Contents

Chapter 1

Thesis Overview

1.1 General Introduction

1.2 Contributions

1.3 Thesis Overview

The rest of this thesis is structured as follows. Chapter 2 provides a introduction to other research that has been undertaken in the field of fluorescence microscopy. General experimental and analysis techniques for fluorescence microscopy of single molecules are introduced, and the work presented in this thesis is contextualised. Chapter 2 also provides an overview of current research in Bayesian statistics and probabilistic analysis, introducing the statistical methods that are used in later chapters. Following this introductory section, we present our results over four separate chapters.

Chapter 3 introduces the pyFRET library, which we developed for analysis of confocal smFRET data. We describe the theory and implementation of different analysis algorithms for smFRET datasets. Data from both continuous and alternating excitation experiments are considered. In the second part of Chapter 3, we provide a comprehensive evaluation of different smFRET analysis algorithms, using a combination of simulated and experimental datasets. We benchmark popular analysis techniques, demonstrating their relative utility under different data collection regimes.

Chapter 4 considers a Bayesian method for the analysis of data from continuous excitation smFRET datasets. First, we introduce a model-based theory of the smFRET experiment. Then, we describe how this parametric model can be used to infer intramolecular distances and population sizes from time-binned smFRET data. We benchmark our Bayesian analysis technique against thresholding techniques used for time-binned data, showing the superior performance of the inference technique.

Chapter 5 extends this Bayesian analysis to sizing of labelled protein aggregates. We describe how a simplified model can be used to describe photon emission from multiple fluorophores. Using a combination of real and simulated datasets, we then show that this model is degenerate, making inference of aggregate sizes undecidable. We further show that a single emission event has multiple sources of heterogeneity, creating a complex, non-linear relationship between aggregate size and the number of photons emitted in a fluorescent event.

Chapter 6, the final results chapter, is somewhat different. This chapter describes a Bayesian analysis tool for error correction in genome assemblies generated using Illumina Nextera mate pairs. Illumina sequencing technology uses fluorescence emission from multiple fluorophores in its base calling algorithm. However, base calling errors and complex repeat structures can make reassembly of short reads challenging. In this chapter, we introduce NxRepair, an error correction tool that can identify mistakes in *d*e novo assemblies of bacterial genomes. We benchmark NxRepair against existing tools, demonstrating its superior performance.

Finally, Chapter 7 provides the conclusion to the thesis. Here, we summarise the overall contribution of this thesis and relate the work described to its wider research context. We also discuss possible extensions of the research described, indicating future applications of the research.