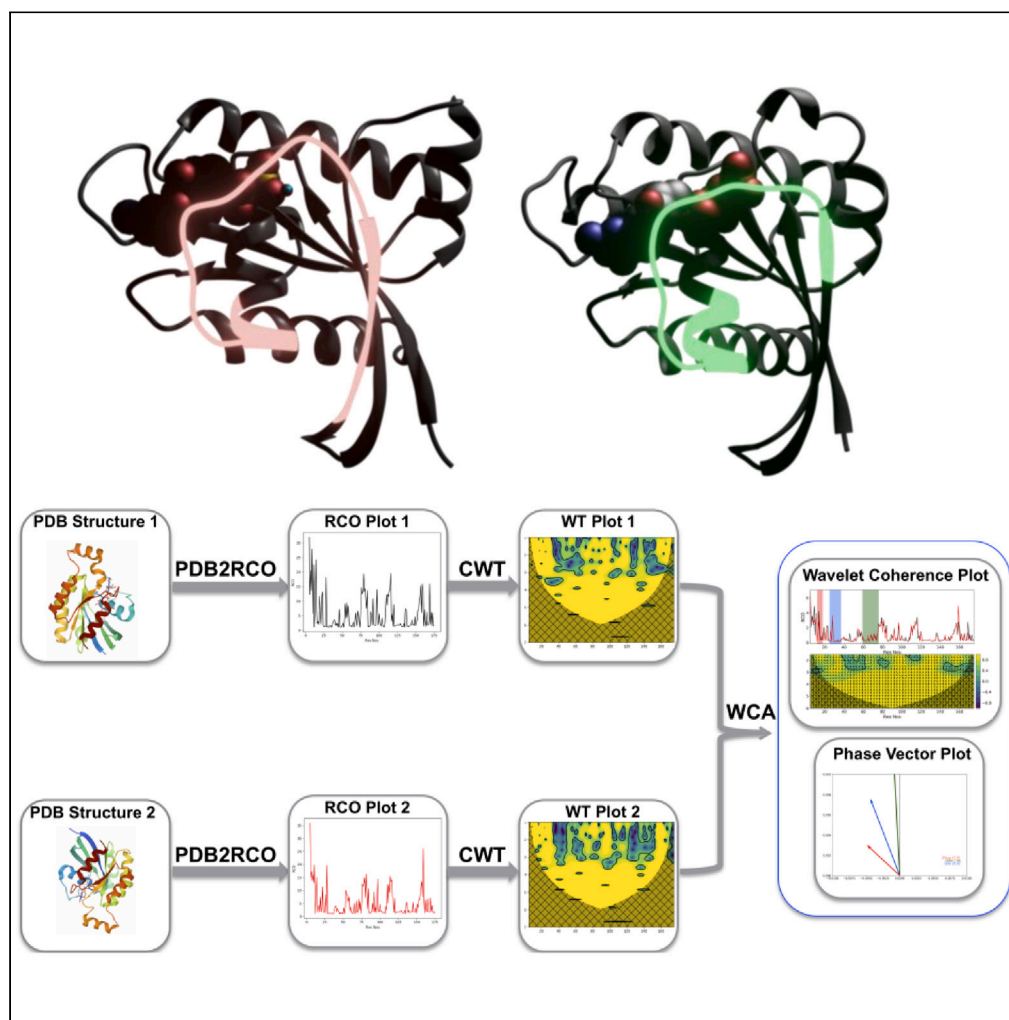


Article

Wavelet coherence phase analysis decodes the universal switching mechanism of Ras GTPase superfamily



Zenia Motiwala,
Anand S.
Sandholu, Durba
Sengupta, Kiran
Kulkarni

ka.kulkarni@ncl.res.in

Highlights

Structure & sequence comparisons fall short to decipher functional states of GTPases

Wavelet coherence analysis (WCA) can be applied to compare protein structures

Wavelet coherence phase couplings could define the functional states of GTPases

Functional aberrations in the oncogenic mutants of GTPase can be unraveled from WCA

Article

Wavelet coherence phase analysis decodes the universal switching mechanism of Ras GTPase superfamily

Zenia Motiwala,^{1,2,4} Anand S. Sandholu,^{1,2,4} Durba Sengupta,^{2,3} and Kiran Kulkarni^{1,2,5,*}

SUMMARY

The Ras superfamily of GTPases regulate critical cellular processes by shuttling between GTP-bound ON and GDP-bound OFF states. This switching mechanism is attributed to the conformational changes in two loops, SWI and SWII, upon GTP binding and hydrolysis. Since these conformational changes vary across the Ras superfamily, there is no generic parameter to define their functional states. A unique wavelet coherence (WC) analysis-based approach developed here shows that the structural changes in switch regions could be mapped onto the wavelet coherence phase couplings (WPCs). Thus, WPCs could serve as unique parameters to define their functional states. Disentanglement of WPCs in oncogenic GTPases shows how breakdown of structural allostery leads to their aberrant function. These observations stand out even for simulated ensemble of switch region conformers. Overall, for the first time, we show that WPCs could unravel the latent structural deviations in Ras proteins to decode their universal switching mechanism.

INTRODUCTION

Proteins utilize reversible transformations, such as covalent modifications (example: phosphorylation) and ligand-induced conformational changes, to exert signal transduction events. For instance, Ras superfamily of GTPases act as biomolecular switches to regulate diverse biological functions such as cell migration, proliferation, and fate. They achieve this switching mechanism by shuttling between GDP bound “OFF” and GTP bound “ON” states (Figure 1A). Comparative analysis of different states of protein structures has found to be a useful technique to gain functional insights. However, in a few cases, simple structure comparison has proved to be inadequate to derive generalized inferences, as the observed differences between structures are either not quantifiable or they exhibit complex incomprehensible patterns. This is well exemplified in the Ras superfamily of GTPases, which comprises of 5 families: Rho, Ras, Rab, Ran, and Arf GTPase.^{1–3} All the members of this superfamily possess the conserved nucleotide binding G domain, which harbors five conserved sequence motifs termed G1–G5 that stabilize the nucleotide-protein interactions. Of these, the P loop, GxxxxGKT (G1), and the two switch regions (SWI and SWII), containing the xT/Sx (G2) and DxxG (G3) motifs, interact with phosphates of the nucleotide to function as sensors for the presence of the gamma-phosphate.^{1–3} Structures of GDP- and GTP-bound proteins have shown that primarily the SWI and SWII regions of the protein undergo conformational changes to deploy the conserved “loaded-spring” mechanism (see Figure S1). Thus, it was inferred that the functional states of Ras superfamily GTPases are the consequence of the conformational changes in their SWI and SWII regions.^{4,5}

However, the quantum of conformational change in SWI and SWII regions, upon change in their bound nucleotides, varies across different sub-families. For example, among GTPases considered here, Rho, Ras, Rab, and Arf, the Rho family exhibits insignificant changes in the conformation of SWI and SWII regions, whereas in the other three, it is little more dramatic (Figures 2A–2D and S2). Therefore, correlating the functional states of GTPases from their SWI and SWII conformations is really challenging. Furthermore, this classical way of ascertaining their functional states would be inaccurate, as GTPases have been proposed to exhibit intermediate conformational states such as “ON-like” state.^{6–12} Furthermore, the correlation of functional states of GTPase with the structural changes in them becomes even more pertinent while deciphering the impact of oncogenic mutants on their functioning, as the crystal structures of mutants at positions 12, 13, or 61 (numbering as per Cdc42 sequence) do not show significant conformational deviation

¹Division of Biochemical Sciences, CSIR-National Chemical Laboratory, Pune 411008, India

²Academy of Scientific and Innovative Research (AcSIR), Ghaziabad 201002, India

³Division of Physical and Material Chemistry, CSIR-National Chemical Laboratory, Pune 411008, India

⁴These authors contributed equally

⁵Lead contact

*Correspondence: ka.kulkarni@ncl.res.in

<https://doi.org/10.1016/j.isci.2023.107031>



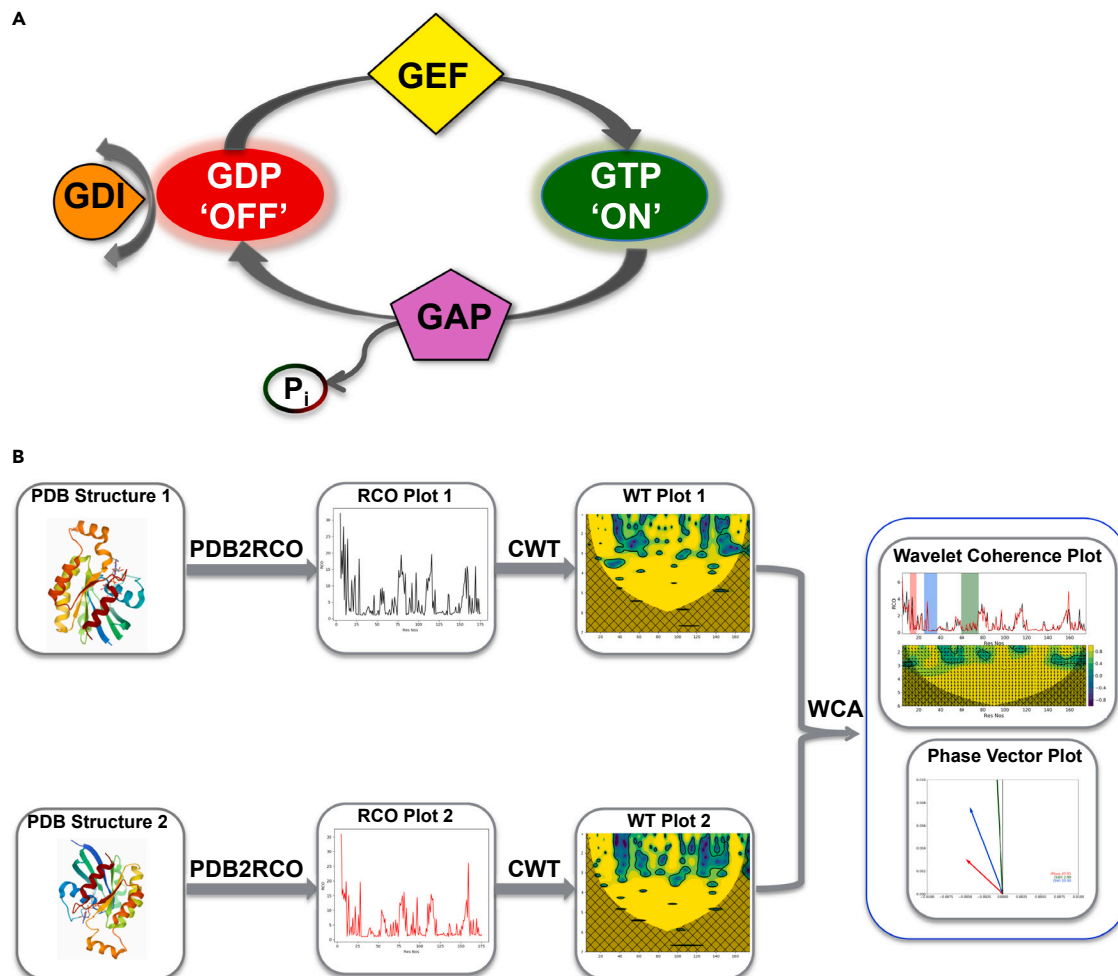


Figure 1. Schematic representation of GTPase switching and wavelet based methodology to study this switching in the current study

(A) Switching mechanism of the Ras superfamily GTPases. The switching between GTP-bound ON and the GDP-bound OFF state is regulated by guanine nucleotide exchange factors (GEFs) and GTPase activating proteins (GAPs); GEFs catalyze the exchange of their bound GDP with GTP and GAPs promote intrinsic GTP hydrolysis. The Rho GTPase subfamily, has an additional level of regulation involving guanine dissociation inhibitors (GDIs), which sequester GDP-bound Rho GTPases in the cytosol.

(B) Flowchart of the wavelet based methodology developed for analyzing the GTPases.

from their wild-type structures (Figures 3A, 3B, and S2) and some of the mutants, despite not falling in either SWI or SWII regions affect their functional states. Molecular dynamics (MD) simulations^{6–8,10,12–15} and statistical coupling analysis of evolutionarily conserved residues have attempted to address these issues by suggesting an allosteric model where the coupling of different structural regions dictate the dynamics and hence the functioning of GTPases and aberrations in the coupling due to mutations would result in their abnormal functioning. However, these techniques do not clearly bring out the unified *modus operandi* of GTPases, manifested in terms of the structural allostery. In this paper, we have developed a unique wavelet-based tool (Figure 1B) to address questions like how the unified “loaded-spring” mechanism manifests across the Ras superfamily, despite exhibiting significant sequence and structural divergence? And how do oncogenic mutations alter the loaded-spring mechanism?

Continuous wavelet transformation (CWT) has emerged as a powerful tool for feature extraction purposes, which can analyze localized and intermittent oscillations in the time series.^{16,17} Coherence analysis of wavelet transform (WT) of two signals, known as wavelet coherence analysis (WCA), simultaneously provides information on the relative changes between the signals and occurrence of these changes in the time domain, with reasonable resolution (see STAR Methods 2 and Figure S5). In the current study, we have mapped the three-dimensional (3D) structures of proteins into one-dimensional(1D) signals and

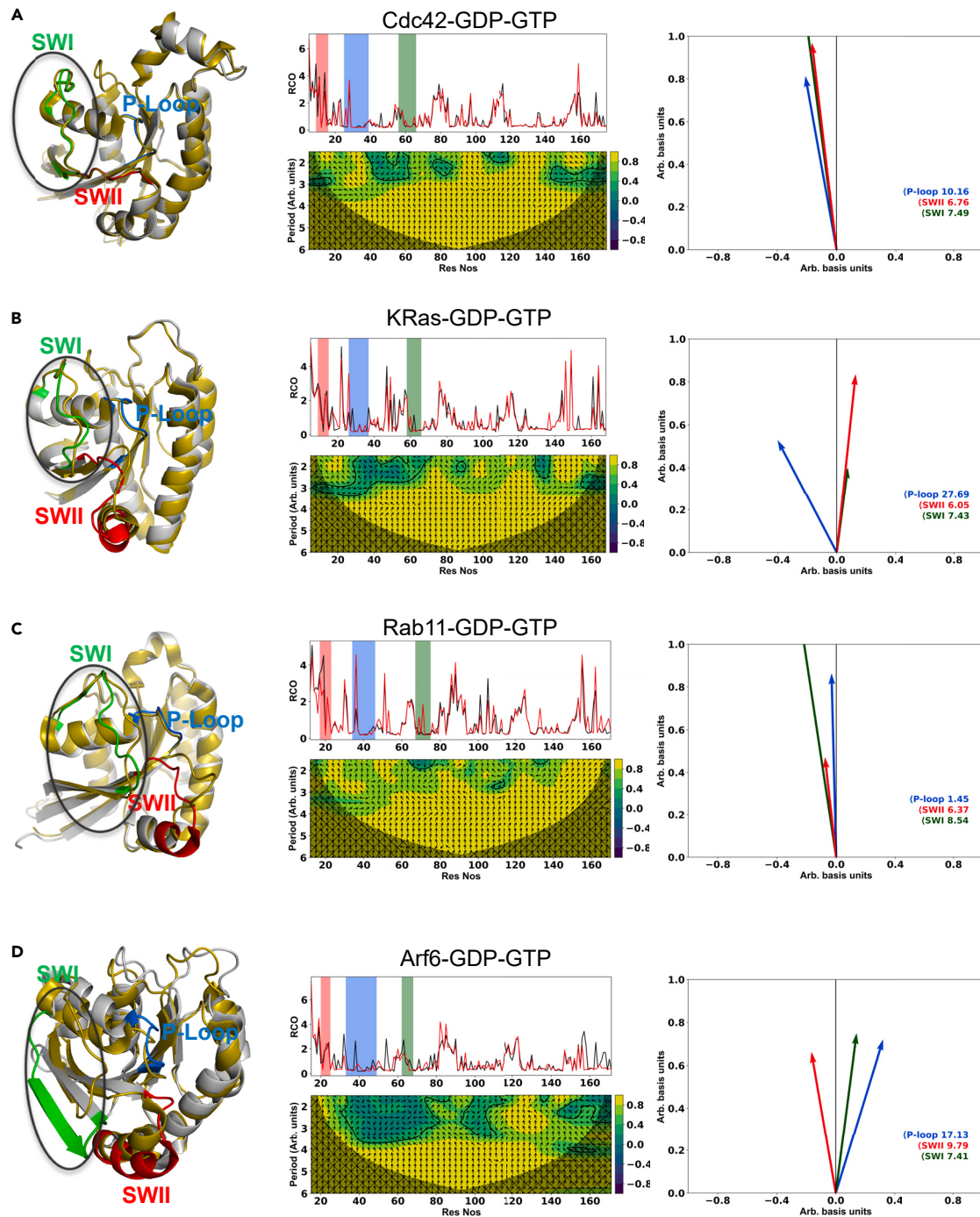


Figure 2. Superposition of GTP- and GDP-bound structures of GTPases (left), residue contact order (RCO) & wavelet coherence (WC) plots (middle), and wavelet coherence phase (WCP) vector plots (right) of GTPase

(A) Cdc42 (PDB: 2QRZ and 1AN0) (B) KRas (PDB: 6G0D and 4OBE) (C) Rab11 (PDB: 1O1W and 1O1V) (D) Arf6 (PDB: 2J5X and 1E0S). In all the structure superimposition figures, GTP-bound structures are shown in olive, the GDP-bound structure is shown in gray, and SWI region is highlighted with a black circle.

applied WCA to unravel the latent conformational changes that influence their functional states. Our results show that indeed there is a unified switching mechanism in the Ras superfamily, which manifests as coupling between wavelet coherence (WC) phase angles corresponding to the SWI and SWII regions.

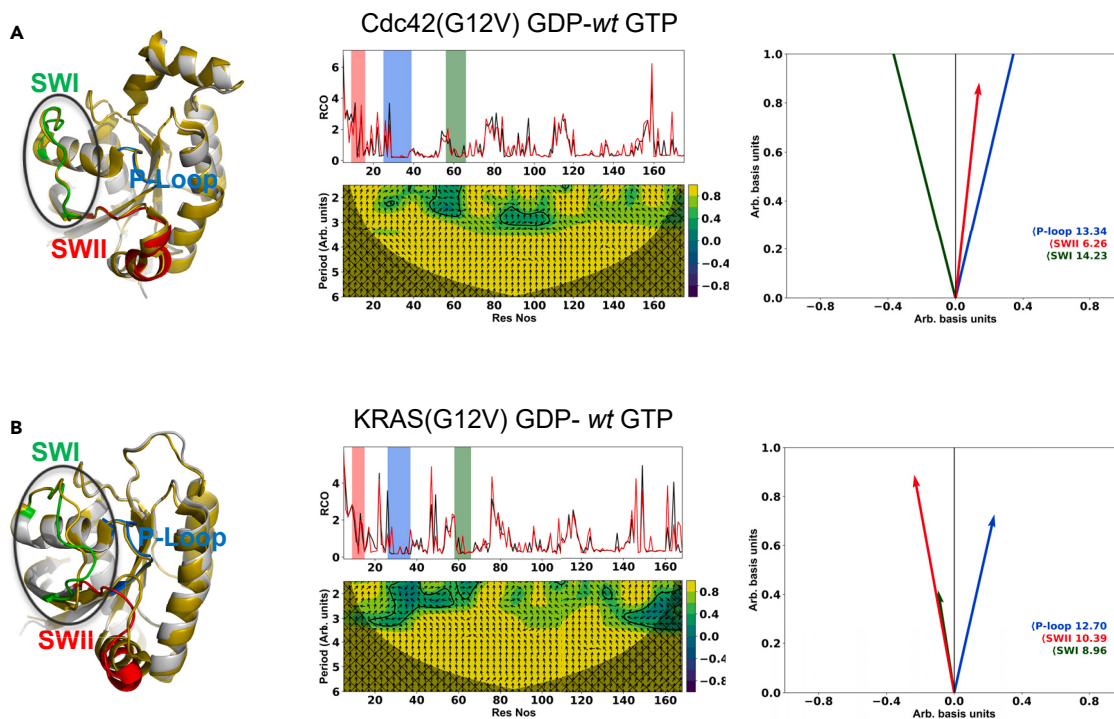


Figure 3. Structure superposition (left), RCO & WC plots (middle), and WCP vector plots (right) of Cdc42 and KRas G12V GDP-bound mutants with their corresponding GTP-bound wild-type structures

(A) Cdc42 (PDB: 2QRZ and 1A4R) (B) KRas (PDB: 6GOD and 4TQ9). Scheme of color coding for structures is same as in Figure 1.

We tested the statistical significance of the observed correlation by populating the SWI and SWII conformers through MD simulation. These tests further support our observations obtained from the comparison of just two (GTP and GDP) structures.

RESULTS AND DISCUSSION

Interpretation of wavelet transform of protein structures

To subject protein structures to WT analysis, we first transformed the 3D structure of the protein to a 1D signal¹⁸ by plotting the residue contact order (RCO) (see STAR Methods 1). RCO essentially represents change in the environment of a particular amino acid as a function of structural modulations and captures both short-range and long-range interactions in protein structures. However, for extended protein structures such as collagen and large helical transmembrane proteins, RCO representation may appear to be skewed. The choice of distance cut-off used in the RCO plot is critical in capturing the structural modulations. For example, a 2.5 Å cut-off would largely capture the short-range interactions, localized to the neighboring amino acids. On the other hand, a higher 10 Å cut-off would largely mask the local modulations. In this study, we found that 3.5 Å cut-off for calculating RCO to be appropriate to map the 3D structure (see Figure S3). This cut-off was found to work well with other globular proteins like Lysozyme and Hemoglobin (see Figures S3B and S3C). Next, we obtained the Fourier transform (FT) of the RCO (henceforth referred to as RCO-FT) as a first step to interpret the frequency content of structural modulations present in RCO (see Figure S3). For example, if 4th residue in the structure of interest has two interactions with 10th residue, and 12th residue has 4 interactions with 24th then the RCO for both the residues would be equal to 3 (Equation 1). On the contrary, there could exist other type of modulation such as 32nd residue interacting with the 36th residue and having 10 contacts. The RCO for this residue (32nd) would be 0.2. Thus, for this structure the FT of RCO would show two peaks, corresponding to 0.2 and 3, however, the positional (residue) information, i.e., where the modulation is happening is lost. Interestingly, from RCO-FT analysis of globular proteins we noticed that the meaningful signal in the Fourier space exists only up to 10 frequency units (see Figure S3). Furthermore, RCO-FT analysis provides hints on the optimal distance cut-off for calculating the RCO (see Figure S3). The frequency information gets quenched for lower (2.5 Å) as well as for very high (10 Å) distance cut-off values (see Figure S3).

In contrast with FT analysis, WT captures both modulation (frequency) and location (time) information simultaneously with reasonable resolution. Therefore, WT of RCO should provide deconvoluted modulations at every residue in terms of scales reflecting different types of interactions (both long-range and short-range) (see Figure S4). Thus, the co-relation between WT of two structures, known as WCA, would provide amount of relative change in the modulations at every amino acid and also the direction of propagation of the changes. Here the quantum of change is represented by the magnitude of correlation (Equation 6) and the direction of change is represented by the WC Phase (Equation 7).

Morlet wavelet captures the variations in the backbone conformation of the protein

Toward WC analysis of GTP- and GDP-bound structures, as a first step, we obtained continuous WT of individual structures from their respective RCO plots. Details of the features that could be captured by the CWT depend on factors like the choice of the mother wavelet and scales chosen for the transformation (see Figures S5 and S6). Here, we choose the Morlet wavelet (with $\omega_0 = 6$) and default scales (see Figure S5) used in the Python module of CWT (<https://pypi.org/project/pycwt/>), to obtain the CWT of the protein 1D signal. Hence, it was essential to check whether the chosen CWT parameters are good enough to extract the structural features of proteins. To achieve this, we sorted all known GTP- and GDP-bound Rac and Cdc42 structures (see Table S1), using unsupervised *Kmeans* clustering, from their CWT plots (STAR Methods 4). The outcome of this exercise would validate whether the CWTs used here are suitable enough to capture the conformational state of the GTPases. Our analysis shows that the chosen CWT parameters are optimal in capturing the conformational states of GTPases, as *Kmeans* clustering sorted the GTP- and GDP-bound structures with more than 80% efficiency (see Table S2).

Cross wavelet phase relationships show unique structural coupling between SWI and SWII regions of small GTPases

Next, we asked whether cross correlation between the wavelet transformations (WC analysis) of GTP- and GDP-bound Ras GTPases provide tangible parameters to correlate the structure and functional states of GTPases. Since for Cdc42, KRas, Rab11, and Arf6 GTPases, complete (without breaks in the chains) GDP- and GTP-bound structures were available, they were chosen for the current analysis (Figure 2). Like the structure superpositions, even the overlay of RCO plots by themselves do not show significant changes in the SWI, SWII, and P loop regions (Figure 2). On the contrary WCA of RCOs unravel the latent differential features of the structures. The y axis of the WC plots corresponds to an arbitrary unit of period, which in the current context is a determinant of the range over which a particular residue has “influence” and the x axis represents the residue numbers. The color gradient of the plot is a measure of the strength of structural correlations between the GTP- and GDP-bound states, which is in the range of -1 (100% inversely correlated) to $+1$ (100% correlated). The arrows (shown in black) in the plot indicate the phase angles, measured in counter-clockwise direction. The arrows pointing right and left correspond to a zero angle (in phase) and 180° (out-of-phase), respectively. Similarly, upward and downward pointing arrows correspond to part of the signal with 90° and 270° phase relationship, respectively.

The WC plots show changes in P loop, SWI, and SWII regions and these changes appear to be propagating from N- to C-terminal regions (Figures 2A–2D [middle]). The quantum of change in these regions, as indicated by the WC-power (color of the region), appears to be the same, which hints at the structural coupling between these regions. Another common feature in all of these plots is deviations at the C-terminus. Even though we have excluded about 5 residues at the N- and C-terminus of the structures, as they are floppy and disordered in many crystal structures, we see significantly correlated conformational differences at extreme regions. From these plots, quantifying absolute structural correlations of the individual regions in terms of WC power (color gradient) appears to be challenging, as they are heterogeneous across the GTPases considered here. However, given the definition of the signal as a 1D map of protein structure, the physical interpretation of arrows in the present context is the propagation of relative conformational change.

Next, we asked whether WC analysis delineates the convergent switching mechanism, as there are no conserved conformational signatures between different GTPase families. To address this, we looked at the collective propagation of phase vectors corresponding to the P loop, SWI, and SWII regions to explore the signatures of conserved structural couplings that result in an orchestrated motion of these functional regions upon a change in the bound nucleotide (Figure 2). For this, we calculated the resultant of the correlation-weighted phase vectors belonging to these regions (Equation 8), as the magnitude coherence

Table 1. Conformational deviations between the two different states of GTPases in terms of angles made by the resultant phase vectors corresponding to P loop, SWI, and SWII regions with y axis

| Structures used for the comparison | P loop | SWI | SWII |
|--|--------|--------------------|--------------------|
| Cdc42 ^{wt} (GTP) – Cdc42 ^{wt} (GDP) | 10.16 | 7.49* | 6.76* |
| KRas ^{wt} (GTP) – KRas ^{wt} (GDP) | 27.69 | 7.43* | 6.05* |
| Rab11 ^{wt} (GTP) – Rab11 ^{wt} (GDP) | 1.45 | 8.54* | 6.37* |
| Arf6 ^{wt} (GTP) – Arf6 ^{wt} (GDP) | 17.13 | 7.41 ⁺ | 9.79 ⁺ |
| Cdc42 ^{wt} (GTP) – Cdc42 ^{G12V} (GDP) | 13.34 | 14.23 [¶] | 6.26 [¶] |
| KRas ^{wt} (GTP) – KRas ^{G12V} (GDP) | 12.70 | 8.96 [¶] | 10.39 [¶] |
| Cdc42 ^{morphed} (GTP) – Cdc42 ^{wt} (GDP) | 13.14 | 21.29 [§] | 11.19 [§] |
| Cdc42 ^{wt} (GDP) – Cdc42 ^{G12V} (GDP) | 7.68 | 22.79 | 11.53 |
| KRas ^{wt} (GDP) – KRas ^{G12V} (GDP) | 0.80 | 1.73 | 3.45 |

The GTPases that show similar trend in their SWI and SWII phase angles are highlighted with *, the divergent Arf6 is highlighted with +. G12V mutants are highlighted with ¶ and the phase angles for the morphed Cdc42 (GTP) is highlighted with §.

provides the strength of change between the structures and the phase vector provides the direction of change. Since from Fourier space, we can see that the meaningful structural modulations exist only up to 10 units of frequency (see Figure S3), the phase vectors corresponding to the frequency levels greater than 10 were omitted for the calculations. Thus the resultant vector for a particular region (set of amino acids) provides a measure of net conformational change and the direction indicates the cumulative correlated change of that region. From the definition of phase vectors, the net conformational propagation, in terms of the angle between the vector corresponding to a particular region and the y axis would be zero if there is absolutely no coordinated conformational change between the GDP- and GTP-bound states and a finite value of this angle serves as an index to compare the conformational changes across the GTPases.

Interestingly, phase vector plots (Figure 2) show that there is a well-defined phase relationship between SWI and SWII regions (Table 1, highlighted with *), irrespective of the conformational changes seen in the crystal structures (Figures 2 and S2). For almost all the structures considered here, the angles between the Y axis and the resultant phase vector corresponding to SWI and SWII regions lie in the range of 7°–8° and 6°–7°, respectively. However, the SWII phase vector of Arf6 (Table 1, highlighted with +) deviates from this trend. Perhaps, this could be due to the role of additional regions in the Arf switching mechanism.¹⁹ Further, to test whether the observed phase vector relationship is a consequence of the switching mechanism or it is an artifact of the methodology, we randomly morphed the SWI region in Cdc42^{GDP} structure and plotted the WC phase vector plot between the wild-type GTP-bound and the GDP-bound morphed structures (Figure S7B). We clearly see that, in this instance, the conserved phase vector relationship is completely abolished, as the conformational changes introduced are random (Table 1, highlighted with §). Thus, the phase vector plots unambiguously bring out the conserved switching mechanism, which is not apparent in otherwise conformationally divergent Ras superfamily GTPases.

Oncogenic mutations in the P loop disrupt the structural coupling between the functional regions of the GTPases

Although it is well known that oncogenic point mutations at the 12th, 13th (P loop), and 61st (SWII), (residue numbering is with respect to the KRas structure) of the GTPases^{20,21} abrogate their GTP hydrolyzing capacity, the exact role of the mutants in disrupting the GTP hydrolysis is not clear. Several studies, employing X-ray crystallography^{22,23} and Nuclear magnetic resonance (NMR)^{6,15,24} have shown that there are no significant structural differences between the wild-type and mutant proteins, including their P loop and switch regions (RMSD between the mutant and wild-type structures is in the range 0.5–0.7 Å). Thus, in the absence of major structural rearrangements (Figures 3 and S2), just from structural analysis, it is difficult to explain how subtle changes at the nucleotide binding pocket alter the intrinsic hydrolyzing capacity of the GTPases. We probed this problem using WCA. The conserved phase coupling between SWI and SWII, present in wild-type proteins, gets completely disrupted in the G12V mutant structure (Figures 3A and 3B, Table 1 values highlighted with ¶, Figures S9A and S9B). However, the extent of disruption varies between Cdc42 and KRas. Here we are tempted to hypothesize that the disentanglement between the P loop

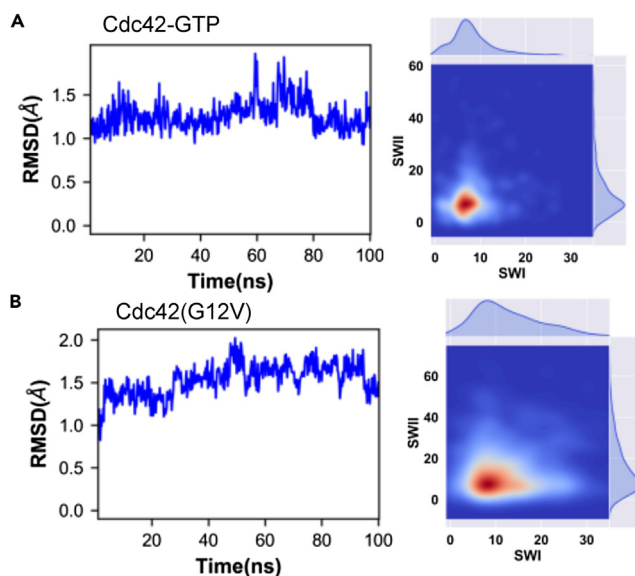


Figure 4. RMSD of MD trajectories of Cdc42 and KRas and bivariate distribution of SWI and SWII Wavelet coherence phase (WCP) angles

RMSD plots from MD simulations, corresponding to the GTP-bound structures are shown in left. Bivariate distribution of SWI and SWII WC phase angles obtained from the MD simulation trajectories sampled at every 50ps with the static GTP- or GDP-bound forms of the respective GTPase are shown in right.

(A) Cdc42 (B) GDP-bound G12V mutants of Cdc42. PDB codes for the structures are same as given in Figures 1 and 2.

and SWII regions, as observed in terms of WC phase coupling, could result in alterations in the solvent configuration of catalytically important residue Q61 (present in the SWII region), which could in turn affect the GTP hydrolysis.

MD simulations validate the WC phase coupling analysis

If we focus on experimentally resolved data, involving the comparison of just two end-point structures, it may introduce bias in our observations due to the limited sampling of the conformational space. Hence, we generated multiple conformers of one of the structures (either GTP- or GDP-bound) and analyzed the distribution of WC phase vectors between the static structure and conformers generated for the other structure. The idea here is not to expect highly correlated distribution of phase couplings but to examine the ability of WCA to detect the coupling in the large conformational samples as, however small may be, a fraction of MD trajectories would contain conformer that resembles their native (initial) states. Therefore, we tested the correlation between the wavelet phase vectors, corresponding to the SWI and SWII regions, for different conformers of both wild-type and mutant structures of KRas and Cdc42, generated through 100ns MD simulations. The trajectories were sampled at every 200 ps to obtain 500 different conformers for which phase vectors were calculated using the WCA tool (see Figure S7). This exercise (Figure 4) clearly shows that even for a larger conformational space, the phase angles corresponding to SWI and SWII regions follow the trend observed with just two structures. The phase coupling pattern could be discerned even from a dynamically larger sample size, obtained from 1 μ s simulation of KRas^{GTP} structure (see Figure S9C). For Cdc42, we also performed reverse analysis by keeping GTP-bound structure static and generating conformers for the GDP-bound structure using MD simulation. Even this exercise showed similar WC phase coupling between SWI and SWII regions (see Figure S9D). However, for mutants (Table 2; Figures S9A and S9B), the joint distribution appears to be getting broadened. The Pearson-R correlation values for the joint (bivariate) distribution of SWI and SWII WCA resultant phases angles for the wild-type proteins lie in the range of 0.4–0.8 [0.10 to 0.20 for mutants] (with p values in the range 0.05–0.9), suggesting a modest linear correlation between the distributions with poor statistical significance (Table 2). However, it is worth noting that these Pearson-R values must be taken with caution, as the exact nature of the correlation (linear or non-linear) between the distributions is not clear. Recently, “distance correlation” (DCORR) and “multiscale graph correlation” (MGC) have emerged as potential statistical tests to unravel the nature (geometry) of the relationship underlying multi-dimensional and/or nonlinear data.^{25–27} Therefore, we

Table 2. p-Values from different correlation tests performed on the wavelet phase angles distributions corresponding to SWI and SWII regions

| Structures used in MD Simulations | Counterpart Structure used for WCA | Pearson (Correlation) | Distance Correlation | MGC |
|--|------------------------------------|-----------------------|----------------------|-------|
| Cdc42 ^{wt} (GTP) | Cdc42 ^{wt} (GDP) | 0.026(0.38) | 0.0001 | 0.009 |
| Cdc42 ^{wt} (GDP) | Cdc42 ^{wt} (GTP) | 0.432(0.44) | 0.0051 | 0.004 |
| KRas ^{wt} (GTP) | KRas ^{wt} (GDP) | 0.212(0.61) | 0.0028 | 0.020 |
| KRas ^{wt} (GTP) with 1 μ s Simulation | KRas ^{wt} (GDP) | 0.05(0.78) | 0.0091 | 0.009 |
| Cdc42 ^{G12V} (GDP) | Cdc42 ^{wt} (GTP) | 0.13(0.11) | 0.2130 | 0.721 |
| KRas ^{G12V} (GDP) | KRas ^{wt} (GTP) | 0.889(0.18) | 0.8840 | 0.851 |

For Pearson, correlation values are shown in the parenthesis.

performed the MGC test (Table 2; Figures S9C and S9D) on our dataset to further validate the correlations between phase angles. For the wild-type protein, the MGC p value lies in the range of 0.004–0.009, with an exception for KRas short (100 ns) MD simulation data. On the contrary, mutant MD data indicate poor correlation between the SWI and SWII phase angle distribution with MGC p values in the range of 0.8 to 0.7 (Table 2). These statistical tests suggest that, indeed the distributions of WCA phase angles have a meaningful correlation for the wild-type proteins, which breaks down in the mutants that show aberrations in their switching action. This observation merits further investigations involving other GTPases.

Conclusions

It is evident that the function of proteins manifests in their sequence, structure and dynamics. However, in some cases, such as for Ras superfamily GTPases, simple structure and sequence comparison would be inadequate to decipher their functional states. Hence, novel computational approaches need to be developed to address this type of problem. Here we report one such approach involving WC analysis to dissect the functional states of Ras superfamily GTPases. These observations even stand out among a larger sample of conformers.

The methodology developed here is powerful enough to bring out the subtle dynamical aspects of proteins and it is computationally inexpensive. Thus, the study presented here could be a stepping-stone in delineating more complex dynamical aspects of protein structures and can be extended to other classes of proteins, such as kinases, which exhibit conformational changes upon ligand binding.

Limitations of the study

This study offers a unique strategy for exploring the operational states of proteins, but its applicability is restricted by the need for multiple structures of the same protein, captured in different functional states. Furthermore, this technique cannot eliminate bias in the analysis arising due to the conformational differences between the structures, introduced by the experimental artifacts. These aspects potentially limit the generalizability of the technique.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead contact
 - Materials availability
 - Data and code availability
- METHOD DETAILS
 - Conversion of 3D protein structures to 1D signal
 - Fundamentals of wavelet transform
 - Wavelet transformation
 - Unsupervised clustering of GTPase structures
 - Wavelet coherence analysis
 - MD simulation methods

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2023.107031>.

ACKNOWLEDGMENTS

K.K. would like to acknowledge grant from Department of Biotechnology, India (BT/PR12502/BRB/10/1387/2015) and Z.M. would like to acknowledge fellowship from University Grants Commission, India. A.S.S. would like to thank Council of Scientific & Industrial Research (CSIR) for fellowship (HCP0008). D.S. thanks DBT Bioinformatics Center NCL for support.

AUTHOR CONTRIBUTIONS

Z.M. and A.S.S. performed data analysis. D.S. assisted with MD simulation analysis. K.K. designed the project, supervised overall work and wrote the paper with inputs from Z.M., A.S.S., and D.S.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: August 16, 2021

Revised: March 6, 2023

Accepted: May 31, 2023

Published: June 7, 2023

REFERENCES

- Goitre, L., Trapani, E., Trabalzini, L., and Retta, S.F. (2014). The ras superfamily of small GTPases: the unlocked secrets. *Methods Mol. Biol.* 1120, 1–18. https://doi.org/10.1007/978-1-62703-791-4_1.
- Valencia, A., Chardin, P., Wittinghofer, A., and Sander, C. (1991). The ras protein family: evolutionary tree and role of conserved amino acids. *Biochemistry* 30, 4637–4648. <https://doi.org/10.1021/bi00233a001>.
- Wennerberg, K., Rossman, K.L., and Der, C.J. (2005). The Ras superfamily at a glance. *J. Cell Sci.* 118, 843–846. <https://doi.org/10.1242/jcs.01660>.
- Pylypenko, O., Hammich, H., Yu, I.M., and Houdusse, A. (2018). Rab GTPases and their interacting protein partners: structural insights into Rab functional diversity. *Small GTPases* 9, 22–48. <https://doi.org/10.1080/21541248.2017.1336191>.
- Vetter, I.R., and Wittinghofer, A. (2001). The guanine nucleotide-binding switch in three dimensions. *Science* 294, 1299–1304. <https://doi.org/10.1126/science.1062023>.
- Pálffy, G., Vida, I., and Perczel, A. (2019). 1H, 15N backbone assignment and comparative analysis of the wild type and G12C, G12D, G12V mutants of K-Ras bound to GDP at physiological pH. *Biomol NMR Assign.* <https://doi.org/10.1007/s12104-019-09909-7>.
- Gorfe, A.A., Grant, B.J., and McCammon, J.A. (2008). Mapping the nucleotide and isoform-dependent structural and dynamical features of ras proteins. *Structure*. <https://doi.org/10.1016/j.str.2008.03.009>.
- Kapoor, A., and Travestet, A. (2014). Mechanism of the exchange reaction in HRAS from multiscale modeling. *PLoS One* 9, e108846. <https://doi.org/10.1371/journal.pone.0108846>.
- Kumawat, A., Chakrabarty, S., and Kulkarni, K. (2017). Nucleotide dependent switching in Rho GTPase: conformational heterogeneity and competing molecular interactions. *Sci. Rep.* 7, 45829. <https://doi.org/10.1038/srep45829>.
- Muraoka, S., Shima, F., Araki, M., Inoue, T., Yoshimoto, A., Ijiri, Y., Seki, N., Tamura, A., Kumasaka, T., Yamamoto, M., and Kataoka, T. (2012). Crystal structures of the state 1 conformations of the GTP-bound H-Ras protein and its oncogenic G12V and Q61L mutants. *FEBS Lett.* 586, 1715–1718. <https://doi.org/10.1016/j.febslet.2012.04.058>.
- Anand, B., Majumdar, S., and Prakash, B. (2013). Structural basis unifying diverse GTP hydrolysis mechanisms. *Biochemistry* 52, 1122–1130. <https://doi.org/10.1021/bi3014054>.
- Prakash, P., and Gorfe, A.A. (2013). Lessons from computer simulations of Ras proteins in solution and in membrane. *Biochim. Biophys. Acta* 1830, 5211–5218. <https://doi.org/10.1016/j.bbagen.2013.07.024>.
- Li, H., Yao, X.Q., and Grant, B.J. (2018). Comparative structural dynamic analysis of GTPases. *PLoS Comput. Biol.* 14, e1006364. <https://doi.org/10.1371/journal.pcbi.1006364>.
- Shurki, A., and Warshel, A. (2004). Why does the ras switch “break” by oncogenic mutations? *Proteins* 55, 1–10. <https://doi.org/10.1002/prot.20004>.
- Chandrashekar, R., Salem, O., Krizova, H., McFeeters, R., and Adams, P.D. (2011). A switch I mutant of Cdc42 exhibits less conformational freedom. *Biochemistry* 50, 6196–6207. <https://doi.org/10.1021/bi2004284>.
- Lee, H.K., and Choi, Y.S. (2019). Application of continuous wavelet transform and convolutional neural network in decoding motor imagery brain-computer Interface. *Entropy* 21, 1199. <https://doi.org/10.3390/e21121199>.
- Sato, T., Kajiwara, R., Takashima, I., and Iijima, T. (2018). Wavelet correlation analysis for quantifying similarities and real-time estimates of information encoded or decoded in single-trial oscillatory brain waves. In *Wavelet Theory and Its Applications (InTech)*. <https://doi.org/10.5772/intechopen.74810>.
- Chernick, M.R. (2001). Wavelet methods for time series analysis. *Technometrics* 43, 491. <https://doi.org/10.1198/tech.2001.s49>.
- Sztul, E., Chen, P.W., Casanova, J.E., Cherfils, J., Dacks, J.B., Lambright, D.G., Lee, F.J.S., Randazzo, P.A., Santy, L.C., Schürmann, A., et al. (2019). Arf GTPases and their GEFs and GAPs: concepts and challenges. *Mol. Biol. Cell* 30, 1249–1271. <https://doi.org/10.1091/mbc.E18-12-0820>.
- Pantsar, T. (2020). The current understanding of KRAS protein structure and dynamics. *Comput. Struct. Biotechnol. J.* 18, 189–198. <https://doi.org/10.1016/j.csbj.2019.12.004>.
- Hodge, R.G., Schaefer, A., Howard, S.V., and Der, C.J. (2020). RAS and RHO family GTPase mutations in cancer: twin sons of different mothers? *Crit. Rev. Biochem. Mol. Biol.* 55, 386–407. <https://doi.org/10.1080/10409238.2020.1810622>.

22. Rudolph, M.G., Wittinghofer, A., and Vetter, I.R. (1999). Nucleotide binding to the G12V-mutant of Cdc42 investigated by X-ray diffraction and fluorescence spectroscopy: two different nucleotide states in one crystal. *Protein Sci.* 8, 778–787. <https://doi.org/10.1110/ps.8.4.778>.
23. Kawazu, M., Ueno, T., Kontani, K., Ogita, Y., Ando, M., Fukumura, K., Yamato, A., Soda, M., Takeuchi, K., Miki, Y., et al. (2013). Transforming mutations of RAC guanine triphosphatases in human cancers. *Proc. Natl. Acad. Sci. USA* 110, 3029–3034. <https://doi.org/10.1073/pnas.1216141110>.
24. Smith, M.J., Neel, B.G., and Ikura, M. (2013). NMR-based functional profiling of RASopathies and oncogenic RAS mutations. *Proc. Natl. Acad. Sci. USA* 110, 4574–4579. <https://doi.org/10.1073/pnas.1218173110>.
25. Lyons, R. (2013). Distance covariance in metric spaces. *Ann. Probab.* 41. <https://doi.org/10.1214/12-AOP803>.
26. Shen, C., Priebe, C.E., and Vogelstein, J.T. (2020). From distance correlation to multiscale Graph correlation. *J. Am. Stat. Assoc.* 115, 280–291. <https://doi.org/10.1080/01621459.2018.1543125>.
27. Vogelstein, J.T., Bridgeford, E.W., Wang, Q., Priebe, C.E., Maggioni, M., and Shen, C. (2019). Discovering and deciphering relationships across disparate data modalities. *Elife* 8, e41690. <https://doi.org/10.7554/eLife.41690>.
28. Shivakumar, D., Williams, J., Wu, Y., Damm, W., Shelley, J., and Sherman, W. (2010). Prediction of absolute solvation free energies using molecular dynamics free energy perturbation and the opl force field. *J. Chem. Theory Comput.* 6, 1509–1519. <https://doi.org/10.1021/ct900587b>.
29. Bowers, K.J., Chow, E., Xu, H., Dror, R.O., Eastwood, M.P., Gregersen, B.A., Klepeis, J.L., Kolossvary, I., Moraes, M.A., Sacerdoti, F.D., et al. (2006). Scalable algorithms for molecular dynamics simulations on commodity clusters. In *Proceedings of the 2006 ACM/IEEE Conference on Supercomputing, SC'06*. <https://doi.org/10.1145/1188455.1188544>.
30. Ivanov, P.C., Amaral, L.A., Goldberger, A.L., Havlin, S., Rosenblum, M.G., Struzik, Z.R., and Stanley, H.E. (1999). Multifractality in human heartbeat dynamics. *Nature* 399, 461–465. <https://doi.org/10.1038/20924>.
31. Goldberger, A.L., Amaral, L.A.N., Hausdorff, J.M., Ivanov, P.C., Peng, C.K., and Stanley, H.E. (2002). Fractal dynamics in physiology: alterations with disease and aging. *Proc. Natl. Acad. Sci. USA* 99, 2466–2472. <https://doi.org/10.1073/pnas.012579499>.
32. Dremmin, I.M., Ivanov, O.V., and Nechitailo, V.A. (2007). Wavelets and their uses. *Phys. Usp.* 44, 447–478. <https://doi.org/10.1070/PU2001v044n05ABEH000918>.
33. Chen, J., and Chaudhari, N.S. (2006). *Statistical analysis of long-range interactions in proteins*. In *Proceedings of the 2006 International Conference on Bioinformatics & Computational Biology, BIOCOMP'06, Las Vegas, Nevada, USA*, pp. 296–302.
34. Kihara, D. (2005). The effect of long-range interactions on the secondary structure formation of proteins. *Protein Sci.* 14, 1955–1963. <https://doi.org/10.1110/ps.051479505>.
35. Cazelles, B., Chavez, M., Berteaux, D., Ménard, F., Vik, J.O., Jenouvrier, S., and Stenseth, N.C. (2008). Wavelet analysis of ecological time series. *Oecologia* 156, 287–304. <https://doi.org/10.1007/s00442-008-0993-2>.
36. Nelli, F., and Nelli, F. (2018). Deep learning with TensorFlow. In *Python Data Analytics*. https://doi.org/10.1007/978-1-4842-3913-1_9.
37. Grinsted, A., Moore, J.C., and Jevrejeva, S. (2004). Application of the cross wavelet transform and wavelet coherence to geophysical time series. *Nonlinear Process Geophys.* 11, 561–566. <https://doi.org/10.5194/npg-11-561-2004>.
38. Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W., and Klein, M.L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79, 926–935. <https://doi.org/10.1063/1.445869>.

STAR★METHODS

KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|--------------------------------|---|---|
| Software and algorithms | | |
| Wave Protein | This paper | https://github.com/drkakulkarni/WaveProtein |
| DOCRR & MGC | hyppo | https://hyppo.neurodata.io/contributing.html |
| Molecular dynamics simulations | Shivakumar et al. ²⁸ Bowers et al. ²⁹ | Schrödinger Software & Desmond v3.6 |
| PyCWT | Python Package Index | https://pypi.org/project/pycwt/ |

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to lead contact, Dr. Kiran Kulkarni (ka.kulkarni@ncl.res.in).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- All original code has been deposited on GitHub (GitHub: <https://github.com/drkakulkarni/WaveProtein>).
- The software and algorithms used in this study is described in the above section and the [key resources table](#) in details.
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

METHOD DETAILS

Conversion of 3D protein structures to 1D signal

Mathematically, a signal is nothing but changes in space or time, in some cases both. For example, an electrocardiogram (ECG) is a representation of a signal, which shows the electrical impulse passing through the heart as a function of time. Determination of fluctuations, such as discontinuities and breakdown points, in the signal will be of prime importance in delineating the information carried by the signal. Mathematical tools like Fourier or Wavelet transform (WT) aid in dissecting the meaningful information from the signals, which are otherwise not easily detectable. For instance WT has proved to be a powerful tool in accurate detection of QRS complex, T and P segments of ECG.^{30,31} These parameters determine how long the electrical wave takes to pass through the heart and pace of the wave to travel from one part of the heart to the other. This type of analysis provides information on the patho-physiology of the heart. Fourier Transform of any signal would give the frequency component present in it. Cardiac arrhythmias clearly reflect in the frequency based features of the ECG and have thus acted as a potential analytical tool for ECG. However, one of the major limitations of Fourier Transform is that the time resolution is lost in the frequency representation of the signal (see [Figure S2](#)). In simpler words, from simple Fourier Transform, it is difficult to know at what time point in the signal did the frequency change. Short term Fourier Transform (STFT) provides some recourse to this problem but suffers from major drawbacks.³² The supplementary section elaborates further on this discussion.

The advantages of WT are that it may decompose a signal directly according to the frequency domains, which would be then represented in the time domain (see [Figure S3](#)). Thus, both time and frequency information of the signal is retained. In simpler words, from WT one can know with reasonable resolution “when and how much” fluctuation occurred in the signal. Here, we have used WT techniques to determine both the quantum and the location of conformational changes in proteins, occurring due to change in their functional states. Proteins are the linear polymers of amino acids with distinct three-dimensional structures,

which define their function. Thus, three-dimensional structures of proteins can be looked as spatial changes of amino acids along the length of the polypeptide chain. For many class of proteins their functional states, like substrate/ligand bound-unbound & active-inactive states, could again be viewed as motions introduced in their polypeptide chains. To apply WT techniques on proteins we first transformed the three dimensional (3D) structure of the protein to a one dimensional signal¹⁸ (1D)¹⁸ by plotting the Residue Contact Order (RCO) for each of the residue against their respective positions. Mathematically, RCO of the i^{th} residue is defined as^{33,34}

$$RCO_i = \frac{1}{n} \sum_j^L |i - j| \delta_{ij} \quad (\text{Equation 1})$$

where n is the number of contacts between residue i and the others (j) belonging the polypeptide chain of length L , which lie within 3.5 Å of the i^{th} residue. Previous reports have defined RCO by excluding the contacts between the immediate neighbours of the residues.^{33,34} However, to generate a continuum model we have included the immediate neighbouring residues, as well, for the calculations. RCO captures the spatial separation between the residues, which may not have covalent interactions, hence, encodes long range interactions present in the protein structure.^{33,34}

Fundamentals of wavelet transform

Dissecting signals into smaller or sub-signals to extract features of interest is at the heart of signal processing. Mathematically, the objective is to decompose the signal and express it as a linear combination of other signals. These other signals are functions (eg. Sine, Cosine, Gaussian) that are often shifted and dilated.

i.e., if $X(t)$ is the signal and $\psi(t)$ is the other function then in terms of transformation the signal can be expressed as,

$$X(t) = \sum_n C_n \psi_n(t) \quad (\text{Equation 2})$$

To put it simply, if a pianist simultaneously presses multiple keys of the piano to produce a particular sound, then, using signal processing, this signal can be decomposed in terms of sound produced by each of the keys pressed.

The numbers of functions depend on the type of signal and the functions used to decompose it. These functions are also known as bases functions and are generally orthonormal in nature. In Fourier Transform, these bases functions are sinusoidal in nature. Thus, Fourier Transform of any function provides information on the frequency content of the signal. However, the localization information of the frequency in the time domain gets diminished.

For example, in [Figure S4](#), there are two waves (A, P) shown in the top panel. The first wave (see [Figure S4A](#)) has lower frequency content at the beginning 2 seconds and subsequently attains higher frequency content. On the contrary, the second wave (see [Figure S4P](#)) is the opposite of first wave. However, the Fourier Transform of both waves ([Figures S2B](#) and [S2Q](#)) is identical, as this method provides the frequency content present in the wave notwithstanding the occurrence of these frequencies in the time domain.

To obtain the localized frequency content, a Fourier Transform combined with a window $w(t - \tau)$ that moves over the signal is necessary. This type of Fourier Transform is termed as Short Time Fourier Transform (STFT). The primary limitation of STFT is that the resolution is bounded with the chosen window size. In simpler words, a smaller window would give better time resolution but compromise the frequency resolution and vice-versa. This problem is circumvented to a great extent in Wavelet Transform (WT).

In WT, the window is combined with the analyzing function and thus the information to be obtained depends on the analyzing function. This function is also known as the mother wavelet and it is not only translated along the signal but can also be dilated ([Figure S5](#)). The dilation of the mother wavelet (analyzing function) is known as scaling (s). The mother wavelets can be of different functions and in our current study we have used Morlet wave ([Equation 3](#)) as the mother wavelet with $\text{scale}=1$.

Wavelet transformation

As explained earlier, the wavelet transform decomposes a signal as a combination of a set of basis functions, obtained by means of dilation (s) and translation (τ) of a proto type wave ψ_0 , called as mother wavelet. As a characteristic, it has sliding windows that expand or compress to capture low- and high-frequency signals, respectively.¹⁸ CWT of a signal $X(t)$ is defined as

$$W_t(s, \tau) = \sum_{t=0}^{T-1} X(t) \psi_0^* \left(\frac{(t - \tau) \delta t}{s} \right) \quad (\text{Equation 3})$$

Where ψ_0^* indicates complex conjugate of ψ_0 . s represents the scale, by varying which the wavelet can be dilated or contracted. By translating along localized time position one can calculate the wavelet coefficients $W_t(s, \tau)$, which describe the contribution of the scales s to the time series $X(t)$ at different time positions t .³⁵

The mother wavelet, which forms the orthonormal basis for the transformation, could be of different forms. For the current study, we have use Morlet mother wavelet, which is defined as

$$\psi_0 = \pi^{-\frac{1}{4}} e^{i\omega_0 t} e^{-\frac{t^2}{2}} \quad (\text{Equation 4})$$

Here ω_0 is the wavenumber, which gives the number of oscillations within the wavelet. Wavelet coefficients are normalized to unit variance in order to allow direct comparisons of the wavelet coefficients. The local wavelet power, which describes how the special range of interaction of the residues in the structure varies along the residue positions, can be computed from wavelet coefficients as

$$P_{w(s, \tau)} = 2^s |W_t(s, \tau)|^2 \quad (\text{Equation 5})$$

If the signal has finite length, as in the present case, wavelet transformation of it introduces zero padding effects. These are the artefacts present in the transformation that get more pronounced at the ends of the signal. Hence the wavelet power spectrum at the edges, in the region known as cone of influence (COI) (Figure S6), is omitted for the analysis.³⁵

The COI is demarcated in Figures 2 and 3 (middle figures) and S6. Other than mother wavelet, choice of scale (s) and wave number (ω_0) also influences the feature detected by WT. Therefore, we obtained WT of RCOs with different s and ω_0 and found that $s=1$ and $\omega_0 = 6$ is optimal for current analysis (Figure S7).

Unsupervised clustering of GTPase structures

Python implementation of KRas image clustering was employed to cluster the wavelet transforms of the GTPase structures.³⁶ For this purpose, only main chain atoms of the proteins were used to calculate the RCOs and subsequent CWTs, as this would be sufficient to delineate the backbone conformation and at the same time mask the deviations in the structures due to mutations present in them. The COI of Wavelet transformation spectrums of the structures were masked to avoid influence of the artefacts in sorting.

Wavelet coherence analysis

Wavelet coherence analysis provides insights on the variations between two signals. As mentioned earlier, wavelets are useful in delineating the “when and how much” of the variations between signals with reasonable resolution. The strength of the covariation between two signals (here conformational changes between two structures of a given protein in different states), represented as, $x(t)$ and $y(t)$ signals, can be quantified from the Wavelet coherence (WC) analysis. WC can be obtained from their cross wavelet transform (w_{xy}), defined as

$$w_{xy}(s, \tau) = w_x(s, \tau) w_y^*(s, \tau) \quad (\text{Equation 6})$$

Where w_x is the CWT of the first structure [$x(t)$] and w_y^* is the complex conjugate of CWT of the second structure [$y(t)$]. The power of the cross-spectrum is modulus of the CWT, $|w_{xy}(s, \tau)|$. The statistical significance of wavelet coherence is obtained by using Monte Carlo randomization techniques.³⁷ The direction of the covariation between the signals can be obtained from the wavelet coherence phase (WCP), which can be calculated from the real, $R(w_{xy}(s, \tau))$, and the imaginary, $I(w_{xy}(s, \tau))$, parts of the cross wavelet transform as

$$\varphi_{xy}(s, \tau) = \tan^{-1} \left[\frac{I(w_{xy}(s, \tau))}{R(w_{xy}(s, \tau))} \right] \quad (\text{Equation 7})$$

In the present situation, WCP provides propagation of the conformational variation for each of the residues as a function of their position in space. To quantify the net phase for a particular region a coherence weighted vector sum of all the phases, corresponding to the residues in that region, was obtained as

$$\vec{\varphi}_{Axy} = \sum_{n=1}^{N \in A} w_{nc} \vec{\varphi}_{nxy} \quad (\text{Equation 8})$$

where $1 \dots N$ are the residues from the region A and $w_{nc} = w_{nx}(s, \tau) w_{ny}^*(s, \tau)$, strength of the coherence at the n^{th} residue.

The algorithm is represented as a flowchart in Figure 1B and it is realized as a C program and a Python script. Both can be accessed on our GitHub page: <https://github.com/drakulkarni/WaveProtein>.

To perform DOCRR and MGC test we used the Hyppo software package (<https://hyppo.neurodata.io/contributing.html>).

MD simulation methods

Molecular dynamics (MD) simulations were carried out using Desmond v3.6,^{28,29} implemented in Schrödinger-maestro 2020 (Schrödinger Release, 2019). Simulation system was prepared by placing the GTPase in an orthogonal box with 10.0Å buffered distance in all the directions, after assigning bonds. The hydrogen atoms were added to the protein keeping pH 7.0 of the model. The system was solvated with TIP3P³⁸ solvent atoms and neutralized by the addition of Na⁺/Cl⁻ ions. Subsequently, the system was subjected to energy minimization followed by 2000 steps of conjugate gradient algorithm with 120.0 kcal/mol/Å convergence threshold energy under NPT ensemble. Finally, simulation of the system was performed for 100 ns. Trajectory recorded in each 200 ps. For KRas GTP, a longer simulation of 1 μs was performed and the trajectories were sampled at every 1ns. All the simulations, except the 1 μs KRas simulation, were performed in triplicates.