

# Extracting the Causality of Correlated Motions from Molecular Dynamics Simulations

Hiqmet Kamberaj<sup>†</sup> and Arjan van der Vaart<sup>†\*</sup>

<sup>†</sup>Center for Biological Physics, Department of Chemistry and Biochemistry, Arizona State University, Tempe, Arizona; and <sup>‡</sup>Department of Chemistry, University of South Florida, Tampa, Florida

**ABSTRACT** The information theory measure of transfer entropy is used to extract the causality of correlated motions from molecular dynamics simulations. For each pair of correlated residues, the method quantifies which residue drives the correlated motions, and which residue responds. The measure reveals how correlated motions are used to transmit information through the system, and helps to clarify the link between correlated motions and biological function in biomolecular systems. The method is illustrated by its application to the Ets-1 transcription factor, which partially unfolds upon binding DNA. The calculations show dramatic changes in the direction of information flow upon DNA binding, and elucidate how the presence of DNA is communicated from the DNA binding H1 and H3 helices to inhibitory helix HI-1. Helix H4 is shown to act as a relay, which is attenuated in the apo state.

## INTRODUCTION

Protein function is intimately related to protein motion. Of particular interest are correlated, or collective motions, which couple the dynamics of different parts of the protein. Such correlated motions are essential for the coupling of binding sites in allosteric regulation (1–3) and the generation of mechanical work in motor proteins (4,5). They also play a role in catalysis, ligand binding, and protein folding (2,6–10). Correlated motions decrease the configurational entropy, which affects the thermodynamics of these processes (9,10). They alter the kinetics by bringing catalytic residues together, by providing the concerted motions necessary to overcome activation barriers, or by modulating tunneling distances in electron or proton transfer reactions (6–10).

Insights into protein-correlated motions are largely obtained from two techniques. The first is nuclear magnetic resonance (NMR). NMR relaxation experiments have been used to measure the order parameter and the internal correlation time for the motion of individual bond vectors (11). These measures give insights into motions on the pico- to nanosecond timescale, and can, in principle, be used to calculate configurational entropies (10). Other NMR experiments probe correlated motions through residual dipolar couplings (12,13) or multiple-quantum spin relaxation (14), or by measuring the change in order parameters for a series of mutants (15), although simulations suggest that the latter merely probes structural changes due to the mutations (16). The second technique, also used in our study, is based on molecular dynamics (MD) simulations. A popular approach is the calculation of the normalized variance-covariance matrix, or Pearson coefficient, from the simulated trajectory

(17,18). The Pearson coefficient quantifies the correlation between pairs of atoms, and can also be used to compute the configurational entropy (19). Although the Pearson coefficient can be calculated in a straightforward manner, the method has several disadvantages. It only captures linear correlations, and ignores nonlinear correlations; moreover, only correlations between co-linear vectors are included, while neglecting all correlations between perpendicular motions (20). To solve for these deficiencies, the information theory measure of mutual information has been employed (20). Mutual information is sensitive to all statistical dependencies and includes all linear and nonlinear pairwise correlations, regardless of the relative spatial orientations of the vectors.

In this article, we will address the causality of correlated motions. If the motion of residues *i* and *j* is correlated, does residue *i* drive the motion of residue *j*, or does residue *i* respond to the motion of *j*? This question is particularly pertinent in biology, where correlated motions often have a direction or causal relationship. For example, in hemoglobin, the correlated motions among heme, helix F, and the rest of the protein are driven by oxygen binding to the heme, which induces the motions to which helix F, and the rest of the protein, respond (21,22). Another example is the GroEL chaperone, in which the correlated motions of the equatorial, intermediate, and apical domains, that lead to the opening of the *cis* cavity, are driven by ATP binding to the equatorial domain (23–25).

We propose to identify the causality of correlated motions from MD simulations, using the information theory measure of transfer entropy (26). Like mutual information, this measure contains all pairwise linear and nonlinear statistical dependencies, and is insensitive to the relative spatial orientation of the motions. Unlike the Pearson coefficient and mutual information, however, transfer entropy is not a symmetric property: the transfer entropy between residues

Submitted April 29, 2009, and accepted for publication July 7, 2009.

\*Correspondence: vandraart@cas.usf.edu

Editor: Nathan Andrew Baker.

© 2009 by the Biophysical Society  
0006-3495/09/09/1747/9 \$2.00

doi: 10.1016/j.bpj.2009.07.019

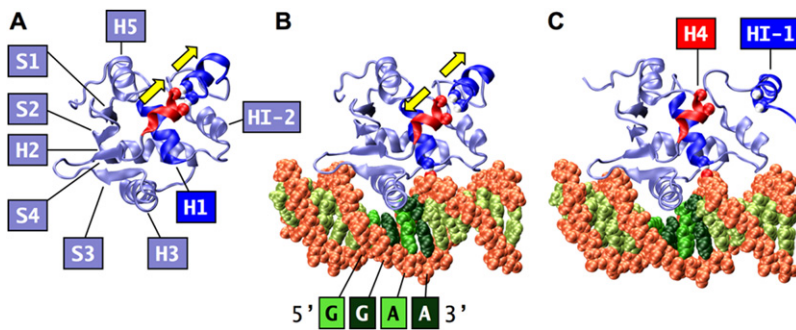


FIGURE 1 Structure of Ets-1  $\Delta 301$ . The figure was prepared with VMD (55) and povray ([www.povray.org](http://www.povray.org)). (A) In apo-Ets-1, the motion between helices H4 and HI-1 is correlated (in-phase), as shown by the parallel arrows. HI-1 is stabilized by the Leu<sup>421</sup>-Phe<sup>304</sup> and Leu<sup>422</sup>-Lys<sup>305</sup> hydrogen bonds with H4 (shown as van der Waals spheres) and macrodipolar interactions with H4. (B) In the metastable, folded Ets-1-DNA complex, the motion between HI-1 and H4 is anticorrelated (out-of-phase), as shown by the arrows. The hydrogen bond between Leu<sup>337</sup> of helix H1 and the T6 phosphate of the complementary DNA strand (shown as van der Waals spheres) acts as a conformational switch. (C) At the onset of unfolding, the anti-correlated motion between H4 and HI-1, respectively, a disruption of the stabilizing macrodipolar interactions between H4 and HI-1, and the outwards motion of HI-1 away from the rest of the protein.

HI-1 lead to the breaking of the Leu<sup>421</sup>-Phe<sup>304</sup> and Leu<sup>422</sup>-Lys<sup>305</sup> hydrogen bonds between H4 and HI-1, respectively, a disruption of the stabilizing macrodipolar interactions between H4 and HI-1, and the outwards motion of HI-1 away from the rest of the protein.

$i$  and  $j$  does not equal the transfer entropy between residues  $j$  and  $i$ . This explicit asymmetry means that the transfer entropy can be used to distinguish between the residues that drive the correlated motion and the residues that respond, between the cause and the effect of correlated motions. The calculation of this information merely requires a standard MD simulation; the transfer entropies are obtained by postprocessing the trajectory. Transfer entropies have been successfully used for signal analysis in other fields, for example for clinical electroencephalography (26–28) and financial data (29), but to our knowledge, no applications to molecular simulation data have been reported. Like other information theory measures, the method is data-intensive and sensitive to noise (27,28). Therefore, we have combined several signal-processing techniques to yield robust and accurate estimates of the transfer entropy, and introduced statistical analyses to optimize the various parameters needed for the calculation. We will summarize our implementation in the methodology section below, and present the technical details in the Appendices.

To illustrate the method, we present its application to Ets-1, a human transcription factor that partially unfolds upon binding its DNA target sequence (30–33). Previous studies revealed that correlated motions are crucial for this highly unusual binding mechanism. Mutation experiments (32) and MD simulations (34) showed that hydrogen bonds between Leu<sup>337</sup> and Gln<sup>336</sup> of helix H1 and the DNA act as a conformational switch, which senses the presence of the DNA and transmits this information to inhibitory helix 1 (HI-1). MD simulations of the folded apo protein, and the metastable, folded Ets-1-DNA complex demonstrated that the conformational switch induces a dramatic change in correlated motions between helix H4 and HI-1 (34). In apo-Ets-1, the motion between H4 and HI-1 is in-phase-correlated, leading to a continuous stabilization of HI-1 through hydrogen bonding and macrodipolar interactions with H4 (Fig. 1 A). In the metastable, folded DNA-bound state, the motion is anti-correlated (out-of-phase), which disrupts the hydrogen bonds and macrodipolar interactions (Fig. 1 B). At the onset of unfolding, the disruptions of these

stabilizing interactions lead to the outward motion of HI-1, away from the rest of the protein (Fig. 1 C). It is thought that ultimately, these motions lead to the experimentally observed unfolding of HI-1.

Here we will apply transfer entropies to investigate the cause of the changes in correlated motions in folded Ets-1 upon DNA binding. Our calculations show how the presence of DNA is transmitted through the protein, providing new mechanistic insights into how DNA destabilizes HI-1 at the onset of unfolding. Given the importance of correlated motions for protein function, we expect that the method will be of general use for a wide range of systems.

## METHODS

The information theory measure of transfer entropy quantifies the statistical coherence between two processes that evolve in time. Since we are interested in the statistical coherence between atomic motions, we take these processes as the positional fluctuations of atoms  $i$  and  $j$ , although in principle, other scalar or vector quantities could be used as well. The positional fluctuations are given by  $r_i(t_n) \equiv |x_i(t_n) - \langle x_i \rangle|$ , where  $\mathbf{x}$  is the atomic position and  $\langle \cdot \rangle$  denotes a time average. Time  $t_n = n\Delta t$  is discrete, with  $\Delta t = M\delta t$  a multiple of the MD integration time step  $\delta t$  and  $M$  the frequency of saving coordinates. By describing the discrete time processes as stationary Markov processes of order  $m$ , the dynamics of the system can be characterized by  $m$ -dimensional state vectors, which we will denote by  $I$ . Using the method of time delayed embedding (35,36), the state vectors are given by

$$\begin{aligned} I_k^{\mu_i} &\equiv (r_i(t_k), r_i(t_k + \tau_i), \dots, r_i(t_k + (m_i - 1)\tau_i))^T, \\ J_k^{\mu_j} &\equiv (r_j(t_k), r_j(t_k + \tau_j), \dots, r_j(t_k + (m_j - 1)\tau_j))^T, \end{aligned} \quad (1)$$

where the time shift  $\tau$  is a multiple of  $\Delta t$ , and the superscript  $T$  indicates the transpose operation. The  $m$  and  $\tau$  embedding parameters can be different for atoms  $i$  and  $j$ ; their specific values are indicated by the superscript  $\mu \equiv (m, \tau)$ . Taking time  $t_{k+(m_i-1)\tau_i}$  as the present,  $I_k^{\mu_i}$  and  $J_k^{\mu_j}$  describe the histories of the processes, which consist of the present and past  $m - 1$  fluctuations. Similarly, the future fluctuations of atoms  $i$  and  $j$  are given by

$$\begin{aligned} I_{k+1} &\equiv r_i(t_{k+m_i\tau_i}), \\ J_{k+1} &\equiv r_j(t_{k+m_j\tau_j}). \end{aligned} \quad (2)$$

To reconstruct the dynamical structure of the time series, the proper choice of  $m$  and  $\tau$  is crucial. We discuss our method to optimally select  $\tau$  and  $m$  from the MD trajectory in the Appendices.

If the fluctuations of atoms  $i$  and  $j$  are independent processes, the conditional probability to observe the future fluctuation of atom  $i$ , given its history, is independent of the history of atom  $j$ ,

$$p(\mathbf{I}_{k+1}|\mathbf{I}_k^{\mu_i}) = p(\mathbf{I}_{k+1}|\mathbf{I}_k^{\mu_i}, \mathbf{J}_k^{\mu_j}) \text{ for independent processes,} \quad (3)$$

where  $p(\mathbf{I}_{k+1}|\mathbf{I}_k^{\mu_i})$  is the conditional probability to observe the future fluctuation of atom  $i$ , given its history, and  $p(\mathbf{I}_{k+1}|\mathbf{I}_k^{\mu_i}, \mathbf{J}_k^{\mu_j})$  is the conditional probability to observe the future fluctuation of atom  $i$ , given the histories of atoms  $i$  and  $j$ . The transfer entropy quantifies the deviation from independence, and is given by the Kullback-Leibler distance (37–39) between these probability distributions (26):

$$T_{j \rightarrow i} = \sum_{k=0}^{N_m} p(\mathbf{I}_{k+1}, \mathbf{I}_k^{\mu_i}, \mathbf{J}_k^{\mu_j}) \log \frac{p(\mathbf{I}_{k+1}|\mathbf{I}_k^{\mu_i}, \mathbf{J}_k^{\mu_j})}{p(\mathbf{I}_{k+1}|\mathbf{I}_k^{\mu_i})}, \quad (4)$$

where  $p(\mathbf{I}_{k+1}, \mathbf{I}_k^{\mu_i}, \mathbf{J}_k^{\mu_j})$  is the joint probability distribution of observing the future atomic fluctuation of atom  $i$ , and the histories of atoms  $i$  and  $j$ . The sum is over all possible state vectors:  $N_m = N - (m - 1)\tau$ , with  $N$  the total number of snapshots. By construction, the transfer entropy is a positive quantity, with a minimum value of zero when the fluctuations of atoms  $i$  and  $j$  are independent, and a maximum of the entropy rate

$$h(\mathbf{I}_{k+1}|\mathbf{I}_k^{\mu_i}) = - \sum_{k=0}^{N_m} p(\mathbf{I}_{k+1}, \mathbf{I}_k^{\mu_i}) \log p(\mathbf{I}_{k+1}|\mathbf{I}_k^{\mu_i})$$

when the fluctuations of atoms  $i$  and  $j$  are completely coupled; this maximum is reached when  $i = j$ , for example. The entropy rate equals the information gained in observing an additional state vector of atom  $i$  when all its previous states are known.

A straightforward derivation shows that the transfer entropy can be reformulated as conditional mutual information ( $I$ ) (27):

$$T_{j \rightarrow i} = I(\mathbf{I}_{k+1}; \mathbf{J}_k^{\mu_j} | \mathbf{I}_k^{\mu_i}). \quad (5)$$

This conditional mutual information equals the amount of information that knowing the history of the fluctuations of atom  $j$  provides about the future fluctuations of atom  $i$ , given the history of fluctuations of atom  $i$ . Moreover, since the transfer entropy is based on conditional probabilities,  $T_{j \rightarrow i} \neq T_{i \rightarrow j}$  in general. This means that the measure can be used to identify causal relationships (26). High values of  $T_{j \rightarrow i}$  indicate that the fluctuations of atom  $j$  are strongly driving the fluctuations of atom  $i$ , whereas low values indicate a smaller dependence. To quantify whether, on average, the motion of atom  $j$  drives the motion of atom  $i$ , or whether the motion of atom  $i$  drives the motion of atom  $j$ , we use the normalized directional index  $D_{j \rightarrow i}$  (27,28),

$$D_{j \rightarrow i} = \frac{T_{j \rightarrow i}}{h(\mathbf{I}_{k+1}|\mathbf{I}_k^{\mu_i})} - \frac{T_{i \rightarrow j}}{h(\mathbf{J}_{k+1}|\mathbf{J}_k^{\mu_j})}, \quad (6)$$

where  $D_{i \rightarrow j} = -D_{j \rightarrow i}$ , with values between  $-1$  and  $+1$ . If  $D_{j \rightarrow i} > 0$ , the fluctuations of atom  $j$  drive the fluctuations of atom  $i$ , or  $j$  is the source of correlations. If  $D_{j \rightarrow i} < 0$ , the motion of atom  $i$  drives the motion of atom  $j$ , while atom  $j$  responds to the motion of atom  $i$ , or  $j$  is the sink of correlations. Although the directional index is a measure of causality, it is not a measure of the independence of the two processes: if  $D_{i \rightarrow j} = D_{j \rightarrow i} = 0$ , there is no overall driving motion between atoms  $i$  and  $j$ , but only if  $T_{j \rightarrow i} = T_{i \rightarrow j} = 0$  the fluctuations of atoms  $i$  and  $j$  are independent. Since  $D_{i \rightarrow j}$  is based on mutual information, the index includes all linear and nonlinear correlations between the two time series.

To limit the effects of statistical noise and to speed up the calculations, we used several specialized techniques. The details of our implementation are described in the Appendices. Although this implementation quantifies the statistical coherence between positional fluctuations, the method could be readily adapted to quantify the statistical coherence between other time-dependent simulation properties, such as displacement vectors, dipole moments, and/or energetic quantities.

## Simulation setup

We analyzed the transfer entropy between the  $C^\alpha$  atoms of apo and DNA-bound Ets-1 and between the P atoms of DNA and the  $C^\alpha$  atoms of Ets-1 using the MD trajectories described in Kamberaj and van der Vaart (34). Since no high-resolution structures are available for the full-length protein, these 15-ns trajectories were generated for Ets-1  $\Delta 301$ , a construct consisting of residues 301–440 with an intact autoinhibitory module and approximately twofold autoinhibition (33). To probe the interactions that lead to the unfolding of HI-1, the simulations were performed on the folded states of Ets-1. The all-atom simulations were performed with the CHARMM force field (40) in the NPT ensemble, coordinates were saved every 500 steps (every ps), and the simulations included explicit water molecules and ions. The simulated DNA sequence in the Ets-1–DNA complex was 5'-AGTGCCGGAAATGTGC-3' with the high affinity 5'-GGAA-3' target sequence. The reader is referred to Kamberaj and van der Vaart (34) for the technical details of the simulations; we used the last 10 ns of the trajectories for the analyses. Given the possibly long timescales of unfolding (33), we did not expect to observe the complete unfolding of HI-1 in the DNA-bound complex in the timescales of the simulation. Instead, we observed the onset of unfolding, consisting of the breaking of the Phe<sup>304</sup>-Leu<sup>421</sup> and Lys<sup>305</sup>-Leu<sup>422</sup> hydrogen bonds between HI-1 and H4, the breaking of the Lys<sup>305</sup>-Glu<sup>428</sup> salt bridge between HI-1 and H5, and the outward motion of HI-1, away from H4 and HI-2 (see Fig. 1). Our objective was to identify the interactions and correlated motions that led to the onset of the unfolding; therefore, the simulation of the complete unfolding process was not necessary and the present simulations met our goals.

## RESULTS

To probe the change in correlated motions that lead to the destabilization of HI-1 upon DNA binding, we analyzed the transfer entropy between the  $C^\alpha$  atoms of Ets-1 in MD trajectories of the folded apo protein in solution, and the metastable, folded Ets-1 in complex with high affinity DNA. Fig. 2 shows the directional index  $D_{j \rightarrow i}$  between the  $C^\alpha$  atoms for apo-Ets-1 (Fig. 2 A) and the Ets-1–DNA complex (Fig. 2 B). In the figure, atom  $j$  is on the vertical, and atom  $i$  on the horizontal axis. Positive values of  $D_{j \rightarrow i}$  (in yellow and red) indicate that the information flow is from  $j$  to  $i$ ; that on average, the fluctuations of atom  $j$  drive the fluctuations of atom  $i$ , and that atom  $j$  is the source of the correlations between atoms  $j$  and  $i$ . Likewise, negative values of  $D_{j \rightarrow i}$  (in blue) signify that atom  $j$  is the sink of the correlations between atoms  $j$  and  $i$ , and that atom  $j$  responds to the motion of atom  $i$ . On the diagonal,  $D_{i \rightarrow i} = 0$ , since  $T_{i \rightarrow i} = h(\mathbf{I}_{k+1}|\mathbf{I}_k^{\mu_i})$  (Eqs. 4 and 6). Fig. 2 C shows which  $C^\alpha$  atoms switched from driver to responder upon DNA binding, and vice versa. Changes from a negative to a positive  $D_{j \rightarrow i}$  upon DNA binding are shown in red; red horizontals indicate  $C^\alpha$  atoms that changed from responder to driver upon DNA binding. The opposite change is shown in blue; atom pairs that did not switch sign of  $D_{j \rightarrow i}$  are shown in white. For clarity, only one shade of red and blue is used in Fig. 2 C. Fig. 2, D and E, show  $D_{j \rightarrow i}$  between the P atoms of each DNA strand and the  $C^\alpha$  atoms of Ets-1.

Fig. 2 shows dramatic changes in the directional index upon DNA binding. For many atom pairs, ~36% of the total, the driver and responder switched roles upon DNA binding

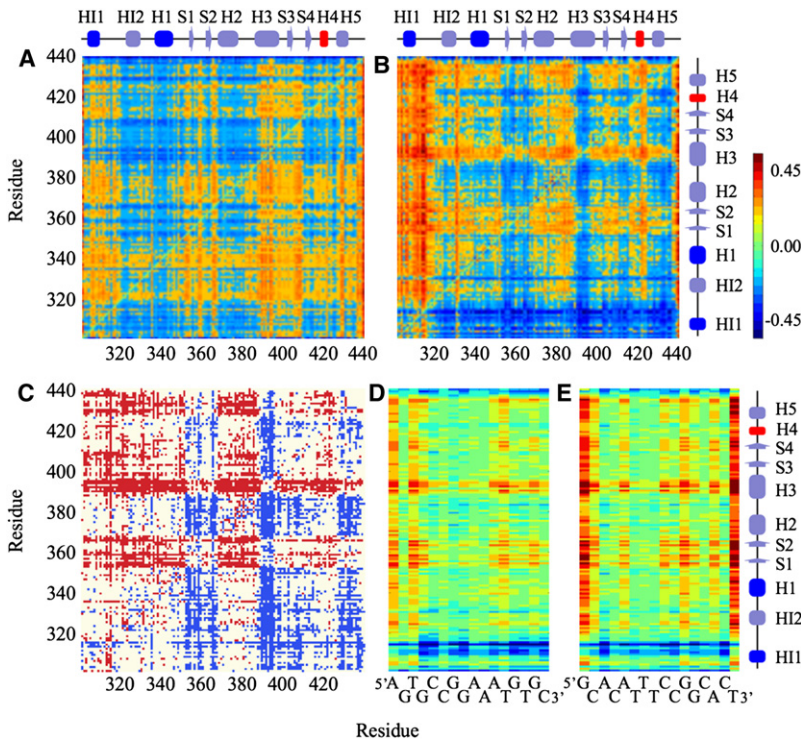


FIGURE 2 The normalized directional index. Atom  $j$  is on the vertical, and atom  $i$  on the horizontal axis. (A)  $D_{j \rightarrow i}$  for the Ets-1 C $^{\alpha}$  atoms in apo-Ets-1 and (B) the Ets-1-DNA complex. (C) Sign switching of  $D_{j \rightarrow i}$  upon DNA binding. Changes from negative to positive values are shown in red, changes from positive to negative values in blue, and changes that did not involve a switch of sign (including changes from  $|D_{j \rightarrow i}| \leq 0.05$ ) are shown in white. (D)  $D_{j \rightarrow i}$  between the P atoms of the DNA main (strand 1) and (E) complementary strand (strand 2) and the C $^{\alpha}$  atoms of Ets-1.

(Fig. 2 C). Especially S1, S2, H3, S3, S4, and H5 showed many switches from responding to driving residues upon binding. For residue pairs that did not switch, DNA binding increased the absolute magnitude of the directional index for 41% of all atom pairs, while decreasing the absolute magnitude for the remaining 25%. Fig. 2, D and E show that Ets-1 drives most of the correlated motions between the protein and DNA; especially large drivers are H3, S1, and S2.

Of particular interest are the transfer entropies of the H1, H3, H4, and HI-1 helices, which are crucial for the binding mechanism (31–34). H3 is the recognition helix, which binds the major groove of DNA (Fig. 1). In the complex, H3 drives the motion of the DNA P atoms in the binding site (the horizontal for H3 is mostly red and yellow in Fig. 2, D and E). For example, Tyr<sup>395</sup> drives the motion of G8 and A9 in strand 1 (the main strand), Arg<sup>391</sup> drives the motion of the T8 and C9 of strand 2 (the complementary strand), and Arg<sup>394</sup> drives the motion of G7, G8, A9, and A10 of strand 1 and T6, T7, C8, and C9 of strand 2. In apo-Ets-1, H3 acts as a sink for the correlated motions (the horizontal for H3 is mostly blue, the vertical is mostly yellow in Fig. 2 A). This means that in the apo-protein, helix H3 responds to the motion of most other protein residues. H3 behaves very differently in the DNA-bound state (Fig. 2 B): in the DNA-complex, H3 is mostly a source of the correlated motions with the other protein residues. This change upon DNA binding makes biological sense: the recognition helix senses the presence of the DNA and transmits this information to the rest of the protein through correlated motions.

Helix H1 contains the conformational switch, consisting of Leu<sup>337</sup> and Gln<sup>336</sup>, which hydrogen-bond with the T6 phosphate of DNA strand 2 (32–34). Fig. 2, D and E show that the correlated motions between Leu<sup>337</sup>, Gln<sup>336</sup>, and this phosphate are driven by the P atom of T6. In apo-Ets-1, H1 is a source for the correlated motions with all other secondary structure elements. In the DNA-bound state, the causality of these correlations is reversed for half the elements. In the complex, H1 is a sink for the correlated motions with S1, S2, H3, S4, and H5, and a source of the correlations with HI-1, HI-2, H2, S3, and H4. H1 becomes a much stronger source for the correlations with HI-1 upon DNA binding, and a slightly stronger source for the correlations with H4.

In the apo-protein, the motion between H4 and HI-1 is in-phase correlated, and H4 stabilizes HI-1 through macrodipolar and hydrogen-bonding interactions. Upon DNA binding, the motion between H4 and HI-1 becomes anti-correlated, and the stabilizing interactions are disrupted (34). H4 is a very weak source for the correlated motions with HI-1, H3, S3, and H5 in the apo protein, and a sink for the correlated motions with the other secondary structure elements. In the DNA-bound complex, H4 is a strong sink for the correlated motions with most residues, and a strong source for the correlations with HI-1. HI-1 is the helix that unfolds upon binding DNA (30–33). In apo-Ets-1, HI-1 is a weak source of the correlations with H3, S3, and H5, and a sink for the correlations with HI-2, H1, S1, S2, H2, S4, and H4. Upon DNA-binding, HI-1 becomes a very strong sink of the correlations with all other protein residues. In addition, parts of

HI-1 and the loop between HI-1 and HI-2 are sinks for the correlations with the DNA.

Other regions of interest are the loop region between S3 and S4, which contact the DNA. Correlations between this loop region and the DNA are partially driven by the DNA and partially by the protein. Lys<sup>404</sup> drives the motion of T3, G4, and C6 and is a sink for the correlation with C5 of strand 1. Thr<sup>405</sup>, Ala<sup>406</sup>, Gly<sup>407</sup>, Lys<sup>408</sup>, and Arg<sup>409</sup> act as a sink for the correlated motions with T3, G4, C5, and C6 of strand 1. We also observed that the correlated motions between Ets-1 and the highly flexible bases at the end of DNA strands are driven by Ets-1.

From these analyses, a clear picture of the information flow within the protein emerges. The presence of DNA is transferred by a network of correlated motions, and the direction of the information flow is from the driving to the responding residues. The network starts with the DNA sensing H1 and H3 helices. DNA binding changes the H3 recognition helix from a sink to a source of correlated motions, transmitting the information that DNA is present to the rest of the protein by driving correlated motions. The T6 phosphate of DNA strand 2 activates the conformational switch, by hydrogen bonding with Leu<sup>337</sup> and Gln<sup>336</sup> of helix H1, and driving the correlated motions with these residues. In addition, H1 becomes a stronger driver of correlated motions with H4 and HI-1 upon DNA binding. H4, the helix that stabilizes HI-1 in the apo state, and destabilizes HI-1 in the DNA-bound state, plays a central role in the correlated network. Whereas in the apo state, H4 is a very weak source for the correlated motions with HI-1, and either a weak source or weak sink for the correlated motions with the rest of the protein, in the DNA-bound state, H4 is a strong source of correlated motions with HI-1 and a strong sink for the correlated motions with the other protein residues. H4 responds to the motion of all non-HI-1 protein residues in the DNA-bound state, and transmits this information to HI-1 by strongly driving the correlations with this helix. Thus, H4 acts as a relay between HI-1 and the other residues, transmitting the information that DNA is present from all other protein residues onto HI-1; moreover, in the apo state, this relay is attenuated. Finally, HI-1 is on the receiving end of all correlated motions in the DNA-bound state, and strongly responds to the motions of all other residues when DNA is bound.

## DISCUSSION

We have used the information theory measure of transfer entropy to extract the causality of correlated motions from MD trajectories. The method dissects each pair of correlated residues into a residue that drives the motion and a residue that responds, and establishes which residues dictate the observed correlations in the system. Combined with existing techniques to calculate the correlations (for example, the Pearson coefficient (17,18) or mutual information (20)),

the methods quantifies the flow of information in biomolecules, and gives new insights into how correlated motions relate to biological function.

We illustrated the method by its application to Ets-1, a protein which unfolds its HI-1 helix upon binding the target DNA sequence (30–33). The transfer entropy analysis revealed that the flow of information drastically changed upon DNA binding, showing how the presence of DNA is transmitted by correlated motions. The T6 phosphate which hydrogen-bonds Leu<sup>337</sup> and Gln<sup>336</sup> drives the correlations with these protein residues, activating the conformational switch. The H3 recognition helix drives the correlated motions with the DNA and with the rest of the protein in the complex. In the apo-protein, H4 is either a weak driver or weak sink for the correlated motions with the other residues, while in the complex H4 strongly drives the correlations with HI-1, and strongly responds to the motions of the rest of the protein (including the DNA-binding H1 and H3 helices). HI-1 is a strong sink for the correlations with all other residues (including the DNA) in the complex; its motion is dictated by the rest of the system. Thus, the analysis showed how the information of the presence of DNA is transmitted through the protein, from the DNA-binding H3 and H1 helices, to helix H4 and HI-1. The calculations identified a central role for helix H4 as a relay of information flow between the DNA-binding helices and HI-1 in the DNA-bound state, and showed that this relay is attenuated in the apo state.

In the Appendices, we describe the technical aspects of the calculation of transfer entropies from MD trajectories. We used several techniques to reduce the effects of noise and to speed up the calculation, and statistical analyses to estimate the optimal values of the  $m$  and  $\tau$  embedding parameters. To verify the accuracy and robustness of our implementation, we have performed extensive tests on a large variety of (noisy) dynamical systems (data not shown). Our implementation provides robust estimates of the transfer entropy, and merely requires a trajectory from a standard MD simulation. Given the importance of correlated motions for biological function, we expect that transfer entropies will be useful for the study of a wide range of biomolecular systems.

## APPENDIX A: TRANSFER ENTROPY

In practice, we calculate the transfer entropy from the Shannon information entropy  $H$ . Using  $\mu \equiv (m, \tau)$ ,  $\mu + 1 \equiv (m + 1, \tau)$ , and the formulation of the transfer entropy as conditional mutual information (Eq. 5),

$$\begin{aligned} T_{j \rightarrow i} &= I(\mathbf{I}_{k+1}; \mathbf{J}_k^{\mu_j} | \mathbf{I}_k^{\mu_i}) \\ &= H(\mathbf{I}_k^{\mu_i+1}) - H(\mathbf{I}_k^{\mu_i}) - H(\mathbf{I}_k^{\mu_i+1}, \mathbf{J}_k^{\mu_j}) + H(\mathbf{I}_k^{\mu_i}, \mathbf{J}_k^{\mu_j}). \end{aligned} \quad (7)$$

The Shannon entropy (41) is given by

$$H(\mathbf{I}_k^{\mu_i}) = - \sum_k p(\mathbf{I}_k^{\mu_i}) \log p(\mathbf{I}_k^{\mu_i}), \quad (8)$$

where the sum is over all states.

In general, the length of the discrete time processes and the number of states are limited by the sampling. To correct for the finite sampling, we use (42)

$$H(\mathbf{I}_k^{\mu_i}) = \ln N - \frac{1}{N} \sum_k n_k \psi(n_k), \quad (9)$$

where sum is over all states,  $n_k$  is the frequency of observing state  $k$ , and  $\psi(x)$  is the derivative of the  $\gamma$ -function (42). Similar expressions can be obtained for the other Shannon entropy terms in Eq. 7. To reduce the effect of noise and to speed up the calculations, we do not use the states  $\mathbf{I}_k^{\mu_i}$  and  $\mathbf{I}_k^{\mu_{i+1}}$  directly. Instead, we use the technique of symbolization (28,43–46) as described in Appendix D.

## APPENDIX B: SHUFFLING OF DATA AND NORMALIZATION OF TRANSFER ENTROPIES

Since the probability distributions of the time series are not known a priori, the computation of the transfer entropy is complicated and subject to noise (27). Moreover, for large time shifts  $\tau_1$  and  $\tau_2$ , a broad joint distribution will be obtained. For such a broad distribution, the transfer entropy can be different from zero even when there is no causal link between the two time series (27). The use of an effective transfer entropy  $\bar{T}_{j \rightarrow i}^E$  based on the shuffling of data removes this effect (27,29). We have adopted this measure in our implementation.

The shuffling of a time series  $I$  or  $J$  completely removes any link between  $I$  and  $J$  without changing their distributions (27,29). This means that the transfer entropy calculated from a shuffled time series can serve as a significance threshold, to distinguish information flow from artifacts due to insufficient sampling. The new measure is the effective transfer entropy (29)

$$\bar{T}_{j \rightarrow i}^E = T_{j \rightarrow i} - \frac{1}{N_{\text{trials}}} \sum_{n=1}^{N_{\text{trials}}} T_{j_{\text{shuffled}} \rightarrow i}, \quad (10)$$

where the second terms represent the averages over all the  $N_{\text{trials}}$  random shufflings of the time series. In this study, we used  $N_{\text{trials}} = 100$ . The effective transfer entropy is normalized by (27)

$$T_{j \rightarrow i}^E = \frac{\bar{T}_{j \rightarrow i}^E}{h(\mathbf{I}_{k+1} | \mathbf{I}_k^{\mu_i})}, \quad (11)$$

where

$$h(\mathbf{I}_{k+1} | \mathbf{I}_k^{\mu_i}) = - \sum_{k=0}^{N_m} p(\mathbf{I}_{k+1}, \mathbf{I}_k^{\mu_i}) \log p(\mathbf{I}_{k+1} | \mathbf{I}_k^{\mu_i})$$

is the entropy rate. In Eq. 11,  $T_{j \rightarrow i}^E$  is distributed in the interval  $[0, 1]$ . The preferred direction of flow is given by the normalized directional index  $D_{j \rightarrow i}^E$  (27,28)

$$D_{j \rightarrow i}^E = T_{j \rightarrow i}^E - T_{i \rightarrow j}^E, \quad (12)$$

where  $D_{i \rightarrow j}^E$  is in the interval  $[-1, 1]$ . All our reported results are based on the effective transfer entropies, and the normalized directional index (Eq. 12).

## APPENDIX C: OPTIMIZATION OF THE $M$ AND $\tau$ EMBEDDING PARAMETERS

The correct choice of the  $m$  and  $\tau$  embedding parameters is crucial for the proper characterization of the structure of the time series (Eq. 1) (47–49).

The objective is to properly describe the dynamics of a lower dimensional space (here, the fluctuations  $r_i$  of a subset of atoms) from the full dynamics of the system, where, as an additional complication, the full dynamics are given as discrete solutions of the continuous equation of motions. For example, very small values of  $\tau$  will result in strongly correlated state vector elements, whereas large values will result in elements that are randomly distributed in the  $m$ -dimensional space; the projected trajectories will cross for small values of  $m$ , whereas for large values of  $m$ , noise will start to dominate and the computation becomes more expensive. This problem has been extensively studied in dynamical systems theory (47–51), and the reader is referred to the original literature for an in-depth treatment of the subject. For example, the mathematical concepts for selecting the state vector dimension  $m$  have been reviewed in Noakes (50) and Sauer et al. (51). Several methods have been proposed for estimating the optimal values of  $m$  and  $\tau$  simultaneously (47,48), as well as a method in which they are individually optimized (49). Since a comparison showed its superiority (49), we largely followed the latter approach.

In this approach, the time lag parameter  $\tau$  is chosen as the time at which the mutual information  $I(\mathbf{I}_k; \mathbf{I}_{k+\tau})$  has its first minimum (49,52,53). For this purpose, the mutual information is written as a function of  $\tau$ ,

$$\begin{aligned} I(\mathbf{I}_k; \mathbf{I}_{k+\tau}, \tau) &= \sum_k p(\mathbf{I}_k, \mathbf{I}_{k+\tau}) \log \frac{p(\mathbf{I}_k, \mathbf{I}_{k+\tau})}{p(\mathbf{I}_k)p(\mathbf{I}_{k+\tau})} \\ &= H(\mathbf{I}_k) + H(\mathbf{I}_{k+\tau}) - H(\mathbf{I}_k, \mathbf{I}_{k+\tau}), \end{aligned} \quad (13)$$

where  $\mathbf{I}_k$  is the value of the time series  $I$  at time  $k$ , and  $\mathbf{I}_{k+\tau}$  at time  $k + \tau$ . Since the calculation of the probability depends on the number of bins used, a dependence of the Shannon entropy on the number of bins is expected (49). In our implementation, we optimize the number of bins, by using the number of bins that maximizes the Shannon information entropy ( $H(\mathbf{I}_k)$ ). As convergence criterion for the number of bins, we use  $|H(N_{\text{bins}} + 1) - H(N_{\text{bins}})| < 0.0001$ ; using this method, we optimized  $\tau_i$  for each atom  $i$ .

The optimized number of bins and the optimized time-lag parameters for the  $C^\alpha$  atoms of apo-Ets-1 and the Ets-1-DNA complex are shown in Fig. 3. For both systems,  $\tau$  varied between 1 and 17, and the number of bins between 300 and 450. The value  $\tau$  is closely correlated to the flexibility of the atoms: the larger the flexibility, the larger the value of  $\tau$ .

After optimizing the time shifts, we used the false nearest-neighbors method (47–49) to obtain the optimal estimate for  $m$ . In this method, the Euclidean distance between two state vectors  $\mathbf{I}_k^\mu$  and  $\mathbf{I}_l^\mu$  in the  $m$ -dimensional space is given by

$$R_i^m = \left( \sum_{n=0}^{m-1} (r_i(t_{k+n\tau}) - r_i(t_{l+n\tau}))^2 \right)^{1/2}. \quad (14)$$

For notational clarity, we have dropped all subscripts  $i$  in  $\mu_i$ ,  $\tau_i$ , and  $m_i$ . The state  $\mathbf{I}_l^\mu$  with the smallest distance is the nearest neighbor of  $\mathbf{I}_k^\mu$ ; we will denote the index  $l$  of the nearest neighbor by  $NN$ . Extending  $\mathbf{I}_k^\mu$  and  $\mathbf{I}_{NN}^\mu$  by one element into  $(m+1)$ -dimensional space, the distance between the extended states is given by

$$R_i^{m+1} = \left( (R_i^m)^2 + (r_i(t_{k+m\tau}) - r_i(t_{NN+m\tau}))^2 \right)^{1/2}. \quad (15)$$

This distance is normalized against the distance in  $m$ -dimensional space (52):

$$\gamma_i^m = \left( \frac{(R_i^{m+1})^2 - (R_i^m)^2}{(R_i^m)^2} \right)^{1/2} = \frac{|r_i(t_{k+m\tau}) - r_i(t_{NN+m\tau})|}{R_i^m}. \quad (16)$$

Here,  $\gamma_i^m$  is compared to a threshold value  $R_{\text{tol}}$ , which is determined a priori (49,52) and recommended to be 15 (49,52). If  $\gamma_i^m$  exceeds  $R_{\text{tol}}$ , then  $\mathbf{I}_{NN}^\mu$  is a false nearest neighbor of  $\mathbf{I}_k^\mu$  and  $f_{NN}$ , the frequency of false nearest

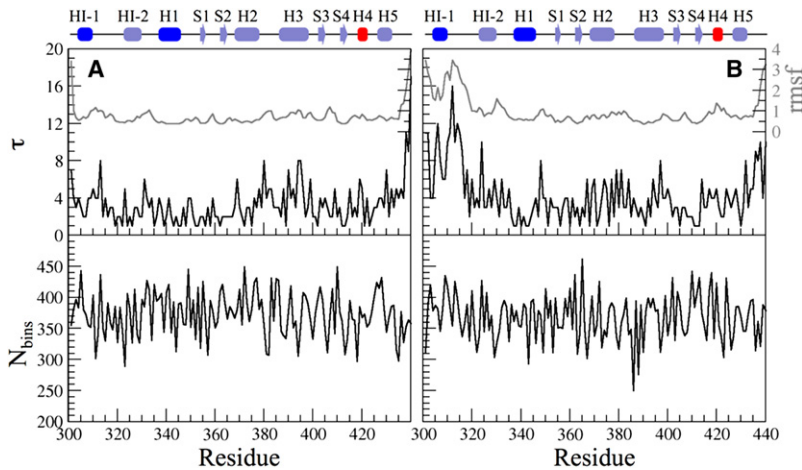


FIGURE 3 Optimized number of bins ( $N_{\text{bins}}$ ) and time lag parameters ( $\tau$ ) for the  $C^\alpha$  atoms of Ets-1. The root mean-square fluctuation (in Å) of the  $C^\alpha$  atoms of Ets-1 is shown in shaded representation. (A) Apo-Ets-1. (B) Ets-1-DNA complex.

neighbors, is increased by one. The value of  $m$  is increased until  $f_{NN}$  approaches zero. The procedure is repeated to optimize  $m_i$  for each atom  $i$ .

Fig. 4 shows  $m$  and  $f_{NN}$  for apo-Ets-1 and the Ets-1-DNA complex. Similar values were obtained for both systems, with  $m$  varying between 6 and 9.

#### APPENDIX D: COARSE-GRAINING OF THE TIME SERIES

The calculation of the transfer entropy can be greatly sped up, and the effects of noise can be substantially reduced by using the technique of symbolization (28,43–46). In this method, a symbolic sequence  $\{\hat{I}_1, \hat{I}_2, \dots, \hat{I}_N\}$  is associated with the time-series  $\{I_1, I_2, \dots, I_N\}$  through coarse-graining (43–45). In the coarse-graining, all information concerning the dynamics of the series is suitably encoded using a partitioning of phase space. The time-series  $\{I_1, I_2, \dots, I_N\}$  is converted into a symbolic sequence by the rule

$$\hat{I}_j = \hat{S}_k \text{ if } X_k < I_j < X_{k+1}, \quad (17)$$

where  $\{X_0, X_1, \dots, X_{N_b}\}$  is a given set of  $N_b + 1$  critical points, and  $\{\hat{S}_0, \hat{S}_1, \dots, \hat{S}_{N_b-1}\}$  is a set of  $N_b$  symbols, here the numbers 0, 1, 2, etc. Concatenation of the symbols of a subsequence of length  $m$  yields the word  $w_i$ ,

$$w_i = \sum_{l=1}^m N_b^{m-l} \hat{I}_{k+(i-1)\tau}, \quad (18)$$

where  $k$  indicates the starting position of the subsequence along the sequence  $\{\hat{I}_1, \hat{I}_2, \dots, \hat{I}_N\}$ . A particular subsequence is uniquely characterized by the word  $w_i$ .

The probability of finding a particular value of  $w_i$  is calculated from the simulation data, and used to compute the Shannon entropy

$$H(\hat{I}_k^\mu) = - \sum_{\{w_i\}} p(w_i) \log p(w_i). \quad (19)$$

Since the time series  $\{I_1, I_2, \dots, I_N\}$  is mapped onto the symbolic sequence  $\{\hat{I}_1, \hat{I}_2, \dots, \hat{I}_N\}$  uniquely (i.e., the symbolic representation is injective), the entropies  $H(\hat{I}_k^\mu)$  and  $H(I_k^\mu)$  coincide (54).

We obtain the critical points  $\{X\}$  for a particular series by maximizing the entropy with respect to all possible partitions. Increasing the number of critical points will initially increase the information entropy, but after a sufficient number of critical points, the information entropy plateaus. At this point the optimum number of critical points has been reached: a further increase will not increase the accuracy of the calculation, but does slow down the computation. In our implementation, we optimize both  $N_b$  and  $\{X\}$  by maximizing the Shannon entropy through a Monte Carlo approach. We chose  $N_b = 2, 3, 4,$  and  $5$  and randomly selected  $N_b + 1$  critical points  $X$  in the range from the minimum to the maximum observed value of the time series. This process is repeated many times (here, 10,000 times) to yield the  $N_b$  and  $X$  for which the information entropy was a maximum. In a similar way, the joint Shannon information entropy of two discrete symbolic processes  $\{\hat{I}_1, \hat{I}_2, \dots, \hat{I}_N\}$  and  $\{\hat{J}_1, \hat{J}_2, \dots, \hat{J}_N\}$  is calculated as

$$H(\hat{I}_k^\mu, \hat{J}_k^\mu) = - \sum_{\{w_i\}, \{w_j\}} p(w_i, w_j) \log p(w_i, w_j), \quad (20)$$

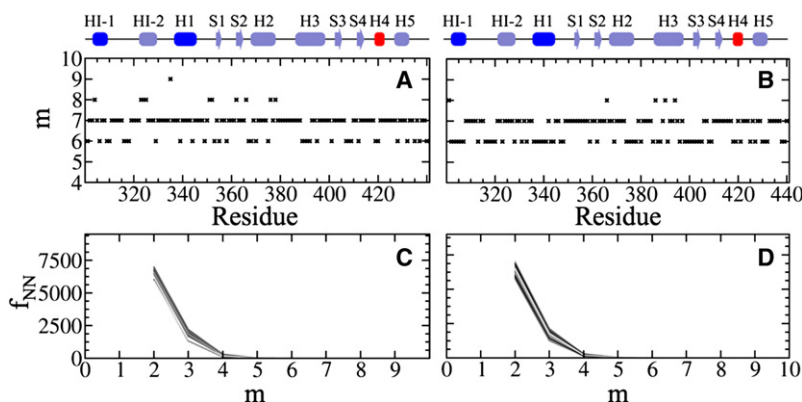


FIGURE 4 Optimized embedding state dimension  $m$  and the frequency of false nearest neighbors  $f_{NN}$  for the  $C^\alpha$  atoms of Ets-1. (A)  $m$  for apo-Ets-1 and (B) the Ets-1-DNA complex. (C)  $f_{NN}$  for apo-Ets-1 and (D) the Ets-1-DNA complex.

where the sum is over all  $\{w_1\}$  and  $\{w_2\}$  states. Usually the entropy is measured in units of bits; in this case, the logarithmic functions in Eqs. 19 and 20 are computed with base  $N_b$ . The Shannon entropies of Eqs. 19 and 20 are used to calculate the transfer entropy of Eqs. 7 and 10.

We thank the TeraGrid and the Fulton High Performance Computing Initiative at Arizona State University for computer time.

This work was supported by the National Science Foundation Career Award No. CHE-0846161.

## REFERENCES

- Goodey, N., and S. Benkovic. 2008. Allosteric regulation and catalysis emerge via a common route. *Nat. Chem. Biol.* 4:474–482.
- Hammes, G. 2002. Multiple conformational changes in enzyme catalysis. *Biochemistry.* 41:8221–8228.
- Kern, D., and E. Zuiderweg. 2003. The role of dynamics in allosteric regulation. *Curr. Opin. Struct. Biol.* 13:748–757.
- Vale, R., and R. Milligan. 2000. The way things move: looking under the hood of molecular motor proteins. *Science.* 288:88–95.
- Schliwa, M., and G. Woehlke. 2003. Molecular motors. *Nature.* 422:759–765.
- Benkovic, S., and S. Hammes-Schiffer. 2003. A perspective on enzyme catalysis. *Science.* 301:1196–1202.
- Hammes-Schiffer, S., and S. Benkovic. 2006. Relating protein motion to catalysis. *Annu. Rev. Biochem.* 75:519–541.
- Bruice, T., and S. Benkovic. 2000. Chemical basis for enzyme catalysis. *Biochemistry.* 39:6267–6274.
- Wand, J. 2001. On the dynamic origins of allosteric activation. *Science.* 293:1395.
- Jarymowycz, V., and M. Stone. 2006. Fast time scale dynamics of protein backbones: NMR relaxation methods, applications, and functional consequences. *Chem. Rev.* 106:1624–1671.
- Lipari, G., and A. Szabo. 1982. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J. Am. Chem. Soc.* 104:4546–4559.
- Tolman, J., J. Flanagan, M. Kennedy, and J. Prestegard. 1997. NMR evidence for slow collective motions in cyanometmyoglobin. *Nat. Struct. Biol.* 4:292–297.
- Bouvignies, G., P. Bernadó, S. Meier, K. Cho, S. Grzesiek, et al. 2005. Identification of slow correlated motions in proteins using residual dipolar and hydrogen-bond scalar couplings. *Proc. Natl. Acad. Sci. USA.* 102:13885–13890.
- Lundström, P., F. Mulder, and M. Akke. 2005. Correlated dynamics of consecutive residues reveal transient and cooperative unfolding of secondary structure in proteins. *Proc. Natl. Acad. Sci. USA.* 102:16984–16989.
- Mayer, K., M. Earley, S. Gupta, K. Pichumani, L. Regan, et al. 2003. Covariance of backbone motion throughout a small protein domain. *Nat. Struct. Biol.* 10:962–965.
- Lange, O., H. Grubmüller, and B. de Groot. 2005. Molecular dynamics simulations of protein G challenge NMR-derived correlated backbone motions. *Angew. Chem. Int. Ed.* 44:3394–3399.
- Ichiye, T., and M. Karplus. 1991. Collective motions in proteins: a covariance analysis of atomic fluctuations in molecular-dynamics and normal mode simulations. *Proteins.* 11:205–217.
- Teeter, M., and D. Case. 1990. Harmonic and quasiharmonic descriptions of crambin. *J. Phys. Chem.* 94:8091–8097.
- Karplus, M., and J. Kushick. 1981. Method for estimating the configurational entropy of macromolecules. *Macromolecules.* 14:325–332.
- Lange, O., and H. Grubmüller. 2006. Generalized correlation for biomolecular dynamics. *Proteins.* 62:1053–1061.
- Baldwin, J., and C. Chothia. 1979. Hemoglobin: the structural changes related to ligand binding and its allosteric mechanism. *J. Mol. Biol.* 129:175–220.
- Perutz, M., A. Wilkinson, M. Paoli, and G. Dodson. 1998. The stereochemical mechanism of the cooperative effects in hemoglobin revisited. *Annu. Rev. Biophys. Biomol. Struct.* 27:1–34.
- Grallert, H., and J. Buchner. 2001. A structural view of the GroE chaperone cycle. *J. Struct. Biol.* 135:95–103.
- Saibil, H., and N. Ranson. 2002. The chaperonin folding machine. *Trends Biochem. Sci.* 27:627–632.
- Karplus, M., Y. Gao, J. Ma, A. van der Vaart, and W. Yang. 2005. Protein structural transitions and their functional role. *Phil. Trans. Roy. Soc. A.* 363:331–356.
- Schreiber, T. 2000. Measuring information transfer. *Phys. Rev. Lett.* 85:461–464.
- Gourévitch, B., and J. Eggermont. 2007. Evaluating information transfer between auditory cortical neurons. *J. Neurophysiol.* 97:2533–2543.
- Staniek, M., and K. Lehnertz. 2008. Symbolic transfer entropy. *Phys. Rev. Lett.* 100:158101.
- Marschinski, R., and H. Kantz. 2002. Analyzing the information flow between financial time series. *Eur. Phys. J. B.* 30:275–281.
- Petersen, J., J. Skalicky, L. Donaldson, L. McIntosh, T. Alber, et al. 1995. Modulation of transcription factor Ets-1 DNA binding: DNA-induced unfolding of an helix. *Science.* 269:1866–1869.
- Skalicky, J., L. Donaldson, J. Petersen, B. Graves, and L. McIntosh. 1996. Structural coupling of the inhibitory regions flanking the ETS domain of murine Ets-1. *Protein Sci.* 5:296–309.
- Wang, H., L. McIntosh, and G. Graves. 2002. Inhibitory module of Ets-1 allosterically regulates DNA binding through a dipole-facilitated phosphate contact. *J. Biol. Chem.* 277:2225–2233.
- Lee, G., L. Donaldson, M. Pufall, H.-S. Kang, I. Pot, et al. 2005. The structural and dynamic basis of Ets-1 DNA binding autoinhibition. *J. Biol. Chem.* 280:7088–7099.
- Kamberaj, H., and A. van der Vaart. 2009. Correlated motions and interactions at the onset of the DNA-induced partial unfolding of Ets-1. *Biophys. J.* 96:1307–1317.
- Packard, N., J. Crutchfield, J. Farmer, and R. Shaw. 1980. Geometry from a time series. *Phys. Rev. Lett.* 45:712–716.
- Takens, F. 1980. Detecting strange attractors in turbulence. In *Dynamical Systems and Turbulence*, Vol. 989 D. Rand and L. Young, editors. Springer-Verlag, New York.
- Kullback, S., and R. Leibler. 1951. On information and sufficiency. *Ann. Math. Stat.* 22:79–86.
- Kullback, S. 1959. *Information Theory and Statistics*. John Wiley and Sons, New York.
- Kullback, S. 1987. The Kullback-Leibler distance. *Am. Stat.* 41:340–341.
- Brooks, B., R. Bruccoleri, B. Olafson, D. States, S. Swaminathan, et al. 1983. CHARMM: a program for macromolecular energy minimization and dynamics calculations. *J. Comput. Chem.* 4:187–217.
- Shannon, C., and W. Weaver. 1949. *The Mathematical Theory of Information*. University of Illinois Press, Urbana, IL.
- Grassberger, P. 1988. Finite sample corrections to entropy and dimension estimates. *Phys. Lett. A.* 128:369–373.
- Lehrman, M., A. Rechester, and R. White. 1997. Symbolic analysis of chaotic signals and turbulent fluctuations. *Phys. Rev. Lett.* 78:54–57.
- Rechester, A., and R. White. 1991. Symbolic kinetic equations for a chaotic attractor. *Phys. Lett. A.* 156:419–424.
- Rechester, A., and R. White. 1991. Symbolic kinetic analysis of two-dimensional maps. *Phys. Lett. A.* 158:51–56.
- Bandt, C., and B. Pompe. 2002. Permutation entropy: a natural complexity measure for time series. *Phys. Rev. Lett.* 88:174102.
- Kennel, M., R. Brown, and H. Abarbanel. 1992. Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Phys. Rev. A.* 45:3403–3411.



48. Abarbanel, H., and M. Kennel. 1993. Local false nearest neighbors and dynamical dimensions from observed chaotic data. *Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics.* 47:3057–3068.
49. Cellucci, C., A. Albano, and P. Rapp. 2003. Comparative study of embedding methods. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 67:066210–066213.
50. Noakes, L. 1991. The Takens embedding theorem. *Int. J. Bifurc. Chaos Appl. Sci. Eng.* 1:867–872.
51. Sauer, T., J. Yorke, and M. Casdagli. 1991. Embedology. *J. Stat. Phys.* 65:579–616.
52. Abarbanel, H. 1996. *Analysis of Observed Chaotic Data.* Springer, New York.
53. Fraser, A. 1989. Reconstructing attractors from scalar time series: a comparison of singular system and redundancy criteria. *Physica D.* 34:391–404.
54. Bonanno, C., and M. Mega. 2004. Toward a dynamical model for prime numbers. *Chaos Solitons Fractals.* 20:107–118.
55. Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD—visual molecular dynamics. *J. Mol. Graph.* 14:33–38.