ACS OMEGA

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

Article

Mutagenic Activation of Glutathione Peroxidase-4: Approaches toward Rational Design of Allosteric Drugs

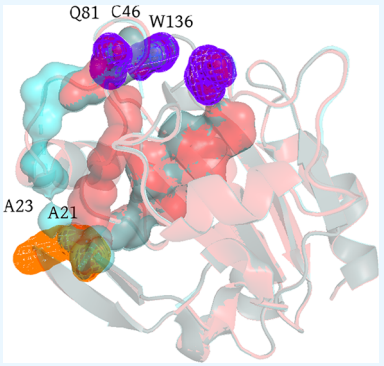
Chunyue Ma,* Daniel J. Chung, Dylan Abramson, David R. Langley, and Kelly M. Thayer

Cite This: *ACS Omega* 2022, 7, 29587−29597

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: Glutathione peroxidase 4 (GPX4) reduces lipid hydroperoxides in lipid membranes, effectively inhibiting iron-dependent cell death or ferroptosis. The upregulation of the enzyme by the mutations at residues D21 and D23 has been suggested to be associated with higher protein activity, which confers more protection against neurodegenerative diseases such as Alzheimer's, Parkinson's, and Huntington's diseases. Therefore, it has become an attractive target for treating and preventing neurodegenerative diseases. However, identifying means of mimicking the beneficial effects of these mutations distant from the active site constitutes a formidable challenge in moving toward therapeutics. In this study, we explore using molecular dynamics simulations to computationally map the conformational and energetic landscape of the wild-type GPX4 protein and three mutant variants to identify the allosteric networks of the enzyme. We present the conformational dynamic profile providing the desired signature behavior of the enzyme. We also discuss the implications of these findings for drug design efforts.



INTRODUCTION

A protein's allosteric behavior is characterized by the modulation of its binding affinity to its ligand as the result of its interaction with a third molecule known as the allosteric effector[1−4] or perturbations to the protein introduced by a point mutation. Allosteric regulation pervades biology as one of the most prominent and significant regulatory mechanisms,[5−7] playing an essential role in the regulation of such diverse functions as transcription activation,[8−10] DNA repair,[11−14] metabolism,[15−19] signal transduction with G-protein coupled receptors,[20−24] neurological function,[25−28] protein kinase regulation,[29−33] regulatory domains of proteins,[34,35] and many more. An allosteric effector may increase or decrease binding affinity for the ligand, making it an allosteric activator or repressor, respectively.[36,37]

The considerable distance between the allosteric effector and the protein active sites raised the question of how the allosteric signal traverses the protein to affect the active site,[5,34,38,39] and several ideas have been put forth.[5,8,21] Many early thoughts on the problem involved a domino effect pathway-like approach in which one residue interacts with the next across a molecule. While this simplistic approach is a helpful zeroth-order model, experimental evidence quickly became too difficult to interpret. A comprehensive description of the allosteric effect may call for characterization of the dynamic interchange of substates.[2,4,40] Such a view has proven difficult to observe in the experiment because capturing it in action requires a time resolved description at the residue level. Ideas of multiple pathways, contributions from evolution, and many other theories have been explored. Cooper and Dryden brought forth the concept of allostery through energetic pathways on a theoretical basis.[41] However, analysis dominated by examining differences in structures has primarily been unsuitable for assessing the validity of these ideas.

With the observation of allosteric effects without significant conformational changes,[42] the idea has become increasingly revitalized. However, asserting a concrete test to determine whether the pathway is correct or not has proven difficult in the absence of a clear observable to measure the allosteric signal propagation. Recently, research efforts have been focusing on combining the power of machine learning and data from molecular dynamics (MD) simulations as a method of studying allosteric effects.[43−45] In addition, network theory is also used to identify the hidden patterns in complex biological systems.[46] Given that the approaches mentioned above provide valuable insights elusive to experiments into the allostery phenomenon, our work aims to describe the dynamics of allostery using network theory and mathematical modeling based on MD simulations.

We selected the glutathione peroxidase 4 (GPX4) system for our inquiries into allostery. It is an enzyme encoded by the human GPX4 gene, and it is known to reduce lipid hydro-

Received: March 3, 2022
Accepted: July 22, 2022
Published: August 16, 2022

https://doi.org/10.1021/acsomega.2c01289
ACS Omega 2022, 7, 29587−29597

peroxides in lipid membranes, effectively inhibiting iron-dependent cell death or ferroptosis.[47−49] It (GPX4) serves as an antioxidant enzyme that reduces the hydroperoxide species, restoring the membrane's integrity. We study the allosteric effect in the GPX4 protein as it illustrates the phenomenon of modulated binding affinity by point mutations. It also serves as a timely target from a human health perspective. Two single-point mutations at residue D21 (GPX4 A) and residue D23 (GPX4 B) as well as a double mutation at both residues (GPX4AB) are all known to activate the GPX4 activity. Specifically, GPX4 A and AB both significantly enhance the protein activity, whereas GPX4 B shows less significant activation.[47] The mutation site residues (green sticks) and active site residues (pink sticks) are shown in Figure 1 (PDB ID: 2OBI).[50]

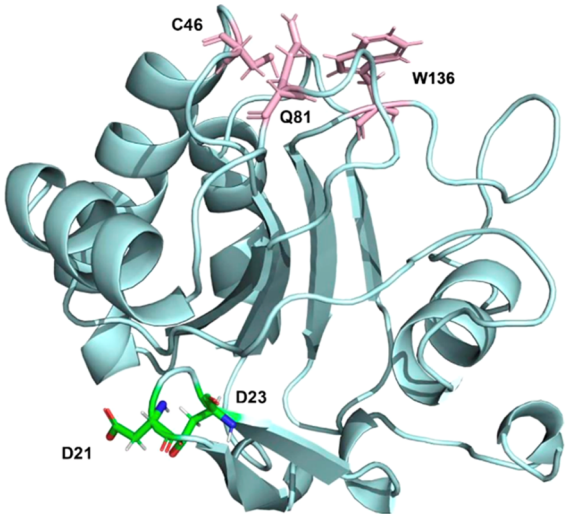


Figure 1. GPX4 protein system: GPX4 structure and engineered mutants used in this study. The wild-type structure (PDB ID: 2OBI) illustrates critical features of the enzyme. The catalytic triad residues collocate three loops to form the active site (pink sticks). Residues D21 and D23 mark the locations of the allosteric active mutations (green sticks).

In this study, we used MD simulations to map the conformational and energetic landscape of the wild-type protein and three mutant variants to identify the allosteric network of the enzyme. We first attempted to explain the hyperactivity of GPX4 mutant variants through changes in their conformations. In addition to root-mean-square deviation (RMSD) and root-mean-square fluctuation (RMSF) analysis, we leveraged conformational dynamics analysis from MD Markov state models (MD-MSMs)[51] when treating the point mutations as allosteric activators. Our research indicated structural differences of a single mutant (GPX4 A) but did not provide sufficient understanding of the particularly strong activating effect of the double mutant (GPX4 AB). We then adopted the energetic approach by comparing the respective energetic networks of various GPX4 mutants. Mapping the network analysis onto the molecule structures, we successfully theorized the dynamic profile associated with hyperactivity of GPX4 that provided the desired signature behavior of the enzyme. Overall, our research demonstrated an exciting possibility that allosteric effects can be considered a superposition of conformational and energetic system changes. We discuss how our insights may be used to develop molecular activators for GPX4 and the prospects for a broader application of this approach for developing therapeutic activators.

RESULTS

RMSD Results. To ensure the convergence of the systems of the simulation, we computed the root-mean square deviation (RMSD) of the backbone atoms of the four constructs concerning their energy-minimized and equilibrated starting structure. We monitored the RMSD as a function of time over the simulation to assess the global stability of the trajectory. An unmodified crystal starting structure will typically deviate about 2 Å from the crystal, and values slightly higher can be expected for engineered systems. We saw that our simulations exhibit an expected amount of dynamic fluctuation on the global scale, with most conformations within about 1.0 Å of their reference when converging over 1 microsecond, as seen in Figure 2.

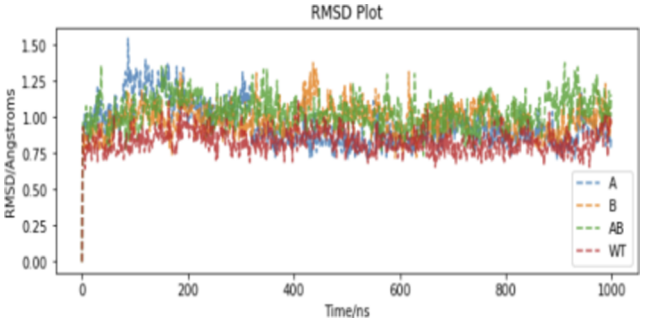


Figure 2. RMSD plot: this plot shows the RMSD deviation, a measure of the average distance between atoms in a molecular structure compared to the energy minimized structure. A snapshot of the protein is taken every 200 ps. This is a 10-point moving average of the RMSD values taken of the backbone C, N, and O atoms in the 160-amino acid system. RMSD values show the similarity between the structures across the course of the simulation.

It is worth mentioning that the RMSD plot from Figure 2 represents only the 5th to the 165th residues of the constructs. This decision was made to remove the tail effect of the first four residues, which caused irregular RMSD behavior and thus affected convergence analysis. In summary, all four systems converged well over the course of the simulation. With the stability of the simulations and apparent convergence of global dynamic properties, we turned our attention to sequence-specific analysis and the detailed characterization of the differences in their dynamic behavior.

RMSF Results. To assess the differences in dynamics by residue, we computed the RMSF of the alpha carbons with the average structure of each simulation as their reference coordinates (Figure 3). GPX4 A experienced larger variations at the residue level with respect to the wild-type structure. Its RMSF plot diverged from the RMSF plot of the wild-type mutant (GPX4 WT) at multiple residues. This indicates structural differences between GPX4 A and GPX4 WT.

Another equally important observation is that the residue mobility for GPX4 B and GPX4 AB, unlike GPX4 A, very closely followed that of GPX4 WT outside the regions of mutation site residues. This lack of structural difference among the three constructs promoted our further investigation into the energetic characteristics of the GPX4 mutations.

MD-MSM Analysis. We applied Molecular Dynamics Markov State Model analysis (MD-MSM)[51,52] to our

https://doi.org/10.1021/acsomega.2c01289
ACS Omega 2022, 7, 29587−29597

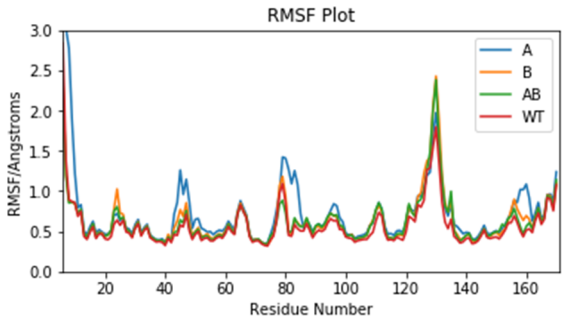


Figure 3. RMSF plot: this plot shows the RMSF of the alpha carbons with the average structure of each simulation as their reference coordinates. RMSF values show the similarity between the systems across the course of the simulation on the residue level.

trajectories to gain more insights into overall dynamic differences. The trajectories of the MD simulations were stripped of ions, water molecules, and side chains to only include the protein backbone atoms. These trajectories were then concatenated into a single long trajectory clustered using a k-means method based on pairwise RMSD values. The categorization of simulation frames into discrete probabilistic states indicates the protein's innate predisposition to adopt specific conformations in the presence or absence of the mutation.

Since the number of clusters was unknown beforehand, the clustering procedure was carried out for two to five clusters. With the increased number of centroids, the distance between snapshots and their respective centroid decreased, as is expected with any parameter fitting calculation. An optimal number of centroids, corresponding to states in the MSM, reduces the error function without overfitting. We chose $K = 3$ for the clustering, confirming the RMSD frequency to centroids distributions.

Because we clustered protein structures from all trajectories simultaneously rather than independently for each trajectory, identified substates were guaranteed to be identical across trajectories. We then assessed the populations in each cluster as the states in the MSM, and the transitions were obtained from the time evolution of the cluster visited by each trajectory. The populations in each substate and the transitions can post priori be associated with the trajectory from which they originated. The MD-MSM provided information about the composition of each cluster by the protein system and frequency of transitions from one substate to the next. To compare the populations in each conformation substate, we showed the frequencies in clusters normalized by simulation as shown in Figure 4.

In Figure 4, we observed that 30.1% of the GPX4 A mutants populated a single substate exclusively in cluster 3, indicating the presence of a distinct conformational state of GPX4 A during the process of simulation. GPX4 WT and GPX4 AB structures populated clusters 1 and 2 with roughly equal proportions. About 54.4% of frames of the GPX WT and 53.1% of frames of the GPX4 AB mutant populated cluster 1, and 45.6% of frames of the GPX WT mutant and 46.8% of the frames of the GPX4 AB mutant inhabited cluster 2. The GPX4 B mutant had an inverse distribution compared to GPX4 AB and GPX4 WT, with 42.6% of the frames in cluster 1 and 56.4% in cluster 2. A small portion of the frames (1.1%) fell into cluster 3. Therefore, GPX4 B had similar but also different conformational states than those of GPX4 AB and GPX4 WT. This hints at a potential explanation for GPX4 B's marginal effect on the protein activity compared to GPX4 AB.

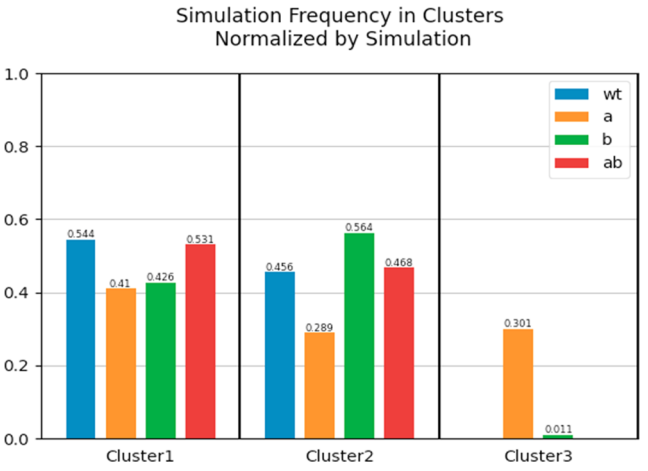


Figure 4. MD-MSM cluster of substates: this histogram shows the contributions of the four GPX4 constructs in each of the three clusters of the MD-MSM. Each of the three clusters (denoted 1−3) is composed of different percentages from each of the four systems.

A more careful look at the clustering of GPX4 A revealed that once GPX4 A adopted a certain conformation in a cluster, it tended to preserve the conformation, as seen from a very high frequency (>70%) of self-transitions (Figure 5). In cluster 3, GPX4 A had a very high self-transition frequency of 96.9%, indicating that it was a very energetically stable structure.

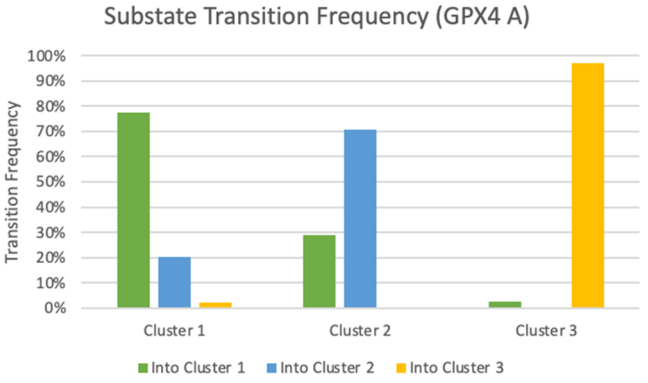


Figure 5. State transition frequency of GPX4 A: this histogram shows the transition frequency of GPX4 A among three different clusters. GPX4 A generally tended to a state within the same conformational state. For cluster 3, GPX4 A had a very high self-transition frequency of 96.9%.

Energetic Network Analysis. We further investigated the energetic differences of the GPX4 systems by comparing the energetic networks constructed with MD simulation data. Energetic networks are created with nodes representing the molecules' residues and edges representing the forces between the pair of residues. Specifically, the edges between each pair of residues are generated as the sum of all forces between pairs of atoms in each residue. Overall, pairwise energetic networks capture the energy profile of the molecule over the course of the simulation, with edge weights representing how much two residues energetically interact with each other.

We built separate energetic networks using both the electrostatic forces and Van der Waals forces among the residues. However, the results from Van der Waals energetic networks among the structures did not produce noticeable

https://doi.org/10.1021/acsomega.2c01289
ACS Omega 2022, 7, 29587−29597

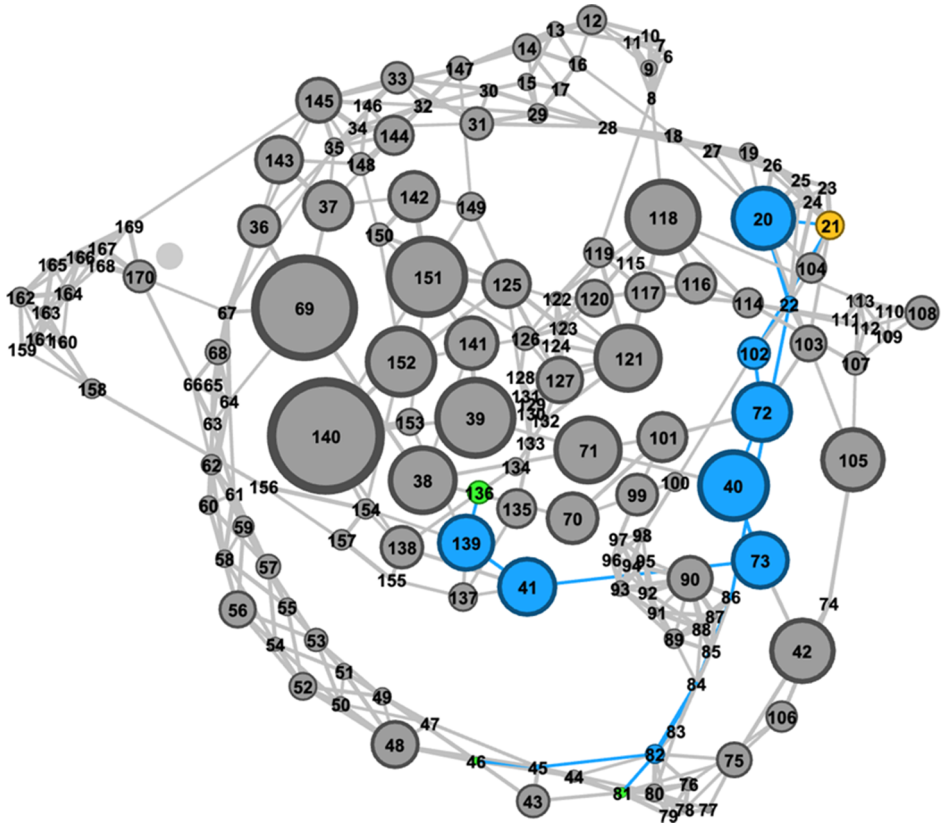


Figure 6. SEPs from D21 of the GPX4 WT network: The SEP between D21 (yellow) and the three active site residues (green) are [21-20-22-86-84-83-81], [21-20-22-86-84-82-45-46], [21-102-72-40-73-41-139-136] respectively. This network shows the top 3.5% of most essential edges.

differences among the structures. We hypothesized that this is due to the short-range nature of such forces given the context of the long-range allosteric effect. Therefore, we focused on comparing the electrostatic energetic network in our study.

The energetic networks were visualized using Gephi software.[52] Since the network was constructed by considering every residue with respect to every other residue, it was a fully connected network. Therefore, to focus on energetically important edges, we kept only 3−5% of the total edges based on their weights. These threshold values were chosen because below 3%, the networks became mostly disconnected. In contrast, beyond 5%, the networks became too densely connected. In both cases, the network was either too disconnected or overconnected for us to record meaningful observations.

Shortest Electrostatic Path Analysis. The shortest electrostatic path (SEP) is a novel approach our lab proposed to examine allosteric effects using an energetic network. We use SEP(i,j) to denote the shortest path from residue i to another residue j on the binarized version of the network.

In this project, we focused on the SEP starting from a mutation site and ending at an active site residue. We hypothesized that the SEPs could model how electrostatic signals propagate from the mutation sites (allosteric sites) to the active sites with minimum signal loss. Following this conjecture, shorter SEPs potentially indicate a more robust electrostatic signaling mechanism between the source and destination residues.

Figures 6 and 7 show SEPs from residue 21 to all three active sites of the protein molecule with 3.5% of total edges kept in the network. As seen from the figures, SEP(21, 46) and SEP(21, 81) were both shorter in GPX4 AB, and SEPs(21, 136) were identical in both GPX4 AB and GPX4 WT. Similar observations persisted among all the SEPs we compared. We observed that all the SEPs of GPX4 AB were shorter than or equal to the ones identified in GPX4 WT. In total, GPX4 AB had the same SEPs as GPX4 WT roughly 70% of the time. In the other 30% of the comparisons, GPX4 AB had shorter SEPs than GPX4 WT (Figure 8).

We also observed that the SEPs of GPX4 AB tended to have more variations than those of GPX4 WT. Figures 9 and 10 show SEPs from residue 23 to all three active sites of the protein molecule with 3.5% of total edges kept in the network. For SEPs starting from residue 23, three distinct paths were identified in GPX4 AB, whereas only two distinct paths could be mapped in GPX WT.

To verify the structural implications of the network analysis, we mapped all the residues participating in the SEPs back onto the molecules using PyMol.[53] From Figures 11 and 12 below, we observed that in GPX4 AB, the residues in the SEPs form three main conduction pathways for electrostatic signals to propagate from mutation sites to the active sites. However, in GPX4 WT, we only identified two such conduction pathways.

The two conduction pathways on the left and right in both GPX4 AB and GPX WT correspond to almost the same set of residues, which reflects the 70% of the SEPs of equal length in both systems shown in Figure 7. However, the middle conduction pathway is unique to GPX4 AB, and it is mapped to the 30% of the SEPs found shorter in GPX4 AB. These residues also only participate in the GPX4 AB network's SEPs. From the energetic point of view, we hypothesize that GPX4 AB

https://doi.org/10.1021/acsomega.2c01289
ACS Omega 2022, 7, 29587−29597

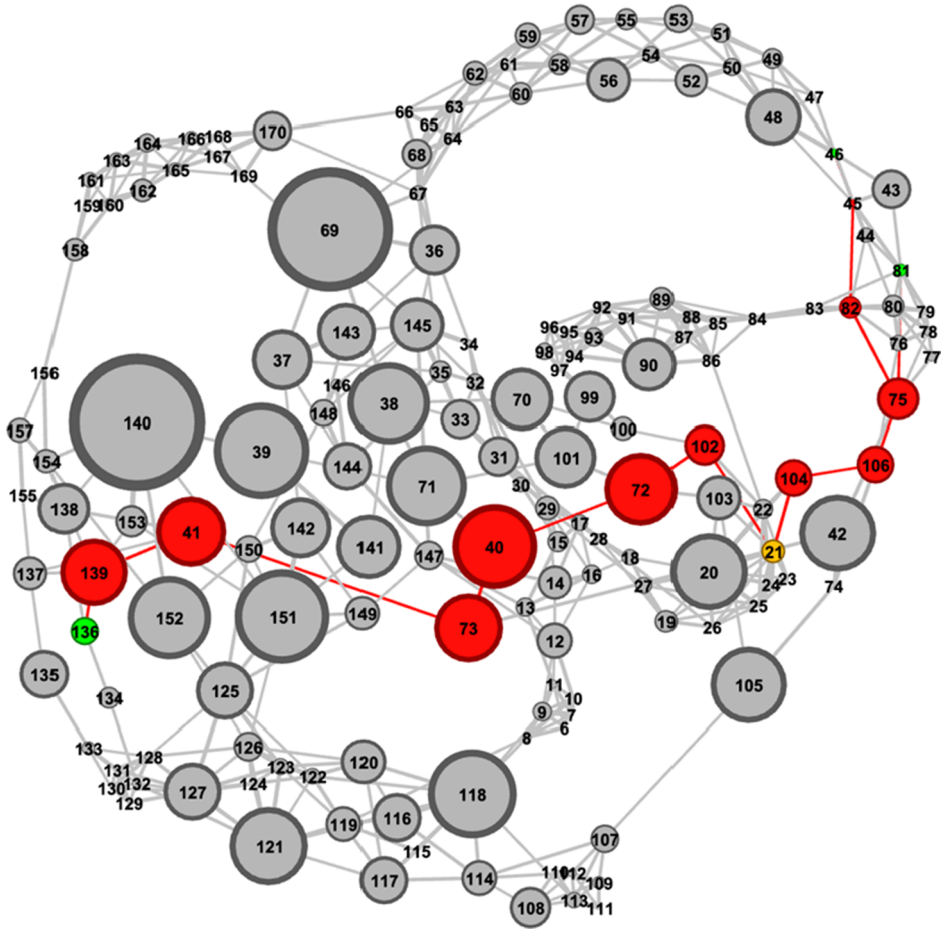


Figure 7. SEPs from D21A of the GPX4 AB network: the SEPs between D21A (yellow) and the three active site residues (green) are [21-104-106-75-82-45-46]/[21-104-106-75-81]/[21-102-72-40-73-41-139-136], respectively. This network shows the top 3.5% of most essential edges. Note that the first two SEPs identified in the GPX4 AB network were shorter than the corresponding SEPs from GPX4 WT, and the third SEP in the GPX4 AB network had the same length as that in the GPC4 WT network.

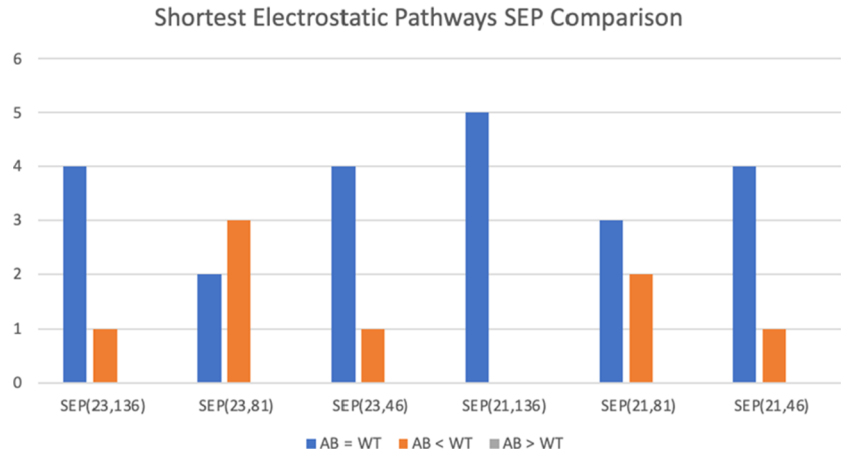


Figure 8. SEP comparisons between GPXWT and GPX4 AB: in all cases, the SEP is either shorter for GPX AB or the same length for both GPX4 AB and GPX4 WT. We did not observe any case when GPX4 WT had a shorter SEP than that of GPX4 AB.

has more conduction pathways with a wider variety of residues involved in its propagation networks.

DISCUSSION

Structural Analysis. The RMSDs were calculated for all the frames in the simulations against their respective energy minimized structure where all four systems converged at roughly 1 Å (Figure 2). Compared to the RMSD plot of the wild-type GPX4 (GPX4 WT), GPX4 A experienced the most significant variation, implying specific structural differences in its molecule structures. On the other hand, GPX4 AB consistently showed higher RMSD values, suggesting higher dispersion of its structure during the simulation. This foreshadows our further

https://doi.org/10.1021/acsomega.2c01289
ACS Omega 2022, 7, 29587−29597

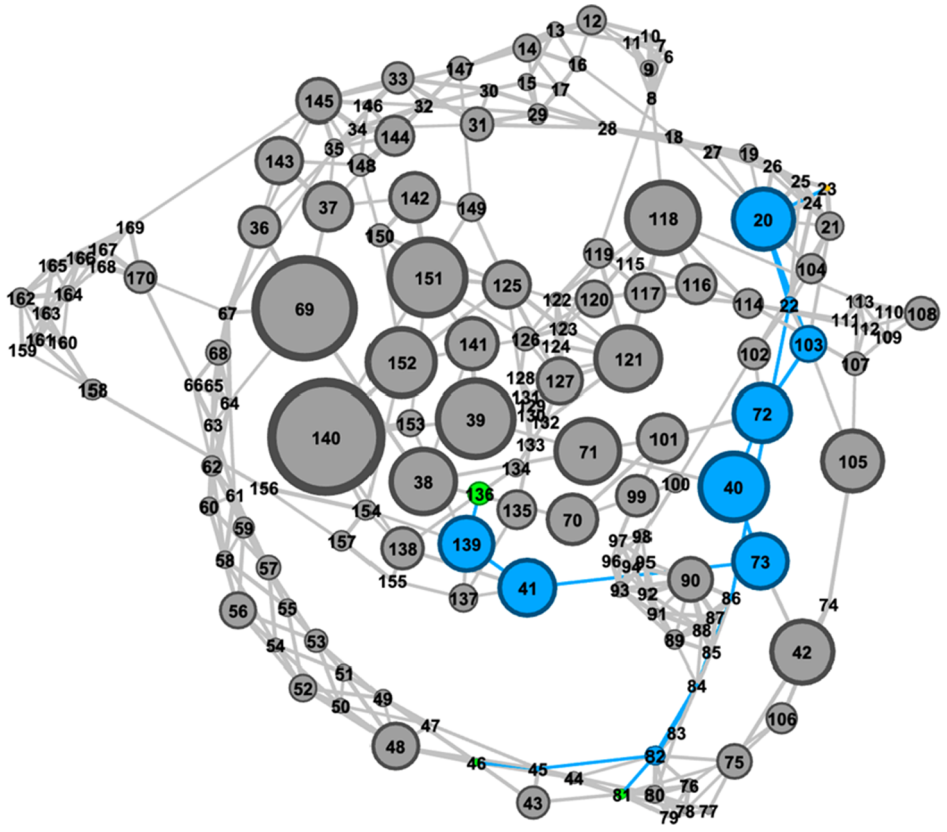


Figure 9. SEPs from D23 of the GPX4 WT network: the SEP between D21 (yellow) and the three active site residues (green). Two distinct paths can be found. SEP(23, 46) and SEP(23, 81) have significant parts of their pathways overlapped with each other.

investigation of the energetic differences between GPX4 AB and GPX4 WT.

The RMSFs represent the positional differences by residues between the structures over time. It calculates the square root of the sum of the Euclidean distance between each residue in the current snapshot with respect to a reference state taken as the energy-minimized conformation. From the RMSF plot of the four systems, we can see the differences in residue flexibility between GPX4 and the other three GPX4 systems (Figure 3). This suggests that GPX4 A is structurally different from GPX WT and further supports the possibility that GPX4 A has a better conformational state, which corresponds to higher protein activity. However, it must be noted that GPX4 B, GPX4 AB, and GPX4 WT all have the RMSF plots closely following each other, indicating a lack of significant structural differences.

Another strategy adopted to examine the structural differentiation between GPX4 systems was MD-MSM analysis. It studies the shift in frequency of the conformational substates of the four GPX4 systems throughout the simulation. A three-cluster model was chosen, with each snapshot of the simulation categorized by its likeness to one of the three centroids. By comparing the snapshot composition of each cluster to empirical activation studies determined in the previous mutagenesis experiment, we characterized cluster 3 as an activated conformational state for GPX4 A. We interpreted that the cluster, which contained about 30% of the total frames of GPX4 A simulation, corresponded to the binding competent conformation of GPX4 A. In addition, the low probability of state transitions from cluster 3 to other clusters and high self-transitioning rate lend further support to the argument that the conformation in cluster 3 was energetically favorable. This analysis shows that the MD-MSM can capture the effect of these "allosteric mutations" on the conformational landscape of GPX4. The self-transition probabilities provide thermodynamic information on the stability of each of the substates. In contrast, the substate transitions allow for the relative comparisons of the activation barriers across these states. Overall, the structural analysis on the mutant systems suggests that GPX4 A increased the protein activity by adopting a more favorable conformation that tends to be preserved.

Given that changes in mutated enzyme activity reported from the experiments were relatively small (e.g., from 100 to 126% for GPX4 A), we also considered the level of resolution of the MD-MSMs approach. In our pilot study of PDZ,[51] we enjoyed good quantitative agreement with NMR data, suggesting a level of resolution appropriate for the applications herein. Our ability to achieve qualitative agreement with the experimental observations validates qualitative interpretation, which allowed us to gain key structural insight into the constructs and conformations for the informed generation of experimentally testable hypotheses. We expect this to be valuable in the iterative process of drug development toward the goal of modulating the GPX4 activity as well.

However, little evidence emerged to construct a coherent argument for the empirical results that show that GPX4 AB had the most substantial activating effect. As we can observe from the similar RMSF plot for both GPX4 AB and GPX4 WT and the almost identical clustering of these two systems in the MD-MSM analysis, GPX4 AB had a very similar conformation with that of the GPX4 WT. To explain the activation of protein without noticeable structural changes, we turned to the energetic network analysis of the systems.

https://doi.org/10.1021/acsomega.2c01289
ACS Omega 2022, 7, 29587−29597

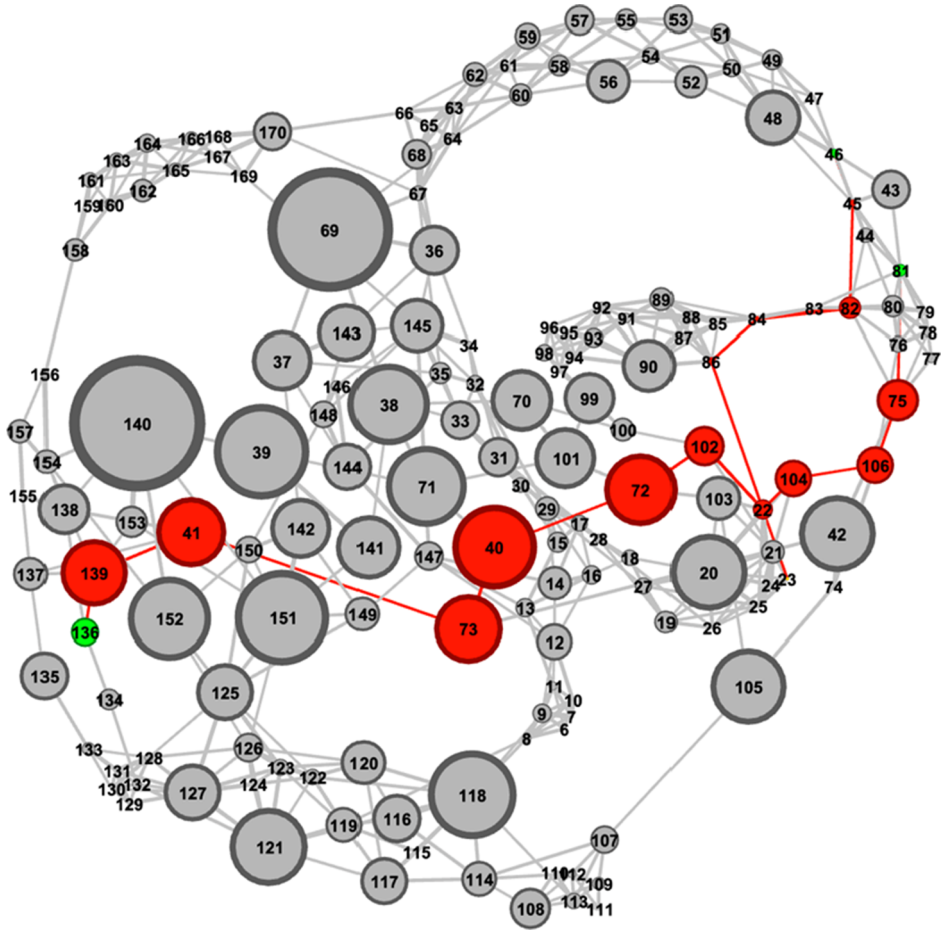


Figure 10. SEPs from D23A of the GPX4 AB network: the SEP between D21 (yellow) and the three active site residues (green). Three, instead of two, distinct paths can be found.

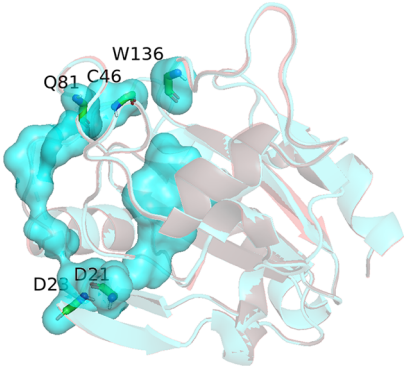


Figure 11. Conduction pathways of GPX4 WT: this graph shows the surface representation of residues participating in the SEPs. In total, the residues identified form two distinct conduction pathways from D21 and D23 to the active sites of the protein molecule.

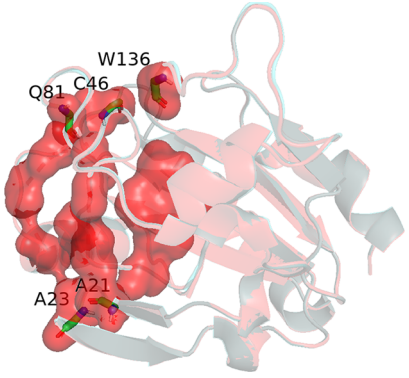


Figure 12. Conduction pathways of GPX4 AB and GPX4 WT: this graph shows the surface representation of residues participating in the SEPs. In total, the residues identified form distinct three conduction pathways from D21A and D23A to the active sites of the protein molecule.

Energetic Network Analysis. Our research tried to identify the differences between mutant AB and the wild-type structure at the energetic level through energetic network analysis. Energetic networks are a relevant new concept connecting biological science with computational science. It has recently been more widely applied to study specific problems of biological interest due to its ability to represent the dynamics of the molecule structures.[46] We constructed the energetics networks of the GPX4 systems to reflect the electrostatic forces of interactions between residues. Given our interest in allosteric signaling, such networks are beneficial as they potentially highlight the conduction pathways on which allosteric signals can propagate.

The most important observation from the electrostatic networks analysis is the improved SEPs in the mutant AB compared to those in the wild-type structure. Holding other variables constant, about one-third of the SEPs we examined were shorter in mutant AB than in the wild-type structure. It is

https://doi.org/10.1021/acsomega.2c01289
ACS Omega 2022, 7, 29587−29597

important to reiterate that shorter SEPs indicate faster conduction and less loss of electrostatic signals.

Since we normalized each edge based on its degree, we had what is known as a transition matrix T such that $T(i,j)$ gives the probability that a random walker starting at node i will transition to node j. Therefore, after thresholding the graph, we only kept the most statistically likely transitions between nodes i and j in a random walk process. Since a random walk process is related to the vibrational dynamics of a network,[54] we believe that the shorter paths could potentially correspond to an improved ability for a perturbation from the mutation site residues to affect the binding site residue activity.

Bearing this in mind, our results provide insights into the seemingly contradictory results between lack of structural differences and empirical evidence of improved protein activity of mutant AB. Despite the lack of positional changes in mutant AB, it can potentially enhance the protein activity by having a better communication channel for the long-range allosteric interactions that can be detected at the energetic level.

Structural Implications of Energetic Networks. The observations from energetic networks can also be corroborated by examining the structural implications. Without any prior assumption of the structures, we could nicely map the SEPs onto structurally plausible conduction pathways for allosteric signals to propagate from the mutation sites to the active sites of GPX4 protein through the medium of electrostatic forces of interaction.

We only identified two conduction pathways from residues D21 and D23 to all the three active site residues in the wild-type structure. In the case of the double-mutant GPX4 AB, we discovered another communication channel. The left and right channels in mutant AB were identical to those of the wild-type structures. However, the middle conduction pathway was unique to mutant AB and corresponded to the residues participating in one-third of the shorter SEPs identified for mutant AB. Hence, this supports our hypothesis that mutant AB improves the protein activity by transmitting electrostatic signals better, faster, and with less loss of information. It also suggests that allosteric signaling in molecules can occur in multiple ways.

Overall, the results indicated that the combined effect of both positional and energetic pathways contribute to the overall allosteric development of the proteins. These combined effects could mean that spatially related allosteric mechanisms can exist alongside energetic pathways, thus providing redundancy in the signaling. Alternatively, alternative means could be accentuated or dampened with specific triggers.

Exploring these ideas is only just beginning, and the approaches from this paper suggest a potential pathway forward to gaining deeper insights into both the coordinate-based and energetics-based allosteric signaling. This may provide needed insights into understanding how allosteric therapeutics may operate. In the absence of a theoretical framework for the operation of allosteric molecules, the design of allosteric therapeutics has mainly been limited and mostly occurred through serendipitous discoveries. Having tools to guide the allosteric design of drugs will open the possibility of designing a new class of drugs, opening the possibility of treating previously undruggable targets.

MATERIALS AND METHODS

Simulation Specifications and Trajectory Analysis. GPX4 starting configurations of the wild-type systems were obtained from the human GPX4 U41C mutant (PDBID# 2OBI).[50] Three activated systems were generated from this crystal structure: two single mutants (D21A and D23A) and one double mutant combining both mutations, all of which are known to enhance the GPX4 activity.[47] Standard lab procedures were adopted to conduct the MD simulations. Energy minimization with decreasing constraints on the protein solute was followed by heating to 300 K, and the temperature was maintained using the Berendsen algorithm. Using explicit counterions and the TIP3P water model, an all-atom 1000 ns (1 ms) MD simulation was executed for each equilibrated GPX4 mutant system in AMBER 16[43,44,55,56] and the CUDA versions of the pmemd routine running on NVIDIA GPUs were parallelized.[45,57] The parm99SB force field[55] was used for proteins and peptides, while the TIP3P potential[58] was used for water. To achieve electroneutrality, minimal salt was added to each simulation cell to electroneutrality. Stability and convergence of the MD were monitored by a standard protocol, such as RMSD. The trajectories were analyzed with AMBER utility CPPTRAJ[59] and the molecular visualization programs PyMol[53] and visual MD.[60,61] The zinc coordination parameterization was achieved using the 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+ counterions were added to the system for electroneutrality, and SHAKE was applied for hydrogen bond motions.

MD Markov State Models. The MD-MSM study of GPX4 is a statistically driven illustration of the difference among the GPX4 mutant activities due to changes at the structural level. It is based on the shift in frequency of the conformational substates of the four GPX4 systems throughout the simulation. The MD-MSM procedure has been described in detail. In summary, the trajectories of the MD simulations were stripped of ions, water molecules, and side chains to only include the protein backbone atoms. These trajectories were then concatenated into a single long trajectory clustered using a K-means clustering method[62] based on pairwise RMSD values implemented in CPPTRAJ. The categorization of simulation frames into discrete probabilistic states indicates the protein's innate predisposition to adopt specific conformations in the presence or absence of the mutation. After the k-means clustering, we identified the substates of the GPX4 mutants and then compared them to analyze the extent to which the wild type differs structurally from the other allosterically liganded protein systems with more active protein activities. MSMs were constructed in terms of the nodes and links of a complex network, with the nodes obtained by clustering the microstates using the K-means clustering algorithm. The links were defined from the Chapman−Kolmogorov-compliant[63] frequency of direct transitions between nodes. In all calculations, atom-based quantities obtained from MD were merged to present results by each residue.

Pairwise Interaction Energy Network. The pairwise interaction energy network is a method adopted to create representations of protein systems at the energetic level. Such networks have nodes representing the residues of the molecules and edges representing the electrostatic/Van der Waals forces between the pair of residues. Every node in the network will have the node degree equal to $n - 1$, where n denotes the number of residues in the molecule. The edges between each pair of residues represent the sum of all electrostatic potentials between pairs of atoms in each residue. We took the absolute value of each of these interactions. Depending on the strength of the force computed, different edges will have varying edge weights,

https://doi.org/10.1021/acsomega.2c01289
ACS Omega 2022, 7, 29587−29597

with higher edge weights representing a higher degree of energetic interactions between residues.

Since computing the pairwise interaction energy for every frame in the trajectory is computationally intensive, we only added pairwise energies for a selected subset of frames from each trajectory. For each 1 ms-long trajectory with 5000 frames, we extracted one frame out of every ten frames starting from the first frame with a regular interval. Collected samples were then concatenated into a new trajectory that serves as the 'trajectory's lightweight representation. We then fed the trajectories through the energy analysis protocol from CPPTRAJ. We computed the energy for every 165 residues and chose two combinations of residue pairs by using a bash script that performed the CPPTRAJ energy command with each pair of residues. Subsequently, Python scripts were written to parse the CPPTRAJ outputs into tensors of shape (N, X, D, D) where N was the number of frames in the trajectory, X is the number of types of energy being studied, and D is the residue number. To compute a representative energy network of the systems, we simply summed over each energy per energy channel. Finally, the network edges were assigned thresholds by normalizing the edge weights on a per-node basis and picking all nodes above that threshold.

Using the open-source program Gephi, we created visualizations of all four networks for electrostatics and van der Waals forces. We focused on examining the visualization of the electrostatic energy network due to its long-range nature that is more closely associated with the allosteric effect happening over a longer distance in the GPX4 protein system. Since the network created is wholly connected, we first filtered the edges in the network based on edge weight to prepare the network for visualization. This step allowed us to only look at energetically essential edges in the system. To further understand the importance of residues, we also denoted the nodes with an average weighted degree. In the images, as shown in Figures 6 and 7, the more prominent nodes are associated with residues that are more energetically important, and more opaque edges represent higher weight.

In general, the drawing of a graph will have no relation to the underlying dynamics of the system as there is no notion of distance on structured chart data. However, these images were generated using the ForceAtlas2 layout algorithm in Gephi, in which nodes are represented as entities repulsing like-charged particles while edges attract their nodes like springs. Therefore, these images highlight some of the essential dynamics in allosteric signaling and thus can offer us good intuition regarding the types of properties on which future experiments may focus.

Shortest Electrostatic Pathways. The SEP is a novel approach our lab proposed to examine allosteric effects in the context of an energetic network. It refers to the shortest path from the mutation site residues to the active site residues on a given network. We examined six SEPs in total at every network snapshot, from two mutation site nodes (allosteric site nodes) to three active site nodes. We selected network snapshots with 2−5% of the edges kept and compared the SEPs between mutated and non-mutated system networks. We hypothesized that the SEPs potentially reflected the most efficient pathways that electrostatic signals propagate from the mutation sites (allosteric sites) to the active sites with minimum signal loss. In the context of the allosteric effect, this would provide clues on how allosteric products work through electrostatic signaling.

AUTHOR INFORMATION

Corresponding Author

Chunyue Ma − *Department of Mathematics & Computer Science, Wesleyan University, Middletown, Connecticut 06459, United States;* orcid.org/0000-0002-9487-6032; Email: cma01@wesleyan.edu

Authors

Daniel J. Chung − *Department of Chemistry and Molecular Biophysics Program, Wesleyan University, Middletown, Connecticut 06459, United States*

Dylan Abramson − *Department of Mathematics & Computer Science, Wesleyan University, Middletown, Connecticut 06459, United States*

David R. Langley − *Department of Chemistry and Molecular Biophysics Program, Wesleyan University, Middletown, Connecticut 06459, United States; Arvinas Inc., New Haven, Connecticut 06511, United States*

Kelly M. Thayer − *Department of Mathematics & Computer Science, Department of Chemistry, and Molecular Biophysics Program, Wesleyan University, Middletown, Connecticut 06459, United States;* orcid.org/0000-0001-7437-9517

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.2c01289

Author Contributions

The manuscript was written through the contributions of all authors.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We would like to thank our Molecules to Medicine consortium colleagues for fruitful discussions, especially David Beveridge, Michael Weir, Ben Cowan, Abhilash Jayaraj, and Mark Han. We would also like to thank Henk Meij for HPCC's technical support. This work was supported by the NIH grant R15 GM128102-01 to K.M.T. The Wesleyan high-performance computing facilities are sponsored by the NSF grant CNS-0959856 and CNS-095985.

REFERENCES

(1) Guarnera, E.; Berezovsky, I. N. Allosteric Sites: Remote Control in Regulation of Protein Activity. *Curr. Opin. Struct. Biol.* **2016**, *37*, 1−8.
(2) Roberts, G. The Role of Protein Dynamics in Allosteric Effects—Introduction. *Biophys. Rev.* **2015**, *7*, 161−163.
(3) Lu, S.; Jang, H.; Muratcioglu, S.; Gursoy, A.; Keskin, O.; Nussinov, R.; Zhang, J. Ras Conformational Ensembles, Allostery, and Signaling. *Chem. Rev.* **2016**, *116*, 6607−6665.
(4) Henzler-Wildman, K.; Kern, D. Dynamic Personalities of Proteins. *Nature* **2007**, *450*, 964−972.
(5) Tsai, C.-J.; Nussinov, R. A Unified View of "How Allostery Works". *PLoS Comput. Biol.* **2014**, *10*, No. e1003394.
(6) Eisenberg, D.; Marcotte, E. M.; Xenarios, I.; Yeates, T. O. Protein Function in the Post-Genomic Era. *Nature* **2000**, *405*, 823−826.
(7) Selvaratnam, R.; Chowdhury, S.; VanSchouwen, B.; Melacini, G. Mapping Allostery through the Covariance Analysis of NMR Chemical Shifts. *Proc. Natl. Acad. Sci.* **2011**, *108*, 6133−6138.
(8) McLeish, T. C. B.; Cann, M. J.; Rodgers, T. L. Dynamic Transmission of Protein Allostery without Structural Change: Spatial Pathways or Global Modes? *Biophys. J.* **2015**, *109*, 1240−1250.
(9) Garcia-Pino, A.; Balasubramanian, S.; Wyns, L.; Gazit, E.; De Greve, H.; Magnuson, R. D.; Charlier, D.; van Nuland, N. A. J.; Loris, R.

https://doi.org/10.1021/acsomega.2c01289
ACS Omega 2022, 7, 29587−29597

Allostery and Intrinsic Disorder Mediate Transcription Regulation by Conditional Cooperativity. *Cell* **2010**, *142*, 101−111.
(10) Shulman, A. I.; Larson, C.; Mangelsdorf, D. J.; Ranganathan, R. Structural Determinants of Allosteric Ligand Activation in RXR Heterodimers. *Cell* **2004**, *116*, 417−429.
(11) Mukherjee, S.; Law, S. M.; Feig, M. Deciphering the Mismatch Recognition Cycle in MutS and MSH2-MSH6 Using Normal-Mode Analysis. *Biophys. J.* **2009**, *96*, 1707−1720.
(12) Lamers, M. H.; Winterwerp, H. H. K.; Sixma, T. K. The Alternating ATPase Domains of MutS Control DNA Mismatch Repair. *EMBO J.* **2003**, *22*, 746−756.
(13) Lakhani, B.; Thayer, K. M.; Hingorani, M. M.; Beveridge, D. L. Evolutionary Covariance Combined with Molecular Dynamics Predicts a Framework for Allostery in the MutS DNA Mismatch Repair Protein. *J. Phys. Chem. B* **2017**, *121*, 2049−2061.
(14) Pieniazek, S. N.; Hingorani, M. M.; Beveridge, D. L. Dynamical Allosterism in the Mechanism of Action of DNA Mismatch Repair Protein MutS. *Biophys. J.* **2011**, *101*, 1730−1739.
(15) Lewis, M. Allostery and the Lac Operon. *J. Mol. Biol.* **2013**, *425*, 2309−2316.
(16) Reynolds, K. A.; McLaughlin, R. N.; Ranganathan, R. Hot Spots for Allosteric Regulation on Protein Surfaces. *Cell* **2011**, *147*, 1564−1575.
(17) Bhabha, G.; Biel, J. T.; Fraser, J. S. Keep on Moving: Discovering and Perturbing the Conformational Dynamics of Enzymes. *Acc. Chem. Res.* **2015**, *48*, 423−430.
(18) Kallen, J.; Welzenbach, K.; Ramage, P.; Geyl, D.; Kriwacki, R.; Legge, G.; Cottens, S.; Weitz-Schmidt, G.; Hommel, U. Structural Basis for LFA-1 Inhibition upon Lovastatin Binding to the CD11a I-Domain. *J. Mol. Biol.* **1999**, *292*, 1−9.
(19) Sharp, K. A. Allostery in the Lac Operon: Population Selection or Induced Dissociation? *Biophys. Chem.* **2011**, *159*, 66−72.
(20) Feng, Z.; Hu, G.; Ma, S.; Xie, X.-Q. Computational Advances for the Development of Allosteric Modulators and Bitopic Ligands in G Protein-Coupled Receptors. *AAPS J.* **2015**, *17*, 1080−1095.
(21) Gentry, P. R.; Sexton, P. M.; Christopoulos, A. Novel Allosteric Modulators of G Protein-Coupled Receptors. *J. Biol. Chem.* **2015**, *290*, 19478−19488.
(22) Wootten, D.; Christopoulos, A.; Sexton, P. M. Emerging Paradigms in GPCR Allostery: Implications for Drug Discovery. *Nat. Rev. Drug Discovery* **2013**, *12*, 630−644.
(23) Chen, A. N. Y.; Malone, D. T.; Pabreja, K.; Sexton, P. M.; Christopoulos, A.; Canals, M. Detection and Quantification of Allosteric Modulation of Endogenous M4 Muscarinic Acetylcholine Receptor Using Impedance-Based Label-Free Technology in a Neuronal Cell Line. *J. Biomol. Screening* **2015**, *20*, 646−654.
(24) Wu, H.; Wang, C.; Gregory, K. J.; Han, G. W.; Cho, H. P.; Xia, Y.; Niswender, C. M.; Katritch, V.; Meiler, J.; Cherezov, V.; Conn, P. J.; Stevens, R. C. Structure of a Class C GPCR Metabotropic Glutamate Receptor 1 Bound to an Allosteric Modulator. *Science* **2014**, *344*, 58−64.
(25) Taly, A.; Corringer, P.-J.; Guedin, D.; Lestage, P.; Changeux, J.-P. Nicotinic Receptors: Allosteric Transitions and Therapeutic Targets in the Nervous System. *Nat. Rev. Drug Discovery* **2009**, *8*, 733−750.
(26) Wang, Y.; Guo, L.; Jiang, H.; Zheng, L.; Zhang, A.; Zhen, X. Allosteric Modulation of Sigma-1 Receptors Elicits Rapid Antidepressant Activity. *CNS Neurosci. Ther.* **2016**, *22*, 368−377.
(27) Anderson, K. W.; Mast, N.; Hudgens, J. W.; Lin, J. B.; Turko, I. V.; Pikuleva, I. A. Mapping of the Allosteric Site in Cholesterol Hydroxylase CYP46A1 for Efavirenz, a Drug That Stimulates Enzyme Activity. *J. Biol. Chem.* **2016**, *291*, 11876−11886.
(28) Maqbool, M.; Mobashir, M.; Hoda, N. Pivotal Role of Glycogen Synthase Kinase-3: A Therapeutic Target for Alzheimer's Disease. *Eur. J. Med. Chem.* **2016**, *107*, 63−81.
(29) Avrahami, L.; Licht-Murava, A.; Eisenstein, M.; Eldar-Finkelman, H. GSK-3 Inhibition: Achieving Moderate Efficacy with High Selectivity. *Biochim. Biophys. Acta* **2013**, *1834*, 1410−1414.
(30) Bhat, R. V.; Budd Haeberlein, S. L.; Avila, J. Glycogen Synthase Kinase 3: A Drug Target for CNS Therapies. *J. Neurochem.* **2004**, *89*, 1313−1317.
(31) Liu, H. C.; Leu, S. J.; Chuang, D. M. Roles of Glycogen Synthase Kinase-3 in Alzheimer's Disease: From Pathology to Treatment Target. *J. Exp. Clin. Med.* **2012**, *4*, 135−139.
(32) Pargellis, C.; Tong, L.; Churchill, L.; Cirillo, P. F.; Gilmore, T.; Graham, A. G.; Grob, P. M.; Hickey, E. R.; Moss, N.; Pav, S.; Regan, J. Inhibition of P38 MAP Kinase by Utilizing a Novel Allosteric Binding Site. *Nat. Struct. Biol.* **2002**, *9*, 268−272.
(33) Shi, Z.; Resing, K. A.; Ahn, N. G. Networks for the Allosteric Control of Protein Kinases. *Curr. Opin. Struct. Biol.* **2006**, *16*, 686−692.
(34) Gerek, Z. N.; Ozkan, S. B. Change in Allosteric Network Affects Binding Affinities of PDZ Domains: Analysis through Perturbation Response Scanning. *PLoS Comput. Biol.* **2011**, *7*, No. e1002154.
(35) Ho, B. K.; Agard, D. A. Conserved Tertiary Couplings Stabilize Elements in the PDZ Fold, Leading to Characteristic Patterns of Domain Conformational Flexibility. *Protein Sci.* **2010**, *19*, 398−411.
(36) Abdel-Magid, A. F. Allosteric Modulators: An Emerging Concept in Drug Discovery. *ACS Med. Chem. Lett.* **2015**, *6*, 104−107.
(37) Dixit, S. B.; Andrews, D. Q.; Beveridge, D. L. Induced Fit and the Entropy of Structural Adaptation in the Complexation of CAP and Lambda-Repressor with Cognate DNA Sequences. *Biophys. J.* **2005**, *88*, 3147−3157.
(38) Ribeiro, A. A. S. T.; Ortiz, V. A Chemical Perspective on Allostery. *Chem. Rev.* **2016**, *116*, 6488−6502.
(39) Thayer, K. M.; Galganov, J. C.; Stein, A. J. Dependence of Prevalence of Contiguous Pathways in Proteins on Structural Complexity. *PLoS One* **2017**, *12*, No. e0188616.
(40) Kern, D.; Zuiderweg, E. R. P. The Role of Dynamics in Allosteric Regulation. *Curr. Opin. Struct. Biol.* **2003**, *13*, 748−757.
(41) Cooper, A.; Dryden, D. T. Allostery without Conformational Change. A Plausible Model. *Eur. Biophys. J.* **1984**, *11*, 103−109.
(42) Louet, M.; Seifert, C.; Hensen, U.; Gräter, F. Dynamic Allostery of the Catabolite Activator Protein Revealed by Interatomic Forces. *PLoS Comput. Biol.* **2015**, *11*, No. e1004358.
(43) 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.
(44) 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.
(45) Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* **2013**, *9*, 3878−3888.
(46) Mishra, S.; Caflisch, A. Dynamics in the Active Site of β-Secretase: A Network Analysis of Atomistic Simulations. *Biochemistry* **2011**, *50*, 9328−9339.
(47) Li, C.; Deng, X.; Zhang, W.; Xie, X.; Conrad, M.; Liu, Y.; Angeli, J. P. F.; Lai, L. Novel Allosteric Activators for Ferroptosis Regulator Glutathione Peroxidase 4. *J. Med. Chem.* **2019**, *62*, 266−275.
(48) Imai, H.; Matsuoka, M.; Kumagai, T.; Sakamoto, T.; Koumura, T. Lipid Peroxidation-Dependent Cell Death Regulated by GPx4 and Ferroptosis. *Curr. Top. Microbiol. Immunol.* **2017**, *403*, 143−170.
(49) Maiorino, M.; Conrad, M.; Ursini, F. GPx4, Lipid Peroxidation, and Cell Death: Discoveries, Rediscoveries, and Open Issues. *Antioxid. Redox Signaling* **2018**, *29*, 61−74.
(50) Scheerer, P.; Borchert, A.; Krauss, N.; Wessner, H.; Gerth, C.; Höhne, W.; Kuhn, H. Structural Basis for Catalytic Activity and Enzyme Polymerization of Phospholipid Hydroperoxide Glutathione Peroxidase-4 (GPx4). *Biochemistry* **2007**, *46*, 9041−9049.
(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) Bastian, M.; Heymann, S.; Jacomy, M. Gephi: An Open Source Software for Exploring and Manipulating Networks. *Proceedings of the*

https://doi.org/10.1021/acsomega.2c01289
ACS Omega 2022, 7, 29587−29597

international AAAI conference on web and social media, 2009; Vol. *3*, pp 361−362.
(53) Schrödinger, L. L. C. *The PyMOL Molecular Graphics System*.
(54) Reuveni, S.; Granek, R.; Klafter, J. General Mapping between Random Walks and Thermal Vibrations in Elastic Networks: Fractal Networks as a Case Study. *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.* **2010**, *82*, 041132.
(55) Pérez, A.; Marchán, I.; Svozil, D.; Sponer, J.; Cheatham, T. E.; 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.
(56) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(57) Le Grand, S.; Götz, A. W.; Walker, R. C. SPFP: Speed without Compromise—A Mixed Precision Model for GPU Accelerated Molecular Dynamics Simulations. *Comput. Phys. Commun.* **2013**, *184*, 374−380.
(58) Roe, D. R.; Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9*, 3084−3095.
(59) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33−38.
(60) Hsin, J.; Arkhipov, A.; Yin, Y.; Stone, J. E.; Schulten, K. Using VMD: An Introductory Tutorial. *Curr. Protoc. Bioinf.* **2008**, *24*. DOI: DOI: 10.1002/0471250953.bi0507s24.
(61) Kanungo, T.; Mount, D.; Netanyahu, N.; Piatko, C.; Silverman, R.; Wu, A. An Efficient K-Means Clustering Algorithm Analysis and Implementation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 881−892.
(62) Prinz, J.-H.; Keller, B.; Noé, F. Probing Molecular Kinetics with Markov Models: Metastable States, Transition Pathways and Spectroscopic Observables. *Phys. Chem. Chem. Phys.* **2011**, *13*, 16912−16927.
(63) Prinz, J.-H.; Wu, H.; Sarich, M.; Keller, B.; Senne, M.; Held, M.; Chodera, J. D.; Schütte, C.; Noé, F. Markov Models of Molecular Kinetics: Generation and Validation. *J. Chem. Phys.* **2011**, *134*, 174105.

https://doi.org/10.1021/acsomega.2c01289
ACS Omega 2022, 7, 29587−29597