

SCIENTIFIC REPORTS



OPEN

Spatio-temporal coordination among functional residues in protein

Sutapa Dutta¹, Mahua Ghosh¹ & J. Chakrabarti^{1,2}

Received: 13 May 2016

Accepted: 07 December 2016

Published: 16 January 2017

The microscopic basis of communication among the functional sites in bio-macromolecules is a fundamental challenge in uncovering their functions. We study the communication through temporal cross-correlation among the binding sites. We illustrate via Molecular Dynamics simulations the properties of the temporal cross-correlation between the dihedrals of a small protein, ubiquitin which participates in protein degradation in eukaryotes. We show that the dihedral angles of the residues possess non-trivial temporal cross-correlations with asymmetry with respect to exchange of the dihedrals, having peaks at low frequencies with time scales in nano-seconds and an algebraic tail with a universal exponent for large frequencies. We show the existence of path for temporally correlated degrees of freedom among the functional residues. We explain the qualitative features of the cross-correlations through a general mathematical model. The generality of our analysis suggests that temporal cross-correlation functions may provide convenient theoretical framework to understand bio-molecular functions on microscopic basis.

Quite often bio-macromolecules undergo cascade of ligand bindings at different sites. Such binding events not only control cellular processes, but also lie at the heart of technological applications with bio-molecules as scaffold¹. The microscopic basis of communication among the binding sites in bio-macromolecules is one of the fundamental questions which have drawn considerable attention, but still remains largely obscure. Motivated by this, we make attempt here to understand communication among functional residues in proteins, incorporating information of microscopic motions.

We consider the case of a small protein, ubiquitin² (Ub) (PDB id: 1UBQ, see Supplementary Fig. S1(a)) involved in ubiquitination^{3,4}, a process ubiquitous among the eukaryotes by which ubiquitin attaches with a target protein to degrade the latter. The process is initiated by covalent attachment of Adenosine mono-phosphate (AMP) to C-terminal Glycine, G76 of Ub. Following this, ubiquitin activation enzyme-E1 binds at different residues of Ub³⁻⁶, (see Supplementary Figs S1(b)–(d)). The question is: How do the spatially distant residues get temporally correlated so that the binding information at one site at a given time affects the binding at other sites at a later time?

Although experimental probes are limited^{7,8}, recent simulation works in this direction emphasize on the covariance (Pearson Correlation Coefficient) between the instantaneous values of microscopic degrees of freedom, like the angles between atomic planes known as dihedral angles in a protein⁹⁻¹⁵. Non-zero but very small values of Pearson Correlation Coefficient have been observed among the dihedral angles of functional but spatially distant residues in Ub⁹. However, information provided by Pearson Correlations is far from complete. Macromolecular binding takes place typically by rotational diffusion ranging in timescales of tens of nano-seconds (ns)¹⁶, so that the binding surfaces are mutually exposed. This means that the changes at sites upon first ligand binding must affect the downstream binding sites till this time. Such temporal information are absent in Pearson Correlation Coefficient. The information entropy transfer^{15,17} approaches has been proposed to causally connect residues depending upon the history of their correlated fluctuations which includes all non-linear coupling between the fluctuating variables. Not only that the computation of information entropy transfer is quite involved, but also the

¹Department of Chemical, Biological and Macro-Molecular Sciences, S. N. Bose National Centre for Basic Sciences, Sector III, Block JD, Salt Lake, Kolkata 700106, India. ²Also at Unit of Nanoscience and Technology-II and The Thematic Unit of Excellence on Computational Materials Science, S. N. Bose National Centre for Basic Sciences, Sector III, Block JD, Salt Lake, Kolkata 700106, India. Correspondence and requests for materials should be addressed to J.C. (email: jaydeb@bose.res.in)

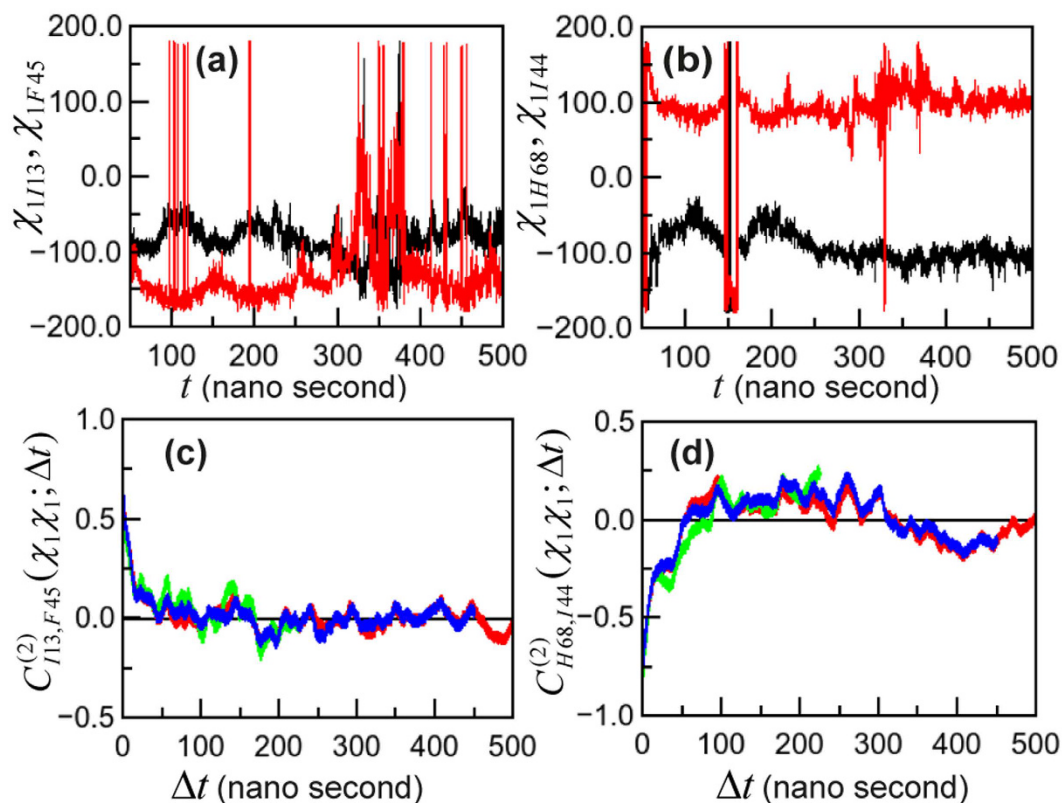


Figure 1. (a) Dihedral angles as functions of time t ; χ_{1I13} (black) and χ_{1F45} (red). (b) χ_{1H68} (black) and χ_{1I44} (red). (c) Convergence of TDCFs for three different $t = 500$ ns (green), 950 ns (blue) and 1.05 μ s (red); $C_{I13,F45}^{(2)}(\chi_1\chi_1; \Delta t)$. (d) $C_{H68,I44}^{(2)}(\chi_1\chi_1; \Delta t)$ as a function of Δt .

physical mechanisms leading to the non-linear couplings is not understood, primarily due to lack of experimental probes. Moreover, time is introduced in the formalism on an ad-hoc basis¹⁷.

Alternatively, temporal cross-correlations of fluctuations of two physical quantities $A(t)$ and $B(t')$ at times t and t' with respect to their mean values, also known as two-point correlations functions^{18,19}, are used to describe time scales of correlated stochastic processes²⁰. The equal time correlation function, $t = t'$, is the statistical Pearson Correlation. Temporal cross-correlation functions can be thought of generalization of Pearson correlation in time domain. Two-point correlation functions do not contain information on non-linear coupling. One major advantage of two-point correlation functions is that they are experimentally accessible by scattering techniques. Interestingly, temporal cross-correlations between fluorescence intensities show asymmetry with inversion in time which has been used to study co-localization of proteins²¹. This observation seems to suggest that temporally causal connections can be extracted from the two-point correlation functions. The power of two-point cross-correlation functions has not been exploited for in-depth understanding of bio-molecular phenomena. Time dependent dihedral cross-correlation functions (TDCF) have been employed for correlation between protein residues only up to a few hundred pico-seconds (ps)²², far too low compared to the bio-molecular binding time scales to have functional relevance. We have established in an earlier work²³ from much longer simulations that TDCF can relate large scale changes in a protein upon ligand binding.

With this backdrop we examine here the TDCF by long computer simulations and mathematical modeling to understand functional co-ordination among residues. We show that the TDCF can explain the causal connection between the functional residues of Ub in the time scale of tens of nano-seconds. We explain the qualitative features of TDCF using simple mathematical model. Interesting aspect of our result is that TDCFs contain temporal information which may be useful to understand biological processes which are orders of magnitude slower than atomic motions without invoking the non-linear effects.

Molecular Dynamics and TDCF

We perform 1.05 microseconds (μ s) long all-atom MD simulations²⁴ (see Methods, Note 1) for Ub with initial input from crystal structure² in explicit water. We analyze data using the portion of the simulated trajectory where the root mean squared deviation (RMSD) of the backbone atoms is saturated (excluding initial 50 ns, see Supplementary Fig. S2). We calculate the dihedral angles for backbone (φ , ψ) and side-chain (χ_1) of the residues of the protein. We plot the dihedrals χ_1 for residue pairs Isoleucine: I13 and Phenylalanine: F45 denoted by χ_{1I13} and χ_{1F45} respectively as functions of time in Fig. 1(a). I13 and F45 have backbone distance ($d_{\alpha-\alpha}$), given by that of their C_α atoms as large as 1.5 nm. Despite that both χ_{1I13} and χ_{1F45} exhibit correlated behavior: The increase in

one is coupled to increase in the other till very long time. Similarly, for Histidine: H68 and I44 with $d_{\alpha-\alpha} \sim 0.5$ nm, the plots of $\chi_{1_{H68}}$ of H68 and $\chi_{1_{I44}}$ of I44 as functions of time (Fig. 1(b)), reveal anti-correlated behavior even at long times.

Now we proceed to quantify temporal correlations between these time series. We extract the TDCF between dihedral θ of residue i and θ' of residue j in time interval Δt from equilibrated trajectory (see Supplementary Fig. S3). TDCF is denoted as $C_{i,j}^{(2)}(\theta\theta'; \Delta t)$, for a time interval $\Delta t = t_2 - t_1$ (see Methods, Eqs 1 and 2, Note 2). We compute TDCF from MD trajectory for three different sets of maximum time up to $t = 500$ ns, 950 ns and 1.05 μ s of the simulated trajectory. Since the correlation functions are computed from the dihedral values at two time intervals over trajectory, averaging at larger time interval gets better with longer time trajectory. We show $C_{I13,F45}^{(2)}(\chi_1\chi_1; \Delta t)$ (Fig. 1(c)) and $C_{H68,I44}^{(2)}(\chi_1\chi_1; \Delta t)$ (Fig. 1(d)) for three different cases. Both Fig. 1(c) and (d) show that data for trajectory up to $t = 500$ ns have differences with respect to larger time trajectories. However, data with trajectory up to $t = 950$ ns and 1.05 μ s are comparable, indicating saturation in the temporal behavior of the TDCFs.

We further report our analysis based on all the data for the longest trajectory, $t = 1.05 \mu$ s. Figure 2(a) and (b) bring out further non-trivial aspect of the TDCFs, exhibited by several dihedral pairs, despite large separation between the residues. We show TDCFs for both forward and reverse direction, obtained by interchanging i and j and θ and θ' for the longest trajectory. $C_{I13,F45}^{(2)}(\chi_1\chi_1; \Delta t)$ (Fig. 2(a)) shows statistical Pearson Correlation, $C_{I13,F45}^{(2)}(\chi_1\chi_1; 0) > 0$ at $\Delta t = 0$. This decays with increasing time interval. $C_{H68,I44}^{(2)}(\chi_1\chi_1; \Delta t)$, (Fig. 2(b)) shows statistical Pearson anti-correlation, $C_{H68,I44}^{(2)}(\chi_1\chi_1; 0) < 0$ at $\Delta t = 0$ following which it decays to zero for large Δt . We also observe in Fig. 2(a) and (b) that $C_{i,j}^{(2)}(\theta\theta'; \Delta t)$ is different in forward and reverse directions, the decay time scales being different indicating asymmetry in TDCFs.

Figure 2(c) shows representative cases of Laplace Transform $F_{i,j}(\theta\theta'; s)$ of $C_{i,j}^{(2)}(\theta\theta'; \Delta t)$, the correlation function (for additional cases see Supplementary Fig. S4). For small s , $F_{I13,F45}(\chi_1\chi_1; s)$ has a maximum where there is a statistical correlation (Fig. 2(a)), while $F_{H68,I44}(\chi_1\chi_1; s) < 0$ and having a minimum in case of statistical anti-correlation (Fig. 2(b)). The asymmetry in $F_{i,j}(\theta\theta'; s)$ under interchanges of i and j and θ and θ' is evident from Fig. 2(c). The peak value of $F_{i,j}(\theta\theta'; s)$ in s , $F_{i,j}^{max}(\theta\theta')$ is a measure of the strength, and the inverse of peak position gives a characteristic time scale $\tau_{i,j}^{\theta\theta'}$ of correlation. These time scales (see Supplementary Table S1) are tens of nanoseconds, in the regime of rotational diffusion time much larger than atomic fluctuation time scales. The log-log plots in Fig. 2(d) show decaying tail ($s^{-\kappa}$) with exponent κ for large s . The values of κ (see Supplementary Table S1) show quasi-universality ($\kappa \sim 1.0$).

Let us now examine the TDCFs for functionally relevant residues⁹ of ubiquitin. In recent simulation studies⁹ statistical Pearson correlations are observed between back-bone dihedrals of residue pairs (I13-F45, Threonine: T14-F45, Lysine: K6-F45 and I13-Leucine: L67, I13-Valine: V5-K6, K6-H68, H68-I44 and I44-F45) of ubiquitin, which belong to the binding surface patch of Ub. We plot $F_{i,j}^{max}(\theta\theta')$ and $\tau_{i,j}^{\theta\theta'}$ for all these residue pairs as functions of $C_{i,j}^{(2)}(\theta\theta'; 0)$ in Fig. 2(e) and (f) respectively. We observe that a strong correlation exists between peak values of TDCFs and $C_{i,j}^{(2)}(\theta\theta'; 0)$. However, the timescales are not correlated to $C_{i,j}^{(2)}(\theta\theta'; 0)$. This is not surprising, for statistical correlation coefficients do not contain temporal information.

Functional relevance

Our analysis yields a detailed map of correlated residue pairs R_i and R_j , as shown in Fig. 3(a). For a particular residue pair R_i and R_j we compute $C_{R_i,R_j}^{(2)}(\theta\theta'; \Delta t)$ for every possible pairs of degrees of freedom (dof), like $\varphi\varphi'$, $\varphi\psi'$, $\varphi\chi_1'$, $\psi\varphi'$, $\psi\psi'$, $\psi\chi_1'$, $\chi_1\varphi'$, $\chi_1\psi'$, $\chi_1\chi_1'$. The dof pairs for which $|F_{R_i,R_j}^{max}(\theta\theta')|$ is maximum is considered to determine the direction of correlation in time domain, namely, if any perturbation at R_i affects R_j at a later time or vice versa. We generate a 76×76 matrix by noting $|F_{R_i,R_j}^{max}(\theta\theta')|$ for all of the 76 residues of ubiquitin. By applying the condition of directionality we obtain the upper triangular matrix showing detailed TDCF map.

This map can be used to understand correlated path among the residues. Let us consider the terminal residue G76 which binds to AMP during Ub activation in ubiquitination. The dihedral φ of G76, φ_{G76} is correlated to R74 by φ_{R74} , ψ_{R74} and χ_{1R74} both in forward and reverse direction. However, among all these correlated dihedrals $F_{G76,R74}^{max}(\varphi\varphi)$ is the largest which we take as an indication that G76 is downstream correlated to R74 via dihedral φ of both the residues. Similarly G76 is downstream correlated to other set of residues, like L73, L67, Glutamine: Q62, Tyrosine: Y59, L56, R54, Aspartate: D52, K48, F45, L43, Q41, Q40, D39, Proline: P38, K33, I30, K27, V26, Glutamate: E24, I23, D21, T12, T7 and V5. Among all the downstream correlated residues to G76 we find that R74 is having the shortest $d_{\alpha-\alpha}$, which is the mean distance between C_α atoms over the entire trajectory. Similarly, the closest downstream correlated residue to R74 is R72. In this way we construct the path of downstream correlated dihedrals of different residues, $\varphi_{G76} - \varphi_{R74} - \varphi_{R72} - \psi_{V70} - \varphi_{L69} - \chi_{1_{H68}} - \chi_{1_{L67}} - \chi_{1_{V5}}$, as shown in a snapshot of ubiquitin obtained from simulation (Fig. 3(b)). Among these temporally correlated residues G76, R74 and R72 belong to the C terminal loop region. The residues V70, L69, H68 and L67 belong to $\beta 5$, while V5 belongs to $\beta 1$ in β strands of the crystal structure². Crystal structure of ubiquitin activation enzyme-E1 loaded with Ub molecules indicates^{4,5} that the hydrophobic surface patch of Ub including L8, I44, V70 and C terminal tail of Ub (G76, R74, R72) interact with the activation enzyme. The temporally correlated path with G76 contains many of the residues, like R74, R72 and V70. Besides, the slowest time scale in this path is that between G76 and R74, around 125 ns. This time scale is comparable to the rotational time scale of the enzyme which is about 90 ns obtained using the Stokes-Einstein²⁵ equation. Thus the path obtained using TDCF analysis is functionally relevant.

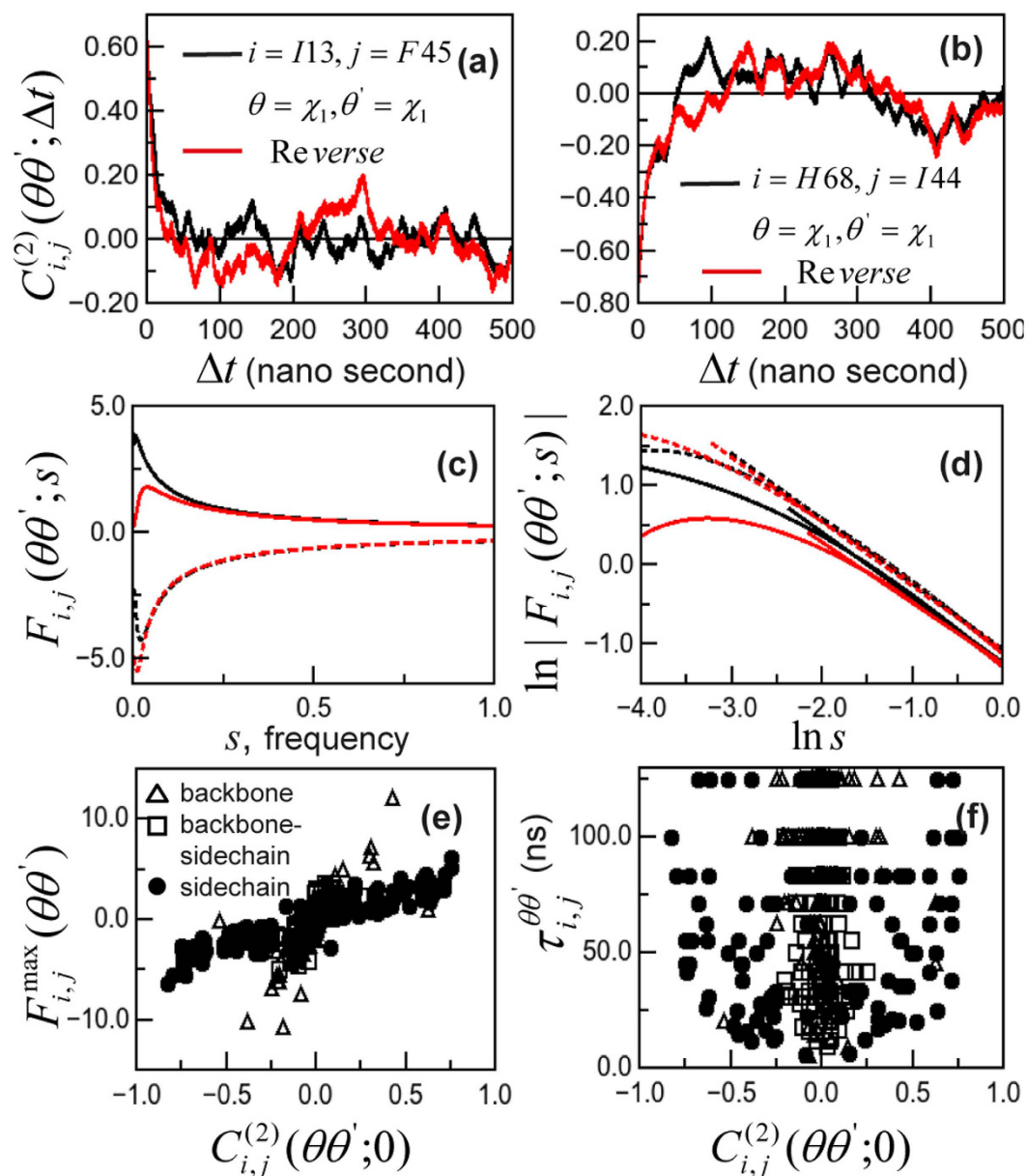


Figure 2. TDCFs between various dihedrals of ubiquitin, (Black: forward direction, red: reverse direction); (a) $C_{I13,F45}^{(2)}(\chi_1\chi_1; \Delta t)$. (b) $C_{H68,I44}^{(2)}(\chi_1\chi_1; \Delta t)$ as function of Δt . (c) Laplace Transform $F_{i,j}(\theta\theta'; s)$ of correlation function, $F_{I13,F45}(\chi_1\chi_1; s)$ (solid line) and $F_{H68,I44}(\chi_1\chi_1; s)$ (dashed line) versus s plots. (d) $\ln |F_{I13,F45}(\chi_1\chi_1; s)|$ versus $\ln s$ (solid line) and $\ln |F_{H68,I44}(\chi_1\chi_1; s)|$ versus $\ln s$ (dashed line) plots showing algebraic tails. (e) Correlations plot between functionally important residues of ubiquitin; $F_{i,j}(\theta\theta'; s)$ versus $C_{i,j}^{(2)}(\theta\theta'; 0)$. (f) $\tau_{i,j}^{\theta\theta'}$ versus $C_{i,j}^{(2)}(\theta\theta'; 0)$ for similar residues. The symbols have the same meaning in (e) and (f).

In order to get mechanistic view of long distance correlations, we calculate variance of the distances between residues belonging to the temporally correlated path. We compute $\text{var}(d_{\alpha-\alpha})$ which represents variance of $d_{\alpha-\alpha}$. Similarly for backbone-side chain distances, we calculate $\text{var}(d_{\alpha-\beta})$, where $d_{\alpha-\beta}$ denotes distance between C_α and C_β atoms of the residue pair. For side chain dihedrals, we compute $\text{var}(d_{\beta-\beta})$, the variance of distance between C_β atoms of the correlated pairs. We plot $|F_{i,j}^{max}(\theta\theta')|$ versus these variances in Fig. 3(c). We observe that the large correlation amplitudes are clustered near smaller values of the variance. This indicates that dihedral dynamical correlations are destroyed by large fluctuations.

We compute transfer entropy from mutual information²⁶ between maximally correlated pairs of fluctuating degrees of freedom which constitute the functionally relevant path. We assign the directionality of entropy transfer between two residues by the larger magnitude of the transfer entropy for the correlated pairs in forward and reverse directions. For instance, in case of $C_{G76,R74}^{(2)}(\theta\theta'; \Delta t)$, φ_{G76} and φ_{R74} are the maximally correlated dof both in forward and reverse directions (see Methods, Note 3). We find that mutual transfer entropy -0.17 for φ_{R74} to φ_{G76} and that for the reverse direction is 0.55 , indicating that the transfer of information takes place from φ_{G76} to

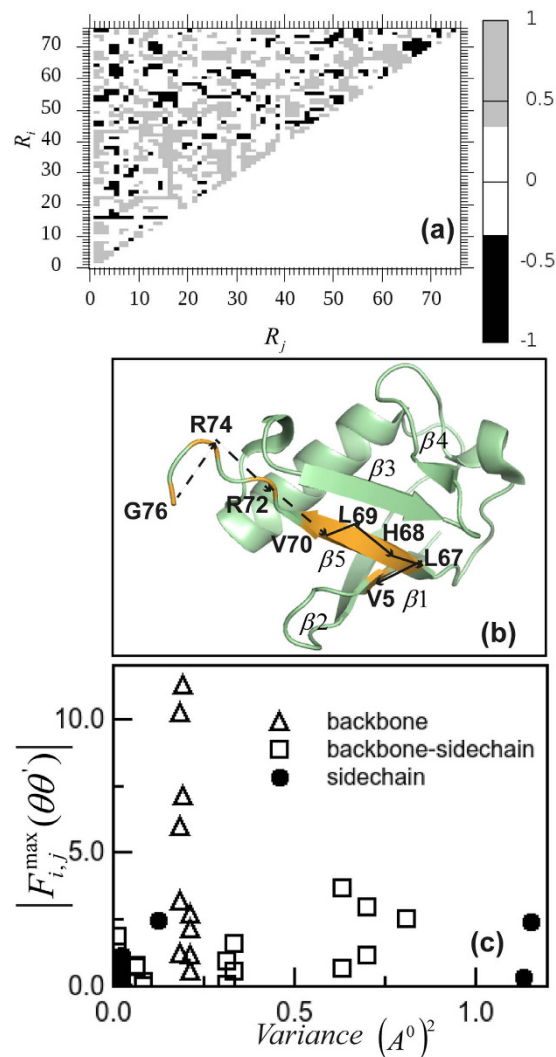


Figure 3. (a) TDCF map for any two residue pair in ubiquitin; Black represents downstream and Grey represents upstream TDCFs. (b) Residues belong to dynamically correlated path of ubiquitin. Solid line connects the residue pairs belong to β -sheets, dashed line connects the pairs belong to the loop region. (c) Correlation peak versus distance fluctuations of residue pairs belong to temporally correlated path.

φ_{R74} similar to experimental observations. We construct in similar way the direction of entropy transfer and time scales of the correlated dof over the path, as given in Supplementary Table S2. It is clear from the table that the directionality of path is not maintained between R74 and R72 where the entropy transfer takes place from R72 to R74. Moreover, the time scale of optimum mutual information is in sub-ns range, orders of magnitude shorter than biologically relevant time scales. Thus the TDCF describes the functionally relevant path in more reliably.

Mathematical Model

We model the qualitative behaviors of the TDCF in terms of equations of motion (see Methods, Eqs 3 and 4, Note 4 and for the details of calculations see Supplementary Note. S1) of two dihedrals $\theta_i(t)$ and $\theta_j(t)$ which are coupled to each other with strengths α' and β' respectively. Let the characteristic frequencies associated with them be ω_i and ω_j . They perform motions in a solvent experiencing drags proportional to $-\Gamma_i \dot{\theta}_i$ and $-\Gamma_j \dot{\theta}_j$ respectively¹⁸. We calculate the Laplace transformed correlation function from the equations of motion, $\langle (\theta_i(s)\theta_j(s)) \rangle$ in frequency, s by averaging over initial conditions on the variables. We find that in $s \rightarrow 0$ limit, $\langle (\theta_i(s)\theta_j(s)) \rangle \sim -C_{i,j}^{(2)}(\theta\theta'; 0)(\Gamma_i^2 \Gamma_j^2 / (\omega_i^2 \omega_j^2))^2 s^2$, $C_{i,j}^{(2)}(\theta\theta'; 0)$ being the statistical correlation coefficient. There is thus maximum in the low s limit if the TDCF shows statistical correlation, $C_{i,j}^{(2)}(\theta\theta'; 0) > 0$, while a minimum for statistically anti-correlated TDCF with $C_{i,j}^{(2)}(\theta\theta'; 0) < 0$. These are qualitatively similar to low s behaviour of the simulated TDCFs. We get an algebraic tail for large s , $\langle (\theta_i(s)\theta_j(s)) \rangle \sim s^{-2}$ where the exponent is universal and independent of the parameters in the model. This universality is revealed by the simulated TDCF, albeit with exponent 1.0. The difference in the exponent may be due to simplicity of the model equations of motion where all effects are neglected except solvent drag and mutual coupling. Moreover, we find that so far as $\alpha' \neq \beta'$, $\langle (\theta_i(s)\theta_j(s)) \rangle \neq \langle (\theta_j(s)\theta_i(s)) \rangle$ as seen in simulations.

Direct probe of dynamical correlation among the dihedrals is difficult due to limitation of probes. However, our analysis suggests an indirect way of probing the dynamical correlations. Our analysis shows that the residues, like R74, R72 and V70 lie in dynamically correlated path with G76 where ubiquitination initiates. We expect these residues to play an important role in the process which can be tested experimentally. R72 is experimentally known to give specificity to the ubiquitin activation enzyme-E1 binding^{4,5}. The role of the other residues needs to be looked into.

To conclude we show with long molecular simulations and mathematical modeling that TDCFs explain the causal connection between binding sites in Ub in the biologically relevant temporal regime. More importantly, our studies indicate that non-linearities are not the primary deciding factor for causal connections between functional sites. Although the simulations are illustrated for ubiquitin activation enzyme-E1 binding to Ub, the generality of our mathematical analysis shows that qualitative features of TDCF can be extended to any microscopic degrees of freedom. On a wider perspective, two point cross-correlation functions between relevant microscopic variables may provide a correct description of bio-molecular function and the related kinetics in terms of underlying microscopic dynamics without invoking the nonlinear effects.

Methods

Note 1: Details of MD Simulation. We perform MD simulation using NAMD at 310 K and 1 atm pressure, following standard protocols for NPT ensemble. We use TIP3P water model, periodic boundary condition and CHARMM27²⁷ force field with 1 femto-second time step. Electronutrality is maintained by adding proper number of mono-valent ions Na⁺ and Cl⁻. Long ranged electrostatic interaction is included by PME²⁸ method. Energy minimization was done for first 10,000 steps and simulation was performed for 1.05 μ s. Equilibration is ensured by RMSD plot over entire simulation time.

Note 2: Details of computation of two-point correlation function (TDCF). Dihedral angle is the intersecting angle between two adjacent planes for four consecutive atoms. In case of polypeptide backbone for $C_{i-1} - N_i - C_{\alpha,i} - C_{\beta,i}$ angle between $C_{i-1} - N_i - C_{\alpha,i}$ and $N_i - C_{\alpha,i} - C_{\beta,i}$ planes is known as dihedral φ . Similarly angle between $N_i - C_{\alpha,i} - C_{\beta,i}$ and $C_{\alpha,i} - C_{\beta,i} - N_{i+1}$ is dihedral ψ where i is residue index. Proteins also possess side-chain dihedrals among which χ_1 defines the angle between $N_i - C_{\alpha,i} - C_{\beta,i}$ and $C_{\alpha,i} - C_{\beta,i} - C_{\gamma,i}$. Thus the dihedral angles can be computed from the specified atomic positions obtained from the simulated MD trajectory.

The computation of TDCF from the MD trajectory has been done as follows: The series of the conformations is ordered in time with a given choice of initial condition. For any given time-difference $\Delta t = |t_2 - t_1|$, we compute the product between fluctuations of dihedral θ of residue i and θ' of residue j for the l -th observation for a given Δt ,

$$D_{i,j}^{(l)}(\theta\theta'; \Delta t) = (\sin[\theta_{i,l}(t_2)] - \langle \sin \theta_i \rangle)(\sin[\theta'_{j,l}(t_1)] - \langle \sin \theta'_j \rangle). \quad (1)$$

The angular bracket signifies ensemble average or mean of the respective quantity over simulation trajectory. The TDCF is given by $C_{i,j}^{(2)}(\theta\theta'; \Delta t) = \frac{1}{N_l} \sum_{l=1}^{N_l} D_{i,j}^{(l)}(\theta\theta'; \Delta t)$, where

$$C_{i,j}^{(2,l)}(\theta\theta'; \Delta t) = D_{i,j}^{(l)}(\theta\theta'; \Delta t) / \sqrt{(\text{var}(\sin \theta_i) \text{var}(\sin \theta'_j))}, \quad (2)$$

var denoting the variance of quantity within parenthesis. Here N_l is the number of observations corresponding to given Δt . For instance, for N number of observations, $N_0 = N$. Similarly, $N_1 = N - 1$ which correspond to the data set ($t_1 = 0, t_2 = 2$), ... ($t_1 = N - 1, t_2 = N$).

The computation is done for sufficiently large Δt until $C_{i,j}^{(2)}(\theta\theta'; \Delta t)$ approaches zero. We compute numerically the Laplace Transform $F_{i,j}(\theta\theta'; s) = \int_0^{\Delta t/2} e^{-st} C_{i,j}^{(2)}(\theta\theta'; \Delta t) dt$.

Note 3: Computation of transfer entropy. We use standard methodology for the computation of transfer entropy¹⁵. First we calculate the minimal embedding dimension (m)²⁶ for two time series by the false nearest neighbors method²⁹. Next we compute mutual information (MI) using the TISEAN package²⁶ for fluctuations of pairs of dihedral angles for different residues belonging to the path. We identify the embedding time intervals (τ) for which MI is optimum^{15,26}. We compute the transfer entropy¹⁵ between two degrees of freedom from the Shannon entropy and joint entropy using the MIToolbox³⁰ for fluctuations of two time series using m and τ .

Note 4: Over damped equations of motion in the long time limit. We have used equation of motion of over damped coupled classical harmonic oscillators.

$$-\Gamma_i \dot{\theta}_i(t) + \omega_i^2 \theta_i(t) + \alpha' \theta_j(t) = 0, \quad (3)$$

$$-\Gamma_j \dot{\theta}_j(t) + \omega_j^2 \theta_j(t) + \beta' \theta_i(t) = 0. \quad (4)$$

Using Laplace Transform of equations (3) and (4) we model behavior of TDCF in s domain. Details are in Supplementary Note S1.

References

- Vallee-Belisle, A. & Plaxco, K. W. Structure-switching biosensors: inspired by Nature. *Current opinion in structural biology* **20**, 518–526 (2010).
- Vijaykumar, S., Bugg, C. E. & Cook, W. J. Structure of Ubiquitin Refined at 1.8 Å Resolution. *J Mol Biol* **194**, 531–544 (1987).
- Schulman, B. A. & Harper, J. W. Ubiquitin-like protein activation by E1 enzymes: the apex for downstream signalling pathways. *Nat Rev Mol Cell Bio* **10**, 319–331 (2009).
- Schafer, A., Kuhn, M. & Schindelin, H. Structure of the ubiquitin-activating enzyme loaded with two ubiquitin molecules. *Acta Crystallogr D* **70**, 1311–1320 (2014).

5. Lee, I. & Schindelin, H. Structural insights into E1-catalyzed ubiquitin activation and transfer to conjugating enzymes. *Cell* **134**, 268–278 (2008).
6. Winget, J. M. & Mayor, T. The Diversity of Ubiquitin Recognition: Hot Spots and Varied Specificity. *Mol Cell* **38**, 627–635 (2010).
7. Wachsmuth, M. *et al.* High-throughput fluorescence correlation spectroscopy enables analysis of proteome dynamics in living cells. *Nat Biotechnol* **33**, 384–389 (2015).
8. Vogeli, B. Comprehensive description of NMR cross-correlated relaxation under anisotropic molecular tumbling and correlated local dynamics on all time scales. *J Chem Phys* **133**, 014501 (2010).
9. Fenwick, R. B. *et al.* Weak long-range correlated motions in a surface patch of ubiquitin involved in molecular recognition. *J Am Chem Soc* **133**, 10336–10339 (2011).
10. Long, D. & Bruschweiler, R. Structural and Entropic Allosteric Signal Transduction Strength via Correlated Motions. *J Phys Chem Lett* **3**, 1722–1726 (2012).
11. Fenwick, R. B., Orellana, L., Esteban-Martin, S., Orozco, M. & Salvatella, X. Correlated motions are a fundamental property of beta-sheets. *Nat Commun* **5**, 1–9 (2014).
12. Kasahara, K., Fukuda, I. & Nakamura, H. A Novel Approach of Dynamic Cross Correlation Analysis on Molecular Dynamics Simulations and Its Application to Ets1 Dimer-DNA Complex. *Plos One* **9**, e112419 (2014).
13. Li, D. W., Meng, D. & Bruschweiler, R. Short-Range Coherence of Internal Protein Dynamics Revealed by High-Precision in Silico Study. *J Am Chem Soc* **131**, 14610–14611 (2009).
14. Black Pyrkosz, A., Eargle, J., Sethi, A. & Luthey-Schulten, Z. Exit strategies for charged tRNA from GluRS. *J Mol Biol* **397**, 1350–1371 (2010).
15. Kamberaj, H. & van der Vaart, A. Extracting the Causality of Correlated Motions from Molecular Dynamics Simulations. *Biophys J* **97**, 1747–1755 (2009).
16. Ryabov, Y. Coupling between overall rotational diffusion and domain motions in proteins and its effect on dielectric spectra. *Proteins* **83**, 1571–1581 (2015).
17. Schreiber, T. Measuring information transfer. *Phys Rev Lett* **85**, 461–464 (2000).
18. Chaikin, P. M. & Lubensky, T. C. *Principles of condensed matter physics*. 361 (Cambridge University Press, 1995).
19. Hansen, J. P. & McDonald, I. R. *Theory of Simple Liquids* (Fourth Edition) (Elsevier, 2013).
20. Kampen, N. G. V. *Stochastic Processes in Physics and Chemistry* (Elsevier Science, 1992).
21. Sisan, D. R., Yarar, D., Waterman, C. M. & Urbach, J. S. Event ordering in live-cell imaging determined from temporal cross-correlation asymmetry. *Biophys J* **98**, 2432–2441 (2010).
22. Garnier, N., Genest, D. & Genest, M. Correlated motions and propagation of the effect of a local conformational change in the transmembrane helix of the c-erb B2 encoded protein and its V659E mutant, studied by molecular dynamics simulations. *Biophysical chemistry* **58**, 225–237 (1996).
23. Das, A., Ghosh, M. & Chakrabarti, J. Time dependent correlation between dihedral angles as probe for long range communication in proteins. *Chem Phys Lett* **645**, 200–204 (2016).
24. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *Journal of computational chemistry* **26**, 1781–1802 (2005).
25. Linegar, K. L., Adeniran, A. E., Kostko, A. F. & Anisimov, M. A. Hydrodynamic radius of polyethylene glycol in solution obtained by dynamic light scattering. *Colloid J+* **72**, 279–281 (2010).
26. Hegger, R., Kantz, H. & Schreiber, T. Practical implementation of nonlinear time series methods: The TISEAN package. *Chaos* **9**, 413–435 (1999).
27. Brooks, B. R. *et al.* CHARMM: the biomolecular simulation program. *Journal of computational chemistry* **30**, 1545–1614 (2009).
28. Essmann, U. *et al.* A Smooth Particle Mesh Ewald Method. *J Chem Phys* **103**, 8577–8593 (1995).
29. Kennel, M. B., Brown, R. & Abarbanel, H. D. Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Physical review. A, Atomic, molecular, and optical physics* **45**, 3403–3411 (1992).
30. Brown, G., Pocock, A., Zhao, M. J. & Lujan, M. Conditional Likelihood Maximisation: A Unifying Framework for Information Theoretic Feature Selection. *J Mach Learn Res* **13**, 27–66 (2012).

Acknowledgements

SD thanks DST for INSPIRE fellowship and Manas Mandal for his help in transfer entropy calculation. JC thanks Ananda Mohan Ghosh for helpful discussions. MG and JC thank DST for funding.

Author Contributions

J.C. designed the research project; S.D. carried out the all the computation and calculations; and S.D., J.C. and M.G. analyzed data and wrote the paper.

Additional Information

Supplementary information accompanies this paper at <http://www.nature.com/srep>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Dutta, S. *et al.* Spatio-temporal coordination among functional residues in a protein. *Sci. Rep.* **7**, 40439; doi: 10.1038/srep40439 (2017).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017