Confocal Single-Molecule Fluorescence as a Tool for Investigating Biomolecular Dynamics *In Vitro* and *In Vivo*

Joseph P. Torella, Keble College

Doctor of Philosophy

Trinity 2010

j.torella1@physics.ox.ac.uk

Acknowledgments

My first thanks go to my parents who, despite my unconventional trajectory, have supported me fully throughout it. They truly put my interests before their own; for their constant love and support, I can never thank them enough.

My thanks also go to those who contributed directly to the work contained in this thesis. Although we are supposed to write in the first person for our theses – “I attempted,” “I identified,” “I developed,” – I feel these phrases are inappropriate. Each chapter in this thesis contains substantial contributions from several others, whether experimental or intellectual, and so my thanks go to Yusdi Santoso, Robert Crawford, Johannes Hohlbein Seamus Holden for their direct contributions to this work. Naturally, I asked these same friends for help in proofreading and commenting on individual chapters, and so my thanks goes double to them for their help in shaping the final version of this thesis.

Great thanks goes to my advisor, Achilles Kapanidis, for his guidance and support. I credit Achilles’ open-door policy, and the freedom and independence he provided in his laboratory, as important elements of my development throughout my DPhil. Achilles went out of his way to be a friend to his students, coming with us on pub trips, and having “Cheersgiving” to make up for the Thanksgiving we American lab members could not economically travel home for. Even when I decided to leave, Achilles seemed more concerned with my future plans and happiness than how my departure might affect the lab. Advisors with this kind of character are rare, and I consider myself lucky to have known Achilles in this capacity.

Many people contribute to our growth as scientists and as human beings. While I cannot hope to credit everyone who played a role in my development during my years at Oxford, I would like to say thank you to those with whom I shared a friendship, mentoring relationship, or both, during my time at Oxford: Richard Branch, Ludovic LeReste, Sonia Trigueros, Alexandra Tomescu, Kris Gryte, Stephan Uphoff, Kieran Finan, Cathy Joyce, Nigel Grindley, Bea Prentiss, Sarah Matthews, Ling Hwang and Kostas Lymperopoulos.

While I’ve thanked them already for their contributions to this thesis, I would be remiss if I did not give a special thanks to Robert Crawford and Seamus Holden, with whom I began my first year as a graduate student at Oxford, and whom I consider myself lucky to call good friends. Some of those pub trips are truly (and perhaps unfortunately) unforgettable!

Yours,
Joe

Confocal Single-Molecule Fluorescence as a Tool for Investigating Biomolecular Dynamics *In Vitro* and *In Vivo*

Joseph P. Torella, Keble College
Doctor of Philosophy
Trinity 2010

Confocal single-molecule fluorescence is a powerful tool for monitoring conformational dynamics[1], and has provided new insight into the enzymatic activities of complex biological molecules such as DNA and RNA polymerases[2-4]. Though useful, such studies are typically qualitative in nature, and performed almost exclusively in highly purified, *in vitro* settings. In this work, I focus on improving the methodology of confocal single-molecule fluorescence in two broad ways: (*i*) by enabling the quantitative identification of molecular dynamics in proteins and nucleic acids *in vitro*, and (*ii*) developing the tools needed to perform these analyses *in vivo*.

Toward the first goal, and together with several colleagues, I have developed three novel methods for the quantitative identification of dynamics in biomolecules: (*i*) Burst Variance Analysis (BVA)[5], which unambiguously identifies dynamics in single-molecule FRET experiments; (*ii*) Dynamic Probability Density Analysis (PDA)[6], which hypothesis-tests specific kinetic models against smFRET data and extracts rate information; and (*iii*) a novel molecular counting method useful for studying single-molecule thermodynamics. We validated these methods against Monte Carlo simulations and experimental DNA controls, and demonstrated their practical application *in vitro* by analyzing the "fingers-closing" conformational change in *E.coli* DNA Polymerase I; these studies identified unexpected conformational flexibility which may be important to the fidelity of DNA synthesis.

To enable similar studies in the context of a living cell, we generated a nuclease-resistant DNA analogue of the Green Fluorescent Protein, or "Green Fluorescent DNA," and developed an electroporation method to efficiently transfer it into the cytoplasm of *E.coli*. We demonstrate *in vivo* confocal detection of smFRET from this construct, which is both bright and photostable in the cellular milieu. In combination with PDA, BVA and our novel molecular counting method, this Green Fluorescent DNA should enable the characterization of DNA and protein-DNA dynamics in living cells, at the single-molecule level. I conclude by discussing the ways in which these methods may be useful in investigating the dynamics of processes such as transcription, translation and recombination, both *in vitro* and *in vivo*.

References

[1] Weiss S (2000) Measuring conformational dynamics of biomolecules by single molecule fluorescence spectroscopy. *Nat Struct Biol* 7: 724-729.

[2] Kapanidis AN, Margeat E, Ho SO, Kortkhonjia E, Weiss S, et al. (2006) Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. *Science* 314: 1144-1147.

[3] Santoso Y, Joyce CM, Potapova O, Le Reste L, Hohlbein J, et al. (2010) Conformational transitions in DNA polymerase I revealed by single-molecule FRET. *Proc Natl Acad Sci U S A* 107: 715-720.

[4] Liu SX, Abbondanzieri EA, Rausch JW, Le Grice SFJ, Zhuang XW (2008) Slide into Action: Dynamic Shuttling of HIV Reverse Transcriptase on Nucleic Acid Substrates. *Science* 322: 1092-1097.

[5] Torella JP, Holden SJ, Santoso Y, Hohlbein J, Kapanidis AN (2010) Identifying molecular dynamics in single-molecule FRET experiments with Burst Variance Analysis. *Biophys J* (accepted).

[6] Santoso Y, Torella JP, Kapanidis AN (2010) Characterizing Single-Molecule FRET Dynamics with Probability Distribution Analysis. *Chemphyschem* 11: 2209-2219.

Table of Contents

1. Introduction

The ability to detect single fluorophores in aqueous solution was first demonstrated by Keller and colleagues in 1990[1]. In the ensuing two decades, single-molecule fluorescence has matured into a set of tools useful for monitoring the dynamics of biological molecules in exquisite detail[2-4]. Whereas conventional biochemical methods operate at the ensemble level, and therefore report the mean of some observable over many molecules, single-molecule fluorescence provides distributions and time courses of these observables, revealing mechanistic details historically unavailable to biochemists[5].

These mechanistic details have provided new insight into the inner workings of biological molecules: single-molecule fluorescence has been used to identify the fluctuations in individual enzymes' activity which underlie Michaelis-Menten enzyme kinetics[6-8], the 'stepping' behaviour of molecular motors[9-11], and the complex catalytic activities of enzymes such as DNA and RNA polymerases[12-15]. The remarkable fidelity of DNA Polymerase I, for instance, is due at least in part to a complex series of conformational changes preceding nucleotide incorporation[16,17]; these conformational dynamics are largely inaccessible to ensemble methods, but can be monitored directly via single-molecule fluorescence [13,18-20]. Being able to monitor the behaviour of individual molecules is perhaps

even more important *in vivo*, where recent studies have shown that single transcription factor binding events are capable of profoundly altering cellular phenotypes[21].

In this work, I focus on confocal single-molecule fluorescence, a simple but powerful technique for monitoring biomolecular dynamics. In confocal fluorescence, a laser is focused into aqueous solution, illuminating molecules that diffuse through it; with an appropriate labelling scheme, the light emitted from these molecules can be used to monitor their structure and dynamics. Confocal single-molecule fluorescence has been used to investigate the conformational dynamics underlying complex enzymatic activities[13,14,18,22], and in biotechnology applications to achieve ultrasensitive detection of microRNAs and transcription factors[23,24]. Though useful, such studies are often qualitative, and performed almost exclusively in highly purified, *in vitro* settings. In this work, I focus on improving the methodology of confocal single-molecule fluorescence in two broad ways: (*i*) by enabling the quantitative identification of molecular dynamics in proteins and nucleic acids *in vitro*, and (*ii*) developing the tools needed to perform these analyses *in vivo*.

1.1 Studying biomolecular dynamics with confocal single-molecule fluorescence

In confocal fluorescence a laser beam is focused into solution, producing a femtolitre-sized "confocal spot" that illuminates molecules as they diffuse through it. If the signal-to-noise ratio of the confocal fluorescence setup is good enough, fluorescence "bursts" from single molecules diffusing through the confocal spot ("single-molecule fluorescence") can be detected. Because diffusion through the spot causes fluctuations in fluorescence intensity during the burst, it is typical to label a given molecule with two spectrally distinct fluorophores, and measure ratiometric observables such as fluorophore Stoichiometry[25,26],

and Förster Resonance Energy Transfer (FRET)[27]. Stoichiometry is a measure of the relative brightness of the two dyes, while FRET efficiency is a measure of the distance between them.

Because the confocal spot is very tiny, molecular observation times for a typical globular protein are ~1 millisecond. FRET and Stoichiometry are measured during these molecular "snapshots," histogrammed, and the resulting distributions used to identify structural or dynamic properties of the given molecule (Fig. 1). One typical application of single-molecule FRET is to measure the conformational change in a molecule, which appears as a shift in the mean value of the FRET histogram (Fig. 1A). Because they can measure conformational changes at biologically relevant size scales (2-8 nanometres for the most common fluorophore pairs[28,29]), such experiments have been instrumental in clarifying structure-function relationships in complex enzymes such as DNA and RNA polymerases[12-15,18], and in elucidating the mechanisms by which nucleic acids and proteins fold[2,30-33]. Recently, much attention has also been paid to the widths and shapes of smFRET distributions, which contain information about static and dynamic structural heterogeneity in the molecule of interest[19,34-39] (Fig. 1B). As such heterogeneities may be relevant to an enzyme's mechanism of action, or to the mechanism by which a protein folds, the width of FRET histograms provides another useful observable for investigating molecular dynamics.

Whereas FRET is typically used to measure intramolecular dynamics, stoichiometry is more frequently employed in studies of intermolecular dynamics. Because stoichiometry measures the intensity ratio of two fluorescent species on a given particle, it has been useful in bimolecular binding assays in which two molecular species, each labelled with a spectrally distinct fluorophore, bind to produce a third, intermediate-stoichiometry species (Fig. 1C). Assays of this kind have been used in biosensors to quantify unlabelled molecules, which influence the binding of the fluorescently labelled molecules; such biosensors have been used to detect transcription factors and microRNAs in cell lysates[23,24]. Stoichiometry has also

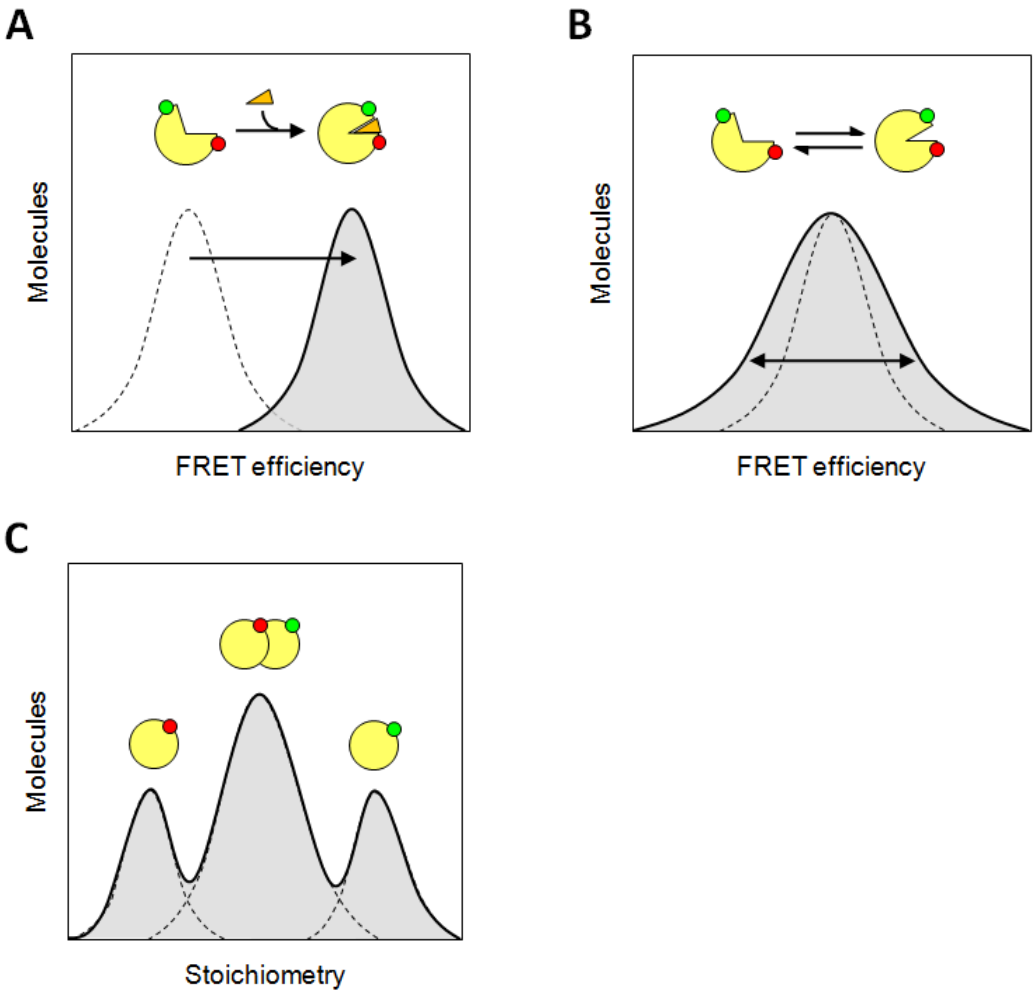


Figure 1. Biological applications of confocal single-molecule fluorescence. Typical observables in single-molecule fluorescence include FRET and stoichiometry. FRET efficiency is a measure of the proximity of two fluorophores. Experiments using FRET have been performed to (A) monitor the change in distance between two fluorophores, for instance to identify conformational changes in a biomolecule upon ligand binding. (B) FRET can also be used to determine if a given molecule exhibits structural heterogeneity. Shown is the effect of molecular dynamics on the width of the FRET histogram; such dynamics cause FRET histogram broadening beyond the "shot noise" distribution expected from photon statistics alone (the "shot noise" distribution indicated by the dashed line, and is discussed in detail in Chapter 4). (C) Stoichiometry is the ratio of intensity between two fluorophores on a single diffusing particle, and can be used to monitor the binding of two molecules labelled with spectrally distinct fluorophores. This has been useful in studies of molecular oligomerization, and in binding-based transcription factor biosensors. Specific examples of these applications are cited in the main text.

been useful in tracking molecular oligomerisation: one such study characterized the particle size distribution and growth kinetics of amyloid fibrils[40], the aggregation of which is relevant to several neurodegenerative diseases[41].

Despite the usefulness and versatility of confocal single-molecule fluorescence, much of the work described is qualitative in nature, or ignores important information present in the measurement. For instance, although FRET histogram broadening can indicate the existence of biologically relevant dynamics (Fig. 1B), this information is typically ignored, or else treated qualitatively[19,34]. Likewise, although stoichiometry-based single-molecule counting has been used as the basis of several biosensing assays[23,24] (Fig. 1C), these often ignore important sources of artefacts, such as the difference in detection efficiency between molecules with different diffusion coefficients[40,42], that complicate confocal molecular counting. In Chapters 3, 4 and 5, I derive and apply methods to identify, model and quantify inter- and intramolecular dynamics in single-molecule FRET experiments.

1.2 Single-molecule fluorescence in living cells

Experiments performed *in vitro* are powerful because they allow a single molecular species to be isolated and studied with careful controls, permitting a simple, reductionist approach to elucidating the molecule's function. In reconstructing how this molecule behaves *in vivo*, however, *in vitro* studies may not be sufficient. Experiments performed *in vitro* are typically carried out in non-physiological buffers that can perturb enzyme activity, and do not recapitulate cellular biophysical properties (e.g. macromolecular crowding) that influence macromolecular structure and binding interactions in living cells[43-45]. Moreover, as such studies include only those components chosen by the experimentalist, *in vivo* experiments may turn up molecular behaviours not seen *in vitro*, e.g. due to unknown binding partners that

influence the molecule's activity or localization. It is therefore important, where possible, to conduct *in vivo* experiments that complement their *in vitro* counterparts.

Unfortunately, single-molecule fluorescence experiments are typically difficult to replicate *in vivo*. One major challenge is fluorescent labelling, as it is not always possible to label intracellular molecules chemically, or to introduce already-labelled molecules into cells[46,47]. The majority of *in vivo* single-molecule fluorescence studies have therefore been limited to using fluorescent proteins (FPs) derived from the Green Fluorescent Protein (GFP) of *Aequorea victoria*[48], which can be genetically fused to the N- and C-termini of a protein of interest[49]. Single-molecule studies performed with FPs in *E. coli* have been used to directly visualize and quantify the dynamics of gene expression[50], to track the diffusion of single mRNA particles[43], and to visualize subunit turnover in the bacterial flagellar motor[51].

Following the lead of the first single-molecule fluorescence experiments, in which detection was achieved by immobilizing fluorophores to improve the signal-to-noise ratio[52,53], single-molecule FP-based studies often use FP immobilization, and the corresponding signal-to-noise increase, as their readout. For example, gene expression has been monitored at the single-molecule level via the appearance of FP molecules localized to the inner membrane[54], while dynamics of the transcription factor LacR (fused to Venus-YFP) were monitored by following the appearance of LacR molecules upon binding to the chromosome[55]. However, many important dynamics, such as the conformational rearrangements of RNA polymerase during initial transcription, or DNA polymerase during DNA replication, do not involve changes in diffusion. Moreover, they occur at tiny length scales (<10 nm) requiring precise FRET measurements at specific sites; this may be difficult or impossible to achieve with FPs, which are large (~4 nm diameter) and typically limited to labelling the N- or C-terminus of a protein. Such dynamics are therefore inaccessible to

current FP-localization technology, and constitute a significant gap in our knowledge about the inner life of cells. In this work, in addition to developing tools for extracting quantitative information about protein and nucleic acid dynamics from confocal experiments performed *in vitro* (Chapters 3-5), in Chapters 6 and 7 I discuss methods for the internalization, and confocal detection, of fluorescent DNA molecules in live *E. coli*.

1.3 Chapter organization

The primary contribution of this thesis is the development of novel analytical tools to quantitatively monitor biomolecular dynamics *in vitro* using confocal single-molecule fluorescence, and accompanying experimental methods to extend this technology to the context of a living cell. In addition to providing an overview of the principles and applications of confocal single-molecule fluorescence (Chapters 1 and 2), I use these tools to investigate single-molecule counting-based biosensors (Chapter 3), the conformational dynamics of DNA hairpins and *E. coli* DNA Polymerase I (Chapters 4-5), and the behaviour of fluorescent DNAs inside *E. coli* (Chapter 6). This work expands the toolbox of methods available for monitoring biomolecular structure and dynamics, both *in vitro* and *in vivo.*

Individual chapters are arranged as follows:

1. **Introduction**
2. **Confocal single-molecule fluorescence: principles and biological applications** Describes the principles behind confocal single-molecule fluorescence, and discusses its most common and interesting applications.
3. **Accurate molecular counting using confocal single-molecule fluorescence** Identifies and resolves artefacts in confocal single-molecule counting, and applies this knowledge to problems in single-molecule thermodynamics and biosensing.

4. **Characterizing single-molecule heterogeneity with Probability Distribution Analysis**

   Generalizes Probability Distribution Analysis (PDA), a method of hypothesis-testing models of static heterogeneity against smFRET data, to include models of dynamic heterogeneity.

5. **Identifying single-molecule dynamics with Burst Variance Analysis**

   Describes Burst Variance Analysis (BVA), a novel method for unambiguously detecting dynamic heterogeneity in smFRET experiments.

6. **A 'Green Fluorescent DNA' for *in vivo* studies of molecular dynamics**

   Describes a novel set of methods for internalizing fluorescent DNA molecules into *E. coli* and observing them via confocal fluorescence, to enable studies of molecular dynamics *in vivo*.

7. **Conclusion**

1.4 Contributions

While the work contained in this thesis is primarily my own, like any research project it includes intellectual contributions from several others. I am therefore intentionally sparing with the use of first person voice throughout this work, instead preferring "we" (as in most publications), to emphasize that the work is of a collaborative nature. Wherever the actual work performed was not mine, however (e.g. experiments performed, or software written), this has been specifically noted in the text; unless so noted, all work is my own. At the end of each chapter is also a "Contributions" section, to summarize the contributions of others, both experimental and intellectual, to that chapter.

References:

[1] Shera EB, Seitzinger NK, Davis LM, Keller RA, Soper SA (1990) Detection of single fluorescent molecules. *Chem Phys Lett* 174: 553-557

[2] Schuler B, Eaton WA (2008) Protein folding studied by single-molecule FRET. *Curr Opin Struct Biol* 18: 16-26

[3] Weiss S (2000) Measuring conformational dynamics of biomolecules by single molecule fluorescence spectroscopy. *Nat Struct Biol* 7: 724-729

[4] Xie XS, Choi PJ, Li GW, Lee NK, Lia G (2008) Single-molecule approach to molecular biology in living bacterial cells. *Annual Review of Biophysics* 37: 417-444

[5] Weiss S (1999) Fluorescence spectroscopy of single biomolecules. *Science* 283: 1676-1683

[6] Xie XS, Lu HP (1999) Single-molecule enzymology. *J Biol Chem* 274: 15967-15970

[7] Lu HP, Xun LY, Xie XS (1998) Single-molecule enzymatic dynamics. *Science* 282: 1877-1882

[8] English BP, Min W, van Oijen AM, Lee KT, Luo GB, et al. (2006) Ever-fluctuating single enzyme molecules: Michaelis-Menten equation revisited. *Nat Chem Biol* 2: 87-94

[9] Peterman EJG, Sosa H, Moerner WE (2004) Single-molecule fluorescence spectroscopy and microscopy of biomolecular motors. *Annu Rev Phys Chem* 55: 79-96

[10] Sowa Y, Rowe AD, Leake MC, Yakushi T, Homma M, et al. (2005) Direct observation of steps in rotation of the bacterial flagellar motor. *Nature* 437: 916-919

[11] Yasuda R, Noji H, Kinosita K, Yoshida M (1998) F-1-ATPase is a highly efficient molecular motor that rotates with discrete 120 degrees steps. *Cell* 93: 1117-1124

[12] Tang GQ, Roy R, Bandwar RP, Ha T, Patel SS (2009) Real-time observation of the transition from transcription initiation to elongation of the RNA polymerase. *Proc Natl Acad Sci U S A* 106: 22175-22180

[13] Santoso Y, Joyce CM, Potapova O, Le Reste L, Hohlbein J, et al. (2010) Conformational transitions in DNA polymerase I revealed by single-molecule FRET. *Proc Natl Acad Sci U S A* 107: 715-720

[14] Kapanidis AN, Margeat E, Ho SO, Kortkhonjia E, Weiss S, et al. (2006) Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. *Science* 314: 1144-1147

[15] Luo G, Wang M, Konigsberg WH, Xie XS (2007) Single-molecule and ensemble fluorescence assays for a functionally important conformational change in T7 DNA polymerase. *Proc Natl Acad Sci U S A* 104: 12610-12615

[16] Joyce CM, Benkovic SJ (2004) DNA polymerase fidelity: Kinetics, structure, and checkpoints. *Biochemistry (Mosc)* 43: 14317-14324

[17] Joyce CM, Potapova O, DeLucia AM, Huang XW, Basu VP, et al. (2008) Fingers-closing and other rapid conformational changes in DNA polymerase I (Klenow fragment) and their role in nucleotide selectivity. *Biochemistry (Mosc)* 47: 6103-6116

[18] Christian TD, Romano LJ, Rueda D (2009) Single-molecule measurements of synthesis by DNA polymerase with base-pair resolution. *Proc Natl Acad Sci U S A* 106: 21109-21114

[19] Santoso Y, Torella JP, Kapanidis AN (2010) Characterizing Single-Molecule FRET Dynamics with Probability Distribution Analysis. *Chemphyschem* 11: 2209-2219

[20] Torella JP, Holden SJ, Santoso Y, Hohlbein J, Kapanidis AN (2010) Identifying molecular dynamics in single-molecule FRET experiments with Burst Variance Analysis. *Biophys J* (accepted):

[21] Choi PJ, Cai L, Frieda K, Xie XS (2008) A stochastic single-molecule event triggers phenotype switching of a bacterial cell. *Science* 322: 442-446

[22] Zhuang XW, Kim H, Pereira MJB, Babcock HP, Walter NG, et al. (2002) Correlating structural dynamics and function in single ribozyme molecules. *Science* 296: 1473-1476

[23] Lymperopoulos K, Crawford R, Torella JP, Heilemann M, Hwang LC, et al. (2010) Single-Molecule DNA Biosensors for Protein and Ligand Detection. *Angew Chem Int Ed Engl* 49: 1316-1320

[24] Neely LA, Patel S, Garver J, Gallo M, Hackett M, et al. (2006) A single-molecule method for the quantitation of microRNA gene expression. *Nat Methods* 3: 41-46

[25] Kapanidis AN, Lee NK, Laurence TA, Doose S, Margeat E, et al. (2004) Fluorescence-aided molecule sorting: Analysis of structure and interactions by alternating-laser excitation of single molecules. *Proc Natl Acad Sci U S A* 101: 8936-8941

[26] Orte A, Clarke R, Balasubramanian S, Klenerman D (2006) Determination of the fraction and stoichiometry of femtomolar levels of biomolecular complexes in an excess of monomer using single-molecule, two-color coincidence detection. *Anal Chem* 78: 7707-7715

[27] Förster T (1946) Energiewanderung und Fluoreszenz. *Naturwissenschaften* 6: 166-175

[28] Sabanayagam CR, Eid JS, Meller A (2005) Using fluorescence resonance energy transfer to measure distances along individual DNA molecules: corrections due to nonideal transfer. *J Chem Phys* 122: 061103

[29] Chan FK, Holmes KL (2004) Flow cytometric analysis of fluorescence resonance energy transfer: a tool for high-throughput screening of molecular interactions in living cells. *Methods Mol Biol* 263: 281-292

[30] Lipman EA, Schuler B, Bakajin O, Eaton WA (2003) Single-molecule measurement of protein folding kinetics. *Science* 301: 1233-1235

[31] Zhao R, Rueda D (2009) RNA folding dynamics by single-molecule fluorescence resonance energy transfer. *Methods* 49: 112-117

[32] Schuler B, Lipman EA, Eaton WA (2002) Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. *Nature* 419: 743-747

[33] Xie Z, Srividya N, Sosnick TR, Pan T, Scherer NF (2004) Single-molecule studies highlight conformational heterogeneity in the early folding steps of a large ribozyme. *Proc Natl Acad Sci U S A* 101: 534-539

[34] Antonik M, Felekyan S, Gaiduk A, Seidel CAM (2006) Separating structural heterogeneities from stochastic variations in fluorescence resonance energy transfer distributions via photon distribution analysis. *J Phys Chem B* 110: 6970-6978

[35] Nir E, Michalet X, Hamadani KM, Laurence TA, Neuhauser D, et al. (2006) Shot-noise limited single-molecule FRET histograms: Comparison between theory and experiments. *J Phys Chem B* 110: 22103-22124

[36] Kalinin S, Valeri A, Antonik M, Felekyan S, Seidel CAM (2010) Detection of Structural Dynamics by FRET: A Photon Distribution and Fluorescence Lifetime Analysis of Systems with Multiple States. *J Phys Chem B* 114: 7983-7995

[37] Kalinin S, Felekyan S, Valeri A, Seidel CAM (2008) Characterizing multiple molecular states in single-molecule multiparameter fluorescence detection by probability distribution analysis. *J Phys Chem B* 112: 8361-8374

[38] Watkins LP, Chang HY, Yang H (2006) Quantitative single-molecule conformational distributions: A case study with poly-(L-proline). *J Phys Chem A* 110: 5191-5203

[39] Gopich IV, Szabo A (2007) Single-molecule FRET with diffusion and conformational dynamics. *J Phys Chem B* 111: 12925-12932

[40] Orte A, Birkett NR, Clarke RW, Devlin GL, Dobson CM, et al. (2008) Direct characterization of amyloidogenic oligomers by single-molecule fluorescence. *Proc Natl Acad Sci U S A* 105: 14424-14429

[41] Koo EH, Lansbury PT, Kelly JW (1999) Amyloid diseases: Abnormal protein aggregation in neurodegeneration. *Proc Natl Acad Sci U S A* 96: 9989-9990

[42] Eigen M, Rigler R (1994) Sorting single molecules - application to diagnostics and evolutionary biotechnology. *Proc Natl Acad Sci U S A* 91: 5740-5747

[43] Golding I, Cox EC (2006) Physical nature of bacterial cytoplasm. *Phys Rev Lett* 96:

[44] Zimmerman SB, Harrison B (1987) Macromolecular crowding increases binding of DNA polymerase to DNA - an adaptive effect. *Proc Natl Acad Sci U S A* 84: 1871-1875

[45] Zhou HX, Rivas GN, Minton AP (2008) Macromolecular crowding and confinement: Biochemical, biophysical, and potential physiological consequences. *Annual Review of Biophysics* 37: 375-397

[46] Kapanidis AN, Weiss S (2002) Fluorescent probes and bioconjugation chemistries for single-molecule fluorescence analysis of biomolecules. *J Chem Phys* 117: 10953-10964

[47] Chen I, Ting AY (2005) Site-specific labeling of proteins with small molecules in live cells. *Curr Opin Biotechnol* 16: 35-40

[48] Tsien RY (1998) The green fluorescent protein. *Annu Rev Biochem* 67: 509-544

[49] Chalfie M, Tu Y, Euskirchen G, Ward WW, Prasher DC (1994) Green Fluorescent Protein as a Marker for Gene Expression. *Science* 263: 802-805

[50] Yu J, Xiao J, Ren XJ, Lao KQ, Xie XS (2006) Probing gene expression in live cells, one protein molecule at a time. *Science* 311: 1600-1603

[51] Leake MC, Chandler JH, Wadhams GH, Bai F, Berry RM, et al. (2006) Stoichiometry and turnover in single, functioning membrane protein complexes. *Nature* 443: 355-358

[52] Moerner WE, Kador L (1989) Optical detection and spectroscopy of single molecules in a solid. *Phys Rev Lett* 62: 2535-2538

[53] Xie XS, Dunn RC (1994) Probing Single-Molecule Dynamics. *Science* 265: 361-364

[54] Cai L, Friedman N, Xie XS (2006) Stochastic protein expression in individual cells at the single molecule level. *Nature* 440: 358-362

[55] Elf J, Li GW, Xie XS (2007) Probing transcription factor dynamics at the single-molecule level in a living cell. *Science* 316: 1191-1194

2. Confocal single-molecule fluorescence: principles and biological applications

Chapter 1 provided a very brief overview of the ways in which single-molecule fluorescence can shine light on biology: using a single fluorophore, studies have identified the conformational dynamics underlying Michaelis-Menten enzyme kinetics[1], and the kinetics of gene expression in living cells[2]. Using two fluorophores, Förster resonance energy transfer (FRET) studies have analyzed the structure, folding and dynamics of nucleic acids and proteins[3-5], while measurements of fluorophore stoichiometry have been useful in studies of protein aggregation[6] and in biosensors to identify transcription factors and miRNAs[7,8]. A meaningful discussion of these experiments, however, requires an understanding of the principles of fluorescence and FRET, and their use as experimental tools to investigate biology.

We focus on solution-phase confocal fluorescence, in which a Gaussian laser beam is focused on a tiny 1fL confocal "spot" in solution, illuminating molecules that diffuse through it. Solution-phase confocal fluorescence has several advantages over other single-molecule fluorescence implementations, such as total internal reflection fluorescence (TIRF)

microscopy. These include the ability to monitor fluorescent molecules without the need for potentially perturbative surface immobilization, and the superior signal-to-noise of confocal fluorescence; the latter is due to the relative insensitivity of conventional CCD cameras as compared to APDs, and their corresponding need for high electron gain [9]. A drawback of confocal fluorescence is its limited observation time; while confocal measurements monitor individual molecules in ~1 ms "snapshots," TIRF measurements can monitor single molecules for minutes, making them more appropriate in studies of long-timescale (>10 ms) dynamics. Nevertheless, confocal fluorescence is a powerful and popular technique for the investigation of molecular dynamics[3,10,11]. We therefore focus on confocal fluorescence this work, while acknowledging its complementarity to TIRF and other surface-based approaches.

In this chapter, I review the physical principles of implementing confocal single-molecule fluorescence and FRET experiments. These principles lay the basis for chapters three, four and five, in which we model the process of confocal fluorescence to improve our quantitative understanding of single-molecule, and especially single-molecule FRET, dynamics. I also briefly discuss fluorescence correlation spectroscopy (FCS), a correlation-based confocal fluorescence method that operates at the small-ensemble level to quantify molecular concentrations and dynamics. Where appropriate, I relate the tools discussed to their most common, and most interesting, applications in biology.

2.1 Fluorescence

Fluorescence is the emission of light by a molecule relaxing from an electronically excited singlet state to its ground state[12]. In biological applications of fluorescence, the light is generally emitted in the visible or near-infrared region of the electromagnetic spectrum, by an aromatic organic fluorophore or fluorescent protein. These fluorophores are conjugated to a biological molecule of interest (Fig. 1A) in order to localize, quantify or track it[13], or in

some cases to better characterize its dynamics[14]. While the fluorophore is typically excited by an external light source, fluorescence has also been achieved in living systems by fluorescent proteins that achieve excitation using chemical energy (e.g. luciferase); this is referred to as bioluminescence[15].

The mechanism of fluorescence can be easily understood using a simplified Jablonski diagram, which depicts the electronic transitions occurring in the fluorophore during excitation, and fluorescence emission[12] (Fig. 1C). In most cases, a fluorophore absorbs a photon from an external source of light, and is excited from its ground state (S_0) to an electronically excited state, which rapidly falls to the singlet state (S_1) via vibrational relaxation. For excitation to occur, the light source must have a wavelength (λ) such that the energy of each photon is comparable to the difference in energy between S_0 and S_1 $\left(hc\lambda^{-1} \approx E_{S1} - E_{S0}\right)$. In the simplest case, the excited state (S_1) will relax back to the ground state with first-order kinetics, the mean waiting time of which is called its fluorescence lifetime (τ_f, on the order of 3 ns for most organic fluorophores in water), and emit a photon (Fig. 1C). Fluorescence lifetimes can be measured experimentally, and provide details about interactions of the fluorophore with both its local environment and other molecules (e.g. quenching, resonance energy transfer); while we do not perform fluorescence lifetime measurements in this work, their usefulness has been discussed elsewhere[16,17].

In addition to fluorescence emission, excited-state fluorophores can return to the ground state via nonradiative relaxation, quenching by other molecules in solution, resonance energy transfer to another chromophore, or spin conversion and intersystem crossing into the triplet state, T_1 (Fig. 1D). Due to spin conversion, the $T_1 \rightarrow S_0$ transition is "spin-forbidden," resulting in triplet state lifetimes much longer than the fluorescence lifetime (~1 μs for many organic fluorophores). Once in the triplet state, a fluorophore typically cannot be further excited; periodic entry into the triplet state therefore results in "blinking" behaviour, which

can produce experimental artefacts. Moreover, fluorophores may exhibit additional complex photophysics, including internal conversion between higher-order singlet and triplet states, and transient dark states[12,18-20]. Such photophysics are an important factor in choosing fluorescent dyes for a given experimental application; in smFRET experiments, for instance, it is ideal to avoid the use of dyes with significant blinking, which can give the appearance of structural dynamics where none are genuinely present[21]. In some cases, undesirable photophysics can be partially mitigated by chemical agents that quench triplet and other fluorophore 'dark' states[18-20]; however, all fluorescent molecules eventually undergo irreversible photobleaching. Photobleaching occurs with a fixed probability for every round of excitation, placing an upper limit on the amount of information ultimately obtainable from a single fluorophore[12].

An important feature of fluorophores is that energy is lost between excitation and emission. In practice, excitation typically occurs into a state slightly more energetic than S_1, which vibrationally relaxes to S_1 prior to emission; likewise, emission can occur into a state slightly more energetic than S_0, which vibrationally relaxes to S_0 (Fig. 1C). In addition to solvent, excited state and other interactions, these vibrational relaxations result in a net loss of energy from excitation to emission, producing the "Stokes shift," an increase in peak emission wavelength relative to the peak excitation wavelength (Fig. 1B). The Stokes shift is crucial in FRET experiments, as it allows the spectral separation of excitation and emission light. In general, fluorophores with greater Stokes shifts are desirable, since they increase the fraction of emitted light recovered after filtering excitation light.

2.2 Förster resonance energy transfer (FRET)

Förster resonance energy transfer (FRET) is a dipole-dipole interaction in which energy is transferred nonradiatively from a donor fluorophore to an acceptor chromophore[22] (Fig. 2A).

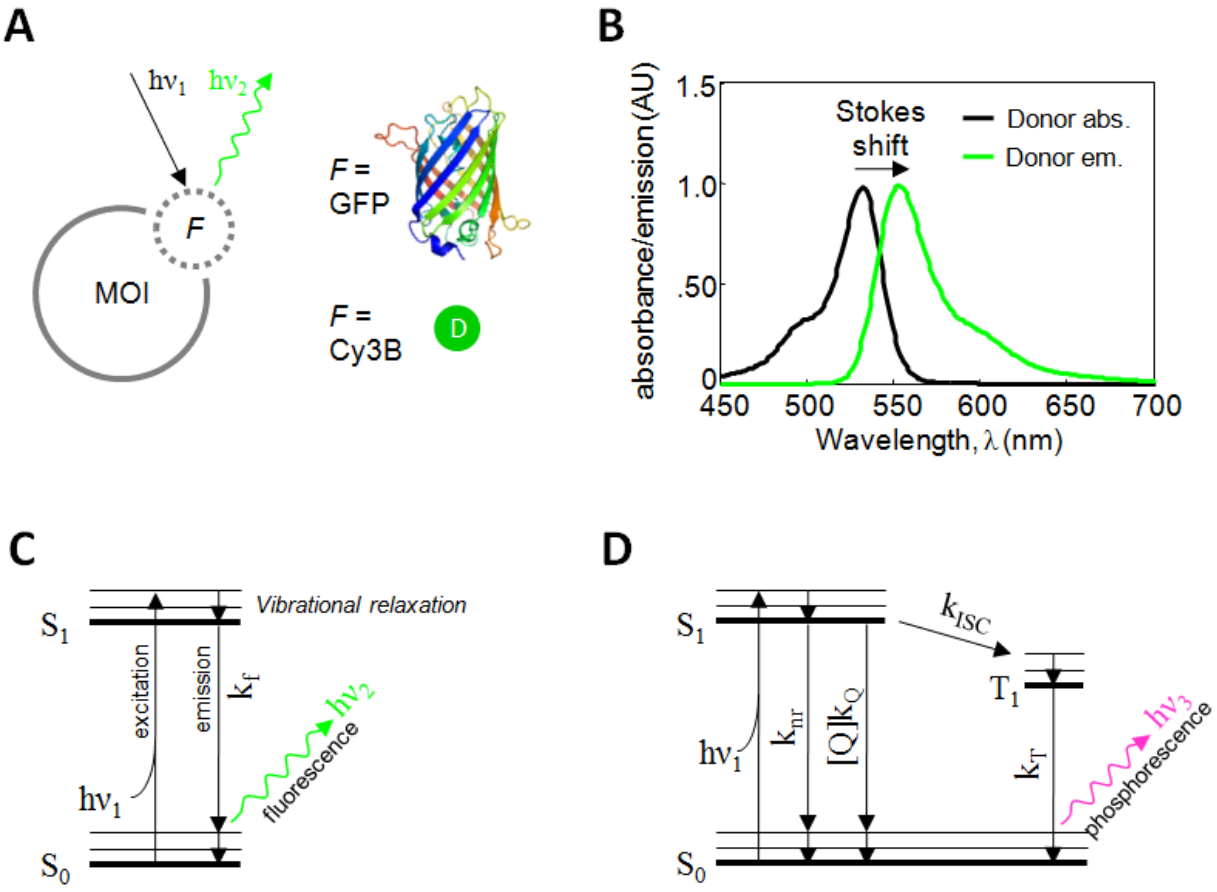

A
hν1
hν2
F
MOI
F =
GFP
F =
Cy3B
D
B
absorbance/emission (AU)
1.5
1.0
.50
0
Stokes
shift
Donor abs.
Donor em.
450
500
550
600
650
700
Wavelength, λ (nm)
C
S1
Vibrational relaxation
excitation
emission
kf
hν2
fluorescence
hν1
S0
D
S1
kISC
T1
hν1
knr
[Q]kQ
kT
hν3
phosphorescence
S0

Figure 1. Principles of fluorescence. (A) In fluorescence experiments on biological systems, a biological molecule of interest (MOI) is labelled with a fluorophore (F), which is typically either a fluorescent protein (e.g. GFP, top; image by Richard Wheeler), or an aromatic organic fluorophore (e.g. Cy3B, bottom). It is excited by a photon of one wavelength (λ_1 , with energy $h\nu_1 = hc / \lambda_1$), and emits a fluorescence photon at a longer wavelength, and with lower energy ($h\nu_2$).(B) Absorbance and emission spectra of Atto532, a common organic fluorophore. The Stokes shift (arrow) is the difference in wavelengths between the absorbance and emission peaks; larger Stokes shifts are desirable since, after filtering out excitation light, more of the emitted light can be collected. (C) Simplified Jablonski diagram of a single round of excitation and emission from a fluorophore. A fluorophore in the ground state (S_0) absorbs a photon ($h\nu_1$), typically exciting it to an energy level above that of the excited singlet state (S_1); the molecule then undergoes rapid vibrational relaxation to S_1. The fluorophore remains at S_1 with mean fluorescence lifetime τ_f , and typically returns to the ground state with first-order rate $k_f = \tau_f^{-1}$, emitting a fluorescence photon ($h\nu_2$). The electron may again undergo vibrational relaxation before returning to the ground state (S_0). The vibrational relaxations occurring during excitation and emission are partially responsible for the net loss of energy during fluorescence, which results in the Stokes shift (panel B). (D) Alternative mechanisms by which fluorophores return to the ground state include nonradiative relaxation (rate k_{nr}), quenching by another molecule (rate $[Q]k_Q$, where $[Q]$ is quencher concentration), and intersystem crossing (rate k_{ISC}) into the triplet state (T_1). Triplet state lifetimes, τ_T , are typically much longer than fluorescence lifetimes. In some molecules, light is emitted during the $T_1 \rightarrow S_0$ transition; this is referred to as phosphorescence ($h\nu_3$), and typically occurs with a wavelength longer than that of fluorescence. Phosphorescence is typically not a feature of commonly-used organic fluorophores, though we include it here for completeness.

Because the efficiency of energy transfer is distance-dependent on the nanometre size scale[23], and linear in the range of 3-7 nm in the case of commonly-used organic fluorophore and fluorescent protein pairs (e.g. Cy3/Cy5[24] and CFP/YFP[25]), FRET is useful as a spectroscopic ruler at the size scale of single biological molecules[26] (e.g. the length of a 10 basepair segment of DNA is 3.4 nm, similar to the diameter of a globular 30kDa protein). Experimentally, FRET is carried out by exciting the donor fluorophore; energy is transferred nonradiatively to the acceptor chromophore with an efficiency depending on distance, and reemitted at a longer wavelength. By collecting and comparing the light emitted from the donor and acceptor fluorophores, the FRET efficiency, and therefore the distance between the two fluorophores, can be determined.

The electronic transitions that occur during FRET are shown in the modified Jablonski diagram in Fig 2B. The ground-state donor fluorophore is excited to the singlet state; when it returns to the ground state, it can either emit a fluorescence photon with rate $k_f = \tau_f^{-1}$, or transfer its energy to the acceptor fluorophore with rate k_{FRET}; this causes excitation of the acceptor fluorophore, which then emits a fluorescence photon and returns to the ground state. The rate of energy transfer is:

$$k_{FRET} = \left(\frac{1}{\tau_D}\right)\left(\frac{R_0}{r}\right)^6 \qquad [1]$$

where τ_f is the donor fluorescence lifetime in the absence of FRET, r is the distance between the donor and acceptor fluorophores, and R_0 is the Förster radius: the distance at which FRET efficiency is 50%. The FRET efficiency E can be calculated from the rates of fluorescence and FRET as:

$$E = \frac{k_{FRET}}{k_{FRET} + k_f} = \frac{\tau_f^{-1}\left(R_0 / r\right)^6}{\tau_f^{-1}\left(R_0 / r\right)^6 + \tau_f^{-1}} = \left(1 + \left(\frac{r}{R_0}\right)^6\right)^{-1} \qquad [2]$$

This result is an important property of FRET: as FRET efficiency *E* varies with r^6 (Eq. 2), there is an essentially linear relationship between FRET efficiency and donor-acceptor distance at scales near the Förster radius, R_0. Because R_0 is typically around 5 nm for commonly used donor-acceptor pairs, the linear FRET range is roughly 3-7 nm, making FRET an effective spectroscopic ruler on the size scale of biological molecules (Fig. 2C).

To complete the discussion of FRET, we note that R_0 can be determined as[12]:

$$R_0^6 = \left(\frac{9000\ln 10}{128\pi^5}\right)\left(\frac{\phi_D \kappa^2 J}{n^4 N_A}\right) \qquad [3]$$

where ϕ_D is the donor quantum yield, n is the refractive index of the medium, N_A is Avogadro's number, κ^2 is the dipole orientation factor, and J is the overlap integral, a measure of overlap between the donor emission and acceptor absorbance spectra[12]. The orientation factor falls in the range $0 \le \kappa^2 \le 4$, depending on the orientational freedom of the donor and acceptor fluorophores; in most cases, the fluorophores undergo sufficient reorientation on the timescale of the fluorescence lifetime to warrant the assumption that $\kappa^2 \approx 2/3$ [27]. The overlap integral, J, is calculated as:

$$J = \int_0^\infty f_D(\lambda)\varepsilon_A(\lambda)\lambda^4 \, d\lambda \qquad [4]$$

where λ is wavelength, $f_D(\lambda)$ is the donor emission spectrum, and $\varepsilon_A(\lambda)$ is the molar extinction coefficient of the acceptor (i.e. its absorbance). The overlap of donor emission and acceptor absorbance is illustrated in Fig. 2D; note, however, that due to the λ^4 term, J is not simply proportional to the visual overlap of the donor emission and acceptor absorbance spectra.

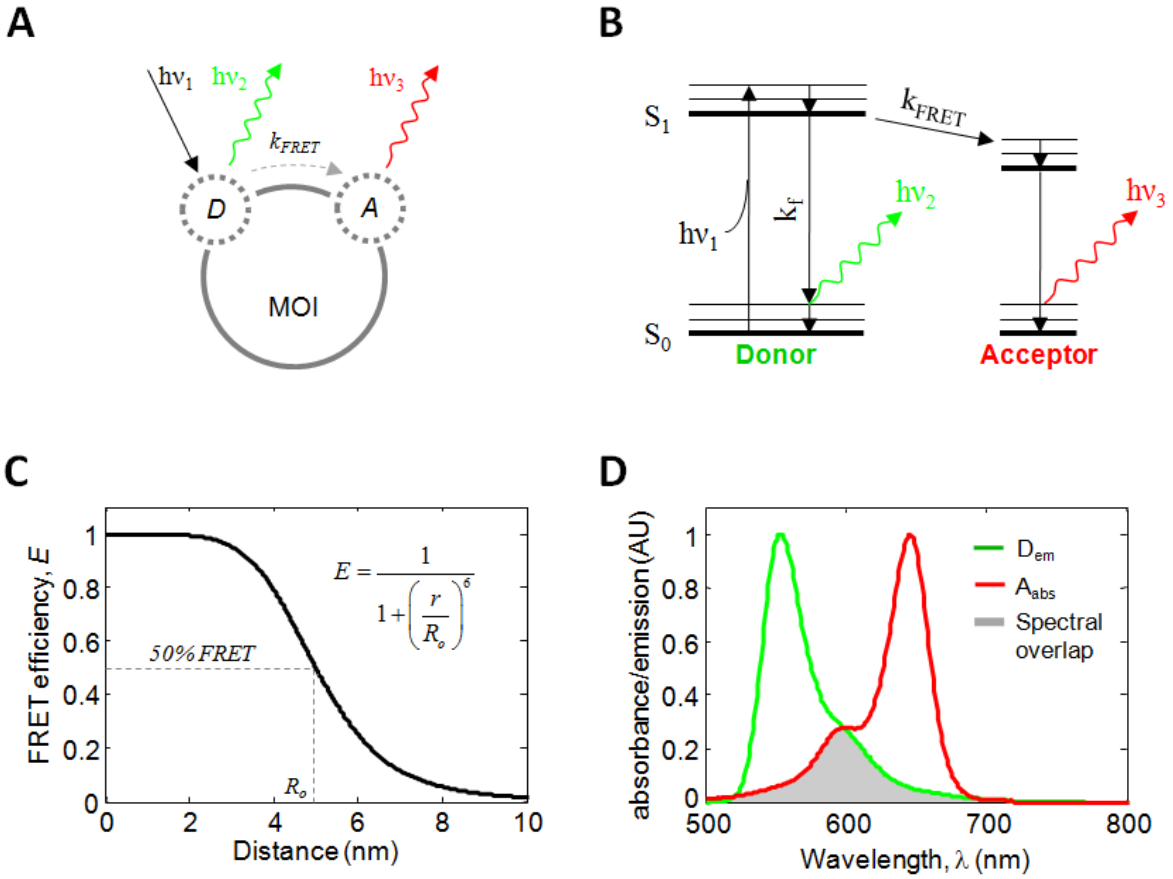

A
hν1
hν2
hν3
kFRET
D
A
MOI
B
S1
kFRET
hν1
kf
hν2
hν3
S0
Donor
Acceptor
C
FRET efficiency, E
E = 1 / (1 + (r/Ro)^6)
50% FRET
Ro
Distance (nm)
D
absorbance/emission (AU)
Dem
Aabs
Spectral overlap
Wavelength, λ (nm)

Figure 2. Principles of Förster Resonance Energy Transfer (FRET). (A) A typical experimental implementation of FRET: a biomolecule is labelled with donor (D) and acceptor (A) fluorophores, and the donor fluorophore is excited by an external light source ($h\nu_1$). Energy is transferred from the donor to the acceptor in a distance-dependent fashion with rate k_{FRET}, which increases with the proximity between them. In an ideal system, photons are emitted by the donor ($h\nu_2$) and acceptor ($h\nu_3$) fluorophores in the proportions *(1-E*) and *E* respectively, where *E* is the overall FRET efficiency. Detected photons can be used to calculate FRET efficiency, and therefore the donor-acceptor distance. (B) Simplified Jablonski diagram illustrating the electronic transitions occurring during FRET. The donor fluorophore absorbs a photon ($h\nu_1$) and is excited to the singlet state (S_1); when it returns to the ground state (S_0), it either emits a fluorescence photon ($h\nu_2$) with rate k_f , or transfers energy to the acceptor fluorophore with rate k_{FRET}; the acceptor then emits a fluorescence photon at a longer wavelength ($h\nu_3$). (C) FRET efficiency as a function of distance for donor-acceptor pair with Förster radius R_0 = 5 nm (the distance at which FRET efficiency is 50%). For R_o = 5 nm, FRET efficiency is effectively linear in the range of 3 – 7 nm, and still significantly distance-dependent (albeit nonlinear) in the range of 2–8 nm. (D) Example absorbance and emission spectra for FRET pair Atto532 and Atto647N. The Förster radius, R_0 (5.9 nm for this pair), depends on the spectral overlap between the donor emission and acceptor absorbance spectra (shaded grey).

2.3 Confocal single-molecule fluorescence

Confocal fluorescence is performed by focusing a Gaussian laser beam into solution, creating a small, ~1fL confocal "spot" (Fig. 3A). The solution into which the beam is focused contains fluorescently-tagged molecules of interest which diffuse stochastically through the confocal volume, giving rise to 'bursts' of fluorescence[28] (Fig. 3B). These bursts are identified computationally, and descriptive statistics, such as total photons, burst duration, FRET and stoichiometry are calculated[29-31].

The effective geometry of the confocal spot is a product of two processes: excitation, and detection. For a well-aligned optical setup, a Gaussian laser beam will produce a Gaussian-Lorentzian excitation profile. Detection is not uniform over space, however, due primarily to the use of a small pinhole that rejects out-of-focus light in the axial direction. Convolution of this effect with the Gaussian-Lorentzian excitation profile yields an overall "collection efficiency function" with a 3D-Gaussian profile[32,33]:

$$CEF(r,z) = I_0 \exp\left[-2\left(\frac{r}{w_o}\right)^2\right]\exp\left[-2\left(\frac{z}{z_o}\right)^2\right] \qquad [5]$$

where r is radial distance from the origin in the x-y planes, z is axial distance along the laser path, and w_0 and z_0 are the radial and axial radii at which maximal laser intensity, I_0, drops by a factor of e^2 (Fig. 3C). This 3D-Gaussian geometry is important in performing numerical simulations of single-molecule fluorescence in order to validate novel analytical methods (Chapters 3-5), and in the derivation of fitting functions for Fluorescence Correlation Spectroscopy (FCS, below).

A significant challenge in confocal single-molecule fluorescence is distinguishing a fluorescence burst from background. Although several methods exist to accomplish this, we use a variant of an algorithm from Seidel and colleagues[31]. For each photon, we ask whether at least *N* other photons have arrived at the detector within time *T* of that photon. If so, we

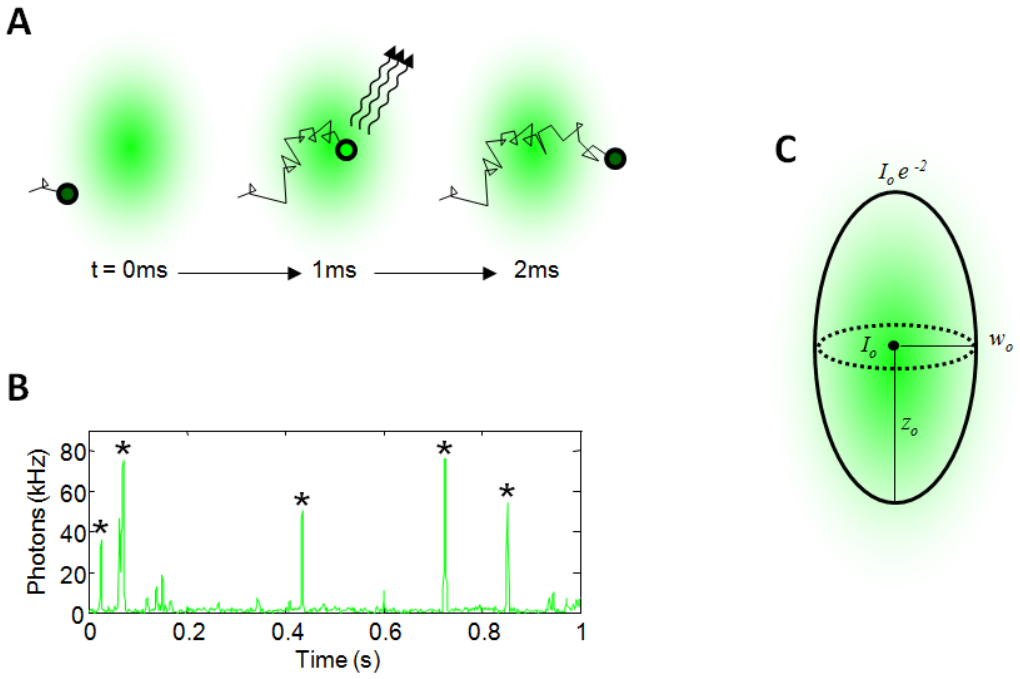


Figure 3. Fluorescence bursts in confocal single-molecule fluorescence. (A) Bursts of fluorescence occur when a fluorescent molecule diffuses into a Gaussian confocal spot (0 ms) and undergoes many rounds of excitation-emission (1 ms) before diffusing away again (2 ms). (B) In a simulation of this process using typical experimental parameters (confocal spot radius 500 nm, molar extinction coefficient 100,000 M^{-1} cm^{-1}, diffusion coefficient 3 x 10^8 nm^2 s^{-1}, background 5kHz), bursts of fluorescence are easily identifiable above the background in a 1 ms-binned fluorescence time trace (bursts indicated with *). (C) 3D-Gaussian collection efficiency function (CEF) of the confocal spot (Eq. 5). At the origin (black dot), the confocal spot has intensity I_0, which drops to I_0 / e^2 at a radial (x-y axis) distance of w_0, or at an axial (z-axis) distance of z_0. This 3D-Gaussian profile results from a convolution of the Gaussian-Lorentzian intensity profile of the laser, with the effects of a small pinhole that restricts detection efficiency in the axial direction; it is therefore the intensity function as imaged on the APD detector (see Fig. 4).

consider it to belong to a fluorescent burst; all consecutive photons matching this criterion are then assumed to belong to the same burst. Because background is typically due to Raman scattering of the incident laser off of water molecules in the confocal volume[34], it is Poisson-distributed. Selection of *N* and *T* can then be chosen rationally to avoid selecting "bursts" that arise from the statistical bunching of background photons. Stricter burst selection criteria can also be employed to improve the signal-to-noise of detected bursts, or to prevent detection of weakly fluorescent impurities in a sample that give rise to non-poisson background counts.

2.4 Alternating laser excitation (ALEX) spectroscopy

In studies of confocal single-molecule FRET, we typically use Alternating Laser Excitation (ALEX) [35]. ALEX refers to the microsecond-timescale alternation of two lasers exciting the donor and acceptor fluorophores independently (Fig. 4A). Information from donor-excitation can be used to monitor the distance between two fluorophores via FRET (E^*), while information from both donor- and acceptor-excitation can be used to determine the relative intensity of donor and acceptor fluorophores on a given molecule (Stoichiometry, S)[30]. ALEX provides three major advantages over smFRET alone: (*i*) in FRET experiments, it allows sorting of molecules along the Stoichiometry axis, allowing the exclusion of donor- or acceptor-only molecules from further analysis[30]; (*ii*) it allows the extraction of correction factors to calculate accurate FRET, E, from the observed FRET, E^* [36]; and (*iii*) it allows single-molecule detection of coincident (i.e. donor- and acceptor-bearing) species that do not exhibit FRET[8]. These technologies are important in Chapter 3, where we use ALEX to detect the binding of transcription factors to DNA[8]; in Chapters 4 and 5, where we use ALEX to exclude donor- and acceptor-only species from FRET analysis of biological molecules[37-39]; and in Chapter 6, where we use ALEX to distinguish intact and degraded DNAs in live *E. coli.*

The laser setup used to implement ALEX is illustrated in Fig. 4A. Two lasers are focused on a ~1fL confocal spot in solution. Fluorescently labelled molecules diffuse through the volume with a mean residence time on the order of 1ms; the donor- and acceptor-excitation lasers are typically alternated with a period of 100 μs such that, on average, a given molecule will undergo 10 alternation cycles while diffusing through the volume. Photons collected throughout a burst while using ALEX are sorted depending on when they were emitted (i.e. during donor- or acceptor-excitation) and in which channel they were collected (donor- or acceptor-emission): F_{Dem}^{Dex}, F_{Aem}^{Dex}, F_{Dem}^{Aex}, F_{Aem}^{Aex}. From these photon counts, descriptive statistics of the molecule's Stoichiometry, *S*, and apparent FRET, or Proximity Ratio *E**, and can be calculated:

$$E^* = \frac{F_{Aem}^{Dex}}{F_{Dem}^{Dex} + F_{Aem}^{Dex}} \quad [6]$$

$$S = \frac{F_{Dem}^{Dex} + F_{Aem}^{Dex}}{F_{Dem}^{Dex} + F_{Aem}^{Dex} + F_{Aem}^{Aex}} \quad [7]$$

where stoichiometry is 1.0 in the case of a donor-only molecule, 0 in the case of an acceptor-only molecule, and 0.5 in the case of a molecule with equally-well-detected donor and acceptor fluorophores. We omit F_{Dem}^{Aex} as it is negligible, and contributes primarily noise to subsequent calculations. To analyze the results of an ALEX experiment, it is typical to plot a 2D burst histogram, where the axes are *E** and *S*[30]; this allows straightforward identification of donor-only, acceptor-only and donor-acceptor subpopulations (along the S-axis), and low- and high-FRET subpopulations (along the *E** axis) (Fig. 4B).

We note that the terms *E** and *S* (Eqs. 6, 7) are approximate FRET and Stoichiometry values, incorporating contributions from background, differences in fluorophore detection efficiency, and spectral leakage of the fluorophores across channels. The latter includes both "leakage," which is donor fluorophore emission into the acceptor detection channel, and

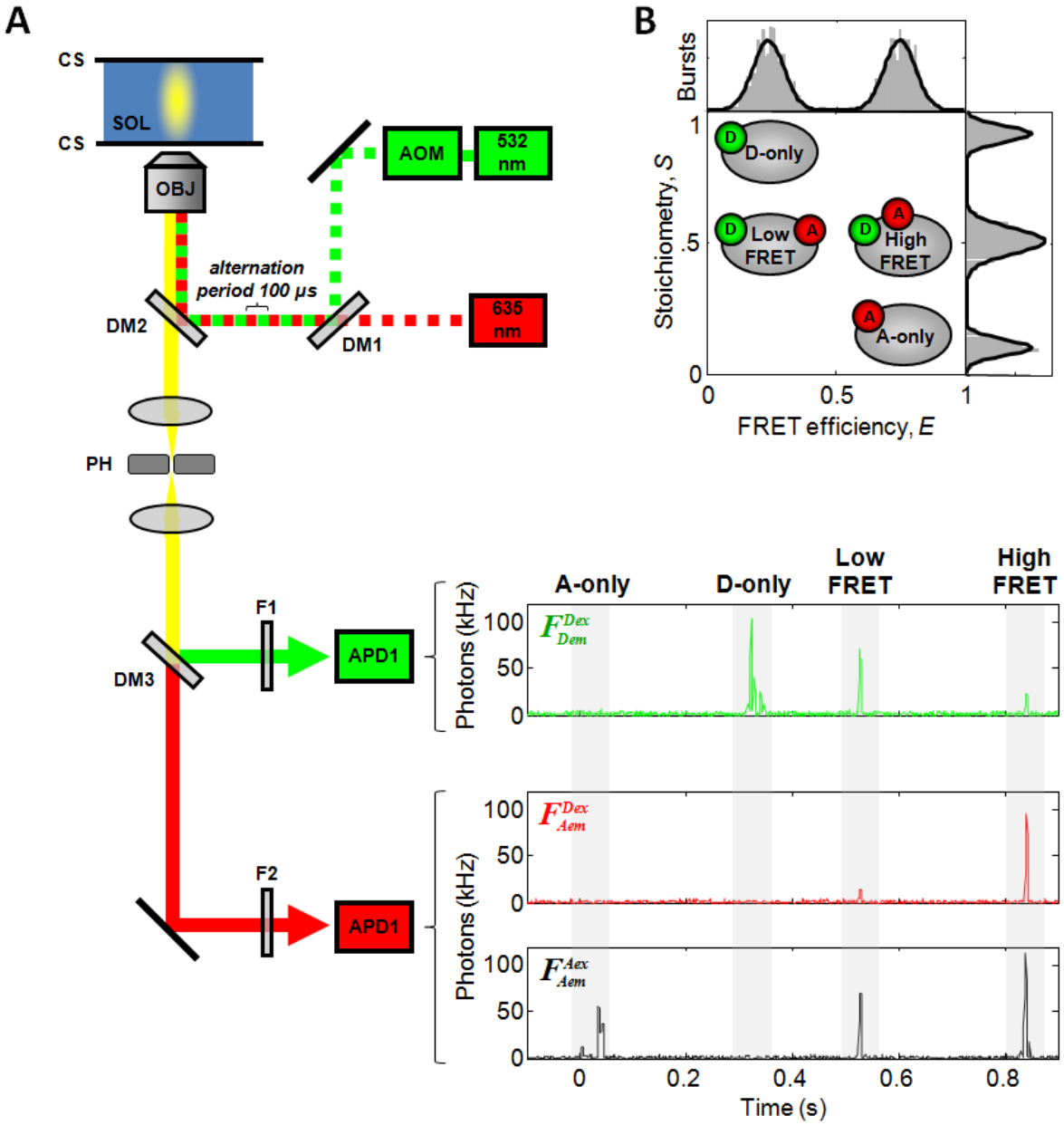

A
CS
SOL
CS
OBJ
AOM
532 nm
alternation period 100 µs
DM2
DM1
635 nm
PH
F1
APD1
DM3
F2
APD1
Photons (kHz)
Photons (kHz)
A-only
D-only
Low FRET
High FRET
Time (s)
B
Bursts
Stoichiometry, S
FRET efficiency, E
D-only
Low FRET
High FRET
A-only

Figure 4. Optical setup and data analysis using ALEX. (A) The optical setup for a confocal single-molecule fluorescence microscope employing Alternating Laser Excitation (ALEX). A pulsed 635 nm diode laser (Cube; Coherent, Santa Clara, CA) and 532 nm Nd:YAG laser (Samba; Cobolt AB, Solna, Sweden) modulated by an acousto-optical modulator (AOM) are combined using a dichroic mirror (DM1) and coupled into an inverted microscope (IX71; Olympus, Tokyo, Japan). The light is focused on a spot in solution (SOL) between two cover slips (CS) by an oil-immersion objective (UPlanSApo 60x, 1.35NA, Olympus). Emitted light is separated from the incident laser light by a second dichroic mirror (DM2) and spatially filtered through a 100 μm pinhole (PH) to reject out-of-focus light in the axial direction. The emitted light is then separated by a third dichroic mirror (DM3) into donor-emission and acceptor-emission channels, filtered (F1 and F2) and collected by two avalanche photodiodes (APDs 1 and 2). In addition to being spectrally separated, the emitted light is separated computationally depending on the time at which it arrives at the detector (and therefore by its origin in a donor- or acceptor-excitation period). This produces four channels: F_{Dem}^{Dex}, F_{Aem}^{Dex}, F_{Dem}^{Aex}, F_{Aem}^{Aex}. While analyzing bursts of fluorescence, F_{Aem}^{Dex} and F_{Dem}^{Dex} can be used to calculate the FRET of a diffusing molecule (E^*, Eq. 6), and F_{Aem}^{Dex}, F_{Dem}^{Dex} and F_{Aem}^{Aex} can be used to compute the fluorophore stoichiometry (S, Eq. 7). (B) 2D ALEX histogram of collected bursts. Donor-only, acceptor-only and donor-acceptor-labelled molecules are separated along the S-axis; doubly-labelled molecules are sorted along the E^* axis according to FRET.

"direct-excitation" of acceptor fluorophores by the donor excitation laser. While these errors can be carefully controlled for by analyzing a standard sample with ALEX[36], most experiments rely only on relative shifts in FRET or Stoichiometry; it is therefore common to analyze approximate, rather than accurate, FRET distributions[29,35,40]. Likewise, we use uncorrected E^* and S values in subsequent chapters.

2.5 Applications of confocal single-molecule fluorescence

Figure 5 illustrates several common applications of confocal fluorescence, along with a schematic 2D-ALEX histogram to show how the data from these experiments is interpreted. Perhaps the most common such application is the use of single-molecule FRET to monitor changes in the structure of a biomolecule in response to a given stimulus (Fig. 5A). This simple technique has been used to identify 'DNA-scrunching' activity during initial transcription by *E. coli* RNA polymerase [10], to identify the unique binding orientations that underlie substrate-mediated regulation of HIV reverse transcriptase activity[11], and to investigate many other interesting and important processes in biomolecular catalysis and folding[4,5,39,41,42].

Illustrated in Fig. 5A is the Klenow Fragment (KF) of *E. coli* DNA Polymerase I, labelled with donor and acceptor fluorophores on its finger and thumb subdomains, respectively. During nucleotide incorporation, KF undergoes a "fingers-closing" transition, which is important for the fidelity of DNA synthesis[43]. Previous work from our lab on this labelled construct was able to clearly resolve fingers-closing upon addition of a complementary nucleotide to DNA-bound KF, by monitoring the increase in FRET (Fig. 5A); this assay served as the basis for further work on the mechanism by which KF discriminates between correct and incorrect nucleotides[39]. We note that ALEX simplifies FRET analysis by sorting along the S axis, making it possible to exclude donor- and acceptor-only species,

produced through photobleaching or imperfect labelling, from analysis (Fig. 5A, pink shading in ALEX plot).

In addition to observing shifts in the mean FRET of a molecule, ALEX can be used to monitor the widths of FRET distributions, which may contain important information about heterogeneity in the molecule's structure[29]. This is typically done by determining the "shot noise" distribution: the FRET distribution expected if broadening is due solely to photon statistics[40,44]. FRET histograms wider than expected from shot noise alone can indicate the presence of static or dynamic structural heterogeneity in the molecule of interest[40,44-46]. Analysis of unliganded KF, for instance (Fig. 5B), identified a very broad distribution, well beyond that expected from shot noise[39]; recent work by our group suggests that this heterogeneity was likely dynamic, and consistent with fingers-closing dynamics on a millisecond timescale[37,38]. Such dynamics may be important to the mechanism by which DNA Polymerase I achieves its impressive fidelity in DNA replication. We discuss this work in depth in Chapters 4 and 5.

ALEX can also be used to monitor changes in fluorophore stoichiometry, making it possible to identify associations between two molecules labelled with spectrally distinct fluorophores, without the need for their fluorophores to be in close proximity. One powerful use of stoichiometry is in biosensing, for instance in the detection of miRNAs in human tissues[7], and transcription factors in whole-cell lysates of *E. coli*[8]. The latter assay is based on transcription factor "half-sites": a transcription factor binding site is split into two DNAs, each labelled with a spectrally distinct fluorophore, and containing short, complementary (~ 6 basepair) overhangs[8,47] (Fig. 5C). These DNAs anneal only transiently on their own; in ALEX experiments, they produce mostly donor- and acceptor-only bursts, and few coincident (or intermediate-stoichiometry) bursts. When an appropriate transcription factor is present, however, it binds to the transiently-annealed half-sites, stabilizing their association; this

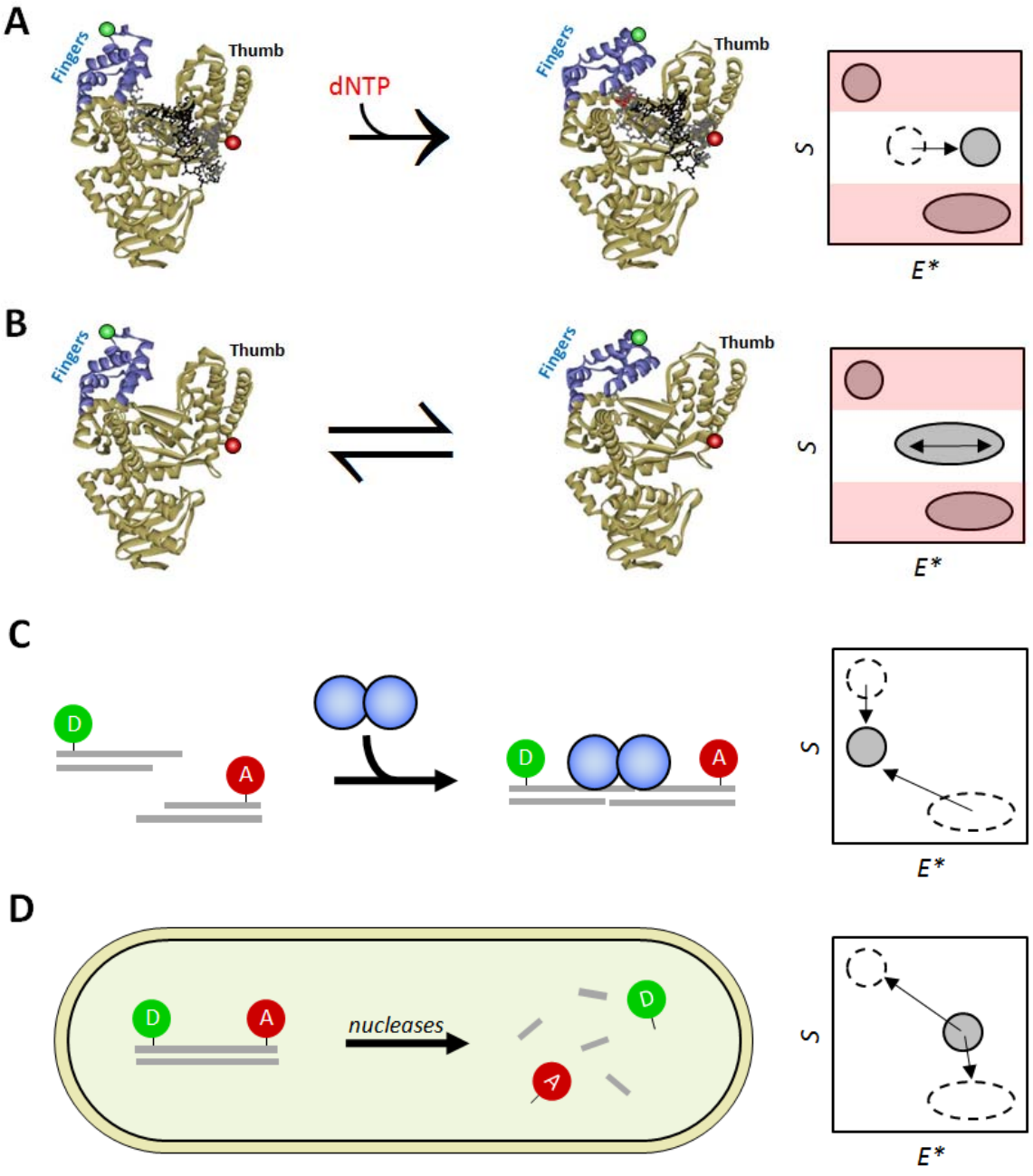

A
Fingers
Thumb
dNTP
Fingers
Thumb
S
E*
B
Fingers
Thumb
Fingers
Thumb
S
E*
C
D
A
D
A
S
E*
D
D
A
nucleases
D
A
S
E*

Figure 5. Applications of confocal single-molecule fluorescence. (A) Monitoring conformational changes with FRET. The fluorescently labelled Klenow Fragment (KF) of DNA Polymerase I (beige structure is a ribbon diagram of the KF backbone; black and grey structures are the top and bottom strands of the DNA template; red structure is the incoming dNTP; red and green spheres indicate positions of donor and acceptor fluorophores). KF undergoes a "fingers-closing" conformational change upon dNTP binding that results in increased FRET (*E*-*S* histogram, right). In addition to monitoring this conformational change on the *E** axis, sorting along the *S* axis allows exclusion of donor-only (upper left) and acceptor-only (lower right) populations from analysis (red shading). We study this transition further in Chapters 4 and 5. (B) The widths of FRET distributions can be used to identify static or dynamic structural heterogeneity in a sample of interest. Here, the unliganded Klenow Fragment exhibits a wide distribution due to dynamic fluctuations in its structure; we discuss this behaviour in Chapters 4 and 5. (C) Monitoring complex formation via changes in stoichiometry. In a transcription factor detection assay, two DNA "half-sites" diffuse independently, producing high-*S* (donor-only) and low-*S* (acceptor-only) populations; in the presence of an appropriate transcription factor, however, the half-sites bind stably, producing coincident species with intermediate *S*; the appearance of a coincident species can therefore be used to detect the transcription factor. We discuss this assay further in Chapter 3. (D) ALEX can also be used to monitor the loss of coincidence in a sample; in a doubly-labelled DNA sample, for instance, the loss of coincidence may be due to nuclease-mediated DNA degradation. In Chapter 6, we use ALEX to observe intact and degraded DNAs in the cytoplasm of *E. coli*.

depletes the donor- and acceptor-only populations and causes an intermediate-stoichiometry species to appear (Fig. 5C). We discuss this biosensor in greater detail in Chapter 3.

Whereas transcription factor detection relies on the appearance of a coincident species on the ALEX histogram, the disappearance of coincidence can also be used to monitor biological processes[6]. For instance, for a DNA labelled with two spectrally distinct fluorophores, the loss of stoichiometry indicates a loss of physical association between the fluorophores, and can therefore be used to track DNA melting, nuclease-mediated DNA degradation, or other processes impacting DNA stability (Fig. 5D). In Chapter 6, we use stoichiometry to distinguish between intact and degraded DNAs in the cytoplasm of *E. coli.*

2.6 Fluorescence correlation spectroscopy (FCS)

Fluorescence-correlation spectroscopy (FCS) is a common method of analyzing confocal fluorescence data to determine molecular concentrations, brightness, diffusive properties and dynamics, both *in vitro* and *in vivo*[48,49]. In FCS, a confocal fluorescent measurement is taken at the small ensemble level (nM concentrations) such that, on average, several molecules are inside the confocal volume at a time. The fluorescence emitted by these molecules fluctuates due to diffusion through the confocal volume, photophysics, and dynamic quenching processes (which may be used as a readout for molecular dynamics); FCS then autocorrelates the fluorescence signal to extract the timescales at which these fluctuations occur[33,50].

The autocorrelation function $G(\tau)$ is the correlation of the intensity time trace with itself, offset by time lag τ *:*

$$G(\tau) = \frac{\langle \delta I(t) \cdot \delta I(t+\tau) \rangle}{\langle I(t) \rangle^2} \qquad [8]$$

where $I(t)$ is the measured fluorescence intensity over time, and $\delta I(t) = I(t) - \langle I(t) \rangle$. Interpreting the autocorrelation curve requires fitting it to an analytical function describing

the types of fluctuations it exhibits. For example, in the common case of a molecule simply diffusing through the confocal volume, the analytical fitting function (derived previously assuming a 3D-Gaussian CEF) is[12,51]:

$$G(\tau) = \left(\frac{1}{CV_{eff}}\right)\left(1+\left(\frac{\tau}{\tau_D}\right)\right)^{-1}\left(1+a^{-2}\left(\frac{\tau}{\tau_D}\right)\right)^{-1/2} \qquad [9]$$

where C is the concentration of the fluorescent species, V_{eff} is the effective volume of the confocal spot, τ_D is the mean diffusion time of the fluorescent species through the confocal volume, and $a = z_o / w_o$, the ratio of the axial and radial radii of the confocal spot (defined as the distances in the axial and radial directions at which laser intensity has decreased by a factor of e^2).

To constrain this fit, it is typical to first analyze a standard sample with well-characterized diffusion coefficient D via FCS, and fit for the parameters τ_D and a. The radial radius of the beam can then be determined as $w_o = (4D\tau_D)^{1/2}$, the axial radius as $z_o = aw_o$ and the effective volume as $V_{eff} = \pi^{3/2} w_o^2 z_o$. The measurement can then be repeated for the sample of interest with only C and τ_D unconstrained, allowing accurate measurement of both molecular concentration and diffusion coefficient (where $D = w_o^2 / 4\tau_D$). This curve is shown in Fig. 6A. When the time lag is zero, the autocorrelation function $G(\tau)$ is simply $1/CV_{eff}$; the y-intercept can therefore be used to monitor the concentration of a fluorescent species (Fig. 6B).

As the time lag increases, the autocorrelation function decays with the diffusion time τ_D; extracting diffusion time allows us to infer various properties of the biomolecule. Since larger particles diffuse more slowly than smaller particles, diffusion time can be used as a

measure of molecular size (Fig. 6C). The hydrodynamic radius of a particle, a, can be calculated using the Einstein-Smoluchowski equation[52]:

$$a = \frac{k_B T}{6\pi\eta D} \qquad [10]$$

where k_B is the Boltzmann constant, T is the temperature, η is the viscosity of the solution, and D is the particle's diffusion coefficient. This ability to measure particle sizes has made FCS popular in studies of intermolecular binding and molecular aggregation[53,54].

In addition to diffusion, which causes fluorescence fluctuations due to the movement of fluorophores into and out of the confocal volume, processes that cause dynamic changes in the fluorescence intensity of the fluorophores themselves will also affect the shape of the autocorrelation function. Detectable dynamic timescales range from microseconds to seconds, making FCS a routine tool in the investigation of dye photophysics [55] (for triplet state blinking, typical timescale of ~10^{-6} s, Fig. 6D), and molecular dynamics (timescales typically ranging from 10^{-5}-10^{-2} s, Fig. 6E). The latter include careful studies on the dynamics of DNA hairpins[50,56,57], of enzyme-substrate binding [58], and of protein folding[59]. As the kinetics of a given source of photophysics or molecular dynamics are not always the same, an appropriate analytical function must be used to fit the autocorrelation curve in each case[55].

Significant drawbacks to FCS include its sensitivity to optical artefacts, which distort the shape of the confocal volume and can give rise to artefactual dynamic signals[60], and its relative insensitivity to diffusion-timescale dynamics, which overlap with the effects of diffusion and can therefore be difficult to identify (Fig. 6F). In addition, FCS is performed on the small-ensemble level, foregoing some of the benefits of single-molecule methods, such as the ability to isolate molecular subpopulations in complex samples, or ignore subpopulations with imperfect fluorescent labelling [30]. Although FCS is a useful and versatile technique, these difficulties form part of our motivation to develop single-molecule methods to quantify

diffusing fluorescent species (Chapter 3), and to detect and quantify diffusion-timescale dynamics (Chapters 4 and 5).

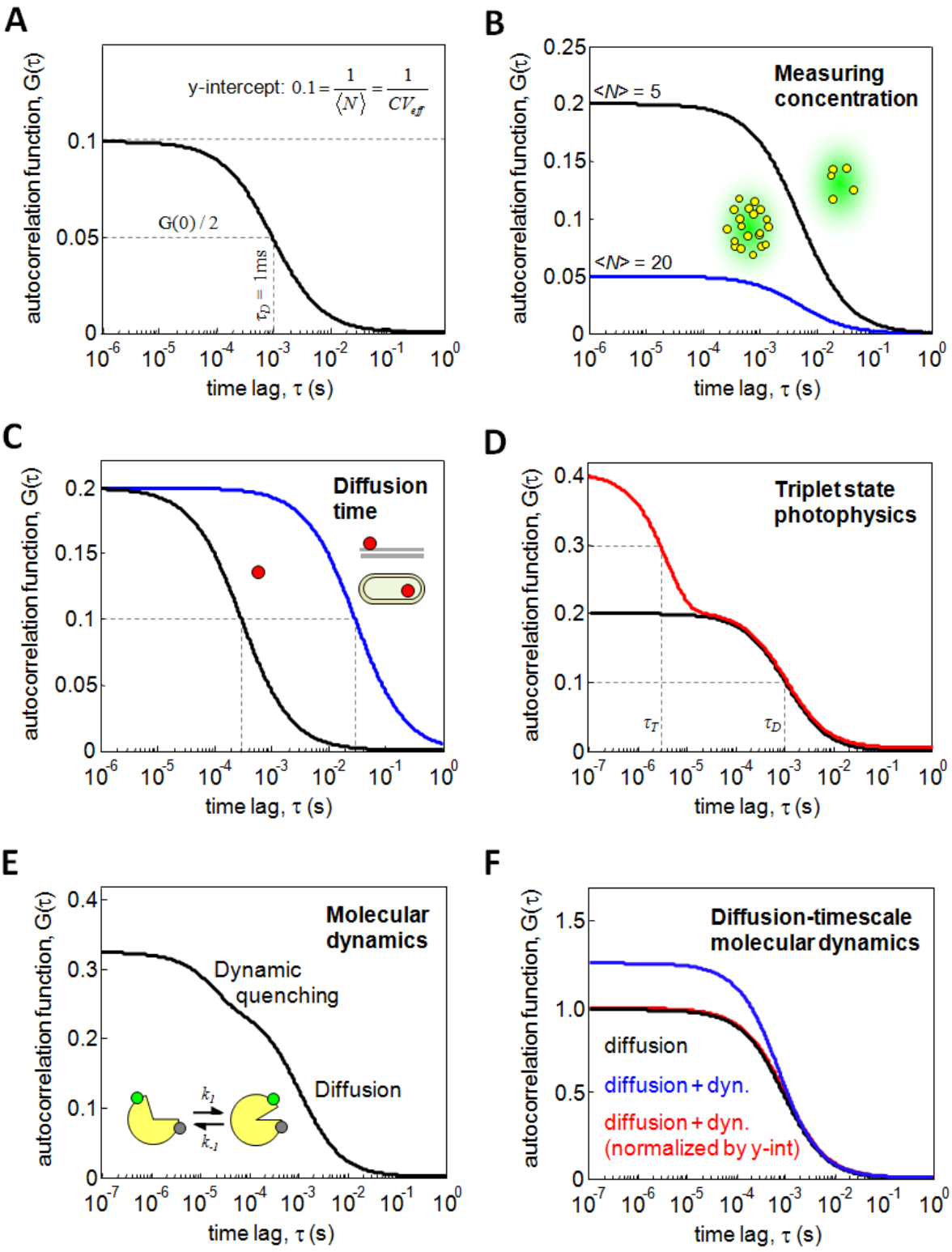


Figure 6. Applications of Fluorescence Correlation Spectroscopy (FCS). (A) The autocorrelation function, $G(\tau)$, for a fluorescent species diffusing through the confocal volume. At the y-intercept, the autocorrelation

function is equal to the reciprocal of the number of particles in the volume: $G(0) = 1/\langle N \rangle$. $G(\tau)$ decays with a characteristic time constant τ_D, the mean diffusion time of a particle through the confocal volume. The mean particle number, $\langle N \rangle$, and diffusion time, τ_D, can be used to calculate the concentration and diffusion coefficient of the fluorescent species; in this example, $\langle N \rangle = 10$ and $\tau_D = 1$ ms. (B) Using FCS to determine concentration. For samples with a higher concentration (blue line), the y-intercept ($G(0) = 1/\langle N \rangle$) decreases; by relating $\langle N \rangle = CV_{eff}$, where C and V_{eff} are the concentration and effective confocal spot volume respectively, the y-intercept can be used to calculate the concentration of a molecular species of interest. (C) Using FCS to determine molecular size and physical interactions. The black line indicates the autocorrelation function for a free fluorophore in solution. Molecules that diffuse more slowly (larger τ_D, blue line) may do so because they are physically larger (e.g. dye attached to labelled DNA) or because their diffusion is slowed by interactions with their environment (e.g. the same dye diffusing in the high-viscosity cytoplasm of a cell). (D) Using FCS to analyze dye photophysics. Shown is the effect of strong triplet-state blinking on the autocorrelation function (red line); triplet blinking creates an additional decay in the autocorrelation curve, distinct from that due to diffusion alone (black line). In addition to $\langle N \rangle$ and τ_D, the fraction of fluorophores in the triplet state, *F*, and the characteristic timescale of triplet blinking, τ_T, can be determined by fitting an appropriate function to this decay curve[55] (in this example, τ_T = 3 μs, *F* = 0.5). (E) Using FCS to analyze molecular dynamics. Similar to photophysics, dynamics resulting in blinking or quenching of the fluorophore on a detectable timescale will result in additional features in the curve; here we show a biomolecule fluctuating between two conformational states with rates k_1 and k_{-1}, and labelled with a dye (green) and quencher (grey). The timescale of the fluctuation is $1/(k_1 + k_{-1})$ = 20 μs, creating a component in the autocorrelation function around 20 μs that is distinct from the effects of diffusion ($\tau_D = 1$ ms). (F) While dynamics far from the diffusion timescale are easily apparent, those near the diffusion timescale are not. The autocorrelation function of a sample with a 1 ms diffusion time and 1 ms-timescale dynamic fluctuations (blue line) is the same as that of a sample with diffusion only (black), but scaled along the y-axis; this is obvious when the diffusion + dynamics autocorrelation function is normalized to the y-intercept of the diffusion-only sample (red). Dynamics near the diffusion timescale are therefore difficult to distinguish from simple changes in the concentration of the fluorescent species.

References

[1] English BP, Min W, van Oijen AM, Lee KT, Luo GB, et al. (2006) Ever-fluctuating single enzyme molecules: Michaelis-Menten equation revisited. *Nat Chem Biol* 2: 87-94

[2] Yu J, Xiao J, Ren XJ, Lao KQ, Xie XS (2006) Probing gene expression in live cells, one protein molecule at a time. *Science* 311: 1600-1603

[3] Schuler B, Eaton WA (2008) Protein folding studied by single-molecule FRET. *Curr Opin Struct Biol* 18: 16-26

[4] Zhao R, Rueda D (2009) RNA folding dynamics by single-molecule fluorescence resonance energy transfer. *Methods* 49: 112-117

[5] Weiss S (2000) Measuring conformational dynamics of biomolecules by single molecule fluorescence spectroscopy. *Nat Struct Biol* 7: 724-729

[6] Orte A, Birkett NR, Clarke RW, Devlin GL, Dobson CM, et al. (2008) Direct characterization of amyloidogenic oligomers by single-molecule fluorescence. *Proc Natl Acad Sci U S A* 105: 14424-14429

[7] Neely LA, Patel S, Garver J, Gallo M, Hackett M, et al. (2006) A single-molecule method for the quantitation of microRNA gene expression. *Nat Methods* 3: 41-46

[8] Lymperopoulos K, Crawford R, Torella JP, Heilemann M, Hwang LC, et al. (2010) Single-Molecule DNA Biosensors for Protein and Ligand Detection. *Angew Chem Int Ed Engl* 49: 1316-1320

[9] Selvin PR, Ha T, editors (2008) Single-Molecule Techniques. New York: Cold Spring Harbor Laboratory Press

[10] Kapanidis AN, Margeat E, Ho SO, Kortkhonjia E, Weiss S, et al. (2006) Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. *Science* 314: 1144-1147

[11] Liu SX, Abbondanzieri EA, Rausch JW, Le Grice SFJ, Zhuang XW (2008) Slide into Action: Dynamic Shuttling of HIV Reverse Transcriptase on Nucleic Acid Substrates. *Science* 322: 1092-1097

[12] Lakowicz JR (2004) Principles of Fluorescence Spectroscopy. New York: Springer

[13] Tsien RY (1998) The green fluorescent protein. *Annu Rev Biochem* 67: 509-544

[14] Elf J, Li GW, Xie XS (2007) Probing transcription factor dynamics at the single-molecule level in a living cell. *Science* 316: 1191-1194

[15] Roda A, Pasini P, Mirasoli M, Michelini E, Guardigli M (2004) Biotechnological applications of bioluminescence and chemiluminescence. *Trends Biotechnol* 22: 295-303

[16] Suhling K, French PMW, Phillips D (2005) Time-resolved fluorescence microscopy. *Photochem Photobiol Sci* 4: 13-22

[17] Kuhnemuth R, Seidel CAM (2001) Principles of single molecule multiparameter fluorescence spectroscopy. *Single Mol* 2: 251-254

[18] Vogelsang J, Kasper R, Steinhauer C, Person B, Heilemann M, et al. (2008) A reducing and oxidizing system minimizes photobleaching and blinking of fluorescent dyes. *Angewandte Chemie-International Edition* 47: 5465-5469

[19] Rasnik I, McKinney SA, Ha T (2006) Nonblinking and long-lasting single-molecule fluorescence imaging. *Nat Methods* 3: 891-893

[20] Dave R, Terry DS, Munro JB, Blanchard SC (2009) Mitigating Unwanted Photophysical Processes for Improved Single-Molecule Fluorescence Imaging. *Biophys J* 96: 2371-2381

[21] Chung HS, Louis JM, Eaton WA (2010) Distinguishing between Protein Dynamics and Dye Photophysics in Single-Molecule FRET Experiments. *Biophys J* 98: 696-706

[22] Förster T (1946) Energiewanderung und Fluoreszenz. *Naturwissenschaften* 6: 166-175

[23] Hillisch A, Lorenz M, Diekmann S (2001) Recent advances in FRET: distance determination in protein-DNA complexes. *Curr Opin Struct Biol* 11: 201-207

[24] Sabanayagam CR, Eid JS, Meller A (2005) Using fluorescence resonance energy transfer to measure distances along individual DNA molecules: corrections due to nonideal transfer. *J Chem Phys* 122: 061103

[25] Chan FK, Holmes KL (2004) Flow cytometric analysis of fluorescence resonance energy transfer: a tool for high-throughput screening of molecular interactions in living cells. *Methods Mol Biol* 263: 281-292

[26] Schuler B, Lipman EA, Steinbach PJ, Kumke M, Eaton WA (2005) Polyproline and the "spectroscopic ruler" revisited with single-molecule fluorescence. *Proc Natl Acad Sci U S A* 102: 2754-2759

[27] Yang H (2009) The Orientation Factor in Single-Molecule Förster-Type Resonance Energy Transfer, with Examples for Conformational Transitions in Proteins. *Israel Journal of Chemistry* 49: 313-321

[28] Shera EB, Seitzinger NK, Davis LM, Keller RA, Soper SA (1990) Detection of single fluorescent molecules. *Chem Phys Lett* 174: 553-557

[29] Deniz AA, Laurence TA, Dahan M, Chemla DS, Schultz PG, et al. (2001) Ratiometric single-molecule studies of freely diffusing biomolecules. *Annu Rev Phys Chem* 52: 233-253

[30] Kapanidis AN, Lee NK, Laurence TA, Doose S, Margeat E, et al. (2004) Fluorescence-aided molecule sorting: Analysis of structure and interactions by alternating-laser excitation of single molecules. *Proc Natl Acad Sci U S A* 101: 8936-8941

[31] Fries JR, Brand L, Eggeling C, Kollner M, Seidel CAM (1998) Quantitative identification of different single molecules by selective time-resolved confocal fluorescence spectroscopy. *J Phys Chem A* 102: 6601-6613

[32] Enderlein J, Robbins DL, Ambrose WP, Goodwin PM, Keller RA (1997) Statistics of single-molecule detection. *J Phys Chem B* 101: 3626-3632

[33] Rigler R, Mets U, Widengren J, Kask P (1993) Fluorescence correlation spectroscopy with high count rate and low background - analysis of translational diffusion. *European Biophysics Journal with Biophysics Letters* 22: 169-175

[34] Affleck RL, Ambrose WP, Demas JN, Goodwin PM, Schecker JA, et al. (1996) Reduction of luminescent background in ultrasensitive fluorescence detection by photobleaching. *Anal Chem* 68: 2270-2276

[35] Kapanidis AN, Laurence TA, Lee NK, Margeat E, Kong XX, et al. (2005) Alternating-laser excitation of single molecules. *Accounts of Chemical Research* 38: 523-533

[36] Lee NK, Kapanidis AN, Wang Y, Michalet X, Mukhopadhyay J, et al. (2005) Accurate FRET measurements within single diffusing biomolecules using alternating-laser excitation. *Biophys J* 88: 2939-2953

[37] Santoso Y, Torella JP, Kapanidis AN (2010) Characterizing Single-Molecule FRET Dynamics with Probability Distribution Analysis. *Chemphyschem* 11: 2209-2219

[38] Torella JP, Holden SJ, Santoso Y, Hohlbein J, Kapanidis AN (2010) Identifying molecular dynamics in single-molecule FRET experiments with Burst Variance Analysis. *Biophys J* (accepted):

[39] Santoso Y, Joyce CM, Potapova O, Le Reste L, Hohlbein J, et al. (2010) Conformational transitions in DNA polymerase I revealed by single-molecule FRET. *Proc Natl Acad Sci U S A* 107: 715-720

[40] Nir E, Michalet X, Hamadani KM, Laurence TA, Neuhauser D, et al. (2006) Shot-noise limited single-molecule FRET histograms: Comparison between theory and experiments. *J Phys Chem B* 110: 22103-22124

[41] Zhuang XW, Kim H, Pereira MJB, Babcock HP, Walter NG, et al. (2002) Correlating structural dynamics and function in single ribozyme molecules. *Science* 296: 1473-1476

[42] Deniz AA, Laurence TA, Beligere GS, Dahan M, Martin AB, et al. (2000) Single-molecule protein folding: Diffusion fluorescence resonance energy transfer studies of the denaturation of chymotrypsin inhibitor 2. *Proc Natl Acad Sci U S A* 97: 5179-5184

[43] Joyce CM, Benkovic SJ (2004) DNA polymerase fidelity: Kinetics, structure, and checkpoints. *Biochemistry (Mosc)* 43: 14317-14324

[44] Antonik M, Felekyan S, Gaiduk A, Seidel CAM (2006) Separating structural heterogeneities from stochastic variations in fluorescence resonance energy transfer distributions via photon distribution analysis. *J Phys Chem B* 110: 6970-6978

[45] Gopich IV, Szabo A (2007) Single-molecule FRET with diffusion and conformational dynamics. *J Phys Chem B* 111: 12925-12932

[46] Watkins LP, Chang HY, Yang H (2006) Quantitative single-molecule conformational distributions: A case study with poly-(L-proline). *J Phys Chem A* 110: 5191-5203

[47] Heyduk T, Heyduk E (2002) Molecular beacons for detecting DNA binding proteins. *Nat Biotechnol* 20: 171-176

[48] Kim SA, Heinze KG, Schwille P (2007) Fluorescence correlation spectroscopy in living cells. *Nat Methods* 4: 963-973

[49] Gurunathan K, Levitus M (2008) Applications of Fluorescence Correlation Spectroscopy to the Study of Nucleic Acid Conformational Dynamics. Progress in Nucleic Acid Research and Molecular Biology, Vol 82. pp. 33-69

[50] Bonnet G, Krichevsky O, Libchaber A (1998) Kinetics of conformational fluctuations in DNA hairpin-loops. *Proc Natl Acad Sci U S A* 95: 8602-8606

[51] Eigen M, Rigler R (1994) Sorting single molecules - application to diagnostics and evolutionary biotechnology. *Proc Natl Acad Sci U S A* 91: 5740-5747

[52] Haustein E, Schwille P (2003) Ultrasensitive investigations of biological systems by fluorescence correlation spectroscopy. *Methods* 29: 153-166

[53] Sengupta P, Garai K, Callaway DJ, Maiti S (2002) Kinetics of aggregation of amyloid beta peptide. *Biophys J* 82: 2466

[54] Pramanik A, Olsson M, Langel U, Bartfai T, Rigler R (2001) Fluorescence correlation spectroscopy detects galanin receptor diversity on insulinoma cells. *Biochemistry (Mosc)* 40: 10839-10845

[55] Schwille P, Haustein E (2001) Fluorescence correlation spectroscopy: an introduction to its concepts and applications. Biophysics Textbook Online. pp. 1-33

[56] Ying LM, Wallace MI, Klenerman D (2001) Two-state model of conformational fluctuation in a DNA hairpin-loop. *Chem Phys Lett* 334: 145-150

[57] Van Orden A, Jung J (2008) Fluorescence correlation spectroscopy for probing the kinetics and mechanisms of DNA hairpin formation. *Biopolymers* 89: 1-16

[58] Yang H, Luo GB, Karnchanaphanurach P, Louie TM, Rech I, et al. (2003) Protein conformational dynamics probed by single-molecule electron transfer. *Science* 302: 262-266

[59] Chattopadhyay K, Elson EL, Frieden C (2005) The kinetics of conformational fluctuations in an unfolded protein measured by fluorescence methods. *Proc Natl Acad Sci U S A* 102: 2385-2389

[60] Hess ST, Webb WW (2002) Focal volume optics and experimental artifacts in confocal fluorescence correlation spectroscopy. *Biophys J* 83: 2300-2317

3. Accurate molecular counting using confocal single-molecule fluorescence

We begin our discussion of confocal single-molecule fluorescence by addressing its simplest possible use: determining the concentration of a fluorescent species of interest. Molecular counting has been used in a number of applications, most notably to quantitatively determine the aggregation state of self-associating biomolecules (e.g. β-amyloid in neurodegenerative diseases[1]), to determine equilibrium binding constants[2], and in biosensor applications to detect transcription factors and microRNAs[3,4]. Despite its diverse applications, attempts to determine concentrations via confocal single-molecule fluorescence suffer from several artefacts. These include multi-molecule events, in which more than one molecule contributes to a single fluorescence burst, decreasing its apparent concentration[5]; confocal volume recrossing, in which a single molecule gives rise to multiple bursts, increasing its apparent concentration[2]; and brightness- and diffusion-dependent burst detection efficiency[1,6]. To address these problems, we derive simple, analytical correction factors which allow us to convert the number of fluorescence bursts in an experiment, to accurate concentrations. We validate these correction factors using numerical simulations of diffusing single molecules,

and use them to improve the accuracy of a confocal fluorescence-based transcription factor biosensor[3]. This method allows the determination of accurate molecular concentrations from confocal single-molecule fluorescence data, and should be useful both in future studies of binding thermodynamics and molecular aggregation, and in biosensing applications.

3.1 Introduction

One of the simplest applications of confocal single-molecule fluorescence is molecular counting: quantifying a given fluorescent species from the bursts it produces. Indeed, counting was one of the earliest applications of confocal fluorescence, where it was used to detect multiply-labelled antibodies and virus particles[7,8], and to generate the statistics needed to convincingly demonstrate the detection of single, freely-diffusing fluorophores[9,10]. In more recent applications, molecular counting has been employed to monitor the thermodynamics of intermolecular binding[11], as a platform for the detection of biomarkers[3,4], and to characterize molecular aggregation[1]. While useful in versatile applications, however, confocal single-molecule counting suffers from several artefacts, such as the unequal detection efficiencies between species of different diffusion coefficients and brightnesses[1,6].

The problem of molecular counting has previously been addressed in two ways: by employing fluorescence correlation spectroscopy (FCS), and by generating calibration curves. FCS is a popular tool for monitoring molecular concentrations (Chapter 2, Refs. [12-14]). Although quite accurate for relatively pure samples at the small-ensemble level, FCS can exhibit poor signal-to-noise on the single-molecule level[15], and does not have certain advantages of single-molecule fluorescence, such as the ability to resolve molecular subpopulations in complex samples, and to ignore the contributions of incompletely-labelled species[16].

Calibration curves involve determining the relationship between bursts and concentration empirically; if generated under the same conditions as the counting experiment to be performed, they will internally control for most experimental artefacts, making them a powerful, and typically very accurate, method of quantitation[4]. However, calibration curves are not appropriate in all cases. If counting is to be employed in complex biological samples or even live cells, where unknown binding partners may affect a given fluorophore's diffusive properties or brightness, it may not be appropriate to compare data collected *in vivo,* against calibration curves generated *in vitro*. Moreover, in analyzing mixtures of species with different burst detection efficiencies, it would be necessary to generate a calibration curve for every species of interest; in some complex samples, such as those exhibiting aggregation[1], this may prove impossible, owing to the difficulty of purifying out aggregates of specific sizes. These difficulties highlight the need for an independent analytical framework for converting fluorescence bursts, to accurate concentrations.

Converting bursts to concentrations is non-trivial. This is due to four properties of solution-phase confocal studies: (*i*) the diffusion-controlled encounter rate of molecules with the confocal volume, (*ii*) diffusive recrossing of the volume, (*iii*) brightness- and diffusion-dependent burst detection, and (*iv*) the potential for multi-molecule bursts.

These four challenges are illustrated in Fig. 1. The diffusion-controlled encounter rate (*i*) means that the number of distinct molecules diffusing into the confocal spot over the course of the measurement is not simply a function of concentration, but also of diffusion coefficient[1,15,17] (Fig. 1A). Recrossing (*ii*) refers to the tendency of molecules that visit the confocal volume to return to it, giving rise to more than one burst and increasing the apparent concentration of a molecule of interest[2] (Fig. 1B). Burst detection (*iii*) refers to the loss of bursts due to detection inefficiencies, resulting in a lower apparent molecular concentration; as detection efficiency is a complicated function of molecular brightness and diffusion

coefficient, the percentage of molecular encounters with the confocal volume that give rise to a detectable fluorescence burst may vary strongly from experiment to experiment, and between different analytes (Fig. 1C). Finally, multi-molecule events (*iv*) occur when one molecule enters the confocal volume before another leaves, resulting in a single burst that represents more than one molecule, and a decrease in the apparent concentration of the species of interest[5,6] (Fig. 1D). We note that while another major error in fluorescence methods is the effect of un- or incompletely-labelled molecules, these will create artefacts in any fluorescence-based quantification method, and must be controlled for externally.

To correct for the errors introduced by the four major effects mentioned, we use a four-step correction procedure:

- We calculate the concentration, c, from the number of encounters over an experiment, N_E, by determining the function: $c = f_1(N_E)$
- We calculate the number of encounters expected for a molecule exhibiting N_{ER} encounters and recrosses, by determining the function: $N_E = f_2(N_{ER})$
- We determine the number of encounters and recrosses that must have occurred to produce B_S, the number of bursts due to single molecules, by determining the function: $N_{ER} = f_3(B_S)$
- We determine B_S from the actual number of bursts, B, which is lower than B_S due to multi-molecule events: $B_S = f_4(B)$.

Finally, we reconstruct the concentration as a function of the number of bursts:

$$c(B) = f_1\left(f_2\left(f_3\left(f_4\left(B\right)\right)\right)\right)$$

Our method enables the calculation of accurate concentrations from the number of bursts in a confocal single-molecule fluorescence experiment by incorporating details about molecular diffusion and brightness, and without requiring either intensive numerical calculations or extensive modeling of the confocal setup[5,9]. We demonstrate the accuracy of this method in

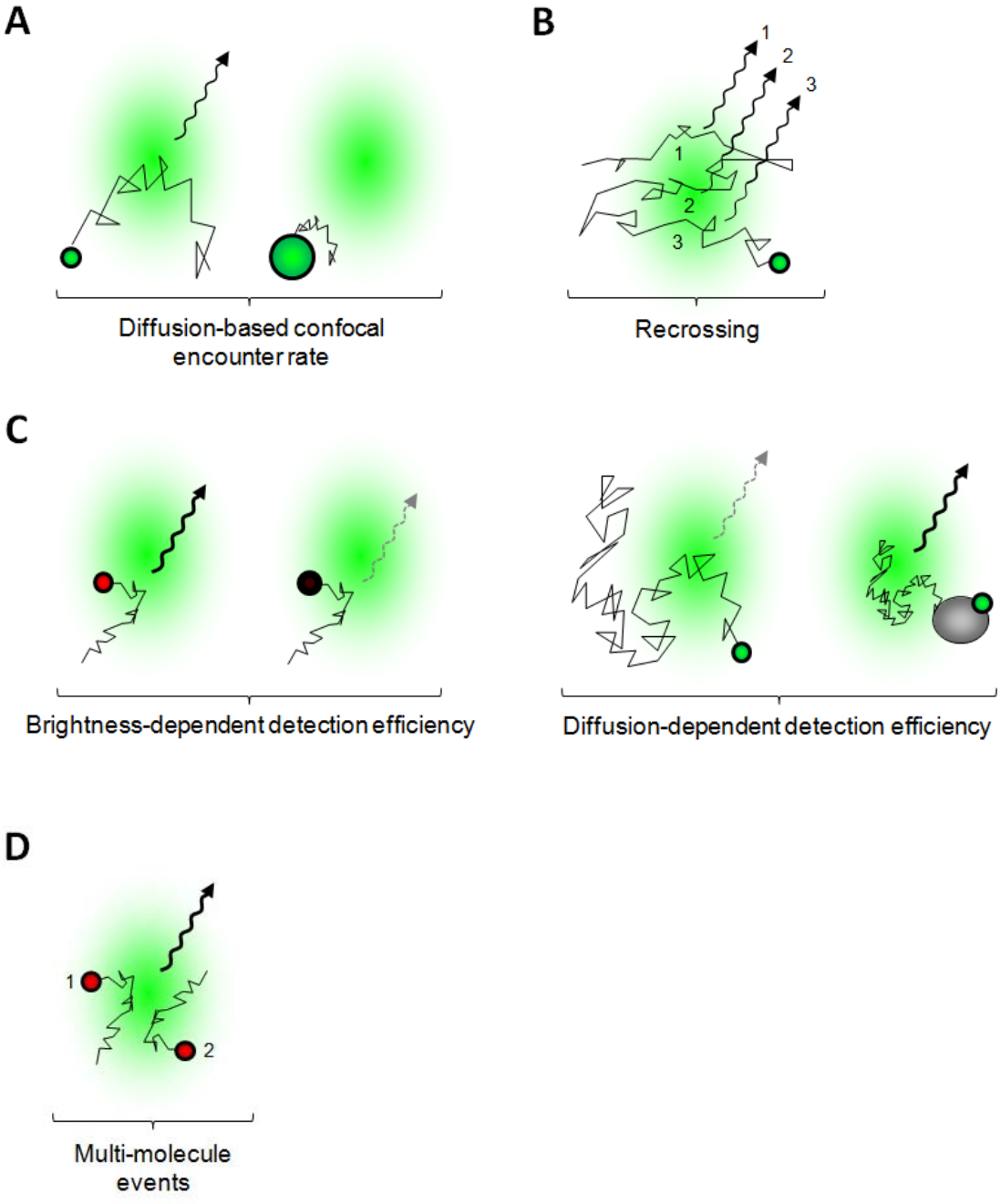

A
Diffusion-based confocal encounter rate
B
1
2
3
1
2
3
Recrossing
C
Brightness-dependent detection efficiency
Diffusion-dependent detection efficiency
D
1
2
Multi-molecule events

Figure 1. Sources of counting errors in confocal single-molecule fluorescence. (A) The number of molecules that encounter the confocal volume during a measurement, and are therefore capable of producing a burst of fluorescence, increases linearly with the diffusion coefficient of the molecule. Smaller molecules diffuse more quickly, and will therefore encounter the confocal volume more frequently than larger molecules. (B) Molecules can also exit, then recross into, the confocal volume, such that individual molecules can produce more than one burst, increasing their apparent concentration. (C) For each encounter with the confocal volume, a molecule has some probability of producing a detectable fluorescence burst. The chance of being detected increases with the brightness of the fluorophore (left), and with its mean diffusion time through the confocal volume (right). (D) If a molecule enters the confocal volume (2) before the previous molecule has left (1), the two bursts they would have produced will instead be detected as a single burst. At high concentrations, frequent multi-molecule events can substantially decrease the overall number of bursts detected.

quantifying molecules for a range of biologically relevant diffusion coefficients, and experimentally relevant fluorophore brightnesses. By using ALEX to separate species along FRET and stoichiometry axes, we can also quantify multiple species from a single confocal experiment; we apply this technique to improve the accuracy of a previously developed transcription factor biosensor[3]. Our method is simple and general, and should be useful in future studies that employ confocal single-molecule counting.

3.2 Theory and results

3.2.1 Effective confocal volume size

We begin by defining the effective size of the confocal volume. This will be required in subsequent calculations, in which we ask how frequently molecules collide with the confocal volume, and how frequently molecules entering the confocal volume are detected. We note that in FCS the confocal volume edge is taken as the position at which maximal laser intensity declines by a factor of e^2; for single-molecule fluorescence, however, this edge is more properly defined as the position at which laser intensity is just high enough for a given fluorophore to produce a detectable burst of photons.

For simplicity, we model the confocal volume as a sphere, with maximal intensity I_o at the origin. Following Chapter 2, we know that the intensity profile of this confocal spot is $I = I_o e^{-2r^2/w_o^2}$, where r is the distance from the origin, and w_o is the distance from the origin at which the intensity declines by a factor of e^2. We model fluorophores diffusing through the volume as having an effective brightness parameter, g: the number of photons they emit per second at the origin of the confocal spot.

We then define the radius of the confocal volume, a, as the position at which a fluorescent molecule would emit photons rapidly enough to be detected as a fluorescence burst. As our burst detection algorithm (Chapter 2) assigns a photon to a burst only if at least

N photons occur within time window T_{detect} of that photon, we approximate the edge of the confocal volume as the position at which a given fluorophore emits photons at a rate of N / T_{detect}. In all subsequent applications, we set $N = 7$ and $T_{detect} = 500$ μs; these parameters were chosen empirically for our confocal setup to eliminate "spurious" bursts due to the random clustering of background photons[11].

As the number of photons produced by a fluorophore per unit time is linear with intensity (under the assumption that the fluorophore is not experiencing substantial saturation or triplet state behaviour), we can determine the effective confocal volume radius, a, at which the intensity declines to N / T_{detect}, as $ge^{-2a^2/w_o^2} = N / T_{detect}$. Rearranging this equation:

$$a = w_o \sqrt{\ln\left(gT_{detect} / N\right) / 2} \qquad [1]$$

where w_o is the distance from the origin at which the intensity has declined by a factor of e^2, and g is the brightness of the fluorophore at the centre of the confocal spot (measured in photons/second). We use Eq. 1 to determine the effective size of the confocal volume in all subsequent applications. We note that although w_o cannot be easily obtained from single-molecule data, it can be easily measured by performing FCS on a standard dye sample (Chapter 2).

Finally, we note that while we consider a spherical confocal volume for simplicity, the edge of the confocal volume is actually spheroidal in most cases. Fortunately, the diffusive processes we consider in this chapter (i.e. encounter rates with adsorbing objects), are far more sensitive to the major than the minor length of a spheroid[18,19]. For a confocal volume with radial radius w_o and axial radius z_o, the latter typically being 3 – 6 times larger than the former[12], the effective confocal volume radius can therefore be approximated by applying Eq. 1 while setting $w_o = z_o$.

3.2.2 Relating concentration to encounter rate

We derive our first correction factor by relating the concentration of molecules in a sample, to their encounter rate with the confocal volume. For simplicity, we model the confocal volume as an adsorbing sphere of radius *a,* such that particles encountering the sphere are absorbed, and do not diffuse away; the encounter rate is therefore the rate at which particles encounter the confocal volume for the first time. The encounter rate of a freely-diffusing species with an adsorbing sphere, τ_E^{-1}, is[17]:

$$\tau_E^{-1} = 4\pi DacN_A \qquad [2]$$

where τ_E is the mean time between encounters, D is the diffusion coefficient, c is concentration, and N_A is Avogadro's number.

The encounter rate scales linearly with diffusion coefficient (Eq. 2) even when the concentration is constant; this is relevant to single-molecule counting because, even in the case of perfect burst detection, the number of bursts will be a function not only of concentration, but of diffusion coefficient. We confirmed this dependence by simulating the diffusion of particles at a concentration of 200 pM over 1 second (Methods; simulation software written by Johannes Hohlbein), and measuring the number of first encounters with a spherical confocal volume of radius a = 250 nm, over a range of diffusion coefficients commonly observed in studies of biomolecules ($3x10^6$ to $3x10^8$ nm^2 s^{-1}, Fig. 2). The fact that encounter rates changed by over an order of magnitude among these biologically relevant diffusion coefficients, despite no difference in concentration, illustrates the importance of accounting for confocal volume encounter rate in confocal single-molecule counting, and the accuracy of Eq. 2 in predicting this encounter rate.

A more general relationship than Eq. 2 can be determined for application in the following sections. By rewriting the diffusion coefficient in terms of the diffusion time of a particle through a sphere, $\tau_D = 3a^2 / D$ [17], replacing encounter rate with the number of bursts

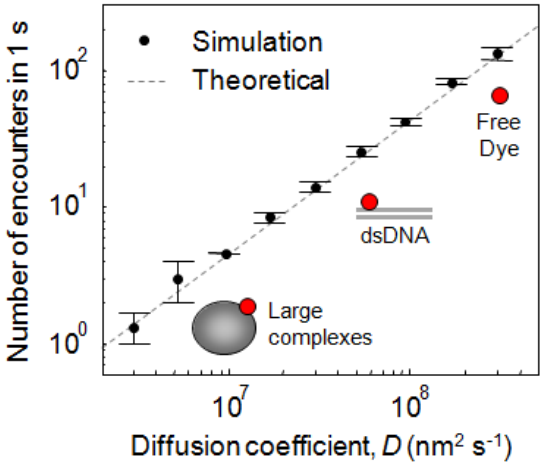


Figure 2. The effect of diffusion coefficient on confocal volume encounter rate. The rate of molecules encountering the confocal volume over the course of an experiment is given in Eq. 2, and is linearly dependent on diffusion coefficient, D. To demonstrate this, we performed numerical simulations of diffusing molecules (Methods; simulation software written by Johannes Hohlbein) interacting with a spherical confocal volume with radius a = 250 nm. We varied the diffusion coefficient over two orders of magnitude, within the ranges encountered in experiments on biological molecules (e.g. a typical organic fluorophore has D = 3x10^8 nm^2 s^{-1}, a 50 bp dsDNA has a D of about 3x10^7 nm^2 s^{-1}, while large protein complexes can have D as low as 10^6 nm^2 s^{-1}). We plotted the number of encounters occurring over the course of 1second. Error bars indicate the standard error of the mean number of bursts for N = 3 simulations. As expected, the number of encounters in each simulation followed the number of encounters predicted by Eq. 2.

detected over the experimental acquisition time, $\tau_E^{-1} = N_E / T_{acq}$, and replacing the volume of the sphere, $\frac{4}{3}\pi a^3$, with V_{eff}, we get:

$$c = f_1(N_E) = N_E \frac{\tau_D}{V_{eff} N_A T_{acq}} \qquad [3]$$

Eq. 3 is our first correction factor, and can be used to convert the number of encounters between a given species and the confocal volume, N_E, to c, the concentration of that species. We note that although derived for a sphere, the encounter rate for a spheroid can be approximated by assuming it is a sphere with a radius equal to the radius of its major axis[18].

3.2.3 Recrossing

In the last section, we determined the encounter rate of a diffusing species with an adsorbing sphere; however, this rate accounts only for the *first* encounter between a particle and the sphere. In practice, a particle may encounter the sphere, leave it, and return – a phenomenon referred to as 'recrossing' [2,20]. The total number of encounters between a diffusing species and the confocal volume (which we denote N_{ER}) is therefore larger than expected from the number of first encounters alone (N_E). To determine the true number of confocal volume encounters, we must therefore correct for the effects of recrossing.

While it is in principle possible, using numerical integration, to calculate the mean number of times a particle recrosses the confocal volume before diffusing away forever[21], most of these recrosses are closely spaced in time; as such, the photons they produce would likely contribute to the same fluorescence burst[2,6]. What matters for our purposes is therefore not how many times a particle re-enters the confocal volume; rather, we want to know how many times it enters the volume, leaves long enough for the burst to end, and only then re-enters.

To determine the relationship between N_E and N_{ER}, we asked what the probability was of a molecule re-entering the confocal volume after having produced a burst of fluorescence. One way of approaching this problem is to ask how far, on average, a molecule must diffuse away from the confocal volume to ensure the burst ends, and what the probability is of the molecule returning to it. As we assume bursts will be detected at the edge of the confocal volume, a, and our burst detection algorithm requires that a certain number of photons arrive in time window T_{detect} for a burst to occur (Methods), we make the assumption that, at the end of the burst, the molecule has diffused an average distance of $\sqrt{2DT_{detect}}$ away from the confocal volume's edge. This reframes the question as follows: for a confocal volume of radius a, and a particle dropped a distance $b = a + \sqrt{2DT_{detect}}$ from the origin (representing the distance at which the burst ends, on average), what is $p_{Recross}$, the probability of the particle diffusing to a before diffusing away forever?

To calculate $p_{Recross}$, we consider a set of three nested spherical shells (Fig. 3A): the confocal volume edge (radius $r = a$), the 'burst-end' shell ($r = b$), and an outer shell ($r = c$). We then ask what the probability is of a particle dropped at b diffusing to a before reaching c. This has previously been modelled by considering a and c as adsorbing spheres, b as a point of maximum concentration, and solving Fick's equations for the steady-state of such a system[18]. Doing this, the probability of diffusing to a before diffusing to c can be found by taking the ratio of the particle current into a, to the current into both a and c, which is equal to[18]:

$$p = \frac{a(c-b)}{b(c-a)} \qquad [4]$$

For recrossing, however, we are interested in the case in which the particle diffuses to a before diffusing away forever:

$$p_{Recross} = \lim_{c \to \infty}(p) = \frac{a}{b} \qquad [5]$$

Equation 5 says that the probability of starting at *b* and recrossing at least once is $p_{Recross} = a/b$. The probability of starting at *b*, and diffusing away forever without reaching *a*, is therefore $(1 - p_{Recross}) = (b-a)/b$.

Since all particles reaching *a* must eventually diffuse back to *b*, the probability of *k* recrosses before diffusing away forever is:

$$P(k) = (p_{Recross})^k (1 - p_{Recross}) \qquad [6]$$

The mean number of recrosses is therefore:

$$\langle P(k) \rangle = \sum_{k=0}^{\infty} (k)(p_{Recross})^k (1 - p_{Recross}) = \frac{p_{Recross}}{1 - p_{Recross}} = \frac{a}{b-a} \qquad [7]$$

To demonstrate the accuracy of this result in predicting the mean number of recrosses for a given 'burst-end' distance *b*, we simulated the diffusion of a particle (Methods) with D = 3 x 10^7 nm^2 s^{-1} between three spherical shells: a confocal volume of radius a = 250 nm, an outer shell of radius c = 10,000 nm (to approximate "diffusing away forever") and a "burst-end" shell of varying radius *b* (Fig. 3A). A simulated particle was dropped at *b* (as expected at the end of a burst), and the number of recrosses into the confocal volume, before diffusing as far as *c*, counted. We plotted both the theoretical (Eq. 7) and simulated recrosses per encounter as a function of $p_{Recross} = a/b$ (Fig. 3B). For all points tested (b = 260 nm to 5000 nm, or $p_{Recross}$ = 0.05 to 0.95), we found good agreement between both the simulated and theoretical numbers of recrosses.

The mean number of recrosses per encounter can be predicted using Eq. 7, and is related to the number of encounters with (N_{ER}) and without recrosses (N_E) by $\langle P(k) \rangle = (N_{ER} - N_E)/N_E$. We also assumed earlier that $b = a + \sqrt{2DT_{detect}}$. The overall correction factor is therefore:

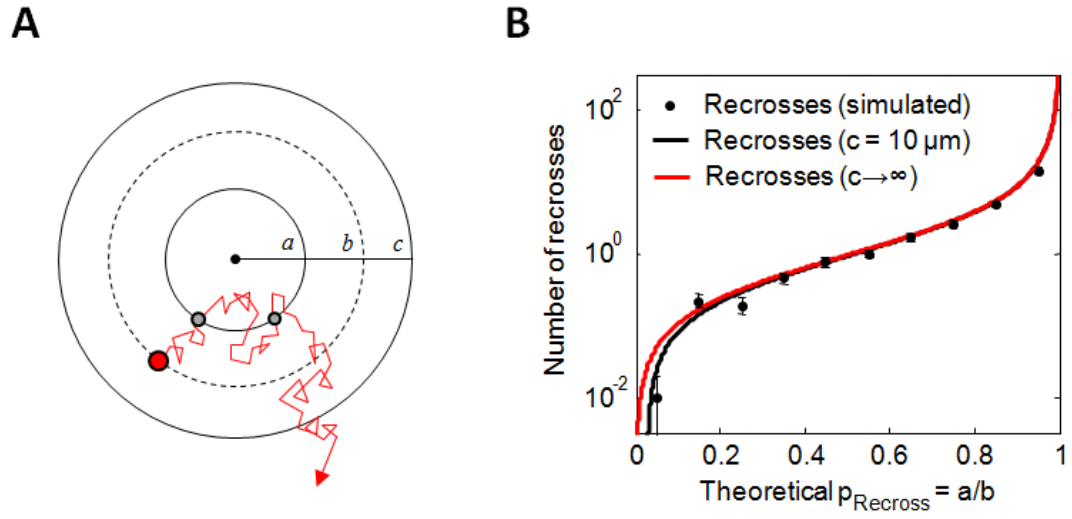


Figure 3. Correcting for confocal volume recrossing. (A) Schematic diagram of the confocal volume, and recrossing process. Following a burst, a particle will diffuse as far as *b,* on average, before the burst ends. At this point, the particle (red circle) can recross into the confocal volume, *a*, then back to *b*, possibly several times, before diffusing away forever. Shown is a single recross, in which a particle dropped at *b* diffuses to *a* (first grey circle) and eventually back to *b*, before diffusing away forever. Although the particle enters *a* a second time (second grey circle), it does not return to *b* first, and so does not produce a distinct fluorescence burst. (B) Simulations of confocal volume recrossing (circles) for a = 250 nm, c = 10,000 nm (to approximate "diffusing away forever"), and different values of b (plotted on the x-axis as $p_{Recross}$ = 250 nm / b). A simulated particle is dropped at b; the number of recrosses is simply the number of return trips between a and b, before reaching c (as illustrated in panel A). The simulations showed good agreement with the predicted number of recrosses (lines; Eq. 7). Each data point represents the mean and standard error of the number of recrosses observed in 100 independent simulations.

$$N_E = f_2\left(N_{ER}\right) = N_{ER}\left(1 - p_{Recross}\right) = N_{ER}\frac{\sqrt{2DT_{detect}}}{a + \sqrt{2DT_{detect}}} \qquad [8]$$

As in the last section, we note that although this correction factor assumes a spherical confocal volume, for spheroidal confocal volumes it can be approximated by assuming the sphere's radius is equal to the radius of the spheroid's major axis[18].

3.2.4 Detection efficiency and thresholding

Once the total number of encounters with the confocal volume (including recrossing), N_{ER}, is known, a major challenge is determining how many of these encounters result in a detectable fluorescence burst. Here we ask what fraction of encounters is detected by the burst selection algorithm.

For any given encounter with the confocal volume, a highly variable number of photons may be produced. If many photons are produced, the encounter will reliably be detected as a burst, while if only a few photons are produced, the encounter may fail to be detected as a burst. As a result, if we were to compare the experimental burst size distribution (BSD) (i.e. the distribution of photons per burst), and the actual BSD (i.e. the distribution of photons per *encounter*), then while the experimental BSD would match the actual BSD for bursts with many photons, it would significantly under-estimate the number of encounters with few photons.

One approach to determining the true number of confocal volume encounters, N_{ER}, from the number of single-molecule bursts, B_S, is to model the shape of the theoretical BSD. By fitting this to the experimental BSD for high photon counts, it would be possible to extrapolate out the actual BSD, including all encounters not successfully detected as bursts. Specifically, we assume that a theoretical BSD is fit to the experimental BSD, for all bursts

with at least C_{min} photons. If the fit is accurate, then the fraction of encounters with more than C_{min} photons, f_D, is:

$$f_D = \frac{B_S}{N_{ER}} = \frac{\sum_{C_{\min}}^{\infty} BSD(C_t)}{\sum_{1}^{\infty} BSD(C_t)} \qquad [9]$$

where C_t is the number of photons per encounter, and BSD(C_t) is the fitted, theoretical distribution of photon counts. By dividing the number of bursts with at least C_{min} photons, B_S, by f_D, we can recover the total number of encounters, N_{ER}. This approach assumes that all confocal volume encounters produce at least one photon detectable as part of a fluorescence burst; this is a consequence of our definition of the effective confocal volume radius, a (see first section of text).

Implementing this method first requires deriving an equation to describe the shape of the BSD. Previous work by Seidel and colleagues provided an analytical solution for the BSD of a 3D-Gaussian collection efficiency function (CEF), for bursts of fixed time length t [6]:

$$BSD(C_t; t, g) = \frac{3\ln\left[gt / C_t\right]^{1/2}}{2C_t \ln\left[gt\right]^{3/2}} \qquad [10]$$

where C_t is the number of photons, and g is a brightness parameter, describing the number of photons emitted per second by a given fluorescent species, at the centre of the confocal spot.

To determine the BSD for bursts of all lengths, we assumed an exponential distribution of residence times in the confocal volume with mean diffusion time τ_D, consistent with previous studies[6,21,22]. The overall BSD is:

$$BSD\left(C_t; g, \tau_D\right) = \frac{A}{C_t} \cdot \int_0^{\infty} \left(\frac{e^{-t/\tau_D}}{\tau_D} \right) \left(\frac{\ln\left[gt / C_t\right]^{1/2}}{\ln\left[gt\right]^{3/2}} \right) dt \qquad [11]$$

where A is a simple scaling parameter. While it does not have a simple numerical solution, Eq. 11 can be integrated numerically and fit to an experimental BSD (Fig. 4).

To demonstrate the fitting of Eq. 11 to an experimental BSD, we first simulated a particle with diffusion coefficient D = 3x10^7 nm^2 s^{-1} and brightness g = 10^5 s^{-1}, diffusing through a spherical confocal spot of radius a = 250 nm. We then performed a burst search on this data, and histogrammed the photon counts per burst to generate a simulated BSD (Fig. 4A). To fit Eq. 11 to this data, we first calculated τ_D by performing a linear fit to a log-transformed histogram of diffusion times (Fig. 4A, inset). We then fit Eq. 11 to the distribution above a threshold of C_{min} = 20 photons, with free parameters g and A (Fig. 4A), and achieved a good fit; the best-fit value of g, 1.2 ± 0.3x10^5 s^{-1}, was within error of the true value, 10^5 s^{-1}. Importantly, this fit was good only for large values of C_t, as the fitted BSD was much larger than the experimental BSD at low photon counts (C_t < 20). This is an important feature of the BSD: although the burst search is least likely to detect small bursts, they are the most frequent (Fig. 4A); as a result, thresholding prevents many, and usually most, confocal volume encounters from being detected. For the BSD in Fig. 4A, for instance, for a threshold of C_{min} = 20 photons, we asked what the fraction was of detected bursts, f_D, according to the theoretical BSD (Eq. 9). The f_D was 0.18; therefore, fewer than 20% of encounters (emitting at least one photon) emitted enough photons to make it above the threshold; the true encounter rate therefore differs from the number of bursts detected by a factor of about 5-fold for our simulation.

We note that while there is no simple criterion by which to choose C_{min}, in fitting this simulated data, we titrated up the size of C_{min} and re-performed the fit until a stable, maximum area under the curve was identified. The smallest value of C_{min} hitting this maximum is then taken as the actual C_{min} – the number of photons in a burst at which detection efficiency is effectively 100%. In all subsequent BSD fits, we use such a procedure to select C_{min}.

To translate confocal volume encounters (including recrosses, N_{ER}) to bursts arising from single molecules (B_S), we calculate the correction factor as:

$$N_{ER} = f_3(B_S) = \frac{B_S}{f_D(\tau_D, g)} \quad [12]$$

The τ_D and g terms are shown to indicate the dependence of this correction factor on the molecular diffusion time and brightness parameter, respectively. To demonstrate the importance of these properties on the detected fraction, we plotted f_D as a function of both typical biological diffusion times through the confocal volume (0.1 – 10 ms), and experimentally relevant brightnesses (10^4 – 10^6 photons s^{-1} at the confocal volume centre) (Fig. 4B). The detected fraction varied from 0.01 to 0.3 for diffusion times and brightnesses over two orders of magnitude. For dim and quickly-diffusing molecules, increases in τ_D and g produced large gains in detection efficiency (e.g. molecules with 1 ms diffusion times were detected ~10x more efficiently than those with 100 μs diffusion times). For brighter and more slowly-diffusing molecules, however, further increases in brightness or diffusion time did not have a very strong impact on the detected fraction (e.g., molecules with 10 ms diffusion times were only detected twice as efficiently as those with 1 ms diffusion times). This is intuitive, as for molecules already bright and slow enough to be frequently detected during confocal volume encounters, further increasing their brightness or diffusion time will produce only a marginal increase in detection efficiency. We note that detection efficiency is generally less than 30%, even for very bright and slowly-diffusing molecules; this is due to the shape of the BSD, which is heavily biased toward few-photon bursts (Fig. 4A; note logarithmic y-axis).

3.2.5 Correcting for multi-molecule events

Once the expected number of single-molecule bursts is known (B_S), a final correction must be made for multi-molecule events, in which two or more bursts are detected as a single, longer

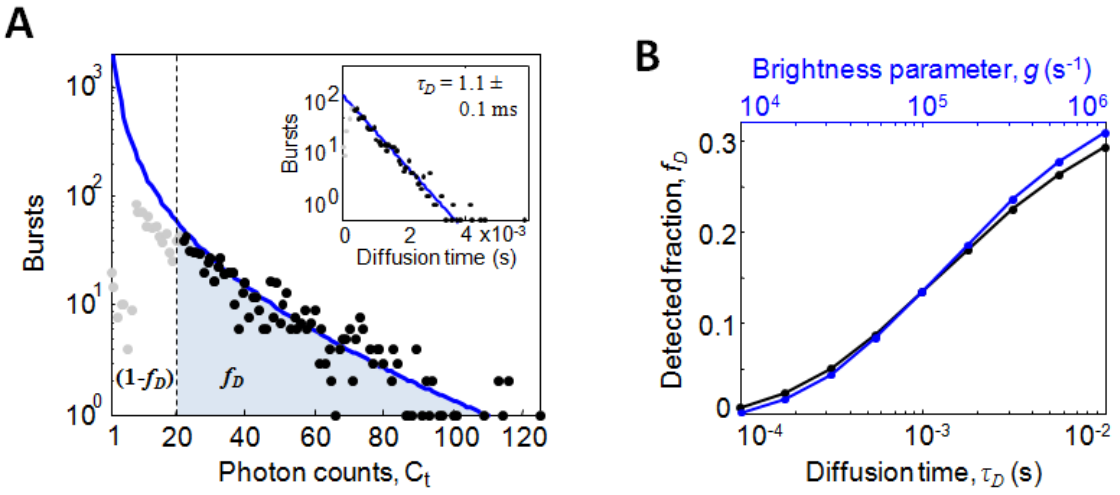


Figure 4. Determining the detected fraction, f_D. (A) Burst size distribution (BSD): the distribution of C_t, the photon counts per burst, from a numerical simulation of diffusing molecules (black circles; D = 3x10^7 nm^2 s^{-1}, a = 250 nm, g = 10^5 s^{-1}, T_{acq} = 10 s). To fit this distribution to our theoretical model, we first determined τ_D as 1.1 ms from a linear-fit to a log-transformed histogram of burst lengths (inset); we then performed a two parameter fit for g and A to data above a threshold of C_{min} = 20 (Eq. 11; grey circles are data points not included in the fit). The theoretical model achieved a good fit to the data, with g = 1.2 ± 0.3 x 10^5 s^{-1}, within error of the true value of 10^5 s^{-1}. The theoretical BSD predicts a large number of small-photon-number counts (for C_t < 20), most of which are not efficiently detected by the burst search algorithm (grey data points). The detected fraction, f_D, for a threshold of C_{min} = 20 photons (dashed line) is simply the theoretical percentage of bursts that have at least 20 photons (Eq. 9); for this simulation we calculated an f_D of 0.18, such that only 18% of all bursts lie above the threshold. (B) The effects of diffusion and brightness on the detected fraction, f_D (for a threshold of C_{min} = 20). Molecules that diffuse more slowly (larger τ_D) and fluoresce more brightly (larger g) are detected more efficiently. As most fluorescently labelled biomolecules, at typical excitation intensities, have a g between 10^5 and 10^6 s^{-1}, and diffusion times between 1- 10ms in a typical confocal volume, the difference in f_D between such molecules will generally be less than 2-fold.

burst, decreasing the apparent concentration of a species of interest[5]. In this section, we derive a correction factor for the loss of bursts due to multi-molecule events.

The probability of a given burst arising from more than one molecule can be calculated by considering the situation in Fig. 1D. A molecule (#1) enters the confocal volume; what is the probability that a second molecule (#2) enters the volume, before molecule #1 leaves? As mentioned previously, the distribution of times spent in the confocal volume is approximately exponential[6,21,22], $P(t_D) = (\tau_D{}^{-1})(e^{-t_D/\tau_D})$, where τ_D is the mean diffusion time of the particle through the confocal volume. The probability that the diffusion time of the first molecule, t_D, is greater than the inter-burst time, t_A, is simply:

$$\int_{t_A}^{\infty} P(t_D)\, dt_D = e^{-t_A/\tau_D} \qquad [13]$$

The inter-burst times (measured as the difference in start times between two consecutive bursts) are also exponentially distributed, $P(t_A) = (\tau_I{}^{-1})(e^{-t_A/\tau_I})$, with mean τ_I. Averaging over the distribution of arrival times for the second burst, the overall probability, q, of a second molecule arriving in the confocal volume before the first leaves is:

$$q = \int_0^{\infty} P(t_A) e^{-t_A/\tau_D}\, dt_A = \frac{\tau_D}{\tau_D + \tau_I} \qquad [14]$$

This equation is intuitive, as it says that for inter-burst times much greater than the diffusion time, the probability of two molecules encountering the confocal volume at the same time is negligible, while for diffusion times on the order of the inter-burst time, the probability of multi-molecule events will be large.

τ_D and τ_I can be obtained by fitting exponential functions to the experimental distributions of diffusion and inter-burst times, respectively, for the species of interest. As these allow us to calculate q, we can now return to our original question: what fraction of

detectable bursts is lost to multi-molecule events? The probability of k molecules contributing to a single burst is simply $q^k(1-q)$. We can therefore calculate this fraction as:

$$\sum_{k=0}^{\infty} q^k(1-q) = \left(\frac{q}{1-q}\right) \qquad [15]$$

Using Eq. 15, we can relate the number of bursts we would detect if there were no multi-molecule events, B_S, to the number of bursts actually detected, B:

$$B_S = f_4(B) = B\left(1-\frac{q}{1-q}\right)^{-1} = \frac{B\tau_I}{\tau_I - \tau_D} \qquad [16]$$

The usefulness of this simple correction is demonstrated in Fig. 5. We simulated diffusing fluorescent molecules at a series of concentrations (10 pM to 100 nM; Methods), and determined the number of bursts detected per second, B (Fig. 5A). For each data point, we obtained τ_D and τ_I by performing a linear fit to log-transformed histograms of diffusion and inter-burst times (Fig. 5B), and calculated B_S from B using Eq. 16. Both B and B_S are plotted in Fig. 5A.

At concentrations at or below 1 nM, the number of bursts scaled linearly with concentration and the correction factor was ~1, such that B and B_S are in agreement; this is because the mean inter-burst time is much longer than the diffusion time, such that two molecules rarely find themselves in the confocal volume at the same time. We note that although 1 nM is a high concentration for single-molecule experiments in practice, this is because the confocal volume in our simulation is a tiny 250 nm-radius sphere, which is significantly smaller than typical experimental volumes. For larger confocal volumes, we would therefore expect this linearity to apply only to concentrations significantly lower than 1 nM.

At concentrations above 1 nM, however, B was lower than expected from a linear prediction, due to multi-molecule events. Application of our correction factor (Eq. 16)

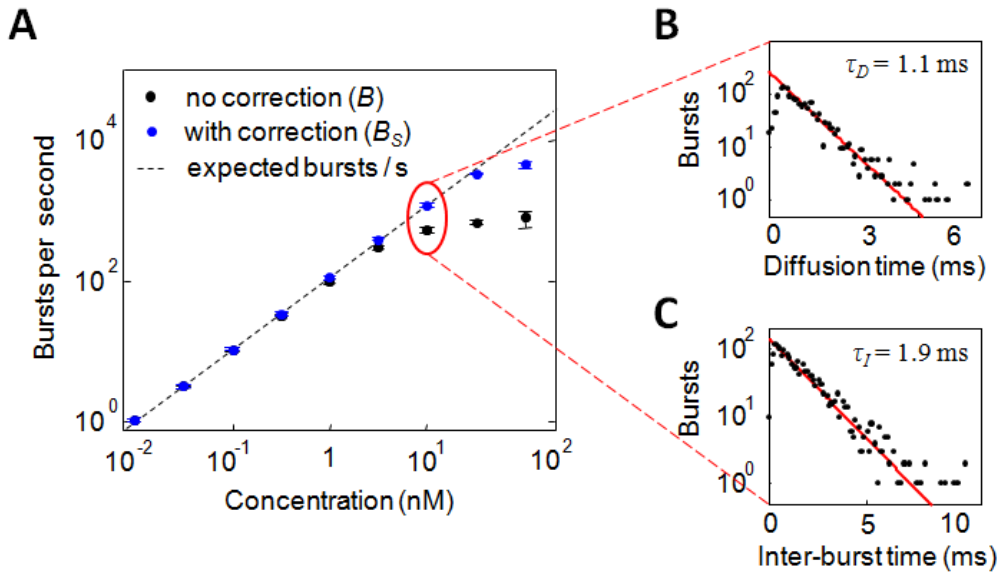


Figure 5. Correcting for the loss of bursts to multi-molecule events. (A) Bursts, B, as a function of concentration (black) for a set of simple numerical simulations (D = 3x10^7 nm^2 s^{-1}, a = 250 nm, g = 10^5 s^{-1}). Whereas B should increase linearly with concentration (dashed line), it is actually sub-linear with concentrations above 1 nM; this is due to multi-molecule events, in which two or more bursts overlap sufficiently to be detected as a single burst. For concentrations up to 36 nM, this effect can be corrected by incorporating an appropriate correction factor to calculate B_S (Eq. 16, blue circles), the number of bursts expected in the absence of multi-molecule events. Above 36 nM, however, the correction factor fails to account fully for the loss of bursts to multi-molecule events, but still performs better than if no correction is used at all. (B, C) Calculating B_S. B_S is calculated as $B_S = B\tau_I / (\tau_I - \tau_D)$, where τ_I and τ_D are the mean inter-burst and diffusion times for the molecule, respectively. As both diffusion and inter-burst times are exponentially distributed, we perform a linear fit to log-transformed histograms of this data to recover τ_I and τ_D. Shown are fits for the 10 nM data point in panel A (red circle). The correction is calculated as $B_S = B(1.9\text{ ms}) / (1.9\text{ ms} - 1.1\text{ ms}) = 2.38$.

accurately recovered B_S, as indicated by its consistency with the linear prediction, for concentrations up to 36 nM, more than an order of magnitude greater than was possible without the correction.

Above 36 nM, B_S was still closer to the linear prediction than B, but no longer matched it within error. This is primarily because, as the number of multi-molecule events increases, burst start times get absorbed into other bursts; as a result, the mean inter-burst time fails to decrease proportionally with the number of encounters, making our correction factor inaccurate for very high concentrations. While a more sophisticated correction can be applied by taking this effect into account, we note that at 36 nM, nearly 90% of all bursts observed were due to multi-molecule events, which is no longer really representative of the "single-molecule" regime, but of a small ensemble. For our purposes, the simple correction factor derived in Eq. 16 is therefore sufficient.

Overall, this correction factor makes it possible to accurately recover the number of single-molecule bursts from the number of actual bursts for concentrations over an order of magnitude higher than is possible without this correction.

3.2.6 Overall correction factor

We have now derived a set of correction factors allowing us to relate concentration to bursts by accounting for encounter rate, the increase in encounters due to recrossing, and the loss of encounters due to detection inefficiencies and multi-molecule events. These correction factors can be summarized as:

$$c = f_1\left(N_E\right) = N_E \frac{\tau_D}{V_{eff} N_A T_{acq}} \qquad [3]$$

$$N_E = f_2\left(N_{ER}\right) = N_{ER} \frac{\sqrt{2DT_{detect}}}{a + \sqrt{2DT_{detect}}} \qquad [8]$$

$$N_{ER} = f_3\left(B_S\right) = \frac{B_S}{f_D\left(\tau_D, g\right)} \quad [12]$$

$$B_S = f_4\left(B\right) = \frac{B\tau_I}{\tau_I - \tau_D} \quad [16]$$

Because each term can be related to the others by a simple analytical function, we can determine the concentration as a function of the number of bursts:

$$c\left(B\right) = f_1\left(f_2\left(f_3\left(f_4\left(B\right)\right)\right)\right) \quad [17]$$

By substitution, the overall correction factor is:

$$c\left(B\right) = B \cdot \left(\frac{\tau_I}{\tau_I - \tau_D}\right) \cdot \left(\frac{1}{f_D\left(\tau_D, g\right)}\right) \cdot \left(\frac{\sqrt{2DT_{Detect}}}{a + \sqrt{2DT_{Detect}}}\right) \cdot \left(\frac{\tau_D}{T_{acq} V_{eff} N_A}\right) \quad [18]$$

Eq. 18 can be used to recover molecular concentrations from the number of bursts observed in an experiment. In addition to containing each of the individual correction factors discussed in previous sections, Eq. 18 reveals a few interesting interactions between these correction factors. For instance, although larger diffusion coefficients increase the number of encounters with the confocal volume (Eq. 3), they decrease the number of confocal volume recrosses (Eq. 8), and decrease the probability that any given encounter is detected as a burst (Eq. 12); the relationship between diffusion coefficient and the number of bursts detected is therefore nontrivial.

To demonstrate the utility of Eq. 18 as an overall correction factor for relating the number of bursts in an experiment to its concentration, we simulated diffusing molecules over a range diffusion coefficients (Fig. 6A, D = 3x10^6 to 3x10^8 nm^2 s^{-1}, while fixing g = 10^5 s^{-1}) and brightnesses (Fig. 6B, g = 10^4 to 10^6 s^{-1}, fixing D = 3x10^7 nm^2 s^{-1}), all at a concentration of 1 nM; we then asked if we could recover the 1 nM concentration from the number of bursts, using the correction factors derived in this chapter. As corrections for multi-molecule bursts have already been illustrated (Fig. 5), the simulations in Fig. 6 are

performed at concentrations low enough (100 pM) to ensure a negligible number of multi-molecule events.

To calculate the overall correction factor (Eq. 18), we did the following:

- We performed a linear fit to the log-transformed histograms of diffusion and inter-burst times to extract τ_D and τ_I, respectively
- Fixing τ_D, we fit Eq. 11 to the BSD to extract *g*, which we then used to calculate f_D
- We determined the effective confocal volume radius, *a*, using Eq. 1, where w_o was set to the axial radius of our confocal setup (z_o), previously determined to be 1.4 μm via FCS (Chapter 2).
- We determine the diffusion coefficient of the molecule as $D = 3a^2 / \tau_D$

As a comparison with these correction-factor-derived concentration estimates, we also plot the concentrations recovered via "calibration." In this case, we assume that the number of bursts detected in the sample with D = 3x10^7 nm^2 s^{-1} and g = 10^5 s^{-1}, corresponds to a 1 nM concentration. For all other samples, we then simply assume that concentration is proportional to the number of bursts detected, regardless of their brightness or diffusive properties; for example, a sample with half as many bursts as the calibration sample would be assumed to have a concentration of 0.5 nM. This mimics the effects of using a single calibration curve in confocal single-molecule studies, where different molecular species may be detected with different efficiencies.

Comparing our correction-factor approach to the calibration-based approach, we found that for differences of two orders of magnitude in both diffusion coefficient and molecular brightness, our correction factor-based approach was capable of determining the true molecular concentration to within a factor of 2 in all cases (Fig. 6A, B). In contrast, the

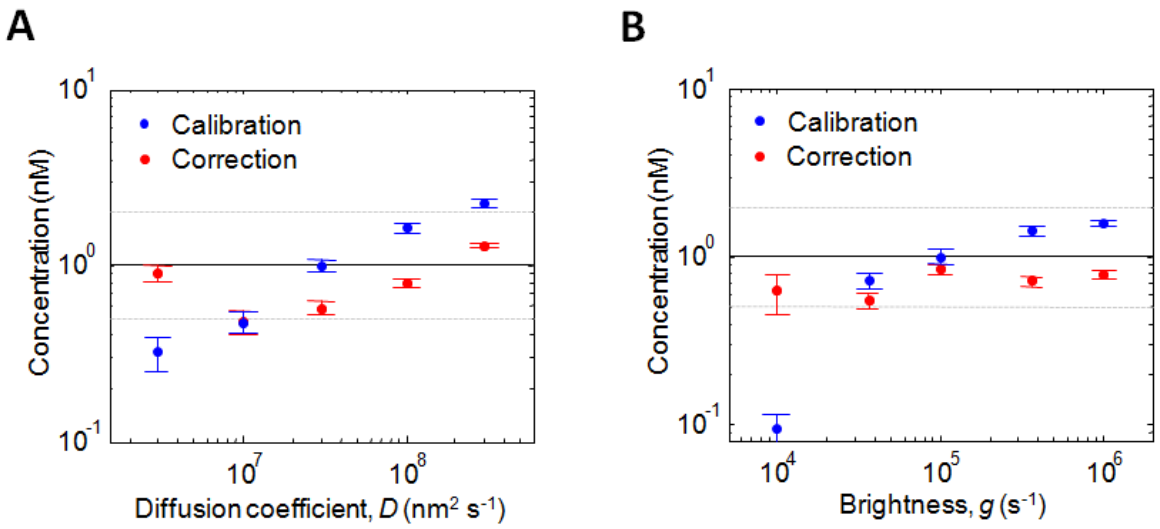


Figure 6. Accurate quantitation of molecular concentration via correction factors or calibration. Simulations of a 1 nM fluorescent species diffusing through the confocal volume were performed (a = 250 nm, T_{acq} = 10 s), and concentrations recovered using either our correction-factor approach (to convert bursts and photon statistics to concentrations), or a calibration approach (we "calibrated" against the D =3x10^7 nm^2 s^{-1}, g = 10^5 s^{-1} sample which has 363 bursts, such that each burst detected, in each sample, is assumed to represent (1/363) nM of concentration, regardless of molecular properties). (A) Fixing g = 10^5 s^{-1}, we simulated molecules with a range of biologically-relevant diffusion coefficients (3x10^8 – 3x10^6 nm^2 s^{-1}, representing the range from single fluorophores, to large protein complexes). Our correction-factor approach (red circles) recovered concentrations to within 2-fold (grey dashed lines) of the actual concentration of 1 nM (black horizontal line), for all diffusion coefficients tested. Calibration (blue circles) performed similarly well for g and D within an order of magnitude of those of the calibration sample (D = 3x 10^7 nm^2 s^{-1}, g = 10^5 s^{-1}), but became less accurate at more extreme diffusion coefficients. (B) Fixing D = 3x10^7 nm^2 s^{-1}, we performed simulations over a range of values for the brightness parameter g (10^4 – 10^6 photons s^{-1}). Again, all concentrations recovered using the correction factor approach fell within two-fold of the true concentration (1 nM). While the calibration approach was similarly effective at recovering concentrations for brightnesses above 3x10^4 s^{-1}, it was dramatically worse at recovering the concentration from dim species (g = 10^4 s^{-1}), underestimating the true concentration by a factor of 10.

calibration approach differed by more than a factor of two-fold in several cases, and by up to an order of magnitude (in Fig. 6B, predicted concentration of ~0.1 nM for $g = 10^4$ s^{-1}, $D = 3x10^7$ nm^2 s^{-1}). However, we note that the calibration approach was generally appropriate for diffusion coefficients or brightnesses within an order of magnitude of those of the sample against which calibration was performed; in cases where molecular species are reasonably similar, a calibration-based approach may therefore be simpler to implement than a full correction-factor based analysis. As our approach was generally more accurate than the calibration approach, however, the extra effort to implement it may be worthwhile in studies where careful quantitation of diffusing molecules or their interactions is desired.

3.2.7 Improving the accuracy of a molecular counting-based biosensor

To demonstrate the utility of our correction factor approach in an experimental context, we re-analyzed results from a previous publication by our group on transcription factor biosensing[3]. Transcription factors (TFs) are proteins that bind to a specific DNA sequence and modulate gene expression in response to environmental signals, such as temperature changes, exposure to light, and the presence of metabolites[23]. TFs are also extremely important in human health and disease[24]; the transcription factor p53, for instance, is the single most frequently mutated gene in cancer, with p53 loss of function mutations appearing in ~50% of all human tumours[25]. Biosensing assays that detect the binding of a TF can therefore be used to quantify the amount of that TF, monitor the stimuli that affect its binding, or identify the faulty TF binding characteristic of disease pathologies.

We built a TF biosensor on the basis of a detection platform first used by Heyduk and colleagues[26]. The DNA binding site for a given TF is split into two 'half-sites,' each of which contains a spectrally distinct fluorophore, and a 6 bp, single-stranded DNA overhang (Fig. 7A). In solution, the half-sites diffuse independently of one another, and associate only

transiently due to annealing between their single-stranded regions. We perform ALEX, and search for bursts in the red-excitation, red-emission channel (Methods); analysis of half-sites alone therefore shows mostly acceptor-only species, and few or no coincident (i.e. donor- and acceptor-bearing) species. In the presence of an appropriate TF, the transiently-annealed half-sites are stabilized by TF binding; this depletes the acceptor-only population and produces a strong coincident signal via ALEX (Fig. 7B). In this way, the appearance of coincident bursts via ALEX can be used to detect a given TF.

We previously applied this method to an *E. coli* transcription factor, the catabolite activator protein (CAP; sequences of half-site DNAs for CAP detection are given in the Appendix, Fig. S1). To quantify the amount of CAP detected in this assay, we built a thermodynamic model of half-site/CAP binding, relating CAP concentration to the bound fraction, F_B (the fraction of half-site bound to TF, rather than freely diffusing, as determined via ALEX; calculation of F_B is explained in Methods). The use of an explicit thermodynamic model allows direct quantitation of TF in a sample – a necessity when working in intracellular or other complex environments, where the generation of a calibration curve may be difficult or impossible.

The system's behaviour can be described by two coupled equilibria (Fig. 7A): one for half-site association (H_1+H_2 ⇌ H_1-H_2, with equilibrium dissociation constant K_{D1}), and another for TF binding to H_1-H_2 (with equilibrium dissociation constant K_{D2}). We used this model to generate an analytical expression for the total CAP concentration as a function of F_B, constants K_{D1} and K_{D2}, and the total concentration H_{tot} of each half-site (equations used to derive this expression are listed in the Appendix). The total CAP concentration as a function of F_B is:

$$[CAP]_{tot} = [H_1]_{tot} F_B + \frac{F_B K_{D2}}{2[H_1]_{tot}(1-F_B)^2} \times \left(2[H_1]_{tot}(1-F_B) + K_{D1} + \sqrt{K_{D1}}\sqrt{4[H_1]_{tot}(1-F_B) + K_{D1}}\right) \quad [19]$$

In a previous publication[3], we applied this model to a simple titration experiment, in which we fixed the concentration of DNA half-sites at 100 nM, and titrated CAP from 0 to 500 nM. These experiments, and the subsequent analysis to determine the number of bursts, were performed by Robert Crawford. To model the titration data, we fixed K_{D1} at 300 nM from ensemble measurements of half-site association (Ref. [3]), then fit Eq. 19 to the titration data to determine K_{D2}. Our model achieved a good fit to the data for a K_{D2} of 8 ± 2 pM; however, previous studies had measured K_{D2} at 24 ± 0.3 pM, a difference of three-fold[27].This is unsurprising: because F_B is meant to represent the relative *concentrations* of freely-diffusing and TF-bound half-sites, the use of F_B is appropriate only if burst detection efficiency is the same for both the half-site and coincident species (which we would expect it not to be, given the significantly larger size of CAP-bound half-sites, relative to free half-sites). We note also that, although relatively modest, three-fold differences in TF concentration or binding affinity are more than that observed in many disease pathologies[24], and greater than the error we observed employing our correction-factor method in numerical simulations (Fig. 6). This suggests that application of our correction factor method to the CAP detection assay would be both useful and appropriate.

To apply our correction-factor method to the titration data, we first determined the per-burst correction factors for both CAP-bound and free half-site DNA by analyzing a sample containing 100 nM half-site DNA and 500 nM CAP, which has an appreciable number of bursts representing both free and CAP-bound DNA. We separated these two populations by thresholding on the basis of stoichiometry, where bursts with $S > 0.4$ are assumed to represent CAP-bound DNAs, and those with $S \leq 0.4$ represent freely-diffusing acceptor-labelled half-sites (Fig. 7B, dashed line). Using photon counts from the red-excitation, red-emission channel, we then performed our correction-factor method (we use

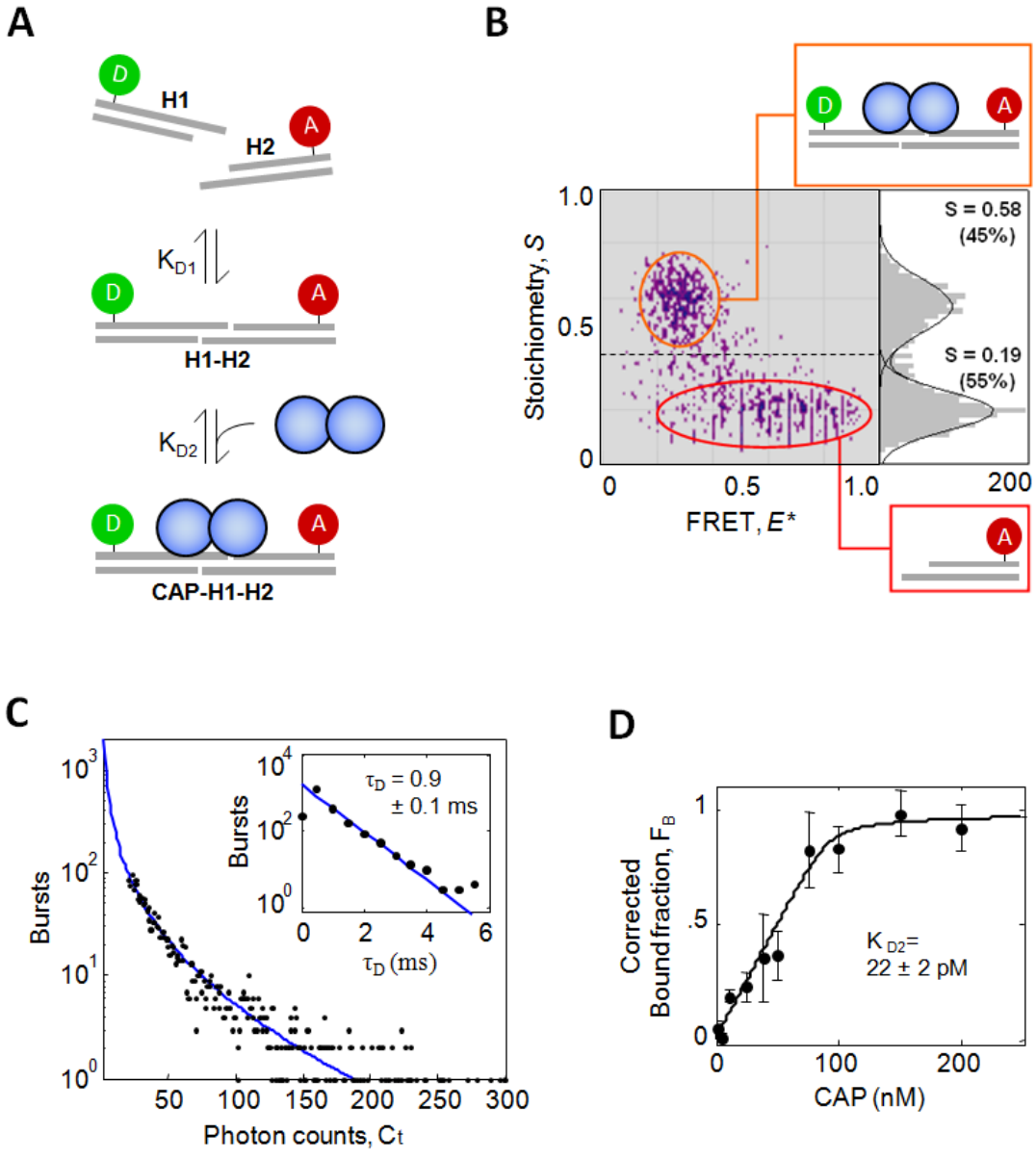

A
D
H1
A
H2
K_D1
D
A
H1-H2
K_D2
D
A
CAP-H1-H2
B
D
A
1.0
0.5
0
Stoichiometry, S
S = 0.58
(45%)
S = 0.19
(55%)
0
0.5
1.0
200
FRET, E*
A
C
10^3
10^2
10^1
10^0
Bursts
10^4
10^2
10^0
Bursts
τD = 0.9
± 0.1 ms
0
2
4
6
τD (ms)
50
100
150
200
250
300
Photon counts, Ct
D
1
.5
0
Corrected
Bound fraction, FB
K D2=
22 ± 2 pM
0
100
200
CAP (nM)

Figure 7. Applying the correction factor to a transcription factor biosensor. (A) Schematic diagram of TF detection. Two DNA half-sites (H_1 and H_2) bearing half a CAP binding site, a spectrally distinct fluorophore (red and green circles) and complementary 6 bp overhangs, anneal transiently, with dissociation constant K_{D1}. In the presence of CAP, binding to the transiently-annealed half-sites (dissociation constant K_{D2}) stabilizes their interaction. This gives rise to (B) a large coincident population (intermediate stoichiometry) as detected by ALEX, representing the TF-bound half-sites. The acceptor-only population represents the free acceptor-labelled half-site. In this example, the bound fraction, F_B, is 45%, as 45% of bursts occur at intermediate, rather than low, stoichiometry. Donor-only bursts are not observed due to the burst search method employed (Methods). For analysis, acceptor-only and coincident bursts are separated computationally by thresholding at $S = 0.4$ (dashed-line). All TF-binding experiments were performed by Robert Crawford. (C) Example BSD and diffusion time (inset) fits for CAP-bound DNA. Using the correction-factor method, we determined that CAP-bound half-sites are detected 34% more efficiently than acceptor-labelled DNA. (D) Using the correction factor approach, we re-calculated F_B for the difference in detection efficiencies between the TF-bound and free half-site DNAs; a single-parameter fit to this data (described in text) yields a K_{D2} of 22 ± 2 pM.

only $R_{ex}R_{em}$ photon counts because both the free and CAP-bound half-sites should have similar brightnesses in this channel).

As in our numerical simulations, we fit the diffusion time distribution and BSD as discussed previously, and achieved good fits to the data (Fig. 7C; τ_D = 0.9 ± 0.1 ms, g = 1.5 ± 0.3 x10^5 s^{-1}). The effective confocal volume radius was determined by employing Eq. 1, and assuming $w_o = z_o$, where z_o = 1.4 μm is the axial e^{-2} radius of our confocal spot as determined by FCS. We then re-calculated F_B (Methods) using the correction factor derived from these parameters, and re-fit our model to the data (Fig. 7D). For the correction-factor-modified data, we obtained a K_{D2} of 22 ± 2 pM, which is within error of the literature value of 24.0 ± 0.3 pM[27]. This demonstrates the usefulness of our technique in improving the accuracy of confocal single-molecule biosensors and, more broadly, in the determination of thermodynamic properties of biomolecules at the single-molecule level.

3.3 Conclusion

Confocal single-molecule fluorescence has been used to quantify diffusing molecules in diverse applications, from profiling molecular aggregation[1], to determining transcription factor concentrations in biosensor assays[3]. In this chapter, we identified and derived simple, analytical correction factors for four common causes of molecular counting artefacts: diffusion-based confocal volume encounter rates, recrossing, detection inefficiency, and multi-molecule events.

Using these correction factors, we were able to translate the number of bursts detected in numerical simulations to within two-fold of the true concentration, over a wide range of common experimental parameters (Fig. 6). This method out-performed a simpler calibration-based approach, which produced results up to an order of magnitude away from the true concentration over this same parameter range. Our approach also allowed us to work at

concentrations an order of magnitude higher than those at which multi-molecule bursts begin to occur, and at which 90% of all bursts are absorbed into multi-molecule events (Fig. 5), improving the robustness of our single-molecule counting approach.

We applied our correction-factor method to characterize the binding of a transcription factor, the *E. coli* catabolite activator protein (CAP), to a DNA biosensor. In a previous publication[3], our group measured the concentration of CAP-bound and freely-diffusing binding sites by counting the number of bursts detected in an ALEX assay, and assuming each species was detected with equal efficiency; under these assumptions, the dissociation constant of CAP for the DNA was calculated to be 8 ± 2 pM, three-fold lower than the literature value of 24 ± 0.3 pM[27]. Using our correction-factor approach, we found that the CAP-bound DNA was detected 34% more efficiently than free DNA. Correcting for this discrepancy, and re-fitting for the K_D of CAP-DNA binding, yielded a dissociation constant of 22 ± 2 pM, within error of the literature value. Though a modest improvement in accuracy (~3-fold), similarly-sized differences in TF binding or concentration often mark the difference between healthy and diseased cells[24] – an important consideration in the design of any biosensor. This demonstrates the potential of our method to improve the accuracy of confocal single-molecule fluorescence as applied to biosensors and, more broadly, to studies of molecular thermodynamics.

We note that our method is at a preliminary stage of development; in addition to further characterizing the effects of experimental parameters such as confocal volume size and geometry on burst detection, it will be necessary to validate the work already performed with simple experimental controls. An ideal experiment would be to generate DNAs of varying lengths, labelled with fluorophores of various brightnesses (or with the laser simply set to different power levels), to experimentally test the effects of different diffusion coefficients and brightnesses on detection efficiency, and determine whether our correction

factors perform as well *in vitro* as they do in our numerical simulations. Moreover, it will be necessary to determine at what point our predictions fail to accurately determine concentration; ideally, we would be able to place a confidence interval around these predictions.

Despite these caveats, the work presented here is a significant step forward in our understanding of burst detection statistics in confocal single-molecule fluorescence, and in its use to accurately quantify diffusing fluorescent species. We expect this work, following further characterization, to be applicable to a range of questions in molecular thermodynamics, aggregation, and biosensing.

3.4 Materials and methods

Data analysis and burst search algorithm

As described in Chapter 2, we use a modified version of the burst detection algorithm previously described by Seidel and colleagues[6]. Briefly, for every photon collected, we ask whether at least N photons have arrived within time T_{detect} of that photon. This is equivalent to asking whether a given photon arrives during a period of fluorescence emission with a rate of at least N / T_{detect}; for our purposes in this chapter, and throughout the rest of this work, we set $N = 7$ and $T_{detect} = 500$ μs and, unless otherwise indicated, perform detection in the $R_{ex}R_{em}$ channel. Following detection, fluorescent bursts are analyzed to determine the proximity ratio, E^*, the donor-excitation photon count, N, the burst duration, T, and fluorophore stoichiometry, S. All data was analyzed using custom software written in MATLAB (MathWorks, USA) or C++.

Monte-Carlo simulations of diffusing molecules

Simulations were carried out using custom software written in C++. Discrete time steps of 1 μs were used, during which single molecules would diffuse, fluoresce and undergo FRET. Unless stated otherwise, the diffusion coefficient was $D = 3\times10^{8}\ \mathrm{m^2\,s^{-1}}$, and the brightness of each fluorophore was $10^{5}\,\mathrm{M^{-1}\,cm^{-1}}$. The confocal spot was modelled with an incident laser intensity of $10^{3}\,\mathrm{W cm^{-2}}$ at the centre, and decaying as a 3D Gaussian with radial width $w_o = 5\times10^{-7}\ \mathrm{m}$, and axial width $z_o = 1.5\times10^{-6}\ \mathrm{m}$. For these parameters, the brightness parameter $g \approx 5\times10^{5}$ (photons/s). Background was simulated by drawing from a Poisson distribution at each time step for each (donor or acceptor) channel.

Single-molecule fluorescence spectroscopy

Confocal single-molecule fluorescence was carried out on a home-built ALEX setup, described in detail in several previous publications[16,28]. Briefly, the microscope is equipped with two lasers: a 532 nm Nd:YAG laser (Samba; Cobolt AB, Solna, Sweden) and a 635 nm pulsed diode laser (Cube; Coherent, Santa Clara, CA). As described in Chapter 2, the two lasers are modulated with a period of 100 μs, either through a TTL pulse (635 nm laser) or using an acousto-optical modulator (AOM; Isomet, Springfield, VA). The lasers are coupled into an inverted microscope (IX71; Olympus, Tokyo, Japan) and focused into solution by an oil-immersion objective (UPlanSApo 60x, 1.35 NA, Olympus). Emitted light is separated from excitation light by a dichroic mirror, and spatially filtered through a 100 μm pinhole to reject out-of-focus light in the axial direction. Another dichroic mirror splits the donor- and acceptor-emission fluorescence, which is directed onto two avalanche photodiodes (APDs; SPCM-AQR14; Perkin Elmer, Waltham, MA) which have single-photon detection sensitivity. Photon arrival times were recorded with a PC counting board (PCI-6602; National

Instruments, Austin, TX). Data processing was performed using custom software written in MATLAB, and as described in both Chapter 2, and in previous publications [16]. For the CAP / half-site binding experiments, laser powers were 60 μW (635 nm laser) and 180 μW (532 nm laser). All experiments carried out in this chapter were performed at room temperature.

Measurement of F_B

Because reactions must be diluted to ~50 pM prior to the ALEX assay, we assume that half-sites with K_{D1} > 1 nM dissociate completely (for our CAP experiments, K_{D1} = 300 nM). The observed R-only bursts, N_A, are therefore proportional to $[H_1]_{eq} + [H_1H_2]_{eq}$, while the coincident bursts, N_C, are proportional to $[CAP \cdot H_1H_2]_{eq}$. This also requires that $CAP \cdot H_1H_2$ does not appreciably dissociate during the experiment, which is reasonable under our conditions: given the slow off-rate (10^{-4} s^{-1}) of the CAP-DNA complex[16], F_B will decrease by no more than ~10% due to dissociation of the $CAP \cdot H_1H_2$ complex during a 15-minute ALEX experiment.

Although the bound fraction in an ideal experiment would be $F_B = N_C / (N_C + N_A)$, in practice, our equation for F_B is:

$$F_B = \frac{N_C}{N_C + N_A - c} \qquad [20]$$

Parameter *c* is a constant describing the average number of "spurious" acceptor-only bursts per experiment. These occur as a result of donor photobleaching of $CAP \cdot H_1H_2$, from improperly annealed half-sites, and other species incapable of participating in TF binding. Parameter *c* can be obtained easily, as it equals N_A when the total TF is in excess of half-site (where all H_1 should, therefore, be bound in $CAP \cdot H_1H_2$ complexes).

Preparation of labelled DNA

Fluorescently labelled DNAs were prepared as described previously[13]. Briefly, 5 nmol of amino-modified, single-stranded DNA (IBA, Germany) were chloroform extracted twice, ethanol precipitated, and resuspended in 95 µL 50 mM sodium phosphate buffer, pH 7.2. 100 nmol of NHS-ester-modified fluorescent dye was resuspended in 5 µL DMSO, added to the DNA, and incubated with shaking for 3 hours. Labelled DNA was ethanol precipitated and purified on a reverse-phase C18 FPLC column (µRPC C2/C18, GE Healthcare, UK), and stored in 1mM EDTA. Annealing of two complementary strands, where necessary, was carried out in 500 mM NaCl, 50 mM Tris-HCl pH 8.0, 1mM EDTA. We note that this method typically produces ≥ 95% labelled DNA (data not shown); we do not typically measure the extent of labelling, however, as the ALEX spectroscopy method used to identify fluorescent bursts separates doubly- from singly- and un-labelled fluorophores analytically. In addition, we explicitly account for unlabelled and other single-fluorophore species as explained in "Measurement of F_B."

Purification of CAP protein

Purification was carried out with Robert Crawford according to the method of Kapanidis *et al.* [29]. Briefly, a BL21 strain of *E. coli* carrying the pAK-CRP plasmid was grown to an OD_{600} of 0.7 in 500 mL LB and induced with 1 mM IPTG for 3 hours. Cells were pelleted, resuspended in 10 mL binding buffer (500 mM NaCl, 50 mM Tris-HCl pH 8.0) and sonicated for 10 minutes. This was spun at 12,000g to pellet debris, and loaded onto a pre-equilibrated NiNTA column. Successive loading and washing steps were carried out with binding buffer and 5 mM imidazole. Elution was performed in binding buffer with 200 mM imidazole. Purified protein was buffer-exchanged with a PD-10 column to remove imidazole and add 10% glycerol for storage at -20°C.

Contributions

Experiments on the CAP biosensor were conducted together with Robert Crawford. Ideas and critical feedback were provided by Seamus J. Holden and Achillefs N. Kapanidis.

References

[1] Orte A, Birkett NR, Clarke RW, Devlin GL, Dobson CM, et al. (2008) Direct characterization of amyloidogenic oligomers by single-molecule fluorescence. *Proc Natl Acad Sci U S A* 105: 14424-14429

[2] Nie SM, Chiu DT, Zare RN (1995) Real-time detection of single molecules in solution by confocal fluorescence microscopy. *Anal Chem* 67: 2849-2857

[3] Lymperopoulos K, Crawford R, Torella JP, Heilemann M, Hwang LC, et al. (2010) Single-Molecule DNA Biosensors for Protein and Ligand Detection. *Angew Chem Int Ed Engl* 49: 1316-1320

[4] Neely LA, Patel S, Garver J, Gallo M, Hackett M, et al. (2006) A single-molecule method for the quantitation of microRNA gene expression. *Nat Methods* 3: 41-46

[5] Enderlein J, Robbins DL, Ambrose WP, Keller RA (1998) Molecular shot noise, burst size distribution, and single-molecule detection in fluid flow: Effects of multiple occupancy. *J Phys Chem A* 102: 6089-6094

[6] Fries JR, Brand L, Eggeling C, Kollner M, Seidel CAM (1998) Quantitative identification of different single molecules by selective time-resolved confocal fluorescence spectroscopy. *J Phys Chem A* 102: 6601-6613

[7] Hirschfeld T (1976) Quantum efficiency independence of the time integrated emission from a fluorescent molecule. *Appl Opt* 15: 3135-3139

[8] Hirschfeld T, Block MJ, Mueller W (1977) Virometer - optical instrument for visual observation, measurement and classification of free viruses. *J Histochem Cytochem* 25: 719-723

[9] Enderlein J, Robbins DL, Ambrose WP, Goodwin PM, Keller RA (1997) Statistics of single-molecule detection. *J Phys Chem B* 101: 3626-3632

[10] Nguyen DC, Keller RA, Jett JH, Martin JC (1987) Detection of single molecules of phycoerythrin in hydrodynamically focused flows by laser-induced fluorescence. *Anal Chem* 59: 2158-2161

[11] Santoso Y, Joyce CM, Potapova O, Le Reste L, Hohlbein J, et al. (2010) Conformational transitions in DNA polymerase I revealed by single-molecule FRET. *Proc Natl Acad Sci U S A* 107: 715-720

[12] Schwille P, Haustein E (2001) Fluorescence correlation spectroscopy: an introduction to its concepts and applications. Biophysics Textbook Online. pp. 1-33

[13] Haustein E, Schwille P (2003) Ultrasensitive investigations of biological systems by fluorescence correlation spectroscopy. *Methods* 29: 153-166

[14] Kim SA, Heinze KG, Schwille P (2007) Fluorescence correlation spectroscopy in living cells. *Nat Methods* 4: 963-973

[15] Laurence TA, Kwon Y, Yin E, Hollars CW, Camarero JA, et al. (2007) Correlation spectroscopy of minor fluorescent species: Signal purification and distribution analysis. *Biophys J* 92: 2184-2198

[16] Kapanidis AN, Lee NK, Laurence TA, Doose S, Margeat E, et al. (2004) Fluorescence-aided molecule sorting: Analysis of structure and interactions by alternating-laser excitation of single molecules. *Proc Natl Acad Sci U S A* 101: 8936-8941

[17] Eigen M, Rigler R (1994) Sorting single molecules - application to diagnostics and evolutionary biotechnology. *Proc Natl Acad Sci U S A* 91: 5740-5747

[18] Berg HC (1993) Random Walks in Biology. Princeton: Princeton University Press

[19] Richter PH, Eigen M (1974) Diffusion controlled reaction rates in spheroidal geometry - application to repressor-operator association and membrane-bound enzymes. *Biophys Chem* 2: 255-263

[20] Peleg G, Ghanouni P, Kobilka BK, Zare RN (2001) Single-molecule spectroscopy of the beta(2) adrenergic receptor: Observation of conformational substates in a membrane protein. *Proc Natl Acad Sci U S A* 98: 8469-8474

[21] Szabo A, Schulten K, Schulten Z (1980) First passage time approach to diffusion controlled reactions. *J Chem Phys* 72: 4350-4357

[22] Ko DS, Sauer M, Nord S, Muller R, Wolfrum J (1997) Determination of the diffusion coefficient of dye in solution at single molecule level. *Chem Phys Lett* 269: 54-58

[23] Latchman DS (1997) Transcription factors: An overview. *Int J Biochem Cell Biol* 29: 1305-1312

[24] Engelkamp D, vanHeyningen V (1996) Transcription factors in disease. *Curr Opin Genet Dev* 6: 334-342

[25] Vogelstein B, Lane D, Levine AJ (2000) Surfing the p53 network. *Nature* 408: 307-310

[26] Heyduk T, Heyduk E (2002) Molecular beacons for detecting DNA binding proteins. *Nat Biotechnol* 20: 171-176

[27] Ebright RH, Ebright YW, Gunasekera A (1989) Consensus DNA site for the Escherichia coli catabolite gene activator protein (CAP) - CAP exhibits a 450-fold higher affinity for the consensus DNA site than for the Escherichia coli Lac DNA site. *Nucleic Acids Res* 17: 10295-10305

[28] Lee NK, Kapanidis AN, Wang Y, Michalet X, Mukhopadhyay J, et al. (2005) Accurate FRET measurements within single diffusing biomolecules using alternating-laser excitation. *Biophys J* 88: 2939-2953

[29] Kapanidis AN, Ebright YW, Ludescher RD, Chan S, Ebright RH (2001) Mean DNA bend angle and distribution of DNA bend angles in the CAP-DNA complex in solution. *J Mol Biol* 312: 453-468

4. Characterizing single-molecule heterogeneity with Probability Distribution Analysis

In Chapter 2, we discussed the use of single-molecule FRET histograms to probe biomolecular structure. Whereas most smFRET studies focus on changes in mean FRET in response to some stimulus, an often-ignored source of information about molecular structure is the widths and shapes of FRET histograms, which can report on biologically relevant structural heterogeneity. Recently-developed methods like Probability Distribution Analysis (PDA) have been used to fit FRET histograms assuming sums of static components (i.e. several species with different, but static, E^* values); however, methods to fit for dynamic sources of FRET histogram broadening have not been developed. We generalized the PDA method to predict the shapes of FRET histograms for molecules interconverting dynamically between multiple states. We validated this method by testing it against a series of two-state kinetic systems, including both Monte Carlo simulations and dynamic DNA hairpins. Fitting of the PDA-predicted FRET histogram to experimental data allowed us to recover the timescale of conformational fluctuations in an experimental DNA hairpin, in good agreement with timescales identified in previous correlation-based studies. We also applied this method

to study the "fingers-closing" transition in the Klenow Fragment (KF) of *E. coli* DNA polymerase I, a conformational change important to the fidelity of DNA synthesis in *E. coli*. We confirmed the consistency of a previously-proposed, two-state kinetic model with the observed FRET histogram of KF, while extracting a dynamic timescale in good agreement with rates measured in a previous study, and using different techniques. We expect this method to be useful in extracting rates from processes exhibiting dynamic FRET, and in hypothesis-testing models of conformational dynamics against experimental data.

4.1 Introduction

Conformational dynamics underlie many biomolecular functions, such as the complex catalytic mechanisms of DNA and RNA polymerases, and the folding of proteins and nucleic acids[1-4]. While these dynamics can give rise to changes in the width and shape of a single-molecule FRET histogram[5-8], most experiments rely only on changes in the mean value of the histogram[1,9-11], which are visually obvious and generally simpler to interpret. Early efforts to interpret the widths of FRET histograms were limited to establishing lower bounds on histogram width, either through computational methods[12-15] or well-characterized static controls[16]; however, none of these methods provided a genuine prediction for the expected width or shape of a given FRET distribution.

Recent work, primarily by Seidel and colleagues[5,8,17-20], laid the groundwork for interpreting the widths and shapes of FRET histograms. These approaches, which we will collectively call Probability Distribution Analysis (PDA) methods, are based on the premise that shot-noise-limited FRET histograms – the widths of which are determined solely by photon statistics – can be recapitulated using only the experimental distribution of photon counts. In the case of an experimental sample with a single, static underlying E^* value, the shot-noise-limited PDA prediction will match the experimental histogram; when the

experimental distribution is broadened by static or dynamic heterogeneity, however, the shot-noise limited distribution will be narrower than the experimental distribution. PDA was subsequently extended to predict the shapes of FRET histograms arising from multiple static species[19]. Until recently, however, no method had been developed to predict the shapes of FRET histograms for species interconverting dynamically between multiple FRET states.

Without such a method, smFRET experiments on diffusing molecules have been limited to time trace-[2,21,22] or correlation-based analyses[23-26] of molecular dynamics. Time trace analysis of diffusing single molecules is mainly qualitative due to the low information content available in short fluorescent bursts (with a duration of ~1 ms for an average globular protein or a short DNA fragment). Correlation-based analyses, on the other hand, provide quantitative information about the timescales of conformational fluctuations; however, they have difficulty resolving diffusion-timescale dynamics, are sensitive to optical artefacts[27], and are typically performed at the small-ensemble level, foregoing the advantages of single-molecule methods, such as the ability to analyze molecular subpopulations and imperfectly-labelled samples[28]. There is therefore a need for quantitative, robust techniques for analyzing dynamics in single-molecule FRET experiments.

Here we derive and apply an extension of the PDA method to predict the shapes of FRET histograms for molecules interconverting dynamically between distinct FRET states, for kinetic schemes of arbitrary complexity. Fitting of PDA-predicted histograms to experimental FRET histograms allows both hypothesis-testing of specific dynamic models, and the extraction of kinetic parameters. We validated this approach using Monte Carlo simulations of dynamic two-state systems, as well as experiments on a dynamic DNA hairpin. We then applied this method to study fingers-closing dynamics in the Klenow Fragment (KF) of *E. coli* DNA polymerase I, and extracted fingers-closing rates in good agreement with those obtained using correlation-based methods[29,30]; this analysis also

confirmed the consistency of the observed smFRET histogram for KF with a simple, two-state model. Together, these results demonstrate the validity and utility of the PDA approach in analyzing molecular dynamics.

4.2 Theory

A biomolecule of interest is labelled with both a donor (D) and acceptor (A) fluorophore. We use solution-phase single-molecule fluorescence to observe bursts of fluorescence from these molecules, and calculate the apparent FRET value, or proximity ratio, as the ratio of acceptor to total photons during the burst, $E^*=(B_A+F_A)/N$, where F_A is the number of photons (excluding background) detected in the acceptor-emission channel, B_A is the background count in the acceptor-emission channel, and N is the total number of photons detected in both donor- and acceptor-emission channels. As discussed in Chapter 2, the proximity ratio, unlike FRET, is affected by the presence of background, spectral cross-talk ("leakage" of donor fluorescence into the acceptor channel, and "direct-excitation" of acceptor fluorophores by the donor-excitation laser), as well as by differences in the detection efficiency and quantum yield between donor and acceptor fluorophores. While it is possible to correct for these effects to obtain accurate FRET values[31], most smFRET experiments are concerned only with relative FRET changes, the interpretation of which is not typically complicated by these artefacts. As such, in this study we focus on the proximity ratio, which is a more common measure of relative distance changes in biomolecules[5,12,32].

We build our dynamic implementation of PDA on the formalism laid out by Antonik *et al.* [17], in which the theoretical distribution of the proximity ratio, E^*, can be calculated as:

$$P(E^*) = \sum_{all\ F_D, F_A, B_D, B_A\ yielding\ E^*} P(F_D, F_A, B_D, B_A) \qquad [1]$$

which is a simple sum over the probabilities of all combinations of F_D, F_A, B_D, and B_A yielding the desired E^* value. Here, F_D and F_A are the number of photons detected in the

donor- and acceptor-emission channels (not including background photons) and B_D and B_A are the background photons arising during the burst. E^* is calculated as:

$$E^* = \frac{B_A + F_A}{B_D + B_A + F_D + F_A} = \frac{B_A + F_A}{B_D + B_A + F} = \frac{B_A + F_A}{N} \quad [2]$$

where F is the total number of fluorescence photons arising from the molecule (and not from the background), and N is the total number of photons in a burst (including background). To extend this analysis to multiple molecular states, we generalize the photon counts to include contributions from each state (i) occupied by the molecule:

$$E^* = \frac{B_A + \sum a_i}{B_D + B_A + \sum d_i + \sum a_i} = \frac{B_A + \sum a_i}{B_D + B_A + \sum f_i} = \frac{B_A + \sum a_i}{N} \quad [3]$$

where f_i is the fluorescence photon count while the molecule is in state i, and d_i and a_i are the subsets of these photons detected in the donor and acceptor channels, respectively. We can then rewrite Eq. 1 as:

$$P(E^*) = \sum_{all\left(B_A + \sum a_i\right)/N = E^*} P\left(\vec{a}, \vec{d}, B_D, B_A\right) \quad [4]$$

where $\vec{a}$ and $\vec{d}$ are vectors of photon counts a_i and d_i, for all molecular states i.

The background in single-molecule experiments on freely diffusing molecules is generally Poisson-distributed with respect to time, and independent of emission from the fluorescent particle[33]:

$$P\left(B_\gamma \mid T, r_\gamma\right) = \frac{\left(r_\gamma T\right)^{B_\gamma} e^{-\left(r_\gamma T\right)}}{B_\gamma !} \quad [5]$$

where r_γ is the rate of background in the Y channel and T is the burst duration. We can thus rewrite Eq. 4 as:

$$P(E^*) = \sum_{all\left(B_A + \sum a_i\right)/N = E^*} P\left(\vec{a}, \vec{d}\right) \cdot P\left(B_D \mid T, r_D\right) \cdot P\left(B_A \mid T, r_A\right) \cdot P(T) \quad [6]$$

where $P(T)$ is the distribution of burst durations.

Although we do not know the joint distribution $P(\vec{a},\vec{d})$ of all a_i and d_i, we can simplify the expression by assuming the total photons emitted in each state, f_i, are known. In this case, each value a_i follows a binomial distribution, with its mean set by the FRET efficiency of each state, E_i:

$$P(E^*) = \sum_{all\left(B_A+\sum a_i\right)/N=E^*} \left[\prod_{i=1}^{m}\left(P(a_i \mid f_i, E_i)\right)\right] \cdot P(\vec{f}) \times P(B_D \mid T, r_D) \cdot P(B_A \mid T, r_A) \cdot P(T) \quad [7]$$

where m is the number of states, $\vec{f}$ is a vector of all f_i, and:

$$P(a_i \mid f_i, E_i) = \binom{f_i}{a_i} (E_i)^{a_i} (1-E_i)^{f_i - a_i} \quad [8]$$

is the binomial distribution describing the probability of detecting a_i photons in state i, where f_i total fluorescence photons have been emitted.

The joint distribution of all f_i is not directly measurable, but can be expressed as a function of known distributions in one of two ways: (*i*) by making the simplifying assumption of uniform excitation intensity, or (*ii*) by incorporating the experimental distribution of arrival times per burst. The former, which we refer to as "Dynamic PDA," is a simple and computationally inexpensive method which we presented in a recent publication[7]; the latter, which we call "Arrival-Time PDA," is more computationally expensive but also more accurate, and can be used in conjunction with Burst Variance Analysis, which I present in Chapter 5 and in an upcoming publication[34]. Since both approaches are useful, we demonstrate the use of both in Figure 3; owing to its increased accuracy, however, we use Arrival-Time PDA preferentially throughout this chapter.

4.2.1 Dynamic PDA

To predict shot-noise-limited E^* distributions using Dynamic PDA, we assume a uniform illumination intensity. Under this assumption, the total number of fluorescence photons f_i arising from each state follows a Poisson distribution, with a mean proportional to the state's brightness, b_i, and to the time spent in that state, τ_i. Estimating the distribution of F by incorporating the experimental distribution of photon counts, $P(F)$ where $F=N-B_A-B_D$ [17], and assuming we know the joint distribution of all τ_i, the distribution of f_i can then be reduced to a simple multinomial:

$$P(E^*) = P(B_D \mid T, r_D) \cdot P(B_A \mid T, r_A) \cdot P(T, F) \cdot P(\vec{\tau} \mid T) \times P(\vec{f} \mid F, \vec{\tau}; \vec{b}) \sum_{all\left(B_A + \sum a_i\right)/N = E^*} \left[\prod_{i=1}^{m} \left(P(a_i \mid f_i, E_i) \right) \right] \quad [9]$$

$$P(f_i \mid F, \vec{\tau}; \vec{b}) = (F!) \left[\prod_{i=1}^{m} \left(\frac{\left(b_i \tau_i / \sum (b_i \tau_i) \right)^{f_i}}{f_i!} \right) \right] \quad [10]$$

where we replace $P(T)$ with the joint distribution $P(T,F)$, since the terms are non-independent and can be obtained jointly from the experimental data. In this work we assume that all states have equal brightness, which is reasonable given previous experimental characterization of our setup. We note that the effects of brightness on the PDA method have been treated in depth elsewhere[19].

The distribution of times spent in each state, $P(\vec{\tau}, T)$, can be determined based on the posited kinetic model and the known experimental distribution of burst durations, $P(T)$ (Fig. 1). We note that while the dwell-time distribution $P(\vec{\tau}, T)$ is analytically solvable in simple (i.e. two-state) cases[15], we use a numerical approach, where $P(\vec{\tau}, T)$ is determined using a Monte Carlo simulation of molecular transitions for the length of the burst, T; such an approach allows the modelling of kinetic schemes of arbitrary complexity.

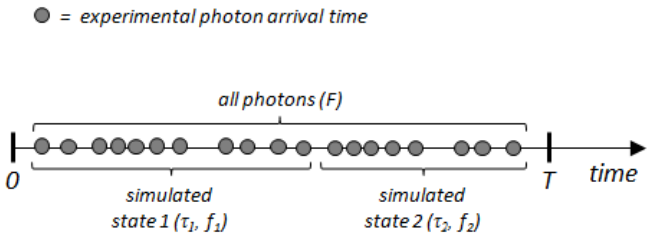


Figure 1. Schematic diagram of Arrival-Time and Dynamic PDA methods. A single burst is shown above, beginning at time 0 and ending at time T, with F photons (arrival times indicated with grey circles). To implement either the Dynamic or Arrival-Time PDA methods, we first simulate kinetic state switching using our Monte Carlo method; for the burst shown, the molecule switches stochastically from state 1 to state 2 about half-way through the burst. In the Dynamic PDA method, the total times spent in each state, τ_1 and τ_2, are recorded; assuming a uniform intensity distribution, the distribution of photons in each state, $P(\vec{f})$, is simply a multinomial with the time spent in each state (Eq. 10). In the Arrival-Time PDA method, however, we directly measure the number of experimental photon arrival times falling in each simulated state, f_1 and f_2. Unlike the original PDA method, this approach effectively draws from the arrival time distribution to calculate $P(\vec{f})$, incorporating details about the illumination geometry that are lost under the uniform-intensity assumption.

4.2.2 Arrival-Time PDA

To improve the accuracy of our method, we also developed Arrival-Time PDA, in which we eliminate the uniform-intensity assumption (or any other confocal volume geometry assumption) used in Dynamic PDA to determine $P(\vec{f})$ (Eq. 10). We did this by incorporating $P(\vec{t})$, the distribution of experimental photon arrival times within each burst, into our calculation. Instead of determining $P(\vec{f} \mid \vec{\tau})$ using a uniform-intensity assumption, in Arrival-Time PDA we determine $P(\vec{f} \mid \vec{t})$ directly.

We do this by drawing a burst from the experimental distribution $P(F,T,\vec{t})$, running our Monte-Carlo state-switching algorithm on it, and simply counting the number of experimental photon arrival times occurring within each simulated state *i* (Fig. 1). Whereas the uniform intensity assumption assumes a uniform arrival-time distribution over time, the updated method draws from the experimental arrival-time distribution, eliminating the need for any assumptions about confocal volume geometry and improving the accuracy of our PDA predictions (Fig. 3). We note that this is similar in concept to the "recolouring" method recently developed by Gopich and Szabo[35]. The predicted *E** distribution using this method is:

$$P(E^*) = \sum_{all\left(B_A+\sum a_i\right)/N=E^*} \left[\prod_{i=1}^{m} P(a_i \mid f_i, \langle E_i^* \rangle)\right] \cdot P(\vec{f} \mid \vec{t}) \times P(F,T,\vec{t}) \cdot P(B_D \mid T, r_D) \cdot P(B_A \mid T, r_A) \qquad [11]$$

We note that this approach is valid as long as all states in the experimental sample have a similar photon arrival-time distribution, $P(\vec{t})$. This is because the true state from which an experimentally observed photon arises is unknown; when our Monte Carlo approach simulates state-switching, however, photons are assigned to a state regardless of their true origin (Fig. 1). If two states have significantly different arrival time distributions (e.g. due to

very different brightnesses), these distributions would be averaged in our PDA predictions, and could produce inaccurate *E** histograms.

4.2.3 Simulating dynamic state-switching

We simulate state switching to calculate the dwell-time distribution for Dynamic PDA, $P(\vec{\tau},T)$, and the distribution of photons in each state for Arrival-Time PDA, $P(\vec{f}\,|\,\vec{t}\,)$. To simulate state switching, we consider the rates of interconversion between each pair of states, and the equilibrium occupancy of each state. For a biological molecule with *m* interconverting states, where the transition from state *i* to state *j* occurs with a single-exponential rate q_{ij}, and this rate is independent of the states previously occupied by the molecule (i.e., the molecule is 'memory-less'), the kinetic fluctuations of the molecule represent a continuous-time Markov process[36]. Defining the *m* x *m* ij^{th} rate matrix *Q* as the matrix of rate constants q_{ij}, the vector containing the equilibrium frequencies of each state, π, can be calculated as:

$$\pi Q = 0 \quad where \quad -q_{ii} = \sum_{i \neq j} q_{ij}, \ \sum \pi_i = 1 \qquad [12]$$

For any given molecule, the probability that it enters the volume while in state *i* is therefore equal to the i^{th} element of π. And, since the transition to any other state *j* is memory-less and occurs with rate constant q_{ij}, the waiting time distribution $P(t_{ij})$ for a transition into any state *j* is given by a single exponential distribution:

$$P(t_{ij}) = q_{ij} e^{-q_{ij} t} \qquad [13]$$

For a given burst of length *T*, we use Eqs. 12 and 13 to simulate switching among the different states. We sum the total time spent in each state to obtain $P(\vec{\tau}\,|\,T)$, and the total experimental photons in each state to obtain $P(\vec{f}\,|\,\vec{t}\,)$.

4.2.4 PDA with static heterogeneity

While this chapter deals primarily with E^* histograms broadened by dynamic heterogeneity, we will also consider cases in which E^* broadening occurs due to static heterogeneity: the coexistence of several species with distinct but static $\langle E^* \rangle$ values. To account for static heterogeneity, Eq. 1 can be generalized to account for k species, where the j^{th} species has concentration c_j and fluorescence intensity distribution $P_j(F)$:

$$P(E^*) = \sum_{all\ F_A, B_D, B_A\ yielding\ E^*} \sum_{j=1}^{k} \left[c_j P\left(F_A \mid \langle E \rangle, F\right) \cdot P_j(F) \right] \times P(B_D) \times P(B_A) \qquad [14]$$

In experimental applications, this equation will frequently require "brightness correction" to calculate the distributions $P_j(F)$ from the overall $P(F)$, a process described in detail in ref. [19]; in our work, we use Eq. 14 only in cases of equally bright fluorescent species, allowing us to avoid the brightness-correction step and assume that all $P_j(F) = P(F)$. Finally, we note that a very useful extension of this method is model-independent extraction of the maximum-likelihood distribution of $\langle E^* \rangle$ from a given distribution $P(E^*)$, which we do not employ in the current work, but has been described in previous publications[19].

4.2.5 Implementation

We based our implementation of PDA on the method of Nir et al.[5]. Specifically, we evaluated equation Eq. 9 (for Dynamic PDA) or Eq. 11 (for Arrival-Time PDA) using a Monte Carlo approach, and taking into account the experimental distribution of $P(T, F, \bar{t})$ directly. For simplicity, we ignored the negligible contribution of background in our measurements (typically <4 kHz, corresponding to a signal-to-noise ratio >10). For typical

confocal measurements, background <6kHz has only a marginal effect on FRET histograms[5]. If needed, background contributions can be included as suggested in ref. [18].

PDA produces a predicted E^* histogram via the following algorithm:

1. Choose an over-sampling factor K (typically K=10) and realistic initial kinetic parameters
2. Repeat K times:
3. Repeat for each burst:
    a. Draw a burst from $P(T,F,\vec{t})$, and run a Monte Carlo simulation of kinetic state switching given the kinetic parameters chosen in (1).
        i. For the "Dynamic PDA" method, determine the distribution of times, $\vec{\tau}$, spent in each state. Determine the number of photons f_i in each state by drawing from the multinomial distribution $P(f_i \mid F, \vec{\tau}, \vec{b})$ (Eq. 10).
        ii. For the "Arrival-Time PDA" method, determine $\vec{f}$ as the number of experimental photons arriving within each simulated state (Fig. 1).
    b. For each state, determine the number of photons emitted by the acceptor, a_i, by drawing from the binomial distribution $P(a_i \mid f_i, E_i)$.
    c. Add the value $\left(\sum a_i\right) / \left(\sum f_i\right)$ to the E^* histogram
4. Divide the final E^* histogram by K
5. Refine the kinetic parameters using a gradient descent method and return to step (2), unless successive parameter estimations are different by less than a given tolerance (we use a tolerance of 0.1%).

We refer to the parameters obtained at the end of step (5) to be the best-fit rates given the proposed kinetic model.

We note that in the first implementations of PDA[5,17], unexplained heterogeneity in the E^* histograms for simple double-stranded DNA molecules was modelled assuming Gaussian, quasi-static heterogeneity (standard deviation σ_r) in the distance between the two fluorophores, r. Where needed, we implement this broadening after step (3) by drawing a modified donor-acceptor distance, $r_{M,i}$, for each state i, from a Gaussian distribution with mean $\langle r \rangle = R_o \left(E_i^{-1} - 1\right)^{1/6}$ and standard deviation σ_r, where R_o is the Förster radius of the

donor-acceptor pair (Chapter 2). We then re-calculate the mean FRET of each state as $E_i = \left(1 + (r_{M,i} / R_o)^6\right)^{-1}$.

4.2.6 Fitting procedure

To calculate the goodness of fit between our PDA-predicted and experimental FRET histograms, we use the reduced chi-square statistic suggested by Kalinin et al.[18]:

$$\chi_r^2 = \frac{1}{z-y} \sum_{i=1}^{n} \frac{\left[Freq(X_i) - Freq_M(X_i) \right]^2}{Freq(X_i)} \qquad [15]$$

where y is the number of fitted parameters, z is the total number of non-zero bins, and $Freq$ and $Freq_M$ represent the frequency of data falling into bin i in either the data or prediction, respectively. For this calculation, we include only those bins containing at least one data point. A perfect fit corresponds to $\chi_r^2 = 1$, while a poor fit corresponds to $\chi_r^2 > 2$ [18].

Although we report the chi-square value as a measure of the goodness of fit between the experimental and predicted E^* histograms, for the actual fitting procedure we use a different objective function, the sum of squared error (SSE):

$$SSE = \sum_{i=all\ bins} \left[Freq(X_i) - Freq_M(X_i) \right]^2 \qquad [16]$$

SSE was chosen over χ_r^2 as the objective function because it incorporates the contribution of bins without any data points, and therefore provides a more stable minimization method in dealing with large differences between the data and the PDA prediction. To fit the data, we minimize SSE using a simple steepest-descent algorithm. To provide a measure of the uncertainty of the fitting process, we repeated the fitting procedure 3 times and reported the standard deviation value of each fitted parameter. We note that a more accurate measure of the uncertainty on each parameter value can be obtained by directly calculating the model objective function in the vicinity of the best-fit parameters obtained.

4.3 Results and discussion

4.3.1 Dynamic timescales and FRET histogram broadening

To illustrate the way in which dynamics give rise to FRET histogram broadening, we first consider a simple, two-state kinetic system (Fig. 2A-1) in which a molecule labelled with donor and acceptor fluorophores fluctuates between two FRET states, *E1* and *E2*, with rates k_{12} and k_{21}. When this molecule diffuses through the confocal volume, it produces a burst of fluorescence representing a ~1 ms "snapshot" of its dynamic FRET trajectory (Fig. 2A-2). During its confocal volume transit, the molecule may occupy state 1 alone, state 2 alone, or switch states dynamically.

In the case of very slow dynamic fluctuations, such that the mean dwell time in a given FRET state, τ_1, is much larger than the mean diffusion time through the confocal volume, τ_D, the molecule will rarely switch states during a burst. This will give rise to two shot-noise limited histograms: one for each FRET state (Fig. 2B, left). Conversely, if τ_1 is much shorter than τ_D, then the molecule will switch states many times during each burst, effectively averaging the two states and giving rise to a single shot-noise limited distribution with a mean FRET value between that of the two states (Fig. 2B, right). For τ_1 on the order of τ_D, however, the resulting FRET histogram will be too narrow to be accounted for by two distinct shot-noise limited distributions, but too broad to be accounted for by only a single shot noise-limited distribution (Fig. 2B, centre). Our goal in this work is to use PDA to accurately fit these intermediate distributions, and in doing so recover their fluctuation timescales. PDA can also be used to hypothesis-test specific dynamic models against them.

4.3.2 Comparison with smFRET simulations

To validate the PDA approach, we first performed Monte Carlo simulations of diffusing single molecules (Methods; simulation software written by Johannes Hohlbein) with two

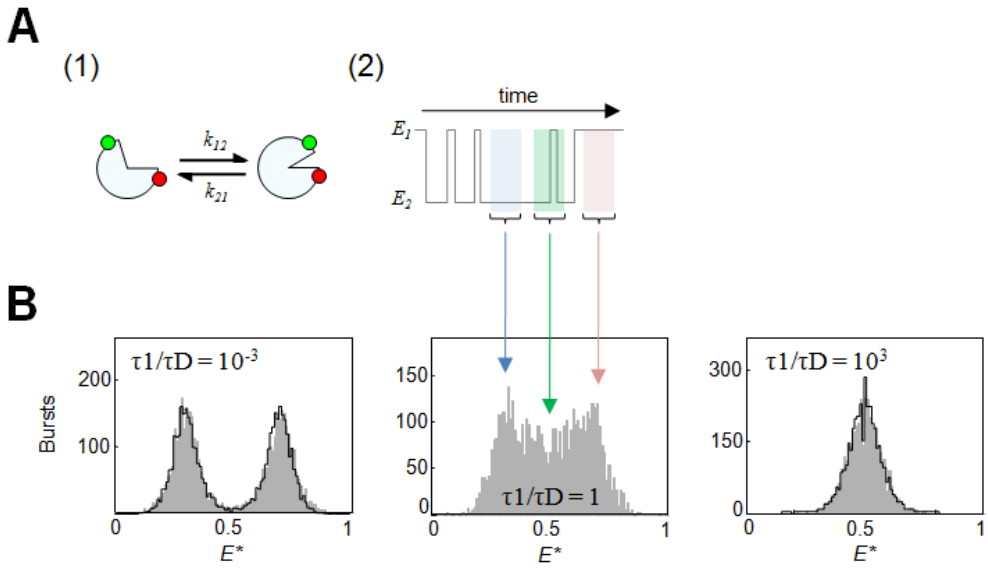


Figure 2. The effect of conformational dynamics on *E histograms.** (A) Schematic diagram of dynamic FRET state switching in a biomolecule. (1) A molecule labelled with donor and acceptor fluorophores undergoes a conformational change between FRET states *E1* and *E2* with rates k_{12} and k_{21}; sub-panel (2) shows the fluctuations in FRET over time. When the molecule diffuses through the confocal volume, "snapshots" of this FRET trajectory (coloured windows in (2)) are taken and added to the FRET histogram. In some cases, the molecule will remain in state 1 or 2 for the duration of its confocal volume transit (pink and blue windows, respectively), while in other cases, the molecule will switch states during its transit (green window), producing a burst with a FRET value intermediate between *E1* and *E2*. (B) Monte Carlo simulations of molecules with a mean diffusion time of 1 ms, two FRET states (*E** = 0.3 and 0.7), and kinetic rates $k_{12} = k_{21}$. (left) If k_{12} and k_{21} are very small, such that the mean dwell time in each state, τ_1, is much greater than the diffusion time through the confocal volume, τ_D, molecules will generally be in either one or the other FRET state during their transit through the volume. This will produce one shot-noise-limited distribution for each FRET state (left panel; black lines indicate shot noise distributions). On the other hand, for $\tau_1 << \tau_D$, switching occurs so rapidly that the two states effectively average out, yielding a single shot-noise limited peak at a FRET value intermediate between *E1* and *E2* (right panel). For $\tau_1 \approx \tau_D$, however, the resulting *E** distributions cannot be adequately fit by either 1 or 2 shot-noise limited distributions (centre panel); we use PDA fits to these distributions to recover their dynamic timescales.

FRET states, $\langle E_1 * \rangle = 0.3$, $\langle E_2 * \rangle = 0.7$, interconverting dynamically with forward and backward rate parameters $k_{1\to 2}$, $k_{2\to 1}$ (Fig. 3). Using the same parameters, we then predicted the E^* histograms using either Dynamic PDA or Arrival-Time PDA, and determined the goodness-of-fit between our predictions, and the simulation results, by calculating a χ_r^2 value.

The kinetic rates we tested included symmetric rates many orders of magnitude above or below the diffusion timescale (Fig. 3 A, B); symmetric rates on the order of the diffusion timescale (Fig. 3 C, D); and asymmetric rates on the diffusion timescale (Fig. 3 E, F). For the Arrival-Time PDA method, we found excellent agreement between our prediction and the data for all combinations of kinetic rates (all χ_r^2 close to 1). In contrast, while Dynamic PDA agreed well with the data in some cases (e.g. it performed well at extreme timescales, Fig. 3 A, B, though at these timescales the fit is essentially a static shot noise prediction), it generally did not achieve good agreement for most parameter sets ($\chi_r^2 > 2$). We note that in a recent publication[7], Dynamic PDA was fit to simulation data, and did a good job matching the shape of simulated FRET histograms; however, it did so at the expense of recovering kinetic rate constants slightly slower than their true values (~20%).

Although we have used Dynamic PDA successfully in the past[7], and it is computationally inexpensive compared to Arrival-Time PDA, we prefer Arrival-Time PDA for its greater accuracy. Moreover, Arrival-Time PDA provides certain advantages over Dynamic PDA when combined with another analytical method, Burst Variance Analysis, which we discuss in the next chapter. We therefore use Arrival-Time PDA for all subsequent analyses, both in this chapter and in Chapter 5.

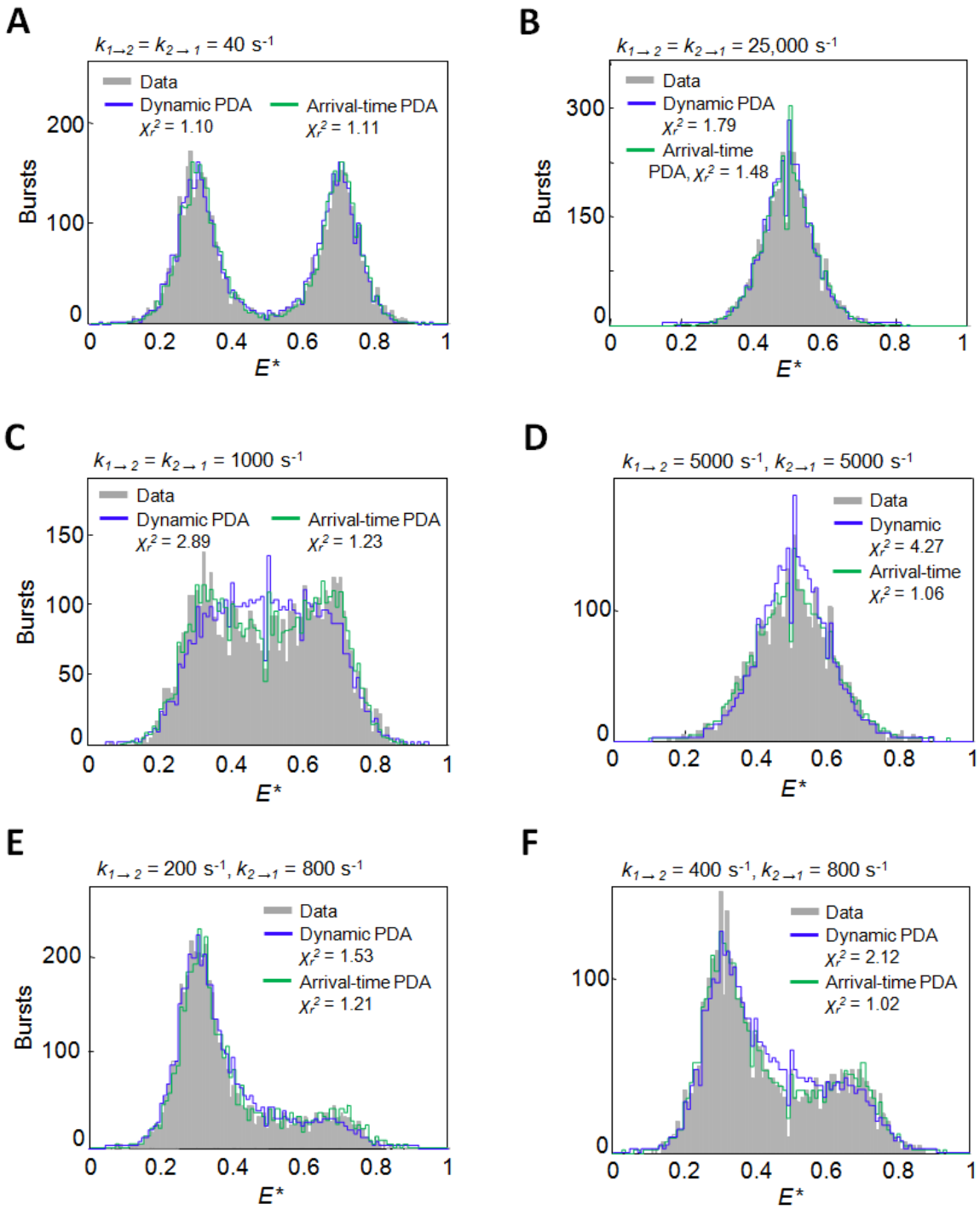

A
$k_{1\to2} = k_{2\to1}$ = 40 s^{-1}
Data
Dynamic PDA
χ_r^2 = 1.10
Arrival-time PDA
χ_r^2 = 1.11
Bursts
E*
B
$k_{1\to2} = k_{2\to1}$ = 25,000 s^{-1}
Data
Dynamic PDA
χ_r^2 = 1.79
Arrival-time
PDA, χ_r^2 = 1.48
Bursts
E*
C
$k_{1\to2} = k_{2\to1}$ = 1000 s^{-1}
Data
Dynamic PDA
χ_r^2 = 2.89
Arrival-time PDA
χ_r^2 = 1.23
Bursts
E*
D
$k_{1\to2}$ = 5000 s^{-1}, $k_{2\to1}$ = 5000 s^{-1}
Data
Dynamic
χ_r^2 = 4.27
Arrival-time
χ_r^2 = 1.06
E*
E
$k_{1\to2}$ = 200 s^{-1}, $k_{2\to1}$ = 800 s^{-1}
Data
Dynamic PDA
χ_r^2 = 1.53
Arrival-time PDA
χ_r^2 = 1.21
Bursts
E*
F
$k_{1\to2}$ = 400 s^{-1}, $k_{2\to1}$ = 800 s^{-1}
Data
Dynamic PDA
χ_r^2 = 2.12
Arrival-time PDA
χ_r^2 = 1.02
E*

Figure 3. Comparison of *E histograms for Monte Carlo simulations, and Dynamic and Arrival-Time PDA predictions.** *E** histograms are shown for a series of Monte Carlo-simulated molecules (Methods; simulation software written by Johannes Hohlbein) with two FRET states, $\langle E_1 * \rangle = 0.3$, $\langle E_2 * \rangle = 0.7$, which interconvert with rates $k_{1\to 2}$, $k_{2\to 1}$ (indicated above each panel). Dynamic PDA (blue) and Arrival-Time PDA (green) predictions are shown in each panel, along with a χ_r^2 value between the prediction and the data. While Arrival-Time PDA achieved a good fit in each case (χ_r^2 close to 1), Dynamic PDA generally showed poor agreement with the simulated data ($\chi_r^2 > 2$). While Dynamic PDA did perform well for rates several orders of magnitude above or below the diffusion timescale (panels A and B), these are essentially static, shot-noise limited populations, and therefore do not fully test the ability of our methods to predict the influence of dynamics on the shape of the *E** histogram.

4.3.3 Static controls: donor-only and doubly-labelled DNAs

To further validate our PDA approach, we applied it to a series of static DNA controls, as well as a dynamic DNA hairpin (sequences of all DNA constructs are shown in Appendix, Fig. S1). Our static DNA controls included a short, double-stranded DNA labelled with either a donor fluorophore alone ($lac^{Cy3B(-5)}$), or both donor and acceptor fluorophores ($lac^{Cy3B(-5),\ A647N(+15)}$), and two static hairpin controls, which mimic the open and closed conformations of the dynamic hairpin without interconverting.

We first analyzed the donor-only DNA control, $lac^{Cy3B(-5)}$, which has no FRET, but a mean proximity ratio of $\langle E^* \rangle \approx 0.15$, due to leakage of donor fluorescence into the acceptor emission channel (Fig. 4A). Leakage is unaffected by dynamic heterogeneity in the DNA itself (assuming this does not result in diffusion-timescale dynamic changes in the donor emission spectrum), but is sensitive to spatial mismatches in the detection efficiencies of the donor- and acceptor-emission channels[5,17]. A shot-noise limited donor-only control therefore rules out both dynamic donor spectral shifts and detector mismatches as causes of E^* histogram broadening. Using PDA, we were able to achieve a good fit to the E^* distribution assuming the presence of shot noise alone (Fig. 4A; $\langle E^* \rangle = 0.151 \pm 0.002$, $\chi_r^2 = 1.34$).

We then analyzed the doubly-labelled DNA construct, $lac^{Cy3B(-5),A647N(+15)}$, whose sequence was identical to that of the donor-only control (Fig. 4 B, C). For this doubly-labelled construct, the E^* distribution was clearly broader than shot noise (Fig. 4B; $\chi_r^2 = 20.52$). In previous studies, analysis of similar double-stranded DNAs has identified broadening in their E^* distributions beyond shot noise; the source of this broadening was suggested to be due to very slow (i.e. quasi-static) photophysics, or the existence of several slowly-interconverting states in which the fluorescent dye assumes different positions and/or orientations with respect to the DNA[5,17,20]. These studies therefore re-fit their data assuming Gaussian, quasi-static broadening of the donor-acceptor distance in their DNA constructs,

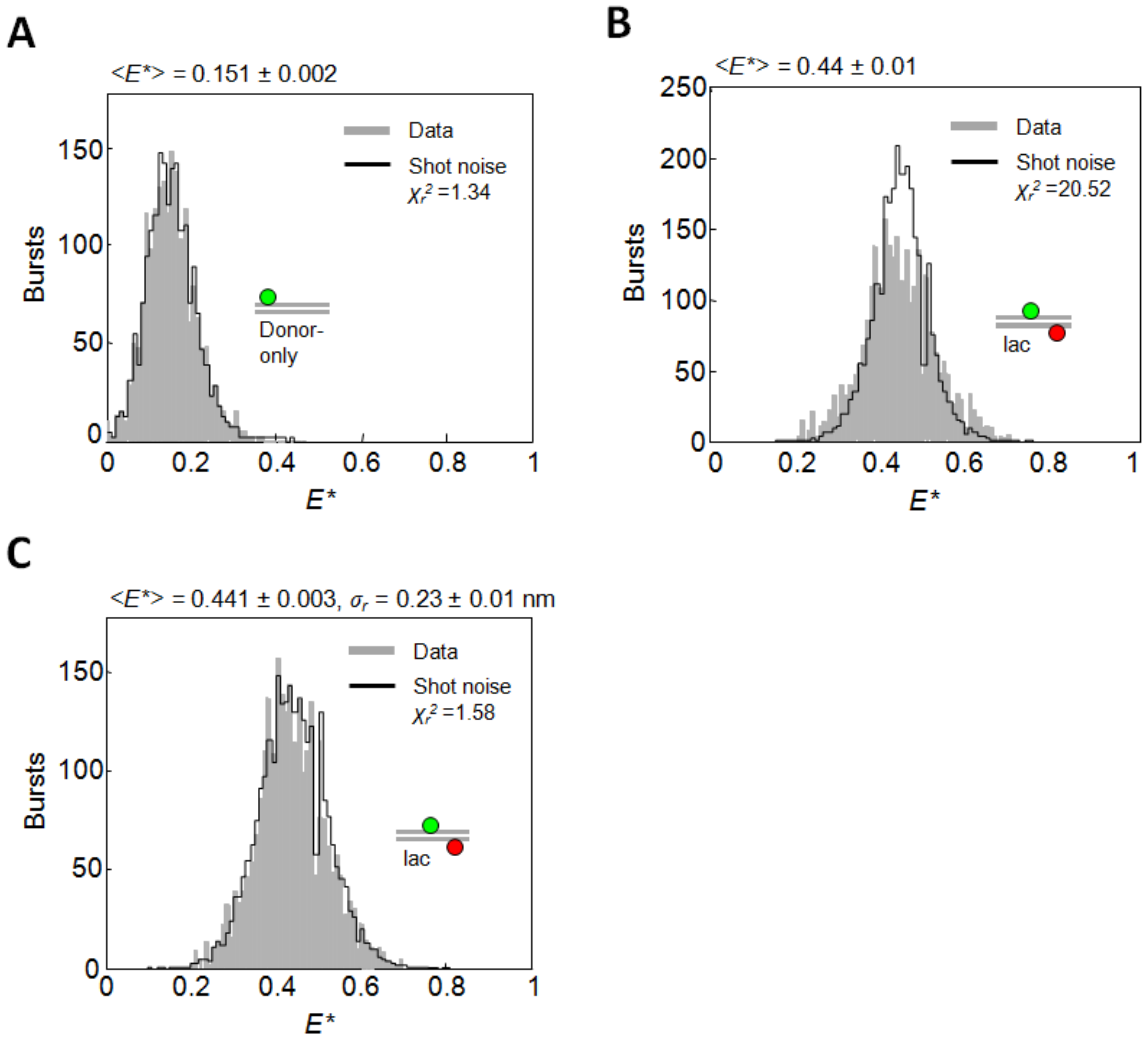


Figure 4. Comparison of *E histograms and PDA predictions for simple double-stranded DNA controls.** (A) PDA analysis of a donor-only-labelled DNA, lac$^{Cy3B(-5)}$, which does not exhibit FRET, but has $\langle E^* \rangle = 0.151$ due to leakage of the donor dye into the acceptor detection channel. PDA assuming the presence of shot noise alone (black line) was sufficient to achieve a good fit to the data ($\chi_r^2 = 1.34$). As leakage is affected only by donor spectral shifts and donor/acceptor detection volume mismatches, the goodness of the shot-noise-only fit indicates that these are not significant sources of experimental error in our experiments. (B) A donor-acceptor-labelled DNA, lac$^{Cy3B(-5),A647N(+15)}$, with the same sequence as the donor-only control. This distribution is clearly broader than shot noise alone (black line; $\chi_r^2 = 20.52$). (C) PDA of the doubly-labelled DNA, assuming a Gaussian, quasi-static distribution of donor-acceptor distances, achieved a good fit to the experimental *E** histogram ($\chi_r^2 = 1.58$) and Gaussian broadening ($\sigma_r = 0.23 \pm 0.01$ nm) similar to that identified in previous studies (0.15 – 0.23 nm[5,17]).

and obtained good fits with standard deviations in the range of 0.15 – 0.23 nm. Consistent with this, we re-fit our data with the assumption of Gaussian broadening of the donor acceptor-distance (Fig. 4C; for our Cy3B-Atto647N pair, R_o = 6.7 nm). We achieved a good fit to the data ($\chi_r^2 = 1.58$), and identified a σ_r similar to those measured in previous studies ($\sigma_r = 0.23 \pm 0.01$ nm , consistent with the 0.15 – 0.23 nm previously measured). This analysis helped us identify the baseline broadening expected for simple, double-stranded DNAs.

4.3.4 Conformational dynamics in a DNA hairpin

We next used PDA to analyze both a dynamic DNA hairpin, and two control hairpins (structures given in Appendix, Fig. S1). The dynamic hairpin is a stem-loop structure that interconverts dynamically between an "open" (melted) and a "closed" (annealed) conformation, giving rise to large structural changes on the timescale of diffusion (Fig. 5A) [21,29,30]. The stem contains a 31-nucleotide double-stranded region, while the loop contains both a single-stranded, 30-adenine sequence, and a 5-nucleotide sequence complementary to the stem that allows closing (annealing) and opening (melting) of the hairpin on a millisecond timescale. The donor and acceptor fluorophores (Cy3B, Atto647N) are placed on the stem and loop respectively, such that opening and closing are accompanied by large changes in FRET (Fig. 5A). The control hairpins are similar to the dynamic hairpin, but have been modified to remain in either the open or closed conformations permanently, without interconverting (Appendix, Fig. S1).

We first analyzed a mixture of control hairpins (Fig. 5B). While each *E** distribution was broader than shot noise alone (Appendix, Fig. S2), each could be adequately fit by assuming Gaussian, quasi-static broadening of the donor-acceptor fluorophore distance: $\langle E_{open}{}^* \rangle = 0.285 \pm 0.002$, $\sigma^r_{open} = 0.18 \pm 0.01$, $\langle E_{closed}{}^* \rangle = 0.905 \pm 0.001$ and $\sigma^r_{closed} = 0.28 \pm 0.02$ ($\chi_r^2 = 1.76$). While the open-state hairpin control showed broadening

similar to that observed in other simple, double-stranded DNAs ($\sigma^r_{open} = 0.18 \pm 0.01$, similar to the 0.15 - 0.23 nm observed previously), the closed complex control was slightly broader ($\sigma^r_{closed} = 0.28 \pm 0.02$). This could be due to a two-nucleotide gap between the double-stranded stem region and the annealed loop in the hairpin (Appendix, Fig. S1), which may allow the DNA or fluorophore to occupy a greater diversity of slowly-interconverting structures not present in the open-complex control hairpin.

Finally, we analyzed the dynamic hairpin (Fig. 5C). On the basis of previous publications characterizing similar hairpins[21,29,30,37-39], we expected our dynamic hairpin construct to switch dynamically, at the millisecond timescale, between the E^* distributions observed in our static hairpin controls. Indeed, the presence of bursts with E^* values intermediate between that of the open and closed forms of the hairpin is readily apparent in the FRET histogram (Fig. 5B). We fit this histogram by fixing $\langle E_1^* \rangle$, $\langle E_2^* \rangle$, σ^r_{closed} and σ^r_{open} to the best-fit parameters obtained from the static control hairpins (Fig. 5B), and performing Arrival-Time PDA to fit for opening and closing rate constants k_{open} and k_{close}. We achieved a good fit to the data ($\chi^2_r = 1.53$), and extracted best-fit hairpin opening and closing rates of $k_{open} = 641 \pm 9\,\mathrm{s}^{-1}$, $k_{close} = 463 \pm 14\,\mathrm{s}^{-1}$, which gave a total reaction time $\tau_R = 1/(k_{open}+k_{close}) =$ 0.91 ms. This result is in good agreement with the 0.5 - 1 ms reaction times measured in previous studies on the same hairpin, but using correlation-based analyses[21,29].

An important feature of PDA is its ability not just to extract rates, but to hypothesis-test the consistency of the observed FRET distribution with a particular kinetic model. While our fit was good, suggesting the general consistency of our hairpin's dynamics with a simple, two-state model, the predicted distribution also shows signs of systematic deviation from the data (Fig. 5C). This could be due to the existence of transient hairpin folding intermediates,

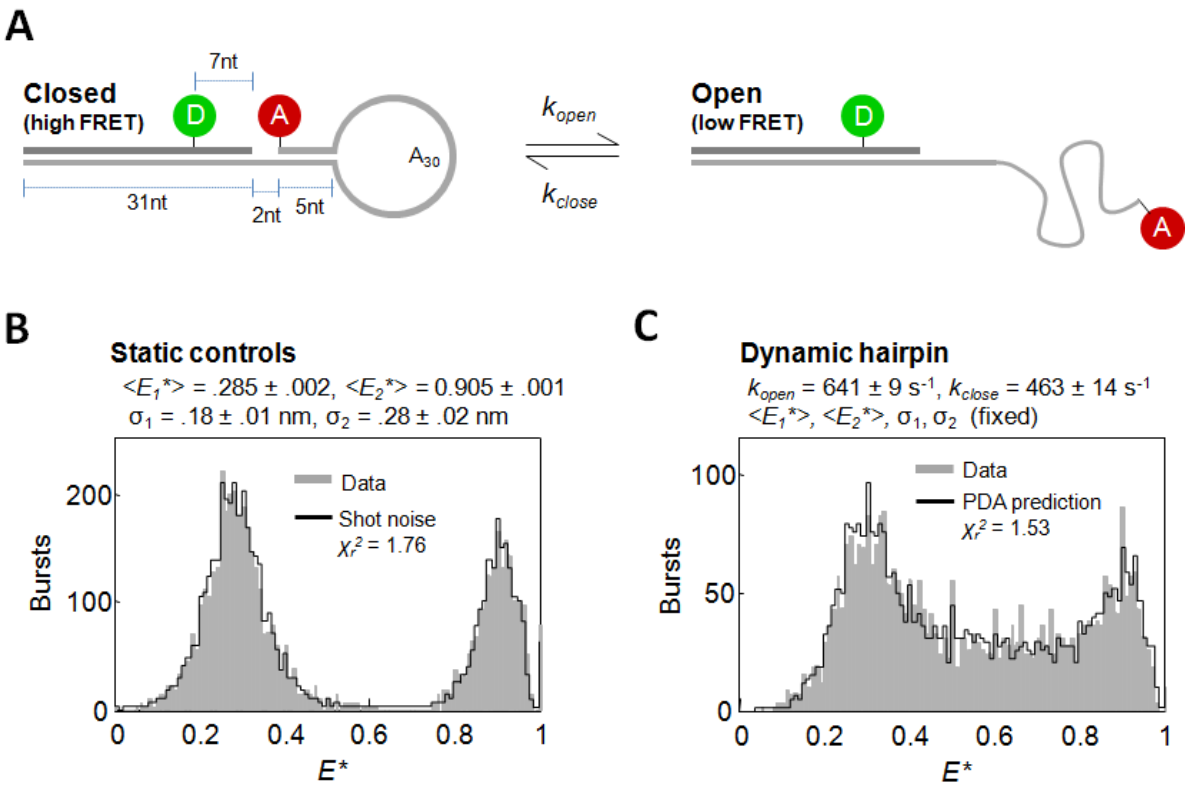

A
Closed
(high FRET)
7nt
D
A
A30
31nt
2nt
5nt
kopen
kclose
Open
(low FRET)
D
A
B
Static controls
<E1*> = .285 ± .002, <E2*> = 0.905 ± .001
σ1 = .18 ± .01 nm, σ2 = .28 ± .02 nm
Data
Shot noise
χr2 = 1.76
Bursts
200
100
0
0
0.2
0.4
0.6
0.8
1
E*
C
Dynamic hairpin
kopen = 641 ± 9 s-1, kclose = 463 ± 14 s-1
<E1*>, <E2*>, σ1, σ2 (fixed)
Data
PDA prediction
χr2 = 1.53
Bursts
100
50
0
0
0.2
0.4
0.6
0.8
1
E*

Figure 5. PDA analysis of static and dynamic hairpin DNAs. (A) Structure and kinetics of a DNA hairpin. The hairpin is a stem-loop structure, where the stem contains a 31-nt double-stranded region, and the loop contains a stretch of 30 adenines followed by a 5-nt region complementary to a single-stranded portion of the stem. The stem and loop are labelled with donor and acceptor fluorophores respectively (green and red circles). In its "closed" conformation (left; 5-nt region is annealed), the hairpin exhibits high FRET; in its "open" conformation (right; 5-nt region is melted), it exhibits low FRET. We model the transition between these two states as a simple two-state kinetic system, with first-order kinetic rates k_{open} and k_{close}. (B) FRET histogram and PDA fit for a mixture of two static hairpin controls, which mimic the open and closed conformations of the dynamic hairpin without interconverting; structures are given in the Appendix (Fig. S1). Each distribution was broader than expected from shot noise alone, but could be adequately fit (black line) by assuming Gaussian quasi-static broadening of the distance between the donor and acceptor fluorophores: $\langle E_{open}^{*} \rangle = 0.285 \pm 0.002$, $\sigma_{open}^{r} = 0.18 \pm 0.01$, $\langle E_{closed}^{*} \rangle = 0.905 \pm 0.001$ and $\sigma_{closed}^{r} = 0.28 \pm 0.02$ ($\chi_r^2 = 1.76$). (C) FRET histogram and PDA fit for the dynamic DNA hairpin. Using Arrival-Time PDA, and fixing $\langle E_{open}^{*} \rangle$, $\langle E_{closed}^{*} \rangle$, σ_{closed}^{r} and σ_{open}^{r} to the best-fit parameters from the static hairpin controls (panel B), we fit the FRET histogram for a dynamic, two-state model. We recovered best-fit first-order kinetic rates of $k_{open} = 641 \pm 9\,\mathrm{s}^{-1}$, $k_{close} = 463 \pm 14\,\mathrm{s}^{-1}$, which yielded a good fit to the data ($\chi_r^2 = 1.53$).

which are just at the detection limit of our technique; indeed, recent work by Van Orden and colleagues on a similar hairpin identified double-exponential hairpin-closing dynamics, suggesting the existence of an improperly-annealed hairpin folding intermediate[38,39]; this or similar behaviours could give rise to the small deviation we observe.

4.3.5 Conformational dynamics in the Klenow Fragment of *E. coli* DNA polymerase I

The bacterial DNA Polymerase I is an essential component of DNA replication and repair; for its role in these crucial genetic processes, the polymerase relies on a remarkable fidelity in selecting the correct template-directed nucleotide for addition to a growing DNA chain[40]. Substantial effort has been invested in studies of how this fidelity is achieved, with several reports pointing to conformational changes in the polymerase that precede successful nucleotide incorporation. In particular, much attention has been focused on the noncovalent "fingers-closing" transition that precedes nucleotide incorporation, in which the polymerase forms a tight pocket around both its DNA substrate and an incoming nucleotide, positioning them for catalysis[41] (Fig. 6A).

Previously, we used single-molecule FRET to monitor the fingers-closing transition in the Klenow Fragment (KF) of *Escherichia coli* DNA Polymerase I, by labelling it with donor and acceptor fluorophores on its "fingers" and "thumb" subdomains, respectively[2]. We showed that fingers-closing and -opening occurs dynamically in the absence of a DNA substrate; this was based on the observation that E^* distributions of the unliganded KF were too wide to be accounted for by either one or two shot-noise-limited distributions, and that the unliganded KF exhibited millisecond-timescale fluctuations via FCS-based methods[2]. Such dynamics were not anticipated on the basis of available crystal structures[42], but may play an important role in the rapid rejection of incorrect nucleotide substrates encountered by

the polymerase *in vivo*[2]. The timescale and potential importance of these conformational fluctuations made KF an attractive model system for our Arrival-Time PDA methodology.

To study fingers-closing conformational dynamics in KF, we analyzed smFRET data from KF site-specifically labelled with two fluorophores: Cy3B as the FRET donor, and Atto647N as the FRET acceptor (Methods; all KF experiments were performed by Yusdi Santoso). The placement of the fluorophores allowed the study of fingers-closing conformational changes through a large change in FRET (from E*≈0.5 in the open state to E*≈0.71 in the closed state).

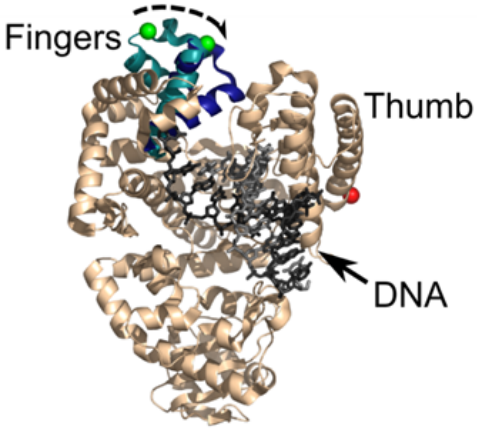
We first used PDA to test whether the observed E^* distribution of KF could be accounted for with a simple, dynamic two-state model (Fig. 6B). To reduce the number of free parameters, we fixed the mean E^* and Gaussian broadening terms for the open and closed states ($\langle E_{open}{}^* \rangle = 0.5$, $\langle E_{open}{}^* \rangle = 0.71$) using fitting results from the Pol-DNA binary complex (which favours the open state) and the Pol-DNA-dNTP ternary complex (which favours the closed state)[7]. We then fit the E^* histogram for first-order rate constants k_{close} and k_{open}. We obtained best-fit rates of k_{open} = 283 ± 8 s^{-1} and k_{close} = 217 ± 8 s^{-1}, in good agreement with the ~166 s^{-1} opening/closing rates identified previously via correlation-based methods[2]. Moreover, the good agreement between the actual and PDA-predicted FRET histograms (Fig. 6B; $\chi_r^2 = 1.54$) suggests that a simple, two-state dynamic model is both sufficient and reasonable in accounting for fingers-closing dynamics, at least at the diffusion timescale, and under our experimental conditions.

We also applied PDA to model dynamics in the binary (DNA-bound) and ternary (DNA- and dNTP-bound) complexes of KF. To generate the binary complex, we incubated KF with an inextensible hairpin substrate DNA in which A is the templating base (structure of DNA is given in Appendix, Fig. S1); dTTP could then be added to generate the ternary complex (Methods). We then analyzed both complexes via smFRET and PDA (Fig. 6 C, D);

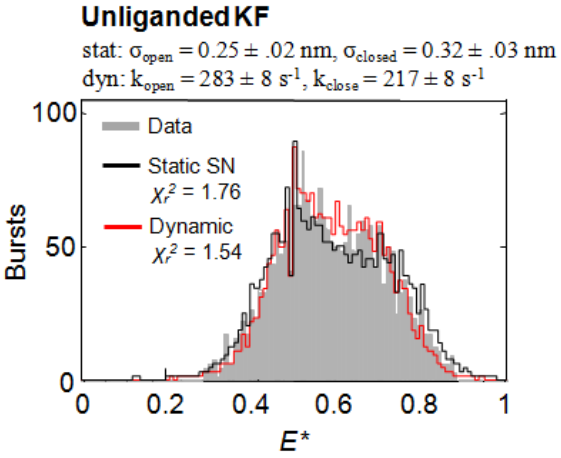
as in the unliganded polymerase, we fit each complex's FRET histogram to a two-state dynamic model (fixing mean FRET and broadening terms as mentioned previously), and extracted k_{open} and k_{close} in the 200 – 500 s^{-1} range, similar to that obtained for the unliganded polymerase. It is important to note, however, that dynamics in these complexes have only been inferred indirectly; for instance, while the closed and open conformations appear to exist in equilibrium in smFRET studies[7], and previous stopped-flow studies have measured millisecond-timescale fingers-closing rates on the transition from binary to ternary complex[43], FCS-based and other methods have failed to identify any direct evidence for dynamics in these complexes[2].

To determine whether the FRET histograms of different KF complexes were better accounted for by static or dynamic heterogeneity, we used PDA to fit a two-species static model to the FRET histograms of our unliganded, binary and ternary KF complexes, and compared these to the experimental data, and to the PDA fits for our two-state dynamic model. For the static model, we assumed the existence of two species with fixed FRET efficiencies ($\langle E_{open}^{*} \rangle = 0.5$, $\langle E_{open}^{*} \rangle = 0.71$), and fit for their Gaussian, quasi-static broadening terms, σ_{open} and σ_{closed} (Fig. 6 B,C,D). In all three cases, we obtained reasonable fits to the experimental FRET histograms, with χ_r^2 values similar to those of the dynamic two-state models. While for the unliganded polymerase we have evidence of dynamics from complementary experimental techniques, and can therefore take the rates extracted by Arrival-Time PDA as legitimate, for the binary and ternary complexes, PDA does little to resolve the question of whether they are static or dynamic. The usefulness of the rates extracted with PDA is therefore questionable. This served as a major motivation for developing Burst Variance Analysis (BVA), on which our next chapter is based. In Chapter 5, we combine BVA with PDA to distinguish between static and dynamic models of FRET histogram heterogeneity.

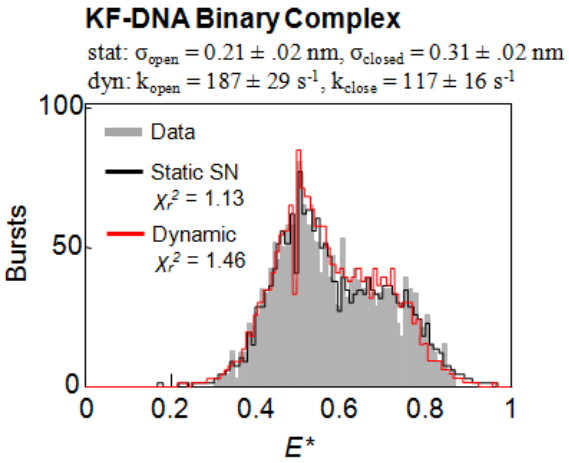
A


Fingers
Thumb
DNA

B


Unliganded KF
stat: σopen = 0.25 ± .02 nm, σclosed = 0.32 ± .03 nm
dyn: kopen = 283 ± 8 s-1, kclose = 217 ± 8 s-1
Data
Static SN
χr2 = 1.76
Dynamic
χr2 = 1.54
Bursts
E*

C


KF-DNA Binary Complex
stat: σopen = 0.21 ± .02 nm, σclosed = 0.31 ± .02 nm
dyn: kopen = 187 ± 29 s-1, kclose = 117 ± 16 s-1
Data
Static SN
χr2 = 1.13
Dynamic
χr2 = 1.46
Bursts
E*

D

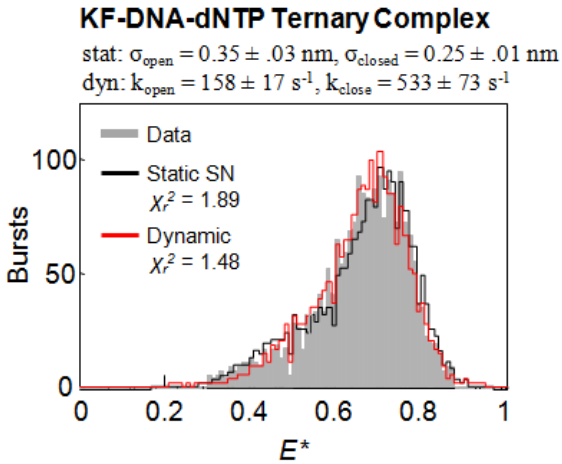

KF-DNA-dNTP Ternary Complex
stat: σopen = 0.35 ± .03 nm, σclosed = 0.25 ± .01 nm
dyn: kopen = 158 ± 17 s-1, kclose = 533 ± 73 s-1
Data
Static SN
χr2 = 1.89
Dynamic
χr2 = 1.48
Bursts
E*

Figure 6. PDA analysis of the Klenow Fragment of *E. coli* DNA Polymerase I. (A) Superposition of the crystal structures of the Pol-DNA binary complex (PDB file 1L3U) and Pol-DNA-dNTP ternary complex (PDB file 1LV5) for *B. stearothermophilus* Pol I, a close homologue of *E. coli* Pol I [46]. The α-carbon backbone of the protein is shown in beige, except for the fingers subdomain, which is shown in teal (fingers open) or dark blue (fingers closed). Green and red spheres indicate the positions of the donor and acceptor fluorophores respectively, while the arrow indicates the structural change in the fingers subdomain during a "fingers-closing" transition. (B) FRET histogram of unliganded Klenow Fragment (KF) of *E. coli* DNA Polymerase I (smFRET distributions of all KF samples were measured by Yusdi Santoso). Using PDA, we fit this histogram to either a dynamic two-state model ($\langle E_{open}^{*} \rangle$, $\langle E_{closed}^{*} \rangle$, σ_{open}^{r}, and σ_{open}^{r} are fixed, and a fit performed for k_{open} and k_{close}) or a static two-species model($\langle E_{open}^{*} \rangle$ and $\langle E_{closed}^{*} \rangle$ fixed, and a fit performed for σ_{open}^{r}, and σ_{open}^{r}). In both cases, a good fit to the data was achieved ($\chi_r^2 < 2$). We performed the same fits for the (B) KF-DNA binary complex and (C) KF-DNA-dNTP ternary complex; in all cases, both static and dynamic models of E^* heterogeneity accounted reasonably well for the experimental FRET histogram ($\chi_r^2 < 2$).

4.4 Conclusion

In this chapter, we derived and implemented a generalization of Probability Distribution Analysis (Arrival-Time PDA) allowing us to predict FRET histograms for molecules exhibiting conformational dynamics, with kinetic schemes of arbitrary complexity. We do this by incorporating the joint distribution of burst sizes, lengths and photon arrival times, $P(N,T,\vec{t})$ into our analysis, and using a Monte Carlo simulation of the proposed kinetic scheme to determine the distribution of photons emitted in each molecular state.

We validated our method using both numerical simulations of diffusing single molecules, and experimental DNA hairpins. Using numerical simulations, we confirmed the ability of our method to accurately reproduce FRET histograms for two-state dynamic systems with a wide range of kinetic rates. We then applied our method to a dynamic DNA hairpin. We were able to fit a simple, two-state dynamic model to the FRET histogram of the dynamic hairpin DNA, and extracted hairpin opening and closing rates in good agreement with those reported previously using the same hairpin, but different experimental and analytical approaches[21,29].

We then used Arrival-Time PDA to analyze "fingers-closing" conformational dynamics in the Klenow Fragment (KF) of *E. coli* DNA Polymerase I, which is important to the fidelity of DNA synthesis[41]. Using PDA, we fit a simple two-state kinetic model to the FRET histogram of unliganded KF and recovered fingers-opening and –closing rates consistent with those obtained in a previous publication, via different experimental methods[2]. Moreover, the goodness of the fit itself (χ_r^2 near 1) suggests that this simple model is sufficient to account for fingers-closing dynamics in the unliganded polymerase, at least on the diffusion timescale. We note that while this model is adequate in accounting for our data, improvements in data quality or time resolution may highlight less obvious dynamic

features of KF. In the future, more complex models may therefore be required to account for the kinetics of fingers-closing in KF.

Fitting the binary and ternary complexes of KF for a two-state dynamic model also produced closing- and opening rates in agreement with those inferred from previous experiments[43]. In these complexes, however, the presence of dynamics is not well-established, with correlation-based studies suggesting the absence of diffusion-timescale dynamics[2]. Compounding this ambiguity, we fit a static two-species model to these FRET histograms, and achieved a goodness of fit comparable to that of the dynamic two-state model. This inability to distinguish between static and dynamic models of FRET histogram heterogeneity motivates our next chapter, in which we discuss Burst Variance Analysis, a novel analytical method which enables the detection of FRET dynamics in smFRET data.

We note that in this work we assumed a Markovian process governing the transition between states in KF (i.e., one in which only the identity of the current state governs the rate of transitioning into any other given state), though non-Markovian dynamics (in which a molecule's kinetics are influenced some "memory" of its previous states) have been observed in other enzymes[44]. While memory effects were not needed to account for the E^* distributions in the present work, it is conceivable that they may be useful in future applications of PDA. As our implementation of PDA uses a Monte-Carlo approach to simulate state-switching, it should be straightforward to incorporate memory effects into PDA predictions, and thereby model non-Markovian enzyme dynamics.

Finally, we note that our PDA method is similar to, and was published concurrently with, a dynamic PDA method developed by Seidel and colleagues[8]; this is not surprising, as our PDA method is built on their original shot noise methodology[17]. Due to their distinct implementations, however, each method has advantages and disadvantages. Relative to theirs, our method has three distinct advantages: it does not require fixed-time-length bursts (and

therefore avoids errors due to time-binning), it can be implemented for arbitrarily complex dynamic schemes without the need for a complex analytical framework, and it incorporates photon arrival times, improving the accuracy of predicted FRET histograms (Fig. 3). Relative to ours, the method developed by Seidel and colleagues goes further in dealing with common sources of experimental artefacts, such as changes in the E^* histogram due to multi-molecule bursts, and differences in brightness among molecular species. As the advantages of each method are essentially orthogonal to one another, our methods can be considered complementary.

In conclusion, our results demonstrate the utility of PDA in modelling and quantifying biomolecular dynamics. Due to its diffusion-timescale sensitivity, Arrival-Time PDA naturally complements correlation-based approaches, which are typically less sensitive to diffusion-timescale dynamics[45]. Moreover, the ability of PDA to hypothesis-test specific dynamic models against FRET data should allow for more rigorous investigations of the nature of molecular dynamics than were possible using early, approximate methods[12]. We expect Arrival-Time PDA to be useful in future studies of diffusion-timescale molecular dynamics, folding and catalysis; as our method does not require prior characterization of the experimental setup, we also expect it to be useful in performing retrospective analyses of already-collected data.

4.5 Materials and methods

DNA

Fluorescently labelled DNAs were prepared as described in Chapter 3. Oligos used to construct simple double-stranded DNAs were purified on a reverse-phase C18 column (μRPC C2/C18, GE Healthcare, UK), while oligos used to generate hairpins were purified via denaturing PAGE. Where necessary, labelled, single-stranded DNA samples were annealed in hybridization buffer (50mM Tris-HCl pH 8.0, 1mM EDTA, 500mM NaCl). DNA sequences are listed in the Appendix, Fig. S1.

Klenow Fragment

Expression, purification, fluorescent labelling and characterization of the exo- Klenow fragment were carried out as previously described[2]. Briefly, Klenow fragment was labelled at positions 550 and 744 with ATTO647N and Cy3B respectively; the specificity of the labelling orientation was ≈88%. The extent of labelling was ≥ 70%.

Single-molecule experiments

Single-molecule measurements were performed on an alternating laser excitation (ALEX) microscope as described in Chapter 3, but using a water-immersion objective (UPLS-APO, 60x, 1.2 NA, Olympus, Japan). The water-immersion objective is necessary to minimize detection volume mismatch, which can result from the refractive index mismatch between an oil-immersion objective and an aqueous sample; this is important because detection volume mismatches can produce false dynamic signals in PDA (Figure 4A and accompanying text). The excitation powers were 180 μW and 60 μW for the 532 nm and 635 nm lasers respectively, and samples were analyzed at a concentration of 50-100 pM. All DNA samples were measured in 400 mM NaCl, 1 mM EDTA, 100 μg/ml BSA, and 5% (v/v) glycerol. KF

samples were measured in 40 mM HEPES-NaOH, pH 7.3, 10 mM $MgCl_2$, 1 mM DTT, 100 μg/ml BSA, and 1 mM mercaptoethylamine. All experiments were performed at room temperature.

Simulations of smFRET experiments

Simulations were carried out as described in Chapter 3.

Data analysis

Single-molecule data (from experiments or simulations) were analyzed using custom software written in MATLAB (Natick, MA), as described in Chapter 3.

Contributions

Work on the Dynamic PDA method was completed together with Yusdi Santoso, and published with joint first authorship in *Chemphyschem*[7]; Chapter 4 therefore contains text from this manuscript. Experiments were conducted together with Yusdi. Simulation software was provided by Johannes Hohlbein. Ideas and critical feedback were provided by Seamus J. Holden, Johannes Hohlbein and Achillefs N. Kapanidis.

References

[1] Kapanidis AN, Margeat E, Ho SO, Kortkhonjia E, Weiss S, et al. (2006) Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. *Science* 314: 1144-1147

[2] Santoso Y, Joyce CM, Potapova O, Le Reste L, Hohlbein J, et al. (2010) Conformational transitions in DNA polymerase I revealed by single-molecule FRET. *Proc Natl Acad Sci U S A* 107: 715-720

[3] Schuler B, Eaton WA (2008) Protein folding studied by single-molecule FRET. *Curr Opin Struct Biol* 18: 16-26

[4] Zhao R, Rueda D (2009) RNA folding dynamics by single-molecule fluorescence resonance energy transfer. *Methods* 49: 112-117

[5] Nir E, Michalet X, Hamadani KM, Laurence TA, Neuhauser D, et al. (2006) Shot-noise limited single-molecule FRET histograms: Comparison between theory and experiments. *J Phys Chem B* 110: 22103-22124

[6] Watkins LP, Chang HY, Yang H (2006) Quantitative single-molecule conformational distributions: A case study with poly-(L-proline). *J Phys Chem A* 110: 5191-5203

[7] Santoso Y, Torella JP, Kapanidis AN (2010) Characterizing Single-Molecule FRET Dynamics with Probability Distribution Analysis. *Chemphyschem* 11: 2209-2219

[8] Kalinin S, Valeri A, Antonik M, Felekyan S, Seidel CAM (2010) Detection of Structural Dynamics by FRET: A Photon Distribution and Fluorescence Lifetime Analysis of Systems with Multiple States. *J Phys Chem B* 114: 7983-7995

[9] Liu SX, Abbondanzieri EA, Rausch JW, Le Grice SFJ, Zhuang XW (2008) Slide into Action: Dynamic Shuttling of HIV Reverse Transcriptase on Nucleic Acid Substrates. *Science* 322: 1092-1097

[10] Margittai M, Widengren J, Schweinberger E, Schroder GF, Felekyan S, et al. (2003) Single-molecule fluorescence resonance energy transfer reveals a dynamic equilibrium between closed and open conformations of syntaxin 1. *Proc Natl Acad Sci U S A* 100: 15516-15521

[11] Weiss S (1999) Fluorescence spectroscopy of single biomolecules. *Science* 283: 1676-1683

[12] Dahan M, Deniz AA, Ha TJ, Chemla DS, Schultz PG, et al. (1999) Ratiometric measurement and identification of single diffusing molecules. *Chemical Physics* 247: 85-106

[13] Deniz AA, Dahan M, Grunwell JR, Ha TJ, Faulhaber AE, et al. (1999) Single-pair fluorescence resonance energy transfer on freely diffusing molecules: Observation of Forster distance dependence and subpopulations. *Proc Natl Acad Sci U S A* 96: 3670-3675

[14] Ying LM, Wallace MI, Balasubramanian S, Klenerman D (2000) Ratiometric analysis of single-molecule fluorescence resonance energy transfer using logical combinations of threshold criteria: A study of 12-mer DNA. *J Phys Chem B* 104: 5171-5178

[15] Gopich IV, Szabo A (2007) Single-molecule FRET with diffusion and conformational dynamics. *J Phys Chem B* 111: 12925-12932

[16] Coban O, Lamb DC, Zaychikov E, Heumann H, Nienhaus GU (2006) Conformational heterogeneity in RNA polymerase observed by single-pair FRET microscopy. *Biophys J* 90: 4605-4617

[17] Antonik M, Felekyan S, Gaiduk A, Seidel CAM (2006) Separating structural heterogeneities from stochastic variations in fluorescence resonance energy transfer distributions via photon distribution analysis. *J Phys Chem B* 110: 6970-6978

[18] Kalinin S, Felekyan S, Antonik M, Seidel CAM (2007) Probability distribution analysis of single-molecule fluorescence anisotropy and resonance energy transfer. *J Phys Chem B* 111: 10253-10262

[19] Kalinin S, Felekyan S, Valeri A, Seidel CAM (2008) Characterizing multiple molecular states in single-molecule multiparameter fluorescence detection by probability distribution analysis. *J Phys Chem B* 112: 8361-8374

[20] Kalinin S, Sisamakis E, Magennis SW, Felekyan S, Seidel CAM (2010) On the Origin of Broadening of Single-Molecule FRET Efficiency Distributions beyond Shot Noise Limits. *J Phys Chem B* 114: 6197-6206

[21] Santoso Y, Kapanidis AN (2009) Probing Biomolecular Structures and Dynamics of Single Molecules Using In-gel Alternating-Laser Excitation. *Anal Chem* 81: 9561-9570

[22] Brasselet S, Peterman EJG, Miyawaki A, Moerner WE (2000) Single-molecule fluorescence resonant energy transfer in calcium concentration dependent cameleon. *J Phys Chem B* 104: 3676-3682

[23] Li HT, Ren XJ, Ying LM, Balasubramanian S, Klenerman D (2004) Measuring single-molecule nucleic acid dynamics in solution by two-color filtered ratiometric fluorescence correlation spectroscopy. *Proc Natl Acad Sci U S A* 101: 14425-14430

[24] Laurence TA, Kwon Y, Yin E, Hollars CW, Camarero JA, et al. (2007) Correlation spectroscopy of minor fluorescent species: Signal purification and distribution analysis. *Biophys J* 92: 2184-2198

[25] Torres T, Levitus M (2007) Measuring conformational dynamics: A new FCS-FRET approach. *J Phys Chem B* 111: 7392-7400

[26] Nie SM, Chiu DT, Zare RN (1995) Real-time detection of single molecules in solution by confocal fluorescence microscopy. *Anal Chem* 67: 2849-2857

[27] Hess ST, Webb WW (2002) Focal volume optics and experimental artifacts in confocal fluorescence correlation spectroscopy. *Biophys J* 83: 2300-2317

[28] Kapanidis AN, Lee NK, Laurence TA, Doose S, Margeat E, et al. (2004) Fluorescence-aided molecule sorting: Analysis of structure and interactions by alternating-laser excitation of single molecules. *Proc Natl Acad Sci U S A* 101: 8936-8941

[29] Wallace MI, Ying LM, Balasubramanian S, Klenerman D (2000) FRET fluctuation spectroscopy: Exploring the conformational dynamics of a DNA hairpin loop. *J Phys Chem B* 104: 11551-11555

[30] Ying LM, Wallace MI, Klenerman D (2001) Two-state model of conformational fluctuation in a DNA hairpin-loop. *Chem Phys Lett* 334: 145-150

[31] Lee NK, Kapanidis AN, Wang Y, Michalet X, Mukhopadhyay J, et al. (2005) Accurate FRET measurements within single diffusing biomolecules using alternating-laser excitation. *Biophys J* 88: 2939-2953

[32] Kapanidis AN, Laurence TA, Lee NK, Margeat E, Kong XX, et al. (2005) Alternating-laser excitation of single molecules. *Accounts of Chemical Research* 38: 523-533

[33] Fries JR, Brand L, Eggeling C, Kollner M, Seidel CAM (1998) Quantitative identification of different single molecules by selective time-resolved confocal fluorescence spectroscopy. *J Phys Chem A* 102: 6601-6613

[34] Torella JP, Holden SJ, Santoso Y, Hohlbein J, Kapanidis AN (2010) Identifying molecular dynamics in single-molecule FRET experiments with Burst Variance Analysis. *Biophys J* (accepted):

[35] Gopich IV, Szabo A (2009) Decoding the Pattern of Photon Colors in Single-Molecule FRET. *J Phys Chem B* 113: 10965-10973

[36] McKinney SA, Joo C, Ha T (2006) Analysis of single-molecule FRET trajectories using hidden Markov modeling. *Biophys J* 91: 1941-1951

[37] Bonnet G, Krichevsky O, Libchaber A (1998) Kinetics of conformational fluctuations in DNA hairpin-loops. *Proc Natl Acad Sci U S A* 95: 8602-8606

[38] Jung JY, Van Orden A (2006) A three-state mechanism for DNA hairpin folding characterized by multiparameter fluorescence fluctuation spectroscopy. *J Am Chem Soc* 128: 1240-1249

[39] Van Orden A, Jung J (2008) Fluorescence correlation spectroscopy for probing the kinetics and mechanisms of DNA hairpin formation. *Biopolymers* 89: 1-16

[40] Joyce CM (2004) Encyclopedia of Biological Chemistry. Elsevier. pp. 720-725

[41] Joyce CM, Benkovic SJ (2004) DNA polymerase fidelity: Kinetics, structure, and checkpoints. *Biochemistry (Mosc)* 43: 14317-14324

[42] Li Y, Korolev S, Waksman G (1998) Crystal structures of open and closed forms of binary and ternary complexes of the large fragment of Thermus aquaticus DNA polymerase I: structural basis for nucleotide incorporation. *EMBO J* 17: 7514-7525

[43] Joyce CM, Potapova O, DeLucia AM, Huang XW, Basu VP, et al. (2008) Fingers-closing and other rapid conformational changes in DNA polymerase I (Klenow fragment) and their role in nucleotide selectivity. *Biochemistry (Mosc)* 47: 6103-6116

[44] Lerch HP, Mikhailov AS, Hess B (2002) Conformational-relaxation models of single-enzyme kinetics. *Proc Natl Acad Sci U S A* 99: 15410-15415

[45] Schwille P, Haustein E (2001) Fluorescence correlation spectroscopy: an introduction to its concepts and applications. Biophysics Textbook Online. pp. 1-33

[46] Johnson SJ, Taylor JS, Beese LS (2003) Processive DNA synthesis observed in a polymerase crystal suggests a mechanism for the prevention of frameshift mutations. *Proc Natl Acad Sci U S A* 100: 3895-3900

5. Identifying single-molecule dynamics with Burst Variance Analysis

In Chapter 4 we used Probability Distribution Analysis (PDA) to model heterogeneity in FRET histograms. While PDA was useful in fitting experimental FRET histograms to specific models of heterogeneity, it could not easily determine whether this heterogeneity arose from dynamic processes, or from the coexistence of several static structures. To address this problem we introduce Burst Variance Analysis (BVA), in which we detect dynamics by comparing the standard deviation of FRET over time for individual molecules, to that expected from theory. Both numerical simulations and experiments on well-characterized DNA hairpins validate the use of BVA to distinguish between static and dynamic sources of heterogeneity in PDA analyses, as well as to hypothesis-test specific models of dynamics against the observed standard deviation information. Using BVA, we revisited our analysis of the fingers-closing transition in the Klenow Fragment of *E. coli* DNA Polymerase I. Whereas PDA could account for the observed E^* distributions with models assuming either static or dynamic heterogeneity, BVA identified clear evidence for dynamics throughout the reaction trajectory of Pol I, which may be important to the mechanism by which it achieves its

impressive fidelity in DNA synthesis. BVA is broadly applicable to single-molecule FRET studies of macromolecular structure, and complements existing tools like PDA and fluorescence correlation spectroscopy in investigating macromolecular dynamics.

5.1 Introduction

In Chapter 4 we discussed heterogeneity in smFRET histograms. Broad FRET distributions may indicate the presence of static heterogeneity, dynamic heterogeneity, or a combination of the two. Static heterogeneity is due to the coexistence of multiple species with static but distinct FRET efficiencies in the same sample, while dynamic heterogeneity is due to a single molecular species that fluctuates between multiple distinct FRET states. Dynamic heterogeneity is of special interest, as it can provide information on the relationship between the conformational states of a biomolecule and its mechanism of action[1,2]. While PDA can be used to fit models of either static or dynamic heterogeneity to broad FRET distributions[3-6], it has difficulty distinguishing between the two cases; it is therefore difficult to determine whether the broadening observed is due to the superposition of many distinct, static structures, or to dynamic heterogeneity resulting from conformational dynamics.

Fluorescence correlation spectroscopy (FCS) methods have been used extensively to identify molecular dynamics through changes in either brightness or FRET[7-11]; as previously discussed, however, these methods suffer from a number of limitations, including their difficulty in resolving fluctuations on the diffusion timescale, their sensitivity to optical aberrations[12], and their use at the small-ensemble level, hindering the study of samples with incomplete fluorescent labelling or complex fluorophore stoichiometries[13]. Although newer FCS-based methods have addressed some of these issues[8,14], these methods typically require additional experimental controls and are meant to analyze the small-ensemble data appropriate to FCS, rather than single-molecule data. There is therefore a need for methods

that can distinguish between static and dynamic sources of heterogeneity in individual single-molecule FRET experiments.

Here we introduce Burst Variance Analysis (BVA)[15], which enables the detection of dynamics in single-molecule FRET data. In BVA we consider not only the mean FRET of each molecule, but also how this FRET fluctuates over time. Whereas the standard deviation of FRET for a perfectly static molecule is a simple analytical function of its mean FRET, for molecules with dynamic fluctuations in FRET the observed standard deviation will increase. BVA compares these expected and experimentally observed standard deviations, using a strict statistical criterion to determine whether a given sample exhibits dynamic FRET fluctuations. We demonstrate the ability of BVA to distinguish between static and dynamic heterogeneity using simulations and experiments on both static double-stranded DNAs, and dynamic DNA hairpins. We also demonstrate that BVA can be used to analyze the shot-noise predictions generated by PDA, providing a second dimension along which to hypothesis-test proposed models of biomolecular dynamics. Finally, we applied BVA to study fingers-closing dynamics in the Klenow Fragment (KF) of *E. coli* DNA Polymerase I; building on our work in Chapter 4, this analysis suggests the presence of fingers-closing dynamics in both KF-DNA (binary) and KF-DNA-dNTP (ternary) complexes at the millisecond timescale, consistent with a model of KF activity in which the nucleotide-bound polymerase may undergo several fingers-opening/closing transitions prior to nucleotide incorporation.

5.2 Theory

As in Chapter 4, the approximate FRET, or proximity ratio E^*, is calculated as the ratio of photons detected in the acceptor channel ($F_A + B_A$), to the total number of donor-excitation photons in the burst (N):

$$E^* = \frac{F_A + B_A}{F_D + B_D + F_A + B_A} = \frac{F_A + B_A}{N} \qquad [1]$$

where B_A is background in the acceptor channel, and F_A is fluorescence from the molecule of interest, including the contributions of leakage and direct-excitation.

PDA is carried out as described in Chapter 4. We note that as BVA analyzes photon arrival-time information, and arrival times are preserved in Arrival-Time PDA but not Dynamic PDA, we use Arrival-Time PDA throughout this chapter. In later sections, this will allow us to analyze not only experimental data, but PDA predictions, with BVA.

5.2.1 Burst Variance Analysis

When the observed FRET distribution is broader than the expected shot-noise distribution, PDA can be used to fit for multiple static components, or multiple dynamically-interconverting states (Refs. [3-6], and Chapter 4); however, it cannot easily discriminate between these two sources of heterogeneity. In its simplest form, we use Burst Variance Analysis (BVA) to determine whether the observed broadening of the E^* distribution can be attributed to dynamic heterogeneity.

Whereas PDA examines the heterogeneity in FRET among all molecules in a sample, BVA analyzes the heterogeneity in the FRET of individual molecules over time (Fig. 1). In the case of static heterogeneity, the width of the E^* distribution will expand beyond shot noise because different molecules have different originating FRET values; however, the FRET distribution over time for any individual static molecule will still be consistent with a shot-noise-limited distribution (Fig. 1A, magenta distribution matches dashed line); in contrast, molecules exhibiting FRET dynamics will show a larger single-molecule standard deviation than expected from shot-noise considerations (Fig. 1A, blue distribution is wider

than the dashed line). In BVA, we therefore test for dynamics by comparing the expected standard deviation for a given mean E^*, σ_{E^*} against the observed standard deviation, s_{E^*}.

For a static species, the expected standard deviation due to shot noise, σ_{E^*}, depends solely on photon statistics. Any set of n consecutive photons will follow a binomial distribution with respect to emission in the donor and acceptor channels. Assuming no background photons are present, the expected standard deviation of F_A is therefore that of a binomial, $\sqrt{nE^*(1-E^*)}$, and the standard deviation of $E^* = F_A / n$ is simply:

$$\sigma_{E^*} = \sqrt{\frac{E^*(1-E^*)}{n}} \qquad [2]$$

To calculate the burst-wide standard deviation, s_i, we segment each burst i into M consecutive (and non-overlapping) windows of n photons each (where M is the maximum number of such windows obtainable from the burst; Fig. 1B), and calculate the standard deviation of all windows within the burst:

$$s_i = \sqrt{\frac{1}{M_i}\sum_{j=1}^{M_i}\left(\varepsilon_{ij} - \mu_i\right)^2}, \quad \text{where} \quad \mu_i = \frac{1}{M_i}\left(\sum_{j=1}^{M_i}\varepsilon_{ij}\right) \qquad [3]$$

where ε_{ij} is the proximity ratio of window j in burst i, and μ_i is the mean FRET of all such windows in burst i.

Individual bursts frequently contain only a few photon windows, resulting in large errors on the calculated s_i. To increase the statistical power of our method, we segment the E^* axis into R bins, each centred on a given value of E^* and bearing a width w. For each bin, we calculate the expected standard deviation, s_{E^*}, of all windows belonging to bursts in the interval $L \le E^* < U$ (triangles in Fig. 1C), where $L = (E^* - w/2)$ and $U = (E^* + w/2)$ are lower and upper bounds on the bin respectively (Fig. 1C),

$$s_{E^*} = \sqrt{\sum_{\substack{i \text{ where} \\ L \le E_i^* < U}} \sum_{j=1}^{M_i} \left[\frac{\left(\varepsilon_{ij} - \mu\right)^2}{\sum M_i} \right]}, \quad \text{where} \quad \mu = \sum_{\substack{i \text{ where} \\ L \le E_i^* < U}} \sum_{j=1}^{M_i} \left(\frac{\varepsilon_{ij}}{\sum M_i} \right) \qquad [4]$$

where E_i^* is the proximity ratio of burst i, and μ is the mean FRET of all windows with $L \le E^* < U$. Unless otherwise indicated we define 20 bins, each with a width of 0.05 along the E^* axis; for instance, the bin centring on $s_{E^*=0.5}$ includes windows from all bursts with $0.475 \le \langle E_i^* \rangle < 0.525$. In this work, we consider s_{E^*} values only from only those bins containing at least 50 bursts, to ensure any dynamics detected are representative of sample behavior.

We note that the impact of photon window size, *n*, on the detection of dynamics is small near the diffusion timescale (Appendix, Fig. S1), though smaller windows are less computationally expensive to work with; unless otherwise indicated, we therefore set $n = 5$. We also note that, like any method of detecting FRET dynamics, BVA may be sensitive to dynamic changes in fluorophore quantum yield or orientation factor that also give rise to dynamic changes in FRET. It is therefore important to ensure that these artefacts are identified and eliminated with proper controls, or else occur on timescales distinct from the dynamics of interest. Finally we note that, as in PDA, we ignore the contribution of background to BVA for simplicity, though such contributions are typically negligible (Appendix, Fig. S2).

5.2.2 Confidence intervals

We calculate upper-limit confidence intervals on σ_{E^*} by considering the sampling distribution of standard deviations, $P(\sigma)$, expected for $M = \sum M_i$ windows of *n* photons. For Gaussian distributions of ε_{ij} with standard deviation σ_{E^*} (Eq. 2), this sampling distribution has an analytical solution[16]:

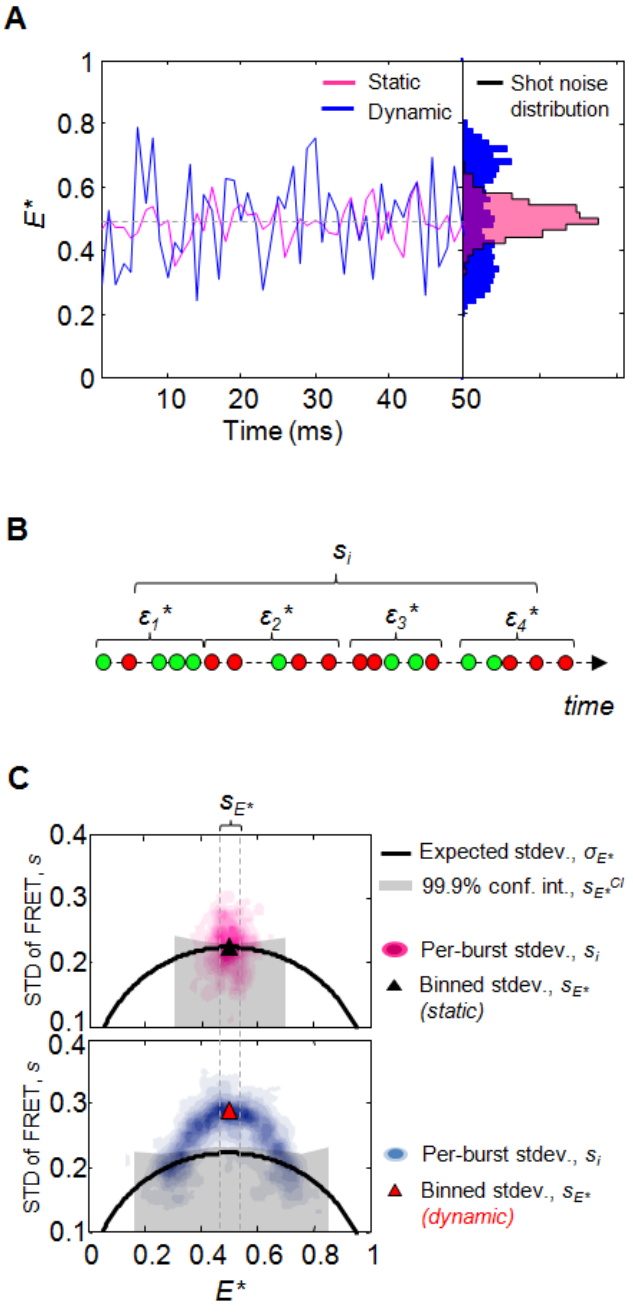

A
Static
Dynamic
Shot noise distribution
E*
Time (ms)
B
s_i
ε_1*
ε_2*
ε_3*
ε_4*
time
C
s_{E^*}
STD of FRET, s
Expected stdev., σ_{E^*}
99.9% conf. int., $s_{E^*}^{CI}$
Per-burst stdev., s_i
Binned stdev., s_{E^*}
(static)
Per-burst stdev., s_i
Binned stdev., s_{E^*}
(dynamic)
E*

Figure 1. Schematic diagram of the Burst Variance Analysis (BVA) method. (A) FRET time traces of single molecules with static (magenta) or dynamic (blue) underlying FRET values. FRET fluctuations in the static case are due only to shot noise, and a histogram of these FRET values (magenta distribution) matches the shot noise prediction (black line). FRET fluctuations in the dynamic case are due both to shot noise and dynamic changes in FRET efficiency; histograms of these FRET values are therefore wider than the shot-noise prediction (blue distribution). (B) Calculation of the standard deviation of FRET over time for a single molecule, s_i. In BVA, M samples of n consecutive photons (in the diagram, $M = 4$, $n = 5$) are taken from each burst. The FRET of each n-photon selection (denoted ε_j) is calculated, and the standard deviation of FRET, s_i is calculated as the standard deviation of ε_j for all M windows (note: there is only a single s_i value per burst; the FRET of the individual windows is not used in BVA other than to calculate s_i). (C) Contour plots of s_i plotted against the proximity ratio E^* for 1000 bursts, in simulations of samples exhibiting static (top, magenta) or dynamic (bottom, blue) heterogeneity. The black line indicates the expected standard deviation, σ_{E^*}, for n photons as a function of E^*. Whereas bursts with s_i values clustering around σ_{E^*} indicate static FRET (top panel), those significantly above σ_{E^*} indicate within-burst dynamics (bottom panel). To determine the significance of these apparent dynamics, we select all bursts within a narrow range of E^* values ($L \le E^* \le U$; dashed lines indicate L and U) and calculate the standard deviation of all photon windows belonging to the bursts within it, s_{E^*} (triangles). We then generate strict confidence intervals, $s_{E^*}^{CI}$ (gray region): for s_{E^*} falling below the confidence interval (top panel, black triangle), the data are consistent with static FRET; for s_{E^*} above the confidence interval, however (bottom; red triangle), the data are inconsistent with static FRET, and indicate the presence of FRET dynamics.

$$P(\sigma) = \frac{2\left(\frac{M}{2\sigma_{E^*}^2}\right)^{(M-1)/2} e^{-M\sigma^2/\left(2\sigma_{E^*}^2\right)} \sigma^{M-2}}{\Gamma\left((M-1)/2\right)} \qquad [5]$$

where Γ is the gamma function: $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$. Because the F_A giving rise to ε_{ij} is binomially distributed, the Gaussian assumption is reasonable for large *n*. In practice, we use a Monte Carlo approach to calculate the sampling distribution; while computationally more expensive than Eq. 5, this approach is more useful because it remains accurate even for small *n* where the Gaussian assumption fails.

To implement the Monte Carlo approach, we simulate the sampling distribution of s_{E^*},

$$\sigma = \sqrt{\sum_{\substack{i \text{ where} \\ L \le E_i^* < U}} \sum_{j=1}^{M_i} \left[\frac{\left(\frac{F_A^{ij}}{n} - \mu\right)^2}{\sum M_i} \right]}, \quad \text{where} \quad \mu = \sum_{\substack{i \text{ where} \\ L \le E_i^* < U}} \sum_{j=1}^{M_i} \left(\frac{\frac{F_A^{ij}}{n}}{\sum M_i} \right) \qquad [6]$$

where F_A^{ij} are random variables drawn from a binomial distribution with *n* trials (i.e. the number of photons per window) and *E** probability of success. We define the resulting Monte Carlo distribution as $P_{MC}(\sigma)$.

We use the distribution $P_{MC}(\sigma)$ to calculate the upper-tail confidence interval on the standard deviation, $s_{E^*}^{CI}$, and test for dynamics by comparing this value to the observed s_{E^*}. To prevent false positives for dynamics due to testing multiple s_{E^*} for dynamics within the same sample, we institute a simple Bonferroni correction for multiple hypothesis testing[17], such that for the desired confidence interval $CI = 100\%(1-\alpha)$ with significance level α, we adjust the significance level to α / R, where R is the number of bins tested. The Bonferroni-corrected upper-tail confidence interval $s_{E^*}^{CI}$ is therefore defined according to $(1-\alpha/R) = \int_0^{s_{E^*}^{CI}} P_{MC}(\sigma) \cdot d\sigma$. As mentioned above, we use $R = 20$ bins, each with a width of

0.05 along the E^* axis. We set our confidence level to $\alpha = 0.001$ so that any indication of significance in the detection of dynamics is highly likely to reflect the presence of dynamics.

5.3 Simulation Results

5.3.1 BVA can distinguish between static and dynamic heterogeneity

We first tested whether BVA can distinguish between static and dynamic heterogeneity as sources of broadening in E^* distributions. In addition to a series of single, static FRET species (Appendix, Fig. S3), we simulated an equimolar mixture of three species with distinct but static FRET efficiencies ($\langle E_1^* \rangle = 0.4$, $\langle E_2^* \rangle = 0.5$, $\langle E_3^* \rangle = 0.6$; simulation software written by Johannes Hohlbein) and analyzed them via PDA (Fig. 2A, top panel). As expected, the PDA prediction assuming the existence of these three static species achieved a good fit to the data (black line; $\chi_r^2 = 1.07$). However, the same data could also be reasonably well-accounted for by fitting for a single species with two dynamically interconverting states. Specifically, we performed a two-parameter fit assuming symmetry about E^* = 0.5 (so that a single parameter defines both $\langle E_1^* \rangle$ and $\langle E_2^* \rangle$), and equal forward and backward kinetic rates (red line; $\langle E_1^* \rangle = 0.373 \pm 0.003$, $\langle E_2^* \rangle = 0.627 \pm 0.003$, $k_{1\to2} = k_{2\to1} = 883 \pm 17\,\mathrm{s}^{-1}$; $\chi_r^2 = 1.19$). The ability of both static and dynamic PDA predictions to account for the observed E^* histogram (Fig. 2A, top panel; $\chi_r^2 < 2$ in both cases) demonstrates the difficulty of resolving static versus dynamic heterogeneity with PDA alone. We then analyzed the sample with BVA. As expected for this static sample, all s_{E^*} fell well within the predicted 99.9% confidence interval (Fig. 2A, bottom panel), correctly suggesting that the observed heterogeneity was due to static, rather than dynamic, sources.

We next simulated a simple, two-state dynamic FRET sample (Fig. 2B), using the FRET values and first-order rate constants from the dynamic PDA prediction in Fig. 2A

($\langle E_1 * \rangle = 0.37$, $\langle E_2 * \rangle = 0.63$, $k_{1\to2} = k_{2\to1} = 883\ \mathrm{s}^{-1}$). Again, PDA predictions assuming either the given two-state dynamic model, or fitting for a three-species static model ($\langle E_1 * \rangle = 0.382 \pm 0.002$, $\langle E_2 * \rangle = 0.5$ (fixed), $\langle E_3 * \rangle = 0.611 \pm 0.001$), could both account for the observed E^* distribution (Fig. 2B, top panel; $\chi_r^2 < 2$ in both cases). BVA, however, showed a clear increase in s_{E^*} beyond the confidence interval for intermediate values of E^*, indicating dynamics (Fig. 2B, bottom panel; triangles lying above the gray region are coloured red to indicate the presence of dynamics). Therefore, despite the similarity of E^* histograms resulting from static or dynamic sources of heterogeneity, BVA could correctly determine the types of heterogeneity present in these samples (Fig. 2B).

For the dynamic sample, the s_{E^*} values at intermediate E^* are above the confidence interval, while those nearer to the E^* of each individual state are not. This occurs because the diffusion time of the molecules (about 1 millisecond) is similar to the timescale of dynamics (in our simulations, molecules fluctuate on a timescale of $1/k_{1\to2} = 1/k_{2\to1} \approx 1.1\,\mathrm{ms}$); some molecules will therefore sample only one state or another during their diffusion through the confocal spot, producing an E^* value near that of one state or another ($E^* \approx 0.3, 0.7$), and giving rise to a s_{E^*} value consistent with static behavior. Molecules appearing at intermediate E^*, however, will have sampled both FRET states, and will therefore show both an intermediate E^* and an increased s_{E^*}.

5.3.2 BVA detects FRET dynamics in a timescale-dependent fashion

In Chapter 4, we demonstrated that while FRET histograms broaden significantly in response to dynamics near the diffusion timescale, they appear shot-noise-limited when dynamics are much faster or much slower than diffusion[5]. In the former case, this is because molecules interconvert so rapidly that the FRET efficiencies of their states effectively average out, and

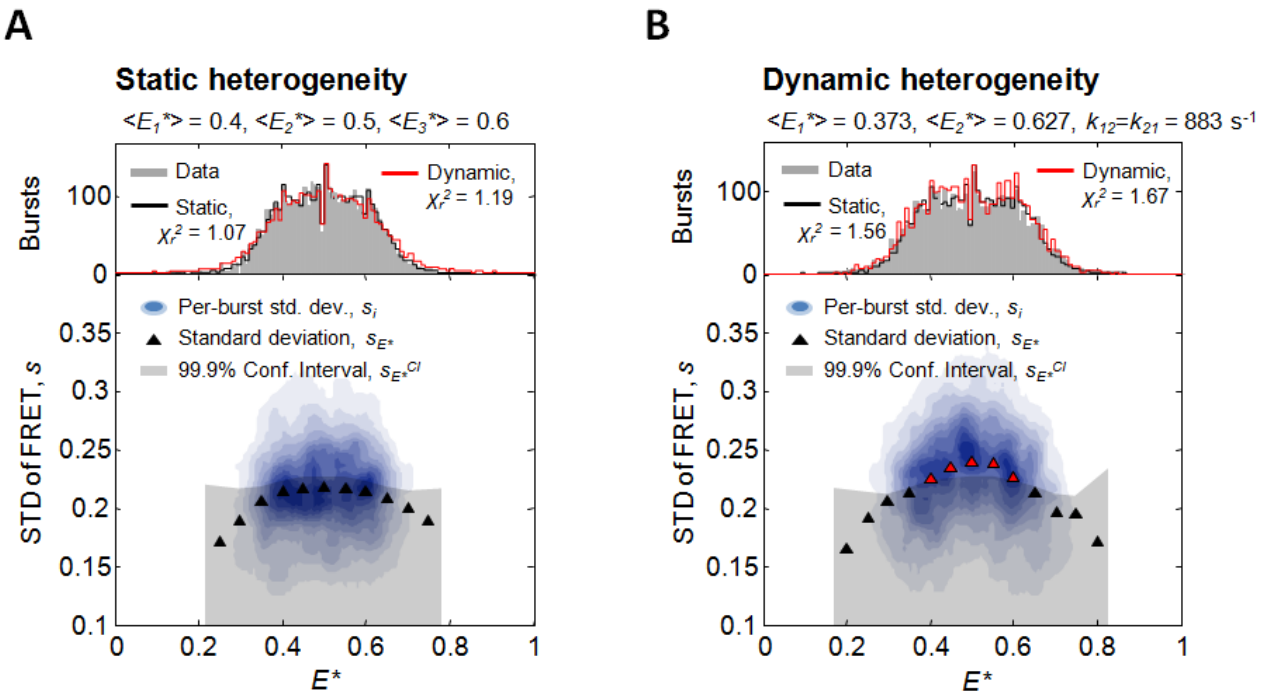


Figure 2. BVA can discriminate between static and dynamic heterogeneity. (A) E^* histogram of three simulated, equimolar static species with $\langle E_1^* \rangle = 0.4$, $\langle E_2^* \rangle = 0.5$, $\langle E_3^* \rangle = 0.6$. PDA predictions (top panel) assuming the presence of either the three known static components (black line), or fitting for a two-state dynamic model (red line; $\langle E_1^* \rangle = 0.373 \pm 0.003$, $\langle E_2^* \rangle = 0.627 \pm 0.003$, $k_{1\to2} = k_{2\to1} = 883 \pm 17\,\mathrm{s}^{-1}$) could both adequately account for the observed E^* distribution ($\chi_r^2 < 2$ in both cases). BVA of the same sample, however (bottom panel), correctly shows no evidence for dynamics (black triangles; all s_{E^*} fall below the confidence interval). (B) E^* histogram of a simulated dynamic species fluctuating with the fitted two-state dynamic parameters from panel A. PDA predictions (top panel) are shown for the three-species static model (black line; $\langle E_1^* \rangle = 0.382 \pm 0.002$, $\langle E_2^* \rangle = 0.5$ (fixed), $\langle E_3^* \rangle = 0.611 \pm 0.001$), and the two-state dynamic model (red line). As in panel A, both sources of heterogeneity could explain the observed E^* distribution ($\chi_r^2 < 2$). BVA, however (bottom panel), showed clear evidence for dynamics (red triangles; s_{E^*} of intermediate E^* fall above the confidence interval).

each burst exhibits an apparently constant (intermediate) FRET efficiency; in the latter, molecules interconvert so slowly that every burst is spent in one state or the other, causing each state to give rise to its own shot noise-limited population. Because BVA detects dynamics through intra-burst FRET fluctuations, it should show a timescale dependence similar to that observed when analyzing *E** histogram broadening via PDA. We therefore asked how well BVA could detect dynamics on different timescales.

We first illustrated the effects of fluctuation timescale on the s_{E^*} calculated with BVA by re-analyzing the dynamic simulations in Chapter 4, Fig. 3, with BVA. We simulated a series of dynamic species fluctuating between two FRET states, $\langle E_1 * \rangle = 0.3$, $\langle E_2 * \rangle = 0.7$, at timescales on the order of diffusion ($k_{1\to2} = k_{2\to1} = 10^3\ \mathrm{s}^{-1}$), or several orders of magnitude above or below it ($k_{1\to2} = k_{2\to1} = 10^6\ \mathrm{s}^{-1}, 1\mathrm{s}^{-1}$). As expected, the species fluctuating on the order of the diffusion times exhibited broadening by PDA, and had a significantly increased s_{E^*} (Fig. 3B). In contrast, molecules fluctuating much slower or faster than the diffusion timescale appeared static by both PDA and BVA (Fig. 3 A, C). Therefore, at least at timescales three orders of magnitude slower or faster than diffusion (given our simulation parameters), BVA is insensitive to the presence of dynamics.

To determine the timescales over which our method can successfully detect dynamics, we simulated the same two-state fluctuation ($\langle E_1 * \rangle = 0.3,\ \langle E_2 * \rangle = 0.7$) at many timescales ranging from 10^6 s^{-1} to 10^0 s^{-1}. To quantify our ability to detect dynamics at each timescale we calculated a "Dynamic Score" (DS), the sum of squared residuals between the observed standard deviation s_{E^*} and the upper-tail confidence interval $s_{E^*}^{CI}$ for all significant s_{E^*} (i.e. those above the confidence interval):

$$DS = \sqrt{\sum_{\left(s_{E^*} - s_{E^*}^{CI} > 0\right)} \left(s_{E^*} - s_{E^*}^{CI}\right)^2} \qquad [7]$$

The DS is similar to a least-squares objective function, but excludes those points consistent with a static hypothesis; this results in an intuitive measure of dynamics such that, when the DS is zero, all s_{E^*} fall within the confidence interval and the molecule appears static; nonzero DS indicates significant evidence for dynamics; and a larger DS indicates stronger evidence for dynamics.

Using the same two-state dynamic species ($\langle E_1 * \rangle = 0.3,\ \langle E_2 * \rangle = 0.7$), we calculated the DS over a large range of timescales, and were able to detect dynamics spanning over four orders of magnitude (black line, Fig. 3D). As expected, BVA was most sensitive to FRET fluctuations at the diffusion timescale. We confirmed this by altering the diffusion coefficient in our simulations ($D = 3.0\times10^6,\ 3.0\times10^7,\ 3.0\times10^8$ nm^2 s^{-1}), and observing the effect on the DS at different fluctuation timescales (Fig. 3D). As expected, we found that decreasing the diffusion coefficient, and therefore increasing the diffusion time (blue, red lines) increased the dynamic timescale at which we achieved the highest DS; in fact, the mean diffusion time per burst (arrows) fell on the fluctuation timescale at which DS was maximal for all diffusion coefficients. As diffusion time depends on the size and geometry of the confocal spot, we also expect features of the experimental setup to affect the sensitivity of BVA to dynamics.

We also tested the effect of fluctuation amplitude, $\Delta\langle E * \rangle = \left|\langle E_2 * \rangle - \langle E_1 * \rangle\right|$, on the sensitivity of BVA to dynamics. Since larger changes in $\Delta\langle E * \rangle$ at the same timescale will necessarily produce larger standard deviations in dynamic samples, the DS should increase significantly with increasing $\Delta\langle E * \rangle$. Indeed we observed a strong increase in DS at all fluctuation timescales with increasing $\Delta\langle E * \rangle$ (Fig. 3E). Our method is therefore sensitive not just to the timescale of fluctuations, but also to their amplitude, and exhibits good sensitivity to dynamics overall: we were able to detect dynamics within at least two orders of magnitude

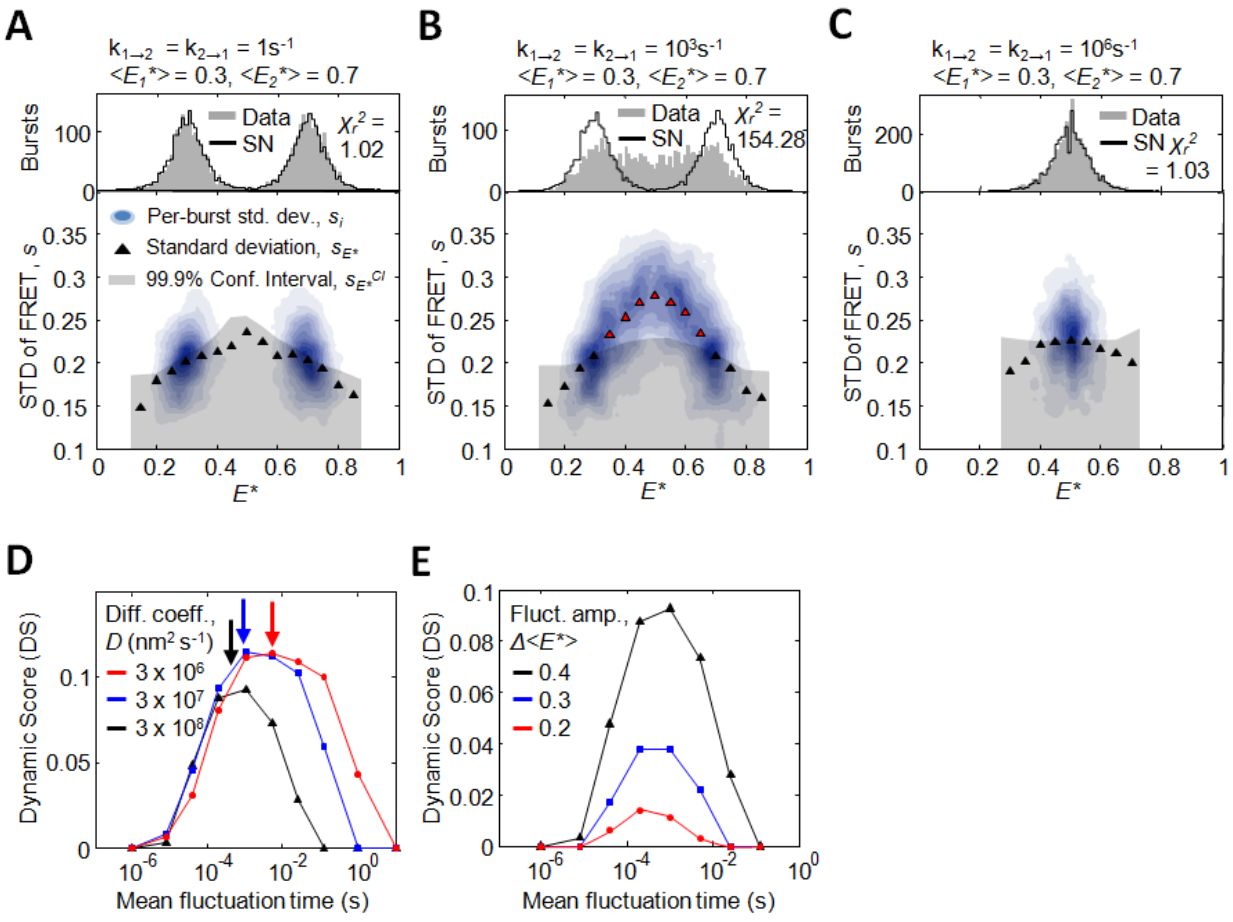

A
$k_{1\to2} = k_{2\to1} = 1s^{-1}$
$\langle E_1^* \rangle = 0.3, \langle E_2^* \rangle = 0.7$
Bursts
Data
SN
χ_r^2 = 1.02
Per-burst std. dev., s_i
Standard deviation, s_{E^*}
99.9% Conf. Interval, $s_{E^*}^{CI}$
STD of FRET, s
E*
B
$k_{1\to2} = k_{2\to1} = 10^3s^{-1}$
$\langle E_1^* \rangle = 0.3, \langle E_2^* \rangle = 0.7$
χ_r^2 = 154.28
C
$k_{1\to2} = k_{2\to1} = 10^6s^{-1}$
$\langle E_1^* \rangle = 0.3, \langle E_2^* \rangle = 0.7$
χ_r^2 = 1.03
D
Diff. coeff., D ($nm^2 s^{-1}$)
3 x 10^6
3 x 10^7
3 x 10^8
Dynamic Score (DS)
Mean fluctuation time (s)
E
Fluct. amp., $\Delta\langle E^* \rangle$
0.4
0.3
0.2

Figure 3. The sensitivity of BVA to dynamics depends on the timescale and amplitude of fluctuations. For a simple two-state dynamic species ($\langle E_1 * \rangle = 0.3$, $\langle E_2 * \rangle = 0.7$), the sensitivity of BVA depends on the timescale of fluctuation. (A) For fluctuation timescales significantly longer than the diffusion time (1 s^{-1}), a dynamic two-state molecule appears as two shot-noise limited distributions (PDA plot, $\chi_r^2 = 1.02$), with no apparent dynamic component (BVA plot, all s_{E*} fall below confidence interval). (B) For fluctuation timescales on the order of the diffusion time (10^3 s^{-1}), FRET distributions show significant broadening beyond shot noise (PDA plot, $\chi_r^2 >> 1$) and evidence for dynamics (BVA plot, s_{E*} fall above the confidence interval). (C) For fluctuation timescales significantly faster than the diffusion time (10^6 s^{-1}), the distribution appears as a single, shot-noise limited population with FRET intermediate between the two states (PDA plot, $\chi_r^2 = 1.03$), and no apparent evidence for dynamics (BVA plot, all s_{E*} fall below confidence interval). (D) Plots of Dynamic Score (DS) over different fluctuation timescales. For molecules with different diffusion coefficients, the DS is maximal near the diffusion time (arrows) in all cases, with greater diffusion times increasing the fluctuation timescale at which sensitivity is greatest. (E) Increasing the amplitude of FRET fluctuation, $\Delta\langle E * \rangle = |\langle E_2 * \rangle - \langle E_1 * \rangle|$, increases the DS at all fluctuation timescales.

of the diffusion constant for fairly low-amplitude fluctuations ($\Delta\langle E^*\rangle = 0.2$), and more than four orders of magnitude for larger fluctuations ($\Delta\langle E^*\rangle = 0.4$).

5.4 Experimental Results

5.4.1 BVA of static and dynamic DNA hairpins

For experimental validation of the BVA method, we used the control DNAs and hairpins previously described (Chapter 4 Appendix, Fig. S1). Briefly, these include: a hairpin that fluctuates dynamically between an "open" and a "closed" conformation, giving rise to large structural changes on the timescale of diffusion[3,11,18], two hairpin controls that occupy the open and closed hairpin conformations without interconverting, and a donor-only-labelled, double-stranded DNA.

We first analyzed the donor-only control, which we previously confirmed to be shot-noise limited (Chapter 4). As expected, all s_{E^*} values fell well within the 99.9% confidence interval, correctly indicating that it is static (Appendix, Fig. S4). We then verified that the individual closed and open conformations of the hairpin are static by analyzing a mixture of our control hairpins (Fig. 4B), which do not fluctuate dynamically but remain in either the closed or open conformations permanently. In Chapter 4 we found that each *E** distribution was broader than expected from shot noise alone (Chapter 4 Appendix, Fig. S2), consistent with a Gaussian distribution of $\langle E^*\rangle$ in the range of 0.15-0.23 nm, as observed previously[3,5,6,15,19,20]. Such heterogeneity has been attributed to either acceptor dye photophysics, or long-lived states in which fluorophores occupy different positions and/or orientations with respect to the DNA[19,20]. Consistent with this proposed quasi-static heterogeneity, BVA showed no evidence for dynamics in either of the two control hairpins (Fig. 4B).

We then used BVA to analyze the DNA hairpin, which is expected to interconvert at the millisecond timescale between the FRET states represented by the two controls, and should therefore produce a large dynamic signal; indeed, in a previous publication we detected dynamics in this hairpin using a simpler but less rigorous form of BVA[21]. As expected, analysis with BVA revealed a dramatic increase in s_{E^*} at intermediate E^* (Fig. 4C, red triangles); in these bursts, the hairpin actively switched between the open and closed conformations during its transit through the confocal volume, giving rise to bursts with intermediate E^* and high s_{E^*}. In contrast, for values of E^* near that of the control hairpins, the bursts remained in one state or the other during their transit through the volume (Fig. 4C, black triangles) producing s_{E^*} values close to those observed in the controls (Fig. 4B).

In several previous publications[3,10,18] hairpin dynamics have been proposed to occur via a simple, two-state kinetic model, in which the hairpin fluctuates between open and closed conformations with first-order rate constants k_{close} and k_{open}. In Chapter 4 we used PDA to fit a two-state model to the data, and recovered best-fit parameters of $k_{open} = 641 \pm 9\,\mathrm{s}^{-1}$, $k_{close} = 463 \pm 14\,\mathrm{s}^{-1}$, which yielded a good fit to the data ($\chi_r^2 = 1.53$) and an overall reaction time ($\tau_R = 1/\left(k_{open} + k_{close}\right) = 0.91$ ms) consistent with previous correlation-based analyses of the same hairpin ($\tau_R = 0.5 - 1.0$ ms)[11,21]. As BVA takes advantage of within-burst dynamical data not considered in PDA, we asked whether this model was still consistent with the data when analyzed by BVA. We therefore performed BVA on the Arrival-Time PDA prediction from Chapter 4 using the best-fit parameters above, and found good agreement between the s_{E^*} of the experimental data, and the s_{E^*} of the dynamic two-state PDA prediction (Fig. 4C, green triangles). This supported the conclusion of earlier PDA-based work suggesting that a two-state model is appropriate in accounting for the dynamics of the hairpin, at least on the

timescale of diffusion (Chapter 4 and Ref. [3]). We note, however, that our best-fit solution exhibits slight but systematic deviations in s_{E^*}, similar to those observed in the PDA analysis; as in Chapter 4, this could be due to the existence of a minor additional state, or states, contributing to dynamics[22,23].

Finally, as a demonstration of the ability of BVA to distinguish between static and dynamic heterogeneity, we tested whether the observed FRET data for the dynamic hairpin could also be accounted for by static heterogeneity. We simulated a number of static molecules with distinct FRET efficiencies, whose overall *E** distribution was very similar to that of the DNA hairpin (Fig. 4C, blue). While the *E** distribution of this static sample was similar to that of the dynamic hairpin ($\chi_r^2 = 1.47$), BVA showed clearly that, unlike the two-state dynamic model, the static model was insufficient to account for the s_{E^*} values of the experimental hairpin. Overall, our analysis of these simple DNA samples demonstrates the ability of BVA both to detect dynamics in experimental samples and, coupled with PDA, to hypothesis-test different models of molecular heterogeneity against experimental data.

5.4.2 Dynamics in the Klenow Fragment of *E. coli* DNA Polymerase I

In Chapter 4, and in a recent publication[3], we showed that the *E** histogram of unliganded KF was consistent with a simple two-state kinetic model using PDA, such that it appears to fluctuate between its open and closed states at the millisecond timescale (Chapter 4, Fig. 6). While PDA also showed consistency with a model in which KF exists as two static but quasi-statically broadened distributions, the dynamic interpretation was supported by data from FCS-based methods, and via a preliminary form of BVA, both of which suggested the presence of millisecond-timescale dynamics in the unliganded polymerase[24].

For the DNA-Pol binary and DNA-Pol-dNTP ternary complexes, however, the results were less clear: FCS-based methods failed to reveal clear dynamics in these samples, despite

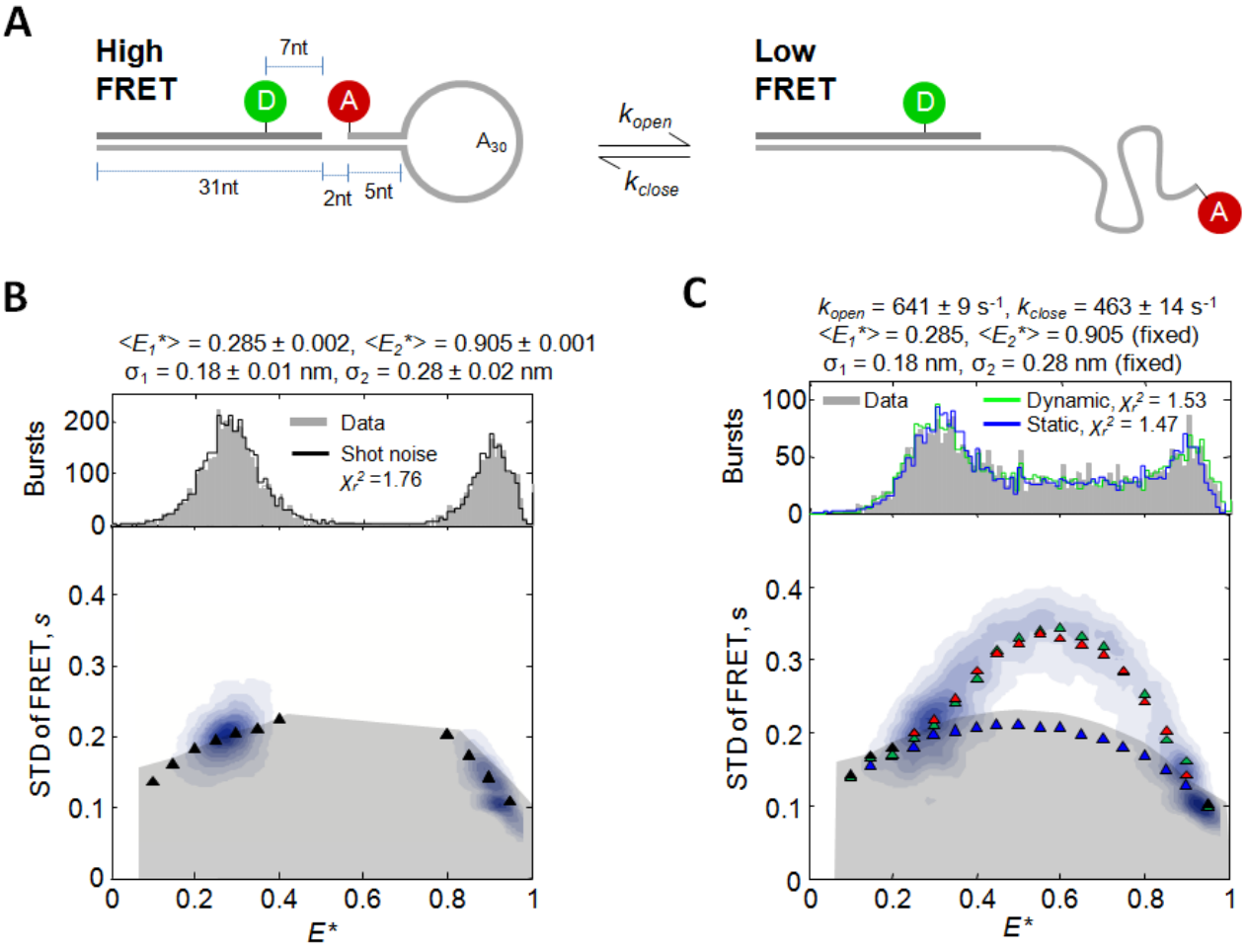

A
High FRET
7nt
D
A
A30
31nt
2nt
5nt
kopen
kclose
Low FRET
D
A
B
<E1*> = 0.285 ± 0.002, <E2*> = 0.905 ± 0.001
σ1 = 0.18 ± 0.01 nm, σ2 = 0.28 ± 0.02 nm
Bursts
Data
Shot noise
χr2 =1.76
STD of FRET, s
E*
C
kopen = 641 ± 9 s-1, kclose = 463 ± 14 s-1
<E1*> = 0.285, <E2*> = 0.905 (fixed)
σ1 = 0.18 nm, σ2 = 0.28 nm (fixed)
Bursts
Data
Dynamic, χr2 = 1.53
Static, χr2 = 1.47
STD of FRET, s
E*

Figure 4. BVA of a dynamic DNA hairpin reveals FRET dynamics. (A) Structure and kinetics of a DNA hairpin; panel is repeated from Chapter 4, Fig. 6A. We model the hairpin as transitioning between open and closed states with first-order kinetic rates k_{open} and k_{close}. (B) A mixture of two static hairpin controls. In Chapter 4, we showed that these FRET distributions could be well-fit assuming a Gaussian distribution of FRET efficiencies, with $\langle E_{open}* \rangle = 0.285 \pm 0.002$, $\sigma^{r}_{open} = 0.18 \pm 0.01$ nm, $\langle E_{closed}* \rangle = 0.905 \pm 0.001$ and $\sigma^{r}_{closed} = 0.28 \pm 0.02$ nm ($\chi^2_r = 1.76$, top panel). BVA suggests that this heterogeneity is static (bottom panel), or else due to dynamics several orders of magnitude slower than the diffusion timescale (Fig. 3D). (C) PDA and BVA of the dynamic FRET hairpin. The *E** histogram could be fit to a two-state dynamic model in which the hairpin fluctuates between the FRET values of the controls at rates $k_{open} = 641 \pm 9\, s^{-1}$, $k_{close} = 463 \pm 14\, s^{-1}$ ($\chi^2_r = 1.53$; top panel, green line); consistent with this model, BVA shows a dramatic increase in s_{E*} at intermediate *E** for the experimental data (bottom panel, black and red triangles). Moreover, BVA of the Arrival-Time PDA prediction generated a s_{E*} profile (green triangles) similar to that of the experimental data (black and red triangles). In contrast, while simulations of several static species could accurately reproduce the *E** histogram of the dynamic hairpin ($\chi^2_r = 1.47$;blue line, top panel), BVA analysis of these simulations (blue triangles), produced s_{E*} values that were clearly static, and diverged strongly from the experimental data.

the apparent existence of both open and closed conformations in equilibrium[3,24], and stopped-flow data showing diffusion-timescale fingers-closing during ternary complex formation[25]. Moreover, these complexes' E^* histograms could be satisfactorily fit by either sums of static distributions(Chapter 4, Ref. [24]) or by a PDA-based two-state dynamic model (Chapter 4, Ref. [3]). Here we revisit the question of whether these samples are dynamic, similar to the unliganded polymerase, or represent mixtures of static (or very-slowly-interconverting) polymerase subpopulations. We use BVA to re-analyze our data from Chapter 4, asking which of these two opposing models is more consistent with the data.

In Figure 5, we show FRET histograms for the unliganded, binary and ternary complex forms of KF (Fig. 5 B, C, D; all smFRET KF data was collected by Yusdi Santoso). As in Chapter 4, we show both the two-species static and two-state dynamic PDA models, all of which achieved a reasonable fit to the data ($\chi_r^2 < 2$). Using BVA, we then analyzed both the experimental FRET data for each form of KF, as well as its static and dynamic PDA predictions, and compared the resulting s_{E^*} values (Fig. 5 B, C, D, bottom panels; full contour plots of s_i values shown in Fig. 6). Whereas the s_{E^*} of the actual data were in good qualitative agreement with the s_{E^*} from the dynamic predictions in all cases, the s_{E^*} of the static predictions were consistently and substantially lower than the s_{E^*} for either the data or the dynamic prediction. To quantify this difference, we calculated the sum of squared residuals (SSR) between the experimental s_{E^*} and those of each prediction, where a smaller SSR indicates a better fit. In all three KF complexes, the SSR of the dynamic prediction was about an order of magnitude smaller than that of the static prediction (Fig. 5 B, C, D), suggesting their E^* distributions were better explained by dynamic, rather than static, heterogeneity.

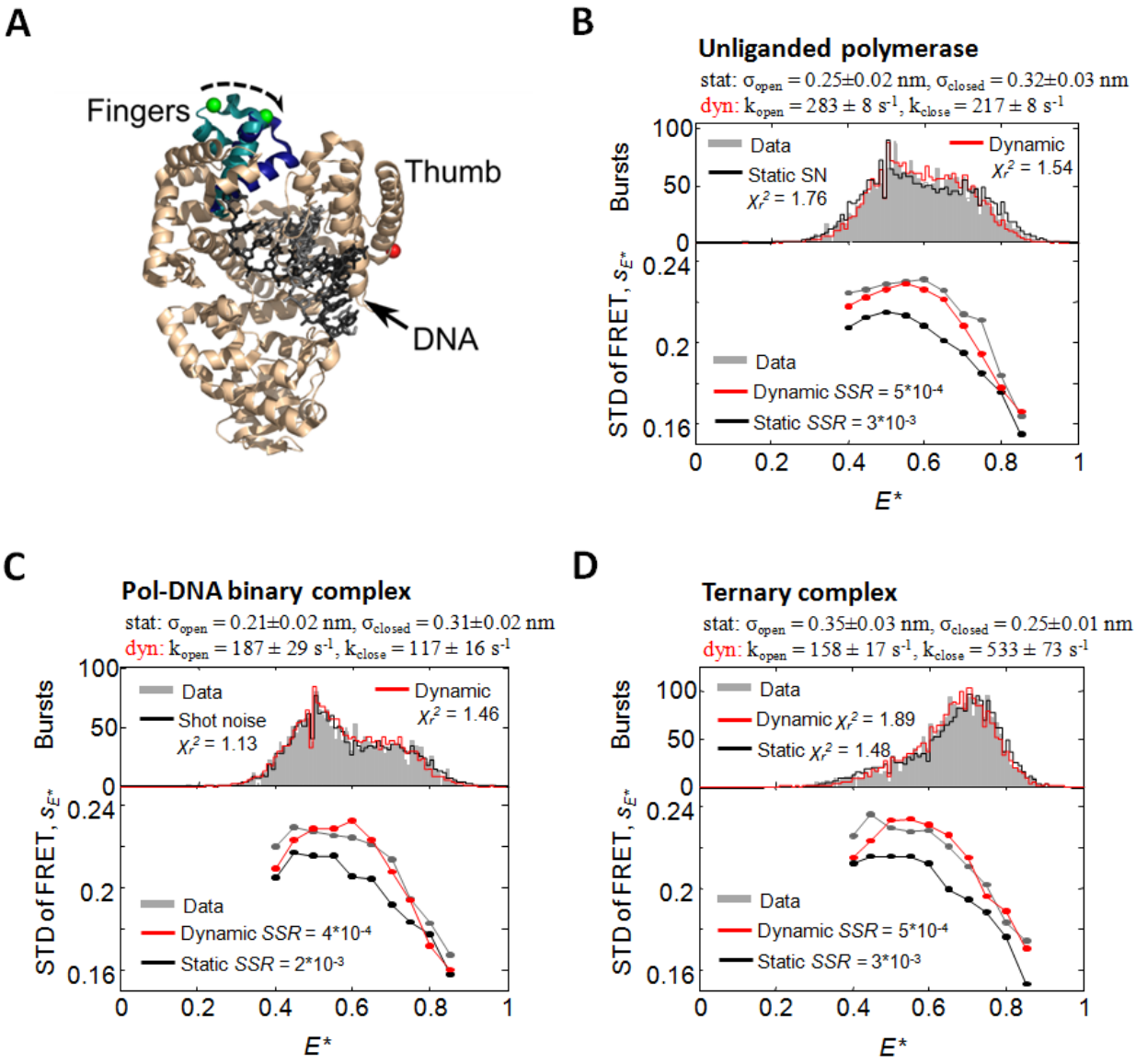

A
Fingers
Thumb
DNA
B
Unliganded polymerase
stat: σopen = 0.25±0.02 nm, σclosed = 0.32±0.03 nm
dyn: kopen = 283 ± 8 s-1, kclose = 217 ± 8 s-1
Bursts
Data
Static SN
χr2 = 1.76
Dynamic
χr2 = 1.54
STD of FRET, sE*
Data
Dynamic SSR = 5*10-4
Static SSR = 3*10-3
E*
C
Pol-DNA binary complex
stat: σopen = 0.21±0.02 nm, σclosed = 0.31±0.02 nm
dyn: kopen = 187 ± 29 s-1, kclose = 117 ± 16 s-1
Bursts
Data
Shot noise
χr2 = 1.13
Dynamic
χr2 = 1.46
STD of FRET, sE*
Data
Dynamic SSR = 4*10-4
Static SSR = 2*10-3
E*
D
Ternary complex
stat: σopen = 0.35±0.03 nm, σclosed = 0.25±0.01 nm
dyn: kopen = 158 ± 17 s-1, kclose = 533 ± 73 s-1
Bursts
Data
Dynamic χr2 = 1.89
Static χr2 = 1.48
STD of FRET, sE*
Data
Dynamic SSR = 5*10-4
Static SSR = 3*10-3
E*

Figure 5. KF and its complexes exhibit conformational dynamics. (A) Superposition of the crystal structures of the Pol-DNA binary complex (PDB file 1L3U) and Pol-DNA-dNTP ternary complex (PDB file 1LV5) for *B. stearothermophilus* Pol I[30], a close homologue of *E. coli* Pol I. The α-carbon backbone of the protein is shown in beige, except for the fingers subdomain, which is shown in teal (fingers open) or dark blue (fingers closed). Green and red spheres indicate the positions of the donor and acceptor fluorophores respectively, while the arrow indicates the structural change in the fingers subdomain during a "fingers-closing" transition. (B) Analysis of the unliganded KF by PDA (top panel) and BVA (bottom panel shows s_{E*} values; for clarity, s_i values are not shown here, but are shown in Fig. 6; all smFRET data collected by Yusdi Santoso). PDA achieved reasonable fits to the data for either a two-species static or two-state dynamic model (fits taken from Chapter 4, Fig. 6; $\chi_r^2 < 2$ in both cases). BVA data, however, strongly favoured the dynamic model: the sum of squared residuals (SSR) between the s_{E*} of the dynamic BVA prediction and experimental data (red) is an order of magnitude lower than the SSR between the s_{E*} of the static BVA prediction and experimental data (black). The same is true of the (C) Pol-DNA binary complex and (D) Pol-DNA-dNTP ternary complex, providing evidence for millisecond-timescale fingers-closing dynamics in all three cases.

To ensure that this result was not dependent on the model of static heterogeneity employed, we also tested a model assuming the existence of a third species with a mean E^* between that of the open and closed complexes; this model, too, showed poor agreement with the data (Fig. 6). We also asked how the specific timescales used in the dynamic prediction affected their agreement with the experimental data. In addition to finding that the rates extracted with PDA (200 – 500 s^{-1}) matched the experimental s_{E^*} very well, we found that these rates could be altered by over an order of magnitude before achieving an SSR comparable to the static prediction (Fig. 7). This strongly suggests the presence of millisecond-timescale fingers-closing dynamics in both the binary and ternary complexes of the Klenow Fragment.

We note that we cannot rule out the possibility that protein dynamics distinct from the fingers-closing transition are affecting the FRET efficiency E^*, and creating, or contributing to, the dynamic signal in BVA. Nevertheless, the consistency of the rates extracted here with those expected from stopped-flow and correlation-based methods suggests that we are indeed observing fingers-closing dynamics. Together these results suggest that the fingers-closing transition of the KF complex may be dynamic throughout its reaction trajectory, whether unliganded, DNA-bound, or poised for dNTP incorporation.

5.5 Conclusion

We have described and implemented Burst Variance Analysis (BVA), a novel analytical method that enables the detection of FRET dynamics in single-molecule FRET experiments; it accomplishes this by comparing the standard deviation of FRET from individual molecules over time, to that expected from theory. In addition to detecting dynamics, BVA can be used

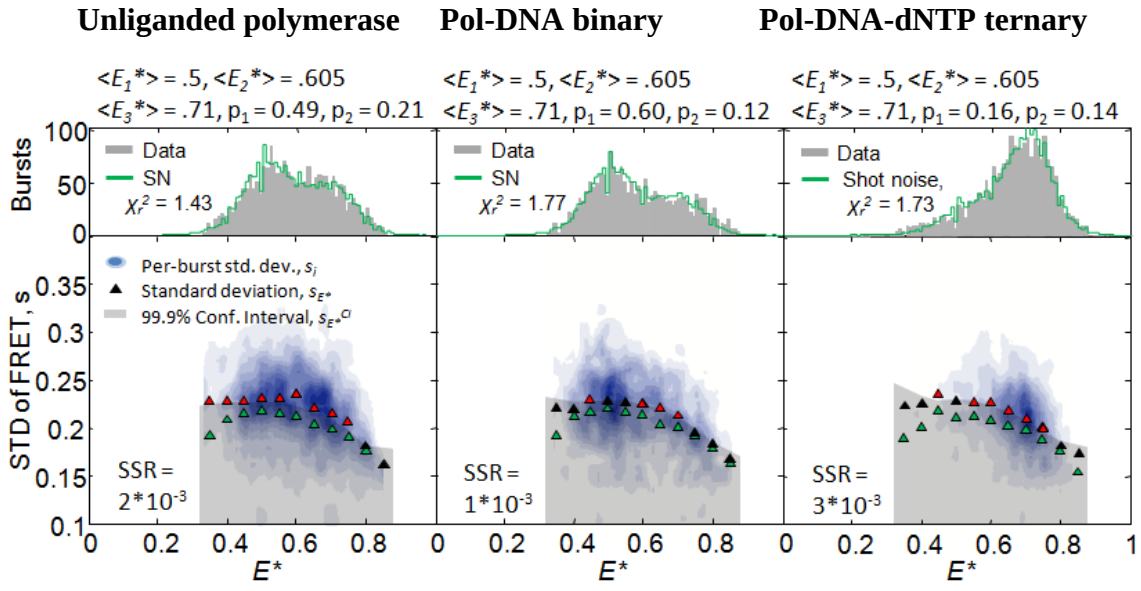


Figure 6. Static three-species models cannot account for DNA Pol I dynamics. As an alternative to the two-species static predictions in Fig. 5, in which we fit for the standard deviation of each of two species, we tested a three-species static model; in this model, all Gaussian widths are fixed (σ^r = 0.18 nm) but relative concentrations of each species are allowed to vary. We then asked whether such a model could account for the data in the (left) unliganded polymerase, (centre) Pol-DNA binary complex or (right) ternary complex. We first used PDA to achieve a fit to the *E** histograms (top panels); we held the FRET of the fingers-open and fingers-closed states fixed ($\langle E_1 * \rangle = 0.5$, $\langle E_3 * \rangle = 0.71$) while adding an intermediate state ($\langle E_2 * \rangle = 0.605$), and fitting using two parameters to describe the frequencies of the first two species (p_1 and p_2; the frequency of the third species, p_3, is simply: $1 - p_1 - p_2$). While in all cases the three-species fit provided a satisfactory match to the *E** histogram ($\chi_r^2 < 2$), BVA plots (bottom panels) showed that the s_{E*} for the static predictions (green triangles) deviate strongly from the s_{E*} observed for the data (black and red triangles). Indeed, the sum of squared residuals, *SSR*, between each set of data and its predicted s_{E*} was similar to that of the static two-state predictions in Fig. 5, being about an order of magnitude larger than the SSR between the data and the dynamic model predictions.

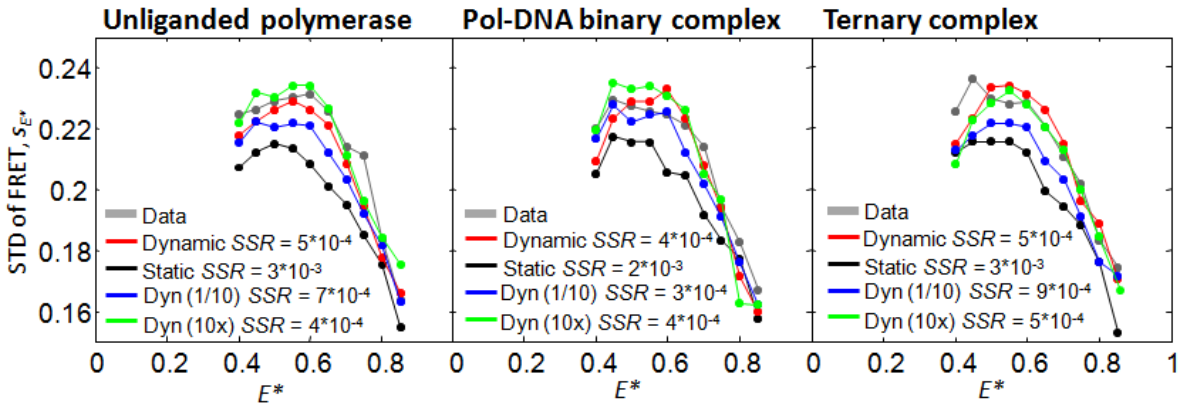


Figure 7. The dynamic model of *E heterogeneity in the Klenow Fragment is robust to fluctuation timescale.** We repeated our BVA analysis of the unliganded Klenow Fragment (left), Pol-DNA binary complex (middle) or Pol-DNA-dNTP ternary complex (right), from Fig. 5. In addition to analyzing experimental data (grey) and best-fit dynamic (red circles) and static (black circles) PDA predictions with BVA, we also analyzed dynamic predictions where both the forward and backward rate constants, k_{open} and k_{close}, had been increased (green circles) or decreased (blue circles) by ten-fold. In all cases, the dynamic predictions achieved a lower SSR with respect to the experimental data, than did the static prediction.

in combination with Probability Distribution Analysis (PDA) to hypothesis-test specific models of static or dynamic heterogeneity against single-molecule FRET data.

Using numerical simulations, we demonstrated that BVA can readily distinguish between static and dynamic sources of heterogeneity in E^* distributions for FRET fluctuations occurring over several orders of magnitude. This ability depends on both the magnitude of the FRET fluctuation and its timescale: BVA was most sensitive to large fluctuations, and to fluctuations on the order of the diffusion time. This unique feature of BVA makes it complementary to existing correlation-based methods of detecting dynamics, which are least sensitive to dynamics on the diffusion timescale[26].

We validated the use of BVA in experimental samples by analyzing both static DNA samples and a well-characterized dynamic DNA hairpin[3,10,11,18,21]. Despite the FRET histogram broadening observed with PDA in simple, double-stranded DNAs, BVA failed to identify dynamics; this is consistent with the presence of slow photophysics, or slow dynamics in the position or orientation of the dye relative to the DNA, as previously suggested[19,20]. In contrast to the static DNAs, the DNA hairpin was clearly dynamic when analyzed by BVA, as expected from previous correlation-based studies[10,11,18,22]; moreover, BVA analysis of a two-state dynamic PDA prediction matched the experimental data well, increasing our confidence in the two-state model as an adequate description of hairpin dynamics, at least at the diffusion timescale. We did, however, observe a small and systematic deviation between BVA analyses of the experimental data and the PDA prediction; as suggested in Chapter 4, such a deviation could be due to the inadequacy of a two-state model in fully accounting for the single-molecule FRET data.

Finally, we used BVA to analyze fingers-closing dynamics in the Klenow Fragment (KF) of *E. coli* DNA Polymerase I in three states: unliganded, DNA-Pol binary complex, and DNA-Pol-dNTP ternary complex. We asked whether the heterogeneity in each FRET

histogram was better explained by static or dynamic factors. We first analyzed the unliganded polymerase; consistent with recent publications from our lab, in which we identified apparent millisecond-timescale dynamics in the unliganded polymerase via both dynamic PDA analysis and correlation-based methods[3,24], BVA provided strong support for the millisecond-timescale dynamic model of heterogeneity over the static model.

Interestingly, we also found evidence for dynamics in the binary and ternary complexes, which were not detected previously using a correlation-based approach[24], and were suggested, but could not be conclusively identified, from PDA analysis and stopped-flow studies (Refs. [3,25] and Chapter 4). These dynamics occurred on a timescale similar to that observed in the unliganded polymerase, and were consistent with rates expected from a combination of previous smFRET[3,24] and stopped-flow experiments[25]. Such dynamics are also consistent with recent studies, which have shown that fingers-closing dynamics are not rate-limiting for dNTP incorporation; this suggests that fingers-closing precedes, but does not commit the polymerase to, dNTP incorporation[25,27] . Together with our results, this is suggestive of a model in which the polymerase binds to a complementary nucleotide and undergoes fingers-closing, possibly several times, prior to successful nucleotide incorporation. As the closed conformation of the polymerase is important in discriminating between complementary and noncomplementary nucleotides[28], dynamic sampling of this conformation may be relevant to the mechanism by which DNA Polymerase I achieves its impressive fidelity.

An important caveat is that BVA, like correlation-based methods, detects apparent FRET dynamics, which can arise from a number of sources other than changes in donor-acceptor distance. FRET dynamics may appear due to photophysical blinking of the acceptor dye, for instance, though this can be detected using methods like ALEX[13], then removed computationally[5]. E^* values can also be affected by changes in the Förster distance R_o,

which may in turn be caused by dynamic variation in orientation factor κ^2, dye spectra, or donor quantum yield; these artefacts are especially problematic since they are difficult to distinguish from dynamics in D-A distance, but can in principle be detected using appropriate measurements of fluorophore lifetime or anisotropy. In these and other cases, it is important to perform appropriate controls to ensure the dynamics observed are occurring in the molecule of interest, and not simply the individual fluorophores attached to it[29].

Our results illustrate the usefulness of BVA for detecting structural dynamics via single-molecule FRET. Due to its single-molecule and diffusion-timescale sensitivity, BVA complements correlation-based methods which operate on the small-ensemble level, and have difficulty detecting dynamics near the diffusion timescale. BVA also complements PDA by adding a dimension along which to hypothesis-test specific models of dynamics; this will be useful in rejecting incorrect models that produce E^* predictions consistent with experimental E^* distributions, but show poor agreement between predicted and observed BVA data. We expect BVA to be broadly useful in single-molecule FRET studies of enzyme structure and dynamics, and protein and nucleic acid folding.

5.6 Materials and methods

Simulations of smFRET experiments

Simulations were carried out as described in Chapter 3.

Single-molecule experiments

Single-molecule measurements were performed on an alternating laser excitation (ALEX) microscope as described in Chapter 3, but using a water-immersion objective (UPLS-APO, 60x, 1.2 NA, Olympus, Japan) to avoid a refractive index mismatch with the sample (see

chapter 4, Methods). The excitation powers were 180 μW and 60 μW for the 532 nm and 635 nm lasers respectively, and samples were analyzed at a concentration of 50-100 pM. All DNA samples were measured in 400 mM NaCl, 1 mM EDTA, 100 μg/ml BSA, and 5% (v/v) glycerol. KF samples were measured in 40 mM HEPES-NaOH, pH 7.3, 10 mM $MgCl_2$, 1 mM DTT, 100 μg/ml BSA, and 1 mM mercaptoethylamine.

Data Analysis

Burst detection and data analysis were performed as described in Chapter 3. Unless otherwise noted, all data was thresholded according to stoichiometry, $S > 0.45$, effectively eliminating acceptor-only fluorescent species from our analysis. BVA was implemented as described (Theory Section), while PDA was carried out as described in Ref. [3]. 2D BVA histograms were smoothed with the *smooth3* function in MATALB, with a convolution kernel of size [7 7 7], and plotted as a filled contour. Contour colours were plotted on a linear scale, and normalized such that the darkest shade represented the densest point on the histogram, while white represented zero density.

DNA

Fluorescently labelled DNAs were prepared as described in Chapter 3. Oligos used to construct simple double-stranded DNAs were purified on a reverse-phase C18 column (μRPC C2/C18, GE Healthcare, UK) running on the AKTA FPLC platform (GEHealthcare, UK), while oligos used to generate hairpins were purified via denaturing PAGE. Where necessary, labelled, single-stranded DNA samples were annealed in hybridization buffer (50 mM Tris-HCl pH 8.0, 1mM EDTA, 500 mM NaCl). DNA sequences are listed in the Chapter 4 Appendix, Fig. S1.

Klenow Fragment

Expression, purification, fluorescent labelling and characterization of the exo- Klenow fragment were carried out as previously described[24,25]. Briefly, Klenow fragment was labelled at positions 550 and 744 with ATTO647N and Cy3B respectively; the specificity of the labelling orientation was ≈88%. The extent of labelling was ≥ 70%.

Contributions

All text, figures and derivations were generated by J.P.T. Experiments on the Klenow Fragment were carried out by Yusdi Santoso, and labelled KF reagents provided by Catherine Joyce and Olga Potapova. Simulation software was provided by Johannes Hohlbein. Ideas and critical feedback were provided by Yusdi Santoso, Seamus J. Holden, Johannes Hohlbein and Achillefs N. Kapanidis.

References

[1] Henzler-Wildman KA, Thai V, Lei M, Ott M, Wolf-Watz M, et al. (2007) Intrinsic motions along an enzymatic reaction trajectory. *Nature* 450: 838-U813

[2] Henzler-Wildman K, Kern D (2007) Dynamic personalities of proteins. *Nature* 450: 964-972

[3] Santoso Y, Torella JP, Kapanidis AN (2010) Characterizing Single-Molecule FRET Dynamics with Probability Distribution Analysis. *Chemphyschem* 11: 2209-2219

[4] Kalinin S, Felekyan S, Valeri A, Seidel CAM (2008) Characterizing multiple molecular states in single-molecule multiparameter fluorescence detection by probability distribution analysis. *J Phys Chem B* 112: 8361-8374

[5] Nir E, Michalet X, Hamadani KM, Laurence TA, Neuhauser D, et al. (2006) Shot-noise limited single-molecule FRET histograms: Comparison between theory and experiments. *J Phys Chem B* 110: 22103-22124

[6] Kalinin S, Valeri A, Antonik M, Felekyan S, Seidel CAM (2010) Detection of Structural Dynamics by FRET: A Photon Distribution and Fluorescence Lifetime Analysis of Systems with Multiple States. *J Phys Chem B* 114: 7983-7995

[7] Chattopadhyay K, Elson EL, Frieden C (2005) The kinetics of conformational fluctuations in an unfolded protein measured by fluorescence methods. *Proc Natl Acad Sci U S A* 102: 2385-2389

[8] Torres T, Levitus M (2007) Measuring conformational dynamics: A new FCS-FRET approach. *J Phys Chem B* 111: 7392-7400

[9] Gurunathan K, Levitus M (2008) Applications of Fluorescence Correlation Spectroscopy to the Study of Nucleic Acid Conformational Dynamics. Progress in Nucleic Acid Research and Molecular Biology, Vol 82. pp. 33-69

[10] Bonnet G, Krichevsky O, Libchaber A (1998) Kinetics of conformational fluctuations in DNA hairpin-loops. *Proc Natl Acad Sci U S A* 95: 8602-8606

[11] Wallace MI, Ying LM, Balasubramanian S, Klenerman D (2000) FRET fluctuation spectroscopy: Exploring the conformational dynamics of a DNA hairpin loop. *J Phys Chem B* 104: 11551-11555

[12] Hess ST, Webb WW (2002) Focal volume optics and experimental artifacts in confocal fluorescence correlation spectroscopy. *Biophys J* 83: 2300-2317

[13] Kapanidis AN, Lee NK, Laurence TA, Doose S, Margeat E, et al. (2004) Fluorescence-aided molecule sorting: Analysis of structure and interactions by alternating-laser excitation of single molecules. *Proc Natl Acad Sci U S A* 101: 8936-8941

[14] Dertinger T, Pacheco V, von der Hocht I, Hartmann R, Gregor I, et al. (2007) Two-focus fluorescence correlation spectroscopy: A new tool for accurate and absolute diffusion measurements. *Chemphyschem* 8: 433-443

[15] Torella JP, Holden SJ, Santoso Y, Hohlbein J, Kapanidis AN (2010) Identifying molecular dynamics in single-molecule FRET experiments with Burst Variance Analysis. *Biophys J* (accepted):

[16] Kenney JF, Keeping ES (1951) The Distribution of the Standard Deviation. Mathematics of Statistics, Pt 2, 2nd ed. Princeton, NJ: Van Nostrand. pp. 170-173

[17] Abidi H (2007) Bonferroni and Sidak corrections for multiple comparisons. Encyclopedia of Measurement and Statistics. Thousand Oaks (CA): Sage. pp. 103-107.

[18] Ying LM, Wallace MI, Klenerman D (2001) Two-state model of conformational fluctuation in a DNA hairpin-loop. *Chem Phys Lett* 334: 145-150

[19] Antonik M, Felekyan S, Gaiduk A, Seidel CAM (2006) Separating structural heterogeneities from stochastic variations in fluorescence resonance energy transfer distributions via photon distribution analysis. *J Phys Chem B* 110: 6970-6978

[20] Kalinin S, Sisamakis E, Magennis SW, Felekyan S, Seidel CAM (2010) On the Origin of Broadening of Single-Molecule FRET Efficiency Distributions beyond Shot Noise Limits. *J Phys Chem B* 114: 6197-6206

[21] Santoso Y, Kapanidis AN (2009) Probing Biomolecular Structures and Dynamics of Single Molecules Using In-gel Alternating-Laser Excitation. *Anal Chem* 81: 9561-9570

[22] Jung JY, Van Orden A (2006) A three-state mechanism for DNA hairpin folding characterized by multiparameter fluorescence fluctuation spectroscopy. *J Am Chem Soc* 128: 1240-1249

[23] Van Orden A, Jung J (2008) Fluorescence correlation spectroscopy for probing the kinetics and mechanisms of DNA hairpin formation. *Biopolymers* 89: 1-16

[24] Santoso Y, Joyce CM, Potapova O, Le Reste L, Hohlbein J, et al. (2010) Conformational transitions in DNA polymerase I revealed by single-molecule FRET. *Proc Natl Acad Sci U S A* 107: 715-720

[25] Joyce CM, Potapova O, DeLucia AM, Huang XW, Basu VP, et al. (2008) Fingers-closing and other rapid conformational changes in DNA polymerase I (Klenow fragment) and their role in nucleotide selectivity. *Biochemistry (Mosc)* 47: 6103-6116

6. A ‘Green Fluorescent DNA’ for *in vivo* studies of molecular dynamics

In Chapters 3-5, we used confocal single-molecule fluorescence to study biomolecular dynamics in highly purified, *in vitro* systems. While *in vitro* experiments simplify the experimental task of studying biological molecules, however, they do so at the expense of cellular context. Here we discuss novel methods to study DNA- and DNA-protein dynamics in live *E. coli* using confocal single-molecule fluorescence. We address several experimental obstacles, such as the internalization of fluorescent DNAs into *E. coli*, and their destruction through photobleaching and nuclease-mediated degradation; we then demonstrate detection of these DNAs via single-molecule fluorescence and FRET in the cytosol of *E. coli.* Like Green Fluorescent Protein did for protein dynamics, these “Green Fluorescent DNAs” may enable the investigation of previously-inaccessible DNA and DNA-protein dynamics in live prokaryotic cells. We discuss these new possibilities both in the conclusion of this chapter and in Chapter 7, where we describe possible extensions to this method in studying nucleic acid dynamics in chromosomal DNA.

6.1 Introduction

Experiments carried out *in vitro* have the advantage that they are well-controlled, being performed in a homogeneous solution containing defined components at known concentrations. For these same reasons, however, *in vitro* experiments often ignore important factors that impact an enzyme's behaviour in its native cellular environment. Inside the cell, physiological concentrations of enzymes and substrates, of important cofactors, and of competitively or noncompetitively inhibitory molecules can profoundly alter molecular behaviour; intermolecular reactions may be dominated by spatial segregation or macromolecular crowding[1-3]; and other proteins may bind to, and modulate the behaviour or localization of these enzymes. Each of these factors may result in large changes in the behaviour of an enzyme *in vivo*, relative to what is known about the activity of the purified enzyme *in vitro*[1,2].

Although more technically challenging, experiments performed *in vivo* include the contributions of these factors, and can contribute greatly to the understanding of an enzyme's mechanism of action. For example, although many details about the mechanism and activity of *E. coli* DNA Polymerase I have been inferred from *in vitro* experiments, including stopped-flow, mutagenesis and single-molecule FRET studies of the Klenow Fragment (Chapters 4 and 5, Refs.[3-8]), how Pol I behaves in its natural cellular context remains poorly understood[2,9]. The cell contains a mixture of matched and mismatched nucleotides at physiological concentrations; how does this impact the distribution of open- and closed-conformations Pol I molecules occupy? How does this distribution, as well as factors like macromolecular crowding, impact the ability of Pol I to search for, bind to, and replicate an appropriate target strand of DNA? How do other protein or small-molecule factors present in the cell alter the kinetics of Pol I activity during DNA repair and replication, relative to what occurs *in vitro*? Answering these questions is necessary to understand the mechanism of

DNA replication and repair in *E. coli,* and requires performing detailed *in vivo* experiments to complement what is known *in vitro.*

In order to ask such questions, it is first necessary to be able to monitor the dynamics of biological molecules *in vivo*. In Chapter 1 we discussed the use of GFP to monitor molecular dynamics, specifically where they can be tracked via changes in diffusion (e.g. binding of a GFP-tagged protein to the chromosome or cell membrane[10,11]). In the case of conformational changes in proteins or DNA, however, as exemplified by our studies of KF in Chapters 4 and 5, no significant change in diffusion occurs. While in principle GFP fusions could be used to monitor conformational changes in KF via FRET, and GFP-based FRET has been monitored on the ensemble level in single cells[12], GFP would be a poor FRET probe for our purposes, due to its large size (~4 nm diameter), the inability to label internal (i.e. non-N/C terminal) amino acid positions on proteins of interest, and the inability to label DNA sequences of interest without substantially altering their sequence[11,13]. Monitoring single-molecule conformational changes *in vivo* therefore requires the ability to either fluorescently label biomolecules inside the cell with organic fluorophores, or to internalize already-labelled molecules into the cell.

Efforts to achieve the former have met with some success[14-19], especially in the development of less-perturbative SNAP tags for labelling the N- and C-termini of proteins[18], and bi-arsenical tags which can be used to label both protein termini and genetically modified loops on the protein's surface[19]. However, there remains no simple method for site-specifically labelling two arbitrary positions on a biological molecule of interest with distinct fluorescent probes *in vivo*. On the other hand, the latter method of internalizing already-labelled biological molecules into cells has the benefit of allowing the molecules of interest to be labelled site-specifically *in vitro*, but requires methods to internalize the molecules into the cell. Internalization of fluorescently labelled molecules for single-molecule FRET analysis

in vivo has been employed successfully in mammalian cells[20,21], largely due to the ease with which mammalian cells will endocytose, or can be microinjected with, the molecules of interest. For the prokaryotic cells studied in our and others' labs, however, analogous methods do not exist.

Here we develop a set of methods to internalize and observe single DNA molecules in *E. coli* via confocal fluorescence. Confocal fluorescence provides a high signal-to-noise ratio and good time resolution, making it an ideal tool for observing FRET dynamics against the strong background inherent in live-cell imaging. We demonstrate that fluorescent DNA molecules can be efficiently electroporated into *E. coli,* are photostable, and can be modified to resist nuclease degradation; we then demonstrate extended confocal detection of these molecules via single-molecule fluorescence and FRET in individual cells, replicating our capabilities in simple, *in vitro* experiments. This new method expands the toolbox for confocal single-molecule fluorescence, and may be considered a DNA analogue of Green Fluorescent Protein, or "Green Fluorescent DNA," which can be used to monitor DNA and DNA-protein dynamics in live prokaryotic cells.

6.2 Results and discussion

6.2.1 Electroporation of *E. coli* with fluorescent DNA

In order to build a Green Fluorescent DNA, the first challenge was to successfully and efficiently internalize them into *E. coli*. Although we tried a number of methods, including detergent-mediated internalization and sonication[22-24], electroporation emerged as the most efficient method for internalizing fluorescent DNAs into *E. coli.* Electroporation is a process in which an external field is applied to a cell suspension, producing aqueous pores in the cell membrane through which DNA and other small molecules can diffuse. While electroporation is commonly used on both prokaryotic and eukaryotic cells for transfection or transformation

with plasmid or linear DNA, we are not aware of previous attempts to electroporate fluorescent molecules into *E. coli* or other prokaryotes.

Principles of electroporation. Electroporation begins by placing a high-resistance (i.e. weakly ionic) cell suspension between two metal plates connected to a DC power supply and a capacitor (Fig. 1, left). The capacitor is charged, then discharged across the cell suspension, producing an initial electrical field of ~20 kV/cm which decays exponentially with a time constant of ~5 ms in most applications, though these two parameters can be varied to optimize electroporation efficiency[25]. Studies on cells undergoing electroporation have shown that, above a certain field strength, the cell suspension increases in conductance[26,27]; this is believed to result from structural rearrangements of the membrane causing the reversible opening of aqueous pores[28,29] (Fig. 1, right), through which biological molecules can travel. Following electroporation, these pores can either reseal or grow; in the former case, the cell will generally survive the electroporation procedure, while in the latter, the cell membrane will lyse, killing the cell. It is worth noting that, consistent with larger pores having both a greater chance of lysing the cell membrane and greater capacity to transport large biological molecules, there is an inverse relationship between the viability of the culture following electroporation, and the percent of surviving cells successfully electroporated with the molecule of interest[25].

Electroporation of* E. coli *with fluorescent DNA. To test whether we could introduce fluorescent DNAs into *E. coli* via electroporation, we first mixed 40 μL electrocompetent cells with 10 pmol of a Cy3B-labelled, double-stranded 64-bp DNA ($lac^{Cy3B(-5)}$; the sequences of this and other DNAs in this chapter are listed in the Appendix, Fig. S1), and electroporated the sample once at 18 kV/cm (Methods); in a control sample, we mixed the cells with DNA,

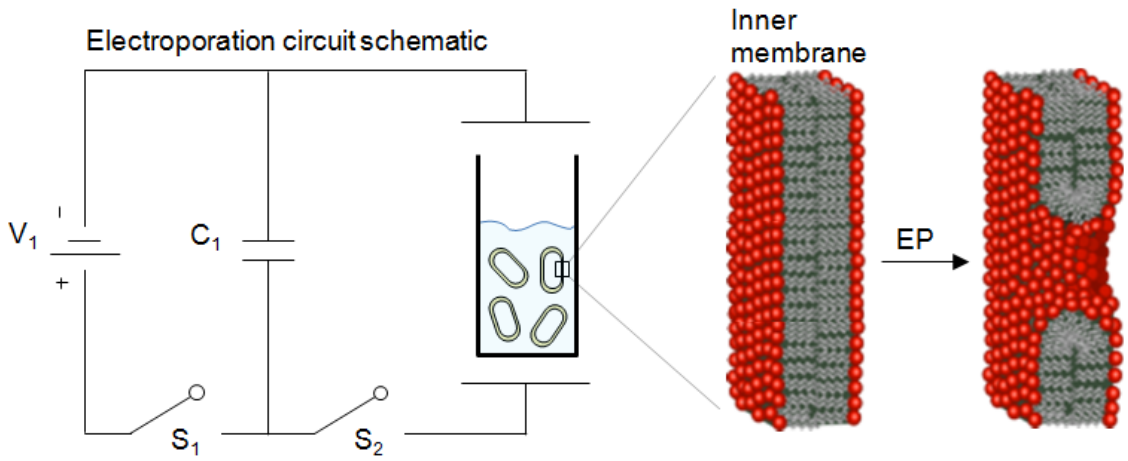


Figure 1. Schematic diagram of the electroporation apparatus and process. Circuit diagram of the electroporation apparatus (left). By closing S1, the capacitor (C1) is charged; closing S2 discharges the capacitor across a cuvette containing a suspension of cells, resulting in an electric field of ~20kV/cm that decays exponentially with time (time constant is typically ~ 5 ms). The electric field is thought to cause a reorientation of the lipid bilayer encapsulating the cell (cross-section shown at the right; "EP" indicates electroporation), producing transient pores that allow the free exchange of DNA and other small molecules between the cell's cytoplasm and the surrounding medium.

but did not electroporate. We then washed both samples and visualized them using widefield fluorescence microscopy, illuminating with a 532 nm laser at 200 μW (Methods), which was low enough to produce only negligible cellular autofluorescence. As expected, we observed strong fluorescence from the cells in the electroporated sample, but no apparent fluorescence in those cells that had simply been mixed with the DNA (Fig. 2A). The fluorescence intensity peaked in the interior, rather than at the edges of the cell, consistent with cytoplasmic, rather than periplasmic or membrane localization[30]. We also found that some cells in the electroporated population appeared not to be successfully electroporated, as they showed no appreciable fluorescence (Fig. 2A); this is consistent with the idea that, in some cells, pores do not grow large enough to permit the transport of DNA or other biomolecules[28].

A significant concern was whether the fluorescence we observed in the cells was genuinely dye-labelled DNA, or simply a fraction of free dye contaminating our DNA sample. We therefore asked whether we could introduce double-stranded DNAs labelled with a FRET pair into the cells; the ability to detect FRET from individual cells would indicate physical association between the two fluorophores, suggesting that the cell contained intact DNA. We electroporated cells with DNA constructs labelled with Cy3B and Att647N, and exhibiting either high or low FRET ($lac^{Cy3B(-15),A647N(-3)}$, $E^* \approx 0.79$, and $lac^{Cy3B(-15),A647N(+15)}$, $E^* \approx 0.23$, respectively), washed them, and visualized them in widefield mode (532 nm laser, 200 μW). As expected, we observed high or low FRET in the cells, according to the DNA internalized (Fig. 2B). Importantly, however, we note that this FRET could only be observed within the first few minutes of electroporation, and that the rate of FRET loss was independent of excitation power (and therefore not due to photobleaching); this suggested that some active process was physically separating the fluorophores. We return to this phenomenon in a later section of this chapter.

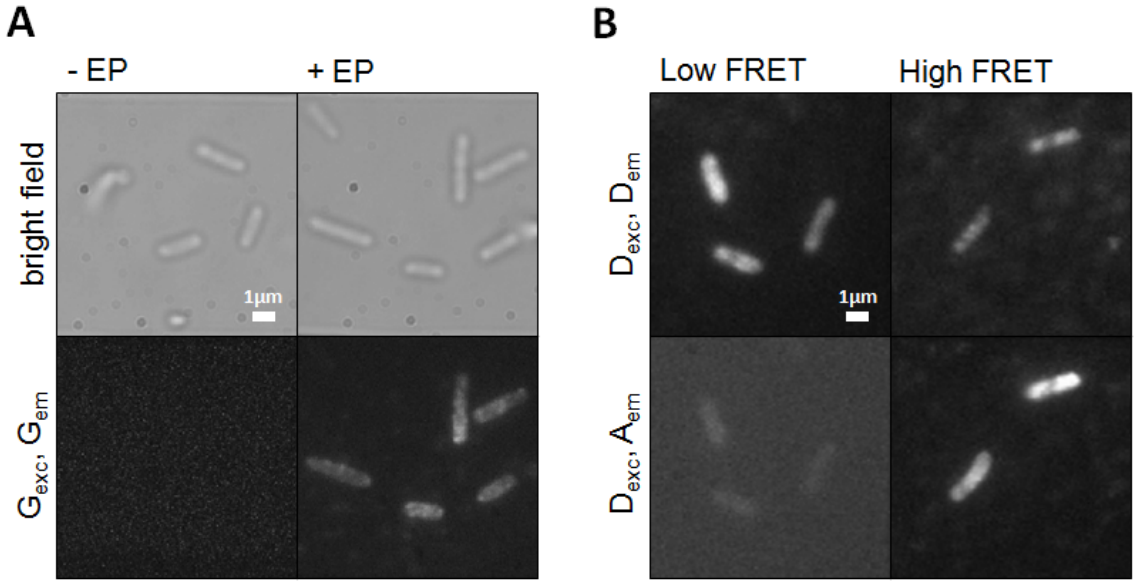


Figure 2. Electroporation of fluorescent DNA into *E. coli*. (A) Electroporation of Cy3B-labelled double-stranded DNA ($lac^{Cy3B(-5)}$) into *E. coli*. 40 µL of an electrocompetent cell suspension (Methods) was mixed with Cy3B-labelled DNA and either electroporated (right, +EP) or not electroporated (left, -EP), then washed and visualized via bright field (top) and widefield fluorescence microscopy (bottom) with a 532 nm laser at 200 µW. Fluorescent DNAs were internalized only when the cells were electroporated. Not all cells are successfully electroporated, however (e.g. the cell in upper-left corner of +EP sample). We note that, in this panel, we allowed brightness to be automatically adjusted by the acquisition software; the "-EP" sample is therefore brightened relative to "+EP", but still shows no apparent fluorescent signal (i.e., autofluorescence was negligible at 200 µW). (B) *E. coli* were electroporated with either low- or high-FRET double-stranded DNAs ($lac^{Cy3B(-15),A647N(-3)}$, $E^* \approx 0.79$, and $lac^{Cy3B(-15),A647N(+15)}$, $E^* \approx 0.23$), and excited as in panel A. In the low-FRET sample, most of the emitted fluorescence is in the D_{em} channel; we again allowed the software to automatically adjust image brightness, making the low amount of A_{em} fluorescence visible against background. Unlike the low-FRET sample, in the high-FRET sample most of the emitted fluorescence is in the A_{em} channel, as expected (here the two channels were set to equal brightness levels for comparison purposes).

Because the pores generated by electroporation allow biological molecules to cross the membrane, but can also lead to cell lysis, we measured both the percentage of cells exhibiting fluorescence following electroporation, and their viability. We first excited the cells in widefield mode at 200 μW (at which autofluorescence and scattering are negligible) and measured the percentage of cells exhibiting visible fluorescence. For $N = 3$ experiments, each of which analyzed ~50 cells, we determined that 11 ± 4% of all cells were successfully electroporated with DNA.

Experiments performed by Robert Crawford determined cell viability, measured as the percentage of viable cells following electroporation at a field strength of 18kV/cm, using the membrane *BacLight* LIVE/DEAD cell-staining kit (Invitrogen, Carlsbad, CA; kit was used according to manufacturers' instructions). The LIVE/DEAD assay identifies dead cells as those permeable to the DNA-intercalating dye SYTO 9. While dead cells should be successfully stained with SYTO 9, it is also possible that, following electroporation, live cells with sufficient membrane damage may take up SYTO 9. LIVE/DEAD staining may therefore overestimate the number of dead cells in an electroporated sample, and provide a lower bound on their viability. Using the LIVE/DEAD assay, we measured the viability of $N = 3$ samples following electroporation as 81 ± 8% relative to a non-electroporated control; this was consistent with an independent colony-counting assay for viability (70 ± 6%), and somewhat higher than the ~50% viability typical of standard electroporation protocols[25].

Experiments currently being performed by Robert Crawford will determine whether successfully electroporated, fluorescent cells are also viable following electroporation. While we do not yet have good statistics regarding this question, preliminary experiments have suggested that most, if not all fluorescent cells, are also viable. This may be because the cells killed by electroporation have large pores in their membranes, allowing any internalized fluorescent species to "leak" out.

6.2.2 Characterization of fluorescent DNA inside *E. coli*

***Fluorescence correlation spectroscopy of probes* in vivo.** Fluorescence correlation spectroscopy (FCS) is a small-ensemble technique that provides information on the concentration and diffusive properties of fluorescent molecules (Chapter 2). Using FCS, we asked two questions: whether the internalized probes were diffusing freely or binding to structures within the cell (determined as an increase in diffusion time), and how many molecules were internalized per electroporation event on average.

We characterized our confocal spot using the standard dye Rhodamine 6G; confocal spot geometry and fitting details are provided in Methods. Following characterization, we electroporated the Cy3B-labelled DNA ($lac^{Cy3B(-5)}$) into *E. coli* and asked how it was interacting with the cell by comparing its diffusion time *in vitro* and *in vivo*. *In vitro*, we used FCS to determine the diffusion times of two species we expected to find *in vivo*: Cy3B-labelled dsDNA, and free Cy3B dye (Fig. 3A). We expected to find species similar to free Cy3B dye *in vivo* because, as mentioned previously, DNA appears to degrade inside the cells. We measured diffusion times of 273 ± 4 μs and 2.36 ± 0.03 ms for Cy3B and Cy3B-DNA respectively. Previous studies have shown that, compared to those measured *in vitro*, diffusion coefficients measured for GFP *in vivo* are slowed by a factor of ~10. While the size of this effect can vary significantly with proteins and tags attached to GFP, the osmolarity of the medium, and even the level of GFP expression[31-34], for simplicity we assumed that the diffusion times of our DNA and dye *in vivo* would be slowed by a factor of 10, to 2.7 ms for Cy3B dye, and 24 ms for Cy3B-DNA.

We then performed FCS on a cell electroporated with Cy3B-DNA. Unlike the autocorrelation curves produced *in vitro*, which decayed at clear timescales (Fig. 3A), the autocorrelation curve from a representative cell showed a steady decay over several orders of magnitude (Fig. 3B). We first assumed the existence of two species with diffusion times of

2.7 ms and 24 ms, representing the Cy3B and Cy3B-DNA species we expected; we therefore fixed the values of the y-intercept and diffusion times of the two species, and performed a one-parameter fit for the ratio of the two species (Fig. 3B, blue line). The fit using this approach was poor, as it could not account for the significant correlation at times greater than 10 ms. We also attempted this fit, including a term describing anomalous diffusion (see Methods), which is characteristic of diffusion in cells due to macromolecular crowding[13]; however, this did not substantially improve the fit (data not shown). We therefore performed a three-parameter fit, assuming the existence of an additional species of unknown diffusion time (the three parameters include the two concentration ratios between the three species, and the diffusion time of the third species). The three-parameter fit achieved good agreement with the autocorrelation curve, and suggested a ratio of 22:25:53 Cy3B:Cy3B-DNA:X, where X is the additional species.

The best-fit diffusion time for the third species was large (173 ms). While the diffusion of DNA inside bacterial cells has not previously been studied, this slow diffusion may be due to binding of our probe to DNA-binding proteins in the *E. coli* cytosol. Significant slowing of diffusion has been observed in previous FRAP and FCS studies of fluorescent proteins in *E. coli*, for instance in the case of GFP fused to histidine tags and to DNA-binding proteins such as TetR and the H-NS regulator[33,35]. For our probe, possible binding partners include large nucleases such as the exonuclease V complex, which would strongly slow diffusion; the particular DNA we use ($lac^{Cy3B(-5)}$) also contains a consensus *lac* promoter, which is likely to bind strongly to RNA Polymerase inside the cell, slowing its diffusion. Future experiments on a similar construct, mutated to weaken its affinity for RNAP, would help discern whether this slowly-diffusing species is indeed due to the binding of RNA Polymerase to this promoter.

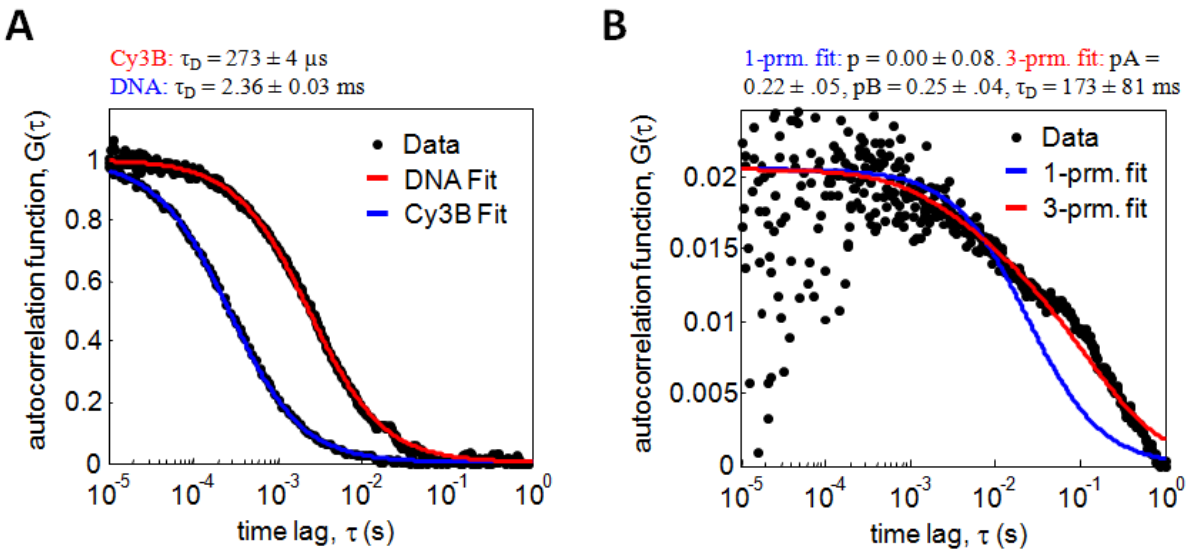


Figure 3. Fluorescence correlation spectroscopy of Cy3B-DNA *in vitro* and *in vivo*. (A) Normalized autocorrelation curves of free Cy3B dye and Cy3B-labelled DNA ($lac^{Cy3B(-5)}$).The Cy3B-DNA species diffuses about 8x more slowly than free Cy3B dye, with diffusion times through the confocal volume of 273 μs and 2.36 ms for the dye and DNA respectively. (B) Autocorrelation curve for a cell electroporated with Cy3B-labelled DNA. The y-intercept of the plot is equal to $1/\langle N \rangle$, where $\langle N \rangle$ is the average number of fluorophores in the confocal volume at a time; given the dimensions of the confocal volume and a typical *E. coli* cell (Methods), the cell analyzed contains ~98 fluorescent molecules. To fit the autocorrelation curve, we assumed the existence of two species: intact Cy3B-DNA and free Cy3B dye produced through DNA degradation; we assume these species to have a diffusion time 10x that of Cy3B-DNA and Cy3B dye *in vitro*, respectively. A one-parameter fit (where the parameter indicates the relative amount of each species) failed to achieve a good fit to the data (blue line), as it showed too significant a decay at long correlation times (> 10 ms). In contrast, a three-parameter fit assuming the existence of an additional species achieved a reasonable fit (red line). The third species is slowly-diffusing (τ_D = 173 ms), and may arise due to the binding of DNA molecules to cytosolic proteins and other structures.

Finally, we asked how many fluorescent DNAs were electroporated into each cell on average. This value can be obtained by determining the mean number of molecules in the confocal spot at any given time (equal to 1 / G(0), the y-intercept of the autocorrelation function G(τ)), and dividing it by f_C, the fraction of the cell occupied by the confocal spot (f_C is calculated as 0.48, see Methods). We obtained 106 ± 21 fluorescent molecules per cell (for N = 4 cells) for the given electroporation protocol (Methods). In principle, it should be possible to adjust this number by increasing or decreasing the concentration of fluorescent DNA in the cell suspension during electroporation. We note that this number of molecules represents all fluorescent species in the cell, whether intact, degraded, or bound to some cellular molecule. Assuming the ratios from the FCS experiment are accurate, this suggests that the cell analyzed in Fig. 3B contained roughly 23 molecules of free Cy3B, 27 molecules of Cy3B-DNA, and 56 slowly diffusing (and possibly protein-bound) Cy3B-DNA molecules.

Photostability of organic fluorophores **in vivo.** We next asked whether the internalized fluorescent DNAs were photostable. Interactions with specific ions, reducing agents and other small molecules are known to alter the photostability of organic fluorophores[36]; however, it is unclear how the complex composition of the *E. coli* cytoplasm might impact the stability of organic fluorophores, and whether they would be more or less stable than biological fluorophores such as GFP in this environment.

To determine the photostability of organic fluorophores *in vivo*, we labelled double-stranded DNAs with either Cy3B or Atto647N ($lac^{Cy3B(-15)}$ or $lac^{A647N(+15)}$) and electroporated them into *E. coli.* For comparison with a fluorescent protein, we also expressed enhanced YFP (eYFP) from a plasmid (Methods). We then photobleached the cells under widefield illumination with a 532 nm laser, and monitored the loss of fluorescence over time (Fig. 4A). Because the rate of photobleaching is linearly dependent on the excitation rate (see

Chapter 2), we used a laser power normalized to each fluorophore's molar extinction coefficient at 532 nm (ε_{532}). Specifically, we excited Cy3B, ε_{532} = 150,000 M^{-1} cm^{-1} at 0.8 mW; Atto647N, with ε_{532} = 6,645 M^{-1} cm^{-1}, at 15.7 mW; and eYFP, ε_{532} = 30,490 M^{-1} cm^{-1}, at 3.5 mW.

To determine the photobleaching lifetime of each fluorescent DNA, we extracted the time constant from a single exponential fit to the fluorescence decay (Fig. 4A) for *N = 3* cells (Methods). We found that Cy3B- and Atto647N-labelled DNA were ~2x and ~6x more photostable than eYFP in *E. coli*, respectively (Fig. 4B). To provide an intuitive sense of the photostability of Cy3B and Atto647N, we note that in a previous publication[10], the fast-maturing fluorescent protein variant, Venus-YFP[37], was used to visualize single molecules in *E. coli* for up to several seconds; as the photostability of eYFP is ~4x greater than that of Venus-YFP[38], organic fluorophores may afford very significant improvements (~8-24 fold) in photostability over those fluorescent proteins currently in use in *in vivo* single-molecule studies.

6.2.3 Degradation of exogenous DNAs in *E. coli*

We noted earlier that when we electroporated double-stranded DNAs exhibiting FRET in *E. coli*, the FRET would disappear within minutes, likely due to physical separation of the two fluorophores. Indeed, in our experience, the most significant barrier to constructing a useful "Green Fluorescent DNA" is preventing its degradation *in vivo*. Due to an evolutionary history of parasitism by DNA viruses[39], and the need to repair breaks in DNA, *E. coli* and other microorganisms express multiple enzymes intended to degrade cytosolic DNA, particularly those recognized as foreign or with single-stranded 5' or 3' ends[40]. These enzymes include promiscuous and site-specific restriction endonucleases, and 3'- and 5'-exonucleases[40-42].

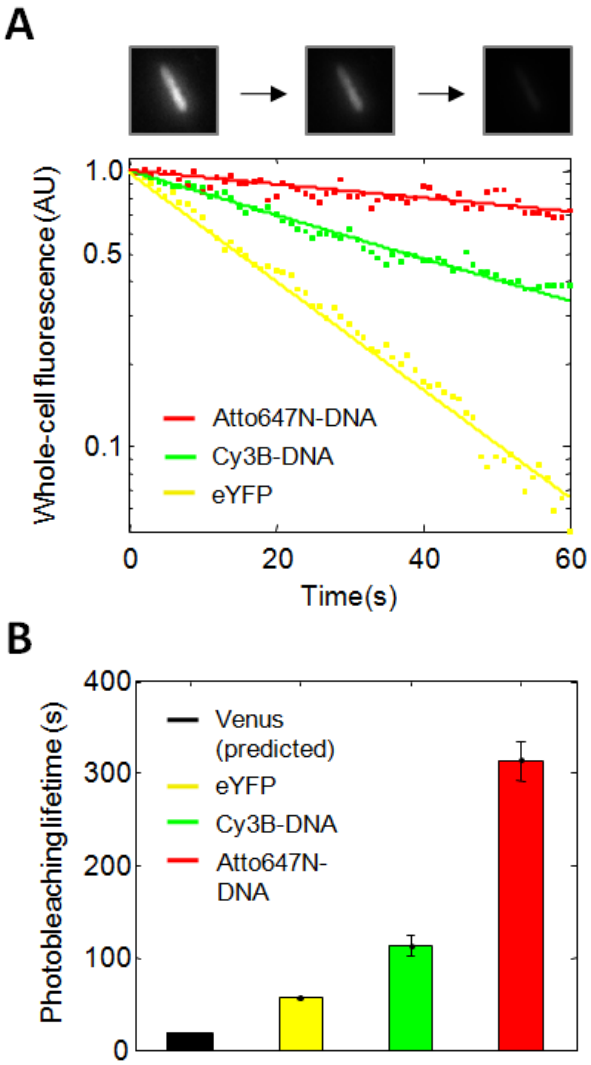


Figure 4. Photobleaching of organic fluorophores and fluorescent proteins in *E. coli*. We electroporated either Cy3B-labelled (green) or Atto647N-labelled (red) double-stranded DNA ($lac^{Cy3B(-15)}$, $lac^{A647N(+15)}$) into *E. coli*, or expressed eYFP from a plasmid (Methods). We then photobleached each population of cells with a 532 nm laser in widefield mode, with powers normalized to the extinction coefficient at 532 nm for each fluorophore (powers and extinction coefficients are listed in main text). (A) Representative fluorescence decay curves from a single cell in each experiment; photobleaching lifetime was measured by performing a single-exponential fit to the data (lines). (B) Photobleaching lifetimes for eYFP, Cy3B-DNA and Atto647N-DNA, calculated from $N = 3$ cells each. eYFP had a photobleaching lifetime of 57 ± 3 s, Cy3B-DNA a lifetime of 112 ± 11 s, and Atto647N-DNA a lifetime of 313 ± 22 s. Also plotted is the theoretical photobleaching lifetime of Venus-YFP, based on the relative photostabilities of eYFP and Venus, reported elsewhere[38].

To demonstrate that nuclease activity was likely responsible for the loss of FRET in our experiments, we first incubated a high-FRET dsDNA construct ($lac^{Cy3B(-15),A647N(-3)}$) in whole-cell lysates of *E. coli,* and monitored the loss of FRET over time via ensemble fluorescence (Fig. 5, top; Methods). We repeated the experiment under two conditions: in the presence of 5 mM Mg^{2+}, an essential cofactor for nuclease activity, or in the presence of 10 mM EDTA, which chelates free Mg^{2+}. In the Mg^{2+}-containing buffer, FRET was almost entirely lost at the end of 60 minutes (solid line). In the buffer containing EDTA, however, no appreciable loss of FRET occurred over the same time period (dashed line).

We also analyzed this degradation reaction via denaturing polyacrylamide gel electrophoresis (PAGE), which separates single strands of DNA on the basis of size and charge, and can therefore directly detect the presence of partially degraded DNA (Fig. 5, bottom); excitation of the acceptor fluorophore was used to visualize one of the DNA strands on the gel. The zero time point shows a sample containing 10 mM EDTA to ensure no degradation took place; the top band on the gel (magenta arrow) shows the intact single strand bearing the acceptor fluorophore. By 10 minutes in the presence of Mg^{2+}, most of this band had degraded at least partially (note loss of highest-MW band, and appearance of lower MW bands). The appearance of a 'fuzzy' band below that of the full-length strand (cyan arrow) is likely due to the accumulation of free (or mono/di-nucleotide-conjugated) Atto647N dye which, though much smaller than DNA, does not have as strong an electrostatic charge, and therefore migrates more slowly during gel electrophoresis (we note that the band's "fuzzy" appearance is characteristic of small species, which diffuse more freely through the PAGE gel than do intact DNAs). By the 60-minute time point, where FRET had been almost entirely lost, the PAGE results showed essentially all free Atto647N. These results are consistent with the notion that DNA degradation is responsible for the loss of FRET in both our *in vitro* assay and in electroporation experiments, and points to the need

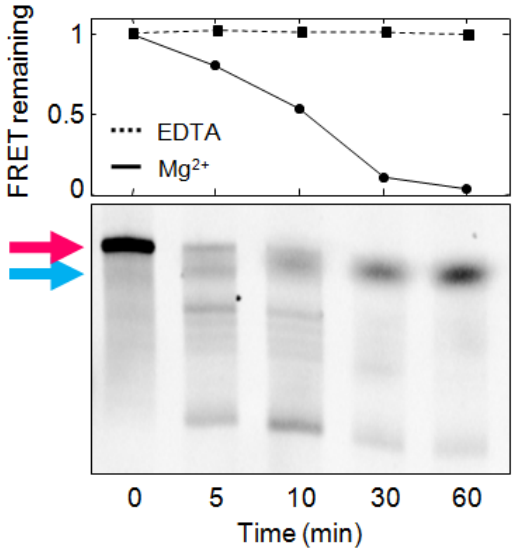


Figure 5. Nuclease activity limits the lifetime of DNAs in *E. coli*. A 5 mg/mL whole-cell lysate of *E. coli* strain DH5α was prepared, mixed with a high-FRET DNA ($lac^{Cy3B(-15),A647N(-3)}$), and the loss of FRET monitored over time with a fluorimeter (top; Methods). Whereas FRET was lost quickly in a buffer containing 5mM Mg^{2+} (solid line), a cofactor necessary for nuclease activity, lysates prepared in a buffer containing 10mM EDTA (dashed line), which stoichiometrically chelates Mg^{2+}, showed no apparent degradation over the course of the experiment. We note that in this plot, FRET was normalized to 1.0 for the first time point, and to 0 for a time point taken at 2 hours, after FRET had ceased to decline. Samples of the DNA degradation reaction were also loaded on a denaturing PAGE gel (bottom, Methods) and visualized via 635 nm excitation of the acceptor fluorophore. Whereas a zero time-point sample containing EDTA showed a single, distinct band (magenta arrow), incubation with Mg^{2+} revealed clear DNA degradation within 5 minutes, and accumulation of another species (cyan arrow, which likely represents an Atto647N-nucleotide conjugate) on the timescale of FRET loss. This supported the idea that the loss of FRET in both whole-cell lysates, and individual cells, is due to primarily to nuclease activity.

for methods to stabilize the DNA electroporated into cells, if extended observation is to be achieved.

Genetic and DNA modifications to prevent DNA degradation. To minimize nuclease degradation of our probes in *E. coli,* we pursued two strategies: genetic modifications of the host, and chemical modifications of the DNA (Fig. 6A). The strains used in this study include:

- DH5α, a K12-derived "cloning" strain of *E. coli* with a mutation in the *endA* endonuclease, which decreases double-stranded DNA degradation
- MG1655 / pKD46, which expresses the Gam protein from phage lambda. Gam inhibits Exonuclease V (RecBCD), which has several activities (exonuclease, helicase, nicking) that impair DNA stability in *E. coli*.
- SURE2 cells from Stratagene, which lack restriction systems (*McrA*$^-$*, McrCB*$^-$*, McrF*$^-$*, Mrr*$^-$*, HsdR*$^-$), endonuclease A (*endA*) and additional exonucleases (*recB, recJ*).

DNA modifications are meant to minimize 5'- and 3'-end-dependent exonuclease activity and include:

- Hairpins, which join the top and bottom strands of DNA, preventing the exposure of one set of 5' and 3' ends of the DNA to exonucleases.
- Dumbbell DNAs: DNAs containing two hairpin ends, such that they have no free 5' or 3' ends[22].
- Unnatural phosphorothioate (PTO) DNA backbones, linking the first five nucleotides (for a total of four PTO linkages) at each end of the DNA, on both strands. Nucleases hydrolyze PTO backbones much more slowly than natural phosphate backbones[43].

Modified and unmodified DNA sequences are shown in the Appendix (Fig. S1). We incubated each of the DNAs in the whole-cell lysate of a given host strain (normalized to

5 mg/mL protein content), and observed roughly exponential decays in FRET, as measured by ensemble fluorescence (Fig. 6B; Methods). We fit each FRET decay curve to an exponential function and extracted the rate of FRET loss, which we assume to be equivalent to the rate of DNA degradation. A summary of extracted DNA degradation rates is shown in Fig. 6A.

The simplest finding was that both genetic and DNA modifications succeeded in abating the degradation of our fluorescent probes, with individual interventions generally achieving a > 40% reduction in the rate of DNA degradation (Fig. 6A).

Among genetic modifications, we were surprised to find that even with deletions in several major restriction systems and nucleases, the SURE2 strain achieved only a 65% decrease in nuclease activity on unmodified double-stranded DNA, relative to the DH5α strain. Interestingly, the SURE2 strain is *RecB*$^{-}$, and performed similarly to the strain expressing Gam (MG1655 / pKD46), which inhibits RecBCD; this is consistent with the possibility of RecBCD being the major limiting factor for the stability of internalized DNA. While it is conceivable that further genetic modifications could better attenuate nuclease activities, such activities are necessary for proper cell growth and division, and cells lacking one or more nucleases are generally less fit than their wild-type counterparts[44,45]. We therefore turned our attention to DNA modifications that increase resistance to nuclease activity.

Among DNA modifications, the dumbbell ($lac^{DB,Cy3B(-5),A647N(-3)}$) and PTO/hairpin hybrid ($lac^{HP,Cy3B(-5),A647N(-3)}$) both decreased the degradation rate, by about 2-3x relative to the unmodified DNA in the DH5α lysate (Fig. 6A). As both interventions limit the access of exonucleases to the DNA, this increase in stability is likely due to decreased exonuclease activity on the DNA constructs. Interestingly, incubating dumbbell DNAs in either SURE2 or Gam-expressing lysates had a synergistic effect on inhibition of nuclease activity. For

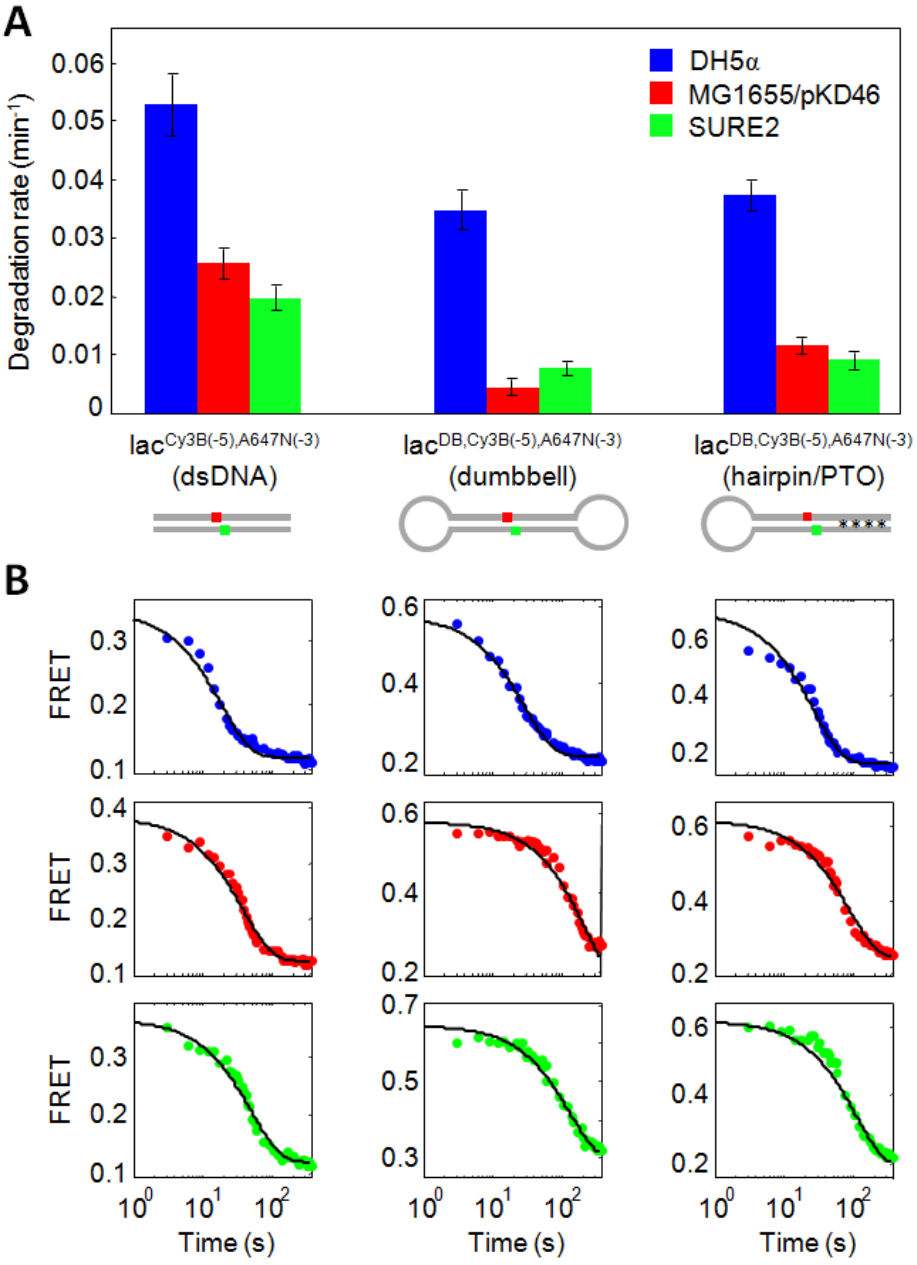

A
Degradation rate (min-1)
0.06
0.05
0.04
0.03
0.02
0.01
0
DH5α
MG1655/pKD46
SURE2
lacCy3B(-5),A647N(-3)
(dsDNA)
lacDB,Cy3B(-5),A647N(-3)
(dumbbell)
lacDB,Cy3B(-5),A647N(-3)
(hairpin/PTO)
B
FRET
Time (s)

Figure 6. Degradation of different DNA constructs in lysates of nuclease-attenuated *E. coli*. (A) First-order kinetic rates of DNA degradation in *E. coli* whole-cell lysates, as measured by ensemble FRET (Methods). Both chemical modifications of DNA, and genetic modifications of the host strain, were capable of attenuating nuclease activity. Relative to the degradation rate of unmodified DNA in a lysate of the standard cloning strain DH5α, a 10-fold increase in DNA stability was achieved by using dumbbell DNA in the MG1655/pKD46 lysate. (B) The degradation rates in panel A were determined by monitoring the loss of FRET over time from each DNA construct, in a given cell lysate, via ensemble fluorescence (Methods); these FRET decay curves were then fit to a single-exponential decay model, $E(t) = a\exp(-bt)$, where b is the degradation rate. Error bars in Panel A were determined as 95% confidence intervals on b. Representative degradation experiments and their exponential fits are shown below the relevant DNA construct in panel A; blue, red and green plots represent degradation in lysates of DH5α, MG1655/pKD46 and SURE 2 respectively, as in panel A.

instance, the dumbbell is degraded ~75% as quickly as unmodified DNA in DH5α lysates, and unmodified DNA in a Gam-expressing lysate is degraded ~ 50% as quickly as it is in the DH5α lysate. While we should therefore expect dumbbells in a Gam-expressing lysate to degrade at 75% * 50% ≈33% as quickly as unmodified DNA in the DH5α lysate, it actually degrades 10% as quickly. It is unclear whether this synergistic effect is due to both DNA and host interventions reducing exonuclease activity, or because inhibition of RecBCD also reduces complementary helicase and DNA-nicking activities, which facilitate access of the exonucleases to the DNA.

6.2.4 Detection of single DNA molecules in *E. coli*

Having internalized fluorescent DNAs into *E. coli*, ensured that they are photostable, and stabilized them to nuclease degradation, we finally asked whether we could successfully detect single fluorescent DNAs inside *E. coli*, and extract descriptive statistics such as stoichiometry and FRET efficiency.

Single-molecule fluorescence and FRET. We first tested whether we could detect fluorescent dsDNA ($lac^{Cy3B(-15)}$) electroporated into *E. coli*. Following electroporation, washing and immobilization on an agarose pad (Methods), we identified a given *E. coli* via brightfield microscopy or widefield fluorescence, switched to confocal mode, and centred the confocal spot on it to monitor diffusing molecules (Fig. 7A). To demonstrate confocal detection of internalized fluorescent DNAs, we identified two nearby cells: one of which was successfully electroporated, and the other of which (which we refer to as the 'blank') was not (Fig. 7B). We first analyzed the 'blank' cell via confocal fluorescence (532 nm laser, 150 μW); a representative time trace is shown in Fig. 7C (top). While the blank cell gave rise to substantial background in the $G_{ex}R_{em}$ channel, this background was stable and therefore

likely due to scattering, rather than autofluorescence; moreover, the background did not produce any apparent bursts. In contrast, the fluorescent cell (Fig. 7C, bottom) produced many large bursts (shaded grey), which were successfully detected by our modified burst search software (Methods). While the bursts shown here lasted from 5 – 25 ms, consistent with the diffusion time expected for Cy3B-labelled DNA, longer bursts were also identified on the order of 100 ms, consistent with the long-timescale correlations observed in Fig. 3B.

We also attempted to detect FRET from fluorescent DNAs inside *E. coli*. In addition to indicating that the molecules observed are intact, and not degraded by nuclease or other activities, successful detection of FRET is necessary for analyzing conformational dynamics in live cells. We therefore repeated the experiment by electroporating cells with a high-FRET DNA construct (Std45-90; sequence given in Appendix, Fig. S1). As expected, confocal analysis of the cell revealed large bursts exhibiting high FRET ($E^* \approx 0.92$; a representative cell trace is shown in Fig. 7D); this is consistent with the $E^* \approx 0.90$ observed *in vitro* for the same DNA construct (data not shown).

***Improving detection statistics via chemical elongation of* E. coli.** A major challenge of performing confocal fluorescence in *E. coli* is their small volume (~0.5 fL). Because a 532 nm diffraction-limited confocal spot is at least 0.1 fL in volume, at least 10-20% of the cell is illuminated in any confocal fluorescence experiment. As a result, confocal single-molecule analysis can only reasonably be performed if there are fewer than 10 fluorescent molecules in the cell at a given time (corresponding to a concentration of ~10 nM). Although many cells can be analyzed to obtain descriptive statistics about a molecule of interest, it may be desirable, both practically and biologically, to observe a large number of molecules in a single cell.

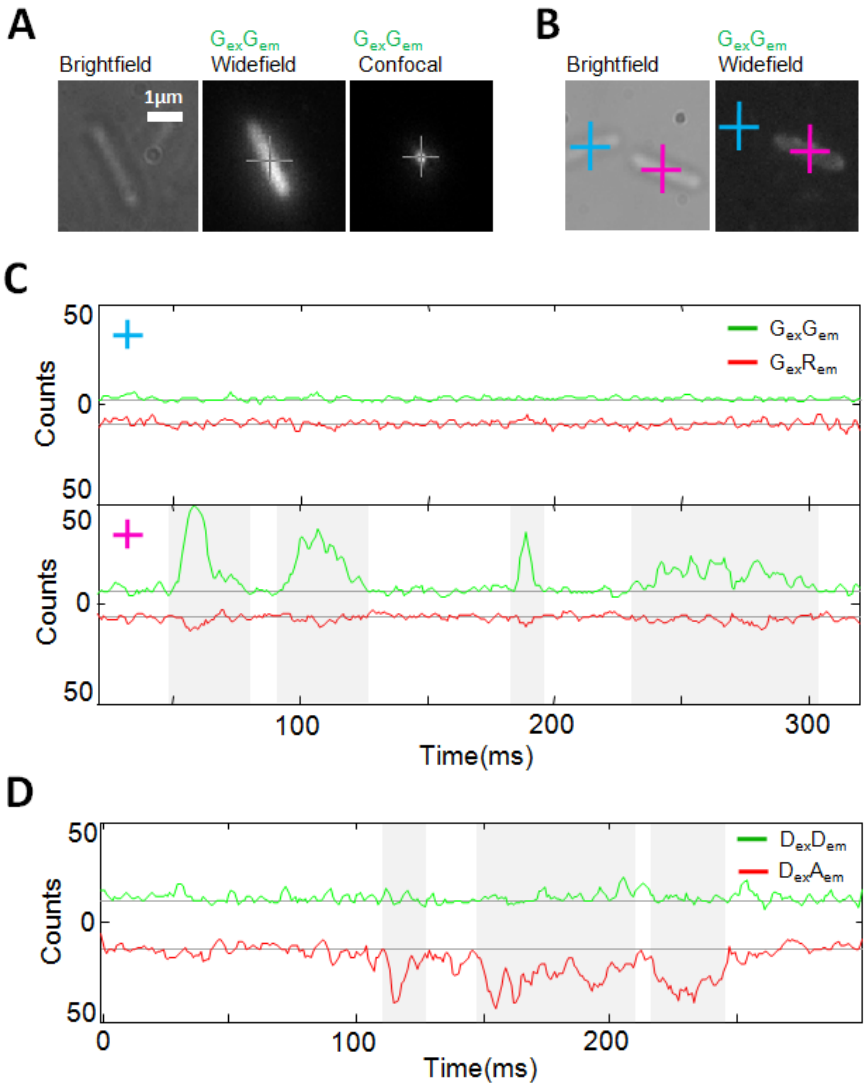

A
Brightfield
GexGem
Widefield
GexGem
Confocal
1μm
B
Brightfield
GexGem
Widefield
C
50
0
50
Counts
GexGem
GexRem
50
0
50
Counts
100
200
300
Time(ms)
D
50
0
50
Counts
DexDem
DexAem
0
100
200
Time(ms)

Figure 7. Detection of single-molecule fluorescence and FRET in E. coli. (A) To monitor single molecules in *E. coli* using confocal fluorescence, we identify a fluorescent cell in widefield mode (left), place the confocal spot over it (crosshairs), then switch to confocal mode (right), illuminating only a ~0.2 fL volume in the cell. (B) Brightfield and $G_{ex}G_{em}$ widefield images of two cells, one of which was successfully electroporated with $lac^{Cy3B(-15)}$ (magenta crosshairs), the other of which, a 'blank' cell, was not successfully electroporated (cyan crosshairs). Crosshair locations indicate the positions at which cells were analyzed via confocal fluorescence. (C) 1 ms-binned time traces of fluorescence during confocal analysis of the blank (top) and successfully-electroporated (bottom) cells from panel B, smoothed over 3 ms; tick marks on time axis indicate 50 ms time intervals. Mean background values in each channel are indicated by a grey line. In the blank cell, background is significant, but there are no apparent bursts of fluorescence in either channel. In the electroporated cell, however, large fluorescent bursts can easily be distinguished above background (bottom panel, grey shaded regions). (D) Fluorescence time trace of a cell electroporated with a high-FRET dsDNA construct (Std45-90); the sample is excited solely at the donor-excitation wavelength (532 nm) but gives rise to bursts in the A_{em} channel, consistent with a high-FRET species diffusing through the confocal volume.

Observing a greater number of molecules per cell requires employing one of two strategies: either decreasing the volume of the confocal spot, or increasing the volume of the cell. While a smaller confocal volume is not possible, it is possible to substantially increase the volume of *E. coli* via mild antibiotic treatments. Elongated cells have been used to study cell signalling dynamics[46] and ion channels[47,48], and to perform fluorescence recovery after photobleaching (FRAP) experiments on *E. coli*[32,33].

Treatment of growing *E. coli* with sub-inhibitory concentrations of the *β*-lactam antibiotic cephalexin results in defects in cell wall biosynthesis, causing cells to elongate up to 100x their normal length without dividing; this results in highly elongated cells with a continuous cytoplasmic space[32]. In addition to providing linear increases in volume with increasing length, elongation also greatly increases confocal observation times, as the mean time for a molecule to diffuse from one end of a cell, to a confocal spot at the other end, increases as the square of cell length.

We prepared electrocompetent, cephalexin-elongated cells (Fig. 8A), and electroporated them with a high-FRET, double-stranded DNA construct ($lac^{Cy3B(-15),A647N(-3)}$; $E^* \approx 0.79$ in *in vitro* experiments). We then monitored individual cells for up to 10 minutes in confocal mode using ALEX spectroscopy, and analyzed the data with a time-binned burst search method (Methods). In a representative cell, we were able to detect over 1000 bursts in the first four minutes; we plotted these bursts on a 2D ALEX contour plot, with apparent FRET, E^*, on the x-axis, and stoichiometry, S, on the y-axis (Fig. 8B). On the plot, we were able to identify the three populations typically expected in a single-molecule FRET experiment: an intermediate-FRET, intermediate-S species corresponding to the intact DNA, and donor- (high S, low E^*) and acceptor-only (low S, high E^*) populations, arising from nuclease-degraded or partially photobleached fluorescent DNA probes. Representative bursts from each subpopulation are shown in Fig. 8C.

As a simple test of whether the DNA behaved similarly *in vivo* to how it behaves *in vitro,* we plotted a 1D FRET histogram of the intermediate-S subpopulation alone (Fig. 8B, top), and asked whether the mean E^* of this population was consistent with the E^* observed *in vitro*. Surprisingly, although *in vitro* the $lac^{Cy3B(-15),A647N(-3)}$ DNA has an E^* of 0.79, the FRET species observed *in vivo* had an E^* of 0.62 (obtained by Gaussian fitting). While the origin of this difference is unclear, three possible reasons are: (*i*) that the difference can be accounted for by simple background considerations – though as we subtract background in our *in vivo* burst detection algorithm, this seems unlikely; (*ii*) that interactions of one or both of the fluorescent dyes (Cy3B and Atto647N) with the *E. coli* cytoplasm caused a change in dye spectrum or quantum yield, altering the value of E^*; (*iii*) that interactions of the DNA with the *E. coli* cytoplasm , and/or DNA-binding proteins inside it, altered the structure of the DNA. Future studies monitoring the behaviour of different DNA constructs and fluorophores *in vivo* will help determine whether the shift in E^* values is due to DNA-protein binding, quenching of the fluorescent DNA by cellular proteins and/or small molecules, or other interactions.

This discrepancy between the behaviour of DNA *in vitro* and *in vivo* (as well as the slow diffusion observed in FCS experiments) demonstrates the need to investigate biomolecular structures and dynamics *in vivo*. While further characterization is needed, overall our work demonstrates the ability to internalize fluorescent DNA probes into *E. coli*, and detect them on the basis of stoichiometry and FRET. With further characterization, this method should make it possible to monitor DNA and DNA-protein dynamics *in vivo.*

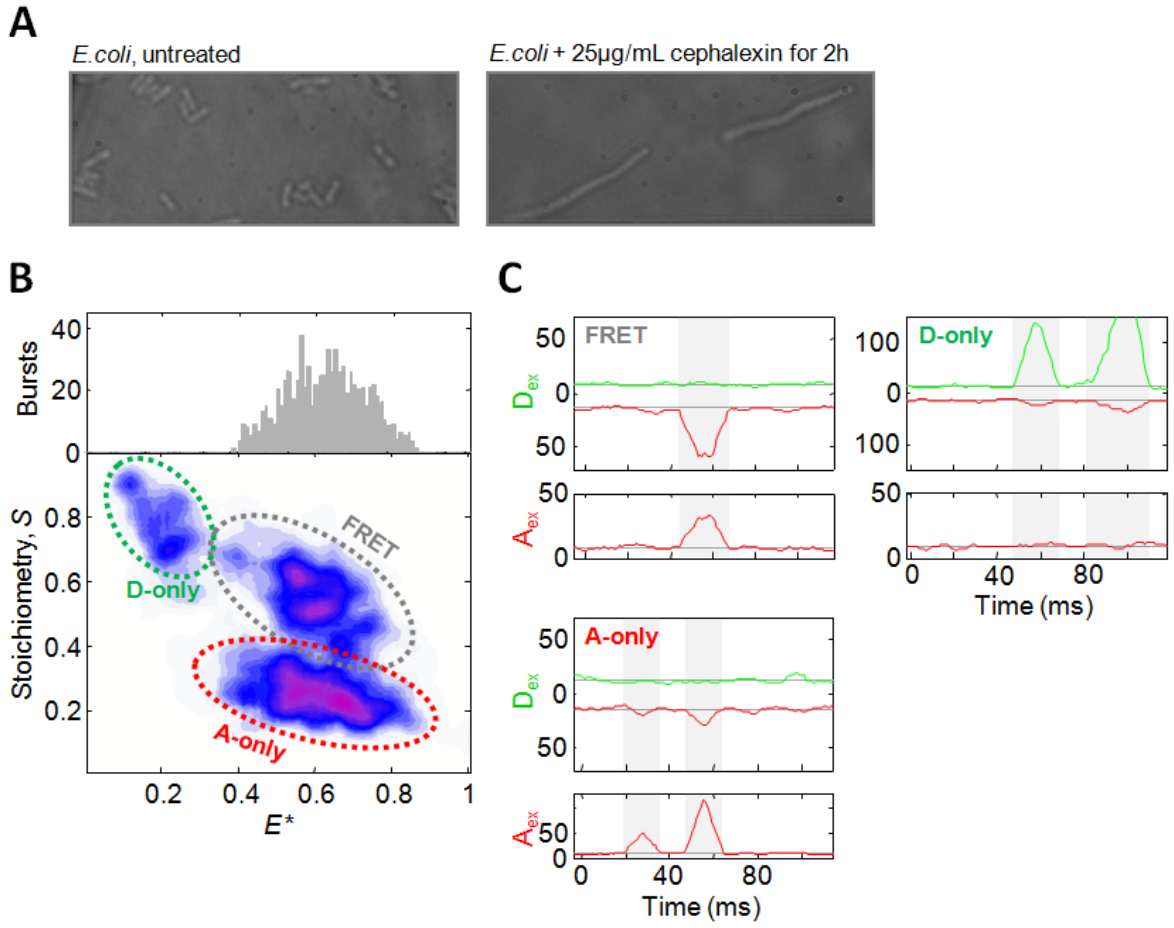

A
E.coli, untreated
E.coli + 25µg/mL cephalexin for 2h
B
Bursts
Stoichiometry, S
E*
D-only
FRET
A-only
C
FRET
D-only
A-only
Dex
Aex
Time (ms)

Figure 8. Extended detection of smFRET in elongated *E. coli*. (A) Cephalexin-mediated elongation of *E. coli*. In the presence of sub-inhibitory doses of cephalexin, *E. coli* continues to grow but does not septate, forming cells up more than 100x longer than the typical cell in the absence of cephalexin. Due to the dependence of diffusion time on the square of distance, a fluorescent molecule 100x farther from a confocal spot will take, on average, 10,000x longer to diffuse into it, dramatically increasing the time over which confocal fluorescence measurements can be made in individual cells of *E. coli*. (B) ALEX histogram of a single cephalexin-elongated cell over the course of 4 minutes. Cells were treated with cephalexin to elongate them as in panel A, then electroporated with a high-FRET DNA ($lac^{Cy3B(-15),A647N(-3)}$). Burst search was performed as described in Methods, and successfully detected distinct donor-only, acceptor-only, and coincident high-FRET species, consistent with the presence of both intact and nuclease-degraded, or partially photobleached, DNA constructs. A 1D histogram of the high-FRET population (grey circle in 2D histogram) is shown on top of the 2D histogram; Gaussian fitting to this population yielded a mean E^* of 0.63, though *in vitro* this DNA has an E^* of 0.79. (C) Representative bursts from the D-only, A-only and high-FRET populations in the ALEX histogram. The time traces are binned into 1 ms intervals, and smoothed over 3 ms to reduce noise. Tick marks along the time axis indicate intervals of 20 ms. In the D- and A-only populations, donor spectral leakage and acceptor direct-excitation result in significant fluorescence appearing in the $D_{ex}A_{em}$ channel, as observed in *in vitro* experiments (Chapter 2).

6.3 Conclusions

In this chapter we overcame several obstacles to performing confocal single-molecule fluorescence in live *E. coli*. These include internalizing fluorescently labelled DNA molecules into *E. coli*, attenuating their degradation by cellular nucleases, and successfully detecting them via confocal fluorescence. Together these methods constitute a GFP analogue, or "Green Fluorescent DNA," capable of monitoring DNA and DNA-protein conformational dynamics in living cells.

The technique in its present form will be most useful for studying the structure and dynamics of nucleic acids that diffuse freely in the *E. coli* cytoplasm, such as the structure and dynamics of messenger RNAs and ribozymes[49-51]. One interesting possibility is that the doubly-labelled Klenow Fragment studied in Chapters 4 and 5 could be electroporated into *E. coli*, and used to ask questions not only about the kinetics of KF opening/closing and nucleotide incorporation, but to learn more about the way in which it searches for its target *in vivo*. From a biotechnology standpoint, our method could also be used to electroporate the transcription factor biosensors described in Chapter 3 into cells, to monitor the activity of transcription factors in real time *in vivo.* We discuss these possibilities further in Chapter 7.

We note that, despite the adequacy of the methodology developed here for observing fluorescent DNAs inside *E. coli*, three improvements would be desirable: greater cell viability following electroporation, increased stability of the DNAs to nucleases, and decreased background. In terms of viability, the electroporation protocol is optimized primarily for the introduction of plasmids into *E. coli*; it may therefore be possible to improve viability and internalization efficiency by altering electroporation conditions. Even if better conditions cannot be found, the use of appropriate organic fluorophores may make it possible to simultaneously stain cells with spectrally distinct viability dyes (e.g., FRET between "blue" and "green" fluorophores Alexa488 and Cy3B to monitor dynamics, while using the 635 nm

far-red *BacLight* LIVE/DEAD cell stain from Invitrogen to monitor cell viability); this would allow us to ensure that only living cells are analyzed following electroporation. Increases in DNA stability *in vivo* might be achieved via the introduction of additional modifications at the 5' and 3' ends of the DNA, such as peptide or locked nucleic acid backbones (PNA or LNA respectively), which are more resistant to nucleases than phosphorothioate linkages in some cases[52,53]. Furthermore, signal-to-noise suffers from high background during confocal fluorescence measurements; the stability of this background suggests that it is caused by scattering off of the cell, the surrounding agarose, and the glass cover slip. Since high background has a broadening effect on FRET histograms[54], using cell immobilization methods that decrease proximity to the cover slide, or a smaller pinhole to minimize the size of the confocal spot in the axial direction, may be desirable in future implementations of this technique.

We also note that while these "Green Fluorescent DNAs" may be attractive in the aforementioned applications, they suffer from several drawbacks. First, unlike fluorescent proteins, which may be expressed continuously in growing cells, fluorescent DNAs electroporated into *E. coli* will be diluted out of the cells in the course of cell division; DNA probes meant to monitor long-term processes during cell growth may therefore have a limited useful lifetime, for reasons beyond photobleaching or nuclease-mediated degradation. Second, these DNAs are freely-diffusing in the cell; while this is appropriate for studying mRNA or ribozymes, it may not be appropriate for studying processes such as transcription or replication, which normally occur on the chromosome. One reason this is a problem is that DNAs electroporated into the cell will not have the topological properties of the chromosome, though chromosomal supercoiling is known to drive, or resist, the activities of DNA-dependent enzymes[55]. In such cases, it will be important to develop methods to

integrate electroporated DNAs into the *E. coli* chromosome – an idea we discuss further in Chapter 7.

Despite these caveats, the methods developed here to internalize and observe fluorescent DNA molecules in *E. coli* fill a significant gap in the *in vivo* single-molecule toolbox, making possible the analysis of DNA and protein-DNA conformational dynamics in live *E. coli* cells; such methods should also be immediately useful in monitoring the structure and dynamics of mRNAs and ribozymes *in vivo*. This method has clear potential in addressing previously inaccessible questions about the behaviour of proteins and nucleic acids in live prokaryotic cells.

6.4 Materials and methods

Fluorescence spectroscopy and microscopy

Confocal single-molecule fluorescence spectroscopy was carried out on a custom ALEX setup as described in Chapter 3. FCS assumes a 3D-Gaussian confocal volume, but the refractive index mismatch between an oil immersion objective and the cell can cause aberrations in the resulting autocorrelation function[56]. We therefore use a water-immersion objective (UPLS-APO, 60x, 1.2 NA, Olympus, Japan) for both FCS and single-molecule fluorescence measurements. For all confocal single-molecule fluorescence experiments, 532 nm and 635 nm laser powers were 160 – 200 μW, and 60 μW, respectively. Fluorescence correlation spectroscopy was performed on the same setup, using the 532 nm laser at laser powers of 10 μW *in vitro*, and 1 μW *in vivo*, with acquisition times of 30 s in both cases. Widefield fluorescence microscopy was carried out by illuminating the sample with the 532 nm laser alone at 200 μW; emitted light was collected as previously described (Chapter 3); donor- and acceptor-emission fluorescence were separated with a dichroic mirror

(630DRLP, Omega, Brattleboro, VT), filtered (HQ585/F70 and HQ650LP, for green and red emission respectively) and directed onto the two the halves of the chip of an EMCCD camera (iXon+, BI-887, Andor, Belfast, UK).

Preparation of electrocompetent cells

Cells were inoculated at an OD_{600} of 0.01 into 200 mL of fresh LB in a 1 L flask, and grown at 37°C, 250 rpm to an OD of 0.5. The culture was pelleted at 1000 g for 15min, washed, and resuspended in an equal volume of ice-cold deionized water; this washing and resuspension step was repeated four times. Cells were then pelleted and resuspended in a final volume of 1.2 mL 7% DMSO, aliquotted into eppendorf tubes, and stored at -80°C. For cephalexin-elongated cells, all steps were repeated as described except that, at OD 0.05, cephalexin was added to the growing culture to a final concentration of 25 μg/mL. Cells were then grown for a further two hours, then washed and stored as described. Unless otherwise indicated, the genetic background of cells used for electroporation and other experiments was DH5α.

Electroporation

Electroporation was performed using the MicroPulser electroporator from Bio-Rad (Hercules, CA) and 0.1 cm gap-width cuvettes. Cuvettes were pre-chilled on ice, 40 μL of electrocompetent cells added, and typically 2 μL of 10 μM fluorescent DNA mixed with the cells. After incubating for 1 minute, the sample was electroporated once at 18 kV/cm. The electroporated sample was immediately mixed with 1mL of LB + 20 mM glucose, and either recovered for 40 min at 37°C, or washed and immediately placed on a slide (see "slide preparation" below). Prior to microscopy, and in order to reduce background, cells were pelleted, washed and resuspended at least four times in 1mL of M9 medium (3.4% Na_2HPO_4,

1.5% KH_2PO_4, 0.25% NaCl, 0.5% NH_4Cl, 0.024% $MgSO_4$, 0.001% $CaCl_2$ wt/vol) containing 0.4% glucose.

Slide preparation

To visualize *E. coli* via widefield or confocal fluorescence microscopy, cells were washed and resuspended in M9 minimal medium + 0.4% glucose, then sandwiched between a pad of M9 containing 0.8% agarose (Bio-Rad, Hercules, CA), and a glass cover slip.

DNA

Unless otherwise indicated, DNA labelling, purification and annealing were performed as described in Chapter 4. For dumbbell and hairpin DNAs, strands were annealed in T4 DNA Ligase buffer (NEB, Ipswich, MA). Nicks between annealed DNA strands were ligated by incubating with 400 units T4 DNA Ligase overnight at room temperature. This ligation mixture was used directly in subsequent DNA degradation and electroporation experiments.

Preparation of *E. coli* whole-cell lysates

40mL of cells were grown to an OD of 0.5 and pelleted, washed and resuspended in 2mL of KG7 buffer (40 mM HEPES buffer pH 7.0, 100 mM potassium-L-glutamate, 10 mM $MgCl_2$, 1 mM DTT, 100 μg/ml bovine serum albumin, 5% glycerol) containing 1mM DTT and EDTA-free protease inhibitor cocktail (Roche Applied Science, Indianapolis, IN). Cells were then sonicated, and spun at 16,000g for 10min. The supernatant was then diluted to 5mg/mL protein (measured by absorbance at 280 nm) and used for further experiments.

DNA degradation assay and DNA electrophoresis

2.5 pmol of DNA was added to 50 μL of a 5 mg/mL whole-cell lysate of *E. coli*. For each time point, 5 μL of lysate was removed, and 0.5 μL of 500 mM EDTA pH 8.0 added to quench the reaction. Each sample was then mixed with 5 μL formamide and run on a denaturing PAGE gel. The PAGE gel contained 1x TBE, 6 M Urea and 20% (29:1) Acrylamide.

DNA degradation assay and ensemble fluorescence

1 pmol of DNA was added to 100 μL of a 5 mg/mL whole-cell lysate of *E. coli*, mixed, and immediately analyzed by ensemble FRET. Fluorescence intensity measurements were performed on a fluorescence spectrophotometer (PTI, Birmingham, NJ); samples were illuminated with a mercury lamp at 525 nm, and donor and acceptor emission measured at 570 nm and 655 nm respectively. Emission was background-subtracted, corrected for detection efficiency and FRET calculated as $E = A / (D+A)$ (where A and D are the corrected fluorescence intensities). In Fig. 5, we normalize FRET to 100% at the start of the experiment, and 0% for the FRET observed at the 2-hour time point (FRET was essentially stable after 1 hour, indicating full degradation of the DNA).

Fluorescence correlation spectroscopy

Using a Rhodamine 6G sample for calibration, we measured the geometry of our confocal volume to have a radius of w_o = 481 nm, and an axial radius of z_o = 3.1 μm (see Chapter 2 for characterization method); the large w_o is due to slight underfilling of the objective. A single cell of *E. coli* can be approximated as a 2 μm-long cylinder with a radius of ~ 400 nm[57] (consistent with the cell sizes observed in Fig. 2), and we assume that the confocal spot completely illuminates the cell on its radial axis, but not along its length. We therefore

approximate the fraction of the cell occupied by the confocal spot, f_C, as (2*481 nm) / (2000 nm) ≈ 0.48.

Expression of eYFP

To express eYFP in *E. coli* for photobleaching studies, we transformed *E. coli* strain BL21(DE3) with a derivative of vector pIvex2.3d (Roche Applied Science, Indianapolis, IN), expressing eYFP from a T7 promoter. A 50 mL culture was grown at 37°C, 250 rpm, to OD_{600} 0.7, induced with 1mM IPTG, and incubated for a further 3 hours. Cells were then pelleted, washed with deionized water, and imaged via widefield fluorescence microscopy (see "slide preparation").

Burst selection

Due to the higher background of confocal measurements performed on *E. coli* than in free solution, we used an alternative burst selection method. Specifically, we used a modified form of the burst detection method previously described by Klenerman and colleagues[58,59] in which fluorescence is time-binned, smoothed, and bursts identified as consisting of consecutive time bins in which some threshold number of photons is exceeded. We used bins of 1 ms in duration, with 3-bin smoothing. Assuming the background is Poisson-distributed, we measured the mean background count per 1-ms time bin, and required that bins belonging to a fluorescence burst have a number of photons greater than that mean, plus two standard deviations (the standard deviation for a Poisson signal is simply the square root of the mean). We also required that any given burst contain at least 60 total photons, after subtracting background.

Contributions

Experiments on electroporation viability and *in vitro* DNA degradation were conducted together with Robert Crawford. Practical assistance in performing confocal fluorescence *in vivo* was provided by Ludovic LeReste. Ideas and critical feedback were provided by Robert Crawford and Achillefs N. Kapanidis.

References

[1] Bedford E, Tabor S, Richardson CC (1997) The thioredoxin binding domain of bacteriophage T7 DNA polymerase confers processivity on Escherichia coli DNA polymerase I. *Proc Natl Acad Sci U S A* 94: 479-484

[2] Maul RW, Sanders LH, Lim JB, Benitez R, Sutton MD (2007) Role of Escherichia coli DNA polymerase I in conferring viability upon the dnaN159 mutant strain. *J Bacteriol* 189: 4688-4695

[3] Santoso Y, Joyce CM, Potapova O, Le Reste L, Hohlbein J, et al. (2010) Conformational transitions in DNA polymerase I revealed by single-molecule FRET. *Proc Natl Acad Sci U S A* 107: 715-720

[4] Santoso Y, Torella JP, Kapanidis AN (2010) Characterizing Single-Molecule FRET Dynamics with Probability Distribution Analysis. *Chemphyschem* 11: 2209-2219

[5] Torella JP, Holden SJ, Santoso Y, Hohlbein J, Kapanidis AN (2010) Identifying molecular dynamics in single-molecule FRET experiments with Burst Variance Analysis. *Biophys J* (accepted):

[6] Joyce CM, Benkovic SJ (2004) DNA polymerase fidelity: Kinetics, structure, and checkpoints. *Biochemistry (Mosc)* 43: 14317-14324

[7] Joyce CM, Potapova O, DeLucia AM, Huang XW, Basu VP, et al. (2008) Fingers-closing and other rapid conformational changes in DNA polymerase I (Klenow fragment) and their role in nucleotide selectivity. *Biochemistry (Mosc)* 47: 6103-6116

[8] Carroll SS, Cowart M, Benkovic SJ (1991) A mutant of DNA Polymerase I (Klenow Fragment) with reduced fidelity. *Biochemistry (Mosc)* 30: 804-813

[9] Markovitz A (2005) A new in vivo termination function for DNA polymerase I of Escherichia coli K12. *Mol Microbiol* 55: 1867-1882

[10] Yu J, Xiao J, Ren XJ, Lao KQ, Xie XS (2006) Probing gene expression in live cells, one protein molecule at a time. *Science* 311: 1600-1603

[11] Elf J, Li GW, Xie XS (2007) Probing transcription factor dynamics at the single-molecule level in a living cell. *Science* 316: 1191-1194

[12] Pollok BA, Heim R (1999) Using GFP in FRET-based applications. *Trends Cell Biol* 9: 57-60

[13] Golding I, Cox EC (2006) Physical nature of bacterial cytoplasm. *Phys Rev Lett* 96:

[14] Adams SR, Campbell RE, Gross LA, Martin BR, Walkup GK, et al. (2002) New biarsenical Ligands and tetracysteine motifs for protein labeling in vitro and in vivo: Synthesis and biological applications. *J Am Chem Soc* 124: 6063-6076

[15] Wang L, Brock A, Herberich B, Schultz PG (2001) Expanding the genetic code of Escherichia coli. *Science* 292: 498-500

[16] Howarth M, Takao K, Hayashi Y, Ting AY (2005) Targeting quantum dots to surface proteins in living cells with biotin ligase. *Proc Natl Acad Sci U S A* 102: 7583-7588

[17] Summerer D, Chen S, Wu N, Deiters A, Chin JW, et al. (2006) A genetically encoded fluorescent amino acid. *Proc Natl Acad Sci U S A* 103: 9785-9789

[18] Mao S, Benninger RKP, Yan YL, Petchprayoon C, Jackson D, et al. (2008) Optical lock-in detection of FRET using synthetic and genetically encoded optical switches. *Biophys J* 94: 4515-4524

[19] Li Y, Lu XY, Li JW, Berube N, Giest KL, et al. (2010) Genetically Engineered, Biarsenically Labeled Influenza Virus Allows Visualization of Viral NS1 Protein in Living Cells. *J Virol* 84: 7204-7213

[20] Murakoshi H, Iino R, Kobayashi T, Fujiwara T, Ohshima C, et al. (2004) Single-molecule imaging analysis of Ras activation in living cells. *Proc Natl Acad Sci U S A* 101: 7317-7322

[21] Sakon JJ, Weninger KR (2010) Detecting the conformation of individual proteins in live cells. *Nat Methods* 7: 203-U256

[22] McArthur M, Bibb MJ (2008) Manipulating and understanding antibiotic production in Streptomyces coelicolor A3(2) with decoy oligonucleotides. *Proc Natl Acad Sci U S A* 105: 1020-1025

[23] Silverman AP, Kool ET (2005) Quenched autoligation probes allow discrimination of live bacterial species by single nucleotide differences in rRNA. *Nucleic Acids Res* 33: 4978-4986

[24] Wyber JA, Andrews J, Demanuele A (1997) The use of sonication for the efficient delivery of plasmid DNA into cells. *Pharm Res* 14: 750-756

[25] Dower WJ, Miller JF, Ragsdale CW (1988) High-efficiency transformation of *E.coli* by high-voltage electroporation. *Nucleic Acids Res* 16: 6127-6145

[26] Hibino M, Shigemori M, Itoh H, Nagayama K, Kinosita K (1991) Membrane conductance of an electroporated cell analyzed by submicrosecond imaging of transmembrane potential. *Biophys J* 59: 209-220

[27] Pavlin M, Kanduser M, Rebersek M, Pucihar G, Hart FX, et al. (2005) Effect of cell electroporation on the conductivity of a cell suspension. *Biophys J* 88: 4378-4390

[28] Weaver JC, Chizmadzhev YA (1996) Theory of electroporation: A review. *Bioelectrochem Bioenerg* 41: 135-160

[29] Gehl J (2003) Electroporation: theory and methods, perspectives for drug delivery, gene therapy and research. *Acta Physiol Scand* 177: 437-447

[30] Santini CL, Bernadac A, Zhang M, Chanal A, Ize B, et al. (2001) Translocation of jellyfish green fluorescent protein via the Tat system of Escherichia coli and change of its periplasmic localization in response to osmotic up-shock. *J Biol Chem* 276: 8159-8164

[31] Nenninger A, Mastroianni G, Mullineaux CW (2010) Size Dependence of Protein Diffusion in the Cytoplasm of Escherichia coli. *J Bacteriol* 192: 4535-4540

[32] Mullineaux CW, Nenninger A, Ray N, Robinson C (2006) Diffusion of green fluorescent protein in three cell environments in Escherichia coli. *J Bacteriol* 188: 3442-3448

[33] Elowitz MB, Surette MG, Wolf PE, Stock JB, Leibler S (1999) Protein mobility in the cytoplasm of Escherichia coli. *J Bacteriol* 181: 197-203

[34] Konopka MC, Sochacki KA, Bratton BP, Shkel IA, Record MT, et al. (2009) Cytoplasmic Protein Mobility in Osmotically Stressed Escherichia coli. *J Bacteriol* 191: 231-237

[35] Kumar M, Mommer MS, Sourjik V Mobility of Cytoplasmic, Membrane, and DNA-Binding Proteins in Escherichia coli. *Biophys J* 98: 552-559

[36] Widengren J, Chmyrov A, Eggeling C, Lofdahl PA, Seidel CAM (2007) Strategies to improve photostabilities in ultrasensitive fluorescence spectroscopy. *J Phys Chem A* 111: 429-440

[37] Nagai T, Ibata K, Park ES, Kubota M, Mikoshiba K, et al. (2002) A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nat Biotechnol* 20: 87-90

[38] Day RN, Davidson MW (2009) The fluorescent protein palette: tools for cellular imaging. *Chem Soc Rev* 38: 2887-2921

[39] Korona R, Levin BR (1993) Phage-mediated selection and the evolution and maintenance of restriction-modification. *Evolution* 47: 556-575

[40] Viswanathan M, Lovett ST (1998) Single-strand DNA-specific exonucleases in Escherichia coli: Roles in repair and mutation avoidance. *Genetics* 149: 7-16

[41] Eskin B, Linn S (1972) DNA modification and restriction enzymes of Escherichia coli B. *J Biol Chem* 247:

[42] Lin JJ (1992) Endonuclease A degrades chromosomal and plasmid DNA of Escherichia coli present in most preparations of single stranded DNA from phagemids. *Proc Natl Sci Counc Repub China B* 16: 1-5

[43] Shaw JP, Kent K, Bird J, Fishback J, Froehler B (1991) Modified deoxyoligonucleotides stable to exonuclease degradation in serum. *Nucleic Acids Res* 19: 747-750

[44] Youngs DA, Bernstei.Ia (1973) Involvement of recB-recC nuclease (exonuclease V) in the process of X-ray-induced DNA degradation in radiosensitive strains of Escherichia coli K-12. *J Bacteriol* 113: 901-906

[45] Capaldok F, Barbour SD (1971) Involvement of recombination genes in growth and viability of Escherichia coli K-12. *J Bacteriol* 106: 204-&

[46] Suel GM, Kulkarni RP, Dworkin J, Garcia-Ojalvo J, Elowitz MB (2007) Tunability and noise dependence in differentiation dynamics. *Science* 315: 1716-1719

[47] Perozo E (2006) Gating prokaryotic mechanosensitive channels. *Nature Reviews Molecular Cell Biology* 7: 109-119

[48] Martinac B, Buechner M, Delcour AH, Adler J, Kung C (1987) Pressure-sensitive ion channel in *E.coli*. *Proc Natl Acad Sci U S A* 84: 2297-2301

[49] Zhuang XW, Kim H, Pereira MJB, Babcock HP, Walter NG, et al. (2002) Correlating structural dynamics and function in single ribozyme molecules. *Science* 296: 1473-1476

[50] Okumus B, Wilson TJ, Lilley DMJ, Ha T (2004) Vesicle encapsulation studies reveal that single molecule ribozyme heterogeneities are intrinsic. *Biophys J* 87: 2798-2806

[51] Xie Z, Srividya N, Sosnick TR, Pan T, Scherer NF (2004) Single-molecule studies highlight conformational heterogeneity in the early folding steps of a large ribozyme. *Proc Natl Acad Sci U S A* 101: 534-539

[52] Kurreck J, Wyszko E, Gillen C, Erdmann VA (2002) Design of antisense oligonucleotides stabilized by locked nucleic acids. *Nucleic Acids Res* 30: 1911-1918

[53] Braasch DA, Corey DR (2002) Novel antisense and peptide nucleic acid strategies for controlling gene expression. *Biochemistry (Mosc)* 41: 4503-4510

[54] Nir E, Michalet X, Hamadani KM, Laurence TA, Neuhauser D, et al. (2006) Shot-noise limited single-molecule FRET histograms: Comparison between theory and experiments. *J Phys Chem B* 110: 22103-22124

[55] Liu LF, Wang JC (1987) Supercoiling of the DNA template during transcription. *Proc Natl Acad Sci U S A* 84: 7024-7027

[56] Hess ST, Webb WW (2002) Focal volume optics and experimental artifacts in confocal fluorescence correlation spectroscopy. *Biophys J* 83: 2300-2317

[57] Pierucci O (1978) Dimensions of Escherichia coli at various growth rates - model for envelope growth. *J Bacteriol* 135: 559-574

[58] Clarke RW, Orte A, Klenerman D (2007) Optimized threshold selection for single-molecule two-color fluorescence coincidence spectroscopy. *Anal Chem* 79: 2771-2777

[59] Orte A, Clarke R, Balasubramanian S, Klenerman D (2006) Determination of the fraction and stoichiometry of femtomolar levels of biomolecular complexes in an excess of monomer using single-molecule, two-color coincidence detection. *Anal Chem* 78: 7707-7715

7. Conclusion

Molecular dynamics underlie crucial processes in biology, from the folding of proteins and nucleic acids[1-7], to the complex enzymatic activities of polymerases and molecular motors[8-14], to stochastic phenotypic switching in individual cells[15]. Whereas conventional biochemical techniques are limited to ensemble-level studies of these processes, confocal single-molecule fluorescence can be used to monitor these processes directly at the nanometre-scale, and at millisecond or faster timescales. In this work, we contributed in two significant ways to the confocal single-molecule fluorescence toolbox: through the development of novel analytical methods for identifying and quantifying molecular dynamics at the single-molecule level, and through the development of novel experimental techniques for the analysis of molecular dynamics in live prokaryotic cells.

The analytical methods we developed to monitor molecular dynamics fall into two categories: intermolecular, and intramolecular. To study intermolecular dynamics, we developed a correction-factor-based analytical method for accurately quantifying diffusing species via confocal single-molecule fluorescence (Chapter 3). We validated this method against a series of numerical simulations, and demonstrated its utility in improving the

accuracy of a single-molecule transcription factor biosensor[16]. In addition to its potential use in other single-molecule biosensors[17], a powerful application of this method may be in monitoring the oligomerisation of amyloid fibrils, which play a role in neurodegenerative diseases such as Alzheimer's and Parkinson's[18]. In a previous study of amyloid aggregation, for example, single-molecule stoichiometry measurements were used to extract aggregate size distributions[19]. While this study correctly identified the need to adjust the size distribution by correcting for the lower encounter rate of large aggregates with the confocal volume, it did not take errors such as recrossing and detection efficiency into account, though these errors are also sensitive to particle size, and can alter apparent concentrations by over an order of magnitude (Chapter 3). Re-analysis of these results using the methods developed in Chapter 3, and potentially a re-evaluation of their biological implications, may therefore be warranted.

To study intramolecular dynamics, we developed two novel analytical methods: a generalized form of Probability Distribution Analysis (PDA) for modelling the contribution of dynamics to single-molecule FRET histograms (Chapter 4, Ref [20]), and Burst Variance Analysis (BVA), which enables the unambiguous detection of FRET dynamics in smFRET experiments (Chapter 5, Ref. [21]). Together, these tools can be used to identify dynamic heterogeneity in smFRET data, hypothesis-test specific kinetic models against it, and extract rate information. Specifically, we used BVA and PDA to identify millisecond-timescale "fingers-closing" dynamics in the Klenow Fragment (KF) of *E. coli* DNA Polymerase I, in both its liganded and unliganded forms. Interestingly, on the basis of crystal structures, binding of a complementary nucleotide was previously thought to 'lock' the polymerase into the closed conformation, poising it for nucleotide incorporation[22]. Consistent with recent stopped-flow studies[23], however, our work showed significant flexibility in KF when bound to a complementary nucleotide, suggesting that substrate-bound KF may open and close

multiple times prior to nucleotide incorporation (Chapters 4 and 5, Refs. [20,21]). As the closed conformation of the polymerase is thought to be crucial to the fidelity of nucleotide incorporation[23,24], dynamic sampling of this conformation prior to nucleotide incorporation may constitute a form of proofreading. This work may therefore be relevant in elucidating the mechanism by which DNA Polymerase I achieves its impressive fidelity in DNA synthesis, and illustrates the potential of PDA and BVA to extract useful information about molecular dynamics from single-molecule FRET data.

In addition to detecting and quantifying molecular dynamics *in vitro*, we also developed experimental methods for performing confocal single-molecule fluorescence in a living cell (Chapter 6). We used electroporation to internalize fluorescent DNAs into live *E. coli*; while nucleic-acid-mediated degradation limited the lifetime of these probes, a combination of genetic modifications of the host, and chemical modifications of the DNA, were capable of decreasing their susceptibility to nuclease-mediated degradation by an order of magnitude. These probes showed superior photostability compared with fluorescent proteins, and could be detected via confocal single-molecule FRET and stoichiometry measurements inside *E. coli*, replicating our abilities *in vitro*. To our knowledge, this is the first example of confocal single-molecule detection of both an exogenous fluorescent probe, and FRET, in a prokaryotic organism. While preliminary, this work should enable the study of DNA conformational dynamics in live prokaryotic cells. This technology complements other recent advances in *in vivo* single-molecule fluorescence, which typically rely on fluorescent protein binding or immobilization for detecting molecular dynamics[15,25-28], and therefore fills a gap in the rapidly growing *in vivo* single-molecule toolbox.

7.1 Future Work

Work in our lab has already begun using PDA and BVA to investigate the relationship between fingers-closing dynamics and the fidelity of DNA Polymerase I. Specifically, KF mutants with reduced fidelity, or with difficulty in distinguishing dNTPs from rNTPs (point mutant E710A[29]) have been fluorescently labelled in on their finger and thumb subdomains. These mutant constructs will make it possible to correlate fingers-closing dynamics with the fidelity of nucleotide incorporation directly, by challenging wild-type and mutant KFs with noncomplementary nucleotides, and observing whether mutants with reduced fidelity also exhibit altered fingers-closing dynamics.

The methodologies of PDA and BVA are also being improved and extended. One major extension of these tools will be the ability to perform PDA and BVA not only on FRET data, but on stoichiometry (*S*). Heterogeneity in an *S* histogram can be used as a reporter for the dynamic quenching of one or both fluorophores[30,31]; this has two attractive applications. First, as both *S* and *E** histograms are sensitive to dye photophysics, but only *E** is sensitive to true FRET dynamics, the width of the *S* histogram could be used as a reporter and internal control in FRET experiments for dye photophysics.

Second, dynamic changes in the brightness of a single fluorophore may be interesting on their own. Dynamic quenching of single dyes has been used in several applications, for instance using resonance energy transfer to monitor conformational changes in T7 DNA polymerase[32] and in DNA hairpins[30]. By labelling a molecule with both an "observer" fluorophore which is not quenched, and a "reporter" fluorophore exhibiting dynamic quenching, PDA and BVA could be used to monitor quenching dynamics in *S* histograms in a manner analogous to their current use on FRET histograms, an idea previously proposed by Klenerman and colleagues[30]. Interestingly, quenching is actually a more versatile molecular ruler than FRET; this is because quenching can be achieved by processes other than

resonance energy transfer (typically, 2–8 nm distance range), such as photo-induced electron transfer (PET, sub-nanometre distance range[33-35]), and nanometal surface energy transfer (NSET, 2–20 nm distance range[36,37]).

While PDA and BVA promise to reveal more about the dynamics of complex enzymes and nucleic acids *in vitro*, the most exciting (but challenging) future work is that performed *in vivo*. An interesting application of confocal single-molecule fluorescence *in vivo,* taking advantage of the techniques developed in Chapter 6, is the study of dynamics in freely-diffusing nucleic acids, such as mRNAs and ribozymes. The structural dynamics of ribozymes, for instance, have been analyzed by smFRET to reveal both significant heterogeneity in ribozyme structure, and a strong correlation between structure and catalytic activity[38,39]. While these studies provided new insight into the structure-function relationship of ribozymes *in vitro*, little is known about the behaviour of these molecules in their natural, cellular context. Our method may be ideal for studying prokaryotic ribozymes such as RNAse P, which diffuses freely in the cytosol (and is therefore appropriate to our confocal detection method), can be easily labelled[38], and should be amenable to electroporation into *E. coli*.

Making *in vivo,* confocal single-molecule fluorescence truly useful in the range of applications mentioned throughout this thesis, however, will require two major developments: the ability to internalize fluorescent proteins into *E. coli*., and the ability to insert fluorescent DNAs into the chromosome. With some optimization the first goal should be achievable, as electroporation has been used to transfer proteins into the cytoplasm of mammalian cells[40]. The ability to introduce proteins tagged with organic fluorophores into *E. coli* would widen the scope of our current technique considerably, for instance by allowing the KF fingers-closing assay to be performed *in vivo*. In such an assay, labelled KF would not simply diffuse through the cytoplasm, but would be recruited to the chromosome for DNA synthesis. This would result in effective immobilization of KF, requiring widefield or TIRF-

based microscopy [26,41]; if good enough signal-to-noise could be achieved, an ideal experimental approach would be to combine confocal and widefield measurements to monitor the dynamics of KF both while searching for DNA (and therefore diffusing, requiring confocal analysis), and while localized to sites of DNA replication. Such studies would greatly improve our understanding of the mechanism by which DNA Polymerase I operates *in vivo* during DNA replication and repair.

The second development, incorporating fluorescent DNA into the chromosome, could be achieved using either the lambda red recombination system[42], or one of many highly efficient recombinases[43,44] to achieve integration. The importance of incorporating fluorophores into the chromosome is difficult to overestimate, as fluorescently-labelled DNA has been used in previous *in vitro* studies of transcription initiation[10,12], DNA replication[9,11] and other processes that, in nature, occur exclusively on a chromosome (or large, relatively immobile plasmid). While in principle the DNA could be left to diffuse freely in the cytoplasm, this may disrupt the search process by which relevant binding proteins find the DNA[26], and could alter the activity of proteins that rely on topological properties of the chromosome for activity. As the point of working *in vivo* is to observe single molecules in their true biological context, it will be important not only to introduce DNAs into the *E. coli* cytoplasm, but to integrate them into the chromosome.

In the last two decades, single-molecule fluorescence has grown from humble origins detecting single fluorophores embedded in solids[45], to beginning to answer important questions about the molecular machines that make life possible. If these trends continue, the next two decades will offer unprecedented opportunities to investigate biology, especially in the context of the living cell[46]; indeed, I believe it will not be long before tasks such as watching DNA or RNA synthesis *in vivo*, with single-nucleotide precision, are not only achieved, but become routine. It is my hope that the methods and ideas presented in this

thesis will contribute in a significant way toward this challenging, but worthwhile, endeavour.

References

[1] Deniz AA, Laurence TA, Beligere GS, Dahan M, Martin AB, et al. (2000) Single-molecule protein folding: Diffusion fluorescence resonance energy transfer studies of the denaturation of chymotrypsin inhibitor 2. *Proc Natl Acad Sci U S A* 97: 5179-5184

[2] Jia YW, Talaga DS, Lau WL, Lu HSM, DeGrado WF, et al. (1999) Folding dynamics of single GCN4 peptides by fluorescence resonant energy transfer confocal microscopy. *Chemical Physics* 247: 69-83

[3] Karymov MA, Chinnaraj M, Bogdanov A, Srinivasan AR, Zheng GH, et al. (2008) Structure, Dynamics, and Branch Migration of a DNA Holliday Junction: A Single-Molecule Fluorescence and Modeling Study. *Biophys J* 95: 4372-4383

[4] Lipman EA, Schuler B, Bakajin O, Eaton WA (2003) Single-molecule measurement of protein folding kinetics. *Science* 301: 1233-1235

[5] Schuler B, Lipman EA, Eaton WA (2002) Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. *Nature* 419: 743-747

[6] Sharma S, Chakraborty K, Mueller BK, Astola N, Tang YC, et al. (2008) Monitoring protein conformation along the pathway of chaperonin-assisted folding. *Cell* 133: 142-153

[7] Zhao R, Rueda D (2009) RNA folding dynamics by single-molecule fluorescence resonance energy transfer. *Methods* 49: 112-117

[8] Yasuda R, Noji H, Kinosita K, Yoshida M (1998) F-1-ATPase is a highly efficient molecular motor that rotates with discrete 120 degrees steps. *Cell* 93: 1117-1124

[9] Christian TD, Romano LJ, Rueda D (2009) Single-molecule measurements of synthesis by DNA polymerase with base-pair resolution. *Proc Natl Acad Sci U S A* 106: 21109-21114

[10] Kapanidis AN, Margeat E, Ho SO, Kortkhonjia E, Weiss S, et al. (2006) Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. *Science* 314: 1144-1147

[11] Santoso Y, Joyce CM, Potapova O, Le Reste L, Hohlbein J, et al. (2010) Conformational transitions in DNA polymerase I revealed by single-molecule FRET. *Proc Natl Acad Sci U S A* 107: 715-720

[12] Tang GQ, Roy R, Bandwar RP, Ha T, Patel SS (2009) Real-time observation of the transition from transcription initiation to elongation of the RNA polymerase. *Proc Natl Acad Sci U S A* 106: 22175-22180

[13] Coban O, Lamb DC, Zaychikov E, Heumann H, Nienhaus GU (2006) Conformational heterogeneity in RNA polymerase observed by single-pair FRET microscopy. *Biophys J* 90: 4605-4617

[14] Peterman EJG, Sosa H, Moerner WE (2004) Single-molecule fluorescence spectroscopy and microscopy of biomolecular motors. *Annu Rev Phys Chem* 55: 79-96

[15] Choi PJ, Cai L, Frieda K, Xie XS (2008) A stochastic single-molecule event triggers phenotype switching of a bacterial cell. *Science* 322: 442-446

[16] Lymperopoulos K, Crawford R, Torella JP, Heilemann M, Hwang LC, et al. (2010) Single-Molecule DNA Biosensors for Protein and Ligand Detection. *Angew Chem Int Ed Engl* 49: 1316-1320

[17] Neely LA, Patel S, Garver J, Gallo M, Hackett M, et al. (2006) A single-molecule method for the quantitation of microRNA gene expression. *Nat Methods* 3: 41-46

[18] Koo EH, Lansbury PT, Kelly JW (1999) Amyloid diseases: Abnormal protein aggregation in neurodegeneration. *Proc Natl Acad Sci U S A* 96: 9989-9990

[19] Orte A, Birkett NR, Clarke RW, Devlin GL, Dobson CM, et al. (2008) Direct characterization of amyloidogenic oligomers by single-molecule fluorescence. *Proc Natl Acad Sci U S A* 105: 14424-14429

[20] Santoso Y, Torella JP, Kapanidis AN (2010) Characterizing Single-Molecule FRET Dynamics with Probability Distribution Analysis. *Chemphyschem* 11: 2209-2219

[21] Torella JP, Holden SJ, Santoso Y, Hohlbein J, Kapanidis AN (2010) Identifying molecular dynamics in single-molecule FRET experiments with Burst Variance Analysis. *Biophys J* (accepted):

[22] Li Y, Korolev S, Waksman G (1998) Crystal structures of open and closed forms of binary and ternary complexes of the large fragment of Thermus aquaticus DNA polymerase I: structural basis for nucleotide incorporation. *EMBO J* 17: 7514-7525

[23] Joyce CM, Potapova O, DeLucia AM, Huang XW, Basu VP, et al. (2008) Fingers-closing and other rapid conformational changes in DNA polymerase I (Klenow fragment) and their role in nucleotide selectivity. *Biochemistry (Mosc)* 47: 6103-6116

[24] Joyce CM, Benkovic SJ (2004) DNA polymerase fidelity: Kinetics, structure, and checkpoints. *Biochemistry (Mosc)* 43: 14317-14324

[25] Cai L, Friedman N, Xie XS (2006) Stochastic protein expression in individual cells at the single molecule level. *Nature* 440: 358-362

[26] Elf J, Li GW, Xie XS (2007) Probing transcription factor dynamics at the single-molecule level in a living cell. *Science* 316: 1191-1194

[27] Xie XS, Choi PJ, Li GW, Lee NK, Lia G (2008) Single-molecule approach to molecular biology in living bacterial cells. *Annual Review of Biophysics* 37: 417-444

[28] Leake MC, Chandler JH, Wadhams GH, Bai F, Berry RM, et al. (2006) Stoichiometry and turnover in single, functioning membrane protein complexes. *Nature* 443: 355-358

[29] Astatke M, Ng KM, Grindley NDF, Joyce CM (1998) A single side chain prevents Escherichia coli DNA polymerase I (Klenow fragment) from incorporating ribonucleotides. *Proc Natl Acad Sci U S A* 95: 3402-3407

[30] Li HT, Ren XJ, Ying LM, Balasubramanian S, Klenerman D (2004) Measuring single-molecule nucleic acid dynamics in solution by two-color filtered ratiometric fluorescence correlation spectroscopy. *Proc Natl Acad Sci U S A* 101: 14425-14430

[31] Kapanidis AN, Lee NK, Laurence TA, Doose S, Margeat E, et al. (2004) Fluorescence-aided molecule sorting: Analysis of structure and interactions by alternating-laser excitation of single molecules. *Proc Natl Acad Sci U S A* 101: 8936-8941

[32] Luo G, Wang M, Konigsberg WH, Xie XS (2007) Single-molecule and ensemble fluorescence assays for a functionally important conformational change in T7 DNA polymerase. *Proc Natl Acad Sci U S A* 104: 12610-12615

[33] Doose S, Neuweiler H, Sauer M (2009) Fluorescence Quenching by Photoinduced Electron Transfer: A Reporter for Conformational Dynamics of Macromolecules. *Chemphyschem* 10: 1389-1398

[34] Yang H, Luo GB, Karnchanaphanurach P, Louie TM, Rech I, et al. (2003) Protein conformational dynamics probed by single-molecule electron transfer. *Science* 302: 262-266

[35] Doose S, Neuweiler H, Barsch H, Sauer M (2007) Probing polyproline structure and dynamics by photoinduced electron transfer provides evidence for deviations from a regular polyproline type II helix. *Proc Natl Acad Sci U S A* 104: 17400-17405

[36] Yun CS, Javier A, Jennings T, Fisher M, Hira S, et al. (2005) Nanometal surface energy transfer in optical rulers, breaking the FRET barrier. *J Am Chem Soc* 127: 3115-3119

[37] Jennings TL, Schlatterer JC, Singh MP, Greenbaum NL, Strouse GF (2006) NSET molecular beacon analysis of hammerhead RNA substrate binding and catalysis. *Nano Lett* 6: 1318-1324

[38] Zhuang XW, Kim H, Pereira MJB, Babcock HP, Walter NG, et al. (2002) Correlating structural dynamics and function in single ribozyme molecules. *Science* 296: 1473-1476

[39] Xie Z, Srividya N, Sosnick TR, Pan T, Scherer NF (2004) Single-molecule studies highlight conformational heterogeneity in the early folding steps of a large ribozyme. *Proc Natl Acad Sci U S A* 101: 534-539

[40] Morgan WF, Day JP (1995) The introduction of proteins into mammalian cells by electroporation. *Methods Mol Biol* 48: 63-71

[41] Tokunaga M, Imamoto N, Sakata-Sogawa K (2008) Highly inclined thin illumination enables clear single-molecule imaging in cells. *Nat Methods* 5: 159-161

[42] Datsenko KA, Wanner BL (2000) One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc Natl Acad Sci U S A* 97: 6640-6645

[43] Marx CJ, Lidstrom ME (2002) Broad-host-range cre-lox system for antibiotic marker recycling in Gram-negative bacteria. *Biotechniques* 33: 1062-1067

[44] Snaith MR, Kilby NJ, Murray JAH (1996) An Escherichia coli system for assay of Flp site-specific recombination on substrate plasmids. *Gene* 180: 225-227

[45] Moerner WE, Kador L (1989) Optical detection and spectroscopy of single molecules in a solid. *Phys Rev Lett* 62: 2535-2538

[46] Xie XS, Yu J, Yang WY (2006) Perspective - Living cells as test tubes. *Science* 312: 228-230

Appendix

Chapter 3

CAP Detection: Kinetic Model. CAP binding was modelled using a set of coupled reversible reactions representing the equilibrium between free and transiently-bound half sites (Eq. S1), and the constant concentration of annealed half-sites at equilibrium in the presence of CAP (Eq. S2). As half-sites in the titration experiment are supplied at equimolar ratios, we assume Eq. S3. Finally, Eq. S4 and Eq. S5 account for conservation of the R-only half-site and the CAP concentration, respectively:

$$[H_1]_{eq}[H_2]_{eq}k_1 = [H_1 \cdot H_2]_{eq}k_{-1} \quad \text{[S1]}$$

$$[H_1]_{eq}[H_2]_{eq}k_1 + [CAP \cdot H_1H_2]_{eq}k_{-2} = [H_1 \cdot H_2]_{eq}(k_{-1} + [CAP]_{eq}k_2) \quad \text{[S2]}$$

$$[H_1]_{eq} = [H_2]_{eq} \quad \text{[S3]}$$

$$[H_1]_{tot} = [H_1]_{eq} + [H_1H_2]_{eq} + [CAP \cdot H_1H_2]_{eq} \quad \text{[S4]}$$

$$[CAP]_{tot} = [CAP]_{eq} + [CAP \cdot H_1H_2]_{eq} \quad \text{[S5]}$$

To relate ALEX data to the model, we determined the bound fraction, F_B, for a given experiment:

$$F_B = \frac{[CAP \cdot H_1H_2]_{eq}}{[H_1]_{eq} + [H_1H_2]_{eq} + [CAP \cdot H_1H_2]_{eq}} = \frac{[CAP \cdot H_1H_2]_{eq}}{[H_1]_{tot}} \quad \text{[S6]}$$

Together, Eq.1-6 allow us to calculate the total TF concentration as a function of F_B (Eq. 7):

$$[CAP]_{tot} = [H_1]_{tot}F_B + \frac{F_B K_{D2}}{2[H_1]_{tot}(1-F_B)^2} \times \left(2[H_1]_{tot}(1-F_B) + K_{D1} + \sqrt{K_{D1}}\sqrt{4[H_1]_{tot}(1-F_B) + K_{D1}}\right) \quad \text{[S7]}$$

We note that practical details of calculating F_B are given in Chapter 3, Methods.

CAP half-site H1

5′-AACGCAATAAATGTGA-3′
3′-TTGCGTTATTTACACTTCATCT-5′

CAP half-site H2

5′-AGTAGATCACATTTTAGGCACCA-3′
3′-AGTGTAAAATCCGTGGT-5′

Figure S1. Sequences of CAP-specific half-site DNAs. Each half-site is composed of two annealed oligos. Half-sites H1 and H2 are labelled with Cy3B and Atto647N fluorophores via their 5'-C6-amino group. Oligo synthesis and labelling details are provided in the Methods section of Chapter 3.

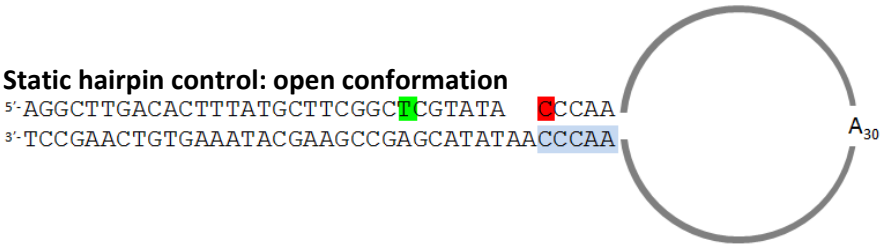
Chapter 4

Donor-only control

5′-AGGCTTGACACTTTATGCTTCGGCTCGTATAATGTGTGGAATTGTGAGAGCGGATAACAATTTC-3′
3′-TCCGAACTGTGAAATACGAAGCCGAGCATATTACACACCTTAACACTCTCGCCTATTGTTAAAG-5′

Dynamic hairpin

Static hairpin control: open conformation



Static hairpin control: closed conformation

5′-AGGCTTGACACTTTATGCTTCGGCTCGTATA CCCAAGAATTGTGAGAGCGGATAACAATTTC-3′
3′-TCCGAACTGTGAAATACGAAGCCGAGCATATAAGGGTTCTTAACACTCTCGCCTATTGTTAAAG-5′

Inextensible hairpin substrate for Klenow Fragment studies

A^{A}GAGTCAACAGGTC$_{H}$-3′
A_{G}CTCAGTTGTCCAGAGATGG-5′

Figure S1. The nucleotides highlighted in green and red were labelled with Cy3B (donor) and Atto647N (acceptor) fluorophores respectively. All DNAs were PAGE-purified before analysis. (A) Donor-only-labelled DNA (lac$^{Cy3B(-5)}$). (B) The hairpin construct is composed of a 31-nucleotide double-stranded region, a two-nucleotide single-stranded gap, another 5-nucleotide double-stranded region and a 30-adenine single stranded loop. The hairpin fluctuates dynamically due to spontaneous melting and re-annealing of the 5-nucleotide double stranded region. (C) As a control for the open (melted) form of the hairpin, we constructed a similar hairpin with a mismatch in the 5-nucleotide double-stranded region (*highlighted in blue*); this hairpin never anneals, and so is always in the open conformation.

(D) As a static control for the closed (annealed) form of the hairpin, we replaced the 30-adenine loop with a 26-basepair double-stranded region. (*E*) Hairpin template-primer used in Klenow Fragment studies. The 3’ terminus contains a dideoxynucleotide to prevent successful nucleotide incorporation, allowing us to study the structure of KF while bound to a complementary nucleotide, and poised for dNTP incorporation.

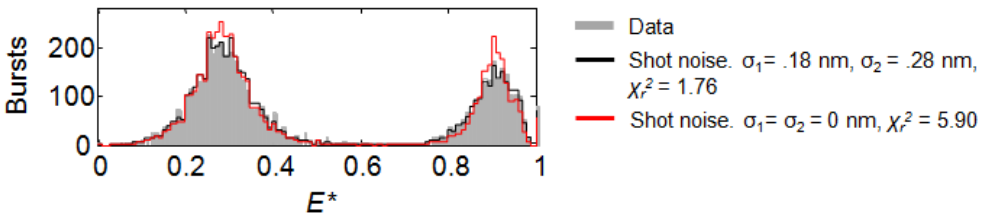


Figure S2. Static hairpin controls show broadening beyond shot noise. We analyzed a mixture of open and closed static control hairpins with smFRET and fit the distributions with PDA (same data as in Fig. 5B). Whereas the distributions could be fit by assuming a Gaussian distribution of originating FRET values (black line; $\langle E_{open}* \rangle = 0.285 \pm 0.002$, $\sigma^{r}_{open} = 0.18 \pm 0.01$ nm, $\langle E_{closed}* \rangle = 0.905 \pm 0.001$ and $\sigma^{r}_{closed} = 0.28 \pm 0.02$ nm; $\chi^{2}_{r} = 1.76$), shot noise alone could not account for the widths of each distribution (red line; $\langle E_{open}* \rangle = 0.285$, $\langle E_{closed}* \rangle = 0.905$, $\chi^{2}_{r} = 5.90$).

Chapter 5

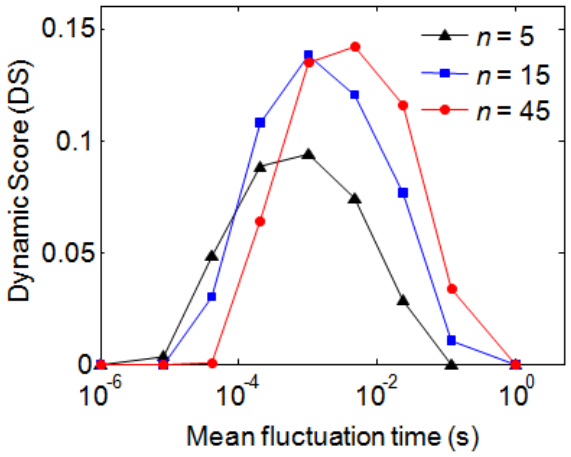


Figure S1. Sensitivity of BVA to dynamics as a function of photon window size, *n*. As in Fig. 3 D,E, we plotted the Dynamic Score (DS) for a series of simulated two-state dynamic species with different fluctuation timescales, while varying the photon window size, *n*, used to perform BVA. DS is a least-squares-like measure of the difference between the expected (σ_{E*}) and observed (s_{E*}) standard deviations over all values of $E*$ for which dynamics are significant, Eq. 7, such that all nonzero points indicate significant evidence for dynamics, and the larger the DS, the greater the evidence for dynamics. We found that increasing the photon window size has two effects: it increases the DS around the diffusion timescale, and slightly translates the DS plot to slower timescales. This translation results in detection of slow dynamics that were not previously detectable (~10^{-1} s timescale), but failure to identify faster dynamics (~10^{-5} s) that had been detected at the smaller window size. As a result, no photon window is optimal for detecting dynamics on all timescales, but faster or slower dynamics may be more easily detected with smaller or larger photon windows, respectively. Altering photon window size may therefore be useful for extending the range of detectable dynamic timescales with BVA.

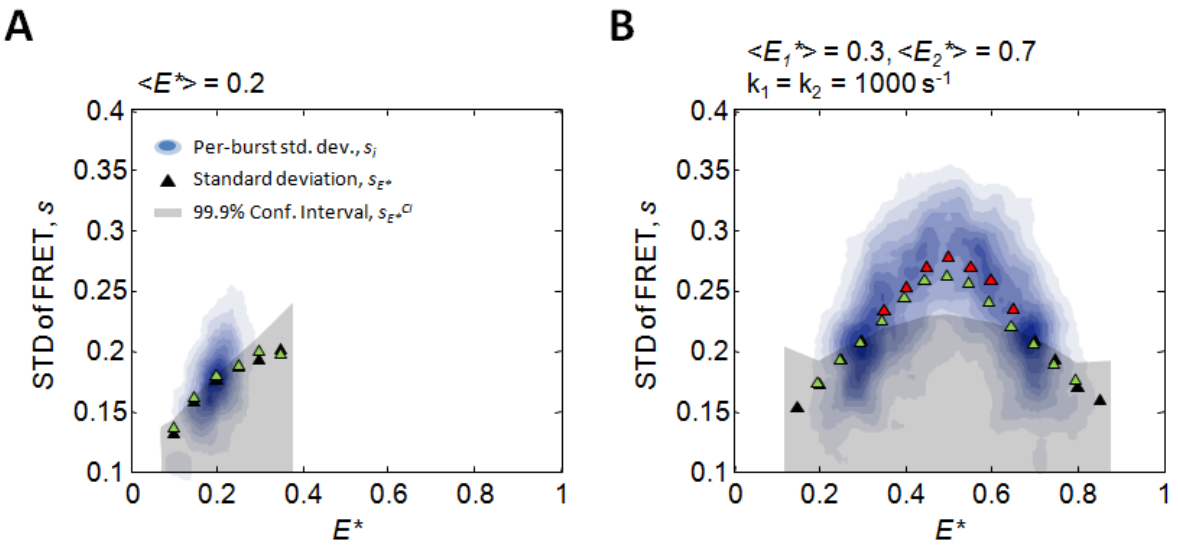


Figure S2. Effect of background on BVA. To illustrate the effects of background on BVA, we first simulated (A) a constant-FRET control sample (<*E**> = *0.2*), and (*B*) a strongly dynamic sample (same as in Fig. 3B). We then repeated these simulations, but added a strong background signal (10kHz in the green-excitation green-emission channel, 6kHz in green-excitation red-emission, more than twice that of typical experimental background counts). We performed BVA on the no-background simulations and plotted them, then overlaid the s_{E*} of the high background simulations (green triangles). Whereas the static simulation showed little difference between the s_{E*} of samples with and without background, the dynamic simulation showed a small but consistent decrease in s_{E*} at intermediate values of *E**, where dynamics were clearest; this is due to the fact that background, like static species, emits photons in a binomial fashion with respect to the donor and acceptor emission channels, lowering dynamic s_{E*} values. When using BVA, high background can therefore produce false negatives for dynamics, though it will not generate false positives. We note that background, if high enough to warrant it, can be incorporated into the "Arrival-Time PDA" approach.

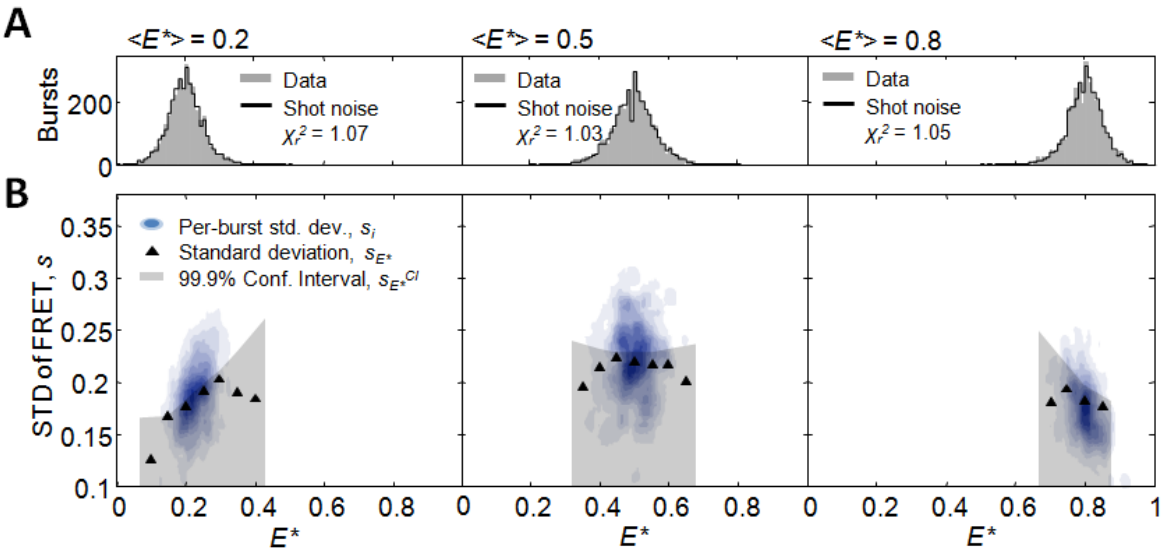


Figure S3. Simulations show good agreement between s_{E^*} and confidence intervals for simulated static FRET species. Simulations were carried out for FRET pairs with $\langle E^* \rangle$ values of either 0.2, 0.5 or 0.8. (A) Probability Distribution Analysis (PDA) was used to confirm that simulated data exhibits no broadening beyond the predicted shot-noise limited histogram for each value of $\langle E^* \rangle$ ($\chi_r^2 \approx 1$ in all cases). (B) These data were subsequently analyzed by BVA, using a photon window size of *n=5*. In all cases, s_{E^*} values (triangles) fall below the strict, 99.9% one-tail confidence interval (gray region), as expected for static FRET species.

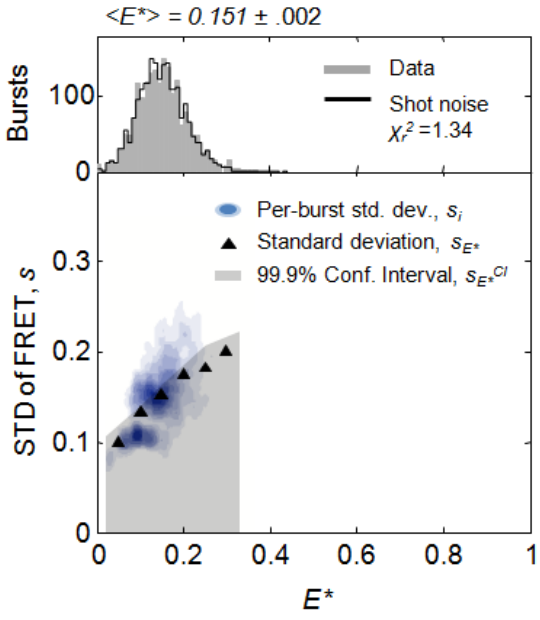


Figure S4. PDA and BVA analysis of a donor-only FRET control. For donor-only-labelled molecules, the apparent proximity ratio is due only to spectral leakage of donor emission into the acceptor detection channel; it should therefore exhibit dynamics only if there are dynamic donor spectral shifts, or mismatches in spatial detection efficiency between the donor and acceptor detection channels. PDA shows a shot noise-limited E^* distribution ($\langle E^* \rangle = 0.15 \pm .0018$, $\chi_r^2 = 1.34$; top panel), while BVA shows consistency of the observed s_{E^*} with a static underlying FRET value, as expected (bottom panel).

Chapter 6

Std45-90

5′-TAAATCTAAAGTAACATAAGGTAACATAACGTAAGCTCATTCGCG-3′

3′-ATTTAGATTTCATTGTATTCCATTGTATTGCATTCGAGTAAGCGC-3′

Lac DNA

-15 -5

5′-AGGCTTGACACTTTATGCTTCGGCTCGTATAATGTGTGGAATTGTGAGAGCGGATAACAATTTC-3′

3′-TCCGAACTGTGAAATACGAAGCCGAGCATATTACACTCCTTAACACTCTCGCCTATTGTTAAAG-5′

-3 +15

Oligo A

-5

5′-TGCTTCGGCTCGTATAATGTGTGGAATTGTGAGAGCGGATAACAATTTCGCGC T T

3′-CGCCTATTGTTAAAGCGCG T

Oligo B

T

T GCGCAGGCTTGACACTTTA-3′

-3

T CGCGTCCGAACTGTGAAATACGAAGCCGAGCATATTACACTCCTTAACACTCT-5′

Oligo C

-3

3′-A*C*G*A*AGCCGAGCATATTACACTCCTTAACACTCT-5′

Figure S1. DNAs used in Chapter 6. Green and red shading indicate labelling with Cy3B and Atto647N respectively; all labels are attached to the DNA via an amino group at the end of a 6-carbon linker. The lac DNA shows two Cy3B and two Atto647N labelling positions, along with a number (-15, +5, -3, +15); DNAs in the text are labelled as $lac^{Cy3B(X),A647N(Y)}$ to indicate the labelled positions. Oligos A and B, or A and C, are annealed and ligated to yield the lac dumbbell ($lac^{DB,Cy3B(-5),A647N(-3)}$) and lac hairpin constructs ($lac^{HP,Cy3B(-5),A647N(-3)}$), respectively. The ligated constructs have sequences identical to that of the unmodified lac DNA, but with either a hairpin (sequence 5'-GCGCTTTGCGC-3') or four phosphorothioate linkages (indicated by *) at each end.

[26] Gansen A, Valeri A, Hauger F, Felekyan S, Kalinin S, et al. (2009) Nucleosome disassembly intermediates characterized by single-molecule FRET. *Proc Natl Acad Sci U S A* 106: 15308-15313

[27] Rothwell PJ, Mitaksov V, Waksman G (2005) Motions of the fingers subdomain of klentaq1 are fast and not rate limiting: Implications for the molecular basis of fidelity in DNA polymerases. *Mol Cell* 19: 345-355

[28] Joyce CM, Benkovic SJ (2004) DNA polymerase fidelity: Kinetics, structure, and checkpoints. *Biochemistry (Mosc)* 43: 14317-14324

[29] Chung HS, Louis JM, Eaton WA (2010) Distinguishing between Protein Dynamics and Dye Photophysics in Single-Molecule FRET Experiments. *Biophys J* 98: 696-706

[30] Johnson SJ, Taylor JS, Beese LS (2003) Processive DNA synthesis observed in a polymerase crystal suggests a mechanism for the prevention of frameshift mutations. *Proc Natl Acad Sci U S A* 100: 3895-3900