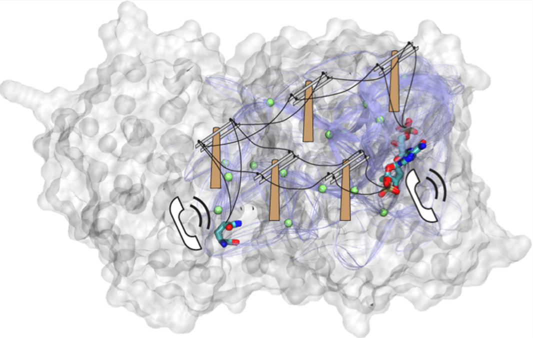
JCTC Journal of Chemical Theory and Computation

pubs.acs.org/JCTC Article

Residue−Residue Contact Changes during Functional Processes Define Allosteric Communication Pathways

Xin-Qiu Yao* and Donald Hamelberg*

Cite This: J. Chem. Theory Comput. 2022, 18, 1173−1187 Read Online



ABSTRACT: Allosteric regulation plays a central role in orchestrating diverse cellular processes. A prerequisite for allostery is a flexible biomolecule within which two distal sites can communicate via concerted or sequential conformational changes. We introduce a computational method to elucidate allosteric communication pathways, comprising critical allosteric residues, in biomolecules by taking advantage of conformational changes during a functional process. Conformational changes are modeled explicitly since they modulate the network of residue−residue interactions, which could propagate allosteric signals between two or more distal sites. The method implements the suboptimal path analysis in the framework of the difference contact network analysis or dCNA. The method identifies key experimentally verified allosteric residues in imidazole glycerol phosphate synthase (IGPS), a well-studied allosteric protein system. By contrast, some of the most important allosteric residues are not captured using methods that do not consider conformational changes, such as those that solely rely on examining the individual bound or unbound state of the protein. Using the dCNA path analysis along with conventional analyses, we gain several new biological insights into IGPS. Interestingly, different binding processes in the thermodynamic cycle generally use a similar group of residues in defining the allosteric communication pathways, with some residues being more specific to a certain binding process. We also observed that the fine-tuning of allosteric coupling depends on the strength of effector binding. Our results are robust against small variations of parameters and details of the network construction. The dCNA path analysis method is general and can be easily applied to diverse allosteric systems.

1. INTRODUCTION

Allosteric regulation is a physiochemical phenomenon prevalent in diverse cellular processes, including enzyme catalysis, signal transduction, cell division, and intracellular trafficking.[1−3] At the molecular level, allosteric regulation manifests a coupling of activities at distal sites within a biomolecule, such as cooperative substrate binding or the modulation of catalysis upon effector binding. Aberrant allosteric regulations underlie many human diseases including cancer.[4] Hence, elucidating the mechanism of allosteric regulation is essential for understanding physiological and pathophysiological processes, and the knowledge obtained may be leveraged in drug discovery.[4,5] The critical importance of allosteric effects has also gained increasing attention in modern rational protein design.[6,7]

Biomolecular internal motions or conformational dynamics are the key to understanding allosteric regulation. Proteins, for example, are intrinsically flexible (malleable), populating a rugged (free) energy landscape around the native folded conformations, consisting of many nearly isoenergetic basins (conformational substates).[8] Protein activity is determined by the ensemble of conformations generated under certain thermodynamic conditions. The protein's malleability and degrees of freedom that are subjected to conformational changes are primarily responsible for the fine-tuning of activity (i.e., allosteric regulation): effector binding and other physical or chemical perturbations alter the energy landscape and shift the conformational ensemble due to the malleable dynamical changes, leading to modified functions of biomolecules. This general description gives a qualitative explanation of allostery but does not provide detailed description and the structural basis of allosteric communication pathways within biomolecules. An important question here is what are the key residues that directly mediate the allosteric coupling (i.e., the allosteric pathway)? Apparently, comparing different conformational ensembles during a functional process (allosteric effector binding, e.g.) is essential to answering this question.

An increasingly popular approach for understanding allosteric regulation applies ideas from network science to biomolecular structures.[9−15] For example, using the network formalism, one can perform the path analysis to map allosteric communication

Received: July 5, 2021
Published: January 20, 2022

https://doi.org/10.1021/acs.jctc.1c00669

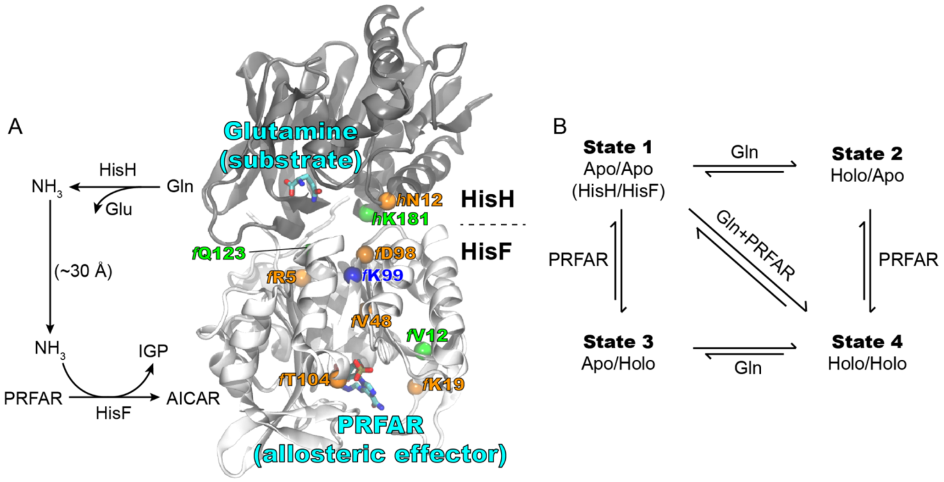


Figure 1. IGPS catalyzes coupled reactions separated by a long (~30 Å) distance. (A) IGPS is represented as cartoon color-coded by subunits. Substrate (Gln) and allosteric effector (PRFAR) are displayed as sticks color-coded by atom types. Key residues identified from the list of experiments in Table S1 are shown as spheres color-coded by the effects of mutations at these sites: orange, at least one mutation causing a strong decoupling between the glutaminase activity and the effector binding; green, moderate decoupling; blue, at least one mutation causing enhanced coupling. The reaction scheme is shown on the left. (B) Thermodynamic cycle to bind Gln and PRFAR in IGPS.

pathways.[9−11,14−16] The hypothesis is that the allosteric signal propagates from one site to another via the shortest (or optimal) and near-shortest alternative (suboptimal) paths. Here, a path is defined by a set of network edges connecting the two end sites or nodes (usually represented by residues) and the path length is the sum of edge weights (defined by geometric, dynamic, or energetic properties depending on the type of the network). Each microscopic path represents a potential route for the allosteric signal to transmit, and the (suboptimal) path analysis, summarizing the ensemble of paths, helps identify critical residues that are on many paths (assumed to be crucial for the allosteric communication).

A major limitation of many conventional network methods, including the representative dynamical network (DN)[11] and protein contact network (PCN)[13] methods, is that each time only a single (experimental) structure, or a conformational ensemble average generated, for example, from a single molecular dynamics (MD) simulation, is used to define the topology of the network. Hence, in these methods, the difference between conformational ensembles (representing the end states of a functional process) is not considered explicitly, which prevents these methods from capturing the full picture of allosteric regulation.

In some of the studies using conventional network methods, networks representing distinct functional states (e.g., ligand-free and -bound) are compared. However, the results of such comparisons either are hard to interpret or can only be understood in terms of changing of the allosteric mechanism. For example, in some studies, shortest path-based network parameters are calculated and compared between different networks.[10,16] Considering the hypothesis underlying the path analysis, such a network comparison means looking for a change of the allosteric pathway or mechanism. On the other hand, if the difference of path analysis results is used to explain allostery, it will be rather elusive to know the meaning of individual microscopic paths and signals they transmit.

In this work, we introduce a new method to delineate allosteric pathways at residue-level resolution that combines the suboptimal path analysis[17−20] with a novel network construction method based on residue−residue contact changes during functional processes. Different from conventional single-ensemble network methods, our method compares two or more conformational ensembles (as a special implementation of our general comparative perturbed-ensemble analysis[21]) and identifies malleable dynamic contacts that change (i.e., modulated) during functional processes. These malleable contacts are hypothesized to mediate the allosteric communication between distal sites. Our method is also fundamentally different from the comparison of network or topological parameters because our method uses "changes" as the input rather than the output of the network analysis; here, the path analysis is regarded as a step in postprocessing the detailed, physically meaningful dynamical input that is derived from the functional process. The same ensemble-comparison principle, but with different network construction procedures, has been applied previously to develop our difference contact network analysis (dCNA) method.[22] The dCNA method has been used to perform the community analysis, which identifies intrinsic residue clusters and outputs a coarse-grained representation of conformational changes that provides an overall picture of the allosteric communication. The new method here, called dCNA path analysis, is designed to identify the residue-wise allosteric communication pathway. The method can be applied to ensembles obtained from both MD simulations and experimental structural data.

We carry out a thorough test of our method on imidazole glycerol phosphate synthase (IGPS), a well-studied allosteric system and an important metabolic enzyme, that catalyzes critical steps in histidine and purine biosynthesis in microorganisms.[23,24] The protein system has two active sites separated by ~30 Å (Figure 1A): the glutamine (Gln) binding site located in the HisH subunit contains the catalytic triad *h*C84, *h*H178, and *h*E180 (here, we follow the convention to name each residue prefixed with the subunit name: *h* for HisH and *f* for HisF) and catalyzes the hydrolysis of glutamine into glutamate and ammonia; the ribonucleotide (PRFAR) binding site in the HisF subunit accepts the ammonia transferred from HisH through a central $(\beta/\alpha)_8$ barrel and catalyzes the synthetic process involving PRFAR. The glutaminase activity of IGPS is enhanced by ~4900 times upon PRFAR binding, suggesting an allosteric coupling between the two distal sites.[25] However, the detailed mechanism of how the two sites communicate remotely remains to be elucidated. To investigate this prototypical allosteric system, we have performed four 10 μs atomistic MD simulations, each for a unique ligand-bound state of IGPS in the thermodynamic cycle (Figure 1B).

The remaining contents of this paper are organized as follows. In Section 2, the algorithm of the new dCNA path analysis, along with other computational technical details, is introduced. In Section 3, the main results are discussed. First, the new method is compared with the conventional DN and PCN methods using an experimental benchmark (Section 3.1). Then, new biological insights obtained by using the dCNA path analysis and conventional analyses are discussed (Section 3.2). The convergence and robustness of the results are also carefully examined (Section 3.3). Concluding remarks are made in Section 4.

2. THEORY AND COMPUTATIONAL METHODS

2.1. Difference Contact Network Suboptimal Path Analysis. We propose a suboptimal path analysis in the framework of dCNA described previously.[22] In dCNA, a key statistical feature is the difference in the probability of the occurrence of residue−residue contacts between two ensembles. Here, a contact is defined by two residues with minimal nonhydrogen atomic distance less than or equal to a distance cutoff, d_c. An optimal cutoff has been found to be d_c = 4.5 Å.[26] Different from our previous dCNA-based community analysis, which builds a consensus network across ensembles first and summarizes contact changes between communities detected based on the consensus network,[22] the new method directly utilizes the residue-wise differential contact probabilities from one ensemble to the other to build the network. The idea to compare contact maps for understanding protein functions has been adopted in previous studies.[27−30] First, for each ensemble, an examination of contact strength is performed for each pair of residues by

$$\Delta g = \begin{cases} -k_B T \ln\left(\dfrac{p}{1-p}\right) & p_c \le p \le 1 - p_c \\ -\varepsilon & p > 1 - p_c \\ \varepsilon & p < p_c \end{cases} \tag{1}$$

where p is the probability of forming a contact between the corresponding residue pair, k_B is the Boltzmann constant, and T is the temperature (300 K). A threshold of probability, p_c, is applied to separate "dynamic contacts" ($p_c \le p \le 1 - p_c$), where the contact strength is proportional to the log-odd of p, from other contacts, where the contact strength is determined by a constant, ε. The introduction of p_c and ε is to resolve the numeric instability of the log-odd function when p goes to 0 or 1. In this work, we set p_c = 0.2 and ε = 2.0 kcal/mol. Tests show that a choice of 0.1−0.2 for p_c and 1.5−2.5 (kcal/mol) for ε produces reasonable and robust results (see Section 3.3). For simplicity, we ignore the parameter n_c that defines nonlocal contacts (i.e., between residues i and $i + n$, $n \ge n_c$). This parameter has been shown to have little influence on the difference contact analysis.[26] Note that by comparing ensembles (see below), covalently bonded residue neighbors are automatically excluded from the analysis because the contact probability (and hence the strength) for such neighbors is always identical (p = 1) among all conformational ensembles.

A protein structural network is then constructed, where each node is represented by a residue, and the edge between nodes is determined by the difference between contact strengths from two distinct conformational ensembles, $\Delta\Delta g = \Delta g_2 - \Delta g_1$ (assuming a process from ensemble 1 to 2). Specifically, an edge is added between two nodes if (1) the corresponding $|\Delta\Delta g| > 0$ or (2) the two residues form a dynamic contact in any of the ensembles under consideration. Each edge of the network is weighted by

$$w = \Delta\Delta g - \min_{\text{all edges}} (\Delta\Delta g) + a \tag{2}$$

where a is an arbitrary constant to make the weights of all edges positive. According to eq 2, a "distance" network is constructed where a larger edge weight represents a longer "distance" (less important) and vice versa. Note that the value of a in eq 2 is unimportant as long as it is positive because in the suboptimal path analysis, only relative distances between edges matter. A noticeable feature of eq 2 is that the direction of contact changes is considered (e.g., the process from ensemble A to B is different from B to A). In particular, we assume in eq 2 that contact strengthening (negative $\Delta\Delta g$ or increase of contact probability) is more important (smaller edge weight or "distance") than contact weakening (positive $\Delta\Delta g$ or decrease of contact probability). A test on the influence of the direction of changes is given in Section 3.3.

Once a network is built, the suboptimal path analysis previously described[19] can be implemented. In brief, a pair of "source" and "sink" residues (nodes) are chosen first; these residues can be, for example, from the allosteric and active sites, respectively. Then, the algorithm described in ref 31 is applied to find k loopless shortest paths between the source and sink. Normalized node degeneracy is calculated by summing up the number of paths going through each residue divided by the total number of paths, k. Higher node degeneracy is assumed to be more critical for the allosteric coupling between the source and sink.

2.2. MD Simulations. MD simulations were performed using Amber20.[32] The initial structure of IGPS was taken from the crystallographic structure (Protein Data Bank[33] or PDB entry ID: 1GPW).[24] The coordinates for the ligands were modeled based on PDB 3ZR4[34] and 1OX5[35] for the substrate (glutamine) and the effector (PRFAR), respectively. All crystal water molecules in 1GPW were kept except for those overlapping with modeled ligands. The Amber ff14SB force field was employed for the protein.[36] For the substrate, the zwitterionic form of glutamine was adopted with parameters from ref 37 obtained from http://amber.manchester.ac.uk/. For PRFAR, the parameters were generated by Antechamber of Amber20 using GAFF2[38] and AM1-BCC.[39,40] The protonation state of ionizable groups at pH = 7.0 was determined based on the pKa value calculated by PROPKA[41,42] called through PDB2PQR.[43,44] In particular, Asp130 and Asp176 were set to neutral. The tautomer state of the imidazole side chain of neutral histidine residues was determined by PDB2PQR along with the visual inspection of the local structural environment of the residue. Specifically, all histidine residues were set to HID (Amber naming convention indicating the protonation at the δ-nitrogen), except for His84, which was set to HIP (the protonated form). Totally, four systems were prepared with both Gln and PRFAR absent (apo/apo or state 1 in Figure 1B),

https://doi.org/10.1021/acs.jctc.1c00669
J. Chem. Theory Comput. 2022, 18, 1173−1187

with Gln present but PRFAR absent (holo/apo or state 2), with Gln absent but PRFAR present (apo/holo or state 3), and with both Gln and PRFAR present (holo/holo or state 4). All missing atoms were added by the LEaP program of Amber. Each system was solvated in an octahedron box filled with pre-equilibrated TIP3P water molecules[45] extended by 10 Å from the surface of the solute to each box face. Counter ions (Na^+ or Cl^-)[46,47] were added to neutralize each system.

During the energy minimization, 2000 steps of the steepest descent followed by 3000 steps of the conjugate gradient were performed, with harmonic positional restraints (force constant of 600 $kcal \cdot mol^{-1} \cdot Å^{-2}$) applied to all solute atoms. Then, each system was heated from 100 to 300 K by eight rounds of simulations totaling 9.3 ns under NVT periodic conditions (Langevin thermostat with collision frequency $\gamma = 1.0\ ps^{-1}$), where position restraints were applied to the solute with the force constant of restraint gradually reduced from 500 to 400, 300, 200, 100, 50, 25, and 5 $kcal \cdot mol^{-1} \cdot Å^{-2}$ from one round to the next. Equilibration was run by 10 ns with a 2 fs time step and no restraint under NPT (300 K, 1 bar; Monte Carlo barostat with coupling constant $\tau_p = 1.0$ ps) periodic conditions. A 9 Å cutoff was used for short-range nonbonded interactions. The particle-mesh Ewald summation method[48] was employed to treat long-range electrostatic interactions. All bonds involving hydrogen atoms were constrained with SHAKE (nonwater)[49] or SETTLE (water).[50] Production simulations were performed under the same conditions as equilibration by 10.2 μs for each system, totaling 40.8 μs, with each trajectory saving every 1 ps. The last 10 μs of each trajectory was used for the analyses.

2.3. Application of the dCNA Path Analysis in IGPS. Six distinct difference contact networks were constructed for the path analysis in IGPS based on the different ligand binding processes in the thermodynamic cycle (Figure 1B): the 1 → 2 network represents Gln binding in the absence of PRFAR, 1 → 3 is PRFAR binding in the absence of Gln, 1 → 4 is the binding of both Gln and PRFAR, 2 → 4 is PRFAR binding in the presence of Gln, 3 → 4 is Gln binding in the presence of PRFAR, and the coupling network was constructed based on the difference of two distinct PRFAR binding processes with and without Gln, respectively (i.e., between 2 → 4 and 1 → 3). The coupling network represents the coupled substrate and effector binding. In the coupling network, the edges were determined by a consideration of all four states and weighted by $\Delta\Delta g = \Delta\Delta g_{2\to4} - \Delta\Delta g_{1\to3}$. Note that equivalent networks can be obtained from $\Delta\Delta g_{3\to4} - \Delta\Delta g_{1\to2}$ or $\Delta\Delta g_{1\to4} - \Delta\Delta g_{1\to3} - \Delta\Delta g_{1\to2}$. For 1 → 2 and 1 → 3 networks, a pseudonode was introduced to represent PRFAR and Gln, respectively, and equally weighted edges were added between the pseudonode and binding-site residues. Binding-site residues were identified by contacts between the ligand and protein residues that occurred over 90% of the simulation trajectory where the ligand was present (i.e., the state 2 simulation for identifying binding-site residues of Gln and state 3 for PRFAR). By this setup, all networks had the same size and so were easier to compare. A representative (2 → 4) network is shown in Figure 2.

Nodes represented by PRFAR and Gln were chosen as the source/sink for the suboptimal path analysis. For each network, $k = 5{,}000$ suboptimal paths were searched. The influence of k on the result is examined in Section 3.3.

2.4. DN Analysis and PCN Analysis. DN analysis was performed as previously described[11] for each of the four simulations. In brief, each node of the network was represented by a residue, and edges between nodes were determined by residue−residue contacts that occurred over 75% of the trajectory under analyses. For covalently linked residue pairs, no edge was created. The edges were weighted by $-\ln(|C_{ij}|)$, where C_{ij} is the dynamic cross-correlation between Cα atoms of residues i and j. The edges were removed for uncorrelated residue pairs ($C_{ij} = 0$). Prior to calculating C_{ij}, simulation frames were superimposed based on all Cα atoms of IGPS excluding flexible N- and C-terminal loops (*f*M1−*f*K4, *f*H244−*f*L253, *h*M1−*h*R2, and *h*C196−*h*R201). For PRFAR, the central C7 atom was used to calculate correlations with protein residues. For systems where one or two ligands were absent, a pseudonode was introduced for each absent ligand in a similar manner as described in Section 2.3. PCNs were constructed the same as DN, except that all network edges were equally weighted. The same path analysis procedure as described in Section 2.3 was performed in each of the DNs and PCNs.

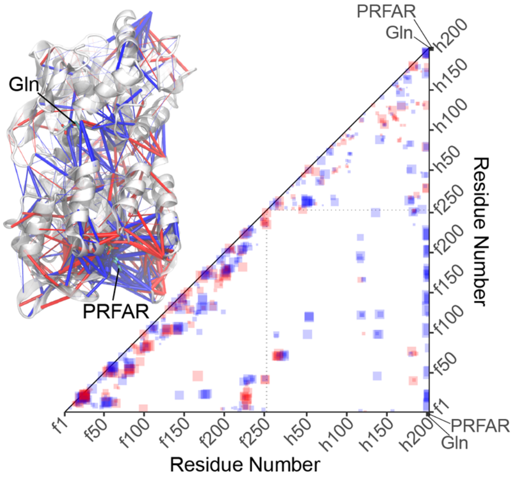


Figure 2. Example of protein structural network for the dCNA path analysis of IGPS. Contact changes from state 2 to state 4 (2 → 4) (see Figure 1B) are used to define network edges, which are mapped as colored cylinders on an IGPS structure and as transparent squares on a 2D plane. Radii of cylinders (or sizes of squares) are scaled by the magnitude of contact changes ($|\Delta\Delta g|$; $\Delta\Delta g$ is in the range of −4.0 to 2.6 kcal/mol) and colors by the sign of changes: blue, contact strengthening (negative $\Delta\Delta g$ or increase of contact probability); red, contact weakening (positive $\Delta\Delta g$ or decrease of contact probability).

2.5. Statistical Errors. Statistical errors of path analysis results were estimated through bootstrapping. Each production trajectory was grouped into 50 (200 ns) chunks. Then, these chunks were randomly picked up (inclusively) to construct a new sample with the same length of data as the original trajectory (10 μs). This random procedure was repeated 10 times, and for each time, the newly generated sample was used to calculate residue−residue contact probabilities and do the path analysis. Node degeneracies for each residue were collected from the 10 samples, and the standard deviation of these degeneracies was adopted as the statistical error of the prediction for the residue (Tables S2−S4).

2.6. Software Used in the Analyses. Trajectories were processed with the CPPTRAJ program of AMBER.[51] Network analyses were performed with Bio3D[19,52,53] and in-house scripts. Molecular graphics were rendered by VMD[54] and PyMol (Schrödinger, LLC). All other figures were generated by ggplot2.[55] All figures were assembled with Illustrator 2021 (Adobe, Inc.).

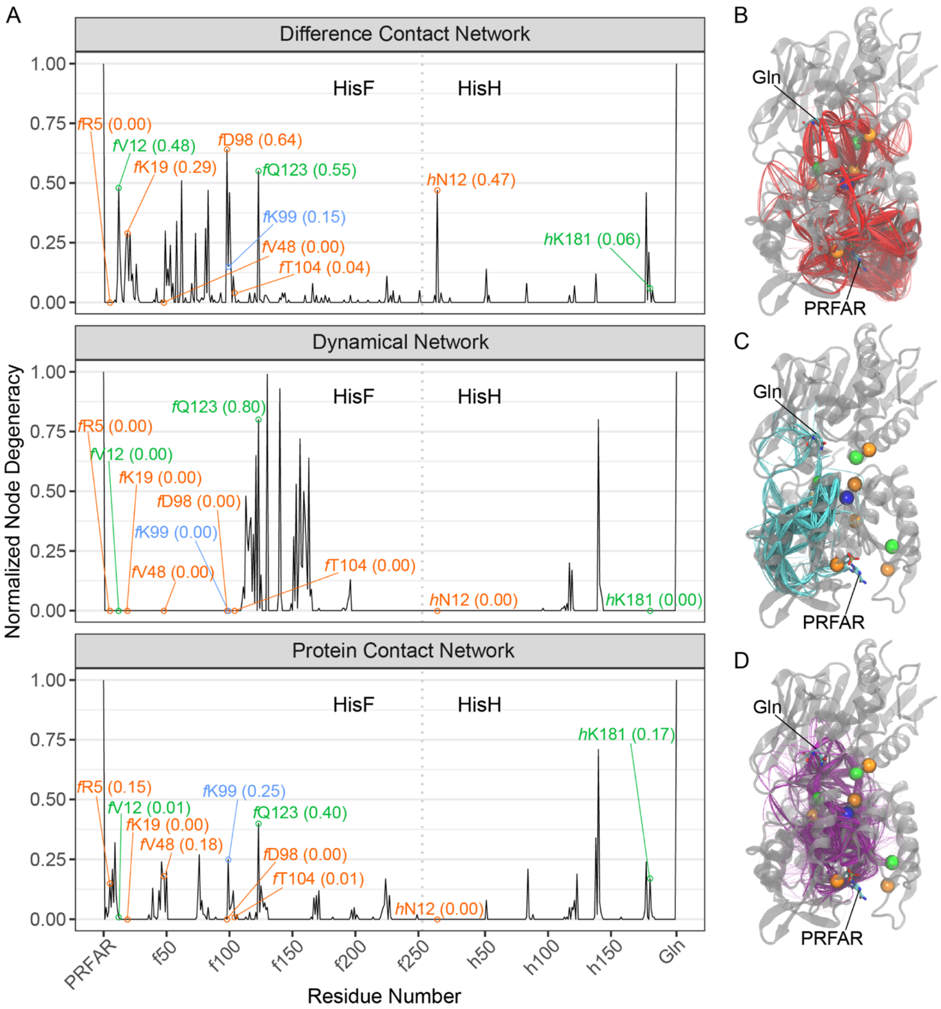


Figure 3. Comparisons of performance between dCNA, DN, and PCN. For dCNA, the 2 → 4 process (see Figure 1B) is used. For the DN and PCN, state 4 is considered. (A) Normalized node degeneracy over 5000 (sub)optimal paths for each method. Experimentally verified key residues are labeled and color-coded by the same scheme as in Figure 1A. (B−D) Optimal and suboptimal paths mapped onto an IGPS structure. Paths are represented as colored lines with the radii of paths scaled by path lengths. Experimental key residues are shown as spheres color-coded the same as in (A). Substrate (Gln) and effector (PRFAR), which are used as the source/sink in the path analyses, are shown as sticks color-coded by atom types.

3. RESULTS AND DISCUSSION

3.1. dCNA Path Analysis Reveals Critical Allosteric Residues Omitted by Conventional Network Analysis Methods. The dCNA path analysis identifies the key experimentally verified allosteric residues in IGPS. Here, the 2 → 4 network (i.e., PRFAR binding in the presence of Gln; Figure 1B) is examined. Changes induced during the 2 → 4 process are expected to be the most relevant to the allosteric activation of the glutaminase activity by PRFAR.[56] The performance is evaluated on a benchmark developed by McCullagh and co-workers,[57] which contains mutations with varying changes (reduction or enhancement) in the activity of the PRFAR-dependent hydrolysis of glutamine, identified by extensive experimental mutagenesis experiments collected from the literature (Table S1 and Figure 1A).[25,58−60] Some of these residues are suggested to be key to mediate the allosteric coupling in IGPS. The path analysis identifies multiple regions exhibiting a high normalized node degeneracy, indicating the importance of these regions for the allosteric coupling between PRFAR and Gln (Figure 3A,B and Table S2). Importantly, three residues including *f*K19, *f*D98, and *h*N12, which have been identified by experiments to be crucial for the allosteric coupling, all show significant (>0.1) node degeneracies. Especially, *f*D98 shows the highest (excluding source and sink) degeneracy (0.64), which has also been found to have the most drastic impact on the allosteric coupling upon mutation by experiments: an alanine substitution for *f*D98 almost completely deletes the allosteric effect of PRFAR binding to activate the glutaminase activity (fold change of the activity of the activated state drops from 4900 of the wild type to 1.8; Table S1). Similarly, *h*N12 has a high degeneracy (0.47), and an alanine substitution for *h*N12 causes a dramatic drop of fold change to 100 (Table S1). For *f*K19, a smaller but notable degeneracy value (0.29) is found, while the mutations of this residue display similar levels of fold-change drop to *h*N12A (Table S1).

The dCNA path analysis also detects *f*V12 (0.48), *f*K99 (0.15), and *f*Q123 (0.55) to be of significant node degeneracies (Figure 3A). The mutations of all three residues have been shown to either moderately reduce or enhance the allosteric coupling (Table S1). Note that the path analysis is to identify critical allosteric residues, not to precisely predict the effect of specific mutations that may span a broad range of severities because of the diverse properties of amino acids. For example, even though the mutation *f*Q123A is found to only reduce the

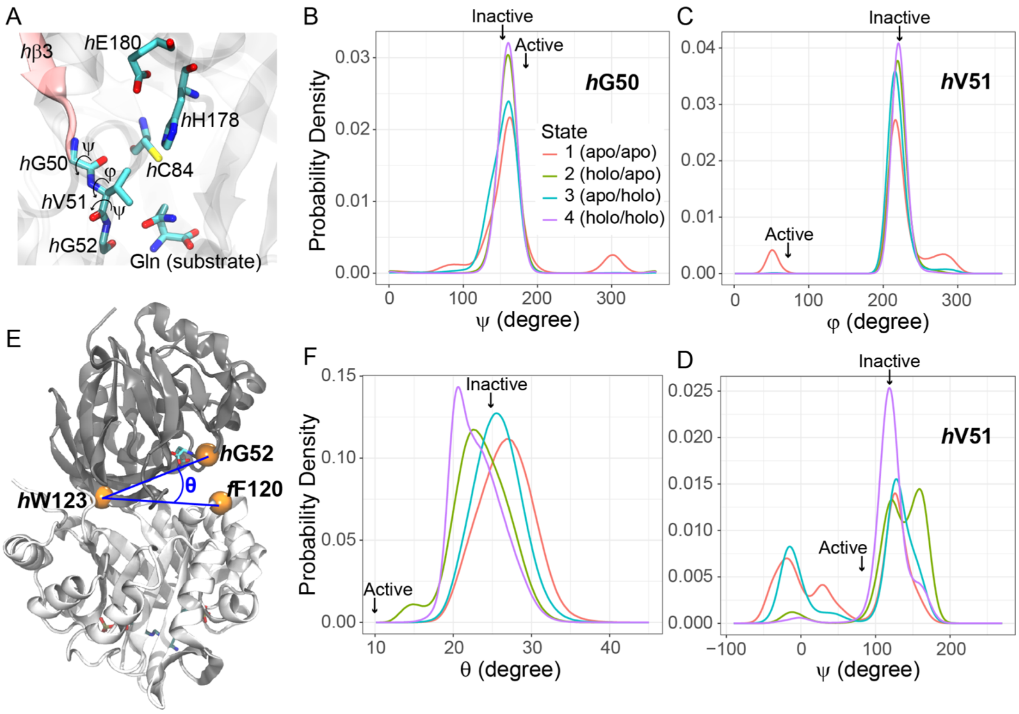


Figure 4. Representative conformational changes during IGPS activation captured by MD simulations. (A) Enlarged view of the HisH active site. The catalytic triad (*h*C84, *h*H178, and *h*E180), the oxyanion strand (*hβ*3; pink) residues (*h*G50, *h*V51, and *h*G52), and the substrate are shown as sticks color-coded by atom types. (B−D) Probability distributions of the backbone dihedral angle ψ of *h*G50 (B), φ of *h*V51 (C), and ψ of *h*V51 (D) for all four simulations. (E) The Cα atoms of *f*F120, *h*W123, and *h*G52 are used to define the angle (θ) describing the inter-subunit "breathing" motion. (F) Probability distributions of θ. In (B−D,F), the "inactive" and "active" labels indicate values calculated based on the chains A/B and E/F, respectively, from the crystallographic structure (PDB: 7AC8).

fold change by about a half of the wild type, other amino acid substitutions may cause a severe impact as suggested by the high degeneracy (0.55) at this residue site.

Four key residues, including *f*R5, *f*V48, *f*T104, and *h*K181, which are found to exhibit either severe or moderate mutational effects by experiments (Table S1), have low or zero degeneracies in dCNA for the 2 → 4 process (Figure 3A). The residues *f*T104 and *h*K181 show a low (<0.1) but nonzero degeneracy, indicating that both are still identified to contribute to the allosteric communication. Residues *f*R5 and *f*V48, however, apparently deviate from the predicted allosteric pathway because of their zero degeneracies for this binding process. Detailed inspections on *f*R5 and *f*V48 are given in Section 3.2.2.

The DN and PCN methods predict node-degeneracy profiles largely different from those predicted by dCNA. The DN-based suboptimal path analysis using the state 4 (Figure 1B) simulation as the input shows an allosteric pathway highly concentrated near two α-helices in HisF, *fα*4 and *fα*5 (Figure 3A,C and Table S3). Surprisingly, almost all experimentally verified key residues are off the predicted pathway, except for *f*Q123 (degeneracy 0.80). Interestingly, in PCN, which is constructed the same as the DN except that all edges are weighted equally, the paths are generally more spread out and some experimentally verified key residues are shown to have significant degeneracies (Figure 3A,D and Table S4). Most notably, *f*R5 and *f*V48, which are not identified by dCNA for the 2 → 4 process, show significant degeneracies (0.15 and 0.18, respectively) in the PCN. Also, *f*K99, *f*Q123, and *h*K181 have significant degeneracies (0.17−0.40) in the PCN. However, several key residues are missed by the PCN, including the most important *f*D98.

Predictions by the DN are sensitive, whereas those by the PCN are invariant, with the input simulation. Both methods are also tested using the state 1, state 2, and state 3 (Figure 1B) simulations as the input. Pathways in the DN vary dramatically (Figures S1 and S4 and Table S3), indicating that the DN method is sensitive to the choice of the ensemble to build the network. In using the state 2 simulation, the DN identifies *f*R5 (degeneracy 0.33) and *f*V48 (0.21), but all other experimentally verified key residues are of low or zero degeneracies. In contrast, largely invariant degeneracy profiles are observed in the PCN, indicating that the PCN is less sensitive to the choice of the ensemble (Figures S2 and S4 and Table S4). Neither the PCN nor the DN detects the key residues, *f*K19, *f*D98, *f*N104, and *h*N12, no matter the ensemble (state) used.

In summary, our results suggest that dCNA has a distinct behavior in identifying allosteric communication pathways compared to the DN and PCN. This is not very surprising considering the distinct principles behind the two types of methods: dCNA takes the difference or changes during a process (i.e., the malleable dynamical residue−residue contacts) to define the network whereas the DN and PCN use stable (mostly invariant during a process) structural features. Both *f*D98 and *h*N12 show significant degeneracies in dCNA but nearly zero degeneracies in the DN and PCN, no matter which ensemble is used for the analysis. We speculate that *f*D98 and *h*N12 directly participate in relaying the allosteric signal by undergoing significant contact changes upon PRFAR binding; their importance is hidden without using dCNA or inspecting the conformational changes directly. Further discussion about the roles of *f*D98 and *h*N12 is given in Section 3.2.2.

https://doi.org/10.1021/acs.jctc.1c00669

3.2. New Biological Insights into the Allosteric Mechanism in IGPS.

In this section, we discuss our main findings about the allosteric regulation in IGPS. A popular model to explain the enhanced glutaminase activity upon PRFAR binding in IGPS is based on the "oxyanion hole" theory of hydrolytic reactions.[23,35] In this model, two amide nitrogen atoms, one from the residue immediately next to the nucleophile (*h*C84) and the other from a β strand (*h*V51 of *h*β3) adjacent to the active site containing the conserved sequence *h*P49−*h*G50−*h*V51−*h*G52 (also called "oxyanion strand"), are assumed to stabilize the negative charge generated during the catalysis (Figure 4A). In the basal, PRFAR-free state, the nitrogen of *h*V51 points away from the active site, resulting in a low catalytic activity. PRFAR binding causes conformational changes involving a flip of the peptide around *h*V51, making the nitrogen of *h*V51 point toward the active site and forming the oxyanion hole that is required by the catalysis.

Despite a tremendous number of experimental and computational studies, the allosteric mechanism in IGPS remains elusive.[12,16,25,34,56−69] This is partly because somewhat controversial conclusions exist in the field. For example, a recent study using a combined approach of nuclear magnetic resonance (NMR) and X-ray crystallography has identified an "active" conformational state of the enzyme upon binding both the substrate and the allosteric effector.[56] This study suggests a significant conformational change occurring during IGPS activation, which contradicts the previous studies showing that effector binding enhances protein dynamics without changing the mean conformation.[59,61] The recent study used the *h*C84S IGPS mutant.[56] Hence, it remains unclear if the observed conformational change was due to the loss-of-function mutation or was an intrinsic property of the wild-type IGPS. As a complement to experiments, MD simulation provides structural information at a high spatiotemporal resolution and is thus an important tool to help gain new insights into the biological system. New computational simulation-based studies are still necessary despite the many existing similar studies on IGPS[12,16,57,58,60,63,65−69] because (1) simulations in the previous studies were usually too short ($\leq 10^{-7}$ s) compared to the biologically meaningful timescale ($>10^{-6}$ s), except for one study that performed ∼2.5 μs simulations but only focused on IGPS/allosteric inhibitor interactions;[67] (2) the many previous studies using conventional network analysis methods may not fully or properly interpret the MD data, and new methods, such as those developed in this work, can help make new discoveries even applied to existing simulations.

In this work, we attempt to address the following specific biological questions:

1. Do the conformational changes described previously[56] occur during the activation of the wild-type IGPS?
2. What are the roles of the experimentally verified key residues (Table S1) in the allosteric communication in IGPS?
3. Does IGPS use the same or different allosteric pathways for distinct ligand binding processes in the thermodynamic cycle (Figure 1B)?
4. Can the allosteric coupling be fine-tuned by the binding pose or strength of the effector?

3.2.1. MD Captures the Representative Conformational Changes during IGPS Activation but Does Not Completely Reproduce the Active Conformation Observed in Experiments.

The 10 μs MD simulations capture the backbone flipping of *h*V51 required for the formation of the oxyanion hole, but the phenomenon is only observed in the ligand-free (state 1 or apo/apo; Figure 1B) simulation. To detect potential conformational changes, we calculate probability distributions of three consecutive backbone dihedral angles, ψ of *h*G50 and φ and ψ of *h*V51, for all four simulations (Figure 4A−D). We use the recent crystallographic structure (PDB: 7AC8) to define reference conformations for the "inactive" and "active" states, as described before.[56] The φ value of *h*V51 is found to be the most relevant, with the reference active- and inactive-state values exhibiting a large, ∼150° separation; by contrast, the references of the other two angles exhibit a separation of ∼35°. The backbone dihedral angles show the largest overall flexibility in the state 1 simulation. Interestingly, in the state 1 simulation, the angle φ-*h*V51 samples, infrequently but significantly, values very close to the reference active-state value (∼50° vs ∼70°; Figure 4C), indicating the backbone flipping of *h*V51 during the simulation. However, binding of either Gln or PRFAR, or both, stabilizes the "inactive" conformation. Similar backbone stabilization by ligands is observed in ψ-*h*G50 (Figure 4B). ψ-*h*V51 is flexible in both state 1 and state 3 simulations and samples values (−50° to +50°) not observed in the crystallographic structure; Gln binding suppresses such angular sampling (Figure 4D).

Another representative conformational change during IGPS activation is the hinge-like bending or "breathing" motion at the inter-subunit interface. The functional relevance of this motion has been discussed before.[56,60,64,67] Here, we measure the motion by calculating probability distributions of the angle, θ, defined by the Cα atoms of *h*G52, *h*W123, and *f*F120 over the simulations (Figure 4E,F). The reference values are 25 and 10°, respectively, for the inactive and active states, suggesting a subunit−subunit closing upon activation.[56] It shows that even though none of the simulations captures the closed conformation suggested by the crystallographic structure, a clear trend of subunit closing upon ligand binding is observed (Figure 4F). Overall, θ is the largest in the state 1 (apo/apo) simulation (∼27° on average) and the smallest in the state 4 (holo/holo) simulation (∼23°); θ in the state 2 (holo/apo) simulation is smaller than that in the state 3 (apo/holo) simulation (∼24° vs ∼26°). Interestingly, even though the average θ in state 2 is slightly larger than that in state 4, θ samples a small value (∼15°, which is close to the active-state reference, 10°) more frequently in the state 2 simulation than in the state 4 simulation (Figure 4F).

In summary, 10 μs MD simulations capture the representative conformational changes during IGPS activation, including the backbone flipping of the oxyanion strand and subunit−subunit closing, but a full reproduction of experimental observations has not been established yet. The backbone flipping of *h*V51 only observed in the ligand-free simulation suggests that the flipped conformation (forming the oxyanion hole) is from an intrinsic functional substrate of IGPS, but the enzyme may need to across a higher free energy barrier to access that conformation in the presence of the substrate and/or the effector. Our results are consistent with the model where the inter-subunit interface is closed upon binding of both the substrate and the effector, although the magnitude of closing is much smaller in simulations. The discrepancy between simulations and experiments may be explained by the timescale problem (IGPS activation happens in milliseconds,[56] whereas the simulations are 2 orders of magnitude shorter), but we cannot exclude the possibility that the largely closed and backbone-flipped "active"

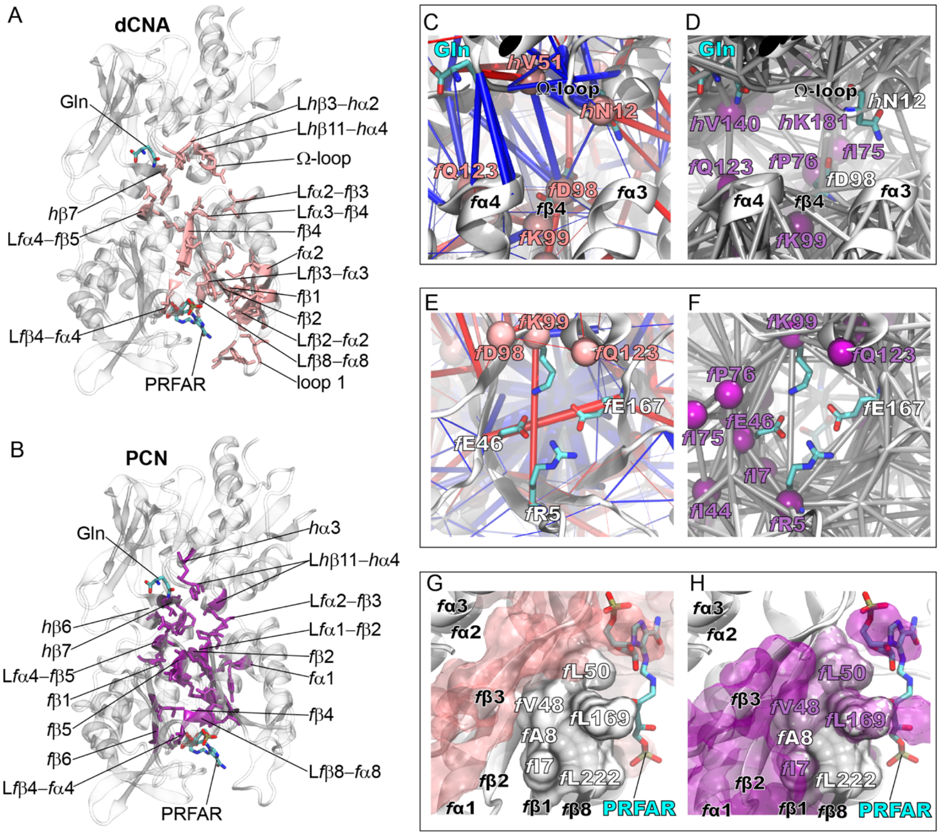


Figure 5. Distinct allosteric pathways identified by dCNA and the PCN. The 2 → 4 process (see Figure 1B) is used in dCNA and state 4 is used in the PCN. (A,B) Residues with a node degeneracy above 0.1 are shown as pink (dCNA) and purple (PCN) cartoon and sticks to exhibit the allosteric pathway, with the secondary structure elements where the residues are located labeled (loops begin with the letter "L" except for loop 1 and Ω-loop). Substrate (Gln) and effector (PRFAR) are shown as sticks color-coded by atom types. (C,D) Enlarged views of the pathways and networks around *f*D98 and *h*N12 (shown as sticks color-coded by atom types) for dCNA (C) and PCN (D), respectively. In (C), the network edges are shown as blue and red cylinders in the same way as in Figure 2. In (D), the network edges are represented by white cylinders. Predicted key residues are shown as spheres color-coded the same as in (A). (E,F) Enlarged views of the subunit interface centered at the central barrel. The four "gating" ionic residues are shown as sticks color-coded by atom types. (G,H) Enlarged views around *f*V48. Hydrophobic residues surrounding *f*V48 are shown as a white surface. Predicted key residues are shown as a transparent surface color-coded the same as in (A). Half of the central barrel (*fβ*4−*fβ*7) is removed for clarity.

conformation is due to the *h*C84S mutation used in the experimental study.[56] The above basic geometric analyses, along with the significant contact changes observed during network constructions in dCNA (e.g., Figure 2), suggest that IGPS activation is not simply an alteration of flexibility or entropy. Hence, our results support the conformational change hypothesis of IGPS activation.

3.2.2. Key Residues Playing Dynamic, Structural, or Dual Roles in the Allosteric Communication in IGPS. In this subsection, we describe in-depth analyses of the two distinct allosteric pathways revealed by dCNA (2 → 4 network)- and PCN (state 4)-based suboptimal path analyses, respectively (Figure 5). Within HisF, the dCNA pathway mainly (defined by node degeneracy $\geq$ 0.1) consists of residues located in loop regions including L*fβ*1−*fα*1 (loop1), L*fβ*2−*fα*2, L*fα*2−*fβ*3, L*fβ*3−*fα*3, L*fα*3−*fβ*4, L*fβ*4−*fα*4, L*fα*4−*fβ*5, and L*fβ*8−*fα*8 (Figure 5A and Table S2). Four major regular secondary structures, *fβ*1, *fβ*2, *fα*2, and *fβ*4, also contribute to the pathway. The side chains of the pathway residues mainly fill the regions between the *α* and *β* elements. In the subunit interface, the pathway residues are more scattered and span a region close to where the breathing motion occurs.[56,60,64,67] In contrast, the PCN pathway goes through more major regular secondary structures, including *fβ*1, *fα*1, *fβ*2, and *fβ*4−6, besides loops L*fα*1−*fβ*2, L*fα*2−*fβ*3, L*fβ*4−*fα*4, L*fα*4−*fβ*5, and L*fβ*8−*fα*8 (Figure 5B and Table S4). The side chains of the pathway residues protrude toward both the inside of the central $(\beta/\alpha)_8$ barrel and between the *α* and *β* elements. In the interface, the pathway residues cluster around the "hinge" (*h*W123−*f*R249) of the breathing motion. In HisH, both pathways go through different catalytic residues: dCNA goes through *h*H178 and *h*E180, whereas the PCN passes through *h*C48 and *h*H178 (Tables S2 vs S4). It is important to point out that both Ω−loop (*h*N12) and the oxyanion strand (*h*V51) are included in the dCNA pathway but not in the PCN pathway. Previous studies suggest that conformational changes around these regions are crucial for the activation of glutaminase activity.[12,56]

Both key residues, *f*D98 and *h*N12, are found to primarily play a dynamic role in the allosteric communication. The divergent allosteric pathways corroborate the distinct predictions by dCNA- and PCN-based suboptimal path analyses. To further understand why some of the experimentally verified key residues are detected by dCNA but not by the PCN and vice versa, we compare local properties of the networks and the structural environment around these key residues. The most drastic differences in the predictions are for the key residues *f*D98 and *h*N12. Both residues show high node degeneracies (0.64 and 0.47, respectively) in dCNA but zero degeneracies in the PCN.

https://doi.org/10.1021/acs.jctc.1c00669
J. Chem. Theory Comput. 2022, 18, 1173−1187

In the dCNA network, multiple (8 and 11, respectively, compared to an average of 2.6 across all nodes) significant contact changes associated with *f*D98 and *h*N12 are observed (Figure 5C). Especially, *f*D98, whose mutational effect is the most drastic, plays a pivotal role in the network integrity by mediating four significant contact changes (all are contact-strengthening) across the subunits. In contrast, in the PCN, the number of edges associated with *f*D98 and *h*N12 is small (3 and 2, respectively, compared to the network average of 6.7), and there is only one inter-subunit edge connecting *f*D98 and *h*K181 (Figure 5D), implying that *f*D98 or *h*N12 is not critical for the structural stability of IGPS. Hence, our results suggest that both *f*D98 and *h*N12 play a dynamic rather than a structural role in the allosteric communication between the effector, PRFAR, and the substrate, Gln.

The key residue *f*R5 may play both structural and dynamic roles in the allosteric communication, whereas the residue *f*V48 primarily plays a structural role. Neither *f*R5 nor *f*V48 is found in the pathway for the 2 → 4 process using dCNA, but both are present in the PCN pathway with significant degeneracies using a single ensemble (0.15 and 0.18, respectively; Figure 5E−H). Inspections on the networks around these residues show that both residues have a higher connectivity (11 and 8 edges, respectively) than the average (6.7) in the PCN, suggesting their important roles in maintaining the structural integrity of the protein. In dCNA, *f*R5 only shows a slightly higher connectivity (3 edges) than the average (2.6). This residue, together with *f*K99, *f*E46, and *f*E167, forms the "gate" of the central $(\beta/\alpha)_8$ barrel for the ammonia transfer.[24] Upon PRFAR binding, apparent contact weakening is observed between the four ionic residues, indicating gate opening (Figure 5E). The mutations of *f*R5 may disrupt the concerted gating motion, hindering the ammonia transfer process, leading to reduced glutaminase activity in the activated state. Alternatively, because of the high node connectivity in the PCN, the mutations of *f*R5 may simply compromise the structural integrity of HisF and decouple the two active sites. In contrast, *f*V48 has a low connectivity (1 edge) in dCNA, indicating its little participation in conformational changes. Examination of the local environment around *f*V48 reveals a hydrophobic cluster at the bottom of the central barrel close to the PRFAR binding site (Figure 5G,H). The mutations of *f*V48 may disrupt the cluster and eventually affect ammonia transfer and/or PRFAR binding. The importance of hydrophobic interactions in a similar region that define an allosteric pathway in IGPS has been stressed previously.[12] Together, we propose that *f*R5 has a dual (structural and dynamic) role, while *f*V48 primarily plays a structural role in the allosteric communication between the effector and the substrate.

The differences in the network approaches discussed in this work and some of the discrepancies with experiments could be explained by the following three possible reasons. First, they suggest the complexity in interpreting the results of experimental mutagenesis. The fidelity of allosteric regulation is conditioned by an optimal balance between the stability and flexibility of a protein. While dCNA primarily identifies residues involved in dynamic residue−residue contacts (e.g., *f*D98 and *h*N12), the PCN and similar approaches mainly identify residues that significantly contribute to the stability (e.g., *f*V48) and may constitute the scaffold to sustain the propagation of conformational changes mediated by the dynamic residues. The mutations of either type would disturb the stability−flexibility balance and compromise the allosteric coupling. In this regard, we argue that simply using mutagenesis data without further interpretations by, for example, MD simulations cannot fully elucidate the allosteric mechanism. Second, they may reflect limitations of each network approach discussed here. In this work, we use residue−residue contacts to describe conformational dynamics. Conformational changes below and beyond the distance cutoff used to define a contact are neglected. This may underlie the omission of *f*R5 by dCNA, which contributes to both the stability and flexibility. This residue may lie on an alternative pathway dominated by subtle structural changes not captured by residue−residue contacts. Third, it is possible that certain residues do not actively participate in certain processes (such as 2 → 4) or a specific ensemble and would not show in the pathway. Each of the binding process in the thermodynamic cycle potentially defines a specific allosteric pathway. Mutations on any of these pathways may affect the final strength of the allosteric coupling. Examining processes other than 2 → 4 may help capture some of the omitted key residues. We discuss this aspect in more detail in Section 3.2.3.

3.2.3. Similar Group of Residues Define the Allosteric Communication Pathways for Distinct Binding Processes in IGPS.

The allosteric pathways are largely similar for the different binding processes in the thermodynamic cycle (Figure 1B), except for some variations in individual residue degeneracies (Figures 3A,B, S3, and S4 and Table S2). A consensus allosteric pathway across all inspected processes is found to be L*f*β3−*f*α3, L*f*α3−*f*β4, *f*β4, L*f*β4−*f*α4, L*f*α4−*f*β5, L*h*β1−*h*α1 (Ω-loop), and L*h*β3−*h*α2 (oxyanion strand). Six critical residues that are shared by all the allosteric pathways are also identified: *f*I83, *f*D98, *f*N103, *f*Q123, *h*N12, and *h*V51. The constantly high degeneracies, especially at *f*D98 and *h*N12 that are far away from both binding sites (with degeneracies 0.51−0.68 and 0.22−0.56, respectively), indicate pivotal roles of these residues in the allosteric regulation of both substrate binding and catalysis. The overall similar pathways from distinct processes suggest a common mechanism exploited by the protein to regulate activities under different conditions.

Differences in exact node degeneracies between the allosteric pathways can help identify process-specific critical allosteric residues. For example, the main differences between the coupling-network (see Section 2.3) pathway and the 2 → 4 pathway occur near the PRFAR binding site, where the former shifts from loop 1 to a region near the loop L*f*β4−*f*α4 (Figure S4 and Table S2). As a result, node degeneracies of the key residues *f*V12 and *f*K19 reduce from 0.48 and 0.29, respectively, to nearly zero, whereas that for the key residue *f*T104 increases dramatically from 0.04 to 0.46, from 2 → 4 to the coupling network (Figure S3E and Table S2). Significant changes of degeneracy are also observed in other regions including the key residues *f*K99 and *f*Q123. Similar biases toward L*f*β4−*f*α4 (*f*T104 degeneracy 0.18−0.46), rather than loop 1 (*f*K19 degeneracy < 0.1), are observed in the 1 → 2, 1 → 3, 1 → 4, and 3 → 4 pathways (Figures S3A−D and S4 and Table S2), suggesting that loop 1 is specific to the 2 → 4 pathway. We speculate that residues specific to the 2 → 4 pathway may be primarily responsible for the regulation of catalysis. A support to this hypothesis is the fact that catalytic residues *h*H178 and *h*E180 are detected by the 2 → 4 network with significant node degeneracies (0.46 and 0.21, respectively) but not by the coupling network (which is assumed to be responsible for the regulation of substrate binding). Also, the relevance of loop 1 to IGPS activation has been discussed before.[56,68] In contrast, *f*T104 (on L*f*β4−*f*α4) is shared by all the pathways except for 2 → 4, suggesting that the residue may be mainly responsible for

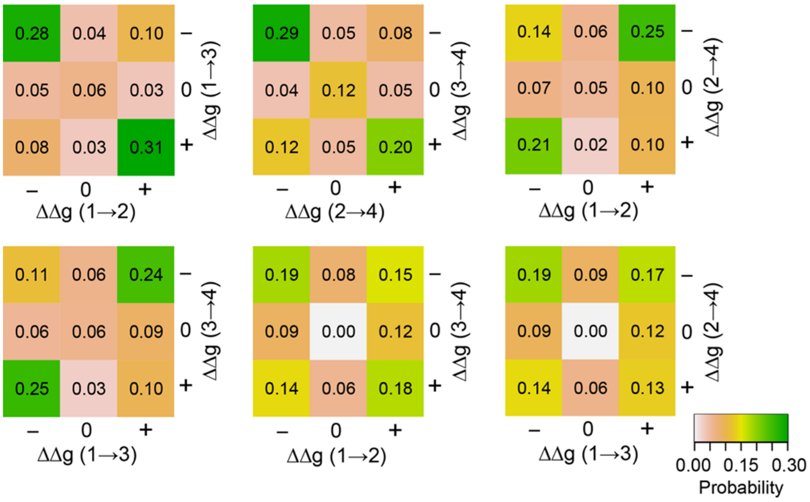


Figure 6. Modulation of the network of malleable residue−residue interactions during distinct ligand binding processes. Differential contact strengths ($\Delta\Delta g$) are categorized based on the sign and magnitude of the change: negative (−), zero (0), or positive (+). Joint probability distributions of the category of contact changes between distinct processes are then calculated and represented by a matrix with cells color-coded and labeled by the probabilities. Residue pairs associated with a union set of network edges across all processes considered (1 → 2, 1 → 3, 2 → 4, and 3 → 4) are included in the calculations. Definitions of states are given in Figure 1B.

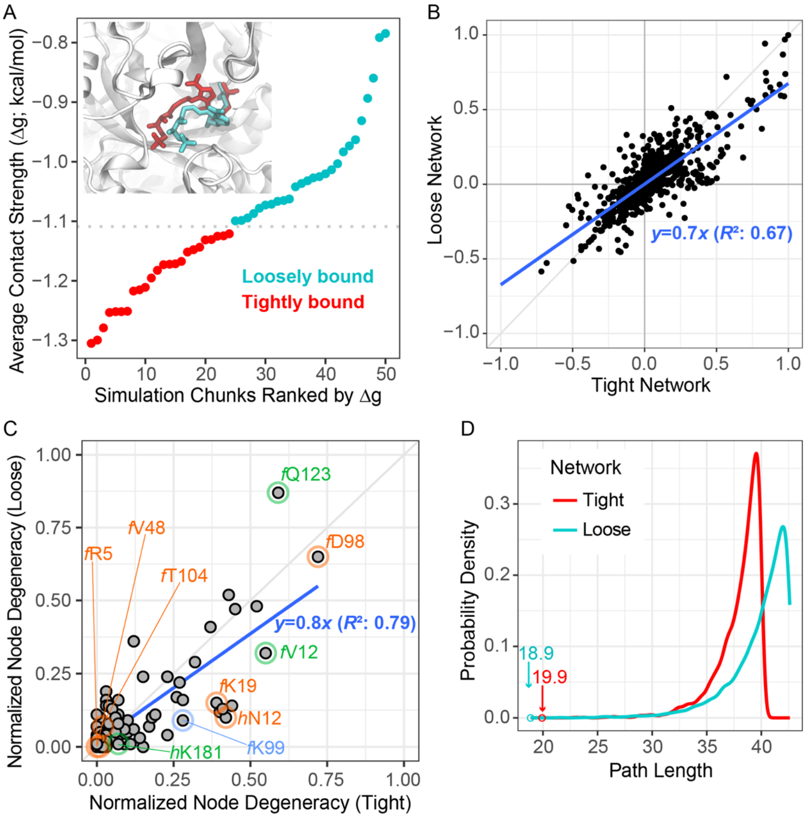


Figure 7. Fine-tuning of allosteric coupling by different levels of effector binding tightness. The 2 → 4 process (effector binding in the presence of the substrate) is considered. (A) Separation of the state 4 (holo/holo) simulation into two groups with the effector bound tightly and loosely, respectively. (B) Correlation of contact probability changes between the tight network (i.e., built based on the portion of simulation where the effector is tightly bound) and the loose network (using the portion of the simulation where the effector is loosely bound). The blue line represents a linear regression analysis with R^2 shown. (C) Comparison of node-degeneracy profiles obtained from the tight and loose networks. (D) Comparison of path length distributions. The shortest path length of each network is indicated by a labeled circle.

regulating Gln binding. However, more evidence is required to validate the hypotheses as the pathway around the PRFAR binding site is found to be slightly sensitive (in terms of the precision of the node degeneracies) to the model parameter p_c, and we cannot exclude the possibility that the difference is simply due to the choice of p_c = 0.2 (see Section 3.3 for more detail). These new insights into process-specific allosteric residues can be useful for future experimental design.

3.2.4. Similar Network Topologies Underlie the Similar Allosteric Pathways across Distinct Binding Processes in

IGPS. Comparisons of residue−residue contacts of different binding processes reveal a similarity in network topologies, which may underlie the similarity of the allosteric pathways revealed by dCNA. The joint probability distributions of the type of contact changes, associated with a union set of network edges across all processes considered (1 → 2, 1 → 3, 2 → 4, and 3 → 4), between two distinct processes are calculated (Figure 6). Differential contact strengths ($\Delta\Delta g$) are categorized into three classes based on their signs and magnitudes: negative (−), zero (0), and positive (+). The class "zero" indicates the absence of an edge in the network defined by the process. It shows that in each comparison, the probabilities of changes that are nonzero in one process and zero in the other are always the lowest (Figure 6), indicating a low population of process-specific network edges (i.e., network edges defined by the processes largely overlap). The result suggests that the modulation of the conformational ensemble by binding the substrate or effector happens via a similar set of residues (or residue−residue contacts), even though the direction and magnitude of the modulation may vary among different processes.

In the comparison of certain network topologies, evident positive or negative correlations between contact changes are found. For example, the distinct processes to form a binary complex (1 → 2 vs 1 → 3) exhibit a positive correlation, where most positive (negative) changes in $\Delta\Delta g$ during the binding of the substrate (1 → 2) also undergo positive (negative) changes during the binding of the effector (1 → 3) and vice versa (Figure 6). This indicates a strong coupling between the two binding processes. A similar but weaker positive correlation is observed between the processes of formation of the ternary complex (2 → 4 vs 3 → 4). The deviation of the network topology between these two processes may be responsible for the unique features of the allosteric pathway derived from the 2 → 4 network. Interestingly, in comparing sequential processes to form the ternary complex (1 → 2 vs 2 → 4 or 1 → 3 vs 3 → 4), a clear negative correlation is observed (Figure 6). This suggests that the modulation of the conformation upon binding the second ligand is mainly to reverse or modulate the changes caused by the binding of the first ligand. In contrast, the separate processes of binding the same ligand (1 → 2 vs 3 → 4 or 1 → 3 vs 2 → 4) exhibit the weakest correlation, indicating that response of the system to the binding of the substrate depends on whether the effector is present or not and vice versa. No matter whether the changes are correlated or not, changes mainly happen in a similar set of contacts, indicating that a common subnetwork of residue−residue interactions in the allosteric system mediates the response to different input signals (ligand binding).

3.2.5. Fine-Tuning of Allosteric Coupling via a Robust Network of Malleable Residue−Residue Interactions. The residue network mediating the propagation of allosteric signals from PRFAR to Gln is robust against different effector binding poses. In the 10 μs MD simulation of state 4 (Figure 1B), the effector PRFAR exhibits a certain degree of flexibility in the binding pocket. Examining a series of 200 ns simulation chunks, we observe divergent effector binding poses, associated with various levels of binding tightness measured by the average contact strength, Δg, between PRFAR and the binding-site residues (Figure 7A). We hypothesize that the magnitude of allosteric coupling depends on the effector binding tightness. To test this hypothesis, we group the simulation chunks by the effector's average Δg and construct two difference contact networks, termed tight (with PRFAR more tightly bound) and loose (PRFAR less tightly bound) networks, respectively, using the same 10 μs state 2 simulation as the reference. As expected, the overall (2 → 4) contact changes between the tight and the loose networks are correlated (R^2: 0.67; Figure 7B), suggesting that a similar subnetwork of residue−residue interactions is responsible for the allosteric communication in both tight and loose networks. A slope of 0.7 obtained from a linear regression analysis suggests that the magnitude of conformational changes in the tight network is overall stronger than that in the loose network.

Suboptimal path analyses further reveal a consistent allosteric pathway, even though the precision of the node degeneracies slightly varies between the two networks (R^2: 0.79; Figures 7C and S5). The pathway is overall similar to that using the full simulation (Figure S5 vs Figures 3A,B and S4). Path lengths in the tight network are overall shorter than those in the loose network, consistent with the larger contact changes in the tight network and implying an enhanced allosteric coupling by tightening the effector binding (Figure 7D).

Collectively, our results suggest that the magnitude of allosteric coupling in IGPS can be tuned by adjusting the effector binding pose without altering the allosteric network and communication pathway. A previous NMR study has tested a series of small-molecule allosteric effectors with varying activating efficacies and found widely dispersed and effector-dependent allosteric pathways.[62] These differing conclusions could be explained as follows. In this work, we only compared different binding poses of the same effector and hence cannot exclude the possibility that more divergent pathways might be observed in using different effectors as in the previous study. Besides, the discrepancy might be because of the timescale limitation in the μs-long MD simulations with respect to the millisecond timescale discussed in the NMR study. On the other hand, in the previous NMR study, only dynamics of backbone amide or methyl groups of Ile, Leu, and Val were inspected. This may limit the sensitivity of the method and reduce the chance to find otherwise overlapped allosteric networks. However, our computational studies fully support the NMR studies that the allosteric regulation in IGPS is not simply an "on or off" mechanism but can be fine-tuned by the identity or binding poses of the effector in the allosteric site. This agreement also aligns with a recent computational work, which has connected the detailed force distribution in binding pockets with the type of allostery (activation or inhibition) and allosteric pathways.[70] These previous and present studies emphasize the importance of inspecting detailed properties of the protein−ligand interaction, in addition to the ligand binding location and affinity, in acquiring desired allosteric output (e.g., in developing new allosteric drugs with the desired efficacy).

3.3. Convergence and Robustness of the Results. All the simulations are expected to be long enough (∼10 μs) to converge. To confirm it, we inspect the variation of cumulative contact statistics over time. We focus on contacts because they are the basis of all the analyses. As shown in Figure S6A, the root-mean-square deviation of contact probability (p) across all dynamic (with p in the range of 0.2−0.8 defined by the full-length simulation) contacts between two time points, t and t + 200 ns, becomes very small (<0.006) when the time t goes to the end of simulation. The standard deviation of square deviations (displayed as error bars in Figure S6A) also becomes small (<0.004) with the increase of t. Additional examinations are performed on the maximal deviation of contact probabilities, which reduces to ∼0.016 at t = 10 μs (Figure S6B). Hence, the result strongly supports that all the simulations have converged at least with respect to contact statistics.

The node degeneracy results are robust against variations of k (number of paths) around the optimal value $k = 5000$. In the suboptimal path analysis, it is hard to have all node degeneracies to truly converge. This is partly because of the equal treatment of all suboptimal paths in the definition of degeneracy: when k becomes very large, distinct areas of the network will start to be visited and formerly "stable" degeneracy values will vary again. A simple solution is to apply a threshold of the path length and only consider paths shorter than the threshold. However, such a threshold is unknown a priori. Here, instead of finding the path-length threshold, we identify an appropriate k by examining the change of path length, L, with the increase of k. We adopt a similar rule of thumb to that used in the principal component analysis of protein structures to identify significant principal components,[21] that is, we choose a k value that lies in the "elbow" of the $L \sim k$ curve, which is roughly around 5000 (Figure S7A). To confirm that the results are robust against variations of k, we perform additional path analyses with k = 1000 and k = 10,000, respectively. It clearly shows that node-degeneracy profiles are almost identical among different k's, proving the robustness of our results (Figure S7B,C). Similar results are obtained for the DN and PCN (Figures S8 and S9).

The allosteric pathway identified by dCNA is invariant with the parameter ε but slightly varies with the contact probability threshold p_c (eq 1). Besides the optimal $p_c = 0.20$ and $\varepsilon = 2.0$, we test different combinations of two alternative values of each parameter: 0.10 and 0.15 for p_c; 1.5 and 2.5 for ε. The node-degeneracy profile is nearly identical with the change of ε (Figure S10). However, slight variations are observed when p_c is lower than 0.20, which mainly contain reductions of degeneracy in loop 1. As discussed in Section 3.2.3, the divergence of pathways between the process 2 → 4 and other processes mainly happens in loop 1. This divergence may have functional implications or simply be a consequence of the particular choice of $p_c = 0.20$. Except for loop 1, degeneracy profiles are overall similar to that with $p_c = 0.20$. Importantly, all parameters tested exhibit high degeneracy values for the key residues *f*D98 and *h*N12. Note that we do not test p_c's above 0.20 as a larger p_c removes more contact changes, some of which are critical for the integrity of the network. In summary, our method is generally robust against small variations of parameters but may generate differences in the precision of the node degeneracies with different choices of p_c.

The pathway by dCNA is also robust against variations in the details of the network construction. We first investigate how the result changes over modifications of eq 2 in Section 2.1. The current dCNA framework discussed above (termed "strengthening" network below) considers the direction of conformational changes, and contact strengthening from one state to the other is treated as more important in the network than contact weakening. Two additional networks are built (for the 2 → 4 process) for comparisons: one weights edges by the magnitude of contact changes (i.e., $|\Delta\Delta g|$ or ignoring signs of changes; termed "nondirectional" network) and the other reverses the importance of direction, treating contact weakening (positive $\Delta\Delta g$ or decrease of contact probability) as more important (termed "weakening" network). All other procedures are the same as in the strengthening network. The pathways derived from the strengthening and the nondirectional networks are very similar, with node degeneracies highly correlated over the entire protein (R^2: 0.87; Figure S11A). A slightly lower but still notable correlation of node degeneracies is also observed between the strengthening and the weakening networks (R^2: 0.61; Figure S11B). This suggests that the result is robust against the details of how the network edges are weighted. However, the strengthening network exhibits a slightly better performance in detecting the experimentally verified key residues. For example, the strengthening network predicts a much higher node degeneracy for the most important *f*D98 than the weakening network (0.64 vs 0.28; Figure S11B), suggesting a more function-relevant allosteric pathway identified in the strengthening network than that in the weakening network. Note that the choice of network also depends on the specific direction of the process considered. For example, if the 4 → 2 process (effector dissociation) is considered, the weakening network should be used as it is equivalent to the strengthening network for the effector binding process.

We further verify the robustness of the node degeneracy result by dCNA using a different strategy of the network construction. As an alternative approach to implement the idea of dCNA for suboptimal path analysis, we build the network by simply adding edges between all residue pairs with a nonzero contact probability change ($|\Delta p|>0$). The edges are weighted in a similar way as in the DN, that is, by $-\ln(|\Delta p|)$. Except for variations in loop 1, the node-degeneracy profile of this method is largely similar to that by the network based on differential contact strengths ($\Delta\Delta g$) (Figure S12 vs Figures 3A,B and S4). Importantly, both *f*D98 and *h*N12 have a high node degeneracy (0.37 and 0.49, respectively). Hence, the general principle of comparing distinct conformational ensembles, rather than the details of how the network is constructed, contributes the most to the unique performance of the dCNA-based suboptimal path analysis. Note that although both versions of dCNA have comparable performance, the network weighted by differential contact strengths or $\Delta\Delta g$'s has a clearer physical interpretation and is therefore preferable.

4. CONCLUSIONS

We introduce a new method for mapping the allosteric communication pathways in biomolecules by implementing the suboptimal path analysis within the dCNA framework. The method exploits a protein structure network where the edges are defined by changes of residue−residue contact statistics during a functional process, such as the binding of an allosteric effector. The method is thoroughly tested on the well-studied allosteric system, IGPS, and is compared to conventional network methods, including the DN and the PCN. Both dCNA and the PCN identify experimentally verified key residues that do not overlap, suggesting distinct behaviors of the two methods in elucidating allosteric communication pathways. The key residues detected by dCNA primarily play a dynamic role, whereas those by the PCN are found to primarily play a structural role in the allosteric communication. In contrast, the DN method is found to be sensitive to the input simulation and does not capture experimentally identified key residues very well.

The dCNA-based path analysis, along with conventional analyses, is also applied to examine distinct substrate/effector binding processes in the thermodynamic cycle. The 10 μs MD simulations identify the representative conformational changes during IGPS activation, even though there are slight differences between the interpretations of the computational and experimental results. The dCNA path analysis reveals largely similar allosteric pathways, except for slight variations in the exact node degeneracies, especially near the effector binding site. The similarity in pathways is a consequence of a similar subnetwork of residue−residue interactions (or a similar group of residues) utilized by IGPS to propagate allosteric signals via modulating the strengths of contacts. Similar subnetwork and allosteric pathway are also utilized by the system to fine-tune the allosteric coupling via changing the effector binding tightness. Our method is generally robust against changes of parameters or the details of the network construction, although slight variations in the precision of the node degeneracy may occur upon selecting different contact probability thresholds. Careful examinations are also performed to assure the convergence of the main results, including contact statistics and node degeneracy calculations. Our study shows that dCNA-based suboptimal path analysis is a powerful tool to identify critical residues for the allosteric communication in IGPS. The dCNA path analysis is general enough to be easily extended to other systems.

Finally, we stress that the general dCNA framework enables not only the community[22] and suboptimal path analyses but also other popular network analyses, such as the calculation of betweenness centralities, which can be used to identify potential key allosteric residues without knowing the coupled binding sites a priori (Figure S13).

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.1c00669.

> Experimental kinetic parameters for basal and activated glutaminase activities of IGPS wild-type and mutant variants; allosteric pathways identified by the dCNA path analysis; allosteric pathways identified by the DN; allosteric pathways identified by the PCN; ; results of DN-based suboptimal path analysis for distinct states; results of PCN-based suboptimal path analysis for distinct states; results of dCNA-based suboptimal path analysis for distinct binding processes; allosteric pathways in the tube representation; fine-tuning of the allosteric coupling by different levels of effector binding tightness; convergence of simulations with respect to contact statistics; determination of the number of paths k and the robustness of the node-degeneracy profile against variations of k for dCNA, DN, and PCN; robustness of node-degeneracy profiles against variations of parameters in dCNA; robustness of the node degeneracy result against different weighting schemes of the network edges; suboptimal path analysis of dCNA with network edges weighted by $-\ln(|\Delta p|)$; and dCNA-based betweenness centrality analysis (PDF)

■ 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.jctc.1c00669

Notes

The authors declare no competing financial interest.

Data Availability: The source code for dCNA path analysis is available at https://github.com/The-Hamelberg-Group/dcna.

■ ACKNOWLEDGMENTS

This research was supported by the National Science Foundation (MCB-2018144). We also acknowledge supports from Georgia State University and the Georgia Research Alliance. This work used the Advanced Research Computing Technology and Innovation Core (ARCTIC) resources, which is supported by the National Science Foundation Major Research Instrumentation (MRI) grant no. CNS-1920024. We acknowledge the use of Georgia State's research computing resources that are supported by Georgia State's Research Solutions.

■ ABBREVIATIONS

IGPS, imidazole glycerol phosphate synthase; PRFAR, *N*′-[(5′-phosphoribulosyl)formimino]-5-aminoimidazole-4-carboxamide ribonucleotide; Gln, glutamine; MD, molecular dynamics; dCNA, difference contact network analysis; DN, dynamical network; PCN, protein contact network; NMR, nuclear magnetic resonance

■ REFERENCES

(1) Goodey, N. M.; Benkovic, S. J. Allosteric Regulation and Catalysis Emerge via a Common Route. *Nat. Chem. Biol.* **2008**, *4*, 474−482.

(2) Smock, R. G.; Gierasch, L. M. Sending Signals Dynamically. *Science* **2009**, *324*, 198−203.

(3) Vale, R. D.; Milligan, R. A. The way things move: looking under the hood of molecular motor proteins. *Science* **2000**, *288*, 88−95.

(4) Nussinov, R.; Tsai, C.-J. Allostery in disease and in drug discovery. *Cell* **2013**, *153*, 293−305.

(5) Wagner, J. R.; Lee, C. T.; Durrant, J. D.; Malmstrom, R. D.; Feher, V. A.; Amaro, R. E. Emerging Computational Methods for the Rational Discovery of Allosteric Drugs. *Chem. Rev.* **2016**, *116*, 6370−6390.

(6) Raman, S. Systems Approaches to Understanding and Designing Allosteric Proteins. *Biochemistry* **2018**, *57*, 376−382.

(7) Pirro, F.; Schmidt, N.; Lincoff, J.; Widel, Z. X.; Polizzi, N. F.; Liu, L.; Therien, M. J.; Grabe, M.; Chino, M.; Lombardi, A.; DeGrado, W. F. Allosteric cooperation in a de novo-designed two-domain protein. *Proc. Natl. Acad. Sci. U.S.A.* **2020**, *117*, 33246−33253.

(8) Frauenfelder, H.; Sligar, S. G.; Wolynes, P. G. The energy landscapes and motions of proteins. *Science* **1991**, *254*, 1598−1603.

(9) 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.

(10) 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.

(11) 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.

(12) 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.

(13) 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.

(14) 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.

(15) Yao, X.-Q.; Malik, R. U.; Griggs, N. W.; Skjærven, 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.

(16) Botello-Smith, W. M.; Luo, Y. Robust Determination of Protein Allosteric Signaling Pathways. *J. Chem. Theory Comput.* **2019**, *15*, 2116−2126.

(17) Eargle, J.; Luthey-Schulten, Z. NetworkView: 3D display and analysis of protein.RNA interaction networks. *Bioinformatics* **2012**, *28*, 3000−3001.

(18) Van Wart, A. T.; Durrant, J.; Votapka, L.; Amaro, R. E. Weighted Implementation of Suboptimal Paths (WISP): An Optimized Algorithm and Tool for Dynamical Network Analysis. *J. Chem. Theory Comput.* **2014**, *10*, 511−517.

(19) Grant, B. J.; Skjærven, L.; Yao, X. Q. The Bio3D packages for structural bioinformatics. *Protein Sci.* **2021**, *30*, 20−30.

(20) Dodd, T.; Yao, X. Q.; Hamelberg, D.; Ivanov, I. Subsets of adjacent nodes (SOAN): a fast method for computing suboptimal paths in protein dynamic networks. *Mol. Phys.* **2021**, *199*, No. e1893847.

(21) Yao, X.-Q.; Hamelberg, D. Detecting Functional Dynamics in Proteins with Comparative Perturbed-Ensembles Analysis. *Acc. Chem. Res.* **2019**, *52*, 3455−3464.

(22) 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.

(23) Chaudhuri, B. N.; Lange, S. C.; Myers, R. S.; Chittur, S. V.; Davisson, V. J.; Smith, J. L. Crystal structure of imidazole glycerol phosphate synthase: a tunnel through a (beta/alpha)8 barrel joins two active sites. *Structure* **2001**, *9*, 987−997.

(24) Douangamath, A.; Walker, M.; Beismann-Driemeyer, S.; Vega-Fernandez, M. C.; Sterner, R.; Wilmanns, M. Structural evidence for ammonia tunneling across the (beta alpha)(8) barrel of the imidazole glycerol phosphate synthase bienzyme complex. *Structure* **2002**, *10*, 185−193.

(25) Myers, R. S.; Jensen, J. R.; Deras, I. L.; Smith, J. L.; Davisson, V. J. Substrate-induced changes in the ammonia channel for imidazole glycerol phosphate synthase. *Biochemistry* **2003**, *42*, 7013−7022.

(26) Yao, X.-Q.; Momin, M.; Hamelberg, D. Establishing a Framework of Using Residue-Residue Interactions in Protein Difference Network Analysis. *J. Chem. Inf. Model.* **2019**, *59*, 3222−3228.

(27) 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.

(28) 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.

(29) 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.

(30) Venkatakrishnan, A. J.; Deupi, X.; Lebon, G.; Heydenreich, F. M.; Flock, T.; Miljus, T.; Balaji, S.; Bouvier, M.; Veprintsev, D. B.; Tate, C. G.; Schertler, G. F. X.; Babu, M. M. Diverse Activation Pathways in Class A GPCRs Converge near the G-Protein-Coupling Region. *Nature* **2016**, *536*, 484−487.

(31) Yen, J. Y. Finding theKShortest Loopless Paths in a Network. *Manage. Sci.* **1971**, *17*, 712−716.

(32) Case, D. A.; Aktulga, H. M.; Belfon, K.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Giambasu, G.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.; Harris, R.; Izadi, S.; Izmailov, S. A.; Jin, C.; Kasavajhala, K.; Kaymak, M. C.; King, E.; Kovalenko, A.; Kurtzman, T.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Machado, M.; Man, V.; Manathunga, M.; Merz, K. M.; Miao, Y.; Mikhailovskii, O.; Monard, G.; Nguyen, H.; O'Hearn, K. A.; Onufriev, A.; Pan, F.; Pantano, S.; Qi, R.; Rahnamoun, A.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C. L.; Skrynnikov, N. R.; Smith, J.; Swails, J.; Walker, R. C.; Wang, J.; Wei, H.; Wolf, R. M.; Wu, X.; Xue, Y.; York, D. M.; Zhao, S.; Kollman, P. A. *Amber 2021*; University of California: San Francisco, 2021.

(33) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.

(34) List, F.; Vega, M. C.; Razeto, A.; Häger, M. C.; Sterner, R.; Wilmanns, M. Catalysis uncoupling in a glutamine amidotransferase bienzyme by unblocking the glutaminase active site. *Chem. Biol.* **2012**, *19*, 1589−1599.

(35) Chaudhuri, B. N.; Lange, S. C.; Myers, R. S.; Davisson, V. J.; Smith, J. L. Toward understanding the mechanism of the complex cyclization reaction catalyzed by imidazole glycerolphosphate synthase: crystal structures of a ternary complex and the free enzyme. *Biochemistry* **2003**, *42*, 7003−7012.

(36) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696−3713.

(37) Horn, A. H. C. A consistent force field parameter set for zwitterionic amino acid residues. *J. Mol. Model.* **2014**, *20*, 2478.

(38) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general Amber force field. *J. Comput. Chem.* **2004**, *25*, 1157−1174.

(39) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic Charges. AM1-BCC model: I. Method. *J. Comput. Chem.* **2000**, *21*, 132−146.

(40) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem.* **2002**, *23*, 1623−1641.

(41) Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J. Chem. Theory Comput.* **2011**, *7*, 525−537.

(42) Søndergaard, C. R.; Olsson, M. H.; Rostkowski, M.; Jensen, J. H. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values. *J. Chem. Theory Comput.* **2011**, *7*, 2284−2295.

(43) Dolinsky, T. J.; Nielsen, J. E.; McCammon, J. A.; Baker, N. A. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res.* **2004**, *32*, W665−W667.

(44) Jurrus, E.; Engel, D.; Star, K.; Monson, K.; Brandi, J.; Felberg, L. E.; Brookes, D. H.; Wilson, L.; Chen, J.; Liles, K.; Chun, M.; Li, P.; Gohara, D. W.; Dolinsky, T.; Konecny, R.; Koes, D. R.; Nielsen, J. E.; Head-Gordon, T.; Geng, W.; Krasny, R.; Wei, G. W.; Holst, M. J.; McCammon, J. A.; Baker, N. A. Improvements to the APBS biomolecular solvation software suite. *Protein Sci.* **2018**, *27*, 112−128.

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

(46) Joung, I. S.; Cheatham, T. E., 3rd Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *J. Phys. Chem. B* **2008**, *112*, 9020−9041.

(47) Joung, I. S.; Cheatham, T. E., 3rd Molecular Dynamics Simulations of the Dynamic and Energetic Properties of Alkali and Halide Ions Using Water-Model-Specific Ion Parameters. *J. Phys. Chem. B* **2009**, *113*, 13279−13290.

(48) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N·log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.

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

(50) Miyamoto, S.; Kollman, P. A. SETTLE: An Analytical Version of the SHAKE and RATTLE Algorithm for Rigid Water Models. *J. Comput. Chem.* **1992**, *13*, 952−962.

(51) Roe, D. R.; Cheatham, T. E., 3rd PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9*, 3084−3095.

(52) Skjærven, L.; Yao, X.-Q.; Scarabelli, G.; Grant, B. J. Integrating Protein Structural Dynamics and Evolutionary Analysis with Bio3D. *BMC Bioinf.* **2014**, *15*, 399.

(53) Grant, B. J.; Rodrigues, A. P. C.; ElSawy, K. M.; McCammon, J. A.; Caves, L. S. D. Bio3d: An R Package for the Comparative Analysis of Protein Structures. *Bioinformatics* **2006**, *22*, 2695−2696.

(54) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14*, 33−38.

(55) Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*, 2nd ed.; Springer-Verlag: New York, 2016; p 260.

(56) Wurm, J. P.; Sung, S.; Kneuttinger, A. C.; Hupfeld, E.; Sterner, R.; Wilmanns, M.; Sprangers, R. Molecular basis for the allosteric activation mechanism of the heterodimeric imidazole glycerol phosphate synthase complex. *Nat. Commun.* **2021**, *12*, 2748.

(57) Lake, P. T.; Davidson, R. B.; Klem, H.; Hocky, G. M.; McCullagh, M. Residue-Level Allostery Propagates through the Effective Coarse-Grained Hessian. *J. Chem. Theory Comput.* **2020**, *16*, 3385−3395.

(58) Myers, R. S.; Amaro, R. E.; Luthey-Schulten, Z. A.; Davisson, V. J. Reaction coupling through interdomain contacts in imidazole glycerol phosphate synthase. *Biochemistry* **2005**, *44*, 11974−11985.

(59) Lisi, G. P.; East, K. W.; Batista, V. S.; Loria, J. P. Altering the allosteric pathway in IGPS suppresses millisecond motions and catalytic activity. *Proc. Natl. Acad. Sci. U.S.A.* **2017**, *114*, E3414−E3423.

(60) Amaro, R. E.; Sethi, A.; Myers, R. S.; Davisson, V. J.; Luthey-Schulten, Z. A. A network of conserved interactions regulates the allosteric signal in a glutamine amidotransferase. *Biochemistry* **2007**, *46*, 2156−2173.

(61) Lipchock, J. M.; Loria, J. P. Nanometer propagation of millisecond motions in V-type allostery. *Structure* **2010**, *18*, 1596−1607.

(62) Lisi, G. P.; Manley, G. A.; Hendrickson, H.; Rivalta, I.; Batista, V. S.; Loria, J. P. Dissecting Dynamic Allosteric Pathways Using Chemically Related Small-Molecule Activators. *Structure* **2016**, *24*, 1155−1166.

(63) Kneuttinger, A. C.; Straub, K.; Bittner, P.; Simeth, N. A.; Bruckmann, A.; Busch, F.; Rajendran, C.; Hupfeld, E.; Wysocki, V. H.; Horinek, D.; König, B.; Merkl, R.; Sterner, R. Light Regulation of Enzyme Allostery through Photo-responsive Unnatural Amino Acids. *Cell Chem. Biol.* **2019**, *26*, 1501−1514.

(64) Kneuttinger, A. C.; Rajendran, C.; Simeth, N. A.; Bruckmann, A.; König, B.; Sterner, R. Significance of the Protein Interface Configuration for Allostery in Imidazole Glycerol Phosphate Synthase. *Biochemistry* **2020**, *59*, 2729−2742.

(65) 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.

(66) 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.

(67) Rivalta, I.; Lisi, G. P.; Snoeberger, N.-S.; Manley, G.; Loria, J. P.; Batista, V. S. Allosteric Communication Disrupted by a Small Molecule Binding to the Imidazole Glycerol Phosphate Synthase Protein-Protein Interface. *Biochemistry* **2016**, *55*, 6484−6494.

(68) Negre, C. F. A.; Morzan, U. N.; Hendrickson, H. P.; Pal, R.; Lisi, G. P.; Loria, J. P.; Rivalta, I.; Ho, J.; Batista, V. S. Eigenvector centrality for characterization of protein allosteric pathways. *Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115*, E12201−E12208.

(69) Gheeraert, A.; Pacini, L.; Batista, V. S.; Vuillon, L.; Lesieur, C.; Rivalta, I. Exploring Allosteric Pathways of a V-Type Enzyme with Dynamical Perturbation Networks. *J. Phys. Chem. B* **2019**, *123*, 3452−3461.

(70) Huang, Q.; Song, P.; Chen, Y.; Liu, Z.; Lai, L. Allosteric Type and Pathways Are Governed by the Forces of Protein-Ligand Binding. *J. Phys. Chem. Lett.* **2021**, *12*, 5404−5412.