bioRxiv preprint doi: https://doi.org/10.1101/084756; this version posted November 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder.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.

# Entropy Transfer between Residue Pairs Shows that Allostery is an Intrinsic Property of Proteins: Quantifying Allosteric Communication in Ubiquitin

Aysima Hacisuleyman, Burak Erman

Chemical and Biological Engineering

Koc University, Istanbul, Turkey

Short title: Quantifying Allosteric Communication using Entropy Transfer

bioRxiv preprint doi: https://doi.org/10.1101/084756; this version posted November 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder.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

It has recently been proposed by Gunasakaran et al. that allostery may be an intrinsic property of all proteins. Here, we apply Schreiber's transfer entropy formulation to the non-allosteric protein Ubiquitin and show that there are indeed systematic pathways of entropy and information transfer between residues that correlate well with the activities of the protein. We use 600 nanosecond molecular dynamics trajectories for Ubiquitin and its complex with human polymerase iota and evaluate entropy transfer between all pairs of residues of Ubiquitin and quantify the binding susceptibility changes upon complex formation. Calculations show that specific residues act as entropy reservoirs in Ubiquitin and others as entropy sinks. Using the plausible conjecture that extracting entropy from a residue makes it more susceptible for interaction with a partner, we explain the ternary complex formation of Ubiquitin in terms of entropy transfer. Finally, we show that time delayed correlation of fluctuations of two interacting residues possesses an intrinsic causality that tells which residue controls the interaction and which one is controlled. Our work shows that time delayed correlations, entropy transfer and causality are the required new concepts for explaining allosteric communication in proteins.

bioRxiv preprint doi: https://doi.org/10.1101/084756; this version posted November 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder.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.

## Author Summary

Allosteric communication is essential for the function of proteins. Recent work shows that allostery results from dynamic processes in the protein associated with atomic fluctuations leading to entropic interactions that involve ensemble of pathways rather than discrete two state transitions. Based on this new picture of allostery, it was proposed that allostery may indeed be an intrinsic property of all proteins. In order to test this hypothesis, we derive the computational tools for quantifying allosteric communication, and explain allostery in terms of entropy transfer, a new concept based on information theory. We use long molecular dynamics simulations of proteins from which we calculate the transfer of entropy between pairs of residues. Results of simulations show that certain residues act as entropy sources while others as entropy sinks. Evaluation of time delayed correlations shows the presence of causality of interactions that allow us to differentiate between residues that are drivers in allosteric activity and those that are driven. Identification of driver-driven relations is important for drug design. Using the example of Ubiquitin, a protein that is not known to be allosteric, we identify paths of information transfer that control its binding to diverse partners in the Ubiquitin-Proteasome System. We conclude that allosteric communication resulting from entropy transfer between residues is an intrinsic property of all proteins.

bioRxiv preprint doi: https://doi.org/10.1101/084756; this version posted November 1, 2016. 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.

## Introduction

Allosteric communication describes the process in which action at one site of a protein is transmitted to another site at which the protein performs its activity. The importance of allostery in biological systems has generated significant experimental and computational research. The basic problem is to identify residues that participate in allosteric communication in the hope of controlling their behavior related to protein function. Allosteric communication first requires the identification of two sites, the effector site, i.e., the site that is acted upon, and the regulatory site where protein's activity is regulated. Although more than 1000 allosteric sites are known [1] many more need to be characterized. In fact several pairs of allosteric endpoints may exist in a protein [2] which increases the number of candidate pairs that communicate. This problem becomes even more important when one considers the fact that most known cancers result from disruption of allosteric communication as a result of single mutations [3, 4] and the number of proteins associated with this phenomenon is very large. Expressed in simple terms, the solution of the problem reduces to finding whether two given residues communicate with each other, and if so what the consequences of this communication are. Various approaches to solve the problem may be found in References [5-16]. The specific aim of the present paper is to develop a rapid computational technique that identifies interaction of residue pairs based on concepts of information transfer and entropy, to scan a given protein and identify pairs of sites that communicate and to determine whether these communicating pairs may be candidates for allosteric activity.

The present work departs from the approaches outlined in the preceding paragraph. We do not focus neither on single allosteric sites nor on allosteric paths. We consider the time trajectory of the fluctuations of two residues, which may be spatially distant, and search for information transfer from the trajectory of one residue to that of the other. The trajectories are obtained from long molecular dynamics (MD) equilibrium simulations that give the fluctuation of each atom at constant temperature. The first requirement for information to be transferred from an atom $i$ to another atom, $j$, is that their trajectories should be correlated. The second requirement is that this transfer should be asymmetric, i.e., information going from $i$ to $j$ should not be equal to information from $j$ to $i$. This requires the use of

bioRxiv preprint doi: https://doi.org/10.1101/084756; this version posted November 1, 2016. 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.

time delayed correlations of fluctuations which may be asymmetric in contrast to time independent correlations which are symmetric by definition and therefore lack information on directionality. If $C_{ij}(t,t+\tau)$ denotes the correlation of fluctuations of $i$ at time $t$ with those of $j$ at time $t+\tau$, then asymmetry requires that $C_{ij}(t,t+\tau) \neq C_{ji}(t,t+\tau)$. This introduces directionality and therefore causality into the problem. If time delayed correlations are asymmetric, then can we quantify the net information that is transferred? The answer is yes if we pose the problem in terms of entropy transfer.

Before going into the discussion of entropy, it is worth pointing out that information transfer is exclusively based on the changes in the amplitudes and frequencies of fluctuations in the system. This was first suggested and modelled by Cooper and Dryden (CD) [17] and reached larger dimensions by the work of Gunasekaran et. al. [18] which suggests that since allosteric communication is a result of correlated fluctuation then allostery should be an intrinsic dynamic property of all proteins. The dynamics aspect of proteins resides in the fluctuations of atoms which may be evaluated by experimentally measuring the B-factors of the atoms. The B-factor of the ith atom is related to its time independent autocorrelation of fluctuations, $\langle (\Delta R_i)^2 \rangle$, by $B_i = \frac{8\pi^2}{3}\langle (\Delta R_i)^2 \rangle$. However, knowledge of them is not sufficient for predicting allosteric communication and cross correlations $\langle \Delta R_i \Delta R_j \rangle$ between the fluctuations of different atoms are needed. Allosteric activity requires not only the modulation of the cross correlations in the system but also on time delayed cross correlations, $\langle \Delta R_i(0) \Delta R_j(t) \rangle$, as will be described below in detail. The CD model is referred to as 'Allostery without conformational change'. In this respect, it goes beyond the classical Monod-Wyman-Changeux (MWC) [19] model and its relative, the Koshland Nemethy Wyman (KNW) model [20] both of which relate allostery to discrete conformational changes at the regulatory site. Sending information by changing the amplitude and frequencies of fluctuations is entropic [21] and depends not only on the value of the entropy but also on the transfer of entropy from residue to residue during communication. Entropy as a source of information transfer is widely used in information theory [22] which is only very recently used for a protein-DNA complex by Kamberaj and van der Vaart [23]. Through analysis of entropy transfer, they determined residues that act as drivers of the fluctuations of other residues, thereby determining causality that is inherent in the correlations. Determining residues

bioRxiv preprint doi: https://doi.org/10.1101/084756; this version posted November 1, 2016. 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.

that act as drivers and those that are driven is important especially from the point of view of drug design. Entropy transfer and causality is a new paradigm for studying allosteric communication in proteins, which we elaborate in detail in the present paper. On a broader scale, our findings show that all proteins may indeed exhibit allosteric communication and therefore supports the hypothesis by Gunasakaran et. al., [18] which states that allostery is an intrinsic property of all dynamic proteins.

The quantitative measure of information flow between two correlated processes is introduced by Schreiber [22] in 2000. In the present work, the processes are generated in the form of trajectories of atomic coordinates using MD simulations from which probabilities of atomic coordinate fluctuations required for evaluating transfer entropy are calculated. We calculate the entropies based on atoms and identify the entropy of a residue with the entropy of its alpha carbon. Denoting the probability of fluctuation of atom $i$ by $p_i$, Callen showed [24] that the Shannon measure of disorder, $-k_B \sum_{i=1}^{N} p_i \ln p_i$ with $N$ and $k_B$ denoting the number of elements of the system and the Boltzmann constant, is the entropy of the system which is maximized at constant energy (See Callen [24], Chapter 17. Entropy and disorder: Generalized canonical formulations). At this point we give here a qualitative explanation of the relationship between information flow and a physical event such as fluctuations of atoms, and continue this discussion on a quantitative basis after we introduce the statistical mechanical basis of the model. Suppose we have two trajectories, one of atom $i$ and the other of atom $j$. If the fluctuations of $i$ and $j$ are independent of each other, than knowledge of the fluctuations of i will not give us information on the fluctuations of $j$ and the uncertainty associated with the two events will be a maximum. The total entropy of i and j will be the sum of the singlet entropies, $S_i + S_j$. If, on the other hand, i and j move in a correlated way, the fluctuations of $i$ controlling the fluctuations of $j$, then we will have more information on the fluctuations of $j$ than if they were uncorrelated. For example, if $i$ and $j$ were perfectly correlated, then we would know exactly what j will do if we know what i is doing. This extra information $I_{ij}$ that we gain because of the physical coupling of $i$ and $j$ is obtained by the Shannon equation and is termed as the mutual information and is always positive. The total entropy, $S_{ij}$ of $i$ and $j$ in this case is

bioRxiv preprint doi: https://doi.org/10.1101/084756; this version posted November 1, 2016. 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.

written as $S_{ij} = S_i + S_j - I_{ij}$ (see Eqs 12 and 13 and also Ref. [25]). Thus, correlation of fluctuations, irrespective of whether they are negative or positive, always decreases the sum of the singlet entropies of $i$ and $j$. These arguments and the Shannon equation have been used to obtain entropy changes in proteins at different levels of approximation [21, 26-38]. However, we need to go beyond the Shannon equation in order to quantify allosteric communication in proteins which requires, as shown by Schreiber in 2000 [22], the knowledge of time delayed conditional probabilities of two trajectories. In the interest of determining which residue drives the correlated motions and which residue responds, van der Vaart applied the Schreiber equation to determine information flow between Ets-1 transcription factor and its binding partner DNA [23] (Also see references [39] and [40] in similar context). Since this first work on entropy transfer in proteins there has been a limited number of studies on information transfer in proteins. Barr et al. [41] quantified entropy transfer among several residues in a molecular dynamics analysis of mutation effects on autophosphorylation of ERK2. Corrada et al. [42] analyzed entropy transfer in antibody antigen interactions. Perilla et al. [43] used the transfer entropy method to analyze barrier crossing transitions in epidermal growth factor receptors. Qi and Im [44] quantified drive-response relations between residues during folding. Jo et al. [45] obtained a causality relationship between intramolecular hydrogen bonds and the conformational states of *N*-glycan core in glycoproteins. Zhang et al. [46] applied the method to understand changes in the correlated motions in the Rho GTPase binding domain during dimerization. An extensive overview of similar techniques is given in reference [47].

In the following section, we define the model on which we build the information theoretical basis of entropy. We then study the problem of time delayed correlation of fluctuations in proteins. Despite its importance in pointing to directionality of events in proteins, as has been shown recently for the allosteric activity od K-Ras [48], time delayed functions have not been studied in detail in the past. We then present a fast and accurate method of calculating entropy changes in proteins subject to pairwise interactions. Calculation of entropy of proteins is not new and has already been investigated by several authors [26-28, 49, 50] at different levels of approximation. Our method of entropy calculation is motivated by the recent finding that the distribution functions for the magnitude of fluctuations of residues in globular proteins can be derived from a universal function [51]. The method that we use for calculating the entropy is fast

bioRxiv preprint doi: https://doi.org/10.1101/084756; this version posted November 1, 2016. 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 accurate, based on histogramming the magnitude of fluctuations of each atom in a protein where the bin number is chosen according to the Sturges' rule of determining the widths of class intervals [52]. We show that the use of Sturges' rule in our computational method leads to results that are in agreement with earlier entropy calculations. We benchmark our method with calculations of Ubiquitin by Fleck et al [38]. The entropy change of Ubiquitin upon binding to human polymerase iota UBM2 that we calculate gives the same value obtained in reference [38] using a different method of entropy estimation. The computational method that we adopt is efficient and plausible.

The association of Shannon equation with statistical mechanical definition of entropy and quantifying transfer entropy by the use of the Schreiber equation allows us to interpret a wide range of events in proteins. If entropy transfer is considered in terms of changes in mobility, then transfer of entropy from $i$ to $j$ implies decrease in the mobility of $i$ due to its correlation with $j$. Stated in another way, residue $j$ extracts entropy from $i$. If binding is considered, one could then say that transfer of entropy from $i$ to $j$ would facilitate binding at $i$ due to lowered mobility of $i$, although this may not be a general trend and may depend on several other factors. We use the model to study the directionality of information flow and entropy transfer in the 76 amino acid protein Ubiquitin which is known to propagate signals allosterically in the cell by binding to a vast number of substrates [53]. Ubiquitin itself is not known to be allosteric, however, its coordinated motions upon binding with a protein modifies the fluctuation patterns on another site that affects the binding of a third protein as in the case of its ternary complex [54]. In order to identify communication patterns leading to such effects, we scan the full Ubiquitin and identify the pairs of residues whose time delayed correlation functions are asymmetric and quantify the amount of entropy transferred between these residues. We then analyze the behavior of Ubiquitin when complexed with the binding partner human polymerase iota UBM2, 2KTF.pdb.

bioRxiv preprint doi: https://doi.org/10.1101/084756; this version posted November 1, 2016. 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.

# Results

## Structure of Ubiquitin

Ubiquitin is a 76 amino acid protein as shown in Fig 1. It consists of 8 distinct secondary structures that actively take part in its interactions with a large number of proteins.

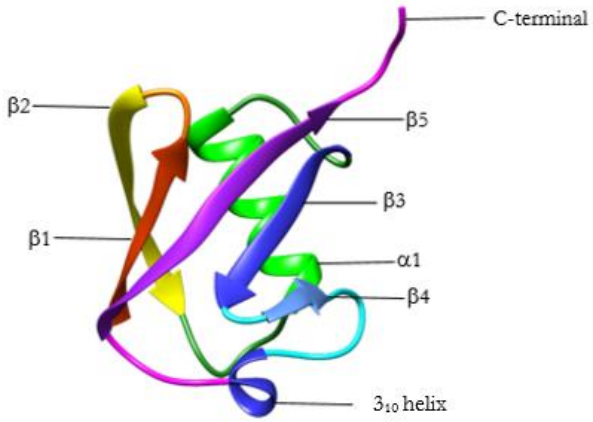


*Fig 1. Structure of Ubiquitin. Ribbon diagram was generated by Chimera.*

The interactions of the secondary structures are strictly coordinated by the correlations in the protein. In Fig 2 we present the results of Pearson correlations of fluctuations, where the negative and positive correlations are shown in the left and right panels, respectively. The correlations with amplitudes (-1,-0.25) and (0.25, 1.0) are shown in the figure. The negative correlation map shows that β1 is correlated with α1, β3 and β4. It is also correlated with the loops between β1-β2, α1-β3 and β3-β4. β2 is correlated with β4 and C-terminal. It is also correlated with the loop between β2-α1 and β4-β5. α1 is correlated with β1, β2, the loop between β2 and α1 and the C-terminal. β3 is correlated with β1, β2, the loop between β4 and β5 and the C-terminal. β4 is correlated with β1, β2, C-terminal and the loop between β4-β5. β5 is correlated with β2, the loop between β4-β5 and the C-terminal. The positive correlation map shows that β1 is positively correlated with β2, β5 and C-terminal. It is also correlated with the loop between β4 and β5. β2 is positively correlated with β1. α1 is positively correlated with β3, β4, β5 and the loop between α1- β3, β3-β4 and β4- β5.β3 is positively correlated with α1, β4 and β5. α1 is also positively correlated with the loop

bioRxiv preprint doi: https://doi.org/10.1101/084756; this version posted November 1, 2016. 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.

between β1-β2, α1-β3 and β3- β4. β4 is positively correlated with α1 and β5. β5 is positively correlated with β1, α1, β3, β4 and the loop between β1-β2.

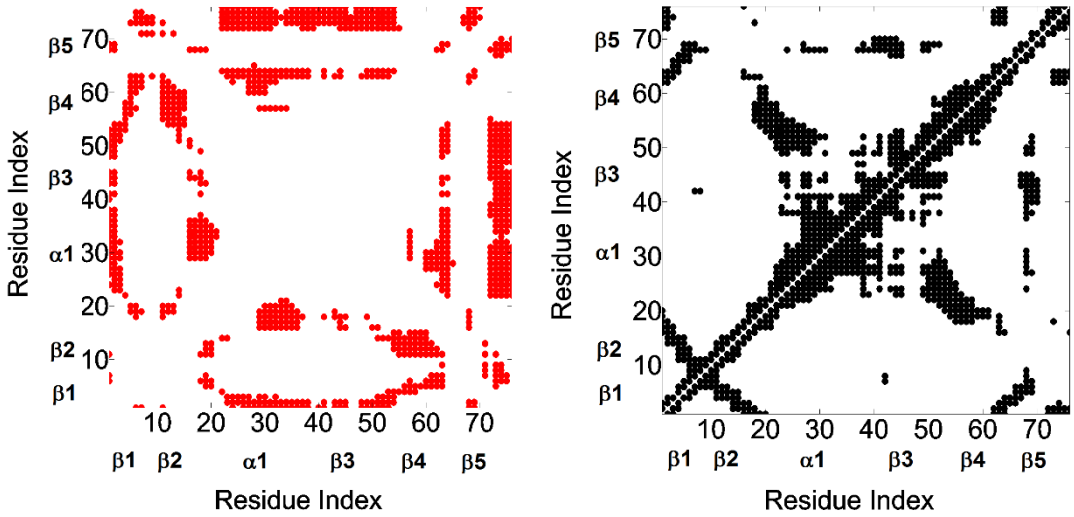


*Fig. 2.* The left panel shows the negative Pearson correlations in the range (-1, -0.25), the right panel shows positive Pearson correlations in the range (0.25, 1.0). Pearson correlations are calculated from Eq. 9.

**Transfer entropy in Ubiquitin**

We present the results of entropy transfer between all residue pairs of Ubiquitin. We consider only the alpha carbons, and the values given are divided by the Boltzmann constant. Using Equation 16 we evaluated the values of entropy transfer from alpha carbon i to j, $T_{i\to j}(\tau)$, for all pairs of i and j for τ = 5 ns. Calculations averaged over several time stations between 0 and 5 ns gave approximately the same values for entropy transfer. In the remaining parts we present τ = 5 ns results only. The characteristic decay time of correlations of fluctuations of alpha carbons, which will be discussed in the following section, is on the average between 5 to 10 ns. The entropy transfer function $T_{i\to j}(\tau)$ that we obtain from fluctuation trajectories of alpha carbons depends on the correlation of two events that are $\tau$ ns apart in time. If $\tau$ is taken very small, i. e., around zero, then the difference between $T_{i\to j}(\tau)$ and $T_{j\to i}(\tau)$ will be very small because $T_{i\to j}(0)=T_{j\to i}(0)$. If $\tau$ is taken much larger than the characteristic decay time,

bioRxiv preprint doi: https://doi.org/10.1101/084756; this version posted November 1, 2016. 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.

then the correlations will have decayed to small values and the differences will be vanishingly small. In agreement with this reasoning, we took $\tau$ =5 ns and calculated entropy transfer at this time. The results are shown in Fig 3. The abscissa, named as entropy donor, denotes the indices of residues that act like entropy reservoirs to other residues. The ordinate, named as entropy acceptor, denotes the indices of residues that act like entropy sinks that absorb entropy from the system.

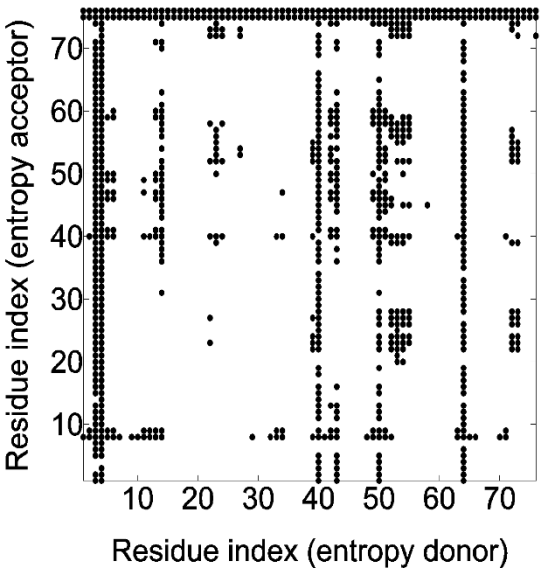


*Fig 3. Entropy transfer from residue i to residue j. Abscissa represents residues which provide entropy to residues shown along the ordinate. Entropy transferred from residue i to residue j is obtained from Eq. 16. Values between 0.0035-0.35 are recorded. Values below 0.0035 are not shown in order not to crowd the figure.* $T_{i\to j}(\tau)$ values are calculated from Eq. 16 with $\tau$=0.5 ns.

The columns of black points in the figure show that specific residues, such as ILE3 and PHE4, ILE13, ILE23, LYS27, GLY53, GLU64, ARG72 provide entropy to several residues of the protein. The rows of black circles indicate residues such as LEU8, THR9, GLY75 and 76, that absorb entropy from several residues of the protein. Residues ILE3 and PHE4, ILE13 and GLU64 form a spatial cluster. Also, the residues ILE23, LYS27 and GLY53 form a spatial cluster. If the allosteric path description is adopted, then we can say that these two spatial clusters lie on the allosteric path.

The net transfer of entropy from residue i, defined by Eq. 17 is presented in Fig 4. Positive values denote net entropy transfer out from a residue, and negative values denote net entropy into a residue. Similar to

bioRxiv preprint doi: https://doi.org/10.1101/084756; this version posted November 1, 2016. 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 pattern observed in Fig 3, we see that certain residues behave as entropy sources for the rest of the protein and some behave as entropy sinks.

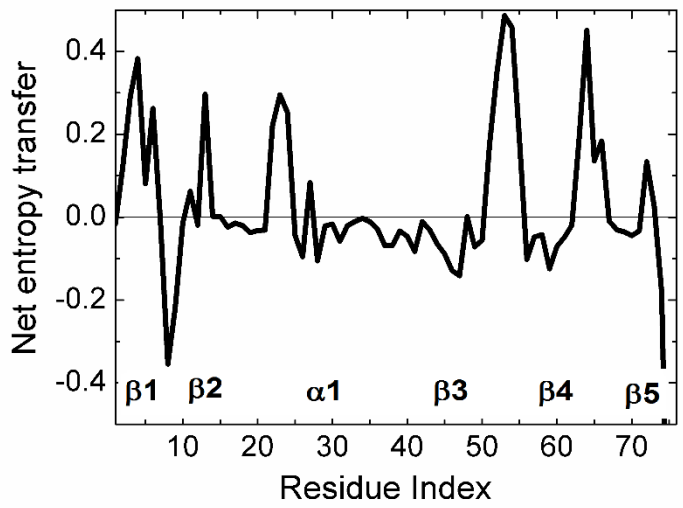


*Fig 4. Net entropy transfer from one residue to the rest of the protein, calculated by Eq. 17.*

We see that $\beta_1$ and $\beta_2$ act as an entropy sources as well as part of the helix $\alpha_1$. The largest amount of entropy is provided by the loops between $\beta_3\beta_4$ and $\beta_4\beta_5$. The two major entropy sinks are the loop between $\beta_1$ and $\beta_2$ and the last two residues of the C-terminal. Entropy sources are located mostly at secondary structures or at their extremities. The three residues PHE4, THR14, GLU64 are spatial neighbors. Similarly, LEU43, LEU50, are spatial neighbors.

In order to have an idea on the mechanism of communication in the system, one needs to know the transfer of entropy among specific pairs of residues. From the data of Fig 4, we can find with which residues a given amino acid interacts entropywise. Fig 5 and 6 summarizes the net entropy exchange, $T_{i\to j}(5)-T_{j\to i}(5)$, between the labeled residue in each panel and the j'th residue of the protein. Fig 5 shows some examples with mostly positive entropy transfer from the labeled residue to others. The top left panel in Fig 5 shows entropy transfer from ILE3 to other residues. Specifically it transfers the largest entropy to LEU8 and GLY75 and GLY76. From the dynamics point of view, we can say that ILE3 contributes to the mobility of these three residues. Both ILE3 and LEU8 are at the opposite extremities of

bioRxiv preprint doi: https://doi.org/10.1101/084756; this version posted November 1, 2016. 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.

$\beta_1$. ILE3 is a spatial neighbor of GLU64. GLU64 is hydrogen bonded to GLN2 which is on a loop, and the mobility of the loop is transferred to ILE3 via the stated hydrogen bond. ILE3 also contributes entropy to several other residues of the protein as may be seen from the figure. Entropic interactions of residues PHE4, ILE13, ILE23 and LYS27 are very similar to those of ILE3 and are not shown in Fig 5. The top right panel in Fig 5 shows the interactions of GLY53 with the rest of the protein. GLY53 is situated on the long loop between $\beta_3$ and $\beta_4$, and is hydrogen bonded to the main alpha carbon of GLU24 which is at the end of $\alpha_1$. Fig 5 shows that GLY53 contributes to the mobility of the segment between VAL17 and LYS29. It also transfers entropy along the chain to LEU56. GLU64 contributes entropy to several residues, in a way similar to that of ILE3. ARG72 has a unique pattern of contribution, specifically to ASP39 which is its spatial neighbor, to the loop between $\alpha_1$ and $\beta_2$, to PHE45 and LEU56, both of which are spatially distant from ARG72. It also contributes to the mobility of the C-terminal. Fig 6 gives two examples for mostly negative values of $T_{i\to j}(5)-T_{j\to i}(5)$. The left panel in Fig 6 shows that LEU8 and GLY76 absorb entropy from most of the residues of the protein.

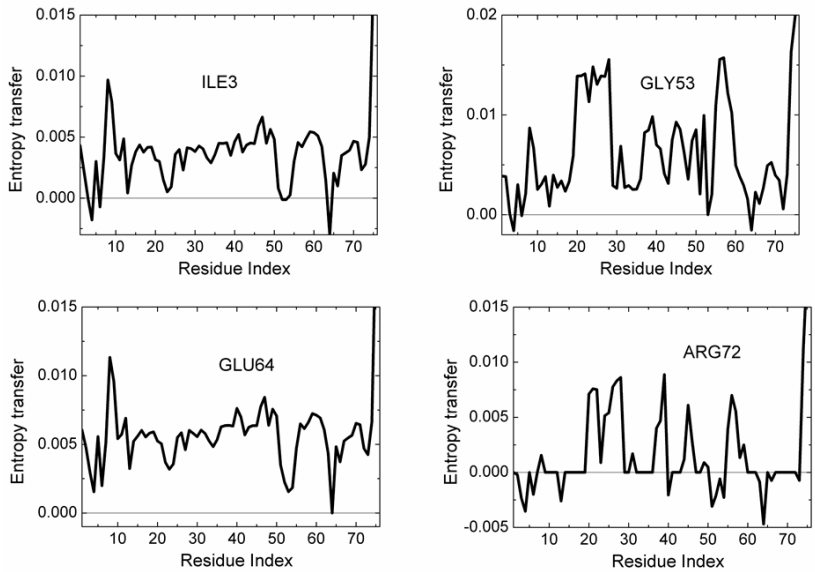


*Fig 5. Entropy transfer from a given residue to other residues of the protein. The residue from which entropy is transferred is marked in each panel. Calculations are based on the relation* $T_{i\to j}(5)-T_{j\to i}(5)$.

bioRxiv preprint doi: https://doi.org/10.1101/084756; this version posted November 1, 2016. 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.

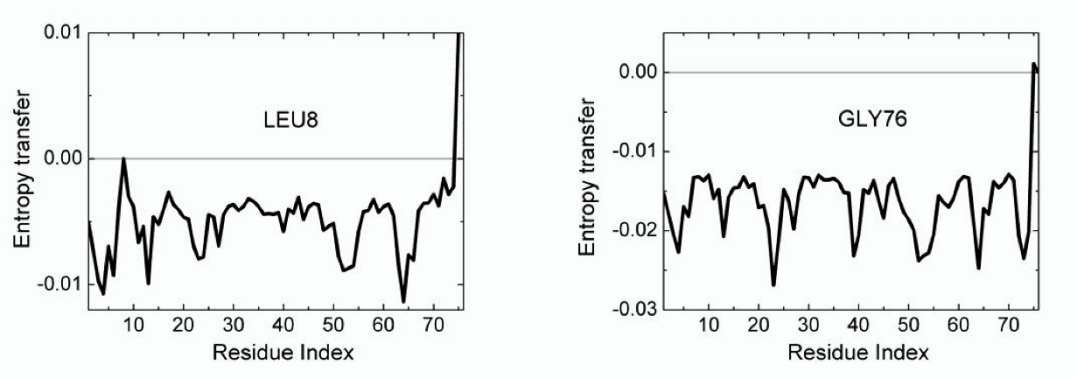


*Fig 6. Entropy transfer from residues of the protein into LEU8 and GLY76. Calculations are based on the relation* $T_{i\to j}(5)-T_{j\to i}(5)$.

## Time delayed correlations of Ubiquitin

Fluctuations of amino acids in Ubiquitin display characteristic decay times that are in the order of 1 to 5 ns. Differences arise from the unique conformational features of the amino acid and its environment. In Fig 7, we show the autocorrelations of LEU7 and THR71.

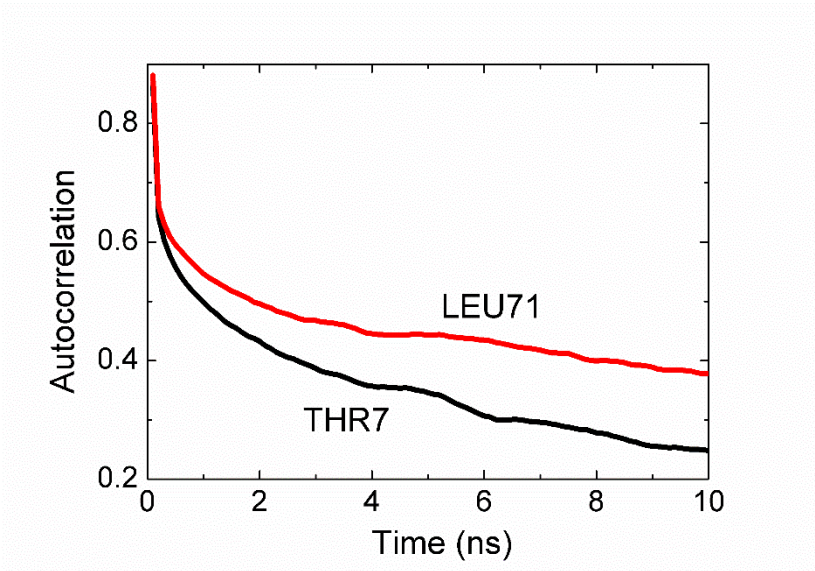


*Fig 7. THR7-LEU71 Autocorrelations calculated from Eq. 7 for i=j.*

bioRxiv preprint doi: https://doi.org/10.1101/084756; this version posted November 1, 2016. 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 autocorrelation function for THR7, i.e., the time required to decay to 1/e of the original value is 5 ns. LEU71 decays slightly slower with a decay time of 10 ns..

The time delayed cross correlations of the fluctuations of two amino acids are of interest because they yield information on the causality of events. The static correlations presented in Fig 2 are symmetric, i.e., $C_{ij}(\tau) = C_{ji}(\tau)$. However, time delayed cross correlations of fluctuations of two amino acids show asymmetries which we discuss in this section.

The strongest asymmetry is between LEU7 and THR71, shown in Fig 8.

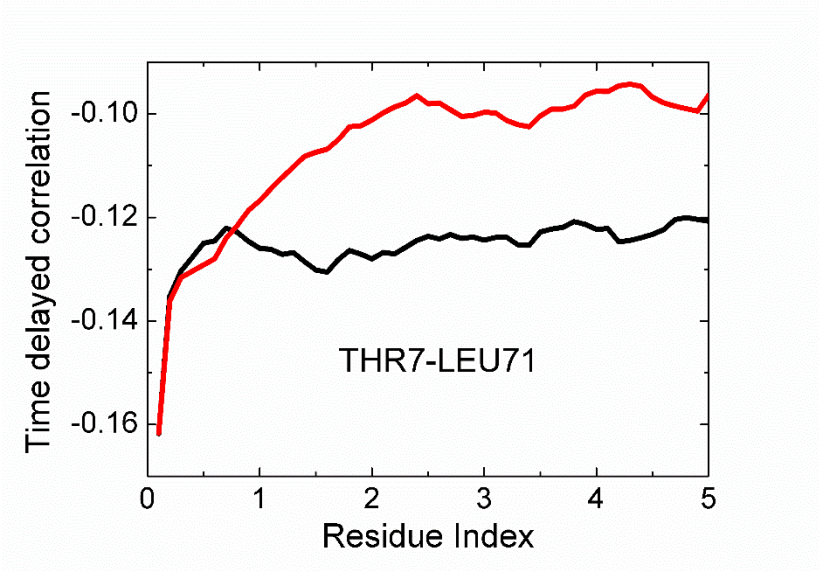


*Fig 8. Cross correlation of fluctuations of THR7 and LEU71. Black line is for correlations where THR7 precedes LEU71. The red line is for correlations where LEU71 precedes THR7. The curves are calculated from Eq. 7.*

In this figure, the black curve is for correlation of THR7 at time t and LEU71 at t+τ. The red curve is for LEU71 at t and THR7 at t+τ. The black curve decays significantly slower than the red curve, indicating that the effect of the fluctuations of THR7 on later fluctuations of LEU71 persists for longer times whereas the converse is not true. We therefore say that the motions of THR7 drive the motions of, LEU71 i.e., THR7 is the driver and is LEU71 driven. Since LEU71 is located on the C-terminal segment, and THR7 is at the end of $\beta_1$, we can say that the $\beta_1$ strand controls the fluctuations of the C-terminal. We

bioRxiv preprint doi: https://doi.org/10.1101/084756; this version posted November 1, 2016. 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.

see that the black curve remains approximately constant after a rapid initial decay. This shows that the driver action of THR7 on LEU71 persists for longer times.

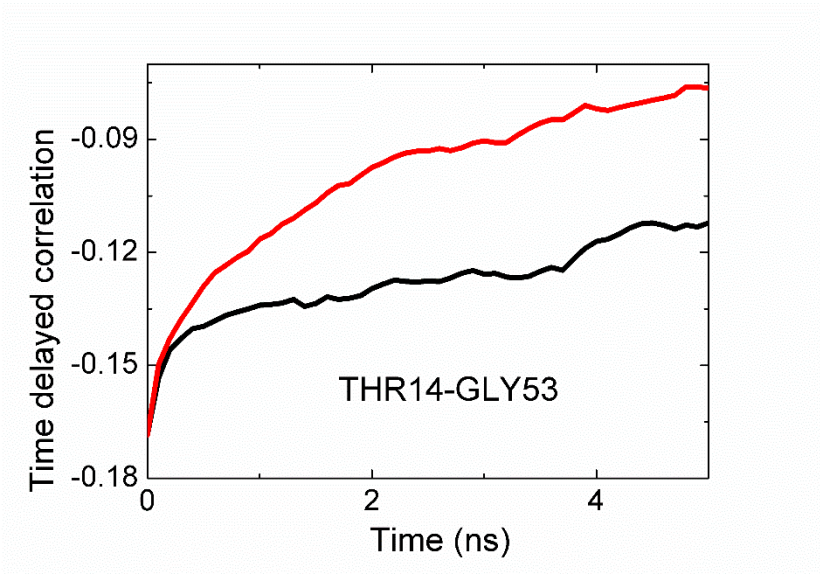


*Fig 9. Cross correlation of fluctuations of THR14 and GLY53. See legend for Figure 8.*

In Fig 9, the black curve is for the correlation of the fluctuations of THR14 with later fluctuations of GLY53. The red curve is for the reverse case, fluctuations of GLY53 affecting later fluctuations of THR14. This figure shows that THR14 is the driver and GLY53 is driven. THR14 is on the $\beta_2$ strand and GLY53 is on the long loop connecting the $\alpha 3_{10}$ helix to $\beta_4$.

## Changes upon complex formation of Ubiquitin

Ubiquitin forms complexes with a multitude of proteins. Here we studied its complex with Human Polymerase Iota (UBM) which is a small protein of 28 amino acids, 2L0G.pdb. UBM binds to Ubiquitin at β1, the loop between β1 and β2, and at β3 and β4. The binding affects the fluctuations of these regions. We compare the B-factors for Ubiquitin in the bound and unbound states in Fig 10. The solid and the dashed curves for Ubiquitin in the bound and free states, respectively. Binding decreases the B-factors of the contact substructures as may be seen from the figure. Additionally, the fluctuations of certain residues

bioRxiv preprint doi: https://doi.org/10.1101/084756; this version posted November 1, 2016. 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.

between 23 and 37 are decreased although UBM binds to a region far from these residues. Interestingly, the amplitude of fluctuations of the C-terminal residues are not affected by the binding.

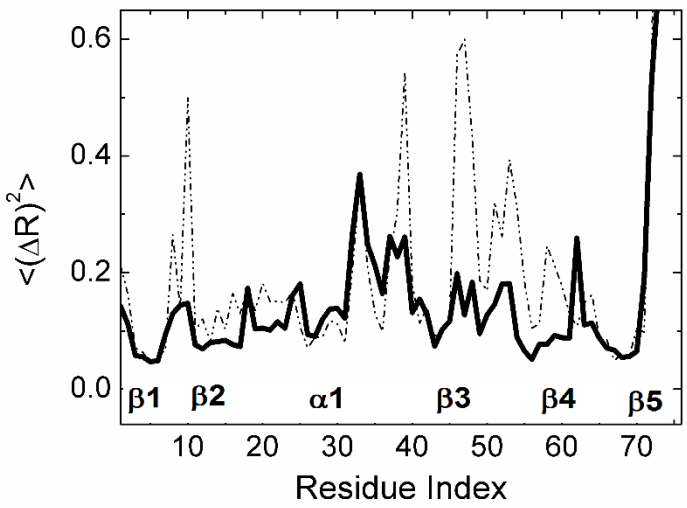


*Fig 10. Mean-square fluctuations of residues in unbound Ubiquitin (dashed curve) and in the complex (solid curve)*

Net entropy transfer in Ubiquitin in the bound and free states is compared in Fig 11. The solid and dashed curves are for the bound and free states, respectively. We see from the two curves that Ubiquitin upon complex formation becomes a totally different structure from entropy point of view. Entropy transfer associated with residues ILE3, PHE4, ILE23, GLY53 and GLU64 is diminished and entropy transfer from LEU8 is markedly increased. ILE13 is not affected. Entropy transfer associated with the last two residues of the C-terminal is strongly decreased although their B-factors are not affected by binding.

bioRxiv preprint doi: https://doi.org/10.1101/084756; this version posted November 1, 2016. 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.

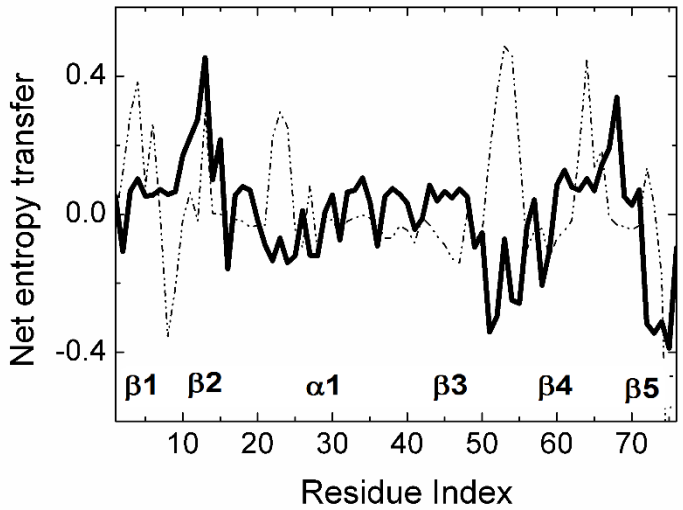


*Fig 11. Net entropy transfer of residues in unbound Ubiquitin (dashed curve) and in the complex (solid curve) Curves are obtained by using Eq. 17.*

Residues on the two loops between $\beta_2$ and $\alpha_1$ (from SER20 to ASN25) and between $\beta_3$ and $\beta_4$, (from LEU50 to ASN60) are not involved in binding with UBM. These two regions, which were entropy sources in the unbound state now become entropy sinks, as may be seen from Fig 11 where the dashed curves have positive values and the solid curves have negative values in these two regions. Based on the arguments presented in this paper, we can say that these regions with entropy sink features will facilitate binding.

## Methods

### Molecular dynamics trajectories

We perform molecular dynamics simulations for a protein in equilibrium and extract stationary trajectories for each atom. The trajectories for the atoms are expressed as

$$R(t_k) = R\left(R_1(t_k), R_2(t_k), R_3(t_k), .., R_N(t_k)\right) \qquad k = 1, 2, .., n_T \qquad (1)$$

Here, $R_i(t_k)$ is the position vector of the ith atom at the kth time $t_k$, expressed in terms of its Cartesian coordinates, $X_i(t_k)$, $Y_i(t_k)$ and $Z_i(t_k)$, $N$ is the total number of atoms and $t_k$ is the time in the kth

bioRxiv preprint doi: https://doi.org/10.1101/084756; this version posted November 1, 2016. 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.

step. $k$ ranges from 1 to $n_T$ , the total number of steps in the simulation. If the total time is $T$ , then the length $\xi$ of each time step is $\xi = T / n_T$. Each atom has a unique equilibrium mean position defined by

$$\bar{R} = R\left(\bar{R}_1, \bar{R}_2, \bar{R}_3, .., \bar{R}_N\right) \tag{2}$$

We denote the instantaneous state of fluctuation of a protein at time $t_k$ by the vector

$$\Delta R\left(t_k\right) = R\left(t_k\right) - \bar{R} \tag{3}$$

which reads in vector form as

$$\Delta R\left(t_k\right) = \Delta R\left(\Delta R_1\left(t_k\right), \Delta R_2\left(t_k\right), \Delta R_3\left(t_k\right), \ldots, \Delta R_N\left(t_k\right)\right) \tag{4}$$

For each $t_k$ , Eq.4 has $N$ entries. For the purposes of the present study, we need the magnitude of the fluctuations only. In the following, we will let $\Delta R_i\left(t_k\right)$ represent the magnitude of the fluctuation at time $t_k$ .

**Evaluation of probabilities**

The most general expression for the probability of fluctuation $\Delta R$ is the joint probability $p\left(\Delta R\right)$ $p\left(\Delta R_1, \Delta R_2, \Delta R_3, .., \Delta R_N\right)$. This expression contains information on all orders of dependence between atoms and is too general for use. In the other extreme, the simplest expression is the singlet probability function $p_i\left(\Delta R_i\right)$ which is obtained from the most general expression by

$$p\left(\Delta R_i\right) = \int_0^\infty \cdots \int_0^\infty \cdots \int_0^\infty \cdots \int_0^\infty p\left(\Delta R_1, \Delta R_2, \Delta R_3, .., \Delta R_N\right) d\Delta R_1, \ldots.., \Delta R_{i-1}, .., \Delta R_{i+1}, .., \Delta R_N \tag{5}$$

$N$ such functions define the probability of fluctuations of the $N$ residues within the singlet approximation.

bioRxiv preprint doi: https://doi.org/10.1101/084756; this version posted November 1, 2016. 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 next simplest probability is the pair probability $p\left(\Delta R_i, \Delta R_j\right)$ obtained from the most general expression by

$$p\left(\Delta R_i, \Delta R_j\right) = \int_0^\infty \cdots \int_0^\infty p\left(\Delta R_1, \Delta R_2, \Delta R_3, .., \Delta R_N\right) d\Delta R_1, .., \Delta R_{i-1}, .., \Delta R_{i+1}, .., \Delta R_{j-1}, .., \Delta R_{j+1}, .., \Delta R_N \tag{6}$$

For $N$ atoms, there are $\frac{N\left(N-1\right)}{2}$ equations for pair probabilities.

In Eqs. 5 and 6, $\Delta R's$ are treated as continuous. In the remaining of the paper, we will adopt a discrete representation for them in terms of histograms. The histograms will be expressed in terms of $n$ bins. We refer to each bin as a state. The variables in the probabilities will then be functions of state variables. Thus we write $p\left(\Delta R_i\left(k\right)\right)$ where k goes from 1 to $n$ where $n$ is the number of states that define $\Delta R_i$. Similarly, $p\left(\Delta R_i\left(k\right), \Delta R_j\left(l\right)\right)$. In order to simplify the notation, we will suppress the state index, and write $p\left(\Delta R_i\left(k\right)\right)$ as $p\left(\Delta R_i\right)$.Similarly, $p\left(\Delta R_i\left(k\right), \Delta R_j\left(l\right)\right) \equiv p\left(\Delta R_i, \Delta R_j\right)$.

## Time delayed correlation functions

We let $p\left(\Delta R_i\left(t\right), \Delta R_j\left(t+\tau\right)\right)$ denote the joint probability of observing the fluctuation $\Delta R_i$ at time t and $\Delta R_j$ at time $t+\tau$. In this simplified notation, $\Delta R_i\left(t\right)$ represents the value of $\Delta R_i$ in state k at time t, which is identical to $\Delta R_i\left(k,t\right)$. $\Delta R_j\left(t+\tau\right)$ may be affected by the earlier fluctuations of $\Delta R_i\left(t\right)$. The extent of this effect may be quantified by the time delayed correlation function

$$C_{ij}\left(\tau\right) = \frac{\left[\sum_{k=1}^{n_T-\tau/\xi} \Delta R_i\left(t_k\right)\Delta R_j\left(t_k+\tau\right)/\left(n_T-\tau/\xi\right)\right]}{\left\langle\left(\Delta R_i\right)^2\right\rangle^{1/2}\left\langle\left(\Delta R_j\right)^2\right\rangle^{1/2}} \tag{7}$$

bioRxiv preprint doi: https://doi.org/10.1101/084756; this version posted November 1, 2016. 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.

This is a conditional correlation where $\Delta R_i$ comes before $\Delta R_j$. In general, $p\left(\Delta R_i(t)\Delta R_j(t+\tau)\right) \neq p\left(\Delta R_j(t)\Delta R_i(t+\tau)\right)$. This leads to directionality in the structure, known as causality, and consequently,

$$C_{ij}(\tau) \neq C_{ji}(\tau) \tag{8}$$

If the fluctuations of residue $i$ control the fluctuations of residue $j$, i.e., if residue $j$ is driven by $i$, then the decay time for $C_{ij}(\tau)$ will be larger than that of $C_{ji}(\tau)$.

When $\tau = 0$ time independent Pearson correlation function is obtained as

$$C_{ij}(0) = \frac{\left\langle \Delta R_i \Delta R_j \right\rangle}{\left\langle (\Delta R_i)^2 \right\rangle^{1/2} \left\langle (\Delta R_j)^2 \right\rangle^{1/2}} \tag{9}$$

**Entropy**

The entropy for a system with pair probabilities is given by

$$S_2 = -k_B \sum_{i>j} \left[ \sum_k \sum_l p_{ij}(k,l) \ln p_{ij}(k,l) \right] = -k_B \sum_{i>j} \left\langle \ln p_{ij} \right\rangle \tag{10}$$

For brevity of presentation, we used the notation $p_{ij}(k,l) = p\left(\Delta R_i(k), \Delta R_j(l)\right)$

In Equation 10, $S_2$ signifies the joint entropy for a system with pair probabilities.

We now divide and multiply the entropy expression by the singlet probabilities:

$$S_2 = -k_B \sum_{i>j} \left\langle \ln \left( \frac{p_{ij}}{p_i p_j} p_i p_j \right) \right\rangle \tag{11}$$

and expand Eq. 11 as

$$S_2 = -k_B \sum_i \left\langle \ln p_i \right\rangle - k_B \sum_{i>j} \left\langle \ln \left( \frac{p_{ij}}{p_i p_j} \right) \right\rangle \tag{12}$$

bioRxiv preprint doi: https://doi.org/10.1101/084756; this version posted November 1, 2016. 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.

$$S_2 = S_1 - I_2 \tag{13}$$

where, $S_1 = -k_B \sum_i \langle \ln p_i \rangle$ is the singlet entropy and $I_2 = k_B \sum_{i>j} \left\langle \ln\left( \frac{p_{ij}}{p_i p_j} \right) \right\rangle$ is the mutual information of the system.

Equations 10 to 13 are written for the full system. Mutual information is more meaningful when two subsystems of the protein are of interest. Using statistical mechanics arguments given by Callen [24], each subsystem may be treated as a canonical ensemble that exchanges energy with its surroundings, represented by the cartoon in Fig 12. The surroundings of Subsystem 1 for example is the protein which

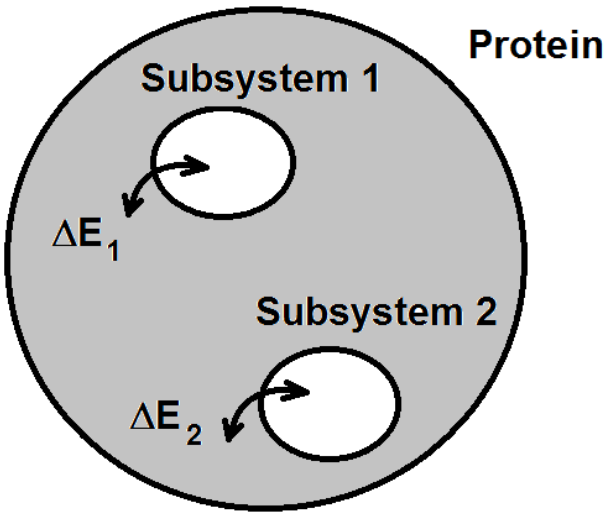


*Fig 12. Energy exchange between protein and its subsystems.*

contains Subsystem 2 also. We may choose the subsystems arbitrarily, an atom, an amino acid, or a secondary structure such as a helix, beta strand, loop or a tail. The subsystem may also be in contact with the surroundings of the protein. Mutual information is zero if the fluctuations of $i$ are independent of the fluctuations of $j$. Otherwise, mutual information is always greater than zero. This leads to the conclusion that correlations always decrease the sum of the individual entropies in a system.

bioRxiv preprint doi: https://doi.org/10.1101/084756; this version posted November 1, 2016. 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.

## Conditional entropy

We consider two trajectories, $\Delta R_i(t)$ and $\Delta R_j(t)$. We now consider two events separated in time by $\tau$, with the condition that $\Delta R_i$ coming before $\Delta R_j$. The conditional entropy for these two events is defined by

$$S\left(\Delta R_j(t+\tau)\middle|\Delta R_i(t)\right) = -\sum p\left(\Delta R_i(t), \Delta R_j(t+\tau)\right) \ln p\left(\Delta R_j(t+\tau)\middle|\Delta R_i(t)\right)$$

$$= \left\langle \ln p\left(\Delta R_j(t+\tau)\middle|\Delta R_i(t)\right)\right\rangle \tag{14}$$

$$= \left\langle \ln \frac{p\left(\Delta R_i(t), \Delta R_j(t+\tau)\right)}{p\left(\Delta R_i(t)\right)} \right\rangle$$

$$= \left\langle \ln p\left(\Delta R_i(0), \Delta R_j(\tau)\right)\right\rangle - \left\langle \ln p\left(\Delta R_i(0)\right)\right\rangle$$

where, the summation is over all states for $i$ and $j$, and the condition of stationarity is used in the last equation.

## Transfer entropy

Following Schreiber's work [22], we write the transfer entropy $T_{i\to j}(\tau)$ from trajectory $i$ to $j$ at time $\tau$ as

$$T_{i\to j}(\tau) = S\left(\Delta R_j(t+\tau)\middle|\Delta R_j(t)\right) - S\left(\Delta R_j(t+\tau)\middle|\Delta R_i(t), \Delta R_j(t)\right) \tag{15}$$

Using the last of Eq. 14, this may be written as

$$\begin{aligned} T_{i\to j}(\tau) = &-\left\langle \ln p\left(\Delta R_j(0), \Delta R_j(\tau)\right)\right\rangle + \left\langle \ln p\left(\Delta R_i(0), \Delta R_j(0), \Delta R_j(\tau)\right)\right\rangle \\ &+ \left\langle \ln p\left(\Delta R_j(0)\right)\right\rangle - \left\langle \ln p\left(\Delta R_i(0), \Delta R_j(0)\right)\right\rangle \end{aligned} \tag{16}$$

bioRxiv preprint doi: https://doi.org/10.1101/084756; this version posted November 1, 2016. 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.

Through the term $p\left(\Delta R_i(0), \Delta R_j(0), \Delta R_j(\tau)\right)$, Eq. 16 requires the evaluation of triple probabilities. If trajectories i and j are independent, then $T_{i\to j}(\tau)=0$ entropy transfer from $i$ to $j$ will be zero. In general, $T_{i\to j}(\tau)\neq T_{j\to i}(\tau)$ and this will determine the net transfer of entropy from one event to another separated in time by $\tau$. Different values of $\tau$ shows how entropy transfer depends on prior interactions. In this study, we will take $\tau = 5ns$ as the representative correlation time of cross correlations.

## Net entropy transfer from an atom

Equation 16 gives the entropy transferred, $T_{i\to j}(\tau)$, from atom i to j. The net entropy transferred from to to all other atoms is obtained by summing over all j as

$$Net\,transfer\,from\,residue\,i = \sum_{j=1}^{N}\left(T_{i\to j}(\tau) - T_{j\to i}(\tau)\right) \tag{17}$$

## Molecular Dynamics simulation

All-atom Molecular Dynamics simulations were performed for unbound (PDB ID: 1UBQ) and bound states (PDB ID: 2KTF ) of Ubiquitin, using NAMD 2.11 simulation program with CHARMM22 All-Hydrogen Parameter set for Proteins and Lipids. TIP3P water model was used to represent water molecules. Counter ions are placed to neutralize the system. Time step of simulations were 2 fs and periodic boundary conditions were applied in an isothermal-isobaric NPT ensemble with constant temperature of 300 K and constant pressure of 1 bar. Temperature and pressure are controlled by Langevin thermostat and Langevin piston barostat, respectively. System coordinates were saved every 1 ps. 1-4 scaling is applied to van der Waals interactions with a cutoff of 12.0 Å. Energy of the system was minimized and system is heated to 300 K for 50 ps and further subjected to MD production run for 600 ns. Frames in trajectories were aligned to the first frame of the simulation by using VMD 1.9.2 to eliminate all rotational and translational degrees of freedom and the analysis is done with the aligned Cartesian coordinates.

bioRxiv preprint doi: https://doi.org/10.1101/084756; this version posted November 1, 2016. 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.

## Entropy Calculations

Amplitude of fluctuations were calculated for each atom from Cartesian coordinates, $R_i(t_k)$, of the trajectory and the mean amplitude of fluctuations, $\overline{R_i}$ was subtracted from each $R_i(t_k)$ and the $\Delta R_{i..N}(t_k)$ matrix was generated with $(t_k, N)$ dimensions, where $N$ is the number of atoms and $t_k$ is the number of frames selected for calculations. All of the calculations in this study will be based on the alpha carbon of each residue unless otherwise stated. Initial data up to equilibration was excluded from the calculations as equilibration. A binning approach was used to calculate configurational entropy, for individual and pairwise dependent atoms. Calculations were performed using MATLAB R2015b. Histogram function of MATLAB was used to cluster data into 8 bins with specified widths and partitioning of data is adaptive according to the maximum and minimum of data. Calculations were performed using 8 discrete bins. Number of bins were selected according to the Sturges' rule . The optimum number, $n_{opt}$, of bins is calculated from the Sturges' rule according to

$$n_{opt} = mean\ fluctuation \cdot (1 + \log_2 N) \tag{18}$$

Here, the mean of fluctuations, i.e., the average fluctuation of the $N$ alpha carbons is calculated from the trajectories and is in the order of 0.5 Å. For a trajectory of 600,000 time steps, the optimum number of bins is obtained as 8 which is used throughout the calculations. After partitioning the fluctuation of each atom into 8 discrete bins, the probabilities were calculated from the frequency of occurrences and entropy was expressed in individual and pairwise mutual information terms. For comparison with benchmark calculations, the configurational entropy was calculated for all heavy atoms by subtracting pairwise mutual information term from individual entropy term as given by Eq. 12. By using Eqn. 16, transfer entropy from ith atom to j was calculated with a delay value of 5 ns for alpha carbons by subtracting the triple conditional entropies from pairwise conditional entropies. Result of configurational entropy calculations were compared with benchmark data and transfer entropy results were used to study changes in entropy transfer patterns when Ubiquitin forms a complex.

bioRxiv preprint doi: https://doi.org/10.1101/084756; this version posted November 1, 2016. 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.

Amount of mutual entropy depends on the distribution of the individual entropies and it is bounded by individual entropy terms.

$$I_2(i,j) \le \min\{S(i), S(j)\} \tag{19}$$

Estimated entropy from a finite sample may be affected by some systematic errors and a correction term is required to get rid of this error [55]. Corrections were applied according to the previous studies [56]

$$S_1^{true} \approx S_1^{estimated} + \frac{M-1}{2N} \tag{20}$$

Where $S_1^{estimated}$ is the raw entropy, $M$ is the number of histogram bins with non-zero probability. Since mutual entropy is a sum of entropies, this formula can also be used to correct $I_2(i,j)$ terms.

$$I_2^{true} \approx I_2^{estimated} + \frac{M_{ij} - M_i - M_j + 1}{2N} \tag{21}$$

Where $M_{ij}$, $M_i$ and $M_j$ represent the numbers of the corresponding histogram bins with non-zero probabilities.

## Benchmark

A benchmark data set which contains one protein-protein complex was used to compare our results with results of MIST(Mutual Information Spanning Trees) method of PARENT [38]. This protein complex is human polymerase iota UBM2 in complex with Ubiquitin, 2KTF.pdb. Unbound state of free Ubiquitin, 1UBQ.pdb, was also simulated. The entropy change upon complex formation was obtained as the difference between configurational entropies of the unbound and bound states.

bioRxiv preprint doi: https://doi.org/10.1101/084756; this version posted November 1, 2016. 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.

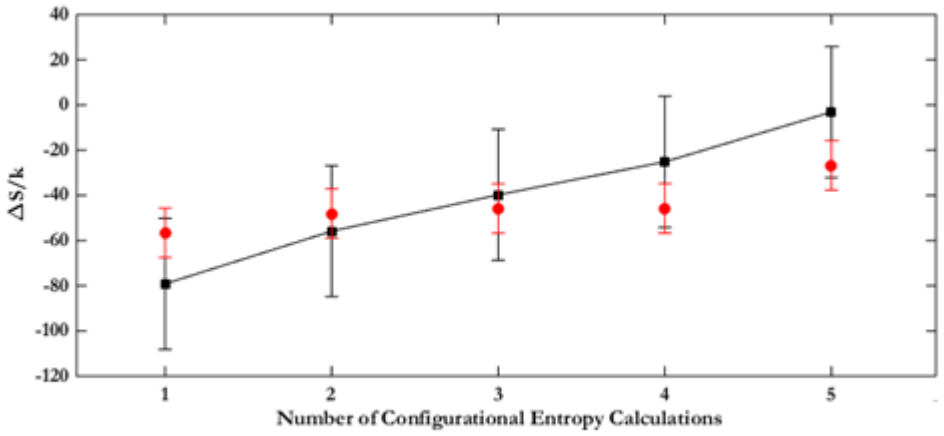


*Fig 13. Comparison of configurational entropy of binding of UBQ and UBM. Black points are from this work, red points are from PARENT. The vertical lines denote the error bars calculated from PARENT.*

The points from this work are in perfect agreement with the results of PARENT.

## Discussion

In the present paper we used the Schreiber's model of entropy transfer and presented a detailed analysis of allosteric communication in Ubiquitin. Based on the analysis of time delayed events, we showed that information may be transferred between pairs of residues. Ubiquitin is not generally known as an allosteric protein but our work shows that there is significant information transfer between residue pairs in this system. From the entropy transfer point of view, all proteins may exhibit allosteric communication. This observation supports the recent hypothesis by Gunasekaran et al [18] that allostery is indeed an intrinsic property of proteins. Our work shows that the knowledge of time delayed correlations and entropy transfer is needed in order to quantify allosteric communication in proteins. Time delayed events have not been widely used in studies of protein function and allosteric communication. Recently, it was shown that causality introduced by time delayed correlations plays significant role on the allosteric communications in K-Ras [48]. In this respect, time delayed correlation functions may be viewed as a new tool for studying allosteric communication in proteins. A three dimensional map of entropy transfer, as shown in Fig 14, which is a three dimensional version of Fig 3, may be useful for visualizing allosteric communication between pairs of residues more easily. The residue index on the 'entropy donor' axis denotes the residues from which entropy goes out, and the residue index on the 'entropy acceptor' axis denotes the residues to

bioRxiv preprint doi: https://doi.org/10.1101/084756; this version posted November 1, 2016. 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.

which entropy goes. For example, residue 76 on the 'entropy acceptor' axis shows that entropy goes into residue 76 from almost all other residues. This means that residue 76 extracts entropy from the protein. Similarly, the peak at residue 53 on the 'entropy donor' axis and residue 23 on the 'entropy acceptor' axis denotes that residue 23 extracts entropy from residue 53.

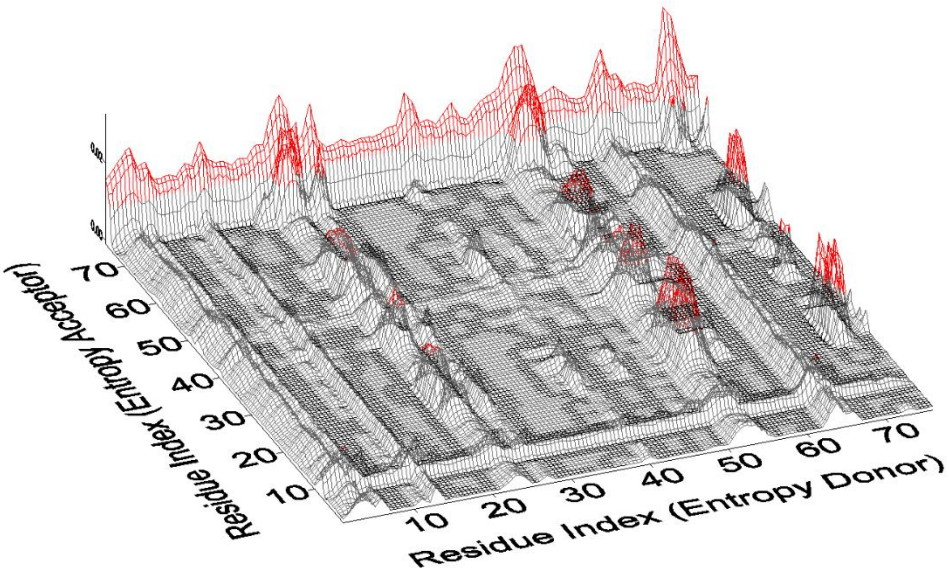


*Fig 14. Three dimensional description of entropy transfer in Ubiquitin. Red regions denote the residues with large contributions to transfer entropy. The figure is a 3-d version of Fig 3.*

Finally, it is worth noting that the present approach which maps the causality, driver-driven relations, and entropy exchange into pairs of residues, as seen in Fig 14, should be of great significance for allosteric drug design because it tells us which residues to manipulate. In this respect, a driver residue is more critical than the driven residue and manipulating the driver will be perturb the existing correlations more efficiently. The effects of mutation on allosteric communication may be quantified by calculating the changes in entropy transfer. As we showed in the UBQ-UBM complex, binding may result in entropy changes in the exposed residues of the complex and change the binding propensities of the complex to other molecules such as another protein, a small molecule ligand or a DNA segment.

bioRxiv preprint doi: https://doi.org/10.1101/084756; this version posted November 1, 2016. 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. Huang, Z., et al., *ASD: a comprehensive database of allosteric proteins and modulators.* Nucleic acids research, 2011. **39**(suppl 1): p. D663-D669.
2. Barbany, M., et al., *Molecular dynamics study of naturally existing cavity couplings in proteins.* PloS one, 2015. **10**(3): p. e0119978.
3. Cerami, E., et al., *The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data.* Cancer Discov, 2012. **2**(5): p. 401-4.
4. Henry, V.J., et al., *OMICtools: an informative directory for multi-omic data analysis.* Database (Oxford), 2014. **2014**.
5. Dror, R.O., et al., *Structural basis for modulation of a G-protein-coupled receptor by allosteric drugs.* Nature, 2013. **503**(7475): p. 295-299.
6. Novinec, M., et al., *A novel allosteric mechanism in the cysteine peptidase cathepsin K discovered by computational methods.* Nature communications, 2014. **5**.
7. Lu, S., W. Huang, and J. Zhang, *Recent computational advances in the identification of allosteric sites in proteins.* Drug discovery today, 2014. **19**(10): p. 1595-1600.
8. Hardy, J.A. and J.A. Wells, *Searching for new allosteric sites in enzymes.* Current opinion in structural biology, 2004. **14**(6): p. 706-715.
9. Panjkovich, A. and X. Daura, *PARS: a web server for the prediction of protein allosteric and regulatory sites.* Bioinformatics, 2014. **30**(9): p. 1314-1315.
10. Collier, G. and V. Ortiz, *Emerging computational approaches for the study of protein allostery.* Archives of biochemistry and biophysics, 2013. **538**(1): p. 6-15.
11. Erman, B., *A fast approximate method of identifying paths of allosteric communication in proteins.* Proteins: Structure, Function, and Bioinformatics, 2013. **81**(7): p. 1097-1101.
12. Shukla, D., et al., *Activation pathway of Src kinase reveals intermediate states as targets for drug design.* Nature communications, 2014. **5**.
13. Tang, S., et al., *Predicting allosteric communication in myosin via a pathway of conserved residues.* Journal of molecular biology, 2007. **373**(5): p. 1361-1373.
14. Kaya, C., et al., *MCPath: Monte Carlo path generation approach to predict likely allosteric pathways and functional residues.* Nucleic acids research, 2013. **41**(W1): p. W249-W255.
15. McClendon, C.L., et al., *Quantifying correlations between allosteric sites in thermodynamic ensembles.* Journal of chemical theory and computation, 2009. **5**(9): p. 2486-2502.
16. Tsai, C.-J. and R. Nussinov, *A unified view of "how allostery works".* PLoS Comput Biol, 2014. **10**(2): p. e1003394.
17. Cooper, A. and D. Dryden, *Allostery without conformational change.* European Biophysics Journal, 1984. **11**(2): p. 103-109.
18. Gunasekaran, K., B. Ma, and R. Nussinov, *Is allostery an intrinsic property of all dynamic proteins?* Proteins: Structure, Function, and Bioinformatics, 2004. **57**(3): p. 433-443.
19. Monod, J., J. Wyman, and J.-P. Changeux, *On the nature of allosteric transitions: a plausible model.* Journal of molecular biology, 1965. **12**(1): p. 88-118.
20. Koshland Jr, D., G. Nemethy, and D. Filmer, *Comparison of experimental binding data and theoretical models in proteins containing subunits*.* Biochemistry, 1966. **5**(1): p. 365-385.
21. Motlagh, H.N., et al., *The ensemble nature of allostery.* Nature, 2014. **508**(7496): p. 331-339.
22. Schreiber, T., *Measuring information transfer.* Physical Review Letters, 2000. **85**(2): p. 461-464.
23. Kamberaj, H. and A. van der Vaart, *Extracting the Causality of Correlated Motions from Molecular Dynamics Simulations.* Biophysical Journal, 2009. **97**(6): p. 1747-1755.
24. Callen, H.B., *Thermodynamics and an introduction to thermostatistics*. Second ed. 1985: Wiley.
25. MacKay, D.J., *Information theory, inference and learning algorithms*. 2003: Cambridge university press.
26. Karplus and Kushick, *Method for Estimating the Configurational Entropy of Macromolecules.*
27. Karplus, M., T. Ichiye, and B.M. Pettitt, *Configurational Entropy of Native Proteins.* Biophysical Journal, 1987. **52**(6): p. 1083-1085.
28. Suarez, E., et al., *CENCALC: A Computational Tool for Conformational Entropy Calculations from Molecular Simulations.* Journal of Computational Chemistry, 2013. **34**(23): p. 2041-2054.

bioRxiv preprint doi: https://doi.org/10.1101/084756; this version posted November 1, 2016. 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.


29. Suarez, E., N. Diaz, and D. Suarez, *Entropy Calculations of Single Molecules by Combining the Rigid-Rotor and Harmonic-Oscillator Approximations with Conformational Entropy Estimations from Molecular Dynamics Simulations.* Journal of Chemical Theory and Computation, 2011. **7**(8): p. 2638-2653.
30. King, B.M., N.W. Silver, and B. Tidor, *Efficient calculation of molecular configurational entropies using an information theoretic approximation.* The Journal of Physical Chemistry B, 2012. **116**(9): p. 2891-2904.
31. Killian, B.J., J. Yundenfreund Kravitz, and M.K. Gilson, *Extraction of configurational entropy from molecular simulations via an expansion approximation.* J Chem Phys, 2007. **127**(2): p. 024107.
32. Killian, B.J., et al., *Configurational Entropy in Protein-Peptide Binding: Computational Study of Tsg101 Ubiquitin E2 Variant Domain with an HIV-Derived PTAP Nonapeptide.* Journal of Molecular Biology, 2009. **389**(2): p. 315-335.
33. Numata, J. and E.W. Knapp, *Balanced and Bias-Corrected Computation of Conformational Entropy Differences for Molecular Trajectories.* J Chem Theory Comput, 2012. **8**(4): p. 1235-45.
34. Hnizdo, V., et al., *Efficient calculation of configurational entropy from molecular simulations by combining the mutual-information expansion and nearest-neighbor methods.* J Comput Chem, 2008. **29**(10): p. 1605-14.
35. Zídek, L., M.V. Novotny, and M.J. Stone, *Increased protein backbone conformational entropy upon hydrophobic ligand binding.* Nature Structural & Molecular Biology, 1999. **6**(12): p. 1118-1121.
36. Suarez, E. and D. Suarez, *Multibody local approximation: Application to conformational entropy calculations on biomolecules.* Journal of Chemical Physics, 2012. **137**(8).
37. Ben-Naim, A., *A farewell to entropy : statistical thermodynamics based on information : S=logW*. 2008, Hackensack, N.J.: World Scientific. xxv, 384 p.
38. Fleck, M., A.A. Polyansky, and B. Zagrovic, *PARENT: A Parallel Software Suite for the Calculation of Configurational Entropy in Biomolecular Systems.* Journal of Chemical Theory and Computation, 2016. **12**(4): p. 2055-2065.
39. van der Vaart, A., *Coupled binding–bending–folding: The complex conformational dynamics of protein-DNA binding studied by atomistic molecular dynamics simulations.* Biochimica et Biophysica Acta (BBA)-General Subjects, 2015. **1850**(5): p. 1091-1098.
40. Karolak, A. and A. van der Vaart, *Importance of local interactions for the stability of inhibitory helix 1 in apo Ets-1.* Biophysical chemistry, 2012. **165**: p. 74-78.
41. Barr, D., et al., *Importance of domain closure for the autoactivation of ERK2.* Biochemistry, 2011. **50**(37): p. 8038-8048.
42. Corrada, D., G. Morra, and G. Colombo, *Investigating Allostery in Molecular Recognition: Insights from a Computational Study of Multiple Antibody–Antigen Complexes.* The Journal of Physical Chemistry B, 2013. **117**(2): p. 535-552.
43. Perilla, J.R., D.J. Leahy, and T.B. Woolf, *Molecular dynamics simulations of transitions for ECD epidermal growth factor receptors show key differences between human and drosophila forms of the receptors.* Proteins: Structure, Function, and Bioinformatics, 2013. **81**(7): p. 1113-1126.
44. Qi, Y.F. and W. Im, *Quantification of Drive-Response Relationships Between Residues During Protein Folding.* Journal of Chemical Theory and Computation, 2013. **9**(8): p. 3799-3805.
45. Jo, S., Y. Qi, and W. Im, *Preferred conformations of N-glycan core pentasaccharide in solution and in glycoproteins.* Glycobiology, 2015: p. cwv083.
46. Zhang, L., T. Centa, and M. Buck, *Structure and Dynamics Analysis on Plexin-B1 Rho GTPase Binding Domain as a Monomer and Dimer.* The Journal of Physical Chemistry B, 2014. **118**(26): p. 7302-7311.
47. Karolak, A., *Application and Development of Computational Methods in Conformational Studies of Biomolecules.* 2015.
48. Vatansever, S., Z. Gumus, and B. Erman, *Intrinsic K-Ras dynamics: A novel molecular dynamics data analysis method shows causality between residue pairs.* bioRxiv, 2016: p. 067496.
49. Kassem, S., et al., *Entropy in bimolecular simulations: A comprehensive review of atomic fluctuations-based methods.* J Mol Graph Model, 2015. **62**: p. 105-17.
50. Numata, J., M. Wan, and E.W. Knapp, *Conformational entropy of biomolecules: beyond the quasi-harmonic approximation.* Genome Inform, 2007. **18**: p. 192-205.
51. Erman, B., *Universal features of fluctuations in globular proteins.* Proteins: Structure, Function, and Bioinformatics, 2016. **84**(6): p. 721-725.
52. Legg, P.A., et al., *Improving accuracy and efficiency of registration by mutual information using Sturges' histogram rule.* Proc. Med. Image Understand. Anal, 2007: p. 26-30.

bioRxiv preprint doi: https://doi.org/10.1101/084756; this version posted November 1, 2016. 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.

53. Liu, J. and R. Nussinov, *The role of allostery in the ubiquitin–proteasome system.* Critical reviews in biochemistry and molecular biology, 2013. **48**(2): p. 89-97.
54. Garner, T.P., et al., *Independent interactions of ubiquitin-binding domains in a ubiquitin-mediated ternary complex.* Biochemistry, 2011. **50**(42): p. 9076-9087.
55. Scarabelli, G., G. Morra, and G. Colombo, *Predicting interaction sites from the energetics of isolated proteins: a new approach to epitope mapping.* Biophysical journal, 2010. **98**(9): p. 1966-1975.
56. Paninski, L., *Estimation of entropy and mutual information.* Neural computation, 2003. **15**(6): p. 1191-1253.