

# Exploring Residue Component Contributions to Dynamical Network Models of Allostery

Adam T. VanWart,<sup>†</sup> John Eargle,<sup>‡</sup> Zaida Luthey-Schulten,<sup>‡</sup> and Rommie E. Amaro<sup>\*,†</sup>

<sup>†</sup>Department of Chemistry and Biochemistry, University of California, San Diego, La Jolla, California 92093, United States

<sup>‡</sup>Department of Chemistry, University of Illinois at Urbana—Champaign, Urbana, Illinois 61801, United States

## S Supporting Information

**ABSTRACT:** Allosteric regulation in biological systems is of considerable interest given the vast number of proteins that exhibit such behavior. Network models obtained from molecular dynamics simulations have been shown to be powerful tools for the analysis of allostery. In this work, different coarse-grain residue representations (nodes) are used together with a dynamical network model to investigate models of allosteric regulation. This model assumes that allosteric signals are dependent on positional correlations of protein substituents, as determined through molecular dynamics simulations, and uses correlated motion to generate a signaling weight between two given nodes. We examine four types of network models using different node representations in Cartesian coordinates: the (i) residue  $\alpha$ -carbons, (ii) the side chain center of mass, (iii) the backbone center of mass, and the entire (iv) residue center of mass. All correlations are filtered by a dynamic contact map that defines the allowable interactions between nodes based on physical proximity. We apply the four models to imidazole glycerol phosphate synthase (IGPS), which provides a well-studied experimental framework in which allosteric communication is known to persist across disparate protein domains (e.g., a protein dimer interface). IGPS is modeled as a network of nodes and weighted edges. Optimal allosteric pathways are traced using the Floyd Warshall algorithm for weighted networks, and community analysis (a form of hierarchical clustering) is performed using the Girvan–Newman algorithm. Our results show that dynamical information encoded in the residue center of mass must be included in order to detect residues that are experimentally known to play a role in allosteric communication for IGPS. More broadly, this new method may be useful for predicting pathways of allosteric communication for any biomolecular system in atomic detail.

## INTRODUCTION

Proteins and enzymes act as catalytic waypoints along a natural products assembly line. They are the molecular machinery of the cell designed to assist complex biological synthesis and mediate biosynthetic regulation. Allostery is broadly categorized as the mechanism, or set of mechanisms, for which an event at a distinct site affects the activity, function, conformational ensemble, or general attributes at another site within an enzyme or protein.<sup>1</sup> As an example, in the case of allosteric modulation, the catalytic efficiency of an allosterically active enzyme or protein is affected through the binding of a metabolite to an allosteric site. The metabolite is said to be an effector molecule that may propagate an activation or inhibition signal to a distant catalytic site that in turn regulates the catalytic activity of the enzyme or protein. Furthermore, a metabolite may take part in the catalytic reaction itself.<sup>2,3</sup> In this way, allostery mechanizes regulation within metabolic pathways as an intrinsic characteristic of a variety of proteins and enzymes. Further understanding of allostery may lead to advances in practical applications, such as rational drug design, protein engineering, and development of synthetic catalysts.

Allosteric regulation usually involves at least two different stereospecific receptor sites.<sup>2–4</sup> As metabolic pathways are critical to living organisms, it follows that communication between both receptor sites must be a protected feature of allosteric modulation. Consequently, homologous proteins often preserve structural elements that conserve communication between receptor sites, the central theme being that

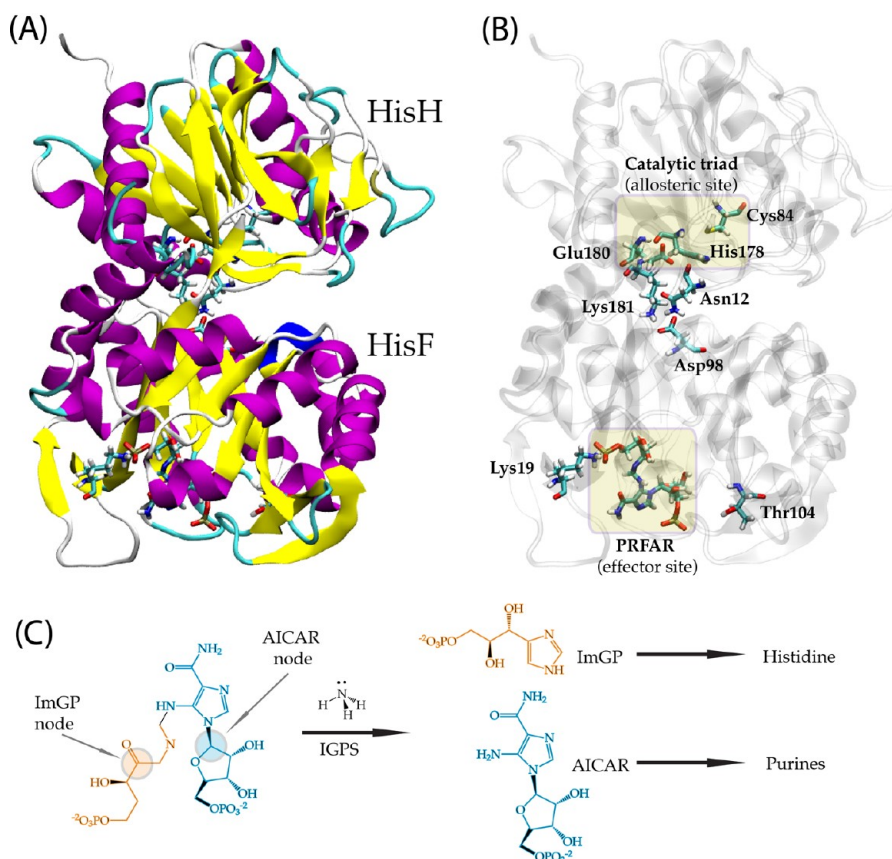
structural elements important to protein function, and thus allosteric signaling, have been evolutionarily conserved.<sup>5–8</sup> A notable proponent of this theme can be found in Statistical Coupling Analysis first introduced by Lockless and Ranganathan.<sup>9</sup>

Much effort has been put forth to develop rigorous theories of allosteric modulation. Some of the early foundations of allostery are proposed in the concerted MWC model (Monod, Wyman, and Changeux<sup>3</sup>) and sequential KNF model (Koshland, Nemethy, and Filmer<sup>10</sup>). These models conceptualize the mechanism within a protein modulated by an allosteric effector as a conformational transition that is either concerted or induced sequentially.<sup>11</sup> Since then, mechanistic details have become rooted in an ensemble view of the protein where cooperativity between sites is governed by a density of states that is intimately connected to entropy and enthalpy.<sup>12,13</sup> The accessibility of an allosteric state may be through an effector substrate inducing a large conformational change within the protein complex as well as small local conformational changes.<sup>14</sup> However, allosteric regulation can also occur in the absence of large conformational changes, as found in works by Cooper and Dryden.<sup>15</sup> In such a situation, an enzyme or protein would have the propensity to facilitate allosteric signals due to small vibrations inherent to the protein's state and not due to a change in shape or fold.

Received: May 13, 2012

Published: July 5, 2012





**Figure 1.** The structure and reaction of HisH–HisF. (A) The two domains, HisH (glutaminase domain) and HisF (cyclase domain), are shown colored by secondary structural elements. (B) The catalytic triad, residues tested experimentally for allostery, and the substrate PRFAR are presented again in ball and stick representation for clarity with the same orientation as A, showing the allosteric and effector sites, which are over 30 Å apart. (C) The overall reaction catalyzed by HisH and HisF produces AICAR and ImGP.

Experimental constructs to explore allostery have included X-ray crystallography with site-directed mutagenesis,  $^1\text{H}$ – $^{15}\text{N}$  heteronuclear single quantum correlation (HSQC) NMR, kinetic assays with site directed mutagenesis, inhibition assays, and others.<sup>16</sup> These tests probe a range of questions associated with small molecule effectors, protein–protein interfaces (hot spots), allosteric pathways, and general protein functionality. NMR provides increasing potential for a direct view into the mechanics of allostery. The  $^1\text{H}$ – $^{15}\text{N}$  amide bond vector is an inherent characteristic of polypeptides and thus has been used to characterize fast backbone and side-chain dynamics. In NMR experiments, thermal motions are present in all proteins above the glass transition temperature, and there is a direct link between molecular motions and local magnetic field fluctuations. These dynamics are commonly associated with the Lipari–Szabo order parameter ( $S^2$ ).<sup>17,18</sup> It is generally accepted that picosecond to nanosecond time scales capture allostery in the absence of large conformational shifts for a majority of proteins (see review<sup>19</sup>). The parameter  $S^2$  is related to the Gibbs free energy through the canonical partition function and provides a means of deriving conformational entropy from bond vector fluctuations.<sup>20</sup>

Molecular dynamics seeks to give accurate virtual results that compete with experiments at a fraction of the cost. With the use of molecular dynamics simulations,<sup>21,22</sup> details of conformational changes are sampled to clarify and categorize mechanisms of allostery that may otherwise be elusive. Quasi-harmonic analysis is a popular analytical tool that has been used

in molecular dynamics to quantify motions in the protein's trajectory ensemble.<sup>23–25</sup> This approach aligns trajectories to a reference frame, often the average of the protein, and can be used to deduce correlations among variables (e.g., dihedral angles, position) that account for the configurational space. However, harmonic motions may be an oversimplified assumption for protein dynamics, and the need for an anharmonic metric arises. Mutual information methods outlined by Lange and Grubmüller provide a more generalized framework that captures anharmonic motions in Cartesian coordinates.<sup>26</sup> Recent works by Killian et al. derive configurational entropies using internal coordinates for a trajectory ensemble.<sup>27</sup> The mutual information values are the second-order terms derived from the entropy expansion and directly link correlated conformations within the protein to molecular configurational entropy values. This work inspired McClendon and co-workers to develop an entropy-based approach to harmonic and anharmonic correlations via mutual information (MutInf) among dihedral angles;<sup>28</sup> this method includes Bayesian filtering as well as sampling penalties to generate more statistically robust values.

Dynamical network analysis is another method used to explore putative allosteric communication pathways.<sup>29,30</sup> Within the realm of biological computation, dynamical network analysis samples correlations among time dependent variables for predefined components of a biological system (i.e., protein). These variables are associated with the trajectory ensemble and provide a distribution at equilibrium. As examples, a variable

could be the position of an atom, center of mass (c.o.m.) of a residue, or a vector modeling the orientation of a residue. These variables serve to model the configuration of the protein at a given time. As developed by Luthey-Schulten and colleagues,<sup>29,30</sup> the variable was the position of each  $\alpha$ -carbon atom within an aminoacyl-tRNA synthetase system. MD simulations sample the system dynamics, which are used to generate correlation values among all variables (nodes) of the biological system. The absolute value of a given correlation is then incorporated into a function that models the signaling between nodes as a distance. The behavior of nodes that are highly correlated and within close physical proximity (aggregates) can be analyzed in terms of community behavior and structure. Allosteric signaling between community aggregates occurs through particular node pairs that exhibit a high degree of network degeneracy (i.e., community cross talk). These “critical nodes” transfer a relatively large degree of information across their edge. As critical node pairs carry out intercommunity cross talk, they may provide favorable sites for signal disruption.

In this work, we extend the dynamical network analysis method and propose allosteric signaling pathways in imidazole glycerol phosphate synthase (IGPS; HisH–HisF), which belongs to the glutamine amidotransferase (GAT) family of enzymes (Figure 1). This multidomain globular protein regulates the fifth step of the histidine biosynthetic pathway in plants, fungi, and microbes. Histidine is considered an essential amino acid, as humans do not have a histidine biosynthetic pathway. In bacteria, HisH–HisF is a protein dimer, whereas in eukaryotes it exists as one single polypeptide chain, with a  $\sim 40$  amino-acid-residue tether connecting the two domains. The HisH–HisF dimer has two receptor sites: one receptor site (effector site) is positioned at the C-terminal cyclase domain of HisF, and the other (allosteric site) is located in HisH (the glutaminase domain) roughly 30 Å away.<sup>31</sup> The triad GAT subfamily strictly conserves these catalytic residues across all members of the set, although the allosteric effector, remote binding site, upregulation, and allosteric site vary between different enzymes.<sup>32–35</sup> Another important conserved subfamily structure in GATs is a  $(\beta/\alpha)_8$  barrel which is located in HisF. The  $(\beta/\alpha)_8$  barrel is believed to assist in transport of the ammonia moiety to the effector or substrate molecule and may also play a role in allostery.<sup>36–38</sup>

Attachment of the primary metabolite N1-[(5'-phosphoribulose)-formimino]-5-aminoimidazole-4-carboxamide ribonucleotide (PRFAR, which is also the allosteric effector) to the cyclase domain induces several global structural changes in the protein dimer. A flexible loop on the C-terminal end of the  $(\beta/\alpha)_8$  barrel appears to modulate access to the PRFAR substrate, and a hinge motion at the dimer interface changes the HisF and HisH domains by roughly 7.1° to form a tighter interface.<sup>8</sup> These large scale conformational changes indicate that the PRFAR-bound (holo) and PRFAR-unbound (apo) states are distinct. Other induced motions consist of smaller conformational changes in the side chain orientations.<sup>37</sup> The catalytic turnover of glutamine to glutamate is increased  $\sim 4900$  fold when PRFAR is bound; however, the details of the allosteric mechanism are not completely understood.<sup>8</sup> All members of the GAT family of enzymes exhibit similar allosteric regulation to control the production and utilization of ammonia in the subsequent catalytic reaction.<sup>39</sup> After the catalytic turnover of glutamine, nascent ammonia is shuttled across the dimer interface and through the center of the  $(\beta/\alpha)_8$  barrel along the

barrel's principal axis to the cyclase domain where PRFAR is broken into 5'-(5-aminoimidazole-4-carbox-amide) ribonucleotide (AICAR) and ImGP (Figure 1).<sup>40</sup> Previous studies have also found that AICAR and ImGP stimulate the glutaminase half reaction, although at a reduction of 45 to 188 times less, respectively, than the PRFAR substrate molecule.<sup>37</sup> Moreover, the two independent products also stimulate the glutaminase half reaction 15 times less than PRFAR. These results suggest that AICAR and ImGP both contribute to allostery, but ImGP may be more involved in the allosteric signaling network. Furthermore, the chemical bond between the AICAR and ImGP moieties is necessary for enhancement of the allosteric state.

We explore extensions of the original dynamical network analysis to ascertain new models for allosteric communication, which are benchmarked against experimental site-directed mutagenesis studies in HisH–HisF. The impact of utilizing different residue contributions (i.e., nodal methods) within the dynamical network analysis framework is investigated using HisH–HisF as a model system. The Floyd Warshall algorithm determines the shortest (optimal) path between two nodes in a network,<sup>41,42</sup> and the Girvan–Newman algorithm,<sup>43</sup> a method without free parameters, generates aggregate nodal clusters, or communities, that optimize intracommunication pathways within the protein. For each nodal method, we determine the optimal signaling pathways. We then compare and investigate the corresponding community structures for the  $\alpha$ -carbon and residue c.o.m. methods. Ultimately, we propose novel allosteric communication pathways for HisH–HisF and show that using the residue c.o.m. provides advantages over other nodal methods within the dynamical network analysis framework.

## METHODS

**System Setup.** The initial coordinates for the (holo) HisH–HisF were obtained as PDB 1GPW (*Thermotoga maritima*) from the Protein Data Bank and included resolved phosphate groups still bound to the active site. Furthermore, the PDB 1GPW included three states of HisH–HisF categorized by chains A–B, C–D, and E–F. In order to prepare an accurate residue environment, PROPKA was used to determine protonation states of titratable residues based on an environment at pH 7.<sup>44</sup> The  $\delta$ -nitrogen was protonated for the HisF domain for residues His84, His209, and His244 and His73, His120, and His141 in HisH. An  $\epsilon$ -nitrogen protonation state was assigned for HisF:His228, and for HisH:His53. Both the  $\delta$ -nitrogen and  $\epsilon$ -nitrogen of the residue HisF:His151 were protonated.

The crystal structure 1OX5 (from *Saccharomyces cerevisiae*) included all PRFAR heavy atoms and was used to build in the PRFAR substrate for 1GPW. Using the VMD plugin MultiSeq,<sup>45</sup> the crystal structure chain A 1OX5 was aligned to 1GPW chains C and D via secondary structures in the HisF domain using the STAMP structural alignment algorithm.<sup>46,47</sup> The result was that the PRFAR substrate aligned correctly to the 1GPW:HisF domain for the holo-state of HisH–HisF. This alignment is consistent with other works.<sup>38,48</sup> The 1GPW:crystal waters, 1OX5:PRFAR metabolite, 1GPW:HisF, and 1GPW:HisH segments were processed by PSFGEN to generate psf and pdb files. The parametrization of the PRFAR substrate followed standards established in the CHARMM protocol<sup>22</sup> and used results from previous works.<sup>49</sup> The formal charge for the protein dimer was  $-8.0e$ , and the formal charge for the primary metabolite, PRFAR, was  $-5.0e$ . The formal charge of



the system was  $-13.0e$  in total. The system was solvated using SOLVATE<sup>50</sup> to add explicit TIP3 water molecules to a box with a water pad of 10 Å around the protein. Thirteen sodium counterions were added with the VMD plugin Autoionize, bringing the overall formal charge of the entire system to zero.

The (apo) HisH–HisF was also derived from 1GPW using chains A and B. The system setup for the (apo) HisH–HisF followed the exact same procedure as (holo) HisH–HisF without the addition of the PRFAR ligand.

**Molecular Dynamics Simulations.** A four-step energy minimization was employed for both apo and holo structures to ensure maximal retention of experimental structural information. The four steps constrained heavy atoms for 1000 fs, heavy atoms excluding water for 1000 fs, backbone atoms excluding all other molecular groups for 2500 fs, and finally a constraint-free motion of all atoms for 5000 fs. Harmonic constraints were induced to allow minimal backbone motion to achieve water equilibration for five time frames of 25 000 fs each and different backbone constraint constants of 1.0, 0.75, 0.50, 0.25, and 0.0 kcal mol<sup>-1</sup> Å.

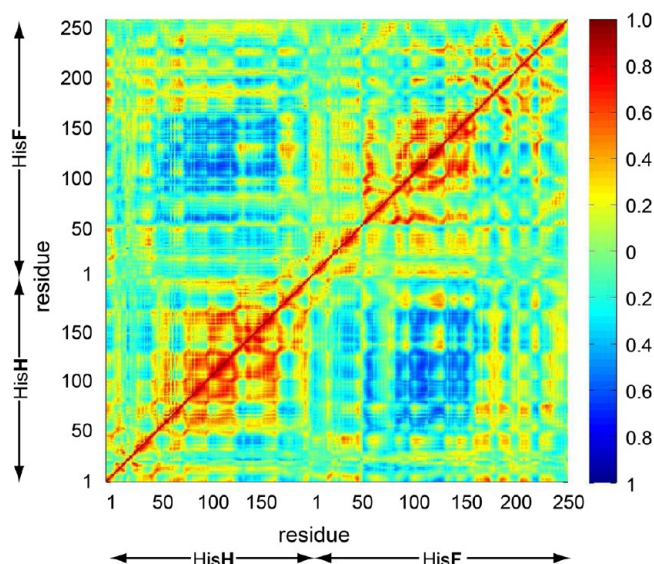
Molecular dynamics for the holo structure were run for 30 ns (including an initial 10 ns to ensure the mutual equilibration of PRFAR and HisH–HisF) with an integration time step of 1 fs using NAMD2.7<sup>21</sup> and the CHARMM27 force field.<sup>21,51</sup> Production molecular dynamics for the (apo) HisH–HisF were run for 20 ns with an initial 1 ns harmonic constraints equilibration phase. All other simulation parameters and benchmarks were the same as the (holo) HisH–HisF simulation. Periodic boundary conditions were used with the NPT ensemble, a pressure of 1 atm using a Langevin piston, and a temperature set to 310 K. A particle mesh Ewald summation was used for the electrostatic approximations.<sup>52</sup> The long-range van der Waals interactions were given a cutoff distance of 12 Å and a switching distance at 10 Å. Simulations were performed on the TACC Ranger cluster using 256 processors. The benchmark for the ~45 000 atom systems were 0.0136156 s/step and 0.157588 days/ns.

Protein snapshots were extracted at 500 fs intervals over the 20 ns production phases of both the holo-state and apo-state, resulting in two sets of 40 000 structures for subsequent analysis. Each snapshot was aligned via RMSD to the average position of all atoms within the protein for both systems over the entire 20 ns simulation of production dynamics, in order to remove any rotational or translational motions.

**Dynamical Network Model.** We explore the use of different sources of correlation information, also known as nodal methods, including (i)  $\alpha$  carbon, and the c.o.m. of (ii) the backbone, (iii) the side chain, and (iv) whole-residue representations (residue c.o.m.). In each case, the covariance matrix is normalized, and as a consequence, the position, and not the mass of the node, is important. In order to model PRFAR as a set of nodes, the PRFAR substrate is broken into two moieties, which represent the reaction end products AICAR and ImGP (Figure 1). The centers of mass for both the AICAR and ImGP moieties are defined as separate nodes for the analysis.

Methods are carried out using a similar protocol to that of refs 29 and 30. A contact map is used to determine Boolean values for node–node interactions. If a residue is within 4.5 Å of another residue for 75% of the 40 000 frame (20 ns) trajectory, then the two residues are considered to be in contact with one another (Supporting Information (SI) Figure 1). Nodes that do not fit these requirements are filtered out. The

correlation values are obtained by trajectory data using the program CARMA.<sup>53</sup> The correlation is defined by eqs 1 and 2, where  $C_{ij}$  is the correlation value between two nodes over a given simulation time (Figure 2). The value of  $C_{ij}$  is unity when



**Figure 2.** Correlation maps. Entries in the lower right triangle show pairwise correlation values using the  $\alpha$ -carbon atoms; the upper left triangle shows the correlation values using the c.o.m. of the entire residue.

$i = j$  and satisfies the condition that a node correlates completely with itself. In eq 2, the position of a given node at time  $t$  is subtracted from the average position of the node over the entire production dynamics (20 ns) and requires that the average position of the node sampled is approximately stationary over time. When comparing the motion of one nodal method to another, eq 1 is also used (Figure 3); these are correlations among different nodal methods within the same residue.

$$C_{ij} = \frac{\langle \Delta \vec{r}_i(t) \cdot \Delta \vec{r}_j(t) \rangle}{(\langle \Delta \vec{r}_i(t)^2 \rangle \langle \Delta \vec{r}_j(t)^2 \rangle)^{1/2}} \quad (1)$$

and

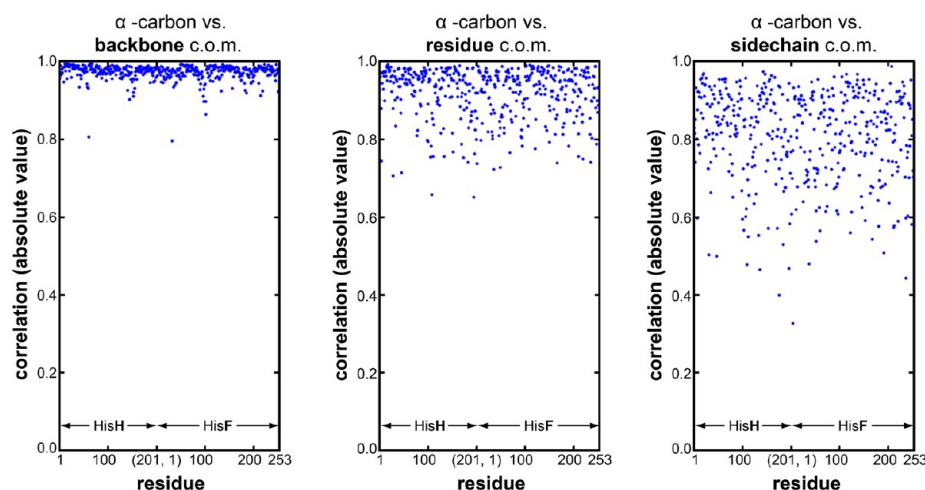
$$\Delta \vec{r}_i(t) = \vec{r}_i(t) - \langle \vec{r}_i(t) \rangle \quad (2)$$

The use of eqs 1 and 2 as well as interaction criteria determined through the contact maps generates a filtered correlation matrix.

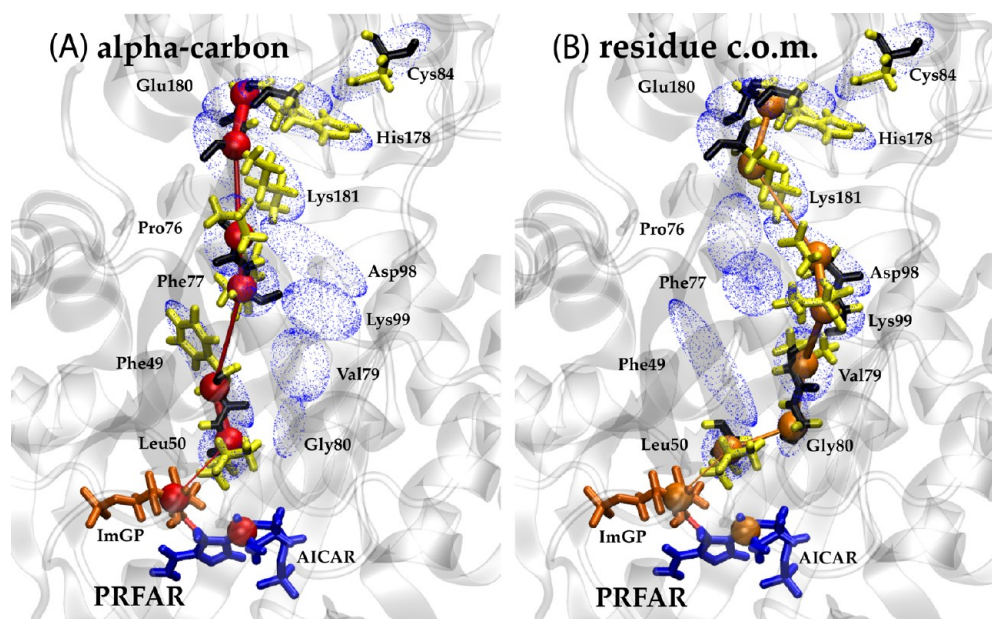
The edge between a pair of nodes is then given a length in network space by functionalizing the correlation values. A positive or negative correlation is viewed equally in that residue interactions are coherent whether they are correlated or anticorrelated, and so the edge lengths take the value  $w_{ij}$  in eq 3 between node  $i$  and node  $j$ :<sup>29,30</sup>

$$w_{ij} = -\log(|C_{ij}|) \quad (3)$$

Although the values of each length have meaning in network diagrams, they can be mapped to the physical protein space by weighting each edge with a thickness or cylindrical diameter. Therefore, thick edges between nodes are indicative of edges close to one another in network space, while thin edges and edges that do not exist are far or infinitely separated,



**Figure 3.** Correlation of different residue components. (A) Here, correlation between the  $\alpha$ -carbon and the backbone c.o.m. is shown for each residue. The  $\alpha$ -carbon is strongly correlated to the backbone c.o.m. and not well correlated to the side chain c.o.m. suggesting that information given by the  $\alpha$ -carbon is highly biased toward backbone motion. (B) Next, correlation between the  $\alpha$ -carbon and the c.o.m. for the entire residue is shown. (C) Lastly, comparisons between the  $\alpha$ -carbon and side chain c.o.m. show the lowest correlation values and greatest spread, which indicates that information from side chain motion may be lost in an  $\alpha$ -carbon representation.



**Figure 4.** Optimal allosteric pathways. (A and B) Here, we show the optimal pathway generated from the Floyd–Warshall algorithm using (A) the  $\alpha$ -carbon method (red spheres) and (B) the residue c.o.m. (orange spheres) method. In both methods, the source is the ImGP moiety, and the sink is catalytic residue HisH:Glu180. The ellipsoids (blue dots) are aligned along the principal axes of a given residue to illustrate a residue's orientation and spatial displacement. Edges are drawn between nodes, and the edge thickness is radially proportional to the weight given by eq 3.

respectively (Figures 4–7). Such edges represent the signaling strength between nodes.

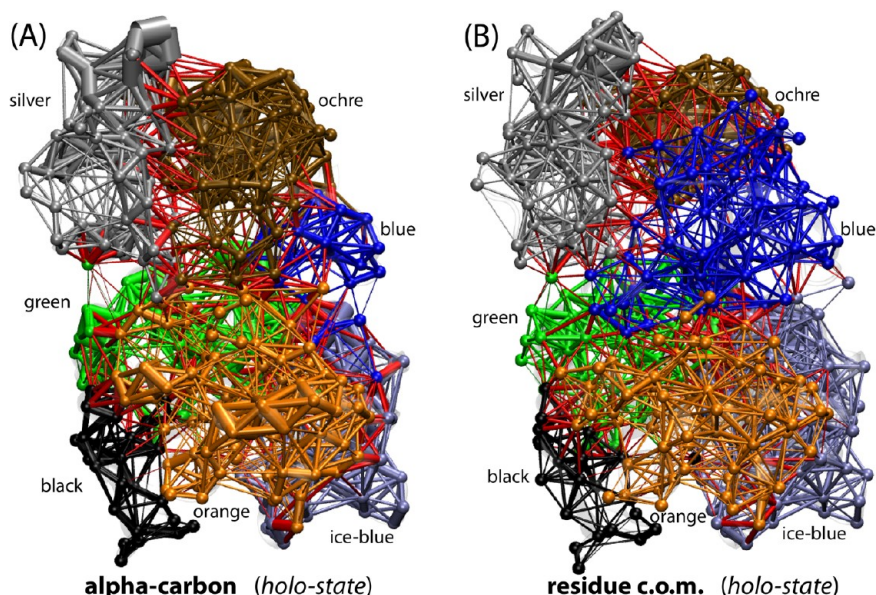
One aim of our analysis is to identify putative allosteric communication pathways in proteins, using HisH–HisF as a model system. Evidence of cooperation between the substrate PRFAR (i.e., the source of allosteric signal) and the catalytic triad (i.e., where the allosteric signal ends) is captured within this framework. In this model, we assume that the most likely or biologically relevant pathways are those that minimize the network distance. Finding a singular optimal path is equivalent to minimizing the distance traveled between nodes in the network. The length of such a path is the sum of the individual path lengths between each node set  $i, j$  (eq 4), starting with the

source nodes in PRFAR and ending with the three sink nodes in the catalytic triad:

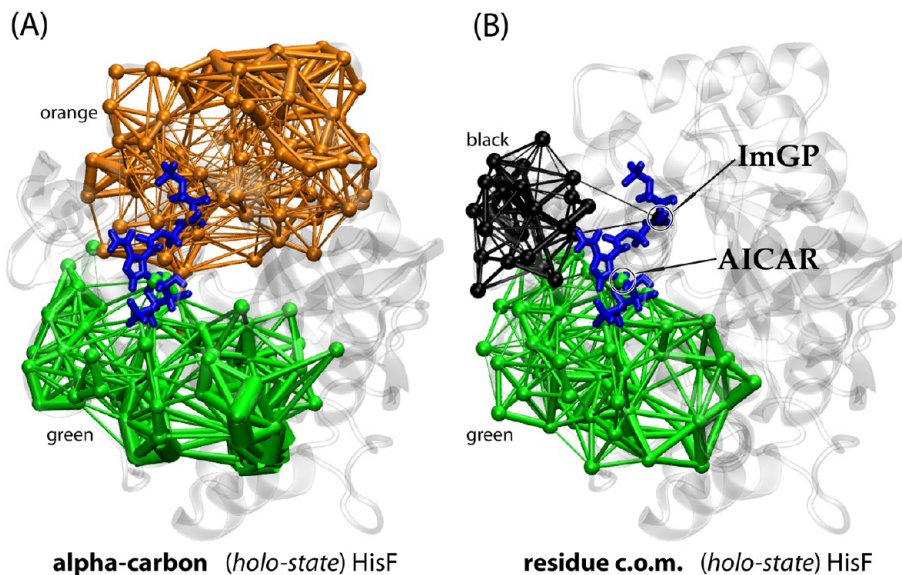
$$D_{ij} = \sum_{k,l} w_{k,l} \quad (4)$$

All other paths are deemed suboptimal if they fall within an acceptable deviation from the optimal path. Such pathways are likely to present highly degenerate signaling along the optimal path. Although it is possible to determine numerous suboptimal pathways within the dynamical network analysis framework, we do not investigate suboptimal pathways in this work. Yet, we acknowledge that their analysis may likely lead to further insight into the complexities of allosteric regulation.





**Figure 5.** Community network analysis. (A and B) The optimal communities for communication among residues are depicted, grouped via the Girvan–Newman algorithm. Note the significant change in community structure between the two methods: residue  $\alpha$ -carbon (A) and residue c.o.m. (B). There are seven communities in the protein dimers: three for HisH (silver, ochre, and blue) and four for HisF (black, orange, green, and ice-blue). The edges that attach critical nodes between communities are shown as red lines and depict high levels of degenerate signaling between community structures. Edge thickness is radially proportional to the weight given by eq 3.



**Figure 6.** Inter-/intracommunity connections of PRFAR. (A and B) The change in community structure for connections between PRFAR and other immediate residues is depicted. The  $\alpha$ -carbon method (A) shows ImGP (blue) as a member of the orange community and AICAR (blue) as a member of the green community. Both the green and orange communities largely include two halves of the  $(\beta/\alpha)_8$  barrel in HisF. The residue c.o.m. method (B) shifts the ImGP moiety (blue) into the black community comprising the loop at the C-terminal end of the  $(\beta/\alpha)_8$  barrel.

**Entropy Calculations.** The entropy for both holo and apo-states was calculated using CARMA,<sup>53</sup> which follows the derivation outlined by Andricioaei and Karplus.<sup>54</sup>

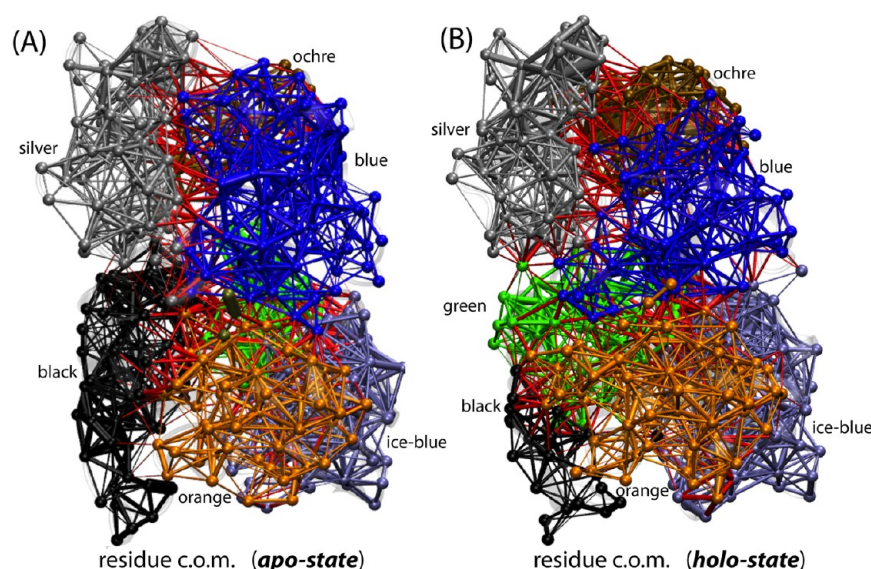
$$S_{ho} = k \sum_i^{3n-6} \frac{\hbar\omega_i/kT}{e^{\hbar\omega_i/kT} - 1} - \ln(1 - e^{-\hbar\omega_i/kT}) \quad (5)$$

The mass weighted covariance matrix was used for the  $\alpha$ -carbons, and each mass was 12.0107 g mol<sup>-1</sup>. A mass weighted covariance matrix was also used for the residue c.o.m. method, and each mass varied depending on the selected residue. The PRFAR molecule was omitted when comparing entropies of the

apo and holo-states in order to maintain identical degrees of freedom between the ligand bound and ligand free systems.

## RESULTS

**MD Simulations.** The MD simulations for HisH–HisF must be sampled over equilibrium conditions in order to use correlations since the average position of a node is required to be approximately stationary over time. To assess each nodal method, 30 ns of MD were run for the holo-state. The RMSD for all 30 ns of molecular dynamics was calculated at each frame for the holo-state and indicates system equilibration after the



**Figure 7.** Comparison of Apo/Holo HisH–HisF for community structure. (A and B) The optimal communities for communication among residues are depicted for both the apo-state and holo-state using the residue c.o.m. method. There are seven communities in the protein dimers for both the apo-state and holo-state: three largely associated with HisH (silver, ochre, and blue) and four associated with HisF (black, orange, green, and ice blue). However, these communities do extend across the dimer interface. The edges that attach critical nodes between communities are shown as red lines. The black community in the apo-state (A) is larger than that in the holo-state (B). The ImGP moiety is a member of the black community for the holo-state, and this may be the direct cause of the community shift from the larger black community associated with the apo-state to the smaller black community in the holo-state.

**Table 1.** Optimal Pathways for Holo ImGP–HisH:Glu180 for Each Nodal Method<sup>a</sup>

ImGP-(HisH:Glu180)		HisF																HisH	
Node Method	time (ns)	ImGP	Leu	Phe	Val	Leu	Ile	Gly	Val	Thr	Phe	Pro	Ile	Asp	Lys	Asp	Lys	Glu	
Side-chain c.o.m.	< 5 >	ImGP	50		48	47							75	74				180	
	< 10 >	ImGP	50	49			73											180	
	< 15 >	ImGP	50		48					78					99	98	181	180	
	< 20 >	ImGP	50		48					78					99	98	181	180	
Residue c.o.m.	< 5 >	ImGP	50					80	79						99	98	181	180	
	< 10 >	ImGP	50					80	79						99	98	181	180	
	< 15 >	ImGP	50					80	79						99	98	181	180	
	< 20 >	ImGP	50					80	79						99	98	181	180	
alpha-carbon	< 5 >	ImGP	50	49	48	47						76					181	180	
	< 10 >	ImGP	50	49						78	77	76					181	180	
	< 15 >	ImGP	50	49						78	77	76					181	180	
	< 20 >	ImGP	50	49							77	76					181	180	
Backbone c.o.m.	< 5 >	ImGP	50	49						78	77	76					181	180	
	< 10 >	ImGP	50	49						78	77	76					181	180	
	< 15 >	ImGP	50	49						78	77						181	180	
	< 20 >	ImGP	50	49						78	77						181	180	

<sup>a</sup>The optimal path from source ImGP to sink HisH:Glu180 is calculated on the basis of 20 ns of production dynamics after an initial 10 ns to allow PRFAR and HisH–HisF to equilibrate. Each entry in the table represents a residue or moiety of PRFAR. The residue name and dimer superset of each residue is indicated above the table. Number entries are associated with the crystal structure 1GPW from the protein data bank. The resulting correlation matrix for each nodal method is calculated at 5 ns, 10 ns, 15 ns, and 20 ns. However, the same contact map is used for each and is based on the 20 ns trajectory. For each nodal method shown, the optimal paths appear to converge to a consistent set of residues. Evolutionarily conserved (100% sequence identity) residues are both underlined and bold, whereas partially conserved (greater than 50% sequence identity) residues are just underlined in reference to previous homology studies.<sup>8</sup>

first 10 ns (SI Figure 2); the remaining 20 ns is considered production dynamics and was utilized in the dynamical network analyses. The RMSD for the 20 ns molecular dynamics of the apo-state also showed convergence. Trajectories were assessed at 5 ns, 10 ns, 15 ns, and 20 ns to determine the time dependence and convergence of the optimal paths.

**Correlation and Contact Maps.** The full correlation matrix, representing both the residue c.o.m. and  $\alpha$  carbon methods, is presented (Figure 2). A contact map indicating which residues were closely interacting was generated over the production MD phase for the holo-state. If residues were within 4.5 Å of another residue for 75% of the 40 000 frame (20 ns) trajectory, the correlation value in the correlation matrix was


kept; otherwise the correlation value was set to zero. This contact-map filtering prevents edges from being drawn between residues that are physically far apart. The same contact map was used for the holo-state across each time assessment at 5 ns, 10 ns, 15 ns, and 20 ns (SI Figure 1).

The contact map for the apo-state was generated from a separate 20 ns trajectory and followed the same criteria as the holo-state. This contact map was used to assess the apo-state over the entire 20 ns trajectory for comparison to the holo-state.

**Correlation of Nodal Methods in Holo HisH–HisF.** The correlation of the different nodal methods was analyzed for the holo-state over 20 ns of production dynamics by determining



Table 2. Optimal Pathways Apo/Holo HisH–HisF Using Residue Center of Mass Method<sup>a</sup>

ImGP–(HisH:Glu180)														distance
HisH–HisF	time (ns)	ImGP	Leu	Phe	Gly	Val	Phe	Pro	Lys	Asp	Lys	Glu		
Holo-residue c.o.m.	< 20 >	ImGP	50		80	79			99	98	181	180	287	
Apo-residue c.o.m.	< 20 >	none	50	49				77	76		181	180	298	

<sup>a</sup>Here, we show the optimal pathway between ImGP and HisH:Glu180 for both the apo-state and holo-state simulation with only the c.o.m. method being used (the preferred method for this paper). The apo-state and holo-state of HisH–HisF is compared here and shows a clear switch between both states. The holo-state traces out the salt bridge HisF:Asp98–HisH:Lys181 that is verified experimentally to be part of the allosteric mechanism.<sup>8</sup> The apo-state does not provide this bridge. The distance values on the right are calculated according to eq 4.

the positional correlation between the  $\alpha$ -carbon and the c.o.m. of the (a) backbone, (b) side chain, and (c) entire residue. This analysis indicates that the motion exhibited by the  $\alpha$ -carbon is most highly correlated to the backbone motion as a general trend for same residue selections (Figure 3). The position of the  $\alpha$ -carbon correlates strongest with the position of the backbone c.o.m. and weakest with the position of the c.o.m. of the side chain. Between these two extremes, the position of the residue c.o.m., which represents both the backbone and side chain atoms, exhibits moderate correlation to the  $\alpha$  carbon position.

**Optimal Pathways. Comparison of Different Nodal Methods within the Holo-State.** The optimal communication path is the shortest path between the source of the allosteric signal, in this case PRFAR, and the end-point of the allosteric signal, or sink, which is the catalytic triad (Figure 4). The catalytic triad is comprised of three residues HisH:Glu180, HisH:His178, and HisH:Cys84 and therefore three potential sink candidates. Each residue within the triad was thus investigated in the search for the optimal paths. For simplicity, we focus on the sink residue HisH:Glu180 (Table 1). However, we note differences in the optimal pathways among all combinations of sources and sinks and provide this additional information (SI Table 1).

The optimal path between source node ImGP and sink node Glu180 captured the same residues at 15 and 20 ns for the different nodal methods, and the optimal path for the residue c.o.m. method was surprisingly the same for all time assessments (Table 1). The pathway between the ImGP moiety and residue Glu180 yielded similar results to the pathway between the AICAR moiety and residue Glu180 when using the  $\alpha$ -carbon nodal method (SI Table 1).

The side chain c.o.m. method gives an optimal path of ImGP, HisF:Leu50, HisF:Val48, HisF:Thr78, HisF:Lys99, HisF:Asp98, HisH:Lys181, and HisH:Glu180. The residue c.o.m. method gives an optimal path of ImGP, HisF:Leu50, HisF:Gly80, HisF:Val79, HisF:Lys99, HisF:Asp98, HisH:Lys181, and HisH:Glu180. In contrast, the pathway predicted by the  $\alpha$ -carbon method suggests an optimal pathway connecting the ImGP source through HisF:Leu50, HisF:Phe49, HisF:Phe77, HisF:Pro76, HisF:Lys181, and HisH:Glu180 (sink). The optimal pathway predicted by the backbone c.o.m. method traces through ImGP (source), HisF:Leu50, HisF:Phe49, HisF:Thr78, HisF:Phe77, HisF:Lys181, and HisH:Glu180 (sink).

**Comparison of Holo- and Apo-States Using the Residue Center of Mass Method.** Both the holo-state, with allosteric effector PRFAR bound, and the apo-state simulations sample equilibrium conformations of the protein and provide a comparison between the ligand activated holo-state and the basal (inactive) apo-state. The optimal path for the holo-state is compared to the optimal path of the apo-state from ImGP

(source) to HisH:Glu180 (sink) using the residue c.o.m. as the preferred nodal method (Table 2). The residue c.o.m. method traces an optimal path for the holo-state as ImGP, HisF:Leu50, HisF:Gly80, HisF:Val79, HisF:Lys99, HisF:Asp98, HisH:Lys181, and HisH:Glu180. In contrast, the apo-state gives an optimal path of HisF:Leu50, HisF:Phe49, HisF:Phe77, HisF:Pro76, HisH:Lys181, and HisH:Glu180. A noteworthy switch between the two states is the HisF:Leu50–HisF:Gly80 in the holo-state versus the HisF:Leu50–HisF:Phe49. The apo-state HisF:Leu50–HisF:Phe49 is clearly a backbone dominated correlation where the holo-state HisF:Leu50–HisF:Gly80 suggests a correlation that may depend on side chain interaction. Another important difference is seen at the interface of the protein dimer, where the apo-state HisH:181Lys–HisF:Pro76 communication path switches to HisH:181Lys–HisF:Asp98 in the holo-state. The HisF:Asp98 residue is a component of the salt bridge verified by experimental results to be a contributing residue in the allosteric mechanism of HisH–HisF.<sup>8,55</sup> Although the optimal apo-state path has fewer nodes than the holo-state path, in part due to the absence of ImGP, it has a longer overall distance. This indicates that the coupled residue–residue motions involved in molecular signaling between the effector site and the active site are stronger when PRFAR is bound.

**Community Structure Analysis. Comparison between Different Nodal Methods (Holo-State).** Community structures depict high regional correlations among nodes within the protein, and these regions are modeled as node aggregates or communities. The community subsets may give insights to allosteric signaling, and in particular, node pairs that perform intercommunity signaling, so-called “critical nodes”, provide possible sites for signal disruption. Although similarities between the  $\alpha$ -carbon method and residue c.o.m. were exhibited in the community structures, the residue c.o.m. method indicated a significant difference regarding the community with which ImGP was associated. Furthermore, certain communities were drastically altered (blue in Figure 5) between the two methods.

The  $\alpha$ -carbon method indicated an intercommunity connection between AICAR and ImGP, and the AICAR moiety was associated with a portion of the  $(\beta/\alpha)_8$  barrel (green in Figure 6A). The  $\alpha$  carbon method suggested AICAR and ImGP to be constituents of two separate communities (orange and green in Figure 6A) and composed of proportional parts of the  $(\beta/\alpha)_8$  barrel. In the  $\alpha$ -carbon method, an intracommunity connection was exhibited between ImGP and HisF:Cys9, HisF:Gly202, HisF:Gly203, HisF:Leu222, HisF:Ala223, HisF:Ala224, and HisF:Ser225. Intracommunity connections were also detected between AICAR and HisF:Gly82, and HisF:Gly81, which was similar to the c.o.m. method.

In the residue c.o.m. method, the ImGP moiety is predicted to belong to the nodal community comprised of a loop near the



Table 3. Critical Nodes in Apo/Holo HisH–HisF Using Residue Center of Mass Method<sup>a</sup>

		HisF Critical Nodes																
		Arg	Ile	Val	Gly	Pro	Leu	Ser	Asp	Glu	Val	Ala	Glu	Ile	Pro	Asp	Lys	Ile
Holo		5		17	30	32	35			46	69		71			99	113	116
Apo		5	7					40	45	46		70		75	76	98	99	
		Gln	Val	Val	Trp	Glu	Asp	Thr	Ser	Leu	Ala	Leu	Val	Asn	Arg	Leu	Glu	Gly
Holo		123	126	127		167	176								249			
Apo			127	156				178	180	211	213	222	246	247		250	251	252

		HisH Critical Nodes																
		Arg	Val	Gly	Met	Leu	Ile	Ser	Glu	Arg	Trp	Tyr	Tyr	Tyr	His	Lys	Lys	Ile
Holo		18	51		58		67	94		117	123	138	143	158				189
Apo		18	51	57		61			96			138	143		178	181	184	185
																		189
																		193
																		none

<sup>a</sup>The critical nodes in a community structure present high signal degeneracy between community aggregates. Here, we compare the critical nodes calculated for the apo-state and holo-state. The preservation of critical nodes across both states (grey columns) may be important to protein structure and function. Residues that are evolutionarily conserved are both bold and underlined, while partial conservation is indicated by just an underline. Residues that are not evolutionarily conserved are written in the column headers as plain text.

C-terminus (black in Figure 6B) with intracommunity connections to HisF:Lys19 and HisF:Asp11. The AICAR moiety for the c.o.m. method belonged to the nodal community composed of a portion of the  $(\beta/\alpha)_8$  barrel (green in Figure 6B) with intracommunity connections to HisF:Gly82, HisF:Gly81, and HisF:Leu50. Although the  $\alpha$ -carbon and residue c.o.m. methods differ in important details of the community partitioning, both show a split between the ImGP and AICAR moieties of PRFAR in agreement with the catalytic mechanism of HisF.

**Community Structure Comparison between the Apo- and Holo-States.** The residue c.o.m. method was used to derive community structures for both the apo and holo conformational ensembles. Considerable shifts in community aggregates are seen in the HisH domain (Figure 7A and B, blue regions). The set of community aggregates also rearranged themselves in the HisF domain between apo and holo states (Figure 7A and B, black regions). Both the apo- and holo-states provided seven community aggregates: three community aggregates in the HisH domain and four in HisF. However, differences in community crossover at the dimer interface are present.

The black community in both the apo- and holo-states contains the flexible loop at the C-terminal end of the  $(\beta/\alpha)_8$  barrel. The black community is clearly altered between the apo- and holo-states (Figure 7). The residues that both states have in common are HisF:Asp11 thru HisF:Glu34 and residue HisF:Ile52. Root mean squared deviation (RMSD) analysis was used to quantify the structural differences for this set of common residues between both states over the course of the trajectories (SI Figures 3 and 4). The average RMSD for the holo-state is 2.166 Å with a standard deviation of 0.412 Å and for the apo-state is 1.416 Å with a standard deviation of 0.322 Å. The addition of PRFAR at the effector site (holo-state) increased the RMSD as well as the fluctuations about the average protein structure of the loop. This corresponds to a decrease in correlations within the black community and provides a distinct shift to the holo-state as seen by the extension of the green community into regions occupied by the black community in the apo-state (Figure 7A and B).

**Critical Node Comparison between the Apo- and Holo-States.** Critical nodes represent strong signaling conduits between communities. Critical nodes were analyzed for both the apo-state and holo-state simulations (Table 3). Four critical nodes were the same across both the apo-state and holo-state in HisF. These nodes were HisF:Arg5, HisF:Glu46, HisF:Lys99, and HisF:Val127. All of these nodes are evolutionarily

conserved, except for HisF:Val127, which is partially conserved.<sup>8</sup> For the HisH domain, six critical nodes were the same. The evolutionarily conserved node is HisH:Tyr138. The partially conserved nodes are HisH:Val51, HisH:Tyr143, and HisH:Val193.<sup>8</sup> The two nodes that did not show conservation are HisH:Arg18 and HisH:Leu189.

#### Entropy Calculations between Apo and Holo-States.

The entropies of the apo- and holo-states were calculated for both the  $\alpha$ -carbon and c.o.m. methods using a quasiharmonic approximation.<sup>54</sup> The  $\alpha$ -carbon method provided entropy values of 13192.54 J mol<sup>-1</sup> K<sup>-1</sup> and 13664.46 J mol<sup>-1</sup> K<sup>-1</sup> for the respective apo- and holo-states. Whereas the residue c.o.m. method provided entropy values of 27540.41 J mol<sup>-1</sup> K<sup>-1</sup> and 27905.33 J mol<sup>-1</sup> K<sup>-1</sup>. These results show a positive entropy gain, upon PRFAR binding, with values  $\Delta S_{ac} = 471.92$  J mol<sup>-1</sup> K<sup>-1</sup> and  $\Delta S_{com} = 364.92$  J mol<sup>-1</sup> K<sup>-1</sup> for the  $\alpha$ -carbon and c.o.m. methods, respectively.

## DISCUSSION

In this work, we have examined four types of networks by varying node representations in Cartesian coordinates: the (i) residue  $\alpha$ -carbons, (ii) the side chain c.o.m., (iii) the backbone c.o.m., and the entire (iv) residue c.o.m. We applied these to imidazole glycerol phosphate synthase (IGPS) in order to test for allosteric signaling pathways within the framework of dynamical network analysis. The analysis was carried out at equilibrium conditions, and the resulting pathways were sensitive to which node representation was used.

In our results, the  $\alpha$ -carbon method failed to indicate interplay between HisH:Lys181 and HisF:Asp98 for the holo-state optimal path ImGP to HisH:Glu180, whereas the residue c.o.m. and side chain c.o.m. methods succeeded in showing this particular interaction at the interface (Table 1). Detecting HisH:Lys181 and HisF:Asp98 is a promising feature of the residue c.o.m. methodology because this salt bridge has been experimentally validated to participate in allostery for the cyclase half reaction.<sup>8,55</sup> Furthermore, HisH:Lys181 and HisF:Asp98 are coupled to the active site of the catalytic triad through close physical proximity and strong correlations. The holo-state of HisH:Lys181 and HisF:Asp98 forms a long-lasting contact across the dimer interface, confirmed by our simulations (data not shown), and both residues are directly adjacent to the glutaminase active site.

Mutations of HisF:Thr104 and HisF:Lys19, assessed in previous works,<sup>8,55</sup> affected the glutaminase half reaction and may be attributed to destroying the functionality of the receptor

site where PRFAR binds. In our simulation, the residue HisF:Lys19 is adjacent to the ImGP moiety of PRFAR and the Girvan–Newman algorithm clusters both into the same community when based on the residue c.o.m. methodology (black, Figure 7). Trajectories are distinct between the apo-state, where HisF:Lys19 is not linked to PRFAR, and the holo-state, where HisF:Lys19 is directly linked to the ImGP moiety of PRFAR and indirectly to the  $(\beta/\alpha)_8$  barrel (green and orange, Figure 7). When analyzing the RMSD for both the apo-state and holo-state (SI Figures 3 and 4), the variations are more stationary for the apo-state and less so for the holo-state over the 20 ns trajectory. The holo-state may comprise specific harmonic and anharmonic motions between the loop structure and components of the  $(\beta/\alpha)_8$  barrel that is mediated by the ImGP moiety of PRFAR. This would imply that HisF:Lys19 plays a role, for the holo-state, via signaling attributes of the loop structure by channeling motion through a direct connection into the optimal path from ImGP to HisH:Glu180. The apo- and holo-state RMSD time series are consistent with previous NMR studies for this loop structure and may indicate signaling attributes enveloped on the millisecond time scale.<sup>56</sup>

Comparison of the different nodal methods indicates that the  $\alpha$ -carbon and backbone c.o.m. methods are strongly correlated (Figure 3). This analysis suggests a clear bias of the  $\alpha$ -carbon to detect the same signals as the residue backbone.

Moreover, using the covariance matrix to calculate quasiharmonic entropy values gives notably different results when using the  $\alpha$ -carbon versus the residue c.o.m. method. Collectively, these results suggest that biologically relevant information may be lost if side chain substituents are not taken into account and emphasize the importance of accounting for side chain dynamics in allosteric signaling captured through the use of the residue c.o.m. Including side chain information in dynamical network models may be essential to providing an accurate analysis of allosteric pathways as well as hierarchical clustering of dynamical networks.

#### Evolved Interdependence of the Two Domains.

Interdependence between HisH and HisF exists as a result of protein function. Elegant biochemical kinetic experiments for HisH–HisF have decoupled the two enzyme reactions *in vitro*. Constructs in bacteria, where the two domains are encoded on separate polypeptide chains, have shown that the cyclase reaction can occur via HisF under nonbiological conditions (50 mM Tris-acetate buffer, pH 8.5, at 25 °C), whereas the glutaminase reaction of isolated HisH has no detectable activity.<sup>57</sup> Consequently, residues that form a foundation for the dependence of the two domains may be highly conserved due to evolutionary pressures. Some of these residues are likely along the signaling pathway.

Every residue along the optimal path determined with our residue c.o.m. analysis, from ImGP to HisH:Glu180, is known to be evolutionarily conserved.<sup>8</sup> This result is in contrast to the  $\alpha$ -carbon method, which also included a nonconserved residue HisF:Thr78 in its optimal path (Table 1). Furthermore, many residues along the  $(\beta/\alpha)_8$  barrel were highly selected for in both the apo-state and holo-state along the optimal paths for all methods. This result is consistent with the conserved nature of the barrel structure across the sub-family of GAT enzymes and is a quantitative argument against possible circuitous routes via other secondary structural elements of the protein. The pathway comparison between the apo- and holo-states (Table 2) may also provide insights to basal turnover as HisH–HisF

produces low levels of glutamate in the half reaction when PRFAR is not bound.

#### Community Structure as an Allosteric Descriptor.

Overall community structure is a function of intracommunication among a set of residues and consequently provides a coarse grained intercommunication between a set of communities. The community results are derived from different nodal methods and indicate that the choice used to determine the correlation matrices is important (Figures 5–7). When using the residue c.o.m. method, a clear shift among community structures is seen between the apo- and holo-states. The loop structure at the base of the  $(\beta/\alpha)_8$  barrel directly influences the ImGP moiety of PRFAR as previously mentioned. However, other shifts in community structure may be more relevant. At the dimer interface, clear changes exist between the apo- and holo-state (Figure 7).

Critical nodes at the interface of the communities may be adequate descriptors of a protein's state. Even though critical nodes may not be part of the optimal pathway, they are assumed to be important to the configurational ensemble. For example, a critical node may represent a pharmaceutically relevant site for allosteric inhibition. The critical nodes between apo- and holo-states also shift (Table 3). However, some residues preserve this critical node quality between both apo- and holo-states. For instance, HisF:Lys99 is evolutionarily conserved and quantified as critical for both states. Whereas, HisF:Asp98 and HisH:Lys181 are also conserved but are only critical nodes for the apo-state. The allosteric importance of HisF:Asp98 and HisH:Lys181 has already been established, and their split into separate communities is a feature of the apo-state that quantifies the separation and extent of influence between the two residues.

From the community analysis, ImGP is also considered a critical node and carries information from the black community to both the green and orange communities, which are commensurate halves of the  $(\beta/\alpha)_8$  barrel. Furthermore, AICAR is also a critical node, and this reflects the flexible characteristics of the PRFAR metabolite as both moieties correlate with different communities (Figure 6B).

**Limitations of the Model.** The optimal pathways for allosteric signaling were based on a contact-map-dependent methodology defined previously.<sup>29,30</sup> Residues within a particular distance of another residue for some percentage of the simulation are assumed to influence the communication pathway directly, and residues that do not satisfy these constraints are removed from consideration. However, other methods for generating a contact map may improve definitive connections within the protein network. For example, a reasonable alternative contact map may use values that are not Boolean, instead relying on a continuous function that varies depending on the distance between nodes. Alternate methods for defining contacts among residues is beyond the scope of this paper, though we note such efforts may impact results and should be considered for future dynamical network models.

The available high-resolution crystal structures of the glutaminase domain have shown that the oxyanion hole required for glutaminase activity is not properly formed in the absence of the cyclase domain or the effector molecule.<sup>31,58–60</sup> In fact, the rearrangement of the active site to a catalytically competent conformation is a hallmark feature of the regulation of nitrogen metabolism across the entire family of glutamine amidotransferases. The energy barrier

required to overcome this oxyanion strand rearrangement in HisH–HisF has been estimated at over 20 kcal/mol.<sup>56</sup> Notably, these crystal structures, as well as the simulations presented here, did not contain the glutamine substrate at the active site of HisH. Our work assesses the effects of PRFAR alone. Further studies are warranted in order to determine whether the inclusion of glutamine within the active site of HisH–HisF, either covalently bound to the active site, cysteine, or simply present locally, would affect the networks and community structures determined from the PRFAR-bound state.

Using a network approach to allostery is both intuitive and instructive for garnering information as to what residues may play crucial roles in an allosteric mechanism. However, there are limitations on how our correlation metrics may provide insight. Not all correlations are taken into account, as there may be correlations that exist but are computationally orthogonal to one another due to bond constraints. Furthermore, the covariance matrix assumes fluctuations about an average structure, but the structural average may not be adequate for anharmonic type motions.

## CONCLUSIONS

In this work, two of five residues (HisH:Lys181 and HisF:Asp98), experimentally found to facilitate allostery in HisH–HisF, were determined to be along the optimal pathway of our network model when using correlation information determined from the residue c.o.m. method. The other three residues appear to be related to the disruption of the allosteric state, for which we find strong qualitative and quantitative arguments that HisF:Lys19 is a crucial link between ImGP and the loop structure at the base of the  $(\beta/\alpha)_8$  barrel. This loop structure is distinctly different for the apo-state and holo-state and is arguably a feature of the allostery exhibited in HisH–HisF. The  $\alpha$ -carbon method was clearly similar to the backbone c.o.m. method and traced out an optimal pathway that did not emphasize the dimer side chain interactions for ImGP (source) through HisH:Glu180 (sink). Using the covariance matrix to calculate quasiharmonic entropy values gives notably different results when using the  $\alpha$ -carbon versus the residue c.o.m. method. In principal, quasiharmonic vibrational energies are also dependent on this choice and may lead to inadequate values for properties of the system, such as heat capacity. As the residue c.o.m. is a composite set of the backbone and side chain groups, our works suggest that this method provides superior information in contrast to the popular  $\alpha$ -carbon approach.

## ASSOCIATED CONTENT

### Supporting Information

Four figures and two tables. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## AUTHOR INFORMATION

### Corresponding Author

\*Tel.: 858-534-9629. E-mail: [ramaro@ucsd.edu](mailto:ramaro@ucsd.edu).

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

This work was funded in part by the National Institutes of Health through the NIH Director's New Innovator Award Program 1-DP2-OD007237 and through the NSF XSEDE Supercomputer resources grant LRAC CHE060073N to REA

Simulations were run at the Texas Advanced Computing Center. This work was also supported by the NIH Resource for Macromolecular Modeling and Bioinformatics P41-RR05969 and NSF grant MCB-0844670 to ZLS We would also like to thank Dr. Robert Swift for his insight and consultation pertaining to general topics of molecular dynamics.

## REFERENCES

- (1) Changeux, J. P. 50th anniversary of the word "allosteric". *Protein Sci.* **2011**, *20*, 1119–24.
- (2) Monod, J.; Changeux, J. P.; Jacob, F. Allosteric proteins and cellular control systems. *J. Mol. Biol.* **1963**, *6*, 306–29.
- (3) Monod, J.; Wyman, J.; Changeux, J. P. On the Nature of Allosteric Transitions: A Plausible Model. *J. Mol. Biol.* **1965**, *12*, 88–118.
- (4) Rubin, M. M.; Changeux, J. P. On the nature of allosteric transitions: implications of non-exclusive ligand binding. *J. Mol. Biol.* **1966**, *21*, 265–74.
- (5) Suel, G. M.; Lockless, S. W.; Wall, M. A.; Ranganathan, R. Evolutionarily conserved networks of residues mediate allosteric communication in proteins. *Nat. Struct. Biol.* **2003**, *10*, 59–69.
- (6) Chen, Y.; Reilly, K.; Chang, Y. Evolutionarily conserved allosteric network in the Cys loop family of ligand-gated ion channels revealed by statistical covariance analyses. *J. Biol. Chem.* **2006**, *281*, 18184–92.
- (7) 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*, 1–12.
- (8) 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*, 2156–73.
- (9) Lockless, S. W.; Ranganathan, R. Evolutionarily conserved pathways of energetic connectivity in protein families. *Science* **1999**, *286*, 295–9.
- (10) Koshland, D. E., Jr.; Nemethy, G.; Filmer, D. Comparison of experimental binding data and theoretical models in proteins containing subunits. *Biochemistry* **1966**, *5*, 365–85.
- (11) Changeux, J. P.; Edelstein, S. J. Allosteric mechanisms of signal transduction. *Science* **2005**, *308*, 1424–8.
- (12) Tsai, C. J.; del Sol, A.; Nussinov, R. Allostery: absence of a change in shape does not imply that allostery is not at play. *J. Mol. Biol.* **2008**, *378*, 1–11.
- (13) Goodey, N. M.; Benkovic, S. J. Allosteric regulation and catalysis emerge via a common route. *Nat. Chem. Biol.* **2008**, *4*, 474–82.
- (14) Chennubhotla, C.; Yang, Z.; Bahar, I. Coupling between global dynamics and signal transduction pathways: a mechanism of allostery for chaperonin GroEL. *Mol. Biosyst.* **2008**, *4*, 287–92.
- (15) Cooper, A.; Dryden, D. T. F. Allostery without Conformational Change - a Plausible Model. *Eur. Biophys. J. Biophys. Lett.* **1984**, *11*, 103–109.
- (16) Arkin, M. R.; Wells, J. A. Small-molecule inhibitors of protein-protein interactions: progressing towards the dream. *Nat. Rev. Drug Discovery* **2004**, *3*, 301–17.
- (17) Lipari, G.; Szabo, A. Model-Free Approach to the Interpretation of Nuclear Magnetic-Resonance Relaxation in Macromolecules. 1. Theory and Range of Validity. *J. Am. Chem. Soc.* **1982**, *104*, 4546–4559.
- (18) Lipari, G.; Szabo, A. Model-Free Approach to the Interpretation of Nuclear Magnetic-Resonance Relaxation in Macromolecules. 2. Analysis of Experimental Results. *J. Am. Chem. Soc.* **1982**, *104*, 4559–4570.
- (19) Sapienza, P. J.; Lee, A. L. Using NMR to study fast dynamics in proteins: methods and applications. *Curr. Opin. Pharmacol.* **2010**, *10*, 723–30.
- (20) Yang, D.; Kay, L. E. Contributions to conformational entropy arising from bond vector fluctuations measured from NMR-derived order parameters: application to protein folding. *J. Mol. Biol.* **1996**, *263*, 369–82.



- (21) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781–802.
- (22) Brooks, B. R.; Brooks, C. L., 3rd; Mackerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: the biomolecular simulation program. *J. Comput. Chem.* **2009**, *30*, 1545–614.
- (23) Horstink, L. M.; Abseher, R.; Nilges, M.; Hilbers, C. W. Functionally important correlated motions in the single-stranded DNA-binding protein encoded by filamentous phage Pf3. *J. Mol. Biol.* **1999**, *287*, 569–77.
- (24) Watney, J. B.; Hammes-Schiffer, S. Comparison of coupled motions in *Escherichia coli* and *Bacillus subtilis* dihydrofolate reductase. *J. Phys. Chem. B* **2006**, *110*, 10130–8.
- (25) Bradley, M. J.; Chivers, P. T.; Baker, N. A. Molecular dynamics simulation of the *Escherichia coli* NikR protein: equilibrium conformational fluctuations reveal interdomain allosteric communication pathways. *J. Mol. Biol.* **2008**, *378*, 1155–73.
- (26) Lange, O. F.; Grubmüller, H. Generalized correlation for biomolecular dynamics. *Proteins* **2006**, *62*, 1053–61.
- (27) Killian, B. J.; Yundenfreund Kravitz, J.; Gilson, M. K. Extraction of configurational entropy from molecular simulations via an expansion approximation. *J. Chem. Phys.* **2007**, *127*, 024107.
- (28) McClelland, C. L.; Friedland, G.; Mobley, D. L.; Amirkhani, H.; Jacobson, M. P. Quantifying Correlations Between Allosteric Sites in Thermodynamic Ensembles. *J. Chem. Theory Comput.* **2009**, *5*, 2486–2502.
- (29) 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*, 6620–5.
- (30) Eargle, J.; Li, L.; Luthey-Schulten, Z. Dynamical Network Analysis, 2012, <http://www.scs.illinois.edu/schulten/tutorials/network/>.
- (31) 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*, 987–97.
- (32) Chaparian, M. G.; Evans, D. R. The catalytic mechanism of the amidotransferase domain of the Syrian hamster multifunctional protein CAD. Evidence for a CAD-glutamyl covalent intermediate in the formation of carbamyl phosphate. *J. Biol. Chem.* **1991**, *266*, 3387–95.
- (33) Roux, B.; Walsh, C. T. p-aminobenzoate synthesis in *Escherichia coli*: kinetic and mechanistic characterization of the amidotransferase PabA. *Biochemistry* **1992**, *31*, 6904–10.
- (34) Nakamura, J.; Straub, K.; Wu, J.; Lou, L. The glutamine hydrolysis function of human GMP synthetase. Identification of an essential active site cysteine. *J. Biol. Chem.* **1995**, *270*, 23450–5.
- (35) Willemoes, M. Thr-431 and Arg-433 are part of a conserved sequence motif of the glutamine amidotransferase domain of CTP synthases and are involved in GTP activation of the *Lactococcus lactis* enzyme. *J. Biol. Chem.* **2003**, *278*, 9407–11.
- (36) Lang, D.; Thoma, R.; Henn-Sax, M.; Sterner, R.; Wilmanns, M. Structural evidence for evolution of the  $\beta/\alpha$  barrel scaffold by gene duplication and fusion. *Science* **2000**, *289*, 1546–50.
- (37) 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*, 7013–22.
- (38) Amaro, R.; Tajkhorshid, E.; Luthey-Schulten, Z. Developing an energy landscape for the novel function of a  $(\beta/\alpha)_8$  barrel: ammonia conduction through HisF. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 7599–604.
- (39) Alifano, P.; Fani, R.; Lio, P.; Lazcano, A.; Bazzicalupo, M.; Carlomagno, M. S.; Bruni, C. B. Histidine biosynthetic pathway and genes: structure, regulation, and evolution. *Microbiol. Rev.* **1996**, *60*, 44–69.
- (40) 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 glycerol phosphate synthase: crystal structures of a ternary complex and the free enzyme. *Biochemistry* **2003**, *42*, 7003–12.
- (41) Marshall, S., A Theorem on Boolean Matrices. *J. Assoc. Comput. Mach.* **1962**, *9*, 11.
- (42) Floyd, R. W. Algorithm-97 - Shortest Path. *Commun. ACM* **1962**, *5*, 345–345.
- (43) Girvan, M.; Newman, M. E. Community structure in social and biological networks. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 7821–6.
- (44) Michal, R.; Mats, O. Graphical analysis of pH-dependent properties of proteins predicted using PROPKA. *BMC Struct. Biol.* **2011**.
- (45) Roberts, E.; Eargle, J.; Wright, D.; Luthey-Schulten, Z. MultiSeq: unifying sequence and structure data for evolutionary analysis. *BMC Bioinf.* **2006**, *7*, 382.
- (46) Russell, R. B.; Barton, G. J. Multiple Protein-Sequence Alignment from Tertiary Structure Comparison - Assignment of Global and Residue Confidence Levels. *Proteins: Struct., Funct., Genet.* **1992**, *14*, 309–323.
- (47) Russell, R. B.; Barton, G. J. The Limits of Protein Secondary Structure Prediction Accuracy from Multiple Sequence Alignment. *J. Mol. Biol.* **1993**, *234*, 951–957.
- (48) Amaro, R. E.; Myers, R. S.; Davisson, V. J.; Luthey-Schulten, Z. A. Structural elements in IGP synthase exclude water to optimize ammonia transfer. *Biophys. J.* **2005**, *89*, 475–87.
- (49) Amaro, R.; Luthey-Schulten, Z. Molecular dynamics simulations of substrate channeling through an  $\alpha$ - $\beta$  barrel protein. *Chem. Phys.* **2004**, *307*, 147–155.
- (50) Grubmüller, H. *Solvate V 1.0*; Theoretical Biophysics Group, Institute for Medical Optics, Ludwig-Maximilians University: Munich, 1996.
- (51) Bjelmar, P. r.; Larsson, P.; Cuendet, M. A.; Hess, B.; Lindahl, E. Implementation of the CHARMM Force Field in GROMACS: Analysis of Protein Stability Effects from Correction Maps, Virtual Interaction Sites, and Water Models. *J. Chem. Theory Comput.* **2010**, *6*, 459–466.
- (52) 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.
- (53) Glykos, N. M. Software news and updates. Carma: a molecular dynamics analysis program. *J. Comput. Chem.* **2006**, *27*, 1765–8.
- (54) Andricioaei, I.; Karplus, M. On the calculation of entropy from covariance matrices of the atomic fluctuations. *J. Chem. Phys.* **2001**, *115*, 6289–6292.
- (55) 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*, 11974–85.
- (56) Lipchock, J. M.; Loria, J. P. Nanometer Propagation of Millisecond Motions in V-Type Allostery. *Structure* **2010**, *18*, 1596–1607.
- (57) Beismann-Driemeyer, S.; Sterner, R. Imidazole glycerol phosphate synthase from *Thermotoga maritima*. Quaternary structure, steady-state kinetics, and reaction mechanism of the bienzyme complex. *J. Biol. Chem.* **2001**, *276*, 20387–96.
- (58) 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.
- (59) Omi, R.; Mizuguchi, H.; Goto, M.; Miyahara, I.; Hayashi, H.; Kagamiyama, H.; Hirotsu, K. Structure of imidazole glycerol phosphate synthase from *Thermus thermophilus* HB8: Open-closed conformational change and ammonia tunneling. *J. Biochem.* **2002**, *132*, 759–765.
- (60) Korolev, S.; Skarina, T.; Evdokimova, E.; Beasley, S.; Edwards, A.; Joachimiak, A.; Savchenko, A. Crystal structure of glutamine

amidotransferase from *Thermotoga maratama*. *Proteins: Struct., Funct., Genet.* **2002**, 49, 420–422.