editing. Marina Vabiltevis: Formal analysis, Writing – review & editing. Kjeld Schmiegelow: Funding acquisition, Resources, Writing – review & editing. Karin Wadt: Data curation, Funding acquisition, Resources, Writing, Writing – review & editing. Matteo Lambrughi: Conceptualization, Formal analysis, Investigation, Validation, Visualization, Supervision, Writing – original draft, Writing – review & editing. Elena Papaleo: Conceptualization, Data curation, Formal analysis, Funding acquisition, Resources, Investigation, Resources, Validation, Visualization, Supervision, Project administration, Writing – original draft, Writing – review & editing.

Acknowledgment

The result of this research has been achieved using the DECI-PRACE 16th HPC Grants on Archer2. The research has been also carried out thanks to the access to the Danish HPC Infrastructure Computerome2. The research has been supported by grants from LEO Foundation (LF17006), Danmarks Grundforskningsfond (DNRF125), Hartmann Foundation (A33877), Danish Cancer Society (R-257-A14720), and the Danish Childhood Cancer Foundation (2019-5934 and 2020-5769). This work is part of Interregional Childhood Oncology Precision Medicine Exploration, across Oresund collaboration between University Hospital Copenhagen, Rigshospitalet, Lund University, RegionSkåne, and Technical University Denmark, supported by the European Regional Development Fund. This work is also part of the nationwide research program Childhood Oncology Network Targeting Research, Organisation & Life expectancy (CONTROL).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jmb.2022.167663.

Received 1 April 2022; Accepted 25 May 2022; Available online 31 May 2022

Keywords: Allosteric Interactions; Therapeutic Target; Molecular Dynamics; Protein Stability; Protein Structure Network

Abbreviations:
acPSN, Atomic Contact Protein Structure Network; CHARMM22*, Chemistry at Harvard Macromolecular Mechanic, forcefield 22*; CONAN, CONtact ANalysis; COSMIC, Catalogue of Somatic Mutations in Cancer; DBD, DNA binding domain; DNA, Deoxyribonucleic acid; FES, Free energy surface; GC (as in GC content), cytosine (C), guanine (G); GOF, Gain of Function; GROMACS, GROningen Machine for Chemical Simulation; LFS, Li, Fraumeni Syndrome; MD, Molecular Dynamics; NMR, Nuclear Magnetic Resonance; OMIM, Online Mendelian Inheritance in Man; PCA, Principal Component Analysis; PDB, Protein Data Bank; PSN, Protein Structure Network; PTM, Post Translational Modification; PUMA, p53 upregulated modulator of apoptosis; TP53, p53, Tumor Protein 53; RMSD, Root Mean Square Deviation; WT, Wild Type

References

1. Vogelstein, B., Lane, D., Levine, J., (2000). Surfing the p53 network. Nature 408, 307–310. https://doi.org/10.1038/35042675.
2. Aylon, Y., Oren, M., (2016). The Paradox of p53: What, how, and why? Cold Spring Harbor Perspect. Med. 6 https://doi.org/10.1101/cshperspect.a026328.
3. Colaprico, A., Olsen, C., Bailey, M.H., Odom, G.J., Terkelsen, T., Silva, T.C., Olsen, A.V., Cantini, L., et al.,
(2020). Interpreting pathways to discover cancer driver genes with Moonlight. *Nature Commun.* **11**, 69. https://doi.org/10.1038/s41467-019-13803-0.

4. Shen, L., Shi, Q., Wang, W., (2018). Double agents: genes with both oncogenic and tumor-suppressor functions. *OncoGenesis* **7**, 25. https://doi.org/10.1038/s41389-018-0034-x.

5. Zhang, Y., Lozano, G., (2017). p53: Multiple Facets of a Rubik’s Cube. *Annual Rev. Cancer Biol.* 1 https://doi.org/10.1146/annurev-cancerbio-050216-121926.

6. Voussen, K.H., Prives, C., (2009). Blinded by the Light: The Growing Complexity of p53. *Cell* **137**, 413–431. https://doi.org/10.1016/j.cell.2009.04.037.

7. Green, D.R., Kroemer, G., (2009). Cytoplasmic functions of the tumour suppressor p53. *Nature* **458**, 1127–1130. https://doi.org/10.1038/nature07986.

8. Speidel, D., (2010). Transcription-independent p53 apoptosis: an alternative route to death. *Trends Cell Biol.* **20**, 14–24. https://doi.org/10.1016/j.tcb.2009.10.002.

9. Taedemir, E., Mauter, M.C., Galluzzi, L., Vitale, I., Djavaheri-Mergny, M., D’Amelio, M., Crivolo, A., Morselli, E., et al., (2008). Regulation of autophagy by cytoplasmic p53. *Nature Cell Biol.* **10**, 676–687. https://doi.org/10.1038/ncl_Cell.2010.170.

10. Kroemer, G., Bravo-San Pedro, J.M., Galluzzi, L., (2015). Novel function of cytoplasmic p53 at the interface between mitochondria and the endoplasmic reticulum. *Cell Death Dis.* **6**, https://doi.org/10.1038/cddis.2015.70 e1698.

11. Luo, Q., Beaver, J.M., Liu, Y., Zhang, Z., (2017). Dynamics of p53: A master decoder of cell fate. *Genes (Basel)* **8** https://doi.org/10.3390/genes8020066.

12. Kandoth, C., McLellan, M.D., Vandin, F., Ye, K., Niu, B., Lu, C., Xie, M., Zhang, Q., et al., (2013). Mutational landscape and significance across 12 major cancer types. *Nature* **502**, 333–339. https://doi.org/10.1038/nature12634.

13. Leroy, B., Anderson, M., Soussi, T., (2014). TP53 mutations in human cancer: Database reassessment and prospects for the next decade. *Hum. Mutat.* **35**, 672–688. https://doi.org/10.1002/humu.22552.

14. Byrjalsen, A., Hansen, T.V.O., Stoltze, U.K., Mehrjouy, M., Barnkob, N.M., Hjalgrim, L.L., Mathiasen, R., Lautrup, M., Bergh, B., et al., (2020). Nationwide germline whole genome sequencing of 198 consecutive pediatric cancer patients reveals a high frequency of cancer prone syndromes. *PLoS Genet.* **16** https://doi.org/10.1371/journal.pgen.1009231.

15. Parsons, D.W., Roy, A., Yang, Y., Wang, T., Scollon, S., Bergstrom, K., Kerstein, R.A., Gutierrez, S., et al., (2016). Diagnostic yield of clinical tumor and germline whole-exome sequencing for children with solid tumors. *JAMA Oncol.* **2**, 616–624. https://doi.org/10.1001/jamaoncol.2015.5699.

16. Waszak, S.M., Northcott, P.A., Buchhalter, I., Robinson, G.W., Sutter, C., Groebner, S., Grund, K.B., Brugliera, L., et al., (2018). Spectrum and prevalence of genetic predisposition in medulloblastoma: a retrospective genetic study and prospective validation in a clinical trial cohort. *Lancet Oncol.* **19**, 785–798. https://doi.org/10.1016/S1470-2045(18)30242-0.

17. Mirabelli, L., Zhu, B., Koster, R., Karlins, E., Dean, M., Yeager, M., Paner, M., Spector, L.G., et al., (2020). Frequency of Pathogenic Germline Variants in Cancer-Susceptibility Genes in Patients with Osteosarcoma. *JAMA Oncol.* **6**, 724–734. https://doi.org/10.1001/jamaoncol.2020.0197.

18. Malkin, D., (2011). Li-fraumeni syndrome. *Genes Cancer* **2**, 475–484. https://doi.org/10.1177/1947601911413466.

19. Bougeard, G., Renaux-Petel, M., Flaman, J.M., Charbonnier, C., Femrey, P., Belotti, M., Gauthier-Villars, M., Stoppa-Lyonnet, D., et al., (2015). Revisiting Li-Fraumeni syndrome from TP53 mutation carriers. *J. Clin. Oncol.* **33**, 2345–2352. https://doi.org/10.1200/JCO.2014.59.5728.

20. Stoltze, U., Skyyte, A.B., Roed, H., Hasle, H., Ejlersen, B., Hansen, T.V.O., Schmiegelow, K., Gerdes, A.M., et al., (2018). Clinical characteristics and registry-validated extended pedigrees of germline TP53 mutation carriers in Denmark. *PLoS ONE* **13** https://doi.org/10.1371/journal.pone.0190050.

21. Joerger, A.C., Fersht, A.R., (2016). The p53 Pathway: Origins, Inactivation in Cancer, and Emerging Therapeutic Approaches. *Annual Rev. Biochem.* **85** https://doi.org/10.1146/annurev-biochem-060815-014710.

22. Hoe, K.K., Verma, C.S., Lane, D.P., (2014). Drugging the p53 pathway: understanding the route to clinical efficacy. *13*, 217–236. https://doi.org/10.1038/ndmb2436.

23. Petty, T.J., Emtiazmadah, S., Costantino, L., Petkova, I., Stavridi, E.S., Saven, J.G., Vauthey, E., Halazonetis, T. D., (2011). An induced fit mechanism regulates p53 DNA binding kinetics to confer sequence specificity. *EMBO J.* **30**, 2167–2176. https://doi.org/10.1038/emboj.2011.127.

24. Khoo, K.H., Andreeva, A., Fersht, A.R., (2009). Adaptive Evolution of p53 Thermodynamic Stability. *J. Mol. Biol.* **393**, 161–175. https://doi.org/10.1016/j.ymol.2009.08.013.

25. Lambrughi, M., de Gioia, L., Gervasio, F.L., Lindorff-Larsen, K., Nussinov, R., Urani, C., Bruschi, M., Papaleo, E., (2016). DNA-binding protects p53 from interactions with cofactors involved in transcription-independent functions. *Nucleic Acids Res.* **44**, 9096–9109. https://doi.org/10.1093/nar/gkw770.

26. Tang, Y., Yao, Y., Wei, G., (2021). Unraveling the Allosteric Mechanism of Four Cancer-related Mutations in the Disruption of p53-DNA Interaction. *J. Phys. Chem. B* **125**, 10138–10148. https://doi.org/10.1021/acs.jpcb.1c05638.

27. Liang, Z., Verkhivker, G.M., Hu, G., (2019). Integration of network models and evolutionary analysis into high-throughput modeling of protein dynamics and allosteric regulation: theory, tools and applications. *Briefings Bioinf.* **00**, 1–21. https://doi.org/10.1093/bib/bbx028.

28. Barros, E.P., Demir, Ö., Soto, J., Cocco, M.J., Amaro, R. E., (2021). Markov state models and NMR uncover an overlooked allosteric loop in p53. *Chem. Sci.* **12**, 1891–1900. https://doi.org/10.1039/d0sc05053a.

29. Rasquinha, J.A., Bej, A., Dutta, S., Mukherjee, S., (2017). Intrinsic Differences in Backbone Dynamics between Wild Type and DNA-Contact Mutants of the p53 DNA Binding Domain Revealed by Nuclear Magnetic Resonance Spectroscopy. *Biochemistry* **56**, 4962–4971. https://doi.org/10.1021/acs.biochem.7b00514.

30. Blanden, A.R., Yu, X., Blayney, A.J., Demas, C., Ha, J.H., Liu, Y., Withers, T., Carpizo, D.R., Loh, S.N., (2020). Zinc shapes the folding landscape of p53 and establishes a pathway for reactivating structurally diverse cancer mutants. *Elife* **9**, 1–26. https://doi.org/10.7554/eLife.61487.
33. Bouaoun, L., Sonkin, D., Ardin, M., Hollstein, M., Byrnes, M.R., Siau, J.W., Kannan, S., Nguyen, M.N., Pradhan, M., (2021). Advances in germline TP53 Variations in neoplasms. Nature Rev. Cancer 21, 122–137. https://doi.org/10.1038/s41568-020-00315-z.

34. Klco, J.M., Mullighan, C.G., (2021). From Somatic TP53 Variants to Be Pathogenic in Patients With Osteosarcoma Are De Novo: A Report From the Children’s Oncology Group. JCO Precis. Oncol. 4, 1187–1195. https://doi.org/10.1200/PO.21.

35. Fortuno, C., Lee, K., Olivier, M., Attardi, L.D., Crowley, S., et al., (2021). Specifications of the ACMG/AMP variant interpretation guidelines for germline TP53 variants. Hum. Mutat. 42, 223–236. https://doi.org/10.1002/humu.24152.

36. Raad, I., Lee, K., Olivier, M., Persan, T., Mai, P.L., de Andrade, K.C., Attardi, L.D., Crowley, S., et al., (2021). Blood functional assay for rapid clinical interpretation of germline TP53 variants. J. Med. Genet. 58, 796–805. https://doi.org/10.1136/jmedgenet-2020-107059.

37. Singhal, D., Hahn, C.N., Feinstein, S., Wee, L.Y.A., Moma, L., Kutyna, M.M., Chhetri, R., Eshraghi, L., et al., (2021). Targeted gene panels identify a high frequency of pathogenic germline variants in patients diagnosed with a hematological malignancy and at least one other independent cancer. Leukemia 35, 3245–3256. https://doi.org/10.1038/s41375-021-01249-w.

38. Douglass, D.P., Stine, K.C., Farrar, J.E., (2021). A Novel Germline TP53 Mutation in a Patient With Li-Fraumeni Syndrome: Resolving a Variant of Uncertain Significance. J. Pediatr. Hematol. Oncol. 43, e1220–e1222. https://doi.org/10.1097/MPH.0000000000002115.
missense mutations. Nucleic Acids Res. 49, D475–D479. https://doi.org/10.1093/NAR/GKA925.

55. Joosten, R.P., Joosten, K., Cohen, S.X., Vriend, G., Perrakis, A., (2011). Automatic rebuilding and optimization of crystallographic structures in the Protein Data Bank. Bioinformatics 27, 3392–3398. https://doi.org/10.1093/bioinformatics/btr590.

56. Nielsen, S.V., Stein, A., Dinitzen, A.B., Papaleo, E., Tatham, M.H., Poulsen, E.G., Kassem, M.M., Rasmussen, L.J., et al., (2017). Predicting the impact of Lynch syndrome-causing missense mutations from structural calculations. PLOS Genet. 13, e1006739. https://doi.org/10.1371/journal.pgen.1006739.

57. Buß, O., Rudat, J., Ochsenreither, K., (2018). FoldX as Protein Engineering Tool: Better Than Random Based Approaches?, Computational and Structural. Biotechnol. J. 16, 25–33. https://doi.org/10.1016/j.bjst.2018.01.002.

58. Pucci, F., Bernaerts, K.V., Giesbers, E., Boljevic, V., (2018). Qualification of biases in predictions of protein stability changes upon mutations. Bioinformatics 34, 3659–3665. https://doi.org/10.1093/BIOINFORMATICS/BTY348.

59. Usmanova, D.R., Bogatyreva, N.S., Bernad, J.A., Eremina, A.A., Gorshkova, A.A., Kanevskiy, G.M., Lonishin, L.R., Meister, A.V., et al., (2018). Self-consistency test reveals systematic bias in programs for prediction change of stability upon mutation. Bioinformatics 34, 3653–3658. https://doi.org/10.1093/BIOINFORMATICS/BTY340.

60. Ghosh, A., Vishveshwara, S., (2007). A study of communication pathways in methionyl- tRNA synthetase by molecular dynamics simulations and structure network analysis. Proc. Natl. Acad. Sci. U. S. A. 104, 15711–15716. https://doi.org/10.1073/PNAS.0704459104.

61. Sora, V., Tiberti, M., Mahdi Robbani, S., Rubin, J., Papaleo, E., (2020). PyInteraph2 and PyInterkne2 to analyze networks in protein structural ensembles. BioRxiv. https://doi.org/10.1101/2020.11.22.381616. 2020.11.22.381616.

62. Weinberg, R.L., Vepринтsev, D.B., Fersht, A.R., (2004). Cooperative binding of tetrameric p53 to DNA. J. Mol. Biol. 341, 1145–1159. https://doi.org/10.1016/j.jmb.2004.06.071.

63. Kitayner, M., Rozenberg, H., Kessler, N., Rabinovich, D., Shaulov, L., Haran, T.E., Shakked, Z., (2006). Structural Basis of DNA Recognition by p53 Tetramers. Mol. Cell 22, 741–753. https://doi.org/10.1016/j.molcel.2006.05.015.

64. H. Tidow, R. Melero, E. Mylonas, S.M.V. Freund, J.G., (2020). BTY348. Bioinformatics 36, 15716–15720. https://doi.org/10.1093/bioinformatics/btaa242.

65. Chen, Y., Dey, R., Chen, L., (2010). Crystal Structure of the p53 Core Domain Bound to a Full Consensus Site as a Self-Assembled Tetramer. Structure 18, 246–256. https://doi.org/10.1016/j.str.2009.11.011.

66. Kitayner, M., Rozenberg, H., Roahs, R., Suad, O., Rabinovich, D., Honig, B., Shakked, Z., (2010). Diversity in DNA recognition by p53 revealed by crystal structures with Hoogsteen base pairs. Nature Struct. Mol. Biol. 17, 423–429. https://doi.org/10.1038/nsmb.1800.

67. R. Melero, S. Rajagopalan, M. Lázaro, A.C. Joerger, T. Brandt, D.B. Vepринtsev, G. Lasso, D. Gil, et al., Electron microscopy studies on the quaternary structure of p53 reveal different binding modes for p53 tetramers in complex with DNA. (n.d.). https://doi.org/10.1073/pnas.1015520107/-/DCSupplemental.

68. Lukman, S., Lane, D.P., 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, https://doi.org/10.1371/journal.pone.0080221 e80221.

69. Liu, Q., Kaneko, S., Yang, L., Feldman, R.I., Nicosia, S.V., Chen, J., Cheng, J.O., (2004). Aurora-A abrogation of p53 DNA binding and transactivation activity by phosphorylation of serine 215. J. Biol. Chem. 279, 52175–52182. https://doi.org/10.1074/jbc.M406802200.

70. Bista, M., Freund, S.M., 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, 15752–15756. https://doi.org/10.1073/pnas.1214176109.

71. Bhattacharjee, S., Mukherjee, S., Roy, S., (2021). DNA-Bound p53-DNA-Binding Domain Interconverts between Multiple Conformations: Implications for Partner Protein Recognition. J. Phys. Chem. B 125, 5832–5837. https://doi.org/10.1021/acs.jpcb.1c03794.

72. Tan, Z.W., Guanana, E., Tee, W.V., Berezovsky, I.N., (2020). AlloSigMA 2: Paving the way to designing allosteric effectors and to exploring allosteric effects of mutations. Nucleic Acids Res. 48, W116–W124. https://doi.org/10.1093/NAR/GKA338.

73. Y. Cho, S. Gorina, P.D. Jeffrey, N.P. Pavletich, Crystal structure of p53 tumor suppressor-DNA complex: understanding tumorigenic mutations, 1994.

74. Papaleo, E., Mereghetti, P., Fantucci, P., Grondori, R., de Gioia, L., (2009). Free-energy landscape, principal component analysis, and structural clustering to identify representative conformations from molecular dynamics simulations: The myoglobin case. J. Mol. Graph. Model. 27, 889–899. https://doi.org/10.1016/j.jmgm.2009.01.006.

75. Daidone, I., Amadei, A., (2012). Essential dynamics: foundation and applications, Wiley Interdisciplinary Reviews: Computational Molecular. Science 2, 762–770. https://doi.org/10.1002/wcms.1099.

76. Bussi, G., Laio, A., (2020). Using metadynamics to explore complex free-energy landscapes. Nature Rev. Phys. 2, 200–212. https://doi.org/10.1038/s42254-020-0153-0.

77. Ioannidis, N.M., Rothstein, J.H., Pejaver, V., Middha, S., McDonnell, S.K., Baheti, S., Musolf, A., Li, Q., et al., (2016). REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. Am. J. Hum. Genet. 99, 877–885. https://doi.org/10.1016/j.ajhg.2016.08.016.

78. Rogers, M.F., Gaunt, T.R., Campbell, C., (2020). CScape- somatic: Distinguishing driver and passenger point mutations in the cancer genome. Bioinformatics 36, 3637–3644. https://doi.org/10.1093/bioinformatics/btaa242.

79. Alford, R.F., Leaver-Fay, A., Jeliazkov, J.R., O’Meara, M. J., DiMaio, F.P., Park, H., Shapovalov, M.V., Renfrew, P. D., et al., (2017). The Rosetta All-Atom Energy Function Foundation and Applications, Wiley Interdisciplinary. https://doi.org/10.1002/wcms.1099.

80. Markthaler, D., Fleck, M., Stankiewicz, B., Hansen, N., (2022). Exploring the Effect of Enhanced Sampling on...
108. Echave, J., (2021). Fast computational mutation-response scanning of proteins. PeerJ 9 https://doi.org/10.7717/peerj.11330.

109. Kamisetty, H., Ovchinnikov, S., Baker, D., (2013). Assessing the utility of coevolution-based residue-residue contact predictions in a sequence- and structure-rich era. Proc. Natl. Acad. Sci. U. S. A. 110, 15674–15679. https://doi.org/10.1073/pnas.1314045110.

110. Hanahan, D., (2022). Hallmarks of Cancer: New Dimensions. Cancer Discovery 12, 31–46. https://doi.org/10.1158/2159-8290.CD-21-1059.

111. Friedel, L., Loewer, A., (2022). The guardian’s choice: how p53 enables context-specific decision-making in individual cells. FEBS J. 289, 40–52. https://doi.org/10.1111/febs.15767.

112. Tuncbag, N., Kar, G., Gursoy, A., Keskin, O., Nussinov, R., (2009). Towards inferring time dimensionality in protein–protein interaction networks by integrating structures: the p53 example. Mol. BioSyst. 5, 1770. https://doi.org/10.1039/b905661k.

113. Bethuyne, J., De Gieter, S., Zwaenepoel, O., Garcia-pino, A., Durinck, K., Verhelle, A., Hassanzadeh-ghassabeh, G., Spelman, F., et al., (2014). A nanobody modulates the p53 transcriptional program without perturbing its functional architecture. Nucleic Acids Res. 42, 12928–12938. https://doi.org/10.1093/nar/gku962.

114. Han, C.W., Lee, H.N., Jeong, M.S., Park, S.Y., Jang, S. B., (2021). Structural basis of the p53 DNA binding domain and PUMA complex. Biochim. Biophys. Res. Commun. 548, 39–46. https://doi.org/10.1016/j.bbrc.2021.02.049.

115. Bhattacharjee, S., Sengupta, J., (2021). Hidden electrostatic energy contributions define dynamic allosteric communications within p53 during molecular recognition. Biophys. J. 120, 4512–4524. https://doi.org/10.1016/bjp.2021.08.037.

116. Nygaard, M., Terkelsen, T., Olsen, A.V., Sora, V., Salamanca, J., Rizza, F., Bergstrand, S., di Marco, M., et al., (2016). The mutational landscape of the oncogenic MZF1 SCAN domain in cancer. Front. Mol. Biol. 3. 1–18. https://doi.org/10.3389/fmolb.2016.00078.

117. Kumar, M., Papaleo, E., (2020). A pan-cancer assessment of alterations of the kinase domain of ULK1, an upstream regulator of autophagy. Sci. Rep. 10. 14874. https://doi.org/10.1038/s41598-020-71527-4.

118. Kenig, S.M., Rissler, V., Terkelsen, T., Lambroughi, M., Papaleo, E., (2019). Alterations of the interactome of Bcl-2 proteins in breast cancer at the transcriptional, mutational and structural level. PLoS Comput. Biol. 15, e1007485. https://doi.org/10.1371/journal.pcbi.1007485.

119. Gao, J., Aksoy, B.A., Dugrsoz, U., Dresdner, G., Gross, B., Sumer, S.O., Sun, Y., Jacobsen, A., et al., (2013). Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci. Signal. 6, pl1. https://doi.org/10.1126/scisignal.2004088.

120. Tate, J.G., Bambord, S., Jubb, H.C., Sondka, Z., Beare, D. M., Bindal, N., Boutselakis, H., Cole, C.G., et al., (2019). COSMIC: The Catalogue Of Somatic Mutations In Cancer. Nucleic Acids Res.. https://doi.org/10.1093/nar/gky1015.

121. Hulsen, T., de Vlieg, J., Alkema, W., (2008). BioVenn – a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. BMC Genomics 9, 488. https://doi.org/10.1186/1471-2164-9-488.

122. Natan, E., Baloglu, C., Pagel, K., Freund, S.M.V., Morgner, N., Robinson, C.V., Persht, A.R., 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, 358–368. https://doi.org/10.1016/j.jmb.2011.03.047.

123. Wang, Y., Rosengarth, A., Luecke, H., (2007). Structure of the human p53 core domain in the absence of DNA. Acta Crystallogr. D Biol. Crystallogr. 63, 276–281. https://doi.org/10.1101/sf090744906048499.

124. N. Eswar, B. Webb, M.A. Marti-Renom, M.S. Madhusudhan, D. Eramian, M.-Y. Shen, U. Pieper, A. Sali, Comparative protein structure modeling using MODELLER., Current Protocols in Protein Science/ Editorial Board, John E. Coligan, et al. Chapter 2 (2007) Unit 2.9. https://doi.org/10.1002/0471140864.ps0209s50.

125. Delgado, J., Radosuy, L.G., Cianferoni, D., Serrano, L., (2019). FoldX 5.0: Working with RNA, small molecules and a new graphical interface. Bioinformatics, 1–2. https://doi.org/10.1093/bioinformatics/btz184.

126. Tiberti, M., Terkelsen, T., Degn, K., Beltrame, L., Canter Cremers, T., da Piedade, I., di Marco, M., Maiani, E., et al., (2022). MutateX: an automated pipeline for in silico saturation mutagenesis of protein structures and structural ensembles. Brief. Bioinf.. 1–16. https://doi.org/10.1093/bib/bbac074.

127. Park, H., Bradley, P., Greisen, P., Liu, Y., Mulligan, V.K., Kim, D.E., Baker, D., Dimaio, F., (2016). Simultaneous Optimization of Biomolecular Energy Functions on Features from Small Molecules and Macromolecules. J. Chem. Theory Comput. 12, 6201–6212. https://doi.org/10.1021/acs.jctc.6b00819.

128. Guerois, R., Nielsen, J.E., Serrano, L., (2002). Predicting changes in the stability of proteins and protein complexes: A study of more than 1000 mutations. J. Mol. Biol. 320, 369–387. https://doi.org/10.1006/jmbi.2002.2836.(02)00442-4.

129. Barducci, A., Bussi, G., Parrinello, M., (2008). Well-tempered metadynamics: A smoothly converging and tunable free-energy method. Phys. Rev. Lett. 100, 1–4. https://doi.org/10.1103/PhysRevLett.100.020603.

130. Abraham, M.J., Murtola, T., Schulz, R., Pål, S., Smith, J. C., Hess, B., Lindahl, E., (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX 2, 19–25. https://doi.org/10.1016/j.softx.2015.06.001.

131. Piana, S., Lindorff-Larsen, K., Shaw, D.E., (2011). How robust are protein folding simulations with respect to force field parameterization? Biophys. J. 100, L47–L49. https://doi.org/10.1016/j.bpj.2011.03.051.

132. Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W., Klein, M.L., (1983). Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. 79, 926. https://doi.org/10.1063/1.445869.

133. Bussi, G., Donadio, D., Parrinello, M., (2007). Canonical sampling through velocity rescaling. J. Chem. Phys. 126, 014101. https://doi.org/10.1063/1.2408420.

134. Bonomi, M., Camilloni, C., (2019). Biomolecular Simulations. Methods Mol. Biol. http://www.springer.com/series/7651.

135. Mercadante, D., Gräter, F., Daday, C., (2018). CONAN: A Tool to Decode Dynamical Information from Molecular
Interaction Maps. *Biophys. J.* **114**, 1267–1273. https://doi.org/10.1016/j.bpj.2018.01.033.

136. Kannan, N., Vishveshwara, S., (1999). Identification of side-chain clusters in protein structures by a graph spectral method. *J. Mol. Biol.* **292**, 441–464. https://doi.org/10.1006/JMBI.1999.3058.

137. Pommie`, C., Levadoux, S., Sabatier, R., Lefranc, G., Lefranc, M.-P., (2004). IMGT standardized criteria for statistical analysis of immunoglobulin V-REGION amino acid properties. *J. Mol. Recognit.* **17**, 17–32. https://doi.org/10.1002/jmr.647.

138. Xin, J., Mark, A., Afrasiabi, C., Tsueng, G., Juchler, M., Gopal, N., Stupp, G.S., Putman, T.E., et al., (2016). High-performance web services for querying gene and variant annotation. *Genome Biol.*. https://doi.org/10.1186/s13059-016-0953-9.

139. Li, J., Zhao, T., Zhang, Y., Zhang, K., Shi, L., Chen, Y., Wang, X., Sun, Z., (2018). Performance evaluation of pathogenicity-computation methods for missense variants. *Nucleic Acids Res.* **46**, 7793–7804. https://doi.org/10.1093/nar/gky678.