ABSTRACT: The development of drugs to restore protein function has been a major advance facilitated by molecular medicine. Allosteric regulation, a phenomenon widely observed in nature, in which a molecule binds to control a distance active site, holds great promise for regulating proteins, yet how to rationally design such a molecule remains a mystery. Over the past few years, we and others have developed several techniques based on molecular dynamics (MD) simulations: MD-Markov state models to capture global conformational substates, and network theory approach utilizing the interaction energy within the protein to confer local allosteric control. We focus on the key case study of the p53 Y220C mutation restoration by PK11000, a compound experimentally shown to reactivate p53 native function in Y220C mutant present tumors. We gain insights into the mutation and allosteric reactivation of the protein, which we anticipate will be applicable to de novo design to engineer new compounds not only for this mutation, but in other macromolecular systems as well.

INTRODUCTION

Network theory when equipped with data from all-atom molecular dynamics (MD) simulations can be a powerful tool to illustrate the underpinnings of how protein function arises from its structure. We are particularly interested in developing computational pipelines using MD simulations to quantify allosteric regulation and signaling as competent assays for rational drug design. Only a few have studied MD and machine learning as conduits for the rational design of allosteric reactivators despite their versatility in predicting ligand–protein interactions. MD time-series data when paired with graph theory and network-centric approaches have shown considerable promise in revealing solvent-accessible allosteric sites.1 Graph embeddings have a wide range of utilities in machine learning including node classification, edge prediction, graph clustering, and network visualization.2

Allosteric regulation may include any perturbation, such as a point mutation or effector-binding to a macromolecule, causing a change at a distal site by long-range signaling.3,4 Tumor suppressor p53 is an ideal system to study this phenomenon—p53 ubiquitously spans the evolutionary tree as vital mediators of cell-cycle homeostasis in mammals, and its activity is largely modulated by allosteric factors such as missense mutations, post-translational modifications, and antibody recognition.5 Disruption in normal p53 function can lead to fatal consequences, where mutations in p53 occur in over 50% of all human cancers, with its vast majority occurring in the DNA-binding domain with high rates of therapeutic resistance.6 Among these mutations, the hotspot Y220C mutation is the ninth most frequent p53 missense mutant accounting for over 100,000 new breast and ovarian cancer cases per year worldwide.

Recent reports of the only known allosteric effector, PK1100, that rescues p53 Y220c mutant present a fitting example of an allosteric modulation at work via mutational disruption and allosteric rescue.7 Despite its therapeutic potential, PK1100 has not been able to pass clinical trials, which is common among many of the hits discovered through high-throughput screening. This further stresses the emergent need for a comprehensive theory on allosteric signaling to better recapitulate protein function. In this study, we used time-series data from MD simulations coupled with ensemble-based statistical mechanics approach and network-based machine learning to uncover the allosteric signaling pathways of tumor suppressor p53. While the earliest models of allostery are based on thermally averaged, “static” crystallographic structures, classical mechanics allows us to compute the deterministic motions of each atom in a protein as the protein traverses its free energy landscape. As such, MD simulations offer a wealth of information on the time evolution of protein...
structures and the energetic interactions of their amino acid residues, making it a critical resource for capturing allosteric signaling.

To date, many machine learning algorithms have proven to be quite resourceful in tackling the technical challenges of MD sampling while highlighting functionally important amino acid residues involved in allosteric interactions.\textsuperscript{4,8–10} Recently, Markov state models (MSMs) have provided a novel vantage point on the ensemble-based analysis of simulated proteins. In an MSM, the data reduction problem is dealt with in terms of clusters which differ in collective conformational changes, rather than individual structural parameters.\textsuperscript{11} By treating systems at equilibrium as probabilistic functions, such as clusters, MSMs link time-resolved microstates to macroscopic thermodynamic substates. MSMs, thereby can provide a global probabilistic view of the allosterically steered conformational selection. Integrating network-based machine learning approaches is another highly strategic way of processing the immense amount of manifold data from MD simulations. As first described by Raganthan and colleagues in 1999,\textsuperscript{12} statistical coupling analysis (SCA) is a method rooted in spectral graph theory which classifies a sparse subgroup of

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure1.png}
\caption{(A) RMSD of the backbone atoms of all atom 1 \(\mu\)s MD simulations. WT p53 in blue, Y220C mutant in red, and PK11000 rescued Y220C mutant in green. (B) Root mean square fluctuations in Angstroms squared of the backbone atoms of all atom 1 \(\mu\)s MD simulations. WT in blue, Y220C mutant in red, and PK11000 rescued Y220C mutant in green. Antibody binding regions are labeled in color dots; PAb1620 (WT specific) in sky blue, PAb246 (WT specific) in light orange, and PAb240 (mutant specific) in purple.}
\end{figure}
Amino acid residues that form cooperative networks known as “sectors.” SCA identifies functional residues from conservation-weighted evolutionary covariance matrices obtained from multiple sequence alignments of homologous families of proteins. The spectral method has been subsequently adapted by our lab to identify allosteric networks in MD-calculated covariance matrices based on two different modes: fluctuations of residue motional displacements, named MD-Sectors and our latest approach using nonbonded interaction energies and named MD-END (energetic network decomposition). While the MD-MSM approach introduces a methodology that can be used to assess the broad landscape of allosteric conformational change, MD-Sectors and the newly introduced MD-END reveal the nuanced signals at the molecular level.

■ Results

All-atom MD simulations including an explicit solvent were applied to an ensemble-based and network-based study of allostery in the p53 DNA-binding domain with each system sampled for 1 µs. We seek to illustrate the linkage between regions of the p53 DNA-binding domain (DBD) as allosteric points of control/rescue. In order to test whether the PK11000 operates by restoring the wild-type (WT) dynamic profile to the Y220C mutant, we undertook MD simulations of the three relevant constructs: p53 WT, p53 having the Y220C mutation (Y220C), and the rescued mutant Y220C bound to PK11000 (Y220C + PK11000). Using this model system, we report on our insights into allostery and present a trio of MD-based machine learning methods as a useful paradigm for the identification of allostERIC signaling networks and structure-based drug design.

Measures of Population and Residue Dynamics. Global Distance Measures. Root Mean Square Deviation (RMSD) Analysis. Figure 1A shows the root mean square deviation (RMSD) of the backbone atoms of the three constructs with respect to their energy minimized and equilibrated starting structure. As expected, most conformations are within about 2 Å of their reference when converged over 1 µs. The mutant shows the greatest structural variation, oscillating between 1.5 and 3.0 Å, foreshadowing conformational substrates which we characterize in greater detail with subsequent lines of inquiry. The WT and PK11000 bound mutant trajectories tend to reside between 1.0 and 2.0 Å from their starting structures. Encouraged by the stability of the simulations and apparent convergence of global dynamic properties, we turned our attention to sequence specific analyses and the detailed characterization of the differences in their dynamic behavior. Root Mean Square Fluctuation (RMSF) Analysis. Global Dynamics: To assess the differences in dynamics by residue, we computed the root mean square fluctuations (RMSFs) of the alpha carbons with the average structure of each simulations as their reference coordinates (Figure 1B). We see that the fluctuations generally coincide with major structural features, with decreased mobility in regions corresponding to the 10 Beta sheets and two alpha helices, and higher fluctuations in loops between these elements. We observe high fluctuations of residues near the DNA direct binding residue K120 in the L1 loop of the Y220C mutant simulation. In contrast, we see the high mobility quenched when PK11000 is covalently introduced to the mutant system, suggesting recapitulation of wild-type dynamics and rescue of the Y220C mutant active site. Surprisingly, drug binding does not affect the mobility of loop 2 dynamics rather we see significant quenching of mobility among residues surrounding the Y220C mutant site in the presence of PK11000.

Antibody Trends: We have also considered the results in light of six known antibody binding regions: PAb1620 (WT specific), PAb246 (WT specific), and PAb240 (mutant specific). Many features of p53 mutant activity can be credited to the discovery of conformation specific antibodies that explicitly distinguish between WT and mutant p53. Among these, only a few have been reported to interact with the DNA-binding region. PAb240 which recognizes regions between amino acids 211–217 was shown to detect mutated p53, while PAb246 and PAb1620 which recognize amino acids 201–204 and 145–157, 206–212, respectively, were shown to detect WT p53. These positions are of interest because of the ability of different mutations of p53 to be differentiated by some of the antibodies that bind in these positions. As expected, the antibodies generally prefer regions in higher mobility regions as opposed to interior core residues. Regions that recognize WT conformation specific antibodies PAb246 and PAb1620 show varying degrees of local displacements between the three systems. In the PAb246 region, the WT and Y220C + PK11000 peaks are quite similar while the Y220C mutant system exhibits higher mobility, confirming the WT conformation specificity of the epitope. PAb1620 shows similar mobility in both systems where the Y220C mutant is present, even in the presence of allosteric effector, PK11000, while the WT system exhibits higher mobility. The inability of the PK11000 to recapitulate the WT conformation in the PAb1620 epitope may be because of its close proximity to the Y220C mutant, which, as reported by many, exposes a hydrophobic pocket. In the mutant specific PAb240 epitope, the Y220C mutant system displays the largest peak and highest mobility, and its mobility is dampened to levels lower than that of the WT system when PK11000 is present. In a structurally focused study, the mutant specific PAb240 epitope site RHSVY sequence has been shown to be buried in the WT conformation and becomes exposed in the mutant state, verifying the mutant specificity of this local region. Having captured these general trends, we were encouraged to examine construct- and position-specific analyses.

PK11000 Effects on Y220C Compared to WT: The RMSF profiles capture several differences between the WT (Figure 1B, blue) as compared to the Y220C (Figure 1B, red). The mutation occurs in the 10-residue connecting loop between beta strands 7 and 8, and it is proximal to the PAb240 epitope region. The region is distant from the DNA and other hotspot mutations. Network analyses and MD-Sectors follow up on this idea by looking for means in which this loop could be allosterically linked to the DNA-binding site. The introduction of the mutation corresponds to changes in dynamics in several regions. Dynamics is increased from around 1 to 2.5 Å in the peak centered around 112–124, the connector between strand 1 and strand 2 of the hairpin site (124–125). Lysine 120 binds at this connector region, a residue capable of discriminately interacting with the DNA and plays an important role in facilitating DNA recognition, and post translational modification. This warrants a closer look at the modulatory behavior of K120 in p53 dynamics. Furthermore, a very high region of fluctuation in both constructs peaks sharply at 208–210, in the center of the PAb1620 epitope site. The WT has a second major peak centered at 225, close to the Y220C mutation at...
approximately 3.8 Å. The mutation markedly amplifies this behavior, increasing the maximum peak height to 4.8 Å.

We next considered the ability of the PK11000 to restore the dynamic profile of the WT to the mutant, using the RSMF data to test our hypothesis that the restorative effect of PK11000 would be evident in the dynamic profile. In the K120-containing region, the presence of the drug largely dampened the increased flexibility introduced with the mutation with a broader peak but a similar extent of dynamics. Interestingly, the covalent binding of PK11000 to C182 did not affect the fluctuation of that specific residue, and its presence rather affected neighboring residues where the high fluctuation of residues in the PAb246 epitope region is dampened to a degree closer to that of WT. This indicates a significant reduction in dynamics at the linker between sheet 5 and sheet 6 which had previously been one of the three most dynamic positions aside from end effects. Additionally, the quenching in the neighboring strand 7-Strand 8 linker harboring the Y220C was restored to a native like state. Thus the PK11000 corresponded to restoration of p53 WT dynamic behavior in the K120, and strand 5-strand 6 linker, while depressing the dynamics in the 220 peak.

**Statistical Network Analysis. MD-MSM.** To gain insight into overall dynamic differences, we applied our MD-based Markov state models to our trajectories. In brief, the method identifies substates in the trajectories from the reference of the protein. Because we cluster protein structures from all trajectories simultaneously rather than independently for each trajectory, identified substates are guaranteed to be identical across trajectories. We then assess the populations in each cluster as the states in the MSM, and the transitions are obtained from the time evolution of cluster visited by each trajectory. The populations in each substate and the transitions are closest to their own centroid, indicating members of each cluster are closest to their own centroid and far from other centroids. The initial guess of centroid selection often lies near the inflection point (S1, Panel B) of the weighted average of each centroid with respect to number of clusters, and that initial guess is confirmed by examining the distributions of distances of each simulation’s snapshots with respect to all possible centroids.

In order to compare the populations in each substate conformation, we show the frequencies in cluster normalized by simulation as depicted in Figure 2B. We see that WT (blue) and drug bound mutant (green) populate a single substate exclusively in cluster 1 (Figure 2B). Conversely, only 27.4% of the Y220C mutant populates this conformational substate all of which are frames residing in the first 20% of the trajectory. The Y220C mutant (orange) populates substate 2 almost exclusively, with 72.6% of its snapshots falling in substate 2 and the rest in substate 1. This indicates that the MD-MSM was sensitive to the differences in dynamics because of the mutation. Examining the Y220C rescued by PK11000, we bear in mind the question as to whether PK11000 restores native dynamics, or if an alternate substate profile could potentially be characteristic of a rescued system which can be further investigated by higher order clustering and the frequency of simulation frames in higher order clusters. This will factor into drug design for rationally crafting novel therapeutics. Thus, for this example of a rescued Y220C mutant, the efficacy of the drug arose through demonstrating that both drugged and WT simulations exclusively adopt the same substate.
Recognition Interface Hydrogen-Bonding Analysis. We next wished to investigate whether the Markov state constructs exhibited differences that propagated into the hydrogen-bonding network. A series of polar contacts between the DNA and residues within the protein were identified among the centroids as shown in Figure 3. Structures represented in cartoon are centroidal, and representative of each cohort and were then laid out to represent their hydrogen bonding in flat representation (Figure 4). Hydrogen-bonding residues in the native substate in green consists of K120, S121, S243, R275, and R282. Hydrogen-bonding residues in the inactive mutant substate in red also consist of the same residues and an additional six more extra DNA-binding residues. The binding mode of K120 is significantly affected where in the native substate K120 directly binds to the major groove while in the mutant inactive substate, and K120 occupies the minor groove of the DNA-binding site.

We have also examined the hydrogen-bonding network of the raw trajectory data and the relative occupancies of unique polar contacts present in the active site. Global hydrogen bond count results show that Y220C mutant p53 has a significantly higher number of hydrogen bonds made throughout the 1 ms simulation, oscillating around 12 unique hydrogen bonds. In contrast, WT exhibits the lowest number of unique bonds fluctuating between 3 and 6 unique hydrogen bonds, and PK11000 present Y220C mutant shows a slightly higher number than that of the WT system at around 7–8 unique hydrogen bonds (S2, Panel A). The relative occupancies are displayed in flat representation focusing on amino acid residues that were previously reported by Cho and colleagues in the crystal structure report of p53 DBD. One factor to take away from this result is the loss of direct hydrogen bond by C277 in the Y220C mutant as a hydrogen bond acceptor. Among the DNA interacting residues reported by Cho and colleagues, C277 is the only amino acid that acts as a hydrogen bond acceptor binding directly to the DNA base pair of the major groove. The fact that this interaction is virtually nonexistent in the Y220C mutant suggests its functional role in signaling and in differentiating between mutant and WT dynamics of p53 DBD.
Measures of Allosteric Communication Networks. MD-Sectors and MD-END. Up to this point we have considered structural dynamics globally (RMSD and MD-MSM) and by sequence (RMSF) and compared between the constructs. However, a mystifying aspect of the mutation is that it confers affinity in the DNA binding without actually being present in the active site itself. This phenomenon is reminiscent of allosteric control, or action at a distance. We cast this quandary in the framework of allostery and we turned to methods to shed light on how the signaling within the protein under these various states may propagate through the protein. For this we turn to tools we developed addressing the networks of residues within the protein achieved through pairwise decomposition based on their cooperative distance (MD-Sectors) and energetic variance (MD-END). We then compared the findings of each of our constructs.

MD-Sectors. A heat map for MD-calculated motional displacements. The indices i and j run over amino acid residues in p53 DBD. The regions in white correspond to neighboring inter-residue contacts, blue correspond to high covariance and yellow to low covariance, based on an interatomic distance cutoff <10 Å. (B) Venn diagram of Sector residues. Venn diagram indicates the coincidences of sector residues obtained from MD-Sectors and their overlap between the three cohorts. WT p53 represented in blue, Y220C mutant in red, and rescued mutant in green. Overlap residues are colored in yellow for all overlapped sectors, WT + Y220C in purple, WT + PK11000 in teal, and PK11000 + Y220C in orange. In the right, the same sector residues are highlighted in the cartoon figure in the same color palette. The Y220C mutation and PK11000 binding amino acid C182 is depicted in black VDW representation.
covariance matrices reveal stark differences in pairwise interactions of all the amino acid residues between the WT and the Y220C mutant. Highest correlations are indicated in blue and low correlations are indicated in yellow. The PK1100 bound Y220C mutant shows a dampened signal that contrasts the highly interactive covariance matrix of the Y220C mutant system. Figure 5B shows the MD sector residues together with the positions observed in the form of a Venn diagram. Sectors that are unique to each system are highlighted in blue, red, green for WT, Y220C mutant, and PK11000 bound Y220C mutant, respectively. Spectral analysis of the MD-calculated motional covariance matrix resulted in an MD sector of 39 residue positions with five residues, R249, I125, R267, N268, and E271, that are consistent among all constructs (Figure 6, Figure 6. MD-END electrostatic network. (A) WT electrostatic network. (B) Y220C mutant electrostatic network. (C) PK11000 bound Y220C mutant electrostatic network.
panel B, in yellow). Each sector amino acid positions are shown as 1 Å spheres centered on the Cα atoms of amino acid residues. We observe that many sector residues coincide with regions of highly scrutinized sites such as hotspot mutations, zinc binding residues, and DNA contact residues (Table 1), verifying that the analysis indeed identifies functionally crucial residues.

### Table 1. Functional Sector Residues Ranked

|       | WT          | Y220C Mutant | Y220C, PK11000 bound |
|-------|-------------|--------------|----------------------|
| R175  | Hotspot Mutations | R175       | Hotspot Mutations | C238 | Zinc binding |
| C242  | Zinc binding | C176         | Zinc binding        | R248 | Hotspot, DNA |
| R249  | Hotspot, DNA | R249         | Hotspot, DNA        | R249 | Hotspot, DNA |
| R273  | Hotspot, DNA | R273         | Hotspot, DNA        |     |              |
| A276  | DNA binding  | C277         | DNA binding         | R280 | DNA binding  |
|       |             | R283         | DNA binding         |     |              |

“Important residues that are mutual in all systems are colored in yellow, WT and Y220C mutual sectors in purple, and WT and PK11000 bound Y220C mutual sectors in teal.

### DISCUSSION

In this study we assess the extent to which a mutant protein can be restored by an allosteric drug by restoring the native protein dynamics. To this end we have explored this with the p53 DBD having the Y220C mutation which has been rescued by the compound PK11000 by an unknown mechanism which we hypothesize to be via allosteric modulation. A recent report of the allosteric effector PK11000 that rescues the Y220C hotspot mutation of p53 presents an ideal test case to study the effect of allosteric disruption and points of control for rescue. We have tackled this assessment in terms of atomic flux and population dynamics (RMSD/RMSF, and MD-MSM, respectively), and measures of allosteric communication networks by SCA of motionally and energetically coupled residues (MD-Sectors and MD-END, respectively). Here we discuss our findings and their implications for allosteric drug design.

Our simulations and statistical analysis show that PK11000 rescues Y220C mutant p53 DBD by recapitulating WT dynamics and hydrogen-bonding network. Y220C exhibits aggregative properties via indiscriminate hydrogen bonding with the consensus DNA sequence. MSMs reveal two clusters with PK11000 present Y220C mutant system in the same cluster as the WT simulation, and the Y220C mutant system in the other, suggesting that PK11000 bound Y220C p53 is thermodynamically akin to that of WT p53. Statistical coupling analysis (SCA) by pairwise decomposition of distance-based and energy-based variables from our simulations reveal an allosteric hub of amino acids that propagate energetic signaling between Y220 and the active site via beta sheet residues 257–273. Sector residues identified 39 residues that are motionally covariant. Among those, very few residues (R249, I251, R267, N268, and E271) are present among all three cohorts, implicating p53’s sensitivity to mutagenesis and their allosteric role in the functional dynamics and pathogenesis of p53.

The RMSF results identified a number of segments of residues whose dynamics are disrupted by the Y220C mutation and are restored by the presence of PK11000. Many of these, interestingly, correspond to the location of the binding sites of antibodies capable of distinguishing between WT and mutant p53, suggesting that they are sensitized to the dynamics in the region and that some similarities in role of destabilization may be shared between mutants. For example, in our results, the average flux of the residues across corresponding to the WT specific antibody epitope PAb246 is identifiably deviant only in the Y220C mutant system. The same epitope regions, where in our prior work on R175H hotspot, points to other mechanisms that could be at play involving the opening of a hydrophobic patch that destabilizes the p53. This combination molecular therapeutics could be valuable particularly with patients suffering from late-stage cancers that have evolved an ensemble of mutations.

We also note that some regions were not reinstated to native dynamics. This suggests that full rescue may not be required for the restoration of essential function. Alternately, some dynamics variation may be unimportant for its functionality. The extent to which p53 may stay functional was measured via the interaction interface between p53 and the DNA. Hydrogen-bonding analysis at the interaction interface of p53 with its cognate DNA provides insight into the changes in H-bond pattern occurring in the presence of the Y220C mutation. The structural consequences of allosteric perturbation and rescue should reflect back onto the direct readout and recognition of the DNA sequence. We analyzed the hydrogen-bonding network in the interface of the centroidal structures of each system (Figures 3 and 4) that were derived from our MD-MSM analysis. The rescue of Y220C mutant by PK11000 is evident by the almost identical hydrogen-bonding patterns (Figure 4A,B) between the WT and drug bound mutant p53 system. DNA recognition has been restored in the presence of
PK11000, raising questions about how the information may be energetically propagated through an allosteric network which we have taken up with the network analysis. Furthermore, we have noticed the residue K120, has an increased residence time interacting with the DNA in the presence of the PK11000. This may provide additional stabilization and its binding orientation can be a good indicator of p53 activation or deactivation. This also suggests a means of differentiation of binding sites for p53, extending the contact region specificity to this position. The finding echoes the observation from the RMSF analysis that restoration need not be exactly the same, and that the level of resolution by which MD-MSMs reveal similar overall conformations is adequate for capturing allosteric activity.

While much of the energetic network stayed consistent throughout all systems, the sectors reveal a network that are vastly different from one another, sharing only very few residues all throughout. The presence of only five conserved residues stresses the precarious nature of p53’s innate stability. The bottleneck of essential residues leaves the protein highly mobile loop regions in proximity to the drug binding site. When PK11000 is covalently introduced to the mutant, we see that many of the sectors begin to cluster in regions involved in DNA contact mainly in the C terminus alpha helix and the L1 loop. The sectors that overlap between the three cohorts present a subnetwork of residues where the kinetic signals from the allosteric perturbation propagate. The overlapping residues of WT and PK11000 bound Y220c mutant (Figure 6B, tea), occupy the same beta sheet region between the L3 loop and Strand loop (residues 257−273) as the five bottleneck residues that are present in all systems. R273, DNA-binding residue and also a hotspot mutation site is conserved motionally in our RMSF results across all three systems, suggesting that this beta sheet region is an overlooked allosteric mediator between the distal Y220 and the DNA-binding site, and therefore is crucial for modulating p53’s functional recognition of the DNA.

Taken together, our findings reveal an allosteric signaling hub at play in the presence of allosteric disruptor Y220C and Y220C reactivator, PK11000. The Y220C mutant is aggregation-prone, where hydrogen bonds are made indiscriminately with the DNA interface with PK11000 recapitulating WT dynamics. Protein wide measures were taken to examine the extent to which communication between the distal points of the mutation and binding could be linked to the readout interface indicated by some level of interaction via an allosteric signaling network. MD sector results reveal a subset of unique residues that are correlated with each other by their motional displacements. MD-END, elucidates the information network and the kinetic transfer based on nonbonded atomic interaction energies, while MD-Sectors unveils only the motional response of individual residues to allosteric disruption, together with MD-END we gain additional insight into the energetic network of covariant residues.

Materials and Methods

Simulation Specifications and Trajectory Analysis. All-atom 1 microsecond MD simulations on the p53 DBD, residues 96−290, with explicit solvent were performed using the standard lab protocol with the AMBER14.0 and 16.0 simulation packages and AMBERTOOLS14 suite. FF14SB force fields for the protein and the TIP3P potential for solvent water were used. p53 DBD starting configurations of the WT (PDB ID:1TUP, Cho et al, 1994) and Y220C + PK11000 (PDB ID: SLAP, Bauer et al. 2016) systems were obtained from the Protein Data Bank (PDB). Mutagenesis of residues of interest was performed using molecular visualization program PyMol. The molecular structure of ligand PK11000 (PDB ID:6SM) was parameterized using AMBER’s Antechamber suite and covalently bonded to Cysteine-182 using t-LeaP. The zinc coordination parameterization was achieved using Zinc AMBER force field. The simulation system was treated under Particle Mesh Ewald periodic boundary conditions with a 10 Å Lennard-Jones cutoff in a truncated octahedral box. Na+ counter-ions were added to the system for electroneutrality and SHAKE was applied for hydrogen bond motions. Energy minimization with decreasing constraints on the protein solute was carried out, followed by heating to 300 K and temperature maintained using the Berendsen algorithm. Simulations were run using the parallelized CUDA version of the PMEMD routine on NVIDIA Graphical processing units. All MD trajectories were analyzed with the AMBER utility “cpptraj” in AMBERTOOLS14and VMD to calculate the RMSD and RMSF of the C-α backbone atoms in AmberTools16. Individual trajectory principal component analysis (PCA) was performed using the Bio3D R package. PCA was employed to examine the exploration of the conformational space of each cohort and the PC contributions to the RMSF of each residue. Hydrogen-bonding analysis was carried out using cpptraj from the Amber14 package, with standard distance and angle cutoff were set to 3.5 Å and 30°, respectively, and visualized using MacPyMol.

MD-MSM. Our MD-MSM study of p53 DBD is a statistically driven illustration of allosteroy and based on a previously done study on the CRIB-PDZ in our lab. A conformational selection may occur when the ensemble of p53-DBD structures of protein, in the absence of ligand, clusters with those of a ligand bound trajectory, or in this case, a small therapeutic molecule PK11000. The categorization into discrete probabilistic states indicates the protein’s innate predisposition to adopt specific conformations in the presence or absence of an allosteric effector or a mutation. MSMS were constructed in terms of the nodes and links of a complex network, with the nodes obtained by clustering the microstates using K-means clustering using Amber’s native cpptraj analysis toolkit. The links were defined from the Chapman−Kolmogorov compliant frequency of direct transitions between nodes. In all calculations, atom-based quantities obtained from MD were merged to present results by each residue.

MD-Sectors. Motional and energy covariance matrices were computed from MD trajectories by standard methods using the AMBER utility cpptraj in AmberTools16. Spectral analysis was applied to the covariance matrices. The first
eigenvalue of a spectral decomposition is by definition unity, and the remaining eigenvalues define modes which comprise a series expansion of the matrix in covariance space. The leading terms make the largest contributions to the expansion and, to be reasonably inclusive, MD-Sectors were defined based on the eigenvectors of the leading 20% of the eigenvalues of the spectral decomposition. Distance covariance was used to measure the correlations of the joint independence of any two vectors. By computing the pairwise covariance between all residues over the simulation time, pairwise covariance was mapped for all residues as defined by Lakhani et al.\(^1\)

**MD-END and Pairwise Interaction Energy Network.** Computing the pairwise interaction energy for every frame in the trajectory is very computationally costly. To limit the computational burden, we computed pairwise energies for a smaller subset of frames from each trajectory. To choose which frames were included in our sample, we implemented a simple Monte Carlo procedure. First, we sample frames at some fixed interval (2 frames per nanosecond). Next, we iterated the sampling over all three trajectories and tested to see if the RMSD of the sample to the Markov state modeled cluster centroid was less than the average RMSD of all points in the cluster to the centroid. This allows us to determine if the sample is representative of structures in its cluster. If the RMSD is greater than the average, we simply skip the frame. After collecting these representative samples, we select our final batch of samples for each cluster proportional to the Boltzmann distribution for each trajectory. For each cluster and trajectory, we picked 10% of frames with the smallest RMSD and randomly sampled the rest 90% to highlight both depth and breadth of the free energy wells for each system. These samples were then used to compile a new trajectory that was fed through cpptraj’s energy analysis protocol. We computed the energy for each of the 195 residues (residue numbers 96–290 of p53) and chose two combinations of residue pairs by using a bash script that performed the cpptraj energy command with each pair of residues. Finally, we used python\(^3\) to parse the cpptraj outputs into tensors of shape (N,X,D,D) where N is the number of frames in the trajectory X is the number of types of energy being studied and D the number of residues. We then simply summed over each energy channel to compute one representative network per energy channel. Finally, the edges of the network are assigned thresholds by normalizing the edge weights on a per node basis and picking all nodes above that threshold.

**Data Sharing.** All of the datasets used in this study are available from the authors upon request. Sample UNIX and python scripts, including those implemented in Jupyter notebook for generating clustering and network graph analyses are available in our supplementary section (S3, Data Sharing: UNIX and Python Scripts).

**ASSOCIATED CONTENT**

Supporting Information. The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.1c05624.

MDMSM RMSD convergence and weighted average distance of clusters; flat representation of the hydrogen-bonding network and number of unique hydrogen bonds over 1 microsecond; data sharing of sample UNIX and python scripts for calculating hydrogen bonding, CPPTRAJ analysis, clustering, and MD-Sectors (PDF)

**AUTHOR INFORMATION**

**Corresponding Author**

In Sub M. Han – College of Integrative Sciences Hall-Atwater Laboratories Middletown, Wesleyan University, Middletown, Connecticut 06459-0180, United States; orcid.org/0000-0003-2791-5975; Phone: (860) 685-2210; Email: ihan01@wesleyan.edu; Fax: (860) 685-2211

**Authors**

Dylan Abramson – College of Integrative Sciences Hall-Atwater Laboratories Middletown, Wesleyan University, Middletown, Connecticut 06459-0180, United States

Kelly M. Thayer – College of Integrative Sciences Hall-Atwater Laboratories Middletown, Wesleyan University, Middletown, Connecticut 06459-0180, United States; orcid.org/0000-0001-7437-9517

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.1c05624

**Funding**

This research was generously funded by NIH-R15 GM128102

**Notes**

The authors declare no competing financial interest.

**ACKNOWLEDGMENTS**

We would like to thank Molecules to Medicine consortium for fruitful discussion, especially Dr. David L. Beveridge, Dr. David R. Langley, Dr. Michael P. Weir and their lab members; Bharat Lakhani and Jonathan Fabry for providing programming structures to successfully perform MD-MSM and MD-Sectors.

**REFERENCES**

(1) Sheik Amamuddy, O.; Veldman, W. M. C.; Manyumwa, C.; Khairelall, A.; Agajan, S.; Oluwemi, O.; Verkhivker, G. M.; Tastan Bishop, O. Integrated Computational Approaches and Tools for Allosteric Drug Discovery. Int. J. Mol. Sci. 2020, 21, 847.

(2) Krieger, N. M.; Johansson, F. D.; Morris, C. A survey on graph kernels. Appl. Netw. Sci. 2020, S, 6.

(3) Wodak, S. J.; Paci, E.; Dokholyan, N. V.; Berezovsky, I. N.; Horovitz, A.; Li, J.; Hilser, V. J.; Bahar, I.; Karanicolas, J.; Stock, G.; Hamm, P.; Stote, R. H.; Eberhardt, J.; Chebaro, Y.; Djeaegere, A.; Ceolin, M.; Chazoux, J. P.; Bolhuis, P. G.; Vreede, J.; Faccioli, P.; Orioli, S.; Ravasio, R.; Yan, L.; Beito, C.; Wyart, M.; Glek, P.; Rivalta, I.; Palermo, G.; McCammon, J. A.;Panecka-Hofman, J.; Wade, R. C.; di Pizio, A.; Niv, M. Y.; Nussinov, R.; Tsai, C. J.; Jang, H.; Padhory, D.; Kozakov, D.; McLeish, T. Allostery in Its Many Disguises: From Theory to Applications. Cell 2019, 27, 566–578.

(4) Jiménez-Osés, G.; Osuna, S.; Gao, X.; Sawaya, M. R.; Gilson, L.; Collier, S. J.; Huisman, G. W.; Yeates, T. O.; Tang, Y.; Houk, K. N. The role of distant mutations and allosteric regulation on LovD active site dynamics. Nat. Chem. Biol. 2014, 10, 431–436.

(5) 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.

(6) Stephen, C. W.; Lane, D. P. Mutant conformation of p53. Precise epitope mapping using a filamentous phage epitope library. J. Mol. Biol. 1992, 225, 577–583.

(7) Bauer, M. R.; Joerger, A. C.; Fersht, A. R. 2-Sulfonylpyrimidines: Mild alkylating agents with anticancer activity toward p53-compromised cells. Proc. Natl. Acad. Sci. U. S. A. 2016, 113, ES271–ES280.

(8) Zerbe, B. S.; Hall, D. R.; Vajda, S.; Whitty, A.; Kozakov, D. Relationship between hot spot residues and ligand binding hot spots in protein-protein interfaces. J. Chem. Info. Model. 2012, 52, 2236–2244.
(11) 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.

(12) Halabi, N.; Rivoire, O.; Leibler, S.; Ranganathan, R. Protein sectors: evolutionary units of three-dimensional structure. Cell 2009, 138, 774−786.

(13) Lakhani, B.; Thayer, K. M.; Black, E.; Beveridge, D. L. Spectral analysis of molecular dynamics simulations on PDZ: MD sectors. J. Biomol. Struct. Dyn. 2020, 38, 781−790.

(14) Yewdell, J. W.; Gannon, J. V.; Lane, D. P. Monoclonal antibody analysis of p53 expression in normal and transformed cells. J. Virol. 1986, 59, 444−452.

(15) Milner, J.; Cook, A.; Sheldon, M. A new anti-p53 monoclonal antibody, previously reported to be directed against the large T antigen of simian virus 40. Oncogene 1987, 1, 453−455.

(16) Wang, P. L.; Sait, F.; Winter, G. The ‘wildtype’ conformation of p53: epitope mapping using hybrid proteins. Oncogene 2001, 20, 2318.

(17) Pieniazek, S. N.; Hingorani, M. M.; Beveridge, D. L. Dynamical allostery in the mechanism of action of DNA mismatch repair protein MutS. Biophys. J. 2011, 101, 1730−1739.

(18) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. The biomolecular simulation programs. J. Comput. Chem. 2005, 26, 1668−1688.

(19) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Biomolecular Simulation Programs. J. Comput. Chem. 2015, 36, 3696−3713.

(20) Jorgensen, W. L. Quantum and statistical mechanical studies of liquids. 10. Transferable intermolecular potential functions for water, alcohols, and ethers. Application to liquid water. J. Am. Chem. Soc. 1981, 103, 335−340.

(21) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. J. Chem. Phys. 1983, 79, 926−935.

(22) DeLano, W. L. The PyMOL Molecular Graphics System; DeLano Scientific: San Carlos, CA, 2002.

(23) Peters, M. B.; Yang, Y.; Wang, B.; Füsti-Molnár, L.; Weaver, M. N.; Merz, K. M., Jr. Structural Survey of Zinc-Containing Proteins and the Development of the Zinc AMBER Force Field (ZAFF). J. Chem. Theory Comput. 2010, 6, 2935−2947.

(24) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems. J. Chem. Phys. 1993, 98, 10089−10092.

(25) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. J. Comput. Phys. 1977, 23, 327−341.

(26) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. J. Chem. Phys. 1984, 81, 3684−3690.

(27) Pérez, A.; Marchán, I.; Svozil, D.; Sponer, J.; Cheatham, T. E., 3rd; Laughton, C. A.; Orozco, M. Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. Biophys. J. 2007, 92, 3817−3829.

(28) Prinz, J. H.; Wu, H.; Sarich, M.; Keller, B.; Senne, M.; Held, M.; Chodera, J. D.; Schütté, C.; Noé, F. Markov models of molecular kinetics: generation and validation. J. Chem. Phys. 2011, 134, 174105.

(29) Grant, B. J.; Rodrigues, A. P.; ElSawy, K. M.; McCammon, J. A.; Caves, L. S. Bio3d: an R package for the comparative analysis of protein structures. Bioinformatics 2006, 22, 2695−2696.