

The quaternary question: Determining allostery in spastin through dynamics classification learning and bioinformatics

Cite as: J. Chem. Phys. **158**, 125102 (2023); <https://doi.org/10.1063/5.0139273>

Submitted: 19 December 2022 • Accepted: 07 March 2023 • Accepted Manuscript Online: 07 March 2023 • Published Online: 28 March 2023

 Maria S. Kelly,  Amanda C. Macke,  Shehani Kahawatte, et al.

COLLECTIONS

Paper published as part of the special topic on [New Views of Allostery](#)



View Online



Export Citation



CrossMark

ARTICLES YOU MAY BE INTERESTED IN

[Allosteric communication in the gating mechanism for controlled protein degradation by the bacterial ClpP peptidase](#)

The Journal of Chemical Physics **158**, 125101 (2023); <https://doi.org/10.1063/5.0139184>

[Predicting allosteric sites using fast conformational sampling as guided by coarse-grained normal modes](#)

The Journal of Chemical Physics (2023); <https://doi.org/10.1063/5.0141630>

[Kinetic effects in directional proteasomal degradation of the green fluorescent protein](#)

The Journal of Chemical Physics **153**, 105101 (2020); <https://doi.org/10.1063/5.0015191>



Time to get excited.

Lock-in Amplifiers – from DC to 8.5 GHz



[Find out more](#)
 Zurich
Instruments

The quaternary question: Determining allostery in spastin through dynamics classification learning and bioinformatics

Cite as: J. Chem. Phys. 158, 125102 (2023); doi: 10.1063/5.0139273

Submitted: 19 December 2022 • Accepted: 7 March 2023 •

Published Online: 28 March 2023



Maria S. Kelly, , Amanda C. Macke, , Shehani Kahawatte, , Jacob E. Stump, , Abigail R. Miller, , and Ruxandra I. Dima^{a)}

AFFILIATIONS

Department of Chemistry, University of Cincinnati, Cincinnati, Ohio 45221, USA

Note: This paper is part of the JCP Special Topic on New Views of Allostery.

^{a)}Author to whom correspondence should be addressed: ruxandra.dima@uc.edu

ABSTRACT

The nanomachine from the ATPases associated with various cellular activities superfamily, called spastin, severs microtubules during cellular processes. To characterize the functionally important allostery in spastin, we employed methods from evolutionary information, to graph-based networks, to machine learning applied to atomistic molecular dynamics simulations of spastin in its monomeric and the functional hexameric forms, in the presence or absence of ligands. Feature selection, using machine learning approaches, for transitions between spastin states recognizes all the regions that have been proposed as allosteric or functional in the literature. The analysis of the composition of the Markov State Model macrostates in the spastin monomer, and the analysis of the direction of change in the top machine learning features for the transitions, indicate that the monomer favors the binding of ATP, which primes the regions involved in the formation of the inter-protomer interfaces for binding to other protomer(s). Allosteric path analysis of graph networks, built based on the cross-correlations between residues in simulations, shows that perturbations to a hub specific for the pre-hydrolysis hexamer propagate throughout the structure by passing through two obligatory regions: the ATP binding pocket, and pore loop 3, which connects the substrate binding site to the ATP binding site. Our findings support a model where the changes in the terminal protomers due to the binding of ligands play an active role in the force generation in spastin. The secondary structures in spastin, which are found to be highly degenerative within the network paths, are also critical for feature transitions of the classification models, which can guide the design of allosteric effectors to enhance or block allosteric signaling.

Published under an exclusive license by AIP Publishing. <https://doi.org/10.1063/5.0139273>

I. INTRODUCTION

Protein allostery is broadly defined as any instance in which a localized event, such as ligand binding, formation of a protein-protein interface, or a mutation at a site in a protein, impacts the dynamics, or the distribution of conformations of another site, located remotely from the original site. In light of the important role played by allostery in proteins, there have been numerous efforts to uncover and characterize allosteric mechanisms of protein regulation and to identify allosteric sites, which can serve as binding sites for novel pharmaceuticals. The methods employed to probe protein allostery stem from principles underlying the transmission of perturbations in protein conformations that take into

account the richness of the conformational space and dynamics of proteins. There are two main avenues that can account for signal (perturbation) transmission in proteins, with clear consequences on the function of the protein:¹ (1) large structural changes that propagate sequentially from the original (allosteric) site to the active site, comprising of processes, such as unfolding and refolding of domains; and (2) localized changes in the dynamic properties of residues, corresponding to the “entropy-driven allostery” model predicted theoretically by Cooper and Dryden.² The first mechanism, termed the “domino model” or induced-fit, corresponds to the view that binding of a ligand triggers structural changes in the neighborhood of the allosteric site, which propagate sequentially, via a single pathway, to an active site, as suggested for the PDZ domain

family.¹ The second mechanism, known as the conformational selection,³ has also been termed the “violin model,” which is based on vibration pattern changes inside the protein. As envisioned by Taylor and collaborators,¹ if a protein is in a particular vibration mode, corresponding to the thermal fluctuations, binding a ligand to an allosteric site can change this mode and the perturbation will spread throughout the protein structure, including to the functional sites, such as nucleotide binding sites in molecular machines, without the need to follow a single/specific pathway.⁴

Experimentally, allosteric sites are identified based on mutational, structural, and thermodynamic characterization with allosteric protein, peptide, or small-molecule modulators using biophysical techniques, such as NMR spectroscopy, protein expression platforms, and deep mutational scanning analysis of protein functions.^{5,6} Computationally, methods span a wide range⁷ from using evolutionary information from multiple protein sequence alignments (MSA), which has been shown to be related to dynamic and functional aspects of proteins,^{8–10} to machine learning approaches designed to extract information about features that characterize transitions between main states sampled in molecular dynamics simulations.¹¹

The original approaches based on the evolutionary information in protein families, such as the Statistical Coupling Analysis,^{12–15} assumed that pairs of residues that tend to be mutated together in MSA are functionally coupled as well and thus are indicative of allosteric communication. While successful in predicting allosteric wiring in certain proteins, such as the PDZ domain, issues related to whether the predicted couplings are relevant to all the proteins in the alignment and if the couplings correspond to structural rather than functional constraints¹⁶ led to the development of new methods that incorporate structural information with the MSA outputs.^{7,17} The majority of the methods used to identify allosteric sites, networks, and pathways in proteins are based on information extracted from protein conformations and dynamics. Several approaches rely on the analysis of an ensemble of structures, built from either atomistic molecular dynamics (MD) or coarse-grained simulations, using cross-correlations between positional fluctuations, as done to extract hot spots,^{18,19} or between atoms' motions,²⁰ or correlations between contacts,²¹ or between the conformations of residues in different site.¹⁴ The graph network science methods complement such approaches by providing information on functional centers and pathways of allosteric communication, being based on the mapping of dynamic fluctuations extracted from MD simulations onto a graph, with nodes representing residues and edges representing weights of dynamic properties.^{16,22,23} The methods based on network science have been applied to both single- and multi-chain proteins, such as chaperones.^{17,24–27} Recently, machine learning based approaches have been developed to address the challenging highly multidimensional problem of the prediction of functional allosteric sites and mechanisms.^{7,11} These methods start from the ensemble of conformations from MD simulations of the target protein, which are classified either based on binding or not to an allosteric effector,²⁸ or as substates corresponding to the metastable states of the protein determined using the Markov state models.²⁹ Successful application of these new techniques, as measured, for example, by the level of agreement with the results of experiments for the identification of allosteric signals, has been reported in various single-chain proteins.^{11,28,29}

Microtubules (MTs) are intrinsically stiff biopolymers found in eukaryotic cells that are responsible for many essential cellular functions. The MT severing enzymes modulate the functions of MTs by breaking the MT lattice along its length.^{30,31} Spastin is an MT severing enzyme that belongs to the AAA+ family (ATPases Associated with diverse cellular Activities), which harnesses the energy from the hydrolysis of adenosine triphosphate (ATP) to perform the severing activity.³² This is a massive protein superfamily that includes ~532 000 sequences and ~23 000 species from Eukaryota.^{33,34} Spastin is a member of the meiotic sub-clade that includes the other MT severing enzymes katanin and fidgetin as well as VPS4.^{35,36} Spastin is crucial for biological activities, such as neurogenesis, mitosis, phototropism, axonal maintenance, and intracellular trafficking.³⁰ A recent study shows that spastin not only severs MTs but also helps their regrowth by increasing the number and the mass of MTs, acting as a dual functional enzyme.³⁷ Mutations in spastin can cause neurodegenerative disorders, such as the hereditary spastic paraplegias (HSP).^{32,38} Although studies have shown the importance of severing enzymes on neuronal disorders, understanding the mechanisms of occurrence of such diseases requires a detailed view of the structures and biochemical actions of the enzyme.^{39,40}

The structure of spastin consists of an MT-interacting and trafficking (MIT) domain, which is connected to the AAA+ ATPase domain through a poorly conserved linker region. The AAA+ domain contains two subdomains: the nucleotide binding domain (NBD) and the helical bundle domain (HBD).³⁸ The presence of ATP and the tubulin substrate allows spastin to oligomerize into hexamers at intracellular concentrations.³² Oligomerization interfaces between the protomers are formed through AAA+ domain contacts. Two spastin hexameric structures have been solved using cryo-electron microscopy (cryo-EM): the spiral and ring conformations. Both structures have been solved in the presence of a glutamate-rich peptide as the substrate.^{32,38} In the spiral conformation, the protomers are arranged in a right-handed spiral, with each protomer having a ~6 Å rise resulting in a ~35 Å gap between the boundary protomers. The nucleotide binding pockets in the AAA+ domain of all the protomers accommodate ATP.³⁸ Upon hydrolysis, protomer F was found to lack ATP and becomes highly flexible, allowing the gap between the boundary protomers to close, resulting in the formation of the ring conformation.³² The substrate binds to the central pore of the hexamer, which consists of two electropositive loops in the NBD of each protomer, namely, pore loop1 (PL1) and pore loop2 (PL2), that form a double helical staircase around the substrate.^{32,38} Structural studies related to spastin suggest that the binding of the nucleotide and substrate activates the allosteric network(s) among the functional regions in the AAA+ domain. For example, R591 is within H-bonding distance to N629 in PL3, which is found in the central pore, and D585, which is within H-bonding distance to E583 of the Walker B motif that is a part of the nucleotide binding pocket. Thus, R591 has been proposed as a center of an allosteric network, which couples the ATP binding sites and the substrate binding loops.³⁸ Mapping out the allosteric networks is thus likely to be essential for gaining an understanding of the severing activity of spastin.

We implemented a variety of the above-reviewed methods utilizing bioinformatics, molecular dynamics, and machine learning to probe the overall allosteric networks for the quaternary structure of

spastin. Using multiple sequence alignments, we addressed the conservation and coevolutionary contributions of identified residues. The direct connectivity of residues was mapped out using protein structure graphs and analyzed with graph network theory to determine optimal and sub-optimal paths of allosteric signal transmission and the positions that control most of this allosteric transmission. We then used a physical descriptor based analysis for secondary structures, coupled with machine learning approaches, to identify the primary contributors for specific ligand binding. Finally, we compared the results of our various approaches to the experimentally identified allosteric positions listed above and our previously established hotspot analysis for lower order oligomers.¹⁹ By taking a multi-method approach, we seek to identify the most consistent regions that indicate a robust allosteric signal for the elusive quaternary allosteric network in spastin vs the monomer case and to pinpoint the best methodology to discern allosteric signals in multimeric protein assemblies.

II. METHODS

A. Bioinformatics

InterPro lists 532 510 sequences across 23 666 species with an AAA+ domain in Eukaryota.^{33,34,36} To assess the role of coevolution on the allosteric networks of spastin, we narrowed down the list of sequences by considering only the metazoa kingdom, which includes the species corresponding to the spastin structures of interest: *Homo sapiens* (6PEN) and *Drosophila melanogaster* (6P07) (203 812 sequences over 2371 species). These AAA+ proteins can be further categorized by considering their other domains.³⁶ The HBD of spastin is categorized by InterPro as the AAA+ lid3 domain (IPRO041569), which is found in 40 109 sequences across 1667 species in Metazoa.^{33,34} Other proteins, which are not biologically similar to the MT severing proteins, contain a domain similar to the AAA+ lid3. Thus, to build the multiple sequence alignment (MSA), we selected only the sequences from the biologically relevant meiotic sub-clade that includes spastin, katanin, fidgetin, and VPS4.^{35,36} From this set, we kept only the intact sequences (no sequences with an “X”) resulting in a set of 3413 sequences. From these sequences, we also built an MSA keeping only the spastin sequences (1022 sequences) in order to address the variance in positions of interest throughout the study. Any residue found to be completely conserved in the spastin-only MSA, but found to lose conservation in the meiotic clade MSA, indicates a position that is of particular importance for spastin. To build the MSAs, we employed Clustal Omega 1.2.4.⁴¹

B. Molecular dynamics simulations

We analyzed the all-atomistic MD simulations of the spastin hexamers from our previous work.⁴² Here, we used the trajectories we ran for the two solved cryo-EM structures of spastin: the spiral (PDB: 6P07) and ring conformations (PDB: 6PEN).^{32,38} The two structures consist of monomers made of NBD and HBD. The spiral conformation has ATP bound to the NBD of every monomer, while the ring conformation has ADP bound to monomers A through E and no nucleotide density in monomer F. Both spastin conformations were also solved with a minimal substrate (polyE for spiral and polyEY for the ring). In order to study allosteric effects due to the

presence of either the nucleotide and/or the substrate, we analyzed simulations with the original cryo-EM composition (which we call the COMPLEX state), as well as perturbed systems, corresponding to the removal of one or both binding partners at a time (the NUC state for the motor with only nucleotides included, the SUB state for the motor with only minimal substrate included, and the APO state for the removal of both the substrate and nucleotide).^{19,42} The residue ranges for the known binding sites are listed in Table S1.^{38,40} In total, we analyzed 4 μ s of simulation time. Simulation details and convergence testing can be found in our previous work.⁴² In addition, we analyzed the atomistic MD simulations of the spastin monomer in the same ligand states used in our most recent work, totaling 600 ns.¹⁹ For the analysis part of all systems, we stripped the ligands, keeping only the protein coordinates, and we excluded the first 5 ns of run time from each trajectory, as in our previous work.⁴²

Following our approach for the extraction of hotspots in spastin oligomers,¹⁹ we labeled the secondary structural elements of each spastin state so that we could identify allostery in terms of a particular alpha helix, flexible loop, or beta strand. These labels were based on the original cryo-EM secondary structure designation, while keeping the rules that each alpha helix has to be at least 4 residues long, and beta strands have to be at least 3 residues long.^{38,43,44} Labels for the spiral and ring conformations are listed in Tables S2 and S3.

C. Markov state models

In order to identify metastable conformations of spastin that can shed light on the allosteric mechanisms between the four simulated setups (COMPLEX, NUC, SUB, and APO), we built a Markov State Model (MSM) using VAMPnets.^{45,46} This is an unsupervised neural network, where the input is all the Cartesian coordinates of spastin, which optimizes the assignment of all sampled trajectory frames into one of a few macrostates. The result is a final transition matrix composed of each trajectory frame's probability of transitioning to each macrostate within a chosen time step or “lag time.” The macrostates and their representative structures are made using crisp assignments, where each frame is assigned to a macrostate based on its highest probability from the output transition matrix. We built MSMs for our spiral, ring, and monomer trajectories separately. Our five-layer, feed-forward network includes an exponential linear unit (ELU) activation function for each layer and Softmax output nodes for a selected number of macrostates. The model was built using 5000 point batch size, 30 epochs, and fit using CPU. A lag time of 400 ps for the hexamer models and 200 ps for the monomer model was selected, tested, and validated by the implied timescale vs lag time plots and the Chapman–Kolmogorov (CK) test (Fig. S3). We attempted to fit our model with 4 through 10 output macrostates and tested their validity through the CK test.

D. Mutual information

In order to extract coevolutionary information from the MSA, we carried out a mutual information analysis with an average product correction, using the MISTIC web server, on the MSA for the meiotic sub-clade according to Eq. (1).^{47–49} Previous studies have,

$$MI_{(i,j)} = \sum_{a,b} P(a_i, b_j) \times \log \left(\frac{P(a_i, b_j)}{P(a_i) \times P(b_j)} \right), \quad (1)$$

shown that coevolutionary relationships can form between residues that are functionally significant.^{49,50} To evaluate such relationships, MISTIC calculates the sequence-based parameter called cumulative mutual information (cMI), listed in Eq. (2). The cMI,

$$cMI_x = \sum_{y, MI(x,y) > t} MI_{(x,y)} \quad (2)$$

plotted per residue in Fig. S6a, is a sum of the MI above a 6.5 threshold, which is used as a score to indicate the degree to which a given amino acid participates in the mutual information network.⁵¹ These values are dependent only on the MSA. To incorporate structural constraints, thereby accounting for the residue connectivity and the order of the protomers in the hexameric assembly of spastin, we used MISTIC to calculate the proximity mutual information (pMI) scores, by residue, for each chain in the representative structures of the seven MSM metastable states for the spiral conformation. We focused on the spiral hexamer conformation because it represents the pre-hydrolysis state of spastin. These scores were then averaged across the seven states per protomer in order to describe the positions that were most crucial for the coevolutionary network of the overall assembly. In order to capture the differences due to the presence of the oligomeric interfaces, the same procedure was used to capture the average pMI scores, by residue, for the representative structures of the 8 MSM metastable states of the monomer. MISTIC calculates the pMI score by averaging the cMI of all the residues within 5 Å of a given residue, according to Eq. (3).⁴⁷ Positions with high pMI correspond to residues

$$pMI_x = \frac{1}{N} \sum_{d(x,y) \leq 5} cMI_y \quad (3)$$

that have significant coevolutionary signals as well as high connectivity.⁵² By taking the top 5% of the cMI and pMI values for comparison, we can assign a relative rank of importance for residues in each structure as seen in Table S5 for the pMI residues.⁵²

E. Network and community analysis

Another approach to evaluate the connectivity and allosteric paths between specific residues was the use of network graphs. We built two types of networks using the Python library NetworkX.⁵³ The nodes were all the residues in a structure. The edges were weighted dependent on the type of network being evaluated. We extracted standard network parameters to characterize the networks and identify positions that were the most important for the passing of information through each protein ensemble.

The first network graph, called a protein structure graph (PSG), was built from the metastable structures for the monomer and the spiral hexamer assemblies used for the pMI values.⁵⁴ These PSGs were built based on the pairwise contacts between residue side chains (no backbone included) that are not covalently bound.^{55,56} The pairwise contacts (n_{ij}) between residues in each chain corresponded to the number of side-chain heavy atoms from the two residues that are found within 4.5 Å, extracted using the VMD measure function.⁵⁷ The interaction strength (I_{ij}) between each pair was calculated according to $I_{ij} = \frac{n_{ij}}{\sqrt{N_i \times N_j}} \times 100$, where the N_i and N_j are the normalization factors for the residues (see Table S6).⁵⁴

We then weighted the edges of the PSG based on the average of the interaction strengths across the metastable states as described in the [supplementary material](#). We evaluated the contribution of the residues to the overall connectivity of a PSG by determining the residues dubbed as “hubs,” meaning a node with more than four edges.⁵⁴ These nodes are structurally important for the protein as many other residues were found to be in direct contact with them. The hubs for each ensemble are listed in Table S7. This type of network characterizes the most structurally significant parts of a protein, but it can fall short in accounting for the protein dynamics.

To address directly the dynamics captured by the MD simulations, we built a second type of network. Here, an edge connects two residues if their C_α 's are within 8 Å in the starting structure used to set up the simulations. Following previous studies,^{55,56,58} the edges were weighted based on the respective pairwise element of the dynamic cross-correlation matrix (DCCM), c_{ij} . Here, we used the average DCCM over the longest trajectories (200 ns each) from our previous work for the COMPLEX and, respectively, the APO setups of the spiral hexamer.⁴² A c_{ij} cutoff of 0.6 was applied so that only highly correlated residues are connected in the network, and the weights were given by $w_{ij} = -\log(|c_{ij}|)$.^{55,59} We chose the betweenness centrality (C_B) to describe the globally significant positions. The C_B of a node in a network is defined as the number of shortest paths between other nodes that pass through the node of interest [see the [supplementary material](#), Eq. (1)].^{17,60} Typically, the networks's top 10% of residues by betweenness centrality are the residues deemed most important to the overall allosteric network.¹⁷ These residues are listed in Table S10 for the monomer and Tables S11 and S12 for each protomer.

While the hubs and centrality analyses of the described networks gave a global indication of structurally and dynamically important residues, they are not adequate to describe how a specific allosteric signal is propagated through the network. To extract detailed descriptions of how an allosteric signal is propagated throughout the network, we performed path analyses between positions of interest in the c_{ij} based network. Based on both experimental work and our results, we used the allosteric center (S589 and R591) as a source. Strikingly, the nodes in the allosteric center were removed from the c_{ij} network entirely in protomer D in the COMPLEX state, indicating a significant change or disruption to the network within the protomer. This disruption was identified in the pMI analysis of the metastates and by the centrality results. To characterize the impact of the propagation of a signal to the positional fluctuations inside the NBD, we chose two positions in the CTT binding region, specifically in pore loop 1 that were identified from the MSA analysis, as the sink (Y556 and E561). To observe how changes at the allosteric center propagate to the HBD, we chose two positions in the CT Hlx (H12) that were also identified from the MSA analysis as the sink (W749 and Y753). The shortest paths, determined with the Dijkstra method in NetworkX as the sum of the weights of the involved edges as described in Fig. S12, were extracted for each pair and used to identify the optimal path for the allosteric signal to follow (Tables S13–S16). We also used sub-optimal path analysis^{55,56} to identify residues frequently sampled by the allosteric signal. We extracted the next 20 000 shortest paths by weight between each pair of source and sink residues with NetworkX using the Yen algorithm.⁶¹ From the sub-optimal paths,

we determined the relative degeneracy (degeneracy per number of paths) of each node visited by the paths between the regions of interest. We then grouped the individual positions with high degeneracy based on their secondary structure elements (Tables S1 and S2) and extracted the average degeneracy of these secondary structures.

F. Descriptor extraction

We sampled the MD simulations trajectories every 200 ps for the hexamers and 100 ps for the monomer systems in order to collect 10 000 frames for each spastin setup. For each frame, we calculated several physical descriptors that would have the potential to identify allosteric sites that are active during the transitions between the simulated spastin setups.²⁸ We extracted these biochemically relevant descriptors at the level of secondary structural elements for all six chains of the hexameric motor. We termed a “feature” as being the descriptor values in each frame of all trajectories for a given structural element. The descriptors tested, which were calculated with either GROMACS, VMD, and/or Python,^{57,62} are listed in Table I.

The CONTACTS descriptor could only be calculated for the hexamer simulations. The fraction of contacts for each system is taken with respect to its starting structure, directly following equilibration. To avoid biasing our models by including features that make no or only brief contacts with neighboring protomers, we kept only features that had at least six initial contacts with neighbors. This resulted in 29 and 23 total CONTACTS features for the spiral and ring hexamer, respectively. We added all the values of a given descriptor listed earlier for trajectories of all four setups (APO, NUC,

SUB, COMPLEX) of spastin in the spiral hexamer so that a final dataset could be made for each descriptor separately. This was then repeated to make similar datasets for spastin in the ring hexamer.

G. Classification of spastin states

Our final lists of descriptors for each system were used to identify important allosteric changes upon either the binding of the nucleotide and/or substrate to the spastin monomer or the perturbation of the COMPLEX spastin hexamers. We first performed classification analysis using XGBoost for each of the three monomer transitions (APO to NUC/SUB/COMPLEX), removing either the nucleotide or the substrate binding regions (Table S1) and any region within 3 Å of the respective binding site, depending on whether the transition added the nucleotide or the substrate.^{28,63} We note that, similar to results from the literature,²⁸ when we included all the regions, irrespective of their location in the structure, our approach always found the binding site and neighboring regions among the top features (data not shown). Due to the large number of features collected per descriptor, we applied early stopping recursive feature elimination using SHAP (ShapRFECV) for base learning tuning and removal of unnecessary features in our final classifications (details in the [supplementary material](#)).⁶⁴

III. RESULTS AND DISCUSSION

A. Sequence conservation analysis

In order to assess the connection between sequence evolution and the allosteric sites in spastin, we built an MSA, including the proteins found in the meiotic clade and another MSA with only spastin, as described in the Methods section. Of the residues identified by Roll-Mecak and collaborators as functionally important in spastin, either through mutation experiments or through the solved cryo-EM structure, listed in Table II and Table S1,^{38,40} the residues that are found to be invariant (>~90% identity) in the spastin-only MSA, but not in the meiotic clade MSA, are E462, F522, Y556, V557, D585, R630, Q632, R633, L743, Y746, and W749 (see also Fig. S5 for the identity plots). These residues are thus significant for spastin since they help distinguish it from the rest of the sub-clade and have been found to play a significant functional role for the motor in experiments.^{38,40} Descriptions of these positions from the literature are further discussed in the [supplementary material](#).

B. Markov state models transitions characterize the nature of allosteric changes in spastin

The analysis of metastable (long-lived) conformations extracted from the MD simulations of the spastin monomer's four setups (APO, NUC, SUB, COMPLEX) can provide insights into the nature of allostery in spastin based on the two views of allostery described in the Introduction: conformational selection and induced fit (Fig. 1). Here, we utilized VAMPnets in order to build a Markov State Model (MSM) that can capture slow conformational changes within a few representative structures, termed macrostates, which can help elucidate the binding mechanisms of ligands (RMSD values for each macrostate are in Table S18). As described in detail in the [supplementary material](#) methods section on MSMs, we found that 8 was the optimal number of macrostates for the spastin monomer, with the most notable

TABLE I. Descriptors extracted at the level of secondary structure for spastin MD trajectories.

DescriptorID	Description
SASA	The average solvent accessible surface area of residues in a secondary structure
RSA	The ratio of solvent accessibility of each residue X to the maximum possible accessibility in a Gly-X-Gly tripeptide
(PHI, PSI)	Phi and Psi angles calculated for each peptide bond and averaged for all residues in a secondary structure
COM	The distance between the center of mass of an element to the center of mass of the spastin motor (nm)
VDW	van der Waals energy associated with the interaction between a secondary structure and the remainder of the protein (kJ/mol)
COULOMB	Coulombic energy associated with the interaction between a secondary structure and the remainder of the protein (kJ/mol)
CONTACTS	The fraction of contacts at the interface between monomers with respect to the first frame of each trajectory. A contact is defined as a C α located within neighboring monomers (monomers $i+1$ and $i-1$) found within 8 Å of any C α within the secondary structure being analyzed

TABLE II. Residues that were found to impact either ATPase activity or MT severing in the literature.^{38,40} The respective secondary structure, according to Table S2, is indicated, together with the functionality they are reported to affect (see Table S1). HSP mutations for human spastin are in Table S4. The amino acid identity is provided for the amino acid from the *D. melanogaster* structure, which was used for the MI and the network analyses. Residues indicated with a (*) are in the top 5% of pMI. Residues indicated with a (#) are found as hubs. Residues indicated with a (+) are in the top 10% betweenness centrality for any setup.

Important residue (meiotic clade %; spastin-only %)	Secondary structure	Functionality from experiments
E462 ⁺ (13%; 89%)	...	Stabilizes PL1 with H-bond with K555
L465 (97%; 97%)	H1	Impairs ATPase and abolishes MT severing
L470 (20%; 62%)	H1	Impairs ATPase and MT severing
I473 (93%; 87%)	L1	Impairs ATPase and MT severing
V474 ⁺ (45%; 85%)	L1	Impairs ATPase and MT severing
W482 ^{+#} (73%; 14%)	H2	Linchpin for a hydrophobic pocket, rearranges upon cofactor binding
F522 ⁺⁺ (37%; 100%)	B1	Covalently bound to the Walker A region—HSP mutation interferes with ATP binding—it is in proximity to W749 from H12 and P631 from PL3
P524 ⁺ (98%; 100%)	L4	Walker A—in proximity to L743 from the H12
K555 ⁺ (100%; 100%)	L6	Found in PL1—forms salt bridges with the substrate—and a H-bond with L462
Y556 ⁺ (73%; 95%)	L6	Abolishes MT severing
V557 ⁺ (32%; 96%)	L6	Impairs ATPase and abolishes MT severing in PL1
E561 ⁺ (100%; 100%)	H5	Impairs ATPase and abolishes MT severing—helix following PL1—In H-bond distance from R601 from PL2 (potentially dependent on the presence of a substrate)
E583 ^{+#} (100%; 100%)	L8	Abolishes ATPase and MT severing—mutated to a Q prevents ATP hydrolysis and stabilizes the structure
D585 ⁺ (81%; 100%)	L8	In H-bond distance with R591
R591 ^{+#} (100%; 100%)	L8	Reported as an allosteric center due to proximity to essential positions in ATP binding region, PL3, and H12
R600 ^{+#} (91%; 100%)	H6	Impairs ATPase and abolishes MT severing
R601 ^{+#} (100%; 100%)	H6	In H-bond distance with E561 following PL1 (potentially dependent on the presence of a substrate)
N629 ^{+#} (99%; 100%)	L10	Abolishes ATPase and MT severing—n H-bond distance with the γ -phosphate in ATP and W749 from H12 and PL3 from a neighboring protomer
R630 ^{+#} (30%; 100%)	L10	Impairs ATPase and abolishes MT severing in PL3—H-bonds with the backbone of R591 in PL2 from $i + 1$ —in proximity to H12
Q632 ⁺ (30%; 99%)	L10	Impairs ATPase and MT severing in PL3
E633 ⁺ (49%; 100%)	L10	Impairs ATPase and MT severing in PL3—R600 from PL2 of $i - 1$ protomer forms a salt bridge
L743 ⁺ (63%; 97%)	H12	CT-Hlx—forms interactions with neighboring protomers—in proximity to Walker A
Y746 ⁺⁺ (84%; 97%)	H12	CT-Hlx—forms interactions with neighboring protomers—in proximity to Walker A
W749 ⁺⁺ (89%; 100%)	H12	CT-Hlx—forms interactions with neighboring protomers—in proximity to Walker A (F522), and PL3
Y753 ⁺⁺ (24%; 40%)	H12	Impairs ATPase and abolishes MT severing—involved directly in the interactions with the neighbor protomers

differences found within its HBD [Fig. 1(a)]. The histograms in Fig. 1(b) show the contributions of each of the four setups, colored as red (APO), blue (NUC), green (SUB), and black (COMPLEX), to the 8 respective numbered MSM macrostates.⁶⁵ The contributions of the setups are normalized such that they each add up to unity across each figure. Figure 1(c) displays the most probable transitions between the 8 macrostates, obtained from the MSM analysis.

For the monomer, the binding of the nucleotide from the APO setup can only be seen in the transitions from macrostate 1 to 7 and 6 to 5. This change is likely to follow an induced fit mechanism due to the lack of NUC setup population overlapping between the starting and ending macrostates of these two pairs. The binding of the substrate to the APO setup follows a conformational selection mechanism since macrostates 1 and 6 contain frames from both APO and SUB setups. Transitions to the COMPLEX-populated

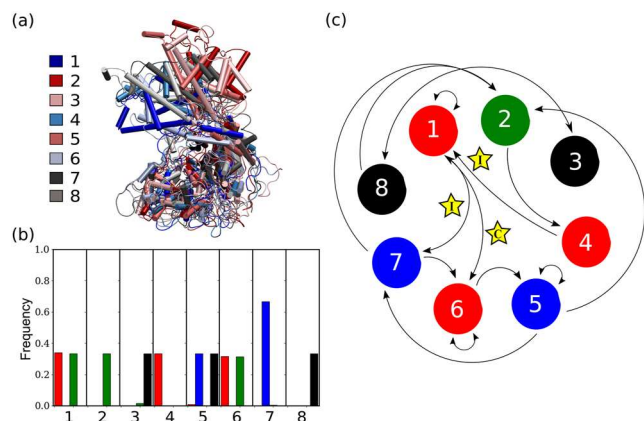


FIG. 1. Summary of MSM results for the spastin monomer. (a) Representative structures for each of the 8 macrostates, aligned to the NBD of the starting structure. The monomer is colored based on the macrostate. (b) Populations of the 4 setups of spastin within each macrostate: red is APO, blue is NUC, green is SUB, and black is the COMPLEX state. (c) Most probable transitions for each macrostate. Macrostates are colored based on the highest populated spastin setup. Stars labeled "I" or "C" indicate transitions found to show induced fit or conformational selection, respectively.

macrostates (3, 5, and 8) are more selective than other transitions, inferring a specific path of ligand binding to the monomer that results in global conformations distinct from other setups. Macrostate 6, containing frames from SUB simulations, is the only transition observed to get to the COMPLEX-populated macrostate 5. This formation of the COMPLEX monomer is through an induced fit mechanism since COMPLEX population is not seen in macrostate 6. When starting from a monomer in the NUC setup, however, macrostate 5 contains both NUC and COMPLEX and, therefore, can become the COMPLEX through the conformational selection. Overall, the MSM built for the spastin monomers support a preferred route to achieving the COMPLEX setup through the conformational selection resulting from substrate binding to the nucleotide setup. From the point of view of statistical mechanics, it is entirely possible to find that a combination of conformational selection and induced fit mechanisms are responsible for the allostery of a protein system.

MSMs were also constructed for the spiral and ring hexamer, separately. The analysis of the global changes between the two conformations can provide mechanistic insight into the functioning of the hexamer (Fig. 2). The spiral hexamer (Fig. S1) gave optimal model results using seven macrostates, while the ring preferred six macrostates (Fig. S2). Perturbation of either hexamer by removal of both ligands resulted in multiple macrostates composed of only APO setup frames, which shows that the removal of both ligands results in higher diversity in conformational changes compared to the other states. Increased flexibility of APO setups was found in other oligomeric proteins.⁵⁶ For the spiral hexamer, the major motion is concentrated in chain F. Namely, macrostates 6 and 7 showed that the HBD of protomer F moves down compared to the starting structure (distances of monomer F's center of mass are listed in Table S19 of the [supplementary material](#)). A similar motion occurs in macrostate 3, which suggests that the removal of

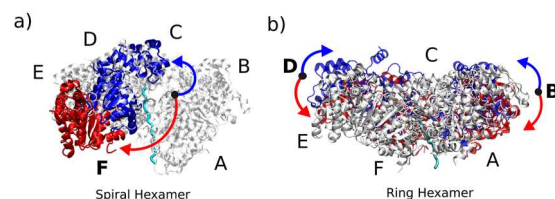


FIG. 2. MSM macrostates showing the most significant conformational shifts for the (a) spiral and (b) ring hexamers. Protomers A through F are labeled for the respective reference structure, which is in white. The minimal substrate (cyan) is located in the central pore of each hexamer. Portions of two macrostates showing the most extreme motions are colored in red and blue. For the spiral hexamer, monomer F is shown for macrostate 1 (blue) and macrostate 4 (red). For the ring hexamer, monomers B and D are shown for macrostates 1 (blue) and 4 (red). The arrows describe the direction of motion of each macrostate vs the reference structure.

the nucleotide causes protomer F to move closer to A, resulting in a configuration resembling the arrangement of these end protomers in the ring hexamer [Fig. 2(a)]. In contrast, macrostates 5 and 4 show a large swing-like motion of the HBD of monomer F, resulting in the opening of the gate to the spiral hexamer's central pore. Interestingly, the ring hexamer's macrostates display an entirely different motion compared to the spiral hexamer [Fig. 2(b)]: a swing-like up and down motion of multiple protomers around a hinge represented by the diagonal between the centers of mass of monomers C and F. Moreover, we found that, for both spiral and ring, the motions described by the representative MSM macrostates recall the motions along principal component 1 from the MD trajectories determined in our previous work,⁴² which resulted in axial fluctuations of the pore loops. Importantly, these coordinated motions resemble the pushing and pulling actions that underlie the proposed death spiral MT severing mechanism of the spastin motor.^{31,66}

C. Mutual information analysis

In order to extract coevolutionary information from the MSA, we computed the mutual information (MI) using the MISTIC web sever.^{47,51} Protein residues coevolve for many reasons including, but not limited to, adapting or accounting for protein stability, ligand binding, or regulation.¹⁷ Studies have shown that residues that are in close proximity to each other often coevolve and form independent functional or allosteric communities.^{67,68} In order to assess the importance of positions in spastin, we focused on the residues corresponding to the top 5% cMI and pMI values.^{17,52} Because residues in close proximity to a protein's binding partner tend to have high cMI values, focusing on positions with high cMI values can help detect catalytic residues from the sequences in the MSA.⁵¹ The top 5% cMI identified residues, represented in Fig. S6b, were K492, P506, P525, G526, V564, D582, Q583, R601, K603, L607, E653, E657, R678, R704, and I758. Strikingly, several of these positions have either been proposed or are known to be important for the function of spastin. For example, Q583 from the Walker B motif in the ATP binding pocket (in L8, according to Table S2) mutation from glutamate (E) to a glutamine (Q) prevents ATP hydrolysis, thereby abolishing severing, as described in Table II.³⁸ The other positions are found in the Walker A motif (L4) from the ATP binding pocket, H6, and

H8–H10 from the HBD and are described more in-depth in the [supplementary material](#).

We analyzed the residues with the highest pMI scores to determine if combining the structural information provides additional insight into spastin's allostery. As an example, the average pMI score for each residue in the monomer and for protomer D are in Fig. S7. The residues corresponding to the top 5% of the pMI scores are presented in Table S5 and Fig. S8 and were compared to the residues identified to be functionally important in experiments (Table II and Table S1). The top 5% pMI residues in the spastin monomer were F522, N527, K529, F568, R572, F580, L602, K603, F606, F610, R641, F642, R656, and Y746. From this list, residues F522, K529, R641, and Y746 were found to be functionally relevant in experiments (Table II and Table S1). These residues correspond to positions in B1, the Walker A motif (L4, H4), H5, B3, H6, and the arginine fingers in H7, H8, and H12, which are further described in the [supplementary material](#).

In the protomers, we found similar residues to the ones in the monomer, with a few important differences that we assign to the formation of the quaternary structure (inter-protomer interfaces). First, residues L567 (in H5) and L607 (in H6) are unique residue signals for the hexamer. Next, residue R641 was present in the monomer but is only present in the hexamer in protomer A. This position is found at the convex interface of protomer i in the assembly, which is the interface that protomer A lacks in the hexamer. In protomer F, there is one unique signal from an additional residue in H12, W749, which was identified in the MSA comparison (Table II). Protomer F has a convex interface but lacks the concave interface, suggesting an enhanced role for H12 in the absence of the concave interface. Finally, protomers D (K517 and L532) and E (V564) have unique residue signals, indicating some significant structural change within those protomers captured by the metastable states.

D. Network graph analysis reveals the role of quaternary interactions in spastin allostery

To map out the connectivity of the residues across spastin structures, we built PSGs for the monomer and the spiral hexamer (see Methods). We analyzed the PSG hubs for the monomer to identify the residues that are the most structurally significant for spastin.⁵⁴ We compared the results with the findings from the PSGs for the full hexamer to identify the changes due to the connectivity unique to the quaternary structures of spastin. We also built a network based on the DCCM pairwise correlations to incorporate the dynamics observed in MD simulations. The betweenness centrality network parameter allowed us to identify nodes that are globally essential to the network connectivity.^{17,60} Finally, optimal and sub-optimal paths from the network were extracted to determine how a perturbation (binding of a ligand or inter-protomer interface formation) might be passed within the networks⁵⁵ from the allosteric center either inside the NBD or to the HBD. The findings could then be compared to the essential residues identified in the literature (Table II, Tables S1 and S4).

1. Analysis of the PSG hubs

The hubs were the nodes throughout the network that were connected by 4 or more edges in the PSGs built with I_{ij} , which means they are keystones for the scaffolding across the meta states

(see Table S7).⁵⁴ The hubs for the monomer were residues from H2, L3, L4, L6, H6, L10, H7, H8, H11, H12, and the β -sheet (B1, B2, and B3), as seen in Fig. S9a and described in detail in the [supplementary material](#). In the PSG, for the full hexamer (Fig. S9b), hubs that are in common with the monomer positions are likely to be important for the passing of perturbations through the tertiary scaffold of the protomers. In contrast, residues that are unique to the hexameric PSG are important for the passing of information through the quaternary ensemble. The unique hexameric hubs are Q583 (in protomers B, C, and D), R591 (in protomers D, E, and F), and R601 (in B, C, D, E, and F). These positions are located in the Walker B motif of the ATP pocket (in L8), in the allosteric center³⁸ (also in L8) and in PL2 of the CTT binding region (in H6), as detailed in the [supplementary material](#). Thus, the unique hubs in the hexamer are all functionally related, being located in the two known active sites and the proposed allosteric site. To understand the factors that make a position a network hub, we analyzed the nature of the highlighted hubs in the monomer and protomer E from Tables S8 and S9, as well as an additional position (465) in protomer E, which is not a hub but it is in the same secondary structure element (H2) as one of the hubs. As detailed in the [supplementary material](#), the hubs are characterized by very high average I_{ij} values (of order 10), which are about an order of magnitude higher than the average I_{ij} value (~ 1) of the non-hub position. This finding complements the representations of the dense structural environment of the main hubs (for F522, R591, R630, W749, Y753) found in the original cryo-EM structure.³⁸ The environment surrounding R591 is also shown in Fig. 5.

2. Analysis of network centrality

The betweenness centrality globally describes how well a residue is connected to the overall network, meaning whether it is essential or not for the passing of information between all possible pairs of nodes.^{17,60} We determined the highly central (top 10% centrality) residues of the network graphs that incorporate the dynamic information captured from the DCCMs of the COMPLEX and APO setups (see Fig. S10). The residues identified for the monomer, listed in Table S10 and depicted in Fig. S11, change between the setups, which is indicative of shifts in the underlying allosteric networks. In the COMPLEX setup, the residues are localized at the concave interface and in proximity to the active sites. Namely, we found positions from the Walker A and B motifs at the concave portion of the ATP pocket, from each of the pore loops, and a residue from the CT Hlx (H12). In the APO setup, the high centrality residues are more spread out across the monomer and are located primarily across the hinge that separates the NBD from the HBD. Importantly, while a few positions populate the ATP pocket, but not nearly as many as in the COMPLEX setup, we did not find any pore loop residues. This depicts a specific organizational shift in the network caused by the binding of ligands, which aligns with the idea that the pore loops become disorganized in the absence of the minimal substrate.⁴⁰ The highly central residues from each protomer in each of the hexameric networks are listed in Tables S11 and S12. We found that there were fewer central residues in protomer A compared to the monomer, and they were exclusively located at the concave interface (see Fig. 3). This indicates that, upon the formation of the concave interface alone, the network will shift to favoring the $i + 1$ neighbor. This is supported by the finding that protomer F, which is the only

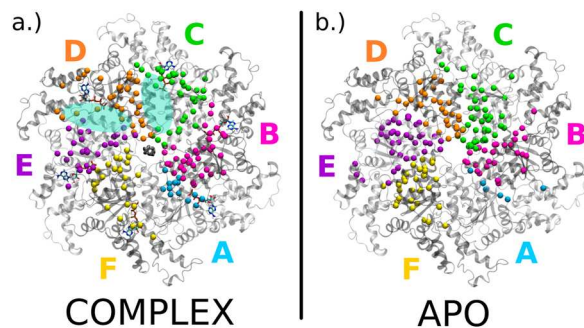


FIG. 3. The structure of the hexamer from the COMPLEX setup (a) and the APO setup (b) indicating the most significant residues from the top 10% of the betweenness centrality (C_B) (see Tables S11 and S12) from the c_{ij} based network. The positions are represented by beads colored by protomer. For the COMPLEX setup, the CTT is shown in the central pore in gray and the ATP molecules are in a licorice representation. The interfaces that were found to be significantly disrupted in the COMPLEX setup between CD and DE are highlighted with teal circles.

one that forms only the convex interface, has a similar distribution of positions to the other protomers. A more even distribution of highly central residues across each protomer, with the exception of A, was observed in the APO setup. In the COMPLEX setup, the network experienced a significant disruption, in particular, at the interfaces involving protomer D. This paints a vivid picture of the asymmetry of the protomers that make up the hexamer in the presence of ligands.

3. Optimal and sub-optimal paths between allosterically important positions illustrate the intrinsic asymmetry of spastin hexamers

We performed a detailed path analysis to create a map of how allosteric signals pass between the two regions of interest through both the monomer and the hexamer in each of the two setups. As detailed in the Methods, the source was the allosteric center identified in experiments (S589 and R591 in L8).³⁸ We note that we identified position R591 as both a hub and as highly central in several of the protomers, but not in the monomer. We selected two different sinks (PL1 and H12), which were also found to be functionally important and previously identified by the various methods (Table II). For the monomer networks, regardless of the destination (PL1 or H12), we found that the absence of the ligands led to significantly longer optimal paths, as listed in Tables S13 and S14, and similar dramatic shifts between the sub-optimal paths distributions (Fig. S13). Moreover, we found that the largest value for the node degeneracy in the APO setup is higher than in the COMPLEX setup. These findings strongly suggest that the allosteric network is more rigid in the APO setup, whereas the presence of the ligands allows the signal to propagate more efficiently through additional regions. The average degeneracies per secondary structure extracted from the sub-optimal paths are in Fig. 4. In the monomer, the source (L8) and the sinks (L6/H5, and H12) are highly degenerate (>0.3) meaning the allosteric signal has to propagate through the entire secondary structures that make up the start and end points in the paths [see Figs. 4(a) and 4(c)]. For the communications across the NBD, we found that H6, which contains PL2, has

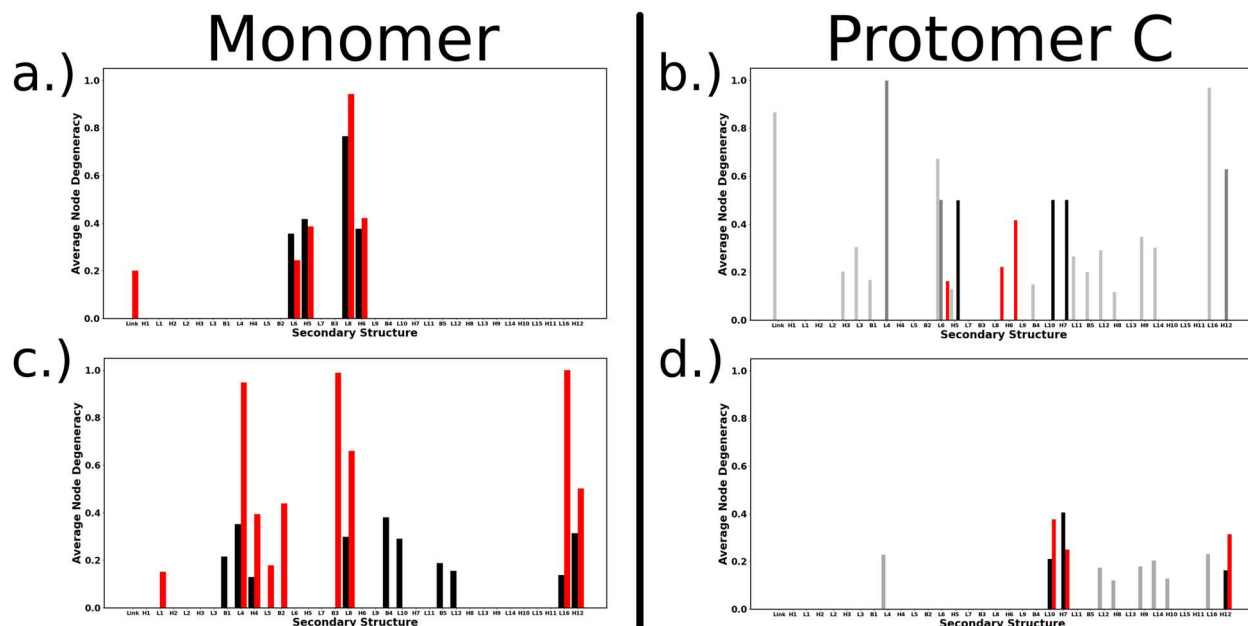


FIG. 4. The relative average node degeneracy per secondary structure element, according to Table S2. The degeneracy for the PL1 sink for the (a) monomer represents the COMPLEX setup in black and the APO setup in red. The degeneracy for the PL1 sink for protomer C from the hexamer (b) represents the COMPLEX setup in gray scale (light gray-A, medium gray-B, and black-C) and the APO setup in red. The degeneracy for the CT helix sink was similarly represented for the monomer (c) and for protomer C (medium gray-B and black-C).

high degeneracy in both the COMPLEX and APO setups, indicating that PL2 is an obligatory region for passing allosteric signals inside the NBD from the allosteric center to PL1. In contrast, for the communications between the NBD and the HBD, we found that the Walker A and Walker B motifs were highly degenerate in both setups. This indicates that the allosteric communication between the two domains in the monomer must pass through the ATP binding site. We also found that B4 was highly degenerate in the COMPLEX setup, while B2, B3, and L16 were highly degenerate in the APO setup.

The results for the hexamer setups [optimal paths in Tables S13 and S14, sub-optimal paths represented by their path length distributions in Figs. S14–S17, and average degeneracies for secondary structure elements in Figs. 4(b) and 4(d) and Figs. S18–S32] show that, while in the monomer the source and the sinks were both highly degenerate, regardless of the destination, in the hexamer, this is true only for the communications inside the NBD. For communications with the HBD, only the sink (H12) is highly degenerate. Thus, upon the formation of inter-protomer interfaces, the entire region that contains the allosteric center remains important for signal propagation only inside the NBD. Instead, for the HBD, the communication is now facilitated by the interfaces. Moreover, similar to the monomer, in the hexamer, we found that, when one setup had a much higher degeneracy than the other setup, both the optimal and the distributions of the sub-optimal paths shifted to larger lengths in the highly degenerate setup, which is indicative of a more rigid allosteric network. In the hexamer, the binding of both ligands enhances the allosteric communications between the NBD and the HBD domains inside and between the four end protomers (A, B, and E, F). For the communications between regions in the NBD, the allosteric paths in the COMPLEX setup are shortened compared to the APO setup primarily inside the two end protomers (A and F) and in their neighbors. These findings support a model where the changes in the terminal protomers due to the binding of ligands play an active role in the force generation in spastin, with the interior protomers providing a mostly scaffolding role.⁶⁹ In addition, the proposed communication between the substrate binding sites (PL1 and PL2) in the hexamer³⁸ occurs from the allosteric center in A to F which, irrespective of the sink, passes through PL1 in protomers B, C, and D, and PL2 in protomer E.

For the propagation of the allosteric signals inside the NBD, we found that PL2, which was crucial in the monomer, continues to be highly important for protomers B and C, but only in the APO setup. PL2 is the most important (highly degenerate) in protomer E for the propagation of allosteric signals to PL1, particularly in the COMPLEX setup. Protomer E is the only chain that behaves similarly to the monomer in the hexamer. Interestingly, the role played by PL2 in the monomer shifts to the ATP pocket at the interface in protomers B, C, and F, and in their pairs with neighboring end chains (B to A, and F to E), as these ATP binding regions show very high degeneracy regardless of the hexamer setup. As discussed earlier, the allosteric center (R591) has been proposed to connect the ATP binding site to the substrate binding site and indirectly to the inter-protomer interfaces based on the hexameric structure.³⁸ Our results provide important quantitative insight into this connectivity: in the monomer and protomer E from the hexameric state, the two binding sites are independently linked with the allosteric center

because the signal from the source is required to pass only through the other main substrate binding site, PL2, to get to PL1. In contrast, for the majority of protomers in the spiral hexameric assembly, the two binding sites become highly dependent on each other. This point is enhanced even further by our finding that H7, which contains the arginine fingers from the ATP binding pocket, and PL3, which connects the substrate binding site to the ATP binding site, as seen in Fig. 5, are also highly degenerate in protomer A, regardless of the ligand setup. The same behavior is present also in protomers B and C, and the pairs made by these protomers with their $i-1$ neighbors in the COMPLEX setup. Finally, we found that protomers A, B, and C have high inter-connectivity in the COMPLEX setup as the propagation of signals from the allosteric center B to its PL1 region passes through A, from the allosteric center of C to its PL1 region passes through both A and B, and from the allosteric center of C to the PL1 of B always passes through A.

For the propagation of allosteric signals between the NBD and the HBD in the spiral hexamer, we found that, similarly to the communication inside the NBD, H7 and PL3 are highly degenerate in paths from protomers A, B, C, and F and from their pairs with the $i-1$ neighbor (B to A, C to B, and F to E), regardless of the ligand setup, and in protomer E in the APO setup. These results and the

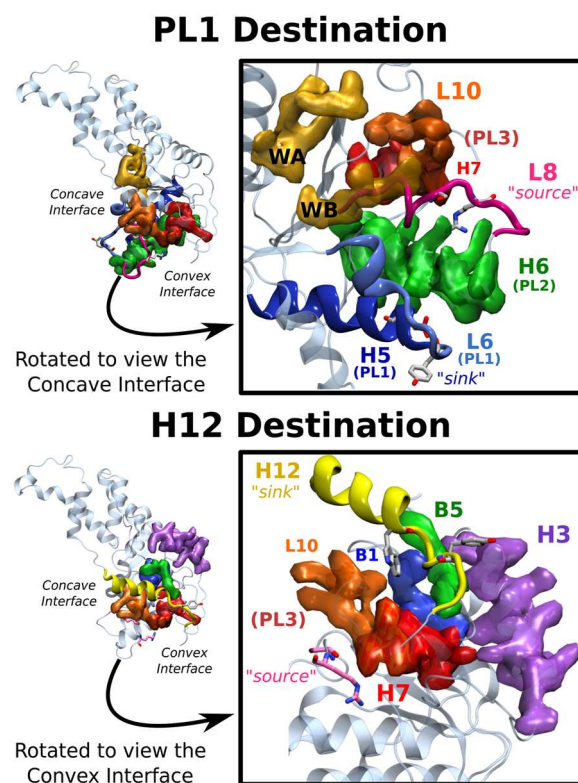


FIG. 5. The regions identified to be essential for passing intra-protomer allosteric signals are depicted and labeled for the CTT binding region destination (top) and the CT Hlx destination (bottom). The source (S589/R591) is in pink. The region was found to be essential for passing information to PL1, but not to CT Hlx. The sink for the PL1 destination (Y556/E561) is in blue, and the sink for H12 (W749/Y753) is in yellow.

fact that these regions are not important in the monomer indicate that, in the hexamer, H7 and PL3 are the obligatory regions for passing allosteric signals both inside the NBD domains and between the NBD and the HBD domains and they are activated by the formation of the inter-protomer interfaces. A region of special interest is H3, identified in Fig. 5, which is highly degenerate in the B to A paths, regardless of the ligand setup, and thus we identify it as necessary for passing information through the interface between the first two protomers. We also found B5 to be highly degenerate, regardless of the ligand setup, in protomer A, and in the APO setup for protomer B. Finally, L16, which we previously identified as highly degenerate in the monomer APO setup and identified in Fig. S33, was also highly degenerate in protomer E. This indicates again that protomer E is the only one that resembles the monomer. Furthermore, L16 was highly degenerate in the i-1 neighbor of protomer B in both setups and of protomer C in the COMPLEX setup. Because L16 forms part of the concave interface between protomers, our results indicate that it facilitates the communication of allosteric signals across the concave interfaces between the NBD and HBD domains of the first three protomers, similar to the above finding for paths ending in PL1. Thus, for the transmission of allosteric signals between the NBD domains and between the NBD and HBD domains, the presence of both ligands induces strong inter-cooperativity between protomers A, B, and C.

E. Classifying binding partner transitions using biochemically relevant descriptors

1. Significant feature identification

We selected descriptors for each spastin system that was effective in recognizing the four major setups (see Results in the [supplementary material](#) for details) in order to determine important regions for spastin's allostery. The descriptors were: *RSA* and *VDW* for all systems, *COULOMB* for both the monomer and spiral hexamer, and *CONTACTS* for both hexamers. We focused on locating allosteric signals through the classification of pairs between the four setups of spastin (APO, NUC, SUB, and COMPLEX) since we know that the binding of ATP and the substrate plays a critical role in spastin's function.⁴⁰ Using XGBoost to learn which secondary structures characterize the transitions from APO to NUC/SUB or COMPLEX in the monomer and from COMPLEX to NUC/SUB or APO in the hexamer, we can determine if the allosteric communication within spastin depends on the type of binding partner.

Overall, we found that the ML methodology identifies almost every one of the known experimental allosteric regions from spastin (see [Table II](#) and [Table S4](#)) as critical for describing the conformational transitions due to adding binding partners. We note that any known allosteric region not identified through this method was due to the initial removal of the feature because of its proximity to the binding pocket of interest (see Methods). Due to the monomer's small size compared to the hexamers, the monomer's transitions between states yielded a larger variety of the experimentally identified regions than the hexamers. Instead, the hexamers identified secondary structures that were important within multiple chains. Importantly, the most prevalent features in all systems, which were found to be important in experiments as well, were L1 and L3. As listed in [Table II](#), mutations to L1 reduced ATPase activity and

MT severing rates, while L3 (L2 in the ring) was critical for spastin oligomerization.

2. Direction of significant feature changes characterizes allosteric transitions for spastin monomer

A powerful aspect of the ML-based approach is that, in addition to its ability to yield allosterically relevant information from select descriptors, it can also reveal the direction of descriptor transitions for each system of spastin when using SHAP beeswarm plots ([Figs. 6–8](#) and [S28–34](#)). The beeswarm plots show the distributions of SHAP values calculated for every feature observation within a set of setup pairs (ex: APO and SUB). Our beeswarm plots are with respect to feature transitions that helped classify the final state; therefore, points found with positive SHAP values were observations that persuaded XGBoost to select the final state of a transition. Each point is also colored according to where the magnitude of an observation's raw descriptor value falls with regard to the average of that feature's total observations. Magenta points with a positive SHAP value mean an increase in that features' descriptor upon the addition or removal of a binding partner, while blue indicates a decrease upon the binding/perturbation.

Allosteric signals identified through changes in the spastin monomer's *RSA* features due to the binding of one or both ligands ([Fig. 6](#)) were overwhelmingly found within the NBD, especially when binding the nucleotide. The APO to NUC transition was characterized by the increase of *RSA* for structures known experimentally to be critical for the ATPase and severing function (H2, H5, H6, and L1). Because H5 and H6 are helices adjacent to PL1 and PL2, respectively, our results suggest that the binding of ATP primes the monomer for substrate binding. Overall, this transition showed more features decrease in *RSA*, seemingly loop regions that become more buried as a result of the movement of the nearby helices that become more exposed. We found that B2 has an allosteric effect in nucleotide binding for *RSA*, *VDW* ([Fig. S38](#)), and *COULOMB* ([Fig. S37](#)), even if it was not identified through experimental mutation studies. Not only does this beta strand, found in the center of the NBD, become more buried but also significantly more stable upon binding of the nucleotide. Both energetic descriptors pulled out more regions in the HBD compared to *RSA* as being important to classify the nucleotide-binding transition. Interestingly, *COULOMB* showed that H10, the helix that makes up the majority of the inner face of the HBD, and H12, which communicates with both inter-protomer interfaces in the hexamer, became more electrostatically stable upon nucleotide binding. This increase in stability infers the nucleotide's role in the formation of the interface between the HBD of a protomer and the NBD of the next protomer, which we have previously found critical for oligomerization.¹⁹

The allosteric communication within the monomer switches in the APO to SUB transition to showing an increase in *RSA* for more features than upon binding the nucleotide. The entire Walker-A motif (L4 and H4) became more exposed to the solvent, indicating that substrate binding affects the allosteric network necessary for regulating ATP binding. Many allosteric signals for the energetic descriptors are also found within regions surrounding the ATP-binding pocket, with the majority showing high levels of instability. Coupling this internal energetic destabilization of the monomer

RSA

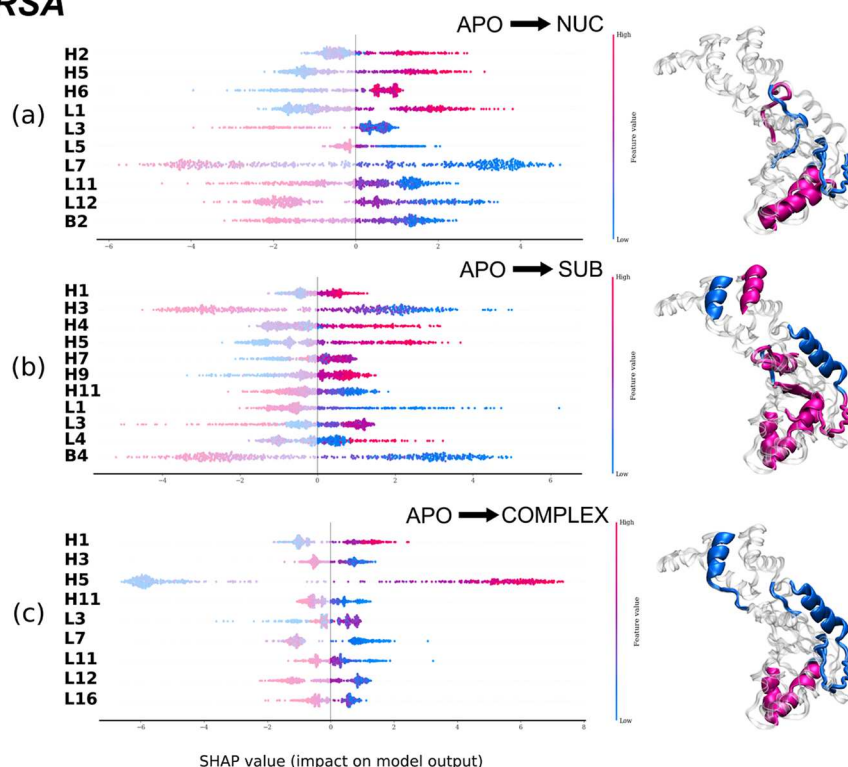


FIG. 6. SHAP beeswarm plots for the three transitions of RSA for the monomer. Features displayed were the common features between transition and/or spastin systems. Feature points calculated to have a positive SHAP value explain whether RSA increases (magenta) or decreases (blue) upon binding of either the nucleotide or the substrate. Representative figures for each transition are shown on the right with colors corresponding to feature magnitudes of positive SHAP values.

upon substrate binding with the opposite effect due to nucleotide binding strongly suggests that the monomer preferentially binds the nucleotide. The formation of the COMPLEX in the beeswarm plots for *VDW* and *COULOMB* also shows that the monomer responds to the binding of both cofactors mostly by the energetic destabilization of both the NBD and HBD. This result suggests that the formation of inter-protomer interfaces through oligomerization is likely needed before a successful binding of the substrate to spastin.

3. Comparing allosteric transitions upon perturbation of spastin hexamers

The analysis of the allosteric signals found when perturbing the spiral and ring hexamer of spastin allowed us to determine the dependencies that each conformation of the hexamer has on the nucleotide and substrate. We found that the connection between monomers B and C strengthened when the substrate was removed in the COMPLEX to NUC transition, while the interface between monomers E and F was weakened. This can be seen in Fig. 7, where almost all the critical features that increased in contacts were located between monomers B and C, including the H10 of monomer B, which forms the HBD-NBD contacts between protomers. The *COULOMB* descriptor (Fig. S39) shows an increase in stabilizing electrostatic interactions between A and B, along with regions between B and C. *VDW* (Fig. S40) and *RSA* (Fig. S41) display contrasting trends

within monomer A, compatible with monomer A's HBD moving away from monomer B. This could be a reaction to monomer F's enhanced motion caused by its separation from E (shown again in *RSA*). The removal of the nucleotide leads to the opening of the central gate in the COMPLEX to SUB transition for *RSA*. The *CONTACTS* show the same event, with multiple structures near the hexamer's pore gained more inter-protomer contacts. This is found, for example, in H6 (PL2) and shows that the loss of the nucleotide may also restrict spastin's interaction with the substrate since more contact between pore loops to the various protomers means less involvement within the central pore. Finally, the loss of both cofactors resulted in more loss of contacts and higher instability in energetic features. Interestingly, *CONTACTS* showed loss of interactions between B and C, C and D, and E and F. Agreeing with the energetic descriptors trends, the spiral hexamer in the APO setup tends to separate into the AB and DE dimers, with C and F left as monomers.

Destabilizing the COMPLEX ring hexamer resulted in varied trends compared to the spiral hexamer. This is crucial to investigate since the ring hexamer gives insight into spastin's allostery following ATP-hydrolysis and loss of nucleotide density within protomer F. The motions of the ring (shown above in the MSM results), along with changes in energy and/or contacts between interfaces, can help us study the motor's potential change in allosteric communication after the severing event has occurred. Overall, the descriptors for

CONTACTS

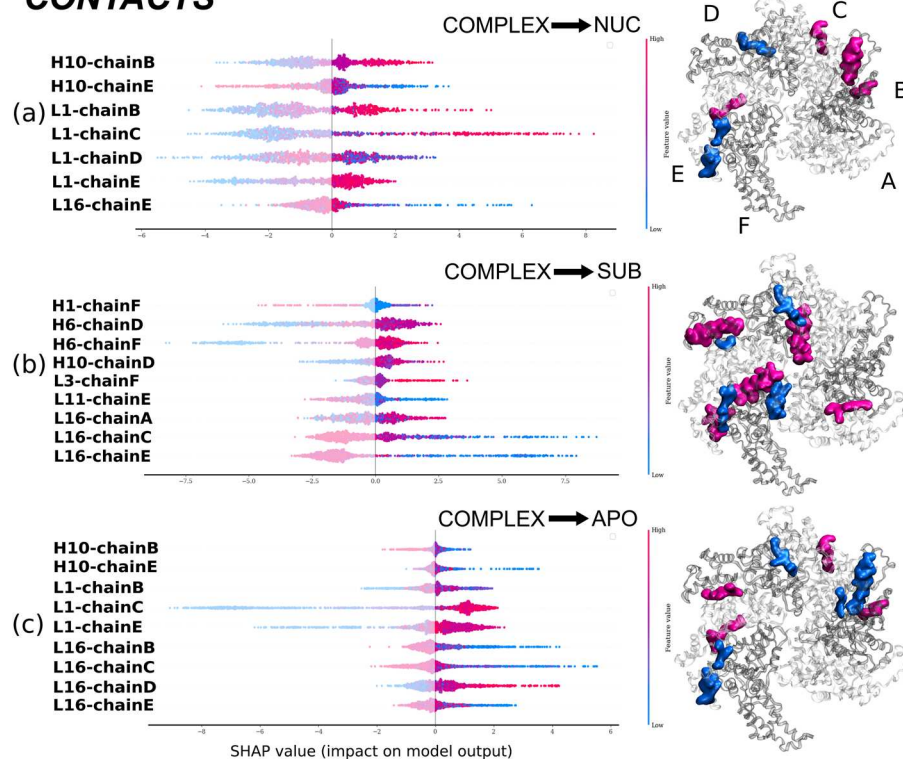


FIG. 7. SHAP beeswarm plots for the three transitions of *CONTACTS* for the spiral hexamer, similar to Fig. 6.

CONTACTS

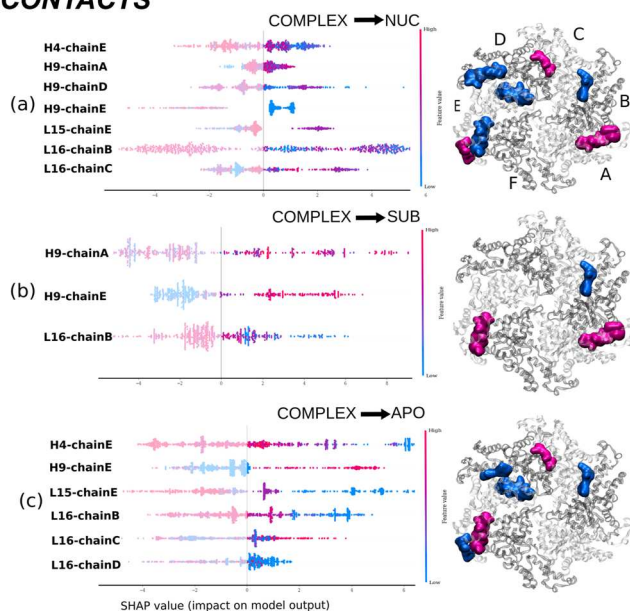


FIG. 8. SHAP beeswarm plots for the three transitions of *RSA* for the ring hexamer, similar to Fig. 6.

the ring hexamer identified L16 as being important for characterizing all transitions within multiple chains. This loop is located along the concave interface of each protomer and, therefore, can signify interfaces that will remain or dissociate upon perturbation of the hexameric structure. While *CONTACTS* focused on the grouping of dimers (AB and CD), energetic descriptors suggest also the grouping of EF in the coupling of dimers formed following hydrolysis of spastin (Figs. 8 and S42-43).

F. Overlap between top features for transitions in the ML models and high degeneracy regions for allosteric communications

An interesting discovery resulting from the path analysis and ML-based top feature extraction methods described above was that multiple secondary structures in spastin are found as both highly degenerate within the network paths and critical for feature transitions of the classification models. This finding is not only further validation of the level of success of the methods, given that many of these regions were already known to be experimentally important, but also allows us to observe what types of descriptor changes can affect the pathways from the allosteric center to either the NBD or HBD. This in turn can guide future efforts to design allosteric effectors that can either enhance or block allosteric signaling along specific paths.⁷ Paths with the sink in the CT-Hlx of monomer A of the spiral hexamer, which pass through PL3, are affected by changes in

solvent accessibility (RSA) and energetics (VDW). For the hexameric paths with the sink in PL1 that pass through PL2, we propose that they will also be affected by ATP binding because we found that the *CONTACTS* showed PL2 moving away from the central pore when the nucleotide was removed. Walker-A and Walker-B were features identified only through *COULOMB* and *VDW*, therefore, showing that paths that pass through these motifs to reach the sink in PL1 will mainly be affected by effectors that induce energetic changes. Notably similar behavior is likely for the spastin monomer. Interestingly, L16 is also highlighted as important in both analyses, with both the spastin spiral hexamer and the monomer showing paths passing through this loop to the sink in the CT-Hlx. L16 makes up the concave interface between protomers and it is highlighted in all four descriptors, meaning that effectors that target the interactions between protomers in the hexamer, or the energetics of the protomers, or the interactions between protomers and the solvent (through the RSA) are all going to be good candidates to induce changes in the propagation of allosteric signals from the NBD to the HBD through L16. Finally, we found that B5 and H3 were commonly found to be highly degenerate in the hexamer paths to the HBD: paths including H3 could be affected by *RSA*, *COULOMB*, and *VDW*, while paths including B5 would be affected only by energetic changes.

IV. CONCLUSIONS

Allostery has been recognized as being important for the proper severing action of the spastin machine on MTs and several positions have been identified experimentally as likely to play a role in the propagation of allosteric signals in the spastin hexameric states.^{38,40} We previously showed that residues from the central pore loops of the spiral and ring hexameric states of spastin, which are involved in the formation of long-lived salt bridges in MD simulations, are important for allostery.⁴² Our DCCM and principal component analysis (PCA) based analysis of MD simulations for lower order oligomers of spastin and of the spastin monomer showed that select helices from the HBD are likely allosteric hotspots, responding to the binding of not only the nucleotide and the minimal substrate but also to the formation of inter-protomer interfaces in oligomers.¹⁹ Here, the application of various levels of methodology to the long MD simulations from our previous work^{19,42} enabled us to characterize the potential allosteric sites, to determine the regions in the spastin monomer and hexamer that are likely obligatory nodes for the propagation of allosteric signals between a source and a sink, and to characterize the direction of structural and energetic changes of spastin regions that are found, by machine learning approaches, to best describe the allosteric transitions.

The results of the features selection using ML approaches show that this methodology is overwhelmingly the best in identifying experimentally proposed allosteric/functional sites: it selects all the regions corresponding to the residues from [Table II](#) and [Table S4](#) in the [supplementary material](#) among the top features. The application of SHAP analysis to understand the top features that drive the learning in XGBoost allows not only the explanation of where certain allosteric signals are occurring but also the biochemical response to those regions described by the descriptor changes in the SHAP beeswarm plots. Here, we were able to connect the critical secondary

structures found upon the addition of cofactors to the monomer and the perturbation of the COMPLEX spiral and ring hexamer. Challenges related to the use of this method include the necessity in optimization of removing less important features to ensure that the classification models are not overfitting and the importance of removing regions that are in direct contact with the ligand so as to not bias the supervised models. The path analysis of the dynamic cross-correlation based graph networks also identified a large portion of the experimentally listed positions. The optimal paths from this network provided a map of the most preferred route for an allosteric signal to travel through the various protomers and setups. However, the path analysis is prone to overinterpretation if relying solely on the shortest path. The analysis of the statistical distribution of signaling pathways and the related node degeneracy holds more weight. Indeed, we found that the sub-optimal paths and the node degeneracies arising from their analysis were instrumental in the identification of the allosteric regions that are most likely to be involved in signal propagation throughout the protein structure.^{55,59} For this type of analysis, special care needs to be taken regarding the number of paths considered for each pair of positions to adequately represent the networks.⁵⁵

The analysis of the MD-MSM composition of the MSM macrostates and the transitions between macrostates in the spastin monomer, combined with the analysis of the direction of change in the top ML features for the monomer transitions, strongly suggest that the simultaneous binding of both the nucleotide and the substrate to the spastin monomer is not the favored route to the formation of the COMPLEX setup as it requires many structural changes. Instead, the monomer favors the binding of ATP, which primes the regions involved in the formation of both the concave and the convex inter-protomer interfaces for binding to other protomer(s). The substrate will then be favored to bind upon the formation of an oligomer. We also analyzed the major structural changes captured by the representative structures corresponding to the MSM macrostates of the hexamer. We found that, for both spiral and ring, these motions are the same as the principal component motions from our previous work,⁴² which resulted in axial fluctuations of the pore loops. This is important because these coordinated motions resemble the pushing and pulling actions that underlie the proposed death spiral MT severing mechanism of the spastin motor.^{31,66}

We used the graph network (PSG) based on the average connectivity extracted from the representative MSM macrostates to pinpoint the positions characterized by high contact density, called hubs, which are fundamental for the scaffolding of the structure. This analysis highlighted positions Q583, R591, and R601 from [Table II](#) as hubs only in the full hexamer, not in the monomer. This indicates that the inter-protomer contacts involving these three positions are particularly tight and long-lived and perturbations leading to the loss of these contacts would destabilize the hexameric structure. The analysis of optimal and sub-optimal paths in the dynamic network connecting one of these residues (the proposed allosteric center, 591, located in the NBD) to two different sinks, taken from [Table II](#), one in the NBD and the other in the HBD can shed light on the essential regions involved in the propagation of the allosteric signal resulting from such perturbations throughout the structure. We employed the graph network weighted based on the directional cross-correlations between residues in either the monomer or the

spiral hexamer APO and COMPLEX setups, which accounts for the longtime dynamics captured by our MD simulations. We found that, in the hexamer, H7, from the ATP binding pocket, and PL3, which connects the substrate binding site to the ATP binding site, are the obligatory regions for passing allosteric signals both inside the NBD domains and between the NBD and the HBD domains and they are activated by the formation of the inter-protomer interfaces. Moreover, based on the inter-protomer path analysis, our findings support a model where the changes in the terminal protomers due to the binding of ligands play an active role in the force generation in spastin, with the interior protomers providing a mostly scaffolding role.⁶⁹ An important discovery resulting from the path analysis and the ML-based top feature extraction methods was that the secondary structures in spastin, which are highly degenerate within the network paths, are also critical for feature transitions of the classification models. This enables the identification of the types of descriptor changes that can affect the pathways from the allosteric center to either the NBD or the HBD. In turn, this new type of analysis can guide future efforts to design allosteric effectors that can either enhance or block allosteric signaling along specific paths.⁷ Finally, our findings show that the best approach to determine the allosteric regions and the intricacies of the signal propagation involving these regions in a large oligomeric molecular machine, such as spastin, is to combine the predictive and explanatory power of the ML-based approaches with the insights into the allosteric signal propagation provided by the analysis of the hubs, centrality, and suboptimal paths of graph networks.

SUPPLEMENTARY MATERIAL

The [supplementary material](#) contains additional methods and analysis from results: Tables S1–S3: secondary structures designation in spastin hexamers. Table S4: list of HSP mutations. Table S5: top pMI residues. Table S6: normalization factors for I_{ij} calculation. Table S7: hubs in the PSG networks. Tables S8 and S9: average I_{ij} values for select hubs. Tables S10–S12: top centrality residues. Table S13–S16: optimal paths. Table S17: convergence data for SHAP analysis. Tables S18 and S19: RMSD and distance changes for MSM states analysis. Figures S1 and S2: MSM results for hexamers. Figure S3: optimization data for MSMs. Figure S4: $I_{critical}$ data. Figure S5: amino acid composition from MSA. Figure S6: Average cMI data. Figure S7: Average pMI data. Figures S8 and S9: Representations of top pMI and hubs positions. Figures S10 and S11: Top centrality positions in the monomer. Figures S12–S17: Optimal paths and sub-optimal paths distributions. Figures S18–S32: Node degeneracy in sub-optimal paths. Figure S33: Representation of high-degeneracy regions for inter-protomer allosteric signaling paths. Figure S34: Descriptor distributions. Figures S35 and S36: Confusions matrices and AUROC data for ML analysis. Figures S37–S43: SHAP Beeswarm plots for monomer and the hexamers.

ACKNOWLEDGMENTS

We thank Rohith Anand Variokti for their help with setting up the GROMACS simulations for spastin. We thank George Stan and Ashan Dayananda for stimulating discussions and interpretation of results. This research was funded by the National Science

Foundation (NSF) [Grant No. MCB-1817948 (to R.I.D.)]. This work used the Extreme Science and Engineering Discovery Environment (XSEDE) through the allocation Grant No. TG-BIO210094 to R.I.D.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Maria S. Kelly: Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Software (equal); Supervision (equal); Validation (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal). **Amanda C. Macke:** Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Software (equal); Supervision (equal); Validation (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal). **Shehani Kahawatte:** Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Software (equal); Validation (equal); Visualization (equal); Writing – original draft (supporting). **Jacob E. Stump:** Software (supporting). **Abigail R. Miller:** Data curation (supporting); Formal analysis (supporting); Investigation (supporting); Software (supporting); Visualization (supporting). **Ruxandra I. Dima:** Conceptualization (lead); Data curation (equal); Formal analysis (equal); Funding acquisition (equal); Investigation (equal); Methodology (equal); Project administration (lead); Resources (lead); Supervision (lead); Validation (equal); Writing – original draft (equal); Writing – review & editing (equal).

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

- 1 A. P. Kornev and S. S. Taylor, *Trends Biochem. Sci.* **40**, 628 (2015).
- 2 A. Cooper and D. T. F. Dryden, *Eur. Biophys. J.* **11**, 103 (1984).
- 3 P. Csermely, R. Palotai, and R. Nussinov, *Trends Biochem. Sci.* **35**, 539 (2010).
- 4 J. Zhang, F. J. Adrián, W. Jahnke, S. W. Cowan-Jacob, A. G. Li, R. E. Jacob, T. Sim, J. Powers, C. Dierks, F. Sun, G.-R. Guo, Q. Ding, B. Okram, Y. Choi, A. Wojciechowski, X. Deng, G. Liu, G. Fendrich, A. Strauss, N. Vajpai, S. Grzesiek, T. Tuntland, Y. Liu, B. Bursulaya, M. Azam, P. W. Manley, J. R. Engen, G. Q. Daley, M. Warmuth, and N. S. Gray, *Nature* **463**, 501 (2010).
- 5 M. Leander, Y. Yuan, A. Meger, Q. Cui, and S. Raman, *Proc. Natl. Acad. Sci. U. S. A.* **117**, 25445 (2020).
- 6 A. J. Faure, J. Domingo, J. M. Schmiedel, C. Hidalgo-Carcedo, G. Diss, and B. Lehner, *Nature* **604**, 175 (2022).
- 7 G. M. Verkhivker, S. Agajanian, G. Hu, and P. Tao, *Front. Mol. Biosci.* **7**, 136 (2020).
- 8 F. Morcos, A. Pagnani, B. Lunt, A. Bertolino, D. Marks, C. Sander, R. Zecchina, J. Onuchic, T. Hwa, and M. Weigt, *Proc. Natl. Acad. Sci. U. S. A.* **108**, 1293 (2011).
- 9 D. Granata, L. Ponzoni, C. Micheletti, and V. Carnevale, *Proc. Natl. Acad. Sci. U. S. A.* **114**, 10612 (2017).
- 10 Q. Tang and K. Kaneko, *Phys. Rev. Lett.* **127**, 098103 (2021).
- 11 J. Zhu, J. Wang, W. Han, and D. Xu, *Nat. Commun.* **13**, 1661 (2022).
- 12 S. W. Lockless and R. Ranganathan, *Science* **286**, 295 (1999).

- ¹³G. M. Süel, S. W. Lockless, M. A. Wall, and R. Ranganathan, *Nat. Struct. Biol.* **10**, 59 (2003).
- ¹⁴C. L. McClelland, G. Friedland, D. L. Mobley, H. Amirkhani, and M. P. Jacobson, *J. Chem. Theory Comput.* **5**, 2486 (2009).
- ¹⁵R. I. Dima and D. Thirumalai, *Protein Sci.* **15**, 258 (2006).
- ¹⁶N. V. Dokholyan, *Chem. Rev.* **116**, 6463 (2016).
- ¹⁷G. Stetz and G. M. Verkhivker, *PLoS Comput. Biol.* **13**, 1 (2017).
- ¹⁸M. T. Posgai, S. Tonddast-Navaei, M. Jayasinghe, G. M. Ibrahim, G. Stan, and A. B. Herr, *Proc. Natl. Acad. Sci. U. S. A.* **115**, E8882 (2018).
- ¹⁹A. Macke, M. S. Kelly, R. A. Varikoti, S. Mullen, D. Groves, C. Forbes, and R. I. Dima, *J. Phys. Chem. B* **126**, 10569 (2022).
- ²⁰O. Lange and H. Grubmüller, *Proteins* **4**, 1053–1061 (2006).
- ²¹M. J. Bradley, P. T. Chivers, and N. A. Baker, *J. Mol. Biol.* **378**, 1155 (2008).
- ²²A. R. Atilgan, P. Akan, and C. Baysal, *Biophys. J.* **86**, 85 (2004).
- ²³L. Astl and G. M. Verkhivker, *J. Chem. Theory Comput.* **15**, 3362 (2019).
- ²⁴G. M. Verkhivker, *J. Phys. Chem. B* **126**, 5421 (2022).
- ²⁵G. Palermo, C. G. Ricci, A. Fernando, R. Basak, M. Jinek, I. Rivalta, V. S. Batista, and J. A. McCammon, *J. Am. Chem. Soc.* **139**, 16028 (2017).
- ²⁶J. Wang, A. Jain, L. R. McDonald, C. Gambogi, A. L. Lee, and N. V. Dokholyan, *Nat. Commun.* **11**, 3862 (2020).
- ²⁷Y. Yuan, J. Deng, and Q. Cui, *J. Am. Chem. Soc.* **144**, 10870 (2022).
- ²⁸H. S. Hayatshahi, E. Ahuactzin, P. Tao, S. Wang, and J. Liu, *J. Chem. Inf. Model.* **59**, 4691 (2019).
- ²⁹H. Zhou, Z. Dong, G. Verkhivker, B. D. Zoltowski, and P. Tao, *PLoS Comput. Biol.* **15**, 1 (2019).
- ³⁰F. J. McNally and A. Roll-Mecak, *J. Cell Biol.* **217**, 4057 (2018).
- ³¹V. Barsegov, J. L. Ross, and R. I. Dima, *J. Phys.: Condens. Matter* **29**, 433003 (2017).
- ³²H. Han, H. L. Schubert, J. McCullough, N. Monroe, M. D. Purdy, M. Yeager, W. I. Sundquist, and C. P. Hill, *J. Biol. Chem.* **295**, P435 (2020).
- ³³A. J. Enright, I. Iliopoulos, N. C. Kyripides, and C. A. Ouzounis, *Nature* **402**, 86 (1999).
- ³⁴T. Paysan-Lafosse, M. Blum, S. Chuguransky, T. Grego, B. L. Pinto, G. A. Salazar, M. L. Bileschi, P. Bork, A. Bridge, L. Colwell, J. Gough, D. H. Haft, I. Letunić, A. Marchler-Bauer, H. Mi, D. A. Natale, C. A. Orengo, A. P. Pandurangan, C. Rivoire, C. J. A. Sigrist, I. Sillitoe, N. Thanki, P. D. Thomas, S. C. E. Tosatto, C. H. Wu, and A. Bateman, *Nucleic Acids Res.* **51**, D418 (2022).
- ³⁵N. A. Lynn, E. Martinez, H. Nguyen, and J. Z. Torres, *Front. Cell Dev. Biol.* **9**, 692040 (2021).
- ³⁶S. N. Gates and A. Martin, *Protein Sci.* **29**, 407 (2019).
- ³⁷Y.-W. Kuo, O. Trottier, M. Mahamdeh, and J. Howard, *Proc. Natl. Acad. Sci. U. S. A.* **116**, 5533 (2019).
- ³⁸C. R. Sandate, A. Szyk, E. A. Zehr, G. C. Lander, and A. Roll-Mecak, *Nat. Struct. Mol. Biol.* **26**, 671 (2019).
- ³⁹A. Roll-Mecak and R. D. Vale, *Curr. Biol.* **15**, 650 (2005).
- ⁴⁰A. Roll-Mecak and R. D. Vale, *Nature* **451**, 363 (2008).
- ⁴¹F. Sievers, A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Soding, J. D. Thompson, and D. G. Higgins, *Mol. Syst. Biol.* **7**, 539 (2011).
- ⁴²M. Damre, A. Dayananda, R. A. Varikoti, G. Stan, and R. I. Dima, *Biophys. J.* **120**, 3437 (2021).
- ⁴³L. Pauling, R. B. Corey, and H. R. Branson, *Proc. Natl. Acad. Sci. U. S. A.* **37**, 205 (1951).
- ⁴⁴L. Pauling and R. B. Corey, *Proc. Natl. Acad. Sci. U. S. A.* **39**, 253 (1953).
- ⁴⁵A. Mardt, L. Pasquali, and F. Noe, *Nat. Commun.* **9**, 4443 (2018).
- ⁴⁶T. Löhr, K. Kohlhoff, G. T. Heller, C. Camilloni, and M. Vendruscolo, *Nat. Comput. Sci.* **1**, 71 (2021).
- ⁴⁷F. L. Simonetti, E. Teppa, A. Chernomoretz, M. Nielsen, and C. M. Buslje, *Nucleic Acids Res.* **41**, W8 (2011).
- ⁴⁸S. D. Dunn, L. M. Wahl, and G. B. Gloor, *Bioinformatics* **24**, 333 (2008).
- ⁴⁹C. M. Buslje, J. Santos, J. M. Delfino, and M. Nielsen, *Bioinformatics* **25**, 1125 (2009).
- ⁵⁰D. Aguilar, B. Oliva, and C. Marino Buslje, *PLoS One* **7**, e41430 (2012).
- ⁵¹C. M. Buslje, E. Teppa, T. D. Domenico, J. M. Delfino, and M. Nielsen, *PLoS Comput. Biol.* **6**, e1000978 (2010).
- ⁵²E. Teppa, A. D. Wilkins, M. Nielsen, and C. M. Buslje, *BMC Bioinform.* **13**, 235 (2012).
- ⁵³A. Hagberg, S. Pieter, and D. S. Chult, in *Proceedings of the 7th Python in Science Conference (SciPy, 2008)*, p. 11.
- ⁵⁴K. V. Brinda and S. Vishveshwara, *Biophys. J.* **89**, 4159 (2005).
- ⁵⁵A. T. Van Wart, J. Durrant, L. Votapka, and R. E. Amaro, *J. Chem. Theory Comput.* **10**, 511 (2014).
- ⁵⁶Q. Cui, J. Deng, and Y. Yuan, *J. Am. Chem. Soc.* **144**, 10870 (2022).
- ⁵⁷W. Humphrey, A. Dalke, and K. Schulten, *J. Mol. Graph.* **14**, 33 (1996).
- ⁵⁸A. Tse and G. M. Verkhivker, *PLoS One* **10**, e0130203 (2015).
- ⁵⁹T. Dodd, X.-Q. Yao, D. Hamelberg, and I. Ivanov, *Mol. Phys.* **119**, 19 (2021).
- ⁶⁰C. Li, J. Liman, Y. Eliaz, and M. S. Cheung, *J. Phys. Chem. B* **125**, 11591 (2021).
- ⁶¹J. Y. Yen, *Manage. Sci.* **17**, 712 (1971).
- ⁶²D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen, *J. Comput. Chem.* **26**, 1701 (2005).
- ⁶³E. Kliuchnikov, E. Klyshko, M. S. Kelly, A. Zhmurov, R. I. Dima, K. A. Marx, and V. Barsegov, *CSBJ* **20**, 953 (2022).
- ⁶⁴P. Zhutovsky, Probatas, 2020. <https://github.com/ing-Bank/probatas/>.
- ⁶⁵K. M. Thayer, B. Lakhani, and D. L. Beveridge, *J. Phys. Chem. B* **121**, 5509 (2017).
- ⁶⁶R. A. Varikoti, H. Y. Y. Fonseka, M. S. Kelly, A. Javidi, M. Damre, S. Mullen, J. L. I. Nugent, C. M. Gonzales, G. Stan, and R. I. Dima, *Nanomaterials* **12**, 1849 (2022).
- ⁶⁷S. Chakrabarti and A. R. Panchenko, *Proteins: Struct., Funct., Bioinf.* **75**, 231 (2009).
- ⁶⁸S. Chakrabarti and A. R. Panchenko, *PLoS One* **5**, e8591 (2010).
- ⁶⁹A. N. Rizo, J. Lin, S. N. Gates, E. Tse, S. M. Bart, L. M. Castellano, F. DiMaio, J. Shorter, and D. R. Southworth, *Nat. Commun.* **10**, 2393 (2019).