

Supporting Information

1 Network Analysis of the Constructed LES/non-LES Network of the BLN Model Protein

We scrutinize the network properties of the state-space network constructed from a scalar time series of the BLN model protein.

Figure 1 shows a linear-log plot of the number (k) of edges emanating from a state (*i.e.*, degree of the state) and the probability (P) the system resides at the LES/non-LES at 0.3–2.0 ϵ/k_B . One can see the existence of significant correlation between the stability and the degree associated with each state for a wide range of temperature. At the high temperature regime of 0.7–2.0 ϵ/k_B above the collapse temperature, one can see that the more the number of edges increases, the more the resident probability also does monotonically and linearly in the logarithmic: $\log P \propto k$. This may be because the state can be entropically more stabilized as the increase of transitions from it. On the contrary, below the collapse temperature where it is known that the kinetics starts to deviate from a single exponential¹ and the dimensionality of the landscape may also start to increase again (see Sec. 4.1 in Text), the relationship between the stability of states and the degree becomes diverse, that is, the same stability does not necessarily imply the same number of edges. For example, at 0.4 ϵ/k_B , states having the resident probability $P \simeq 0.01$ can be split into two groups, one is of a small number of edges and the other a large number of edges. This may indicate the former group as enthalpically stabilized but the latter entropically stabilized.

Most of all studies on protein conformation networks have been performed by means of computer simulation. Our technique is expected to “bridge” protein conformation network studies in computer simulation and single molecule experiments, and to provide new insights on the topological and topographical feature of the weighted directed network for complex biological systems.

2 LES/non-LES networks over a range of temperature

How are the LES/non-LES networks at different temperatures related with each other? For example, as temperature is increased a pair of two LESs may merge into a single LES. Hence, we can expect a hierarchical organization of LES/non-LES as a function of temperature T . Here we present a new visualization scheme for constructing a unified picture to capture the relationship among different energy landscapes across a parameter such as temperature. This is regarded as an ex-

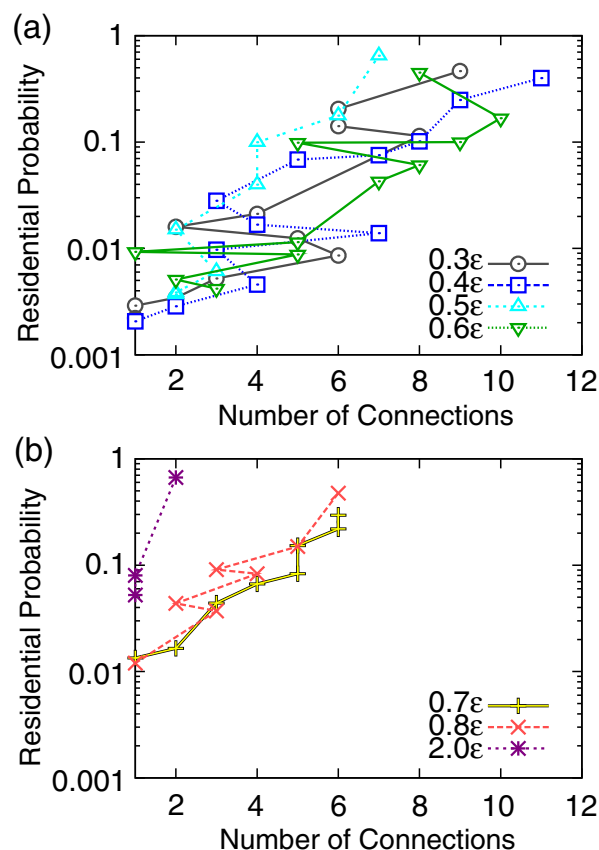


Fig. 1 The relationship between the resident probability P and the number of edges (degree k) of LES/non-LES obtained from the BLN model protein at a wide range of temperature. (a) At 0.3–0.6 ϵ/k_B , (b) At 0.7–2.0 ϵ/k_B . Note that $-k_B T \log P$ corresponds to the free energy of the state if the system actually experiences the free energy landscape. The colors of the lines denote the different temperatures: gray, blue, light blue, green, yellow, orange, and purple are 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, and 2.0 ϵ/k_B , respectively.

ension of the method proposed previously,² which explicitly involves adding a third type of information to the two dimensional space by using a multidimensional scaling (MDS) technique in the metric space.

Figure 2(a) shows two-dimensional projections of the LESs/non-LESs obtained from the time series of the model protein at different T . The projection is performed by principal coordinate analysis which preserves the mutual relationship in the full dimensional metric space among LESs/non-LESs as much as possible. One can see from the figure that the horizontal axis is almost proportional to the averaged end-to-end distance of the LES/non-LES (*i.e.*, the first-moment

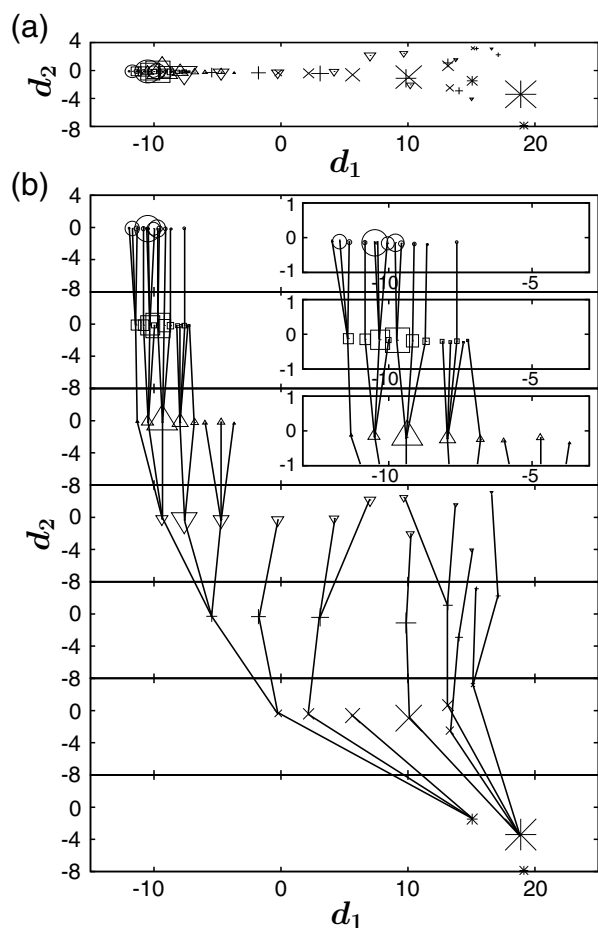


Fig. 2 The state space network across T of the model protein. **(a)** A two dimensional projection of the states distributed in the metric space. Here a multidimensional scaling (MDS) by a principal coordinate analysis was adopted. The horizontal and the vertical axes are the first and the second principal coordinates, respectively. **(b)** The relationship among the states at different T . The temperatures are 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, and $2.0 \text{ } \epsilon/k_B$ from the top to the bottom, respectively. Note that between 2.0 to $0.8 \text{ } \epsilon/k_B$, whose difference is very large, *ca.*, $1.2 \text{ } \epsilon/k_B$, the connections were assigned not by the direct relationship between the LES/non-LES at these two distant temperatures but by a set of several sequential assignments using the direct relationships at temperatures in between 2.0 and $0.8 \text{ } \epsilon/k_B$.

of the LES/non-LES distribution). The vertical axis approximately corresponds to the difference in variance (the second-

moment) of the distribution functions of the LES/non-LES. The LESs/non-LESs corresponding to the collapsed state are localized in the leftmost horizontal axis at $T < 0.5\epsilon/k_B$ while those corresponding to the denatured state are more scattered and shifted to the right at $T > 0.8\epsilon/k_B$. In the vertical axis, the variance of the data is smaller in the compact state than in the denatured state. This may imply the fact that the compact state is much more localized than the denatured state and, thus, the difference in the second and higher moments tends to be less significant even if there exist differences in the moments in the compact state.

Fig. 2(b) shows how the LES/non-LES observed at different temperatures are related with each other, by transforming Fig. 2(a) to a tree diagram with respect to T . In this sense, we can explicitly include a third information into the two dimensional representation. Each node represents the LES/non-LES in the metric space whose area is proportional to the residence probability at each T . A line connects from each node at a particular T to the closest node at the adjacent higher T (with respect to the proximity in shape of the probability density functions).

The biggest non-LES at $2.0 \text{ } \epsilon/k_B$ becomes split into one large state and three small states as T decreases to $0.8 \text{ } \epsilon/k_B$. The centroids of a set of the LESs/non-LESs at $0.8 \text{ } \epsilon/k_B$ migrate from larger values to smaller values in the horizontal axis, manifesting the existence of the collapse temperature between them. One can see in the insets of the figures that there exists a change or shift of the most stable superbasin between 0.5 to $0.3 \text{ } \epsilon/k_B$, as reported previously.² This might reflect the existence of the folding temperature although the statistics may not be so reliable for such a low T region.

It was found that the previous visualization scheme based on a nonlinear MDS method² sometimes fails to construct a meaningful tree diagram because the nonlinear MDS method depends on the choice of an initial guess in the procedure. The present scheme is based on the linear algebra of principal coordinate analysis that produces a unique diagram to reveal the hierarchical organization of the state space network over the third information (temperature in this work).

References

- 1 H. Nymeyer, A. E. Garcia and J. N. Onuchic, *Proc. Natl. Acad. Sci. USA*, 1998, **95**, 5921–5928.
- 2 A. Baba and T. Komatsuzaki, *Proc. Natl. Acad. Sci. USA*, 2007, **104**, 19297–19302.