

Variable Regions of p53 Isoforms Allosterically Hard Code DNA Interaction

Isabel Armour-Garb, In Sub Mark Han, Benjamin S. Cowan, and Kelly M. Thayer*

 Cite This: *J. Phys. Chem. B* 2022, 126, 8495–8507

 Read Online

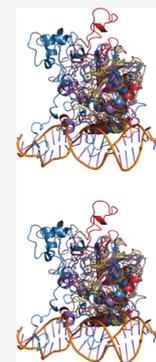
ACCESS |

 Metrics & More

 Article Recommendations

 Supporting Information

ABSTRACT: Allosteric regulation of protein activity pervades biology as the “second secret of life.” We have been examining the allosteric regulation and mutant reactivation of the tumor suppressor protein p53. We have found that generalizing the definition of allosteric effector to include entire proteins and expanding the meaning of binding site to include the interface of a transcription factor with its DNA to be useful in understanding the modulation of protein activity. Here, we cast the variable regions of p53 isoforms as allosteric regulators of p53 interactions with its consensus DNA. We implemented molecular dynamics simulations and our lab’s new techniques of molecular dynamics (MD) sectors and MD-Markov state models to investigate the effects of nine naturally occurring splice variant isoforms of p53. We find that all of the isoforms differ from wild type in their dynamic properties and how they interact with the DNA. We consider the implications of these findings on allostery and cancer treatment.



INTRODUCTION

Allosteric activation of a protein occurs when an allosteric effector binds to a site conferring a change at a distant location.^{1,2} It was first studied in the homotetrameric hemoglobin,^{3,4} where the binding of the first oxygen to one of the tetramers induces a conformation change in the quaternary structure, increasing the binding affinity in the subsequent three binding sites as an example of cooperativity. Allostery has been attributed to play a pervasive role in diverse biological functions including bacterial promoter control,^{5,6} transcription activation,⁷ DNA repair,^{8,9} signaling with G-coupled proteins,^{10,11} protein kinase regulation,¹² and many more. Despite the fact that the original studies occurred some half a century ago, allostery remains an active area of research because it is so widespread^{13,14} and because of the prospects for the rational design of allosteric drugs to treat currently undruggable targets.^{15–18}

The ability to capture the allosteric effect in a molecular dynamics simulation^{8,19,20} illustrates the fact that it can be described by first principles. Nevertheless, a comprehensive theory as to how signal transduction traverses the separation continues to be investigated.^{15,21} Several ideas have been put forth,^{22–25} and these can be broadly classified as pathway^{26–31} and energy landscape^{32,33} models. The pathway approach in simplest form can be understood as a mechanical domino effect, where a perturbation at the binding site of the effector causes sequential effects to subsequent neighboring residues constituting a path between the allosteric and active sites. Such models generally rely upon contact networks, where one residue is within some contact distance cutoff of the next

residue, which is often chosen to be 6 Å.³⁴ Other criteria to establish networks could be applied with a similar analysis.

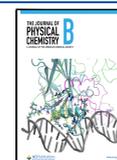
However, such an explanation is becoming increasingly difficult to reconcile in cases for which no allosteric conformational change can be measured readily, manifesting, for example, as a change in thermal fluctuations,³⁵ B-factors and NMR data,^{22,36} or dynamical dispersion.³⁷ A mounting number of cases of this type beg an alternate explanation. The energy landscape model purported on theoretical grounds by Cooper and Dryden offers such a viewpoint. The essence of the model suggests that communication may occur in difference in dynamic behavior resulting in a different distribution of energy states in the protein ultimately leading to observable changes in the active site. In a simple case, one can envision a change in the width of the dispersion about a common mean, for example, but of course, many other possibilities could exist. Cyclic AMP receptor protein (CAP) exemplifies such changes in NMR spectra, supporting this idea.

While categorizing allostery into distinct frameworks conveniences the researcher, in fact, the mechanisms may not be mutually exclusive;³⁸ what is actually observed may be the result of the superposition of many pathways and energy landscape changes. For small globular proteins, the emergence of contact pathways is not surprising because the surface-to-

Received: August 31, 2022

Revised: October 6, 2022

Published: October 17, 2022



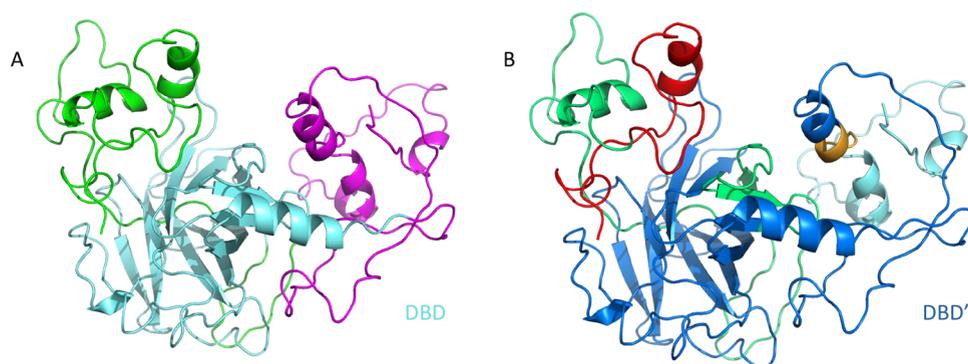


Figure 1. Structural mapping of isoform variable regions relative to major p53 domains. (A) The regions are colored as follows: N-terminus in green, DBD in cyan, and C-terminus in magenta. (B) Structural representation of p53 Phyre folded structure with the different regions colored based on IVRs. Red is IVR 1, green is IVR 2, dark blue is the DBD', yellow is IVR 3, and light blue is IVR 4. Conformations for wild type are shown.

volume ratio is so high.³⁹ In the absence of a full consideration of the system, underlying energetic networks could in fact be overlooked. In larger proteins, however, the ratio decreases, amplifying the effect. The complicated nature of allostery reflects the many years of evolution perfecting robust signaling mechanisms that are also tunable by even slight perturbations to the system.

The p53 tumor suppressor protein,^{40,41} the system chosen for this study, is involved in 50% or more of all human cancers. As such, it is a timely and important target of study. The wild-type protein (Figure 1A) consists of 393 amino acid residues. The central portion folds into a well-defined classic β barrel structure, which has been crystallized (PDB ID 1TUP).⁴⁰ However, the N- and C-terminal regions, accounting for about half of the protein's total length, are intrinsically disordered.^{42–44} For this reason, their structural determination has proven exceedingly difficult, and many studies of the protein work with constructs excluding these regions.

Nevertheless, the N- and C-terminal regions play a critical role in the regulation of the activity of p53.⁴⁵ In its tumor suppressor role, it must integrate cellular signals in the form of post-translational modifications to the N- and C-terminal regions. Programmed cell death (apoptosis) or DNA repair^{41,46,47} are the two major pathways it may trigger, depending on the severity of DNA damage. How such changes in intrinsically disordered regions modulates the activity of the p53 binding interface is largely unknown.⁴⁴

In this study, we consider a zeroth order model relative to the PTM modifications: the isoforms. Many eukaryotic cells are capable of modifying the protein to be expressed through a complex series of enzymes known as the spliceosome.⁴⁸ From a single gene, many variants can be produced by alternately splicing the mRNA before it becomes expressed as a protein. A certain subset of isoforms tends to be present in different tissue types (Table SI) and in characteristic proportions.^{49,50}

In human p53, there are nine major isoforms that we consider in this study. The C-terminal side is described as α , β , or γ moving from fully intact to truncated down to the compact highly structured DNA binding domain (DBD). The N-terminal composition is indicated by the delta symbol and the residue of the truncation, following common notation from genetics. We term the four blocks of sequence that are present or absent in the isoforms as isoform variable regions (IVRs) (Table 1 and Figure 1B, multiple sequence alignment Table SII). Of note, the DBD is customarily considered to be

Table 1. Organization of p53 Isoform Variable Regions^a

Δ Notation	Isoform	IVR 1	IVR 2	DBD	IVR 3	IVR 4
$\Delta 0p53\alpha$	Isoform 1 (wt)	Y	Y	Y	Y	Y
$\Delta 0p53\beta$	Isoform 2	Y	Y	Y		
$\Delta 0p53\gamma$	Isoform 3	Y	Y	Y	Y	
$\Delta 40p53\alpha$	Isoform 4		Y	Y	Y	Y
$\Delta 40p53\beta$	Isoform 5		Y	Y		
$\Delta 40p53\gamma$	Isoform 6		Y	Y	Y	
$\Delta 133p53\alpha$	Isoform 7			Y	Y	Y
$\Delta 133p53\beta$	Isoform 8			Y		
$\Delta 133p53\gamma$	Isoform 9			Y	Y	

^aThis chart lists each isoform and whether it contains or is missing each of the four IVRs. The other names listed correspond to the presence or absence of the IVRs. An α suffix indicates that an isoform has both IVR 3 and IVR 4. A β suffix indicates that an isoform has neither IVR 3 nor IVR 4. A γ suffix indicates that an isoform has IVR 3 but not IVR 4. A $\Delta 40$ prefix indicates that an isoform has IVR 2 but not IVR 1 (meaning that the first 40 amino acids are deleted). A $\Delta 133$ prefix indicates that an isoform has neither IVR 1 nor IVR 2 (meaning that the first 133 amino acids are deleted).

composed of residues 94–312, but because this is encroached upon by IVR2 up to residue 133, we define the DBD' to begin at residue 134 for this study. Generally, this consideration preserves the important elements of the DBD and encompasses the secondary element structures. The notable exception is K120 which may interact with DNA⁴⁰ and falls within the IVR2 region. The $\Delta 0p53\alpha$ represents the 393 amino acid full-length wild-type protein, and $\Delta 133p53\beta$ represents the most extensively truncated protein which roughly corresponds to the DBD used for crystallization (PDB ID 1TUP).⁴⁰

In this study, we extend the concept of allostery to encompass the presence and absence of the IVRs. We consider the central DBD', defined as the constant amino acids across the set, to be acted upon by the N- and C-terminal regions cast as the allosteric effectors. Within this framework, we then ask questions about the allosteric nature of the system using molecular dynamics (MD) simulations and MD-based measures. We characterize the effect of IVRs on p53 activity as "hard-coded" allosteric effectors. We then consider the ramifications of our observations on the development of molecular-based therapies and considerations for cancer

treatments depending on what tissues, and therefore what isoforms, are present.

METHODS

To address key questions regarding the isoforms of p53, we have undertaken Phyre folding of the sequences to produce starting structures for molecular dynamics simulations. We have determined the dynamic properties of the various isoforms. To investigate if and how dynamics differ, we have analyzed our resulting trajectories using both well-established means of assessment, and emergent techniques from our lab.

DBD vs DBD'. The canonical DBD has traditionally been based on the 1TUP crystal structure.⁴⁰ While much of the key secondary structure was determined in this early study, the disordered N- and C-terminal regions, now known to be important for post-translational modification for regulation of p53 function, were not structurally observed. Many experimental studies have revealed that the C-terminus, while being highly modified during its activation, in fact, anchors to the DNA directly and stabilizes p53 for tetramerization.^{51,52} This highlights the crucial role for both signaling and DNA recognition in this uncrystallized region. Much of the traditional DBD and this C-terminal region is conserved among all of the isoforms and therefore, we have chosen to encompass a modified DBD, DBD' as a basis for our analyses.

PHYRE2 Folding for Prediction of Molecular Dynamics Starting Structure. General steps for running an MD simulation may depend on the system that is being studied; systems based on known crystal structures are inherently more stable and, thus, may not require extensive energy minimization or equilibration. Because we do not know the structure of the full-length p53, we must use software to engineer an approximate structure. Thus, we refer to these structures as “highly engineered.” This is a description of the standard MD procedure used by the Thayer Lab, adjusted to accommodate for the highly engineered structures of the isoforms of p53.

We began by downloading the FASTA file for each isoform from the UniProt Knowledgebase (UniProtKB).⁵³ To generate a *.pdb file based on this sequence, we used the PHYRE2 web portal for protein prediction and analysis.

PHYRE2 allows us to fold a protein based solely on the protein's amino acid sequence as user input. Using advanced remote homology detection methods, PHYRE2 will analyze the structure effects of amino acid variants for a given protein.⁵⁴ This technology seeks to close the gap between the ever-growing number of known protein sequences and the convergent number of known protein structures.⁵⁵ Each isoform was analyzed using the “intensive” pipeline available through PHYRE2.

The models resulting from the PHYRE2 “intensive” pipeline execution all had confidence of greater than 90% for most residues. According to PHYRE2 standards, this confidence level is high enough for the model to be considered useful.⁵⁴ Thus, the PHYRE2 folded structures were chosen as the starting point for our MD simulations. The accession numbers for the isoforms from Uniprot⁵³ are P04637-1 (wild type) through P04637-9 (Δ 133p53 γ).

Molecular Dynamics Protocol. Each isoform was passed through the general Thayer Lab MD simulation protocol, modified to account for the highly engineered nature of the system, and optimized to make full use of parallelization. We used the AMBER 18^{56,57,54,55} suite of biomolecular simulation

programs to carry out the following steps: setup of parameters with tLEaP, Energy Minimization, Heating/equilibration, Dynamics, and cpptraj trajectory analysis. The resulting models from PHYRE2 were parameterized in tLEaP. The TIP3P potential was used for solvent water and the ff99SB force field to model atomic interactions.^{58–60} The chelated zinc was modeled using AMBER ZAFF parameters per previously established protocol.^{58,59}

Because we were working with highly engineered structures, we extended our usual equilibration phase and carefully monitored the AMBER energy of the constructs. For isoforms 2–9 we performed 5 rounds of equilibration with 2,500,000 steps in each round, starting with a restraint weight of 10 kcal/mol/Å² and decreasing by 2.5 in each round. The energy minimization plots are available in Supporting Figure S1. Isoform 1, the wild type, was taken from a previous study.⁵⁹ We opted to discard the first 300 ns of simulation as equilibration to sample a trajectory with a level of convergence comparable with the lengthy equilibration used for the other isoforms. Production dynamics was carried out to 200 ns using pmemd.cuda.MPI.⁶¹ A GitHub page for ongoing updates on relevant code is maintained at <https://github.com/iarmourgarb/p53-isoforms>.

Trajectory Analysis. Trajectory analysis was carried out using AMBER's cpptraj module.⁶⁰ We analyzed whether we have reached a relatively stable structure by analyzing the results of a root-mean-squared deviation (RMSD) calculation, which allows us to view a stabilization of the structure's deviation from the average structure of the isoform through the course of the simulation using backbone atoms. We also performed root-mean-squared fluctuation (RMSF) analysis to inspect the fluctuation of each individual residue from the average structure.

Molecular Visualization. Molecular dynamics simulations were animated for visual inspection using the visual molecular dynamics^{61,62} (VMD) software package. All structural figures were rendered using PyMOL⁶³ or VMD.

Linear Regression Analysis. Linear regression models were run in R to model the average RMSF of the DBD based on the presence or absence of the IVRs (Figure S3). In this case, the presence/absence of each IVR was a factor variable. This model reports if there is a significant change in the average RMSF of the DBD based on the presence or absence of an IVR and the percentage of the variance of the average RMSF of the DBD that can be explained by the presence or absence of the IVRs.

Statistical Network Analysis—Molecular Dynamics-Based Markov State Models. To discern the similarities and differences in conformational dynamics, MD-Markov state models (MD-MSMs)¹⁹ were employed. Briefly, the method identifies conformational substates of the trajectory using a *k*-means cluster approach. The method involves concatenating snapshots from all trajectories of interest and clustering them simultaneously, guaranteeing that centroids have been defined with respect to all involved structures. The key implication from this is that structural subsets common across constructs can be identified; thus, snapshots in a particular substate have been attracted by exactly the same centroid.¹⁹

For MD-MSM analysis, all trajectories were preprocessed to contain only the DBD' residues stripped to backbone atoms and concatenated for *k*-means cluster analysis. Because *k*, the number of clusters, is not known *a priori*, we tried two clusters through seven clusters and chose to work with *k* = 4 clusters

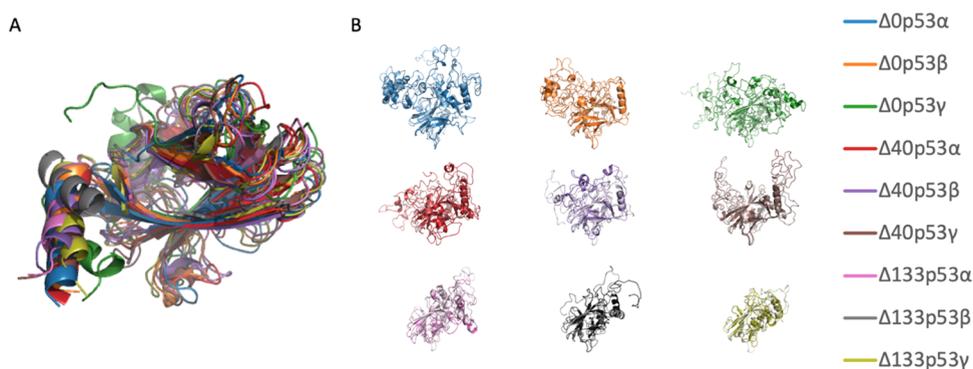


Figure 2. Structural alignment of IVR variant isoforms. (A) Alignment of the DBD' after 200 ns of MD simulation. The isoforms align in the β barrel region and in the C-terminal α helix. (B) Each isoform's initial structure obtained from Phyre folding aligned with the final snapshot from MD simulation. Here, it is clear that there is still complete alignment in the DBD', but there has been movement in the N- and C-terminal regions. The initial structure is in the darker color, and the final structure is in the lighter color for each isoform.

(Figure S2A). We have determined a protocol for assessing the optimal number of clusters;⁶⁴ generally, k is optimized when it corresponds to a good fit to the data that is not substantially improved by incrementing k to $k + 1$. We also check that the RMSD of cluster membership is low to its own centroid and high to all other centroids (Figure S2B). We are using backbone heavy atoms for clustering and the related RMSD measurements following recently published methods.⁵⁹

MD Sectors. Molecular dynamics sectors seek to identify networks of motionally covarying residues in a molecular dynamics simulation. MD Sectors is based on the work of Ranganathan *et al.* to identify evolutionarily covarying networks.^{28,65} We adapted the method to identify residues covarying with the motion of the protein. In short, the method employs a spectral analysis technique to remove the high-frequency noise to reveal the concerted motion of the protein.⁶⁶ The extent to which the positional information covaries is quantitated in the vpica metric. The measurement is pairwise but integrated into a matrix approach to quantitate the covariance of all residues. The method identifies a group of residues exhibiting such covariance. In addition to the PDZ system in which the method was developed and validated, it has been applied to GPX4 previously.^{66,67}

H-Bond Analysis. Trajectories obtained from MD simulation were subjected to hydrogen-bond analysis in the standard implementation of the AMBER package's cpptraj suite of programs. The analysis reports the number of hydrogen bonds made between the mentioned donor and acceptor, the identity of the atoms and residues involved, and the duration of time for which the interaction was present in the simulation. Here we have applied filters to simplify the data. We have kept frames for which the interaction is present for at least 5% of the time. In particular, we monitored the eight well-characterized residue interactions of p53 with its binding site DNA.⁴⁰ We also screened for spurious interactions arising from other H-bonds formation at the protein–DNA interface from the cpptraj H-bond analysis.

Electrostatic Network Generation. To investigate the role of IVRs on the electrostatic dynamics of the p53 protein's binding domain we implement molecular dynamics-energetic network decomposition (MD-END)⁶⁸ analysis on each of the nine p53 isoforms' DBD'. To capture the residue-residue electrostatic interactions for each p53-isoform, the energetic network for each of the 100 regularly sampled frames across the 200 ns trajectory of that system is generated. As the

conserved p53 protein-binding domain across all p53 isoforms is composed of 199 residues, each of these networks is represented as pairwise 199×199 matrices with pairwise values representing an “edge” weight between each residue pair, describing a normalized level of the electrostatic interaction energy between each residue pair which accounts for the effects of each residue's local energetic environment. Subsequently, the pairwise interaction edge weights undergo a localized thresholding protocol. This procedure results in a more sparse network, pruning edges to emphasize regions in the network where such environmental energetic contributions significantly occur. Networks are visualized using the Gephi software package.⁶⁹

Heat Kernel and Kernel PCA. For each of the nine p53 isoforms, the 100 sampled frames of locally thresholded energetic interaction networks are utilized to compute 100 different heat kernels using a previously described heat kernel analysis.⁶⁸ The heat kernel of a network⁷⁰ is a function simulating the diffusion of heat and/or information across the network over a parameter of time, measuring how information tends to “flow” within different communities of nodes and weighting significant topological features of network structure.

By calculating each frame's heat kernel over the matrix representation of the weighted network, the edge weights of each electrostatic network's significant topological properties are further biased to emphasize node connectivities, interactions, and the structure of the network on both local and global scales.⁶⁸ We capture the statistical variations in node associations across the heat kernels by projecting the 100 heat kernels onto a shared set of eigenbasis vectors through kernel PCA (KPCA). Each of the 199 residues is therefore represented by 100 node embeddings in the shared PCA space. As we aim to visually interpret differences between isoform heat kernel embeddings, we choose to project the nodes in R3 space across the three eigenvectors with the highest eigenvalues (PC1, PC2, and PC3). To achieve this, a scree plot⁷¹ of the mean kernel is generated to determine the time parameter t necessary to generate the heat kernel such that the significant number of factors (i.e., eigenvalues) to explain the variance in the heat kernel is 3. As such, we obtain a lower-dimensional representation of the most important modes of topological variation among node associations. Embedding each residue's electrostatic dynamics with other residues across simulation time endows insight into the dynamics of node clusters and associations in the context of

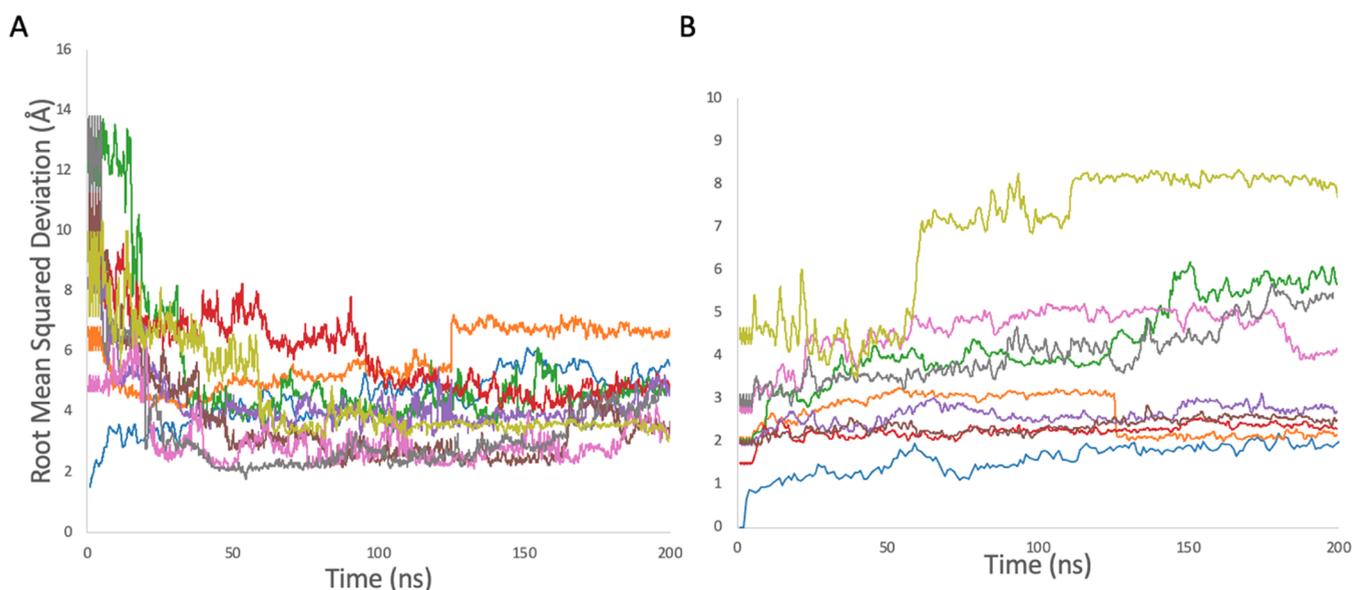


Figure 3. RMSD calculations. (A) RMSD plot for the full-length isoforms of p53. (B) RMSD plot for the DNA binding domain of the isoforms of p53. For this RMSD calculation, we use the DNA binding domain as the reference structure. Colors are as in Figure 2.

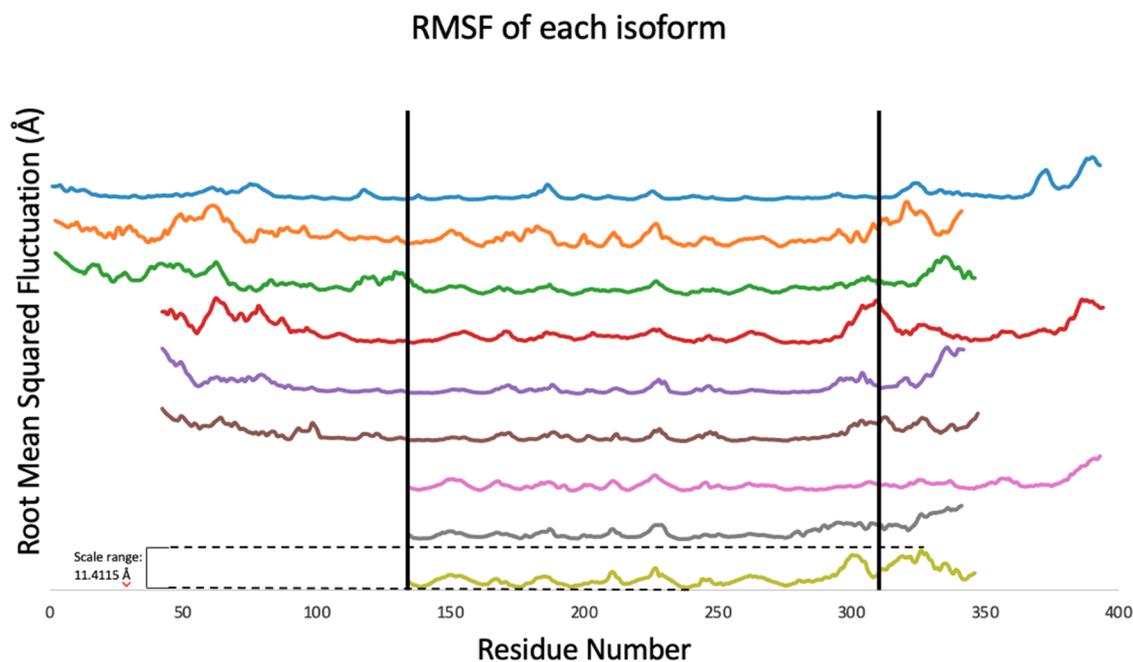


Figure 4. RMSF plot for the isoforms of p53. RMSF values for each isoform through 200 ns of simulation with residue numbers adjusted to account for the missing IVRs at the beginning of the sequence. The RMSF traces are separated out to analyze patterns of variation in certain residues. The vertical black bars indicate the beginning and end of the DBD'. All traces are to scale.

residue's energetic interactions over the course of the simulation. Thus, by analyzing the shared kernel PC space of each isoform's 100 heat kernels, we investigate how the electrostatic network of the protein evolves over the course of the MD simulation and how the topologies of these networks differ and parallel the electrostatic dynamics of other isoforms.

RESULTS

The major goal of the study is to characterize the influence of the IVRs on the dynamics and binding of the DBD' to the DNA. Constructs for each of the isoforms were obtained by PHYRE2 folding (Table SIII) the respective sequences for

each of the isoforms. These were then used for subsequent MD simulations generating the trajectories for the analysis. Figure 2A shows the local structural alignment of the DBD' with N- and C-terminal intrinsically disordered regions with considerable variation, and Figure 2B illustrates the initial PHYRE2 folded structure compared to the respective final MD snapshots of each isoform. Local structural alignments before and after simulation were also examined (Figures S4 and S5).

We begin our trajectory analysis with the root-mean-square deviation (RMSD) measure (Figure 3A) as a function of time, a typical way of assessing the overall stability of the system, and to ensure that the trajectory has reached convergence.

Table 2. Hydrogen Bonding by Isoform^a

residue	$\Delta 0p53\alpha$	$\Delta 0p53\beta$	$\Delta 0p53\gamma$	$\Delta 40p53\alpha$	$\Delta 40p53\beta$	$\Delta 40p53\gamma$	$\Delta 133p53\alpha$	$\Delta 133p53\beta$	$\Delta 133p53\gamma$
THR118					19.35				
<u>LYS120</u>		23.52							
SER121						8.76			
VAL122	38.31								
CYS135			8.86						
GLN136			15.89						
LEU137			24.05						
ALA138			7.3						
ASP138			8.17						
ASP184			16.29						
SER185									
SER240						17.43			10.97
<u>SER241</u>						9.26		48.63	7.56
MET243			6.55						
CYS275			29.84						
<u>ALA276</u>		32.08	22.39		6.55	17.48			
<u>CYS277</u>			8.52						
<u>ARG280</u>	18.66								
HIE296					9.13				
GLY302	43.57								
SER303	40.63								
ASN310	76.45						37.83		
ASN311	43.67								
THR312	36.26								
SER313							7.65		
SER314							23.66		
SER315							61.64		
GLU336				23.2					
ARG337				41.7					
THR377	18.99								
SER378	23.78								

^aEach table entry represents the percent of frames of simulation that a residue made hydrogen bonds with for each isoform. Only residues that make contact with DNA in more than 5% of frames are included in our analysis. The underlined residues (yellow spheres, Figure 5) are those that we would expect to make significant contact according to the 1TUP crystal structure to the same DNA sequence.

When simulating a basic, known crystal structure, we usually expect to see RMSD values near 2 Å.⁵³ However, the isoforms of p53 do not have known crystal structures, so we are simulating highly engineered structures. Thus, we can anticipate seeing higher RMSD in our simulations.

By the time we reach 200 ns of simulation, most of the isoforms have reached a plateau in their RMSD measurement. This indicates that we have sufficiently equilibrated our systems, which may have taken a bit longer than would be expected due to their highly engineered structure. At this point, we can be confident that these trajectories are valid to address our questions.

All of the isoforms, with the exception of $\Delta 0p53\beta$, level out to an RMSD value between 3 and 5 Å. Though this is higher than the RMSD value, we seek when simulating a known crystal structure, we still view this as sufficient. This is about what we would expect, given the highly engineered nature of the structures that resulted from PHYRE2 and also taking the structure of the N- and C-terminal regions into consideration. The highly mobile N- and C-terminal regions of the isoforms cause the RMSD of the entire structure to be higher on average.

In addition to analyzing the RMSD of the full-length isoforms, we also looked at the RMSD for the DBD of each isoform (Figure 3B). We chose to use the DBD, rather than

DBD', for consistency with previous studies. We expect each isoform to have a DBD that closely resembles that of the wild type; thus, we would expect Phyre to give us a nearly perfect approximation of this region. The DBD is tightly folded compared to the highly mobile N- and C-terminal regions. We also wanted to analyze the RMSD of the DBD to confirm that the core domain would remain tightly packed and that the higher RMSD values for the full-length isoforms were originating from motion in the highly mobile N- and C-terminal regions. It is evident that the isoforms did not unfold during the simulation. From this, we were encouraged by the quality of the simulations, prompting us to explore the residue specific properties of the system.

Root-mean-squared fluctuation (RMSF) analysis (Figure 4) provides information about the fluctuation of an individual residue of a protein with respect to the average structure throughout the simulation. We use RMSF analysis to understand which residues or areas of a protein are structurally in flux through simulation. Because we are working with a highly engineered structure with unknown structure in the N- and C-terminal regions, we expect our RMSF values to be relatively higher through the N- and C-terminal regions. Peaks and valleys in the RMSF plot of the DBD' can also be indicative of secondary structural elements in this area. The *x*-axis of an RMSF plot represents the residue number. In the

case of these simulations, one frame of simulation is 0.002 ns. On the y -axis, we measure fluctuation from the average structure in Ångströms.

Overall, in the DBD', most of the isoforms level out to a lower RMSD value compared to their full-length RMSD value. The average RMSF values of residues in the DBD remain below 3.5 Å for all of the isoforms. Notably, the isoforms that have the highest RMSF values at the end of the DBD' are isoforms $\Delta 40p53\beta$, $\Delta 133p53\beta$, and $\Delta 133p53\gamma$. These are the only isoforms that are missing more than two IVRs, which suggests to us that the missing portions of the N- and C-terminal regions may act as structural regulators. Additionally, isoforms $\Delta 0p53\beta$ and $\Delta 0p53\gamma$ are the only isoforms that contain both IVR 1 and IVR 2, but they have the highest average RMSF in the DBD by more than 0.4 Å. From this, we drew the major observation that the isoforms differ widely from the wild type, and also from each other.

To explore whether the effects of adding and subtracting the IVRs may have an additive effect on RMSF, we constructed a linear regression model. The first model measured the effect of the presence or absence of the IVRs on the average RMSF of the DBD. While none of the IVRs proved to be statistically significant at the 0.05 level, we did observe a multiple R^2 value of 0.750 (Figure S3A), indicating 75.0% of the variance in the average RMSF of the DBD can be explained by the presence or absence of IVRs.

The second linear regression model measured the effect of the average RMSF value of each IVR on the average RMSF of the DBD (Figure S3B). Like the previous model, none of the average IVR values proved to be statistically significant at the 0.05 level. However, we observed a multiple R^2 value of 0.718. This allows us to conclude that 71.8% of the variance in the average RMSF of the DBD can be explained by the average RMSF of the IVRs. We thus have quantitative RMSF observations and a quantitative regression model result, both indicating that while the isoforms clearly differ from the wild type, the simplistic view of additivity seems insufficient to describe the effects.

We then turned our attention to the conformational dynamics of the system to assess the substates experienced by the DBD' when acted upon by the various IVRs. Visual inspection of the trajectories with visual molecular dynamics (VMD)⁶¹ confirms our hypothesis that there is little variation in the DBD', with the most variability and motion in the N- and C-terminal regions of the isoform. The N- and C-terminal regions exhibit a high degree of dynamic motion during the simulations. This paves the way for further hypotheses regarding the N- and C-terminal regions as allosteric regulators if the N- and C-terminal regions are closer to the DBD' than we initially thought, there is more opportunity for interaction between the N- and C-terminal regions, the DBD', and other transcription factors.

The major findings once again pointed to a complex regulation of p53 active site. We then conducted an analysis of the hydrogen bonds in the system to determine whether changes in the interaction of the p53 binding interface with its cognate DNA might be affected by the IVRs.

After running our hydrogen bonding analysis on each isoform, we analyze the actual contacts of each isoform with the DNA as the acceptor compared with the expected contacts of the wild type, $\Delta 0p53\alpha$. Based on past studies, the wild type makes significant contacts with K120, S241, R248, K273, A276, C277, R280, and R283.⁴⁰ Table 2 and Figure 5 show

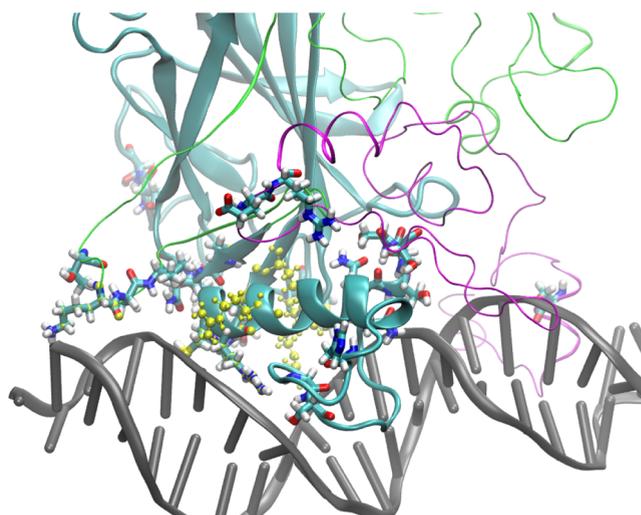


Figure 5. Hydrogen-bond contacts. Residues making H-bond contact (CPK, colored thick sticks) with DNA (gray) for more than 5% of any individual simulation are aggregated onto the p53 wild-type MD simulation structure (cartoon; N-terminal region (green), DBD (cyan), C-terminal region (cyan)). Residues making contact with the same DNA sequence in the 1TUP crystal structure are shown as a reference (yellow spheres).

that some of the isoforms make contact with these expected residues. All of the isoforms make contact with additional residues that the wild-type p53 does not significantly contact. None of the isoforms behave in the exact same way nor do they have all of the same contacts. Importantly, we also note that none of the isoforms have significant contact with more than two of the expected residue contacts.

While the trend of different H-bonds was readily apparent, the behavior of the isoforms differed widely.

To investigate how allosteric signaling may take place in p53 and its isoforms, we conducted MD-based sector analysis.⁶⁶ This method uses spectral decomposition to extract residues covarying in coordinate space over the course of the MD trajectory (Figure 6). Such residues may form a cohesive network of residues capable of allosterically transmitting information within a macromolecule. Table 3 shows the sector residues obtained for each isoform. The residues involved in the sector do indeed change as the IVR composition is varied, but consistent with our prior observations, the lack of a predictable pattern suggests nonlinear effects of the complex interactions among the residues of the system. To gain insight into which residues were most frequently involved in the sector, we summed the number of times a residue appeared in a sector list across all isoforms, and plotted these on the α carbons of the structure. Residues most often involved in sectors tend to be among the β sheets and only sometimes include residues along the recognition helix interacting with the DNA.

To examine how the IVRs affect the conformational substates of the DBD', we performed cluster analysis with concatenated trajectories from our MD-Markov state model protocol^{19,59} (Figure 7A). Because we perform clustering simultaneously for all trajectories, the cluster centroids are shared across all trajectories as opposed to being relevant to a particular construct. Thus, the population counts in each cluster may be directly compared between the constructs. We see that $\Delta 0p53\alpha$ and $\Delta 133p53\beta$ cluster-independently in

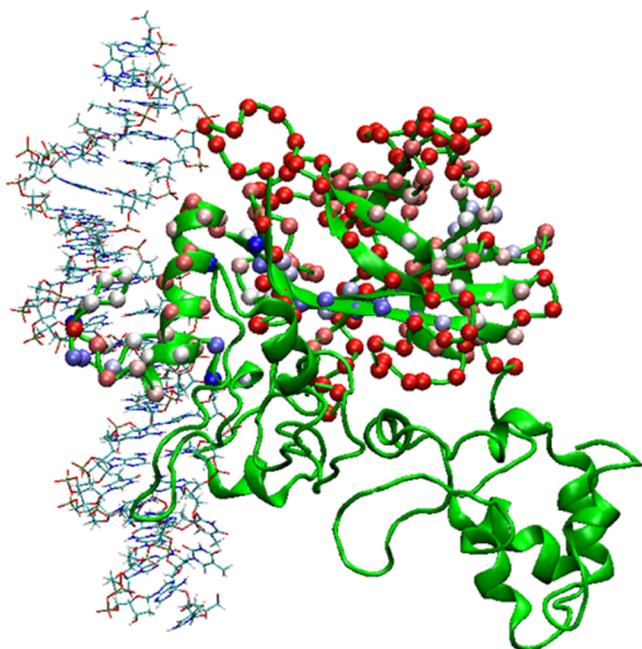


Figure 6. Sector residue frequency structural representation. The frequency of a residue's inclusion in a sector across the nine constructs is mapped to the structure. DBD' is shown in cartoon green. Sector residues are represented in VDW where colors range from red (low incidence, minimum = 0) to blue (high incidence, maximum = 6).

Cluster 3 and Cluster 4, respectively. $\Delta 0p53\beta$, $\Delta 0p53\gamma$, $\Delta 40p53\alpha$ (14.3%), $\Delta 40p53\beta$ clusters together in Cluster 1, and $\Delta 40p53\alpha$ (85.7%), $\Delta 40p53\gamma$, $\Delta 133p53\alpha$, $\Delta 133p53\gamma$ cluster in Cluster 2. $\Delta 40p53\alpha$ uniquely shares its residency in two clusters. To probe the structural differences between these centroidal structures, each centroid of DBD' from our clustering results was also further analyzed for their solvent-accessible surface area in the β sheet region (Figure 7B). We have omitted the C-terminus tail due to its intrinsically disordered nature. Cluster 3, which contains only isoform 8 substates, had the lowest solvent-accessible surface area at 8112.424 \AA^2 , while cluster 4 centroid, containing only WT full-

length p53 substates, had the highest at 8493.668 \AA^2 (Table SIV). The β sheet is largely constricted and compact in $\Delta 133p53\beta$ dynamics (gray) while in others, the β sheets remain open and solvent-accessible.

Finally, we wished to extend our analysis to include an energetic component.

By projecting each of the 100 sampled frames' electrostatic network heat kernels onto a set of R^3 eigenbasis vectors, we obtain an interpretable low-dimensional representation of how the electrostatic interactions between binding-domain residues are dynamically structured across simulation time for each IVR isoform. Moreover, we can gain a visual sense of how the organization of electrostatic interactions in the protein-binding domain differs between the isoforms. We generate plots of the PCA-embedded heat kernel across PC1 and PC2 for each of p53 isoform's 199 conserved protein-binding domain residues (Figure 8). Each plot contains $199 \times 100 = 19,900$ unique node embeddings, with each embedding corresponding to a single residue's electrostatic interactions at a single time frame. The shared KPCA electrostatic heat kernels are populated by node embeddings which are colored by each node's degree in the electrostatic network system. Nodes with higher degrees of connectivity display a more "hub-like" relation to other nodes in the graph by exhibiting a higher number of connections (i.e., significant electrostatic interactions with other residues) to other nodes/residues within the graph structure of the electrostatic network. The level of compactness in node embedding clusters within the shared PCA space reveals information about the strength of the energetic relationships between node embeddings and the residues they represent both across each frame's electrostatic interaction networks and the networks across simulation time. Notably, the topological distribution of both node embedding and connectivity values for the majority of the nine isoforms' electrostatic network heat kernels appears to be correlated with their isoform type (i.e., α , β , or γ).

DISCUSSION

Casting the isoform variable regions of p53 allosteric regulators, we have explored the effects the combination of regions confers on the p53 DBD' constant region. MD

Table 3. Sector Residues^a

isoform (DBD')	$\Delta\alpha,\beta,\gamma$	sector residues
full-length WT	p53 α	133, 162, 163, 166, 167, 169, 170, 171, 172, 173, 174, 175, 194, 195, 203, 204, 211, 212, 213, 217, 245, 246, 247, 249, 250, 251, 252, 308, 309, 310, 311, 312, 314, 315, 316, 317, 318, 319, 331
isoform 2	p53 β	133, 134, 144, 145, 146, 147, 148, 149, 177, 178, 179, 255, 256, 257, 258, 263, 264, 265, 266, 267, 268, 269, 270, 271, 280, 282, 297, 310, 311, 312, 313, 315, 316, 317, 318, 319, 320, 321, 322
isoform 3	p53 γ	141, 163, 164, 197, 203, 204, 205, 234, 235, 247, 248, 249, 250, 251, 274, 299, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 311, 312, 321, 322, 323, 324, 325, 326, 327, 328, 329, 330, 331
isoform 4	p53 $\Delta 40\alpha$	133, 134, 146, 147, 148, 151, 154, 155, 156, 157, 204, 205, 218, 219, 252, 253, 254, 255, 256, 257, 266, 267, 268, 269, 270, 271, 272, 273, 282, 285, 299, 300, 301, 302, 303, 304, 305, 306, 311
isoform 5	p53 $\Delta 40\beta$	133, 249, 250, 251, 252, 253, 254, 255, 256, 259, 263, 264, 265, 267, 268, 269, 270, 271, 272, 273, 274, 282, 303, 304, 305, 306, 307, 312, 313, 314, 315, 316, 317, 327, 328, 329, 330, 331
isoform 6	p53 $\Delta 40\gamma$	134, 135, 136, 142, 192, 193, 208, 209, 210, 211, 212, 233, 235, 236, 273, 274, 275, 278, 279, 280, 281, 282, 283, 284, 290, 291, 292, 309, 310, 312, 313, 314, 315, 316, 317, 327, 328, 329, 330
isoform 7	p53 $\Delta 133\alpha$	133, 134, 145, 146, 147, 155, 156, 213, 214, 219, 227, 228, 229, 238, 240, 242, 243, 244, 245, 246, 247, 248, 249, 250, 251, 254, 266, 267, 268, 269, 270, 271, 272, 273, 322, 323, 324, 325, 326
isoform 8	p53 $\Delta 133\beta$	252, 253, 254, 277, 278, 279, 282, 285, 286, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 299, 300, 302, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 315, 316, 317, 330
isoform 9	p53 $\Delta 133\gamma$	133, 134, 144, 151, 155, 156, 157, 158, 159, 204, 205, 216, 217, 218, 219, 220, 229, 230, 231, 252, 253, 254, 255, 256, 257, 258, 259, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273, 282, 304

^aThe sector residues for each isoform were calculated using the DBD' region.

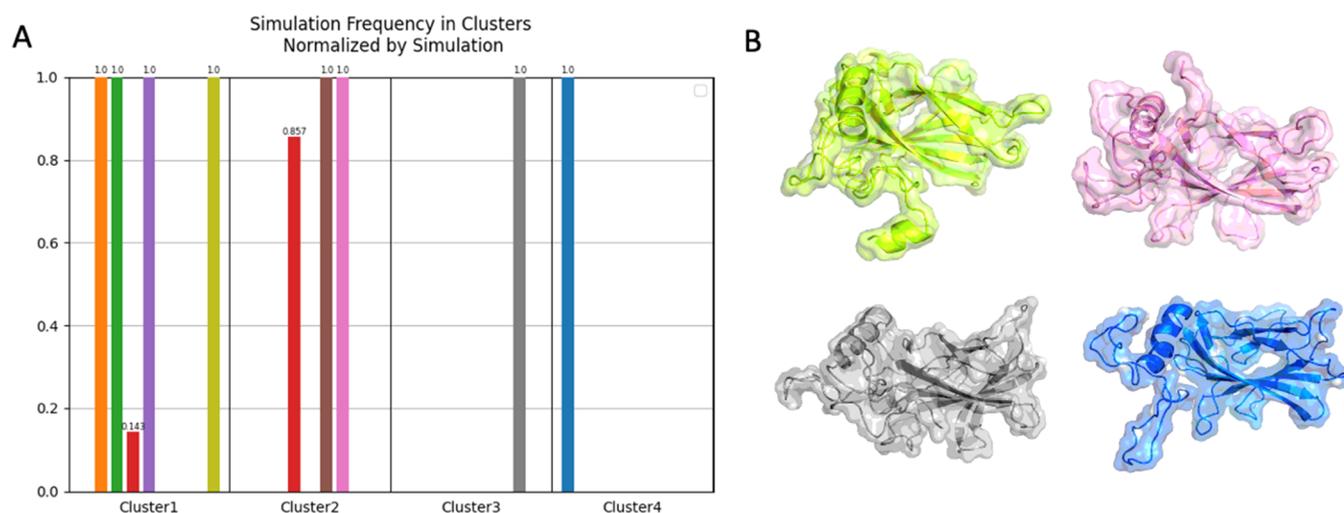


Figure 7. *K*-means ($n = 4$) clustering and structural representation. (A) Simulation frequency in clusters normalized by simulation with $n = 4$. Cluster 1 contains $\Delta 0p53\beta$, $\Delta 0p53\gamma$, $\Delta 40p53\alpha$, $\Delta 40p53\beta$, and $\Delta 133p53\gamma$. Cluster 2 contains $\Delta 40p53\alpha$, $\Delta 40p53\gamma$, and $\Delta 133p53\alpha$. Cluster 3 contains $\Delta 133p53\beta$. Cluster 4 contains $\Delta 0p53\alpha$. (B) Centroid structural representations for each cluster. Lime green is cluster 1, pink is cluster 2, gray is cluster 3, and blue is cluster 4.

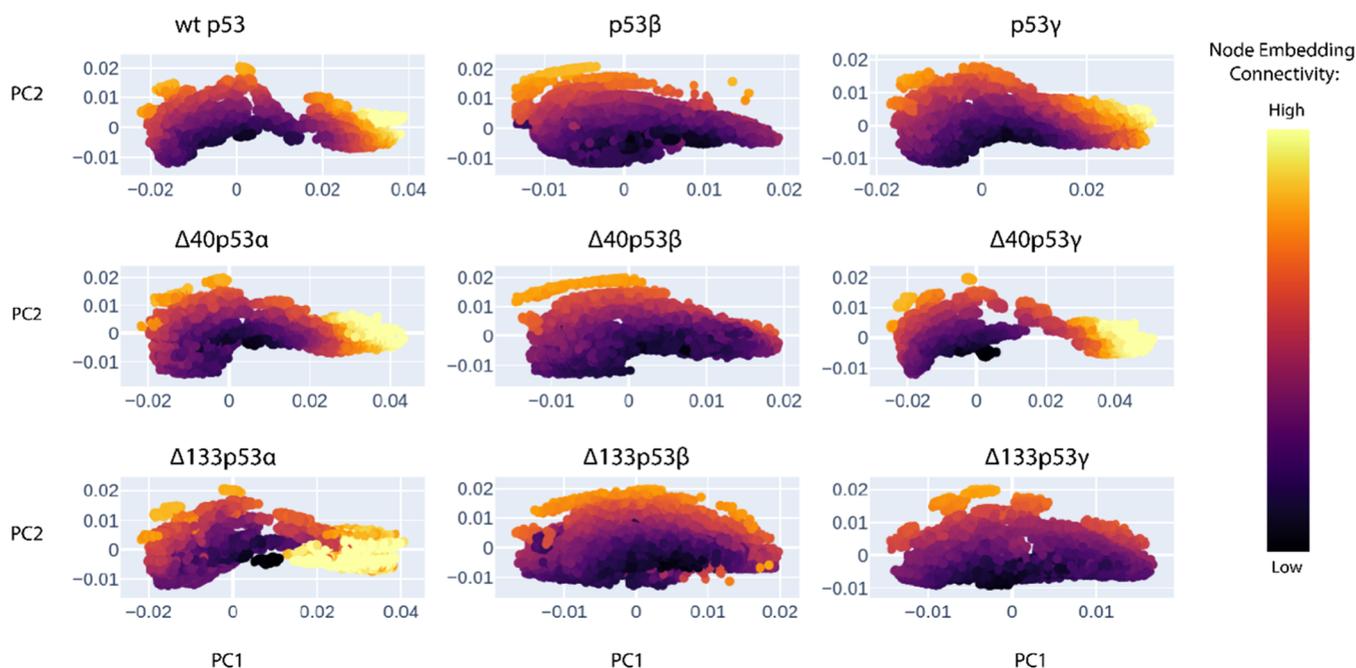


Figure 8. Electrostatic heat kernels. Electrostatic heat kernel PCA projections are shown for the 199 DBD' residues of each p53 IVR isoform on axes PC1 and PC2 (R_2 space). In total there are $199 \times 100 = 19,900$ embedded points for each plot. Color mapping represents the degree of node connectivity in the protein system.

simulations of the major nine naturally occurring isoforms including wild type provide molecular-level insight into the process. Our simulations characterize the structure and dynamics of the full-length p53 protein. This captures the β barrel structure of p53 well as highly mobile N- and C-terminal regions, consistent with well-established Levant characterization, now with time resolution at the atomic level. While the N- and C-terminal regions exhibit high fluctuations, they are not devoid of secondary structure. Rather, we see that they are connected by highly flexible loop regions. Furthermore, we do observe some extent of these regions folding back against the DNA and DBD' introducing an extension of the binding site.

Equipped with these simulations in hand, we endeavored to first consider whether any measurable difference between the isoform variants and wild-type p53 isoform could be detected. An extensive literature purports the idea the Post Translational Modifications in the IVR regions control p53 DNA sequence-specific affinity to modulate the target sequence in accordance with the current condition of the cell. A more fundamental question, and the central one driving this study, is whether the IVR sequence itself modulates the binding properties. On all counts measured, we observed an unequivocal response in the affirmative. The RMSF identified residues of high level of fluctuation, the MD-MSMs indicated unique clustering of snapshots from the wild type distinct from two conformational

substates of the isoforms. Hydrogen-bond analysis indicated only a single common DNA contact residue between the wild type and the other isoforms. We thus observed completely new modes of p53–DNA interactions. Similarly, the MD sector analysis revealed a shift to new network residues in the isoforms, and the AMBER energy analysis suggested the same. Therefore, MD simulations of p53 isoforms indicate that the different IVRs of p53 themselves intrinsically possess the ability to modulate the p53 binding activity.

We then turn our attention to the task of understanding how the presence or absence of the IVRs affects the DBD'. In the design phase of the project, we had been encouraged by the seemingly systematic way in which the splice variants relate to each other, leading us to hypothesize expected outcomes on the basis of linearly independent additivity as a convenient way to ascribe functionality to each region. However, our data repeatedly induced nonlinear effects that predominantly govern the behavior of p53 isoforms. The first indication was observed in our attempt to create a linear regression model on the RMSD data. The resulting equation and coefficients indicate linearity at about 80%. This is well below the threshold of statistical significance. In the practical level, we interpret general linearity exists, but this is largely overshadowed by nonlinear effects. These may occur due to the complex overlay of interactions between protein residues. Separating the behavior of the wild type from the other isoforms was readily achievable, but discerning the pattern of systematic linear regulation by IVR was not apparent by the metrics applied. The MD-MSMs characterized two conformation substates aside from the wild type. However, the partitioning between those two groups further indicated nonlinearity. Following a logic line that some residues may be highly fluctuating but not contributing to the core dynamics, we additionally considered construction of MD-MSMs on subsets of residues identified as important sector residues, and residues involved in H-bond recognition of the DNA. These continued to emphasize the importance of nonlinear effects.

The MD-based sector residues also provided additional support to the presence of nonlinear effects. The sectors of the isoforms were distinct from the wild type but were also unique for each combination. The same trend furthermore was borne out in the AMBER energy analysis.

This diversification of p53 activity repeatedly observed in our results suggests consideration when dealing with cancers in different tissues. Clinical reports regularly separate discussion of the manifestation of the cancers and their treatments on the basis of the tissue that is involved. Perhaps the underlying reason for the diversity observed in tumors is caused at least in part by unique underlying populations of isoforms that our work suggests may be expected to behave differently.

According to the model proposed by Jorruiz and Bourdon,⁷² cell response to the same damage can be altered by manipulating the expression of only a subset of p53 isoforms using p53 isoform-specific small interfering RNA (siRNA). Experiments have found that after depletion of $\Delta 0p53\beta$ and $\Delta 0p53\gamma$, and using the siRNA siSplice, the cell response to the same damage is to induce apoptosis, while after the depletion of $\Delta 133p53\alpha$ and $\Delta 133p53\beta$ using the siRNA si133, the cell response to the same damage results in senescence (Figure S6). We see similar trends in the results of our Markov state model-based k-means clustering. $\Delta 133p53\beta$, which is depleted when cell response to damage undergoes both senescence and

apoptosis, distinctively resides in its own cluster (Cluster 3, Figure 7A). $\Delta 40p53\alpha$ and the full-length wild type are coactivated and rather stay untouched when undergoing both senescence and apoptosis. While the wild type resides in its own cluster (Cluster 4, Figure 7A), $\Delta 40p53\alpha$ shares most of its cluster residency with $\Delta 133p53\alpha$ in Cluster 2 (85.7% in Cluster 2, and 14.3% in cluster 1, Figure 7A). Interestingly, isoforms $\Delta 40p53\alpha$ and $\Delta 133p53\alpha$ are both preserved with the WT cells undergoing apoptosis. $\Delta 0p53\alpha$ and $\Delta 0p53\gamma$ are uniquely preserved in senescent cells that also cluster exclusively in Cluster 1. $\Delta 40p53\alpha$, which is the only isoform exclusively preserved along with the full-length wild-type p53 when undergoing both senescence and apoptosis, shares its cluster residency with Clusters 1 and 2, suggesting that its expression is versatile and necessary for moderating normal cell function in parallel with the full-length wt. When extending the number of clusters, we see the full-length wild-type $\Delta 133p53\beta$ and largely $\Delta 133p53\gamma$ residing in their own clusters. All the while, $\Delta 40p53\alpha$ fluctuates its residency in between the two clusters that we hypothesize to resemble both apoptosis and senescence-inducing p53 dynamics.

While many of the sectors on all of our systems co-localized also in the same β sheet region where much of the global structural differences were observed in our clustering, $\Delta 133p53\beta$ is the only isoform with sector loci concerted in the C-terminal region. Our results suggest that its dynamics serve a specific purpose for its elusive activity when cells undergo both senescence and apoptosis, but their exact role for their differentiation in dynamics and structure remains largely unstudied.

Furthermore, the diversified dynamics of the isoforms bears particular importance on the development of allosteric modulators of p53. To date, most studies consider either the full 393 amino acid protein or the DBD. In light of our observations, we might expect a range of behaviors within the ensemble of p53 isoforms in each tissue type. One may expect that mutations may not affect isoforms in the same way, and similarly, we may expect all allosteric effectors and drugs to interact in the same manner. While the combinatoric space this opens appears prohibitive, computational approaches make broaching such studies feasible to at least gain critical initial insight to ascertain why it is important and what may be omitted with minimal detriment. These details of course remain largely to be worked out in the next phase of experimentation.

Machine learning (ML)⁷³ is a methodology well suited for pattern recognition and extraction of information easily overlooked by humans. Molecular-based inquiry such as that we have carried out to some extent relies on a guess-and-check approach and is therefore limited by the imagination of the experimenter generating the hypotheses. ML on the other hand does not require an explicit hypothesis to check, and all available data can be considered as input. The model, of course, is limited by the availability of the data but has the advantage of being able to identify what is important, from which one can work backward to ascribe an appropriate hypothesis. In our drug design lines of inquiry, we have been engaged with generative adversarial neural networks (GANNs)⁷⁴ to create drugs to restore native function to p53 dynamics. Variants of these, or other deep neural networks, may prove useful not only for understanding the dynamics of p53 and its isoforms but also for uncovering the mechanism of

allosteric signal transmitting an exclusive effect arising from nonadditive collective motions of residues in a protein.

CONCLUSIONS

In conclusion, our study provides comprehensive molecular details of the time-resolved conformational dynamics of p53 including eight alternate splice variants involving 2 N-terminal and 2 C-terminal isoform variable regions. Our observations through analysis of our trajectories capture a substantive difference between the wild-type variant and all other isoforms. However, the way in which the IVRs confer alternate dynamics and DNA interaction likely to ultimately alter regulation properties are largely nonlinear. The properties governing each isoform's behavior are hard-coded into the sequence of the isoforms particularly in the IVRs, resulting in a complex milieu of interaction of p53 DBD' with its cognate DNA target. This raises important considerations for drug design to rescue mutant p53, suggesting that thinking of p53 DBD' activity as an ensemble of isoforms appropriate to the issue of interest may be important to make headway for the *de novo* design of allosteric reactivators of mutant p53. Given the ubiquitous nature of splicing in eukaryotes, these principles are likely to be broadly applicable in other protein systems, especially where allosteric regulation is called for. While our study perhaps raises more questions than it solves, it does provide a perspective that may finally advance the purposeful design of therapeutics from chemical principles to provide customized medicine to restore the body's natural defense by reactivating p53 to control tumors. While challenges of the complex nature of p53, and perhaps all proteins, raise, we are also in an age where machine learning is equipped to take on these challenges with major advances in computing power by graphical processing units (GPUs). This dataset will be important for such trials and light the path, and the next chapter in allosteric drug discovery.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jpcb.2c06229>.

Energy minimization graph for each of the isoforms (Figure S1); optimization of number of centroids (Figure S2); linear regression using average RMSF of the DBD (Figure S3); alignment of Phyre-folded structures before simulation (Figure S4); alignment of all of the isoforms of p53 after simulation (Figure S5); isoform partitioning resulting from cell damage (Figure S6); tissue-specific isoform expression (Table SI); multiple sequence alignment of the isoforms of p53 (Table SII); results from "Intensive Mode" PHYRE2 folding (Table SIII), and calculation of solvent-accessible surface area (SASA) for MD-MSM determined centroids (Table SIV) (PDF)

AUTHOR INFORMATION

Corresponding Author

Kelly M. Thayer – Department of Mathematics and Computer Science, Department of Chemistry, and College of Integrative Sciences, Wesleyan University, Middletown, Connecticut 06457, United States; orcid.org/0000-0001-7437-9517; Email: kthayer@wesleyan.edu

Authors

Isabel Armour-Garb – Department of Mathematics and Computer Science, Wesleyan University, Middletown, Connecticut 06457, United States

In Sub Mark Han – Department of Chemistry, Wesleyan University, Middletown, Connecticut 06457, United States; orcid.org/0000-0003-2791-5975

Benjamin S. Cowan – College of Integrative Sciences, Wesleyan University, Middletown, Connecticut 06457, United States

Complete contact information is available at: <https://pubs.acs.org/10.1021/acs.jpcb.2c06229>

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors thank the Molecules to Medicine consortium, especially David L. Beveridge and Michael P. Weir for fruitful discussion. They also thank James Lipton and Dan Licata for reading early versions of the manuscript. This work was supported by the NIH grant #R15 GM128102-02 to K.M.T. This work was also supported by NSF grants CNS-0619508 and CNS-095985 to Wesleyan University for high-performance computing facilities.

REFERENCES

- Guarnera, E.; Berezovsky, I. N. Allosteric Sites: Remote Control in Regulation of Protein Activity. *Curr. Opin. Struct. Biol.* **2016**, *37*, 1–8.
- McLeish, T. C. B.; Rodgers, T. L.; Wilson, M. R. Allostery without Conformation Change: Modelling Protein Dynamics at Multiple Scales. *Phys. Biol.* **2013**, *10*, No. 056004.
- Monod, J.; Wyman, J.; Changeux, J. P. On the Nature of Allosteric Transitions: A Plausible Model. *J. Mol. Biol.* **1965**, *12*, 88–118.
- Koshland, D. E.; Nemethy, J. G.; Filmer, D. Comparison of Experimental Binding Data and Theoretical Models in Proteins Containing Subunits. *Biochemistry* **1966**, *5*, 365–385.
- Lewis, M. Allostery and the Lac Operon. *J. Mol. Biol.* **2013**, *425*, 2309–2316.
- Sharp, K. A. Allostery in the Lac Operon: Population Selection or Induced Dissociation? *Biophys. Chem.* **2011**, *159*, 66–72.
- Garcia-Pino, A.; Balasubramanian, S.; Wyns, L.; Gazit, E.; de Greve, H.; Magnuson, R. D.; Charlier, D.; van Nuland, N. A. J.; Loris, R. Allostery and Intrinsic Disorder Mediate Transcription Regulation by Conditional Cooperativity. *Cell* **2010**, *142*, 101–111.
- Lakhani, B.; Thayer, K. M.; Hingorani, M. M.; Beveridge, D. L. Evolutionary Covariance Combined with Molecular Dynamics Predicts a Framework for Allostery in the MutS DNA Mismatch Repair Protein. *J. Phys. Chem. B* **2017**, *121*, 2049–2061.
- Gupta, S.; Gellert, M.; Yang, W. Mechanism of Mismatch Recognition Revealed by Human MutS β Bound to Unpaired DNA Loops. *Nat. Struct. Mol. Biol.* **2012**, *19*, 72–79.
- Gentry, P. R.; Sexton, P. M.; Christopoulos, A. Novel Allosteric Modulators of G Protein-Coupled Receptors. *J. Biol. Chem.* **2015**, *290*, 19478–19488.
- Feng, Z.; Hu, G.; Ma, S.; Xie, X. Q. Computational Advances for the Development of Allosteric Modulators and Bitopic Ligands in G Protein-Coupled Receptors. *AAPS J.* **2015**, *17*, 1080–1095.
- Shi, Z.; Resing, K. A.; Ahn, N. G. Networks for the Allosteric Control of Protein Kinases. *Curr. Opin. Struct. Biol.* **2006**, *16*, 686–692.
- Nussinov, R.; Zhang, M.; Liu, Y.; Jang, H. AlphaFold, Artificial Intelligence (AI), and Allostery. *J. Phys. Chem. B* **2022**, *126*, 6372–6383.

- (14) Huang, Z.; Mou, L.; Shen, Q.; Lu, S.; Li, C.; Liu, X.; Wang, G.; Li, S.; Geng, L.; Liu, Y.; Wu, J.; Chen, G.; Zhang, J. ASD v2.0: Updated Content and Novel Features Focusing on Allosteric Regulation. *Nucleic Acids Res.* **2014**, *42*, D510–D516.
- (15) Nussinov, R.; Zhang, M.; Maloney, R.; Liu, Y.; Tsai, C. J.; Jang, H. Allosteric Cancer Drivers and Innovative Allosteric Drugs. *J. Mol. Biol.* **2022**, *434*, No. 167569.
- (16) Wagner, J. R.; Lee, C. T.; Durrant, J. D.; Malmstrom, R. D.; Feher, V. A.; Amaro, R. E. Emerging Computational Methods for the Rational Discovery of Allosteric Drugs. *Chem. Rev.* **2016**, *116*, 6370–6390.
- (17) Abdel-Magid, A. F. Allosteric Modulators: An Emerging Concept in Drug Discovery. *ACS Med. Chem. Lett.* **2015**, *6*, 104–107.
- (18) Amamuddy, O. S.; Veldman, W.; Manyumwa, C.; Khairallah, A.; Agajanian, S.; Oluyemi, O.; Verkhivker, G. M.; Bishop, Ö. T. Integrated Computational Approaches and Tools for Allosteric Drug Discovery. *Int. J. Mol. Sci.* **2020**, *21*, No. 847.
- (19) Thayer, K. M.; Lakhani, B.; Beveridge, D. L. Molecular Dynamics-Markov State Model of Protein Ligand Binding and Allosteric in CRIB-PDZ: Conformational Selection and Induced Fit. *J. Phys. Chem. B* **2017**, *121*, 5509–5514.
- (20) Hertig, S.; Latorraca, N. R.; Dror, R. O. Revealing Atomic-Level Mechanisms of Protein Allosteric with Molecular Dynamics Simulations. *PLoS Comput. Biol.* **2016**, *12*, No. e1004746.
- (21) Nussinov, R.; Tsai, C. J.; Jang, H. Allosteric, and How to Define and Measure Signal Transduction. *Biophys. Chem.* **2022**, *283*, No. 106766.
- (22) Petit, C. M.; Zhang, J.; Sapienza, P. J.; Fuentes, E. J.; Lee, A. L. Hidden Dynamic Allosteric in a PDZ Domain. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 18249–18254.
- (23) Lockless, S. W.; Ranganathan, R. Evolutionarily Conserved Pathways of Energetic Connectivity in Protein Families. *Science* **1999**, *286*, 295–299.
- (24) del Sol, A.; Tsai, C. J.; Ma, B.; Nussinov, R. The Origin of Allosteric Functional Modulation: Multiple Pre-Existing Pathways. *Structure* **2009**, *17*, 1042–1050.
- (25) Dixit, S. B.; Andrews, D. Q.; Beveridge, D. L. Induced Fit and the Entropy of Structural Adaptation in the Complexation of CAP and λ -Repressor with Cognate DNA Sequences. *Biophys. J.* **2005**, *88*, 3147–3157.
- (26) Reynolds, K. A.; McLaughlin, R. N.; Ranganathan, R. Hot Spots for Allosteric Regulation on Protein Surfaces. *Cell* **2011**, *147*, 1564–1575.
- (27) Halabi, N.; Rivoire, O.; Leibler, S.; Ranganathan, R. Protein Sectors: Evolutionary Units of Three-Dimensional Structure. *Cell* **2009**, *138*, 774–786.
- (28) Rivoire, O.; Reynolds, K. A.; Ranganathan, R. Evolution-Based Functional Decomposition of Proteins. *PLoS Comput. Biol.* **2016**, *12*, No. e1004817.
- (29) Süel, G. M.; Lockless, S. W.; Wall, M. A.; Ranganathan, R. Evolutionarily Conserved Networks of Residues Mediate Allosteric Communication in Proteins. *Nat. Struct. Biol.* **2003**, *10*, 59–69.
- (30) Zhuravlev, P. I.; Papoian, G. A. Protein Functional Landscapes, Dynamics, Allosteric: A Tortuous Path towards a Universal Theoretical Framework. *Q. Rev. Biophys.* **2010**, *43*, 295–332.
- (31) McLeish, T. C. B.; Cann, M. J.; Rodgers, T. L. Dynamic Transmission of Protein Allosteric without Structural Change: Spatial Pathways or Global Modes? *Biophys. J.* **2015**, *109*, 1240–1250.
- (32) Cooper, A.; Dryden, D. T. F. Allosteric without Conformational Change - A Plausible Model. *Eur. Biophys. J.* **1984**, *11*, 103–109.
- (33) Vijayabaskar, M. S.; Vishveshwara, S. Interaction Energy Based Protein Structure Networks. *Biophys. J.* **2010**, *99*, 3704–3715.
- (34) Noel, J. K.; Whitford, P. C.; Onuchic, J. N. The Shadow Map: A General Contact Definition for Capturing the Dynamics of Biomolecular Folding and Function. *J. Phys. Chem. B* **2012**, *116*, 8692–8702.
- (35) McLeish, T. C. B.; Cann, M. J.; Rodgers, T. L. Dynamic Transmission of Protein Allosteric without Structural Change: Spatial Pathways or Global Modes? *Biophys. J.* **2015**, *109*, 1240–1250.
- (36) Rafferty, J. B.; Somers, W. S.; Saint-Girons, I.; Phillips, S. E. V. Three-Dimensional Crystal Structures of Escherichia Coli Met Repressor with and without Corepressor. *Nature* **1989**, *341*, 705–710.
- (37) Popovych, N.; Tzeng, S. R.; Tonelli, M.; Ebricht, R. H.; Kalodimos, C. G. Structural Basis for CAMP-Mediated Allosteric Control of the Catabolite Activator Protein. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 6927–6932.
- (38) Lakhani, B.; Thayer, K. M.; Hingorani, M. M.; Beveridge, D. L. Evolutionary Covariance Combined with Molecular Dynamics Predicts a Framework for Allosteric in the MutS DNA Mismatch Repair Protein. *J. Phys. Chem. B* **2017**, *121*, 2049–2061.
- (39) Thayer, K. M.; Galganov, J. C.; Stein, A. J. Dependence of Prevalence of Contiguous Pathways in Proteins on Structural Complexity. *PLoS One* **2017**, *12*, No. e0188616.
- (40) Cho, Y.; Gorina, S.; Jeffrey, P. D.; Pavletich, N. P. Crystal Structure of a P53 Tumor Suppressor-DNA Complex: Understanding Tumorigenic Mutations. *Science* **1994**, *265*, 346–355.
- (41) Joerger, A. C.; Fersht, A. R. Structural Biology of the Tumor Suppressor P53. *Annu. Rev. Biochem.* **2008**, *77*, 557–582.
- (42) Pagel, K.; Natan, E.; Hall, Z.; Fersht, A. R.; Robinson, C. V. Intrinsically Disordered P53 and Its Complexes Populate Compact Conformations in the Gas Phase. *Angew. Chem., Int. Ed.* **2013**, *52*, 361–365.
- (43) Lum, J. K.; Neuweiler, H.; Fersht, A. R. Long-Range Modulation of Chain Motions within the Intrinsically Disordered Transactivation Domain of Tumor Suppressor P53. *J. Am. Chem. Soc.* **2012**, *134*, 1617–1622.
- (44) Dunker, A. K.; Lawson, J. D.; Brown, C. J.; Williams, R. M.; Romero, P.; Oh, J. S.; Oldfield, C. J.; Campen, A. M.; Ratliff, C. M.; Hipps, et al. Intrinsically Disordered Protein. *J. Mol. Graphics Modell.* **2001**, *19*, 26–59.
- (45) Gu, B.; Zhu, W. G. Surf the Post-Translational Modification Network of P53 Regulation. *Int. J. Biol. Sci.* **2012**, *8*, 672.
- (46) Feng, L.; Hollstein, M.; Xu, Y. Ser46 Phosphorylation Regulates P53-Dependent Apoptosis and Replicative Senescence. *Cell Cycle* **2006**, *5*, 2812–2819.
- (47) Tsang, W. P.; Ho, F. Y. F.; Fung, K. P.; Kong, S. K.; Kwok, T. T. P53-R175H Mutant Gains New Function in Regulation of Doxorubicin-Induced Apoptosis. *Int. J. Cancer* **2005**, *114*, 331–336.
- (48) Lamond, A. I. The Spliceosome. *BioEssays* **1993**, *15*, 595–603.
- (49) Fei, P.; Bernhard, E. J.; El-Deiry, W. S. Tissue-Specific Induction of P53 Targets in Vivo. *Cancer Res.* **2002**, *62*, 7316–7327.
- (50) Bouvard, V.; Zaitchouk, T.; Vacher, M.; Duthu, A.; Canivet, M.; Choisy-Rossi, C.; Nieruchalski, M.; May, E. Tissue and Cell-Specific Expression of the P53-Target Genes: Bax, Fas, Mdm2 and Waf1/P21, before and Following Ionising Irradiation in Mice. *Oncogene* **2000**, *19*, 649–660.
- (51) Natan, E.; Baloglu, C.; Pagel, K.; Freund, S. M. V.; Morgner, N.; Robinson, C. V.; Fersht, A. R.; Joerger, A. C. Interaction of the P53 DNA-Binding Domain with Its N-Terminal Extension Modulates the Stability of the P53 Tetramer. *J. Mol. Biol.* **2011**, *409*, 358–368.
- (52) Kitayner, M.; Rozenberg, H.; Kessler, N.; Rabinovich, D.; Shaulov, L.; Haran, T. E.; Shakked, Z. Structural Basis of DNA Recognition by P53 Tetramers. *Mol. Cell* **2006**, *22*, 741–753.
- (53) Bateman, A.; Martin, M. J.; O'Donovan, C.; Magrane, M.; Apweiler, R.; Alpi, E.; Antunes, R.; Arganiska, J.; Bely, B.; Bingley, M.; et al. UniProt: A Hub for Protein Information. *Nucleic Acids Res.* **2015**, *43*, D204–D212.
- (54) Götz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *J. Chem. Theory Comput.* **2012**, *8*, 1542–1555.
- (55) Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* **2013**, *9*, 3878–3888.
- (56) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J.

The Amber Biomolecular Simulation Programs. *J. Comput. Chem.* **2005**, *26*, 1668–1688.

(57) Salomon-Ferrer, R.; Case, D. A.; Walker, R. C. An Overview of the Amber Biomolecular Simulation Package. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2013**, *3*, 198–210.

(58) Yu, Z.; Li, P.; Merz, K. M. Extended Zinc AMBER Force Field (EZAFF). *J. Chem. Theory Comput.* **2018**, *14*, 242–254.

(59) Han, I. S. M.; Abramson, D.; Thayer, K. M. Insights into Rational Design of a New Class of Allosteric Effectors with Molecular Dynamics Markov State Models and Network Theory. *ACS Omega* **2022**, *7*, 2831–2841.

(60) Roe, D. R.; Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9*, 3084–3095.

(61) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33–38.

(62) Hsin, J.; Arkhipov, A.; Yin, Y.; Stone, J. E.; Schulten, K. Using VMD: An Introductory Tutorial. *Curr. Protoc. Bioinf.* **2008**, *24*, 1–48.

(63) *The PyMOL Molecular Graphics System*, version 2.0; Schrödinger, LLC.

(64) Thayer, K. M.; Lakhani, B.; Beveridge, D. L. Molecular Dynamics-Markov State Model of Protein Ligand Binding and Allostery in CRIB-PDZ: Conformational Selection and Induced Fit. *J. Phys. Chem. B* **2017**, *121*, 5509–5514.

(65) Lockless, S. W.; Ranganathan, R. Evolutionarily Conserved Pathways of Energetic Connectivity in Protein Families. *Science* **1999**, *286*, 295–299.

(66) Lakhani, B.; Thayer, K. M.; Black, E.; Beveridge, D. L. Spectral Analysis of Molecular Dynamics Simulations on PDZ: MD Sectors. *J. Biomol. Struct. Dyn.* **2019**, *38*, 781–790.

(67) Ma, C.; Chung, D. J.; Abramson, D.; Langley, D. R.; Thayer, K. M. Mutagenic Activation of Glutathione Peroxidase-4: Approaches toward Rational Design of Allosteric Drugs. *ACS Omega* **2022**, *7*, 29587–29597.

(68) Abramson, D. The Colors of a Protein: Protein Dynamics through the Lens of Spectral Graph Theory. Undergraduate Thesis, Wesleyan University: Middletown, CT, 2001.

(69) Bastian, M.; Heymann, S.; Jacomy, M. In *Gephi: An Open Source Software for Exploring and Manipulating Networks*, Proceedings of the International AAAI Conference on Web and Social Media, 2009; pp 361–362.

(70) Elghawalby, H.; Hancock, E. R. Heat Kernel Embeddings, Differential Geometry and Graph Structure. *Axioms* **2015**, *4*, 275–293.

(71) Lewith, G. T.; Jonas, W. B.; Walach, H. *Clinical Research in Complementary Therapies: Principles, Problems and Solutions*; Elsevier Health Sciences, 2010.

(72) Jorruiz, S. M.; Bourdon, J. C. P53 Isoforms: Key Regulators of the Cell Fate Decision. *Cold Spring Harbor Perspect. Med.* **2016**, *6*, No. a026039.

(73) Badillo, S.; Banfai, B.; Birzele, F.; Davydov, I. I.; Hutchinson, L.; Kam-Thong, T.; Siebourg-Polster, J.; Steiert, B.; Zhang, J. D. An Introduction to Machine Learning. *Clin. Pharmacol. Ther.* **2020**, *107*, 871–885.

(74) Aggarwal, A.; Mittal, M.; Battineni, G. Generative Adversarial Network: An Overview of Theory and Applications. *Int. J. Inf. Manage. Data Insights* **2021**, *1*, No. 100004.