Fast Step Transition and State Identification (STaSI) for Discrete Single-Molecule Data Analysis

Bo Shuang
Landes Research Group
Rice University

QBI
Jan. 08, 2015
Acknowledgements

Landes Group
Dr. Christy Landes - PI
Dr. Jixin Chen
Lydia Kisley
David Cooper
Wenxiao Wang

Dr. J. Nick Taylor
Dr. Chun-Biu Li
Dr. Tamiki Komatsuzaki

Link Group
Motivation: When Histogram is not Enough

FRET: Förster (fluorescence) resonance energy transfer

1 Cooper, D. et. al. unpublished image
Bronson, et. al. *Biophys. J.* 2009
• Targeting binned data
• 1) Fitting of different number of states
   – Student’s t test
• 2) Determine the optimum number of states
   – Minimum Description Length Principle (MDL)
Step 1: Step Detection using Student’s $t$-test:

$$R(t_i) = \frac{|I_2(t_{i+1}, t_N) - I_1(t_1, t_i)|}{\sigma \sqrt{\frac{1}{i} + \frac{1}{N-i}}}$$

Pre-selection: Step Detection and State Identification for Different Number of States

**Step 2:** Group similar states using a greedy algorithm

\[
M(i, j) \propto (m_i + m_j) \times I_{i,j}^2 - (m_i \times I_i^2 + m_j \times I_j^2)
\]

Step 3: Determine the Optimum Number of States

$$\text{MDL} = F + G$$

$$F = \frac{\sum_{i=1}^{N} |y(t_i) - y_{fit}(t_i)|}{2\sigma}$$

$$G = \frac{k}{2} \ln \frac{1}{2\pi} + k \ln \frac{V}{\sigma} + \frac{N_{tp}}{2} \ln N + \frac{1}{2} \left( \sum_{i=1}^{k} \ln n_i + \sum_{j=1}^{N_{tp}} \ln \frac{T_j^2}{\sigma^2} \right)$$

---

Performance: More Accurate

Gray: Simulated FRET trace  Red: StaSI fit  Green: vbFRET fit\(^1\)  Blue: True states

Performance: Faster in Computation
The Importance of L1 norm

\[ F = \frac{\sum_{i=1}^{N} |y(t_i) - y_{fit}(t_i)|}{2\sigma} \]

\[ F = \frac{\sum_{i=1}^{N} (y(t_i) - y_{fit}(t_i))^2}{2\sigma^2} \]
Instantaneous diffusion coefficient

Photobleaching steps

The Program with GUI

Status
Done analysis.

View Results
x axis controller
Left Position
Range

Use 5 States
Save
Conclusion

- STaSI provides comprehensive, objective analysis of multiple traces requiring few user inputs about the underlying physical models, and is faster and more precise in determining the number of states than established and cutting-edge methods for single-molecule data analysis.
- This method improves the state determination for noisy data or data with fast dynamics, which potentially increases the temporal resolution in single-molecule measurements.
- We plan to expend this method to data with variance change or frequency change.
- Our program (in MATLAB) can be downloaded at: http://lrg.rice.edu/Content.aspx?id=96
Questions