

# TBA

Mike Jones and Andrew L. Ferguson\*

*Pritzker School of Molecular Engineering, University of Chicago, Chicago, Illinois 60637*

E-mail: [andrewferguson@uchicago.edu](mailto:andrewferguson@uchicago.edu)

## Abstract

<https://www.overleaf.com/project/5e9e5110c524b8000192c548>

TEST CITE<sup>1</sup>

## 1 Introduction

DNA motivation

1. Connection to Tokmakoff results
2. Some origami, sensor, other nanotech applications

SRV motivation

1. First time studying a two body problem
2. Compare with other TICA-MSM approaches to DNA
3. Compare with Noe nucleotide paper

## 2 Methods

Simulation set up

Featurization

1. Vamp-2 score for permutable distances is slightly lower but should not contain degenerate information (SI data)
2. Inverse distances have higher vamp-2 score than regular distances (SI data)

SRV validation

1. Plot loss scores and self-consistency check for each sequence (fig 1 and 2)
2. Select number of slow modes (different for each) and lag time (same for each) based on this analysis.

3. Only use off diagonals and add sequence visuals along axes (fig 3)

MSM construction

1. hyperparameters include: number of microstates, lag (use same as SRV), number of macrostates (usually  $n+1$ ) but we use  $n$  states for GC-core and GC-mid. I have not done a full optimization sweep over these parameters, but can explain the motivation for choosing these.
2. Can show comparison data with TICA – timescales converge faster and state assignments tend to be more concrete (although this has varied) (fig 4)

### 3 Discussion

Brief description and diagrams for collective variables:

1. Inverse average shifting distance
2. Fourth basepair distances

AT-all

1. Dynamics dominated by slow shifting modes
2. sources to verify this would be the case
3. Minimal data from Tokmakoff on this, but can look at relative hybridization rates
4. SI Calculations to prove preferred thermo stability of 5' shifted state

GC-end

1. Similar dynamics to AT-all, but with overall lower stability and timescales

2. Difficult to make any claims because the thermodynamics of external mismatches with dangling ends have not been measured experimentally, and 3spn2 does not parameterize these mismatches explicitly
3. According to SantaLucia external mismatches and dangling ends tend to have weakly stabilizing effect, and there are 4-6 intact basepairs for these conformations
4. Might account for unknown signal in 2018 Tokmakoff paper with timescale between the fraying and dehybridization signal, but not sure...

#### GC-core

1. Frayed states requires the fourth base pair to be broken, most partially frayed states are not included in metastable clustering
2. 2nd slow mode correlates well with expected fraying behavior, can sample along this mode and generally find increasing frayed states (although the SRV coords are spiky)
3. 3rd mode is reflective of changes in orientation in dissociated behavior. This could be interesting for the SI, but this state does not inform the macrostate clustering

#### GC-mid

1. Evidence of fraying correlations in higher order modes, but these are not consistent and their inclusion does not generate a metastable third mode
2. Treat this as a two-states model, where fraying is an expected feature of the hybridized state but not does not characterize a metastable state

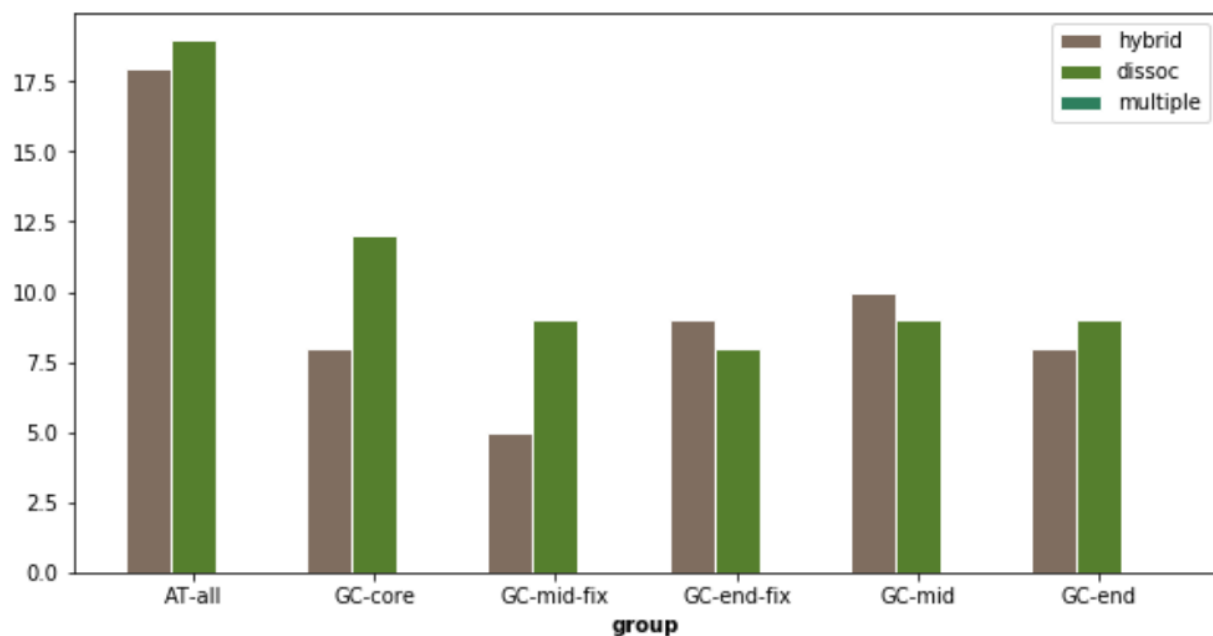


Figure 1: Number of events per sequence. I would only included the four sequences in the actual pub. (this could be supplemental

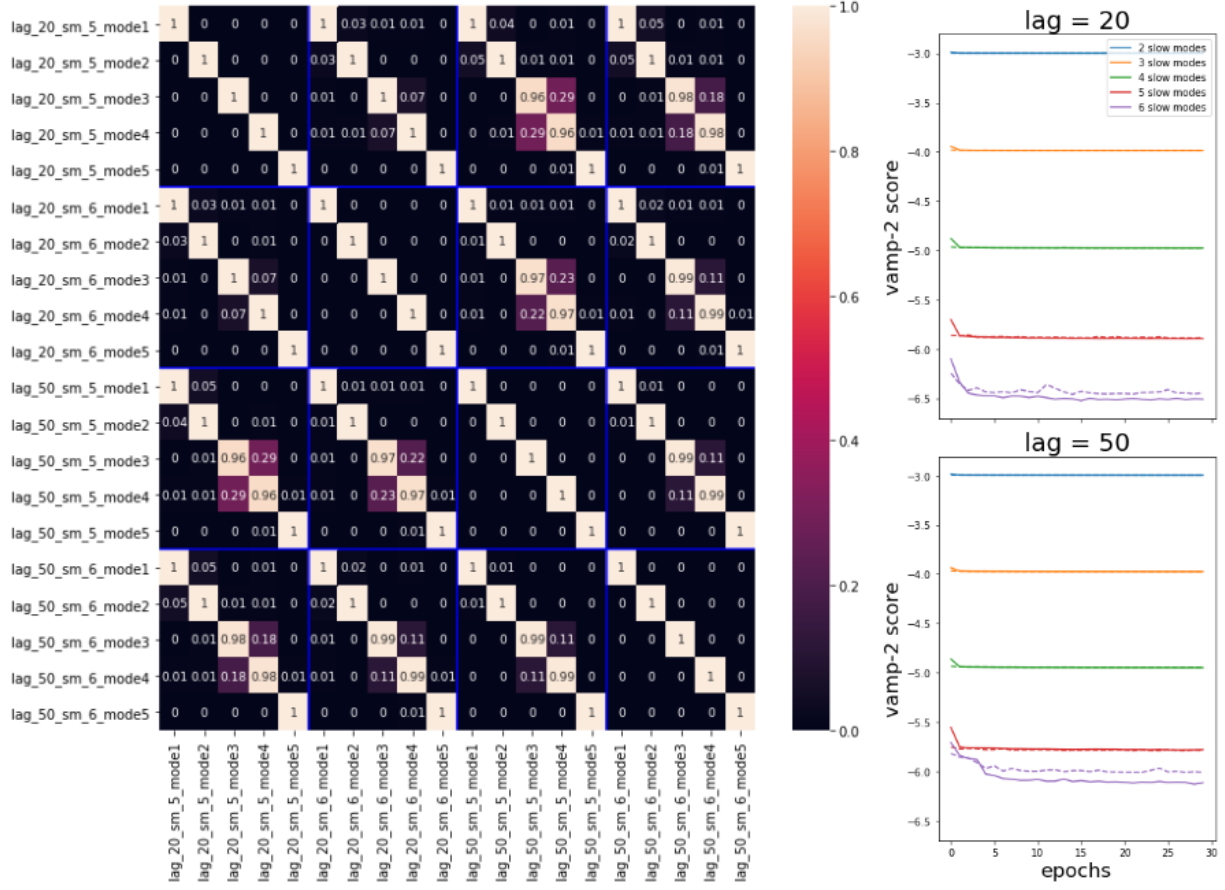


Figure 2: For Supplemental info: Selection process for AT-all slow modes

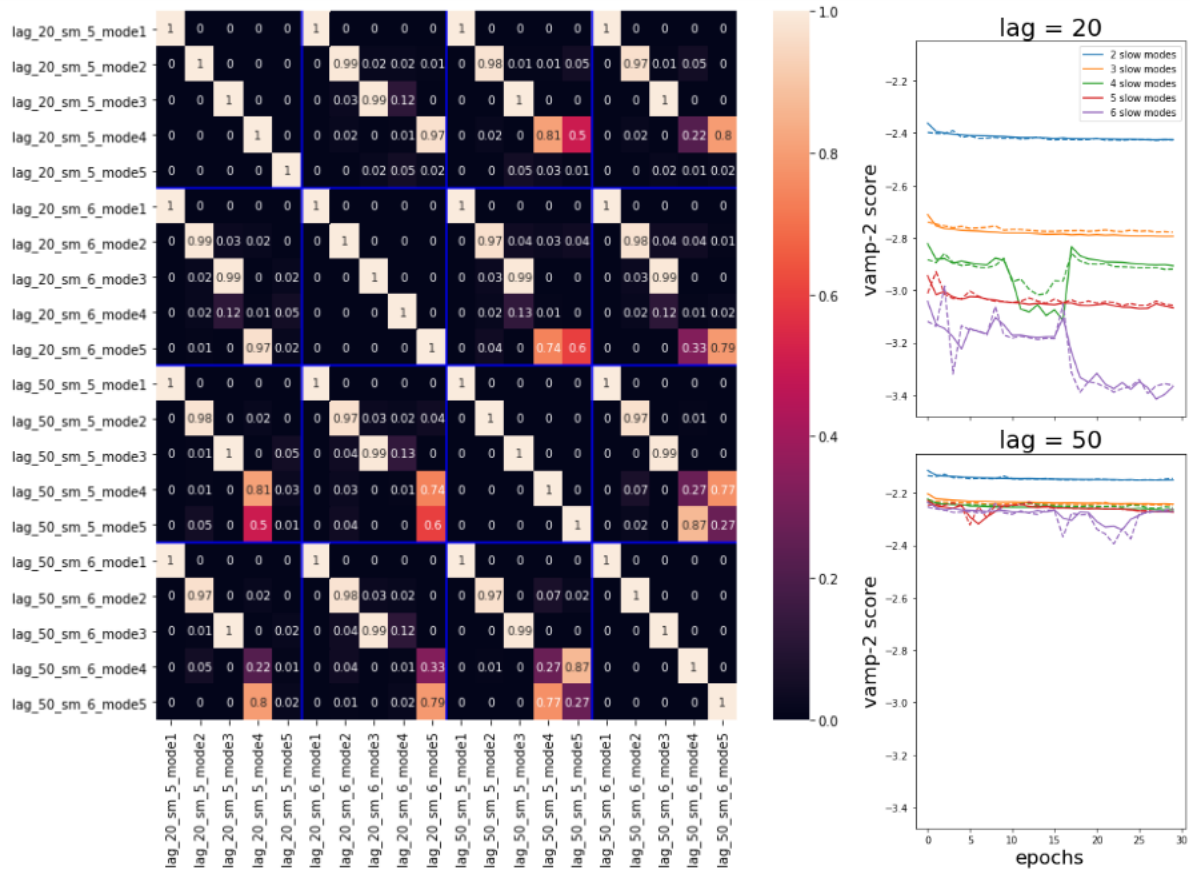


Figure 3: For Supplemental info: Selection process for GC-core slow modes



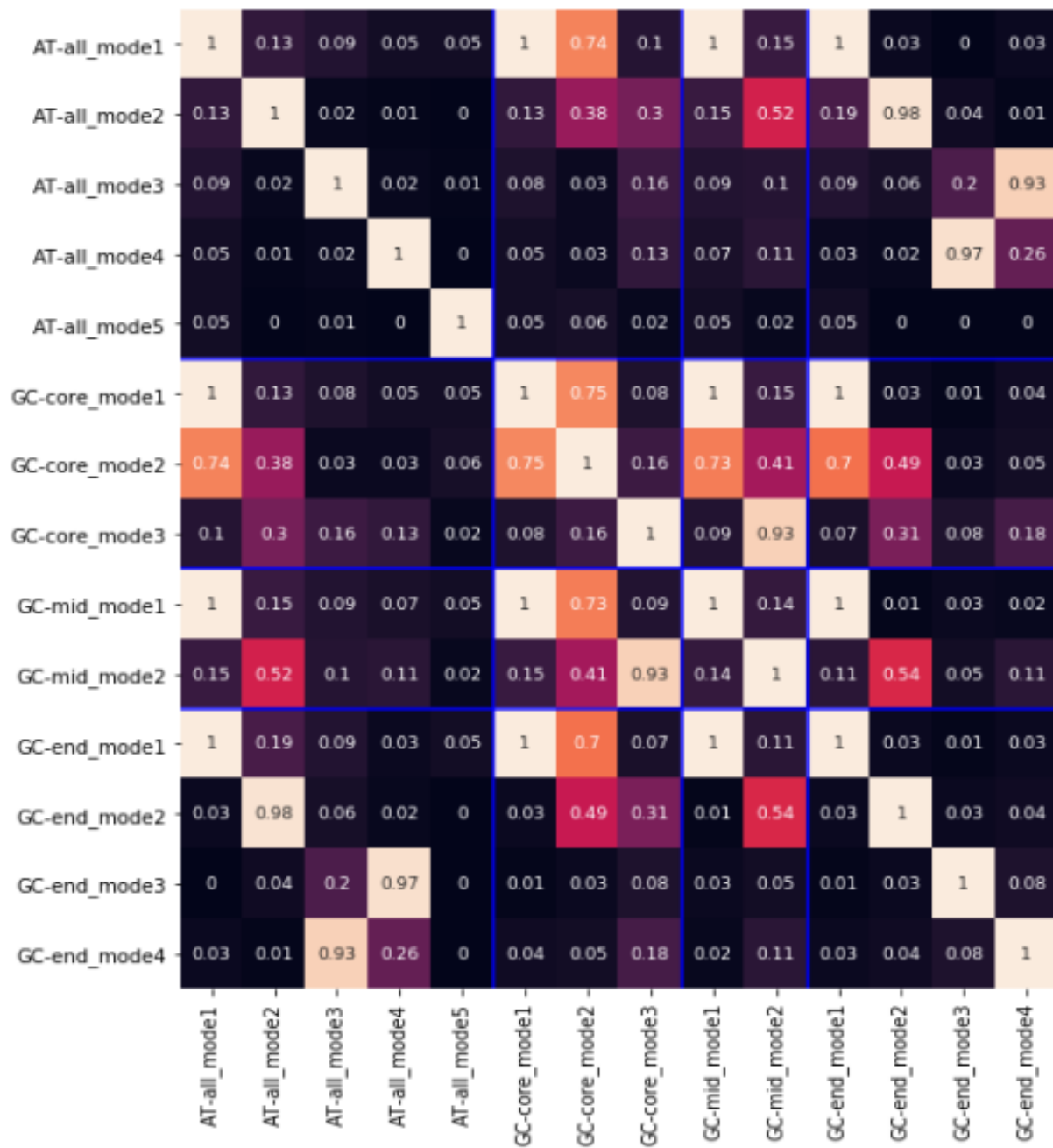


Figure 4: For Supplemental info: Correlations between all leading modes

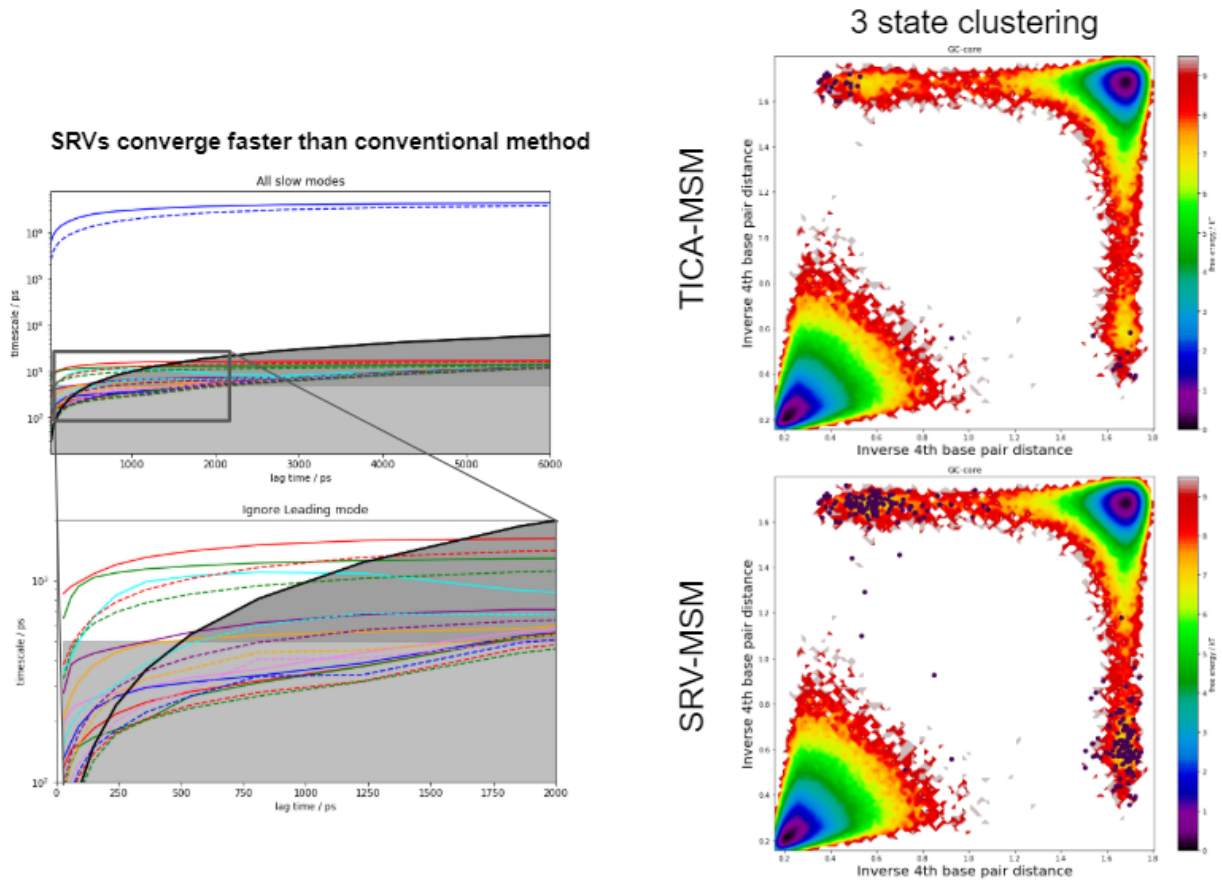


Figure 5: SRV vs. TICA MSMs and resulting cluster memberships

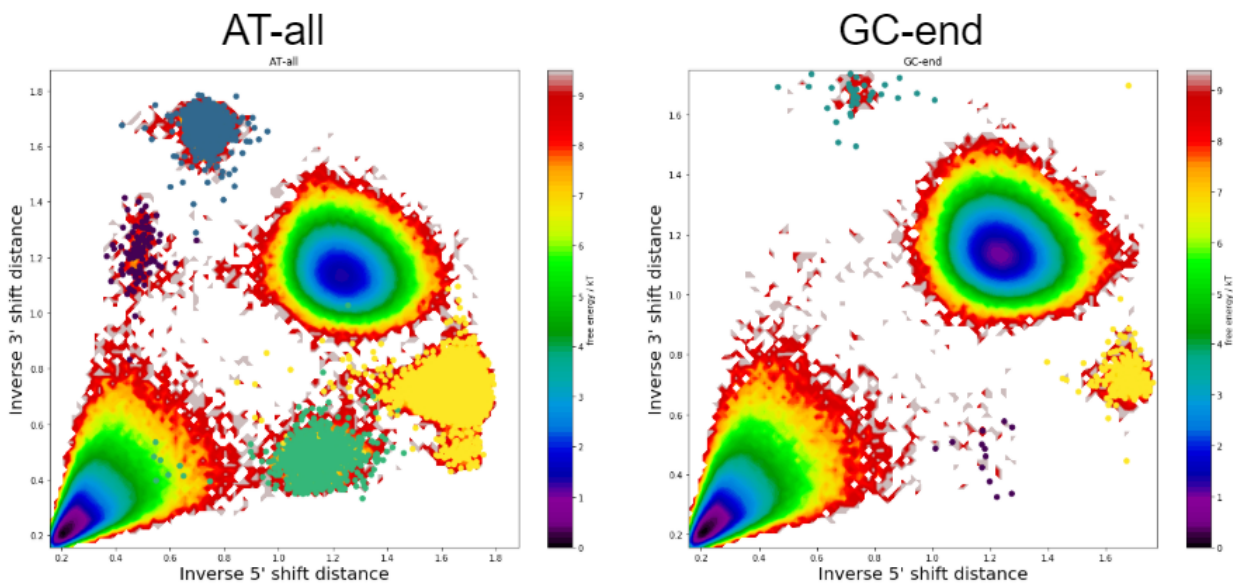


Figure 6: Shows similarities between shifting modes and state memberships for AT-all and GC-ends

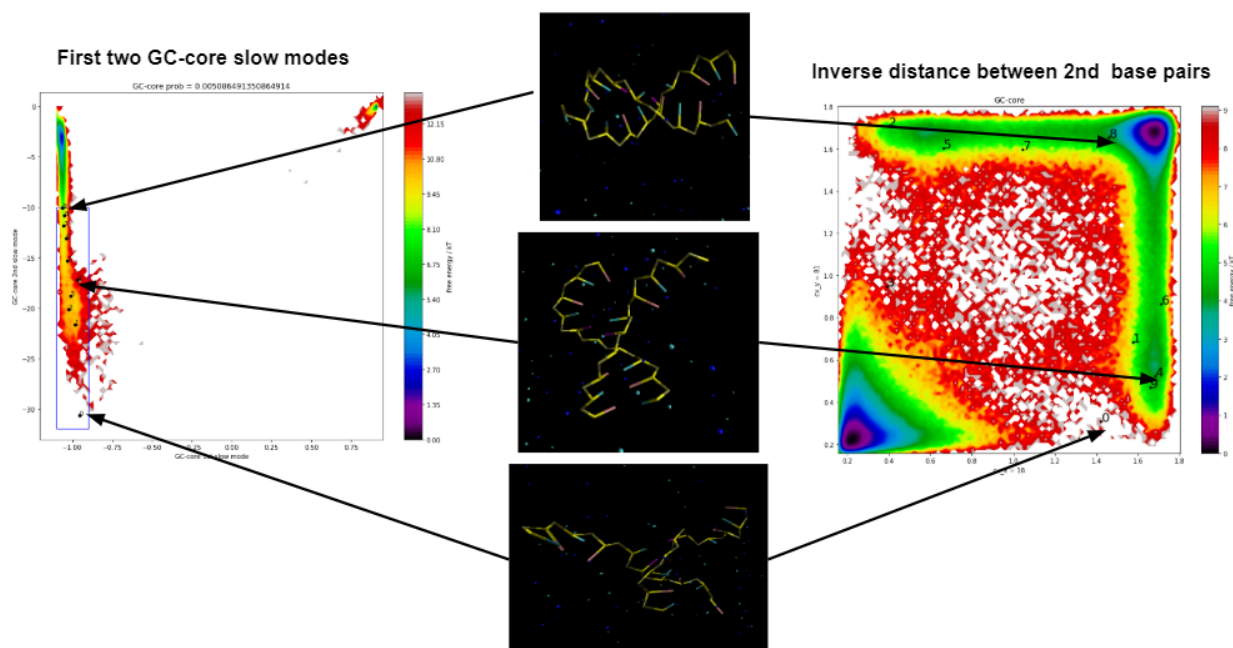


Figure 7: Sample along 2nd GC-core SRV mode and show that these have increasing amounts of fraying

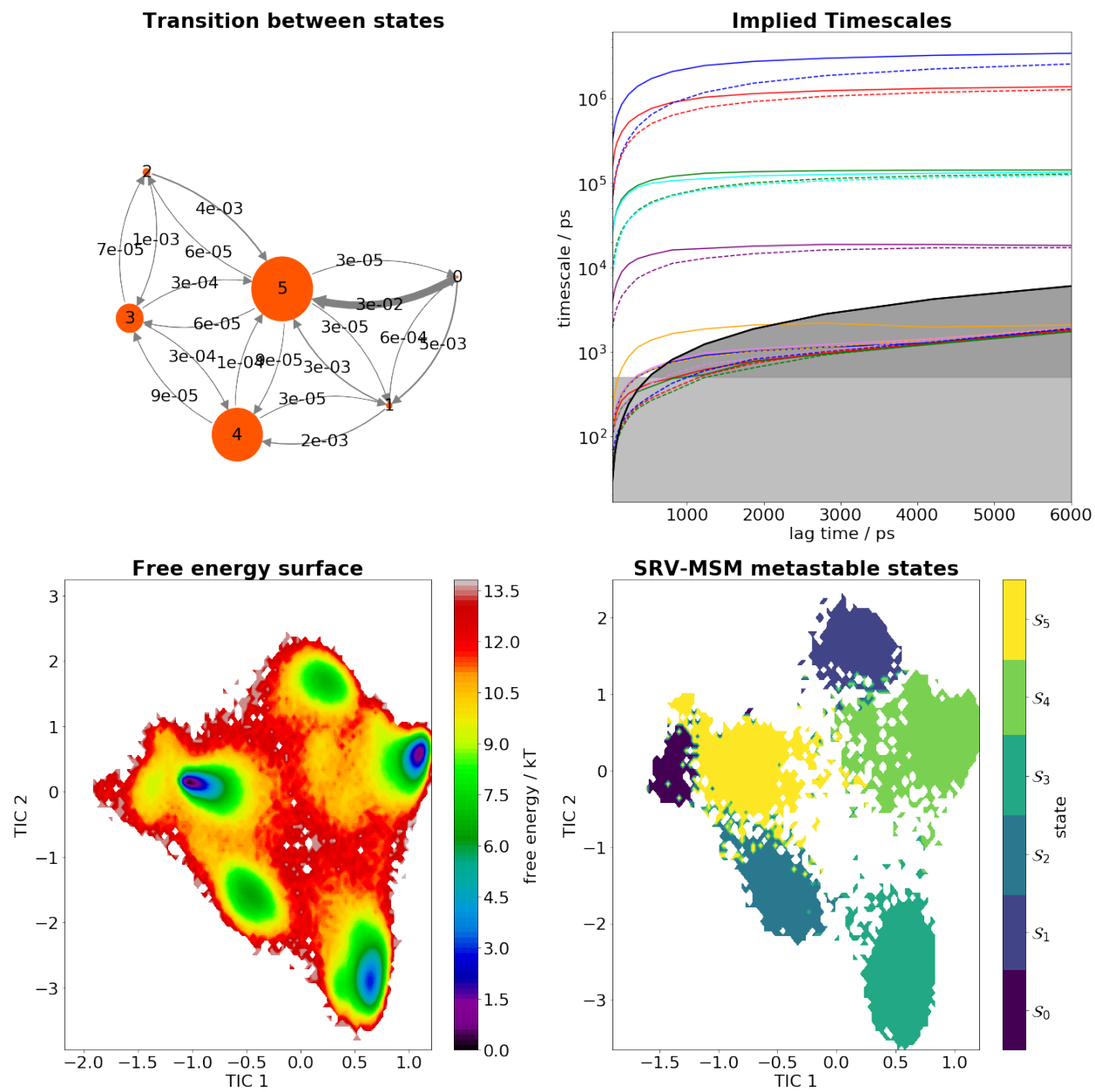


Figure 8: AT-all transition diagram, implied timescales between SRV-MSM and TICA-MSM, free energy surface and metastable state assignment.

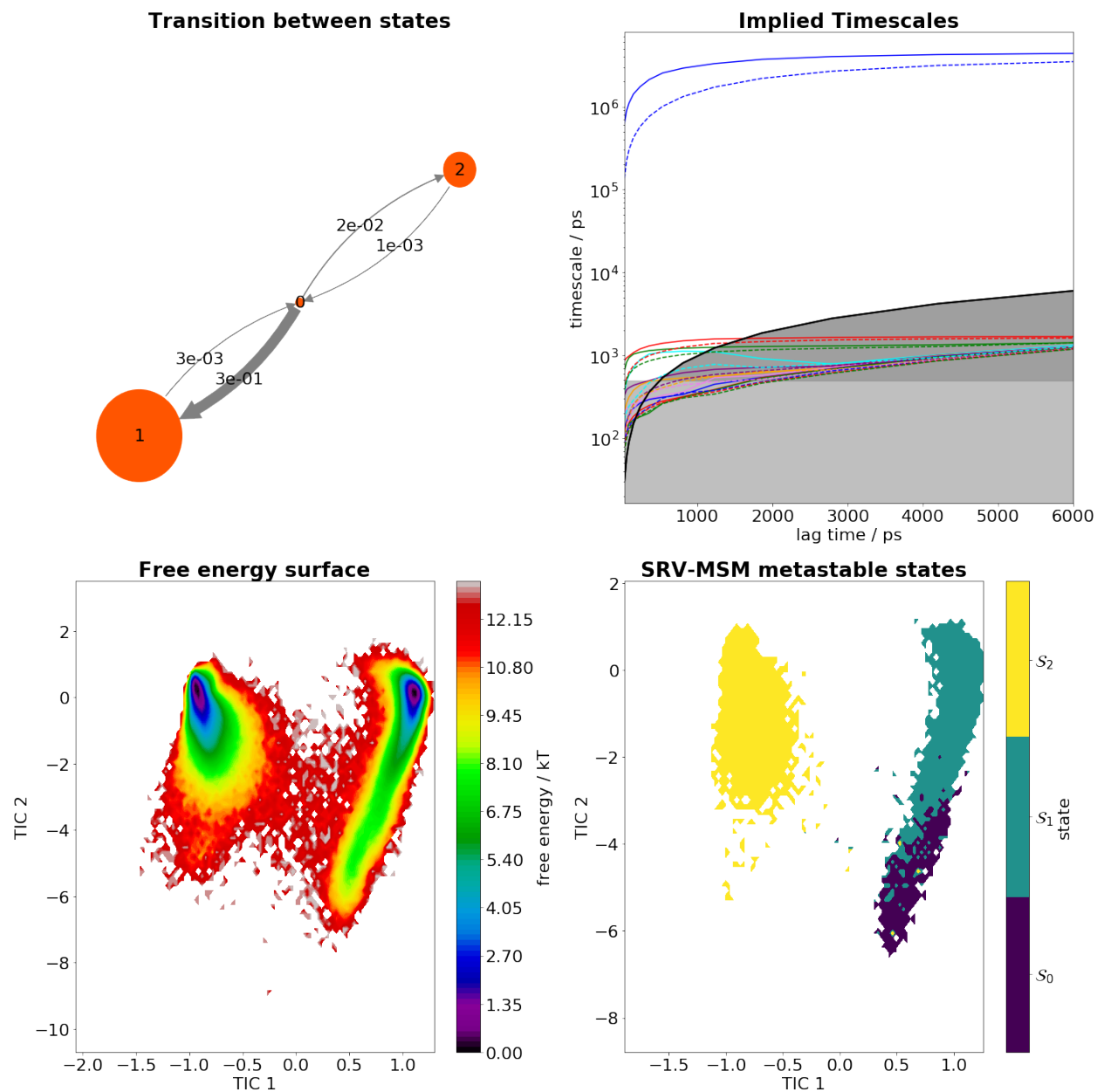


Figure 9: GC-core transition diagram, implied timescales between SRV-MSM and TICA-MSM, free energy surface and metastable state assignment.

## References

- (1) Phys, J. C.; Hinckley, D. M.; Lequieu, J. P.; Pablo, J. J. D. Coarse-grained modeling of DNA oligomer hybridization : Length , sequence , and salt effects. **2014**, *035102*.