A. The detailed procedure to construct local equilibrium state (LES) from scalar time series

Here we present the detailed procedure of constructing local equilibrium state (LES) from which one can evaluate its effective free energy landscape.

1. Short-time probability distribution function

Figure 1 illustrates how the short-time distribution function is constructed for a fragment of time series. Suppose that we have ten points recorded every an equal interval Δt_S as shown in Figure 1 (a). In the figure, each cross symbol denotes the observed value s_i at each time step t_i. *How can one construct the local distribution function from such a fragment of time series s_i ?* The discrete time series one obtains in actual experiments can be interpreted as a series of the observable s averaged over each time window Δt_S (which corresponds to the resolution in the experiment). We transform $\{s_i\}$ to the "continuous" time series $s(t)$ by setting that $s(t)$ is constant within each Δt_S (as indicated by the red line in Figure 1 (a)). As depicted in Figure 1 in the main text for the sake of brevity, one may usually plot the histogram with respect to s by defining a certain bin size. However, the histogram crucially depends on the chosen bin size. In our actual procedure we define the local distribution function without determining the bin size: we first construct the "probability density function" $p(s)$ from $s(t)$ as shown in Figure 1 (b). Rigorously speaking, this is called *probability mass function* (*pmf*) that gives the probability of finding the *discrete* variable exactly equal to a certain value of s. The $p(s)$ is defined for all s including the cases that s could never take by assigning such values a probability of zero. In short, $p(s)$ corresponds to the histogram with an "infinitesimally" small bin size.

We define the short-time distribution function $P(s)$ centered at time t with a range of $(t-\tau/2, t+\tau/2]$ by *probability distribution function* (*pdf*) with respect to s given by

$$P(s) = \int_{s_{\min}}^{s} p(s')\ \delta(s')\ ds', \quad \text{with} \quad s_{\min} \le s \le s_{\max} \tag{1}$$

where $\int_{s_{\min}}^{s_{\max}} P(s) ds = 1$ and $s_{\min}$ and $s_{\max}$, respectively, mean the minimum and maximum values of s observed in $(t-\tau/2, t+\tau/2]$ (see Figure 1 (c)). The reason why Dirac delta function δ appears is that we simply ignored shot noise and any broadening effects not dependent on the interdye distance. Here note that Eq. 1 does not require to specify any bin size in defining the short-time pdf. Moreover, the time window Δt_S is not necessarily the same along the time series and one can straightforwardly generalize the short-time pdf for any variable time window.

Here we describe the mathematical representation of the pmf with a constant time resolution Δt_S,

$$p(s) = (1/n_S) \sum_{i=1}^{n_S} \begin{cases} 1 & \text{if} \quad s = s(t_i), \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

Here n_S denotes the total number of data points, with which one constructs a pmf. The probability density function $f(s)$ corresponding to $p(s)$ is given by

$$f(s) = (1/n_S) \sum_{i=1}^{n_S} \delta(s - s(t_i)). \tag{3}$$

Here the relationship $f(s) = p(s)\ \delta(s)$ holds under the assumption of ignoring any broadening effects. Eq. 3 corresponds to Eq. 6 in the main text. The probability distribution function $P(s)$ is obtained by the integration of $f(s)$:

$$\begin{aligned} P(s) &= \int_{s_{min}}^{s} f(s')ds', \\ &= (1/n_S) \sum_{i=1}^{n_S} \Theta(s - s(t_i)), \end{aligned}$$

where Θ denotes the Heaviside function obtained by the integration of the δ function.

Figure 1 (d) illustrates the Kantorovich metric $d_K(p_i||p_j)$ between two pmfs p_i and p_j. In the figure, the red, and blue lines correspond to the short-time pdf defined at $t = 5$ and $\tau = 10$ (composed of all ten points), and that at $t = 1.5$ and $\tau = 3$ (of the first three points), respectively. The Kantorovich metric $d_K(p_i||p_j)$, that is, $\int_{-\infty}^{\infty} ds \left| \int_{-\infty}^{s} ds' \left(p_i(s') - p_j(s')\right) \delta(s') \right|$, corresponds to the sum of all shaded areas enclosed by these two pdfs.

2. The assignment procedure of local equilibrium state (LES) candidates

As described in the main text, *conceptually*, one could partition $\{p_i\}$ into a union of 'clusters (subsets),' the candidates of LES, in the Kantorovich metric space by computing

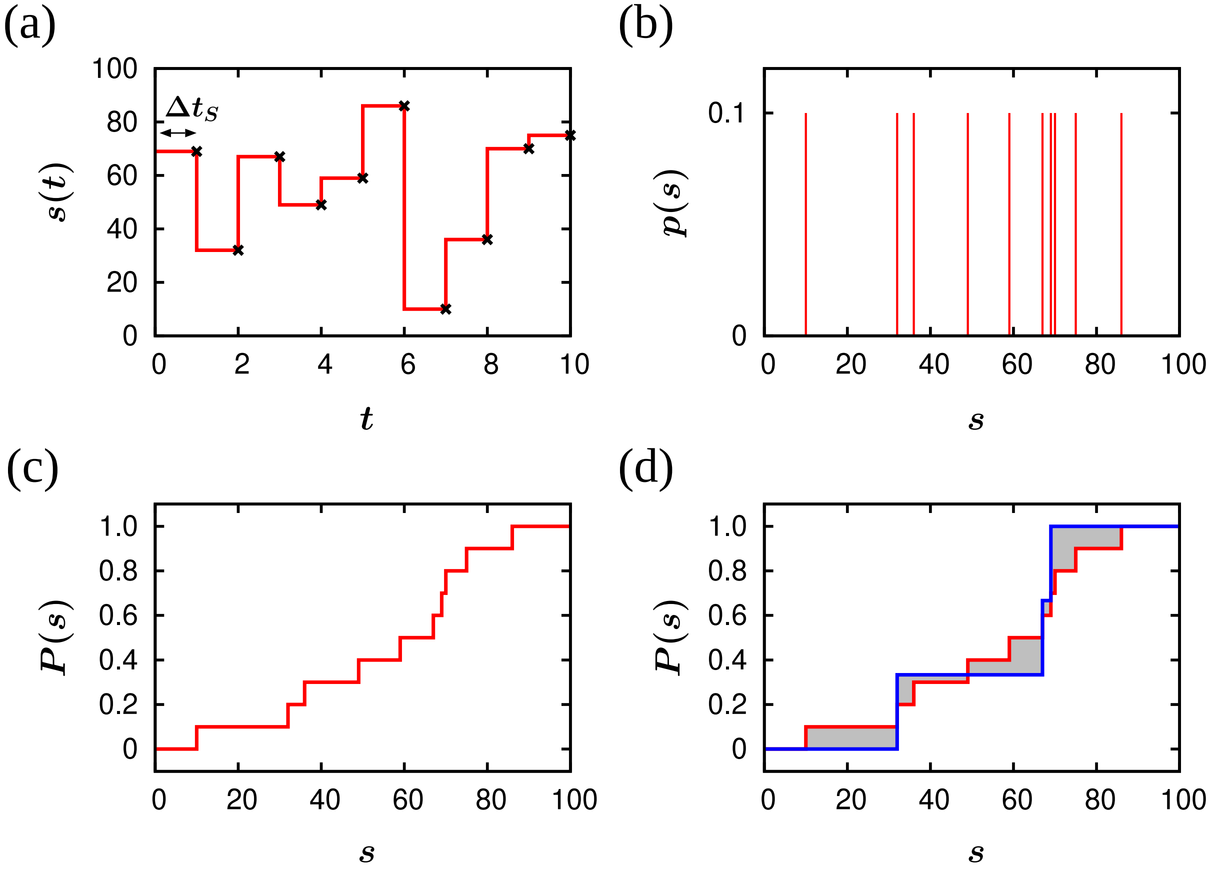

FIG. 1: The definition of the short-time distribution function and the Kantorovich metric. (a) A time series segment of the observable $s(t)$. The '×' symbol denotes each sampled point with the time resolution of the observation Δt_S. (b) The pmf $p(s)$ of the whole time series in (a). (c) The pdf $P(s)$ of the whole time series in (a). (d) The Kantorovich metric $d_K(p||p')$ between the pmf of the whole time series $p(s)$ and the short-time pmf of the first three data points $p'(s)$ whose corresponding pdf are represented by red and blue lines, respectively. The sum of the shaded areas enclosed by the two pdfs is equal to the Kantorovich metric $d_K(p||p')$.

the metric for all possible pairs of short-time pmfs. However, this procedure requires $n \times n$ distance matrix between short-time pmfs. Here n is the total number of short-time pmfs, namely, roughly equal to the total time steps in the time series. In this section, we present our actual clustering algorithm to partition the whole set of the short-time pmfs into a union of clusters without such an elaborate computation of order n^2.

In general, one does not know *a priori* the total number of states (more precisely, LES) of the system at a given (experimental) condition with a chosen time scale. Therefore, we chose a clustering algorithm which extracts each cluster or subset step by step from the ensemble of short-time pmfs without fixing *a priori* the total number of LES candidates. This algorithm is designed to assign sequentially from the largest LES candidate, where the system resides with highest probability, to the smaller one with lower residential probability.

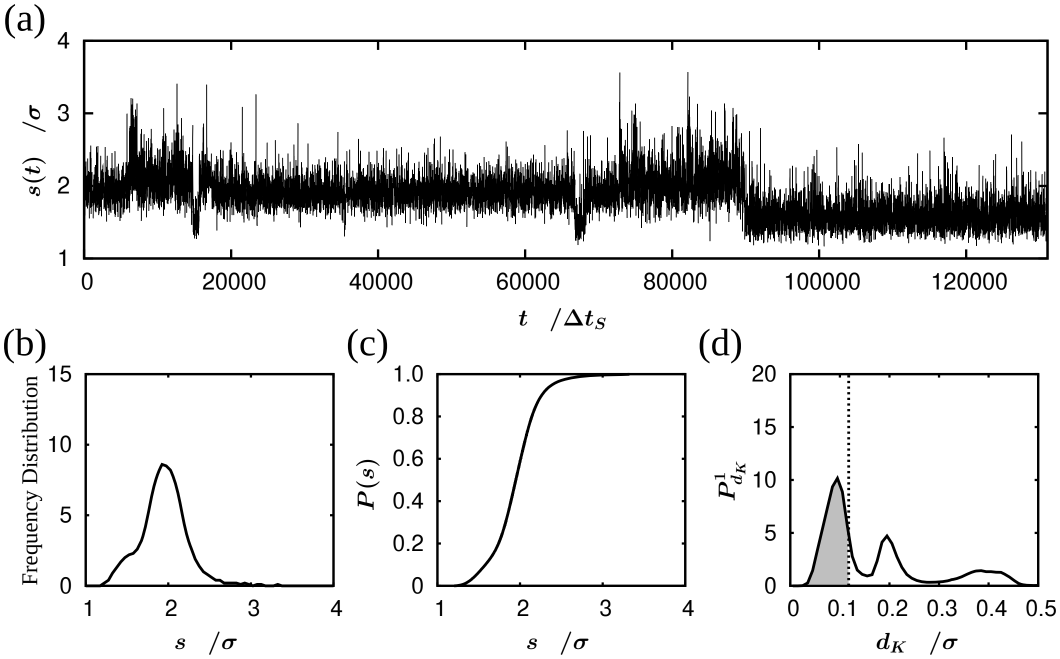

FIG. 2: The assignment procedure of the LES candidate from a given time series $s(t)$. **(a)** A time series $s(t)$ taken from the end-to-end distance of the 46-bead BLN model protein at $T = 0.3\epsilon$. The time resolution of the observation Δt_S corresponds to 1/1.8 of one vibrational period of the bond. The σ denotes the equilibrium bond length. **(b)** An initial guess frequency distribution of the first LES candidate. The unit of the vertical axis is $10^{-2}[-]$. The bin size is 0.01σ. Note that this figure and figures (e) and (h) are solely for making easier to understand the procedure. **(c)** The initial guess pdf $P(s)$. **(d)** The frequency distribution $P^1_{d_K}$ of the Kantorovich metric d_K between the individual short-time pmfs and the initial guess pmf of the first LES candidate. The unit of the horizontal axis and the vertical axis are σ^{-1} and $10^{-2}[-]$, respectively. The vertical dotted line denotes the threshold value chosen so as to cover the first peak from $d_K = 0$. The shaded area is composed of an ensemble of the short-time pmfs close to the initial guess pmf.

Figure 2 illustrates our procedure to assign the first, largest LES candidate from a

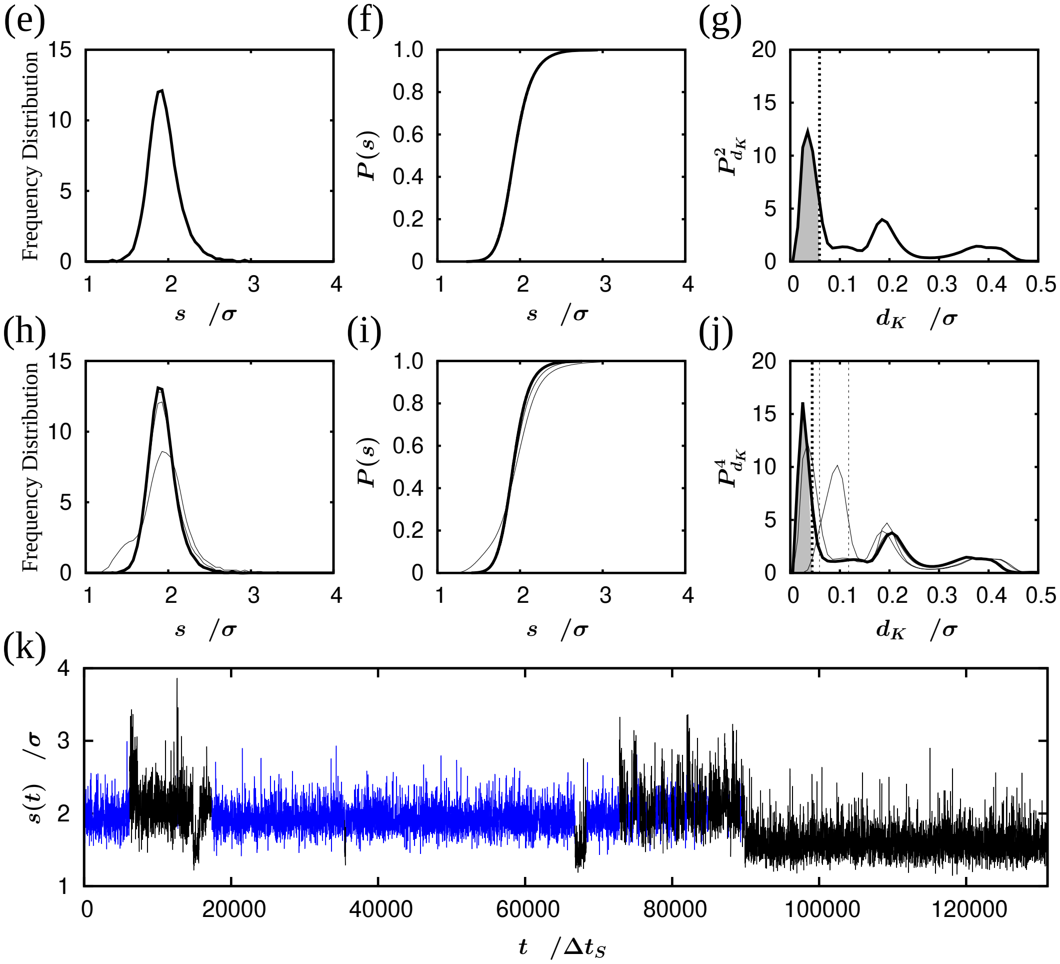

FIG. 2: (Continued) **(e)** The second guess frequency distribution of the first LES candidate constructed in terms of the set of the short-time pmfs belonging to the shaded area in (d). **(f)** The second guess pdf of the first LES candidate. **(g)** The frequency distribution $P^2_{d_K}$ of the Kantorovich metric d_K between the individual short-time pmfs and the second guess pmf of the first LES candidate. **(h)** The converged frequency distribution of the first LES candidate through the iteration process to reach the convergence (denoted by the bold line). All the other thin lines represent all the transient frequency distributions obtained through the iteration. **(i)** The converged pdf of the first LES. **(j)** The frequency distribution $P^4_{d_K}$ of the Kantorovich metric d_K between the individual short-time pmfs and the converged pmf of the first LES candidate (denoted by the bold line). **(k)** The time series fragments assigned as belonging to the first LES candidate is denoted by the blue lines. The other residual fragments besides the blue lines are to be used for the next assignment procedure of the second largest LES candidate.

given time series $s(t)$. Provided that we have an ensemble of the short-time pmfs with a τ, *how do the short-time pmfs corresponding to one LES candidate distribute in the Kantorovich metric space?* Suppose that the system can move about "ergodically" faster than the chosen τ within a LES candidate. Then, all short-time pmfs centered at each time t, denoted by $g_t^{(\tau)}(s)$ hereinafter, taken from the time period when the system resides in the LES candidate, are expected to be almost the same as the pmf with respect to s of the LES candidate (in general, irrespective of discreteness of s, there exists no guarantee that this is a Gaussian function). However, in practice, $g_t^{(\tau)}(s)$ are constructed in terms of a *finite* number of the sampled points n_S and, hence, $g_t^{(\tau)}(s)$ must, to some extent, deviate from the pmf of the LES candidate. If one could have an "infinitely" large number of sampled points within the same time interval $\tau(>> \tau_{eq})$, all $g_t^{(\tau)}(s)$ belonging to that LES candidate would fall into a single 'pin point' in the Kantorovich metric space. However, due to the finiteness of sampled points within τ, the set of $g_t^{(\tau)}(s)$ belonging to the same LES candidate should be broaden or diffuse in that space to some extent dependent on n_S (and also due to the contamination of some noise in actual experiments).

How can one estimate or identify a union of the short-time pmfs $g_t^{(\tau)}(s)$ which are expected to belong to the same LES candidate ? Here we present our iterative clustering procedure. First, we define an initial guess pmf of the first LES candidate with respect to the observable s, to which we compute the Kantorovich distances d_K from all $g_t^{(\tau)}(s)$ extracted from the whole time series. Here, we chose the pmf of the whole time series with respect to s as the initial guess. Figure 2 (c) shows the corresponding probability distribution function one can compute from the initial guess pmf of the first LES candidate by integrating over s (Figure 2 (b) presents the frequency distribution with respect to s of the whole time series with a bin size of 0.01σ in order to make easier to capture the procedure). In principle, irrespective of the detailed functional form of the initial guess, one can expect that some peaks appear in the frequency distribution with respect to d_K between the chosen initial guess pmf of the LES candidate and all $g_t^{(\tau)}(s)$, denoted by P_{d_K}, if several sets of $g_t^{(\tau)}(s)$ are very close each other (in the sense of Kantorovich metric) in their own subset. The reasons of our choice of this initial guess are: (1) a set of $g_t^{(\tau)}(s)$ to form the most *dominant* and *stable* LES may be close to the pmf of the whole time series and, if so, it results in a large peak around $d_K \sim 0$ in P_{d_K} (see the shaded region in Figure 2 (d)). Namely, one can assign a (plausible) set of $g_t^{(\tau)}(s)$ that mostly contribute to the dominant state; (2) the computation of the probability distribution function of the whole time series is most straightforward and simplest without determining the bin size.

As an initial guess, one can also use the most significant Gaussian function by fitting the frequency distribution of the whole time series by a combination of Gaussian functions. Especially, if distinct multi-modal peaks exist in the frequency distribution of the whole time series, the most dominant Gaussian function should be a better initial guess. However, at least, in our case study of 46-bead BLN model protein, both initial guesses gave rise to almost the same consequence after we performed our iterative clustering scheme we will describe below.

Now let us describe the iterative scheme to extract a set of $g_t^{(\tau)}(s)$ to compose the largest LES candidate from $s(t)$ after choosing an initial guess pmf of the LES candidate. We evaluate the frequency distribution of the Kantorovich metric d_K between the chosen initial guess pmf and all $g_t^{(\tau)}(s)$, denoted by $P_{d_K}^1$ (see Figure 2 (d)). Here, the superscript "1" in $P_{d_K}^1$ emphasizes that the P_{d_K} is evaluated for the *first*, initial guess. One can see three peaks in Figure 2 (d). What does the first, largest peak mean in the figure (i.e., the shaded peak in Figure 2 (d))? The first peak implies that there exists an ensemble of $g_t^{(\tau)}(s)$ close to the chosen initial guess pmf of the LES candidate. Here, we define a threshold of d_K to cover the first peak (indicated by the dotted line in the figure). There exist several criteria to define the threshold, e.g., a d_K to correspond to the first minimum which is expected to mostly cover the first peak (c.f., the computation of the delay time in embedology [1, 2]). However, if one picks up the first minimum in a rigorous sense, one might end up with an undesired value of d_K not to cover the first peak because P_{d_K} is not necessarily smooth. Therefore, in the present article, for the sake of brevity, we determined the threshold empirically in $P_{d_K}^i$ at every i^{th} iteration process and confirmed that the outcome was not so sensitive to the details of the chosen threshold (we will develop an objective scheme to determine the threshold value in the future).

The second guess pmf of the largest LES candidate shown in Figure 2 (f) (c.f., Figure 2 (e)) is constructed in terms of a set of $g_t^{(\tau)}(s)$ corresponding to the shaded area in Figure 2 (d). The shaded area corresponds to a union of $g_t^{(\tau)}(s)$ close to the initial guess pmf, which can be regarded as the most plausible candidates $g_t^{(\tau)}(s)$, at this stage, to form the most dominant LES candidate. Next we evaluate $P_{d_K}^2$ between the second guess pmf of the largest LES candidate and all $g_t^{(\tau)}(s)$ in the whole time series (see Figure 2 (g)), and determine the threshold to cover the new largest and sharper peak shifted closer to $d_K \sim 0$. Likewise, one can iteratively refine the pmf of the desired LES candidate.

Figures 2 (h) and (i) present the converged frequency distribution and pdf of the desired, first LES candidate indicated by bold lines. It was found for the first LES candidate that the convergence can be brought about by ~ 4 iterations. One can see a general trend of how $P^i_{d_K}$ evolves along the iteration process in Figure 2 (j). That is, along the process of the iteration, the first peak appeared in $P^i_{d_K}$ becomes sharper and closer to $d_K \sim 0$, while the other peaks slightly shift away from $d_K \sim 0$. Finally, the converged pdf (shown by the bold line in Figure 2 (i)) along the above procedure is considered as that of the desired, first LES candidate. The time fragments in $s(t)$ depicted by blue color in Figure 2 (k) indicate time regimes in which all $g_t^{(\tau)}(s)$ defined at time t were turned out to compose the first LES candidate. We use the residual part of time series for identifying the next LES candidate, after subtracting all fragments belonging to the first LES candidate. We can assign all sets of $g_t^{(\tau)}(s)$ to form their LES candidates by repeating the above procedure up to the stage such that all time fragments to be used for extracting the next LES candidate become shorter than the chosen τ.

The projection of the short-time probability density functions to a two-dimensional plane such as Figure 1(b) in the main text is not required in our iterative clustering algorithm. The visualization, however, sometimes helps us to capture the procedure. We made the figure as follows: the positions of short-time probability distribution functions on the two-dimensional plane was optimized the closeness centrality of Kantorovich metric between every pair of the short-time distribution functions by using the software named 'visone' (http://www.visone.info). The initial positions are not randomly selected but manually chosen because of the poor convergence of the default position supplied by the software visone. We manually divided these short-time probability distribution functions to several clusters by reflecting the consequence of the assignment of our LES analysis.

Clustering can be considered as NP-complete combinatorial optimization problem, for which optimal solutions can be found by branch-and-bound technique but in exponential time. For instance, the well-known k-means method of clustering algorithm requires the total number of the clusters *a priori* [3]. The other methods such as an agglomerative hierarchical clustering which constructs a dendrogram of the aggregation pathways from all the isolated data points to the single merged group may be more efficient but more complicated [4]. Our iterative clustering algorithm free from fixing the total number of clusters provides the converged solution for each clustering within ~ 10 iterations (at least in the current example of the end-to-end distance time series of 46-bead model protein). We believe that our algorithm is simplest so that one can easily implement the algorithm.

3. The transition sequence between LES candidates, the escape time from LES candidates, and the reaction rate between LES candidates

After the assignment of all LES candidates in the time series, the transitions between the LES candidates are assigned as follows: as exemplified in the blue line in Figure 2 (k), one can assign when the system visits each LES candidate obtained in our iterative clustering algorithm by checking which LES candidate is the closest (in the sense of Kantorovich metric) to the short-time pmf defined at the time t in question. Then, we check if the time window τ is shorter than the escape time $\tau_{esc}(i)$ from the ith LES candidate. If a LES candidate satisfies $\tau < \tau_{esc}(i)$ we assign the candidate of state as an LES, otherwise as a non-LES at the chosen τ. The escape time $\tau_{esc}(i)$ is evaluated as follows: The survival probability distribution of each LES candidate is estimated as a histogram of the residential time in the corresponding LES candidate. We fit the survival probability distribution in terms of a single exponential function $\propto e^{-t/\tau_{esc}(i)}$ for a region such that $t > \tau$. The reason why the escape time $\tau_{esc}(i)$ is estimated for fragments of longer residential times than the time window τ is the following: (1) When extracting a segment of time window, $(t_m - \tau/2, t_m + \tau/2]$, the time segments overlap from adjacent m. This is aimed at maximally using the given time series with persisting the experimental time resolution Δt_S. This should result in a certain apparent correlation within a time scale of τ. In elucidating quantities relevant to the time evolution such as the computation of the escape time of the system from the LES candidate and the identification of LES transition sequence, we omitted the time domain shorter than τ.

The other condition of supporting the concept of LES, the equilibrium time $\tau_{eq}(< \tau)$, was not tested explicitly. The state classified as an LES should, in principle, provide us with a *unique* short-time distribution of the observable whenever the system revisits the same state along the course of time evolution. However, when τ_{eq} is not fast enough to result in the unique short-time distribution, compared with τ, the corresponding short-time pmfs $g_t^\tau(s)$ may much diffuse in the Kantorovich metric space. One can expect that $g_t^\tau(s)$ such that $\tau_{eq} \gtrsim \tau$ must not be assigned as the same cluster, i.e., the same LES candidate.

One can evaluate the transition probabilities P_{ij} from the ith, to the jth LES candidate, that is, how often the system escapes or reacts from the ith, to the jth LES candidate per unit time. From them, one can compute an effective free energy landscape with checking if $P_{ij} \simeq P_{ji}$ is satisfied in a given time series. The reaction rate between the LESs are elucidated by the average number of the reaction events per unit time. Here, because of the same reason in computing the escape time, we count the reaction events between different LESs after subtracting (from the time series) a piece of the time fragments in which the system resides for a certain time duration shorter than τ.

B. Kantorovich metric in comparison with Kullback-Leibler divergence and Hellinger distance

What is the most appropriate measure to describe the 'distance' between two probability density functions evaluated for a short time window τ? In order to compare with the actual distance between two superbasins composed of a number of protein conformations we define the averaged distance R_S between all the pairs of the conformations belonging to the ith and jth LES/non-LES in $3N$-dimensional conformation space (N is the number of particles):

$$R_S(i\|j) = \frac{1}{M_i M_j} \sum_{\alpha \in i}^{M_i} \sum_{\beta \in j}^{M_j} (\sum_{k=1}^{3N} (r_{\alpha,k} - r_{\beta,k})^2)^{\frac{1}{2}}, \tag{4}$$

where M_i and M_j denote the number of (transient) configurations classified into the ith and jth LES/non-LES, respectively. $r_{\alpha,k}$ means the kth Cartesian coordinate of the αth configuration where McLachlan 's 'best fit' prescription [5] was employed to remove the total translational and rotational degrees of freedom to remove uncertainty in the definition of the coordinate system. Figs. 3 and 4 reveal how the Kantorovich metric d_K, and Kullback-Leibler divergence (relative entropy) d_{KL} and Hellinger distance d_H can capture the actual distances between the two LES/non-LES in $3N$-dimensional conformation space. In the Fig. 3, for the sake of comparison, the average of the end-to-end distance difference

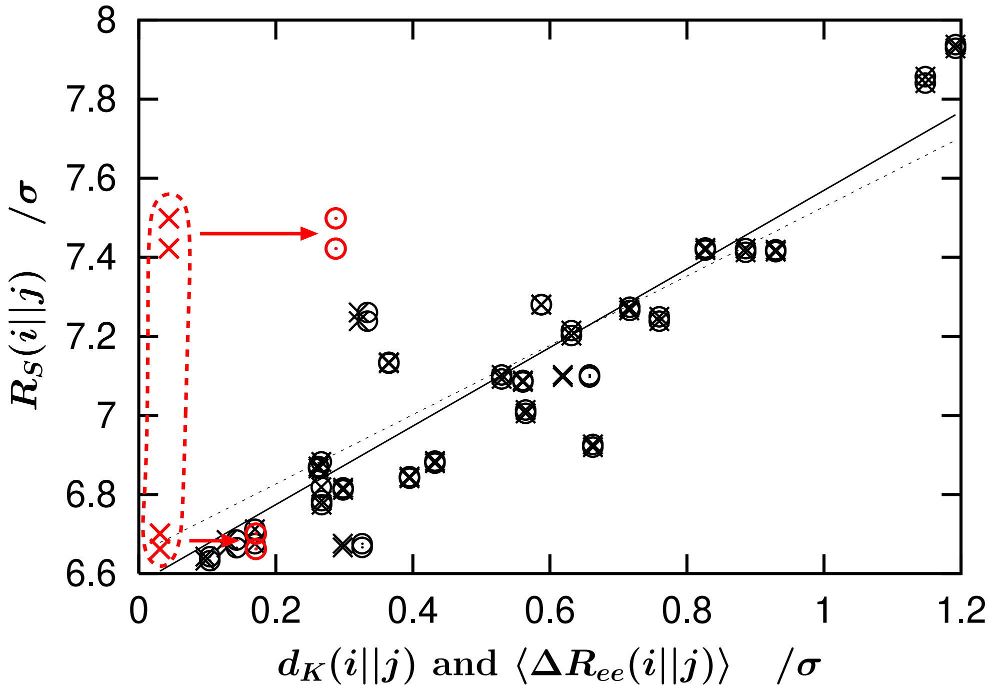

FIG. 3: The relationship between the structural distance $R_S(i\|j)$ and $d_K(i\|j)$ in terms of the end-to-end distance time series between the ith and jth LES/non-LES at 0.4ϵ ($\tau = 100$). Here, the '×' denotes the average of differences of the end-to-end distances R_{ee} between the (transient) configurations composed of the ith and jth LES/non-LES $\langle \Delta R_{ee}(i\|j) \rangle$ defined by $1/(M_i M_j) \sum_{\alpha\in i}^{M_i} \sum_{\beta\in j}^{M_j} ((R_{ee}(\alpha) - R_{ee}(\beta))^2)^{\frac{1}{2}}$. The degeneracies in $\langle \Delta R_{ee}(i\|j) \rangle$ among the pairs of LES/non-LES denoted by red × are partially lifted in $d_K(i\|j)$ as shown by red ○. The dotted and solid lines represent the least square fitted lines for $\langle \Delta R_{ee}(i\|j) \rangle$ and $d_K(i\|j)$, respectively.

$\langle \Delta R_{ee} \rangle$ between all the pairs of the conformations belonging to the two LES/non-LES are also plotted.

Figures manifestly demonstrate that the Kantorovich metric d_K is much superior to the others d_{KL} and d_H in capturing the actual conformational distance. The most striking consequence is this: in the region where the two LES probability density functions $f_i(s)$ and $f_j(s)$ have some overlaps in shorter separations, d_K partially lift degeneracy that exists in $\langle \Delta R_{ee} \rangle$. This implies that the Kantorovich metric d_K can more differentiate the underlying (multidimensional) morphological feature associated with LES/non-LES than $\langle \Delta R_{ee} \rangle$ in addition to Kullback-Leibler divergence and Hellinger distance.

It should be noted that Kullback-Leibler divergence (relative entropy) is not true metric

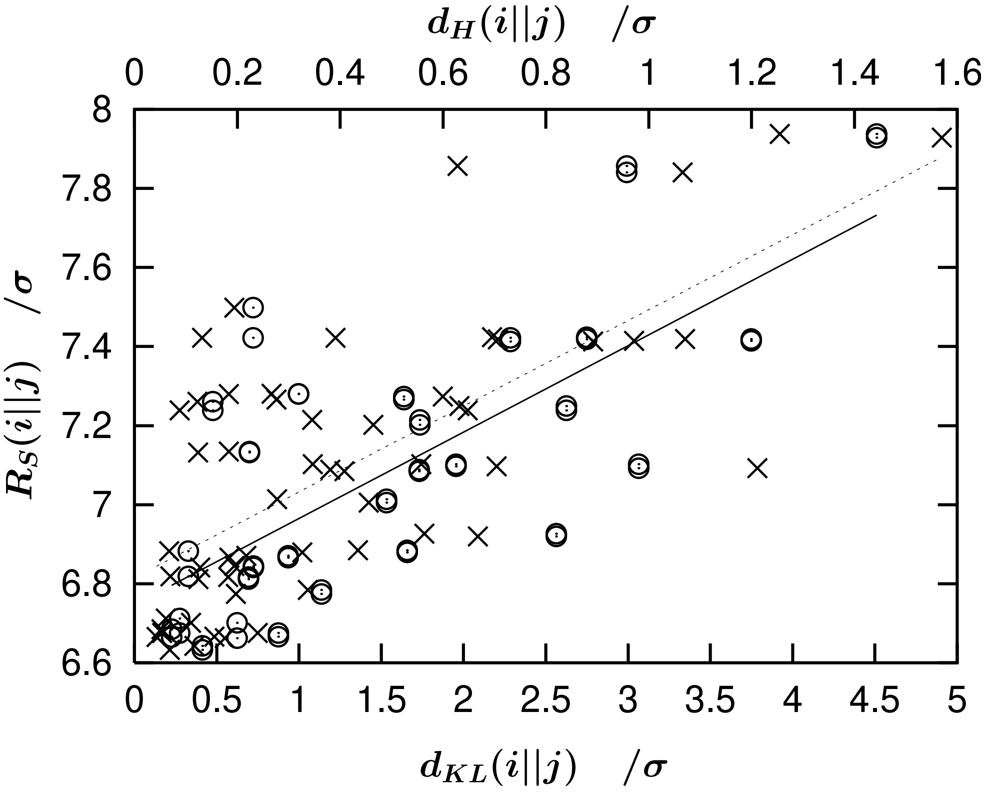

FIG. 4: The relationship between the structural distance $R_S(i\|j)$ and Kullback-Leibler divergence $d_{\mathrm{KL}}(i\|j)$ (×) and Hellinger distance $d_{\mathrm{H}}(i\|j)$ (○) for the end-to-end distance time series of BLN model protein at 0.4 ϵ. The i and j stand for ith and jth LES/non-LES constructed from $\tau = 100$. The dotted and solid lines represent the least square fitted straight lines for d_{KL} and d_{H}, respectively. The horizontal bottom and upper axes denote the scales of d_{KL} and d_{H}, respectively. From the definitions of d_{KL} and d_{H}, $d_{\mathrm{KL}} = \int_{-\infty}^{\infty} f_i(s)\log_2(f_i(s)/f_j(s))ds$ and $d_{\mathrm{H}} = \int_{-\infty}^{\infty}(f_i(s)^{1/2} - f_j(s)^{1/2})^2 ds$ (where $f_i(s)$ and $f_j(s)$ are the ith and jth probability density function), one can easily see $d_{\mathrm{KL}} \to \infty$ and $d_{\mathrm{H}} \to 2$ when the overlap between two probability density function diminishes.

although it has some properties of metric; it is always non-negative and is zero if and only if $f_i(s) = f_j(s)$. Hellinger distance satisfies the metric condition but the value of d_H converges to two for large separation of no-overlap between two distributions. Here we show that $d_K(f_i\|f_j)$ is equal to the actual difference between the average values of s of the individual $f_i(s)$ and $f_j(s)$ $|(\int_{-\infty}^{\infty} ds\ s\ f_i(s)) - (\int_{-\infty}^{\infty} ds\ s\ f_j(s))|$ when they have no overlap in the variable s, which neither the Kullback-Leibler divergence nor Hellinger distance is not so. The d_K (Eq. [1]) satisfies the triangle inequality,

$$d_K(f_i\|f_j) \leq d_K(f_i\|f_k) + d_K(f_k\|f_j), \tag{5}$$

which is one of the most important properties of metric [6]. In addition, this metric $d_K(f_i\|f_j)$ is simply equal to the Euclidean distance between the average values of s of the individual probability density functions $f_i(s)$ and $f_j(s)$ if the following condition is satisfied:

$$\forall s \in \mathbb{R}; P_i(s) \geq P_j(s), \tag{6}$$

In such cases,

$$\begin{aligned} d_K(f_i\|f_j) &= \int_{-\infty}^{\infty} ds \, |P_i(s) - P_j(s)| \\ &= \int_{-\infty}^{\infty} ds \, (P_i(s) - P_j(s)) \\ &= [s \ P_i(s)]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} ds \ s \ f_i(s) \\ &\quad - [s \ P_j(s)]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} ds \ s \ f_j(s) \\ &= \langle s \rangle_i - \langle s \rangle_j \, , \end{aligned}$$

where

$$\langle s \rangle_i = \int_{-\infty}^{\infty} ds \ s \ f_i(s). \tag{7}$$

The second equality results from Eq. (6) (the order of i and j has no meaning because this measure satisfies the symmetric property of metric, i.e., $d_K(f_i\|f_j) = d_K(f_j\|f_i)$) and the fourth equality from $P_i(-\infty) = P_j(-\infty) = 0, P_i(\infty) = P_j(\infty) = 1$. One can easily see that Eq. (6) is satisfied when the two probability density functions have no overlap in the variable s.

C. The relation between the LES/non-LES vs TRDG basins

One can easily assign which LES/non-LES the system traces along the scalar time series after the clusters (subsets) are extracted from a set of the short time distribution for time window τ. The TRDG basins were constructed in terms of a subset of 1.6×10^4 inherent structures quenched along an isothermal MD trajectory of $2.2\times10^8\Delta t$. Thus one can assign to which TRDG basins the system was quenched along the time series of the full set of dimension [7]. Fig. 5 demonstrates the residential probabilities in each LES/non-LES while the system is quenched in the lowest to tenth lowest TRDG basins of the BLN model at 0.4 ϵ. Figure tells us, for example, that the highest ratio in the residential probabilities at TRDG basin 1, 2, 3, and 4 were 76.9% in LES 1, 89.8% in LES 2, 55.5% in non-LES 3, and 58.2% in LES 4, respectively. This suggests that the LES analysis fairly well captures the underlying multidimensional free energy landscape because the most dominant contribution of LES to TRDG basin i are ith LES for $i = 1, 2, 4$ (namely, the relative order of stability does not change). Note, however, that such one-to-one correspondence starts to cease at higher than TRDG basin 5. For instance, although the TRDG basin 3 is mainly composed of the non-LES 3, the non-LES 3 also contributes in the TRDG basins 5 and 6. The most dominant contribution to the TRDG basin 5 is not LES 5 but non-LES 3.

In principle, for any scalar finite time series, it is inevitable that some short-time distribution functions with time window τ (which should belong to distinct free energy basins) still degenerate. One may conjecture that the discrepancy between LES/non-LES and TRDG basins observed at TRDG basin i (≥ 4) arise from such a degeneracy which could not be 'lifted' by the short time distributions. It should be noted however that there exist several implicit assumptions for the TRDG procedure: for example, it assumes transition state theory (TST) based on the concept of local equilibrium (and no-return ansatz) for the elucidation of the escape rates from all potential minima irrespective of kinds of potential minima (e.g., passages through some shallow potential minima does not necessarily guarantee the validity of TST). Therefore, there is no source to yield non-LES within the framework of TST. In addition, the two TRDG basins are unified when both the evaluated TST rate constants from one basin to another and vice versa are faster than a chosen threshold [8]. It was turned out that, with a threshold larger than $\tau/2$ ($\sim$22 oscillations of the individual bond stretching in 46 bead model protein), all the TRDG free energy basins are unified as the single LES at 0.4 ϵ above the folding temperature. This might too exaggerate the morphological change of multidimensional free energy landscape in time scale of observation.

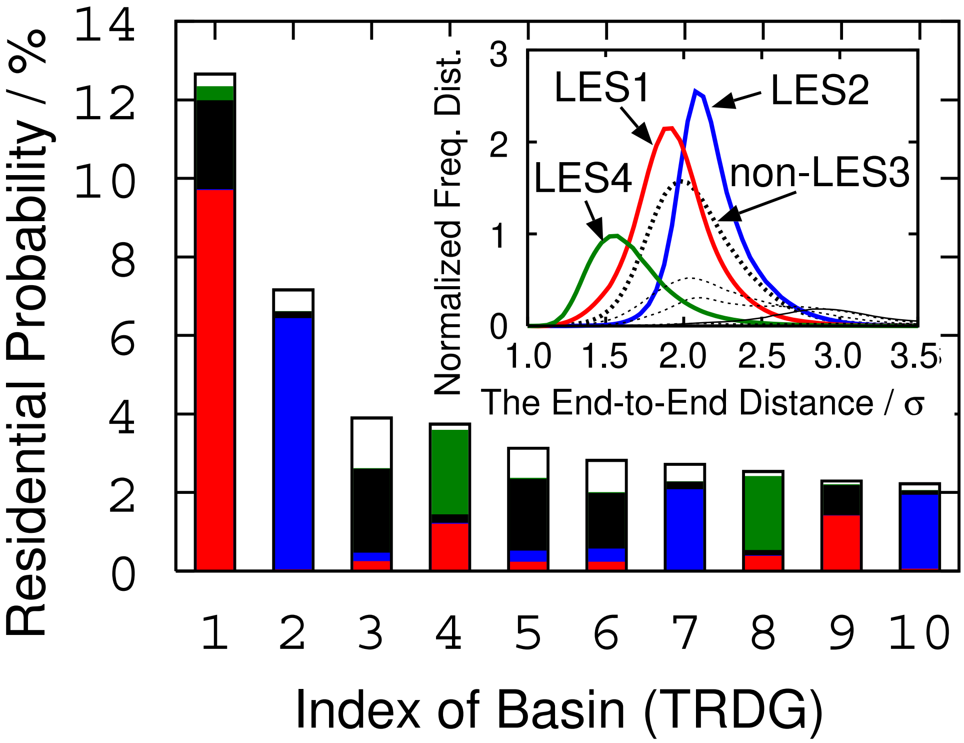

FIG. 5: The ratio of the residential probabilities in each LES/non-LES at the system traversing the ith lowest TRDG basins of the BLN model at 0.4 ϵ. The inset is the same as Fig. 5 (b). The red, blue, black, green and while color in the histogram corresponds, respectively, to the ratio of LES 1, LES 2, non-LES 3, LES 4 and all LES/non-LES i $(i > 4)$.

The TRDG is constructed from inherent structures obtained by being quenched along the dynamical evolution. This treatment should be appropriate at least in a certain low temperature regime where the system can actually trace the inherent structures. However, in principle, there is no firm foundation of how much the system actually 'experiences' the underlying inherent structures at higher temperature where the system starts to see lots of hierarchical superbasins with higher-rank saddles.

[1] Abarbanel, HD-I, (1995) *Analysis of Observed Chaotic Data.* (Springer-Verlag, New York).

[2] Kantz, H, & Schreiber, T, (1997) *Nonlinear Time Series Analysis.* (Cambridge Univ Press, Cambridge, UK).

[3] MacKay, DJC, (2003) *Information Theory, Inference, and Learning Algorithms* (Cambridge

Univ Press, Cambridge, UK).

[4] Larsen, J, Szymkowiak, A, & Hansen, LK, (2002) *International Conference on Knowledge-Based Intelligent Information Engineering Systems & Allied Technologies* 6:56–62.

[5] McLachlan, AD, (1979) *J Mol Biol* 128:49–79.

[6] Vershik, A, (2006) *J Math Sci* 133:1410–1417.

[7] Krivov, SV, & Karplus, M, (2004) *Proc Natl Acad Sci USA* 101:14766–14770.

[8] Evans, DA, & Wales, DJ, (2003) *J Chem Phys* 118:3891–3897.