Biophysical Journal Volume 99 October 2010 2245–2254 

# Allosteric Response Is both Conserved and Variable across Three CheY Orthologs

James M. Mottonen,[†] Donald J. Jacobs,[†*] and Dennis R. Livesay[‡*]
[†]Department of Physics and Optical Science, and [‡]Department of Bioinformatics and Genomics, University of North Carolina at Charlotte, Charlotte, North Carolina

ABSTRACT A computational method to identify residues likely to initiate allosteric signals has been developed. The method is based on differences within stability and flexibility profiles between wild-type and perturbed structures as computed by a distance constraint model. Application of the approach to three bacterial chemotaxis protein Y (CheY) orthologs provides a comparison of allosteric response across protein family divergence. Interestingly, we observe a rich mixture of both conservation and variability within the identified allosteric sites. While similarity within the overall response parallels the evolutionary relationships, >50% of the best scoring putative sites are only identified in a single ortholog. These results suggest that detailed descriptions of intraprotein communication are substantially more variable than structure and function, yet do maintain some evolutionary relationships. Finally, structural clusters of large response identify four allosteric hotspots, including the $\beta4/\alpha4$ loop known to be critical to relaying the CheY phosphorylation signal.

## INTRODUCTION

Long-range intramolecular communication within protein structures (i.e., allostery) is essential for life. Allosteric regulation is ubiquitous within both metabolic and cell signaling pathways because it provides for rapid responses to external stimuli (1). Consequently, allostery has been referred to as "a defining principle of life" and as the "second secret of life," second to the genetic code (2). Nevertheless, the physical mechanisms leading to allostery remain ambiguous (3). Additionally, the extent of conservation of allosteric mechanisms, or lack thereof, across a protein family remains largely undetermined (4).

To fill in the mechanistic gaps from experimental studies, there have been a large number of computational approaches developed to map out allosteric mechanisms and to identify putative allosteric sites. Because of the detail provided, methods based on molecular dynamics (MD) simulation are desirable; however, MD remains too computationally intensive to systematically interrogate all putative allosteric sites. As such, several coarse-grained methods spanning an array of theories have been developed. Despite overall diversity, each is fundamentally based upon perturbation/response analyses. For example, Hilser et al. (5) have mapped out how energetic couplings propagate from sites of energetic perturbation using COREX, their ensemble-based model of protein structure. While model details vary significantly, Ming and Wall (6,7) have developed an analogous approach based on changes within an elastic network model. Herein, surface points are systematically added to the network; points that substantially change the normal mode spectrum are then identified as allosteric. Similarly, Nussinov et al. (8) have developed a method based on network centrality where nodes corresponding to residues are systematically deleted from the network. Deletions with the largest effect on network centrality properties are predicted to be allosteric, which tend to cluster between the active and ligand-binding sites.

In this report, we introduce a related method based upon our distance constraint model (DCM). The DCM is an all-atom statistical mechanical model of protein structure that integrates thermodynamic and mechanical viewpoints (9,10). Previously, we have used the DCM to demonstrate that, while protein backbone flexibility is generally conserved across protein families, intrinsic intramolecular residue-to-residue couplings display pronounced variability (11–13). We now expand upon our earlier work by comparing allosteric couplings identified through a systematic set of external perturbations applied across two mesophilic and one thermophilic chemotaxis protein Y (CheY) structures. Specifically, using a mechanical perturbation method (MPM), additional constraints are applied to one residue at a time to mimic the reduction in conformational entropy that would occur upon the hypothetical situation of a ligand binding exclusively to that residue. Subsequently, we recalculate the DCM descriptions of stability and flexibility, which are compared to those from the unperturbed structure. Perturbed positions with large changes are thus identified as good candidates for residues likely to initiate allosteric signals. Conservation and variability across the three CheY structures are then evaluated to identify evolutionarily conserved allosteric pathways, and to assess the frequency of distinct intramolecular couplings.

CheY is a 128-residue $\alpha/\beta$ protein that is part of the bacterial chemotaxis signal transduction pathway (14). The inactive CheY is activated upon phosphorylation of conserved $Asp^{57}$ by CheA, which results in a 20-fold increase in affinity for the flagellar switch protein FliM (15). Upon

*Submitted January 18, 2010, and accepted for publication July 22, 2010.*
*Correspondence: djacobs1@uncc.edu or drlivesa@uncc.edu
Editor: Ruth Nussinov.

0006-3495/10/10/2245/10 $2.00
doi: 10.1016/j.bpj.2010.07.043

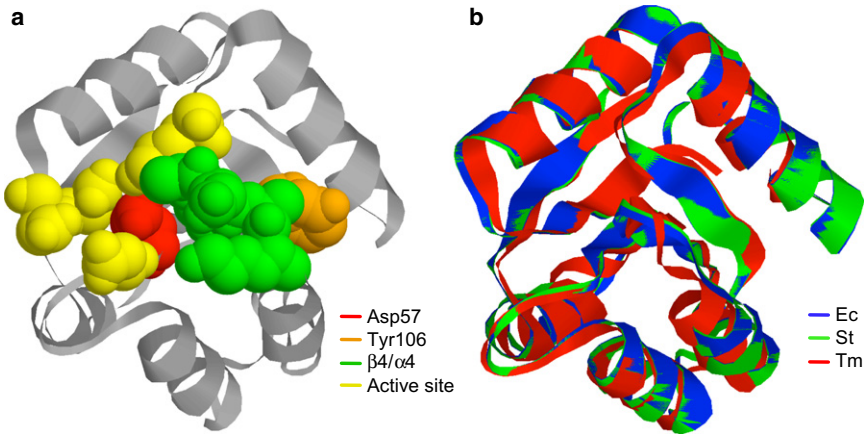


FIGURE 1 (*a*) The structure of CheY. The $Asp^{57}$ phosphorylation site, the other active site residues ($Asp^{12}$, $Asp^{13}$, $Asn^{59}$, and $Lys^{109}$), $Tyr^{106}$ and the $\beta4/\alpha4$ loop are highlighted in spacefill. (*b*) Structural superposition of the *E. coli*, *S. typhimurium*, and *T. maritima* CheY orthologs. The two panels are shown in the same orientation.

CheY/FliM association, the flagellum switches to a clockwise rotation resulting in random tumbles that reorient the bacterium. It is believed that phosphorylation of $Asp^{57}$ is coupled to an exposed-to-buried conformational change within $Tyr^{106}$ that facilitates the increased affinity for FliM (16). As shown in Fig. 1 *a*, there is no direct contact between $Asp^{57}$ and $Tyr^{106}$. Consequently, it is generally believed that the intervening $\beta4/\alpha4$ loop is, in one way or another, linked to the propagation of an allosteric signal between the two end-points. In particular, coupling between $Thr^{87}$ that is within $\beta4/\alpha4$ and $Tyr^{106}$ has been observed experimentally (16,17). In addition, a computational perturbation/response analysis (18) and molecular dynamics simulations (19–21) have also emphasized the importance of the $\beta4/\alpha4$ loop.

## METHODS

### The distance constraint model

The DCM is fundamentally based on a free energy decomposition scheme. Starting with a description of molecular structure as a mechanical framework that defines a network of distance constraints (22,23), each constraint is associated with an enthalpy and entropy component. In most free energy decompositions, the total enthalpy and entropy is determined by simply summing all components present in the network. However, this additive approach is incorrect because conformational entropy is a global nonadditive property. Nonadditivity originates from the fact that only independent entropy components should be summed (24,25). The DCM employs efficient network rigidity graph algorithms (23) to identify independent and redundant distance constraints within the network to obtain accurate estimates of the conformational entropy. In addition, the DCM calculates a number of ensemble averaged mechanical properties. In particular, network rigidity provides many mechanical descriptions of protein structure, such as local flexibility, rigid cluster decomposition and correlated motions (all of which are discussed below).

Herein, and in all of our work to date analyzing protein structure stability and flexibility (11–13,26), we employ a minimal distance constraint model (mDCM) that is based on two types of fluctuating constraints:

1. H-bonds.
2. Torsion angle forces.

Salt bridges are considered a special case of H-bonds, and torsion angles are segregated into native and disordered states. To construct the partition function, an ensemble of network rigidity graphs must be constructed. Even after invoking the common Gō-like assumption where only native contacts are allowed (27), the number of frameworks within the ensemble remains much too large (~$2^{780}$ for CheY) to directly calculate the partition function. As such, a two-dimensional free energy landscape is defined by order parameters associated with the numbers of H-bonds and nativelike torsion angles. Monte Carlo sampling is used to generate realizations that satisfy the macrostate definition by stochastically deleting interactions found in the native structure. This process is based on a Fermi-Dirac probability distribution where the chemical potential, which is described by a Lagrange multiplier, is adjusted to yield $N_i$ interactions of type $i$. The free energy of a given macrostate, $G(N_{hb}, N_{nat})$, is calculated by Monte Carlo sampling within a subensemble of frameworks with $N_{hb}$ H-bonds and $N_{nat}$ native torsions using the free energy functional,

$$\begin{aligned} G(N_{hb}, N_{nat}) \;=\; & U_{hb}(N_{hb}) - N_{hb}u_{sol} + N_{nat}v_{nat} \\ & - RTS_{cnf}(N_{hb}, N_{nat}|\delta_{nat}, \delta_{dis}, \gamma_{hb}) \\ & - RTS_{mix}(N_{hb}, N_{nat}), \end{aligned}$$

where $R$ is the universal gas constant and $T$ is temperature. The enthalpy and entropy parameters associated with each constraint describe, respectively, the depth and breadth of the potential energy basin for a particular interaction. For example, a native torsion force has a lower enthalpy than a disordered torsion (i.e., $v_{nat} < v_{dis}$), and the amount of phase space associated with a native torsion force is less than its disordered counterpart (i.e., $\delta_{nat} < \delta_{dis}$). Note that $v_{dis}$ is the reference energy, which is set to zero. $U_{hb}$ is calculated using an empirical potential (28) that the H-bond entropy $\gamma_{hb}$ is linearly related to (9,10). While intramolecular H-bond energies are calculated using $U_{hb}$, $u_{sol}$ accounts for the loss of stabilizing H-bonds to solvent. The net stability of a single H-bond is thus given by

$$\varepsilon_{net}^{hb} \;=\; U_{hb} - u_{sol}.$$

The entropy parameter $\delta_{dis}$ was fixed in previous work, leaving only $\{u_{sol}, v_{nat}, \delta_{nat}\}$ as adjustable. The mixing entropy, $S_{mix}$, is a $p_i \cdot \ln p_i$ term that arises from the mean-field sampling procedure. Typically 200 realizations are sufficient to achieve good statistics in $G(N_{hb}, N_{nat})$, which is needed to estimate the conformational entropy, $S_{\mathrm{cnf}}$, through an averaging over different constraint topologies, each involving a network rigidity calculation (11).

In this work, we consider three CheY orthologs. The mesophilic orthologs from *E. coli* and *S. typhimurium* are very similar, whereas the thermophilic *T. maritima* structure is quite divergent. Table S1 in the Supporting Material summarizes the differences and similarities between the three proteins (i.e., percent sequence identity and structural root-mean-square distance). During the comparison of the three CheY orthologs, we find that the H-bond potential energy function is overly sensitive. For example,

the H-bond between $Asp^{12}$ and $Arg^{18}$ is present in all three structures. While it is stabilizing in both the *E. coli* and *S. typhimurium* structures (−5.38 and −6.15 kcal/mol, respectively), it is destabilizing in the *T. maritima* structure (+0.45 kcal/mol). Careful examination reveals the *T. maritima* H-bond to be unfavorable (positive energy) because the donor/acceptor distance is very short, and the radial component of the H-bond potential is from a 10-12 Lennard-Jones potential, $V_{LJ}(R) \rightarrow \infty$ as $R \rightarrow 0$. In this case, it is obvious that the unfavorable H-bond is actually a very strong interaction.

Energy minimization of each input structure is a viable option to improve consistency. However, the sensitivity between corresponding H-bonds is not fully eliminated. To further reduce H-bond energy sensitivity, we alter the radial component such that for all $R < R_{eq}$, $V_{radial}(R) = V_{LJ}(R_{eq})$, and for $R \geq R_{eq}$, $V_{radial}(R) = V_{LJ}(R)$. With the updated potential, the *T. maritima* $Asp^{12}$-$Arg^{18}$ H-bond is now (−5.56 kcal/mol) consistent with the other two. Note that the *E. coli* and *S. typhimurium* values do not change because the radial distance is not less than $R_{eq}$. Several other H-bonds were brought into correspondence using this approach.

CheY binds $Mg^{+2}$, which is critical to its phosphotransfer activity (29). The $Mg^{+2}$ is chelated by $Asp^{13}$, $Asp^{57}$, and the backbone carbonyl of $Asn^{59}$ (30). Using a realistic total protein-metal binding energy (−4.50 kcal/mol), we model the three CheY-$Mg^{+2}$ interactions as H-bonds, each with $U_{Mg} = -1.50$ kcal/mol.

## Parameterization of the mDCM

Table S2 lists the values for all enthalpy and entropy parameters. In nearly all of our earlier work, these parameters were determined by fitting to experimental $C_p$ curves. Unfortunately, $C_p$ curves for the three CheY orthologs investigated herein are not available. Following the approach in Livesay and Jacobs (12), $\delta_{nat}$ is held constant across the three CheY structures because it is assumed to be dependent upon protein fold. Three different values of $\delta_{\text{nat}}$ are considered (0.4, 0.8, and 1.2), each equally spaced across the normal range identified from previous works. Conversely, $u_{sol}$ and $v_{\text{nat}}$ are allowed to vary. A grid search in $u_{sol}$ and $v_{nat}$ space is performed to identify putative values, conditional upon $\delta_{nat}$, that approximately match the experimental to the predicted $T_m$ as determined by the peak of the predicted $C_p$ curve. Multiple parameter sets are found in this way. Therefore, we demand that similar free energy profiles for the three orthologs are present at their respective $T_m$. After this second level of discrimination, the combination of parameters that provide the sharpest $C_p$ curves (meaning, cooperative-looking) are used to finally select the parameter sets (compare to Fig. S1). While this approach may not provide ideal parameters, it is expected to be sufficient to the work herein based on our prior demonstration that DCM predictions, especially for mechanical descriptions of the native basin, are robust to modest parameter variations. Fig. S2 and Fig. S3 in the Supporting Material demonstrate that the MPM results are largely insensitive across the three resultant parameter sets. Therefore, discussions hereafter will focus on $\delta_{nat} = 0.8$, keeping in mind that the same general conclusions occur for the other two parameter sets as well. Moreover, Table S3 quantifies the variation in the parameters $\{u_{sol}, v_{nat}\}$ for a given $\delta_{nat}$ value. The $\Delta\Delta G$ results are remarkably similar across the considered parameter space, whereas the $\Delta FI$ and $\Delta CC$ results are somewhat variable at the edges of the considered space that identifies acceptable parameter boundaries. Nevertheless, Fig. S4 clearly demonstrates that the minima corresponding to the best scoring positions are mostly conserved despite local fluctuations within the $\Delta FI$ and $\Delta CC$ results.

The considered structures are: *E. coli* = 3CHY (31), *S. typhimurium* = 2CHE (30), and *T. maritima* = 1TMY (32), all of which are the inactive forms of the protein. Herein, residue numbering is based on the 3CHY structure. Fig. 1 *b* highlights the structural similarity between the three structures. As an all-atom model, the mDCM requires that hydrogen atoms be added to the x-ray structures. Using our previous protocol (13), hydrogen atoms are added using the program H++ (33), which adds hydrogen atoms consistent with $pK_a$ values calculated from Poisson-Boltzmann electrostatics. Employed electrostatic parameters include a salinity of 0.15 M and external/internal dielectrics of 80 and 6, respectively. Output structures are protonated assuming a pH of 7.0. After hydrogen atoms have been added, Fig. S5 compares the resultant H-bond topologies across the three CheY pairs.

## The mechanical perturbation method

To test for allosteric response, the MPM systematically introduces quenched constraints as a mechanical perturbation to one residue at a time. The perturbation is defined by locking in place the $\phi$, $\psi$, and $\chi_1$ of the target residue, which will alter the thermodynamic and mechanical properties of the protein. Per target residue, all metrics using the perturbed input structure are recalculated, and subsequently mechanical and thermodynamic response is tracked via changes detected relative to the unperturbed structure. After each residue is targeted, putative allosteric sites are identified as outliers within the statistics of the response in stability and flexibility profiles. While different types of mechanical perturbations have been tested, clamping in place the $\phi$, $\psi$, and $\chi_1$ dihedral angles in the target residue removes undesired bias toward greater response from larger residues.

Note that, while not normally defined as a $\chi$-angle, we include a quenched constraint to lock the torsion angle defined by N-CA-CB-H in Ala. Quenching this Ala side-chain torsion in a molecular dynamics simulation would only locally prevent rotation of the methyl group that is generally not important. However, this additional constraint allows the mechanical perturbation to be consistent across all residue types, except Gly. The exception Gly, where only $\phi$ and $\psi$ are constrained, shows no significant differences in response compared to other residues. Consequently, the response from any target residue reported by the MPM reflects the intrinsic network properties of interactions.

The mDCM provides a high dimensional description of protein stability and flexibility that we collectively refer to as quantified stability/flexibility relationships (QSFR). Following our previous work (11–13), we track three informative QSFR quantities: $\Delta\Delta G_{\text{fold}}$, change in backbone flexibility, and residue-residue couplings. Backbone flexibility is described by a flexibility index, *FI*, where positive values indicate an excess density of degrees of freedom, and negative values indicate an excess density of redundant constraints. Residue-to-residue couplings are described using a symmetric cooperativity matrix, *CC*, to quantify the degree at which two residues are mutually rigid or flexible within a correlated motion. The *FI* and *CC* descriptions are calculated within the native free energy basin to characterize native state response, which also suppress variations due to parameterization differences (12).

The rank of each response type varies, where $\Delta\Delta G$, $\Delta FI$, and $\Delta CC$ reflect a difference of a scalar, vector, and matrix, respectively. Changes in thermodynamic response are expressed as $\Delta\Delta G$ (perturbed, wild-type), whereas the two mechanical metrics are more difficult to quantify with a single value because they describe per-residue and residue-pair properties. $\Delta FI$ is calculated as the root-mean-square distance (RMSD) between vectors composed of the perturbed and wild-type structures. However, upon mechanical clamping of a target residue, it and its flanking residues along the backbone become much more rigid, but this response is not allosteric. Therefore, *FI* differences within $\pm 3$ residue positions from the target residue are not considered in the RMSD. Although a RMSD metric for $\Delta CC$ could be used, a more sensitive measure given by the Pearson correlation coefficient between the wild-type and perturbed structures is employed. Large differences in the metrics used to quantify $\Delta\Delta G$, $\Delta FI$, and $\Delta CC$ globally across the protein reflect an accumulative effect that occurs only when many residues are affected by the perturbation, indicating a high degree of allosteric communication.

Although it is desirable to identify allosteric mechanisms that trace the response at site *j* upon ligand binding at site *i*, where the ligand binding site may contain multiple residues, such an analysis greatly complicates the investigation of evolutionary relationships within allostery. The MPM eliminates the complexity of modeling particular protein-ligand

interactions. By uniformly applying a localized perturbation that is small in magnitude (only three applied distance constraints) to all residues, a linear response regime is approached. That is, the identified long-range response is an intrinsic property of the protein in its native state. As such, the nature of the response need not be assessed in terms of different types of disturbances of the interaction network. For the same reasons, intrinsic response is best described by global metrics (i.e., $\Delta\Delta G$, $\Delta FI$, and $\Delta CC$) to reduce variance among local differences between specific residue pairs. Tracking global variance builds a consensus from all residue pairs to highlight when a target residue is likely to transmit an allosteric signal, albeit the response location is lost. The MPM is therefore well suited for comparing the residues across a family of proteins that is most active in generating allosteric response.

## RESULTS

### Intrinsic CheY properties

It is well known that most thermophilic proteins are more stable and rigid than mesophilic orthologs at a constant temperature. On the other hand, it is also generally believed that the balance between stability and rigidity is conserved across the family at some appropriately shifted temperature (i.e., $T_m$ or optimal grown temperature) (34–37). These characteristics were sought during parameterization by looking for similar free energy as a function of global flexibility at each respective $T_m$. The one-dimensional free energy landscape for the three CheY orthologs in Fig. 2 *a* shows similar overall stability/flexibility characteristics within the native basins. Moreover, Fig. 2 *b* plots the rigid cluster susceptibility curves that quantify the amount of mechanical fluctuations within the ensemble. The overlapping curves demonstrate that the rigid-to-flexible transition is nearly identical across the three structures. The location of the native free energy basin along the global flexibility order parameter, $\theta$, precedes the rigidity transition, which indicates that the native structure is globally rigid in all three cases.

Therefore, it is somewhat surprising that the N-terminal region of the thermophilic *T. maritima* ortholog (specifically, the first 53 residues) is observed to be substantially more rigid than its mesophilic counterparts (compare to Fig. 2 *c*). This result is especially puzzling considering the fact that the *T. maritima* ortholog is known to not have the normal signatures of thermophilic proteins (i.e., increased numbers of H-bonds/salt bridges, better packing, greater compactness, etc.) (32). In fact, Fig. S6 highlights that there

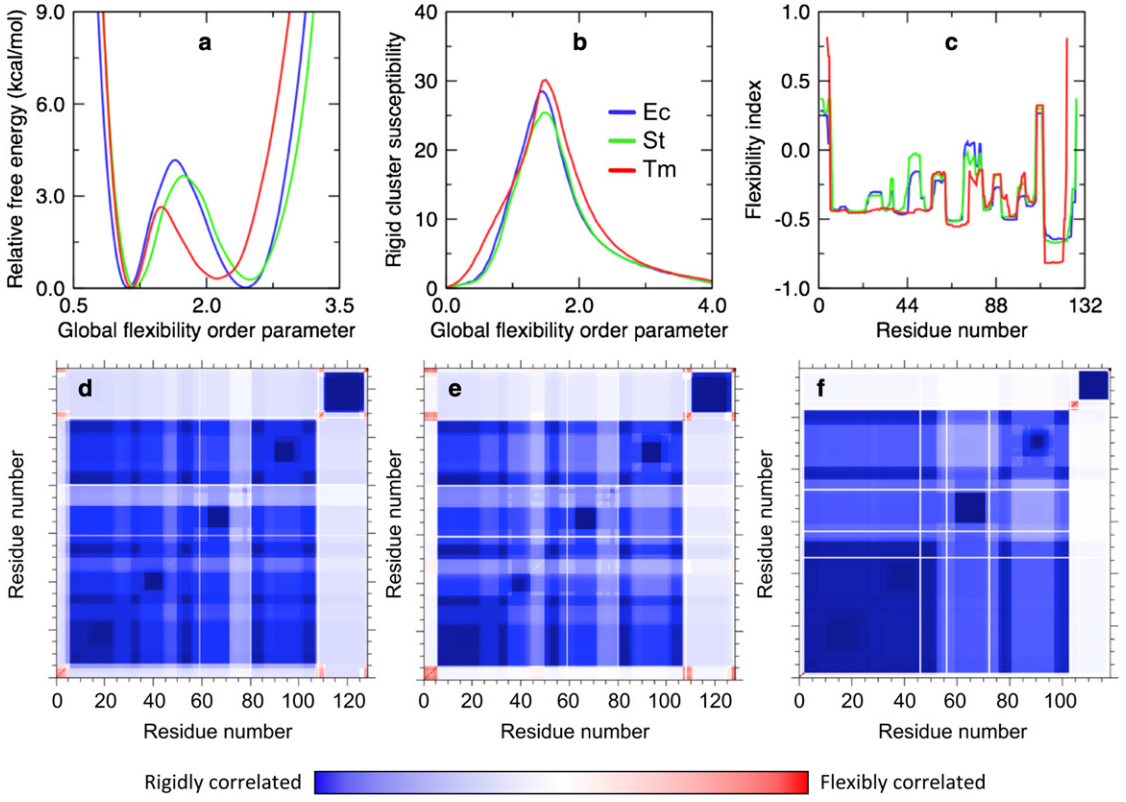


FIGURE 2 Various QSFR quantities of wild-type CheY structures. The (*a*) free energy landscape, (*b*) rigid cluster susceptibility, and (*c*) flexibility index of the *E. coli*, *S. typhimurium*, and *T. maritima* CheY orthologs. Cooperativity correlation plots that report residue-to-residue pairwise mechanical couplings are provided for the same three, respectively, in panels *d–f*. Cooperativity correlation plots are symmetrical matrices that indicate whether each residue pair is likely to be within the same rigid cluster or correlated motion. A color scale for the cooperativity correlation plots is provided. In panels *c–f*, the flexibility index and cooperativity correlation values are from an average over the native basin (versus full ensemble) only.

are actually fewer total H-bonds in the *T. maritima* structure compared to the other two. The decrease in H-bond numbers is partially offset by an increase in the average H-bond strength. As a consequence of the differences in the H-bond networks, we find that when using a constant set of mDCM parameters, the *T. maritima* structure is less stable (lower $T_m$) than its mesophilic counterparts. As such, parameter differences are used to reestablish the correct $T_m$ trends. This situation is the same as our previous work on a mesophilic/thermophilic RNase H pair (12), where phenomenological parameter diversity were interpreted in terms of differences within hydrophobic packing. While parameter variability is suggestive of differences due to hydrophobic packing, this interpretation is not clear-cut in this work because the experimental solvent conditions that were constant in the RNase H example were not controlled for. Regardless of interpretation, the parameter differences represent an added cohesive force within the *T. maritima* structure that is supported experimentally by a greater $T_m$.

The *CC* plots shown in Fig. 2, *d–f*, parallel the above results. In all three structures, the protein predominantly includes two large rigid clusters that are respectively composed of the protein core and the C-terminal helix. In the mesophilic structures, there is a small amount of correlated flexibility between the N- and C-terminal coil regions. The dark blue within the N-terminal region of the *T. maritima* structure highlights that this portion of the structure is almost always part of a large rigid cluster, whereas the lighter shades in the other two structures indicate that they have perceptible rigidity fluctuations within this region. As indicated by the lighter-blue color, there are more rigidity fluctuations within the C-terminal end of the core in all three structures, especially near residues 70–80.

## Overview of the MPM results

The results for the three MPM response functions that systematically assess the consequences of rigidifying one residue at a time are plotted in Fig. 3, *a–c*. As expected, mechanical clamping of a residue increases protein stability and tends to rigidify the structure, but it typically will increase flexibility in some localized regions. To facilitate comparisons across the three CheY structures and the response functions, the raw data have been converted to *z*-scores that quantify the number of standard deviations a given signal is away from the mean for that ortholog/response function pair. For consistency within the plots, a negative of the *z*-score is used to characterize $\Delta FI$, which makes the minimums in all panels indicate positions with the most extreme change, whereas peaks correspond to minimal change. For example, consider $\Delta CC$. Positions predicted to initiate allostery are those that produce a low Pearson correlation coefficient between perturbed and wild-type structures, which implies the set of pairwise couplings throughout the protein is substantially altered. Note that no drastic peaks are possible for $\Delta CC$ because the greatest Pearson correlation coefficient that can be obtained is 1. For $\Delta FI$, there can be more interesting peak structure, but not much is observed.

Upon comparing the three quantities, it is clear that there is a rich mixture of both conservation and variability across the three MPM response functions. Fig. 3 *d* is a scatter plot comparing each *E. coli* response function pair. Table S4 lists the correlation coefficients comparing the various response function pairs for a given ortholog, demonstrating the orthologs share similar results. Specifically, we observe no appreciable correlation between changes in protein stability to the mechanical responses $\Delta CC$ and $\Delta FI$. However, considerable

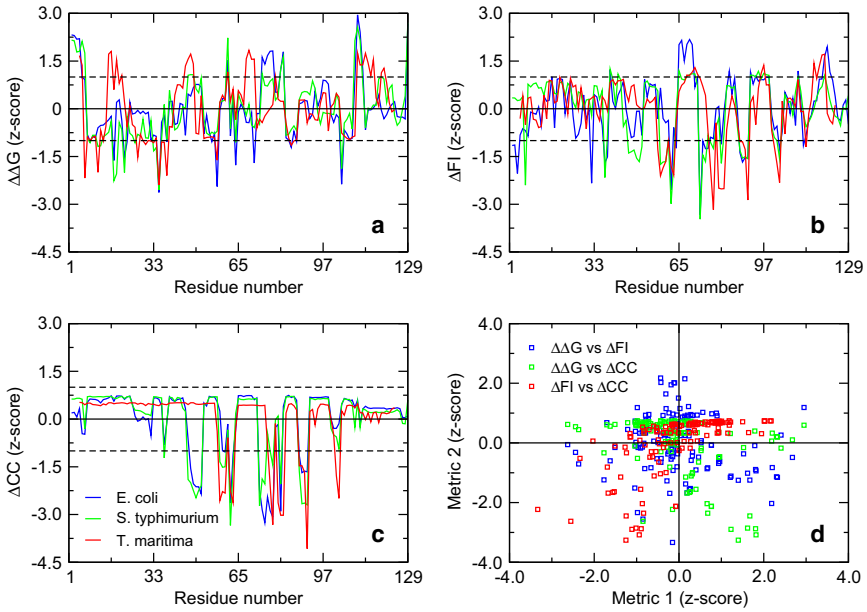


FIGURE 3 Summary of the MPM results. Upon mechanically clamping a given residue, relatively large changes in (*a*) $\Delta G_{fold}$, (*b*) flexibility index (*FI*), and (*c*) cooperativity correlation (*CC*) are used to identify allosteric sites. To facilitate comparisons, raw values have been converted to statistical *z*-scores (except for $\Delta FI$, which is negative of the *z*-score). The solid $z = 0$ line indicates the average change, whereas the dashed lines indicate $z = \pm 1$ standard deviation. In panel *d*, all three of the *E. coli* metrics are compared in a scatter plot. While there is clearly some agreement within the scores (especially between *FI* and *CC*), there is also appreciable diversity within the responses. Correlation coefficients for all metric pairs are provided in Table S4.

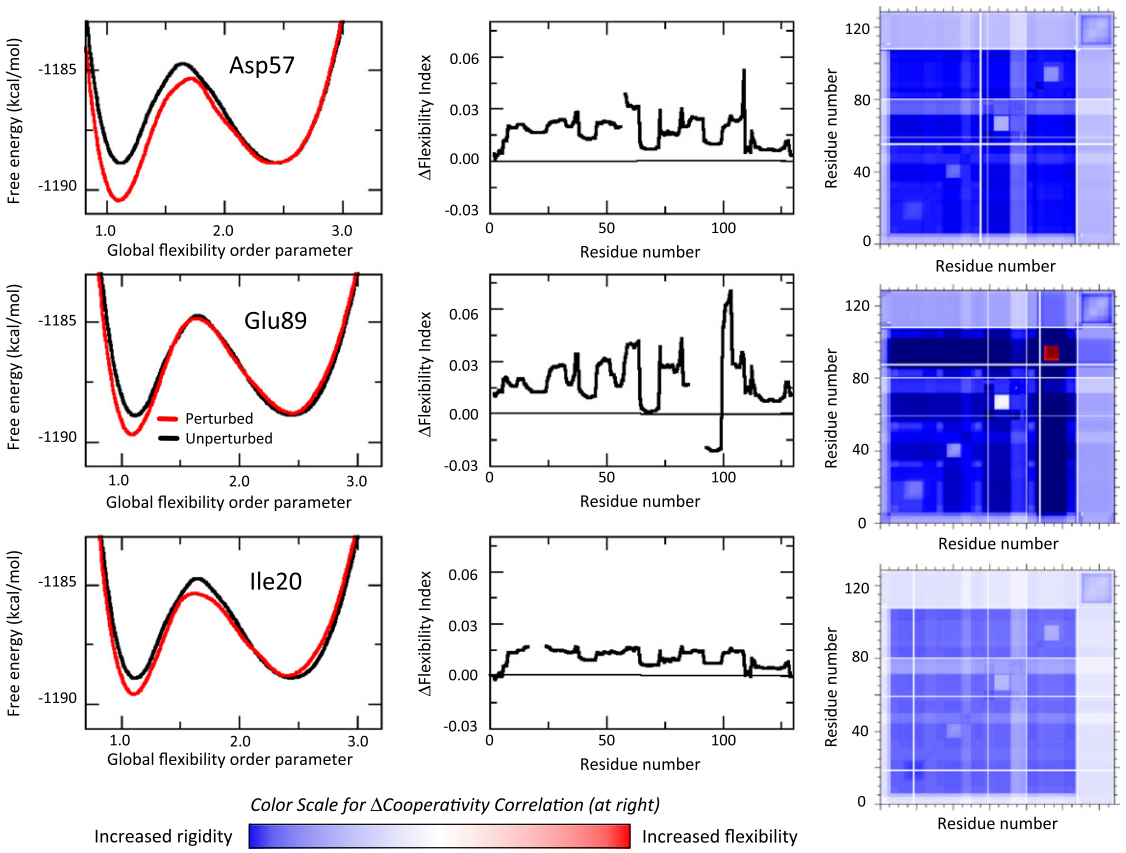


FIGURE 4 Overview of the MPM response. The $\Delta\Delta G$ (left column), $\Delta FI$ *(middle column)*, and $\Delta CC$ *(right column)* response is shown for three exemplar cases. From top to bottom, the rows correspond to $Asp^{57}$, $Glu^{89}$, and $Ile^{20}$. Relative to the other positions, $Asp^{57}$ displays a large $\Delta\Delta G$ response, whereas its mechanical response is not within the top 20%. Nevertheless, small changes in $\Delta FI$ and $\Delta CC$ are observed throughout the structure. Conversely, $Glu^{89}$ has a relatively small $\Delta\Delta G$ response, whereas its mechanical response is large. $Ile^{20}$ has a relatively small response in all three. The break in the $\Delta FI$ response corresponds to the three residues on either side of the perturbation that are not considered in the MPM calculation. A color scale for the $\Delta CC$ plots is provided.

correlation between $\Delta CC$ and $\Delta FI$ is observed, which indicates backbone flexibility plays an important role in mediating cooperativity between residue pairs, albeit not a controlling factor. Fig. 4 shows the QSFR effects upon perturbation of three exemplar positions: $Asp^{57}$, $Glu^{89}$, and $Ile^{20}$, which represent positions with relatively large $\Delta\Delta G$ response and/or large $\Delta CC$ and $\Delta FI$ responses. Positions with limited response in all three metrics are not regarded as being prone to activate an allosteric response. Nevertheless, it is worth mentioning that positions that initiate the smallest response (i.e., $\Delta\Delta G$ = residues 5, 61, and 110; $\Delta FI$ = residues 68 and 120) can also be evolutionarily conserved.

## Evolutionary conservation of predicted allosteric positions

Table S5 lists the correlation coefficients of a given MPM response function across the three CheY orthologs. As expected based on their evolutionary relationships, the MPM results are well conserved across the *E. coli* and *S. typhimurium* pair, whereas each of the response functions is significantly different than *T. maritima*. To facilitate comparisons, the MPM results are coarse-grained in order to focus on residues giving the strongest signals. For simplicity, we consider signals within the top 10, 15, and 20% of a given ortholog/response function pair. Positions with MPM values outside of the top 20% are considered to have no significant ability to initiate allostery. The probability distributions for each pair are shown in Fig. S7. In all cases, the considered cutoffs are within a satisfactory distance to the left of the distribution peak. Fig. S8 color-codes the CheY alignment by the percentile cutoff value at which each CheY residue is, if at all, identified. Table S6 provides summary statistics regarding evolutionary conservation of the MPM results. Specifically, the percentage of identified allosteric sites for a given response function/cutoff pair are provided. Regardless of cutoff, >52% of the predicted allosteric positions only occur in a single ortholog. The percentages found in any two and all three CheY orthologs expectedly diminishes in each case. Despite numerous permutations, no correlation was found between the sequence variability and the raw MPM data or the coarse-grained results using the cutoff percentages.

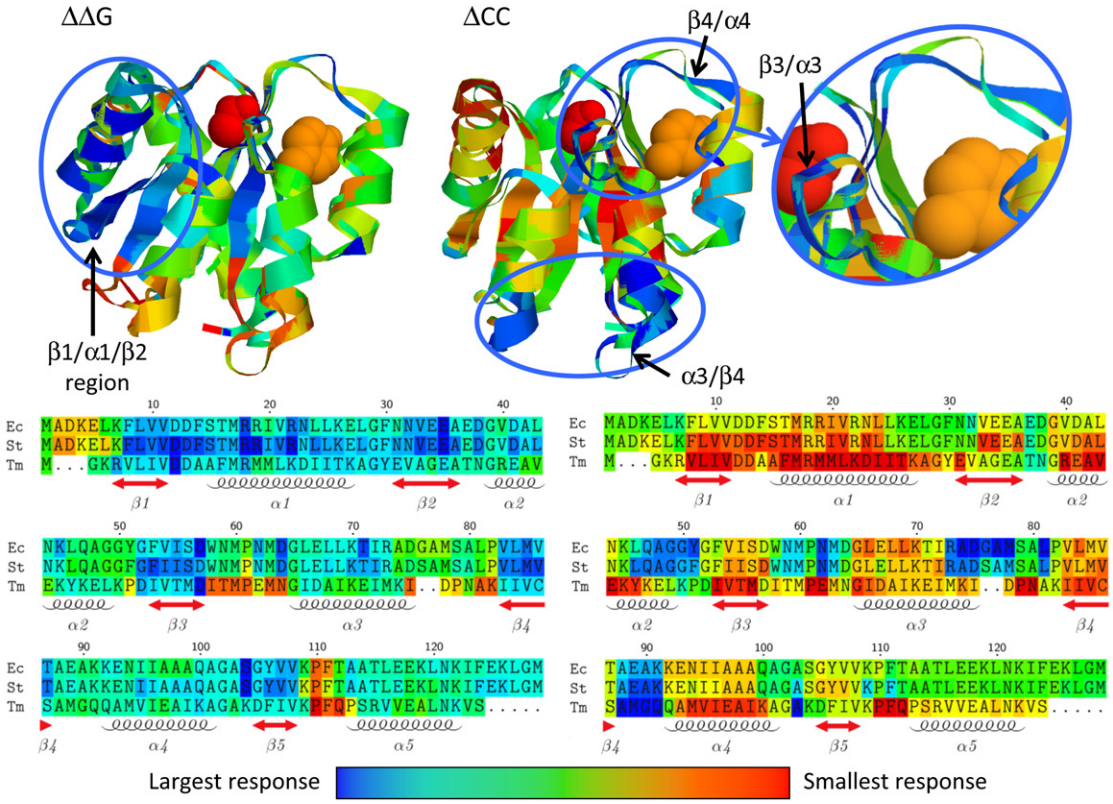


FIGURE 5 Structural and sequence descriptions of the mechanical perturbation results. The three CheY ortholog structures are superimposed and color-coded by $\Delta\Delta G$ and $\Delta CC$. The two figures are in nearly the same orientation, but slightly shifted to highlight respective key features. $Asp^{57}$ and $Tyr^{106}$ are shown in spacefill to assist orientation using the same coloring as Fig. 1. Interestingly, our results identify four allosteric hotspots, including the $\beta4/\alpha4$ loop that is known to be important to CheY allostery. The $\Delta FI$ results are not shown because they are very similar to the $\Delta CC$ results. The multiple alignments are color-coded in the same way (*left* = $\Delta\Delta G$ and *right* = $\Delta CC$). In each case, the sequence ordering is (*top* to *bottom*) *E. coli*, *S. typhimurium*, and *T. maritima*. Note: Color is available only in the online version.

Fig. 5 plots the results from two of the MPM response functions ($\Delta\Delta G$ and $\Delta CC$) to a structural superposition of the three CheY proteins and also shows the alignments color-coded by the same quantities. Note that Fig. S9 plots a statistically binned view of the same data where a single structure is color-coded by the number of times a given position is identified as allosteric. While there is considerable evolutionary variability within the position-specific ability to initiate allostery, scrutiny of both figures reveals that structurally isolated allosteric sites are rarely observed. Instead, there is extensive clustering of positions with large responses in both, suggesting that residue-specific differences within intramolecular communication pathways are both possible and common. Nevertheless, structure and/or function does impose evolutionary constraints that limit the global variation.

## The CheY active site and allosteric hotspots

Table 1 lists the percentile cutoff at which known active site residues (31) are identified as likely to initiate an allosteric response. Because phosphorylation of $Asp^{57}$ initiates a structural rearrangement that causes a conformational change within the gating residue $Tyr^{106}$, it is natural to expect $Asp^{57}$ to be highly allosteric. In fact, this is the case for two of the orthologs. The $\Delta\Delta G$ response function at $Asp^{57}$ is among the strongest signals for the *E. coli* and *T. maritima* orthologs. $Asp^{57}$ from *T. maritima* is also identified by $\Delta CC$. $Asp^{57}$ is not identified here within the *S. typhimurium* structure. Across the set of active site residues, three residues are identified by a single response function and only identified within a single ortholog. The remaining three are identified by two response functions in one or two of the CheY orthologs. None of the active site residues are predicted to initiate an allosteric response across all three CheY orthologs. Interestingly, MD simulations suggest that rotation within $Tyr^{106}$ actually initiates allosteric response by activating the $\beta4/\alpha4$ loop (20). In this work, $Tyr^{106}$ is only predicted to initiate an allosteric response at the 20% cutoff in one ortholog, which supports the traditional view that conformational changes in $Tyr^{106}$ is the outcome of the allosteric signal, and not an initiating factor.

Table 1 also includes the MPM results for the $\beta4/\alpha4$ loop residues, which initiate the largest $\Delta FI$ and $\Delta CC$ responses. In fact, loop residues 88–91 initiate global mechanical changes in all three orthologs. Considering the diversity in

**TABLE 1 Comparison of the three MPM response functions within known CheY functional sites using $\delta_{nat} = 0.8$**

| Residue | Functional role | $\Delta\Delta G$ | $\Delta FI$ | $\Delta CC$ |
|---|---|---|---|---|
| Active site residues | | | | |
| $Asp^{12}$ | Part of active site H-bond network, which is critical to function | –, 10, 10 | –,–, 20 | –,–,– |
| $Asp^{13}$ | $Mg^{+2}$ binding site | –, 10,– | –,–,– | –,–,– |
| $Asp^{57}$ | Phosphorylation and $Mg^{+2}$ binding site | 10,–, 10 | –,–, 15 | –,–, 15 |
| $Asn^{59}$ | $Mg^{+2}$ binding site | –,–,– | –, 10, 10 | –, 10, 10 |
| $Tyr^{106}$ | Conformational switch | –,–, 20 | –,–,– | –,–,– |
| $Lys^{109}$ | Part of active site H-bond network, which is critical to function | –,–,– | –,–,– | 10,–,– |
| The $\beta4/\alpha4$ loop | | | | |
| $Thr^{87}$ | Allosteric response | 15,–, 15 | –,–,– | –,–,– |
| $Ala^{88}$ | Allosteric response | –,–,– | 10,–, 15 | –, 15, 15 |
| $Glu^{89}$ | Allosteric response | –,–,– | 10, 15, 15 | 10, 10, 15 |
| $Ala^{90}$ | Allosteric response | –,–,– | 10, 15, 15 | 15, 15, 15 |
| $Lys^{91}$ | Allosteric response | –,–,– | 10, 15, 10 | 15, 10, 10 |

The MPM results for the six active site residues identified within Volz and Matsumura (31) are indicated for the *E. coli* structure. Likewise, the MPM results for the $\beta4/\alpha4$ loop are also indicated. Within each MPM results entry, the order is: *E. coli*, *S. typhimurium*, and *T. maritima*, respectively. Values indicate which of the three percentile cutoffs each particular site is first identified.

response of global change due to perturbation, this pronounced conservation is compelling. Clearly, our results support the notion the $\beta4/\alpha4$ loop is critical to function in all three considered CheY structures. Interestingly, while perturbation at $Glu^{89}$ is generally rigidifying, it also leads to an atypical increase in flexibility around $Ala^{98}$ (compare to Fig. 4). $Thr^{87}$ is also predicted to be allosteric in the *E. coli* and *T. maritima* structures by $\Delta\Delta G$. It is noteworthy that there is a clear enrichment of conserved residues that initiate allostery within the known functional sites. Across the entirety of our results, only 25% of the CheY positions are found to be allosteric in two or more orthologs using the 15% cutoff (compare to Fig. S7), whereas 73% of the functional residues indicated in Table 1 meet this threshold.

Other regions where the MPM identifies hotspots of structural positions with a conserved ability to produce allosteric responses are also observed in Fig. 5 and Fig. S9; however, the following regions are generally only conserved across two of the three orthologs. Specifically, there is a large clustering of positions with large $\Delta\Delta G$ values at a site on the $\beta1/\alpha1/\beta2$ face of the protein, which is roughly perpendicular to the active site region. Conversely, $\Delta FI$ and $\Delta CC$ also identify the $\beta3/\alpha3$, which is not surprising owing to its structural proximity to the sites discussed above (including $Asp^{57}$, $Tyr^{106}$, and the $\beta4/\alpha4$ loop). In addition, $\Delta FI$ and $\Delta CC$ also identify a large site distal to the active site that is defined by $\alpha2$ and the subsequent loop is also identified. To the best of our knowledge, these two sites have never been identified as playing a role in CheY allostery. Of particular excitement is that these two sites are at locations distant from the CheY active and FliM binding sites. As such, we encourage the CheY experimental community to test the accuracy of our predictions in the near future through structure/function analyses within these sites.

As discussed further below, the $\Delta FI$ and $\Delta CC$ response functions tend to identify sites defined by loop regions, which can extend into the flanking secondary structure elements. A direct comparison of Fig. 2 *c* and Fig. 3, *b* and *c*, suggests that mechanical response functions closely parallel the protein's intrinsic flexibility. Nonetheless, normalization of the $\Delta FI$ and $\Delta CC$ response by residue flexibility demonstrates that $\Delta FI$ and $\Delta CC$ are not simply recapitulating *FI*. Note that even after normalization by FI, the $\beta4/\alpha4$ loop provides the strongest signal (results not shown). The inset in Fig. 5 highlights that the identified $\beta3/\alpha3$ loop is within the active site region. In particular, $Asn^{62}$, which is identified within the site record of the *E. coli* PDB file, is also identified within all three structures.

## DISCUSSION

### Allosteric response is both conserved and variable

There is a growing consensus that allosteric mechanisms can be quite variable across a protein family. For example, cooperativity within hemoglobins, an archetype of long-range communication, includes a diverse array of mechanistic pathways (38). Similarly, allosteric control of the G-protein coupled receptor family has been shown to vary significantly in regulator choice and binding site locations (39). Moreover, quantitative differences in allosteric regulation have been used as a molecular basis of taxonomic assignments within enzyme families (40). In all three examples, stark differences within allosteric mechanisms are observed across the family. However, conserved allosteric responses across taxonomic groups are also observed in all cases, highlighting that conservation of allostery can occur over short evolutionary distances.

Our computational results on the CheY orthologs studied here closely parallel all three of these general trends. We observe a substantial amount of position-specific diversity across the three CheY structures. In fact, approximately half of the positions predicted to initiate allostery using the 20% cutoff occur in only a single ortholog, highlighting the position-specific variability within the ability to initiate allosteric changes. Nevertheless, allosteric positions do tend to be structurally clustered into hotspot regions, indicating that while sequence variability allows for position-specific differences, structure and/or function does somewhat constrain the location of residues that initiate allostery. In addition, strong taxonomic effects are observed. The MPM results are strongly conserved across the *E. coli* and *S. typhimurium* orthologs (compare to Table S5), whereas

each is much less conserved to the evolutionary divergent *T. maritima* ortholog. These relationships reflect the similarity, or lack thereof, between the three CheY proteins. For example, the *E. coli* and *S. typhimurium* sequences are 97.7% identical and their backbone structure RMSD is 0.2 Å. Nevertheless, depending upon which MPM function is considered, their MPM results only correlate at $R = 0.74$ to 0.92. The attenuated similarity is explained by the fact, even though their sequences and backbone structures are quite similar, their H-bond networks are slightly more divergent (compare to Fig. S5). As a consequence of the long-range nature of rigidity, the allosteric signal is globally propagated throughout the altered networks in varying and unforeseen ways. This point is underscored by the large amount of variability within the positions that are identified within the top 20% of the MPM responses.

### Interpreting thermodynamic versus mechanical response

Fig. 3 *d* and Table S4 demonstrate that there is very little overlap between $\Delta\Delta G$ and the two mechanical response functions. The peaks and minima in $\Delta FI$ (Fig. 3 *b*) are interpreted the same, but in this case local flexibility is being tracked (versus residue-to-residue couplings in $\Delta CC$). Again, changes are observed throughout the protein, including small deviations within the nonallosteric positions (compare to Fig. 4). Using the same signal threshold, the number of minima is far greater than the number of peaks. However, a few peaks do occur in this case, indicating atypically small abilities to initiate changes in FI upon perturbation. The most obvious is the peak around residue 68 within the *S. typhimurium* structure. Curiously, this peak cannot be attributed to anything simple (i.e., H-bond strength, intrinsic flexibility, etc.). Nevertheless, we note a general tendency for positions with the smallest abilities to alter FI to occur within rigid secondary structure elements, whereas positions with the largest ability to alter FI tend to occur at the ends of secondary structure elements and within intervening loops.

The $\Delta\Delta G$ spectra (Fig. 3 *a*) are somewhat noisier than $\Delta CC$ or $\Delta FI$. The general interpretation of the peaks and minima is the same as the two mechanical response functions. Herein, all perturbations have some stabilizing effect on the structure. The minima indicate positions that have the largest stabilizing effect, whereas the peaks indicate positions with the smallest stabilizing effect. The latter positions tend to correspond to intrinsically flexible positions. This is because flexible regions tend to have an excess number of degrees of freedom. Upon introduction of the native torsion, there is a concomitant reduction in entropy, thus attenuating the $\Delta\Delta G$ response. Conversely, structural regions that are innately rigid tend to display increased $\Delta\Delta G$ response because the region is more likely to have independent constraints (before the perturbation) that no longer need to pay an accompanying entropic price after perturbation.

Taken together, the MPM response functions are shown to be somewhat orthogonal to each other. That is, $\Delta\Delta G$ generally tends to identify rigid sites, whereas $\Delta CC$ and $\Delta FI$ tend to identify flexible sites. However, this is a clearly an overly simplistic description of the results. There are exceptions to these general trends, and the strongest abilities to initiate allostery in all three cases remain even after normalization for flexibility. In fact, there are many loops that are not identified by $\Delta FI$ or $\Delta CC$, thus highlighting the complex nature of network rigidity (23). What is important to distill is how the perturbation signal is propagated through the structure, which is related to the constraint network topology and entropy nonadditivity. Despite many attempts to do so, these complex relationships could not be recapitulated using simple sequence-base descriptions, because the structural network plays an essential role in allosteric communication. While each MPM response function provides a glimpse into how network properties change, none provide a complete view. Consequently, we take a more Gestalt view of the results by simultaneously considering all three and also by looking for consensus across the three CheY orthologs—from which we predict four allosteric hotspots within CheY (compare to Fig. 5), including the $\beta4/\alpha4$ loop.

## CONCLUSIONS

Using an in silico screen for allosteric sites, we demonstrate that residues likely to initiate allosteric changes are both conserved and variable across three CheY orthologs, highlighting the complex nature of allostery evolution. While our results identify a large number of residue-specific differences across the family, they also indicate that allosteric response is more conserved across short evolutionary distances and that structure and/or function imposes appreciable constraints on allosteric response.

Based on this rich mixture of conservation and variability, we suggest that hard and fast rules regarding conservation of allostery are unlikely to materialize. Moreover, the strongest predicted allosteric site strengthens the view that the $\beta4/\alpha4$ loop is a critical link in the intramolecular communication within CheY. In addition, limited conservation within three additional hotspots is also identified. We hope that future experimental studies will target these putative sites to test our predictions.

## SUPPORTING MATERIAL

Six tables and nine figures are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(10)00915-X.

The authors thank the anonymous reviewers for several useful suggestions. Key to the distance constraint model is the use of graph-rigidity algorithms. This algorithm is claimed in U.S. Patent No. 6,014,449, which has been

assigned to the Board of Trustees Michigan State University and is used with permission.

This work is supported by National Institutes of Health grant No. R01 GM070382.

## REFERENCES

1. Kuriyan, J., and D. Eisenberg. 2007. The origin of protein interactions and allostery in colocalization. *Nature.* 450:983–990.
2. Fenton, A. W. 2008. Allostery: an illustrated definition for the 'second secret of life'. *Trends Biochem. Sci.* 33:420–425.
3. Whitley, M. J., and A. L. Lee. 2009. Frameworks for understanding long-range intra-protein communication. *Curr. Protein Pept. Sci.* 10:116–127.
4. Hwang, P. K., and R. J. Fletterick. 1986. Convergent and divergent evolution of regulatory sites in eukaryotic phosphorylases. *Nature.* 324:80–84.
5. Pan, H., J. C. Lee, and V. J. Hilser. 2000. Binding sites in *Escherichia coli* dihydrofolate reductase communicate by modulating the conformational ensemble. *Proc. Natl. Acad. Sci. USA.* 97:12020–12025.
6. Ming, D., and M. E. Wall. 2005. Allostery in a coarse-grained model of protein dynamics. *Phys. Rev. Lett.* 95:198103.
7. Ming, D., and M. E. Wall. 2005. Quantifying allosteric effects in proteins. *Proteins.* 59:697–707.
8. del Sol, A., H. Fujihashi, …, R. Nussinov. 2006. Residues crucial for maintaining short paths in network communication mediate signaling in proteins. *Mol. Syst. Biol.* 2:2006–0019.
9. Jacobs, D. J., and S. Dallakyan. 2005. Elucidating protein thermodynamics from the three-dimensional structure of the native state using network rigidity. *Biophys. J.* 88:903–915.
10. Livesay, D. R., S. Dallakyan, …, D. J. Jacobs. 2004. A flexible approach for understanding protein stability. *FEBS Lett.* 576:468–476.
11. Livesay, D. R., D. H. Huynh, …, D. J. Jacobs. 2008. Hydrogen bond networks determine emergent mechanical and thermodynamic properties across a protein family. *Chem. Cent. J.* 2:17.
12. Livesay, D. R., and D. J. Jacobs. 2006. Conserved quantitative stability/flexibility relationships (QSFR) in an orthologous RNase H pair. *Proteins.* 62:130–143.
13. Mottonen, J. M., M. Xu, …, D. R. Livesay. 2009. Unifying mechanical and thermodynamic descriptions across the thioredoxin protein family. *Proteins.* 75:610–627.
14. Eisenbach, M. 2007. A hitchhiker's guide through advances and conceptual changes in chemotaxis. *J. Cell. Physiol.* 213:574–580.
15. Dyer, C. M., and F. W. Dahlquist. 2006. Switched or not? The structure of unphosphorylated CheY bound to the N terminus of FliM. *J. Bacteriol.* 188:7354–7363.
16. Zhu, X., C. D. Amsler, …, P. Matsumura. 1996. Tyrosine 106 of CheY plays an important role in chemotaxis signal transduction in *Escherichia coli*. *J. Bacteriol.* 178:4208–4215.
17. Cho, H. S., S. Y. Lee, …, J. G. Pelton. 2000. NMR structure of activated CheY. *J. Mol. Biol.* 297:543–551.
18. Atilgan, C., and A. R. Atilgan. 2009. Perturbation-response scanning reveals ligand entry-exit mechanisms of ferric binding protein. *PLOS Comput. Biol.* 5:e1000544.
19. Formaneck, M. S., L. Ma, and Q. Cui. 2006. Reconciling the "old" and "new" views of protein allostery: a molecular simulation study of chemotaxis Y protein (CheY). *Proteins.* 63:846–867.
20. Ma, L., and Q. Cui. 2007. Activation mechanism of a signaling protein at atomic resolution from advanced computations. *J. Am. Chem. Soc.* 129:10261–10268.
21. Knaggs, M. H., F. R. Salsbury, Jr., …, J. S. Fetrow. 2007. Insights into correlated motions and long-range interactions in CheY derived from molecular dynamics simulations. *Biophys. J.* 92:2062–2079.
22. Jacobs, D. J., A. J. Rader, …, M. F. Thorpe. 2001. Protein flexibility predictions using graph theory. *Proteins.* 44:150–165.
23. Jacobs, D. J., and M. F. Thorpe. 1995. Generic rigidity percolation: the pebble game. *Phys. Rev. Lett.* 75:4051–4054.
24. Dill, K. A. 1997. Additivity principles in biochemistry. *J. Biol. Chem.* 272:701–704.
25. Mark, A. E., and W. F. van Gunsteren. 1994. Decomposition of the free energy of a system in terms of specific interactions. Implications for theoretical and experimental studies. *J. Mol. Biol.* 240:167–176.
26. Jacobs, D. J., D. R. Livesay, …, M. L. Tasayco. 2006. Elucidating quantitative stability/flexibility relationships within thioredoxin and its fragments using a distance constraint model. *J. Mol. Biol.* 358: 882–904.
27. Taketomi, H., Y. Ueda, and N. Gō. 1975. Studies on protein folding, unfolding and fluctuations by computer simulation. I. The effect of specific amino acid sequence represented by specific inter-unit interactions. *Int. J. Pept. Protein Res.* 7:445–459.
28. Dahiyat, B. I., D. B. Gordon, and S. L. Mayo. 1997. Automated design of the surface positions of protein helices. *Protein Sci.* 6:1333–1337.
29. Lukat, G. S., A. M. Stock, and J. B. Stock. 1990. Divalent metal ion binding to the CheY protein and its significance to phosphotransfer in bacterial chemotaxis. *Biochemistry.* 29:5436–5442.
30. Stock, A. M., E. Martinez-Hackert, …, G. A. Petsko. 1993. Structure of the $Mg^{2+}$-bound form of CheY and mechanism of phosphoryl transfer in bacterial chemotaxis. *Biochemistry.* 32:13375–13380.
31. Volz, K., and P. Matsumura. 1991. Crystal structure of *Escherichia coli* CheY refined at 1.7-Å resolution. *J. Biol. Chem.* 266:15511–15519.
32. Usher, K. C., A. F. de la Cruz, …, S. J. Remington. 1998. Crystal structures of CheY from *Thermotoga maritima* do not support conventional explanations for the structural basis of enhanced thermostability. *Protein Sci.* 7:403–412.
33. Gordon, J. C., J. B. Myers, …, A. Onufriev. 2005. $H^{++}$: a server for estimating $pK_a$s and adding missing hydrogens to macromolecules. *Nucleic Acids Res.* 33(Web Server issue):W368–W371.
34. Hollien, J., and S. Marqusee. 1999. A thermodynamic comparison of mesophilic and thermophilic ribonucleases H. *Biochemistry.* 38:3831–3836.
35. Wrba, A., A. Schweiger, …, P. Závodszky. 1990. Extremely thermostable D-glyceraldehyde-3-phosphate dehydrogenase from the eubacterium *Thermotoga maritima*. *Biochemistry.* 29:7584–7592.
36. Jaenicke, R., and P. Závodszky. 1990. Proteins under extreme physical conditions. *FEBS Lett.* 268:344–349.
37. Zavodszky, P., J. Kardos, …, G. A. Petsko. 1998. Adjustment of conformational flexibility is a key event in the thermal adaptation of proteins. *Proc. Natl. Acad. Sci. USA.* 95:7406–7411.
38. Royer, Jr., W. E., H. Zhu, …, J. E. Knapp. 2005. Allosteric hemoglobin assembly: diversity and similarity. *J. Biol. Chem.* 280:27477–27480.
39. Jensen, A. A., and T. A. Spalding. 2004. Allosteric modulation of G-protein coupled receptors. *Eur. J. Pharm. Sci.* 21:407–420.
40. Whitaker, R. J., G. S. Byng, …, R. A. Jensen. 1981. Comparative allostery of 3-deoxy-D-arabino-heptulosonate 7-phosphate synthetase as an indicator of taxonomic relatedness in pseudomonad genera. *J. Bacteriol.* 145:752–759.