Supporting Information

A. The procedure to determine the optimal length $L_{\mathbf{past}}$

In this paper, we employed the equal probability partition, in which each symbol has the same occurrence, to symbolize the observable $\boldsymbol{\tau}$ and its decomposed components $\mathbf{A}^{(j)}$ and $\mathbf{D}^{(j)}$. This aims at obtaining good statistics with the finite length of the time series. The state constructed from CM is originally termed as *causal* state as its definition relies on the transition from a finite *history* of the observable to a future one. In this paper, the so-called causal state splitting reconstruction algorithm (1) was employed to construct the SSN in which $L_{\text{future}} = 1$ and therefore a transition from one state to another corresponds to a one step shift in the symbolic sequence. On the other hand, the number of partitions and the length of the past sub-sequence L_{past} were chosen so that the topological and topographical features of the inferred SSN do not change significantly. The topological and topographical features of the inferred SSN can be quantified by the so-called topological complexity C_{top} and statistical complexity C_μ in the framework of information theory defined by

$$C_{\text{top}} = \log_2 N_{\text{S}}, \tag{1}$$

and

$$C_\mu = -\sum_{i=1}^{N_{\text{S}}} P(S_i)\ \log_2 P(S_i), \tag{2}$$

where N_{S} and $P(S_i)$ denote the number of the states and the resident probability of the state S_i, respectively (2). The statistical complexity of the network is interpreted as the average amount of information (in bits) in the past, i.e. memory content, to predict the future. In the present article, the statistical complexity measure C_μ was used to examine the convergence of the topographical feature of the SSN as the past sub-sequences length L_{past} increases from zero.

For example, suppose that one can identify the conformational state in which the protein resides, namely, either the unfolded (U) or native (N) state, at each sampling time in a single-molecule experiment of protein folding-unfolding process at the folding temperature. The time series looks like "...UUNNUNU...". How can one assign the correct L_{past} and construct the underlying SSN? For the sake of brevity, we assume that the experimental

sampling rate are slow enough so that the switching between U and N in the recorded time series is Markovian, i.e., $P(s_{i+1}|s_i s_{i-1} s_{i-2} \dots) = P(s_{i+1}|s_i)$ where $s =$ U or N and the subscripts label the time t_i.

Let us start with a zero length of past sub-sequences, having $P(\mathrm{U}|\mathrm{null}) = P(\mathrm{U}) = 0.5$ and $P(\mathrm{N}|\mathrm{null}) = P(\mathrm{N}) = 0.5$. This leads us to assign only a single state containing the null sequence as shown in Fig. 6(A) with statistical complexity equal to zero. Next one proceeds to examine the case of past sub-sequences of length one. Suppose that we obtain $P(\mathrm{U}|\mathrm{U}) = 0.9, P(\mathrm{N}|\mathrm{U}) = 0.1$ and $P(\mathrm{U}|\mathrm{N}) = 0.1, P(\mathrm{N}|\mathrm{N}) = 0.9$ with the given sampling rate. Since the transition probabilities to the future are different, the past sub-sequences U and N are not grouped together into a same state and one ends up with two states as shown in Fig. 6(B) with statistical complexity larger than zero. This implies that $L_{\mathrm{past}} = 0$ is not appropriate to define the states because the topographical feature of the SSN does not converge. Note that, because the time series s_i is Markovian, one should expect that the structure of the SSN remains the same as Fig. 6(B) for all past sub-sequences length $L_{\mathrm{past}} \geq 2$. One can easily check if this is the case, for example, with $L_{\mathrm{past}} = 2$ as follows: the possible candidates of states made from $s_i s_{i-1}$ are UU, NU, UN, and NN. The transition probabilities $P(s_{i+1}|s_i s_{i-1})$ are $P(\mathrm{U}|\mathrm{UU}) = 0.9$, $P(\mathrm{N}|\mathrm{UU}) = 0.1$, $P(\mathrm{U}|\mathrm{NU}) = 0.1$, $P(\mathrm{N}|\mathrm{NU}) = 0.9$, $P(\mathrm{U}|\mathrm{UN}) = 0.9$, $P(\mathrm{N}|\mathrm{UN}) = 0.1$, $P(\mathrm{U}|\mathrm{NN}) = 0.1$, and $P(\mathrm{N}|\mathrm{NN}) = 0.9$. It is apparent that one can group UU and NU, and separately UN and NN, together into a same state as shown in Fig. 6(C). One can easily verify that this holds for all $L_{\mathrm{past}} \geq 1$. As a consequence, the optimal L_{past} in this case can be chosen to be one since the statistical complexity of the SSN does not change as the length of past sub-sequence increases from one.

Note here that the structure of the SSN depends on the choice of the sampling time in the observation. For instance, if the sampling time scale is short enough to result in non-Markovian dynamics in the folding-unfolding process, the length L_{past} should be longer than one in defining the states. On the contrary, one can expect that the two states U and N should merge as one state (just as Fig. 6(A)) when the sampling time is extremely longer than the characteristic time scale of the folding-unfolding. It is because each transition probability $P(s_{i+1}|s_i)$ with $L_{\mathrm{past}} = 1$ just reduces to the resident probability $P(s_{i+1}) = P(s_{i+1}|\mathrm{null})$. That is, the next state should be statistically independent from the previous state the system visited very long time ago.

It has been proved that the SSN with the convergence of statistical complexity is not only a hidden Markovian model but also a minimal 'optimally-predictive model' capable of statistically reproducing the original time series based on information-theoretic considerations (2). It is minimal because the complexity of network structure quantified by the statistical complexity C_μ is minimum only when states are constructed according to the CM scheme. Moreover, the optimity is based on the fact that the conditional entropy $H(\text{future}|\text{past})$ which measures the uncertainty in predicting the future once we know the past, is also minimized if and only if the past sub-sequences with same future transition probabilities are grouped to the same state as described by CM (2).

Fig. 7 exemplifies the statistical complexity as a function of L_{past} for the time series of $\mathbf{A}^{(3)}$ and $\mathbf{D}^{(j)}$ $(j = 1, 2, 3)$ of the delay-time time series $\boldsymbol{\tau}$ of ET single-molecule measurement for Fre/FAD complex. The figure shows that the network structures for the SSN of $\mathbf{D}^{(2)}$ and $\mathbf{D}^{(3)}$ converge at $L_{\text{past}} = 1$, while an apparent convergence appears at $L_{\text{past}} = 2$ for $\mathbf{A}^{(3)}$ (the C_μ of $\mathbf{A}^{(3)}$ increases again for larger L_{past} that is not shown in the figure). This shows that the $\mathbf{D}^{(2)}$ and $\mathbf{D}^{(3)}$ time series are Markovian, while the $\mathbf{A}^{(3)}$ time series is non-Markovian such that the future values depend on the past sub-sequences with length longer than one.

B. Dependence of the localization properties of a state on the transition timescales

For deterministic systems, with transition time steps much smaller than the Lyapunov time (the timescale for nearby trajectories of the system to diverge), closeness in transition probability should give rise to states with subsequences localized in position since the trajectories do not have time to separate yet. However, for transition time steps equal or longer than the Lyapunov time, two trajectories which are close in position may end up in two very different regions after the transition or two distinct trajectories may merge into a nearby region after the transition. Therefore, one can then expect that subsequences of a state with closeness of transition probabilities may have very different positions when the transition time is long compared with the Lyapunov time.

One can also carry out a similar argument for stochastic systems when comparing the transition time with the correlation time (the timescale for two trajectories to be correlated). As shown in the case of Langevin dynamics, states constructed with a short

transition time compared with the correlation time are localized in position (e.g., the case of $m = 1/\lambda$ in Fig. 7(B) in the main text), while states constructed with a long transition time are not localized (e.g. the single state in Fig. 7(B) with $m = 5/\lambda$ contains subsequences with all positions.)

In the case of complex system (e.g. protein), however, instead of a single correlation time as in the simple Langevin case, a broad range of correlation timescales may exist in different regions of the highly inhomogeneous state space. When a certain transition timescale is chosen, there may exist some regions with long correlation time which still give rise to states with subsequences localized in position, and there are some other regions with short correlation time where the localization of the subsequences is lost. Therefore, a study of the localization properties of the states may provide us with an atlas of timescales on the underlying state space.

C. Haar wavelet and down-sampling problem

The approximation and details of the Haar wavelet have simple statistical interpretations (3). For the Haar wavelet $\mathbf{A}^{(j)}$ and $\mathbf{D}^{(j)}$ of the time series τ_i are given by

$$\begin{aligned} A_i^{(j)} &= \left(\sum_{k=i}^{i+2^j-1} \tau_k \right) / 2^j, \\ D_i^{(j)} &= \left(\sum_{k=i}^{i+2^{j-1}-1} \tau_k - \sum_{k=i+2^{j-1}}^{i+2^j-1} \tau_k \right) / 2^j, \quad j \geq 1. \end{aligned} \tag{3}$$

Therefore, $A_i^{(j)}$ and D_i^j are simply the mean and the average fluctuation over a bin of 2^j time steps, respectively. Note that $A_i^{(j)}$ and $A_{i'}^{(j)}$ (or $D_i^{(j)}$ and $D_{i'}^{(j)}$) with $|i - i'| \leq 2^j$ are unphysically correlated since some common data points are used in evaluating the two $A^{(j)}$'s (or $D^{(j)}$'s). In fact, only $N/2^j$ points, e.g. $(A_1^{(j)}, A_{1+2^j}^{(j)}, \cdots, A_{1+n(2^j)}^{(j)}, \cdots)$, in the j-level should be taken to avoid artificial correlations. This down-sampling problem, which may lead to poor statistics in constructing the multiscale SSN for process with long-term memory effect (i.e., large j), can be improved by treating the set .fd .fd $\{(A_i^{(j)}, A_{i+2^j}^{(j)}, \cdots, A_{i+n(2^j)}^{(j)}, \cdots), i = 1, \cdots, 2^j\}$ (and similarly for the details) as an ensemble of 2^j time series (each with $N/2^j$ data points) from which the SSN is built.

The stationarity of the approximation and details can be inspected by evaluating their autocorrelations. The autocorrelations of the $\mathbf{D}^{(j)}$ and $\mathbf{A}^{(j)}$ are presented in Fig. 8. One

can see that the autocorrelation of $\mathbf{D}^{(j)}$ decays rapidly on a timescale of 2^j time steps with small oscillations for longer time, while those of $\mathbf{A}^{(j)}$ remains approximately constant for 2^j steps and shows similarly behavior as those of $\boldsymbol{\tau}$ for timescale longer than 2^j. These indicate that $\mathbf{D}^{(j)}$'s are approximately stationary with timescale of 2^j and $\mathbf{A}^{(j)}$ capture all the nonstationarity of $\boldsymbol{\tau}$ with timescale longer than 2^j.

D. Relation between the delay-time probability density and the lifetime spectrum (Eq. 4 in the main text)

The conditional probability density for finding the delay-time τ given the lifetime γ^{-1} is,

$$P(\tau|\gamma^{-1}) = (\gamma^{-1})^{-1} e^{-\tau/\gamma^{-1}} = \gamma e^{-\gamma\tau}. \tag{4}$$

Let $\alpha(\gamma^{-1})$ be the spectrum of lifetime, then the normalized probability density of finding a particular lifetime γ^{-1} is given by

$$P(\gamma^{-1}) = \frac{\alpha(\gamma^{-1})}{\int d\gamma^{-1}\alpha(\gamma^{-1})}. \tag{5}$$

Therefore, to calculate the probability density of finding τ regardless of the lifetime γ^{-1}, one integrates out the lifetime from the joint probability, i.e.,

$$\begin{aligned} P(\tau) &= \int d\gamma^{-1} P(\tau, \gamma^{-1}) = \int d\gamma^{-1} P(\tau|\gamma^{-1}) P(\gamma^{-1}) \\ &= \int d\gamma^{-1} \big[(\gamma^{-1})^{-1} e^{-\tau/\gamma^{-1}}\big] \frac{\alpha(\gamma^{-1})}{\int d\gamma^{-1}\alpha(\gamma^{-1})} \\ &= \frac{\int d\gamma^{-1}\alpha(\gamma^{-1})(\gamma^{-1})^{-1} e^{-\tau/\gamma^{-1}}}{\int d\gamma^{-1}\alpha(\gamma^{-1})}, \end{aligned} \tag{6}$$

where Eq. 4 and 5 have been used in the second line.

E. The derivation of Eq. 7 in the main text

First note that the correlation function can be expressed in terms of the joint probability

$$\begin{aligned} C(t) &= \left\langle \delta\gamma^{-1}(t)\delta\gamma^{-1}(0)\right\rangle \\ &= \left\langle \gamma^{-1}(t)\gamma^{-1}(0)\right\rangle - \left\langle \gamma^{-1}\right\rangle^2 \\ &= \sum_{\gamma_0^{-1},\gamma_t^{-1}} \gamma_t^{-1} P(\gamma_t^{-1}, \gamma_0^{-1})\gamma_0^{-1} - \left(\sum_{\gamma^{-1}} \gamma^{-1} P(\gamma^{-1})\right)^2 \end{aligned} \tag{7}$$

where γ_0^{-1} and γ_t^{-1} are the lifetime at the current time and at t steps later, respectively. $\sum_{\gamma_0^{-1},\gamma_t^{-1}}$ means the summation over all possible pairs of γ_0^{-1} and γ_t^{-1}. $P(\gamma_t^{-1},\gamma_0^{-1})$ denotes the joint probability of γ_0^{-1} and γ_t^{-1}. As shown in the inset of Fig. 3(D) in the main text, each state S_I in the SSN has its own decay-time distribution $\sim \int d\gamma^{-1}\alpha_I(\gamma^{-1})\, e^{-\tau/\gamma^{-1}}$ with lifetime γ^{-1}. For a given SSN with timescale $t = 2^n$, the joint probability can be expressed as

$$P(\gamma_{2^n}^{-1},\gamma_0^{-1}) = \sum_{I,J} P\left(\gamma_{2^n}^{-1},\gamma_0^{-1},S_J,S_I\right), \tag{8}$$

where $P\left(\gamma_{2^n}^{-1},\gamma_0^{-1},S_J,S_I\right)$ is the joint probability of visiting the state S_I at the current time with lifetime γ_0^{-1} and visiting the state S_J at $t = 2^n$ time steps later with lifetime $\gamma_{2^n}^{-1}$. Therefore for $t = 2^n$ the first term on the right hand side of Eq. 7 becomes $\sum_{I,J}\sum_{\gamma_0^{-1},\gamma_{2^n}^{-1}} \gamma_{2^n}^{-1} P\left(\gamma_{2^n}^{-1},\gamma_0^{-1},S_J,S_I\right)\gamma_0^{-1}$.

Note here that, in terms of the chain rule of joint probability, the joint probability can be decomposed as

$$\begin{aligned} P\left(\gamma_{2^n}^{-1},\gamma_0^{-1},S_J,S_I\right) &= P\left(\gamma_{2^n}^{-1}|\gamma_0^{-1},S_J,S_I\right) \\ &\quad \times P(\gamma_0^{-1}|S_J,S_I)P_{2^n}(S_J,S_I). \end{aligned} \tag{9}$$

The current lifetime γ_0^{-1} does not depend on the future state S_J from causality consideration. Thus, we have $P(\gamma_0^{-1}|S_J,S_I) = P(\gamma_0^{-1}|S_I)$. If $\gamma_{2^n}^{-1}$ depends solely on the state S_J where the system resides at that time such that $P(\gamma_{2^n}^{-1}|\gamma_0^{-1},S_J,S_I) \approx P(\gamma_{2^n}^{-1}|S_J)$, the first term of Eq. 7 can be then estimated as

$$\begin{aligned} &\sum_{\gamma_0^{-1},\gamma_{2^n}^{-1}} \gamma_{2^n}^{-1} P(\gamma_{2^n}^{-1},\gamma_0^{-1})\gamma_0^{-1} \\ \approx &\sum_{I,J}\sum_{\gamma_0^{-1},\gamma_{2^n}^{-1}} \gamma_{2^n}^{-1} P(\gamma_{2^n}^{-1}|S_J)P(\gamma_0^{-1}|S_I)P_{2^n}(S_J,S_I)\gamma_0^{-1} \\ = &\sum_{I,J} \overline{\gamma_J^{-1}} P_{2n}(S_J,S_I)\overline{\gamma_I^{-1}}, \end{aligned} \tag{10}$$

where $P_{2^n}(S_J,S_I)$ is the joint probability of visiting S_I followed by S_J after 2^n steps, and

$$\overline{\gamma_J^{-1}} = \sum_{\gamma_{2^n}^{-1}} \gamma_{2^n}^{-1} P(\gamma_{2^n}^{-1}|S_J), \tag{11}$$

$$\overline{\gamma_I^{-1}} = \sum_{\gamma_0^{-1}} \gamma_0^{-1} P(\gamma_0^{-1}|S_I). \tag{12}$$

The average lifetime $\overline{\gamma_I^{-1}}$ for the given state S_I can be calculated using the corresponding delay-time probability density $P_I(\tau)$ (see Eq. 4 in the main text) as

$$\overline{\gamma_I^{-1}} = \frac{\int d\gamma^{-1}\gamma^{-1}\alpha_I(\gamma^{-1})}{\int d\gamma^{-1}\alpha_I(\gamma^{-1})} = \int d\tau\tau P_I(\tau) \tag{13}$$

The second term on the right hand side of Eq. 7 can also be evaluated as follows:

$$\begin{aligned}\left(\sum_{\gamma^{-1}}\gamma^{-1}P(\gamma^{-1})\right)^2 &= \left(\sum_I\sum_{\gamma^{-1}}\gamma^{-1}P(\gamma^{-1}|S_I)P(S_I)\right)^2 \\ &= \sum_I\sum_J\overline{\gamma_I^{-1}}\,\overline{\gamma_J^{-1}}P(S_I)P(S_J).\end{aligned} \tag{14}$$

By combining Eqs. 10 and 14, one can obtain Eq. 6 in the main text.

One can also generalize the above procedure to evaluate the multi-time correlation functions. For example, one can estimate the three-time correlation function $\langle\gamma^{-1}(2t)\gamma^{-1}(t)\gamma^{-1}(0)\rangle$ as $\sum_{I,J,K}\overline{\gamma_K^{-1}}\,\overline{\gamma_J^{-1}}\,\overline{\gamma_I^{-1}}P(S_K,S_J,S_I)$,where $P(S_K,S_J,S_I)$ is the joint probability of visiting the states S_I, S_J, and S_K at the current time, t steps later, and $2t$ steps later.

F. The degree-k dependence on the stability of states in the multiscale SSNs

By means of computer simulations, Rao and Caflish (4) and Gfeller et al. (5) revealed the conformational space network for simulated folding of beta3s (20-residue antiparallel β-sheet peptide). The conformational space network (CSN) are composed of nodes and links between them: the nodes represent a set of snapshots recorded along the trajectory grouped according to secondary structure and the links are the transitions between them. It was found that the CSN of beta3s exhibits scale-free characteristics (6,7) (power-law behavior of the degree distribution), similar to World-Wide Web, protein interaction network, and metabolic network (8). Furthermore, the denatured state ensemble on the CSN is very heterogeneous—including low-entropy, low-enthalpy traps as well as high-entropy, high-enthalpy conformations. It was also found that the more the number of links of a node in the CSN, the more the node tends to be populated.

Fig. 9 presents the normalized degrees of the I-th state S_I, $k_I/k_{\max}$, and the stability elucidated by $\log P_I$ of the multiscale SSNs at three different timescale of 2^4, 2^6, and 2^8 steps (where P_I is the resident probability of the I-th state). The quantity $-\log P_I$ can

be interpreted as the relative free energy of the I-th state if the concept of free energy landscape is validated at the chosen timescale. One can see the existence of positive correlation between the degrees and the stability in the network. That is, the state tends to be more stabilized when there exist more transition paths from the state. However for the SSNs at 2^4 and 2^6 time steps, two distinct slopes exist in the region of $\log P_I \gtrsim -10$. This suggests that the stable states belonging to this region can be classified into the two classes, namely, either with many transition paths (larger $k_I/k_{\max}$) or with relatively fewer transition paths (smaller $k_I/k_{\max}$). The number of transition paths might reflect the entropic stabilization of the state. If so, such stable states associated with more (fewer) paths of transitions might be interpreted as high-entropy, high-enthalpy states (low-entropy, low-enthalpy traps) (4). A further investigation to support this conjecture is required by analyzing time series based on computer simulations of small proteins. At the timescale of 2^8 time steps, the correlation between $k_I/k_{\max}$ and $\log P_I$ becomes slightly 'diffuse.' In particular, the degree becomes 'saturated' for the stable states with $\log P_I \gtrsim -9$ and the measure of the (normalized) degree can no longer differentiate the diversity in transitions for such stable states. The measure of degree is simply based on the topology of the network. In a forthcoming article, we will present a new measure based on transition probability in the SSNs, which can further differentiate the diversity in transitions on the complex network even after the degree is saturated.

1. Shalizi, C. R. (2003) *http://www.cscs.umich.edu/ crshalizi/CSSR/.*

2. Shalizi, C. R & Crutchfield, J. P. (2001) *J. Stat. Phys.* **104**, 819–881.

3. Masuda, N & Aihara, K. (2001) *Electronics and Communications in Japan* **84**, 50.

4. Rao, F & Caflisch, A. (2004) *J. Mol. Biol.* **342**, 299–306.

5. Gfeller, D, Rios, P. D. L, Caflisch, A, & Rao, F. (2007) *Proc. Natl. Acad. Sci. USA* **104**, 1817–1822.

6. Barabási, A.-L & Albert, R. (1999) *Science* **286**, 509–512.

7. Albert, R & Barabási, A.-L. (2002) *Rev. Mod. Phys.* **74**, 47–97.

8. Gallos, L. K, Song, C, Havlin, S, & Makse, H. A. (2007) *Proc. Natl. Acad. Sci. USA* **104**, 7746–7751.

FIG. 6: The SSN for a simple Markovian protein folding-unfolding example with past sub-sequence length equal to (A) zero, (B) one and (C) two. The past sub-sequences that have the same transition probabilities are grouped as the same state, represented by circles. The directed links between states are labeled with the resulting conformation and with its transition probability shown inside the parenthesis.

FIG. 7: The statistical complexity as a function of L_{past} for the time series of $\mathbf{A}^{(3)}$ and $\mathbf{D}^{(j)}$ $(j = 1, 2, 3)$ of the delay-time time series $\boldsymbol{\tau}$ of ET single-molecule measurement for Fre/FAD complex. This analysis clearly shows that $\mathbf{D}^{(2)}$ and $\mathbf{D}^{(3)}$ are Markovian whereas $\mathbf{A}^{(3)}$ is non-Markovian.

FIG. 8: Autocorrelation functions of $\mathbf{A}^{(3)}$ and $\mathbf{D}^{(j)}$ $(j = 1, 2, 3)$ of the delay-time time series $\boldsymbol{\tau}$ of ET single-molecule measurement for Fre/FAD complex. The results indicate that $\mathbf{A}^{(j)}$ capture all the nonstationarity of $\boldsymbol{\tau}$ with timescale longer than 2^j.

FIG. 9: $k_I/k_{\max}$ vs $\log P_I$ for each multiscale SSNs, $\varepsilon^{A^{(4)},D^{(4)},D^{(3)}}$, $\varepsilon^{A^{(6)},D^{(6)},D^{(5)}}$ and $\varepsilon^{A^{(8)},D^{(8)},D^{(7)}}$ for the protein conformational fluctuation of the Fre/FAD complex. Notice the divergence into two classes of $k_I/k_{\max}$ for $\log P_I \gtrsim -10$, and the consistency of such divergence for time steps less than 2^8.