ACS OMEGA

http://pubs.acs.org/journal/acsodf

Article

Network Analysis of Molecular Dynamics Sectors in the p53 Protein

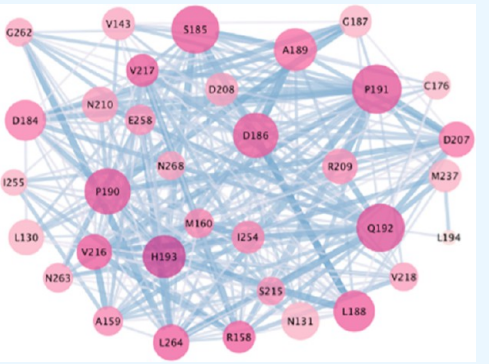
Jonathan D. Fabry and Kelly M. Thayer*

Cite This: *ACS Omega* 2023, 8, 571−587

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Design of allosteric regulators is an emergent field in the area of drug discovery holding promise for currently untreated diseases. Allosteric regulators bind to a protein in one location and affect a distant site. The ubiquitous presence of allosteric effectors in biology and the success of serendipitously identified allosteric compounds point to the potential they hold. Although the mechanism of transmission of an allosteric signal is not unequivocally determined, one hypothesis suggests that groups of evolutionarily covarying residues within a protein, termed sectors, are conduits. A long-term goal of our lab is to allosterically modulate the activity of proteins by binding small molecules at points of allosteric control. However, methods to consistently identify such points remain unclear. Sector residues on the surfaces of proteins are a promising source of allosteric targets. Recently, we introduced molecular dynamics (MD)-based sectors; MD sectors capitalize on covariance of motion, in place of evolutionary covariance. By focusing on motional covariance, MD sectors tap into the framework of statistical mechanics afforded by the Boltzmann ensemble of structural conformations comprising the underlying data set. We hypothesized that the method of MD sectors can be used to identify a cohesive network of motionally covarying residues capable of transmitting an allosteric signal in a protein. While our initial qualitative results showed promise for the method to predict sectors, that a network of cohesively covarying residues had been produced remained an untested assumption. In this work, we apply network theory to rigorously analyze MD sectors, allowing us to quantitatively assess the biologically relevant property of network cohesiveness of sectors in the context of the tumor suppressor protein, p53. We revised the methodology for assessing and improving MD sectors. Specifically, we introduce a metric to calculate the cohesive properties of the network. Our new approach separates residues into two categories: sector residues and non-sector residues. The relatedness within each respective group is computed with a distance metric. Cohesive sector networks are identified as those that have high relatedness among the sector residues which exceeds the relatedness of the residues to the non-sector residues in terms of the correlation of motions. Our major finding was that the revised means of obtaining sectors was more efficacious than previous iterations, as evidenced by the greater cohesion of the networks. These results are discussed in the context of the development of allosteric regulators of p53 in particular and the expected applicability of the method to the drug design field in general.



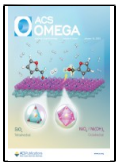
1. INTRODUCTION

Allosteric signaling in proteins occurs when an effector molecule binds at one site and confers a functional change at a topologically distinct location.[1−4] The phenomenon has been known for over half a century and pervades biology to the extent that it has been dubbed "The second secret of life."[5] Nevertheless, the mechanism by which it traverses the protein remains actively debated. Models broadly categorize mechanisms as pathways,[6−8] in which signals travel along spatially proximal residues connecting the two sites, or energy landscapes proposed by Cooper and Dryden[9] acting via a reorganization of the energy landscape over a long range.[10−12]

The emergence of allosteric drugs[2,13−15] provides a glimpse into an alternate modality of protein control and engineering over current strategies. While the mechanism widely controls biochemistry, most are discovered serendipitously because a requisite understanding of the principles governing it remains largely obscured. Allosteric drugs offer the promise to control an active site distally, leaving it unoccluded to continue carrying out a function. This is highly desirable when activity needs to be modulated as opposed to abolished, the very crux of many yet uncured diseases. Cancer is among these;[2] reactivation of mutant p53 to wild-type behavior as a pathway to tumor eradication constitutes one such challenge undertaken by our lab.

The full-length p53 protein is made up of 393 residues[16] and can be divided into three regions: the *N*-terminal, the C-terminal and the core domain.[17] The core domain of p53 has been crystallized (PDB ID 1TUP[18]) and illustrates the binding

Received: August 31, 2022
Accepted: December 5, 2022
Published: December 20, 2022



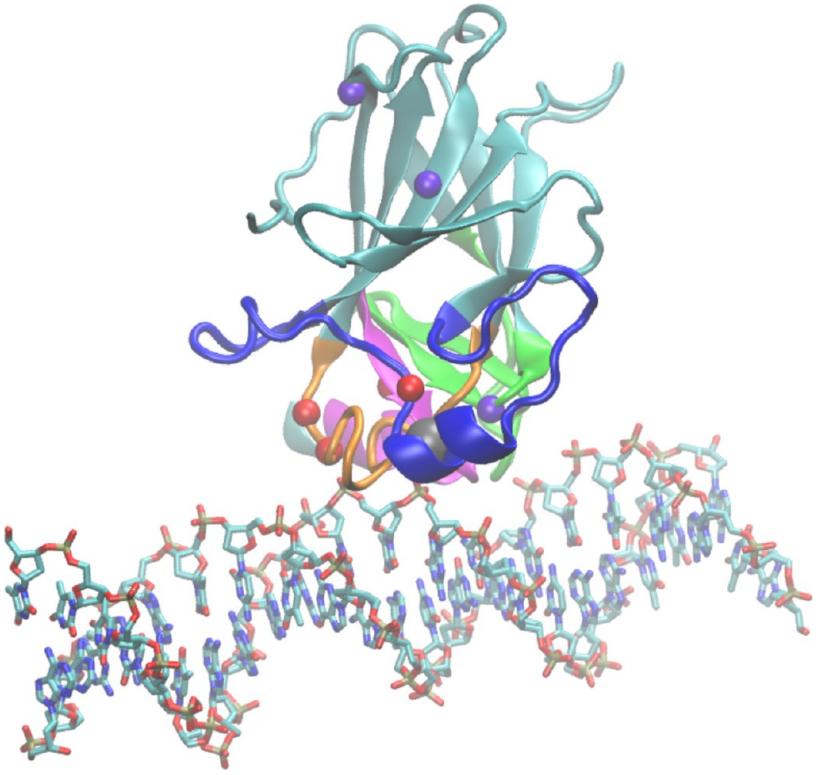


Figure 1. p53 Hospots: 3-D visualization of p53 bound to DNA with hotspot residues 175, 248, 249, and 273 highlighted as red spheres. The 1TUP[18] crystal structure highlights structural features of the p53 structure: backbone trace (cyan), bound DNA (thick sticks), coordinated essential zinc (gray sphere), loop L1 residues 113−140 (green), loop L2 residues 163−195 (royal blue), loop L3 residues 236−251 (orange), strand loop helix residues 271−286 (magenta), and Ala mutation scan residues 158, 264, and 137 (purple spheres).

interface between p53 and its cognate DNA. It encompasses the most important hotspot mutations[19−21] 175, 248, 249, and 273, all of which negatively affect p53's tumor suppressing capabilities (Figure 1).

Toward restoring native activity of hotspot mutations, the Thayer Lab is developing a pipeline incorporating mechanistic details of allostery. Our ideas presently involve identification of residues involved in allosteric control using machine learning to generate small molecules, docking the molecules, and computationally observing the extent to which native dynamics of the protein is restored in simulations. A major bottleneck in the process is accurately identifying residues which allosterically control the binding specificity of p53 to its target DNA sequences. Such residues are of significant interest because the change in binding specificity they cause can affect apoptosis or cellular repair in p53, thereby preventing tumor growth. We have turned to the idea of sectors to provide such information. In their first incarnation, sectors characterized evolutionarily covarying groups of residues hypothesized to transmit allosteric signals, among other possible functions.[22−24] We reimagined them as molecular dynamics (MD)-based sectors (MD sectors),[25] motionally covarying residues based on the time evolution of positional coordinates in MD simulations. By focusing on the motion of residues, we ground MD sectors in physical principles deriving from statistical mechanics of the Boltzmann ensemble of thermally accessible structures. The procedure follows that of the evolutionary sectors but substitutes the motional covariance matrix for the original evolutionarily covarying version. This then allows us to study the conformational interchange of substates using Markov state models[26] derived from the trajectories obtained from MD simulations (MD-MSM). In our pipeline, MD-MSMs constitute a validated means to assess the extent to which an allosteric effector has restored the target activity.[27−29] We envision utilizing MD sector residues as allosteric points of control from which we will design small-molecule therapeutics. The efficacy of these potential drugs will be further explored with MD-MSMs.

While the aforementioned developments have made significant inroads to link sector residues to first principles, the claim that the sectors constitute a cohesive network, or a network at all, has been an assumption rather than a fact. By cohesive, we refer to the strength to which motions are correlated. We aim to categorize residues as either in a network or not in a network based on motional correlation to introduce objective quantitation of cohesiveness. The measure of cohesiveness will be used as the basis for our analysis on the extent to which MD sectors comprise a network. The MD sector calculation is based upon the pairwise correlation matrix, and residues are chosen for sector membership based on the overall correlation of its motions to all other residues, subject to a spectral analysis decomposition. We note, however, that therein lies an inconsistency; a cohesive sector by definition ought to be more correlated within the group than to residues outside the group. Methods up to this point have not taken that into consideration. In this work, we address the two key points raised, namely, quantitating the extent of cohesion among residues in an MD sector network and reformulating the cohesion measure to reflect strength of correlation of members of the sector groups among itself as compared to the expected lack of cohesion of residues not in the sector.

https://doi.org/10.1021/acsomega.2c05635
ACS Omega 2023, 8, 571−587

2. METHODS AND COMPUTATIONAL DETAILS

2.1. MD Simulation. MD simulations provide an estimate of the Boltzmann ensemble of molecular structures with atomic detail. The AMBER suite of programs has been developed for the purpose of simulating the behavior of biomolecules and has undergone extensive development and testing.[30−33] The introduction of graphics processing unit (GPU) computing has enabled routine simulations on the order of 100s of nanoseconds to microseconds.[30] The crystal structure of p53[18] was used as the starting point, considering the central protein bound to the DNA. MD simulations were carried out according to standard protocol as follows: The AMBER 16 molecular simulation package was implemented for parameterization and modeling.[33,34] The TIP3P potential was used for solvent water,[35,36] the force field ff99SB was used to model atomic interactions,[37] ions08 was used to model monovalent Cl^- and K^+,[38] and particle mesh Ewald summation with a cutoff of 10 Ångströms was implemented to maintain the accuracy of electrostatic interactions.[39] The SHAKE algorithm applied the necessary constraints on hydrogen bond motions.[40] The system was energy-minimized and then heated to 300 K, and Berendsen coupling to a heat bath maintained the temperature.[41] After an equilibration phase, the simulation was carried out to 200 nanoseconds using the pmemd code with the parallel computing architecture.[42−45]

In this case, four simulations using the above protocol were carried out on various forms of p53: wt and the three alanine mutations R158A, L264A, and L137A. All simulations were within 2 Ångströms of dispersion from the crystal structure as expected,[46] and the convergence of each simulation was monitored with standard tools such as the Euclidean distance-based measure root-mean-square deviation (rmsd) as well as ensemble-based analysis.[47] Various lab analysis methods were used: concatenating multiple trajectories for MD-MSMs and calculating covariance values for MD sectors.[48−51] MD-MSMs have been developed and validated previously, and we follow the protocol as described therein. The results are analyzed in terms of the population of trajectory snapshots assigned to the centroids, and the time evolution of the transition of the snapshots between centroids. To avoid over-interpretation, we chose not to analyze the transition matrix because it was observed to be very sparse and instead focus on the cluster membership.

2.2. SCA Sectors. Intra-protein interactions between amino acids provide insight into how proteins fold, bind, and adapt. An underlying assumption of SCA sectors is that groups of residues—termed sectors—which change together across evolution may provide insight into protein functions. The innate characteristics (evolutionary covariance and evolutionary conservation) defining an SCA sector do not necessitate a functionally cohesive network of residues; thus, the practical significance of sectors obtained via SCA sectors remains an active area of research. MD sectors, our sector methodology, is innovative due to its ability to identify and test for residues of allosteric control in silico through an analysis based strictly on the ensemble of dynamic structures adopted by a protein.

2.3. MD Sectors. To perform MD sectors,[50] pairwise distance covariance was computed from MD simulation trajectories. Two residues covary with one another if their positions tend to deviate in any direction at the same time over the course of an entire trajectory. A pairwise motion covariance matrix (MCM) can be computed (Figure S1). We set the covariance of bonded residues to 0 (white cells in the matrix), which screens this information out as not meaningful. High covariance (darker blue cells) between the bonded residues would be expected and likely significant enough to overshadow the information gleaned by observing non-bonded residue covariance. Note that the matrix is symmetric about the diagonal.

Spectral decomposition analysis was applied to the covariance matrix to identify a network of residues that share covarying motions. This measure of covariance is distinct from pairwise covariance which involves only two residues; here, we observe covariance on a more global scale for groups of residues. The symmetric motion correlation matrix is decomposed into a product of its diagonal eigenvalue matrix and corresponding eigenvector matrices

$$\mathrm{MCM} = V \cdot D \cdot V^{\mathbb{T}} \tag{1}$$

where MCM stands for the motion correlation matrix, D is the diagonal eigenvalue matrix, V is the eigenvector matrix corresponding to D, and $V^{\mathbb{T}}$ is the transpose of V. The resulting eigenvalues represent the magnitude of information captured; larger eigenvalues are indicative of the most important linear combinations of variables. The eigenvalue distribution was compared against a randomized eigenvalue distribution to screen out noise and determine the number of significant eigenvalues. The further use of independent component analysis (ICS) ensured statistical independence between selected residues. ICS takes the eigenvalues and outputs *V*pica values, the final 'scores' assigned to each residue.[48] *V*pica values report on the degree to which a residue covaries motionally in the context of the entire system. The MD sector is defined by taking the 20% of residues with the highest *V*pica values, mirroring the SCA methodology. At this point, we have a complete sector, and analysis of the sector residues follows.[50]

In contrast to slow and costly high-throughput methods such as screen-based assays, MD sectors can efficiently identify potentially functionally significant residues at a minimal cost.[52] We feel, however, that analyzing the MD sector as a network of residues would further validate the sector hypothesis and improve our ability to pinpoint allosteric pathways and pockets to target. Such network analysis would allow MD sectors to offer a robust means of identifying not only a group of allosteric residues but also the network properties of such residues, which is valuable information to be carried forth into the drug design process.

Furthermore, using MD sectors in conjunction with our MD-MSM conformational analysis enables us to precisely detect the impact of manipulating sector residues by casting mutations or binding small-molecule ligands. An updated implementation of MD sectors allows for broad and simplified use which fits seamlessly into our drug discovery pipeline. This implementation of MD sectors represents a substantive deviation from screen-based allosteric drug discovery. In the realm of drug design, MD sector analysis holds a unique promise of generating a rich dynamics-based explanation of specific allosteric control that is not available in other network-finding methods.

2.4. Graph Theory and Networks. Sectors are hypothesized to function as cohesive networks capable of influencing protein functions. However, formal network theory has not yet been applied to sectors obtained from the various methods. To test the sector hypothesis, we reformatted biological systems into graph-centric systems or networks. Residues were taken as

https://doi.org/10.1021/acsomega.2c05635
ACS Omega 2023, 8, 571−587

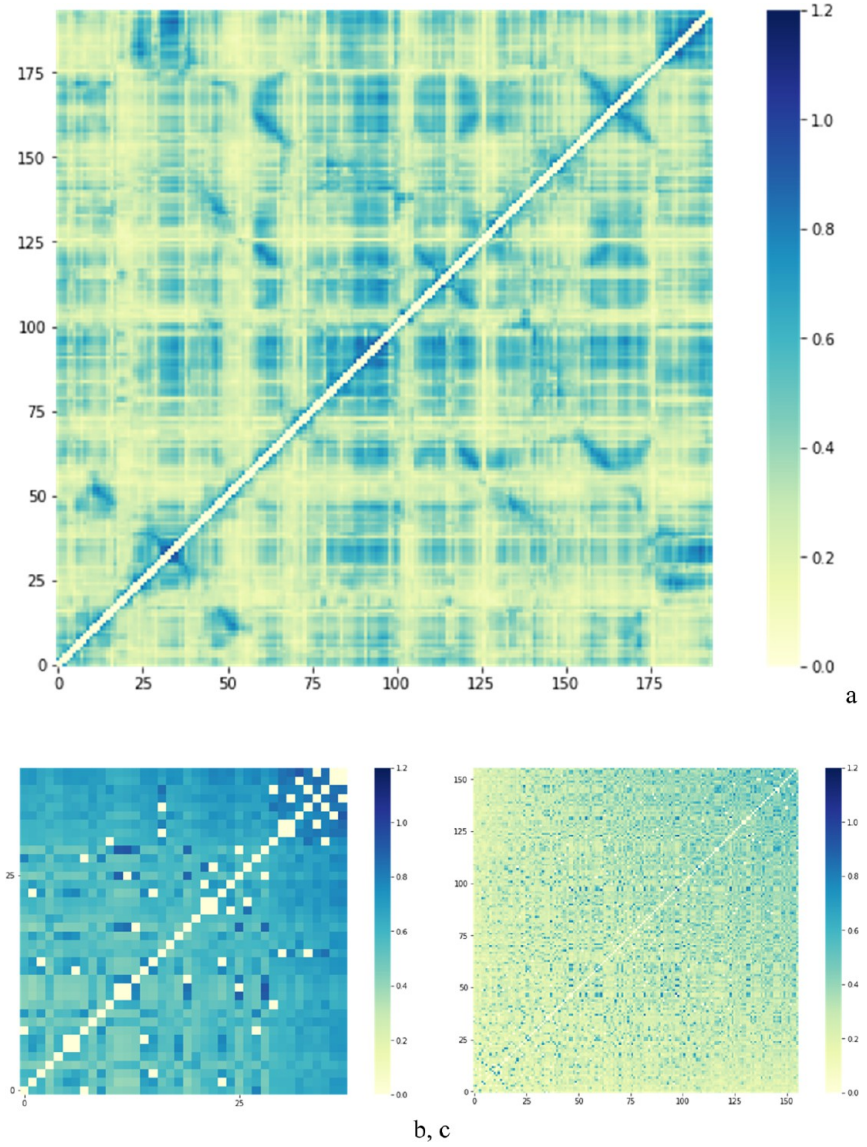


Figure 2. Covariance Heat maps: (a) pairwise covariance values between p53 residues are shown. (b) Heat map of pairwise covariance values between sector residues in p53. (c) Heat map of pairwise covariance values between non-sector residues in p53.

nodes, and bidirectional edges were drawn between all possible edges to represent covariance interactions.

A graph G is a tuple (V, E) where V is a set of vertices, or nodes, and E is a set of edges. Typically, we say $N = |V|$ and $M = |E|$. A simple graph where $M = N(N - 1)/2$ is said to be a complete graph since the graph contains all possible edges (Figure S2).[53]

Graphs are defined to be directed if edges are given direction; for each $e_i = (v_i, v_j)$, the pair of vertices is ordered.[53] A subgraph of $G = (V,E)$ is defined as a graph $G' = (V',E')$ in which $V' \subseteq V$ and $E' \subseteq E$.[54]

A common property observed in a graph is the degree of a node. For a node v_i in an undirected graph, the degree $d(v_i)$ of a node is the number of edges connected to, or incident to, the node. In Figure S2a, $d(a) = 2$ and in 2b, $d(a) = 4$.[53]

Two nodes v_i and v_j are connected if there exists a path which can be traversed from v_j to v_i. A connected component C of some graph $G = (V,E)$ is a set of vertices $\{v_1, v_2, ..., v_k\} \subseteq V$ in which all $v_i \in C$ are connected to one another and not connected to any $v_i \notin C$ (Figure S3b).[53] We say a graph is connected if it contains at least one connected component.[55] A bridge is any edge e whose deletion splits the graph into more components than it previously had (Figure S3c).[56]

The probability that a pair of nodes v_j and v_k are directly connected given they are both directly connected to some other node v_i is known as the clustering coefficient of v_i. The clustering coefficient c_i of node v_i is calculated using the following formula[53]

$$c_i = \frac{2\eta_i}{d_i(d_i - 1)} \text{ for } d_i \geq 2;$$

$$\eta_i = \#\text{ of edges between nodes in the neighborhood of } v_i \quad (2)$$

The average clustering coefficient of a network, denoted by $\overline{c}$, is given by

$$\overline{c} = \frac{1}{N}\sum_{i=1}^{N} c_i \quad (3)$$

https://doi.org/10.1021/acsomega.2c05635
ACS Omega 2023, 8, 571−587

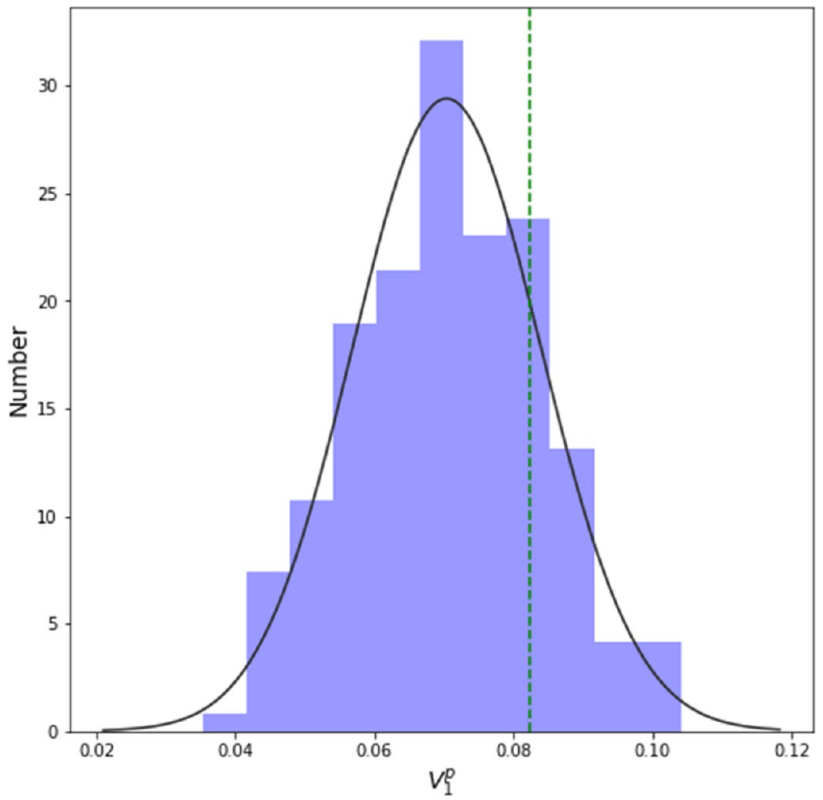


Figure 3. *V*pica Histogram: Histogram of number of p53 residues having various *V*pica values. The vertical green dotted line represents the cutoff for sector membership, while the black line overlaid on the histogram models a normal distribution.

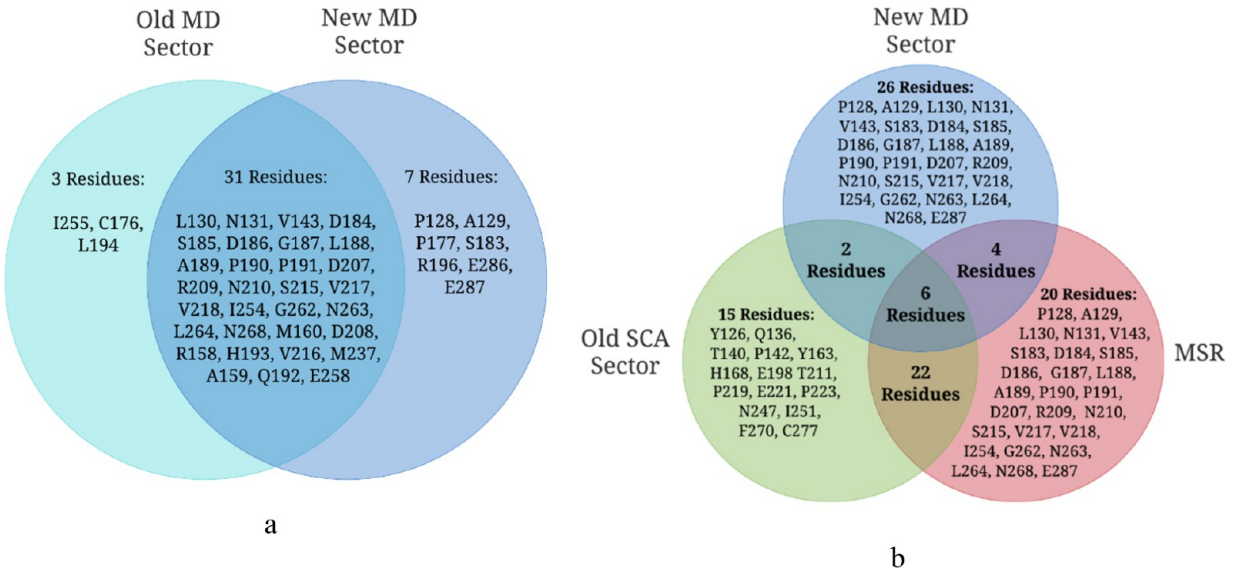


Figure 4. Sector Venn Diagrams: (a) Venn diagram of MDS1.0 and MDS2.0 membership. The high level of overlap confirms that the two implementations are reporting on similar protein behaviors. (b) Venn diagram of MDS2.0 (38 residues total), the SCA sector (45 residues total), and MSR (52 residues total) membership in p53.

To utilize network theory statistics and render related images, the open source network visualization tool Cytoscape was used.[57]

3. RESULTS AND DISCUSSION

3.1. MD Sector Analysis on p53. We will now examine the results of running the updated implementation of MD sectors on the p53 tumor suppressor protein. The main results to be analyzed here relate to pairwise covariance values, *V*pica values, sector overlaps, and sector cohesion.

Figure 2 presents a heat map of pairwise covariance from 200 ns MD simulations of the p53 DNA binding domain. Lighter yellow sections correspond to low correlation, while darker blue sections correspond to high correlation. The bottom-left to top-right diagonal contains only values of 0, reflecting a step in which the pairwise covariance values between all neighboring residues

https://doi.org/10.1021/acsomega.2c05635
ACS Omega 2023, 8, 571−587

were set to 0. A wide distribution in the correlated motions that tend to have higher pairwise covariance and lower pairwise covariance exists within p53. The pairwise matrix is spectrally decomposed to pick out the 20% of residues which covary the most. Their level of covariance is quantified as a *V*pica value. A chart of *V*pica values by position is shown in Figure S4. A histogram of *V*pica values (Figure 3) was used to ascertain sector residue membership. Residues to the right of the vertical line, which signals higher *V*pica values, comprise the sector.

Taking the top 20% of residues in the p53 core domain based on *V*pica values, the following 38-member sector constituting the MD sector is obtained: P128, A129, L130, N131, V143, R158, A159, M160, P177, S183, D184, S185, D186, G187, L188, A189, P190, P191, Q192, H193, R196, D207, D208, R209, N210, S215, V216, V217, V218, M237, I254, E258, G262, N263, L264, N268, E286, E287.

Next, the pairwise covariance matrix was broken into two smaller heat maps, one showing pairwise covariance values for sector residues and one for non-sector residues (Figure 9). As expected, the sector residues have noticeably higher pairwise covariance values. This suggests that the sector conserves highly correlated residue motions.

We now observe the overlap between various residue sectors of p53. We will reference two versions of the MD sectors procedure throughout the remainder of this paper. MDS1.0 represents our original method which was based on a previous implementation of SCA sectors which involved the manual removal of specific residues to better match the range of residues observed in SCA sectors on the same protein. MSD2.0, which is the focus of this paper, is based on an updated implementation of SCA sectors and does not omit any additional residues. To verify that the MDS2.0 implementation is working as intended, we will compare MDS1.0 and MDS2.0. MDS1.0 contains 34 residues, and MDS2.0 contains 38 residues. The overlap of MDS1.0 and MDS2.0, which contains any residue present in both sectors, has 31 residues out of a maximum of 34 (Figure 4a). The results of the updated method are consistent with those of the previous iteration despite a fully revamped implementation. The differences between the sectors were negligible and can likely be attributed to the slight changes in our pre-processing techniques.

Figure 4b displays the overlaps of MDS2.0 with the SCA sector and the mutationally sensitive residues (MSRs) of p53. MSRs are residues that confer a change to the protein structure. Notice the level of overlap between MDS2.0 and the other two groups of residues; 12 out of the sector's 38 residues (31.6%) are found to be potentially significant via functionally unrelated methods. The difference in methodologies leads to findings based off of different characteristics. With a highly dynamic and unstable protein such as p53, we did not expect an overwhelming overlap between the identified sectors.

To test the cohesion of the sector, we developed a novel procedure termed *V*pica ratio analysis (VRA). In our procedure, we compute the covariance matrix C_0 as described (Ranganathan et al., Lakhani et al.) using the motional correlation of residues as the basis for covariance measures. The spectral decomposition algorithm, represented as the decomposition operator $\hat{D}$ (Ranganathan et al.), is applied to the covariance matrix C_0 to obtain $V\text{pica}_i$, the measure of each residue's extent of covariance with all other residues. We then divide the total set of residues in the protein $\boldsymbol{T}$ into sets $\boldsymbol{S}$, the sector residues, and $\boldsymbol{N}$, the non-sector residues, such that $\boldsymbol{T} = \boldsymbol{S} \cup \boldsymbol{N}$ and $\boldsymbol{S} \cap \boldsymbol{N} = \emptyset$. To select which residues constitute the sector, we rank order the residues according to their associated *V*pica values. The number of residues selected as the sector is equal to 20% of the total number of residues $\boldsymbol{N}$. Consequently, the sector $\boldsymbol{S}$ contains the residues mapped to the largest *V*pica values.

We then generate submatrices C_s and C_N by filtering C_0 subject to the conditions that C_S solely contains entries of covariance between sectors and C_N solely contains entries of covariance between non-sector residues. Notice that the cross terms are discarded by the filter. In other words, the filtering follows the following criteria

$$C_{ij} \in C_S \text{ iff } i \in \boldsymbol{S} \text{ and } j \in \boldsymbol{S} \text{ for } i, j = \{1,2, \ldots, N\} \tag{4}$$

$$C_{ij} \in C_N \text{ iff } i \in \boldsymbol{N} \text{ and } j \in \boldsymbol{N} \text{ for } i, j = \{1,2, \ldots, N\} \tag{5}$$

A second round of *V*pica analysis via the spectral decomposition procedure $\hat{D}$ is carried out to obtain *V*pica values reporting on the strength of the interaction of those residues within the sector and within the non-sector, in contrast to the $V\text{pica}_0$ values reporting on the interaction of the residue with the rest of the matrix. The method is applied to each of the submatrices $\hat{D}(C_S)$ and $\hat{D}(C_N)$ yielding a list of *V*pica values corresponding to each residue, $V\text{pica}_{s_1}$ to $V\text{pica}_{s_N}$ and list $V\text{pica}_{N_1}$ to $V\text{pica}_{N_N}$; in other words, there is one *V*pica value mapped to each residue, and those residues are in either sector list $\boldsymbol{S}$ or non-sector list $\boldsymbol{N}$. For residue $i \in \boldsymbol{S}$, its *V*pica_in value equals $V\text{pica}_{S_i}$, the list component value, which reports on its covariance within the sector and excludes consideration of the interaction with residues not in the sector.

To calculate its *V*pica_out value for each sector residue i, we suppose the residue were not in the sector and append it to set $\boldsymbol{N}$. Correspondingly, this adds the row and column of residue i interaction with all other non-sector residues; the values $C_{i,j}$ from C_0 for $j = \{1, 2, 3, \ldots, N_N\}$ are appended to matrix C_N to create the revised matrix C_N'. The spectral decomposition procedure is then applied to the revised matrix to observe how the *V*pica values change: $\hat{D}(C_N')$.

The *V*pica ratio analysis then consists of computing the ratio

$$\text{VRA} = \frac{V\text{pica\_in}}{V\text{pica\_out}} \tag{6}$$

In the case of p53, we found that the average ratio value for residues in the MD sector was ∼1.55, while the average ratio for residues out of the sector was ∼1.41. Thus, there is a higher level of covariance between sector residues with themselves as compared to the covariance between non-sector residues and sector residues. This suggests that the sector residues form a cohesive group which conserves motional covariance.

To put these ratio values in perspective, *V*pica ratio analysis was also performed on the other residue groups of interest. We will call a group of 38 random residues in the DNA binding domain of p53, the random 38 sector (38SR). The average ratio values for residues in the 38SR was ∼2.10, and the average ratio for residues out of 38SR was ∼2.03. The average ratio for residues in the SCA sector was ∼1.85, while the average for residues outside the SCA sector was ∼1.81.

Let us now compare, for these three groups of residues, how the average in-sector ratio compares to the average out-of-sector ratio. In other words, how much more do sector residues covary with one another than with non-sector residues? To quantify this difference, we divide the average ratio for residues in the sector

https://doi.org/10.1021/acsomega.2c05635
ACS Omega 2023, 8, 571−587

by the average ratio for residues outside the sector; let us call this value the sector cohesion level (SCL).

The SCL for the MD sector is ∼1.12, the SCL for the 38SR is ∼1.04, and the SCL for the SCA sector is ∼1.03; the MD sector has a greater SCL than the other two sectors, suggesting a higher level of cohesion. Further related data are shown in Figure S5.

Another point of comparison relating to VRA is the overlap of sector residues with residues which have the highest ratio values. This overlap examines whether certain residues have high motional correlation in the context of the entire core domain of p53 as well as a higher correlation to residues within the sector. The overlaps between the three sectors discussed above and the corresponding residues with the highest ratio values are displayed in Figure 5. Note that the MD sector and 38SR overlaps were between groups of 38 residues, while the SCA sector overlap was between groups of 45 residues; these values are the sizes of each sector. Thus, 47.4% of MD sector residues, 34.2% of 38SR residues, and 42% of SCA sector residues overlap with high-ratio residues. MD sectors have the highest percentage overlap, exhibiting a higher success rate at identifying residues with high overall covariance and high covariance localized within the sector.

Sector	Number of Overlapping Residues	Percent of Overlapping Residues	List of Overlapping Residues
MD	18	47.4%	158, 159, 160, 177, 191, 193, 207, 210, 215, 216, 217, 218, 237, 254, 258, 262, 264, 268
R38S	13	34.2%	116, 120, 122, 123, 124, 125, 130, 132, 136, 180, 213, 273, 276
SCA	19	42.0%	126, 132, 138, 163, 173, 175, 179, 198, 208, 211, 213, 239, 242, 246, 247, 250, 251, 270, 273

Figure 5. Overlapping Ratio Residues. Overlapping residues between MDS2.0, 38SR, SCA sector, and residues with the highest ratio values.

3.2. Network Analysis on p53 Sectors. We will now relate the findings to the structure of the p53 tumor suppressor protein as a network of residues and interactions. The 1TUP crystallized DNA binding domain of p53 contains 194 residues, which is the section of p53 on which network analysis was performed. Figure 6 shows a complete, undirected graph containing all residues in the DNA binding domain of p53. Each residue is represented by a node (circles). Residue interactions are represented by the blue edges. There is an edge between all pairs of residues, creating a vastly dense network. In total, there are $194 \times 193/2 = 18{,}721$ edges in this network. We will break down this network to focus on a variety of groups of residues. All networks discussed are assumed to be simple, undirected, and edge-weighted with motional pairwise covariance values found using the procedures described in the methods section, unless otherwise specified.

Subgraphs visualizing the new MD sector (MDS2.0), old MD sector (MDS1.0), SCA sector (SCAS), random 38 sector (38SR), and the mutationally sensitive residues (MSR) were created using a standardized procedure. Note that the random sectors were created by selecting 38 random residues from the DNA binding domain of p53. The steps to create the MDS2.0

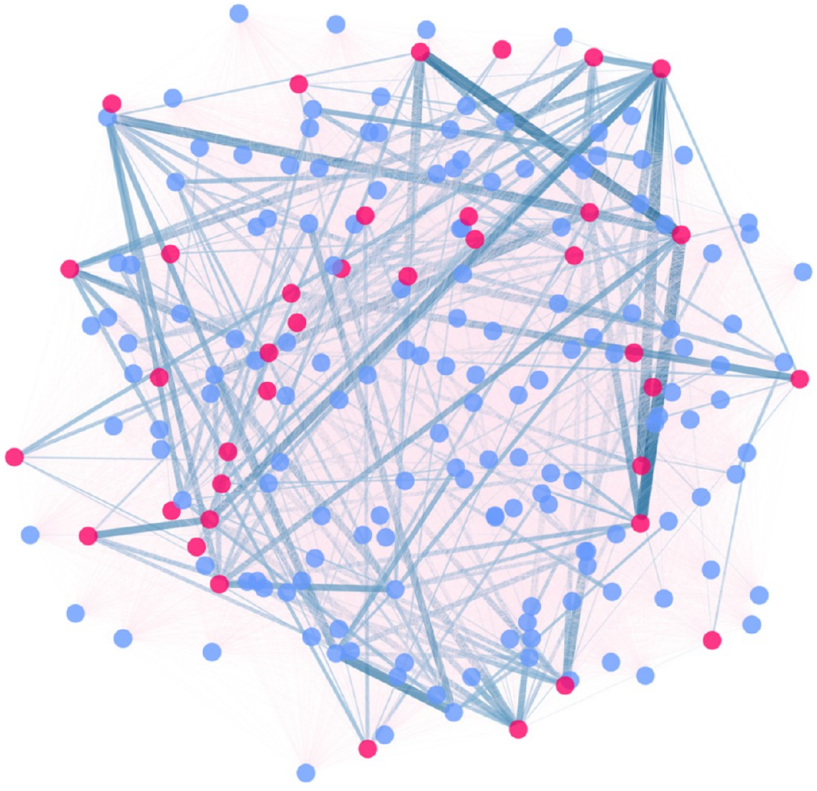


Figure 6. Complete Graph of p53. Complete graph visualization of p53. MDS2.0 residues are highlighted in red, and non-sector residues are highlighted in blue. Additionally, higher pairwise covariance values map to thicker, bluer edges to emphasize the relationship between sector residues and strong pairwise interactions.

https://doi.org/10.1021/acsomega.2c05635
ACS Omega 2023, 8, 571−587

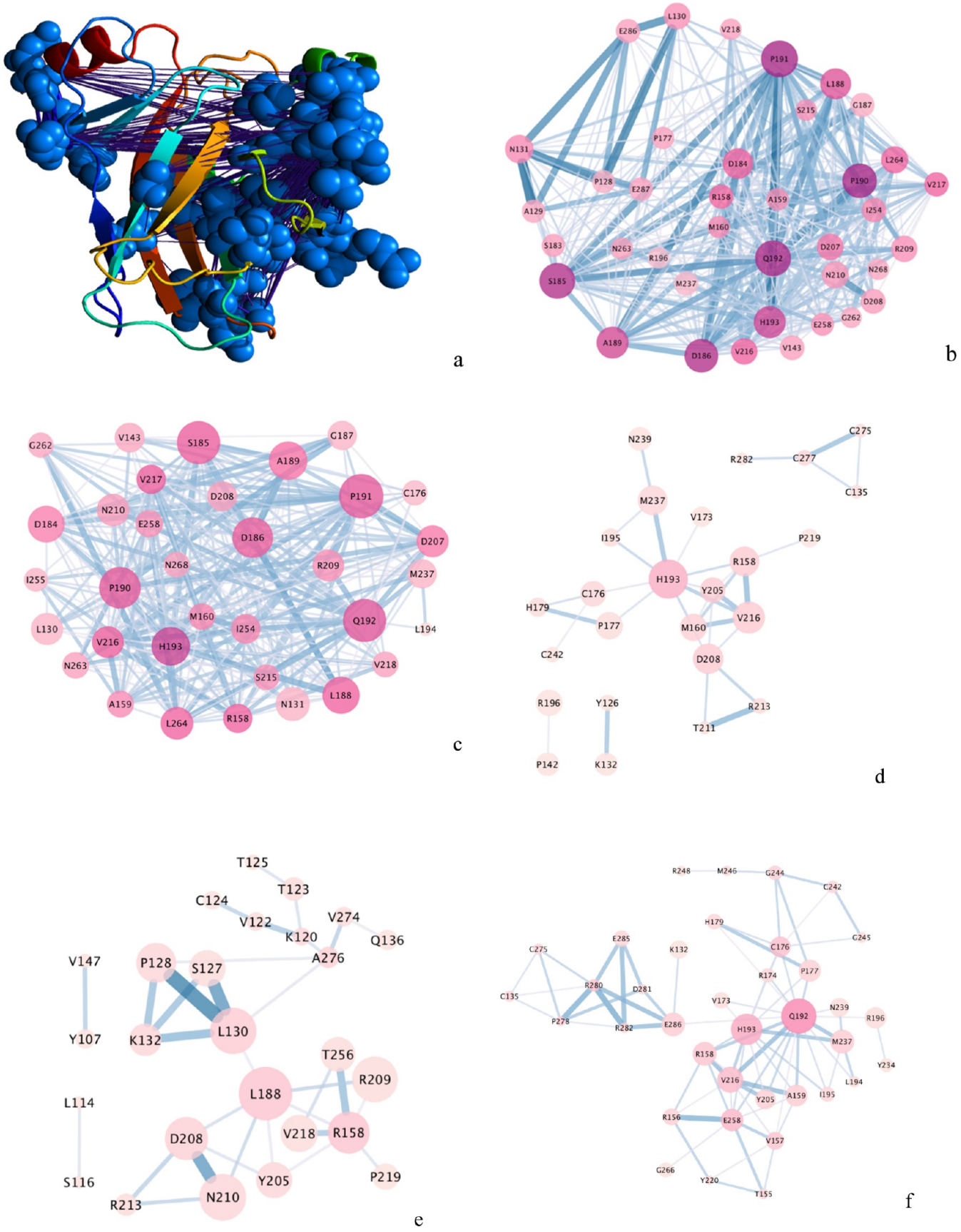


Figure 7. Subgraphs of High Motional Covariance (SHMC) (a) MDS2.0 visualization as a 3-D model using PyMOL. (b) MDS2.0 SHMC: $N = 38$, $M = 345$, one component. Sector visualization as a graph using Cytoscape. (c) MDS1.0 SHMC: $N = 34$, $M = 280$, one component. (d) SCAS SHMC: $N = 25$, $M = 33$, four components. (e) 38SR SHMC: $N = 26$, $M = 32$, three components. (f) MSR SHMC: $N = 37$, $M = 78$, one component.

https://doi.org/10.1021/acsomega.2c05635
ACS Omega 2023, 8, 571−587

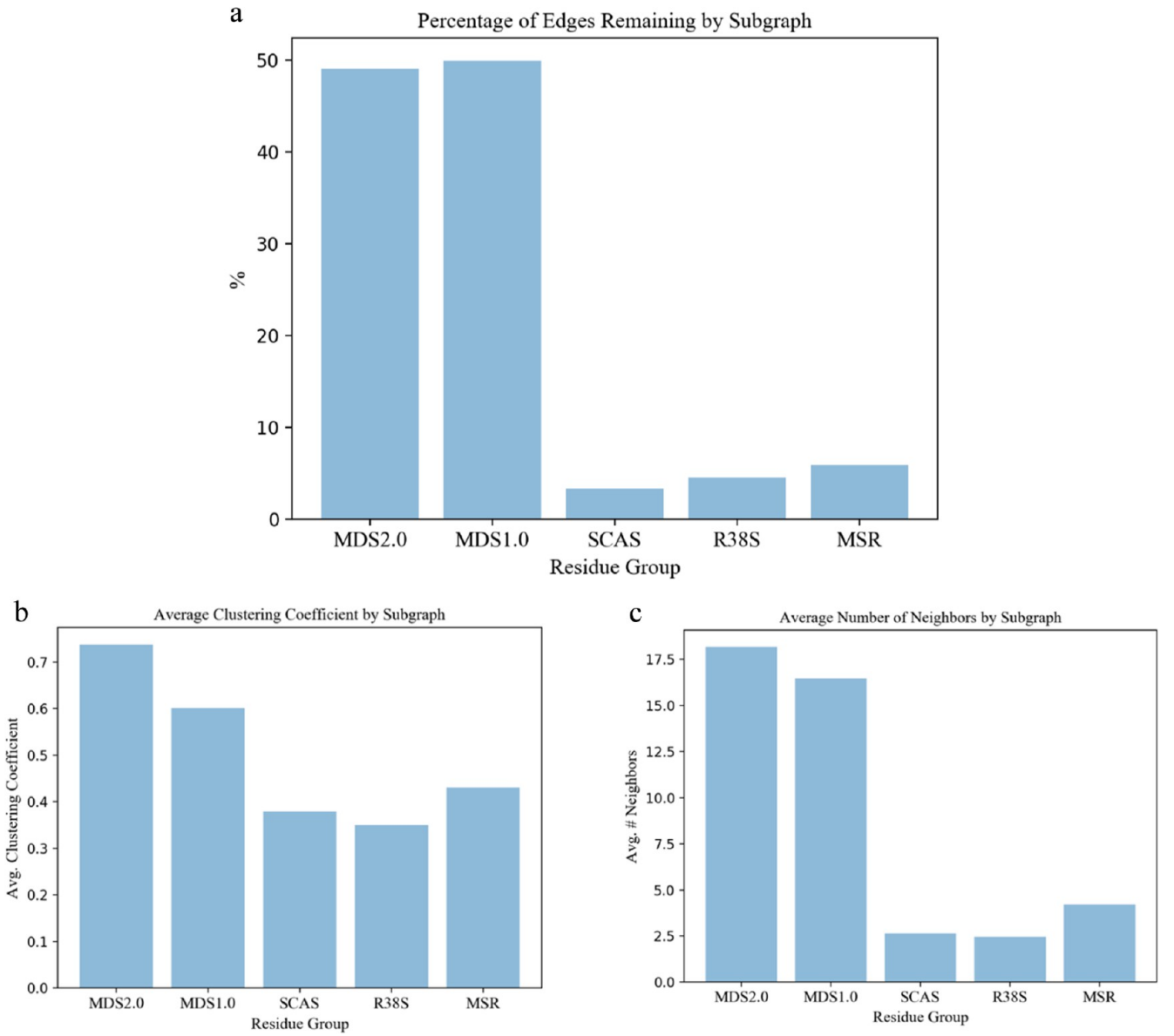


Figure 8. Network Statistics. (a) Bar chart of the percent of edges remaining in each SHMC, as compared to a complete subgraph containing the nodes of the associated residue group. The MD sectors SMHCs have a strikingly higher percentage of edges remaining. (b) Bar chart of the average clustering coefficient in each SHMC. While the distribution of this chart is not overly variable, MD sectors SHMCs do have high average clustering coefficients. (c) Bar chart of the average number of neighbors in each SHMC. The chart suggests that nodes within the MD sectors SHMCs have many more neighbors.

subgraph will be described; analogous steps were applied to create subgraphs for each of the residue groups.

A subgraph of all MDS2.0 residues and all edges between such residues was created. From this subgraph, a pairwise cutoff value of ∼0.61 was set; it represents the 95th percentile of pairwise covariance values in the motion correlation matrix. Thus, only edges which represent prominent interactions remain. All edges with pairwise covariance values above this cutoff, and nodes with any such adjacent edges, remain in the subgraph. Mappings between the subgraphs' characteristics and their visual representations were implemented; higher *V*pica values give larger nodes, higher node degrees give darker-pink nodes, and higher pairwise covariance values give bluer, thicker edges between pairs of nodes. We take the resulting network to be our subgraph to analyze (Figure 7b). We call each of these subgraphs, shown in Figure 7, subgraphs of high motional correlation (SHMCs). These are the subgraphs of the complete graph shown in Figure 6.

The SHMCs are significantly less dense than the initial complete graph. They represent the strongest remaining motional interactions within their respective group of residues. Presumably, SHMCs which have more remaining edges contain more potential pathways for transducing signals via physical interactions. Figure network statistics: Figure 8a shows the percentage of edges remaining in each SHMC; to get this percentage for the MDS2.0 SHMC, for example, we did the following calculation

$$\frac{\#\text{edges in MDS2.0 SHMC}}{\#\text{edges in complete subgraph containing all MDS2.0 residues}} = \frac{345}{703} \approx 0.491$$

https://doi.org/10.1021/acsomega.2c05635
ACS Omega 2023, 8, 571−587

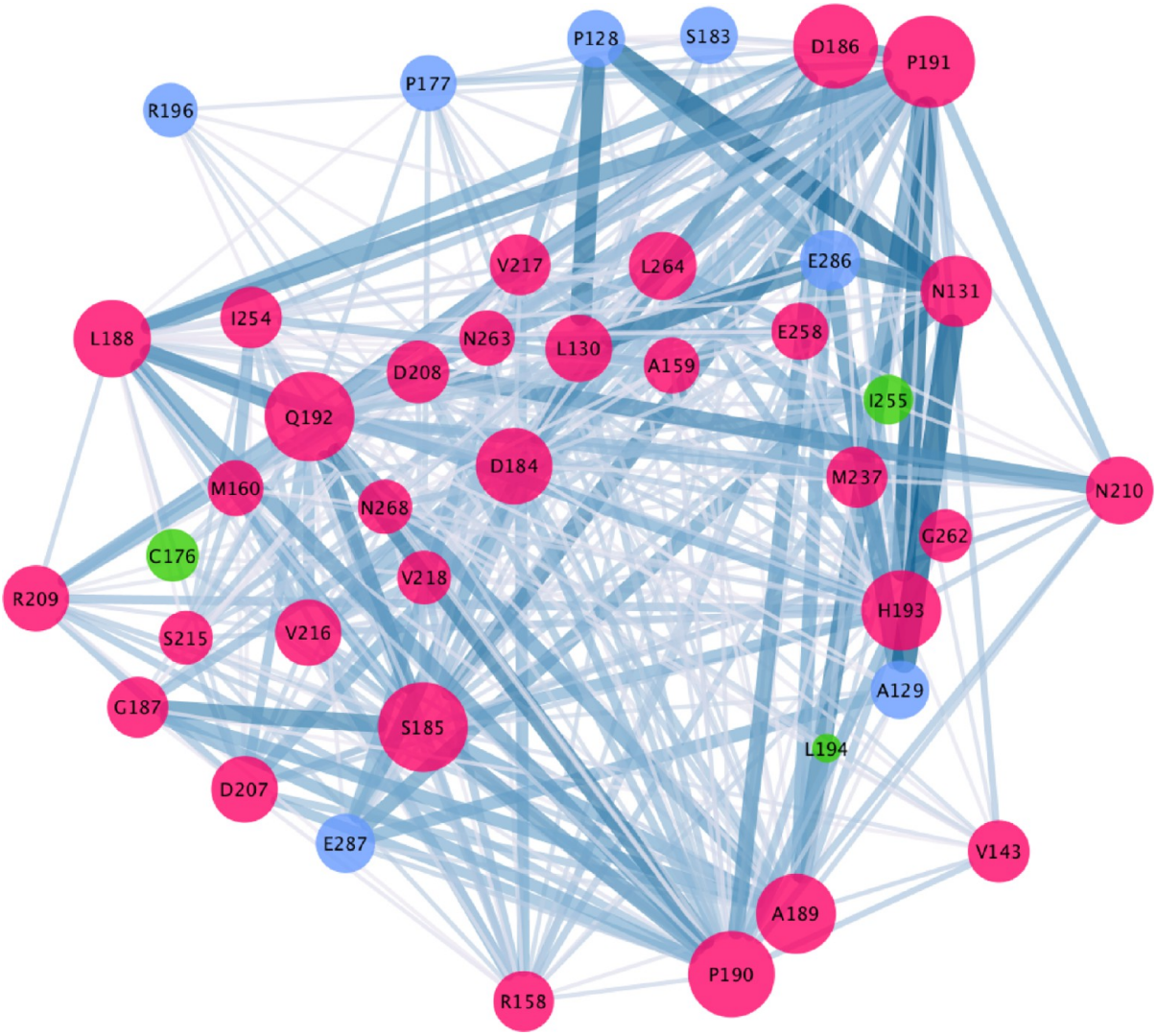


Figure 9. Graph of MDS Overlaps. Overlap between MDS1.0 and MDS2.0. Red nodes are in both sectors, blue nodes are only in MDS2.0, and green nodes are only in MDS1.0. *V*pica values are mapped to node size; lower *V*pica values correspond to smaller nodes, while higher *V*pica values correspond to larger nodes.

Quite clearly, the MD sectors SHMCs have a higher percentage of edges remaining, meaning that within MD sectors, a larger quantity of strong motional correlations between residues persists. We observe the differences between the MDS2.0 and 38SR SHMCs (Figures 7b,e), which can both have a maximum of 703 edges. The 38SR SHMC has only 32 remaining edges, while the highly connected NMDS SHMC has 345 remaining edges; the MDS2.0 SHMC retains over 10 times the number of edges a random sector of the same size does.

Using the 38SR as a control makes apparent the significance of developing a criterion for deciding what kind of a node is "important." When forgoing a real measure of importance and instead picking randomly, we end up with a fragmented network. There are quite a few degree-one residues in the 38SR as well as some bridges holding together multiple components; the edge between L130 and L188 is a bridge, for example. These types of residues and edges create a very tenuous network. The probability of a signal propagating across the network is greatly reduced by dead ends and a lack of pathways. The MDS2.0 SHMC provides a picture of a network which is dense and cohesive; even between residues physically far apart in the protein chain, there are dozens of pathways connecting them.

Figure 7 tells a similar story; the MD sectors SHMCs remain fuller, while the other SHMCs become sparse. Only the MD sectors SHMCs and MSR SHMC have one component, while all the other SHMCs are broken down into as many as five components. Even the SCAS SHMC, which we believe to capture some level of functional value, contains four components. If allosteric signaling does in fact rely on cooperated motions between residues, sending a signal between residues in separate components would likely be infeasible.

Looking at some basic network characteristics further emphasizes the dichotomy in the fullness of the networks. The clustering coefficient of a network characterizes the level of connected triangles present.[58] A signal being sent between two residues in a network rich with triangles may have a variety of potential pathways to follow, as seen in the MD sectors SHMCs (Figure 8b).

As made apparent in Figure 8c, there is a stark contrast between the average number of neighbors in MD sectors SMHCs and the other SMHCs. An increased average number of neighbors further suggests the presence of viable pathways to transduce signals. Finding groups of residues in which signals can be sent is precisely the goal of running MD sectors analysis.

https://doi.org/10.1021/acsomega.2c05635
ACS Omega 2023, 8, 571−587

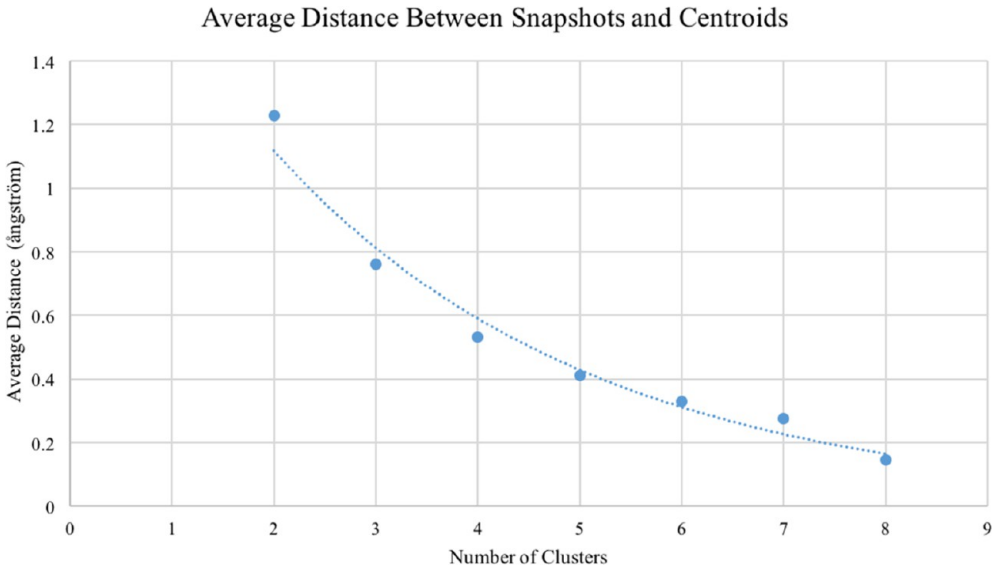


Figure 10. Average Centroid Distance. This graph displays the average distance between snapshots and their corresponding centroids, calculated using rmsd values. The average distance for a variety of number of clusters is shown to help optimize the number of clusters chosen.

The findings of analyzing these SHMCs using network properties further give credibility to the lab's hypothesis that MD sectors are able to find a functionally significant network of residues capable of playing a role in allosteric signaling.

We will now focus in on the MD sectors SHMCs, specifically that of MDS2.0. Notice that the MDS2.0 SHMC contains all 38 residues originally identified in the sector, making the MDS2.0 SHMC a spanning subgraph of the complete graph containing all MDS2.0 residues. When applying the pairwise cutoff to the MDS2.0, no residues were lost. All residues in the network have high *V*pica values as well as some level of significant pairwise interactions with other sector residues. Note that these residues are not guaranteed to form a fully cooperative network. Rather, we are investigating to what extent a fully cooperative network has been uncovered. We believe an approach rooted in first-principles statistics, such as MD sectors, is a cogent way to find such a network.

Figure 9 displays the overlaps of residues in the MDS2.0 and MDS1.0. The red nodes are in both sectors, the blue nodes are only in MDS2.0, and the green nodes are only in the MDS1.0. Note that in this network, the standard pairwise cutoff of ∼0.61 is still applied. The network once again displays a high level of overlap between the new and old methods.

Another important point this network displays is the robustness of MD sectors as a methodology. Notice that the three nodes which are only in the MDS1.0 are quite small, signaling lower *V*pica values. The largest nodes in this network, on the other hand, tend to be in both sectors. This pattern indicates that truly important residues will be incorporated in the MD sector regardless of slight implementation differences. Human error and researcher discretion are relevant factors in any analysis method; minimizing the effects of these factors preserves the legitimacy of the analysis. The robustness of MD sectors is encouraging in that it leaves room for further improvements to the method without introducing significant risks of invalidating previous findings.

3.3. MD-Based Markov State Models on Mutant p53.

The use of MD-based Markov state models allows us to compare the conformational dynamics of related systems.[51] In this section, we select three residues to focus on based upon their importance relative to the *V*pica analysis and test the hypothesis that they affect the dynamic conformations of p53 by observing their conformational substates in MD-MSM. Using computational alanine mutagenesis in silico, we mutated three specific residues in p53 and simulated a trajectory for each version of mutated p53 as well as the wild-type p53. Computational alanine mutagenesis is a process in which amino acids are mutated from their native identity to alanine, as motivated by the small size of the alanine side chain.

Here, we summarize key features of three residues mutated via computational alanine mutagenesis:

- **R158** has the highest ratio score of any MD sector residue. R158 also happens to be one of six residues which is in the MDS2.0, SCAS, and MSR.
- **L264** has a high *V*pica value and a low average shortest path length and is in the MD sector.
- **L137** is not in the MD sector, so it is being used as a control. It is a medium-sized residue with a middling *V*pica value and ratio score. Additionally, it is a leucine residue, so it offers an analytical comparison to L264.

A network with these residues highlighted is shown in Figure S6, and they are shown in the p53 structure relative to key features and hotspot mutations in Figure 1. The goal of mutating these residues is to observe whether MD sector mutations to p53 have a larger impact on structural behavior than non-MD-sector mutations.

After running the four simulations, *k*-means clustering was applied, testing between two and eight clusters. We observed the average distance between each snapshot and its corresponding centroid (Figure 10). Choosing a number of clusters at this point is not exact; we look for a selection that has a relatively low average distance between snapshots and centroids, but that does not use more clusters than necessary. These cluster number candidates tend to reside around the inflection point of the overlaid exponential curve in Figure 10.

The decision to split the conformational space into five clusters was made after observing the rmsd distribution plots outputted when using five clusters. These plots can be viewed in Figure 11. The distribution of rmsd distances of snapshots in

https://doi.org/10.1021/acsomega.2c05635
ACS Omega 2023, 8, 571−587

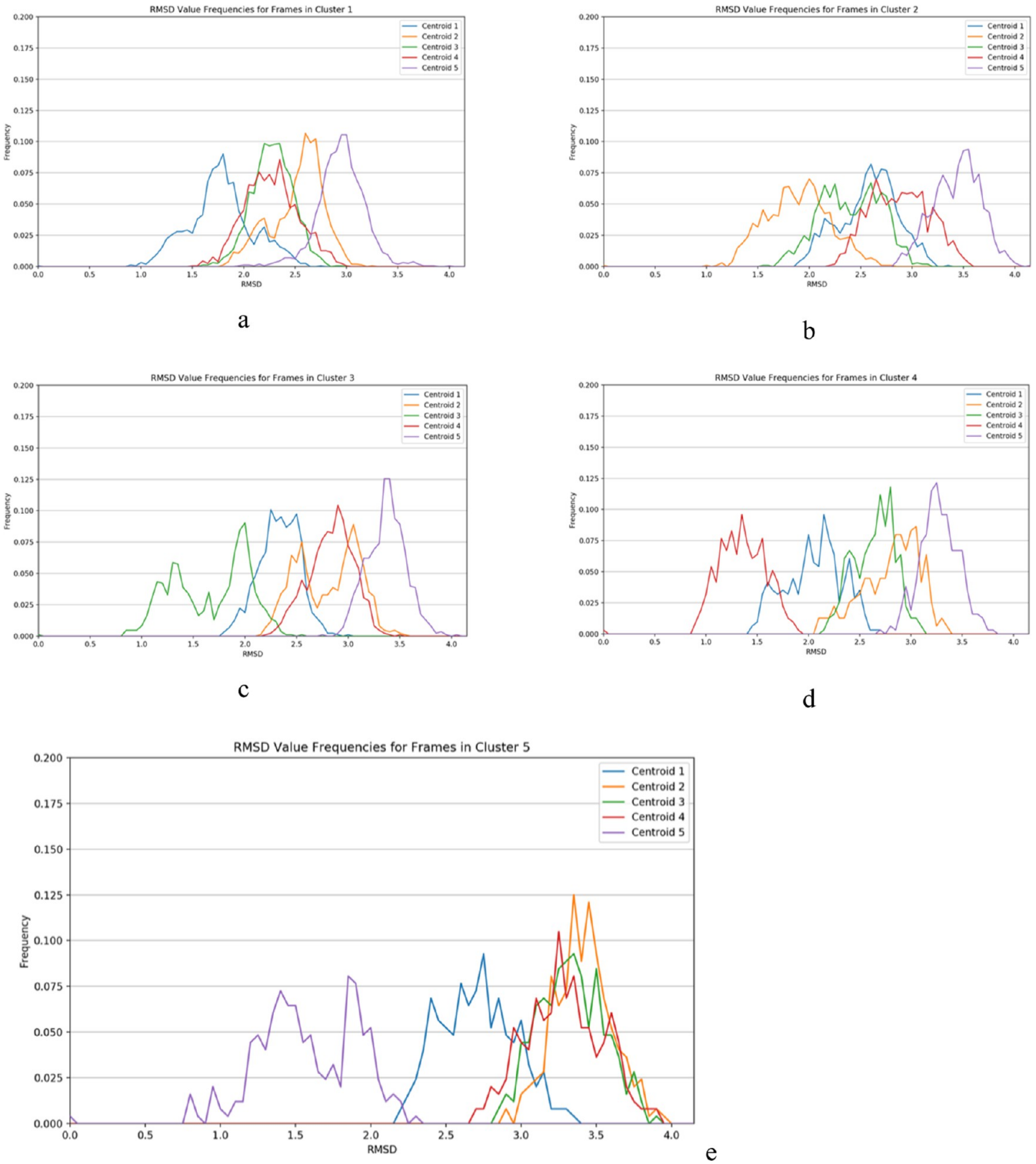


Figure 11. p53 rmsd Plots. The rmsd frequency plot between frames in each cluster and all five centroids. rmsd values to each centroid are split up into their respective distributions. These results come from running four trajectories on various forms of p53. (a) Cluster 1, (b) Cluster 2, (c) Cluster 3, (d) Cluster 4, and (e) Cluster 5.

each cluster is taken with respect to each centroid in turn. An optimal number of clusters is not known but rather is the input to the k-means clustering algorithm. This necessitates some trial and error in establishing a reasonable number of clusters, and we have screened 2 to 8 as possible choices (Figure 10). A desirable feature of an optimally chosen number of clusters is that snapshots are closest to their respective centroid and distant from all others, seen by left shifting of each cluster's distribution curve with its respective centroid in the rmsd frequency plots (Figure 11). This was observed with reasonable separation from other curves, supporting our choice to work with five centroids. While some of the frequency plots contained bimodal distributions (Figure 11c,e), testing with alternate numbers of clusters did not produce better results. A visualization of the five

https://doi.org/10.1021/acsomega.2c05635
ACS Omega 2023, 8, 571−587

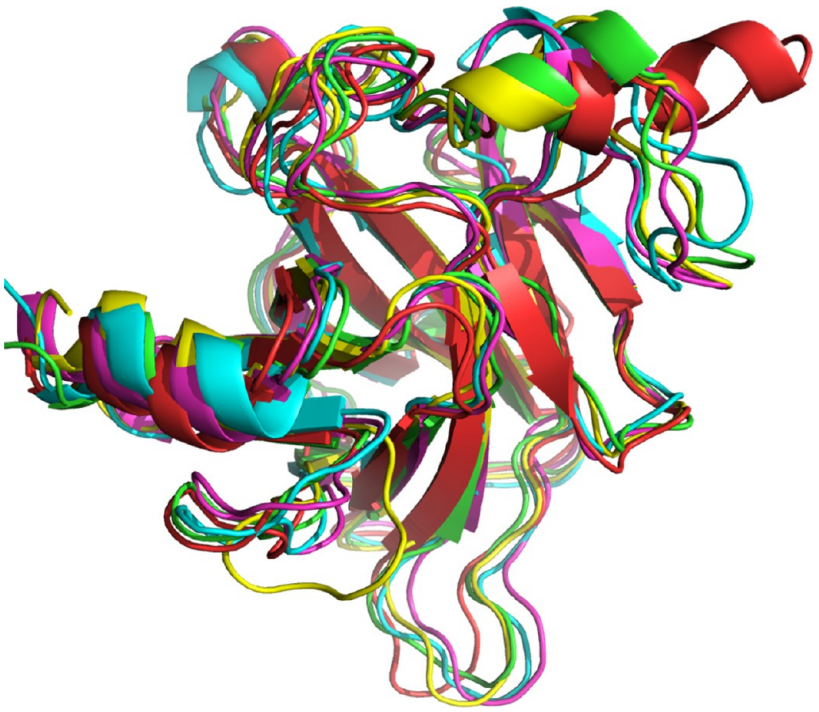


Figure 12. Centroid Overlay. Visualization of the average structures (centroids) found via MD-MSMs on p53 assuming global alignment. Centroid 1 (green), Centroid 2 (cyan), Centroid 3 (magenta), Centroid 4 (yellow), and Centroid 5 (red) are overlaid on one another.

centroid molecular structures overlaid on one another can be seen in Figure 12.

Now, we will examine the substate overlaps of the trajectories from the various forms of p53 (Figure 13). The wt simulation samples two clusters significantly, and one of the clusters is made up entirely of frames from the wt simulation. The other cluster, which is Cluster 1, has over 90% of its frames from two simulations, the wt simulations and the L137 simulation. L137 happens to be the only mutated residue tested that is not in the sector. In other words, the wt simulation had no significant cluster overlaps with any of the sector-mutated simulations, but it did have a significant overlap with a non-sector-mutated simulation. It is of note that a mutation could cause a local unfolding of the protein, thereby causing effects that are not necessarily involved in allosteric linkage, and would not necessarily be a significant enough change to be detected in the substate overlap from a global measure such as rmsd. The trajectories from the L264 and R158 mutations spent the majority of time in clusters which were mostly sampled by themselves.

By making mutations to the above residues, we observed in silico whether sector membership has an effect on the dynamics of the mutated system. The findings hint that when mutating residues within the sector, the conformational dynamics of the protein may be more affected than when mutating residues outside of the sector. The conformational dynamics of the sector-mutated simulations were appreciably different from those of the wild-type and non-sector mutated simulations. If a pathway of residues necessary to the transduction of allosteric signaling is broken by a mutation, we would expect the conformational states in the trajectory to be noticeably impacted. Potentially, such a pathway was broken in these experiments, causing a shift in substate dynamics.

3.4. MD Sectors to Identify Allosteric Points of Control for Drug Design. In light of the fact that allosteric drug design holds excellent promise for creating a novel class of therapeutic drugs, we tackled the most formidable challenge to this approach's success, which is identifying allosteric points of control. Sector residues show great promise in being the conduit by which allosteric signals are conveyed across proteins and are thus a likely target for allosteric drugs. To better understand their functionality, which is heretofore poorly conceptualized, we applied network theory to gain a rigorous analysis of their efficacy. We verified that our MD-MSM models based on motional covariance generate more cohesive networks than SCA networks based on evolution. The approach suggested revising the method with our *V*pica ratio analysis, which enabled us to more accurately measure the cohesion of the DNA binding domain.

The MD sectors allow a systematic and broadly applicable analysis pipeline for the development of allosteric drugs:

1) Run MD simulations on a system of interest
2) Pinpoint candidate residues of functional significance via MD sectors
3) Visualize the sector as a 3-D model and 2-D graph
4) Perform *V*pica ratio analysis on system residues
5) Perform network analysis on the sector graph
6) Apply MD-MSMs on sector-mutated variations of a system to directly assess the dynamic impacts such residues have on the system

The importance of a standardized method cannot be understated. Standardization allows for clear one-to-one comparisons across individual research projects and molecular systems.

A focus of this project was developing techniques for understanding a sector as a graph, or network. Doing so provided novel insight into the physical dynamics of the case study protein, p53. We found that MD sectors fostered networks characterized by high density and the preservation of covarying motional undercurrents. Such properties are presumably

https://doi.org/10.1021/acsomega.2c05635
ACS Omega 2023, 8, 571−587

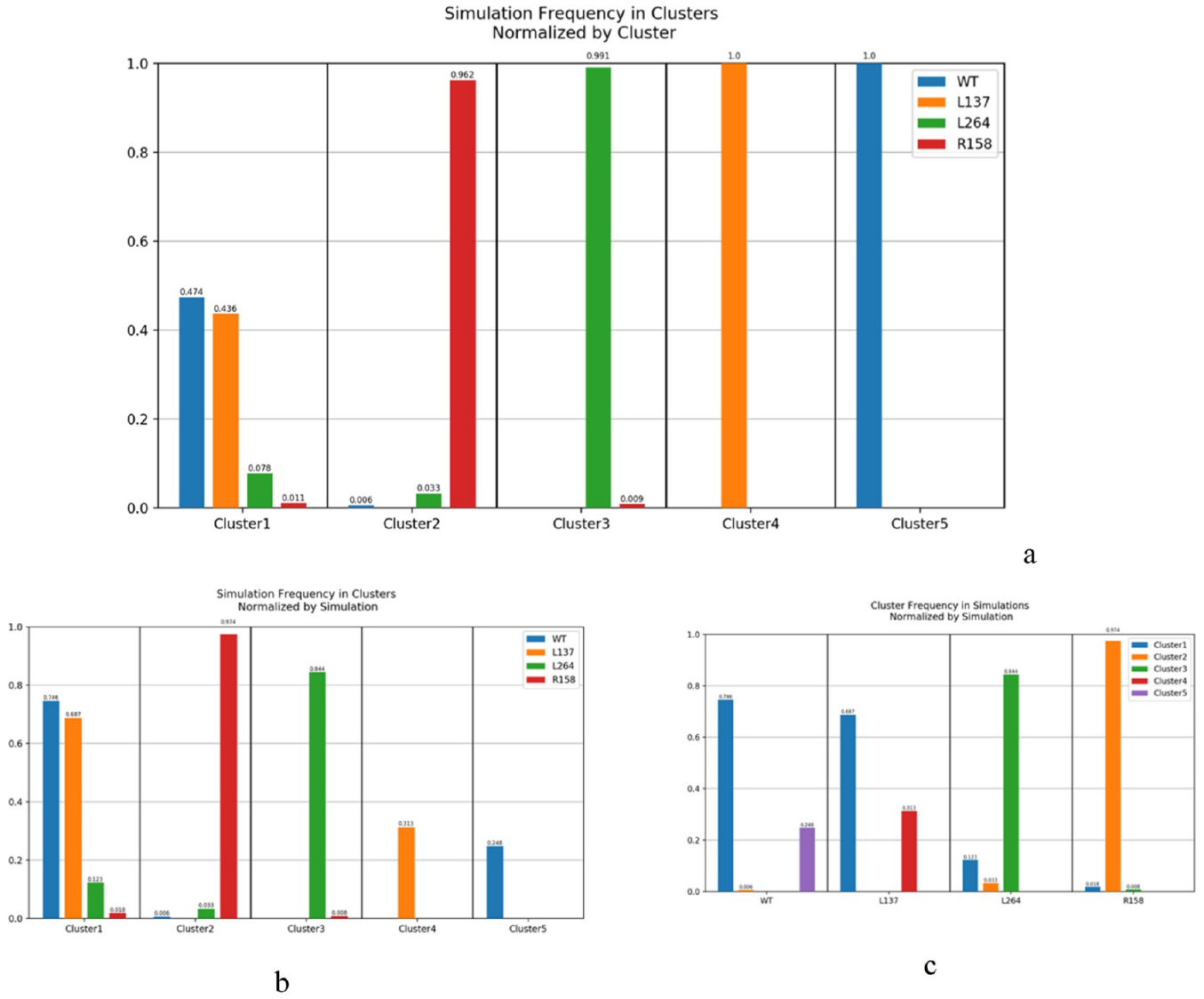


Figure 13. Cluster-Simulation Frequency Bar Charts. (a) Bar chart displaying the frequency of simulation snapshots within each cluster, *normalized by cluster*. The leftmost bar, for example, denotes that 47.4% of snapshots within Cluster 1 come from the wild-type simulation. Notice that over 90% of Cluster 1 is sampled by the wild-type simulation and the L137-mutated simulation. (b) Bar chart displaying the frequency of simulation snapshots within each cluster, *normalized by simulation*. The leftmost bar denotes that 74.6% of snapshots from the wild-type simulation are mapped to Cluster 1. Both the wild-type simulation and the L137-mutated simulation have over 68% of their frames mapped to Cluster 1. (c) Bar chart displaying the frequency of cluster snapshots within each simulation, *normalized by simulation*. The tall green bar indicates that 84.4% of the snapshots from the L264-mutated simulation are within Cluster 3. The L264 and R158-mutated simulations tend to mostly sample clusters by themselves.

inherent to a group of residues capable of transducing allosteric signals via physical means. Furthermore, *V*pica ratio analysis underscored the cohesion of an MD sector, providing an apt comparison of sector and non-sector covariances. Finding dense networks of residues with high levels of cohesion that preserve motional covariance was precisely what we hoped to do with MD sectors. These analysis methods are new in the realm of MD sectors and are only the start for building a stronger understanding of the significance of a sector.

Convincingly, MD sectors is a viable option for ascertaining drug target candidates. Each implementation and analysis step discussed in this paper are small puzzle pieces in the ever-complicated science of drug design. As discussed, MD sectors have proven to be robust and relatively impervious to small changes in implementation details. Thus, the possibility of further improvements to the method remains strong. As MD sectors become widely applied, it will surely undergo further iterations, but each iteration will be strongly linked to the fundamental characteristics found in a sector as observed in this study.

We note that *V*pica analysis with sectors is not unique in analyzing the motions of the protein to identify potentially significant residues. Normal mode analysis (NMA) and principal component analysis (PCA) are two additional methods analyzing the motional dynamics of protein structure. NMA[59] is based upon local harmonic displacements of residues about an equilibrium structure and typically involves an elastic rod interpretation of the structure. PCA[60] considers motional modes displayed as joint probability maps of conformers from which the substate structure is inferred from the projection of motions onto the principal axes of the modes. These methods involve similar but distinct analyses of how the modes and/or correlation of the motions evolve in time. Our method stands out in its roots with the statistical coupling analysis[61] to identify

https://doi.org/10.1021/acsomega.2c05635
ACS Omega 2023, 8, 571−587

the most significantly correlated residues. With the *V*pica analysis, our method is further strengthened to report on a cohesive network, which we believe is the most expedient means of identifying points of allosteric control. These similar methods, however, may provide equally insightful results if applied to identifying networks of residues.

3.5. Further Network Analysis. The network analysis applied to various protein sectors in this work was an exploratory foray into a new perspective of understanding sectors. Hopefully, this approach can be furthered with the standardization of a sophisticated network analysis procedure. A specific addition to our current procedure could be some sort of pathway analysis. The Thayer lab has previously looked at the presence of contiguous pathways within protein systems, including p53.[62] Checking for the presence of contiguous pathways within protein sectors would be a logical addition to our network analysis techniques. Using Euclidean distance measures as edge weights, we could visualize sector pathways and offer another scheme for constructing networks.

3.6. Targeting Sector Residues. Identification of sector residues plays a critical role in the drug design process because it suggests positions of allosteric control. Residues in the sector that also appear at or near the surface of the protein constitute the ideal profile to target with allosteric effectors. Estimating the solvent accessible surface area of a residue can be completed from a given static structure in visualization software such as VMD, and it can be followed over time in a simulation using standard utilities in the AMBER suite's cpptraj. Similar analysis of protein structures has been used to aid in the identification of so-called "cryptic" binding pockets that evolve in the dynamics but otherwise may not be detected from static structures.[63] Once residues have been identified, measures to develop an effector to bind to it emerges as the next challenge, but this is not without precedence; rational drug design has provided many examples and rules of thumb which can guide the systematic development and refinement of molecules with improved affinity and functionality. While allosteric regulator sites may prove to be on flat surfaces rather than within crevices, alternate drug modalities beyond the classic small molecule have been under development. Alpha helical peptides currently offer promising ability to bind to broad, flat surfaces. Developing these with a gambit of natural and artificial residues allows for considerable variability to adapt the drugs to the target area of interest. Stapled varieties of these show promise for aiding in traversing the cell membrane, and several examples have been successfully engineered to date. Our lab is currently working on developing and refining such allosteric drugs.

3.7. Implications for Drug Design. The nature of drug design is speculative; if the specific targets of a potential therapeutic were known a priori, methods like MD sectors would be deemed unnecessary. Of course, this is not the case. Thus, mathematical and computationally intensive algorithms are central to modern rational drug discovery. Taking on a spectrum of perspectives in understanding biomolecular systems allows for novel and innovative strategies used to influence behavior. MD sectors, as part of the lab's research pipeline, employs a valuable cross-disciplinary approach; network analysis, computer simulations, spectral decomposition, clustering analysis, and biological theory all enable this method to be successful.

The impact of MD sectors could prove to be widespread. Having the ability to identify residues of allosteric control through purely computational means would vastly improve the prospects of allosteric drug design. Understanding the network dynamics and physical dynamics of targeting specific residues within a system allows for a more precise and efficient means of manipulating such systems. Allosteric drugs have the potential to be instrumental in curing diseases traditional drugs have been unable to treat because they do not require binding at an active site. This opens up the possibility of drugging protein targets previously deemed 'undruggable', with MD sectors playing a large role in such allosteric drug design.

Starting with a protein composed of hundreds, if not thousands, of amino acids is daunting. MD sectors takes the important first step of narrowing the pool of functionally significant residues. This step often proves to a major bottleneck in drug design. Researchers often have to turn to high-throughput methods, which may constitute screening hundreds of thousands of compounds.[64] Target identification is vital to the drug design process; targets must have a means for functionally manipulating a system safely and efficaciously.[65] Methods such as MD sectors must be developed to advance target identification, which in turn enables the entire allosteric drug design process to progress faster and more effectively. The wide application of MD sectors could prove to help overcome a major limiting factor in allosteric drug design, aiding in a more rapid and economical process of developing innovative therapeutics.

4. CONCLUSIONS

In this work, we successfully analyze MD sectors through the development of a cohesiveness metric, *V*pica ratio analysis. This allows the quantitation of MD sectors' cohesiveness in the biological context, which we have demonstrated in the p53 tumor suppressor protein. This approach will continue to not only inform our own p53 allosteric drug design efforts but also provide the first example of a new paradigm informing rational design of allosteric drugs in general.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.2c05635.

Motion covariance, graphs for network analysis, network analysis, *V*pica distribution, ratio values, and network with highlighted mutants (PDF)

Script for clustering auto_clust.py, script for drawing links draw_all_links_params.py, script for calculating MD sectors MD_Sectors.py, script containing Python utilities p53_util.py, script to write graph edges write_all_edges_excel.py, script to write graph nodes write_all_nodes_excel.py, script for processing p53 edges p53_residue_edges_all.xlsx, and script for processing p53 nodes p53_residue_nodes_all.xlsx (ZIP)

■ AUTHOR INFORMATION

Corresponding Author

Kelly M. Thayer − *Department of Mathematics and Computer Science, Wesleyan University, Middletown, Connecticut 06457, United States; 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

https://doi.org/10.1021/acsomega.2c05635
ACS Omega 2023, 8, 571−587

Author

Jonathan D. Fabry − *Department of Mathematics and Computer Science, Wesleyan University, Middletown, Connecticut 06457, United States; Department of Chemistry, Wesleyan University, Middletown, Connecticut 06457, United States;* Present Address: Mathworks, Inc., 1 Apple Hill Rd., Natick, MA 01760, United States; orcid.org/0000-0003-3110-4630

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.2c05635

Funding

This work was supported by NIH R15 GM128102 to KMT and by NSF grants CNS-0619508 and CNS-0959856 to Wesleyan University for the development of the High Performance Computing Cluster.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors gratefully acknowledge fruitful discussion with the Molecules to Medicine group, especially David L. Beveridge, David R. Langley, and Michael P. Weir. The authors acknowledge commentary on manuscript drafts by Daniel Krizanc and Saray Shai. The authors wish to thank Henk Meij for technical assistance with the High Performance Computing Cluster.

■ REFERENCES

(1) Nussinov, R.; Tsai, C. J.; Jang, H. Allostery, and How to Define and Measure Signal Transduction. *Biophys. Chem.* **2022**, *283*, 106766.

(2) Nussinov, R.; Zhang, M.; Maloney, R.; Liu, Y.; Tsai, C. J.; Jang, H. A. Allosteric Cancer Drivers and Innovative Allosteric Drugs. *J. Mol. Biol.* **2022**, *434*, 167569.

(3) Liu, J.; Nussinov, R. A. An Overview of Its History, Concepts, Methods, and Applications. *PLoS Comput. Biol.* **2016**, *12*, e1004966.

(4) Berezovsky, I. N.; Nussinov, R. Multiscale Allostery: Basic Mechanisms and Versatility in Diagnostics and Drug Design. *J. Mol. Biol.* **2022**, *434*, 167751.

(5) Brown, C. J.; Lain, S.; Verma, C. S.; Fersht, A. R.; Lane, D. P. Awakening Guardian Angels: Drugging the P53 Pathway. *Nat. Rev. Cancer* **2009**, *9*, 862−873.

(6) Tang, S.; Liao, J. C.; Dunn, A. R.; Altman, R. B.; Spudich, J. A.; Schmidt, J. P. Predicting Allosteric Communication in Myosin via a Pathway of Conserved Residues. *J. Mol. Biol.* **2007**, *373*, 1361−1373.

(7) Stolzenberg, S.; Michino, M.; LeVine, M. v.; Weinstein, H.; Shi, L. Computational Approaches to Detect Allosteric Pathways in Transmembrane Molecular Machines. *Biochim Biophys Acta Biomembr* **2016**, *1858*, 1652−1662.

(8) 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.

(9) Cooper, A.; Dryden, D. T. F. Allostery without Conformational Change - A Plausible Model. *European Biophysics Journal* **1984**, *11*, 103−109.

(10) Tsai, C.-J.; Nussinov, R. A Unified View of "How Allostery Works. *PLoS Comput. Biol.* **2014**, *10*, e1003394.

(11) Nussinov, R. Introduction to Protein Ensembles and Allostery. *Chem. Rev.* **2016**, *116*, 6263−6266.

(12) Liu, J.; Nussinov, R. Energetic Redistribution in Allostery to Execute Protein Function. *Proc Natl Acad Sci U S A* **2017**, *114*, 7480−7482.

(13) 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.

(14) Abdel-Magid, A. F. Allosteric Modulators: An Emerging Concept in Drug Discovery. *ACS Med Chem Lett* **2015**, *6*, 104−107.

(15) Sheik Amamuddy, O. S.; Veldman, W.; Manyumwa, C.; Khairallah, A.; Agajanian, S.; Oluyemi, O.; Verkhivker, G. M.; Tastan Bishop, Ö. T. Integrated Computational Approaches and Tools for Allosteric Drug Discovery. *Int. J. Mol. Sci.* **2020**, *21*, 847.

(16) Linzer, D. I. H.; Levine, a. J. Characterization Tumor Antigen and Uninfected of a 54K Dalton Cellular SV40 Present in SV40-Transformed Cells. *Cell* **1979**, *17*, 43−52.

(17) Xue, B.; Brown, C. J.; Dunker, A. K.; Uversky, V. N. Intrinsically Disordered Regions of P53 Family Are Highly Diversified in Evolution. *Biochim Biophys Acta Proteins Proteom* **2013**, *1834*, 725−738.

(18) 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.

(19) Liu, D. P.; Song, H.; Xu, Y. A Common Gain of Function of P53 Cancer Mutants in Inducing Genetic Instability. *Oncogene* **2010**, *29*, 949−956.

(20) Butler, J. S.; Loh, S. N. Structure, Function, and Aggregation of the Zinc-Free Form of the P53 DNA Binding Domain. *Biochemistry* **2003**, *42*, 2396−2403.

(21) Thayer, K. M.; Quinn, T. R. P53 R175H Hydrophobic Patch and H-Bond Reorganization Observed by MD Simulation. *Biopolymers* **2016**, *105*, 176−185.

(22) Smock, R. G.; Rivoire, O.; Russ, W. P.; Swain, J. F.; Leibler, S.; Ranganathan, R.; Gierasch, L. M. An Interdomain Sector Mediating Allostery in Hsp70 Molecular Chaperones. *Mol Syst Biol* **2010**, *6*, 414.

(23) Teşileanu, T.; Colwell, L. J.; Leibler, S. Protein Sectors: Statistical Coupling Analysis versus Conservation. *PLoS Comput. Biol.* **2015**, *11*, e1004091.

(24) Halabi, N.; Rivoire, O.; Leibler, S.; Ranganathan, R. Protein Sectors: Evolutionary Units of Three-Dimensional Structure. *Cell* **2009**, *138*, 774−786.

(25) 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**, *1*, 781.

(26) Chodera, J. D.; Noé, F. Markov State Models of Biomolecular Conformational Dynamics. *Curr. Opin. Struct. Biol.* **2014**, *25*, 135−144.

(27) Han, I. S.; Abramson, D.; Thayer, K. 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.

(28) Jasuja, R.; Spencer, D.; Jayaraj, A.; Peng, L.; Krishna, M.; Lawney, B.; Patel, P.; Jayaram, B.; Thayer, K. M.; Beveridge, D. L.; Bhasin, S. Estradiol Induces Allosteric Coupling and Partitioning of Sex-Hormone-Binding Globulin Monomers among Conformational States. *iScience* **2021**, *24*, 102414.

(29) 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.

(30) Salmon-Ferrer, R.; Goetz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER - Part II: Particle Mesh Ewald. *J. Chem. Theory Comput.* **2013**, *9*, 3878−3888.

(31) Case, D.; Darden, T.; Cheatham, T. I.; Simmerling, C.; Roitberg, A.; Want, J.; Duke, R.; Luo, R.; Roe, D.; Walker, R.; LeGrand, S.; Swails, J.; Cerutti, D.; Kaus, J.; Al, E.*Amber 14 Reference Manual*, 2014.

(32) Götz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C.; 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. I. Generalized Born. *J. Chem. Theory Comput.* **2012**, *8*, 1542−1555.

(33) 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.

(34) Case, D. A.; Berryman, J. T.; Betz, R. M.; Cerutti, D. S.; Cheatham, T. E.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Homeyer, N.; Izadi, S.; Janowski, P.; Kaus, J.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Li, P.; Luchko, T.; Luo, R.; Madej, B.; Mer, K. M.*AMBER 2014*; University of California: San Francisco, 2014.
(35) Jorgensen, W. L. Transferable Intermolecular Potential Functions. Application to Liquid Methanol Including Internal Rotation. *J. Am. Chem. Soc.* **1981**, *103*, 341−345.
(36) 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.
(37) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C.*Comparison of Multiple AMBER Force Fields and Development of Improved Protein Backbone Parameters HHS Public Access*2006, *65*.
(38) Joung, I. S.; Cheatham, T. E. Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *J. Phys. Chem. B* **2008**, *112*, 9020−9041.
(39) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N·log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089.
(40) 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.
(41) 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.
(42) Case, D. A.; Darden, T. A.; Cheatham, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Götz, A. W.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wolf, R. M.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M.-J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Sagui, R.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A.*Amber12*; University of California: San Francisco, 2012.
(43) Jin, G.; Mellor-Crummey, J.; Adhianto, L.; Scherer III, W. N.; Yang, C.Implementation and Performance Evaluation of the HPC Challenge Benchmarks in Coarray Fortran 2.0*Proceedings - 25th IEEE International Parallel and Distributed Processing Symposium*, 2011, pp 1089−1100.
(44) Storti, D.; Yurtoglu, M.*CUDA for Engineers: An Introduction to High-Performance Parallel Computing*; Addison-Wesley Professional: Boston, MA, 2015.
(45) Götz, a W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S. L.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with Amber - Part i: Generalized Born. *J. Chem. Theory Comput.* **2012**, *8*, 1542−1555.
(46) Rueda, M.; Ferrer-Costa, C.; Meyer, T.; Pérez, A.; Camps, J.; Hospital, A.; Gelpí, J. L.; Orozco, M. A Consensus View of Protein Dynamics. *Proc Natl Acad Sci U S A* **2007**, *104*, 796−801.
(47) Zuckerman, D. M. Equilibrium Sampling in Biomolecular Simulations. *Annu Rev Biophys* **2011**, *40*, 41−62.
(48) Rivoire, O.; Reynolds, K. A.; Ranganathan, R. Evolution-Based Functional Decomposition of Proteins. *PLoS Comput. Biol.* **2016**, *12*, No. e1004817.
(49) Reynolds, K. A.; Russ, W. P.; Socolich, M.; Ranganathan, R.; Evolution-Based Design of Proteins*Methods Enzymology*; Elsevier Inc., 2013; Vol. *523*, pp 213−235.
(50) Lakhani, B.; Thayer, K.; Black, E.; Beveridge, D. Spectral Analysis of Molecular Dynamics Simulations on PDZ: MD Sectors. *J. Biomol. Struct. Dyn.* **2019**, *38*, 781.
(51) 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.
(52) Ekins, S.; Waller, C. L.; Bradley, M. P.; Clark, A. M.; Williams, A. J.; Four Disruptive Strategies for Removing Drug Discovery Bottlenecks*Drug Discovery Today*; Elsevier Current Trends March, 2013; Vol. *18*, pp 265−271.
(53) Stumpf, M., Wiuf, C., Eds. In *Statistical and Evolutionary Analysis of Biological Networks*; Imperial College Press: London, 2010.
(54) Novak, L.. In *Hybrid Graph Theory and Network Analsysis*; Gibbins, A., Ed.; Cambridge University Press: New York, 1999.
(55) Rahman, M.. *Basic Graph Theory*, 1st ed.; Springer International Publishing: New York, NY, 2017.
(56) Wilson, R.*Introduction to Graph Theory*; John Wiley & Sons, Inc.: Boston, MA, 1986.
(57) Cytoscape Open Source Newwork Visualization Software.
(58) Masuda, N.; Sakaki, M.; Ezaki, T.; Watanabe, T. Clustering Coefficients for Correlation Networks. *Front Neuroinform* **2018**, *12*, 7.
(59) Gerek, Z. N.; Keskin, O.; Ozkan, S. B. Identification of Specificity and Promiscuity of PDZ Domain Interactions through Their Dynamic Behavior. *Proteins: Struct., Funct., Bioinf.* **2009**, *77*, 796−811.
(60) Shlens, J.*A Tutorial on Principal Component Analysis: Derivation, Discussion and Singular Value Decomposition*2003, *2*, 1−16.
(61) Potters, M.; Bouchaud, J.Chapter 40 Financial Application. In *The Oxford Handbook of Random Matrix Theory*; Akemann, G., Baik, J., FrancescoDi, P., Eds.; Oxford University Press: Oxford, 2011.
(62) Thayer, K. K. M.; Galganov, J. C. J.; Stein, A. J. A. Dependence of Prevalence of Contiguous Pathways in Proteins on Structural Complexity. *PLoS One* **2017**, *12*, No. e0188616.
(63) Horn, J. R.; Shoichet, B. K. Allosteric Inhibition Through Core Disruption. *J. Mol. Biol.* **2004**, *336*, 1283−1291.
(64) Smith, A.Screening for drug discovery: The leading question*Nature*Macmillan Magazines Ltd July, 2002; Vol. *418*, pp 453−455.
(65) Hughes, J. P.; Rees, S. S.; Kalindjian, S. B.; Philpott, K. L. Principles of early drug discovery. *Br J Pharmacol March* **2011**, *162*, 1239−1249.

https://doi.org/10.1021/acsomega.2c05635
ACS Omega 2023, 8, 571−587