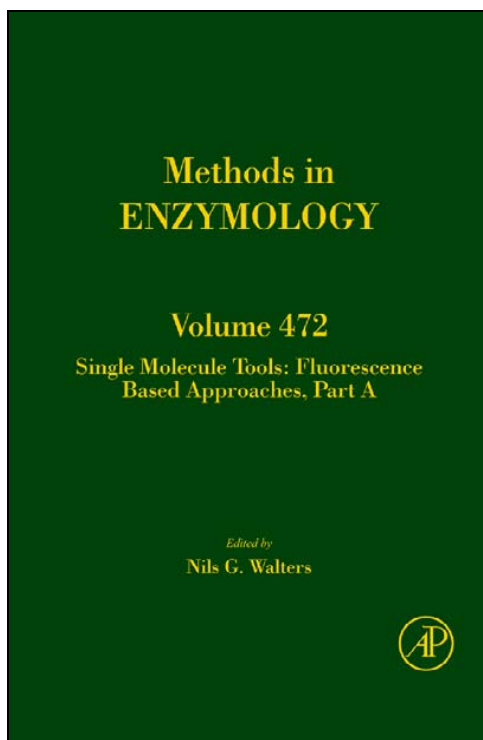


**Provided for non-commercial research and educational use only.
Not for reproduction, distribution or commercial use.**

This chapter was originally published in the book *METHODS IN ENZYMOLOGY*, Vol. 472, published by Elsevier, and the attached copy is provided by Elsevier for the author's benefit and for the benefit of the author's institution, for non-commercial research and educational use including without limitation use in instruction at your institution, sending it to specific colleagues who know you, and providing a copy to your institution's administrator.



All other uses, reproduction and distribution, including without limitation commercial reprints, selling or licensing copies or access, or posting on open internet sites, your personal or institution's website or repository, are prohibited. For exceptions, permission may be sought for such use through Elsevier's permissions site at: <http://www.elsevier.com/locate/permissionusematerial>

From: Mario Blanco and Nils G. Walter, Analysis of Complex Single-Molecule FRET Time Trajectories

In Nils G. Walters editor: *METHODS IN ENZYMOLOGY*, Vol. 472, Burlington: Academic Press, 2010, pp.153-178.

ISBN: 978-0-12-374954-3

© Copyright 2010, Elsevier Inc.
Academic Press.

ANALYSIS OF COMPLEX SINGLE-MOLECULE FRET TIME TRAJECTORIES

Mario Blanco^{*,†} and Nils G. Walter^{*}

Contents

1. Introduction	154
2. Analysis of Simple Trajectories	156
2.1. Selection of trajectories for analysis	157
2.2. Analysis of FRET state distribution	158
2.3. Kinetic analysis with thresholding algorithms	158
3. Analysis of Complex Trajectories	160
3.1. Hidden Markov analysis of complex FRET trajectories	160
3.2. Overview of HMM software available for smFRET analysis	163
3.3. Preprocessing trajectories for analysis by HMM	164
3.4. Selecting the appropriate number of FRET states	167
4. Post-HMM Processing and Data Visualization	168
4.1. Local detection of correlation based on HMM	168
4.2. Data condensation and visualization	169
4.3. Applications to single-molecule studies of yeast pre-mRNA splicing	171
4.4. Summary of a detailed strategy for analysis of complex trajectories with QuB	174
Acknowledgment	176
References	176

Abstract

Single-molecule methods have given researchers the ability to investigate the structural dynamics of biomolecules at unprecedented resolution and sensitivity. One of the preferred methods of studying single biomolecules is single-molecule fluorescence resonance energy transfer (smFRET). The popularity of smFRET stems from its ability to report on dynamic, either intra- or intermolecular interactions in real-time. For example, smFRET has been successfully used to characterize the role of dynamics in functional RNAs and their protein

^{*} Department of Chemistry, Single Molecule Analysis Group, University of Michigan, Ann Arbor, Michigan, USA

[†] Program in Cellular and Molecular Biology, University of Michigan, Ann Arbor, Michigan, USA

complexes, including ribozymes, the ribosome, and more recently the spliceosome. Being able to reliably extract quantitative kinetic and conformational parameters from smFRET experiments is crucial for the interpretation of their results. The need for efficient, unbiased analysis routines becomes more evident as the systems studied become more complex. In this chapter, we focus on the practical utility of statistical algorithms, particularly hidden Markov models, to aid in the objective quantification of complex smFRET trajectories with three or more discrete states, and to extract kinetic information from the trajectories. Additionally, we present a method for systematically eliminating transitions associated with uncorrelated fluorophore behavior that may occur due to dye anisotropy and quenching effects. We also highlight the importance of data condensation through the use of various transition density plots to fully understand the underlying conformational dynamics and kinetic behavior of the biological macromolecule of interest under varying conditions. Finally, the application of these techniques to studies of pre-mRNA conformational changes during eukaryotic splicing is discussed.

1. INTRODUCTION

One of the most significant advances in the single-molecule field was the advent of measuring fluorescence resonance energy transfer (FRET) between a single FRET fluorophore pair (Ha *et al.*, 1996). FRET relies on the distance-dependent interaction between two fluorophores, a donor and acceptor, so termed because the former “donates” its energy through space to the latter, therefore decreasing in emission intensity while the acceptor’s intensity increases. Due to the nature of the transition dipole interaction between the two fluorophores, energy transfer is more efficient when they are in close proximity than when they are further apart, allowing one to measure relative distances of up to ~ 10 nm between labeled sites on one or more biomolecules (Michalet and Weiss, 2002). A simple FRET ratio on a scale from 0 to 1 as a relative measure of the interfluorophore distance can be calculated as

$$\text{FRET} = I_A / (I_A + I_D), \quad (9.1)$$

where I_A and I_D are the fluorescence intensities of the acceptor and donor, respectively. Changes in this FRET ratio over time are then used as a measure of the conformational changes of the labeled biomolecule(s) over the course of an experiment. While absolute interfluorophore distances can also be estimated from this FRET ratio (Pereira *et al.*, 2008), purely structural studies of this kind are not the focus of this chapter.

FRET measurements have long been performed in the ensemble (Stryer, 1978); however, advances in microscopy and sample preparation

have allowed FRET to be performed at the single-molecule level with relative ease and reliability (Roy *et al.*, 2008; Walter *et al.*, 2008). In a commonplace single-molecule FRET (smFRET) experiment, the emissions of the fluorophores attached to an immobilized biomolecule are monitored by wide-field video fluorescence microscopy in real-time (Fig. 9.1A; for a description of the necessary instrumentation, the reader is referred to recent reviews of Roy *et al.* (2008) and Walter *et al.* (2008)). Corresponding donor and acceptor spots are identified postacquisition by mapping and pattern recognition of the two appropriately color-filtered video images, their signals integrated, and the FRET ratio calculated at the sampling rate of the detector used to collect the signal (Fig. 9.1B; Roy *et al.*, 2008). The most common fluorophores used for smFRET are Cy3 (donor) and Cy5 (acceptor) because of their relative brightness and photostability when compared to other fluorophores (Aitken *et al.*, 2008; Kapanidis and Weiss, 2002). The Cy3–Cy5 FRET pair has a Förster radius of ~ 54 Å, allowing it to report on distance changes on the 20–100 Å scale (Ishii *et al.*,

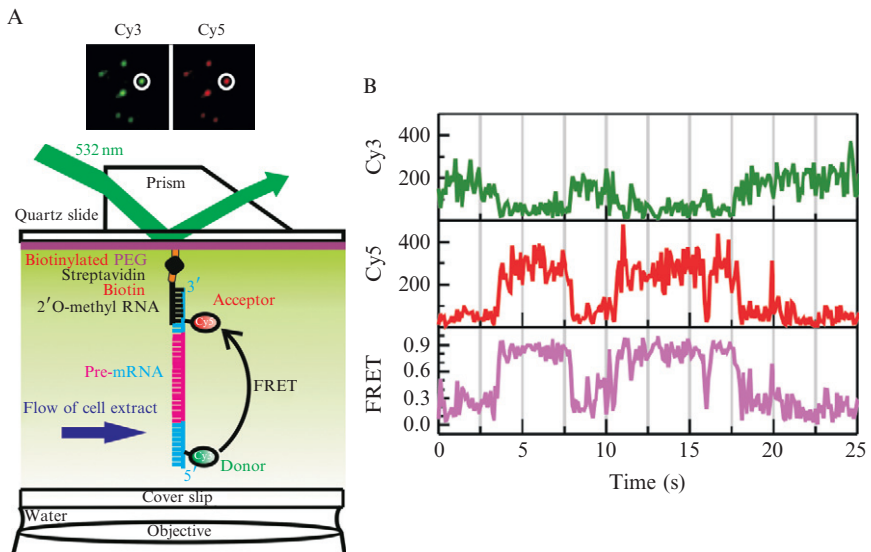


Figure 9.1 Capturing the conformational dynamics of single pre-mRNA molecules through smFRET in real-time. (A) A pre-mRNA molecule is immobilized through a 2'-O-methylated capture oligonucleotide, and is bound to a PEG passivated quartz slide through a biotin–streptavidin interaction. The inset depicts a portion of a field of view captured by the I-CCD camera, and how the donor (Cy3) and acceptor (Cy5) fluorophores can be captured simultaneously. Peak-finder algorithms are used to automatically find and match corresponding fluorophores from single molecules (white circles). (B) Exemplary fluorescence intensity and FRET changes of a single molecule. Modified in part from Abelson *et al.* (2010).

1999). Detection by smFRET has distinct advantages as it directly observes the kinetics of conformational changes of single molecules at (or away from) equilibrium without the need to synchronize and thus perturb them as ensemble FRET methods require; even highly heterogeneous biomolecular samples with rare and/or transient conformational states can thus be analyzed by smFRET in depth, revealing structural dynamics at unprecedented detail (Roy *et al.*, 2008; Walter *et al.*, 2008).

The additional information gathered by smFRET over ensemble FRET approaches depends on the ability to detect subtle changes in inherently noisy data from individual molecules. The contribution of static and dynamic heterogeneity (or disorder) to biological systems, which is typically lost by ensemble-averaging, is being unraveled now by single-molecule detection, contradicting the assumption that all molecules in a population are behaving exactly the same (Ditzler *et al.*, 2008; Fiore *et al.*, 2008; Min *et al.*, 2005). In addition, rare and/or short-lived intermediates often go undetected (or are ignored) in ensemble experiments because of their minimal contribution to the overall signal, yet in smFRET they are readily detectable as long as they are longer lived than the inverse of the rate of collection, usually on the order of 25–100 ms. Through the observation of a large population of molecules during wide-field video fluorescence microscopy, sufficient statistics can be built up quickly to recapitulate the ensemble behavior of molecules, while not masking subpopulations. These features have made smFRET experiments popular, but the ability to extract all possible information from single-molecule time trajectories in an unbiased fashion depends on (semi-)automated data analysis routines, the development of which so far lags behind that of the experimental techniques. We faced a particularly daunting task when analyzing complex smFRET trajectories from a pre-messenger RNA (pre-mRNA) during splicing *in vitro* (Abelson *et al.*, 2010) and describe here the resulting, practical strategy for the largely unbiased extraction of kinetic information from smFRET trajectories with three or more states. Scripts for implementation of this strategy are available upon request.

2. ANALYSIS OF SIMPLE TRAJECTORIES

The practical value of smFRET experiments depends greatly on the ability to extract kinetic and conformational information about the biomolecule of interest. A priority for smFRET assays should be the design of a system (in terms of composition and fluorophore labeling sites) that leads to changes over time and differences between molecules that are significant enough to discern them in single-molecule trajectories. Once established, such a system allows for the application of FRET distribution analyses and

thresholding algorithms that, when incorporated with information about dye position, yield information regarding the structural dynamics of the molecule(s) of interest and how experimental conditions affect them. We will first describe a systematic set of analysis tools that are suitable for simple (two- or three-state) trajectories, and often can be used for the initial characterization of more complex trajectories.

2.1. Selection of trajectories for analysis

Even in the case of simple trajectories with only a few (two or three) FRET states, the amount of information that is collected throughout a time series of images from an smFRET experiment requires the use of (semi-)automated scripts (often written for the programs MATLAB or IDL) to process the raw imaging data. First, peak-finder and signal integration algorithms (Roy *et al.*, 2008) are usually performed in IDL to select molecules from a field of view and match peaks collected in different color-filtered channels (Cy3 and Cy5) (inset of Fig. 9.1A). For a wide-field microscope this approach is necessary since an average field of view of perhaps $50 \times 100 \mu\text{m}^2$ often contains over 50 molecules, and five fields of view or more may be observed per experiment. Second, the initial pool of candidate signals identified by the peak-finder algorithm is filtered to reject background noise and singly labeled molecules and select for bona fide FRET-labeled molecules. This selection is typically done by eye and may be subject to user bias; to minimize such bias, it is helpful to establish a set of defined selection criteria. The more complex the trajectories the more relevant the establishment of objective selection criteria, since the behavior of molecules will often be heterogeneous and a single molecule may go from little or no (FRET ≈ 0) to high energy transfer efficiency (FRET ≈ 1). Perhaps the two simplest criteria to implement are the presence of photoactive fluorophores and some level of FRET between them. These criteria are satisfied when molecules exhibit significant anticorrelation in donor and acceptor fluorophore intensities (detected by eye or using mathematical correlation analysis), followed by single-step photobleaching of one or both fluorophores. For molecules locked in low-FRET states, anticorrelation may not be easily observed, but the presence of an active acceptor fluorophore can be established upon direct acceptor excitation by a laser pulse at the end of the desired observation window (or intermittently). A photoactive acceptor is characterized by a sharp increase in emission upon such illumination (above a threshold to be established), regardless of the FRET state of the molecule; this feature helps distinguish molecules that are doubly labeled but reside in a low-FRET state from those that are singly labeled. It is also helpful to define as a criterion an expected signal threshold based on the experimental background and signal-to-noise ratio, and to exclude molecules whose donor and acceptor trajectories are significantly positively correlated (Munro *et al.*,

2007). The remaining time trajectories will provide the pool of molecules from which quantitative data can be extracted.

2.2. Analysis of FRET state distribution

One of the first and most straightforward analysis steps is to create a FRET probability distribution by sampling the FRET values from an ensemble of molecules. Such analysis can be accomplished, for example, by binning all FRET values into a histogram that arise from a 10-s segment of all molecule trajectories observed (sometimes a single long molecule trajectory can be analyzed similarly). This simple procedure yields the relative occupancy of FRET states within the molecule population, as well as the associated relative interfluorophore distances. To this end, fitting with a sum of Gaussian functions as implemented in graphing software such as Microcal Origin is used to model the FRET distribution and obtain quantitative information on the mean FRET values, distribution widths (which is largely shot noise derived), and relative abundances of each molecule conformation (or state) detected (Fig. 9.2A). In Microcal Origin, this is accomplished using the “Fit MultiPeaks” analysis routine and selecting approximate peak centers and widths for initial guesses. For Fig. 9.2, a two-state system was simulated wherein 100 molecules reversibly interconvert between FRET states of 0.2 and 0.8 with rate constants of 0.1 and 0.6 s⁻¹ (inset of Fig. 9.2A) over 1000 data points each. The result in Fig. 9.2A shows that the multipeak fitting procedure is able to identify both FRET states used in the simulation. Even though this procedure works reasonably well, it has been shown to be subject to bias from arbitrary values such as the chosen bin size (Okamoto and Terazima, 2008). Such inaccuracies become more problematic with an increasing number of states and a poor signal-to-noise ratio since the overlap between Gaussian distributions will become more severe. The utility of analyzing FRET distribution histograms is thus limited in the case of complex, multistate single-molecule trajectories.

2.3. Kinetic analysis with thresholding algorithms

FRET distribution histograms yield only the identity of FRET states in a population of molecules (or possibly a single long molecule trajectory), but do not extract any kinetic information on their dynamics. For this purpose, one needs to first reliably and efficiently identify the FRET states a molecule is sampling, then calculate the dwell times, or the amount of time spent in a state before transitioning to a specific other state, for all observed state transitions. In systems with FRET states that are easily distinguished by eye and/or histogram analysis, it is possible to perform kinetic analysis with the use of simple thresholding. Thresholding is performed by implementing an algorithm wherein a FRET value that has little to no occupancy

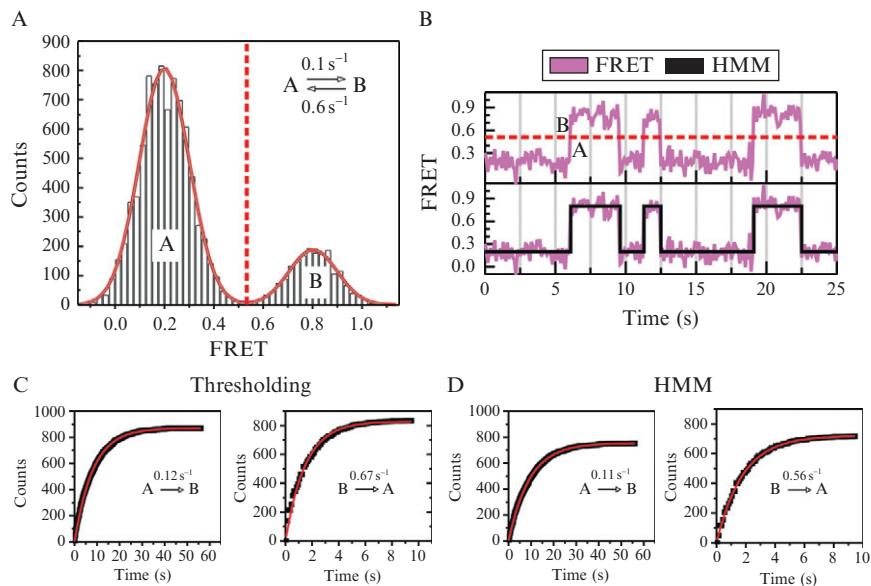


Figure 9.2 Histogram distribution analysis, thresholding, and hidden Markov modeling are well suited for simple trajectories. (A) Simulated FRET probability distribution created by simulating 10 s at 100 ms repetition rate (1000 data points) each of 100 molecules based on a two-state model (inset). The distribution was fit using two Gaussians whose sum models the distribution well (red outline). The centers of the Gaussians as determined by the fitting routine are situated at 0.203 and 0.802. (B) A sample trajectory of the simulated two-state system with discrete FRET states determined through either thresholding (top) or hidden Markov modeling (HMM) (bottom). (C) Dwell time analysis of the simulated two-state system after determining the dwell times using the thresholding method or HMM. Measured dwell times were binned and plotted as a cumulative distribution and then fit with single-exponential functions using Microcal Origin.

(as determined by state distribution analysis, Fig. 9.2A) is assigned as the threshold point. The algorithm then scans the FRET values within a trajectory and assigns each to a state based on its position relative to the threshold value (Fig. 9.2B, top panel). To avoid noise-induced assignment of artificial transitions, an additional requirement may be imposed that a new state be only assigned if there are two or more consecutive data points in a new position relative to the threshold value. After each data point within a trajectory is assigned a state, dwell times can be calculated, allowing for the determination of kinetic rate constants for state transitions. Rate constants are calculated directly by plotting dwell times as a cumulative probability distribution (or as a noncumulative probability density function) that is then fit with a single- or, if necessary, multiexponential growth curve (left side of Fig. 9.2C). An example where such analysis has been successful

is the two- to three-state system of the hairpin ribozyme (Bokinsky *et al.*, 2003; Ditzler *et al.*, 2008; Liu *et al.*, 2007; Rueda *et al.*, 2004; Zhuang *et al.*, 2002). Thresholding can be used reliably only for systems with three or fewer states since with more states threshold values become difficult to assign due to increasing overlap of the Gaussian distributions representing each state.

3. ANALYSIS OF COMPLEX TRAJECTORIES

Due to the complex conformational behavior of many biomolecules, simple two- or three-states FRET systems may be rather the exception than the rule. Complex trajectories may arise from the presence of multiple interconverting structural or chemical states, and/or from transacting factors in solution that interact with the labeled molecule. Limited time resolution of the detector relative to the FRET state dwell times can also complicate trajectories (Lee, 2009), as well as photophysical effects such as blinking and dye anisotropy changes that lead to abrupt changes in FRET signal, often lacking anticorrelation of the donor–acceptor signal. Blinking can be detected with relative ease since the fluorescence intensity of the affected fluorophore drops to background level, and this portion of a trajectory can be ignored in (removed from) the analysis. Dye anisotropy changes due to local fluctuations in dye environment are more subtle effects that are harder to detect and are often disguised if relying solely on the FRET ratio for analysis. Distinction of true FRET changes from these photophysical artifacts becomes increasingly difficult with an increasing number of states that occupy more of the limited FRET value range of 0–1. We therefore found that it is not sufficient to rely solely on the FRET ratio in the analysis of more complex trajectories, and that additional confidence can be gained from concomitantly analyzing the donor and acceptor signals (Abelson *et al.*, 2010).

3.1. Hidden Markov analysis of complex FRET trajectories

Hidden Markov modeling (HMM) is a statistical algorithm that has been used for applications as varied as speech recognition, sequence alignment, and now smFRET analysis (Eddy, 2004; McKinney *et al.*, 2006; Poritz, 1988). HMM is well suited for smFRET analysis because of its ability to find discrete states within noisy time series data and reliably find the most probable path through these states (Schuster-Böckler and Bateman, 2007). A hidden Markov model has three main parameter sets—the probability matrices of transition, emission, and initiation—that are optimized through an iterative process to find the set of parameters that best describe the data. The transition probability matrix contains the probabilities of any one FRET state changing to any other state in the subsequent time step.

The emission probability matrix contains the probabilities of a specific FRET signal being emitted by each discrete FRET state. The initiation probability matrix gives the probabilities of starting at each of the possible discrete FRET states. Typically, an HMM is “trained” by using expectation maximization algorithms such as the Baum–Welch algorithm to iteratively modify these three parameter sets to better describe the data. The Viterbi algorithm, for example, can then be used to find the most likely sequence of states (MLSS) based on the trained Markov model parameters. For an in-depth discussion of the mathematical foundations underlying HMM, we refer the reader to [Lawrence and Rabiner \(1989\)](#).

HMM has been successfully applied to smFRET analysis through the use of programs such as HaMMy ([McKinney *et al.*, 2006](#)), QuB ([Qin and Li, 2004](#)) and vb-FRET ([Bronson *et al.*, 2009](#)). These publicly available programs allow for the application of HMM algorithms to smFRET trajectories, and the extraction of FRET states and rate constants of their inter-conversion. These HMM programs present the most accessible form of data analysis that produces the most reliable results with minimal *a priori* assumptions required from the user. Additionally, they perform as well if not better than thresholding algorithms. This point is demonstrated by applying HMM to the simulated system of [Fig. 9.2A](#), resulting in the representative hidden Markov model shown in the lower panel of [Fig. 9.2B](#). Even for this simple two-state system, HMM achieves the best agreement with the raw data in terms of both kinetic rate constants and identity of the underlying FRET states, with less need for user defined (and possibly biased) input (compare [Fig. 9.2C and D](#) with inset of [Fig. 9.2A](#)). For more complex smFRET trajectories, such as those of the simulated five-state system of [Fig. 9.3A](#) (100 molecules, 1000 data points per molecule), it is even clearer that thresholding and distribution analysis do not suffice. A sample trajectory in [Fig. 9.3B](#) (top panel) shows occupancy of nearly every point in the FRET range, making choosing a threshold value nearly impossible and largely arbitrary. The data are modeled well, however, with a five-state hidden Markov model that accurately identifies the underlying FRET states despite the data noise and correctly detects when the states are changing ([Fig. 9.3B](#), lower panels). The FRET probability distribution of these same data ([Fig. 9.3C](#)) shows what appears to be a simple two-state system, obscuring the three additional states used in the simulation, due to the effects of binning and noise overlap between states.

It should be noted that Markov models generally assume that each stochastic transition is governed by a single rate constant with exponentially distributed waiting times. Molecular heterogeneity and limited observation windows often are, however, inherent to smFRET trajectories and cause nonexponentially distributed passage times and thus non-Markovian dynamics ([Bokinsky *et al.*, 2003](#); [Ditzler *et al.*, 2008](#); [Fiore *et al.*, 2009](#); [Rueda *et al.*, 2004](#); [Zhuang *et al.*, 2002](#)). In situations where non-Markovian

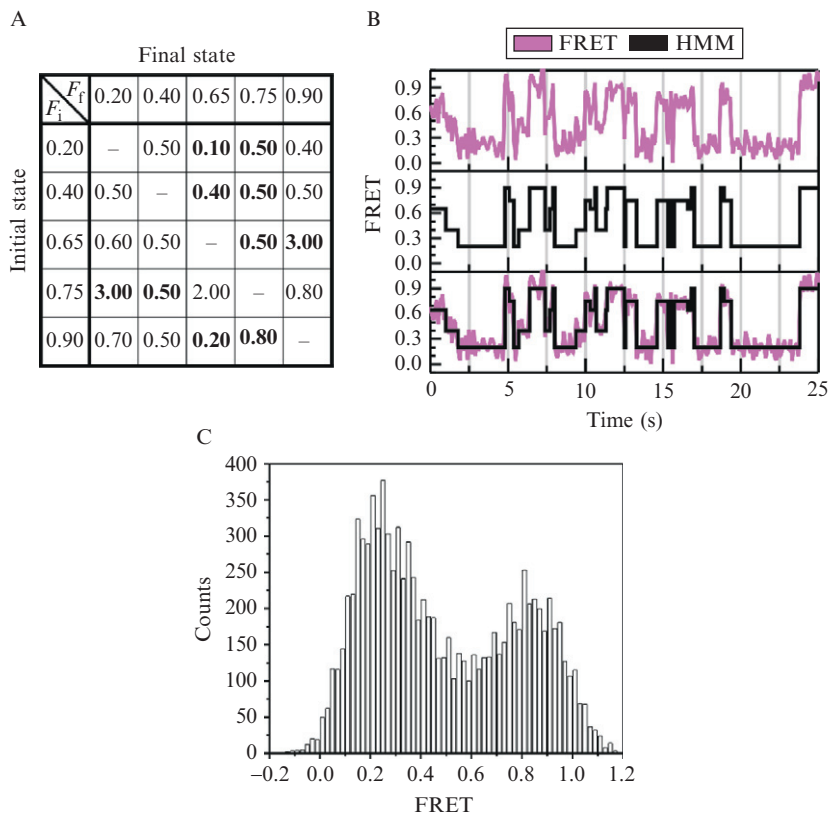


Figure 9.3 Complex trajectories are not amenable to thresholding and distribution analysis, but require hidden Markov modeling. (A) Parameters used for simulating a five-state system. FRET states and their transition rate constants are depicted as a transition matrix with the slowest and fastest rate constants (all in s^{-1}) in bold. F_i = initial FRET state; F_f = final FRET state. (B) Raw FRET trajectory of one exemplary molecule simulated with the parameters in (A) (top). The idealized trajectory as determined by HMM executed in QuB (middle) and both trajectories overlaid (bottom). (C) FRET probability distribution of the simulated five-state system. This histogram shows that even in simulated data the five underlying states are obscured by noise, rendering distribution analysis less informative as in simple two-state systems.

behavior is detected, one can still apply HMM as a practical solution for the reliable identification of FRET states and quantification of the associated transition kinetics (Talaga, 2007).

The remainder of this chapter is dedicated to the practical aspects of using the programs HaMMY, QuB, and vb-FRET, and presents advances in their use derived from their application to the smFRET characterization of the conformational dynamics of a yeast pre-mRNA during splicing *in vitro* (Abelson *et al.*, 2010).

3.2. Overview of HMM software available for smFRET analysis

The program HaMMY by Ha and coworkers presented the first readily available software specifically tailored for HMM analysis of smFRET trajectories (McKinney *et al.*, 2006). The program employs HMM algorithms with a simple graphical user interface that processes the donor and acceptor intensities to determine the underlying discrete FRET states as well as the MLSS through a trajectory. HaMMY's user interface requires minimal input from the user, mainly guesses as to the number and mean FRET values of the states, and a choice of reestimation algorithm. The simplicity of this program is perhaps its best attribute. One simply loads the single-molecule trajectories formatted as *.dat files with a time column, followed by donor and acceptor intensity columns. The program will then analyze each trajectory and output three files that include the idealized FRET path, the dwell times, and a count of the total number of transitions between FRET states. This information can be utilized with the accompanying transition density plot (TDP) program to extract dwell times and rates for transitions of interest. We have successfully applied HaMMY, for example, to three-state smFRET trajectories of VS ribozyme folding (Pereira *et al.*, 2008). The HaMMY and TDP programs and manuals are freely available at <http://bio.physics.illinois.edu/>.

The program QuB (available at <http://www.qub.buffalo.edu/>) was not originally designed for smFRET trajectories, but instead for single-ion channel measurements (Qin and Li, 2004). The electrophysiology field has been performing single-molecule measurements on ion channels for over 30 years (Neher and Sakmann, 1976), and the signals acquired from such experiments have the same underlying features as smFRET trajectories with discrete states obscured by noise, and of interest are similarly the values of these states and their rate constants of interconversion. QuB has been successfully applied, for example, to smFRET experiments on the ribosome (Munro *et al.*, 2007) and the spliceosome (Abelson *et al.*, 2010). The graphical user interface is somewhat more challenging to master than that of HaMMY, but it provides the user with greater flexibility and control over the model parameters. For example, it provides the user with a variety of algorithms for reestimation, MLSS calculations, and the ability to impose constraints on a model. One current major advantage of QuB is that it performs HMM much more rapidly than either HaMMY or vb-FRET.

The program vb-FRET by Gonzalez, Wiggins, and coworkers (available at <http://vbFRET.sourceforge.net>) is yet another option for HMM analysis tailored for smFRET trajectories (Bronson *et al.*, 2009). It provides an easy-to-use graphical interface similar to that of HaMMY, but with more customizable options similar to QuB. It is also a MATLAB executable file that for our use has been more stable than other choices. In addition to determining optimal parameters for the hidden Markov model, vb-FRET

uses an approach known as “maximum evidence” to select the most likely model from a set of models being tested (Bronson *et al.*, 2009). This approach alleviates the common problem of overfitting data as a consequence of the use of a maximum likelihood approach, which often leads to the introduction of uninformative (superfluous) parameters to model the data, as well as the need for postprocessing efforts.

Table 9.1 lists the major differences between the three HMM programs. In the following, we provide an example protocol for the use of each program, which may not necessarily be the best choice for all possible datasets but has worked effectively in our hands for both simple and more complex smFRET trajectories.

3.3. Preprocessing trajectories for analysis by HMM

3.3.1. Removal of outliers

Often single-molecule trajectories exhibit a state with a FRET value near the boundary of the range of 0–1 that will show transient excursions to FRET values either below 0 (when the acceptor intensity briefly dips into negative territory due to noise in the data) or above 1 (when the donor signal goes into negative territory). For trajectories dwelling extensively in low- or high-FRET states this phenomenon may lead to a significant occupancy of a virtual state below zero or above unity, which in turn may result in it being fit with a state by the HMM software. Consequently,

Table 9.1 Summary of features for HMM analysis programs

Program	Model selection	Time to analyze five-state system with 10^5 data points	Data range
HaMMy	Yes, BIC	~3 h 50 min	[0, 1]
QuB	No, postidealization BIC by user	~15 min	[0, 32,767]
vb-FRET	Yes, maximum evidence	~1 h	[0, 1], but can accept beyond this range

A five-state system was modeled to simulate a series of complicated smFRET trajectories. The simulation included 10^5 total data points (100 molecules with each 1000 frames of data collected) which is on par with what is expected from a series of sm-FRET experiments. This dataset was independently analyzed with the various HMM analysis programs. Model selection was carried out with the indicated techniques, and the total time of analysis was recorded. The column “Data range” indicates the acceptable input for the FRET, donor, and acceptor trajectories for each program. Because HaMMy and vb-FRET were designed to accept input of the range [0, 1] the donor and acceptor trajectories need to be scaled accordingly.

the analysis will be complicated by introducing irrelevant states and by interrupting dwell times of more relevant states. The removal of outliers can be accomplished by the implementation of a simple algorithm that finds outliers and scales them back to within the FRET range of 0–1. Generally this algorithm involves identifying the points beyond the allowable FRET range and normalizing them to within the 0–1 range by taking an average of adjacent FRET values. A script containing this algorithm is available upon request.

3.3.2. Smoothing (noise reduction)

Single-molecule trajectories are sometimes smoothed to help average out the inherent noise of the data collection process and emphasize the discrete states present. Although this is often useful for simple trajectories, complications arise if a larger number of states with rapid interconversion kinetics are present. One simple method of smoothing, rolling (point) averaging, may obscure transitions with dwell times that are shorter than the averaging window and introduce false FRET states in molecules where two or more states are rapidly interconverting. Rolling point averaging can work well with simple trajectories, but breaks down for more complex trajectories.

A nonlinear forward–backward filter introduced by Haran has been used to smooth single-molecule trajectories while minimizing their distortion, offering a clear advantage over rolling point averaging (Haran, 2004). We have applied this nonlinear filter, for example, to trajectories of VS ribozyme folding where it reduces noise and emphasizes conformational changes between the three, clearly separated states (Pereira *et al.*, 2008). One possible problem with this filter is the fact that the noise profile will no longer be Gaussian in shape, which is an assumption for analysis by HMM. In practice, this feature may lead to the fitting of small FRET changes that are within the noise of the raw data.

3.3.3. Formatting for HMM analysis

To perform HMM analysis, it is necessary that smFRET trajectories be formatted in a manner amenable for manipulation with the desired analysis program. It should be noted that we are presenting only a limited, most commonly used set of input formats for each program, and for a more extensive list of input and output formats one should refer to the downloadable manuals for HaMMMy (<http://bio.physics.illinois.edu/HaMMMy.html>), QuB (http://www.qub.buffalo.edu/wiki/index.php/Main_Page), and vb-FRET (<http://vbfret.sourceforge.net/>). HaMMMy loads files in ASCII format (e.g., *.dat) with the tabulated structure “time, donor intensity, acceptor intensity”. This structure is convenient as it is commonly the form in which the data are extracted from experiment. They can be loaded individually or in batch. HaMMMy expects data values to fall within the range of [0, 1] and the data should be scaled to fit as needed. Based on the standard FRET ratio (9.1), values outside of 0 and 1 do not have any physical

meaning and should be considered outliers that can be corrected as described above.

vb-FRET accepts data in ASCII format as does HaMMY, allowing one to load the same data into both programs for comparison. Although not formally supported, we have tested the program with data scaled outside of [0, 1] and found idealizations to work well.

QuB loads data in a variety of formats, including ASCII, without a need for normalization. QuB batch analysis can be performed by loading all molecules into a single *.txt file with each molecule separated from its neighbors by line breaks. QuB allows for the analysis of multiple channels in parallel, so one can visualize the donor, acceptor, and FRET ratio into the same file for observation and analysis. Scripts for the stitching and resegmenting of idealized trajectories are available upon request. We have also written a simple script that converts HaMMY formatted data into a *.txt file for input into QuB and comparison.

3.3.4. Stitching trajectories

It is often the case that a single trajectory does not display all possible transitions due to the limitations of photobleaching and the presence of some long-lived states that occupy much of the available time window. We have found that this scenario can limit the effectiveness of HMM programs in determining the discrete states for each molecule because there will be little to no occupancy of some states in some molecules. Additionally, the analysis of trajectories separately leads to HMM algorithms fitting the same conformational states with multiple idealized FRET values since subtle differences in background between trajectories can shift the value of the FRET state up or down. The variability of the values of discrete FRET states found when fitting trajectories separately requires that postprocessing steps are carried out to group similar states before kinetic rates can be calculated for dwell time analysis. Such grouping of similar FRET states, however, requires a user-established criterion that is independent of the HMM analysis, which slows down the analysis routine and diminishes the objectivity of the analysis.

A better solution to this problem we found to be to “stitch” trajectories together so that a global analysis can be performed on an entire dataset at once, which provides more data for the HMM algorithm to calculate reliable transition probabilities. This approach can be particularly helpful for shorter trajectories, but care should be taken to resegment trajectories from one another after Markov modeling to prevent the introduction of false transitions at the molecule boundaries. Additionally, it is good practice to independently analyze a subset of molecules, large enough to recapitulate the behavior of the entire population, in both manners and compare the fitting results and HMM rates and states to ensure convergence. Scripts for the stitching and resegmenting of idealized trajectories are available upon

request. HaMMMy will only analyze segments containing up to 50,000 data points, whereas vb-FRET has been tested for up to 150,000 data points and the stability of the program is not affected. For both of these programs the idealization of such large trajectories will take a long time (Table 9.1), and in our hands has affected the stability of HaMMMy. QuB has performed best with the stitched trajectories since it performs Markov modeling much more quickly than either of the currently available versions of HaMMMy or vb-FRET (Table 9.1).

3.4. Selecting the appropriate number of FRET states

Perhaps the most difficult task in HMM analysis is deciding on the number of states to fit a dataset with, a general problem when using statistical models to approximate an experimental distribution. For simple trajectories the number of states is often easily discernable by eye, but for complex trajectories this task becomes more difficult. Finding an objective and reproducible manner of choosing the appropriate number of states requires the use of a “rule.” The HaMMMy manual suggests to allow for two more states than the total numbers of states one assumes to be present in the dataset. The program then has the flexibility to sample higher order models, at the sacrifice of calculation speed. Inherently, a likelihood score based on the total probability calculated will continue to increase with an increase in the number of model parameters (Bronson *et al.*, 2009). Therefore, the goal should not be to maximize the likelihood score, but instead to strive for model parsimony, that is, to maintain the simplest model that best describes the data. HaMMMy aims for model parsimony by calculating the total probability for each model tested and then using the Bayesian Information Criterion (BIC) to decide on the most appropriate model. BIC corrects for overfitting, a common problem when using maximum likelihood approaches for determining model parameters, by introducing a penalty for complexity (Wasserman, 2000):

$$\text{BIC} \equiv -2\ln(\text{LL}) + k\ln(N), \quad (9.2)$$

where LL is the maximum likelihood reached by the model, k is the number of parameters, and N is the number of data points used in the analysis.

QuB in addition outputs a log-likelihood (LL) score for each model that can be used to compare the results of two independent idealizations and, when used to calculate the BIC, can help select the most appropriate model. The BIC calculation has to be performed outside of QuB as there is currently no nested model selection within the QuB algorithm, which speeds it up relative to the other analysis programs.

vb-FRET takes a different approach to model selection, mainly the use of maximum evidence instead of maximum likelihood for idealization

(Bronson *et al.*, 2009). Maximum evidence aims to select the most likely model (and not only the most likely parameter values) through the assumption that the model with the correct number of states has the highest probability of reproducing the experimental data. Maximum evidence can be thought of as the probability of selecting the best parameterized model from the pool of all possible Markov models of a defined order (number of states). In a simple example, it is highly unlikely that a system with three distinct states will ever accurately be reproduced by a two-state Markov model. Conversely, a four-state model can reproduce the three-state system, but is overly complex, thus lowering the confidence in such a higher order model. This leaves the three-state model as the best choice since it can both reproduce the experimental data and is simple enough that the properly parameterized model is more likely to be chosen from the set of all possible outcomes.

4. POST-HMM PROCESSING AND DATA VISUALIZATION

Traditionally when analyzing smFRET trajectories, the FRET ratio has been the only metric by which transitions have been detected, under the assumption that the donor and acceptor traces are always anticorrelated. When dealing with more complex trajectories, however, it becomes meaningful to maximize the observables to better understand the underlying dynamics of the molecules being studied. Incorporating the donor and acceptor trajectories, in particular, becomes necessary for systems with more than three states since uncorrelated changes between the dyes will often lead to a FRET value that exhibits some occupancy and therefore may appear as a bona fide FRET state. Such uncorrelated changes may arise from subtle variations in the local environment around the fluorophores such as changes in their rotational diffusion behavior (anisotropy) and fluorescence decay pathways.

4.1. Local detection of correlation based on HMM

Although the FRET ratio provides us with a bounded metric for conformational change, and can help correct for problems such as focal drift of the microscope, it is sensitive to unilateral changes in fluorescence intensity in either fluorophore. In complex systems, where cofactors may be acting on the molecule of interest in trans or the fluorophore is placed in a central position relevant for activity but exposed to a complex environment, the local environment around one fluorophore may not remain constant, which in turn may lead to changes in FRET ratio that are not caused by changes in FRET efficiency. Such apparent FRET changes uncorrelated in donor and

acceptor signal may be intermixed with true changes in FRET efficiency, making their distinction within an smFRET trajectory nearly impossible. As a practical solution to this problem, we have adopted a strategy that analyzes by HMM the donor and acceptor trajectories alongside the corresponding smFRET trajectory, and subsequently scores each FRET change observed based on the presence or absence of corresponding donor and acceptor signal changes. Implementation of this algorithm requires that first the corresponding donor, acceptor, and FRET trajectories are analyzed as follows.

HaMMy requires that the input data be scaled to fit between 0 and 1 so that the donor and acceptor intensities need to be normalized accordingly. In addition, HaMMy was designed to analyze FRET trajectories and therefore computes the FRET ratio (Eq. (9.1)) from its input (“time, donor intensity, acceptor intensity”). As a consequence, if one is interested in analyzing the donor channel, for example, the data should be formatted as: “time, 1-donor intensity, donor intensity.”

For vb-FRET, the same input format as HaMMy can be used, except that no normalization is required.

QuB does not require that the data values lie between 0 and 1 so that no normalization is necessary. In addition, QuB's data input allows for flexibility so that the donor, acceptor, and FRET trajectories can all be loaded at once and analyzed independently by specifying how many channels are present in the dataset. Alternatively, it also allows for each channel to be loaded separately.

Once each channel has been idealized using HMM algorithms, we employ a simple algorithm in MATLAB to detect donor-acceptor uncorrelated changes. The algorithm first scans the idealized FRET trajectory for each FRET change, then finds the corresponding time point in both the donor and acceptor trajectories to determine if there was a substantial change in idealized fluorescence intensity, and finally scores the transition using the simple scoring system outlined in Fig. 9.4. Each FRET transition is scored based using the same criteria, and the dwell times of FRET states are corrected after donor-acceptor uncorrelated FRET transitions have been identified and removed from consideration. Dwell time correction is important since uncorrelated changes in FRET otherwise distort (accelerate) the kinetics measured for a system since they shorten the apparent dwell times in specific conformational states.

4.2. Data condensation and visualization

HMM techniques provide an efficient method for extracting quantitative measures of FRET states and rate constants of transitions between them from single-molecule trajectories. With increasing complexity of smFRET trajectories, the need to assess these parameters in a condensed and comprehensive representation becomes more evident. Of particular interest are

transitions between pairs of sequential FRET states. Along with the release of HaMMy, Ha and coworkers released a program that visualizes FRET transitions as so-called TDPs, where the number of times a transition occurs

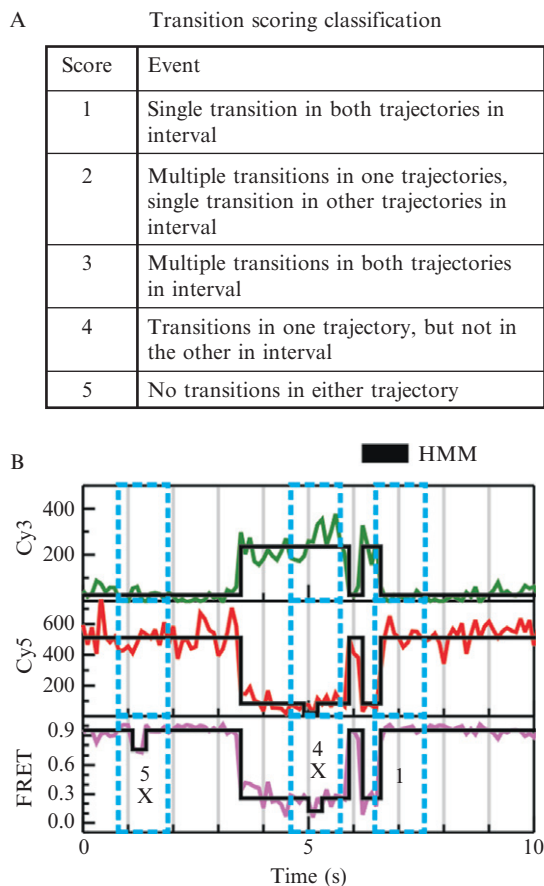


Figure 9.4 Local correlation analysis utilizing HMM algorithms and transition scoring. (A) Transition quality scoring used to exclude artificial FRET transitions caused by unilateral changes in one fluorophore. In our studies we are interested in only transitions with scores 1–3, since these transitions exhibit anticorrelated changes in the fluorescence intensity of both fluorophores simultaneously, as indicated. The time window used to search for transitions in the donor and acceptor trajectories is set by examining the FRET trajectory. The size of the time window we chose to relate to the dwell time immediately before the FRET transition being scored. (B) Experimental trajectory and the scores of highlighted transitions after idealizing the donor signal, acceptor signal, and FRET using hidden Markov modeling (black lines). Transitions marked with an x are not characterized by donor–acceptor anticorrelation and not considered in any further analysis. Modified in part from [Abelson *et al.* \(2010\)](#).

is plotted as a heat map on a two-dimensional grid of final versus initial FRET state (Fig. 9.5; McKinney *et al.*, 2006). TDPs highlight the most prevalent transitions within a population of molecules and require that the transition rate constants be represented in a secondary plot (Joo *et al.*, 2006; Pereira *et al.*, 2008). The mirror symmetry relative to the main diagonal often observed in this type of plot is a result of reversible conformational changes (Fig. 9.5). It is important to note that the number of possible transitions that can be mapped onto a TDP is $N*(N - 1)$, where N is the number of discrete states found in the HMM analysis. For example, a trajectory that contains five states can have up to 20 distinct positions on the TDP, leading to a possibly quite complex plot. Importantly, transitions with slow kinetics will show up only infrequently in the trajectories due to their long dwell times relative to the limited observation window imposed by the photobleaching rate of the fluorophores.

To help minimize this relative underrepresentation of slow compared to fast transitions in a TDP, we have developed complementary transitional occupancy density plots (TODPs) that scale transitions based on the fraction of molecules that exhibit them at least once, rather than the number of times they are observed over all molecules. This approach makes slow transitions that many molecules exhibit more visible on the heat map of a TODP than a TDP, and thus guards against a visual overrepresentation of unrepresentative fast transitions only few molecules exhibit (Fig. 9.5).

To also incorporate the kinetics of each transition into one and the same plot, we have further developed POPulation-weighted and Kinetically Indexed Transition density (POKIT) plots. POKIT plots thus provide two additional, comprehensive pieces of information compared to TDPs. First, they present as a number of concentric circles the fraction of molecules in the entire dataset that exhibits a specific FRET transition at least once (the information represented in TODPs as heat map). Transitions that are common in a majority of molecules can help identify, for example, conformational changes that are important in a reaction with intermediates leading to products, even if this reaction is irreversible and the transition occurs only once per molecule. Second, POKIT plots provide the average dwell time for each transition in the form of circle colors, facilitating the rapid visual comparison of the kinetics of various datasets (Fig. 9.5).

4.3. Applications to single-molecule studies of yeast pre-mRNA splicing

Eukaryotic pre-mRNA splicing is a dynamic process that involves the precise recognition and excision of intervening sequences from in between coding regions (exons) of transcribed pre-mRNAs. Splicing is carried out by a multimega Dalton ribonucleoprotein complex known as the spliceosome. Although rivaling the ribosome in size, the spliceosome is unique in

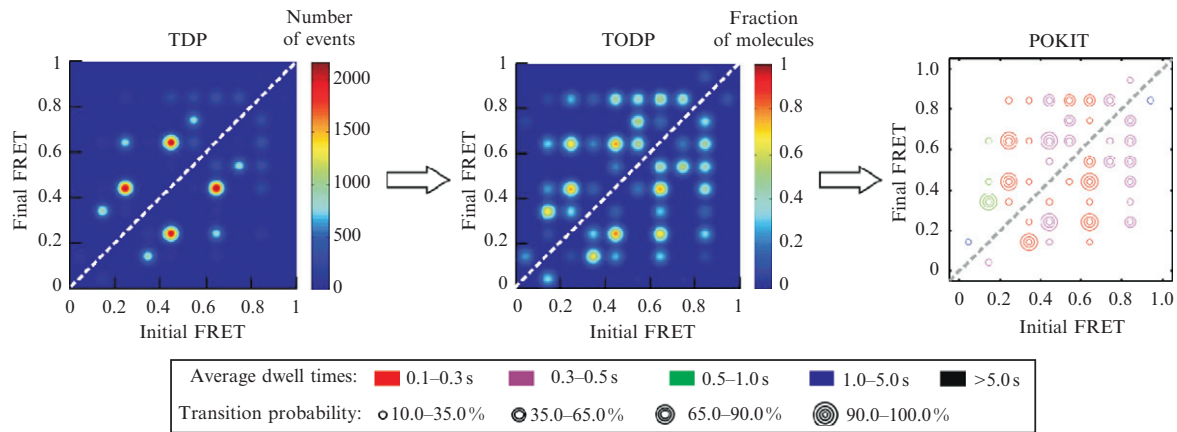


Figure 9.5 Data visualization for complex trajectories, including TDP and POKIT plot analysis. A side-by-side comparison of the same dataset using three representations. Traditional TDPs are scaled by the number of times a transition is observed over all molecules, regardless of whether in only a small subpopulation of molecules with rapid transitions or commonly in all molecules. TODP and POKIT plots are scaled by the fraction of all molecules within a population that exhibits a particular transition. POKIT plots additionally provide kinetic information encoded in the color of the concentric circles. Modified from [Abelson *et al.* \(2010\)](#).

that it lacks a preformed catalytic core. Instead, assembly proceeds in a stepwise manner, and is influenced by the ATPase activity of several RNA-dependent ATPases. Some of the rearrangements involved include the disruption of RNA–protein, RNA–RNA interactions, and the binding of recognition sequences and accessory proteins that lead to the formation of the catalytically competent complex necessary for the two transesterification reactions of intron excision and exon ligation. Yeast genetics and the development of *in vitro* yeast splicing assays have allowed for the biochemical characterization of the components involved in splicing as well as the key assembly steps required for splicing.

The introduction of smFRET has now allowed us to directly observe the conformational dynamics of the pre-mRNA substrate during spliceosome assembly and catalysis in real-time (Fig. 9.1A; Abelson *et al.*, 2010). The complexity of smFRET trajectories from these experiments (Fig. 9.1B) reflects the overall complexity of the rearrangements needed for splicing activity, as well as the asynchronous behavior of splicing in total yeast cell extract. These features lead to a diversity of FRET states, heterogeneous kinetics, and many occupied transitions between the states as a reflection of the range of conformations through which the pre-mRNA substrate is shuttling. To highlight those FRET transitions that are relevant to splicing it was necessary to compare the wild-type substrate with mutant substrates that are blocked at different stages of splicing (Abelson *et al.*, 2010). Spliceosome assembly on a branchpoint mutant (BP) is impaired, leading to a complete lack of splicing activity. In the 3' splice site mutant (3'SS) the second step of splicing and thus exon ligation is blocked. A detailed kinetic and conformational analysis of the substrates in ATP depleted or ATP-supplemented extract allowed us to identify conformational states that are required for splicing activity (Abelson *et al.*, 2010).

The analysis of the smFRET data goes as follows: First, trajectories to be studied are prefiltered by searching for the presence of any substantial anticorrelation by visual inspection. The raw donor, acceptor, and FRET trajectories are independently Markov modeled to determine transition boundaries, using a global fitting routine (stitched trajectories) that simultaneously analyzes all data points taken under a given experimental condition. The entire dataset for each condition is analyzed by the iterative application of the Forward–Viterbi and Baum–Welch algorithms in the QuB program to generate idealized trajectories. The number of states assumed in the idealization is varied from 5 to 11, with all states initially being assigned equal probabilities and rate constants that then are iteratively optimized; the resulting fits are evaluated using the BIC. The model that results in the best BIC score is selected for further analysis. This process is performed on all three corresponding trajectories (donor, acceptor, and FRET). After idealization, a post-HMM processing algorithm (coded and executed in MATLAB) classifies each FRET transition by counting the number and

direction of transitions found in the donor and acceptor trajectories within a time window defined to begin one quarter of the immediately preceding dwell time before the FRET transition in question and ending one quarter of the following dwell time after this FRET transition, with a minimum set to 0.3 s in either direction from the transition. Each transition is scored based on the metric shown in Fig. 9.4A, and transitions with scores of one, two, or three are used for TDPs. Additionally, FRET transitions with a FRET change smaller than 0.1 are not considered significant and consequently removed.

The idealized scored trajectories are used to create TDPs and POKIT plots to examine the conformational and kinetic differences between the mutants under the various experimental conditions. Both TDPs and POKIT plots show substantial differences between the mutants and various conditions, with the POKIT plots highlighting even more subtle differences. The POKIT plots show, for example, that after extended incubation of only the WT substrate in ATP-supplemented extract a population of molecules with a relatively stable high-FRET conformation arises that resembles that of the mature mRNA after splicing (Abelson *et al.*, 2010).

4.4. Summary of a detailed strategy for analysis of complex trajectories with QuB

Molecule selection; requirements for an accepted trajectory are

- (1) donor–acceptor anticorrelation; and
- (2) presence of each a single donor and acceptor fluorophore as verified by the observation of: (I) single-step photobleaching and (II) emission of Cy5 upon direct excitation.

4.4.1. Preparing data for distribution analysis and HMM

- (1) *Background correction*: A straightforward method of background correction involves subtracting the mean value of signal after photobleaching from each value in the trajectory.
- (2) Remove blinking events, and truncate the trajectory at the point of photobleaching.
- (3) Output raw data in *.dat format (format “time, donor signal, acceptor signal”).
- (4) Sample 10 s worth of signal from each molecule and create a histogram from binned data using Microcal Origin.
- (5) Stitch (concatenate) all trajectories into a single trajectory, and use the frequency count function in Microcal Origin with a bin size of 0.05 or less.

4.4.2. Preprocessing

- (1) *Outlier removal*: Trajectories are inspected for data points outside of the FRET range of $[0, 1]$. These points are corrected based on the values of the adjacent points in the trajectory. If the background correction is performed properly there should be few outliers. If the number of outliers is roughly $\sim 10\%$ or more of the data points of a trajectory this hints at incorrect background correction, or that blinking or photobleaching events have not been properly removed.
- (2) *Stitching*: All FRET trajectories of a single experimental condition are stitched together, and formatted for QuB. This procedure is also performed for the donor and acceptor trajectories. The QuB input format is a *.txt file with the data columns “time, donor intensity, acceptor intensity, and FRET” that can be truncated to two columns such as “time, donor intensity” to read out the donor or acceptor trajectories.

4.4.3. HMM analysis with QuB

- (1) *Import data into QuB*: Specify the number of columns in the data file and whether or not there is a time column. We typically choose to analyze the donor signal, acceptor signal, and FRET ratio as separate files to allow for better visualization of idealized trajectories that are overlaid on the raw data.
- (2) *Create a hidden Markov model*: In the modeling window of QuB begin by creating a model with the lowest number of states assumed to be possible (typically we start our models for complex trajectories with five states).
- (3) *Idealization*: The amplitudes and standard deviations of the states are then estimated by using the “Amps” function in QuB. This will initiate the model, and then the Baum–Welch and Forward–Viterbi algorithms are used under the “Idl/Base” menu for idealization. There are several other algorithms available for idealization; SKM has been used successfully, for example, to model smFRET data from ribosomes (Munro *et al.*, 2007) and perform idealizations faster than the Baum–Welch, Forward–Viterbi algorithms presented here. SKM did not fit our smFRET data well, perhaps because of the lower signal-to-noise ratio due to the use of crude yeast cell extract. After each idealization add a state to the model and repeat the amplitude initiation and idealization procedures. Save each idealization as a *.dwt file (QuB file format with FRET states and dwell times), and then copy the LL score for the model from the reports panel and paste it into a spreadsheet for the calculation of BIC.

4.4.4. Postprocessing

- (1) *Model selection*: The LL score output from QuB is used to calculate the BIC for each model tested, and the model with the lowest BIC score is selected for further analysis.
- (2) *Parsing idealized data*: The *.dwt file for the selected model is read and the idealized FRET states are then matched with the raw data to create a path file that has the format “time, donor signal, acceptor signal, FRET, idealized FRET.” This path file is then segmented back into the individual molecule trajectories that were initially used to generate the stitched data.
- (3) *Transition scoring*: The transition scoring routine is run by loading the path files into a MATLAB script that then finds transitions in the FRET channel, and takes note of the directionality and number of transitions at the corresponding time point in the donor and acceptor trajectories. It then scores each transition based on the scale in Fig. 9.4A, and a scored path file is created. Transitions of scores 1–3 are considered true FRET transitions and thus chosen for further analysis.

4.4.5. Data visualization

The scored path file is then input into MATLAB scripts written to recognize the score of the transition, and create TDPs, TODPs, or POKIT plots based on this information. From the TDP and TODPs, dwell time information can be extracted from our scripts by boxing a region within MATLAB around the transition(s) of interest. For POKIT plots, the dwell times are automatically extracted for each transition, and an average is calculated and encoded by the color scheme in Fig. 9.5. The percent of molecules exhibiting a transition is calculated by counting how many molecules within the dataset exhibit that particular transition at least once, and then dividing by the total number of molecules used in the analysis.

ACKNOWLEDGMENT

The authors would like to acknowledge the work of Franklin Fuller in writing and developing several of the MATLAB scripts used for transition scoring.

REFERENCES

- Abelson, J., Blanco, M., Ditzler, M. A., Fuller, F., Aravamudhan, P., Wood, M., Villa, T., Ryan, D. E., Pleiss, J. A., Maeder, C., Guthrie, C., and Walter, N. G. (2010). Conformational dynamics of single pre-mRNA molecules during *in vitro* splicing. *Nat. Struct. Mol. Biol.* doi:10.1038/nsmb.1767.

- Aitken, C. E., Marshall, R. A., and Puglisi, J. D. (2008). An oxygen scavenging system for improvement of dye stability in single-molecule fluorescence experiments. *Biophys. J.* **94**, 1826–1835.
- Bokinsky, G., Rueda, D., Misra, V. K., Rhodes, M. M., Gordus, A., Babcock, H. P., Walter, N. G., and Zhuang, X. (2003). Single-molecule transition-state analysis of RNA folding. *Proc. Natl. Acad. Sci. USA* **100**, 9302–9307.
- Bronson, J. E., Fei, J., Hofman, J. M., Gonzalez, R. L. Jr., and Wiggins, C. H. (2009). Learning rates and states from biophysical time series: A Bayesian approach to model selection and single-molecule FRET data. *Biophys. J.* **97**, 3196–3205.
- Ditzler, M. A., Rueda, D., Mo, J., Hakansson, K., and Walter, N. G. (2008). A rugged free energy landscape separates multiple functional RNA folds throughout denaturation. *Nucleic Acids Res.* **36**, 7088–7099.
- Eddy, S. R. (2004). What is a hidden Markov model? *Nat. Biotechnol.* **22**, 1315–1316.
- Fiore, J. L., Hodak, J. H., Piestert, O., Downey, C. D., and Nesbitt, D. J. (2008). Monovalent and divalent promoted GAAA tetraloop-receptor tertiary interactions from freely diffusing single-molecule studies. *Biophys. J.* **95**, 3892–3905.
- Fiore, J., Kraemer, B., Koberling, F., Erdmann, R., and Nesbitt, D. (2009). Enthalpy-driven RNA folding: Single-molecule thermodynamics of tetraloop-receptor tertiary interaction. *Biochemistry* **48**, 2550–2558.
- Ha, T., Enderle, T., Ogletree, D. F., Chemla, D. S., Selvin, P. R., and Weiss, S. (1996). Probing the interaction between two single molecules: Fluorescence resonance energy transfer between a single donor and a single acceptor. *Proc. Natl. Acad. Sci. USA* **93**, 6264–6268.
- Haran, G. (2004). Noise reduction in single-molecule fluorescence trajectories of folding proteins. *Chem. Phys.* **307**, 137–145.
- Ishii, Y., Yoshida, T., Funatsu, T., Wazawa, T., and Yanagida, T. (1999). Fluorescence resonance energy transfer between single fluorophores attached to a coiled-coil protein in aqueous solution. *Chem. Phys.* **247**, 163–173.
- Joo, C., McKinney, S. A., Nakamura, M., Rasnik, I., Myong, S., and Ha, T. (2006). Real-time observation of RecA filament dynamics with single monomer resolution. *Cell* **126**, 515–527.
- Kapanidis, A. N., and Weiss, S. (2002). Fluorescent probes and bioconjugation chemistries for single-molecule fluorescence analysis of biomolecules. *J. Chem. Phys.* **117**, 10953–10964.
- Lawrence, R., and Rabiner, A. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE* **77**, 257–286.
- Lee, T. (2009). Extracting kinetics information from single-molecule fluorescence resonance energy transfer data using hidden Markov models. *J. Phys. Chem. B* **113**, 11535–11542.
- Liu, S., Bokinsky, G., Walter, N. G., and Zhuang, X. (2007). Dissecting the multistep reaction pathway of an RNA enzyme by single-molecule kinetic “fingerprinting” *Proc. Natl. Acad. Sci. USA* **104**, 12634–12639.
- McKinney, S. A., Joo, C., and Ha, T. (2006). Analysis of single-molecule FRET trajectories using hidden Markov modeling. *Biophys. J.* **91**, 1941–1951.
- Michalet, X., and Weiss, S. (2002). Single-molecule spectroscopy and microscopy. *Compt. Rend. Phys.* **3**, 619–644.
- Min, W., English, B. P., Luo, G., Cherayil, B. J., Kou, S. C., and Xie, X. S. (2005). Fluctuating enzymes: Lessons from single-molecule studies. *Acc. Chem. Res.* **38**, 923–931.
- Munro, J. B., Altman, R. B., O’Connor, N., and Blanchard, S. C. (2007). Identification of two distinct hybrid state intermediates on the ribosome. *Mol. Cell* **25**, 505–517.
- Neher, E., and Sakmann, B. (1976). Single-channel currents recorded from membrane of denervated frog muscle fibres. *Nature* **260**, 799–802.

- Okamoto, K., and Terazima, M. (2008). Distribution analysis for single molecule FRET measurement. *J. Phys. Chem. B* **112**, 7308–7314.
- Pereira, M. J. B., Nikolova, E. N., Hiley, S. L., Jaikaran, D., Collins, R. A., and Walter, N. G. (2008). Single VS ribozyme molecules reveal dynamic and hierarchical folding toward catalysis. *J. Mol. Biol.* **382**, 496–509.
- Poritz, A. B. (1988). Hidden Markov models: A guided tour. In “International Conference on Acoustics, Speech, and Signal Processing.” ICASSP-88, Vol. 1. IEEE Press, pp. 7–13.
- Qin, F., and Li, L. (2004). Model-based fitting of single-channel dwell-time distributions. *Biophys. J.* **87**, 1657–1671.
- Roy, R., Hohng, S., and Ha, T. (2008). A practical guide to single-molecule FRET. *Nat. Methods* **5**, 507–516.
- Rueda, D., Bokinsky, G., Rhodes, M. M., Rust, M. J., Zhuang, X., and Walter, N. G. (2004). Single-molecule enzymology of RNA: Essential functional groups impact catalysis from a distance. *Proc. Natl. Acad. Sci. USA* **101**, 10066–10071.
- Schuster-Böckler, B., and Bateman, A. (2007). An introduction to hidden Markov models. *Curr. Prot. Bioinform.* A.3A.1, A.3A.9 Appendix 3A.
- Stryer, L. (1978). Fluorescence energy transfer as a spectroscopic ruler. *Annu. Rev. Biochem.* **47**, 819–846.
- Talaga, D. S. (2007). COCIS: Markov processes in single molecule fluorescence. *Curr. Opin. Colloid Interface Sci.* **12**, 285–296.
- Walter, N. G., Huang, C.-Y., Manzo, A. J., and Sobhy, M. A. (2008). Do-it-yourself guide: How to use the modern single-molecule toolkit. *Nat. Methods* **5**, 475–489.
- Wasserman, L. (2000). Bayesian Model Selection and Model Averaging. *J. Math. Psychol.* **44**, 92–107.
- Zhuang, X., Kim, H., Pereira, M. J. B., Babcock, H. P., Walter, N. G., and Chu, S. (2002). Correlating structural dynamics and function in single ribozyme molecules. *Science* **296**, 1473–1476.