

## Preface

Steady progress in experimental techniques and imaging capabilities has made the observation of individual atoms and molecules a routine practice in many laboratories worldwide. Despite this fact, single-molecule measurements remain difficult to perform and interpret. The exceedingly weak signals arising from individual molecules are more readily degraded and distorted by macroscopic-scale noise sources than are most ensemble averaged signals. We accept these difficulties since single-molecule experiments offer qualitatively new information about the behavior of physical systems, information unobtainable by traditional means. Observations of individual molecules, nanoparticles and other nano-objects reveal the full extent of their heterogeneity, be it for enzymes at physiological conditions, or for fluorescent probes in supercooled molecular liquids. Even more importantly, single-molecule signals open a window to directly observe dynamical fluctuations at the nanoscale.

A variety of modern techniques providing single-molecule sensitivity have been reviewed in the literature. The first methods to observe single molecules were scanning probe microscopies: STM (scanning tunnelling microscopy)<sup>1</sup> and AFM (atomic force microscopy).<sup>2</sup> In biophysics, electrical conductivity measurements can monitor the function of single ion-channels.<sup>3</sup> More recently, optical spectroscopy and fluorescence microscopy<sup>4,5</sup> have reached single-molecule sensitivity. In spite of the differences in the experimental approaches, the obtained single-molecule data share the same characteristic features; the data is noisy and contains large and irreproducible fluctuations. Depending on the experiment, these fluctuations may be of primary interest or may frustratingly serve to obscure a more interesting dynamics. Stated differently, single-molecule data streams are inherently stochastic — a given single molecule data

trajectory is apparently random and is irreproducible. However, as we may extract statistically meaningful information from these data streams, their fundamental irreproducibility is an asset as well as a curse. While complex data treatments, analysis methods, statistical indicators, etc., are required to ensure the robustness of our conclusions, the information gained is unique and is inaccessible by ensemble-averaged measurements.

Statistical methods of data analysis applied to single molecules should, to the extent possible, be unbiased and model-free. They should eliminate the effects of noise, while retaining the significant features essential to understanding molecular dynamics. A number of procedures are currently in use to aid in this analysis, as summarized in Refs. 6 and 7. The simplest way to evaluate the time-trace of a single-molecule signal is to bin the signals by a low-frequency pass filter, and compare the average signal value to thresholds. This works well for slow variations, but the choice of bins and thresholds is somewhat arbitrary. More systematic analysis tools, either time-dependent (such as delay distributions, lifetimes, correlation functions) or signal-strength dependent (such as photon-counting histograms) do not involve any arbitrary parameter, but do average the data in ways which can sometimes erase the useful information. Finally, simple statistical indicators can sometimes answer a question about the data. Examples are the Mandel parameter indicating deviations from Poisson statistics in photon-counting traces, or indicators of the renewal character of series of events. (Is the data consistent with all events having been drawn from the same probability distribution?) Of course, an ideal analysis scheme would let the data “speak for itself”, i.e. determine a model for the molecular dynamics based solely on the collected data without prior hypotheses influencing the final interpretation. This is seldom possible.

Since it is typically not possible to directly invert single-molecule data into a complete model for molecular dynamics, it is important to be able to model single-molecule experiments theoretically and computationally. Theoretical predictions for a given model may be compared to observables obtained experimentally, which allows for the refinement of hypothetical models via the scientific process. While most of the models currently used in the description of single-molecule experiments are based on traditional physical pictures (kinetics, diffusion, quantum mechanics, etc.), the observables

available in single-molecule measurements are inherently different from those found in ensemble-averaged experiments. The tools used to predict ensemble-averaged experimental observables for a given model's dynamics are typically not well suited to determining the statistical information obtained via single-molecule experiments.

The aim of this book is to discuss some of the theoretical approaches to single-molecule data evaluation and interpretation developed over the last 10 years. We focus on fluorescence and mechanical measurements. We do not explicitly consider the case of ion channels, which have been considered in some detail previously.<sup>8,9</sup> The following chapters consider specific theoretical and statistical problems unique to single-molecule systems and single-molecule experiments. These problems range from the conditioning of single-molecule data to enable effective analysis, to the analysis of single-molecule data streams to the modelling of single-molecule systems by simulation and/or analytical theory.

Faced with a raw data stream, the first problem encountered in analysis is to identify which features are statistically significant. For example, how many distinct levels or states are resolvable in fluorescence traces or mechanical displacement measurements. Yang (Chapter 1) explains how to apply maximum likelihood criteria to objectively assess experimental time series.

A similar problem is discussed by Plakhotnik in Chapter 2. In particular, he outlines the procedure of Bayesian probability analysis to estimate the probability that a given hypothesis or model is correct, given experimental data and additional background information. This contribution stresses the importance of a proper statistical analysis in single-molecule measurements, where this step is much more crucial than with ensembles.

In Chapter 3, Brown presents the generating function formalism, which allows for the calculation of single-molecule experimental observables using tools that are only slightly modified from traditional ensemble-averaged theories. The generating function method applies to kinetic, stochastic and quantum-mechanical dynamics schemes.

A similar approach is proposed in Chapter 4 by Sanda and Mukamel to access multiphoton event probabilities and multipoint correlation functions, which are connected to the susceptibilities found in nonlinear optics. Here too, the formalism may be applied

to classical kinetic schemes or to quantum-mechanical problems described by a Liouville equation. Sanda and Mukamel also discuss how to extend the formalism to non-Markovian and continuous-time random walk processes.

The contribution of Hummer and Szabo (Chapter 5) discusses thermodynamic and kinetic properties of single molecules in mechanical force measurements. They show how to extract information on the free energy profile through non-equilibrium measurements, and how potential energy landscapes translate into time-dependent rates for breaking or unfolding reactions.

A common result of single-molecule measurements is long time traces, in which transitions between different levels can be distinguished with suitable methods. The next step in the evaluation is to interpret these changes as random transitions between states, for example between conformational substates of a protein. Describing these transitions as Markovian jumps between levels leads to kinetic schemes, which can be evaluated theoretically to derive statistical properties for comparison to experimental data.

The next chapters focus on relating experimental results to classical kinetic schemes. Gopich and Szabo (Chapter 6) show how the kinetic-scheme formulation can be generalized to include the many processes present in a single-molecule measurement. Among these processes are: the complex photophysics of fluorescence, conformational evolution of single molecules, and translational diffusion of the molecules through the focus of the exciting beam. They show how experimental observables — the photon counting intensity distribution, waiting time distribution and related correlation functions — can be generated in a consistent way accounting for all these effects.

Cao (Chapter 7) examines the general solution of kinetic schemes represented by rate matrices and focuses on specific features in these solutions, such as echoes, and other indicators, which can directly give *a priori* information on the system, without the need for a full solution.

Kinetic schemes obeying detailed balance give rise to exponential relaxation, as do overdamped stochastic models. Several experiments have found surprising oscillating behavior in the dynamics of single protein molecules, which suggest that these simple models may be inappropriate for the description of proteins. For example Lu *et al.*<sup>10</sup> found accumulations of diagonal correlations in the consecutive

on-times of cholesterol oxidase enzyme, Edman and Rigler found oscillations in the fluorescence of labeled horse radish peroxidase,<sup>11</sup> and Baldini *et al.* in the dynamics of fluorescent proteins excited with two photons.<sup>12</sup> Vlad and Ross, in Chapter 8, discuss two possible theoretical models that may explain these intriguing experimental results.

Chapter 9, by Kolomeisky, considers the motion of molecular motors. The stepping of the motors is represented as a sequence of transitions in an open kinetic scheme, which leads to prediction of their dynamical and kinetic properties.

In the preceding chapters, the complex dynamics of biomolecules is described by kinetic schemes, i.e. a network of states connected with rate constants. In most cases, however, even the topology of this network is unknown. The determination of the relevant scheme from experimental data alone is difficult, even impossible because the same experimental data can correspond to several kinetic schemes with different topologies and connectivities. This follows from the information loss taking place when the molecule's evolution in a multidimensional space is projected onto the few dimensions accessible to experiment (e.g. intensity in the case of a fluorescence trace). In Chapter 10, Flomenbom and Silbey discuss how to classify kinetic schemes and reduce them to canonical forms, so as to select the simplest class of schemes compatible with the data.

Single molecules provide unique insights into heterogeneities and fluctuation phenomena which cannot be seen directly in ensemble measurements. As Barkai discusses in Chapter 11, ergodicity breaking is a perfect example of a property which directly appears from the behavior of blinking quantum dots and molecules in disordered systems. Ergodicity is broken when the time-average of a single molecule's property, e.g. its emission rate, does not coincide with the ensemble average of this property over many equivalent molecules. These differences, obviously stand out in single-molecule data. Ergodicity breaking can be strong, when each individual molecule explores only a very small, disconnected region of the phase space, or weak, when a large part of the phase space is explored by each molecule, but sojourn times in microstates of the system obey fractal power law statistics.

Statistical analysis, modeling and understanding of single-molecule fluorescence and mechanical experiments have expanded

spectacularly in scope and power during the last ten years. We hope that the theoretical chapters collected in this book will be of use to experimentalists as well as to theoreticians, and will facilitate the necessary dialogue between them. Only a back-and-forth interaction between experiments and theory can lead to a deeper understanding of fluorescence intermittency, conformational fluctuations, and many other fascinating signatures of individual molecules and nano-objects.

**E. Barkai, F. Brown, M. Orrit and H. Yang**

## References

1. C. Julian Chen, *Introduction to Scanning Tunneling Microscopy*, 2nd edn. (Oxford University Press, 2007).
2. J. Howard, *Mechanics of Motor Proteins and the Cytoskeleton* (Sinauer Press, Sunderland, Massachusetts, 2001).
3. *Single-Channel Recording*, B. Sackmann and E. Neher, eds. (Plenum, New York, 1995).
4. *Single-Molecule Optical Detection, Imaging and Spectroscopy*, Th. Basché, W. E. Moerner, M. Orrit and U. P. Wild, eds. (Wiley-VCH, 1997).
5. *Single-Molecule Detection in Solution*, Ch. Zander, J. Enderlein and R. A. Keller, eds. (Wiley-VCH, 2002).
6. M. Lippitz, F. Kulzer and M. Orrit, Fluctuating fluorescence of single nano-objects, *Chem. Phys. Chem.* **6** (2005) 770.
7. E. Barkai, Y. Jung and R. Silbey, Theory of single molecule spectroscopy: Beyond the ensemble average, *Ann. Rev. Phys. Chem.* **55** (2004) 457.
8. D. Colquhoun and F. J. Sigworth, Fitting and statistical analysis of single channel records, in *Single-Channel Recording*, Chap. 19, B. Sackmann and E. Neher, eds. (Plenum, New York, 1995), pp. 483–587.
9. L. Venkataramanan and F. J. Sigworth, Applying hidden Markov models to the analysis of single ion channel activity, *Biophys. J.* **82** (2002) 1930–1942.
10. H. P. Lu, Y. Xun and X. S. Xie, Single-molecule enzymatic dynamics, *Science* **282** (1998) 1877.
11. L. Edman and R. Rigler, Memory landscapes of single-enzyme molecules, *Proc. Nat. Acad. Sci. USA* **97** (2000) 8266.
12. G. Baldini, F. Cannone and G. Chirico, Pre-unfolding resonant oscillations of single green fluorescent protein molecules, *Science* **309** (2005) 1096.