New Frontiers in Ultrasensitive Bioanalysis

Advanced Analytical Chemistry Applications in Nanobiotechnology, Single Molecule Detection, and Single Cell Analysis

Edited by

XIAO-HONG NANCY XU

Old Dominion University

Wiley-Interscience
A John Wiley & Sons, Inc., Publication
New Frontiers in Ultrasensitive Bioanalysis
CHEMICAL ANALYSIS

A SERIES OF MONOGRAPHS ON ANALYTICAL CHEMISTRY AND APPLICATIONS

Editor
J.D. WINEFORDNER
VOLUME 172

A complete list of the titles in this series appears at the end of this volume.
New Frontiers in Ultrasensitive Bioanalysis

Advanced Analytical Chemistry Applications in Nanobiotechnology, Single Molecule Detection, and Single Cell Analysis

Edited by

XIAO-HONG NANCY XU

Old Dominion University

Wiley-Interscience
A John Wiley & Sons, Inc., Publication
CONTENTS

PREFACE vii

CONTRIBUTORS xi

CHAPTER 1. IS ONE ENOUGH? 1
Andrew C. Beveridge, James H. Jett, and Richard A. Keller

CHAPTER 2. DISSECTING CELLULAR ACTIVITY FROM SINGLE GENES TO SINGLE mRNAs 29
Xavier Darzacq, Robert H. Singer, and Yaron Shav-Tal

CHAPTER 3. PROBING MEMBRANE TRANSPORT OF SINGLE LIVE CELLS USING SINGLE-MOLECULE DETECTION AND SINGLE NANOPARTICLE ASSAY 41
Xiao-Hong Nancy Xu, Yujun Song, and Prakash Nallathamby

CHAPTER 4. NANOPARTICLE PROBES FOR ULTRASENSITIVE BIOLOGICAL DETECTION AND IMAGING 71
Amit Agrawal, Tushar Sathe, and Shuming Nie

CHAPTER 5. TAILORING NANOPARTICLES FOR THE RECOGNITION OF BIOMACROMOLECULE SURFACES 91
Mrinmoy De, Rochelle R. Arvizo, Ayush Verma, and Vincent M. Rotello

CHAPTER 6. NANOSCALE CHEMICAL ANALYSIS OF INDIVIDUAL SUBCELLULAR COMPARTMENTS 119
Gina S. Fiorini and Daniel T. Chiu

CHAPTER 7. ULTRA SENSITIVE TIME-RESOLVED NEAR-IR FLUORESCENCE FOR MULTIPLEXED BIOANALYSIS 141
Li Zhu and Steven A. Soper

CHAPTER 8. ULTRASENSITIVE MICROARRAY DETECTION OF DNA USING ENZYMATICALLY AMPLIFIED SPR IMAGING 169
Hye Jin Lee, Alastair W. Wark, and Robert M. Corn

CHAPTER 9. ULTRASENSITIVE ANALYSIS OF METAL IONS AND SMALL MOLECULES IN LIVING CELLS 195
Richard B. Thompson
CHAPTER 10. ELECTROCHEMISTRY INSIDE AND OUTSIDE SINGLE NERVE CELLS
   Daniel J. Eves and Andrew G. Ewing
   215

CHAPTER 11. ELECTROCHEMILUMINESCENCE DETECTION IN BIOANALYSIS
   Xiao-Hong Nancy Xu and Yanbing Zu
   235

CHAPTER 12. SINGLE-CELL MEASUREMENTS WITH MASS SPECTROMETRY
   Eric B. Monroe, John C. Jurchen, Stanislav S. Rubakhin,
   and Jonathan V. Sweedler
   269

CHAPTER 13. OUTLOOKS OF ULTRASENSITIVE DETECTION IN BIOANALYSIS
   Xiao-Hong Nancy Xu
   295

INDEX
   301
PREFACE

This book aims to provide an overview of current exciting research topics in ultrasensitive bioanalysis; to provide an in-depth understanding of objectives, motivations, and future directions of the new frontiers of ultrasensitive bioanalytical research; and to introduce new ideas, new technologies, and new applications of ultrasensitive bioanalysis to a wide spectrum of research communities, including biological, biomedical, clinical, chemical, environmental, and materials science. The book strives to provide sufficient fundamentals and research background for readers to learn and apply these tools. The book is also structured in a way that can serve as a textbook or a primary reference book for an advanced analytical chemistry course or a special topic course for graduate and senior undergraduate students. I discovered the difficulty of finding effective teaching materials and suitable reference books for my students and collaborators while I taught the course and engaged joint research projects with colleagues in the department of biological sciences and engineering.

Ultrasensitive detection plays a vital role in advancing analytical chemistry and has been a primary driving force for the development of new analytical instrumentation and methodology. As analysis of biological samples (e.g., genomics, proteomics) and living systems (e.g., cellular and subcellular function) becomes more demanding, new platforms of ultrasensitive analysis using multiplexing, single nanoparticle sensing, nano-fluidics, and single-molecule detection have emerged and have become indispensable tools to advance biological, biomedical, and biomaterials sciences and engineering. Furthermore, the emerging research field of nanoscience and nanotechnology provides new possibilities for development of new analytical tools and instrumentation for bioanalysis. For example, the unique properties of nanoparticles offer enormous opportunities to develop new probes for real-time in vivo and in vitro imaging and sensing of individual biomolecules with sub-100-nm spatial resolution and millisecond to nanosecond time resolution. Such powerful capabilities will lead to the development of new analytical techniques to improve disease diagnosis and treatment, as well as to advance our fundamental understanding of important phenomena such as membrane transport, gene expression, enzyme activities, and intracellular and intercellular signaling. It is undeniable that analytical chemistry has evolved to be a vital player in all cutting-edge scientific research fields.

In this book, a diverse group of analytical chemists working in the forefront of ultrasensitive bioanalysis share their insights, visions, and latest results. The book is structured to offer appreciable background on the fundamentals and to provide brief overviews of current status of ultrasensitive bioanalysis. More importantly, this book focuses on showcasing the exciting opportunities of the new frontiers of ultrasensitive bioanalysis. Thus, this book is much more for the future than for the summary of the past and present.

This book includes 13 chapters and starts, in Chapter 1, by addressing the most fundamental question of single-molecule detection (SMD): How many measurements of a single...
molecule are required to obtain representative molecular properties of ensemble distributions? It describes the advantages of SMD and its applications to virtual sorting and DNA fingerprinting of single genomes. Chapter 2 focuses on presenting a representative application of SMD to directly observing gene expression in living cells, illustrating potential applications of SMD for better understanding of living cellular mechanism and function. Chapter 3 offers an overview of SMD using fluorescence microscopy and spectroscopy, and it gives an example of studying membrane transport of living cells using single fluorophor molecules and single-nanoparticle optics to illustrate the distinguished advantages of SMD and non-photobleaching noble metal nanoparticles for probing living cellular mechanism and function. This chapter aims to introduce the basic detection schemes and experimental configurations of SMD and single-nanoparticle optics for bioanalysis and single-living-cell imaging.

Chapter 4 is devoted to the recent development of multifunctional nanoparticle probes, especially quantum dots (QDs), for ultrasensitive analysis of biomarkers in living cells and in vivo imaging, providing the latest advances and future possibility of this exciting research arena. To develop nanoparticle probes for living-cell imaging, smart drug delivery, and in vivo diagnosis, the primary challenge is to rationally design the functional nanoparticle surfaces that can specifically recognize biomolecules of interest and be biocompatible to living systems. Thus, Chapter 5 offers examples of the latest advances in designing surface functionality of nanoparticles that can specifically recognize protein and DNA surfaces. This chapter presents an overview of this exciting research topic and provides a wealth of experimental details and new insights into the future research direction.

Miniaturization and multiplexing have become important technology platforms in ultrasensitive analysis of biomolecules and profiling contents of individual living cells. Chapters 6 through 8 present several interesting research projects, reporting the recent development of this particularly exciting research field. Chapter 6 describes the new platform of using the state-of-the-art research tools, such as a laser-based system and droplet-based microfluidics, to selectively extract organelles from living cells for subsequent analysis. Chapter 7 presents the development of near-IR fluorescent dyes, time-resolved fluorescence spectroscopy, and multiplexing approach for DNA sequencing, offering more sensitive and informative genomic measurements. In contrast, Chapter 8 describes a new approach of directly determining genomic DNA samples with no need of labeling or PCR amplification. In this approach, surface plasmon resonance imaging (SPRI) is used as a label-free detection means, while nucleic acid microarrays serve as a multiplexing platform, which is coupled with surface enzymatic method to amplify SPRI imaging signal.

Like proteins and DNA, metal ions and small molecules play a key role in cellular function. However, quantitative and qualitative analysis of small molecules, especially metal ions, in living cells in real time remains extremely challenging. Chapter 9 gives a glimpse of research activities and emerging approaches in this extremely important research area and offers new prospects of the future research possibility and direction.

Chapter 10 describes the development and application of electrochemical methods to monitor the release of neurotransmitters in real time for better understanding communication among individual nerve cells, demonstrating the unique and powerful detection capability of electrochemistry. Chapter 11 introduces electrochemiluminescence (ECL) detection for ultrasensitive bioanalysis, describes its latest development and applications, and presents the potential challenges and exciting future research opportunities of ECL detection in ultrasensitive bioanalysis.
Mass spectrometry offers an unparalleled detection capability for identifying unknown analytes (e.g., tumor markers) and has demonstrated the possibility of profiling the chemical compositions of individual living cells. However, mass spectrometry currently provides very limited detection sensitivity. Chapter 12 illustrates the latest research development of using state-of-the-art mass spectrometry for single-cell analysis, highlights several representative examples, and presents the unique research challenges and opportunities in this exciting emerging new research field. Finally, the book is concluded with Chapter 13: it summarizes current cutting-edge ultrasensitive detection means in bioanalysis, along with their unique features and potential applications. This final chapter also offers a brief outlook on ultrasensitive bioanalysis in terms of emerging technologies and methodologies, along with new challenges and opportunities.

I hope that this book will serve as a valuable tool for beginning and well-established scientists, especially graduate students and the researchers who consider learning about and using ultrasensitive detection tools. I have selected one example of particular detection techniques and research approaches to illustrate the concepts, ideas, experimental design, and potential applications of research field, rather than providing a comprehensive review of an entire research field. It is hoped that through these selected examples, the readers will feel the vibrant and excitement of entire research field and will pursue so many other distinguished research works that are unable to be included in this book. Thus, I hope that I will have the forgiveness and understanding of my colleagues whose research work has not been presented here.

I am very grateful to all contributors for their enthusiastic support and for taking precious time from their demanding busy schedule to prepare the chapters. I would like to acknowledge each and every one of them for their generous effort and inputs. I wish to extend my gratitude to James Winefordner at University of Florida for his persistent invitations over years, as well as to Heather Bergman at Wiley-Interscience for her valuable assistance.

It has been a very pleasant and valuable experience for me to work on this book. I gained the first-hand appreciation and better understanding of the effort, time, energy, and persistence that are essential for an editor of a book. With an already saturated daily schedule of young academics, I had to devote all my weekends and holidays this past year to working on this book in the hope that it would reach readers in time to meet the demands of this very rapidly expanding research arena. I will be extremely delightful if readers, especially students and new investigators, find that the book stimulates their fascination in ultrasensitive bioanalysis and helps them to steer toward new research directions that ultimately lead to new scientific discoveries.

Despite my best effort, I am mindful that I could have done more in editing this book. Therefore, I am eager to hear the constructive comments and suggestions of my colleagues and readers. I am determined to continue to engage in vigorous research and teaching activities in this thriving research field, and I plan to prepare follow-up volumes in the coming years.

It is fitting to take this opportunity to acknowledge several funding agencies for their generous support of my research program: NIH (R21 RR15057; R01 GM0764401), NSF (BES 0507036; DMR 0420304), DoD-MURI (AFOSR #F49620-02-1-0320), DOE (DE-FG02-03ER63646), and Old Dominion University. This financial support allows me to actively participate in this exciting cutting-edge research field, acquire the first-hand research experience, widely interact with colleagues in the field, closely follow the advance of
research field, and gain in-depth understanding of the research direction, which are essential to the successful construction of this book.

Finally, I would like to thank my family, especially my parents, for their tireless support of this endeavor. Without their patience, love, and guidance, this mission would have been impossible.

XIAO-HONG NANCY XU

Norfolk, Virginia
January 2007
CONTRIBUTORS

Amit Agrawal, Departments of Biomedical Engineering and Chemistry, Emory University and Georgia Tech, Atlanta, Georgia

Rochelle R. Arvizo, Department of Chemistry, University of Massachusetts, Amherst, Massachusetts

Andrew C. Beveridge, Bioscience Division, Los Alamos National Laboratory, Los Alamos, New Mexico

Daniel T. Chiu, Department of Chemistry, University of Washington, Seattle, Washington

Robert M. Corn, Department of Chemistry, University of California—Irvine, Irvine, California

Mrinmoy De, Department of Chemistry, University of Massachusetts, Amherst, Massachusetts

Xavier Darzacq, Biologie Cellulaire de la Transcription, Ecole Normale Supérieure, Paris, France

Daniel J. Eves, Department of Chemistry, Pennsylvania State University, University Park, Pennsylvania

Andrew G. Ewing, Department of Chemistry and Department of Neural and Behavioral Sciences, Pennsylvania State University, University Park, Pennsylvania

Gina S. Fiorini, Department of Chemistry, University of Washington, Seattle, Washington

James H. Jett, Bioscience Division, Los Alamos National Laboratory, Los Alamos, New Mexico

John C. Jurchen, Department of Natural Sciences, Concordia University, Seward, Nebraska

Richard A. Keller, Bioscience Division, Los Alamos National Laboratory, Los Alamos, New Mexico

Hye Jin Lee, Department of Chemistry, University of California—Irvine, Irvine, California

Eric B. Monroe, Department of Chemistry, University of Illinois, Urbana, Illinois

Prakash Nallathamby, Department of Chemistry and Biochemistry, Old Dominion University, Norfolk, Virginia

Shuming Nie, Departments of Biomedical Engineering and Chemistry, Emory University and Georgia Tech, Atlanta, Georgia
Vincent M. Rotello, Department of Chemistry, University of Massachusetts, Amherst, Massachusetts

Stanislav S. Rubakhin, Department of Chemistry, University of Illinois, Urbana, Illinois

Tushar Sathe, Departments of Biomedical Engineering and Chemistry, Emory University and Georgia Tech, Atlanta, Georgia

Yaron Shav-Tal, The Mina and Everard Goodman, Faculty of Life Sciences, Bar-Ilan University, Ramat-Gan, Israel

Robert H. Singer, Departments of Anatomy and Structural Biology, Albert Einstein College of Medicine, Bronx, New York

Yujun Song, Department of Chemistry and Biochemistry, Old Dominion University, Norfolk, Virginia

Steven A. Soper, Department of Chemistry, Louisiana State University, Baton Rouge, Louisiana

Jonathan V. Sweedler, Department of Chemistry, University of Illinois, Urbana, Illinois

Richard B. Thompson, Department of Biochemistry and Molecular Biology, School of Medicine, University of Maryland, Baltimore, Maryland

Ayush Verma, Department of Chemistry, University of Massachusetts, Amherst, Massachusetts

Alastair W. Wark, Department of Chemistry, University of California—Irvine, Irvine, California

Xiao-Hong Nancy Xu, Department of Chemistry and Biochemistry, Old Dominion University, Norfolk, Virginia

Li Zhu, Department of Chemistry, Louisiana State University, Baton Rouge, Louisiana

Yanbing Zu, Department of Chemistry, The University of Hong Kong, Hong Kong, China
CHAPTER

1

IS ONE ENOUGH?
ANDREW C. BEVERIDGE, JAMES H. JETT, AND RICHARD A. KELLER

1.1. INTRODUCTION

A concern often expressed about single-molecule measurements can be stated as, “Single-molecule measurements are ‘cool’ but what good are they if you have to measure a million single molecules to estimate the properties of the system?” This is a reasonable concern. Thus, a major focus of this chapter is modeling the effect of sample size (typically a few molecules) on the accuracy of analytical measurements and the extrapolation of properties measured at the single-molecule level to properties derived from ensemble measurements. We include a summary of DNA fingerprinting of single viral particles and bacterial cells that demonstrates that only a few measurements are required. The contrast between single-molecule and bulk measurements is that single-molecule measurements yield physical properties of individual molecules whereas individual molecular properties are difficult or impossible to obtain from bulk measurements; often the heterogeneous nature of single measurements is masked in bulk systems. We finish with (a) a description of “virtual sorting,” where the data stream is sorted instead of the sample stream, and (b) the implementation of virtual sorting to statistical chemistry to either eliminate or reduce the necessity to purify a sample.

1.1.1. Significance of a Single-Molecule Measurement

There is a fundamental postulate of statistical mechanics that states, “The (long) time average of a mechanical property $M$ in the thermodynamic system of interest is equal to the ensemble average of $M$, in the limit as $N$ goes to infinity” (Hill, 1960). (Here and throughout this chapter, $N$ refers to the number of molecules.) If this true, and indeed it must be, why do single-molecule experiments? It is clear that the average fluorescence intensity of a single molecule recorded over many cycles, as it undergoes changes between a fluorescent species and a nonfluorescent species, is equivalent to the average intensity from multiple molecules over one cycle.

We address two different but related aspects of this postulate: (1) How many times should the same single molecule be analyzed for the measured value(s) to be a good representation of the ensemble average? (2) Can single-molecule properties be extracted from bulk data?

The first question relates to studying the properties of single molecules free from averaging effects associated with bulk measurements. A good example is the quantification of the
thermodynamics and dynamics of protein folding where subtle differences in the response curves indicate the presence of multiple folding pathways, barriers in the exit channel, and deviations from a simple two-state model, which are hidden in ensemble measurements (Schenter et al., 1999; Yang and Xie, 2002a,b; Yang and Cao, 2001). A related question is, How many identical single molecules are needed to estimate a sample’s composition? For example, DNA fingerprinting for pathogen detection and identification by gel electrophoresis typically takes several days. Most of the time is spent culturing the sample to acquire sufficient material for analysis. The ability to size accurately single DNA fragments reduces sample preparation time considerably (Ferris et al., 2005; Larson et al., 2000). A similar problem exists with nonculturable bacteria where it is difficult to grow enough sample for standard analyses (Leadbetter, 2003; Pennisi, 2004).

The second question—How can you extract single-molecule properties from data collected from probe volumes that contain more than one emitting molecule?—is more difficult. The magnitude of the problem is apparent by looking at Figure 1.1, which shows time-dependent fluorescence intensity fluctuations as a function of the number of molecules in the probe volume. Even for two molecules, it is impossible to deconvolute the signal into the contribution from each molecule.

It is often stated that the difference between single-molecule and ensemble measurements is just a synchronization problem; but it is more than a synchronization problem. Due to the stochastic nature of fundamental processes, following a synchronized start, the molecules

![Figure 1.1. Simulated data stream for the fluorescence intensity from $N$ molecules in the probe volume. The parameters for the simulation are listed in Table 1.1. Note that the intensity in each plot is normalized and that all the molecules start in the “on” state. The number of molecules present in the probe volume varies from 1 to 5000.](image-url)
Table 1.1. Values for the Simulations Computed Here

<table>
<thead>
<tr>
<th>Species</th>
<th>$k$ (s$^{-1}$)</th>
<th>$\tau$ (s)</th>
<th>Intensity</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1.111</td>
<td>0.9</td>
<td>1.00</td>
</tr>
<tr>
<td>B</td>
<td>3.333</td>
<td>0.3</td>
<td>0.06</td>
</tr>
</tbody>
</table>

$k$ is the rate constant, $\tau$ is the inverse of $k$ (that is, the lifetime of the state), and the intensity represents the fluorescence intensity of each state. The average cycle time, $\tau_{\text{cyc}}$, is 1.2 s and the mean fluorescence intensity is 0.765.

get “out of step” in approximately one cycle and distributions of the measured parameters of the sample are difficult to discern. Fortunately, properties of the system can be measured without a synchronous start.

1.2. MODEL SYSTEM

1.2.1. Equilibrium and Kinetics

We choose a simple system of a nondiffusing single molecule oscillating between a strongly fluorescent state and a weakly fluorescent state to characterize the particularities associated with the study of single molecules in small probe volumes. We focus on the problems and opportunities that result from using probe volumes less than or equal to a femtoliter.

Equations (1.1a) and (1.1b) describe the equilibrium behavior of two conformations of the same molecule, A and B:

\[
\begin{align*}
A \underset{k_2}{\overset{k_1}{\rightleftharpoons}} B \\
K_{\text{eq}} &= \frac{[B]}{[A]} = \frac{k_1}{k_2},
\end{align*}
\]

(1.1a, 1.1b)

where $[A]$ and $[B]$ are the respective concentrations of each species, $k_1$ and $k_2$ are the rate constants, and $K_{\text{eq}}$ is the equilibrium constant. For a small displacement from equilibrium, the differential rate law is

\[
\frac{d\Delta f_A(t)}{dt} = -(k_1 + k_2)\Delta f_A(t) = -\frac{d\Delta f_B(t)}{dt},
\]

(1.2)

where $\Delta f_A(t)$ and $\Delta f_B(t)$ denote the deviation of the concentration from equilibrium for species A and B at time $t$. For the closed, two-state system modeled here, the sum of $[A]$ plus $[B]$ is constant; thus $\Delta f_A(t)$ equals $-\Delta f_B(t)$. The solution to Eq. (1.2) is

\[
\Delta f_A(t) = \Delta f_A^0 \exp[-(k_1 + k_2)t],
\]

(1.3)

where $\Delta f_A^0$ is the initial deviation from equilibrium (Cantor and Schimmel, 1980). The probability of finding the system in A or B is

\[
P_A = \frac{k_2}{k_1 + k_2}, \quad P_B = \frac{k_1}{k_1 + k_2},
\]

(1.4)
where $P_A$ and $P_B$ are the probabilities for A and B, respectively. The mean fluorescence intensity, $\langle I_f \rangle$, is given by

$$\langle I_f \rangle = \alpha \phi_A[A] + \beta \phi_B[B] = \alpha \phi_A N \frac{k_2}{k_1 + k_2} + \beta \phi_B N \frac{k_1}{k_1 + k_2},$$

where $\alpha$ and $\beta$ depend upon the absorption, emission, and detection factors for A and B, and $\phi_A$ and $\phi_B$ are the fluorescence quantum yields for A and B.

Single-molecule measurements of the system give $k_1$, $k_2$, $I_A$, and $I_B$; therefore the parameters $\alpha$ and $\beta$ can usually be determined by calibration of the apparatus allowing one to obtain values for $\phi_A$ and $\phi_B$.

### 1.2.2. Generation of Synthetic Data

The parameters used in the following Poisson simulation of a data stream (Demas, 1983; Matthews and Walker, 1964), chosen to represent a typical photophysical process, are listed in Table 1.1. The conclusions drawn from data synthesized from these particular values may change significantly for a different choice of parameters. Here a cycle is defined as the total of the time period that a molecule spends in each state.

$$\tau_{\text{cyc}} = \tau_1 + \tau_2,$$

where $\tau_1$ and $\tau_2$ are the inverse of $k_1$ and $k_2$, and $\tau_{\text{cyc}}$ is the cycle time. The probability for the forward reaction in Eq. (1.1), $P(t)$, that A changes its conformation at time $t$ within $dt$ is

$$P(t) \approx k_1 \, dt.$$
parameters that can be studied and improves the accuracy of the parameters extracted from fits of the data to curves derived from the reaction mechanism and eliminates complications due to diffusion (Baldini et al., 2005; Jia et al., 1999; Rhoades et al., 2003; Talaga et al., 2000). In practice, the sampling time is limited to less than 100,000 photon absorptions by photobleaching (Soper et al., 1993) and other considerations (Vazquez et al., 1998; Zander et al., 2002).

We describe three methods for data analysis when only one molecule is in the probe volume: (1) averaging, (2) autocorrelation, and (3) distribution analyses.

1. The time averages of the “on and off” periods are measured for a chosen number of cycles; the averages are then inverted to give the rate constants, $k_1$ and $k_2$. The rate constants then can be summed to give the relaxation time, $(k_1 + k_2)^{-1}$, of the system. Here, the assumption is made that differences between states are discernible such that their respective time dependence can be accurately determined (Verberk and Orrit, 2003).

2. The fluorescence autocorrelation function for a simple first-order kinetic process, undergoing fluctuations around the equilibrium concentration, exhibits an exponential decay with a decay rate that is the sum of the rate constants $(k_1 + k_2)$. This type of analysis is called fluorescence correlation spectroscopy (FCS). Autocorrelation requires an analysis time long enough for both states to be sampled such that their average is independent of the time—that is, constant within statistical variation. Note that the time to reach a stationary system is independent of the number of molecules in the system. The fluorescence intensity autocorrelation function is defined as

$$C(t) = \frac{\langle \delta I(t) \delta I(t + \tau) \rangle}{\langle I \rangle^2},$$

(1.8)

where $\delta I$ are the zero-mean intensity fluctuations, $\delta I(t) = I(t) - \langle I \rangle$, at time $t$ and $t + \tau$, respectively (Aragon and Pecora, 1976, Elson, 1985, Krichevsky and Bonnet, 2002, Zander et al., 2002).

3. Distribution analyses include both binning and cumulative distribution functions. In the binning method, “on and off” time periods are binned separately and plotted individually in histogram format; rate constants are determined from each histogram. For the simulations here, a minimum of 10 cycles is needed to define a rudimentary histogram. Alternatively, the data can be fit to the cumulative distribution function for the probability density function that describes the data. The cumulative distribution method has an advantage over binning in that as few as two to three points are needed. However, for cases more complicated than considered here, the cumulative distribution function may be difficult to compute analytically.

### 1.2.3.2. Multiple Molecules in the Probe Volume

In the study of biological systems within a fixed probe volume (cells, lipid vesicles, cell membranes, etc.), it is often of interest to determine the number of a particular species in the probe volume. The ability to measure the number of molecules and determine individual
molecular properties leads to new insights regarding basic intracellular functions. Here, two limits of multiple molecules in the probe volume are examined: a bulk system, \( N > 10 \), where discrete contributions from individual molecules are not apparent, and a system where discrete contributions from individual molecules are obvious, \( 2 \leq N \leq 10 \), herein referred to as a “prebulk” system.

Figure 1.1 illustrates the change from single-molecule system to prebulk system to bulk system as \( N \) is increased from 1 to 5000. This corresponds to a concentration range from 1 nM to 10 µM for a probe volume of 1 fL. Although each molecule is in either A or B, the system is not constrained to a number normalized intensity of either 1 or 0.06 because the molecules get out of phase. For example, in the case where the number of molecules is five, all of the molecules must be in B for a normalized intensity of 0.06; this is improbable, approximately one out of a thousand, for the simulation parameters used here. For five molecules, six normalized intensities are possible with frequencies determined by probability theory. The corresponding plateaus are clearly visible for \( N \) equals 2 and 5, and occasionally 10 in Figure 1.1. Discrete steps are not evident for \( N \) greater than 10.

**Bulk Systems and Synchronous Starts.** Consider the simulations in Figure 1.1 for the bulk case. Here, at \( t = 0 \) all of the molecules are in A and relax to the equilibrium distribution in less than the average of one cycle. As stated previously, the simulations are analogous to rapid perturbation experiments—for example, optical pumping, temperature jump, pH jump, or the addition of an appropriate chemical. For this type of experiment, the time dependence of the signal relaxation gives the sum of the rate constants, \( k_1 + k_2 \), whereas the long-time limit of the signal gives the average fluorescence intensity as given in Eq. (1.5). In the special case where one of the states is “dark,” the equilibrium constant is determined from the ratio of the intensity at time zero to the intensity at times greater than the chemical relaxation time. Depending on the type of experiment, bulk measurements give the equilibrium fluorescence intensity and the sum of the rate constants; individual values of \( I_A, I_B, k_1 \), and \( k_2 \) are typically difficult to measure.

**Prebulk System.** For the prebulk system we refer to two distinct states of the system: all on and all off. (There are a total of \( N + 1 \) distinct states; the number of configurations is \( (A + B)^N \).) Here all on refers to the case where every molecule in the probe volume is A concurrently; correspondingly, all off indicates that every molecule is B at the same time. Assuming that all the fluorescing molecules in the probe volume are identical, probability theory predicts that the individual rate constants can be determined by the average of the period when all the molecules are either all on or all off; the average period multiplied by the number of molecules gives the inverse of the corresponding rate constant (Riley et al., 1997). This measurement depends on explicit knowledge of the number of molecules in the probe volume. Figure 1.2 gives an example for 2 to 10 molecules. Conversely, if the rate constants are known and assuming identical molecules, the average of either the all on or all off periods can be used to determine the number of molecules in the probe volume. If neither the rate constant nor the number of molecules is known, the ratio of the average lifetime of the all off state to the all on state gives the equilibrium constant. For all types of measurements in a prebulk system the system state must be known—that is, whether the system is in the all on or all off state. There are two possible experimental procedures that
Figure 1.2. Determination of $k_1$ in a prebulk system from 1 to 10 identical molecules in the probe volume. The average of all on period is computed for a data stream where the average number of all on periods is 10 in the data stream; see Eq. (1.9). The average all on period is then multiplied by the number of molecules and inverted to give $k_1$. The error bars represent the standard deviation from 10 independent numeric simulations. The dashed line indicates the true value for $k_1$.

can be used with this method, and both methods can determine the system state:

1. If there is a distribution of $N$ in the sample, then one can look for a probe volume where $N = 1$; then use the data to determine the rate constants. Once the rate constants are known, $N$ can be determined for other probe volumes when $N$ is less than or equal to 10.

2. After sufficient data are collected, the light intensity can be increased to facilitate photobleaching. For noninteracting molecules, photobleaching occurs in discrete steps; the steps are counted to determine $N$ (Park et al., 2005). The rate constants can be calculated once $N$ is known.

Table 1.2 gives the probabilities of finding the system in either all on or all off states as a function $N$ for the parameters in Table 1.1. The time trajectory for either a simulation or an

<table>
<thead>
<tr>
<th>Molecules</th>
<th>Probability (All On)</th>
<th>Probability (All Off)</th>
<th>Simulation Time (s) (All On)</th>
<th>Simulation Time (s) (All Off)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.56</td>
<td>$6.3 \times 10^{-2}$</td>
<td>8.0</td>
<td>14.0</td>
</tr>
<tr>
<td>4</td>
<td>0.32</td>
<td>$3.9 \times 10^{-3}$</td>
<td>7.2</td>
<td>192.0</td>
</tr>
<tr>
<td>6</td>
<td>0.18</td>
<td>$2.4 \times 10^{-4}$</td>
<td>8.5</td>
<td>$2.1 \times 10^3$</td>
</tr>
<tr>
<td>8</td>
<td>0.10</td>
<td>$1.5 \times 10^{-5}$</td>
<td>11.3</td>
<td>$2.5 \times 10^4$</td>
</tr>
<tr>
<td>10</td>
<td>0.06</td>
<td>$9.5 \times 10^{-7}$</td>
<td>16.0</td>
<td>$3.2 \times 10^5$</td>
</tr>
</tbody>
</table>

The simulation time is the time needed for either a simulation or real experiment, assuming one makes 10 measurements of either the “all on or all off” periods. Figure 1.2 is the determination of $k_1$ using the simulation times in this table.
experiment can be computed from the following formulae:

\[ T_A = \left( \frac{m}{N k_1} \right) (P_A)^{-N}, \]  

(1.9a)

\[ T_B = \left( \frac{m}{N k_2} \right) (P_B)^{-N}, \]  

(1.9b)

where \( T_A \) and \( T_B \) are the times needed to measure a specific number, \( m \), of all on and all off periods. For example, we anticipate 10 separate periods of all on (all off) for \( m = 10 \). For the all on case where \( N = 3 \) and \( m = 10 \), along with the parameters in Table 1.1, \( T_A \) is 7.2 s; 100 independent simulations with the time set to 7.2 s give \( m = 10.1 \pm 1.9 \), which agrees with the initial premise of 10 independent measurements in the data stream. The interesting part of this method is that times are measured rather than intensities. Finally, autocorrelation analysis can also be applied to this system to determine the sum of the rate constants. Neither method requires a synchronous start.

1.2.4. Comparison of Data Analysis Techniques

Both averaging and distribution methods give the mean value of the histogram and the variance; thus there is little difference between the two. Because averaging analysis and distribution analysis are closely related, only the differences between the averaging method and autocorrelation analysis are considered; the differences between distribution and autocorrelation analyses are assumed to be the same as the former case.

The major difference between distribution analysis and averaging is that distribution analysis gives a visual representation of the values that may help distinguish between two distinct states, whereas averaging gives a single point. For example, the distinction between two closely separated peaks in a distribution analysis is evident, whereas the average would only give one point, the mean of the two peaks. In the limit of a normal distribution, averaging and distribution analysis give identical answers.

1.2.4.1. Single-Molecule Case

There is a fundamental difference between an autocorrelation analysis and measuring the distributions of on and off periods for the case where both states are fluorescent. The time dependence of the autocorrelation of the fluorescence intensity fluctuations provides the sum of the forward and reverse reaction rate constants; in the special case, where only one state is fluorescent and the other state is dark, the zero-time offset gives the equilibrium constant (Zander et al., 2002). However, the averaging method gives rate constants for each state even if both states are fluorescent. In addition, the averaging method is computationally simpler to implement than autocorrelation analysis. For a series of Poisson events, as in the reactions modeled here, probability theory predicts that maximum likelihood for the series is the average value of the Poisson process (Matthews and Walker, 1964). Thus, a reasonable estimation of the real value results from measuring only a few cycles; however, autocorrelation analysis depends on the system being in a stationary state; therefore, measurement of only a limited number of cycles may not meet this requirement.

A graphical comparison of the two data analyses methods is shown in Figure 1.3 for simulated data sets of the same single molecule. Figure 1.3 clearly shows that as the
Figure 1.3. Comparison of the sum of the rate constants for a single molecule calculated either by autocorrelation analysis or averaging the “on and off” times for a given number of cycles as indicated on the x axis. For clarity, the points for the autocorrelation analysis are on one side of the points for the averaging method; the points for the averaging method give the true number of cycles for the other method. The circles (○) are the autocorrelation fit values and the triangles (▲) are the values determined by averaging for the rate constant sum. The error bars are the standard deviations for five independent simulations. The dashed horizontal line at 4.444 \( s^{-1} \) is from the parameters used in the simulation of the data stream in Table 1.1 and represents the true value.

As the number of cycles increases, both methods give similar values for the sum of the rate constants \((k_1 + k_2)\). For example, at 150 cycles, the values are \(4.58 \pm 0.53\) and \(4.46 \pm 0.19\) for autocorrelation and averaging, respectively. However, for a limited number of cycles, generally less than 20, the accuracy of the averaging method is usually greater than that for autocorrelation analysis. For example, in Figure 1.3, the averaging method for the five cycles gives the sum of the rate constants within 10% of the true value, whereas autocorrelation analysis gives a value within 18%; in addition, the standard deviations of the simulations are greater for autocorrelation analysis.

Another possible advantage of the averaging method is the analysis of long-time data sets. A “moving” average, a nonweighted average over successive subsets of the data, which is commonly used in assembly line quality control, identifies any slow processes, such as conformation changes, occurring on a time scale long with respect to changes between states. Because the averaging method gives a reasonable result with between 5 and 15 measurements, a moving average can be applied to a subset within this range to determine where any changes are occurring. Once any changes are identified, the data can be reanalyzed by adjusting the subset size to give favorable results. Autocorrelation analysis of a system with slow conformation changes gives deviations from simple exponential decay; however, it may be difficult to determine when the deviations occurred. More complicated autocorrelation functions are needed to fit the deviations (Schenter et al., 1999; Yang and Xie, 2002b). To our knowledge, the use of a moving average method—a simple, powerful technique—to single molecule studies has not been reported previously.

**Noise Considerations.** The noise considerations for each type of analysis are quite different. The averaging method measures time widths whereas autocorrelation analysis measures
Figure 1.4. (A) Simulated data for an on and off process. (B) Identical data as in part A except noise and an RC constant of 0.04 time units are added; the amount of noise added gives a signal-to-noise ratio of approximately 5. Because the averaging method requires distinguishable states, the signal-to-noise ratio is important; a threshold amplitude value can be chosen to determine the molecule’s state. If the noise level approaches the threshold value, discrimination between states is difficult. For example, in part B, if the threshold value for being in part A is set to greater than 0.5, then time spent in part B for a cycle, as indicated by the arrow, is overlooked. Alternatively, setting the threshold value to a smaller value can lead to noise signals interpreted as changes in state.

intensity fluctuations. Assuming a reasonable signal-to-noise ratio (i.e., the ability to distinguish states), the uncertainty for measurement of the widths of the on and off pulses depends largely on the RC constant of the detector and measurement system (see Figure 1.4). A large RC constant has the effect of noise reduction and data smoothing, but fast transitions between states are integrated and are potentially not resolvable. Because the autocorrelation function analyzes fluorescence intensity fluctuations as the molecule cycles between A and B, it is subject to shot noise associated with photon counting, white noise, 1/f noise, and other technical noise; however, most noise is uncorrelated, and thus there is a reduction of noise proportional to the square root of the number of time bins.

The method of choice for data analysis depends largely on the data itself:

1. If the signal-to-noise ratio is low (i.e., it is difficult to distinguish individual states), autocorrelation analysis is the preferred method.
2. If individual states are distinguishable and the amount of data is limited, then the averaging method generally yields more accurate results.
3. If individual states are distinguishable and the amount of data is not limited, then both the averaging method and autocorrelation analysis give equivalent results.

1.2.4.2. Prebulk Case

Both the averaging method and autocorrelation analysis assume that the chemical system is known in the probe volume. In addition to the former requirement, the averaging of the periods when all the molecules are either all on or all off involves knowledge of the molecular brightness if either individual rate constants or $N$ is to be ascertained from the system.
best type of analysis for the prebulk system, for two fluorescent states, is a combination of both methods. For this case, autocorrelation analysis gives $k_1 + k_2$, whereas the averaging method gives $K_{eq}$. By combining the results of both types of analysis, individual rate constants and $N$ are determined. The noise considerations are analogous to those for the $N = 1$ case.

1.2.4.3. Bulk Case

In the bulk case, data analysis is limited to either the time dependence of the fluorescence intensity or autocorrelation analysis. Generally, for under 100 molecules, autocorrelation analysis is preferred because of its noise reduction feature. Because the relative intensity fluctuations become smaller as the number of molecules in the probe volume increases, autocorrelation analysis of bulk systems is not appropriate due to signal-to-noise considerations. Thus, time analysis of the relaxation following a perturbation jump experiment for bulk systems for large numbers of molecules may be the only method to determine the sum of the rate constants. In this case, technical noise is the main source of noise.

1.3. PROBE VOLUMES

In many interesting biological cases—for example, cells and lipid vesicles—choosing the probe volume is not possible. The determination of $N$ is often the object of the experiment. In the previous section we discuss the treatment of data for multiple molecules in the probe volume. Here we examine the importance of the choice of a probe volume for single-molecule measurements and methods for attaining probe volumes smaller than can be obtained with diffraction optics.

The requirement that the average occupancy of the probe volume be less than one limits the range of reactions that can be studied by single-molecule studies (Laurence and Weiss, 2003). Consider the reaction

$$A \xrightleftharpoons{kd}{ka} A + B \quad (1.10a)$$

$$K_D = \frac{kd}{ka} = \frac{[A][B]}{[AB]} \quad (1.10b)$$

where $K_D$ is the equilibrium constant, $ka$ is the association rate constant, and $kd$ is the dissociation rate constant.

The magnitude of $K_D$ is inversely related to the strength of the complex; $K_D$ is often called the affinity to reflect this relationship. A small affinity value indicates a tightly bound complex. Here, affinity values of AB where the complex is half-dissociated are examined:

$$[B] = [AB] \quad (1.11a)$$

$$[B] = C_0/2 = K_D, \quad (1.11b)$$

where $C_0$ is the concentration corresponding to one analyte molecule in the probe volume. At equilibrium, $kd$ equals $(ka \times C_0/2)$ and the recombination rate is $(ka \times C_0/2)$. Assuming
that the forward reaction is diffusion controlled, $k_a$ is approximately $10^8$ M$^{-1}$ s$^{-1}$ for a typical protein in water (Cantor and Schimmel, 1980).

A particular probe volume is appropriate for a limited range of affinities. For example, a volume of 0.2 fL, typically used in single-molecule confocal spectroscopy, requires a concentration less than 8.3 nM for an average occupancy of less than 1. This corresponds to an affinity of approximately 4 nM. It would be difficult to study molecules with larger affinities in this probe volume because the lower $C_0$ would result in unacceptable dissociation times resulting in fewer observable events. (See Table 1.3.) Two ways around this problem are smaller probe volumes and multiple occupancy of the probe volume.

Various ways of attaining probe volumes smaller than 1 fL are described below and are listed in Table 1.3; the volumes cover the range from picoliters to zeptoliters. This table illustrates several important points. Using the correct probe volume is critical when studying probe/target binding. While the range of affinities for a particular probe volume is limited, affinities from $10^{-3}$–$10^6$ nM can be studied by choosing the appropriate probe volume.

It is mistakenly perceived that the diffraction of light limits the smallest probe volume to approximately a femtoliter (1 μm × 1μm × 1μm). Indeed, the most commonly used probe volume is 0.2 fL attained by diffraction limited, confocal optics. The most common way to exceed the limits imposed by diffraction optics is to use an evanescent field (Ambrose et al., 1999). When light strikes a transparent surface at an angle that exceeds the critical angle essentially all of the light is reflected, this effect is called total internal reflection (TIR). However, a small fraction of light “leaks” into the media; this is the evanescent wave. The intensity of the evanescent wave decreases exponentially with an $e^{-1}$ length in the range of 100–500 nm. Probe volumes smaller than 1 fL are attained by a combination of confocal detection, spatial confinement into volumes with dimensions smaller than the wavelength of light, and evanescent excitation (Foquet et al., 2002; Ha et al., 1999; Hassler et al., 2005; He et al., 2005; Mukhopadhyay et al., 2004; Schwille; 2003; Starr and Thompson, 2001). TIR techniques are well-suited for studying single molecules at interfaces such as surfaces and cellular membranes (Schwille, 2003; Starr and Thompson, 2001). Foquet et al. (2002) reported a probe volume of 0.05 fL by confocal excitation of a sample confined between two

### Table 1.3. Representative Parameters for the Different Probe Volumes

<table>
<thead>
<tr>
<th>Method</th>
<th>Volume (fL)</th>
<th>$C_0$ (nM)</th>
<th>Dissociation Time (s)</th>
<th>Affinity (nM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Focused flow</td>
<td>$1.0 \times 10^3$</td>
<td>$1.7 \times 10^{-3}$</td>
<td>$1.2 \times 10^4$</td>
<td>$8.3 \times 10^{-4}$</td>
</tr>
<tr>
<td>Droplets</td>
<td>1.0</td>
<td>1.7</td>
<td>12.1</td>
<td>0.8</td>
</tr>
<tr>
<td>Two-photon</td>
<td>1.0</td>
<td>1.7</td>
<td>12.1</td>
<td>0.8</td>
</tr>
<tr>
<td>Nanochannels</td>
<td>0.3</td>
<td>5.5</td>
<td>3.6</td>
<td>2.8</td>
</tr>
<tr>
<td>Confocal</td>
<td>0.2</td>
<td>8.3</td>
<td>2.4</td>
<td>4.2</td>
</tr>
<tr>
<td>TIR</td>
<td>0.1</td>
<td>16.6</td>
<td>1.2</td>
<td>8.3</td>
</tr>
<tr>
<td>TIR/FCS</td>
<td>$5.0 \times 10^{-2}$</td>
<td>33.2</td>
<td>0.6</td>
<td>16.6</td>
</tr>
<tr>
<td>Two-plate</td>
<td>$5.0 \times 10^{-3}$</td>
<td>$3.3 \times 10^2$</td>
<td>0.1</td>
<td>$1.7 \times 10^2$</td>
</tr>
<tr>
<td>Waveguide</td>
<td>$1.0 \times 10^{-6}$</td>
<td>$1.7 \times 10^6$</td>
<td>$1.2 \times 10^{-5}$</td>
<td>$8.3 \times 10^5$</td>
</tr>
<tr>
<td>Human cell</td>
<td>$6.6 \times 10^4$</td>
<td>$2.5 \times 10^{-5}$</td>
<td>$1.6 \times 10^6$</td>
<td>$1.3 \times 10^{-5}$</td>
</tr>
</tbody>
</table>

*a Column 1 lists the method used to attain the probe volume listed in column 2. Column 3 lists the concentration, $C_0$, where there is, on average, one molecule in the probe volume. The dissociation time is the time needed for the complex to split—that is, the inverse of $k_d$. The affinity, $K_D$, is defined as the concentration where the complex is half-dissociated, $[B] = [AB]$, and the affinity is $C_0/2$. The rate constant for the forward reaction is $k_a$ is approximately $10^8$ M$^{-1}$ s$^{-1}$ for a typical protein in water (Cantor and Schimmel, 1980). The rate constant for the forward reaction is $k_a$ is approximately $10^8$ M$^{-1}$ s$^{-1}$ for a typical protein in water (Cantor and Schimmel, 1980).
glass plates separated by 100 nm. Craighead’s group recently combined spatial confinement and evanescent excitation; the sample was placed in a 200-nm inner-diameter quartz tube and excitation light was focused into the tube, resulting in an evanescent wave in the tube that gave a probe volume of $10^{-6}$ fL (Levene et al., 2003).

Two-photon excitation is another attractive way to define a small probe volume (Mertz et al., 1995). In two-photon spectroscopy, molecules are excited by the simultaneous absorption of two photons. Because the absorption is a two-photon process, the excitation probability is proportional to the square of the irradiance yielding probe volumes smaller than 1 fL (Schwille, 2001). Excitation of visible fluorophores is confined to the high irradiance region reducing both photobleaching of molecules outside of the probe volume and background luminescence from intracellular material.

1.4. STATISTICS OF SINGLE-MOLECULE MEASUREMENTS

1.4.1. Application to DNA Fingerprinting

This section describes our work on DNA fragment sizing (Ferris et al., 2005; Goodwin et al., 1993; Habbersett and Jett, 2004; Larson et al., 2000) that we include as a demonstration of the validity of extracting ensemble properties from less than ten molecules. Previously, we have measured the fluorescence intensity of individual DNA restriction fragments intercalated with a fluorescent dye to obtain a DNA fingerprint of a chromosome from a single bacterial cell or viral particle (Ferris et al., 2005). DNA intercalating dyes react stoichiometrically with DNA, resulting in a fluorescence signal that is proportional to the number of base pairs in the fragment. A histogram of the fluorescence intensity from individual restriction fragments is a DNA fingerprint.

1.4.1.1. Flow Cytometry

For analytical measurements, both an orderly delivery of analyte to the detection volume and efficient detection of analyte molecules are desired. One approach to this problem is detection in flowing sample streams. This method is based upon the use of hydrodynamic focusing to position sample streams in the center of a square-bore flow cell. When care is taken to ensure that the sample stream passes through the probe volume, such that all of the molecules experience similar excitation and optical detection efficiencies, the molecular detection efficiency is greater than 90% at a detection rate of approximately 100 molecules/s (Keller et al., 2002). Our flow approach to DNA sizing is roughly 100 times faster (from days to minutes), one million times more sensitive (from micrograms to femtograms), five times more certain (from 10% to 2%), and more accurate in the range of 125 bp to 450 kbp than gel electrophoresis (Larson et al., 2000). Figure 1.5 shows a DNA fingerprint from a HindIII digest of λ phage DNA (Habbersett and Jett, 2004). Fluorescence from approximately 1800 fragments is used to construct this histogram. How many were really needed?

Figure 1.6 illustrates how accurately the mean of a Gaussian distribution can be determined from the average of $m$ measurements where $m = 2, 5, 10,$ and $20$ (Ferris et al., 2004). Only two measurements are required to reasonably estimate the mean and $m$ greater than or equal to five does remarkably well for a Gaussian distribution. Figure 1.6 can be thought of as a graphical representation of the maximum likelihood theorem; that is, the most probable
outcome for a random sequence of numbers from a Gaussian distribution is the average value of the distribution (Matthews and Walker, 1964).

Figure 1.7 shows the effect of reducing the sample size in an *S. aureus* restriction digest assay from 14,163 fragments to 570 fragments (Ferris et al., 2004). The sample size of 570 ensures that there are at least 10 replicate measurements of each fragment. The smaller-sized sample results in a loss of resolution, but still gives a representative fingerprint.

### 1.4.1.2. Optical Mapping

We use optical mapping techniques developed by David Schwartz and colleagues (Aston et al., 1999; Meng et al., 1995; Schwartz et al., 1993; Schwartz and Samad, 1997) to obtain DNA fingerprints from single viral genomes deposited on a surface (Ferris et al., 2005). Briefly, the experimental procedure involves embedding cells in an agarose matrix, lysing the cells, and processing to yield intact DNA. The DNA is then deposited on a derivatized glass substrate, which stretches out the genome. The elongated genome is digested with a restriction enzyme and stained with YOYO-1, an intercalating dye. The DNA relaxes upon cutting, leaving gaps at the restriction sites and intact DNA between restriction sites; after digestion, the order of the fragments is preserved. We use a concentration of 25 ng/mL...
Figure 1.6. Distribution of randomly chosen single molecules within a Gaussian profile representative of an intact λ phage DNA fragment generated using a mean of 48.502 kbp and a CV of 10%. The solid curve represents the Gaussian distribution whereas the vertical lines denote the positions of the random picks and the filled circle denotes the average of the $m$ picks. [Reprinted from Ferris et al. (2004), with permission of Wiley-Liss.]

of intact viral genomes to obtain one genome per $10^4 \text{µm}^2$. Fluorescence from individual fragments is collected with a CCD camera mounted on a sensitive microscope.

Figure 1.8 shows fluorescence from a PmeI digest of a mixture of λ phage and bacteriophage T4, strain GT7. The fragment patterns are clearly different and easily distinguished from each other. We use the known sizes of T4 GT7 restriction fragments and a CV of 8%, characteristic of our measurements, to construct a virtual fingerprint of T4 GT7. DNA fragment lengths measured from seven T4 GT7 genomes, using λ phage DNA as an intensity standard, are plotted in Figure 1.8. Fragments from individual genomes are identified
Figure 1.7. Comparison of the DNA fragment size distribution from a *S. aureus* Mu50 digested with *Sma*I derived from the measurement of 14000 fragments to a distribution derived from a measurement of 570 fragments. [Reprinted from Ferris et al. (2004), with permission of Wiley-Liss.]

by a color code. The fragment lengths from the individual cells are closely grouped, and a fingerprint from any one of the seven genomes represents the actual fingerprint.

Figure 1.9A shows a fluorescence image from a *Pmel* digest of a mixture of adenovirus, λ phage, and T4 GT7. Again the fingerprints are easily identified by their pattern. Figure 1.9B shows the fragment lengths measured in Figure 1.9A plotted on top of a virtual digest of adenovirus and T4 GT7. The λ digest is used to calibrate the fragment sizes. These data show that optical mapping techniques can identify the composition from a mixed sample. This is in contrast to restriction fragments analyzed by either gel electrophoresis or flow

Figure 1.8. (A) A portion of a fluorescence image displaying stained DNA fragments of a *Pmel* digest of a mixture of λ phage DNA, falsely colored red, and T4 GT7, falsely colored green. The different genomes are easily identified by their respective patterns. See insert for color representation of figure A. (B) Seven distinct T4 GT7 genomes from the complete image shown in part A plotted on top of a virtual digest of T4 GT7. The data are offset from each other on the y axis for viewing. The two T4 GT7 genomes in part A are denoted by the square (■) and the open circle (O) in the plot. The fluorescence intensity of the T4 GT7 genomes is calibrated with data from the λ phage DNA. [Reprinted from Ferris et al. (2005), with permission of Elsevier.]