Model-Free Analysis of Time-Dependent Single-Molecule Spectroscopy

Dynamics of Biological Macromolecules

Kevin Welsher and Haw Yang
Department of Chemistry
Princeton University
Princeton, NJ 08544, USA
E-mail: hawyang@princeton.edu

***Abstract*—Rigorous and unbiased statistical analysis of single-molecule data is now recognized as a crucial component to substantiate any discovery made through such experiments. This presentation summarizes many such tools, and ventures some future directions in the context of the dynamics of biological macromolecules.**

Keywords-Fisher information; maximum information; change point; maximum entropy; protein dyamics; single-molecule FRET

I. INTRODUCTION

The successful detection and spectroscopy of individual molecules [1, 2] has fundamentally changed the way by which scientific and technological problems are approached and resolved. Conceptually, single-molecule spectroscopy enables us to think about complex systems from the most elementary, molecular perspective, completely doing away with the convoluted ensemble averaging. The two unique pieces of information it directly provides (which will be impossible to acquire by any other means)—the distribution of molecular properties and the long-term dynamics of individual molecules—are recognized as the crucial missing links relating the microscopic dynamics of a complex system to its macroscopic transformation. In addition to testing theories and hypotheses, single-molecule spectroscopy holds great promise for new and unexpected discoveries, which can only be substantiated via quantitative and unbiased experimental analyses. This overview summarizes the steps taken by our group toward realizing the full potential of single-molecule spectroscopy. To make the discussion more concrete, the dynamics of biological macromolecules is used as an example to illustrate the application to systems of multiple time and length scales.

II. A CHEMICAL PHYSICS VIEW OF PROTEIN DYNAMICS PROBLEM

Proteins are molecular machines that accomplish and mediate much of cellular functions. In sharp contrast to macroscopic machines, which require a frictionless operation to avoid unwanted energy dissipation, protein machines operate in a highly dissipative environment. Thus, the chemical physics problem can be formulated as follows: How do proteins utilize the dissipative thermal fluctuations, rather than be frustrated by them, to achieve their function? Specifically, one is interested in understanding the *operation* and the *design* principles of protein machines [3]. This is an extremely challenging problem because the dynamics of a protein span from the femtosecond side-chain fluctuation to the seconds inter-domain conformational changes. While no single technique can cover the entire ~15 orders of time scale, time-dependent single-molecule spectroscopy affords following the molecular movements in real time on the sub-millisecond to tens of second time scale, which overlaps with that of enzyme catalysis [4]. It is important to note that real-time monitoring of a single protein transitioning from one state to the other is essential for establishing the *causality* of its functional dynamics. As such, time-dependent single-molecule spectroscopy is anticipated to contribute greatly to solving the protein dynamics problem.

III. CHALLENGES AND THE MAXIMUM-INFORMATION FRAMEWORK FOR CONTINUOUSLY CHANGING STATES

The fundamental difference between single-molecule experiments and the conventional ensemble-averaged experiments is that the experimental observable is time averaged for the former but number averaged for the latter [5]. The requirement to time-average the experimental data makes it very difficult to quantitatively and objectively extract from the data the two most important pieces of information, the distribution and the dynamics, because their values will depend on the time it takes to average. The issues are illustrated in Fig. 1. The apparent distribution and dynamics change with different "bin time," leaving open a great deal of arbitrariness to the choice of the bin width, which in turn is likely to impact on the interpretation.

One way to resolve this challenge was to use the "maximum information" framework [6]. For a time-averaged observable, it is possible to find an analytical expression relating the uncertainty of the observable and the time resolution (the aforementioned bin time) through the use of Fisher information. The relationship allows one to find the optimal time resolution when the observables were estimated using a maximum-likelihood estimator (the Cramér-Rao bound [7, 8]). An important consequence of the maximum-information analysis is that each observation (now averaged

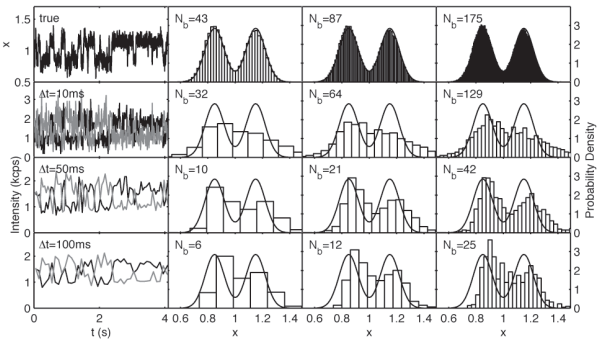


Figure 1. The apparent dynamics and state distribution are affected by the bin time used in the analysis of single molecule time traces. The first row shows the "true" distance trajectory of a simulated two-state system stochastically switching between two conformations. The second to the fourth rows display the apparent intensity traces (first column) from a single-molecule Förster-type resonance transfer experiment at different bin times, as well as the resulting apparent distributions with different histogram width (columns 2-4). (Reproduced from [9] with permission from the American Chemical Society.)

over a maximum-information time period) would now have uncertainties, providing a way of quantitatively evaluating and following the dynamic changes of the observable.

The maximum-information treatment of time-dependent single-molecule data readily lends itself to a model-free evaluation of the *distribution* [9]. This is because the uncertainty of a maximum-likelihood estimate is asymptotically Gaussian which, in turn, allows for the use of a Gaussian-kernel density estimate to evaluate the distribution of the observable in a non-parametric way. The resulting distribution (particularly the width) inevitably includes the measurement uncertainties, primarily arising from the Poisson photon-counting noise. The uncertainties can be removed quantitatively using the maximum-entropy method without the use of any distribution model. It is important to note that the maximum-entropy deconvolution is valid only when the data curve is smooth (in this case, the distribution estimated by the Gaussian-kernel method). The quantitatively extracted distribution has been used to show that the flexural rigidity along a series of poly-proline molecules is contour-length dependent [10].

As mentioned earlier, the shape of the distribution will depend on the time resolution with which the distribution is constructed, especially when the time scale of the molecular dynamics is commensurate with the experimental average time scale. The distribution becomes sharper as the time resolution becomes lower (longer average time) because the dynamical information is lost—a phenomenon known in NMR as motional narrowing [11, 12]. The kinetic rates of two-state conformational changes can be extracted by analyzing the manner by which the noise-removed distribution changes with time resolution [13, 14].

The methods discussed above allow the quantitative evaluation of the distribution and dynamics from time-dependent single-molecule spectroscopic data. Applying these unbiased, data-driven tools to the study of protein dynamics has enabled us to uncover several unexpected operation principles. For example, the conformational transition rate of a protein was found to increase upon interaction with its substrate [15]. Furthermore, the partial folding and unfolding of a local segment in a protein was found to play a critical role in controlling the conformation-transition rates [16]. Additional operation principles for proteins are expected to emerge as these tools are applied to a wider array of different systems. A viable strategy to uncover the design principles for protein dynamics, especially with the use of single-molecule spectroscopy, will be the next challenge.

IV. The Change-Point Method for Abruptly Changing States

In cases when the time scale of molecule-state change is much faster than the experimental data acquisition time, the transitions will show up in the single-molecule data time trace as abrupt changes. Examples include the electron transfer / trapping step in single quantum dots (blinking) or the chemical step in a catalytic system (burst). The conventional way of determining these transitions has been to use some threshold to decide which state the molecule is in. This intuitive way of analyzing the data, however, may introduce serious artifacts [17, 18].

The change-point method is a data-driven method for determining and locating the abrupt changes [19, 20]. It uses a generalized likelihood-ratio test to rigorously determine both the location of the state changes and the observable value for the state—thus both the distribution and dynamics can be quantitatively determined. Since the change-point method does not require any kinetic model or adjustable parameters, it provides the researcher with an objective evaluation of the data. The use of the change-point method has enabled, for instance, unraveling of the continuous-state nature of single-quantum dot emission [21], observation of the memory effects of organic chromophore blinking [22], and resolution of the conformation-coupled photo-dynamics of single fluorescent proteins [23]. The change-point method appears to be more general than it was originally intended, as a recent paper [18] shows that it can handle the burst-detection problem [24] as well. This encouraging result hints upon the possibility to correlate enzyme conformation dynamics (through single-molecule Förster-type energy transfer) with catalytic reactivity (through burst analysis of fluorescent product, for example.

V. Outlook

The study of molecular interactions in biological systems has long been hampered by the inability to observe molecular phenomena on their native length scale (nanometers) and on a time scale relevant to molecular dynamics (sub-milliseconds), while simultaneously placing these interactions and dynamics in their macromolecular biological context, which require a minimal length scale of microns and time scale of seconds to minutes. As the cellular environment is expected to be highly heterogeneous both temporally and spatially, the combination of imaging with single-molecule *spectroscopy*, while yet achieved, is recognized as an important step forward [25, 26].

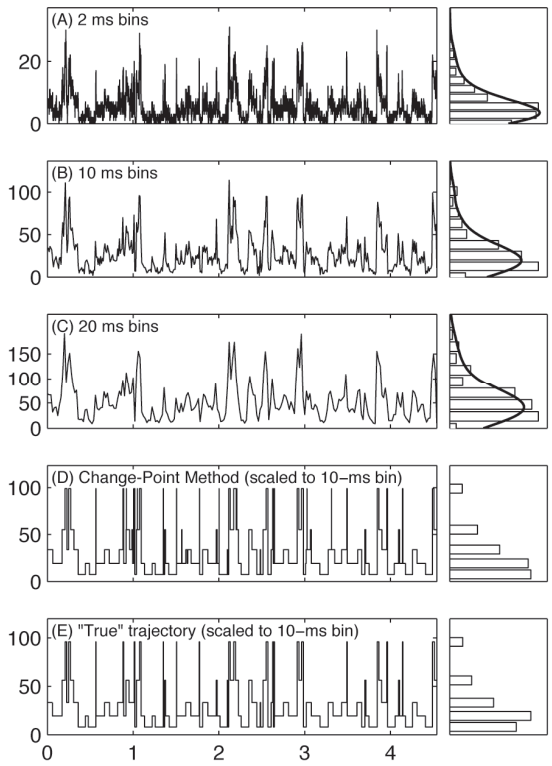


Figure 2. Simulated single-molecule time traces with abrupt changes, illustrating the challenges in extracting the dynamical changes and the state distribution from such data (panels A-C). Also shown is how the change-point method (panel D) can recover the "true" dynamics (panel E). (Reproduced from [19] with permission from the American Chemical Society.)

With the quantitative analysis tools outlined above, and the capacity to perform time-dependent spectroscopy on a single nanoprobe while it is moving in three dimensions [27-30], the contextual single-molecule spectroscopy and imaging is expected to be within reach.

REFERENCES

[1] W. E. Moerner and L. Kador, "Optical-Detection and Spectroscopy of Single Molecules in a Solid," *Phys. Rev. Lett.,* vol. 62, pp. 2535-2538, May 22 1989.

[2] M. Orrit and J. Bernard, "Single Pentacene Molecules Detected by Fluorescence Excitation in a Para-Terphenyl Crystal," *Phys. Rev. Lett.,* vol. 65, pp. 2716-2719, Nov 1990.

[3] H. Yang, "The Orientation Factor in Single-Molecule Förster-Type Resonance Energy Transfer with Examples for Conformational Transitions in Proteins," *Israel J. Chem.,* vol. 49, pp. 313-322, 2009.

[4] Y. W. Tan and H. Yang, "Seeing the forest for the trees: fluorescence studies of single enzymes in the context of ensemble experiments," *Phys. Chem. Chem. Phys.,* vol. 13, pp. 1709-1721, 2011.

[5] H. Yang, "Change-Point Localization and Wavelet Spectral Analysis of Single-Molecule Time Series," in *Single-Molecule Biophysics: Experiment and Theory*. vol. 146, T. Komatsuzaki, M. Kawakami, S. Takahashi, H. Yang, and R. J. Silbey, Eds., ed New York: Wiley, 2011, pp. 129-143.

[6] L. P. Watkins and H. Yang, "Information bounds and optimal analysis of dynamic single molecule measurements," *Biophys. J.,* vol. 86, pp. 4015-4029, Jun 2004.

[7] H. Cramér, *Mathematical Methods of Statistics*. Princeton, NJ: Princeton University Press, 1946.

[8] C. R. Rao, "Sufficient Statistics and Minimum Variance Estimates," *Proc. Camb. Philos. Soc.,* vol. 45, pp. 213-218, 1949.

[9] L. P. Watkins, H. Chang, and H. Yang, "Quantitative Single-Molecule Conformational Distributions: A Case Study with Poly-L-proline," *J. Phys. Chem. A,* vol. 110, pp. 5191-5203, 2006.

[10] J. A. Hanson, J. Brokaw, C. C. Hayden, J.-W. Chu, and H. Yang, "Structural distributions from single-molecule measurements as a tool for molecular mechanics," *Chem. Phys. (in press),* p. http://dx.doi.org/10.1016/j.chemphys.2011.06.014, 2011.

[11] R. Kubo, "Note on the Stochastic Theory of Resonance Absorption," *J. Phys. Soc. Jpn.,* vol. 9, pp. 935-944, 1954.

[12] P. W. Anderson, "A Mathematical Model for the Narrowing of Spectral Lines by Exchange or Motion," *J. Phys. Soc. Jpn.,* vol. 9, pp. 316-339, 1954.

[13] E. Geva and J. L. Skinner, "Two-state dynamics of single biomolecules in solution," *Chem. Phys. Lett.,* vol. 288, pp. 225-229, May 22 1998.

[14] I. Gopich and A. Szabo, "Theory of photon statistics in single-molecule Forster resonance energy transfer," *J. Chem. Phys.,* vol. 122, p. 014707, Jan 1 2005.

[15] J. A. Hanson, K. Duderstadt, L. P. Watkins, S. Bhattacharyya, and H. Yang, "Illuminating the Mechanistic Roles of Enzyme Conformational Dynamics," *Proc. Natl. Acad. Sci. U.S.A.,* vol. 104, pp. 18055-18060, 2007.

[16] E. M. Flynn, J. A. Hanson, T. Alber, and H. Yang, "Dynamic Active-Site Protection by the M. tuberculosis Protein Tyrosine Phosphatase PtpB Lid Domain," *J. Am. Chem. Soc.,* vol. 132, pp. 4772-4780, 2010.

[17] C. H. Crouch, O. Sauter, X. H. Wu, R. Purcell, C. Querner, M. Drndic, and M. Pelton, "Facts and Artifacts in the Blinking Statistics of Semiconductor Nanocrystals," *Nano Lett.,* vol. 10, pp. 1692-1698, May 2010.

[18] T. G. Terentyeva, H. Engelkamp, A. E. Rowan, T. Komatsuzaki, J. Hofkens, C.-B. Li, and K. Blank, "Dynamic Disorder in Single-Enzyme Experiments: Facts and Artifacts," *ACS Nano, article ASAP,* 2012.

[19] L. P. Watkins and H. Yang, "Detection of intensity change points in time-resolved single-molecule measurements," *J. Phys. Chem. B,* vol. 109, pp. 617-628, Jan 13 2005.

[20] D. Montiel, H. Cang, and H. Yang, "Quantitative Characterization of Changes in Dynamical Behavior for Single-Particle Tracking Studies," *J. Phys. Chem. B,* vol. 110, pp. 19763-19770, 2006.

[21] K. Zhang, H. Chang, A. Fu, A. P. Alivisatos, and H. Yang, "Continuous Distribution of Emission States from Single CdSe/ZnS Quantum Dots," *Nano Lett.,* vol. 6, pp. 843-847, 2006.

[22] K. L. Wustholz, E. D. Bott, B. Kahr, and P. J. Reid, "Memory and Spectral Diffusion in Single-Molecule Emission," *J. Phys. Chem. C,* vol. 112, pp. 7877-7885, 2008.

[23] R. H. Goldsmith and W. E. Moerner, "Watching conformational- and photodynamics of single fluorescent proteins in solution," *Nat. Chem.,* vol. 2, pp. 179-186, Mar 2010.

[24] K. Zhang and H. Yang, "Photon-by-photon determination of emission bursts from diffusing single chromophores," *J. Phys. Chem. B,* vol. 109, pp. 21930-21937, Nov 2005.

[25] S. J. Lord, H. L. D. Lee, and W. E. Moerner, "Single-Molecule Spectroscopy and Imaging of Biomolecules in Living Cells," *Anal. Chem.,* vol. 82, pp. 2192-2203, Mar 2010.

[26] H. Yang, "Progress in single-molecule spectroscopy in cells," *Curr. Opin. Chem. Biol.,* vol. 14, pp. 3-9, 2010.

[27] H. Cang, C. M. Wong, C. S. Xu, A. H. Rizvi, and H. Yang, "Confocal Three Dimensional Tracking of A Single Nanoparticle with Concurrent Spectroscopic Readouts," *Appl. Phys. Lett.,* vol. 88, p. 223901, 2006.

[28] H. Cang, C. S. Xu, D. Montiel, and H. Yang, "Guiding a Confocal Microscope by a Single Fluorescent Nanoparticle," *Opt. Lett.,* vol. 32, pp. 2729-2731, 2007.

[29] C. S. Xu, H. Cang, D. Montiel, and H. Yang, "Rapid and Quantitative Sizing of Nanoparticles Using Three-Dimensional Single-Particle Tracking," *J. Phys. Chem. C,* vol. 111, pp. 32-35, 2007.

[30] H. Cang, D. Montiel, C. S. Xu, and H. Yang, "Observation of Spectral Anisotropy of Gold Nanoparticles," *J. Chem. Phys.,* vol. 129, p. 044503, 2008.