Hidden Complexity in the Isomerization Dynamics of Holliday Junctions

Changbong Hyeon1*, Jinwoo Lee2, Jeseong Yoon1, Sungchul Hohng2 & D. Thirumalai3*

1School of Computational Sciences, Korea Institute for Advanced Study, Seoul 130-722, Korea
2Department of Physics and Astronomy, Seoul National University, Seoul 151-747, Korea
3Institute for Physical Science and Technology, University of Maryland, College Park, MD 20742, USA

Contents

1. Single molecule FRET experiments ............................................................... 2
2. Memory of initial condition ................................................................. 3
3. Probing the ergodicity breaking ............................................................ 4
4. Clustering the stationary FRET distribution functions .......................... 5
5. Clustering at the optimal values of K ...................................................... 6
6. Calculation of electrostatic potential maps of HJ structures .................. 6
7. Calculation of radial distribution function of Mg2+ ions around HJ ........ 7
   • References and Notes ............................................................ 8
   • Figures S1 - S10 .................................................................. 9-18
1 Single molecule FRET experiments

To assemble the Holliday junction, the following DNA sequences were purchased from IDTDNA (Coralville, IA).

R branch: 5’-biotin-TTTTTTTT CCCACCGCTCGGCTCAACTGGG-3’
H branch: 5’-Cy3-CCGTCAGCAGCGCGAGGGTG-3’
X branch: 5’-GGGCCGGCGACCT CCCAGTTGAGCGTGGCTAGGG-3’
B branch: 5’-Cy5-CCCTAGCAAGCCGCTACGG-3’

where the sequences with identical font style represent the pair of complementary DNA strands. The DNA strands were annealed by heating the mixture of DNAs (1 µM, 20 µl) to 90 ºC in TN buffer (10 mM Tris with 50 mM NaCl, pH 8.0), and slowly cooling down to 4 ºC with 1 ºC/min cooling rate. For single-molecule FRET (smFRET) measurements, a sample chamber was made between a cleaned quartz microscope slide (Finkenbeiner) and a cover slip using double-sided adhesive tape. DNAs were immobilized on quartz surface by successive additions of biotinylated BSA (40 µl, 1 mg/ml, Sigma-Aldrich), streptavidin (40 µl, 0.2 mg/ml, Invitrogen), and DNA in TN buffer. The concentration of DNA was adjusted to achieve proper single molecule density. Each injection was incubated for 2 minutes, which was followed by washing with TN buffer. Experiments were performed at room temperature in a conventional imaging buffer (10 mM Tris-HCl with 0.4% (w/v) glucose (Sigma-aldrich), 1% (v/v) trolox (Sigma-aldrich), 1mg/ml glucose oxidase (Sigma-aldrich), 0.04 mg/ml catalase (Roche, Nutley, NJ), and designated magnesium ion) by using a home-built prism-type total internal reflection single-molecule FRET setup (Fig. S1). Specifically, a green laser (532 nm, Compase 215 M, Coherent) was used as an excitation source. Fluorescence signals of donor and acceptor were collected through a water-immersion objective (Olympus, UPlanoSApo 60x/1.2w), divided using dichroic mirror (Chroma, 635 dcxr) as a wavelength, and finally focused on different areas.
of an electron multiplier charge-coupled-device camera (Andor, Ixon DV897ECS-BV). To reliably detect individual transitions of HJ, optimum exposure time of the camera was selected at each Mg\textsuperscript{2+} concentration (30 ms for 5 mM and 10 mM Mg\textsuperscript{2+} concentrations, and 50 ms for other concentrations). Data acquisition and selection of single molecule traces were done by using home-made programs written in VC++6.0, and IDL (ITT), respectively.

For Mg\textsuperscript{2+} pulse experiment (Fig. S2), a teflon tube was connected to the exit hole of a sample chamber filled with 50 mM Mg\textsuperscript{2+} buffer, and a pipette tip filled with Mg\textsuperscript{2+}-free imaging buffer was carefully plugged into the entrance hole. After starting a data acquisition, the buffer solution in the detection chamber was rapidly replaced with a Mg\textsuperscript{2+}-free buffer solution in the pipette tip by pulling a syringe tube connected to the Teflon tube. Then, remaining solution in the pipette tip was carefully replaced with a 50 mM Mg\textsuperscript{2+} buffer, and the new solution containing 50 mM Mg\textsuperscript{2+} was rapidly introduced into the detection chamber by suction. During the whole process, data acquisition was maintained. The measured smFRET trajectories, under a range of counterion (Mg\textsuperscript{2+}) concentrations, were analyzed using concepts in glass physics and bioinformatic tools.

2 Memory of initial condition

Another evidence for the inadequacy of using $\Sigma(t)$ and $P_{ens}(t)$ calculated over ensemble in unraveling the rugged folding landscapes can be given by comparing the patterns of FRET trajectories and the probability of observing such trajectories based on two-state picture. A series of dwell times of a trajectory generated from a system that precisely exhibits two-state kinetics should obey a renewal (more precisely a Poisson) process with dwell time distribution $p_{dwell}(t) = \tau^{-1} e^{-t/\tau}$, where $\tau$ is the mean dwell time. For such a system, the probability of observing a time trace with successively long multiple ($n$) dwells ($\tau^* \gg \tau$), i.e.,
\[ \left( \int_{\tau}^{\infty} dtp_{\text{dwell}}(t) \right)^n = \exp \left( -n\tau^* / \tau \right) \] is essentially zero (For \( n = 10 \) and \( \tau^* = 5\tau \), the probability of observing such a time trace is less than \( 10^{-22} \)). Nevertheless, our smFRET trajectories contain a preponderance of such cases, in which successively long multiple dwells are observed (see Fig. 2a). Thus, even during frequent isomerization, a molecule behaves as if it retains “memory” of its basin of attraction [1], which implies that a given molecule repeatedly visits exactly the same states in iso-I and iso-II.

3 Probing the ergodicity breaking

For an ensemble of smFRET trajectories that reports the FRET efficiency \( (E) \) as a function of time \((t)\), the fluctuation metric \( (\Omega_E(t)) \) is defined [2]:

\[
\Omega_E(t) = \frac{1}{N} \sum_{i=1}^{N} \left( \varepsilon_i(t) - \overline{\varepsilon}(t) \right)^2
\]

where \( \varepsilon_i(t) \equiv \frac{1}{t} \int_0^t ds E_i(s) \) is the time average of \( E \) for molecule \( i \), and \( \overline{\varepsilon}(t) \equiv \frac{1}{N} \sum_{i=1}^{N} \varepsilon_i(t) \) is the ensemble average of \( \varepsilon_i(t) \). In accord with the ergodic hypothesis we expect that, \( \varepsilon_i(t \rightarrow \infty) \equiv \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t ds E_i(s) = \frac{1}{N} \sum_{i=1}^{N} E_i(t) = \langle E \rangle \), which leads to \( \lim_{t \rightarrow \infty} \varepsilon(t) \equiv \frac{1}{N} \sum_{i=1}^{N} \varepsilon_i(t \rightarrow \infty) = \langle E \rangle \). Therefore, the necessary condition, \( \lim_{t \rightarrow \infty} \Omega_E(t) \rightarrow 0 \), should be fulfilled for ergodic systems. Eq.(1) for \( \Omega_E(t) \) can be rewritten as

\[
\Omega_E(t) = \frac{1}{t^2} \int_0^t ds_1 \int_0^t ds_2 \frac{1}{N} \sum_{i=1}^{N} (E_i(s_1) - \langle E \rangle)(E_i(s_2) - \langle E \rangle) = \frac{1}{t^2} \int_0^t ds_1 \int_0^t ds_2 C(s_1, s_2)
\]

where \( C(s_1, s_2) \) is the equilibrium time correlation function. If \( (E_i(t) - \langle E \rangle) \) is self-averaging, one can put \( C(s_1, s_2) = C(|s_2 - s_1|) \). Since in equilibrium there is no preferred origin of time \( C(s_1, s_2) \) depends only on the difference between the two times \( s_1 \) and \( s_2 \). With straightforward algebra \( \int_0^t ds_1 \int_0^t ds_2 C(|s_2 - s_1|) = 2 \int_0^t ds_1 \int_{s_1}^{t} ds_2 C(s_2 - s_1) = 2 \int_0^t d\tau(t - \tau)C(\tau) \) one gets
where $D_E = \lim_{t \to \infty} \left[ 2 \int_0^t ds \frac{C(s)}{C(0)} \right]^{-1}$. Note that for ergodic systems, the equilibrium time correlation function $C(s)/C(0)$ is a fast decaying function; thus the integral $2 \int_0^\infty \frac{C(s)}{C(0)} ds$ yields a constant. Next, the slope of the inverse of $\Omega_E(t)/\Omega_E(0)$ yields effective diffusion coefficient, $D_E$, the rate at which the molecule navigates the accessible conformational space. However, if the ergodicity of the system is effectively broken, $[\Omega_E(t)/\Omega_E(0)]^{-1}$ is no longer linear in time, but saturates into a constant value, which we observed for the ensemble of time trajectories that probe the isomerization dynamics of HJs.

4 **Clustering the stationary FRET distribution functions**

The set of stationary distributions calculated for the individual time traces of HJ isomerization, \( \{p_s(E; i) | i = 1, 2, \ldots, N \} \), can be partitioned into the K-clusters, \( S(E) = (S_1(E), S_2(E), \ldots, S_K(E)) \) with $K < N$, by employing K-means clustering algorithm that is widely used as a bioinformatics tool in analyzing the patterns of gene expression [3, 4]. We treated each stationary distribution \( p_s(E; i) \) as a vector in the $E$-space and iteratively minimized the sum of squared Euclidean distance from each $p_s(E; i)$ to its cluster centroid, i.e.,

$$
\sum_{k=1}^K \sum_{p_s(E; i) \in S_k(E)} ||p_s(E; i) - \mu_k(E)||^2
$$

where $\mu_k(E)$ is the mean of $S_k(E)$. After the random assignment of the cluster centroid to a vector in $E$-space, the K-means procedure iteratively refined partitioning of the $p_s(E; i)$ among the K clusters by assigning each vector to the nearest cluster centroids, then recalculating cluster centroids if cluster membership (\( S(E) \)) was changed from the previous iteration. If no change in any cluster membership occurred then the K-means algorithm was terminated. To avoid local
minima, the optimization was initiated from 200 random cluster centroids. We selected the best set of clusters that reaches the minimum value among the 200 trials.

The quality of clustering result can be assessed by using a silhouette function:

\[ s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \tag{4} \]

where \(a(i)\) is the average distance (dissimilarity) of a datum \(i\) with all other data within the same cluster, and \(b(i)\) is the lowest average dissimilarity of the datum \(i\) with the data belonging to other clusters. The optimal K value is determined by examining the mean silhouette function \(\langle s \rangle = \frac{1}{N} \sum_{i=1}^{N} s(i)\) against K.

### 5 Clustering at the optimal values of K

At \([\text{Mg}^{2+}] = 50 \, \text{mM}\), \(\langle s \rangle\) becomes optimal at \(K=3, 5, \text{ and } 13\). However, the criteria using Euclidean distance or Pearson’s correlation do not guarantee the ergodicity of each cluster. To this end we evaluated the fluctuation metric \(\Omega_E(0)/\Omega_E(t)\) for each cluster to ensure its ergodicity (Fig. S7). Although maximal \(\langle s \rangle\) was found at \(K=3\) (Fig. S7A), \(\Omega_E(0)/\Omega_E(t)\) of the cluster \(k=2\) (red), which represents the 36.5 % of the time traces, was not linear in \(t\) (Fig. S7B). Consequently, we had to inspect the next optimal values of \(K\) (\(K=5\) and 13).

### 6 Calculation of electrostatic potential maps of HJ structures

Electrostatic potential maps of the three Holliday Junctions were calculated by solving the non-linear Poisson-Boltzmann equation using the APBS software [5] at 200 mM monovalent ion condition with the dielectric constants for nucleic acids and solvent being 2.0 and 78.5, respectively. The calculated electrostatic potentials were visualized (Fig. 5a in the Main text) using
the PyMol program (http://www.pymol.org).

7 Calculation of radial distribution function of Mg\(^{2+}\) ions around HJ

To calculate radial distribution function of Mg\(^{2+}\) around HJ, we performed all-atom molecular dynamics simulations using the crystal structure of HJ with PDB ID 1DCW. After solvating the HJ molecule in a 66 Å × 66 Å × 66 Å box containing 8369 TIP3P water molecules, 37 Na\(^{+}\), 41 Cl\(^{-}\), and 20 Mg\(^{2+}\) ions were randomly placed. The system was minimized by using 1 fs time step for 2000 steps with constraints on DNA coordinates, then for an additional 3000 steps without constraints. During equilibration stage we gradually heated the system from 0 K to 310 K using 2 fs time step for 620 ps with constraint, and equilibrated for an additional 5 ns by using the NPT ensemble at 310 K and 1 atm. After equilibration, we generated a 100 ns trajectory in the NPT ensemble at 310 K and 1 atm. The integration step during production run is 2 fs. The simulations were performed by using NAMD with CHARMM force field.
References and Notes


Figure S1: A schematic diagram of experiments. DNA molecules were immobilized on a quartz surface via streptavidin-biotin interaction. Single-molecule images were taken in a prism-type TIRF (Total Internal Reflection Fluorescence) microscope. The fluorescence signals of Cy3 and Cy5 were collected by a water-immersion objective, and imaged on a separate areas of an EM-CCD camera (M: mirror, L: lens, F: filter, D: dichroic mirror).
Figure S2: A schematic diagram of the magnesium pulse experiments. A sample chamber filled with 50 mM Mg$^{2+}$ was prepared, and a pipette tip with 0 mM Mg$^{2+}$ buffer inside was carefully plugged into a injection hole. While single-molecule images were taken, 0 mM Mg$^{2+}$ buffer was rapidly introduced into the detection chamber by pulling a syringe pump connected to the exit hole of the detection chamber via a Teflon tube. Remaining solution in the pipette tip was carefully replaced with a 50 mM Mg$^{2+}$ buffer, and the new solution containing 50 mM Mg$^{2+}$ was rapidly introduced into the detection chamber by suction. During the whole process, data acquisition was maintained.
Figure S3: (Top) Donor ($I_D$), acceptor ($I_A$) signals and (bottom) the corresponding FRET efficiency ($E(t) = I_A(t)/(I_A(t) + I_D(t))$) as a function of time.
Figure S4: Histogram of FRET value distribution with varying Mg$^{2+}$ concentrations. Each $P_{ens}(E)$ is fit to double-Gaussian distribution (blue line), $P_{ens}(E) := \phi G_L(E) + (1-\phi)G_H(E)$ where $G_\lambda(E) = (2\pi\sigma_\lambda^2)^{-1/2}e^{-(E-E_\lambda)^2/2\sigma_\lambda^2}$ ($\lambda = L, H$) with the set of parameters shown in the box. Note that population of iso-I and iso-II are equally likely. Biomodal to uni-modal transition at a certain Mg$^{2+}$ ion concentration ($\leq 50\mu M$, see Ref.[6]) is reminiscent of second order phase transition. In this case, Mg$^{2+}$ ions is analogous to inverse temperature in an Ising spin system (see Mg$^{2+}$ ion dependent free energy and phase diagram on the right).

<table>
<thead>
<tr>
<th>[Mg$^{2+}$]</th>
<th>$\phi$</th>
<th>$E_L$</th>
<th>$\sigma_{E_L}$</th>
<th>$E_H$</th>
<th>$\sigma_{E_H}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0mM</td>
<td>1</td>
<td>0.30</td>
<td>0.051</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>0.5 mM</td>
<td>0.42</td>
<td>0.21</td>
<td>0.074</td>
<td>0.46</td>
<td>0.106</td>
</tr>
<tr>
<td>1mM</td>
<td>0.46</td>
<td>0.20</td>
<td>0.069</td>
<td>0.48</td>
<td>0.102</td>
</tr>
<tr>
<td>5mM</td>
<td>0.51</td>
<td>0.20</td>
<td>0.073</td>
<td>0.53</td>
<td>0.108</td>
</tr>
<tr>
<td>10mM</td>
<td>0.48</td>
<td>0.20</td>
<td>0.078</td>
<td>0.52</td>
<td>0.123</td>
</tr>
<tr>
<td>20mM</td>
<td>0.51</td>
<td>0.20</td>
<td>0.075</td>
<td>0.52</td>
<td>0.125</td>
</tr>
<tr>
<td>30mM</td>
<td>0.45</td>
<td>0.19</td>
<td>0.061</td>
<td>0.53</td>
<td>0.093</td>
</tr>
<tr>
<td>50mM</td>
<td>0.55</td>
<td>0.20</td>
<td>0.061</td>
<td>0.54</td>
<td>0.093</td>
</tr>
</tbody>
</table>
Figure S5: The probability of remaining in high (green) and low $E$ values (blue). Red lines correspond to single exponentials. The mean life time in each state can be obtained using $\langle \tau \rangle = \int_0^\infty d\tau \Sigma(\tau) = \int_0^\infty d\tau \Sigma(\tau)$ where $\Sigma(t) = 1 - \int_0^t d\tau \Sigma(\tau)$.
Figure S6: Analysis of the transition dynamics of HJ trajectories at \([\text{Mg}^{2+}] = 50\) mM shows that the ergodicity is still broken over an extended observation time \(T_{\text{obs}} \approx 70\) sec. \(\varepsilon_i(t)\), \(\varepsilon(t)\) and \(\Omega_E(t)/\Omega_E(0)\) (or \([\Omega_E(t)/\Omega_E(0)]^{-1}\)) are shown in grey, green, and red lines, respectively.
Supplementary Information

Figure S7: Clustering analysis of FRET trajectories at [Mg^{2+}]=50 mM. A. Mean silhouette function \( \langle s \rangle \) as a function of the number of clusters, K. Optimal \( \langle s \rangle \) values are found at K=3, 5, 13 as indicated by the arrows. B. Partitioning of \( \{ p_s(E; i) \} \) with K=3 yields three clusters (k=1 (12.7 %), k=2 (36.5 %), k=3 (50.8 %)) but the dynamics within the cluster, k=2, is still non-ergodic. \( \Omega_E(0)/\Omega_E(t) \) and scatter plot of \( \langle \tau_{L,i}, \tau_{H,i} \rangle \), partitioned into three clusters, are shown on the right. C. Partitioning of \( \{ p_s(E; i) \} \) with K=13 are shown. The cluster k=6 is still non-ergodic but its percentage (5.1 %) is small. \( \Omega_E(0)/\Omega_E(t) \), \( D_E \) calculated for each \( k(=1, 2, \ldots, 13) \), and scatter plot of \( \langle \tau_{L,i}, \tau_{H,i} \rangle \), partitioned into 13 clusters, are shown on the right.
Figure S8: Mg$^{2+}$ pulse-induced transition frequency matrix among 13 kinetically disjoint states based on 148 FRET trajectories. The indexes at the sides of matrix denote the cluster number $k = 1, 2 \ldots 13$. The number in the matrix element represents the number of transitions from state $\xi$ to state $\eta$ induced by the sequence of Mg$^{2+}$ pulses (see Fig.S2 and Fig. 6a).
Figure S9: All the histograms of FRET efficiency, $P(E)$, under 150, 500, 1000 mM monovalent salt condition using NaCl but in the absence of Mg$^{2+}$ ions are uni-modal, suggesting that monovalent ions do not induce the two-state like isomerization dynamics in HJs. This control experiment suggests that isomerization dynamics and heterogeneity of this dynamics are related to specific DNA-Mg binding events, not the screening effects due to the high ionic strength in 50 mM MgCl$_2$ condition.
Figure S10: Control experiment – Mg\(^{2+}\)-pulse experiments using donor-only-tagged HJs. The Mg\(^{2+}\) is switched at 40 s (50 mM → 0 mM) and 65 sec (0 mM → 50 mM). It is of note that although the photon intensity itself varies from molecule-to-molecule and there are weak gradual drift in the data, the overall intensities from the donor dyes are constantly maintained irrespective of the Mg\(^{2+}\) concentration. It is hard to see the change in the pattern of data as in Fig.6a when only the donor dyes are tagged to the molecule. This observation reinforces our hypothesis that the HJ dynamics observed with sm-FRET are originated from the conformational dynamics of HJs, not due to the dye-nucleotide stacking or heterogeneous interaction with surface.