SUPPORT INFORMATION

Quantification of Drive-Response Relationships Between Residues During Protein Folding

Yifei Qi and Wonpil Im
Department of Molecular Biosciences and Center for Bioinformatics, The University of Kansas, 2030 Becker Drive Lawrence, Kansas 66047, United States

S1. Transfer entropy (TE) in three numerical systems

Our implementation of TE calculation was verified with the following three numerical systems.[1]

#1. A unidirectional liner process

$$x_i = 0.8x_{i-1} + n_i^{(x)}(\sigma_x^2) + ey_{i-1}$$
$$y_i = 0.4y_{i-1} + n_i^{(y)}(\sigma_y^2)$$

where $e \in [0,1]$ is the coupling strength; $n_i^{(x)}$ and $n_i^{(y)}$ are independent Gaussian random process with zero mean and variance of $\sigma_x^2 = \sigma_y^2 = 0.2$. The system was simulated for 2×10^4 steps and only the last 10^4 steps were used for calculation. By construction, y_i drives x_i, and the calculated $D_{y\to x}$ is shown in Figure S1A.

#2. A unidirectional Henon map

$$x_i = a - x_{i-1}^2 + b_x x_{i-2}$$
$$y_i = a - \{ex_{i-1} + (1-e)y_{i-1}\}y_{i-1} + b_y y_{i-2}$$

where $e \in [0,1]$ is the coupling strength; coefficients a, b_x, and b_y were set to 1.4, 0.3 and 0.3. We simulated 2×10^5 steps and only kept the last 10^5 steps for analysis. By construction, x_i drives y_i, and the calculated $D_{y\to x}$ is shown in Figure S1B.

#3. A unidirectional Ulam map lattice

$$x_i^{(l)} = f(ex_{i-1}^{(l-1)} + (1-e)x_{i-1}^{(l)})$$
$$f(x) = 2 - x^2, l = 1,....,L$$

where $e \in [0,1]$ is the coupling strength, and L, the number of maps in the lattice, was fixed to 100. The last map is connected to the first one using the periodic boundary condition. We simulated 2×10^5 steps and used the last 10^5 steps to analyze the interactions from the second map $x_i^{(2)}$ to the first one $x_i^{(1)}$. By construction, $x_i^{(1)}$ drives $x_i^{(2)}$, and the calculated $D_{x_i^{(2)}\to x_i^{(1)}}$ is shown in Figure S1C.

For all TE calculations, the embedding dimension was set to 1 and the time series were symbolized to 4 symbols using even partition. In all the numerical systems, TE captured the drive/response relationship successfully (**Figure S1**). In the linear process (#1), the driving interaction increases as the coupling strength increases. In the Henon map (#2), the driving interaction drops quickly after e = 0.7, due to the synchronization of the two maps. In the Ulam map lattice (#3), the lattice undergoes two bifurcations where the driving interaction from $x_i^{(1)}$ to $x_i^{(2)}$ disappears.[2]

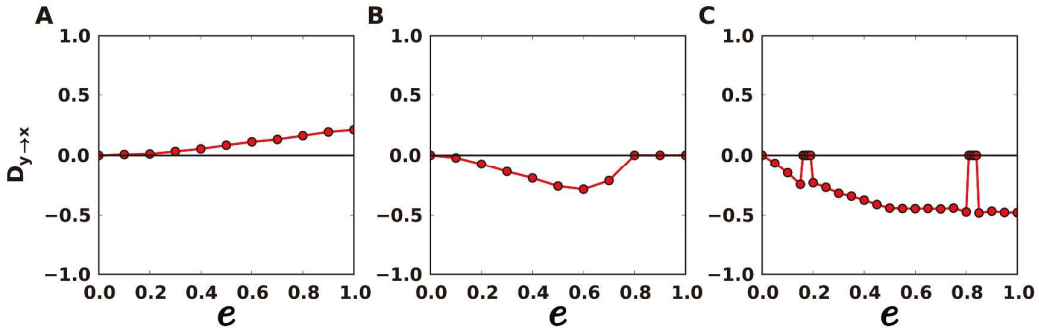

Figure S1. TE from y_i to x_i for (A) unidirectional liner process and (B) Henon map, from $x_i^{(2)}$ to $x_i^{(1)}$ for (C) unidirectional Ulam map lattice. A positive $D_{y\to x}$ value indicates a drive interaction, while a negative value indicates a response interaction.

S2. Driving and responding residues and MI for BBL, Villin and BBA

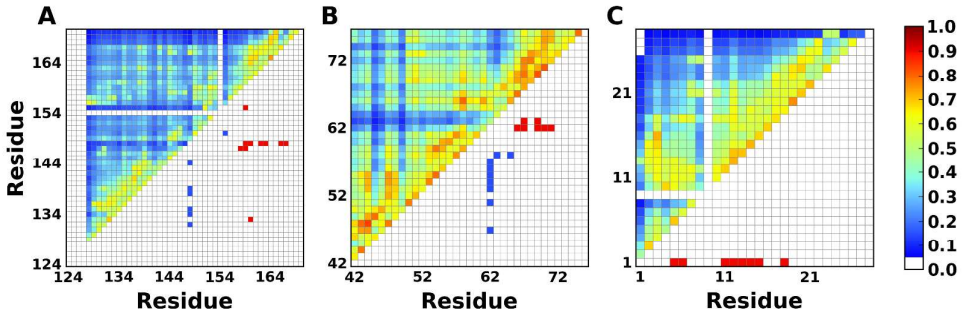

Figure S2. Driving and responding residues and MI of residue pairs in (A) BBL, (B) Villin and (C) BBA. In the lower-half triangle, red indicates that a residue from x-axis drives a residue from y-axis, blue indicates that a residue from x-axis responds to a residue from y-axis. MI was calculated from the whole trajectory (upper-half triangle). Values are colored from blue to red. MI values smaller than 0.05 are colored white.

S3. Time-delayed mutual information (TDMI) in three numerical systems and folding trajectories

Introducing a time delay to equation (6) in the main text, we get the normalize TDMI:

$$MI'(x-t,y)=\sqrt{1-e^{-2MI(x-t,y)}}\in[0,1]$$

$$MI'(x,y-t)=\sqrt{1-e^{-2MI(x,y-t)}}\in[0,1]$$

Where $MI()$ is the commonly used mutual information as in equation (5) in the main text. Similar to TE, we calculated the difference of TDMI between two directions,

$$TDMI_{y\to x}=MI'(x-t,y)-MI'(x,y-t)\in[-1,1]$$

We used the same partition method as TE calculation and the time delay t was set to 1. In all the three numerical systems in Section S1, TDMI yielded very similar results to TE (**Figure S3**). However, when we applied TDMI to residues in the folding trajectories, it failed to give any information on the drive/response relationship, i.e., the information from both direction were too close and they canceled each other (**Figure S4**).

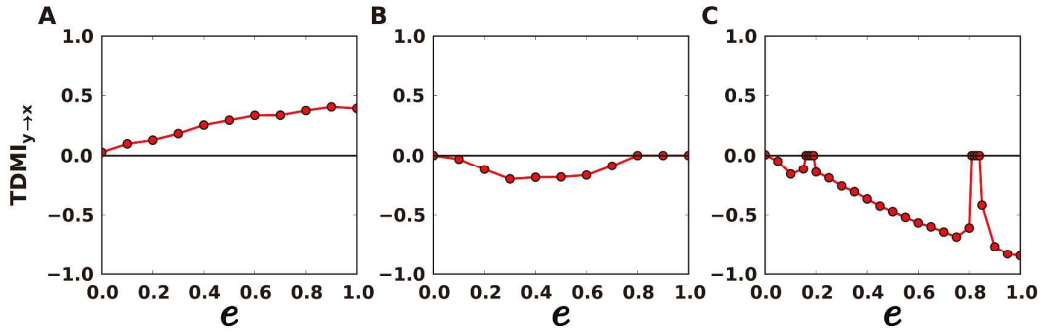

Figure S3. TDMI from y_i to x_i for (A) unidirectional liner process and (B) Henon map, from $x_i^{(2)}$ to $x_i^{(1)}$ for (C) unidirectional Ulam map lattice.

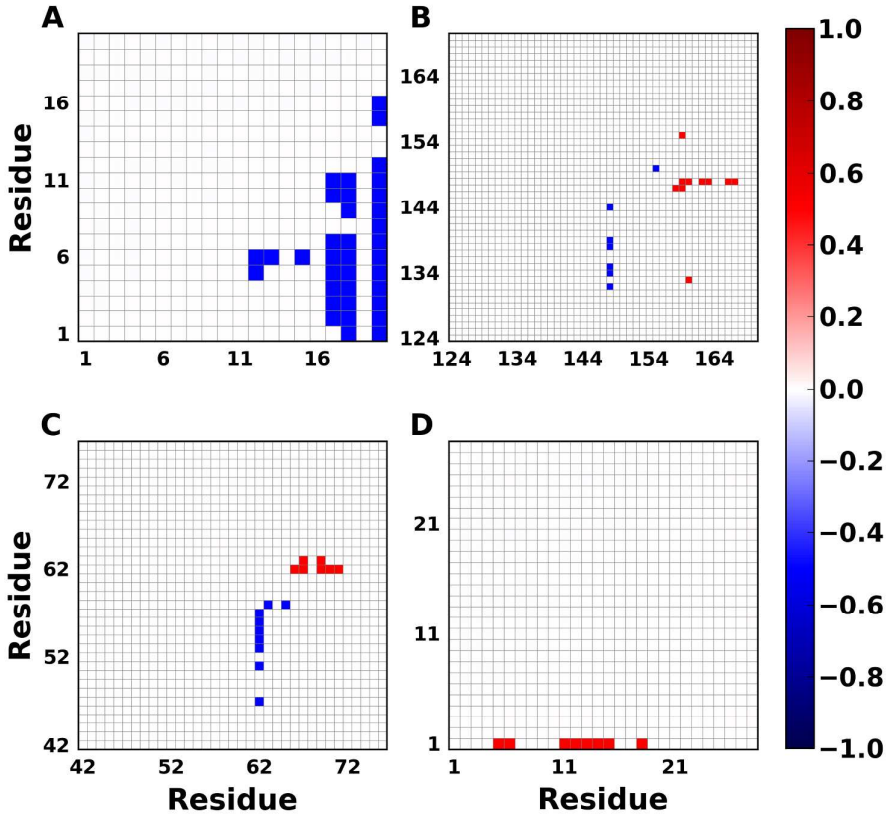

Figure S4. Driving and responding residues and TDMI of residue pairs in (A) Trp-cage, (B) BBL, (C) Villin and (D) BBA. In the lower-half triangle, red indicates that a residue from x-axis drives a residue from y-axis, blue indicates that a residue

from x-axis responds to a residue from y-axis. TDMI is colored from blue (-1) to red (1), but the values are too close to 0 and are all white (upper-half triangle).

S4. Coarse-grained simulation of protein L

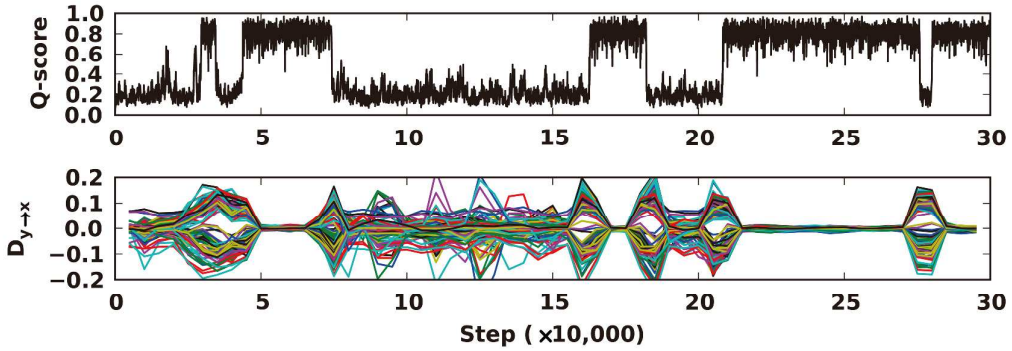

Figure S5. Q-score and $D_{y\to x}$ profiles of protein L from a coarse-grained simulation.

References

(1) Lungarella, M.; Ishiguro, K.; Kuniyoshi, Y.; Otsu, N., Methods for quantifying the causal structure of bivariate time series. *Int. J. of Bifurcation and Chaos* **2007,** *17*, 903-921.

(2) Schreiber, T., Measuring information transfer. *Phys. Rev. Lett.* **2000,** *85*, 461-464.