bioRxiv preprint doi: https://doi.org/10.1101/2024.01.15.575718; this version posted January 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Dynamically driven correlations in elastic net models reveal sequence of events and causality in proteins

Albert Erkip[a] and Burak Erman[b]

[a]Faculty of Engineering and Natural Sciences, Sabanci University, Orhanli 34956 Tuzla, Istanbul, Turkey

[b]Department of Chemical and Biological Engineering, Koc University, Rumelifeneri Yolu, Sariyer 34450 Istanbul, Turkey

Running Title: Dynamically driven correlations, sequence of events and causality in proteins

Keywords: Protein dynamics, allosteric regulation, causality, Gaussian Network Model, information flow, Cyclophilin A, Circadian rhythm.

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.15.575718; this version posted January 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Abstract: Protein dynamics orchestrate allosteric regulation, but elucidating the sequence of events and causal relationships within these intricate processes remains challenging. We introduce the Dynamically Perturbed Gaussian Network Model (DP-GNM), a novel approach that uncovers the directionality of information flow within proteins. DP-GNM leverages time-dependent correlations to achieve two goals: identifying driver and driven residues and revealing communities of residues exhibiting synchronized dynamics. Applied to wild type and mutated structures of Cyclophilin A, DP-GNM unveils a hierarchical network of information flow, where key residues initiate conformational changes that propagate through the protein in a directed manner. This directional causality illuminates the intricate relationship between protein dynamics and allosteric regulation, providing valuable insights into protein function and potential avenues for drug design. Furthermore, DP-GNM's potential to elucidate dynamics under periodic perturbations like the circadian rhythm suggests its broad applicability in understanding complex biological processes governed by environmental cycles.

Introduction

Proteins are dynamic and complex molecular systems that carry out their diverse functions through orchestrated thermal fluctuations. These fluctuations are not random but rather organized and consequential, enabling proteins to adopt conformational flexibility and engage in intricate functional interactions with other molecules. Proteins constantly experience perturbations, both intrinsic and extrinsic, arising from both stochastic and deterministic sources. Intrinsic perturbations include mutations and post-translational modifications, while extrinsic perturbations stem from ligand or protein binding events. Both types of perturbations modulate the protein's energy landscape, resulting in alterations to its dynamics and equilibrium properties. Static effects of perturbations in proteins are widely studied [1, 2]. The present work is on the dynamic allosteric effects of perturbations in proteins, based on the Langevin equation solution of the dynamics of a Gaussian Network Model (GNM)[3]. The dynamic version of the GNM that we use here was introduced[4] in 2004, where a single-frequency cosine perturbation was applied to a specific mode of the GNM and the response showed significant synchronicity between different modes. In 2021, the modal cosine perturbation approach was extended[5] to perturbation of single residues, allowing one to probe the dynamic response of the entire protein upon perturbing a specific residue. This mimics perturbation-response dynamics observed in allosteric regulation. The present paper introduces the 'Dynamically Perturbed GNM Model' (DP-GNM), which builds upon the two previous models where we now focus on the causal relationships introduced into

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.15.575718; this version posted January 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

the dynamics. Causality, defined by the difference in influence between residue i on j and j on i over time, is a critical aspect of protein dynamics as it forms the foundation for information transfer and allostery. Recently, Scaramozzino et al.[6, 7] extended the Langevin equation solution to another elastic network model, the Anisotropic Network Model (ANM), and analyzed protein fluctuations in response to random external forces, validating our previous results for the GNM.

An essential aspect of causality is the sequence of events or the directionality of allostery. A diverse array of methods has emerged to determine the key residues involved in allosteric modulation[8-17], only few computational papers elaborated on the sequence of events and causality in this process[18-22]. Interestingly, recent experimental work[23-26] clearly shows the importance of causality in dynamic allostery.

We show that by externally perturbing a protein, and calculating time correlation function of residue fluctuations, we can identify the residues that exert a stronger influence on the response of other residues, i.e. we can determine the direction of information flow. An extreme analogy of directional information flow would be the electric switch and the lightbulb where on/off information flows from the switch to the lightbulb but no information flows from the lightbulb to the switch. Direction of information flow cannot be determined with the use of the static picture of proteins. The 'driver-driven' terminology aptly captures the directional nature of information flow in proteins, with drivers acting as switches and driven residues receiving the signals.

The DP-GNM, which we introduce in this study focusses on phase differences between fluctuations of residues, hence it becomes possible to answer the question ‘Which residues move in phase with the fluctuations of the perturbed residue?’ The answer to this question helps in the clear identification of communities and networks of allosteric residues in the protein.

Crucially, the directionality identified by DP-GNM is not an inherent property of the unperturbed protein, but rather emerges as a consequence of the applied perturbation and its interaction with the protein's dynamic landscape. Moreover, this directionality exhibits a periodic reversal between successive time intervals of $(n-1)\pi < \omega\tau < n\pi, \quad n = 1, 2, 3, \ldots$ where $\omega\tau$ represents the product of applied frequency ω and delay time τ. This periodicity highlights the dynamic nature of information flow within proteins, where the roles of driver and driven residues can shift depending on the frequency of perturbation and timescale of observation.

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.15.575718; this version posted January 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

To illustrate how one can use time-dependent perturbations to study allostery and the sequence of events in proteins, we examine the system Cyclophilin A (CypA) which is an important allosteric protein that catalyzes the conversion of cys-trans transition of proline residues. We study the wild type cyclophilin A and two mutated structures.

Theory

In this section, we introduce the Dynamically Perturbed Gaussian Network Model, DP-GNM, and Causal relations. The details of the mathematical derivations are presented in the Supplementary material section. Here we give only the results of the detailed derivations.

1. Formulation of the problem: The time correlation $C_{ij}(\tau)$ of the fluctuations ΔR of two residues i and j is defined as

$$C_{ij}(\tau) = \left\langle \Delta R_i(t)\, \Delta R_j(t+\tau) \right\rangle \quad (1)$$

where τ is the time delay, i and j denote the residue indices and angular brackets denote an average. The fluctuations are vectorial quantities and the product represents the scalar product. GNM assumes that fluctuations are spherically symmetric, therefore considering only one component is sufficient. $C_{ij}(\tau)$ represents the effect of the fluctuations of i on subsequent fluctuations of j. If $C_{ij}(\tau)$ is larger than $C_{ji}(\tau)$, for example, then one can assign directionality to the system and say that fluctuations of i drive the fluctuations of j. This is how the DP-GNM introduces causality into allostery.

The time evolution of the protein of n residues obeys the Langevin equation

$$\eta \Delta R(t)' + \gamma \Gamma \Delta R(t) = F_r(t) + F_0 \cos(\omega t) \quad (2)$$

where, $\Delta R(t)$, the instantaneous fluctuation vector, is an nx1 vector with the ith entry referring to the instantaneous fluctuation of the ith residue. For anisotropic fluctuations, each entry of $\Delta R(t)$ will have three components in three coordinate directions. Here, we consider a single value, which may be assumed as the average of the three since we adopt the isotropic representation[3] of the GNM for simplicity. The prime in Eq. 2 shows differentiation with respect to time, η is the friction coefficient, γ is a characteristic spring constant and Γ is the nxn dimensionless Kirchoff matrix where the ij'th term is -1

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.15.575718; this version posted January 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

if residues i and j are within a cutoff distance of each other and interacting, or zero otherwise. Each diagonal term equates to the negative sum of the corresponding row. In the theory, the ratio $\frac{\eta}{\gamma}$ will appear, which we call the effective viscosity, and denote by ζ .The right hand side represents the sum of a random force $F_r(t)$ and a deterministic force $F_0\cos(\omega t)$ on the protein. In our previous work, we showed[4] that the solution of the Langevin equation where the random force is white noise that obeys $\langle F_i(t)F_j(t')\rangle = \eta k_B T\delta_{ij}\delta(t-t')$ leads to the GNM model. Here, k_B is the Boltzmann constant, T is the temperature, δ_{ij} is the Kronecker delta which is equal to 1 if i and j are equal and zero otherwise, and $\delta(t-t')$ is the Dirac function which is zero if the argument is nonzero and infinity if the argument is zero. In the present work we superpose on the random force a deterministic component, $F(t)=F_0\cos\omega t$, where ω is the frequency of the applied force and F_0 is a nx1 vector whose ith entry is the force on the ith residue. In this study, we perturb the GNM model by simultaneously applying a white noise generated random force and a time-varying force with a cosine profile. The solution of Eq. 2 is obtained as the superposition of three functions, $\Delta R(t)=\Delta R_r(t)+\Delta R_p(t)+Ce^{(-\zeta t\Gamma)}$ where $\Delta R_r(t)$ is the solution for the random force and $\Delta R_p(t)$ is the solution for the deterministic force. The solution to the random force is given in Reference [4] which is also presented in the Supplementary Material. The solution to the deterministic force is

$$\Delta R_p(t)=\cos(\omega t)L\Gamma F_0+\sin(\omega t)\eta\omega LF_0 \qquad (3)$$

where L is an nxn matrix given by

$$L=\left(\Gamma^2+\zeta^2\omega^2 I\right)^{-1} \qquad (4)$$

where, the ij'th component is

$$L_{ij}=\frac{1}{\gamma^2}\sum_k\frac{1}{\zeta^2\omega^2+\lambda_k{}^2}u_{ik}u_{jk} \qquad (4)'$$

with $\zeta=\eta/\gamma$, i.e., the effective vicosity, and λ_k and u_{ik} showing the kth eigenvalue and the i'th entry of the kth eigenvector of the Γ matrix, respectively. We are interested not in the fluctuations $\Delta R_r(t)$

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.15.575718; this version posted January 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

and $\Delta R_p(t)$ in the solution of the Langevin equation, but in the time correlation function obtained from them:

$$C_{ij}(\tau) = \langle \Delta R_i(0) \Delta R_j(\tau) \rangle = \langle \Delta R_i(0) \Delta R_j(\tau) \rangle_r + \langle \Delta R_i(0) \Delta R_j(\tau) \rangle_p \quad (5)$$

The subscripts r and p refer to $\Delta R_i(t)_r$ and $\Delta R_i(t)_p$, respectively. The averages are performed on the assumption that the system is ergodic and time averages are equal to ensemble averages. In Eq. 5, the cross terms $\langle \Delta R_i(0)_r \Delta R_j(\tau)_p \rangle, \langle \Delta R_i(0)_p \Delta R_j(\tau)_r \rangle$ are zero since the average of a zero mean random and a deterministic continuous function is zero. Moreover, the exponential term $Ce^{(-\zeta t\Gamma)}$ does not contribute as it decays exponentially as $t \to \infty$. We give the derivation of the explicit forms of these functions in the Supplementary Material Section. Here, we show only the results:

i. Time correlation for the unperturbed protein: The expression for time delayed pair correlations in an unperturbed GNM is given[4, 27]

$$\langle \Delta R_i(0) \Delta R_j(\tau) \rangle_r = \frac{1}{2} \frac{k_B T}{\gamma} \sum_k \frac{u_{ik} u_{jk}}{\lambda_k} e^{-\frac{\lambda_k}{\zeta}\tau} \quad (6)$$

This correlation function is symmetric in i and j.

ii. Time correlation in perturbed network: When residue p is perturbed, the correlation between i at time 0 and j at time τ is given as (See Supplementary material section for derivation)

$$\langle \Delta R_i(0) \Delta R_j(\tau) \rangle_p = \left(A_i^2 + B_i^2\right)^{1/2} \left(A_j^2 + B_j^2\right)^{1/2} \cos\left(\omega\tau + \theta_{ij}\right) \quad (7)$$

where,

$$A_i \equiv \left(L\Gamma F\right)_i = \frac{1}{\gamma} \sum_k \frac{\lambda_k}{\zeta^2\omega^2 + \lambda_k^2} u_{ik} u_{pk} F_p \quad (8)$$

$$B_i = \left(LF\right)_i = \frac{1}{\gamma} \sum_k \frac{\zeta\omega}{\zeta^2\omega^2 + \lambda_k^2} u_{ik} u_{pk} F_p \quad (9)$$

and

$\theta_{ij} = \theta_j - \theta_i$ represents the phase shift with $\theta_i = \arctan\frac{B_i}{A_i}$, and related to A_i and B_i with

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.15.575718; this version posted January 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

$$\tan\theta_{ij} = \frac{B_j A_i - B_i A_j}{A_i A_j + B_i B_j} \qquad (10)$$

Since θ_{ij} is antisymmetric, $\left\langle \Delta R_i(0)\Delta R_j(\tau)\right\rangle_p$ will not be equal to $\left\langle \Delta R_j(0)\Delta R_i(\tau)\right\rangle_p$. When p is perturbed, response of i will contain a phase difference and response of j will have another phase difference. Therefore correlation of i at time zero and of j at time τ will contain a phase difference θ_{ij}. However, if j is observed first and i after a time τ the phase difference will be $-\theta_{ij}$. It is this difference that leads to causality. In the calculations, the perturbed residue p may be the residue i or j.

2. Dimensional analysis and choice of system variables: The theoretical model is a coarse-grained model and the correspondence of its parameters to the real system should be considered. Denoting the dimensions of time, length and force by T, L and F, respectively, the dimensions of the friction constant η, the spring constant γ and the applied frequency ω respectively are FT/L, F/L and $1/T$. The effective viscosity ζ will then have the dimension of time and the eigenvalues will be dimensionless. Equations 7-10 which we use will all be expressed in terms of three dimensionless parameters, $\omega\tau$, $\zeta\omega$ and the set of eigenvalues λ_i. The average of the eigenvalues $\bar{\lambda}$ is related to the sum of the diagonal elements of Γ by the relation

$$\bar{\lambda} = \frac{n}{n-1}\sum_i \Gamma_{ii} \qquad (11)$$

As an example, taking a cutoff value 7.0 Å for Cyclophilin A in Eq. 11 gives an average number of 8.2 for both sides. If dynamics close to the unperturbed state is of interest, then values of $\zeta^2\omega^2$ should be taken such that $\zeta^2\omega^2 << \bar{\lambda}^2$. The average of the smallest 10 eigenvalues of Γ, which shows large scale collective motions, is approximately 1.0. Therefore a value of $\zeta^2\omega^2 \le 1$ is representative of large scale, slow dynamics.

Relaxation of correlations in the random force case may be written in terms of the dimensionless parameters as

$$\left\langle \Delta R_i(0)\Delta R_j(\tau)\right\rangle_r = \frac{1}{2}\frac{k_B T}{\gamma}\sum_k \frac{u_{ik}u_{jk}}{\lambda_k} e^{-\frac{\lambda_k}{\zeta\omega}\omega\tau} \qquad (6')$$

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.15.575718; this version posted January 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Equation 6' may be used as a benchmark for comparing the dynamics of the perturbed and the unperturbed protein.

In the formulation, γ has dimensions of FL^{-1} . In order to make the model easily scalable and to remove the dimensionality, we equate γ to unity in all the calculations below.

3. Factors affecting the causality of information transfer between a residue pair

The time correlation given by eq. 7 is a cosine function, which is shifted with respect to the applied force. The amount of shift depends on the parameters of the system. The numerator on the right hand side which induces the shift is antisymmetric in i and j and depends on the effective friction coefficient, frequency, the eigenvectors of the Γ matrix and the applied force. Due to the presence of the antisymmetric θ_{ij} , $C_{ij}(\tau)$ and $C_{ji}(\tau)$ are not equal. The inequality $C_{ij}(\tau) \neq C_{ji}(\tau)$ is the source of causality between i and j. If $\langle \Delta R_i(t) \Delta R_j(t+\tau) \rangle > \langle \Delta R_j(t) \Delta R_i(t+\tau) \rangle$, then the fluctuations of residue i at time t has more influence on the future values of the fluctuations of j at $t+\tau$. Equation 7 is a periodic function. The front factor on the right hand side does not depend on the order ij or ji, the only dependence being on θ_{ij} . Therefore, examination of the $\cos(\omega\tau \pm \theta_{ij})$ term in Eq.7 will shed light on directionality of correlations. The two cosine curves will be shifted relative to each other by an amount of $1 \pm \cos(\theta_{ij})$, where the $\pm$ sign depends on the sign of θ_{ij} . For positive θ_{ij} , $C_{ij}(\tau) = \cos(\omega\tau + \theta_{ji})$ will be larger than $C_{ji}(\tau) = \cos(\omega\tau + \theta_{ij})$ for the intervals $0 < \omega\tau < \pi$, $2\pi < \omega t < 3\pi$ ... This is sketched in the inset in Figure 1. For illustrative purposes we use the following values for $\theta_{ij} = 1$ and $\theta_{ji} = -1$. In the figure, the abscissa shows the values of $\omega\tau$ and the ordinate values for $C_{ij}(\tau)$ (solid curve) and $C_{ji}(\tau)$ (dot dashed curve). In the present study, we assume low frequencies such that the system is in the first interval $0 < \omega\tau < \pi$ as sketched in the main part of Figure 1.

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.15.575718; this version posted January 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

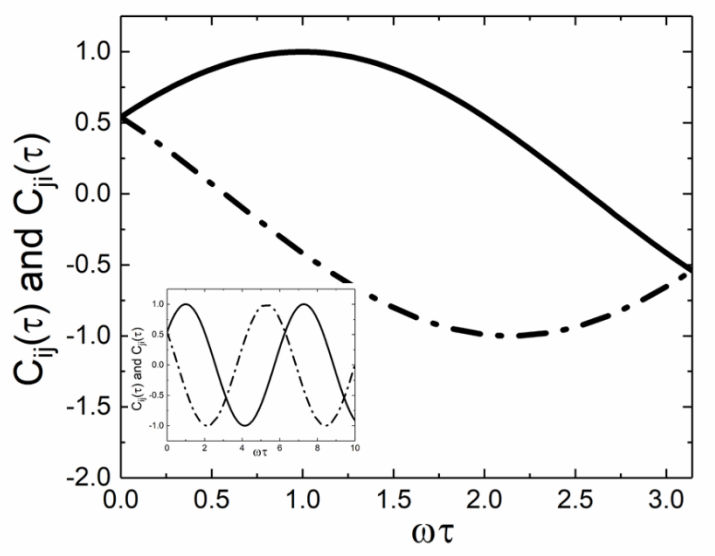


Figure 1. Sketch of $\cos\left(\omega\tau+\theta_{ji}\right)$ (solid curve) and $\cos\left(\omega\tau+\theta_{ij}\right)$ (dot dashed curve) both in the inset and the main figure. The solid curve shows the values of the correlation $C_{ij}\left(\tau\right)$ when residue i is perturbed.

We now consider the low frequency region, i.e. the region $0<\omega\tau<\pi$ in Figure 1. Examination of equations 8 and 9 shows that the shift is proportional to the difference $B_jA_i-B_iA_j$. The sign of this function which determines the direction of information transfer between i and j depends on the eigenvalues and the eigenvectors of the gamma matrix, i.e., the topology of the protein structure. Finally, we note that the direction of information transfer through correlations reverses periodically as a function of frequency and delay time. Therefore, there is no absolute information transfer direction between two residues. In the present study, we consider slow frequencies and short delay times, i.e., the regime where $0<\omega\tau<\pi$.

4. Identifying switch and receptor residues

If $C_{ij}\left(\tau\right)-C_{ji}\left(\tau\right)>0$ then residue i acts as the driver of j. A large positive value of the sum $\sum_j\left[C_{ij}\left(\tau\right)-C_{ji}\left(\tau\right)\right]$ indicates that residue i drives several residues and hence acts as a switch in the protein. Conversely, a large negative value of this sum indicates that the residue i is a driven residue of the protein, analogous to the lightbulb example. Identification of switch and lightbulb residues of the protein gives significant information on the allosteric behavior of the protein. We show the significance and importance of this in the results section.

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.15.575718; this version posted January 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

5. Identifying residues that move synchronously; Communities

If two residues i and j move synchronously, then the phase difference between their motions, given by Eq. 10, will vanish. Identifying the set of residues that move synchronously with the perturbed residue may reveal valuable information about protein dynamics and potentially allosteric pathways. These synchronously moving residues form a network directly or indirectly connected to the perturbed residue. Their movements are dynamically correlated, meaning a change in one affects the others, often due to physical interactions or conformational changes propagating through the protein. This network may be involved in protein functions. Residues directly coupled to the perturbation site might be critical for substrate binding, enzyme activity, or other mechanisms. Synchronous movements suggest communication pathways within the protein structure. The network likely transmits information or signals from the perturbation site to other functional regions. If the perturbation site is functionally relevant, its connected network potentially forms the communication route for allosteric effects. Alternatively, we can consider the relationship between synchronously moving parts and communities. Proteins often exhibit a modular architecture, where semi-autonomous sub-units called communities perform specific functions. Identifying synchronously moving residues can tell us which communities are functionally coupled. For example, communities involved in substrate binding and catalysis might move together when the substrate binds. Communities are not static entities; they dynamically interact and communicate to fulfill complex protein functions. By analyzing synchronous movements, we can learn about the communication pathways between communities, crucial for understanding how different functional modules work together. Allosteric regulation relies on long-range signal propagation within the protein. Communities can serve as "communication hubs" within this network. Identifying synchronously moving residues can help us map these allosteric pathways, highlighting the communities mediating signal transmission from the effector site to the functional output. Allosteric effects can result in the reorganization of communities. For example, ligand binding might reconfigure interactions between communities, altering their dynamics and influencing overall protein function. Therefore, studying synchronous movements can reveal these dynamic changes and their functional consequences.

Application of the DP-GNM to Cyclophilin A

1. The structure of Cyclophilin A and the allosteric network residues: In this section we apply the results of the DP-GNM to a widely studied protein Cyclophilin A, a 163 residue protein, which binds its substrate at its active site and catalyzes the cys-trans transition of the proline residue of its substrate. The protein

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.15.575718; this version posted January 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Data Bank code is 3k0n. The residues participating in catalytic activity were determined using nmr relaxation experiments[28] and studied using molecular dynamics simulations[29, 30]. Simulations showed that catalytic activity of CypA could be explained by changes in motions of residue Phe113 on a ns–µs timescale. The three dimensional structure of CypA is shown in Figure 2.

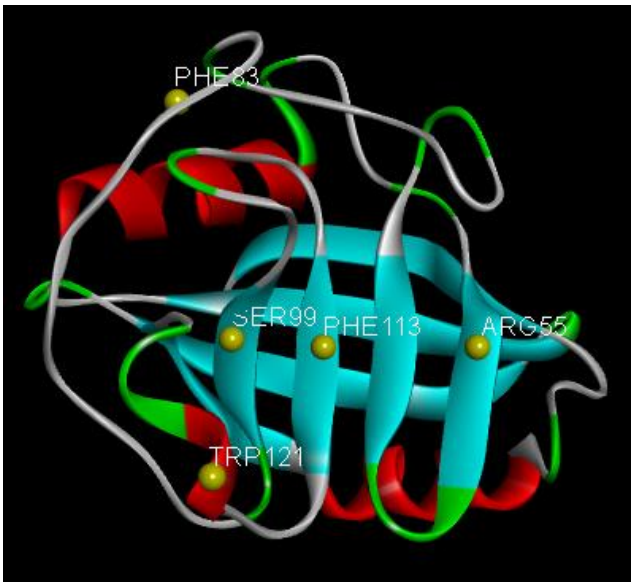


Figure 2. Ribbon diagram of Cyclophilin A, 3K0N.pdb, where yellow spheres show the alpha carbons of the allosteric network residues.

The residues that are known[28, 29] to constitute the allosteric network, Arg55, Phe83, Ser99, Phe113, Trp121 are identified by their alpha carbons as yellow spheres. Arg55 is the key active site residue, which has the catalytic activity of converting a proline of a substrate between cis and trans states. The allosteric interactions between Arg55 and the remaining allosteric network residues are the key components of the catalytic activity. Although the allosteric activity of Cyclophilin A has been addressed in several papers[9, 29-38] causality and directional information transfer have not been considered in detail.

2. Determining the driver and driven residues in pair interactions: According to Eq. 7 when residue i is perturbed, its correlation with j at a later time will differ from the correlation of j with the perturbed i at a later time. If $\left|C_{ij}\left(\omega\tau\right)\right| > \left|C_{ji}\left(\omega\tau\right)\right|$, we call residue i the driver and j the driven. In order to understand the driver-driven relations between residue Arg55, the major player in catalysis, and the residues of the allosteric network, we perturbed residue Arg55 and calculated the ensuing correlations. In Figure 3, we show the time correlation functions between Arg55 and other members of the allosteric pathway. It is

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.15.575718; this version posted January 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

interesting to see that the perturbed residue does not always become the driver as would normally be expected. Calculations with the remaining residues of the protein support this observation.

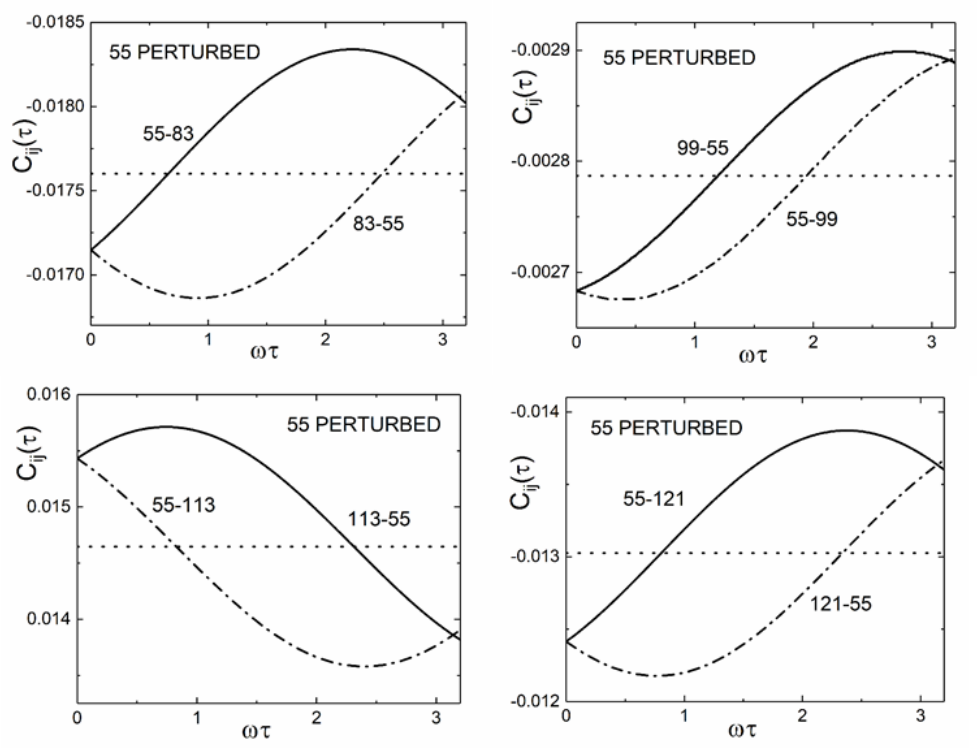


Figure 3. Time correlations between residue Arg55 and residues Phe83, Ser99, Phe113 and Trp121 as function of $\omega\tau$. In all panels, residue Arg55 is perturbed. The dimensionless parameter in Equations 7-10 are taken as $\zeta\omega = 1.$ and $F_0 = 1$. Cutoff radius for the GNM is taken as 7.0 Å. The dotted horizontal line shows the equilibrium pair correlation of unperturbed protein given by $\left\langle \Delta R_i(0)\Delta R_j(0)\right\rangle_r = \sum_k \frac{u_{ik}u_{jk}}{\lambda_k}$.

The upper left panel in Figure 3 compares the time correlation function between Arg55 and 83 when residue 55 is perturbed. The solid curve shows the correlation $C_{55,83}(\omega\tau)$ between the fluctuation of residues Arg55 and Phe83 where the fluctuations of Phe83 are observed $\omega\tau$ units later than those of Arg55. The lower dot-dashed curve on the other hand shows the correlation $C_{83,55}(\omega\tau)$. The abscissa covers the $0 \le \omega\tau \le \pi$ range . At fixed $\omega\tau$, fluctuations of residue Arg55 have larger effect on later values of Phe83. The trend is reversed in the next interval $\pi \le \omega\tau \le 2\pi$ as discussed in the theory section. Throughout this study, we concentrate on the first range only, which represents the slow perturbation range. The $\omega\tau = 0$ intercept shows the case of static perturbation, which is weaker than the unperturbed correlation between Arg55 and Phe83. As $\omega\tau$ increases, correlation between the pair

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.15.575718; this version posted January 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

increases. In the remaining three panels, correlations of residue Arg55 with the other members of the allosteric pathway are shown. In the lower left panel, for the Arg55-Phe113 pair, an increase in static perturbation at $\omega\tau = 0$ is stronger than the unperturbed correlation. If residue i is perturbed and $\left|C_{ij}\left(\omega\tau\right)\right| > \left|C_{ji}\left(\omega\tau\right)\right|$ then the residue i drives j and residue j is driven by i. With this definition, we see from Figure 3 that when Arg55 is perturbed, residues Arg83, Phe113 and Trp121 drive Arg55, but Arg55 drives only Ser99, with a strength that is an order of magnitude weaker.

Next, we analyze the difference $C_{ij}\left(\omega\tau\right) - C_{ji}\left(\omega\tau\right)$ which shows the directionality of correlations upon perturbation of a residue. For this purpose, we calculated the values of $C_{ij}\left(\omega\tau\right)$ for all values of i and j where the residue indicated by the first index is perturbed. Thus, $C_{ij}\left(\omega\tau\right)$ is the correlation when i is perturbed and j is observed at a time $\omega\tau$ later, and $C_{ji}\left(\omega\tau\right)$ is the correlation when j is perturbed and i is observed $\omega\tau$ later . A positive value of the difference shos that i is the driver residue. Positive values of this difference are shown in Figure 4. Hence, the horizontal axis indicates the driver residue index i

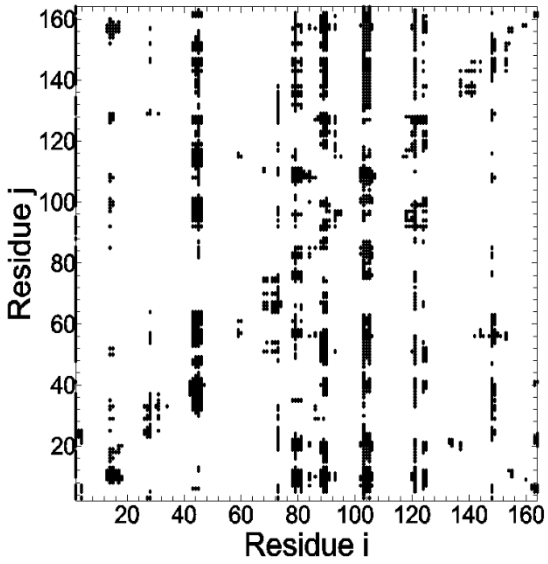


Figure 4. Values of $\left|C_{ij}\left(\omega\tau\right) - C_{ji}\left(\omega\tau\right)\right|$ presented in terms of residue indices i and j. Black dots show values of this difference in the range $0.003 < C_{ij} - C_{ji} < 0.07$ which covers the top 15% of all points obtained with $F_0 = 1$, $\omega\tau = 0.1,\ \zeta\omega = 0.1.$

which is perturbed by a cosine force of magnitude $F_0 = 1$ and $\omega\tau = 1.0$. The vertical axis indicates the driven residue index j which is affected by the perturbation. The calculations are performed with

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.15.575718; this version posted January 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

$\zeta\omega = 0.1$. The values of $C_{ij} - C_{ji}$ indicate the level of causality between residue pairs. In this example, $0 < C_{ij} - C_{ji} < 0.07$. There are 13,366 points with positive values. Only the largest 15% of those points are shown in the figure in the interest of identifying important pairs with causality. Interestingly, driver residues are clearly visible as specific vertical patterns in the figure. Accordingly, residues between 42-46, 78-81, 88-91, 100-106, 121, 123-125 and 147-149 act as strong drivers.

Each residue is under the influence of all other residues. This is reflected in the sum $\sum_j \left[C_{ij}(\tau) - C_{ji}(\tau) \right]$. A positive value of this sum indicates that the residue i acts as a driver, and conversely a negative value indicates a driven residue. In Figure 5, the values of this sum are plotted as a function of residue index.

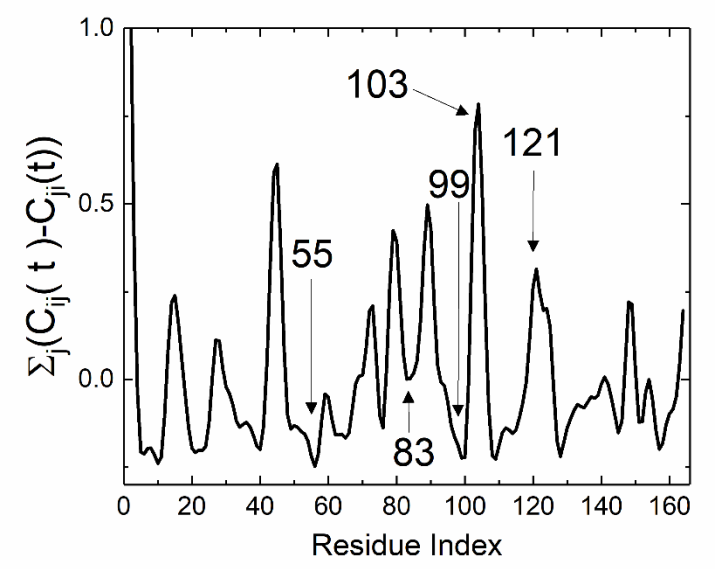


Figure 5. Values of the sum $\sum_j \left[C_{ij}(\tau) - C_{ji}(\tau) \right]$ presented as a function of the residue index i. See legend to Figure 4.

The peaks in the figure correspond to strong driver and the minima to driven residues. The deepest minimum is for residue Arg55, which is the active site residue that induces the cis-trans isomerization of the substrate. Among the maxima, residues Glu43, Phe83, Ala101 and Trp121, which are known to play a role in the allostery, are observed.

3. Determining the community residues: Understanding how proteins function often boils down to deciphering the intricate networks of interactions between their amino acids. These interactions extend

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.15.575718; this version posted January 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

beyond direct physical contacts, forming dynamic communities of residues that act in concert. While the term "community" is widely used, how do we actually identify and understand these hidden groups within a protein? Perturbing a specific residue and observing the other residues which fluctuate with a small phase difference, i.e., with small values of θ_{ij} is a plausible approach. This synchronized movement suggests a protein community, a cluster of residues that share a functional response. Thus, perturbing a single residue acts as a probe, uncovering the existing communities within the protein. These communities might not be static groups, but rather dynamic ensembles that fluctuate depending on the surrounding environment and cellular needs. However, some communities exhibit remarkable persistence. Despite the location of the perturbation, these core groups consistently show synchronized movements, hinting at their crucial roles in protein function. Identifying these persistent communities through perturbation may help to understand the protein's functional landscape. In Figure 6, we present the values of $\theta_{ij} = \theta_j - \theta_i$, as a function of residue index where the number in each figure shows the residue which is perturbed. The calculations are performed with F_0 being the force vector with a 1.0 in the component corresponding to the forced residue and zero otherwise. Below, we used the notation $F_0 = 1$ for brevity. $\omega\tau = 1.0$, $\zeta\omega = 1.0$.

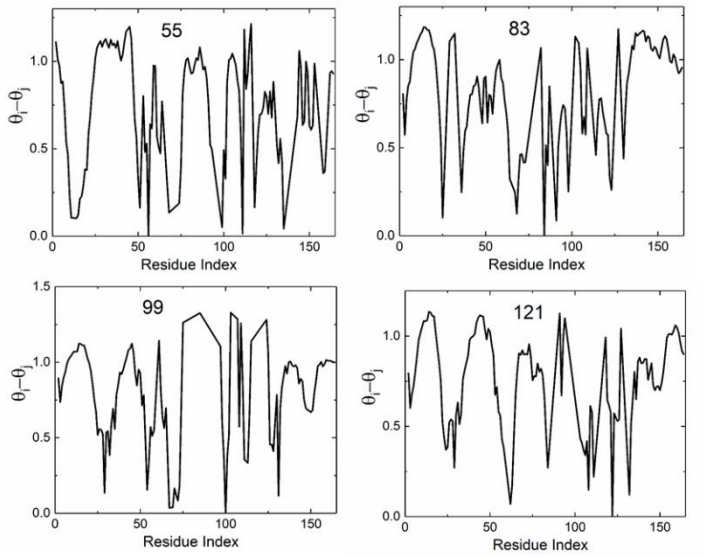


Figure 6. Phase difference between a perturbed residue (identified in each figure) and the remaining residues of the protein, plotted as a function of residue index. Calculations are performed with $F_0 = 1$, $\omega\tau = 1.0$, $\zeta\omega = 1.0$.

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.15.575718; this version posted January 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The curves in Figure 6 do not show long stretches of residues for which θ_{ij} is close to zero, which would indicate same phase with the perturbed residue. However, the curves identify single residues that are in phase with the perturbed residue. For example, in the upper left panel of the Figure, residues Ser99, Phe113 and Val132 move in phase with Arg55. Lower left panel shows that residues Arg55 and Glu74 move in phase with Ser99.

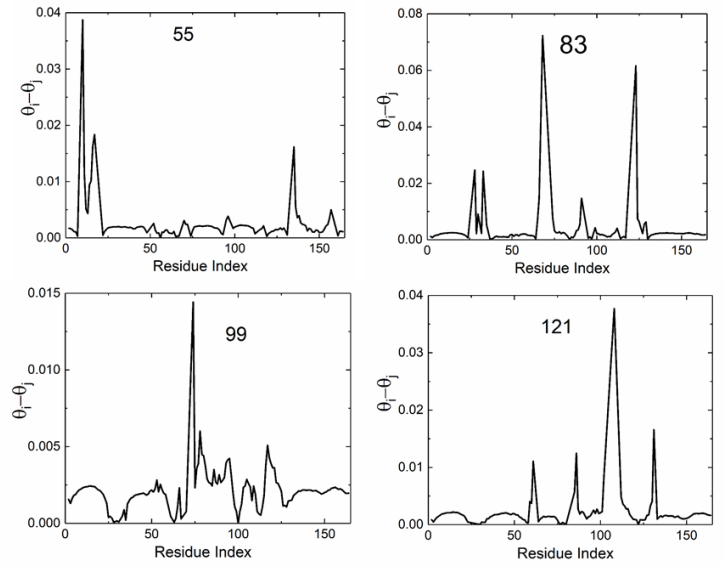


Figure 7. Phase difference between a perturbed residue (identified in each figure) and the remaining residues of the protein, plotted as a function of residue index. Calculations are performed with $F_0 = 1,\ \omega\tau = 1.0,\ \zeta\omega = .01.$

While perturbation with values of $\zeta\omega = 1.0$ is unable to identify communities, calculations with a lower frequency perturbation, $\zeta\omega = .01$ clearly can identify communities, as may be seen from Figure 7. Several regions of the protein that fluctuate with the same or similar phases are seen. Thus, for all panels the range of residues 30-60 which is the helix-coil-beta motive that contains the catalytic residue Arg55 forms a community commonly observed when residues Arg55, Phe83, Ser99 and Trp121 are perturbed.

4. Effect of mutation on Cyclophilin activity. In order to see the effects of mutation on the catalytic activity of Cyclophylin A, (i) the distal residue Ser99 was mutated to Thr99 (PDB code 3K0O) and (ii) the central catalytic residue Arg55 was mutated to Arg (PDB code 3K0R). The result was approximately 60 fold decrease of the rate of conversion of cis to trans of the ligand for both structures[28, 29, 39]. In Figure 8

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.15.575718; this version posted January 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

we show the changes in driver-driven relations for the two mutations. A positive value for a residue indicates that mutation makes that residue a stronger

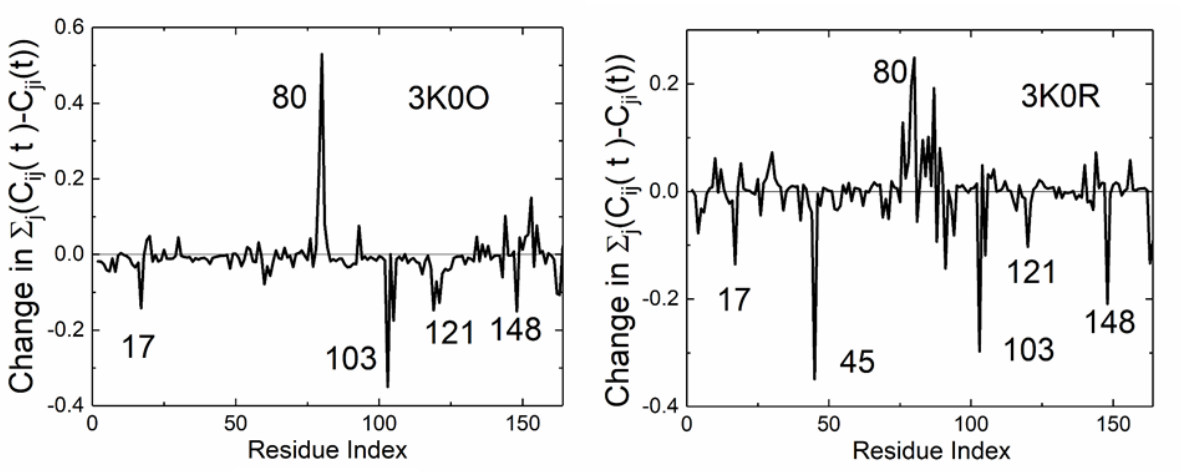


Figure 8. The differences $\sum_j \left[C_{ij}(\tau)-C_{ji}(\tau)\right]_{mutated} - \sum_j \left[C_{ij}(\tau)-C_{ji}(\tau)\right]_{WT}$ in driver-driven relations for the mutated and wild type structures. The left panel is for the Ser99Thr mutation, the right panel is for the Arg55Lys mutation. Largest changes are indicated in each panel.

driver relative to the wild type. In both panels of Figure 8, we see that Gly80, which is already a driver residue in the wild type structure becomes a stronger driver with either mutation. On the contrary, residues Ala103, Trp121 and Asn149 which are drivers in the wilt type structure, lose this property upon mutation of either Ser99 and Arg55. Overall, it appears that mutation upsets the driver-driven balance in the protein and leads to loss of function.

Discussion and Conclusion

The application of DP-GNM to Cyclophilin A has facilitated a comprehensive analysis of its dynamic behavior and provided insightful details into the temporally ordered events within its allosteric network. Our results reveal a complex interplay between driver and driven residues, organized within a hierarchical framework of information flow. Key residues such as Arg55 and Trp121 exhibit a strong driver character, exerting significant influence on others, while some, like Phe113, demonstrate a dynamic role, transitioning between driver and driven depending on the context. This suggests a cascading pattern of information transfer within the protein, where perturbations at specific sites trigger a sequential propagation of conformational changes throughout the network.

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.15.575718; this version posted January 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Intriguingly, residue 55, traditionally viewed as the primary driver of allosteric regulation due to its active site role, primarily appears as a driven residue in our analysis. This suggests that its functionality is largely governed by conformational changes initiated elsewhere in the protein, highlighting the intricate web of influence between distinct regions within Cyclophilin A. This finding underscores the limitations of the single-driver model of allosteric regulation and emphasizes the need to account for the complexity of hierarchical information flow within proteins.

Furthermore, DP-GNM has identified the presence of dynamically coupled communities—groups of residues exhibiting synchronized movements upon perturbation. These communities are likely to represent functionally relevant modules within Cyclophilin A, responsible for specific aspects of its activity. Identifying these communities enhances our understanding of protein function and the underlying dynamic pathways. However, it is crucial to acknowledge that these communities are not static entities. They may reconfigure under different conditions such as frequency of perturbation or upon binding ligands, highlighting the dynamic nature of protein structure and function. Further research is necessary to explore the plasticity of these communities and their relationship to protein function under various environments.

While directionality exhibits a periodic reversal within intervals of $(n-1)\pi < wt < n\pi$, it's important to note that the first interval, $0 < wt < \pi$, holds particular relevance for understanding slow dynamics. If perturbations primarily reside within this range, the observed directionality remains consistent, and concerns regarding potential reversals within subsequent intervals become less pertinent.

Limitations of the current study include the use of a simplified GNM model that assumes isotropic fluctuations and neglects specific intramolecular interactions. Additionally, the chosen frequency range focuses on slower processes and might not capture the full spectrum of allosteric dynamics. Future work can incorporate anisotropic models and explore a wider range of frequencies to achieve a more comprehensive understanding of protein dynamics and the complete repertoire of information flow within proteins.

DP-GNM has emerged as a powerful tool for elucidating the intricacies of protein dynamics and allosteric regulation. By revealing the temporal sequence of events, directionality of information flow, and dynamic interplay between different regions, DP-GNM provides valuable insights into the mechanisms underlying protein function. Our analysis of Cyclophilin A demonstrates the hierarchical organization of the allosteric network, the driven nature of the active site residue Arg55, the existence of dynamically coupled communities, and the importance of considering perturbation frequency ranges for interpreting

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.15.575718; this version posted January 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

directionality. This approach has the potential to advance our understanding of protein function and provide a foundation for the development of novel therapeutic strategies targeting allosteric processes. Future research directed towards refining the model, exploring a wider range of proteins, and integrating experimental data will further solidify DP-GNM as a useful tool in the field of protein dynamics and open new avenues for understanding and manipulating protein function.

While this study focused on DP-GNM's application to a single case of catalytic activity of Cyclophilin A, its potential extends to analyzing systems under various perturbations, including ligand binding, for example. A ligand docked at a specific site on the protein with its dominant frequency overlapping with the protein's vibrational modes, could be efficiently absorbed, effectively "pumping" energy into the protein's internal network. This vibrational energy transfer could act as a periodic perturbation, triggering a ripple effect through the protein structure. DP-GNM could identify the key residues that initially "catch" the vibrational energy from the ligand, the communities of residues that synchronize their movements in response, and ultimately, the downstream functional consequences of this allosteric excitation. This detailed understanding of ligand-induced dynamics could offer valuable insights into ligand efficacy, selectivity, and even potential off-target effects. Moreover, by analyzing the directionality of information flow under ligand binding, DP-GNM could inform the design of ligands with targeted vibrational properties. By precisely engineering ligand vibrations to resonate with specific protein modes, we could potentially steer allosteric responses towards desired outcomes, paving the way for novel strategies in drug design and protein engineering.

Intriguingly, the shifts in driver/driven residues, emergence of synchronized communities, and coordinated responses under dynamic loading bear resemblance to the dynamic nature of the circadian clock. While the circadian clock primarily relies on biochemical processes for its rhythm generation, several types of mechanical perturbations can influence its function in indirect ways. This raises the question: could DP-GNM, applied to protein networks during circadian oscillations, illuminate the interplay between environmental cues and internal protein dynamics?

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.15.575718; this version posted January 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

References

1. Atilgan, C., Computational methods for efficient sampling of protein landscapes and disclosing allosteric regions. *Advances in protein chemistry and structural biology* **2018,** *113*, 33-63.
2. Atilgan, C.; Atilgan, A. R., Perturbation-response scanning reveals ligand entry-exit mechanisms of ferric binding protein. *PLoS computational biology* **2009,** *5* (10), e1000544.
3. Bahar, I.; Atilgan, A. R.; Erman, B., Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Fold Des* **1997,** *2* (3), 173-181.
4. Erkip, A.; Erman, B., Dynamics of large-scale fluctuations in native proteins. Analysis based on harmonic inter-residue potentials and random external noise. *Polymer* **2004,** *45* (2), 641-648.
5. Hacisuleyman, A.; Erkip, A.; Erman, B.; Erman, B., Synchronous and Asynchronous Response in Dynamically Perturbed Proteins. *The Journal of Physical Chemistry B* **2021,** *125* (3), 729-739.
6. Scaramozzino, D.; Khade, P. M.; Jernigan, R. L., Protein Fluctuations in Response to Random External Forces. *Applied Sciences* **2022,** *12* (5), 2344.
7. Scaramozzino, D.; Piana, G.; Lacidogna, G.; Carpinteri, A., Low-Frequency Harmonic Perturbations Drive Protein Conformational Changes. *International Journal of Molecular Sciences* **2021,** *22* (19), 10501.
8. Botello-Smith, W. M.; Luo, Y., Robust determination of protein allosteric signaling pathways. *J Chem Theory Comput* **2019,** *15* (4), 2116-2126.
9. Haliloglu, T.; Hacisuleyman, A.; Erman, B., Prediction of allosteric communication pathways in proteins. *Bioinformatics* **2022,** *38* (14), 3590-3599.
10. Dokholyan, N. V., Controlling allosteric networks in proteins. *Chemical reviews* **2016,** *116* (11), 6463-6487.
11. Kaya, C.; Armutlulu, A.; Ekesan, S.; Haliloglu, T., MCPath: Monte Carlo path generation approach to predict likely allosteric pathways and functional residues. *Nucleic acids research* **2013,** *41* (W1), W249-W255.
12. Negre, C. F.; Morzan, U. N.; Hendrickson, H. P.; Pal, R.; Lisi, G. P.; Loria, J. P.; Rivalta, I.; Ho, J.; Batista, V. S., Eigenvector centrality for characterization of protein allosteric pathways. *Proceedings of the National Academy of Sciences* **2018,** *115* (52), E12201-E12208.
13. Wu, N.; Yaliraki, S. N.; Barahona, M., Prediction of protein allosteric signalling pathways and functional residues through paths of optimised propensity. *Journal of Molecular Biology* **2022,** *434* (17), 167749.
14. Lu, S.; Huang, W.; Zhang, J., Recent computational advances in the identification of allosteric sites in proteins. *Drug discovery today* **2014,** *19* (10), 1595-1600.
15. Stolzenberg, S.; Michino, M.; LeVine, M. V.; Weinstein, H.; Shi, L., Computational approaches to detect allosteric pathways in transmembrane molecular machines. *Biochimica et Biophysica Acta (BBA)-Biomembranes* **2016,** *1858* (7), 1652-1662.
16. Yao, X.-Q.; Hamelberg, D., Residue–residue contact changes during functional processes define allosteric communication pathways. *J Chem Theory Comput* **2022,** *18* (2), 1173-1187.
17. Ribeiro, A. A.; Ortiz, V., Determination of signaling pathways in proteins through network theory: importance of the topology. *J Chem Theory Comput* **2014,** *10* (4), 1762-1769.
18. Guarnera, E.; Berezovsky, I. N., Structure-based statistical mechanical model accounts for the causality and energetics of allosteric communication. *PLoS computational biology* **2016,** *12* (3), e1004678.
19. Kamberaj, H.; van der Vaart, A., Extracting the Causality of Correlated Motions from Molecular Dynamics Simulations. *Biophysical Journal* **2009,** *97* (6), 1747-1755.

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.15.575718; this version posted January 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

20. Hacisuleyman, A.; Erman, B., Causality, transfer entropy, and allosteric communication landscapes in proteins with harmonic interactions. *Proteins: Structure, Function, and Bioinformatics* **2017**.
21. Vatansever, S.; Erman, B.; Gumus, Z. H., Identifying Causality in Mutant KRas Residue Pairs from Molecular Dynamics Data Analysis. *Biophysical Journal* **2018,** *114* (3), 397a.
22. Vatansever, S.; Gumus, Z.; Erman, B., Intrinsic K-Ras dynamics: A novel molecular dynamics data analysis method shows causality between residue pairs. *bioRxiv* **2016**, 067496.
23. Bozovic, O.; Zanobini, C.; Gulzar, A.; Jankovic, B.; Buhrke, D.; Post, M.; Wolf, S.; Stock, G.; Hamm, P., Real-time observation of ligand-induced allosteric transitions in a PDZ domain. *Proceedings of the National Academy of Sciences* **2020,** *117* (42), 26031-26039.
24. Hyeon, C.; Lorimer, G. H.; Thirumalai, D., Dynamics of allosteric transitions in GroEL. *Proceedings of the National Academy of Sciences* **2006,** *103* (50), 18939-18944.
25. Buchenberg, S.; Sittel, F.; Stock, G., Time-resolved observation of protein allosteric communication. *Proceedings of the National Academy of Sciences* **2017,** *114* (33), E6804-E6811.
26. Wolf, S.; Sohmen, B.; Hellenkamp, B.; Thurn, J.; Stock, G.; Hugel, T., Hierarchical dynamics in allostery following ATP hydrolysis monitored by single molecule FRET measurements and MD simulations. *Chemical science* **2021,** *12* (9), 3350-3359.
27. Haliloglu, T.; Bahar, I.; Erman, B., Gaussian dynamics of folded proteins. *Physical Review Letters* **1997,** *79* (16), 3090-3093.
28. Fraser, J. S.; Clarkson, M. W.; Degnan, S. C.; Erion, R.; Kern, D.; Alber, T., Hidden alternative structures of proline isomerase essential for catalysis. *Nature* **2009,** *462* (7273), 669-673.
29. Wapeesittipan, P.; Mey, A. S.; Walkinshaw, M. D.; Michel, J., Allosteric effects in cyclophilin mutants may be explained by changes in nano-microsecond time scale motions. *Communications Chemistry* **2019,** *2* (1), 41.
30. Wapeesittipan, P.; Mey, A.; Walkinshaw, M.; Michel, J., Allosteric effects in catalytic impaired variants of the enzyme cyclophilin A may be explained by changes in nano-microsecond time scale motions. *bioRxiv* **2018**, 224329.
31. Doshi, U.; Holliday, M. J.; Eisenmesser, E. Z.; Hamelberg, D., Dynamical network of residue–residue contacts reveals coupled allosteric effects in recognition, catalysis, and mutation. *Proceedings of the National Academy of Sciences* **2016,** *113* (17), 4735-4740.
32. Lv, M.; Shi, T.; Mao, X.; Li, X.; Chen, Y.; Zhu, J.; Ni, S.; Shen, X.; Jiang, H.; Li, J., 1-(2, 6-Dibenzyloxybenzoyl)-3-(9H-fluoren-9-yl)-urea: a novel cyclophilin A allosteric activator. *Biochem Bioph Res Co* **2012,** *425* (4), 938-943.
33. Vu, P. J.; Yao, X.-Q.; Momin, M.; Hamelberg, D., Unraveling allosteric mechanisms of enzymatic catalysis with an evolutionary analysis of residue–residue contact dynamical changes. *ACS Catalysis* **2018,** *8* (3), 2375-2384.
34. Blackburn, E. A.; Wear, M. A.; Landré, V.; Narayan, V.; Ning, J.; Erman, B.; Ball, K. L.; Walkinshaw, M. D., Cyclophilin40 isomerase activity is regulated by a temperature-dependent allosteric interaction with Hsp90. *Bioscience reports* **2015,** *35* (5), e00258.
35. Holliday, M. J.; Camilloni, C.; Armstrong, G. S.; Vendruscolo, M.; Eisenmesser, E. Z., Networks of dynamic allostery regulate enzyme function. *Structure* **2017,** *25* (2), 276-286.
36. Dujardin, M.; Madan, V.; Gandhi, N. S.; Cantrelle, F.-X.; Launay, H.; Huvent, I.; Bartenschlager, R.; Lippens, G.; Hanoulle, X., Cyclophilin A allows the allosteric regulation of a structural motif in the disordered domain 2 of NS5A and thereby fine-tunes HCV RNA replication. *Journal of Biological Chemistry* **2019,** *294* (35), 13171-13185.
37. Rodriguez-Bussey, I.; Yao, X.-Q.; Shouaib, A. D.; Lopez, J.; Hamelberg, D., Decoding allosteric communication pathways in cyclophilin a with a comparative analysis of perturbed conformational ensembles. *The Journal of Physical Chemistry B* **2018,** *122* (25), 6528-6535.

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.15.575718; this version posted January 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

38. Lu, M.; Hou, G.; Zhang, H.; Suiter, C. L.; Ahn, J.; Byeon, I.-J. L.; Perilla, J. R.; Langmead, C. J.; Hung, I.; Gor'kov, P. L., Dynamic allostery governs cyclophilin A–HIV capsid interplay. *Proceedings of the National Academy of Sciences* **2015,** *112* (47), 14617-14622.
39. Eisenmesser, E. Z.; Millet, O.; Labeikovsky, W.; Korzhnev, D. M.; Wolf-Watz, M.; Bosco, D. A.; Skalicky, J. J.; Kay, L. E.; Kern, D., Intrinsic dynamics of an enzyme underlies catalysis. *Nature* **2005,** *438* (7064), 117-121.

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.15.575718; this version posted January 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Supplementary Material

Dynamically driven correlations in elastic net models reveal sequence of events and causality in proteins

Albert Erkip[a] and Burak Erman[b]

[a]Faculty of Engineering and Natural Sciences, Sabanci University, Orhanli 34956 Tuzla, Istanbul, Turkey

[b]Department of Chemical and Biological Engineering, Koc University, Rumelifeneri Yolu, Sariyer 34450 Istanbul, Turkey

We consider the system

$$\zeta \Delta R(t)' + \Gamma \Delta R(t) = F_r(t) + F_0 \cos(\omega t) \qquad \text{(SM-1)}$$

where $F_r(t)$ is a random force and F_0 is a constant vector. ζ is the friction coefficient, and Γ is the spring matrix of the Gaussian Network Model, GNM[3]. $F_r(t)$ is a zero mean white noise satisfying $\left\langle (\Delta F_r)_i (\Delta F_r)_j \right\rangle = \frac{\zeta}{\beta} \delta_{ij} \delta(t - t')$ where, δ_{ij} is the Kronecker delta, $\delta(t-t')$ is the Dirac Delta function and $\beta = 1/k_B T_0$, k_B being the Boltzmann constant and T_0 the absolute temperature.

Solutions of (SM-1) are of the form

$$\Delta R(t) = \Delta R_r(t) + \Delta R_p(t) + C e^{-\frac{t}{\zeta}\Gamma} \qquad \text{(SM-2)}$$

with $\Delta R_r(t)$ and $\Delta R_p(t)$ being solutions corresponding to the random force $F_r(t)$ and deterministic forces, respectively. The term $Ce^{-\frac{t}{\zeta}\Gamma}$ with an arbitrary constant vector C corresponds to the solution of the homogeneous equation, i.e., with zero force . As we are interested in the correlations

$$\left\langle \Delta R_i(0) \Delta R_j(\tau) \right\rangle = \lim_{T\to\infty} \frac{1}{T} \int_0^T \left\langle \Delta R_i(t) \Delta R_j(t+\tau) \right\rangle dt \qquad \text{(SM-3)}$$

the term $e^{-\frac{t}{\zeta}\Gamma}$ will not contribute as it decays rapidly as $t \to \infty$, hence we will neglect it.

To evaluate ΔR_r, we consider the Fourier series expansion for F_r in the

interval [0; 2T]

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.15.575718; this version posted January 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

$$F_r(t) = A_0 + \sum_{p=1}^{\infty} \cos\left(\frac{p\pi}{T}t\right) A_p + \sum_{p=1}^{\infty} \sin\left(\frac{p\pi}{T}t\right) B_p \quad \text{(SM-4)}$$

with the coefficients $A_0 = \frac{1}{2T}\int_0^{2T} F(t)dt$, $A_p = \frac{1}{T}\int_0^{2T} \cos(\omega_p t) F(t)dt$, $B_p = \frac{1}{T}\int_0^{2T} \sin(\omega_p t) F(t)dt$.

The random solution $\Delta R_r(t)$ then turns out to be (see 2004 for further details)[4, 5]

$$\Delta R_r(t) = U_0 + \sum_{p=1}^{\infty} \cos\left(\frac{p\pi}{T}t\right) U_p + \sum_{p=1}^{\infty} \sin\left(\frac{p\pi}{T}t\right) V_p \quad \text{(SM-5)}$$

where,

$$U_p = L_p\left(\Gamma A_p - \zeta\omega_p B_p\right) \quad \text{(SM-6)}$$

$$V_p = L_p\left(\zeta\omega_p A_p + \Gamma B_p\right) \quad \text{(SM-7)}$$

where $\omega_p = \frac{p\pi}{T}$ and

$$L_p = \left(\Gamma^2 + \zeta^2 {\omega_p}^2 I\right)^{-1} \quad \text{(SM-8)}$$

Since $F_r(t)$ has zero mean, it follows that the coefficients A_p, B_p, U_p, V_p and thus $\Delta R_r(t)$ are all zero mean random processes. To compute the correlation, we first simplify the terms

$$\cos\frac{p\pi}{T}(t+\tau) = \cos\left(\frac{p\pi}{T}t\right)\cos\left(\frac{p\pi}{T}\tau\right) - \sin\left(\frac{p\pi}{T}t\right)\sin\left(\frac{p\pi}{T}\tau\right)$$

$$\sin\frac{p\pi}{T}(t+\tau) = \cos\left(\frac{p\pi}{T}t\right)\cos\left(\frac{p\pi}{T}\tau\right) + \sin\left(\frac{p\pi}{T}t\right)\sin\left(\frac{p\pi}{T}\tau\right)$$

to get

$$\frac{1}{2T}\int_0^{2T} \Delta R_r(t)\Delta R_r^{\,T}(t+\tau)\,dt = U_0 U_0^{\,T} + \sum_{p=1}^{\infty} \cos\left(\frac{p\pi}{T}\tau\right)\left(U_p U_p^{\,T} + V_p V_p^{\,T}\right) \quad \text{(SM-9)}$$

We take ensemble averages

$$\frac{1}{2T}\int_0^{2T} \left\langle \Delta R_r(t)\Delta R_r^{\,T}(t+\tau)\right\rangle dt = \left\langle U_0 U_0^{\,T}\right\rangle + \sum_{p=1}^{\infty} \cos\left(\frac{p\pi}{T}\tau\right)\left(\left\langle U_p U_p^{\,T}\right\rangle + \left\langle V_p V_p^{\,T}\right\rangle\right) \quad \text{(SM-10)}$$

We then have

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.15.575718; this version posted January 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

$$\begin{aligned}\left\langle U_p {U_p}^T \right\rangle &= L_p \left\langle \left(\Gamma A_p - \zeta\omega_p B_p\right)\left({A_p}^T - \zeta\omega_p B_p T\right)\right\rangle L_p \\ &= L_p\left[\Gamma\left\langle A_p {A_p}^T\right\rangle\Gamma - \zeta\omega_p\Gamma\left\langle A_p {B_p}^T\right\rangle - \zeta\omega_p\left\langle B_p {A_p}^T\right\rangle\Gamma + \zeta^2{\omega_p}^2\left\langle B_p {B_p}^T\right\rangle\right]L_p\end{aligned} \quad \text{(SM-11)}$$

We now compute the ensemble averages $\left\langle A_p {A_p}^T \right\rangle$, etc.

$$\begin{aligned}\left\langle A_p {A_p}^T \right\rangle &= \frac{1}{T^2}\int_0^{2T}\int_0^{2T}\cos\left(\omega_p t\right)\cos\left(\omega_p s\right)\left\langle F_r\left(t\right){F_r}^T\left(s\right)\right\rangle dsdt \\ &= \frac{1}{T^2}\left(\int_0^{2T}\int_0^{2T}\cos\left(\omega_p t\right)\cos\left(\omega_p s\right)\delta\left(t-s\right)dsdt\right)I \\ &= \frac{1}{T^2}\left(\int_0^{2T}\cos^2\left(\omega_p t\right)dt\right)I = \frac{\zeta}{\beta T}I\end{aligned} \quad \text{(SM-12)}$$

with the identity matrix I. Similarly,

$$\left\langle A_0 {A_0}^T \right\rangle = \frac{\zeta}{2\beta T}I \quad \left\langle B_p {B_p}^T \right\rangle = \frac{\zeta}{\beta T}I \quad \left\langle A_p {B_p}^T \right\rangle = 0 \quad \left\langle B_p {A_p}^T \right\rangle = 0 \quad \text{(SM-13)}$$

This in turn gives

$$\left\langle U_p {U_p}^T \right\rangle = \left\langle V_p {V_p}^T \right\rangle = \frac{\zeta}{\beta T}L_p\left(\Gamma^2 + \zeta^2{\omega_p}^2 I\right) \text{ and } \left\langle U_0 {U_0}^T \right\rangle = \frac{\zeta}{2\beta T}\Gamma^{-2} \quad \text{(SM-14)}$$

where,

$$L_p = \left(\Gamma^2 + \zeta^2{\omega_p}^2 I\right)^{-1} \quad \text{(SM-15)}$$

The second equality follows similarly. Then

$$\lim_{T\to\infty}\frac{1}{2T}\int_0^{2T}\left\langle \Delta R_r\left(t\right)\Delta {R_r}^T\left(t+\tau\right)\right\rangle dt = \lim_{T\to\infty}\sum_{p=1}^{\infty}\frac{\zeta}{2\beta T}\cos\left(\frac{p\pi}{T}\tau\right)\left(\Gamma^2 + \zeta^2{\omega_p}^2 I\right)^{-1} \quad \text{(SM-16)}$$

Noting that $\Delta s = \frac{\left(p+1\right)\pi}{T} - \frac{p\pi}{T} = \frac{\pi}{T}$ as $T\to\infty$ the sum converges to the integral on the inteval $\left(0, 2T\right)$ and then on $\left(0,\infty\right)$. Summing up we get

$$\left\langle \Delta R_r\left(0\right)\Delta {R_r}^T\left(\tau\right)\right\rangle = \frac{1}{\pi}\int_0^{\infty}\cos\left(s\tau\right)\left(\Gamma^2 + \zeta^2 s^2 I\right)^{-1}ds = \frac{1}{2\beta}\Gamma^{-1}e^{-\frac{\tau}{\zeta}\Gamma} = \frac{1}{2\beta}\sum_{k=1}^{N}{\lambda_k}^{-1}e^{-\frac{\tau}{\zeta}\lambda_k}u_k {u_k}^T \quad \text{(SM-17)}$$

We now look at the periodic, deterministic term. Similar to the computations above, we get

$$\Delta R_p\left(t\right) = \cos\left(\omega t\right)L\Gamma F_0 + \sin\left(\omega t\right)\zeta\omega L F_0 \quad \text{(SM-18)}$$

with

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.15.575718; this version posted January 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

$$L = \left(\Gamma^2 + \zeta^2\omega^2 I\right)^{-1} \tag{SM-19}$$

It will be more informative to look at the components

$$\begin{aligned}\left(\Delta R_p\right)_i(t) &= \cos(\omega t)\left(L\Gamma F_0\right)_i + \sin(\omega t)\zeta\omega\left(LF_0\right)_i = A_i\cos(\omega t) + B_i\sin(\omega t) \\ &= D_i\cos\left(\omega t + \theta_i\right)\end{aligned} \tag{SM-30}$$

with $D_i = \sqrt{A_i^2 + B_i^2}$ and $\theta_i = \arctan\left(\frac{B_i}{A_i}\right)$. The correlation integrals then can be computed to yield

$$\begin{aligned}\left\langle \Delta R_i(t)\Delta R_j(t+\tau)\right\rangle &= \lim_{T\to\infty}\frac{1}{T}\int_0^T D_iD_j\cos\left(\omega t + \theta_i\right)\cos\left(\omega\left(t+\tau\right) + \theta_j\right)dt \\ &= \frac{1}{2}D_iD_j\cos\left(\omega t + \theta_j - \theta_i\right) = \frac{1}{2}D_iD_j\cos\left(\omega t + \theta_{ij}\right)\end{aligned} \tag{SM-21}$$

with $\tan\theta_{ij} = \frac{\tan\theta_j - \tan\theta_i}{1+\tan\theta_j\tan\theta_i} = \frac{B_jA_i - B_iA_j}{A_iA_j + B_iB_j}$. (SM-22)

Finally, as $\Delta R_r(t)$ is a random process with zero mean, all the mixed terms in the correlation will vanish to give:

$$\left\langle \Delta R_i(0)\Delta R_j(\tau)\right\rangle = \frac{1}{2\beta}\sum_{k=1}^{N}\lambda_k^{-1}e^{-\frac{\tau}{\zeta}\lambda_k}u_{ik}u_{jk}^T + D_iD_j\cos\left(\omega t + \theta_{ij}\right) \tag{SM-23}$$

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.15.575718; this version posted January 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

References

1. Atilgan, C., Computational methods for efficient sampling of protein landscapes and disclosing allosteric regions. *Advances in protein chemistry and structural biology* **2018,** *113*, 33-63.
2. Atilgan, C.; Atilgan, A. R., Perturbation-response scanning reveals ligand entry-exit mechanisms of ferric binding protein. *PLoS computational biology* **2009,** *5* (10), e1000544.
3. Bahar, I.; Atilgan, A. R.; Erman, B., Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Fold Des* **1997,** *2* (3), 173-181.
4. Erkip, A.; Erman, B., Dynamics of large-scale fluctuations in native proteins. Analysis based on harmonic inter-residue potentials and random external noise. *Polymer* **2004,** *45* (2), 641-648.
5. Hacisuleyman, A.; Erkip, A.; Erman, B.; Erman, B., Synchronous and Asynchronous Response in Dynamically Perturbed Proteins. *The Journal of Physical Chemistry B* **2021,** *125* (3), 729-739.
6. Scaramozzino, D.; Khade, P. M.; Jernigan, R. L., Protein Fluctuations in Response to Random External Forces. *Applied Sciences* **2022,** *12* (5), 2344.
7. Scaramozzino, D.; Piana, G.; Lacidogna, G.; Carpinteri, A., Low-Frequency Harmonic Perturbations Drive Protein Conformational Changes. *International Journal of Molecular Sciences* **2021,** *22* (19), 10501.
8. Botello-Smith, W. M.; Luo, Y., Robust determination of protein allosteric signaling pathways. *J Chem Theory Comput* **2019,** *15* (4), 2116-2126.
9. Haliloglu, T.; Hacisuleyman, A.; Erman, B., Prediction of allosteric communication pathways in proteins. *Bioinformatics* **2022,** *38* (14), 3590-3599.
10. Dokholyan, N. V., Controlling allosteric networks in proteins. *Chemical reviews* **2016,** *116* (11), 6463-6487.
11. Kaya, C.; Armutlulu, A.; Ekesan, S.; Haliloglu, T., MCPath: Monte Carlo path generation approach to predict likely allosteric pathways and functional residues. *Nucleic acids research* **2013,** *41* (W1), W249-W255.
12. Negre, C. F.; Morzan, U. N.; Hendrickson, H. P.; Pal, R.; Lisi, G. P.; Loria, J. P.; Rivalta, I.; Ho, J.; Batista, V. S., Eigenvector centrality for characterization of protein allosteric pathways. *Proceedings of the National Academy of Sciences* **2018,** *115* (52), E12201-E12208.
13. Wu, N.; Yaliraki, S. N.; Barahona, M., Prediction of protein allosteric signalling pathways and functional residues through paths of optimised propensity. *Journal of Molecular Biology* **2022,** *434* (17), 167749.
14. Lu, S.; Huang, W.; Zhang, J., Recent computational advances in the identification of allosteric sites in proteins. *Drug discovery today* **2014,** *19* (10), 1595-1600.
15. Stolzenberg, S.; Michino, M.; LeVine, M. V.; Weinstein, H.; Shi, L., Computational approaches to detect allosteric pathways in transmembrane molecular machines. *Biochimica et Biophysica Acta (BBA)-Biomembranes* **2016,** *1858* (7), 1652-1662.
16. Yao, X.-Q.; Hamelberg, D., Residue–residue contact changes during functional processes define allosteric communication pathways. *J Chem Theory Comput* **2022,** *18* (2), 1173-1187.
17. Ribeiro, A. A.; Ortiz, V., Determination of signaling pathways in proteins through network theory: importance of the topology. *J Chem Theory Comput* **2014,** *10* (4), 1762-1769.
18. Guarnera, E.; Berezovsky, I. N., Structure-based statistical mechanical model accounts for the causality and energetics of allosteric communication. *PLoS computational biology* **2016,** *12* (3), e1004678.
19. Kamberaj, H.; van der Vaart, A., Extracting the Causality of Correlated Motions from Molecular Dynamics Simulations. *Biophysical Journal* **2009,** *97* (6), 1747-1755.

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.15.575718; this version posted January 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

20. Hacisuleyman, A.; Erman, B., Causality, transfer entropy, and allosteric communication landscapes in proteins with harmonic interactions. *Proteins: Structure, Function, and Bioinformatics* **2017**.
21. Vatansever, S.; Erman, B.; Gumus, Z. H., Identifying Causality in Mutant KRas Residue Pairs from Molecular Dynamics Data Analysis. *Biophysical Journal* **2018,** *114* (3), 397a.
22. Vatansever, S.; Gumus, Z.; Erman, B., Intrinsic K-Ras dynamics: A novel molecular dynamics data analysis method shows causality between residue pairs. *bioRxiv* **2016**, 067496.
23. Bozovic, O.; Zanobini, C.; Gulzar, A.; Jankovic, B.; Buhrke, D.; Post, M.; Wolf, S.; Stock, G.; Hamm, P., Real-time observation of ligand-induced allosteric transitions in a PDZ domain. *Proceedings of the National Academy of Sciences* **2020,** *117* (42), 26031-26039.
24. Hyeon, C.; Lorimer, G. H.; Thirumalai, D., Dynamics of allosteric transitions in GroEL. *Proceedings of the National Academy of Sciences* **2006,** *103* (50), 18939-18944.
25. Buchenberg, S.; Sittel, F.; Stock, G., Time-resolved observation of protein allosteric communication. *Proceedings of the National Academy of Sciences* **2017,** *114* (33), E6804-E6811.
26. Wolf, S.; Sohmen, B.; Hellenkamp, B.; Thurn, J.; Stock, G.; Hugel, T., Hierarchical dynamics in allostery following ATP hydrolysis monitored by single molecule FRET measurements and MD simulations. *Chemical science* **2021,** *12* (9), 3350-3359.
27. Haliloglu, T.; Bahar, I.; Erman, B., Gaussian dynamics of folded proteins. *Physical Review Letters* **1997,** *79* (16), 3090-3093.
28. Fraser, J. S.; Clarkson, M. W.; Degnan, S. C.; Erion, R.; Kern, D.; Alber, T., Hidden alternative structures of proline isomerase essential for catalysis. *Nature* **2009,** *462* (7273), 669-673.
29. Wapeesittipan, P.; Mey, A. S.; Walkinshaw, M. D.; Michel, J., Allosteric effects in cyclophilin mutants may be explained by changes in nano-microsecond time scale motions. *Communications Chemistry* **2019,** *2* (1), 41.
30. Wapeesittipan, P.; Mey, A.; Walkinshaw, M.; Michel, J., Allosteric effects in catalytic impaired variants of the enzyme cyclophilin A may be explained by changes in nano-microsecond time scale motions. *bioRxiv* **2018**, 224329.
31. Doshi, U.; Holliday, M. J.; Eisenmesser, E. Z.; Hamelberg, D., Dynamical network of residue–residue contacts reveals coupled allosteric effects in recognition, catalysis, and mutation. *Proceedings of the National Academy of Sciences* **2016,** *113* (17), 4735-4740.
32. Lv, M.; Shi, T.; Mao, X.; Li, X.; Chen, Y.; Zhu, J.; Ni, S.; Shen, X.; Jiang, H.; Li, J., 1-(2, 6-Dibenzyloxybenzoyl)-3-(9H-fluoren-9-yl)-urea: a novel cyclophilin A allosteric activator. *Biochem Bioph Res Co* **2012,** *425* (4), 938-943.
33. Vu, P. J.; Yao, X.-Q.; Momin, M.; Hamelberg, D., Unraveling allosteric mechanisms of enzymatic catalysis with an evolutionary analysis of residue–residue contact dynamical changes. *ACS Catalysis* **2018,** *8* (3), 2375-2384.
34. Blackburn, E. A.; Wear, M. A.; Landré, V.; Narayan, V.; Ning, J.; Erman, B.; Ball, K. L.; Walkinshaw, M. D., Cyclophilin40 isomerase activity is regulated by a temperature-dependent allosteric interaction with Hsp90. *Bioscience reports* **2015,** *35* (5), e00258.
35. Holliday, M. J.; Camilloni, C.; Armstrong, G. S.; Vendruscolo, M.; Eisenmesser, E. Z., Networks of dynamic allostery regulate enzyme function. *Structure* **2017,** *25* (2), 276-286.
36. Dujardin, M.; Madan, V.; Gandhi, N. S.; Cantrelle, F.-X.; Launay, H.; Huvent, I.; Bartenschlager, R.; Lippens, G.; Hanoulle, X., Cyclophilin A allows the allosteric regulation of a structural motif in the disordered domain 2 of NS5A and thereby fine-tunes HCV RNA replication. *Journal of Biological Chemistry* **2019,** *294* (35), 13171-13185.
37. Rodriguez-Bussey, I.; Yao, X.-Q.; Shouaib, A. D.; Lopez, J.; Hamelberg, D., Decoding allosteric communication pathways in cyclophilin a with a comparative analysis of perturbed conformational ensembles. *The Journal of Physical Chemistry B* **2018,** *122* (25), 6528-6535.

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.15.575718; this version posted January 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

38. Lu, M.; Hou, G.; Zhang, H.; Suiter, C. L.; Ahn, J.; Byeon, I.-J. L.; Perilla, J. R.; Langmead, C. J.; Hung, I.; Gor'kov, P. L., Dynamic allostery governs cyclophilin A–HIV capsid interplay. *Proceedings of the National Academy of Sciences* **2015,** *112* (47), 14617-14622.
39. Eisenmesser, E. Z.; Millet, O.; Labeikovsky, W.; Korzhnev, D. M.; Wolf-Watz, M.; Bosco, D. A.; Skalicky, J. J.; Kay, L. E.; Kern, D., Intrinsic dynamics of an enzyme underlies catalysis. *Nature* **2005,** *438* (7064), 117-121.