Physical Biology

PURPOSE-LED PUBLISHING™

PAPER • **OPEN ACCESS**

Fluctuation-driven synergy, redundancy, signal to noise ratio and error correction in protein allostery

To cite this article: Burak Erman 2025 *Phys. Biol.* **22** 026003

View the article online for updates and enhancements.

You may also like

- The physical logic of protein machines
John M McBride and Tsvi Tlusty
- Nonlinear excitations match correlated motions unveiled by NMR in proteins: a new perspective on allosteric cross-talk
Francesco Piazza
- Roadmap on biology in time varying environments
Arvind Murugan, Kabir Husain, Michael J Rust et al.

This content was downloaded from IP address 88.255.99.39 on 19/08/2025 at 11:55

IOP Publishing Phys. Biol. 22 (2025) 026003 https://doi.org/10.1088/1478-3975/adb9af

Physical Biology

CrossMark

OPEN ACCESS

RECEIVED
3 January 2025

REVISED
7 February 2025

ACCEPTED FOR PUBLICATION
24 February 2025

PUBLISHED
5 March 2025

Original content from this work may be used under the terms of the Creative Commons Attribution 4.0 licence.

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.



PAPER

Fluctuation-driven synergy, redundancy, signal to noise ratio and error correction in protein allostery

Burak Erman

Department of Chemical and Biological Engineering, Koc University, Istanbul, Turkey

E-mail: berman@ku.edu.tr

Keywords: fluctuations, driven, synergy, redundancy, signal to noise ratio, error correction

Abstract

This study explores the relationship between residue fluctuations and molecular communication in proteins, emphasizing the role of these dynamics in allosteric regulation. We employ computational tools including the Gaussian network model, mutual information, and interaction information, to analyze how stochastic interactions among residues contribute to functional interactions while also introducing noise. Our approach is based on the postulate that residues experience continuous stochastic bombardment from impulses generated by their neighbors, forming a complex network characterized by small-world scaling topology. By mapping these interactions through the Kirchhoff matrix framework, we demonstrate how conserved correlations enhance signaling pathways and provide stability against noise-like fluctuations. Notably, we highlight the importance of selecting relevant eigenvalues to optimize the signal-to-noise ratio in our analyses, a topic that has yet to be thoroughly investigated in the context of residue fluctuations. This work underscores the significance of viewing proteins as adaptive information processing systems, and emphasizes the fundamental mechanisms of biological information processing. The basic idea of this paper is the following: given two interacting residues on an allosteric path, what are the contributions of the remaining residues on this interaction. This naturally leads to the concept of synergy, redundancy and noise in proteins, which we analyze in detail for three proteins CheY, tyrosine phosphatase and β-lactoglobulin.

1. Introduction

Evolution has endowed proteins with a remarkable capacity for efficient and robust nanoscale communication and information transmission. This capability is rooted in several fundamental features, including large-scale spatially correlated fluctuations of residues that facilitate dynamic interactions within the protein structure. A network of interactions, characterized by small-world scaling topology [1] with spectral characteristics in the terahertz region [2], conveniently probed through molecular dynamics simulations, establishing effective pathways for information transfer. Additionally, the synergistic effects of specific residues enhance signaling along these pathways, while redundancy in the fluctuations of certain residues provides stability and acts as a corrective mechanism to ensure reliable signal propagation. Together, these factors contribute to the overall resilience and efficiency of intramolecular signal transmission in proteins, highlighting their role and functional significance in allosteric processes [3–5]. The spatial and temporal replication of the same protein with high precision across all species clearly demonstrates that the aforementioned fundamental features are universal determinants of protein function [6].

The network topology of a protein is mathematically represented by the Kirchhoff matrix, K, a computational construct that transforms our understanding of molecular communication. Defined as $\beta = D - A$, where D is the degree matrix and A is the adjacency matrix, this framework captures protein residue connectivity. Computational validation studies have demonstrated that these matrices can map intricate molecular communication networks that traditional methods overlook [7]. The pseudoinverse of the Kirchhoff matrix, Γ^{-1}, generates the covariance

© 2025 The Author(s). Published by IOP Publishing Ltd

matrix Σ, a dynamic representation of molecular motion that gives residue fluctuation correlations [8]. Biological validation studies have shown that these matrices can predict functional mutations with up to 70% accuracy, bridging theoretical models with experimental observations [9, 10]. By employing a linear spring model, we can describe molecular movements through a multivariate Gaussian probability density function. This probabilistic approach forms the basis of information theory of proteins that we adopt here [11].

Entropy quantifies the uncertainty in residue fluctuations, which might play a role in allosteric regulation and information processing within proteins. It reflects how conformational changes at one site can influence distant regions, facilitating communication essential for protein function. This entropic interplay enables proteins to adaptively respond to stimuli, allowing for precise control over biological processes [12, 13]. Recent work by Cecconi *et al* [14] has highlighted entropy transfer as a key mechanism for directional information flow in proteins, distinguishing causal interactions from mere correlations and providing insights into how perturbations propagate through the molecular structure.

Mutual information provides a nuanced understanding of residue interactions that predicts functional relationships between residues with remarkable computational accuracy. Computational studies have shown that residues with high mutual information are often involved in critical cellular processes, such as signal transduction and enzymatic catalysis. In G-protein coupled receptors, for example, mutual information analysis has uncovered communication pathways explaining how distant mutations can dramatically alter protein function [15].

Conditional mutual information represents an advanced approach to understanding molecular communication mechanisms. This metric has been particularly powerful in explaining allosteric phenomena, such as hemoglobin's cooperative binding—where oxygen binding to one subunit induces conformational changes in distant protein regions, demonstrating the complex, interconnected nature of protein dynamics [16, 17].

Interaction information emerges as a sophisticated tool for understanding molecular relationships beyond simple pairwise interactions. Positive interaction information suggests redundancy, potentially representing backup pathways and molecular error-correction mechanisms, while negative interaction information reveals synergistic communication pathways [18].

Interactions between residues are inherently stochastic, either enhancing each other's functional integrity or contributing to noise. Each residue experiences continuous stochastic bombardment from impulses generated by all other residues, whether they are nearby or distant, depending on the connectivity of the interaction network topology. A subset of these impulses consists of evolutionarily conserved and essential correlations for the protein's function and stability, while the remaining impulses are random and noise-like [19–22]. In either case, the information-theoretic tools we employ in this paper shift our perspective and treatment of proteins from stochastic molecular structures to adaptive information processing systems, especially regarding their spectral properties. Notably, while the signal-to-noise ratio in these analyses depends strongly on the number of eigenvalues considered, highlighting the importance of selecting relevant modes for accurate predictions, the specific problem of signal-to-noise ratio related to residue fluctuations has not yet been thoroughly investigated. Recent breakthrough studies have utilized these methods to design novel proteins with enhanced stability and to develop targeted therapeutic interventions aimed at understanding the molecular basis of genetic diseases [23, 24]. By mapping intricate information networks, the present study paves the way for predicting mutation-induced communication disruptions, elucidating protein misfolding mechanisms, and advancing sophisticated protein engineering strategies [25, 26]. The implications extend far beyond basic structural biology, offering significant insights into cellular signaling, drug design, and the fundamental mechanisms of biological information processing.

In this paper, we utilize three well-studied cases of allosteric activity, CheY, tyrosine phosphatase and β-lactoglobulin. We use the Gaussian d network model, (GNM), to demonstrate the role of residue fluctuations in information transfer and to explain the strategies employed to achieve synergy and redundancy in allosteric communication.

2. Materials and methods

2.1. Allostery in proteins and metrics of information theory

The problem that we are interested in is the following: we assume that two residues, say A and B, in a protein interact, and this interaction is essential for allosteric activity. If the two residues are within the coordination shell of each other, they will interact through typical Lennard-Jones forces, including hydrogen bonds, electrostatic interactions, and steric effects. They may be spatially distant. In this case, the graph structure of the protein significantly influences their interaction [27, 28].

If a third residue C enhances the interaction between A and B, its effect is considered synergistic, forming what we term the allosteric channel. Conversely, C may provide overlapping information about the interaction between A and B that is already conveyed by either A or B. In information theory, this

overlapping information is referred to as 'redundant'. However, in the context of proteins, C can be viewed as a backup residue that supports the *A*–*B* interaction, particularly if that interaction is weakened due to mutations or dynamic changes within the protein. Therefore, we will use the term 'redundant information' to mean 'overlapping information'. The residues that provide this information will be called 'redundant residues', but not in a negative sense.

Entropy of a residue A is represented by the Shannon equation [29]

$$H(A) = -\sum p(A)\log p(A) \tag{1}$$

where, A represents the set of fluctuations of residue A, $p(A)$ is the probability and $H(A)$ quantifies the degree of uncertainty associated with those fluctuations. For correspondence between the thermodynamic entropy and the Shannon entropy, please see chapter 17 of the book by Callen [30].

Mutual information, $I(A;B)$ between A and B is defined as

$$I(A;B) = \sum_{\Delta R_A}\sum_{\Delta R_B} p(\Delta R_A, \Delta R_B)\log\frac{p(\Delta R_A, \Delta R_B)}{p(\Delta R_A)p(\Delta R_B)} \tag{2}$$

where, $p(\Delta R_A, \Delta R_B)$ is the joint probability of fluctuations of residue A and B. The summation is carried out over all fluctuations of A and B. For brevity, the notation will be simplified by replacing $p(\Delta R_A, \Delta R_B)$ with $p(A,B)$ and $p(\Delta R_A)$ with $p(A)$, etc, in the following. $I(A;B)$ quantifies the reduction in uncertainty in the fluctuations of B that is achieved by knowing the fluctuations of A. If A and B are independent, meaning that knowing A gives no information about B, the mutual information is zero. However, if there is some relationship—whether linear or non-linear—mutual information will be positive.

Joint entropy $H(A,B)$ is defined as,

$$H(A,B) = -\sum p(A,B)\log p(A,B). \tag{3}$$

Conditional entropy $H(B|A)$ is defined as,

$$H(B|A) = -\sum p(A,B)\log\frac{p(A,B)}{p(A)}. \tag{4}$$

$H(B|A)$ measures the amount of uncertainty remaining in the fluctuations of residue B when the fluctuations of residue A are known. It quantifies how much additional information is needed to describe B given that we already have information about A. When we observe the fluctuations of residue A, we may gain insights into the behavior of residue B. If knowing the state of A significantly reduces our uncertainty about B, then $H(B|A)$ will be low. This indicates that the fluctuations of A and B are correlated; when A fluctuates, B tends to fluctuate in a predictable manner.

Mutual information is related to entropy by the equation

$$I(A;B) = H(A) + H(B) - H(A,B). \tag{5}$$

Conditional mutual information in terms of entropy is

$$I(A;B|C) = H(A|C) + H(B|C) - H(A,B|C). \tag{6}$$

Conditional mutual information measures the amount of information that two random variables, A and B, share in the presence of a third variable, C. And finally, the interaction information

$$I(A;B;C) = I(A;B) - I(A;B|C) \tag{7}$$

which represents a measure of how the information shared between variables A and B changes under the effect of another variable C. If the presence of C decreases the interaction between A and B, then $I(A;B) > I(A;B|C)$ and $I(A;B;C)$ is positive. The amount of information provided by C to the interaction between A and B in this case has an overlapping component. The lowering of $I(A;B|C)$ due to the presence of C may be due to two different effects that may be operating in the system. First, the presence of C may create mechanical noise, which may affect the direct correlation between A and B. Calculations show that the noise created by neighboring residues C along the primary chain increase the interaction information between A and B. The second effect occurs when C acts similarly to A or B in their interaction. In this case, C provides overlapping or duplicate information. This causes the conditional mutual information $I(A;B|C)$ to decrease because C does not add unique information but rather repeats what A and B are already doing. This second case may indeed be useful for the protein to carry its allosteric function. It will be needed when the protein experiences mechanical stress, when one allosteric pathway is compromised, helps to ease-down negative effects of mutation, provides robustness against genetic variations, creates alternative information transmission routes, compensates for local structural fluctuations, maintains interaction integrity under perturbations, and prevents catastrophic information loss. The key point is that redundancy occurs when an additional element (C in this case) provides information that is already largely captured by the existing interaction between A and B. It is interesting to note that both redundancy and noise can lead to similar mathematical outcomes (positive interaction information), but they represent fundamentally different processes: redundancy strengthens or confirms existing relationships. Noise, on the other hand, introduces uncertainty and irrelevant information.

If the presence of C increases the interaction between A and B, which we call synergy, then $I(A;B) < I(A;B|C)$ and the interaction information is $I(A;B;C)$ is negative. By identifying synergetic residues, it becomes possible to map potential allosteric paths. Such synergetic residues are essential in understanding long-range communication mechanisms.

The interaction between A and B is subject to impulses from all other residues of the protein. The information received by A–B is $I(A;B;C)$ where C includes all residues other than A and B. The standard deviation (SD), of the information that comes to $\underline{A}$–B is representative of the information transferred by C. Given this description of the signal $I(A;B;C)$, the noise naturally comes from the SD of the conditional mutual information, SD($I(A;B|C)$) showing the variability or uncertainty in the information transfer between A and B when accounting for all other residues. This SD quantifies the fluctuations and inconsistencies in how the other residues modulate the information exchange between A and B. Higher SD indicates more unpredictable or scattered information transfer, suggesting less stable or more complex interactions between A and B in the context of the entire protein. Thus, the signal to noise ratio, S/N becomes

$$S/N = \frac{\mathrm{SD}\left[I(A;B;C)\right]}{\mathrm{SD}\left[I(A;B|C)\right]}. \tag{8}$$

The signal-to-noise ratio (S/N) strongly depends on the spectral properties of the protein, particularly with larger eigenvalues of the Kirchhoff matrix being more representative of noise, while noise-like eigenvalues contribute more significantly to the denominator of equation (8).

Intramolecular noise provides insights into the robustness and adaptability of allosteric pathways, highlighting how internal redundancy or synergy can maintain effective communication despite perturbations. This perspective aligns with the principles of information theory, where variability in pathways can serve as a basis for error correction and signal reliability.

Finally, we consider the error correction capability of C. C provides overlapping information about the interaction between A and B, where transmission errors may occur. Here, we use the term 'transmission' in its broadest sense. For example, a sporadic impulse that tends to break a necessary hydrogen bond between A and B may be regarded as a transmission error that can be masked by the information sent by C. In this context, C acts as a molecular error correction code. Alternatively, if the A-B interaction is compromised or weakened, the overlapping information provided by residue C may reduce the uncertainty introduced. This suggests a natural error correction mechanism in protein communication pathways that compensates for imperfections in primary signal transmission and restores the integrity of the signals.

2.2. The GNM and information transfer

The GNM is a coarse-grained representation of proteins, where each $C\alpha$ atom is modeled as a node in a graph, and edges connect nodes that are either covalently bonded or spatially proximate. The graph's connectivity is described by its Laplacian or Kirchoff matrix, Γ, which assigns unit entries to connected nodes and diagonal elements equal to the negative sum of each row. This Laplacian encodes the protein's local structural features, i.e. the residue interaction matrix, while its pseudoinverse is proportional to the fluctuation correlation matrix

$$\left(\Gamma_{ij}\right)^{-1} \propto \left\langle \Delta R_i \Delta R_j \right\rangle = \sum\nolimits_{k=2}^{N} {\lambda_k}^{-1} {u_i}^k {u_j}^k \tag{9}$$

where, angular brackets denote average, i and j refer to residue indices, λ_k is the kth eigenvalue of β, and ${u_i}^k$ is the kth eigenvector of Γ for the ith residue. The model approximates fluctuations as isotropic, and despite this simplicity, it remarkably aligns with experimental observations of protein dynamics [7].

The GNM is based on the Laplacian matrix, which represents local interactions as linear springs, a characteristic feature of Gaussian distributions. A key strength of GNM lies in the inversion of the Laplacian matrix, which transforms local structural information into a global interaction map. In this transformed space, nonzero entries indicate indirect coupling, capturing generic correlations rather than strictly physical interactions between neighboring residues. In the original Laplacian, residue pairs that are physically distant are represented with zero entries due to the lack of direct interactions. However, the inverse Laplacian introduces nonzero entries for these pairs, capturing the indirect interactions mediated through the network, which automatically contributes to the mutual information between distant residue pairs. These nonzero values reflect dynamic couplings between residues, offering insight into long-range relationships within the protein. This capability to account for allosteric interactions in a single computational step distinguishes GNM from other methods.

In addition to providing structural insights, GNM captures protein dynamics by analyzing the eigenmodes of motion derived from the Laplacian [8]. These eigenmodes explain collective movements and pathways that underlie functional processes such as allosteric signaling. The Laplacian has a zero eigenvalue, which corresponds to rigid body motion. Smaller eigenvalues characterize slow and large scale motions and larger eigenvalues correspond to fast motions that are more noise-like. By integrating both structural and dynamic information, GNM offers a computationally efficient framework for studying information transfer and long-range communication

within proteins, making it the simplest yet powerful approach for such analyses.

Entropy for n variables for a multivariate Gaussian distribution of fluctuations reads as

$$H(\Delta R_1, \Delta R_2, \ldots, \Delta R_n) = \frac{1}{2}\ln\left(\det\langle \Delta R \Delta R^{\mathrm{T}}\rangle_n\right) + \frac{n}{2}\ln(2\pi e). \quad (10)$$

Expressions for mutual information, conditional mutual information and interaction information are obtained from the entropy using equations (3)–(7).

Substituting from equation (10), which works for a Gaussian system, leads to the following [31]

$$I(A;B) = -\frac{1}{2}\ln\left[1 - \frac{\langle \Delta R_A \cdot \Delta R_B\rangle^2}{\left\langle (\Delta R_A)^2\right\rangle \left\langle (\Delta R_B)^2\right\rangle}\right] \quad (11)$$

$$\begin{aligned} I(A;B|C) = &-\frac{1}{2}\ln\left(\left\langle (\Delta R_C)^2\right\rangle\right) + \frac{1}{2}\ln\left(\left\langle (\Delta R_A)^2\right\rangle \left\langle (\Delta R_C)^2\right\rangle - \langle \Delta R_A \cdot \Delta R_C\rangle^2\right) \\ &+ \frac{1}{2}\ln\left(\left\langle (\Delta R_B)^2\right\rangle \left\langle (\Delta R_C)^2\right\rangle - \langle \Delta R_B \cdot \Delta R_C\rangle^2\right) \\ &- \frac{1}{2}\ln\left[\left\langle (\Delta R_A)^2\right\rangle \left\langle (\Delta R_B)^2\right\rangle \left\langle (\Delta R_C)^2\right\rangle + 2\langle \Delta R_A \cdot \Delta R_B\rangle \langle \Delta R_A \cdot \Delta R_C\rangle \langle \Delta R_B \cdot \Delta R_C\rangle \right. \\ &\left. - \langle \Delta R_A \cdot \Delta R_C\rangle^2 \left\langle (\Delta R_B)^2\right\rangle - \langle \Delta R_A \cdot \Delta R_B\rangle^2 \left\langle (\Delta R_C)^2\right\rangle - \langle \Delta R_B \cdot \Delta R_C\rangle^2 \left\langle (\Delta R_A)^2\right\rangle\right]. \end{aligned} \quad (12)$$

The derivation of equations (11) and (12) are explained in [18] and the supplementary information section of that reference. All entries in equations (11) and (12) follow from the inverse of the Laplacian and their use in the interaction information expression, equation (9). Scanning over the protein by keeping A and B fixed and changing C leads to the interaction information profile for the protein in which positive peaks show the backup residues that provide overlapping information and negative peaks show the synergistic residues for the interaction between residues A and B.

When C equals A or B, then the right hand side of equation (7) is undefined, but since $I(A;B|B) = 0$ we have

$$\underset{C=B}{I(A;B;C)} = I(A;B). \quad (13)$$

3. Results

3.1. CheY bacterial chemotaxis

CheY is a signaling protein in the bacterial chemotaxis signaling pathway. When phosphorylated at D57, the protein travels to a large flagellar motor complex and binds to FliM, which is part of the motor complex. When CheY binds to FliM, it causes the switch complex to change its rotation direction. The primary effect is the increased interaction of D57 with T87, enabling CheY to bind to FliM. The interaction is in the form of a hydrogen bond which forms when D57 is phosphorylated. It helps maintain the allosteric activity effectively. Thus, the allosteric mechanism of CheY occurs through changes in its dynamical properties upon the interaction of D57 with T87 rather than a simple, mechanistic structural rearrangement. The problem we investigate here is the following: given the interaction of D57 and T87, how do the remaining residues of the protein affect this interaction. We are particularly interested in the remaining residues of the protein that enhance this interaction or serve as a backup node for the maintenance of this interaction. We also investigate the level of information and noise coming to this interaction.

3.1.1. Determining the synergistic and overlapping residues controlling the interaction between D57 and T87

The interaction information profile for CheY, calculated according to equations (7)–(12) for $I(57;87;C)$ where C represents each residue scanned across the entire protein is presented in figure 1. The negative synergistic and positive redundant peaks in the figure are supported by experimental observations [32, 33].

The conserved D12 and D13 residues, though located on the opposite side of the protein from D57 and the FliM binding site, i.e. residues A90, K91, K92, G105, Y106, and K109, contribute redundant information to the D57-T87 interaction and act as backup nodes. This means that they can compensate for the primary interaction pathway, can help maintain the functional relationship between D57 and T87 even if the original interaction is compromised and can provide an alternative route for information, or signal transmission. Both D12 and D13 coordinate the binding of the Mg(II) ion, which is necessary for the activation of CheY. This ion is essential for stabilizing the structure of CheY and facilitating its interaction with other proteins during allosteric activity [35].

The first ten residues of the protein act synergistically on the interaction of D57 and T87 and

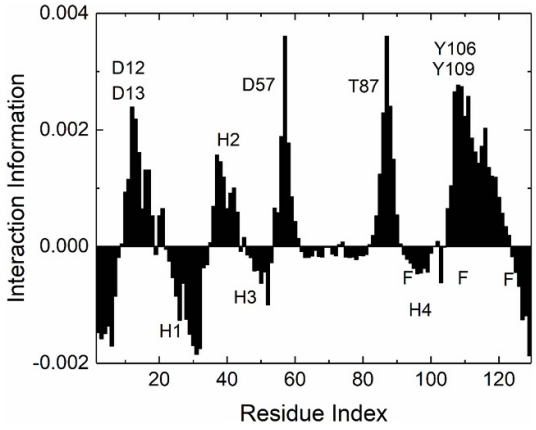


Figure 1. Interaction information $I(57;87;C)$ where C is scanned over all residues of the protein shown along the abscissa. The Protein Data Bank structure 1F4V is used in the calculations with a cutoff distance of 7.3 Å in obtaining the Γ matrix. The proportionality constant for equation (9) is taken as unity. All modes are used in the negative peaks show the synergistic contribution of the corresponding residue to the interaction between D57 and T87. Positive peaks correspond to residues that provide overlapping information to this interaction. D12, D13, D57, T87 and Y106 and K109 are conserved residues [34]. Binding sites of Film are indicated by F. Helices are indicated by *H*. Interaction information values are obtained keeping all modes of the Kirchoff matrix.

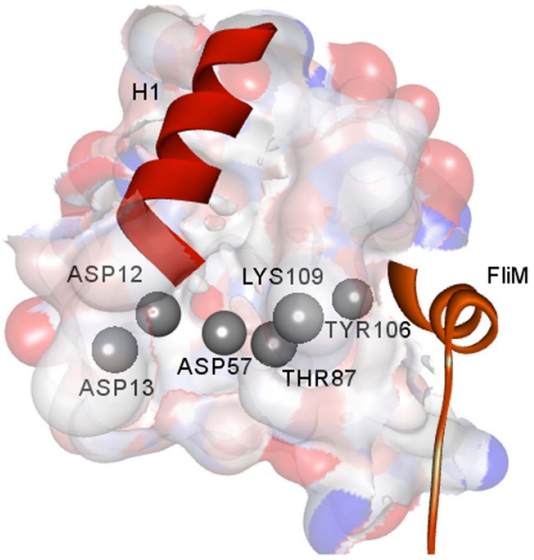


Figure 2. The allosteric path obtained from peaks of figure 1. Alpha carbons of the six conserved residues are shown with gray spheres. Allosteric activity starts at D57 when it is phosphorylated by the protein CheA. This is followed by the rearrangements of the conformations of the residues shown such that FliM becomes able to bind to CheY as a result of which the flagellar motor attached to FliM is activated. Helix1 and the conserved residues shown are derived from the peaks of the interaction information profile $I(57;87;C)$, which is calculated for each residue of CheY.

are critical for establishing the overall structure of CheY. They help define the protein's conformation, which is essential for its interaction with other proteins and for transmitting signals upon phosphorylation. Phosphorylation of D57 is a conditional requirement for changes in the conformations of the first ten residues, modulating CheY's interaction with the flagellar motor. Mutations in these residues are shown to lead to altered conformations that affect CheY's ability to activate flagellar rotation [35, 36].

The first helix, H1, shown in figure 2, is the strongest synergistic region strengthening the interaction between D57 and T87. The formation of hydrogen bonds between residues in helix 1 and the neighboring beta strand contributes to the overall stability of CheY which is the source of communication between H1 and D57 [37]. According to figure 1, the second helix, H2, acts redundantly and the third helix, H3, acts synergistically on the interaction between D57 and T87. Previous work [31, 38] identified the importance of the helices H1–H3 in allosteric communication. The present model is able to differentiate the mechanism of action between these helices as to whether they are synergistic or redundant. Synergy amplifies effects and leads to emergent properties in signaling pathways, while redundancy tends to maintain baseline functionality despite perturbations.

Interaction information between D57 and T87 is represented by the two vertical lines at positions 57 and 87 in figure 1. The positive peaks in the neighborhood of D57 and T87 show that these residues provide overlapping information to the 57–87 interaction.

The nature of the information transmitted from D57 to FliM can be visualized in three distinct forms:

(a) Biochemical state information: phosphorylation at D57 alters the charge and chemical properties of this residue. This modification serves as a biochemical signal that switches CheY from an inactive to an active state.
(b) Conformational information: the phosphorylation event induces a conformational shift in the structure of CheY, particularly in residues surrounding the phosphorylation site, including D12 and D13. These residues play a critical role in stabilizing the phosphorylated state and facilitating the propagation of the activation signal to CheY's FliM-binding interface, as illustrated by F in figure 1. This process effectively 'transmits' information about CheY's activation state.
(c) Functional information: the final 'output' of this information transfer is CheY's ability to bind FliM. The phosphorylation and subsequent conformational changes, supported by D12 and D13, create a high-affinity binding interface for FliM, determining whether CheY can influence the rotational state of the flagellar motor (clockwise or counterclockwise).

Explaining the allosteric activity of CheY in terms of information theory adds a new dimension to our previous allostery analysis of this system [11].

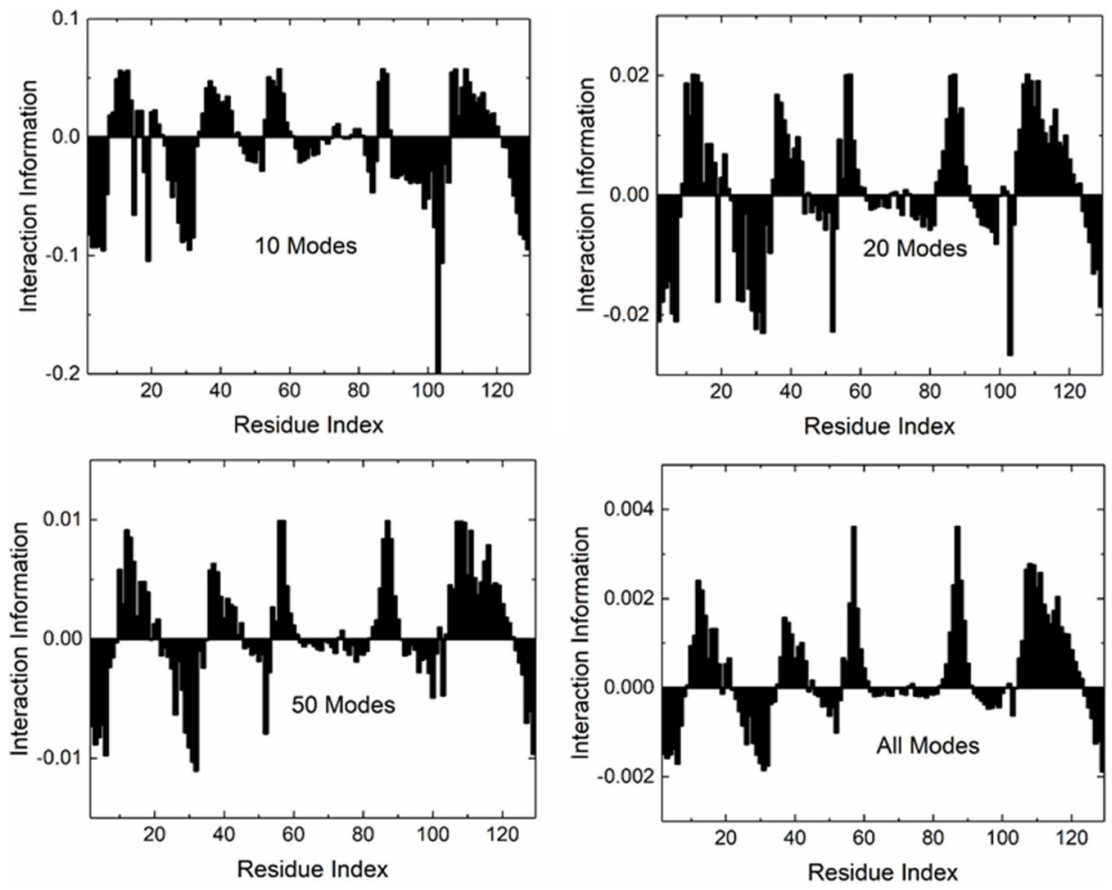


Figure 3. Interaction information profiles obtained by using different number of modes in equation (9). The number of modes used is shown in each panel.

3.1.2. Effect of noise on information transmission in CheY and the signal to noise ratio, S/N

The effect of a third residue on the interaction between a given pair of residues depends strongly on the number of modes of equation (9) used in the analysis. As the number of modes increases, the magnitudes of the interaction information profiles tend to decrease significantly because additional modes contribute predominantly to noise rather than meaningful interactions. This dilution of signal obscures the significance of the main interactions in the allosteric process. However, the shape of the profile remains approximately independent of the number of modes used, showing that residues contributing to synergy and redundancy are identified irrespective of the number of modes. In figure 3, profiles are shown for four different cases where 10, 20, 50 and all modes are used in calculations. One sees that while the general shape remains approximately the same for the four different cases, the amplitudes drop down significantly as the number of modes becomes larger.

The question of 'how many modes should be used for a representative case' can be answered by calculating the signal-to-noise ratio S/N as a function of the number of modes, as shown in the following figure.

Figure 4 illustrates that the use of the lowest eight modes results in signal-to-noise ratio S/N values greater than 0.5, which can be accepted as contributing to the allosteric signal affecting the interactions between residues D57 and T87. It is reasonable to conclude that higher modes correspond to such small amplitudes that they collectively contribute to noise in the context of information propagation. An order of magnitude analysis of the mean-square fluctuations of CheY residues, along with their decomposition into modes and comparison with experimental B-factors, indicates that the first ten modes are associated with displacements greater than approximately 0.1 angstroms, while higher modes correspond to displacements below this threshold. It is interesting to note that the plot of S/N versus the number of modes exhibits fluctuations as the number of modes increases. These fluctuations indicate that using n modes yields a high information value, while adding an additional mode $(n + 1)$ decreases the S/N. Detailed calculations show that the numerastor SD $[I(A;B;C)]$ in equation (8) decreases monotonically with the number of modes, whereas the denominator SD $[I(A;B|C)]$ fluctuates, resulting in the variations observed in figure 4.

The analysis of the signal-to-noise ratio (S/R) shows that when only the first 10 or 20 modes are considered, the S/R remains above or close to unity. Beyond this range, the S/R decreases significantly,

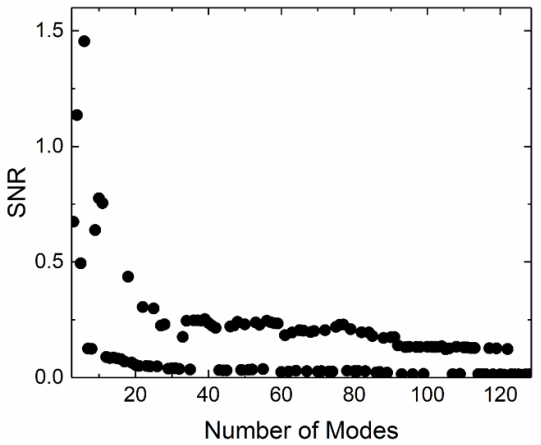


Figure 4. *S*/*N* graph for CheY. The points are obtained using equations (11) and (12) for the total number of modes used, shown along the abscissa. See legend to figure 1 for details of calculation.

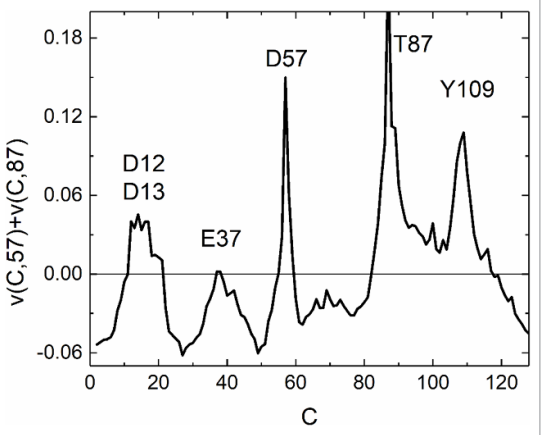


Figure 5. Sum of correlations of *C* with D57 and T87 for different values of residue *C*. The correlations are obtained from equation (9). See legend for figure 1 for parameters used.

indicating that higher-order modes primarily contribute to noise rather than meaningful correlations. This suggests that the first 10–20 modes capture the most relevant structural correlations, while higher modes introduce fluctuations that obscure the signal.

Figure 3 presents Interaction Information profiles for different mode selections (10, 20, 50, and all modes). Notably, for 10 and 20 modes, the relative heights of positive interaction information peaks remain approximately the same, indicating that the ranking of backup sites remains stable within this range. However, as more modes are included (50 modes and beyond), fluctuations introduced by higher-order modes alter peak magnitudes, potentially affecting rankings.

These observations suggest that selecting 10–20 modes provides a reasonable balance, capturing essential correlations while minimizing noise-driven distortions. A more rigorous criterion for determining the optimal number of modes could further refine this analysis in future studies.

3.1.3. Error correction by redundant residues

Finally, in a rather heuristic manner, we discuss a hypothetical but plausible error correction mechanism in CheY, facilitated by redundant residues that support the interaction between residues D57 and T87. These redundant residues provide overlapping information that can stabilize the interaction between D57 and T87, which may be perturbed by structural changes such as mutations or ligand binding. A residue, denoted as *C*, can help maintain the interaction by compensating for any loss of stability when it fluctuates in a manner that is positively correlated with D57 and/or T87. This means that when D57 and T87 are experiencing unfavorable conditions, residue *C* should ideally be in a state that supports their interaction. A strong positive correlation between *C* and these residues can help preserve the necessary interactions or structural integrity despite perturbations caused by mutations or other factors.

In CheY, there are three redundant peaks at residues D12 or D13, E37, and Y109. In figure 5 we present the sum $v(C,57)+v(C,87)$ of covariance of residue *C* with both D57 and T87, where the covariances are calculated for different residues *C* across the entire protein. Notably, we observe that the two redundant peaks, D12/D13 and Y109, exhibit strong positive correlations with both D57 and T87. In contrast, the third redundant peak, E37, which is a smaller redundant peak, does not show a significant positive correlation with D57 and T87.

The concept of redundancy in this context does not necessarily imply a functional backup mechanism in the strictest sense, but rather that certain residues provide overlapping information about the interaction between D57 and T87. This overlap suggests that if perturbations (such as mutations or ligand binding) disrupt the D57–T87 interaction, these redundant residues may still carry correlated signals that help stabilize or modulate the interaction. The strong positive correlation of D12/D13 and Y109 with both D57 and T87 suggests that these residues participate in a network of cooperative interactions, which may help maintain structural integrity under perturbations.

While our data show that residues D12/D13 and Y109 exhibit strong correlations with D57 and T87, this does not necessarily mean they function as direct backup sites in a strict error correction sense. Instead, their high positive correlations suggest they contribute to a cooperative network that reinforces the stability of D57–T87 interactions. This may be due to electrostatic effects, hydrogen bonding networks, or shared structural roles within the fold of CheY. If mutations or ligand binding perturb the primary D57–T87 interaction, the influence of these

redundant residues could still help preserve aspects of the interaction through indirect coupling.

It is worth noting that the idea of redundancy as a backup mechanism is a heuristic interpretation rather than a direct functional claim. However, the observed positive correlations suggest that these residues contribute in a distributed manner to the interaction stability of D57 and T87. A more rigorous test of this hypothesis would require perturbation studies, such as mutational analysis or simulations, to determine whether disrupting these residues leads to greater instability of the D57–T87 interaction. While such an analysis is beyond the scope of this work, it may be interesting to explore this further.

3.2. Tyrosine phosphatase

Tyrosine phosphatase 1B (PTP1B) is a negative regulator of insulin receptor phosphorylation and signaling which is a therapeutic target for type 2 diabetes [39]. Its allosteric activity is described in detail [40].

Two residues whose interaction is crucial for allosteric activity of PTPB1 are W291 and L192. In this section, we find residues of PTP1B that effect this interaction. PTP1B has eight helices. W291 is located within helix H7 and is crucial for the allosteric modulation of PTP1B. When an allosteric inhibitor binds to the enzyme around H7, it can displace W291, which disrupts the hydrogen bond network essential for the closure of the WPD loop (residues 179–181) necessary for catalytic activity. This displacement keeps the WPD loop in an 'open' conformation, thereby inhibiting the enzyme's ability to catalyze dephosphorylation reactions [41].

Leu192 is situated in helix H3 and plays a significant role in stabilizing interactions with allosteric inhibitors. Its position helps maintain the protein remain in its inactive conformation, preventing the transition of PTP1B to its active form.

Results of calculations using equations (11) and (12) show that PTPB1 consists mostly of backup residues, as presented in figure 6. Helix 2 is known to connect to the allosteric site, and its conformation is linked to the movement of the WPD loop keeping the WPD loop in an open state upon binding of an allosteric inhibitor. Figure 6 shows that this effect takes place by the synergistic action of H2 on W291 and L192.

Q127 is in the catalytic site, which also has synergistic effect on the interaction of W291 and L192. In a similar way, the catalytic residue C215 has a synergistic effect on the W291 and L192 interaction.

Signal to noise ratio for PTPB1 is presented in figure 7 which shows that first 60 modes give S/N values larger than 1.

3.3. Bovine β-lactoglobulin

β-lactoglobulin is involved in transmitting allosteric signals especially in the context of biosensors as shown by the experimental work of Yang *et al* [42].

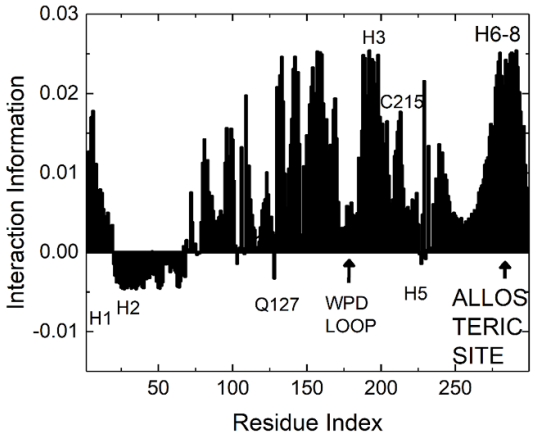


Figure 6. Interaction information profile for PTPB1 for the interaction between W291 and L192. The points are calculated using 60 modes of equation (9) (Se explanation in figure 7). See legend to figure 1 for details of calculations.

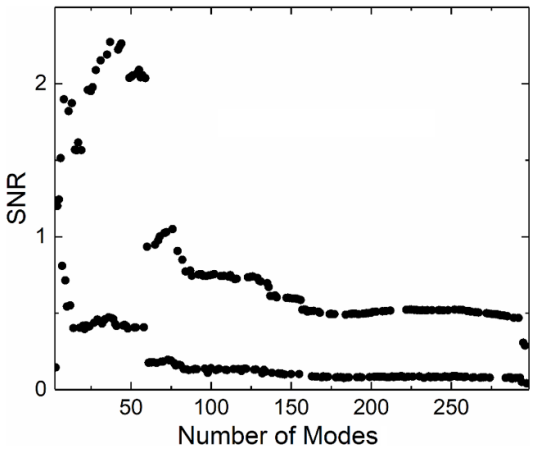


Figure 7. S/N values for PTP1B.

Changes in the environment around W19, such as ligand binding such as fatty acids or retinol, influence the conformation of E108 and facilitate allosteric communication within the protein. Changes in conformation of E108 either enhances or inhibits the protein's functionality as a transporter. In this subsection, we search for residues of that affect the interaction between W19 and E108.

In figure 8 the asterisks denote the conserved residues [43–45] which act as complementary or backup residues that are on the path between W19 and E108. R124 and F136 are two residues, synergistic and redundant, respectively, that affect the ligand binding features of β-lactoglobulin.

S/N values for β-lactoglobulin, presented in figure 9, are significantly different from those in the two previous cases, with ratios exceeding unity even for large eigenvalues. Additionally, the fluctuations in S/N values with alternating successive eigenvalues are substantial. When the signal remains large irrespective of the number of eigenvalues considered,

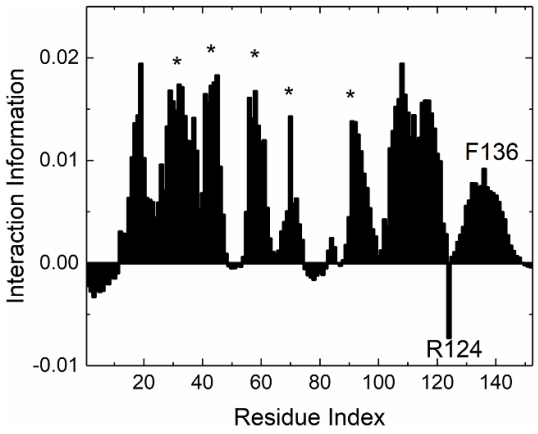


Figure 8. Interaction information profile for bovine β-lactoglobulin. The conserved residues denoted by asterisks are backup residues for the W19-E108 interaction.

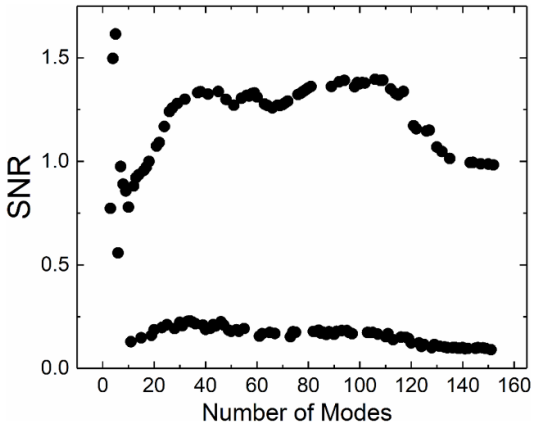


Figure 9. S/N values for bovine β-lactoglobulin.

it indicates a robust system with strong signal characteristics that effectively dominate the allosteric communication.

4. Discussion

The findings presented in this paper contribute significantly to our understanding of protein allostery by explaining the relationships between residue fluctuations, molecular communication, and the roles of synergy and redundancy. By employing the coarse-grained GNM and various information-theoretic metrics, we have demonstrated how these dynamics influence functional interactions within proteins.

Significance of residue fluctuations: our analysis reveals that residue fluctuations are not merely random events but are integral to the protein's ability to communicate effectively. The continuous stochastic bombardment from neighboring residues creates a complex network that facilitates both signal transmission and noise introduction. This perspective shifts the traditional view of proteins from static structures to adaptive systems capable of processing information dynamically. Understanding these fluctuations can lead to insights into how proteins maintain functionality under varying conditions, including mutations and environmental stresses.

Synergy and redundancy in allosteric communication: the concepts of synergy and redundancy are pivotal in our exploration of allosteric mechanisms. Synergistic interactions enhance the communication pathways between residues, while redundant interactions provide a backup that ensures stability and reliability in signal propagation. This duality is crucial for maintaining the integrity of protein function, especially in scenarios where primary interactions may be compromised. Our findings indicate that residues contributing to redundancy should not be viewed negatively; rather, they play a vital role in ensuring robustness against perturbations.

Signal-to-noise ratio optimization: a key aspect of our study is the emphasis on optimizing the signal-to-noise ratio (S/N) through careful selection of relevant eigenvalues from the Kirchhoff matrix. This approach highlights an often-overlooked area in protein dynamics research—the need to balance effective communication with noise management. By focusing on specific spectral properties, we can enhance our predictive capabilities regarding how mutations or external factors may disrupt communication pathways.

Implications for protein engineering and therapeutics: the insights gained from this study extend beyond theoretical implications; they offer practical applications in protein engineering and therapeutic development. By mapping intricate information networks within proteins, we can identify potential targets for intervention in diseases caused by misfolding or dysfunctional signaling pathways. Understanding how redundancy and synergy operate at a molecular level can guide the design of novel proteins with enhanced stability and functionality.

Future research directions: future research should aim to explore the interplay between residue fluctuations and allosteric regulation further. Investigating how different environmental conditions affect these dynamics could provide deeper insights into protein adaptability. Additionally, expanding this framework to include more diverse protein systems may reveal universal principles governing molecular communication across various biological contexts.

5. Conclusion

In conclusion, this study underscores the importance of viewing proteins as adaptive information processing systems rather than static entities. By integrating concepts from information theory with molecular dynamics, we pave the way for a more comprehensive understanding of protein functionality and its implications for biological processes. The findings not only advance our theoretical knowledge but also open avenues for practical applications in drug design and synthetic biology, emphasizing the critical role of residue interactions in maintaining cellular homeostasis and facilitating robust signaling mechanisms. Future work will focus on expanding these insights to encompass a broader range of proteins, further elucidating the principles governing allosteric regulation and molecular communication.

Data availability statement

No new data were created or analysed in this study.

ORCID iD

Burak Erman https://orcid.org/0000-0002-2496-6059

References

[1] Atilgan A R, Akan P and Baysal C 2004 Small-world communication of residues and significance for protein dynamics *Biophys. J.* **86** 85–91
[2] Elayan H, Eckford A W and Adve R 2022 Selectivity of protein interactions stimulated by Terahertz signals *IEEE Trans. Nanobiosci.* **22** 318–28
[3] Yang H, Dong Y, Bian Y, Huo C, Zhu C, Qin T, Chen S, Peng D and Liu X 2023 The synergistic effect of residues 32T and 550L in the PA protein of H5 subtype avian influenza virus contributes to viral pathogenicity in mice *PLoS Pathogens* **19** e1011489
[4] Hamborg L, Granata D, Olsen J G, Roche J V, Pedersen L E, Nielsen A T, Lindorff-Larsen K and Teilum K 2021 Synergistic stabilization of a double mutant in chymotrypsin inhibitor 2 from a library screen in E. coli *Commun. Biol.* **4** 980
[5] Blacklock K and Verkhivker G M 2014 Computational modeling of allosteric regulation in the hsp90 chaperones: a statistical ensemble analysis of protein structure networks and allosteric communications *PLoS Comput. Biol.* **10** e1003679
[6] Tang Q-Y and Kaneko K 2021 Dynamics-evolution correspondence in protein structures *Phys. Rev. Lett.* **127** 098103
[7] Bahar I, Atilgan A R and Erman B 1997 Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential *Fold. Des.* **2** 173–81
[8] Haliloglu T, Bahar I and Erman B 1997 Gaussian dynamics of folded proteins *Phys. Rev. Lett.* **79** 3090–3
[9] Mishra S K 2022 PSP-GNM: predicting protein stability changes upon point mutations with a gaussian network model *Int. J. Mol. Sci.* **23** 10711
[10] Zheng W 2024 Predicting hotspots for disease-causing single nucleotide variants using sequences-based coevolution, network analysis, and machine learning *PLoS One* **19** e0302504
[11] Hacisuleyman A and Erman B 2022 Information flow and allosteric communication in proteins *J. Chem. Phys.* **156** 185101
[12] Hacisuleyman A and Erman B 2017 Causality, transfer entropy, and allosteric communication landscapes in proteins with harmonic interactions *Prot. Struct. Funct. Bioinf.* **85** 1056–64
[13] Cooper A and Dryden D 1984 Allostery without conformational change *Eur. Biophys. J.* **11** 103–9
[14] Cecconi F, Costantini G, Guardiani C, Baldovin M and Vulpiani A 2023 Correlation, response and entropy approaches to allosteric behaviors: a critical comparison on the ubiquitin case *Phys. Biol.* **20** 056002
[15] Demirtaş K, Erman B and Haliloğlu T 2024 Dynamic correlations: exact and approximate methods for mutual information *Bioinformatics* **40** btae076
[16] Daura X 2019 Advances in the computational identification of allosteric sites and pathways in proteins *Protein Allostery in Drug Discovery* (Springer) pp 141–69
[17] LeVine M V and Weinstein H 2014 NbIT-a new information theory-based analysis of allosteric mechanisms reveals residues that underlie function in the leucine transporter LeuT *PLoS Comput. Biol.* **10** e1003603
[18] Hacisuleyman A and Erman B 2024 Synergy and anti-cooperativity in allostery: molecular dynamics study of WT and oncogenic KRAS-RGL1 *Prot. Struct. Funct Bioinf.* **92** 665–78
[19] Ormos P 2008 Dynamic fluctuation of proteins watched in real time (https://doi.org/10.2976/1.2978985)
[20] Zheng W and Tekpinar M 2009 Large-scale evaluation of dynamically important residues in proteins predicted by the perturbation analysis of a coarse-grained elastic model *BMC Struct. Biol.* **9** 1–17
[21] Erkip A and Erman B 2004 Dynamics of large-scale fluctuations in native proteins. Analysis based on harmonic inter-residue potentials and random external noise *Polymer* **45** 641–8
[22] Tang Q-Y, Hatakeyama T S and Kaneko K 2020 Functional sensitivity and mutational robustness of proteins *Phys. Rev. Res.* **2** 033452
[23] Leaver-Fay A, Tyka M, Lewis S M, Lange O F, Thompson J, Jacak R, Kaufman K W, Renfrew P D, Smith C A and Sheffler W 2011 *Methods in Enzymology* (Elsevier) pp 545–74
[24] Dantas G, Kuhlman B, Callender D, Wong M and Baker D 2003 A large scale test of computational protein design: folding and stability of nine completely redesigned globular proteins *J. Mol. Biol.* **332** 449–60
[25] Hollmann F, Sanchis J and Reetz M T 2024 Learning from protein engineering by deconvolution of multi-mutational variants *Angew. Chem., Int. Ed.* **63** e202404880
[26] Loutchko D and Flechsig H 2020 Allosteric communication in molecular machines via information exchange: what can be learned from dynamical modeling *Biophys. Rev.* **12** 443–52
[27] Clementel D, Del Conte A, Monzon A M, Camagni G F, Minervini G, Piovesan D and Tosatto S C 2022 RING 3.0: fast generation of probabilistic residue interaction networks from structural ensembles *Nucleic Acids Res.* **50** W651–W6
[28] Del Sol A, Fujihashi H, Amoros D and Nussinov R 2006 Residues crucial for maintaining short paths in network communication mediate signaling in proteins *Mol. Syst. Biol.* **2** 2006–0019
[29] Cover T M 1999 *Elements of Information Theory* (Wiley)
[30] Callen H B 1985 *Thermodynamics and an Introduction to Thermostatistics* (Wiley)
[31] Haliloglu T, Hacisuleyman A and Erman B 2022 Prediction of allosteric communication pathways in proteins *Bioinformatics* **38** 3590–9
[32] McDonald L R, Boyer J A and Lee A L 2012 Segmental motions, not a two-state concerted switch, underlie allostery in CheY *Structure* **20** 1363–73
[33] Cho H S, Lee S-Y, Yan D, Pan X, Parkinson J S, Kustu S, Wemmer D E and Pelton J G 2000 NMR structure of activated CheY *J. Mol. Biol.* **297** 543–51

[34] Lowry D F, Roth A F, Rupert P B, Dahlquist F W, Moy F J, Domaille P J and Matsumura P 1994 Signal transduction in chemotaxis. A propagating conformation change upon phosphorylation of CheY *J. Biol. Chem.* **269** 26358–62
[35] Bourret R B, Drake S K, Chervitz S A, Simon M I and Falke J J 1993 Activation of the phosphosignaling protein CheY. II. Analysis of activated mutants by 19F NMR and protein engineering *J. Biol. Chem.* **268** 13089–96
[36] Quax T E, Altegoer F, Rossi F, Li Z, Rodriguez-Franco M, Kraus F, Bange G and Albers S-V 2018 Structure and function of the archaeal response regulator CheY *Proc. Natl Acad. Sci.* **115** E1259–E68
[37] Wilcock D, Pisabarro M T, López-Hernandez E, Serrano L and Coll M 1998 Structure analysis of two CheY mutants: importance of the hydrogen-bond contribution to protein stability *Acta Cryst.* D **54** 378–85
[38] Wang J, Jain A, McDonald L R, Gambogi C, Lee A L and Dokholyan N V 2020 Mapping allosteric communications within individual proteins *Nat. Commun.* **11** 3862
[39] Sharma S, Ebrahim A and Keedy D A 2023 Room-temperature serial synchrotron crystallography of the human phosphatase PTP1B *Acta Cryst.* F **79** 23–30
[40] Hardy J A and Wells J A 2004 Searching for new allosteric sites in enzymes *Curr. Opin. Struct. Biol.* **14** 706–15
[41] Li S, Zhang J, Lu S, Huang W, Geng L, Shen Q and Zhang J 2014 The mechanism of allosteric inhibition of protein tyrosine phosphatase 1B *PLoS One* **9** e97668
[42] Yang H, Zhu L, Wang X, Kang S, Li T, Chen K, Dong Y and Xu W 2024 A label-free fluorescent magnetic dual-aptasensor based on aptamer allosteric regulation of β-lactoglobulin *Talanta* **271** 125664
[43] Ragona L, Pusterla F, Zetta L, Monaco H L and Molinari H 1997 Identification of a conserved hydrophobic cluster in partially folded bovine β-lactoglobulin at pH 2 *Folding Des.* **2** 281–90
[44] North A 1989 Three-dimensional arrangement of conserved amino acid residues in a superfamily of specific ligand-binding proteins *Int. J. Biol. Macromol.* **11** 56–58
[45] Katakura Y, Totsuka M, Ametani A and Kaminogawa S 1994 Tryptophan-19 of β-lactoglobulin, the only residue completely conserved in the lipocalin superfamily, is not essential for binding retinol, but relevant to stabilizing bound retinol and maintaining its structure *Biochim. Biophys. Acta* **1207** 58–67