THE JOURNAL OF PHYSICAL CHEMISTRY B
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCB

Perspective

From Distinct to Differential Conformational Dynamics to Map Allosteric Communication Pathways in Proteins

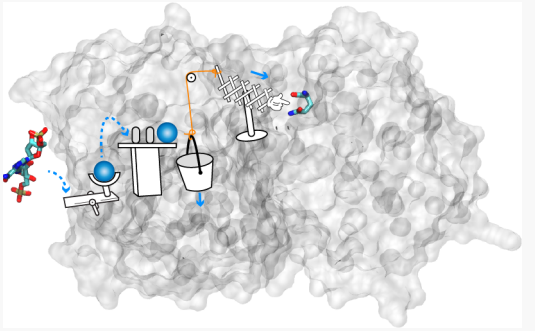
Xin-Qiu Yao* and Donald Hamelberg*

Cite This: *J. Phys. Chem. B* 2022, 126, 2612−2620

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: Initiation of biological processes involving protein−ligand binding, transient protein−protein interactions, or amino acid modifications alters the conformational dynamics of proteins. Accompanying these biological processes are ensuing coupled atomic level conformational changes within the proteins. These conformational changes collectively connect multiple amino acid residues at distal allosteric, binding, and/or active sites. Local changes due to, for example, binding of a regulatory ligand at an allosteric site initiate the allosteric regulation. The allosteric signal propagates throughout the protein structure, causing changes at distal sites, activating, deactivating, or modifying the function of the protein. Hence, dynamical responses within protein structures to stimuli contain critical information on protein function. In this Perspective, we examine the description of allosteric regulation from protein dynamical responses and associated alternative and emerging computational approaches to map allosteric communication pathways between distal sites in proteins at the atomic level.



1. INTRODUCTION

Proteins are intrinsically flexible, not only "jiggling and wiggling" around the native conformation in a particular state (e.g., ligand-free) but also can undergo conformational changes in response to stimuli such as ligand binding, protein−protein interactions, and amino acid modifications. The dynamic property of proteins is described by the collection of the conformations, also known as the conformational ensemble, under certain conditions (temperature, pressure, ligand concentration, etc.) that obeys the Boltzmann distribution. Upon receiving a stimulus, the protein relaxes to a new thermodynamic equilibrium (described by an altered conformational ensemble) and the transition is, to some extent, to oppose the changes caused by the stimulus (Le Chatelier's principle). Many physiological stimuli on proteins are in general related to the initiation and fine-tuning of biological processes, including enzyme catalysis, signal transduction, and allosteric regulation. Hence, understanding the mechanism of how a protein system responds to changes by external conditions is essential for understanding protein function and related biological processes.

The fact that a protein adjusts its conformations (shifting of the ensemble) in response to stimuli indicates the importance of comparing distinct conformational ensembles (e.g., before and after regulatory ligand binding) for understanding protein function. In other words, approaches exploring only a single conformational ensemble derived from, for example, a single molecular dynamics (MD) simulation (termed single-ensemble approach here and below) may be insufficient to understand protein function. In extreme cases, only a single protein conformation (e.g., crystal structure) may be considered. While issues associated with using a single conformation are obvious and are related to the omission of protein motions (Figure 1A), the use of a single ensemble of conformations presents its own challenges and problems in trying to understand protein function in certain situations that may not be so obvious.

In this Perspective, we identify issues that could arise from using single ensemble (or conformation) approaches and discuss solutions, with a focus on protein allosteric regulation. We start by discussing the fundamentals of the conformational selection and population shift model[1−3] and how it may be generalized. The comparative perturbed-ensembles analysis approach[4] developed under the generalized view is described and serves as a general solution to addressing issues arising from using single ensemble (or conformation) approaches to study allostery at the atomic level. We then discuss, at a more technical level, potential issues of building allosteric networks using single-ensemble-based computational methods and describe an alternative difference contact network analysis (dCNA) method[5,6] as a potential solution. We conclude that the general framework of the comparative perturbed-ensembles analysis and

Received: January 10, 2022
Revised: February 17, 2022
Published: March 23, 2022

https://doi.org/10.1021/acs.jpcb.2c00199
J. Phys. Chem. B 2022, 126, 2612−2620

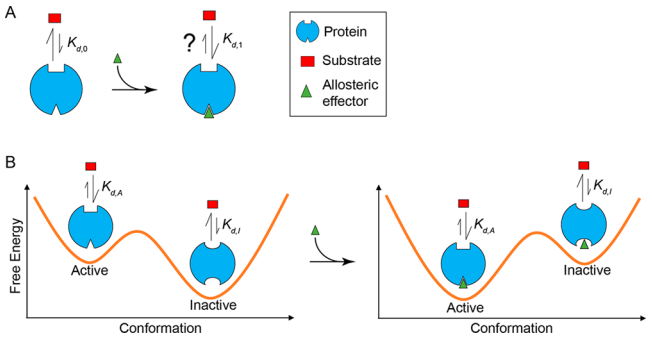


Figure 1. A flexible protein molecule is a prerequisite for allosteric regulation. (A) A rigid protein conformation does not explain, for example, the change of the substrate binding affinity by the binding of the allosteric effector (e.g., $K_{d,0} > K_{d,1}$, where $K_{d,x}$ is the dissociation constant for the state x and $x = 0$ or 1 representing the effector free or bound state, respectively). (B) A simple two-state example showing how the conformational selection and population shift model can explain allosteric regulation. The orange curve indicates the free energy landscape of the protein, which is altered by the binding of the allosteric effector. $K_{d,y}$, $y = A$ or I, is the microscopic dissociation constant for the active or inactive conformational substate, respectively.

dCNA are sensible tools for mapping allosteric communication pathways in proteins.

2. CONFORMATIONAL SELECTION AND POPULATION SHIFT MODEL AND AN EXTENDED VIEW

A popular phenomenological theory about protein motions and their relationship to function is the conformational selection and population shift model.[1−3] This model is rooted in the (free) energy landscape theory that was originally proposed to explain protein folding.[7−9] A major assumption of the model is that all functional relevant conformations pre-exist in a single functional state of the protein, such as the ligand-free state, with relative free energies dictated by the local free energy minima defined as conformational substates.[10] In accordance with this model, ligand binding, for example, does not create any new conformation but simply selects one of the pre-existing conformations, causing a population shift of the conformational ensemble toward this selected conformation.

The conformational selection and population shift model can be used to explain allosteric regulation: in the simplest case, a protein possesses two conformational substates, represented by an active and an inactive conformation, where the former has a higher binding affinity for both the substrate and the regulatory ligand or allosteric effector (assuming a positive cooperativity) than the latter (Figure 1B). In the ligand-free state, the protein primarily populates around the inactive conformation but can access the active conformation with a small probability, thus exhibiting an overall low apparent binding affinity when only the substrate is present. In the presence of the allosteric effector, ligand binding causes a population shift toward the active conformation and so the apparent binding affinity for the substrate is enhanced (Figure 1B).

Although the conformational selection and population shift model emphasizes the shift or "change" of the conformational ensemble (the population shift description) and hence the importance of ensemble comparison, single-ensemble approaches seem to ignore the shift in the ensemble and rely heavily on the assumption of the preexistence of all relevant conformations (the conformational selection description) and the knowledge of their roles. The idea of these approaches is to "decode" the functional information that is already present in the single (ligand-free state, for example) ensemble. This type of approach may be subjected to the following three limitations. (1) It is not generally guaranteed that all relevant conformations are accessible in one state. The conformation relevant to the ligand-bound state, for example, may be of high free energy in

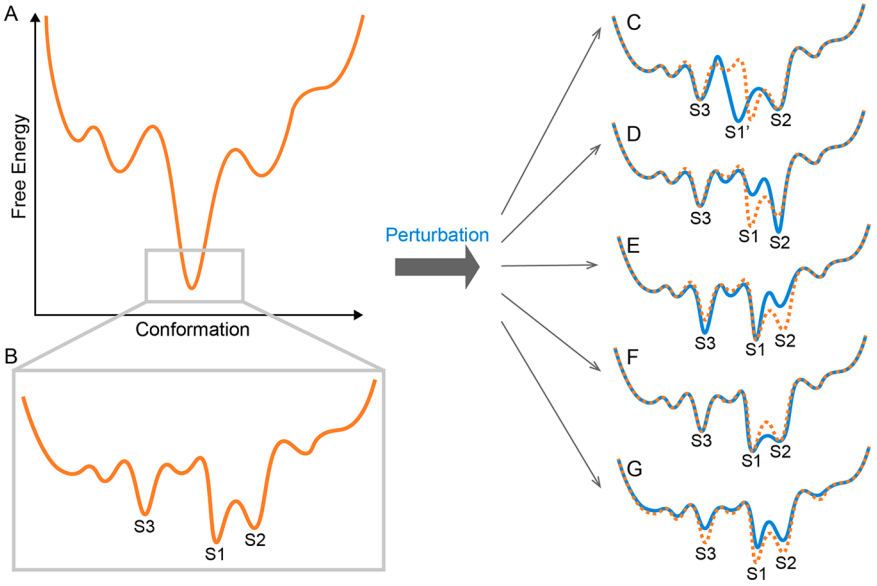


Figure 2. Schematic of a general case of the free energy landscape and its responses to stimuli or perturbation. (A) The "funnel" like free energy landscape of a protein, where the lowest-energy well represents the native folded conformation. (B) Enlarged view of the native well showing multiscale ruggedness. The three lowest-energy wells (S1, S2, and S3) represent function-related conformational substates. (C) A new substate (S1′) is created upon perturbation, exhibiting an "induced-fit" scenario. (D) Population shift causing a switch of the most stable substate. (E) Population shift that does not change the most stable substate. (F) Fine tuning of the transition between substates (kinetics) and the configurational entropy. (G) Modulation of the overall configurational entropy without changing the mean conformation.

https://doi.org/10.1021/acs.jpcb.2c00199
J. Phys. Chem. B 2022, 126, 2612−2620

the ligand-free state. Examples are the binding of large ligand molecules as suggested previously.[11,12] In these cases, the induced-fit mechanism[13] may dominate. Also, even if all functional conformations are thermodynamically (meta)stable (reside in local free energy minima) in the free state, the access to them may be restricted by the limited computational sampling. (2) Even if the sampling is exhaustive and all relevant conformations are accessible in the free state, the dynamic information for a specific functional activity (or process, such as binding) can be elusive if the underlying free energy landscape is rugged, containing multiple conformational substates related to distinct protein activities. For example, these substates may be related to binding of different ligands or the protein activation by post-translational modifications. Simple statistics, such as ensemble average, of the free state mixes these distinct conformational substates, resulting in either misleading conclusions or signals too weak to decode. Advanced statistics, such as a conformational clustering followed by a functional annotation of identified clusters (from prior experiments, for example) may alleviate the issue, but it would only be applicable to a few very well studied biological systems. (3) Even if all relevant conformations are accessible and all substates are functionally annotated, the information obtained by comparing relevant conformational substates in the free state does not tell how exactly the protein responds to specific stimuli. For example, the protein may retain the same set of conformational substates but exhibit different directions and magnitudes of the population shift upon binding regulatory ligands with varying efficacy. Without comparing distinct ensembles directly, it is hard to explain the direction (activator or inhibitor) and magnitude (full or partial agonist) of the allosteric effect of a regulatory ligand.

In general, we should make no prior assumption about the shape of the free energy landscape and how it may change upon responding to stimuli (see Figure 2 for a schematic of a general case of the free energy landscape, with multiscale ruggedness and multiple conformational substates near the native-folded conformation, and diverse types of the response to a stimulus or perturbation). The mechanism of a complex system is best understood by experimenting on the response of the system to various stimuli. Similarly, to understand protein function, the best way is to put the protein system under various conditions and observe the response of the conformational ensemble of the protein, analogous to how many experimental approaches work in understanding biological function. For example, one approach (i.e., comparative perturbed-ensembles analysis) is to perform multiple MD simulations under distinct ligand binding, mutation, phosphorylation, protonation, *cis*−*trans* isomerization, or other modification conditions (in this case, simulations are treated as doing computer experiments); the generated conformational ensembles are then compared as postsimulation analyses.[4] This approach has been applied to understand the allosteric regulation in diverse biological systems.[14−18]

An important concept here is that it is the change or difference between two or more ensembles that defines the functional mechanism, providing additional details to supplement the traditional "structure-activity" view of protein function. The change can be large-scale backbone conformational changes (change of the mean conformation), side-chain reorientations, subtle conformational or energetic twisting, or increase/decrease of conformational fluctuations. Irrespective of what metric is used, it is always the change of the quantity that should be examined to elucidate the mechanism of protein function. By directly sampling the ensembles of interest, relevant conformational substates automatically stand out. Meanwhile, by ensemble comparison, irrelevant quantities (e.g., structural properties shared by all substates or specific to irrelevant substates) may be canceled out. Hence, in this general framework of the approach, no conformational clustering or prior functional annotation of substates is required, although conducting such a type of analysis is not prohibited by the framework. This suggests that the complexity of free energy landscape has little influence on the performance of the approach, unlike single-ensemble approaches as discussed above.

Note that we do not oppose the statement that a single ensemble (from any state) could determine all functional activities of the protein. What we argue is that it is very difficult to understand these activities, at least for some of them including allosteric regulation, without directly comparing relevant ensembles. Theoretically, the ensemble of any state can be reconstructed from an existing ensemble by the reweighting process,[19] no matter if the system follows the conformational selection hypothesis or not, as long as the perturbation can be modeled as a term in the energy function, which depends only on the protein conformation. In practice, however, a "sufficient" sampling for the unperturbed state is usually insufficient with respect to the perturbed state. So, approximations would have to be made during the reweighting or the perturbation should be very small. No matter how it is implemented, the approach that reproduces an ensemble from another is in essence the same as direct sampling: both are to generate distinct ensembles to be compared as a follow-up study. Hence, single ensemble approaches, we worry about here, are those that only characterize (structurally or energetically) a single ensemble without an apparent process of generating altered ensembles using physical methods. In our view, ensemble comparison is the first step of decoding the structure−dynamics−function relationship. Once the (atomic) conformational changes are established through comparison of the ensembles, the same analysis tools that are employed in single-ensemble approaches (such as those from graph theory) can be used in a more effective manner because the inputs (conformational changes) are now more functionally relevant.

3. COMPUTATIONAL METHODS FOR MAPPING ALLOSTERIC COMMUNICATION PATHWAYS

In this section, we discuss specific examples of the potential problem of using single-ensemble approaches to map allosteric communication pathways. We focus on computational methods that use concepts and algorithms from network science to describe and analyze protein structures in understanding the mechanism of allosteric communications. Popular methods include the protein structure network and similar methods.[20−33] The philosophy behind these methods is that functional activities, including allosteric regulation, can be learned solely from the "wiring diagram" or connectivity of the protein structure defined by, for example, residue−residue contacts. The implementation of these methods to interpret allosteric regulation can be categorized into two main types: (1) building a single network from a single conformational ensemble (or a single conformation) and (2) comparing two or more networks, with each network built from a single, distinct ensemble (or conformation).

3.1. Single Network to Interpret Allosteric Regulation.

In most studies using traditional network analysis methods, the

https://doi.org/10.1021/acs.jpcb.2c00199
J. Phys. Chem. B 2022, 126, 2612−2620

network is built from a single protein conformation or a single conformational ensemble. These studies assume the structure−activity relationship that is rooted in the organic chemistry. In organic chemistry, the chemical structure of a molecule is used to predict the biological activity of the molecule (the so-called SAR or QSAR analysis). We argue that proteins are fundamentally different from small organic molecules and so the analogy is inappropriate. The major difference is that protein molecules are much more complex and flexible, possessing diverse conformational substates near the native-folded conformation. The use of a single X-ray crystallographic structure (possibly corresponding to the substate of the lowest free energy), for example, cannot capture the global picture of the free energy landscape. A typical example of the problem is when the perturbation (e.g., mutation) does not change the conformation of the lowest free energy but alters the population distribution among the substates (i.e., the population-shift scenario; see Figure 2E). In this case, the (lowest energy) structure will be the same, and the single-conformation approach cannot explain, for example, the different binding affinities of the same ligand toward the wildtype and the mutant with very similar crystallographic structures.

Even with the entire conformational ensemble, the allosteric mechanism may still be elusive. As we discussed in Section 2, even though all the protein activities are determined by the conformational ensemble of the protein in one state, relevant information is not readily obtained by inspecting just one ensemble. The current expectation of graph theory and the so-called "contact-network language" is probably too optimistic.[34] In a typical contact network analysis, the residue−residue contact is calculated for each individual sampled conformational snapshot, and contacts with high frequency (above a certain threshold) are chosen to define the network edges, whereas contacts with low frequency are removed. Hence, the network mainly represents the stable structural features of the protein that are largely invariant among different functional states. An alternative strategy is to build the network for each frame of conformation in the ensemble and summarize over all network parameters.[23] In both cases, the critical dynamical information about the structural basis for the allosteric transition, which is mainly defined by the malleable degree of freedoms that are either excluded because of the instability or hidden among the many stable, ensemble-invariant contacts, is obscure. In a sense, this single-ensemble approach follows a "dressed up" structure−activity view, where the structure here represents the core stable protein conformation within an ensemble. Using this approach automatically assumes that the allosteric transition occurs without a conformational change (the so-called dynamic or entropy driven allostery).[35] Clearly, this is not generally true because many proteins undergo significant conformational changes (and so alterations of the topology of the contact network) to some extent (e.g., as depicted by the MWC and KNF models of allostery).[36,37] Even in some special cases (if they exist) where a conformational change is not required for allostery, the network mainly represents a stable structure most likely relevant to the protein folding stability. Hence, it may only make sense to use these methods if the system uses the same set of degree of freedoms to perform both folding and function, which is not necessarily true in general.

3.2. Comparing Networks to Interpret Allosteric Regulation. Sometimes, in traditional network analysis-based studies, distinct ensembles are compared but the comparison is made at the network level. In this type of studies, each network is built based on a single conformation or a single conformational ensemble, and some network parameters or descriptors are compared to understand the allosteric mechanism. In this case, the network analysis methods "translate" protein structures to the "graph language" that, when compared between different states, is believed to better describe the allosteric regulation. A typical example is the comparison of the partition of communities (or parameters defined based on the partition) between two networks.[38,39] One opinion suggests that the "translation" into communities can help get global changes from subtle changes of the conformation.[39] We argue that this statement is overly optimistic and neglects the alternative possibility that the "translation" may amplify errors. The source of error can be insufficient sampling of limited-time MD simulations or uncertainty of the empirical parameter chosen to define the partition. An example of the latter is the use of modularity to determine the optimal partition of communities. In some cases, the separation of the maximal modularity from other values may not be very sharp, and so distinct partitions with the optimal and suboptimal modularity scores may all be relevant. Hence, comparing communities without carefully checking the robustness of the results by, for example, comparing multiple MD simulation replicas and alternative suboptimal partitions, is risky and may lead to overinterpretation of the results.

Another issue with comparison of communities is the difficulty in understanding the meaning of such a comparison. In many cases, the partitions are so different (in terms of both the number and the definition of communities) that it is hard to derive any useful information from the comparison. A special note here is the potential pitfall of associating the pattern of residue clustering with physical interactions. Some may easily consider the split of a large community to be due to the weakened inter-residue interactions or coupling within the community. This could be misleading because the pattern of residue clustering is purely dependent on relative residue−residue interactions within the network; residues belonging to two separate communities may have stronger interactions than the same set of residues grouped in one community in another network simply because the overall residue−residue interactions in the latter are weaker.

There may also be a paradox in comparing network parameters. For example, some studies compared shortest path-based network parameters to delineate the allosteric communication pathway.[23,33] Here, a path is defined by consecutive edges connecting the two end nodes in the network and represents a potential route for the allosteric signal to transmit within the protein. In accordance with this definition, the comparison of path-based parameters can only be explained as the change of the allosteric mechanism or communication pathway, which goes back to the discussion in Section 3.1. On the other hand, if the comparison, or the resulting differential path parameters, is to define the allosteric pathway, it would be rather elusive to know the meaning of individual microscopic paths and signals they transmit.

3.3. Solutions by Difference Contact Network Analysis. The above discussion about traditional network analysis methods indicates the importance of comparing conformational ensembles derived from distinct states. We stress that it is the change in the ensemble between the active and inactive states, for example, that defines the allosteric mechanism, not that the allosteric mechanism is changing. Such an ensemble change should be investigated at the atomic/residue level with simple,

https://doi.org/10.1021/acs.jpcb.2c00199
J. Phys. Chem. B 2022, 126, 2612−2620

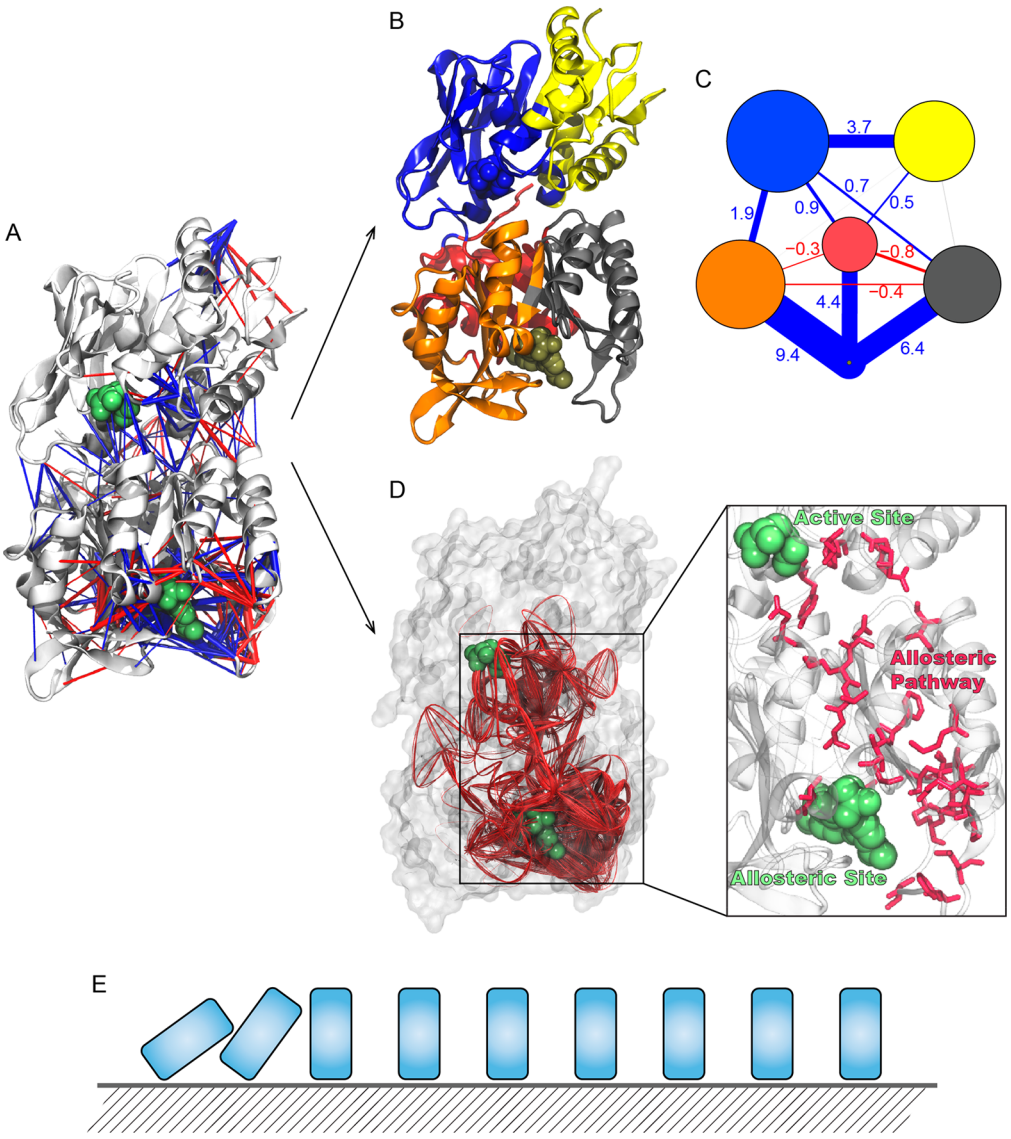


Figure 3. The dCNA-based community and path analyses identify the allosteric pathway defined by dynamic residues. (A) The most important input of dCNA is a set of residue−residue contact changes displayed as colored cylinders mapped on the protein (IGPS) structure shown as a white cartoon. The two ensembles compared are the ensemble where the substrate (Gln; top green spheres) is present and the allosteric effector (PRFAR; bottom green spheres) is absent and the ensemble where both ligands are present. Radii of cylinders are scaled by the magnitude of contact changes (df) and colors by the sign of changes: blue, contact strengthening; red, contact weakening. (B,C) The dCNA-based community analysis. Communities are defined by colored regions on the protein structure (B). The community level network (C) is a coarse-grained representation of the contact changes. Communities are indicated by colored vertices (as in B), with the radius of vertex proportional to the number of residues in the corresponding community. The line linking vertices describes the net contact change ($\sum$df) between the communities. Blue and red lines indicate positive and negative changes (with $\sum$df labeled), respectively, and the line width is determined by $|\sum \mathrm{d}f|$. Gray lines indicate the magnitude of the net contact change smaller than the threshold (e.g., $|\sum \mathrm{d}f| < 0.1$). (D) The dCNA-based path analysis. The protein is represented as white transparent surface and (sub)optimal paths from the allosteric ligand to the substrate as red lines. The enlarged view shows the key dynamic residues (red licorices) defining the allosteric communication pathway, identified by high (≥0.1) node degeneracies from the path analysis, where the protein is represented as a white transparent cartoon. (E) The "Domino model" of allostery. The pattern of the domino chain reaction is defined by the dominos themselves, not the sustaining ground or table (shaded area).

easy to understand physiochemical descriptions. Examples of this type of study include investigations on differential residue−residue contact maps (or the alternative atomic overlap-and-relief analysis using alternative conformations).[40−45] Following the general framework of the comparative perturbed-ensembles analysis approach and prior contact-based methods, we recently developed a new type of network analysis method called dCNA.[5,6] The main difference between dCNA and the traditional network methods is that dCNA uses multiple ensembles as the input of the network formalism. Similar ideas, but implemented in different ways, have been reported by other groups.[46−48] In the following, we discuss two typical network calculations using dCNA: the community analysis and the suboptimal path analysis. We show fundamental conceptual differences between dCNA and the traditional network methods and how the differences may help resolve some of the issues in the traditional methods.

A fundamental difference between the dCNA-based community analysis and the traditional methods lies in how the network is viewed. In dCNA, residue communities are simply

https://doi.org/10.1021/acs.jpcb.2c00199
J. Phys. Chem. B 2022, 126, 2612−2620

considered as a coarse-grained representation of the protein structure. There is no special "functional" meaning of the specific pattern of the partition of communities. Rather, it is the edge between communities that matters. In dCNA, communities are detected from a consensus network defined by stable contacts across all relevant ensembles. In this sense, the general definition of community here is kept the same, i.e., it represents an invariant collective part of the protein structure, similar to a subdomain. However, different from the traditional methods, a community−community edge in dCNA is defined by the net intercommunity contact probability change from one ensemble to the other, calculated from all residues belonging to the two communities. Hence, in dCNA, structural and dynamic parts are explicitly separated, and the community-level network provides an overview of the conformational change underlying the allosteric transition. By contrast, in traditional single-ensemble community analyses, either the two parts are mixed or only the structural part is considered.

The partition of communities in dCNA is consistent across all ensembles under consideration. This consistency is important for an effective ensemble comparison as illustrated before.[32,49,50] This consistent partition also allows a hierarchical examination of the conformational changes. For example, by tuning up or down the number of communities, one can zoom in or out of the protein structure and obtain dynamic information at various levels of resolution. This type of analysis is impossible in traditional community calculations, because in the traditional methods the partition of communities is inconsistent across ensembles and so there is no universal definition of the level of resolution.

An example of results by dCNA-based community analysis is illustrated in Figure 3A−C, which is performed on the well-studied imidazole glycerol phosphate synthase (IGPS) using the simulation data described previously.[6] IGPS has two ligand binding sites: the active site that catalyzes the hydrolysis of glutamine (Gln or substrate binding site) and the allosteric site (PRFAR or effector binding site), located in two different subunits and separated by ∼30 Å. Effector binding largely enhances the enzyme catalysis in the active site by ∼4900 times. Here, two conformational ensembles are compared, which were generated from simulations with and without the allosteric effector in the presence of the substrate. Six communities are detected: two are in the substrate binding subunit (including Gln), three are in the effector binding subunit, and one contains the effector only. The community analysis results clearly show that upon effector binding, there is a global conformational rearrangement involving residue−residue contact strengthening (blue edges in Figure 3C) within the substrate binding subunit and in the subunit−subunit interface, along with slight contact weakening (red edges in Figure 3C) within the effector binding subunit. By contrast, using the traditional community analysis would result in two separate community networks, where both the community partition and edges between communities are subject to change from one network to the other.

The network for the dCNA-based suboptimal path analysis is directly built on residue−residue contact probability changes from one ensemble to the other (e.g., Figure 3A). The central idea is that the allosteric signal propagates throughout the protein structure via concerted or sequential local conformational changes described by contact breakages and formations. By specifying two end sites (e.g., the allosteric and active sites) and searching for the optimal (shortest in terms of the number of edges and associated contact changes composing the path) and multiple suboptimal paths connecting the end sites, the structural basis for the allosteric communication between the two end sites can be established. Compared to traditional path analysis methods, dCNA is more sensible and has a clearer physical basis because it incorporates the critical dynamic information about the allosteric transition, and the result of dCNA is more robust and easier to interpret.

An example of results by dCNA-based path analysis is illustrated in Figure 3D, which identifies the residue wise allosteric pathway in IGPS. The results are from our previous study.[6] In brief, we have calculated 5000 (sub)optimal paths connecting the allosteric and the active site. Each residue is evaluated by the fraction of the number of paths going through the corresponding node in the network (i.e., normalized node degeneracy). The allosteric communication pathway is defined by residues with high ($\geq$0.1) degeneracy values. Similarly, we have identified allosteric communication pathways using conventional methods (e.g., the protein contact network). An important finding is that dCNA identifies key residues that are experimentally shown to disrupt the allosteric communication upon mutations; these residues are not detected by the conventional protein contact network method no matter the conformational ensemble used. These key residues are found to primarily play a dynamic role in the allosteric communication, which explains their omissions in the conventional method that, by its nature, mainly identifies residues crucial for the structural integrity (i.e., playing a structural role).

A special note is given here to the definition of the allosteric communication pathway. Proteins need both stable and dynamic parts to perform their functions. For allosteric regulation, a protein needs both residues that actively participate in the propagation of the allosteric signal (i.e., playing a dynamic role) and residues that are stable and mainly to sustain the structural scaffold where the propagation is carried out (structural role). Both types of residues are crucial for the allosteric regulation; mutations of both may cause a disruptive effect. Hence, experimental mutagenesis data do not tell if the identified residues are of a structural or dynamic role in the distal communication. Similar arguments exist for the sequence conservation or coevolution analysis: highly conserved residues or coevolving residue pairs may be due to interactions of multiple factors that determine the fitness of the molecule and so may not be specifically related to allosteric regulation.

We argue that even though structural residues are equally important, only dynamic residues define the communication pathway. With the use of the "Domino model" of allostery[51] as an example (Figure 3E) here, an analogy is made between the domino chain reaction and the propagation of the allosteric signal in protein structures. Dominos can be regarded as "dynamic residues," which have two (standing or lying) states and directly define the pattern of the chain reaction, whereas the ground or table is the group of "structural residues" that sustain the dominos but do not directly participate in the propagation of the signal. This definition of pathway has an apparent advantage in pharmaceutical applications: novel allosteric drugs may be developed to provide direct intervention of the pathway residues so that fine-tuning of the protein activity can occur without disturbing the native active and allosteric sites or the structural basis of the communication and, hence, may have a minimal adverse drug effect.

We stress that dCNA is compatible with an error analysis to distinguish significant observations from noises. One example is to use bootstrapping to estimate the statistical errors, as

https://doi.org/10.1021/acs.jpcb.2c00199
J. Phys. Chem. B 2022, 126, 2612−2620

previously described.[6] In brief, simulation trajectories can be split into multiple nonoverlapping chunks. By randomly picking up (inclusively, i.e., the same chunk can be picked up more than once) the chunks, ensembles of the same size as those from the original trajectories can be reconstructed with slightly different statistics. The same, for example, path analysis procedure can be performed on the new ensembles and a new set of node degeneracy can be generated. This random process can be repeated multiple times and the standard deviation of degeneracy over the random experiments can be used as an estimate of the error. A similar method can be used to estimate the errors of the community−community edges in the dCNA-based community analysis. Besides, it is essential to perform an accompanying checking on the convergence of, for example, residue−residue contact statistics. Simulations must be long enough to ensure that the variation of cumulative contact probabilities over time is smaller than some predefined threshold, which usually requires microsecond or longer simulations as previously shown.[6]

4. CONCLUSIONS AND FUTURE DIRECTIONS

In this Perspective, we contrast, philosophically, two types of computational approaches to understanding allosteric regulation in biomolecules. One relies on the "structure−activity" way of thinking and considers only a single conformation, or a conformational ensemble generated under the same conditions. The reason to put single-ensemble approaches into the "structure−activity" category is because in many applications the most stable structural properties (e.g., high-frequency residue−residue contacts) are extracted (through ensemble average, for example) for further processing, which follow the same principle in using a single conformation. Although more advanced statistical analyses may help dissect the structure of the ensemble and enable direct comparisons between conformational substates, these analyses do not explicitly tell how the ensemble is altered when responding to the change of external conditions, preventing a full understanding of the functional process. This type of approach assumes the preexistence of all relevant functional information in a single ensemble that is readily derived (the "conformational selection" hypothesis). Examples of methods developed along this type include the protein structure or contact network analysis and the dynamical network analysis.

By contrast, an alternative type of approach is to compare two or more ensembles, each generated under distinct conditions, explicitly and stress the critical importance of "difference" or "change" during a process (e.g., ligand binding) for understanding function and allosteric regulation. The change during the process upon binding the allosteric effector is akin to the chain reaction in a Rube Goldberg machine or falling dominos. Examples include the comparative perturbed-ensembles analysis, which values more on "population shift" and is applicable to general cases of the free energy landscape and its responses to stimuli. The network analysis method developed under this general framework (i.e., dCNA) applies graph theory to the elementary, physically meaningful local conformational changes at the atomic or residue level. The dCNA is fundamentally different from the traditional ensemble comparison at the network level, and the result generated is more robust and easier to interpret. Importantly, the allosteric pathway identified by dCNA consisting of primarily dynamic residues is more sensible and more useful in pharmaceutical applications.

There are multiple directions for future development and applications of dCNA. One straightforward extension of dCNA is to calculate many other commonly used network parameters, including various types of residue centrality that have been used to identify functionally important residues without any prior knowledge about, for example, the active and allosteric sites (a test of concept on the betweenness centrality has been performed in our previous study[6]). A user-friendly web application of dCNA will allow the use of the method for the broader community with little to zero programming experience, which is under construction. The dCNA-based path analysis identifies new potential "hotspots" to bind small molecules and hence can be combined with, for example, molecular docking and virtual screening for drug discovery. Finally, dCNA is general enough to be applied to all biomolecular allosteric systems and diverse types of allosteric effects elicited by not only allosteric effector binding but also mutations, post-translational modifications, and other physiochemical modifications and to handle not only simulation data but also experimentally resolved conformational ensembles (e.g., by X-ray crystallography, nuclear magnetic resonance, and cryogenic electron microscopy).

AUTHOR INFORMATION

Corresponding Authors

Xin-Qiu Yao − *Department of Chemistry, Georgia State University, Atlanta, Georgia 30302-3965, United States;* orcid.org/0000-0003-2706-2028; Phone: (404) 413-5564; Email: xyao4@gsu.edu

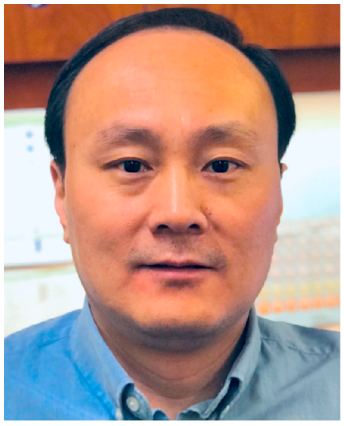
Donald Hamelberg − *Department of Chemistry, Georgia State University, Atlanta, Georgia 30302-3965, United States;* orcid.org/0000-0002-3785-3037; Email: dhamelberg@gsu.edu

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpcb.2c00199

Notes

The authors declare no competing financial interest.

Biographies



Xin-Qiu Yao is an Adjunct Assistant Professor and a Postdoctoral Associate in the Hamelberg lab at the Department of Chemistry, Georgia State University. His research focuses on developing and applying computational tools to solve fundamental biological and biomedical problems. His main approach is to combine molecular simulations and structural bioinformatics to elucidate the atomistic mechanism for how biomolecular conformational dynamics determine function. Before joining Georgia State University, he also worked as a

https://doi.org/10.1021/acs.jpcb.2c00199
J. Phys. Chem. B 2022, 126, 2612−2620

research fellow at the University of Michigan and Kyoto University. He received his Ph.D. in Mechanics (Biomechanics and Medical Engineering) from Peking University in 2008.

Donald Hamelberg is a Professor of Chemistry at Georgia State University. The overarching goals of his research program are to use theory and develop computational methods to study allosteric mechanisms and probe long time scale dynamics of biomolecules in subcellular processes. Prior to joining the faculty at Georgia State University in 2008, he was a Research Scientist at the Howard Hughes Medical Institute and a Postdoctoral Research Fellow at the University of California, San Diego. He received his Ph.D. in Computational Biophysical Chemistry from Georgia State University in 2001.

■ ACKNOWLEDGMENTS

This research was supported by the National Science Foundation (MCB-2018144). We also acknowledge support from Georgia State University and the Georgia Research Alliance.

■ REFERENCES

(1) Tsai, C. J.; Ma, B.; Nussinov, R. Folding and binding cascades: shifts in energy landscapes. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 9970−9972.
(2) Ma, B.; Kumar, S.; Tsai, C. J.; Nussinov, R. Folding funnels and binding mechanisms. *Protein Eng.* **1999**, *12*, 713−720.
(3) Kumar, S.; Ma, B.; Tsai, C. J.; Sinha, N.; Nussinov, R. Folding and binding cascades: dynamic landscapes and population shifts. *Protein Sci.* **2000**, *9*, 10−19.
(4) Yao, X. Q.; Hamelberg, D. Detecting Functional Dynamics in Proteins with Comparative Perturbed-Ensembles Analysis. *Acc. Chem. Res.* **2019**, *52*, 3455−3464.
(5) Yao, X.-Q.; Momin, M.; Hamelberg, D. Elucidating Allosteric Communications in Proteins with Difference Contact Network Analysis. *J. Chem. Inf. Model.* **2018**, *58*, 1325−1330.
(6) Yao, X. Q.; Hamelberg, D. Residue-Residue Contact Changes during Functional Processes Define Allosteric Communication Pathways. *J. Chem. Theory Comput.* **2022**, *18*, 1173−1187.
(7) Bryngelson, J. D.; Wolynes, P. G. Spin glasses and the statistical mechanics of protein folding. *Proc. Natl. Acad. Sci. U. S. A.* **1987**, *84*, 7524−7528.
(8) Leopold, P. E.; Montal, M.; Onuchic, J. N. Protein folding funnels: a kinetic approach to the sequence-structure relationship. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89*, 8721−8725.
(9) Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G. Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins* **1995**, *21*, 167−195.
(10) Frauenfelder, H.; Sligar, S.; Wolynes, P. The energy landscapes and motions of proteins. *Science* **1991**, *254*, 1598−1603.
(11) Okazaki, K.; Takada, S. Dynamic energy landscape view of coupled binding and protein conformational change: induced-fit versus population-shift mechanisms. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 11182−11187.
(12) Cui, Q.; Karplus, M. Allostery and cooperativity revisited. *Protein Sci.* **2008**, *17*, 1295−1307.
(13) Koshland, D. E. Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proc. Natl. Acad. Sci. U. S. A.* **1958**, *44*, 98−104.
(14) Adeniran, C.; Hamelberg, D. Redox-Specific Allosteric Modulation of the Conformational Dynamics of kappaB DNA by Pirin in the NF-kappaB Supramolecular Complex. *Biochemistry* **2017**, *56*, 5002−5010.
(15) Momin, M.; Xin, Y.; Hamelberg, D. Allosteric Fine-Tuning of the Binding Pocket Dynamics in the ITK SH2 Domain by a Distal Molecular Switch: An Atomistic Perspective. *J. Phys. Chem. B* **2017**, *121*, 6131−6138.
(16) Rodriguez-Bussey, I.; Yao, X. Q.; Shouaib, A. D.; Lopez, J.; Hamelberg, D. Decoding Allosteric Communication Pathways in Cyclophilin A with a Comparative Analysis of Perturbed Conformational Ensembles. *J. Phys. Chem. B* **2018**, *122*, 6528−6535.
(17) Souffrant, M. G.; Yao, X.-Q.; Momin, M.; Hamelberg, D. N-Glycosylation and Gaucher Disease Mutation Allosterically Alter Active-Site Dynamics of Acid-β-Glucosidase. *ACS Catal.* **2020**, *10*, 1810−1820.
(18) Kumutima, J.; Yao, X. Q.; Hamelberg, D. p53 Is Potentially Regulated by Cyclophilin D in the Triple-Proline Loop of the DNA Binding Domain. *Biochemistry* **2021**, *60*, 597−606.
(19) Shen, T.; Hamelberg, D. A statistical analysis of the precision of reweighting-based simulations. *J. Chem. Phys.* **2008**, *129*, 034103.
(20) Amitai, G.; Shemesh, A.; Sitbon, E.; Shklar, M.; Netanely, D.; Venger, I.; Pietrokovski, S. Network analysis of protein structures identifies functional residues. *J. Mol. Biol.* **2004**, *344*, 1135−1146.
(21) del Sol, A.; Fujihashi, H.; Amoros, D.; Nussinov, R. Residues Crucial for Maintaining Short Paths in Network Communication Mediate Signaling in Proteins. *Mol. Syst. Biol.* **2006**, *2*, 2006.0019.
(22) Chennubhotla, C.; Bahar, I. Markov propagation of allosteric effects in biomolecular systems: application to GroEL-GroES. *Mol. Syst. Biol.* **2006**, *2*, 36.
(23) Ghosh, A.; Vishveshwara, S. A Study of Communication Pathways in Methionyl- tRNA Synthetase by Molecular Dynamics Simulations and Structure Network Analysis. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 15711−15716.
(24) Sethi, A.; Eargle, J.; Black, A. A.; Luthey-Schulten, Z. Dynamical Networks in tRNA:Protein Complexes. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 6620−6625.
(25) Vijayabaskar, M. S.; Vishveshwara, S. Interaction Energy Based Protein Structure Networks. *Biophys. J.* **2010**, *99*, 3704−3715.
(26) Rivalta, I.; Sultan, M. M.; Lee, N. S.; Manley, G. A.; Loria, J. P.; Batista, V. S. Allosteric Pathways in Imidazole Glycerol Phosphate Synthase. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, E1428−E1436.
(27) VanWart, A. T.; Eargle, J.; Luthey-Schulten, Z.; Amaro, R. E. Exploring Residue Component Contributions to Dynamical Network Models of Allostery. *J. Chem. Theory Comput.* **2012**, *8*, 2949−2961.
(28) Gasper, P. M.; Fuglestad, B.; Komives, E. A.; Markwick, P. R.; McCammon, J. A. Allosteric Networks in Thrombin Distinguish Procoagulant vs. Anticoagulant Activities. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 21216−21222.
(29) Di Paola, L.; De Ruvo, M.; Paci, P.; Santoni, D.; Giuliani, A. Protein Contact Networks: An Emerging Paradigm in Chemistry. *Chem. Rev.* **2013**, *113*, 1598−1613.
(30) Ribeiro, A. A. S. T.; Ortiz, V. Determination of Signaling Pathways in Proteins through Network Theory: Importance of the Topology. *J. Chem. Theory Comput.* **2014**, *10*, 1762−1769.
(31) Scarabelli, G.; Grant, B. J. Kinesin-5 Allosteric Inhibitors Uncouple the Dynamics of Nucleotide, Microtubule, and Neck-Linker Binding Sites. *Biophys. J.* **2014**, *107*, 2204−2213.
(32) Yao, X. Q.; Malik, R. U.; Griggs, N. W.; Skjaerven, L.; Traynor, J. R.; Sivaramakrishnan, S.; Grant, B. J. Dynamic Coupling and Allosteric Networks in the Alpha Subunit of Heterotrimeric G Proteins. *J. Biol. Chem.* **2016**, *291*, 4742−4753.

https://doi.org/10.1021/acs.jpcb.2c00199
J. Phys. Chem. B 2022, 126, 2612−2620

(33) Botello-Smith, W. M.; Luo, Y. Robust Determination of Protein Allosteric Signaling Pathways. *J. Chem. Theory Comput.* **2019**, *15*, 2116−2126.
(34) Di Paola, L.; Giuliani, A. Protein Contact Network Topology: a Natural Language for Allostery. *Curr. Opin. Struct. Biol.* **2015**, *31*, 43−48.
(35) Cooper, A.; Dryden, D. T. Allostery without Conformational Change. A Plausible Model. *Eur. Biophys. J.* **1984**, *11*, 103−109.
(36) Monod, J.; Wyman, J.; Changeux, J. P. On the Nature of Allosteric Transitions: A Plausible Model. *J. Mol. Biol.* **1965**, *12*, 88−118.
(37) Koshland, D. E., Jr; Nemethy, G.; Filmer, D. Comparison of Experimental Binding Data and Theoretical Models in Proteins Containing Subunits. *Biochemistry* **1966**, *5*, 365−385.
(38) De Ruvo, M.; Giuliani, A.; Paci, P.; Santoni, D.; Di Paola, L. Shedding Light on Protein-Ligand Binding by Graph Theory: the Topological Nature of Allostery. *Biophys. Chem.* **2012**, *165*−*166*, 21−29.
(39) Gadiyaram, V.; Dighe, A.; Ghosh, S.; Vishveshwara, S. Network Re-Wiring During Allostery and Protein-Protein Interactions: A Graph Spectral Approach. *Methods Mol. Biol.* **2021**, *2253*, 89−112.
(40) Gorfe, A. A.; Grant, B. J.; McCammon, J. A. Mapping the Nucleotide and Isoform-Dependent Structural and Dynamical Features of Ras Proteins. *Structure* **2008**, *16*, 885−896.
(41) Sayyed-Ahmad, A.; Prakash, P.; Gorfe, A. A. Distinct Dynamics and Interaction Patterns in H- and K-Ras Oncogenic P-loop Mutants. *Proteins* **2017**, *85*, 1618−1632.
(42) Johnson, Q. R.; Lindsay, R. J.; Nellas, R. B.; Fernandez, E. J.; Shen, T. Mapping allostery through computational glycine scanning and correlation analysis of residue-residue contacts. *Biochemistry* **2015**, *54*, 1534−1541.
(43) Doshi, U.; Holliday, M. J.; Eisenmesser, E. Z.; Hamelberg, D. Dynamical Network of Residue-Residue Contacts Reveals Coupled Allosteric Effects in Recognition, Catalysis, and Mutation. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 4735−4740.
(44) Venkatakrishnan, A. J.; Deupi, X.; Lebon, G.; Heydenreich, F. M.; Flock, T.; Miljus, T.; Balaji, S.; Bouvier, M.; Veprintsev, D. B.; Tate, C. G.; et al. Diverse Activation Pathways in Class A GPCRs Converge near the G-Protein-Coupling Region. *Nature* **2016**, *536*, 484−487.
(45) van den Bedem, H.; Bhabha, G.; Yang, K.; Wright, P. E.; Fraser, J. S. Automated Identification of Functional Dynamic Contact Networks from X-Ray Crystallography. *Nat. Methods* **2013**, *10*, 896−902.
(46) Daily, M. D.; Upadhyaya, T. J.; Gray, J. J. Contact Rearrangements Form Coupled Networks from Local Motions in Allosteric Proteins. *Proteins* **2008**, *71*, 455−466.
(47) Daily, M. D.; Gray, J. J. Allosteric Communication Occurs via Networks of Tertiary and Quaternary Motions in Proteins. *PLoS Comput. Biol.* **2009**, *5*, No. e1000293.
(48) Li, Q.; Luo, R.; Chen, H. F. Dynamical important residue network (DIRN): network inference via conformational change. *Bioinformatics* **2019**, *35*, 4664−4670.
(49) Li, H.; Yao, X. Q.; Grant, B. J. Comparative structural dynamic analysis of GTPases. *PLoS Comput. Biol.* **2018**, *14*, No. e1006364.
(50) Kim, J. I.; Chang, H.; Lee, M.; Na, S. Internal interaction changes within the mutation of SLC26A4 STAS domain. *Chem. Phys. Lett.* **2018**, *710*, 226−233.
(51) Kornev, A. P.; Taylor, S. S. Dynamics-Driven Allostery in Protein Kinases. *Trends Biochem. Sci.* **2015**, *40*, 628−647.

https://doi.org/10.1021/acs.jpcb.2c00199
J. Phys. Chem. B 2022, 126, 2612−2620