Single-Molecule Mapping and Heterogeneous Dynamics of Epigenetic Modifications in Live Microbes

by

Ziyuan Chen

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biophysics)
in The University of Michigan
2023

Doctoral Committee:

Professor Julie S. Biteen, Chair
Professor Lyle A. Simmons
Assistant Professor Jonathan Terhorst
Professor Sarah L. Veatch
Associate Professor Qiong Yang

Ziyuan Chen

ziyuanc@umich.edu

ORCID iD: 0000-0003-2832-945X

This Thesis is dedicated to my parents and my sister.

ACKNOWLEDGEMENTS

It was a long long journey to this final end. During these 5 years, I have learned a lot, experienced a lot, and grown a lot. I would like to use this chance to acknowledge colleagues, family and friends for your company and help during these years.

First of all, I want to thank my advisor Dr. Julie Biteen for her guidance in science and in life. I first met Julie 6 years ago when I was an undergraduate visiting student. Her kindness and passion for science encouraged me to come back to Michigan and pursue this degree. Under her help, I have grown from a nervous international student into an independent scientist. Julie understands and respects my needs, my puzzle and my passion for the science I am interested in, and she has done everything she can to help me succeed. The freedom and support Julie gives me enabled me to explore all different fields. She is willing to spend hours and hours with me talking about anything from statistical models to my grammar errors in writing. I would also like to use this chance to thank my collaborators Dr. Lyle Simmons, Dr. Kaushik Ragunathan and Dr. Jonathan Terhorst. Thank you all for letting me understand the knowledge in these interesting fields outside our lab, and for the help and enlightening discussions we had during the pushing of our collaborated projects. Meanwhile I would like to thank my other committee members Dr. Sarah Veatch and Dr. Qiong Yang for the feedback and care on my science and career development.

I would like to thank all my labmates, previous and current ones, for your company and help during these 5 years. I will miss our afternoon casual talks or morning intense scientific discussions, I will miss the happy times we spent together at all the different activities. I would like to first thank Yilai who was my mentor and still a very good friend. I would like to thank current Biteen lab members, Anna, Zech, Chris, Saaj, Kate, David, Xiaofeng, Lauren, Luis and Daniel for the help and fun we bring each other. I would also like to use this chance to thank all the previous members which I am really fortunate to have the chance to work with. I remember when I was new in the lab, how Josh, Stephen and Laurent helped me deal with all the lab puzzles, also I remember Curly, Hannah and Aathmaja's kindness and patience with me to help me grow into an independent graduate student. It is my pleasure to spend my graduate school years with these wonderful labmates.

I also would like to thank the Biophysics program which make me feel like home. I would like to thank all the faculties and staffs, especially our student coordinator Sara Grosky for her help from the begining of my enrollment to this final end near graduation. I would like also to thank other Biophysics graduate students for the inspiring discussions and fun program activities together.

At this very special time I would like also to thank all my friends, it is hard to live abroad for such a long time, and COVID has disrupted everything in the middle of my graduate school years, without the companionship with all of you, I could not walk to this end. Thank you Guoming, Sicong, Katie, Minjun, Renjie, Zhenyu and so many more.

Finally, and most importantly, I want to thank my parents and my sister. They support me unconditionally and from them I find the love and courage to keep moving forward. I am eight thousand miles away from home, but when I hear their voice on the phone, I am never alone.

TABLE OF CONTENTS

CHAPTER

LIST OF FIGURES

Figure

LIST OF TABLES

Table

LIST OF ABBREVIATIONS

3C chromosome conformation capture

ACF autocorrelation function

BM Brownian Motion

B. theta *Bacteroides thetaiotaomicron*

BSL Bayesian synthetic likelihood

B. subtilis *Bacillus subtilis*

CD chromodomain

CSD chromoshadow domain

ChIP-qPCR chromatin immunoprecipitation–quantitative polymerase chain reaction

ChIP-Seq Chromatin Immunoprecipitation Sequencing

CTRW Continuous Time Random Walk

DP Dirichlet Process

dsDNA double-stranded DNA

E. coli *Escherichia coli*

EMCCD Electron Multiplying Charge-Coupled Device

EC50 50% effective concentration

EMSAs electrophoretic mobility shift assays

EOP efficiency of plaquing

FBM Fractional Brownian motion

FCS Fluorescence Correlation Spectroscopy

FRET Fluorescence Resonance Energy Transfer

FRAP fluorescence recovery after photobleaching

H3K9me Histone H3 lysine 9 methylation

HDP Hierarchical Dirichlet Process

HMM Hidden Markov Models

HDP-HMM Hierarchical Dirichlet Process Hidden Markov Model

HPUra 6-(p-hydroxyphenylazo)-uracil

IP-MS Immunoprecipitation followed by mass spectrometry

LSTM long short-term memory

LW Lévy Walk

MCMC Markov Chain Monte Carlo

MTase methyltransferases

MM minimal medium

m6A N6-methyladenosine

MSD mean squared displacement

NHD Normalized Hamming Distance

NOBIAS NOnparametric Bayesian Inference for Anomalous diffusion in Single-molecule tracking

NIW Normal-inverse-Wishart distribution

OD Optical Density

REase restriction endonuclease

RM restriction modification

RNN Recurrent Neural Network

SMAUG single-molecule analysis by unsupervised Gibbs

SMRT single-molecule real-time

S. pombe *Schizosaccharomyces pombe*

SPT Single-Particle Tracking

sptPALM single-particle tracking Photoactivated Localization Microscopy

ssDNA single-stranded DNA

ssRNA single-stranded RNA

WAIC widely applicable information criterion

WT wild-type

ABSTRACT

Single-Particle Tracking (SPT) in living cells informs the dynamics of target molecules and enables the investigation of their functions and interactions with other components in the cell. The quantitative analysis of SPT trajectories from living cells traditionally relies on the Brownian diffusion model. For molecules with homogenous dynamics or in well-studied biological systems whose biophysical mobility states are predictable, SPT analysis is robust and the mobility states of the molecules can be related to their biological functions in cells. However, for complex or poorly understood biological systems such as epigenetic modification systems, an objective SPT analysis method that quantify the heterogeneous dynamics of target molecules is keenly needed to investigate their functions and interactions *in vivo*.

In this Dissertation, I develop a single-molecule tracking analysis framework with nonparametric Bayesian statistics and anomalous diffusion models to investigate epigenetic modifications in live bacterial and yeast cells. Chapter II presents a new SPT analysis method combining nonparametric Bayesian statistics and a supervised recurrent neural network. The method is named NOnparametric Bayesian Inference for Anomalous diffusion in Single-molecule tracking (NOBIAS). The performance of NOBIAS is validated with simulated datasets of heterogeneous dynamics, asymmetric diffusion, and a mixture of anomalous diffusion models. NOBIAS is also applied to experimental datasets from live cells and identifies anomalous diffusion and asymmetric diffusion in the systems.

DNA methylation in bacterial cells is a marker for specific protein-DNA interactions. DnmA is a recently characterized DNA methyltransferases (MTase) in *Bacillus subtilis*, responsible for all detectable N6-methyladenosine DNA methylation. In Chapter III, I use single-molecule tracking and spatial mapping to study of DnmA in live *Bacillus subtilis*. The results show that DnmA is regulated by the DNA substrate and correlates with DNA replication and DNA-RNA hybrid cleavage. This work combines single-molecule imaging of DnmA and phage predation assays to identify that DnmA is functionally an orphan MTase regulating gene expression.

Epigenetic modifications in eukaryotes regulate the chromatin structure and the gene expression level. Histone H3 lysine 9 methylation (H3K9me) is a conserved epigenetic marker for heterochromatin and gene silencing. Epigenetic modifications rely on writer, reader, and eraser proteins to establish, maintain and remove modifications. In Chapters IV and V, I use single-molecule tracking and nonparametric Bayesian statistical analysis to understand the behaviors of these modifiers *in vivo*. In Chapter IV, I focus on the H3K9me reader protein, Swi6, in the fission yeast cell. I present the dynamics of Swi6 following different perturbations including knockouts of related proteins and the engineering of the Swi6 protein itself. I map Swi6's distinct mobility states onto their biological roles in living cells and show a high-specificity binding mechanism through Swi6 oligomerization. Chapter V presents the single-molecule dynamics of multiple H3K9me modifiers in fission yeast. By comparing the dynamics of these modifiers and centering on the two H3K9me reader proteins Swi6 and Chp2, I propose that chromatin plays an important role to reinforce interaction and complex assembly on its site rather than just an inserted platform for interaction.

Through this dissertation, I show a powerful and informative methodology combining live-cell single-molecule tracking and advanced statistical inference. The dissertation provides quantitative analysis, detailed statistical models, and their application to epigenetic modifications in bacterial and yeast cells. This methodology is also applicable to any system where *in vivo* single-molecule tracking is feasible.

CHAPTER I

Introduction

This dissertation expands the analysis of single-molecule tracking with nonparametric Bayesian statistics and an anomalous diffusion model and applies single-molecule tracking microscopy to investigate epigenetic modifications in live bacterial and yeast cells. Within this introduction chapter, I include the principles of single-molecule microscopy and how it can provide the *in vivo* dynamics of target molecules. I comment on current single-molecule tracking analysis methods and the Brownian diffusion model. I explain the principles of nonparametric Bayesian analysis and anomalous diffusion models, and how they can be applied to trajectory analysis to extract advanced and objective dynamics information. I provide biological background about epigenetics and the two specific systems I study in this dissertation, DNA methylation in *Bacillus subtilis (B. subtilis)* and Histone H3 lysine 9 methylation (H3K9me)in *Schizosaccharomyces pombe (S. pombe)*. Finally, I summarize the objective of each chapter within the dissertation.

1.1 Single-molecule tracking experiment and analysis

1.1.1 Fluorescence microscopy and single-molecule imaging

The motion and spatial organization of proteins in live cells enable and indicate their function. A fluorophore is a fluorescent molecule that can emit light upon light excitation. Conventional fluorescence microscopy is limited by the diffraction limit of light as emission fluorescence from a single fluorophore is a diffraction-limited spot, and typically only provides static overlapping snapshots of the target molecule, not the dynamics [1]. Quantitative fluorescence microscopy techniques such as Fluorescence Correlation Spectroscopy (FCS), fluorescence recovery after photobleaching (FRAP), and Fluorescence Resonance Energy Transfer (FRET) provide robust and quantitative data for investigating protein properties *in vivo* and *in vitro* [2]. With the progress of hardware in microscopes, fluorophore performance [3–5], ways to label fluorophores [6], and especially new designs of microscopy [7–10], higher spatiotemporal resolution for proteins in live cells is acquirable. Within these methods, single-molecule tracking *in vivo* is especially powerful to reveal dynamic information for the protein of interest.

Single-molecule tracking is achieved by imaging sparse densities of target molecules to localize one molecule at a time to beat the diffraction limit and avoid overlapping this cycle (Figure 1.1) [11]. In conventional fluorescence microscopy, the sample is excited by a laser, and the emission light from the fluorophores in the sample is collected by the camera. To achieve sparse densities, I use the single-particle tracking Photoactivated Localization Microscopy (sptPALM) method with an additional activation laser with a connected optical shutter to turn on the activation laser for a short activation time meanwhile turning off the excitation laser (Figure 1.1). The emission of sparse single-molecule fluorescence is collected by an Electron Multiplying Charge-Coupled Device (EMCCD) camera to be converted into digital images and movies.

To localize single molecules, the fluorescence images from collected movies are first screened in a guessing algorithm that looks for local regions with high intensity that are potentially a single molecule, then the regions of interest are fitted by a 2D Gaussian to determine if the candidate is a single molecule. The location of the molecule is acquired as the fitted peak position of the 2D Gaussian fitting, and the uncertainty of this fit location is the precision or localization error of the single-molecule position [12]. After this step, the original fluorescence images series has been turned into very sparse point sets with

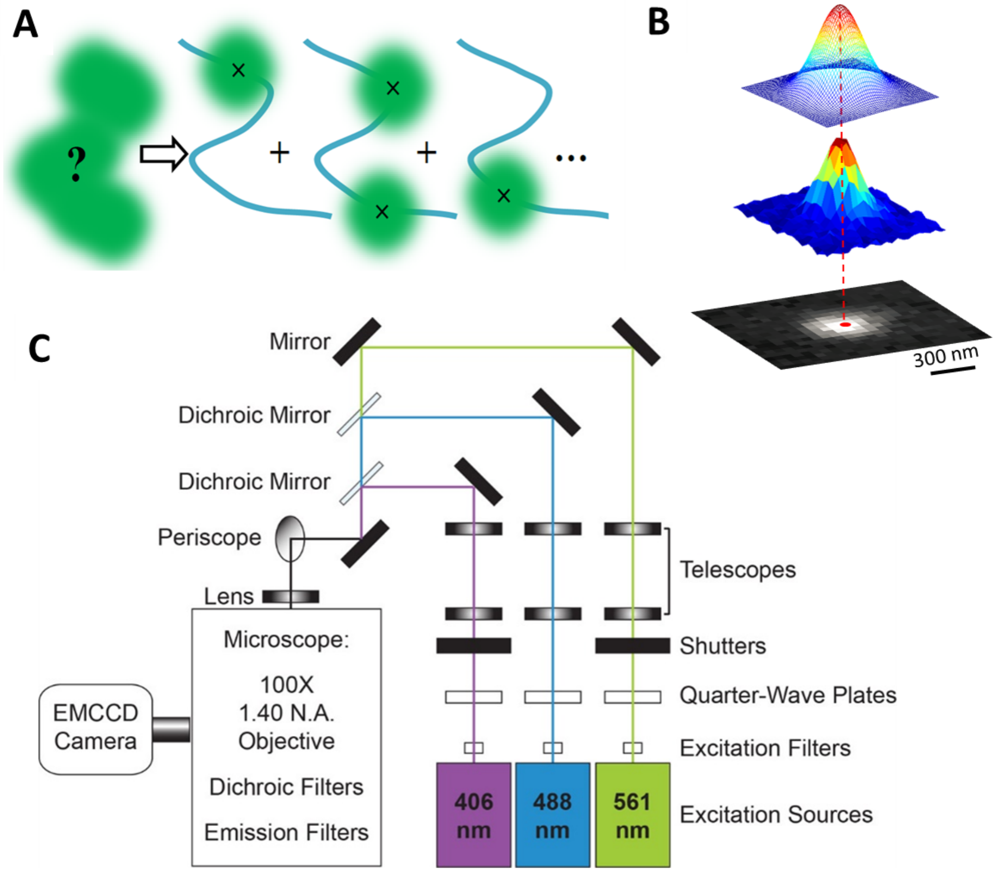


Figure 1.1: (A) imaging sparse single molecules to beat the diffraction limit of bulk fluorescence imaging. (B) Fitting single-molecule fluorescence signal to a 2D Gaussian to precisely localize peak position. (C) Schematic experimental layout of single-molecule microscopy used in this dissertation, images from Tuson and Biteen [1]

time and locations of well-fit molecules. Then tracking algorithm connects this 3D set of points with (x, y, t) into trajectory datasets based on how close the points are in space and in time and within the bounds of a preset threshold [12].

The final step of the quantification of the single-molecule dynamics of a target molecule is the analysis of the dynamics within these tracks. Despite the different biochemical functions molecules carry out within the crowded and complicated live cellular environment, conventionally Single-Particle Tracking (SPT) trajectories are analyzed with the assumption of Brownian motion [13,14]. The diffusion coefficientss(D) is defined by Fick's law of diffusion $J = -D\nabla\phi$. Where J denotes the diffusion flux, ϕ is the concentration. In Brownian motion, the mean squared displacement (MSD) of the track is linearly proportional to the time lag (τ) and D with the following relation: $MSD = 2nD\tau$, where n is the dimensionality of the trajectories ($n = 2$ in the two-dimensional experiments described in this thesis). The simplicity and elegant statistical properties of Brownian motion enable fast and easy-to-implement analysis of SPT datasets. The oldest but still very powerful approach for SPT analysis is to analyze each trajectory: a linear regression of the time-averaged MSD of the trajectory and τ is an estimator of D [13] (Figure 1.2). The distribution of D shows the occupancy of molecules with different diffusion coefficients, which could correspond to distinct diffusive patterns of the molecules. For example, the target protein could be bound or unbound in the live cells, and the single-trajectory D histogram would be a two-peak distribution with a lower D peak for bound molecules and a higher D for unbound ones. The comparison of the D distribution upon perturbations to the system—for example mutating the molecule or changing its environment— allows us to monitor the change in the dynamic behavior of the molecule. Because it addresses the issue of correlation between adjacent steps in the MSD estimator, the covariance-based estimator (CVE) is also widely applied for trajectories collected in camera systems [15].

Single-trajectory MSD calculation of diffusion coefficients is straightforward but greatly constrained by the assumption that one D value can represent the diffusion of the molecule

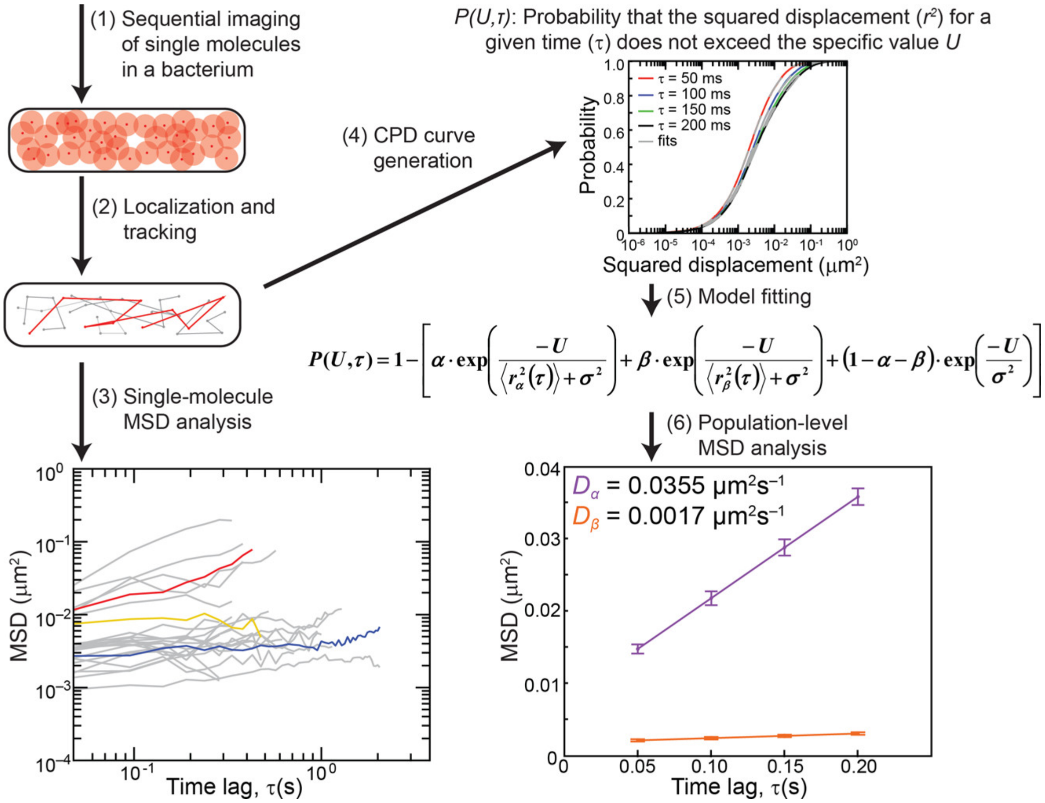

(1) Sequential imaging of single molecules in a bacterium
(2) Localization and tracking
(3) Single-molecule MSD analysis
(4) CPD curve generation
$P(U,\tau)$: Probability that the squared displacement (r^2) for a given time (τ) does not exceed the specific value U
τ = 50 ms
τ = 100 ms
τ = 150 ms
τ = 200 ms
fits
Probability
Squared displacement (μm²)
(5) Model fitting
$P(U,\tau)=1-\left[\alpha\cdot\exp\left(\frac{-U}{\langle r_\alpha^2(\tau)\rangle+\sigma^2}\right)+\beta\cdot\exp\left(\frac{-U}{\langle r_\beta^2(\tau)\rangle+\sigma^2}\right)+(1-\alpha-\beta)\cdot\exp\left(\frac{-U}{\sigma^2}\right)\right]$
(6) Population-level MSD analysis
D_α = 0.0355 μm²s⁻¹
D_β = 0.0017 μm²s⁻¹
MSD (μm²)
Time lag, τ(s)

Figure 1.2: 1) Isolated fluorophores are sequentially imaged in a live cell. (2) Each single molecule is localized in each imaging frame by fitting with a 2D Gaussian to estimate its peak position, and localizations are linked into single-molecule trajectories. (3) The MSD for each time lag (τ) is calculated; the slope of the MSD vs. τ curve is proportional to the average diffusion coefficient of each trajectory. (4-6) Single-molecule displacements at different τ are considered together in probability models and fit into a two-state kinetics model to reveal a slow and a fast mobility state. This figure is from Tuson and Biteen [1].

within this track. This assumption could be easily violated as the trajectory could be capturing a molecule that undergoes a transition between different functions. For example, a DNA-binding protein could unbind from DNA during the monitored time window of a trajectory. Under similar cases, the mobility of the molecule could undergo great transitions and cannot be captured by one D value. Within the context of live cell imaging, these trajectories with mobility transitions could provide crucial information about reaction kinetics. The other limitation of the single-trajectory MSD method is that for every single trajectory once fitted through the MSD linear regression, the uncertainty from the linear fitting is very hard to further take into consideration for the distribution of all D. Linear regression typically applies R-squared (R^2) of fitting as the threshold of fitting quality. Due to the limited track length, the arbitrary R^2-based threshold leads to loss of data and biased D value. To overcome these limitations, biophysicists come up with the ensemble-averaging method [16–18]. Instead of analyzing each trajectory individually, displacements of all tracks at different τ are measured and pooled together to be fit for the probability density distribution of the Brownian motion. These methods are typically referred to as kinetics model fitting (Figure 1.2). The ensemble averaging methods have overcome two limitations in the single-trajectories analysis: first, it can fit into multiple mobility states with additive terms in the probability model without assuming that one track has only one mobility state; second, it fits all SPT data together with probability model inference and gives robust uncertainty measurement.

Although kinetics model fitting is powerful and robust, one challenge for it is that the probability model of displacement has to be predetermined [18]. Specifically, the probability distribution of the single-molecule displacement needs prior knowledge of the number of mobility states, which is often unknown for biomolecules in live cells. To consider the question statistically, if Brownian motion is assumed, the displacements of single-molecule tracks from the same mobility state follow a 2D Gaussian distribution with a mean value of (0,0), and the variance of this 2D Gaussian is $4D\tau$. The potential heterogeneity within the tracks with an unknown number of mobility states leads to a Gaussian mixture model for the entire displacement distribution. The tracks are temporal data which allows us to apply time-series and Markov models to solve this Gaussian mixture model. Considering the statistical features of the SPT dataset, nonparametric Bayes statistics will be a powerful tool to objectively determine the number of mobility states and infer the associated parameters.

In addition to these mentioned localization and trajectory based methods, there are also many other methods to quantify dynamics, such as FCS based methods and mapping between diffusivity and cellular structure [19–22]. Other experimental efforts, such as structured illumination modulation and minimal photon fluxes, have also improved the spatiotemporal resolution of single-molecule tracking experiments [23–27]. Within this dissertation I focus on the sptPALM approach and analysis, and the analysis principle could be applied to related SPT experiments.

1.1.2 Nonparametric Bayesian statistics for SPT

In model fitting with frequentist inference, the unknown parameter value is fixed. For example, in the single-molecule tracking analysis, the single-trajectory MSD linear regression and the ensemble average kinetics model both belong to frequentist inference, where there are fixed parameters to be estimated. However, the model for the dynamics of the motion is not certainly known. For example, the number of mobility states and the diffusion model for each diffusive state are often unknown to us. Bayesian statistics is a different philosophy that considers parameters as random variables and assigns probabilities to parameters. Bayesian statistics estimate parameters' distribution given the data and some prior information about the system. The basis of Bayesian statistics is Bayes theorem as below:

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} \tag{1.1}$$

Where θ is the parameter in the model and y is observed data. The posterior distribu-

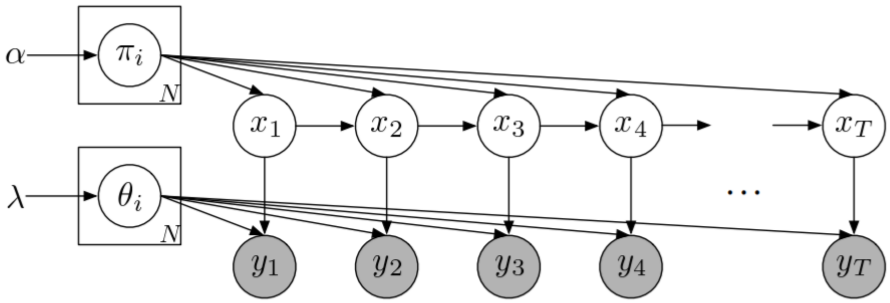


Figure 1.3: x_t in the framework is the hidden layer, which denotes the state at the time, y_t is the emission layer with the observed data. The transition between x_t is determined by π_i, the transition matrix, and the emission layer and hidden layer are connected with emission parameter θ_i. The illustration is from Johnson, et al. [32]

tion for the parameter $p(\theta|y)$ is determined based on the data likelihood $p(y|\theta)$ and the prior information $p(\theta)$. The parameter is estimated through sampling from the posterior distribution based on prior data.

When pooling 2D displacements from all trajectories at different time lags in the kinetic model fitting method, the temporal information that some steps are adjacent is lost. To make full use of the temporal information, the Hidden Markov Models (HMM) can be applied for heterogeneous single-molecule tracking datasets under the Brownian motion assumption [28]. HMM is a Bayesian model for time series data with two-layer: the emission layer and the hidden layer (Figure 1.3). In the scenario of single-molecule tracking, the emission layer is the displacement or the observed data, and the hidden layer is the corresponding Brownian diffusion mobility states. The emission at each step only depends on that state's diffusion coefficient, and the transition between hidden states in time is captured by a transition matrix [29–31].

HMM alone provides a time-series Bayesian approach to estimate the posterior for D of each state, the transition between states, and the assignment of each single-molecule displacement their states, however, it still needs a predetermined number of states. To tackle this problem, the Dirichlet process is introduced as the prior in the HMM to build a Hierarchical Dirichlet Process Hidden Markov Model (HDP-HMM) which can objectively determine the number of states in the HMM model give the emission dataset [33,34]. In an HDP-HMM model, the number of states in the prior is unbound and to be estimated from the data, and is flexible to be further learned providing more data. HMM can be sampled through Gibbs sampler or mean field variational inference through iterative forward and backward message-sending algorithms [35,36]. For HDP-HMM, a sticky parameter can be added to prevent over-splitting and fast switching of states, and weak-limit approximation can be applied to accelerate the algorithm [37].

1.1.3 Anomalous diffusion and neural network classification

Usually, Brownian Motion (BM) is assumed for SPT datasets due to the limitation of data quality and limited prior knowledge about the biological system. The mathematical simplicity of Brownian motion enables robust and fast quantification of SPT data [13,14], however, in living cells, the motion of target molecules is typically under specific functions and affected by the crowding cellular environment. More informative and realistic motion models should be applied to understand the dynamics of biomolecules in live cells. Anomalous diffusion models contain more information and capture the living molecules better than simplified BM [38]. MSD of anomalous diffusion has the following relation with τ:

$$MSD = 2nD\tau^{\alpha} \tag{1.2}$$

Where different from the linear relation of BM, there is exponent index $0 < \alpha < 2$. Anomalous diffusion with $\alpha < 1$ is subdiffusion and diffusion with $\alpha > 1$ is referred to as superdiffusion (Figure 1.4). BM is a special case of anomalous diffusion with $\alpha = 1$. For living molecules, the crowded cellular environment or their specific functions in the cell result in the finding that nearly all molecular motion is anomalous [39–41].

The characterization of anomalous diffusion models from SPT trajectories is a challenging task. Physicists try to address this challenge with developments of diffusion models and advances in computer science specifically in neural networks [43–45]. In a 2021

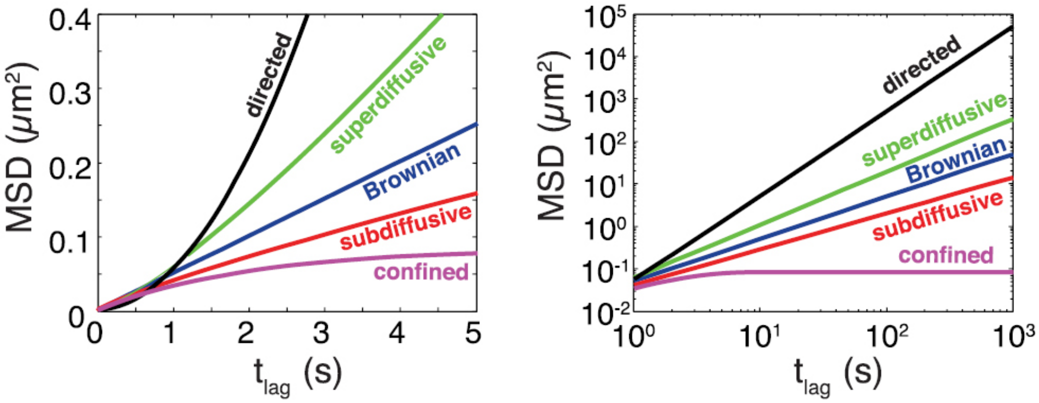


Figure 1.4: The relation between MSD and the time lag for different diffusion models is shown in left (normal scale) and right (log-log scale). Different diffusion models are in different colors: Directed (black), superdiffusion (green), Brownian (blue), subdiffusion (red), and confined diffusion (purple). This figure is from Manzo et al. [42]

competition for anomalous diffusion classification, ANDI, neural network-based methods showed excellent performance [46, 47]. This competition result indicates the direction for the future of anomalous diffusion analysis: the combination of physics and stochastic process theory with exceptional computational power provided by the neural network [46, 47]. The recurrent neural network could not only be applied to classify the diffusion type given the tracks data but also can potentially provide a regression model to estimate the anomalous exponent α and to detect the transition step of diffusive models within one trajectory [45]. The performance of a supervised neural network greatly depends on the training of the network. Simulation of specific anomalous diffusion models can provide huge training sets. The performance of the trained model on experimental data greatly depends on the length of each single-molecule trajectory, which can range from hundreds of steps for the SPT dataset of polymer or material science to only tens of steps for biosamples and live cell systems. The expertise regarding the biological systems to predict candidate anomalous diffusion models of the trajectories could narrow the training set down to increase the neural network performance.

1.2 Epigenetic modifications

Gene or DNA sequence encoding protein information is the fundamental basic unit of an organism, however, the behavior of the living system is not completely determined through genes. There exist phenotype changes that are independent of the DNA sequence alteration, and these changes play important roles in the living systems and also could be inheritable through generations [48]. The regulation of gene expression without altering DNA sequence is called epigenetics, and it is conserved from prokaryotes to eukaryotes and in multiple forms including DNA modifications and histone modifications [49, 50]. There are various functions epigenetics modifications play in different organisms, in general DNA or histone methylation is a silencing marker while acetylation on the opposite way is an active transcription marker [51]. Specifically in bacterial cell DNA N6-methyladenosine (m6A) is well known as part of the Restriction-Modification systems for bacterial phage defense [52], but also m6A is known to participate in Bacterial DNA replication, repair, and transcription as well [53]. In eukaryotic cells, there are both DNA modifications and histone modifications. Long DNA is wrapped around histone octamers to form nucleosomes and further form chromatins, including gene transcription active euchromatin and silenced heterochromatin. The histone epigenetic modifications are the key to the regulation of the structure of euchromatin and heterochromatin thus regulating the structure of the entire genome [54]. All these different epigenetics modifications are carried out by epigenetics modification proteins, which can come in three types, writer (like methyltransferase, acetylase), reader, and eraser (demethylase) [55]. The study of how these modification proteins behave in the living cell would be crucial for us to understand the epigenetic modification mechanism. Conventional studies of epigenetic modifications are carried out with *in vitro* biochemical assays, targeting the modified substrates and phenotypes to understand the role of each modification [56–58]. The advances in the single-molecule microscopy and the fluorophore labeling methods enable the direct *in vivo* observation of epigenetic modification proteins motions. The dynamics of these modifiers inform on their function and regulatory mechanism for the corresponding epigenetic modifications. In this dissertation, I focus on epigenetic modifications in two different organisms: prokaryote bacteria *Bacillus subtilis* DNA m6A methylation and the eukaryote *S. pombe* H3K9me.

1.2.1 DNA methyltransferase DnmA in *Bacillus subtilis*

As prokaryotic genomes are not organized in the units of nucleosomes, DNA methylation is the predominant form of epigenetic modifications, which specifically includes N6-methyladenine (m6A), 5-methylcytosine (m5C), and N4-methylcytosine (m4C) [59]. m5C is well-studied in eukaryotes regarding gene regulation and cell development. Bacterial genomes harbor all these 3 types of DNA modifications, and currently, the most well-studied DNA methylation system is the restriction-modification (RM) system for the defense of bacteria against bacteriophage infection [52]. DNA methylation is modified through DNA methyltransferases (MTase) by adding a methyl group to a nitrogenous base in a sequence-specific context [60]. The unmethylated DNA sequence is a marker for the restriction endonuclease to degrade the target ectopic gene [52]. MTases outside RM systems which don't have a paired endonuclease are called orphan MTases, and two well-studied orphan MTases, the Dam methylase of *Escherichia coli* and CcrM methylase of *Caulobacter crescentus*, have been shown to methylate DNA which regulates many physiological processes including the cell cycle, DNA replication, mismatch repair, and epigenetic gene expression [53, 61]. The development of the single-molecule real-time (SMRT) sequencing method provides a powerful platform for the detection of methylation sites [62]. Bioinformatic approaches are then used to identify the gene that encodes the methyltransferase. The function and regulatory mechanism of m6A in gram-positive bacteria are much less studied compared with gram-negatives. SMRT sequencing was used to show that the newly identified MTase DnmA is responsible for all the detectable m6A methylation in the genome of *Bacillus subtilis*, a prototypical gram-positive bacterium [63]. DnmA specifically recognizes nonpalindromic 6 base pair sequence 5′GACGAG-3′ and methylates the 5th adenine. DnmA is responsible for all detectable m6A methylation within *Bacillus subtilis*. The 5′-GACGAG-3′ m6A methylation has been shown to regulate promoters in *Bacillus subtilis* PY79 strain, and the absence of this m6A results in increased binding of ScoC, a transcriptional repressor, to the promoter of several genes including *scpA* [63]. This mechanism indicates that m6A in *B. subtilis* could regulate gene expression, however, other roles of m6A remains unclear. Whether DnmA is part of the RM system or is an orphan MTase is also unknown. It is shown that 99.7% of the 5th adenine in 5′-GACGAG-3′ motifs has already been methylated. Considering the active replication of *B. subtilis* DNA, how the MTase DnmA cooperates with the DNA replication to maintain this high methylation level remains unknown. Direct observation of DnmA localization and dynamics at the single-molecule level within the live cell could provide crucial *in vivo* information to understand how DnmA and other MTases works in the live cell. Single-molecule imaging has been applied to study DNA polymerases and DNA repair proteins in *B. subtilis* [64–66], and the multicolor imaging method could help understand the correlation between DnmA and the DNA replication machinery.

1.2.2 H3K9me modification proteins in fission yeast

The chromatins of eukaryotes are divided into euchromatin and heterochromatin based on the density of nucleosomes. DNA sequences within euchromatin are transcription active and genes within heterochromatin are silenced as compact heterochromatin structure blocks RNA polymerase out. H3K9me is one of the most important conserved epigenetics markers for the formation of heterochromatin throughout all eukaryotes from fungi to mammals [54]. H3K9me is directly related to the regulation of chromatin structure maintenance and post-transcriptional silencing [48,67]. The successful establishment and maintenance of H3K9me play a central role in the integrity of the genomic structure [54]. The dysfunctional H3K9me pathway is shown to cause multiple diseases including cancers and inheritable gene diseases [68]. The H3K9me modification, like all the epigenetic modifications, is carried out through three types of histone modification proteins: writer, reader, and eraser proteins. Despite the H3K9me's importance and the studies that have already been carried out, the complexity of the system makes it still challenging for us to completely understand the H3K9me regulatory pathway. Various experimental tools including genetic functional assays, Chromatin Immunoprecipitation Sequencing (ChIP-Seq), chromosome conformation capture (3C) sequencing, biochemical pull down, *in vitro* calorimetry binding assays, and structure of key proteins, have been applied to investigate the H3K9me pathway [56, 58, 69, 70]. However, in addition to these efforts, *in vivo* information is keenly needed to understand the H3K9me regulation in living cells.

Schizosaccharomyces pombe or fission yeast is a commonly used model system to understand epigenetic modifications [71]. As a single-cell eukaryote, the H3K9me regulatory system of *S. pombe* is highly conserved. As *S. pombe* has no detectable DNA methylation, it is an ideal system to study histone methylation in eukaryote cells. Clr4, as a family conserved SUV49 SET histone methyltransferase family, is the solely H3K9 MTase in *S. pombe* and thus serves as the writer protein [72]. H3K9me reader proteins are the Heterochromatin Protein 1 family (HP1) proteins and there are two HP1 proteins in *S. pombe* Swi6 and Chp2 [73]. Swi6 is the more abundant HP1 protein in *S. pombe* and played an important role in ensuring chromatin integrity and inheritance. Swi6 in the cells exists in the form of a dimer, where each monomer consists of a chromodomain (CD), chromoshadowdomain (CSD), and a hinge region [74]. CD domain is the H3K9me recognition domain, and CSD is the dimerization interface and reported to mediate an oligomerization of Swi6 [75]. The hinge region of Swi6 does not have a fixed structure but its positively charged and could bind to nuclei acid like DNA or RNA [76]. Previous *in vivo* and *in vitro* fluorescence imaging data has shown HP1α proteins form foci that could colocalize with heterochromatin regions in the nucleus [77, 78]. Swi6 is also reported to potentially go through a liquid-liquid phase separation mechanism [79]. Swi6 can bind to Epe1, a putative H3K9me demethylase eraser and an anti-silencing factor [80,81]. Genetic sequencing and functional assays both show that Swi6 is necessary for the maintenance of epigenetic silencing, but how its function is realized in live cells remains barely known.

The other HP1 protein in *S. pombe*, Chp2, is 100-fold less highly expressed than Swi6 despite the shared structure: both have a CD domain for H3K9me recognition and a CSD domain as the dimerization interface [82]. Chp2's primary known function is to recruit the Snf2/Hdac Repressive Complex (SHREC), which is the fission yeast Nucleosome Remodeling and Deacetylase (NuRD) complex equivalent [83, 84]. The NuRD complex in mammalian cells is one of the remodeler complexes that play an important role in post-transcriptional silencing, genome integrity, and cell cycle regulation [85]. The SHREC complex in the *S. pombe* consists of two modules: the remodeler module which includes histone remodeler Mit1, and the deacetylase module which includes the H3K14 deacetylase Clr3 [83, 86, 87]. Chp2 is considered part of SHREC as well, along with two other proteins, Clr1 and Clr2, which connect the two modules of SHREC [87]. The structural information about the SHREC complex and the interaction between Chp2 and Mit1 have been thoroughly studied [83, 87], however, how these histone-modification proteins interact *in vivo* is still poorly understood. Clr3 and Epe1 are reported to be competitors for their pro-silencing and anti-silencing roles, and they are regulated through two different HP1 proteins Swi6 and Chp2 [88,89]. The relation between these two structurally similar HP1 proteins and their regulatory roles in H3K9me is not understood at the live cell level with high spatiotemporal resolution. In the scale of the entire chromatin, how the histone modification complexes assemble within the crowding chromatin environment and how multiple components from multiple complexes cooperate or compete with each other remains unknown. Studying the H3K9me-centered histone modification mechanism *in vivo* would open up a methodology for other similar complexes regarding epigenetic modifications and understand the role of heterochromatin in these complexes' assembly and function.

1.3 Thesis outline

This thesis aims to expand single-molecule tracking analysis to include nonparametric, asymmetric, and anomalous diffusion models. Applications of some features of the new analysis to several biological systems regarding epigenetic modifications demonstrate how powerful and informative the live-cell single-molecule tracking combined with statistical inference is.

In Chapter II, I present a new SPT analysis framework NOnparametric Bayesian Inference for Anomalous diffusion in Single-molecule tracking (NOBIAS) which combines nonparametric Bayesian statistics, anomalous diffusion models, and a neural network to analyze SPT datasets with heterogeneous, asymmetric dynamics and anomalous diffusion. I explain the statistical model for this framework which consists of a Bayesian module and a neural network module. Then I validate the performance of the framework with simulated SPT datasets and an experimental SPT dataset.

In Chapter III, I use single-molecule tracking and localization to understand the dynamics and spatial pattern of the DNA methyltransferase DnmA in *Bacillus subtilis*. I characterize the role of DnmA methylation and its relation with DNA replication with single-molecule tracking and mapping experiments under different disruptions, including replication arrest, aberrant DNA-RNA hybrids, and DNA-binding deficient mutants. My results suggest DnmA is regulated through the DNA methylation and RNA-DNA hybrid level.

In Chapter IV, I present the nonparametric Bayesian statistical analysis of the single-molecule dynamic of the HP1 protein Swi6 in the yeast model system *Schizosaccharomyces pombe*. Nonparametric Bayesian statistics reveal four biophysical mobility states. Through the deletion of related epigenetic proteins and the engineering of Swi6, the biochemical meanings of these four mobility states are determined. In the end, I conclude that a multivalence mechanism through Swi6 oligomerization enables its low-affinity high-specificity binding to H3K9me.

In Chapter V, I apply single-molecule tracking and the NOBIAS framework to understand the role of heterochromatin in HP1 protein complex assembly. I perform single-molecule tracking for the two HP1 proteins in *S. pombe*, Swi6 and Chp2, together with other proteins that form complexes with them. With NOBIAS dynamics analysis and spatial mapping of these proteins, I conclude that there exists a general mechanism for HP1 protein complex formation: H3K9 methylation enforces interactions and complex assembly at heterochromatin sites and attenuates off-site interactions.

Finally, in Chapter VI, I summarize the conclusions of this dissertation and present promising future research directions regarding the SPT analysis development and the application of SPT to study epigenetics in different organisms and systems beyond those mentioned in this dissertation.

CHAPTER II

Analyzing Anomalous Diffusion in Single-Molecule Tracks with Nonparametric Bayesian Inference

The work presented in this chapter was previously published in

Frontiers in Bioinformatics.

Chen, Z., Geffroy, L., and Biteen, J.S.
NOBIAS: Analyzing anomalous diffusion in single-molecule tracks
with nonparametric Bayesian inference
Frontiers in Bioinformatics, **1**: 742073 (2021). DOI: 10.3389/fbinf.2021.742073

In this work, I conceptualized the proposed statistical framework in the project. I implemented the algorithm of the framework into open access software. I validated the framework with my simulated data and experimental data provided by Lauren Geffroy.

2.1 Introduction

The biophysical dynamics of biomolecules reflect the biochemical interactions in the system, and these dynamics can be quantified within a dataset of single-particle trajectories obtained by tracking individual molecules. The invention of the super-resolution microscope [7–10, 90] and Single-Particle Tracking (SPT) methods [11, 91–93] have made possible investigations of biomolecular dynamics at a high temporal and spatial resolution both *in vitro* and *in vivo*. Moreover, quantitative algorithms can connect the real-time dynamics from biophysical trajectories to biochemical roles to uncover whether a molecule interacts with other cellular components [94], freely diffuses [95], is actively transported [96], or is constrained to a certain region [97].

Conventionally, SPT trajectory datasets have been assumed to be Brownian, such that the mean squared displacement (MSD), of each track is linearly proportional to the time lag, τ, and the diffusion coefficient, D, can be calculated from a linear fit to this curve [13, 14]. This Brownian motion assumption works accurately for freely diffusing molecules in solution. Despite the accessibility of this method, it has a simplified assumption that the molecule is freely diffusing with a single diffusive state (a single D value) for each trajectory. In the complicated cellular environment, however, multiple diffusive states, each characterized by an average D, can exist—for instance due to binding and unbinding events— and molecules can transition between different states to produce heterogeneity even within single trajectories. To reveal these heterogeneous dynamics, probability distribution-based methods such as cumulative probability distribution [16, 98], have been applied. Probability distribution-based models use kinetic modeling with a predetermined number of diffusive states and are fit to histograms of displacements calculated at different time lags. These probability-based kinetic models pool displacements from the SPT dataset to estimate the D and weight fraction for each diffusive state in the model. Probability distribution-based analytical tools [17, 18] have been widely applied to SPT datasets with extra corrections that consider the experimental microscopy data collection process. These corrections include localization error [15], confinement [99], motion blur [100, 101], and out-of-focus effects [102] in the probability model.

For some well-studied biological systems in which the biochemical states of molecules have been determined through other methods, a fixed-state number analytical tool can be suitable for quantifying the dynamics and weight for each state [103, 104]. However, SPT can also be used as the beginning step to investigate biomolecule dynamics without prior knowledge of how many diffusive states there supposed to be [30,105,106]. In these cases, how to objectively determine the number of diffusive states is a great challenge. Moreover, these models provide a D value for each subpopulation, but they do not assign the diffusive state to each individual single-molecule step, nor do they quantify the transition probability between distinct diffusive states within one trajectory. However, these transition probabilities can reveal important biological meaning such as the presence of critical biochemical intermediates [106].

Bayesian statistics and Hidden Markov Models (HMM) have been applied to analyze SPT datasets without assuming a predetermined number of diffusive states and to access the probabilities of transitioning between distinct states [29–31, 107]. vbSPT, which was one of the first applications of HMM for SPT analysis [29], uses a maximum-evidence criterion to select between models with different numbers of diffusive states; within each model, a fixed-order HMM is used to infer the diffusion coefficient, weight fraction, and transition probabilities for each state. More recently, nonparametric Bayesian models based on Dirichlet processes were combined with HMM to recover the number of diffusive states from SPT trajectory datasets, such as in single-molecule analysis by unsupervised Gibbs (SMAUG) [31] and DSMM [107]. In these models, the motion of the molecule is approximated to be symmetric and Brownian, which is an oversimplification considering the crowded environment and various interaction partners for biomolecules in cells.

To move beyond Brownian motion, here we consider a more general random walk family: anomalous diffusion. In anomalous diffusion, MSD and τ are related by a power law distribution, $MSD \sim \tau^{\alpha}$, where α is the anomalous diffusion exponent [38]. Brownian motion is a special case of anomalous diffusion (α=1), and other cases can be further divided into subdiffusion (α>1) and superdiffusion (α<1). Biomolecules have been reported to diffuse anomalously in many situations, such as constrained membrane protein motion [39], the facilitated diffusion of DNA binding protein [108], and active transportation of cargoes [40]. Different designs of neural networks effectively classify the diffusion type of trajectories [43–45, 109], however these analyses typically assume that each track is dynamically homogeneous and is characterized by a single type of diffusion and a single *D* value. It remains a challenge to classify the diffusion type within a trajectory when considering the possibility of changes in dynamics or diffusion types within a single track.

Here we introduce the NOnparametric Bayesian Inference for Anomalous diffusion in Single-molecule tracking (NOBIAS) framework to address the assumptions and simplifications discussed above and provide a more physiologically relevant analysis algorithm to quantify the dynamics encoded in SPT datasets (Fig. 2.1). In particular, NOBIAS recovers the number of diffusive states and predict the diffusion type for each diffusive state, even in heterogeneous trajectories. The NOBIAS framework consists of two modules. The first module uses an Hierarchical Dirichlet Process Hidden Markov Model (HDP-HMM) with multivariate Gaussian emission to recover the number of diffusive states and infer their corresponding diffusion coefficients and weight fractions. This module also assigns each single-molecule step a diffusive state label to provide the state label sequence and the matrix of transition probabilities. In the second module, the original trajectories are segmented by diffusive state label and a pre-trained Recurrent Neural Network (RNN) is used to classify these segments and assign the diffusion type (Brownian motion, Fractional Brownian motion, Continuous Time Random Walk, or Lévy Walk) for each diffusive state. We simulated trajectory datasets with mixtures of heterogeneous dynamics and diffusion types to validate the NOBIAS framework, and we analyzed the SPT dataset from experimental measurements of the SusG outer-membrane protein in living *Bacteroides thetaiotaomicron (B. theta)* to access its dynamics and anomalous diffusion behaviors, which are consistent with its role in starch catabolism in gut microbiome. This framework uses nonparametric Bayesian statistics and Deep learning to thoroughly analyze a single-molecule tracking dataset. It provides an objective method to determine the number of diffusive states in an SPT dataset and accesses the multidirectional dynamics of each state. A further diffusion type classification for each diffusive state is also included in the framework. The NOBIAS framework overcomes some oversimplified assumptions in SPT analysis and provides a powerful tool to fully make use of single-molecule tracking data.

2.2 Methods

2.2.1 Hidden Markov model

A HMM infers a system with a discrete-valued sequence of unobservable states that can be modeled as a Markovian process [28]. The HMM assumes that the observed data have a hidden discrete-valued state sequence, and at each observed time, the observed data only depends on its hidden state. In our NOBIAS application of the HMM model, the observed data is the single-molecule displacements and the hidden state is the molecule's distinct biophysical diffusive state.

Suppose z_t is the hidden state of the Markovian chain at time t and y_t is the observed data at time t, the HMM follows the following generative process:

$$z_1 \sim \pi^{(0)} \quad , \quad z_{t+1}|z_t \sim \pi^{(z_t)} \quad , \quad y_t|z_t \sim f(\theta^{(z_t)}) \tag{2.1}$$

Here, π refers to the transition matrix of a HMM and $\pi^{(z_t)}$ is the z_t row of the transition matrix and is the transition distribution for state z_t. Given z_t and the corresponding emission parameter $\theta^{(z_t)}$, y_t is independently generated from the emission function $f(\theta^{(z_t)})$. In NOBIAS, the observed data, y_t, is the vector of single-step displacements, $\Delta \boldsymbol{x_t}$, and the emission function is a zero-mean multivariate Gaussian distribution, and the emission parameter is the set of diffusion coefficients, $\boldsymbol{D}^{z_t}$:

$$\Delta \boldsymbol{x_t}|z_t \sim Norm(0, 4\boldsymbol{D}^{z_t}\tau)$$

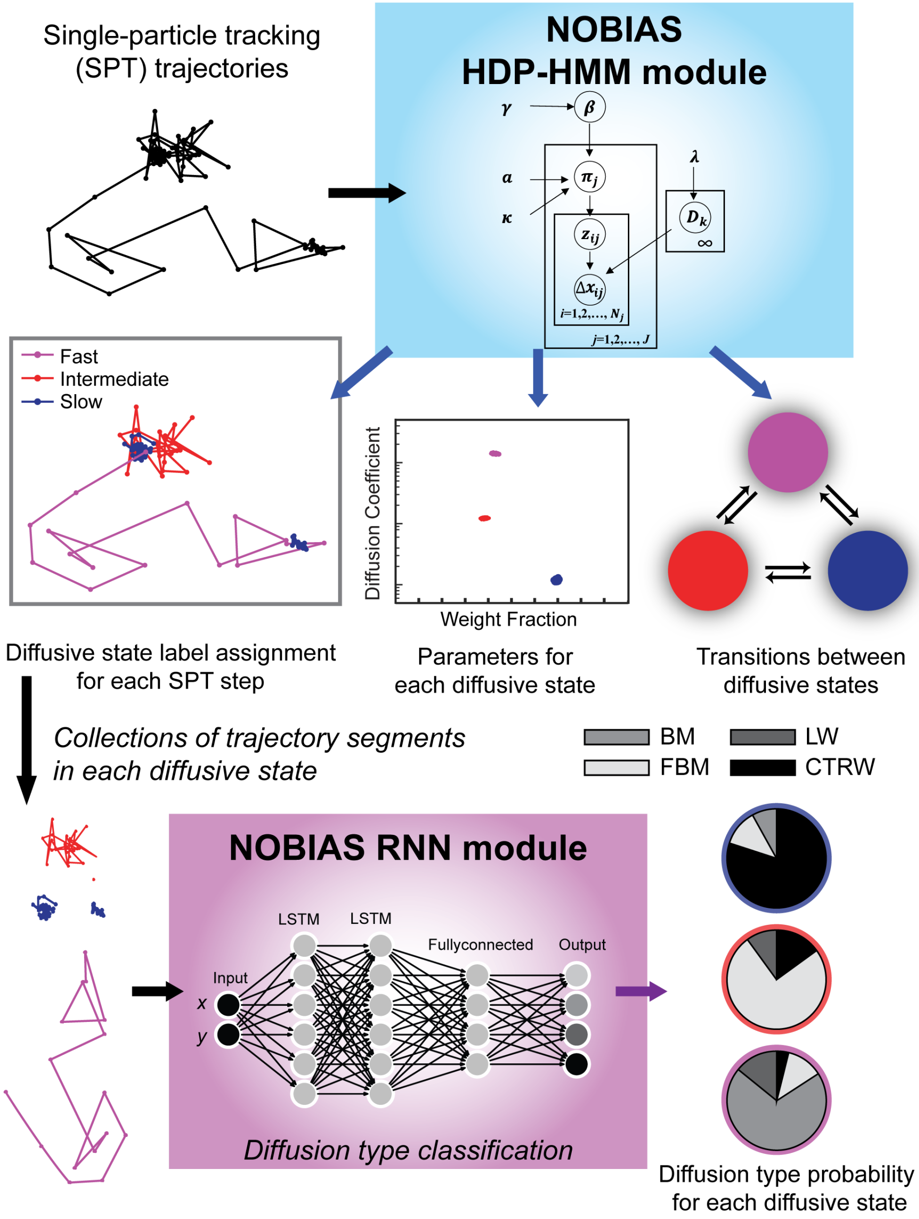


Figure 2.1: NOBIAS workflow(1) SPT trajectory datasets are processed in the NOBIAS HDP-HMM module: the observed data (the displacements, Δx) are analyzed in the context of the emission parameters (the diffusion coefficients, D). The state sequence,z, indicates the diffusive state corresponding to each step, and the transition matrix, π, is estimated with a Hierarchical Dirichlet process prior using concentration hyperparameters α and γ and the sticky parameter, κ. The HDP-HMM module provides D and the weight fraction for each diffusive state, the π for transition probabilities between these states, and a state label assignment for each SPT step. (2) In the NOBIAS RNN module, trajectory segments of the same diffusive state are collected and put in a pre-trained RNN with two long short-term memory (LSTM) layers to classify the diffusion type for each diffusive state.

2.2.2 Dirichlet process for Nonparametric Bayesian

In NOBIAS, Dirichlet Process (DP) is used in the prior for the parameters of a mixture model with an unknown number of components. A random probability measure, G_0, on a measurable space, Θ, is distributed according to a DP when [33]:

$$(G_0(B_1), ..., G_0(B_n))|\gamma, H \sim Dir((\gamma H(B_1), ..., \gamma H(B_k))) \tag{2.2}$$

Here, Dir is a Dirichlet distribution, H is a base measurement, γ is a positive concentration parameter, and $\{B_i\}_{i=1}^{n}$ is a finite partition of Θ. In this case, we write $G_0 \sim DP(\gamma, H)$.

From this definition follow two properties of DP. First, if$G_0 \sim DP(\gamma, H)$, then G_0 is atomic and can be written as:

$$G_0 = \sum_{i=1}^{\infty} \beta_i \delta_{\theta_i} \tag{2.3}$$

Here, β_i is a weight and δ_{θ_i} is a unit-mass measure at observation$\theta_i|H \sim H$.

Second, based on the conjugacy of the finite Dirichlet distribution, given a set of observations $\theta_1, .., \theta_N$ where $\theta_i \sim G_0$, the posterior distribution for a Dirichlet process G_0 is:

$$G_0|\theta_1, ..., \theta_N, H, \gamma \sim DP(\gamma + N, \frac{\gamma}{\gamma + N} H + \frac{1}{\gamma + N} \sum_{i=1}^{N} \delta_{\theta_i}) \tag{2.4}$$

A stick-breaking process is used to construct the weight parameter β_i as follows:

$$\beta_i = v_i \prod_{l=1}^{i} (1 - v_l), \quad v_l|\gamma \sim Beta(1, \gamma), \quad i = 1, 2, ...$$

In this process, the weight β_i comes from a unit stick according to a weight that is beta-distributed based on the remaining stick length after the last breaking. The weights from this construction, which is denoted $\beta \sim GEM(\gamma)$, have been proven [110] to be the weights β_i of a Dirichlet process as in Eq2.3.

For each value of θ_i, a random indicator variable z_i is used to denote that $\theta_i = \theta'_{z_i}$, and then a predictive distribution of z can be written as:

$$p(z_{N+1} = z|z_1, ..., z_N, \gamma) = \frac{\gamma}{\gamma + N}\delta(z, K + 1) + \frac{1}{\gamma + N}\sum_{i=1}^{N} N_k\delta(z, k)) \tag{2.5}$$

Where K is the current unique number of values of z and N_k is the number of z_i that take value k. This predictive distribution implies that a new observation takes the value of a seen observation θ_{z_k} with probability proportional to N_k or takes a unseen value θ_{K+1} with probability proportional to concentration parameter γ. When a seen observation θ_{z_k} is chosen for the new observation, the indicator $z_{N+1} = k$, or if unseen value θ_{K+1} is taken, the indicator $z_{N+1} = K + 1$. This 'the rich get richer' property is essential for inferring a finite generated mixture model. Because the DP posterior nonparametrically converges to parameters of a mixture model for a finite mixture dataset [111], the DPis an appropriate prior for the parameters of a mixture model with an unknown number of components.

2.2.3 Hierarchical Dirichlet Process and Sticky Extension

In NOBIAS, the different single-molecule trajectories of multiple molecules under different biological condition and from different cells, so the groups of data are related but generated independently. Therefore, the DP is extended to a Hierarchical Dirichlet Process (HDP) [34]. In the HDP, a first DP, G_0, is the base measure of a new DP, G_j:

$$G_j \sim DP(\alpha, G_0), \quad G_0 \sim DP(\gamma, H)$$

To apply anHDPas prior for an HMM model, an HDP-HMM model is generated according to:

$$\beta \sim GEM(\gamma), \quad \pi_j \sim DP(\alpha, \beta), \quad \theta^{(j)}|\lambda \sim H(\lambda), \quad j = 1, 2, ...$$

$$z_t|\{\pi\}, z_{t-1} \sim \pi_{z_{t-1}}, \quad y_t|\{\theta\}, z_t \sim F(\theta^{(z_t)}) \quad t = 1, 2, ..., T$$

In the NOBIAS parameter setting, the observed data y_t would be the single-step displacement $\Delta \boldsymbol{x_t}$, the emission parameter θ would be the diffusion coefficient D, and the hyperparameter λ for θ would be the Normal-inverse-Wishart distribution (NIW) with four prior hyperparameters $\{\kappa, \vartheta, \nu, \Delta\}$ as stated below in the Multivariate Normal Model section.

A common issue for the HDP-HMM model is that if the algorithm artificially divides a set of observations into an alternating pattern of rapid switching between several different states, then this alternating pattern will be reinforced by the DP [37]. This assignment would result in an artificial over-splitting of one state into multiple substates characterized by a high probability of transitions between the substates. Because we would not expect such rapid transitions back and forth between two distinct but similar dynamical states in the single-molecule trajectory data studied here, a sticky parameter, κ, is introduced which enforces self-transitions and avoids this over-splitting of states. With this new hyperparameter, the π_j can be sampled as:

$$\pi_j \sim DP(\alpha + \kappa, \frac{\alpha\beta + \kappa\delta_j}{\alpha\kappa}), \quad j = 1, 2, ... \tag{2.6}$$

Which add a self-transition bias to the j^{th} components of the DP. The effects of κ on the results are shown in Figure 2.2D: if κ is too small, the over-splitting of states still occurs and if κ is too large, the model will underestimate the number of states.

Different Markov Chain Monte Carlo (MCMC) sampling methods such as Direct Assignment Sampling, Beam Sampling, and Blocked Sampling have been developed for the HDP-HMM model [34, 112, 113]. In NOBIAS, we apply the most computationally efficient Blocked Sampling method [112], which uses a fixed-order truncation with weak-limit approximation HDP-HMM. In this approach, the DP is L-degree approximated as:

$$\beta \sim GEM_L(\gamma) \sim Dir(\gamma/L, ..., \gamma/L) \tag{2.7}$$

$$\pi_j \sim DP_L(\alpha + \kappa, \frac{\alpha\beta + \kappa\delta_j}{\alpha\kappa}) \sim Dir(\alpha\beta_1, ..., \alpha\beta_j + \kappa, ..., \alpha\beta_L) \tag{2.8}$$

with a truncation level, L, that is larger than the expected total number of mixture components. Increasing L does not affect the posterior results, but L does affect the running time (Fig. 2.2C).The Blocked Sampling method algorithm is detailed in [112], which describes how the state sequence is generated and how the parameters for each state are sampled.

2.2.4 Multivariate Normal Model

Bayes' rule states that the posterior distribution is proportional to the product of the prior probability and the likelihood, i.e., $P(\theta|y) \sim P(\theta)P(y|\theta)$. It is crucial to build conjugacy in order to elegantly and concisely express the posterior distribution. If we choose an appropriate prior distribution class for $P(\theta)$ given a known sampling distribution $P(y|\theta)$, then the posterior distribution $P(\theta|y)$ will have the same distribution class as the prior distribution. This choice of a prior distribution is called a conjugate prior, and this property that the posterior and prior distributions are in the same class is called conjugacy.

In NOBIAS HDP-HMM module, we assume 2D Brownian motion trajectories. In this case, the displacements follow a zero-mean 2D Gaussian and the diffusion coefficients D determine the variance, Σ, of the 2D Gaussian. Without loss of generality, the mean, μ, is also included in the model, $\theta = \{\mu, \Sigma\}$, and the data distribution is written as:

$$p(y|\theta) = \frac{1}{2\pi|\Sigma|^{1/2}} exp(-\frac{1}{2}(\Delta\boldsymbol{x} - \mu)^T|\Sigma|^{-1}(\Delta\boldsymbol{x} - \mu)) \tag{2.9}$$

In the 2D case, the observed data, $\Delta\boldsymbol{x}$, is a 1×2 vector of the 2D displacements, μ is a 1×2 vector and Σ is the 2×2 covariance matrix

As derived in reference [114], the general conjugate prior model for this multivariate normal model is the prior for the mean and the variance of the step displacement follow

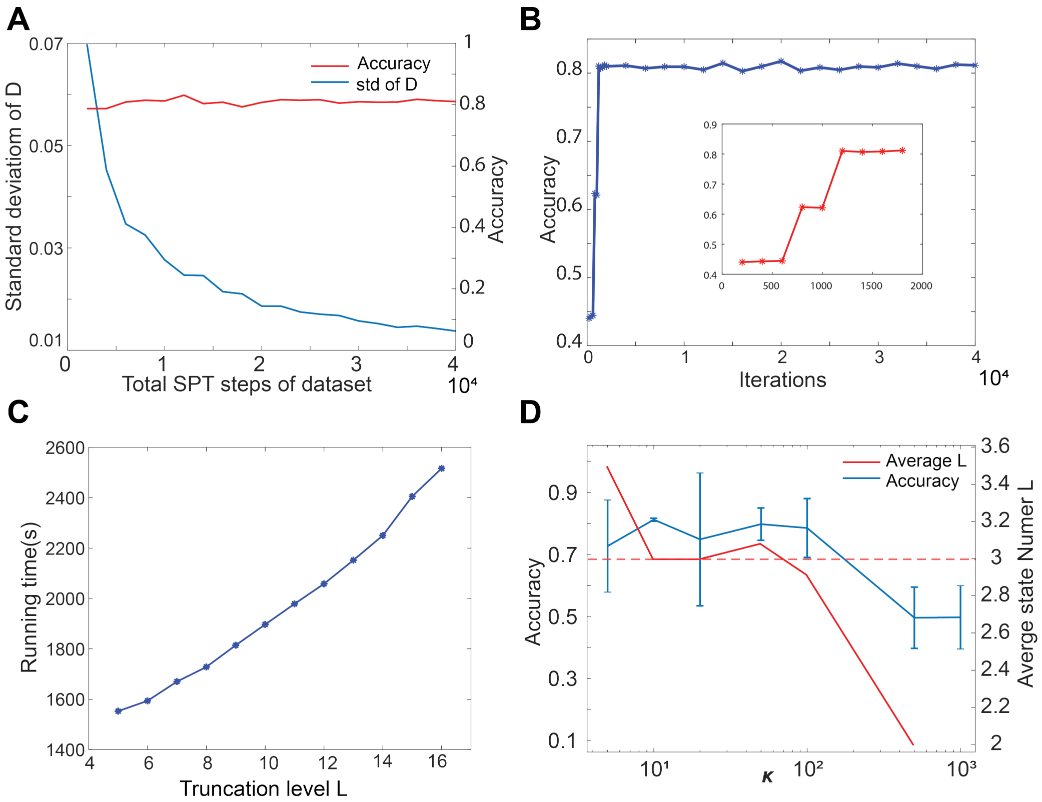


Figure 2.2: All evaluations use the simulated 3-state motion blur sparse data with the parameter settings as in Table 2.66, aside from Figure 2.8C which uses the standard abundant 3-state data. (A) The state label accuracy (red) is largely insensitive to the total number of steps in the SPT trajectories, while the posterior parameter sample uncertainty improves with an increase in the amount of data amount. All tracks used for this plot are 10 steps long. (B) For the same 3-state motion blur sparse dataset, the NOBIAS accuracy is independent of the final number of iterations beyond 2000 iterations. Inset: zoom in on iterations 0 − 2000. (C) The running time (blue) increases with the truncation level, L. where the final number of states (red) is not affected. (D) Tuning the sticky parameter, κ, affects the HDP-HMM module performance. Red solid line: average final number of states. The red dashed line indicates the true number of states. Blue line: average state label accuracy (error bars: standard deviation of accuracy over the 12 chains). All results are averaged over 12 chains.

a NIW:

$$p(\mu, \Sigma) \sim NIW(\kappa, \vartheta, \nu, \Delta) \tag{2.10}$$

Specifically, the variance, Σ, follows an inverse-Wishart prior distribution $IW(\nu, \Delta)$, and the mean, μ, has a conditional Normal distribution: $p(\mu|\Sigma) \sim N(\vartheta, \Sigma/\kappa)$.

The posterior updates for this normal model with NIW prior follows [114]:

$$p(\mu^{(z_t)}, \Sigma^{(z_t)}|\Delta\boldsymbol{x}^{(z_t)}) \sim NIW(\bar{\kappa}, \bar{\vartheta}, \bar{\nu}, \bar{\Delta}) \tag{2.11}$$

Where $\Delta\boldsymbol{x}^{(z_t)}$ is the entire displacement dataset in state z_t, and for each state z_t, we update these parameters as:

$$\bar{\kappa} = \kappa + N, \quad \bar{\kappa}\bar{\vartheta} = \kappa\vartheta + \sum_{n=1}^{N} \Delta\boldsymbol{x_n}$$

$$\bar{\nu} = \nu + N, \bar{\nu}\bar{\Delta} = \nu\Delta + \sum_{n=1}^{N} \Delta\boldsymbol{x_n}\Delta\boldsymbol{x_n}^T + \kappa\vartheta\vartheta^T - \bar{\kappa}\bar{\vartheta}\bar{\vartheta}^T$$

To decrease the running time, we apply the conjugate prior for the Multivariate Normal Distribution, though a non-conjugate prior is permissible. For further discussion of choice of prior see [114].

2.2.5 Trajectory Simulation

A state label sequence was firstly simulated with a given transition matrix through a Markov chain process. Then according the state label and the D of corresponding diffusive state, the 2D displacement step is generated, and cumulatively summed to get a single trajectory. Standard trajectory datasets are simulated by generate 2D Gaussian random variable where mean is 0 and variance is determined by the set diffusion coefficients with symmetry and no correlation in two directions.

Motion blur trajectory datasets are generated by simulating a state label sequence that is T_{exp} times of the desired length with a transition matrix that self-transit enhanced T_{exp} times. Also according to the label of this T_{exp} times longer label sequence a true trajectories with T_{exp} times more steps can be generated as in the standard dataset case. 2D localization error is added to the average position of every T_{exp} steps in the true trajectory and saved to create a motion-blur trajectory with desired length. In the motion blur trajectory datasets used in this study, T_{exp} was set to 10.

2.2.6 Anomalous Diffusion

In the NOBIAS RNN module, trajectory segments of the same diffusive state (identified by the HDP-HMM module) are evaluated to classify the diffusion type for each diffusive state. In Brownian Motion, the MSD is linearly proportional to the time lag, τ. In anomalous diffusion, MSD is related to τ according to a power law [38]:

$$MSD \propto \tau^{\alpha} \tag{2.12}$$

Here, α is the anomalous exponent. When $\alpha = 1$, this relation describes Brownian motion; when $\alpha > 1$,Equation 2.12 describes superdiffusion; and when $\alpha < 1$, Equation 2.12 describes subdiffusion. The NOBIAS framework includes the three specific types of anomalous diffusion types that are most common in biology: Fractional Brownian motion (FBM) [115], Continuous Time Random Walk (CTRW) [116], and Lévy Walk (LW) [117].

FBM is a Gaussian process with correlated increments such that MSD is related to τ according to: $MSD = 2D_H\tau^{2H}$ [115, 118]. Here, the Hurst exponent, H, is related to α in Equation 2.12 by $\alpha = 2H$. The D_H is the generalized coefficients with physical dimension m^2s^{-2H}. The correlation between two time points for FBM is $\langle x(t_1), x(t_2)\rangle = D_H(t_1^{2H}+t_2^{2H}-|t_1 - t_2|^{2H})$. When this correlation is positive, $H > 0.5$ and the motion is superdiffusive; when the correlation is negative, $H < 0.5$ and the motion is subdiffusive.

CTRW defines a random walk family in which the particle displacement, Δx, follows a wait at its current position for a random waiting time t that is a stochastic variable [116]. NOBIAS considers the case where t follows a power-law distribution, $\psi(t) \propto t^{-\sigma}$, and the following displacement is sampled from a zero-mean Gaussian with fixed variance. In this case, the σ in CTRW is related to α in Equation 12 by $\alpha = \sigma - 1$. This CTRW can only be subdiffusion, i.e., $0 < \alpha < 1$.

LW is a special case of CTRW in which the waiting time, t, still follows power law, but the displacement is not Gaussian, and is instead determined by the waiting time [117]. The displacement will have a constant speed, $v = |\Delta \boldsymbol{x}|/t$, and this process can only be superdiffusive with exponent $1 \leq \alpha \leq 2$.

We simulated these three types of anomalous diffusion with the open-source Python package from the recent AnDi challenge [46].

2.2.7 Recurrent Neural Network for NOBIAS

All segments 40 steps or greater identified in the HDP-HMM module were further analyzed by the NOBIAS Recurrent Neural Network (RNN) consisting of two LSTM layers [119]. We trained this RNN to classify trajectory segments identified to have the same diffusive state from the HDP-HMM module. We implemented this architecture, which is based on the design of the RANDI package classification task [43, 45] with the MATLAB Deep Learning Toolbox. The two LSTM layers have 100 and 50 units, respectively, and these two LSTM layers are followed by a fullyconnected layer, and the output classification layer order is given in Figure2.1.

The input to the network is the set of 2D coordinates from the track segments; these coordinates are normalized to have zero mean and unit variance. Despite a much higher classification performance when using tracks > 50 steps long to train and validate [45,47,109], we trained two networks with 20-step tracks and with 40-step tracks, respectively, after considering the typical segment lengths from real biological trajectories. The training data of 750,000 trajectories were simulated with the open-source Python package from the

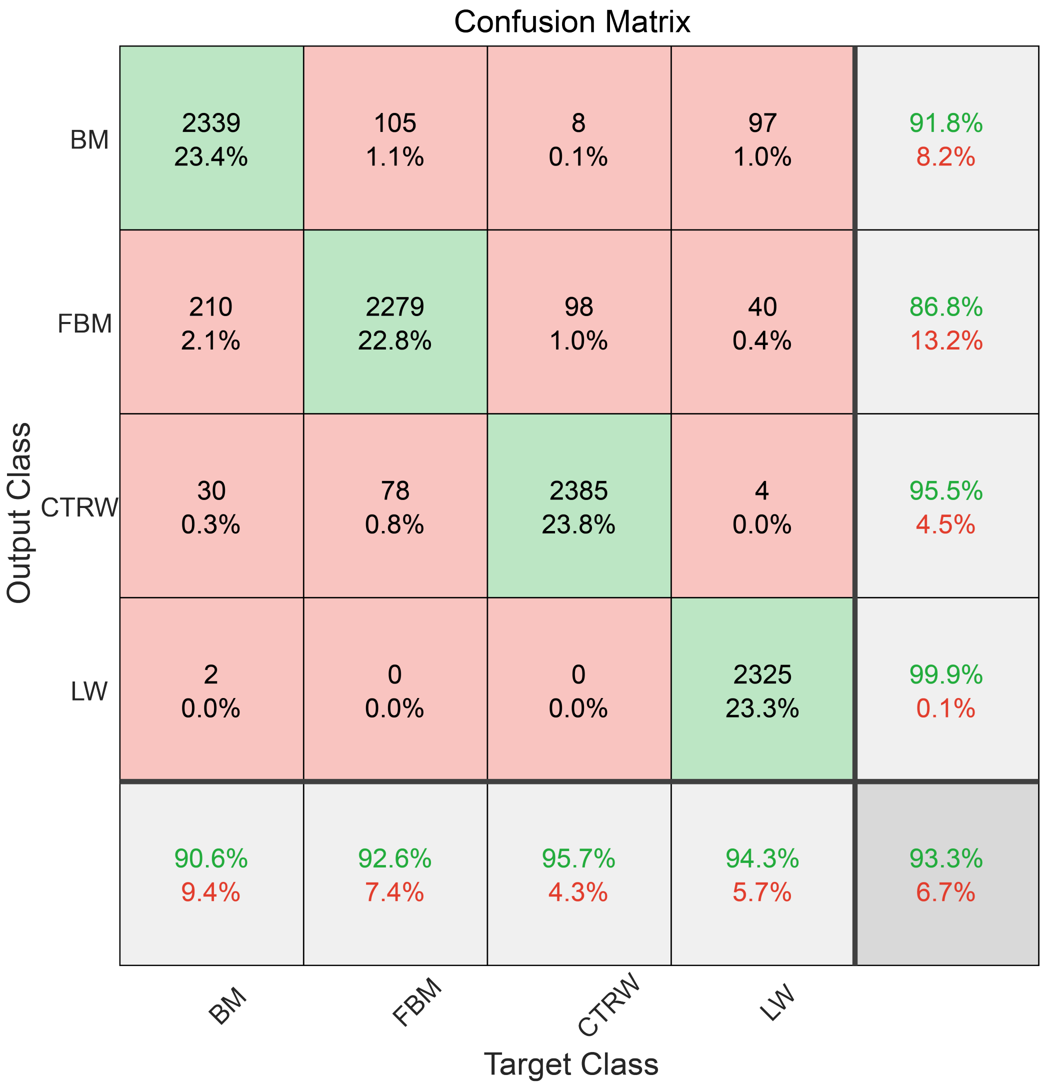
AnDi challenge [46]. Regression networks with similar 2 LSTM layers architecture were also trained for FBM and CTRW to estimate the anomalous exponent α for the experimental data. The performance of the classification network with 40-step data is shown in the confusion matrix which was made with 10000 test trajectories . However, although the RNN module can classify CTRW and LW motion Figure2.3, because our HDP-HMM module assumes Brownian motion, this first module cannot predict the correct state label for these two diffusion types. We therefore test a mixture of FBM and Brownian Motion (BM) motion in Figure2.6.

2.2.8 Single-Molecule Tracking in Living *Bacteroides thetaiotaomicron* Cells

B. theta cells expressing SusG-HaloTag fusions at the native SusG promoter were grown as previously described [120]. Briefly, cells were cultured overnight in 0.5% tryptone-yeast-extract-glucose medium and incubated at 37 °C under anaerobic conditions (85 % N2, 10 % H2, 5 % CO2) in a Coy chamber. Approximately 24 h before imaging, cells were diluted into *B. theta* minimal medium (MM) [121] containing 0.25% (wt/vol) amylopectin. On the day of the experiment, cells were diluted into fresh MM and carbohydrate and grown until reaching OD600nm 0.55 – 0.60 [122].

Before labeling, 900 μL of cells were washed twice by pelleting (6000 G, 2 min) followed by resuspension in MM. Cells were then incubated in MM supplemented with 100 nM PA-JF549 dye [4] for 15 min in the dark. Cells were then washed five times in MM, transferring to a new tube on every step, to remove excess dye [123]. Finally, 100 μL cells were resuspended in MM containing 0.25% (wt/vol) amylopectin for 30 min in the dark. 1.5 μL labeled cells were pipetted onto a pad of 2% agarose in MM and placed between a large and a small coverslip. The two coverslips were sealed together with epoxy (Devcon 31345 2 Ton Clear Epoxy, 25 mL) to keep the media anaerobic [120].

Cells were imaged on an Olympus IX71 inverted epifluorescence microscope with a 1.45 numerical aperture, 100× oil immersion phase-contrast objective (Olympus UP-



Confusion Matrix

Output Class \ Target Class	BM	FBM	CTRW	LW	
BM	2339 23.4%	105 1.1%	8 0.1%	97 1.0%	91.8% 8.2%
FBM	210 2.1%	2279 22.8%	98 1.0%	40 0.4%	86.8% 13.2%
CTRW	30 0.3%	78 0.8%	2385 23.8%	4 0.0%	95.5% 4.5%
LW	2 0.0%	0 0.0%	0 0.0%	2325 23.3%	99.9% 0.1%
	90.6% 9.4%	92.6% 7.4%	95.7% 4.3%	94.3% 5.7%	93.3% 6.7%

Figure 2.3: Confusion matrix for classification of the diffusion type by the NOBIAS RNN module. A total of 750,000 40-step tracks of the four diffusion types were used to train the network, and 10,000 tracks were tested to get the confusion matrix.

LXAPO100XOPH) and a 3.3× beam expander. Frames were collected continuously on a 512 × 512 pixel electron-multiplying charge-coupled device camera (Photometrics Evolve 512) at 50 frames/s. In this microscopy geometry, 1 camera pixel corresponds to 48.5 nm. PA-JF549 dyes were photo-activated one at a time with a 200 − 400 ms exposure by a 406-nm laser (Coherent Cube 405-100; 0.1 $\mu W/\mu m^2$) and imaged with a 561-nm laser (Coherent-Sapphire 561-50; 1 $\mu W/\mu m^2$)) using appropriate filters as previously described [122].

In each movie, each cell was analyzed separately by using an appropriate mask. The collected frames were processed with SMALL-LABS [12] to detect single molecules frame-by-frame and localize their position with typically 30 nm uncertainty. Single molecules were identified as non-overlapping punctuate spots of diameter larger than 7 pixels and with pixel intensities larger than the 92$^{\text{nd}}$ percentile intensity of the fame. The punctate spots were fit to a 2D Gaussian and true single-molecule localizations satisfied the following conditions: (1) standard deviation > 1 pixel and (2) fit error ≤ 0.06 pixel. Localizations in each cell over time were connected into trajectories using a merit value: trajectories were selected for further analysis based on their highest merit ranking.

2.3 Results

2.3.1 The NOBIAS HDP-HMM module recovers the number of diffusive states and the associated diffusion parameters

We first validated the NOBIAS HDP-HMM module with simulated single-molecule tracks, beginning from the most basic case: a mixture of Brownian motion trajectories. Figure 2.4A-D depicts the results for a mixture of two distinct diffusive states with $D_1 = 0.135 \mu m^2/s$ and $D_2 = 1.8 \mu m^2/s$ Table 2.1. A sequence of state labels (1 or 2) was first simulated with a given transition matrix (probability of transitioning from state 1 to 2 or from state 2 to 1) through a Markov chain process 2.2. Then, according the state label and the apparent diffusion coefficient, D, of the corresponding diffusive state, each 2D

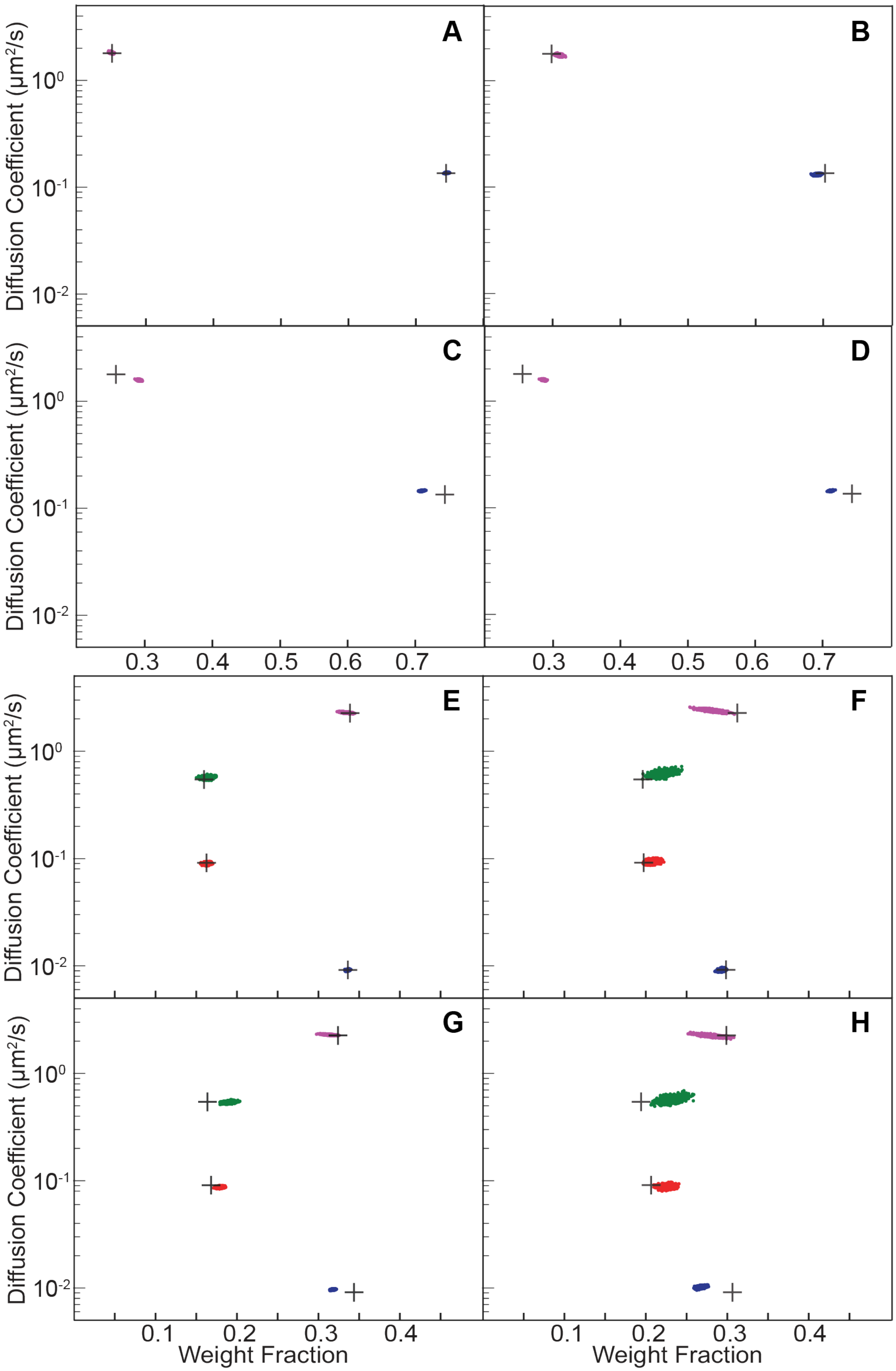

A
B
C
D
E
F
G
H
Diffusion Coefficient (μm²/s)
10⁰
10⁻¹
10⁻²
0.3
0.4
0.5
0.6
0.7
0.1
0.2
0.3
0.4
Weight Fraction

Figure 2.4

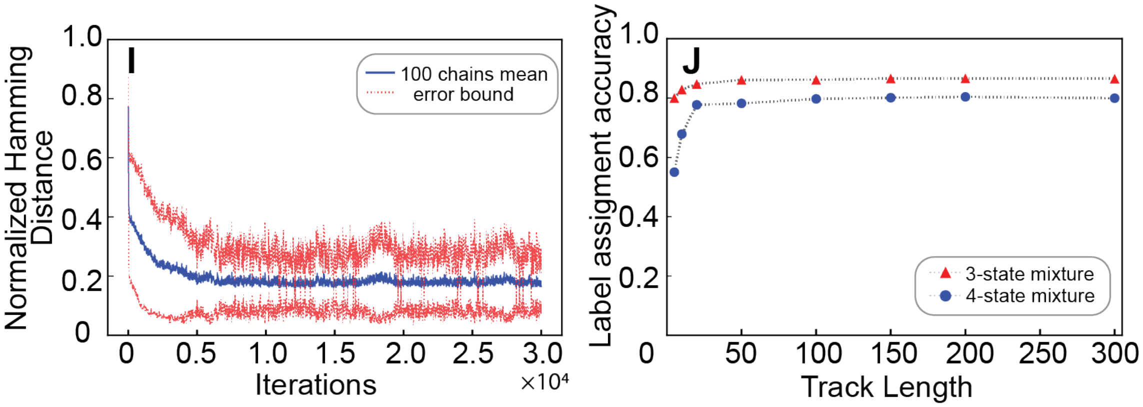


Figure 2.4: Validation of the NOBIAS HDP-HMM module with simulated trajectories. (A-H) The HDP-HMM module identifies distinct mobility states (colored clusters). All scatter plots include at least 500 uncorrelated samples. Each point represents the average apparent single-molecule diffusion coefficient vs. weight fraction in each distinct mobility state at each iteration of the Bayesian algorithm saved after convergence. The black crosses indicate the ground truth input for these simulated trajectories. (A-D) Results for two-state mixture simulated trajectories results: (A) Standard (no motion blur) and abundant (500 100-step trajectories) simulations, (B) Standard and sparse (2000 10-step trajectories) simulations, (C) Motion blur and abundant simulations, and (D) Motion blur and sparse simulations. (E-H) Results for four-state mixture simulated trajectories results: (E) Standard (no motion blur) and abundant (500 100-step trajectories) simulations, (F) Standard and sparse (2000 10-step trajectories) simulations, (G) Motion blur and abundant simulations, and (H) Motion blur and sparse simulations. (I) The Normalized Hamming Distance (NHD) decreases and converges with the number of iterations. All 100 chains use the same dataset under the settings in panel (E). (J) The final label assignment accuracy increases with the track length for three- and four-state mixture datasets. The number of trajectories decreases as the track lengths increase such that the total amount of steps is 30,000 for all track lengths.

displacement step was generated, and cumulatively summed to get a single trajectory. Similar state label sequences were simulated to generate other trajectory datasets with 4 diffusive states (Figure 2.4E-G, Table 2.2).

The posterior results of the HDP-HMM module are shown in scatter plots of the inferred D and weight fraction from each iteration after the inferred number of states converges. Figure 2.4A shows the result for a dataset of 500 trajectories each with 100 steps. Here, the black crosses indicate the ground truth diffusion coefficient and weight fraction for each diffusive state; the posterior samples of the HDP-HMM model for the two states after convergence are distributed around the true values. Based on the posterior sample autocorrelation function (ACF) analysis (Figure 2.5), the posterior samples are thinned by saving every 10 iterations; this setting is the same for all results in this paper and was chosen by considering the effective sample sizes and the ACF analysis for all the diffusive states. The number could be updated accordingly depending on the correlation of posterior samples from output. The mean values and standard deviations for the estimation of D and weight fractions for the two states are listed in Table 2.1. The estimated number of unique states for this simulated dataset converges quickly over the course of iterations to the true number of states and remains mostly stable at that number (Figure 2.8). Next, we considered the less ideal case that often occurs experimentally: much shorter trajectory lengths (10 steps) and many fewer total steps (2000 10-step trajectories). We refer to the 2000 10-step trajectories as a sparse dataset and the 500 100-step trajectories are an abundant dataset. Figure 2.4 B shows that the HDP-HMM model still successfully converges to the true number of states (two) for this dataset, and the posterior samples of the diffusive parameters are still distributed near the true inputs (black crosses).

We further considered the true form of collected microscope experimental data by including the localization error due to finite photon counts and noise and motion blur due to the finite image acquisition time (2.2). We refer these datasets ‘Motion blur dataset’ in contrast with the more ideal ‘Standard’ dataset. In the case of motion blur, the sticky parameter is increased to avoid oversampling a single diffusive state into multiple state with similar dynamics. The hyperparameter settings for this sticky HDP-HMM model are listed in Table 2.3. For both the abundant dataset (Figure 2.4C: 500 100-step trajectories) and the sparse dataset (Figure 2.4D: 2000 10-step trajectories), the true number of states (two) is recovered with our sticky HDP-HMM model, and despite these added errors, the estimated parameters deviate only slightly from the true inputs (black crosses).

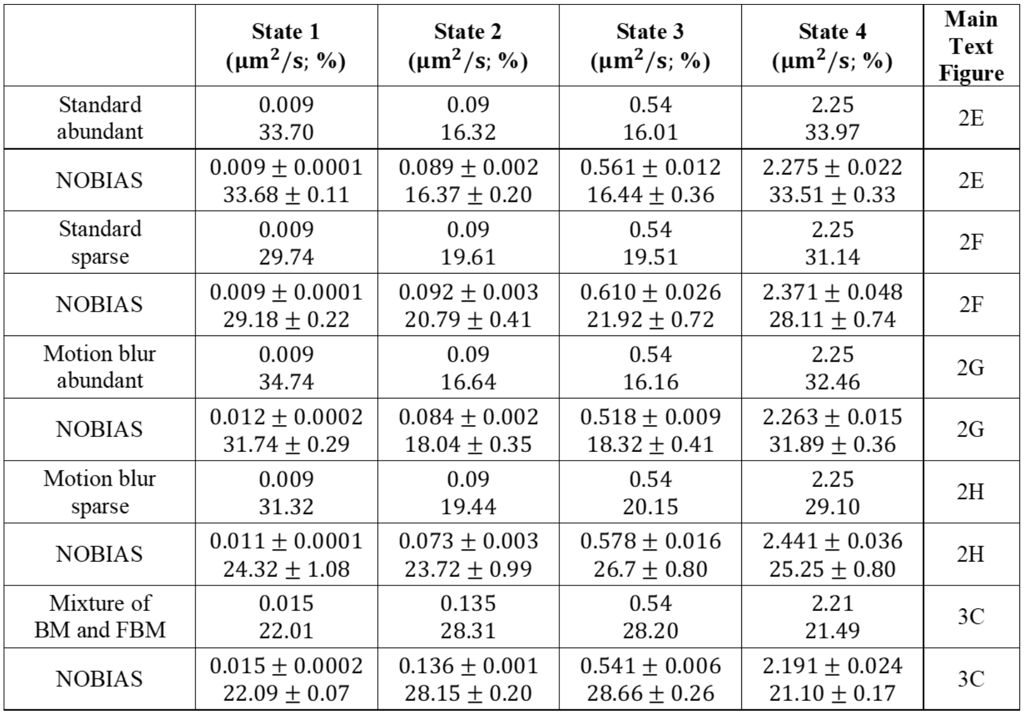
We extended our simulations of standard and motion blur Brownian motion track mixtures to a more complicated 4-state scenarios for abundant (500 100-step trajectories) and sparse (2000 10-step trajectories) datasets (Figure 2.4E-H). Even with 4 diffusive states, the performance of the HDP-HMM module is still excellent for the standard mixture (Figure 2.4E-F). For the 4-state mixture simulation that includes localization error and motion blur, the HDP-HMM still successfully recovers the true number of states, and the parameters for the four distinct states are still estimated well, though the posterior samples have increased variance and deviation from the true value (Figure 2.4G-H). The statistics of the posterior samples for estimated parameters of the 4-state simulation result are listed in Table 2.2, and the transition matrices for all the simulations in Figure 2.4 are shown in Table 2.1-2.2.

The NOBIAS HDP-HMM module also assigns diffusive state labels to each single-molecule step within the trajectories dataset; we call this the state sequence for each track. We quantified the performance of the state sequence assignment relative to the ground truth simulated state sequence with the Hamming distance: the Hamming distance between two 1D sequences with equal length is the number of points where the components are different [124]. The resulting distances were normalized to the total length to demonstrate the Normalized Hamming Distance (NHD) convergence over iterations (Figure 2.4I). The NHD decreases with increasing iteration number and converges to approximately 0.18. This final converged NHD depends on the dataset size, the true transition matrix, and how separable the diffusive state are from one another.

	State 1 **($\mu m^2/s$; %)**	**State 2** **($\mu m^2/s$; %)**	**Main Text Figure**
Standard abundant	0.135 74.75	1.8 25.25	2A
NOBIAS	0.136 ± 0.001 74.82 ± 0.14	1.824 ± 0.019 25.18 ± 0.14	2A
Standard sparse	0.135 70.23	1.8 29.78	2B
NOBIAS	0.132 ± 0.001 69.12 ± 0.26	1.754 ± 0.026 30.88 ± 0.26	2B
Motion blur abundant	0.135 74.32	1.8 25.68	2C
NOBIAS	0.143 ± 0.001 71.35 ± 0.18	1.579 ± 0.010 28.65 ± 0.18	2C
Motion blur sparse	0.135 70.24	1.8 29.76	2D
NOBIAS	0.142 ± 0.001 67.20 ± 0.32	1.580 ± 0.016 32.80 ± 0.32	2D
Mixture of BM and FBM	0.045 50.76	0.90 49.24	3A
NOBIAS	0.044 ± 0.0003 49.20 ± 0.08	0.901 ± 0.006 50.80 ± 0.08	3A

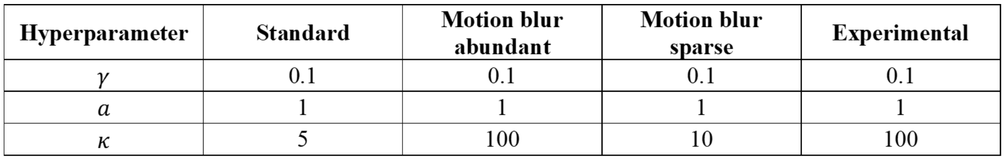
Table 2.1: Two-state mixture results for simulations of the standard abundant, standard sparse, motion blur abundant, motion blur sparse, and a mixture of Brownian and subdiffusive fractional Brownian motion models. Diffusion coefficients (in $\mu m^2/s$) and weight fractions (in %) are given for the ground truth inputs and the NOBIAS HDP-HMM module outputs for each of the 2 states. These data correspond respectively to the main text figures as indicated. Errors represent standard deviation.

The true number of diffusive states can be recovered for datasets of both abundant and sparse tracks, but the HDP-HMM module performance depends strongly on the length of the individual tracks. Using the overall state sequence assignment accuracy (1 - NHD) as a performance evaluator for datasets with the same total amount of steps (30000), we found that the assignment accuracy is considerably worse for tracks shorter than 20 steps and almost linearly increases with the track length till asymptotes for longer tracks (> 20 steps; Figure 2.4J). This trend is shared for a 3-state and 4-state dataset, but the overall accuracy for 3-state dataset is higher than 4-state one for all the track length.



	State 1 (μm^2/s; %)	State 2 (μm^2/s; %)	State 3 (μm^2/s; %)	State 4 (μm^2/s; %)	Main Text Figure
Standard abundant	0.009 33.70	0.09 16.32	0.54 16.01	2.25 33.97	2E
NOBIAS	0.009 ± 0.0001 33.68 ± 0.11	0.089 ± 0.002 16.37 ± 0.20	0.561 ± 0.012 16.44 ± 0.36	2.275 ± 0.022 33.51 ± 0.33	2E
Standard sparse	0.009 29.74	0.09 19.61	0.54 19.51	2.25 31.14	2F
NOBIAS	0.009 ± 0.0001 29.18 ± 0.22	0.092 ± 0.003 20.79 ± 0.41	0.610 ± 0.026 21.92 ± 0.72	2.371 ± 0.048 28.11 ± 0.74	2F
Motion blur abundant	0.009 34.74	0.09 16.64	0.54 16.16	2.25 32.46	2G
NOBIAS	0.012 ± 0.0002 31.74 ± 0.29	0.084 ± 0.002 18.04 ± 0.35	0.518 ± 0.009 18.32 ± 0.41	2.263 ± 0.015 31.89 ± 0.36	2G
Motion blur sparse	0.009 31.32	0.09 19.44	0.54 20.15	2.25 29.10	2H
NOBIAS	0.011 ± 0.0001 24.32 ± 1.08	0.073 ± 0.003 23.72 ± 0.99	0.578 ± 0.016 26.7 ± 0.80	2.441 ± 0.036 25.25 ± 0.80	2H
Mixture of BM and FBM	0.015 22.01	0.135 28.31	0.54 28.20	2.21 21.49	3C
NOBIAS	0.015 ± 0.0002 22.09 ± 0.07	0.136 ± 0.001 28.15 ± 0.20	0.541 ± 0.006 28.66 ± 0.26	2.191 ± 0.024 21.10 ± 0.17	3C

Table 2.2: Four-state mixture results for simulations of the standard abundant, standard sparse, motion blur abundant, motion blur sparse, and a mixture of Brownian, subdiffusive fractional Brownian, and superdiffusive fractional Brownian motion models. Diffusion coefficients (in $\mu m^2/s$) and weight fractions (in %) are given for the ground truth inputs and the NOBIAS HDP-HMM module outputs for each of the 4 states. These data correspond respectively to the main text figures as indicated. Errors represent the standard deviation.



Hyperparameter	Standard	Motion blur abundant	Motion blur sparse	Experimental
γ	0.1	0.1	0.1	0.1
α	1	1	1	1
κ	5	100	10	100

Table 2.3: NOBIAS HDP-HMM module hyperparameter settings for the simulations and experimental data.

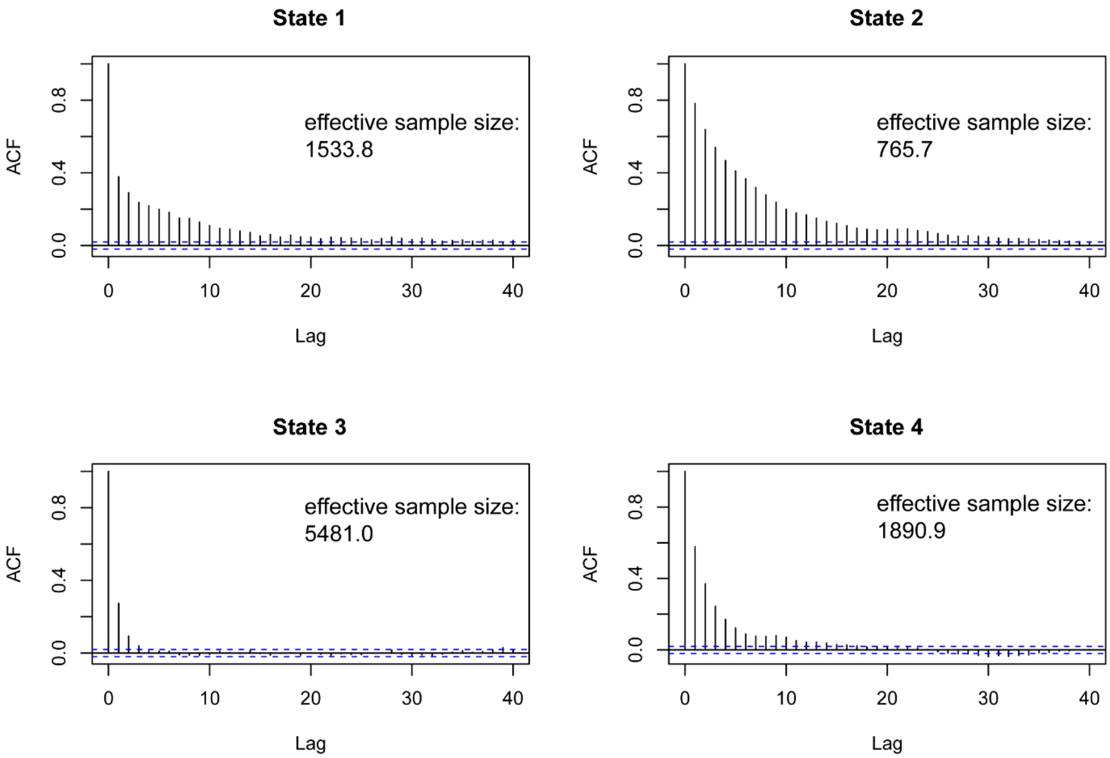


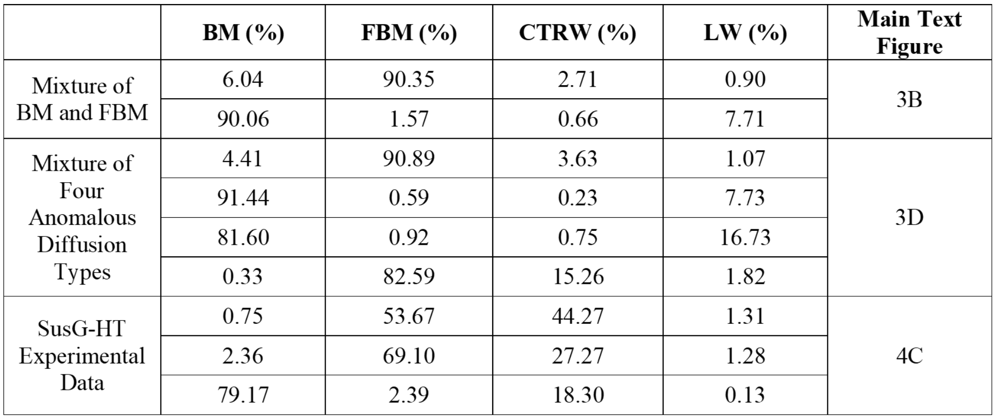
Figure 2.5: ACF analysis for posterior samples of the diffusion coefficient of the four-state standard abundant simulation described in main text Figure2.4E. Over 20,000 iterations, the number of states converges to 4 with a 2000 burn-in. The final 10,000 samples are used for further analysis.

2.3.2 The NOBIAS RNN module predicts the diffusion type for each diffusive state

To analyze anomalous diffusion in an SPT dataset, NOBIAS includes a second module: we built an RNN to classify the type of motion (Brownian motion (BM), Fractional Brownian motion (FBM), Continuous Time Random Walk (CTRW), or Lévy Walk (LW)) corresponding to the track segments within each diffusive state identified by HDP-HMM module. The RNN consists of two LSTM layers, a fullyconnected layer, and data input/output layer (2.2). Although the HDP-HMM module is based on BM, for some anomalous diffusion types, for example FBM, if the dynamics level for each state is distinct, the HDP-HMM module still performs well.

We simulated a mixture of BM and FBM with distinct apparent diffusion coefficients for the two states ($D_1 = 0.045\mu m^2/s$ and $D_2 = 0.90\mu m^2/s$) to validate the performance of NOBIAS on mixtures of different diffusion types. Figure 2.6A shows the HDP-HMM posterior results for this 2-state BM-FBM mixture (500 100-step trajectories) where the FBM state is anomalous subdiffusion with $\alpha = 0.5$ (Eq. 2.12) and with lower diffusion coefficient. Then, based on the state sequence labels from the HDP-HMM module, we generated track segments for the two diffusive states and put them into the trained NOBIAS RNN network to predict the diffusion types. NOBIAS RNN successfully predicts the diffusion types for both states (Figure 2.6B, Table 2.4).

We further simulated a 4-state mixture (500 100-step trajectories) corresponding to subdiffusive FBM, BM, BM, and superdiffusive FBM (in order of increasing D). The HDP-HMM module still successfully recovers the 4 states and make excellent estimations for D and weight fraction for each state (Figure 2.6C). The NOBIAS RNN module also predicts the true diffusion type for the segments from each of the four states (Figure 2.6D, Table 2.4). Note that all track segments are normalized before being put into the RNN to avoid dynamics information bias in the diffusion type prediction (2.2). One limitation for this RNN classification analysis methodology is that only track segments with



	BM (%)	FBM (%)	CTRW (%)	LW (%)	Main Text Figure
Mixture of BM and FBM	6.04	90.35	2.71	0.90	3B
	90.06	1.57	0.66	7.71	
Mixture of Four Anomalous Diffusion Types	4.41	90.89	3.63	1.07	3D
	91.44	0.59	0.23	7.73	
	81.60	0.92	0.75	16.73	
	0.33	82.59	15.26	1.82	
SusG-HT Experimental Data	0.75	53.67	44.27	1.31	4C
	2.36	69.10	27.27	1.28	
	79.17	2.39	18.30	0.13	

Table 2.4: NOBIAS RNN module diffusion type classification probabilities for BM, FBM,CTRW, or LW. These data correspond respectively to Figure 2.6 as indicated.

at least certain length (20 or 40 in our analysis depending on the trained network) could be classified with high accuracy; it is very challenging to use very short track segments to identify these modes of diffusion. Therefore, when the overall trajectory length is short (10 steps), the network classification module might not be usable. Another limitation of the HDP-HMM module is that the current implementation is based on BM displacement distributions, thus it would fail for anomalous diffusion types like LW, which does not have a similar Gaussian distribution of displacements.

2.3.3 The Performance of NOBIAS on experimental data for the diffusion of SusG-HaloTag in *B. theta* cells

After validating the performance of the two NOBIAS modules on simulated data, we applied this framework to experimental single-molecule trajectories. The SusG amylase recognizes and binds starch on the surface of *B. theta* cells to enable starch catabolism [125]. We measured the motion of 7897 trajectories (minimum length of 6 and average length of 64) of single SusG molecules in 226 *B. theta* cells based on imaging photoactivatable fluorescently labeled SusG-HaloTag fusions (2.2).

We analyzed this data with NOBIAS to infer the number of diffusive states and to

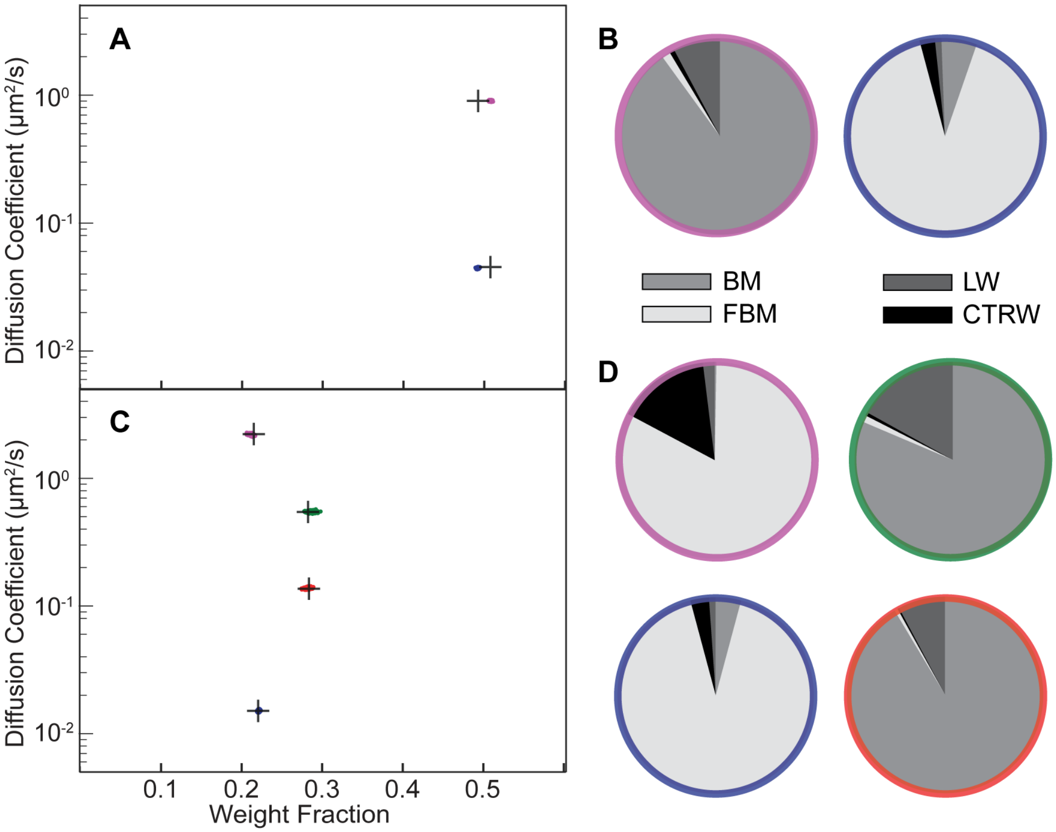


Figure 2.6: (A, C) The HDP-HMM module identifies distinct mobility states (colored clusters). Each point represents the average apparent single-molecule diffusion coefficient, D, vs. weight fraction in each distinct mobility state at each iteration of the Bayesian algorithm saved after convergence. The black crosses indicate the ground truth input for these simulated trajectories. (A) Two-state mixture comprising a subdiffusive FBM state with lower D and a BM state with higher D. (B) The NOBIAS-RNN determines the probability that the diffusion type for each diffusive state in (A) is classified as BM, FBM, CTRW, or LW. The final probability for each diffusive state is the average of the classification probability of its track segments weighted by the segment length. The color of each pie chart indicates the diffusive state corresponding to the color in (A). (C) Four-state mixture comprising a subdiffusive FBM state, two BM states, and a superdiffusive FBM state with D in ascending order. (D) Diffusion type classification probability pie chart for each diffusive state in (C). The final probability for each diffusive state is the average of the classification probability of its track segments weighted by the segment length and the color of each pie chart indicates the diffusive state corresponding to the corresponding color in (C).

estimate the diffusion coefficient, weight fraction, and type of motion for each state as was done for the simulated data (Figure 2.4 and 2.6). Additionally, NOBIAS analyzes 2D trajectories with a 2D Gaussian function and can therefore infer the diffusion coefficients for the x and y directions separately and estimate the potential correlation between the two directions. Though the simulations used symmetric tracks in an unbound domain, the experiments measure motion on the surface of cells with a long axis and a short axis, which may create an asymmetry in the diffusion. We rotated the cell orientations to orient the long axis in the x direction without rescaling (Figure 2.7A). We analyzed this rotated dataset with NOBIAS and found that it converged to a 3-state model, with a very small (1.8%) fast state fraction (Figure 2.7B). Interestingly, we found that the D_x and D_y values were similar for each of the two slower states (Table 2.5), while they were significantly different for the fastest state ($D_x = 0.68 \mu m^2/s vs. D_y = 0.45 \mu m^2/s$). This asymmetry for the fast state indicates that it corresponds to free diffusion that is constrained by the cell shape (and therefore is more constrained in the short-axis y direction), while the symmetry for the two slower states implies molecules that only diffuse regionally and are not affected by the cell shape. Compared with previous SPT analysis methods, NOBIAS provides a two-dimensional analysis of the dynamics of experimental single-molecule trajectories.

We separated the track segments by the state sequence label from the HDP-HMM module and placed each group into the RNN classification module. The fastest state was predicted with high probability (80%) to be Brownian motion (Figure 2.7C, Table 2.4), consistent with the asymmetry between D_x and D_y that was attributed to free diffusion (Figure 2.7B). The two slower states were predicted to be either FBM or CTRW. We used a RNN regression network (2.2) to estimate the anomalous exponent α for the track segments of the two slower states and both were found to be subdiffusion ($\alpha_1 = 0.38, \alpha_2 = 0.46$), consistent with the symmetry between D_x and D_y found (Table 2.5). This finding of subdiffusion is also consistent with the role of SusG in starch catabolism: we have previously found that SusG motion slows in the presence of its amylopectin substrate, as well as when

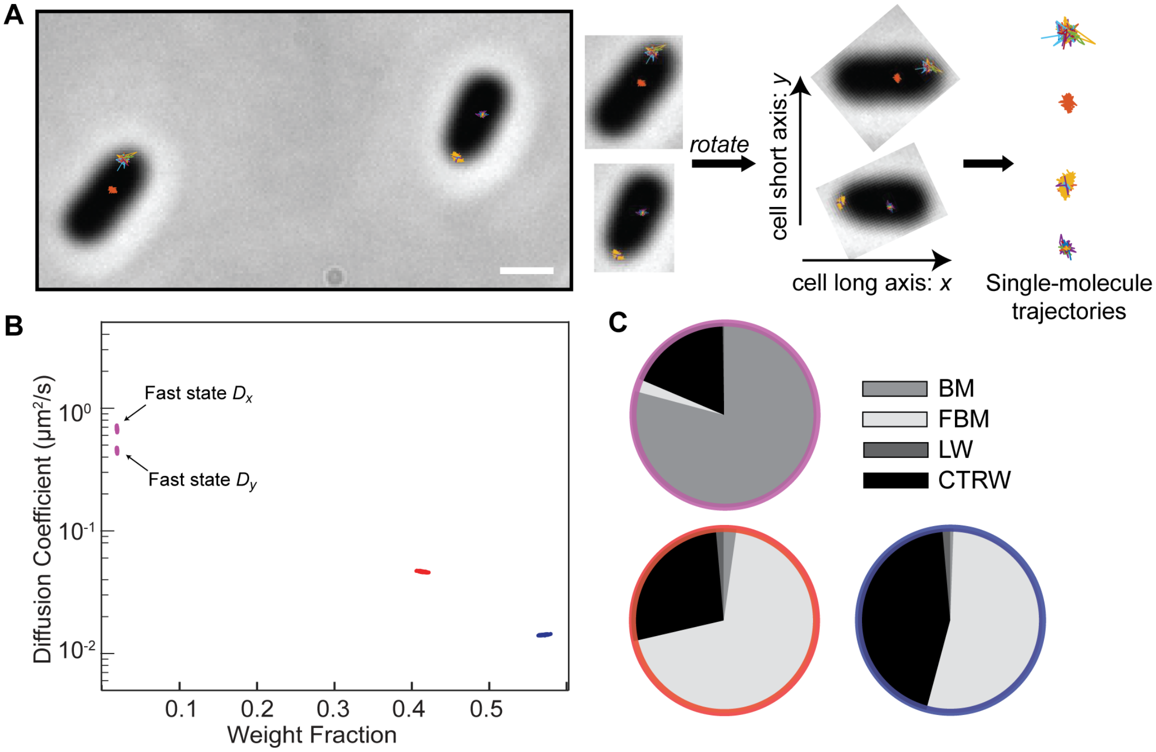


Figure 2.7: (A)Single-molecule trajectories of SusG-HaloTag overlaid on the phase-contrast image of the corresponding *B. theta* cells, scale bar: 1 μm. The long axis of the phase mask for each cell was detected and a rotation transform was applied to all the trajectories in each cell such that the x-axis is the cell long axis for all cells. (B) The NOBIAS HDP-HMM module identifies three diffusive states for SusG (colored clusters). Each point represents the average apparent single-molecule diffusion coefficient vs. weight fraction in each distinct mobility state at each iteration of the Bayesian algorithm saved after convergence. The blue and red points clusters average the x- and y- diffusion coefficients as they are symmetric (Table 2.4); the asymmetric fast state (purple) shows a different D_x and D_y. (C) The NOBIAS RNN determines the probability that the diffusion type for each diffusive state in (B) is classified as BM, FBM,CTRW, or LW. The color of each pie chart indicates the diffusive state corresponding to the color in (B). The fast state (purple) is predicted with high probability to be BM; the two slower states (red and blue) are predicted to be FBM or CTRW.

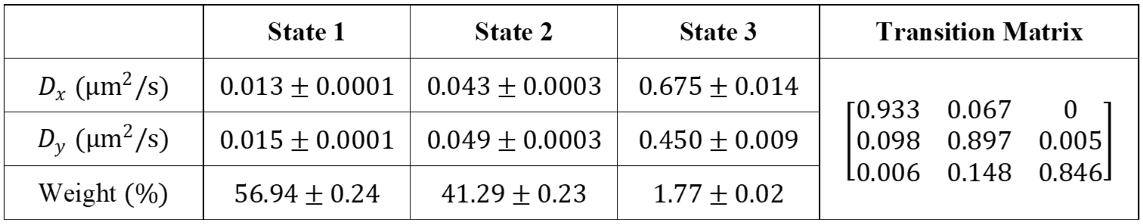
it transiently associates other outer-membrane proteins, indicating starch-mediated Sus complex formation [120].

2.4 Discussion

Single-molecule tracking measures dynamics in biological systems at high spatial and temporal resolution, but how to make the best use of these tracking data for a broad set of experimental conditions remains an analysis challenge in the field [126, 127]. Here, we have introduced NOBIAS to quantify single-molecule dynamics and to associate these biophysical measurements with the underlying biochemical function and biological processes. NOBIAS handles complicated live-cell SPT datasets for which: (1) the number of diffusive states is unknown, (2) mixtures of different diffusive populations may exist, even within single trajectories, (3) symmetry cannot be assumed between the x and y directions, and (4) anomalous diffusion is possible. These features are enabled based on applying Nonparametric Bayesian statistics [32, 34, 37] to SPT datasets that have the same means but different variance with a HDP-HMM module that has a 2D Gaussian as the emission function and then by further investigating the anomalous diffusion types in the RNN module of NOBIAS .

Compared with previous applications of nonparametric Bayesian statistics in this field [29, 31, 107], the NOBIAS HDP-HMM module is more robust and has high computational efficiency (Table 2.6). NOBIAS and SMAUG both consider motion blur effects and their estimation of D for each state is closer to the ground truth then other methods. As Bayesian method with similar principle NOBIAS is almost 10 times faster than SMAUG.

This HDP-HMM module also provides a multivariate output to quantify and correlate dynamics in multiple directions instead of assuming symmetry (Table 2.7). We observed that for asymmetric simulated trajectories, vbSPT overestimates the true number of states, and SMAUG can only provide the average D of for each diffusive state while NOBIAS provides the respective diffusion coefficients in two directions. The high accuracy of step

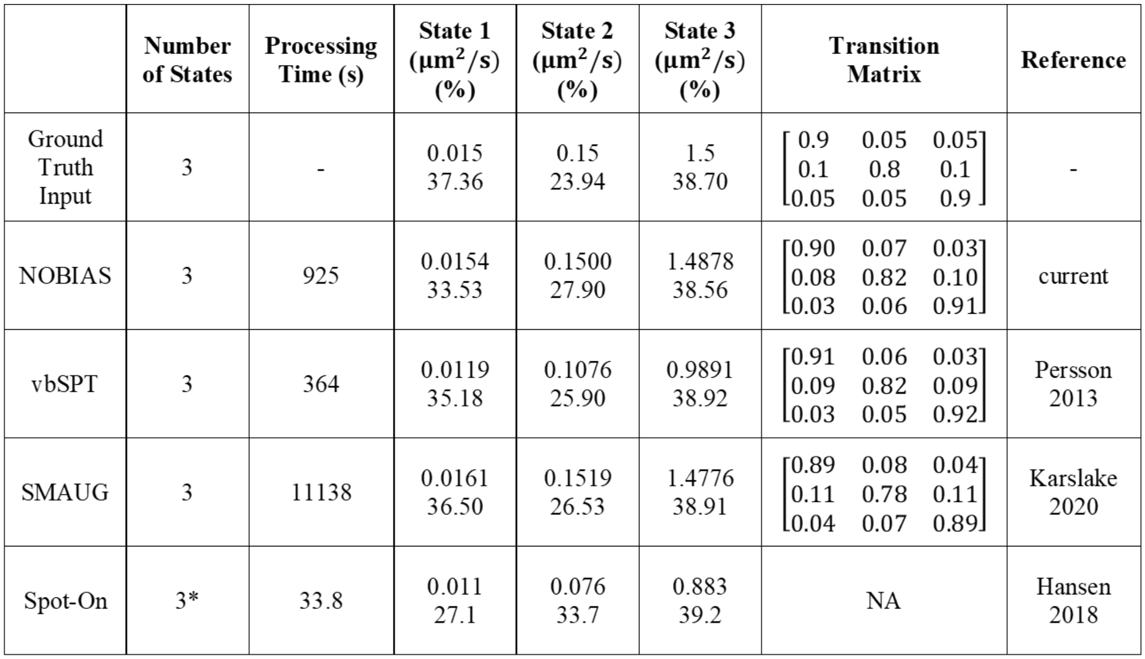


	State 1	State 2	State 3	Transition Matrix
D_x (μm^2/s)	0.013 ± 0.0001	0.043 ± 0.0003	0.675 ± 0.014	$\begin{bmatrix} 0.933 & 0.067 & 0 \\ 0.098 & 0.897 & 0.005 \\ 0.006 & 0.148 & 0.846 \end{bmatrix}$
D_y (μm^2/s)	0.015 ± 0.0001	0.049 ± 0.0003	0.450 ± 0.009	
Weight (%)	56.94 ± 0.24	41.29 ± 0.23	1.77 ± 0.02	

Table 2.5: NOBIAS HDP-HMM module results for analysis of the experimental measurements of SusG-HT diffusion in the *B. theta* outer membrane corresponding to main text Figure 2.7B. The x-axis and y-axis diffusion coefficients (Dx and Dy, respectively) are evaluated separately in NOBIAS.

state sequence prediction also enables the classification of anomalous diffusion type in the NOBIAS RNN module. We also applied SMAUG and vbSPT on the experimental dataset (Table 2.8): SMAUG ran slow on large datasets and suggested four diffusive state, while vbSPT suggested the best model to be 10 diffusive state which is hard to explain their corresponding biological meanings.

A further advantage of NOBIAS lies in its ability to treat sets of relatively short trajectories (10-step trajectories in the simulated data of Figures 2.4 and 2.6 and minimal 6-step trajectories in the experimental data of Figure 2.7). The recent AnDi (Anomalous Diffusion) Challenge [47] demonstrated that Deep Learning and Neural Network methods are currently the most powerful tools to study anomalous diffusion [45, 109]. However, in this challenge, the target dataset was an ideal collection of simulated anomalous diffusion trajectories with 100-1000 steps, and only the simple case of one state transition in the middle part of a track was considered. There are also probability-based models that consider confinement and anomalous diffusion [128] and Bayesian methods that directly predict the diffusion type [129, 130], but these analyses, like the neural network-based methods, are used for very long trajectories or assume a single diffusive state in each track. To apply a deep learning-based diffusion type classifier to realistic simulated trajectories and real experimental trajectories, NOBIAS segments the raw trajectories into collections of track segments that belong to the same diffusive state (as identified by the HDP-HMM module) and then predicts the diffusion type of the long segments in the RNN module. Since differ-



	Number of States	Processing Time (s)	State 1 ($\mu m^2/s$) (%)	State 2 ($\mu m^2/s$) (%)	State 3 ($\mu m^2/s$) (%)	Transition Matrix	Reference
Ground Truth Input	3	-	0.015 37.36	0.15 23.94	1.5 38.70	$\begin{bmatrix} 0.9 & 0.05 & 0.05 \\ 0.1 & 0.8 & 0.1 \\ 0.05 & 0.05 & 0.9 \end{bmatrix}$	-
NOBIAS	3	925	0.0154 33.53	0.1500 27.90	1.4878 38.56	$\begin{bmatrix} 0.90 & 0.07 & 0.03 \\ 0.08 & 0.82 & 0.10 \\ 0.03 & 0.06 & 0.91 \end{bmatrix}$	current
vbSPT	3	364	0.0119 35.18	0.1076 25.90	0.9891 38.92	$\begin{bmatrix} 0.91 & 0.06 & 0.03 \\ 0.09 & 0.82 & 0.09 \\ 0.03 & 0.05 & 0.92 \end{bmatrix}$	Persson 2013
SMAUG	3	11138	0.0161 36.50	0.1519 26.53	1.4776 38.91	$\begin{bmatrix} 0.89 & 0.08 & 0.04 \\ 0.11 & 0.78 & 0.11 \\ 0.04 & 0.07 & 0.89 \end{bmatrix}$	Karslake 2020
Spot-On	3*	33.8	0.011 27.1	0.076 33.7	0.883 39.2	NA	Hansen 2018

Table 2.6: acNOBIAS Comparison of results for analyzing simulated Brownian motion data with symmetric diffusion coefficient inputs in NOBIAS and three established non-parametric Bayesian statistics algorithms. The truncation level of NOBIAS, the max state number of vbSPT, and the starting number of states in SMAUG were all set to 10.

ent biophysical diffusive states correspond to different biochemical functions which will exhibit different diffusion types due to interactions like confinement, binding, directional motion, NOBIAS enables a thorough investigation of these biochemical roles by revealing the diffusion coefficients, the transition probabilities between states, and the anomalous diffusion behaviors. Ultimately, NOBIAS will enable investigators to extract a complete information set from SPT data and to understand the role of each tracked molecule, even in the living cell.

Despite these strengths, NOBIAS has several limitations. Firstly, as an HMM-based method, NOBIAS is limited by the length of each track. Under the extreme case where only very short trajectories (2-5 steps) are available, the HDP-HMM module may suggest a number of states and posterior results with extremely high uncertainty; probability-based models [17] or the histogram-based Bayesian method DPMM [107] should be applied for these short trajectories. The track length also limits the RNN module, as the trained net-

	Number of States	State 1 (µm²/s) (%)	State 2 (µm²/s) (%)	State 3 (µm²/s) (%)	Transition Matrix	Reference
Ground Truth Input	3	D_x: 0.021 D_y: 0.009 36.65	D_x: 0.21 D_y: 0.09 24.49	D_x: 2.1 D_y: 0.9 38.86	$\begin{bmatrix} 0.9 & 0.05 & 0.05 \\ 0.1 & 0.8 & 0.1 \\ 0.05 & 0.05 & 0.9 \end{bmatrix}$	-
NOBIAS	3	D_x: 0.0216 D_y: 0.0094 32.82	D_x: 0.2025 D_y: 0.0918 28.11	D_x: 2.0657 D_y: 0.8868 39.06	$\begin{bmatrix} 0.90 & 0.07 & 0.03 \\ 0.08 & 0.82 & 0.10 \\ 0.03 & 0.06 & 0.91 \end{bmatrix}$	current
vbSPT	4	0.0115 34.64	0.1010 24.31	0.5191 9.27 †	$\begin{bmatrix} 0.91 & 0.05 & 0.03 & 0.01 \\ 0.09 & 0.80 & 0.07 & 0.04 \\ 0.07 & 0.03 & 0.56 & 0.34 \\ 0.01 & 0.06 & 0.16 & 0.77 \end{bmatrix}$	Persson 2013
SMAUG	3	0.0153 34.04	0.1493 27.32	1.4996 38.64	$\begin{bmatrix} 0.878 & 0.086 & 0.036 \\ 0.11 & 0.75 & 0.14 \\ 0.035 & 0.09 & 0.875 \end{bmatrix}$	Karslake 2020
Spot-On	3*	0.011 27.4	0.073 33.2	0.83 39.5	NA	Hansen 2018

Table 2.7: NOBIAS Comparison of results for analyzing simulated Brownian motion data with asymmetric diffusion coefficient inputs in NOBIAS and three established nonparametric Bayesian statistics algorithms. The truncation level of NOBIAS, the max state number of vbSPT, and the starting number of states in SMAUG were all set to 10.

	Number of States	Running time (s)	State 1 (µm²/s) (%)	State 2 (µm²/s) (%)	State 3 (µm²/s) (%)	Transition Matrix	Reference
NOBIAS	3	28790.8	D_x: 0.013 D_y: 0.015 56.94	D_x: 0.043 D_y: 0.049 41.29	D_x: 0.675 D_y: 0.450 1.77	$\begin{bmatrix} 0.933 & 0.067 & 0 \\ 0.098 & 0.897 & 0.005 \\ 0.006 & 0.148 & 0.846 \end{bmatrix}$	current
vbSPT	10	10067.7	†	†	†	†	Persson 2013
SMAUG	4	264930.6	0.0025 33.1	0.0040 49.2	0.4121 1.4 *	$\begin{bmatrix} 0.92 & 0.08 & 0 & 0 \\ 0.05 & 0.88 & 0.07 & 0 \\ 0.01 & 0.20 & 0.79 & 0 \\ 0.01 & 0.02 & 0.03 & 0.94 \end{bmatrix}$	Karslake 2020

Table 2.8: Comparison of results for analyzing experimental data in NOBIAS and two established nonparametric Bayesian statistics algorithms. The truncation level of NOBIAS, the max state number of vbSPT, and the starting number of states in SMAUG were all set to 10.

work need tracks with at least 20 steps for good classification performance because some anomalous diffusion types are characterized by memory of previous steps [38]. Therefore the application of the RNN module is limited for short experimental tracks. Secondly, NOBIAS performs the diffusive state estimation based on apparent diffusion coefficient in the HDP-HMM module and then carries out the anomalous diffusion classification in the RNN module. NOBIAS therefore assumes that each biochemical state has a unique average apparent diffusion coefficient. Although the RNN module can classify the diffusion types of two different diffusive states with the same diffusion coefficient, the HDP-HMM module would fail to separate these processes. Furthermore, for some diffusion types like LW, the trajectory displacements may exhibit different types of dynamics even though the trajectories are generated from one process. Finally, even for Brownian trajectories, a single biochemical state might not be represented by a single diffusion coefficient value. Thus, the actual number of biochemical states may not be equal to the number of diffusive states. Future development of NOBIAS could use spatial filtering to distinguish between these similar biochemical states.

NOBIAS provides a pioneering and compatible framework for the analysis of dynamical mixtures that also classifies the anomalous diffusion types. Future development of NOBIAS could include more types of diffusion and could integrate the anomalous distributions directly into the Bayesian framework for more accurate prediction of the stepwise state labels and the diffusion types. Furthermore, extra experimental corrections corresponding to the specific microscope setting [18, 100, 102] could also help adapt NOBIAS more broadly to different types of SPT datasets. Overall, NOBIAS has provided a powerful framework to analyze of SPT dataset with unknown number of diffusive states and potential asymmetric diffusion, and to access the anomalous diffusion type for each diffusive state. The combination of nonparametric Bayesian statistics and Deep learning enables NOBIAS to fully extract the rich dynamics information from the SPT dataset.

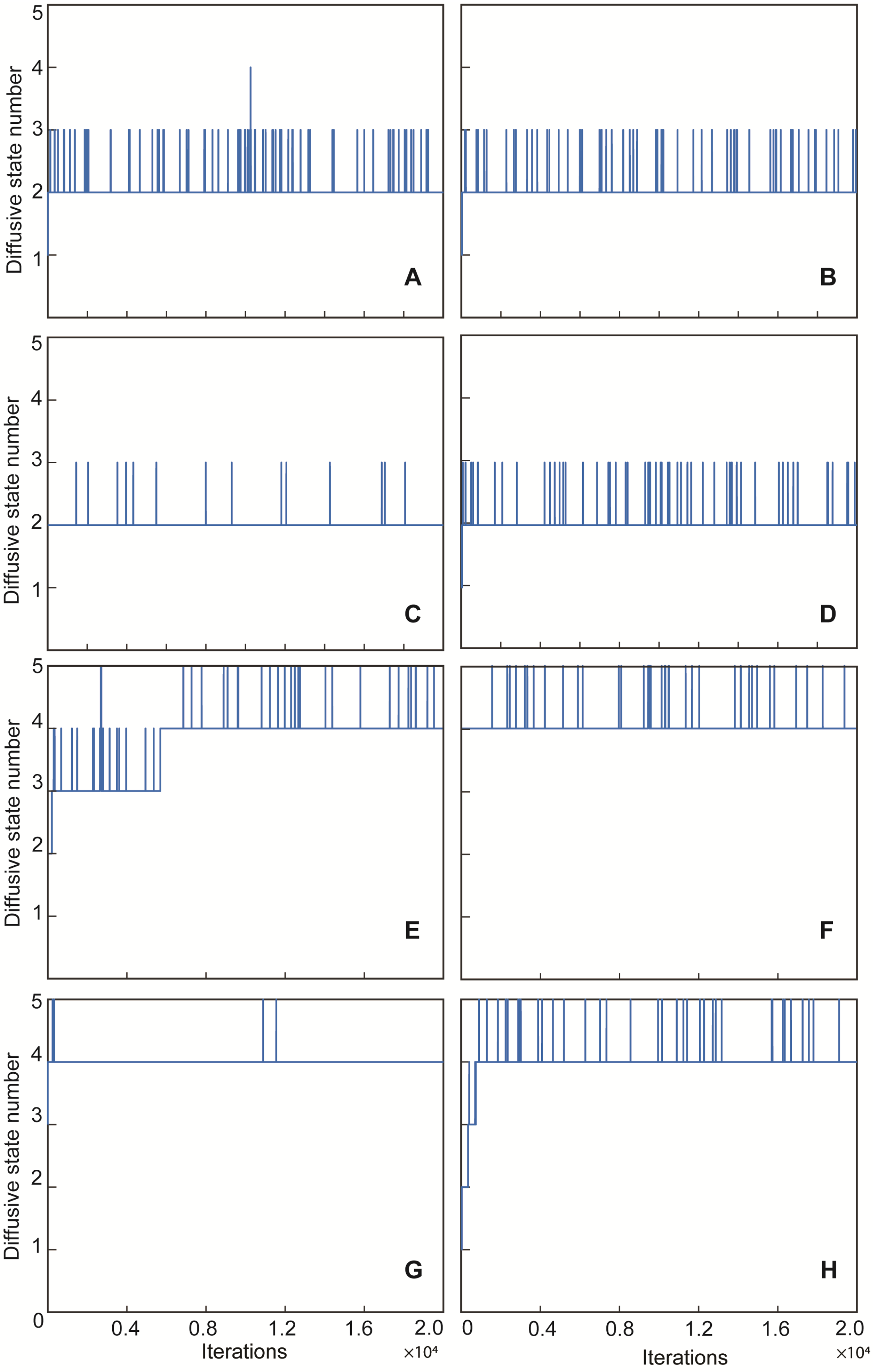

5
4
3
2
1
0
Diffusive state number
A
B
C
D
E
F
G
H
0
0.4
0.8
1.2
1.6
2.0
Iterations
×10⁴

Figure 2.8

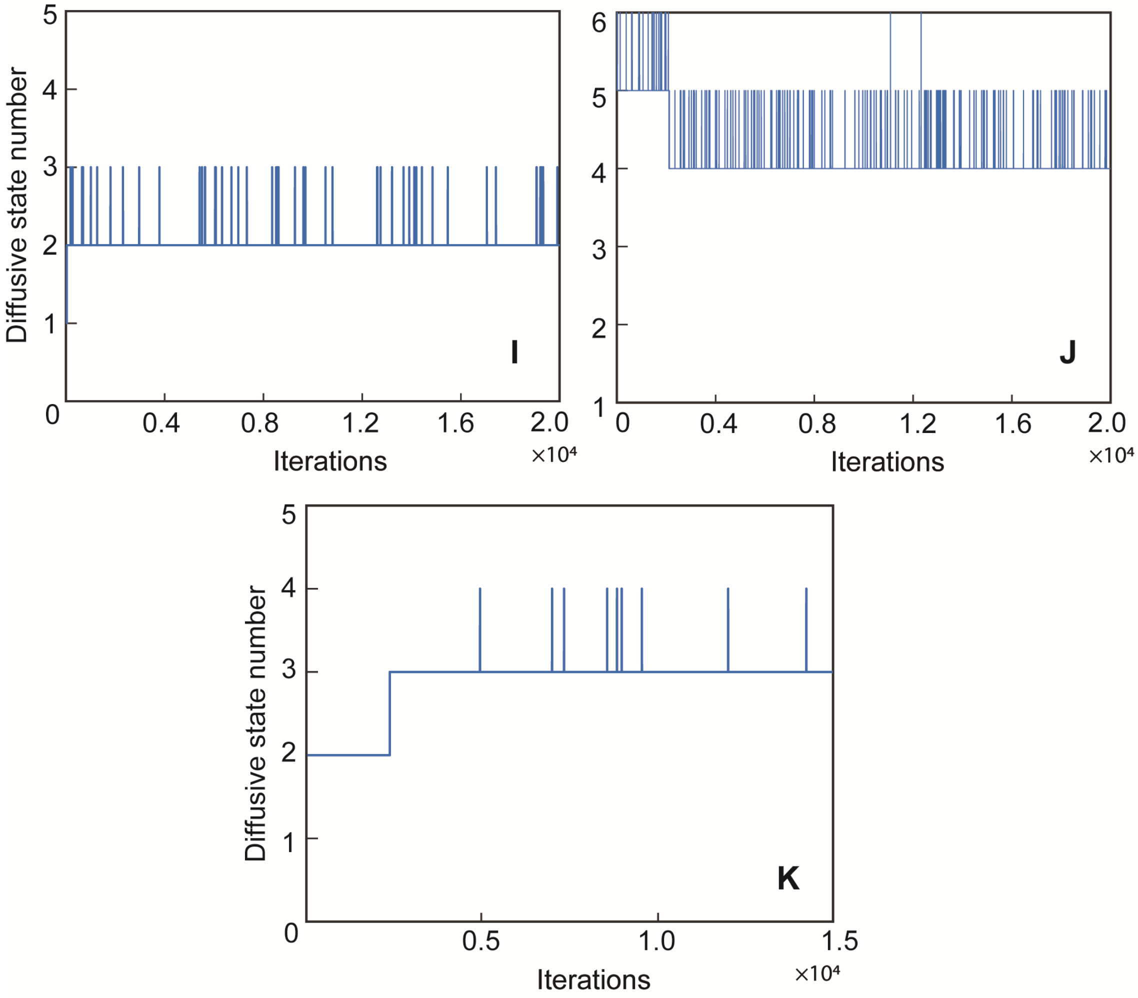


Figure 2.8: Convergence of the number of diffusive states in the NOBIAS HDP-HMM module with iteration number. The number of states convergence plots in A-H correspond to the analysis of simulated tracks in main text Figure 2.4A-H. The number of states convergence plots in I-J correspond to the analysis of simulated tracks in main text Figure 2.6A,C. The number of states convergence plot in K corresponds to the analysis of experimental tracks in main text Figure 2.7B.

CHAPTER III

Single-Molecule Localization of a DNA Methyltransferase DnmA on the Bacterial Nucleoid

The work presented in this chapter was previously published in mBio.

Fernandez, N. L*., Chen, Z.*, Fuller, D. E., van Gijtenbeek, L. A., Nye, T. M.,
Biteen, J. S., and Simmons, L. A.
DNA Methylation and RNA-DNA Hybrids Regulate the Single-Molecule
Localization of a DNA Methyltransferase on the Bacterial Nucleoid.
mbio, e03185-22(2023). DOI: 10.1128/mbio.03185-22

In this work, I participated in the conceptualization of the project and the design of used biological perturbations. I performed single-molecule imaging experiments and dynamics analysis. I designed and implemented the normalized heatmap correlation analysis, plotted the analysis of the single-molecule dynamics, and interpreted the comparison between different datasets. *: equal contribution.

3.1 Introduction

The restriction modification (RM) systems were one of the first recognized defense mechanisms that bacteria use to thwart bacteriophage infection [131, 132]. Initial bacteriophage studies identified that only phage that have been modified by a host can successfully infect the host [132]. This modification was later identified as DNA methylation from enzymes called DNA methyltransferases (MTase) (Reviewed in [60,61]). MTases from RM systems modify DNA by adding a methyl group in a sequence-specific context to form either N6-methyladenosine (m6A), N4-methylcytosine, or 5-methylcytosine [59]. Genes encoding MTase function are often adjacent to genes encoding restriction endonuclease (REase) activity [133]. If a cell encodes an RM system and unmethylated DNA enters the host cell, for example from a phage, REase will degrade the invading DNA before it can be replicated [133].

In addition to functioning in RM systems, DNA methylation regulates other processes including DNA replication, DNA repair, and transcription [53]. Many gammaproteobacteria encode dam, which is referred to as an orphan DNA MTase because it lacks a cognate REase enzyme [53, 61, 134]. In *Escherichia coli (E. coli)*, DNA methylation by Dam influences the timing of replication and aids in the excision of mismatched bases from the new DNA strand following replication during methyl-directed mismatch repair [61]. Alphaproteobacteria also encode the conserved orphan MTase ccrM that regulates the timing of DNA replication and is essential for *Caulobacter crescentus* growth in rich media [61].

Epigenetic regulation of gene expression in bacteria results from the interaction between certain DNA-binding proteins and methylated DNA [135]. Orphan and RM-associated MTase enzymes influence gene expression and bacterial behaviors through DNA methylation, with examples ranging from pili expression in *E. coli*, eukaryotic cell adhesion in Campylobacter jejuni, and virulence regulation in Streptococcus pyogenes [136–138]. Studies have benefited from the use of single-molecule real-time (SMRT) sequencing analysis to characterize the methylome and identify sites of methylation followed by predicting the MTase enzymes responsible for the corresponding modification [134].

Previously, we used SMRT sequencing to characterize the methylome of the Gram-positive soil bacterium *Bacillus subtilis (B. subtilis)* [63]. We identified the DNA MTase DnmA (M.BsuPY79I), which recognizes the six base-pair, non-palindromic sequence 5′-GACGAG-3′ and methylates adenine to form m6A [63]. *In vitro* methylation assays with DnmA demonstrated substrate specificity: double-stranded DNA (dsDNA) harboring the methylation site was identified as the optimal substrate, followed by single-stranded DNA (ssDNA) and single-stranded RNA (ssRNA) [63]. DNA substrate compositions heavily influence DNA and MTase interactions *in vitro* for some well characterized MTases, but how these *in vitro* experiments inform *in vivo* activity is not well understood [139, 140]. The *dnmA* gene is flanked by *yeeB* and *yeeC*, two genes with putative REase functions, in a genetic structure suggestive of an operon from a horizontally acquired element. While deletion of *dnmA* alters the expression of a subset of genes, the growth rate and restriction of plasmid uptake are unchanged. Therefore, it remains unclear if *dnmA-yeeB-yeeC* are functional under stress conditions, such as bacteriophage infection.

In this study, we identify how different substrates influence the *in vitro* DnmA binding kinetics and how that affects *in vivo* DnmA dynamics. We also investigate the conservation of the gene synteny and architecture between dnmA and its genetic neighbors across many bacterial species, and we assess the role of *dnmA* in response to bacteriophage infection. We show that the association of DnmA with DNA *in vitro* and *in vivo* is regulated by prior DNA methylation and formation of RNA-DNA hybrids. We also show that DnmA searches the entire nucleoid but localizes more strongly at the replisome position, suggesting that binding site recognition can occur anywhere on the chromosome with preference for positions near the replisome. Furthermore, we find that *dnmA* and the flanking genes *yeeB* and *yeeC* do not function as an active RM system and fail to protect *B. subtilis* from phage predation. Our work demonstrates how substrate specificity alters the *in vivo* localization of an MTase that arises from a restriction modification relic, causing DnmA to function as an orphan MTase in the regulation of gene expression in *B. subtilis*.

3.2 Results

3.2.1 Localization of DnmA-PAmCherry *in vivo*

Our prior work showed that DnmA is both necessary and sufficient to methylate dsDNA *in vitro* and *in vivo* [63]. Given the role of DnmA in altering gene expression [63], it is important to understand how DnmA interacts with DNA *in vivo*. To this end, we generated a *B. subtilis* strain in which the wild-type (WT) *dnmA* allele was replaced with a gene encoding DnmA fused to a photoactivatable fluorescent protein, PAmCherry, at the C-terminus (*dnmA-PAmCherry*). To ensure DnmA-PAmCherry retained methyltransferase activity *in vivo*, we measured the activity of a transcriptional reporter that is dependent on DnmA [63]. We found that reporter activity is the same between WT and DnmA-PAmCherry, indicating that the C-terminal tag does not interfere with DnmA function. Further, Western blot analysis demonstrated that the DnmA-PAmCherry fusion is not degraded *in vivo* (Figure 3.1B). Based on photoactivation and tracking of single copies of DnmA-PAmCherry in living cells (3.4) (Figure 3.2B) [12], we observed the localization of this protein in its native environment in N = 1766 single-molecule trajectories in n = 275 *B. subtilis* cells growing exponentially in defined minimal medium. We categorized the motion of these molecules based on fitting each single-molecule trajectory to a linear mean-square displacement model for normal diffusion (3.4); the histogram of the log diffusion coefficients for DnmA-PAmCherry trajectories weighted by the track length is given in Figure 3.2D. As a positional reference for nascent DNA, we imaged fusions of the replisome component DnaX to the fluorescent protein mCitrine in a separate fluorescence channel (Figure 3.2A).

The overlay of the super-resolution images of DnaX-mCitrine (grayscale) and DnmA-PAmCherry (jet) shows some spatial overlap for DnmA and DnaX, although the DnmA positions are more spread out over the region of the cell occupied by the nucleoid (Figure 3.2C). To further quantify their spatial correlation at the population level, we generated a normalized localization density map of DnmA to determine the localization pattern of DnmA in 275 WT cells (Figure 3.2F). We also generated a normalized localization density map of the replisome by analyzing DnaX-mCitrine (Figure 3.2E) [65]. The Pearson correlation between the two heatmaps is 0.37, showing that DnmA has a positive spatial correlation with the replisome. Due to the non-palindromic nature of the DnmA recognition site, nascent DNA will be unmethylated post-replication, acting as a substrate for methylation by DnmA. Thus, our data suggests that binding and methylation of nascent, unmethylated DNA drive the correlative positioning of DnmA and DnaX, although DnmA does explore much more of the nucleoid region in the cell.

3.2.2 Manipulating Available Substrate *in vivo* Disrupts DnmA Localization

DNA binding is heavily influenced by substrate, where most N6-DNA MTase enzymes tend to have lower binding affinities toward substrates that are not dsDNA *in vitro* [139, 140]. We hypothesized that the position of DnmA can be explained by the availability of unmethylated substrate near the replisome, where unmethylated dsDNA would be enriched shortly after DNA replication. To test this hypothesis, we first set out to establish how altering DNA substrate influences DnmA binding *in vitro* using electrophoretic mobility shift assays (EMSAs). In addition to unmethylated dsDNA substrate, we utilized methylated dsDNA and an RNA-DNA hybrid as candidate substrates for possible *in vivo* DNA modifications or perturbations. Methylated dsDNA is the primary DNA species in *B. subtilis* grown under standard conditions, while RNA-DNA hybrids are transiently found throughout the genome from DNA replication and highly transcribed regions [63, 141]. DnmA binds to unmethylated dsDNA with the greatest estimated affin-

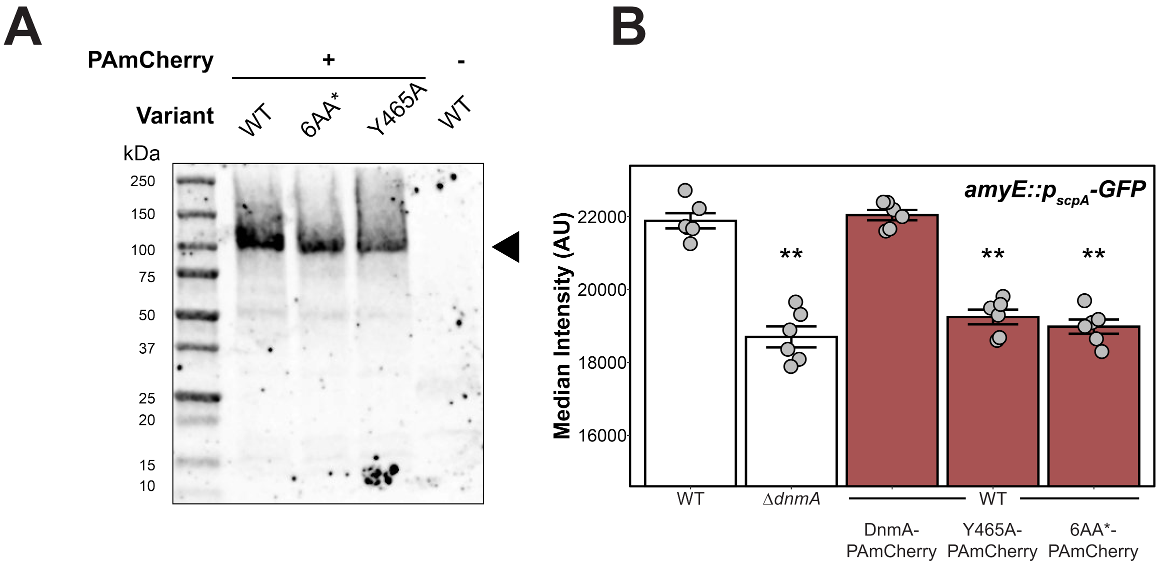


Figure 3.1: (A) Western blot analysis of DnmA-PAmCherry variants. Black arrow points to DnmA-PAmCherry. (B) Flow cytometry analysis of a GFP transcriptional reporter that is regulated by DnmA ($amyE :: P_{scpA} - GFP$) [63]. White bars indicate either WT DnmA or *dnmA* deletion, while filled bars indicate either WT DnmA, Y465A DnmA, or 6AA DnmA fused to PAmCherry in an otherwise WT background. Bars represent the mean from six biological replicates (grey filled circles) and error bars represent standard deviation. Asterisks indicate statistical significance ($p < 0.05$) when compared to the WT background using the Wilcox Test.

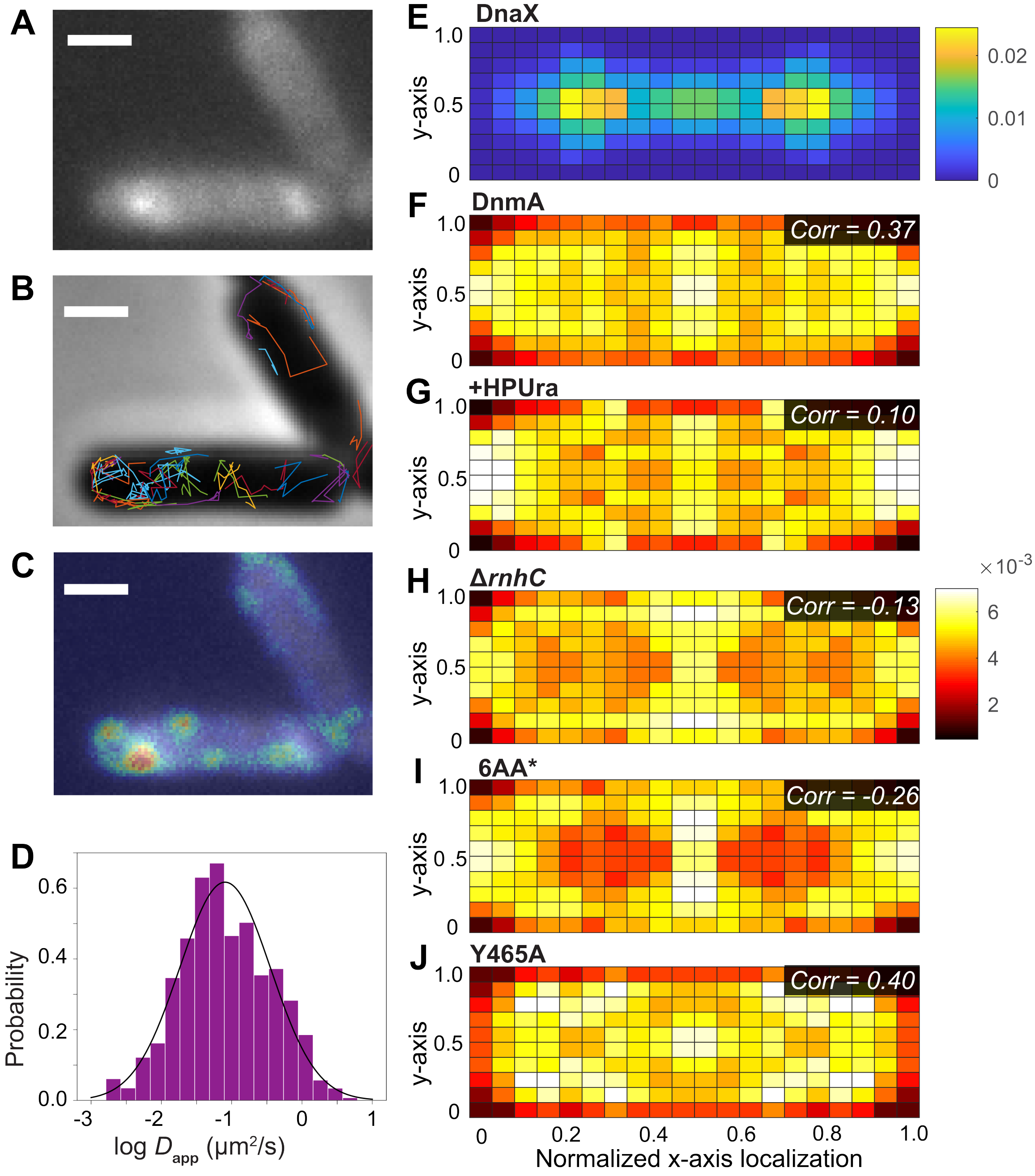


Figure 3.2: (A) Fluorescence image of DnaX-mCitrine. Scale bars = 1μm for panels A to C. (B) False-colored single-molecule trajectories of DnmA-PAmCherry in two representative WT cells overlaid on the phase-contrast image of the *B. subtilis* cells. (C) Overlay of single-molecule localizations of DnmA-PAmCherry (jet heatmap) and fluorescence image of DnaX-mCitrine (grayscale). (D) Normalized histogram showing the distribution of the log diffusion coefficients of the single-molecule trajectories of DnmA-PAmCherry. Black line, Gaussian fit to the log diffusion coefficient distribution. The histogram and fit curve are weighted by track length. catption continues

Figure 3.2: (E to J) Normalized localization probability density maps of (E) DnaX-mCitrine, (F) WT DnmA-PAmCherry, (G) + 6-(p-hydroxyphenylazo)-uracil (HPUra) DnmA-PAmCherry, (H) Δ*rnhC* DnmA-PAmCherry, (I) DnmA[6AA*]-PAmCherry, and (J) DnmA[Y465A]-PAmCherry, all within a normalized cell. Single-molecule localizations are projected along the long and short axes of the cell, normalized to their relative position, and resymmetrized along the axes. Colormaps show localization probability. *Corr* in panels F to J, Pearson's correlation of that DnmA variant's localization heatmap with the DnaX localization heatmap. Each single-molecule data set was acquired from 4 distinct days from independent cultures.

ity 50% effective concentration (EC50) = 36.8 ± 14.2 nM (mean ± sd) and has much lower estimated affinities for methylated dsDNA and RNA-DNA hybrids (EC50 = 156.4 ± 76.0 nM and EC50 = 321.4 ± 16.5 nM, respectively, Figure 3.3A-C,F). Though the range of DnmA concentrations for the methylated dsDNA and RNA-DNA substrates makes affinity calculations less accurate, we conclude that DnmA binds preferentially to unmethylated dsDNA relative to methylated dsDNA or RNA-DNA hybrids *in vitro*.

Since DnmA binds to unmethylated dsDNA with the greatest affinity *in vitro*, we reasoned that changing the pool of this substrate *in vivo* would alter DnmA dynamics and localization *in vivo*. We measured DnmA-PAmCherry localization after treating *B. subtilis* with the replication inhibitor HPUra which depletes the pool of available unmethylated dsDNA substrate in the cell [142]. We acquired single-molecule tracking data from N = 1047 trajectories within n = 233 cells. We found that HPUra treatment decreases the average diffusion coefficient (Figure 3.4A). The decreased affinity for methylated dsDNA *in vitro* suggests that the weight fraction of slow-moving DnmA molecules in HPUra-treated cells should decrease. However, we observe a slight increase in the weight fraction of slow-moving molecules in HPUra-treated cells compared to untreated cells with a concomitant decrease in the weight fraction of fast-moving molecules (Figure 3.4G, H). We also found that the DnmA correlation with DnaX decreases from 0.37 to 0.10 in HPUra-treated cells (Figure 3.2g). These data suggest that HPUra treatment likely does not decrease DNA binding throughout the nucleoid but does negatively influence DNA binding

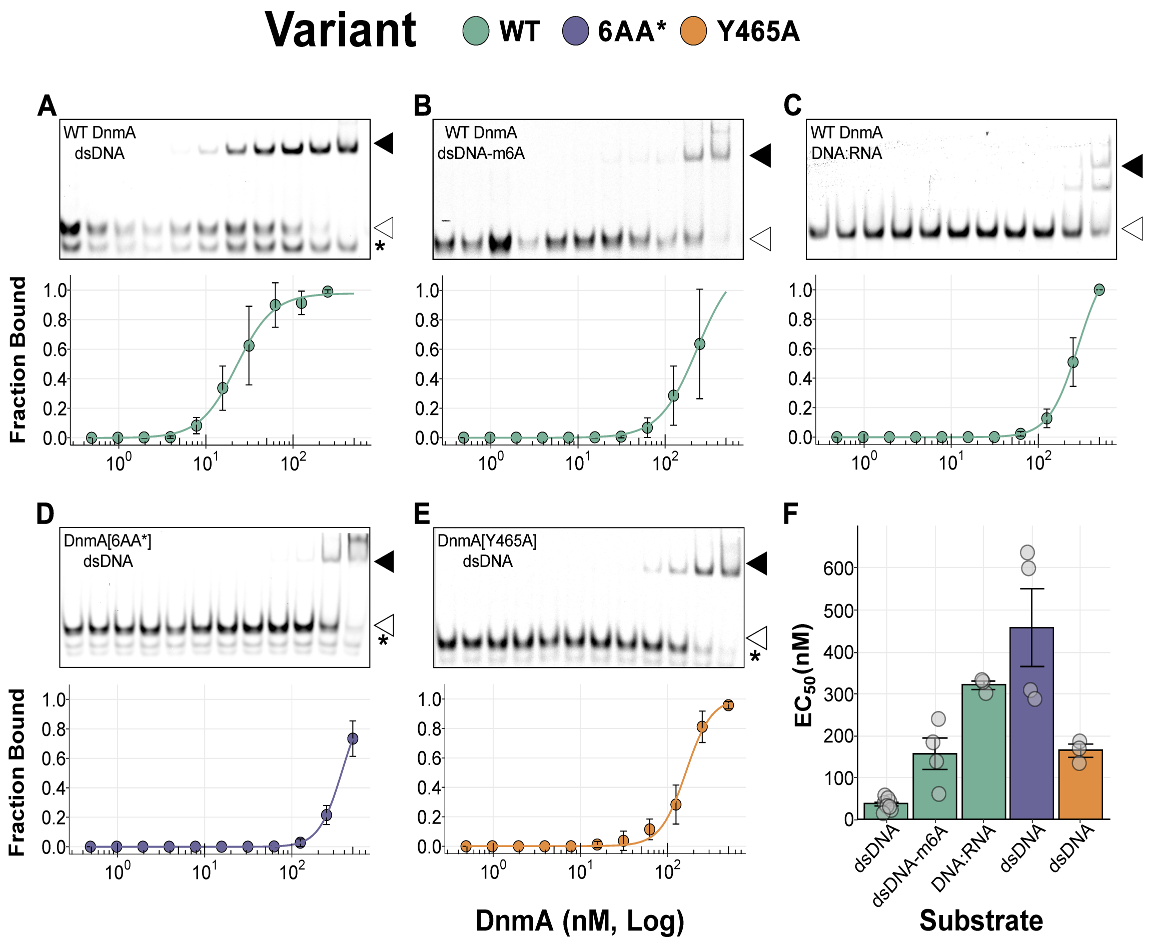


Figure 3.3: (A to E) EMSAs experiments with DnmA variants and different DNA substrates. Representative gels showing unshifted bands (white arrows), shifted bands (black arrows), or unannealed ssDNA (asterisks) (top) and quantitation of fraction bound with increasing concentrations of DnmA (bottom), where points represent the average, error bars represent standard deviations, and lines are modeled from four-parameter log-logistic equations. DnmA variant and DNA substrate are in the top-left corner of the representative gel. (F) Average ± standard deviation of estimated half maximal concentrations (EC50) for DNA binding calculated from the binding curves. Points represent individual experiments, and bar fill colors represent the DnmA variant.

near the replisome. Of note, this marked change in the DnmA-DnaX spatial correlation is observed *in vivo* even though 99.7% of DnmA recognition sites are methylated during exponential growth [63].

Next, we measured DnmA–PAmCherry localization in *B. subtilis* cells lacking the RNAase HIII gene *rnhC*, which is suggested to remove RNA–DNA hybrids in the genome [141]. We acquired the $\Delta rnhC$ single–molecule tracking data from N = 1348 trajectories within n = 226 $\Delta rnhC$ *B. subtilis* cells. Unlike the WT cells, in which DnmA and DnaX are positively spatially correlated, the localization density map of DnmA in $\Delta rnhC$ cells has a negative spatial correlation with DnaX (-0.13; Figure 3.2H). Further, more of the DnmA–PAmCherry molecules move slowly in $\Delta rnhC$ than in WT (50% slow population for WT compared to 60% for $\Delta rnhC$, Figure 3.4G). In summary, this mutation has a marked effect in decreasing the co–localization of DnmA with DnaX, and causes a subtly reduced average diffusion coefficient, resulting in an increase in the fraction of molecules diffusing slowly.

3.2.3 The DNA Binding Variant DnmA[6AA*] localizes away from the replisome and the nucleoid

Our data suggests that DNA binding and methylation explains DnmA–PAmCherry localization *in vivo*. To test this hypothesis, we generated variants of DnmA with amino acid substitutions at key residues involved in DNA binding and catalysis. DnmA is 57% similar to MmeI, a Type II DNA MTase for which a structure is available [143]. We structurally aligned DnmA with MmeI and identified putative residues important for DnmA interaction with its cognate sequence. Interestingly, single alanine substitutions in MmeI or other methyltransferases are often unable to completely abrogate DNA binding in vitro and can sometimes cause recognition of a different sequence [143–145], likely due to the high number of contacts between the residues in the DNA binding pocket and DNA (Figure 3.5). Therefore, we designed a six amino acid alanine substitution vari-

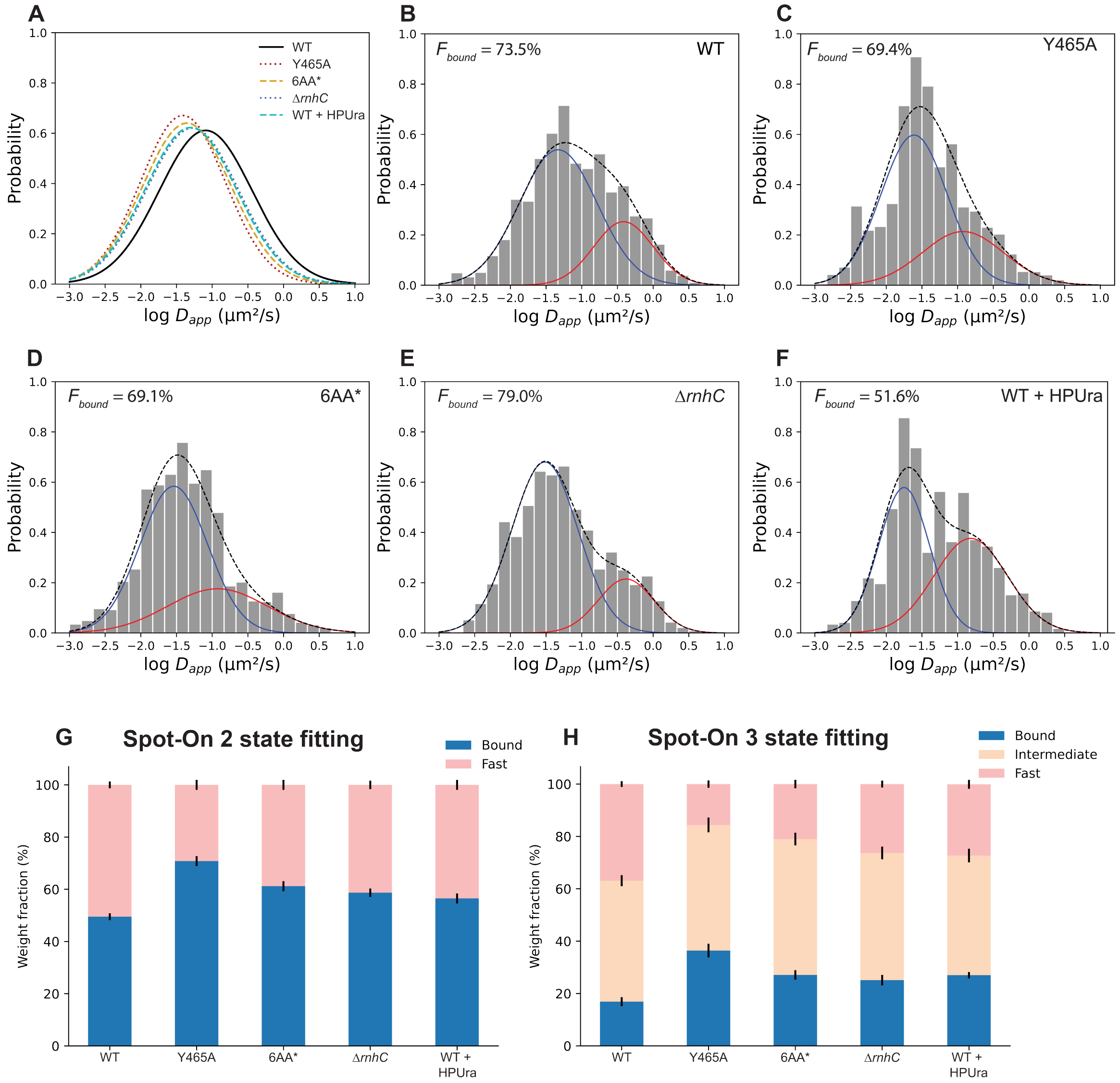


Figure 3.4: (A) Single-component Gaussian fit to the log diffusion coefficients distribution. The fit curves are weighted by track length. (B-F) Normalized histogram showing the distribution of the log diffusion coefficients of the single-molecule trajectories of DnmA-PAmCherry. Black dashed line: Two-component Gaussian fit to the histogram; blue line: fit of the slower component; red line: fit of the faster component. F_{bound} indicates the weight fraction for the slower component in each fit. The histograms are weighted by track length. (G) 2-state and (H) 3-state weight fraction bar plots from Spot-On analysis [18] of the five DnmA single-molecule tracking datasets. In the Spot-On fitting, the average diffusion coefficient for each component is kept within the confidence interval of the WT diffusion coefficient to directly compare weight fractions (see 3.4).

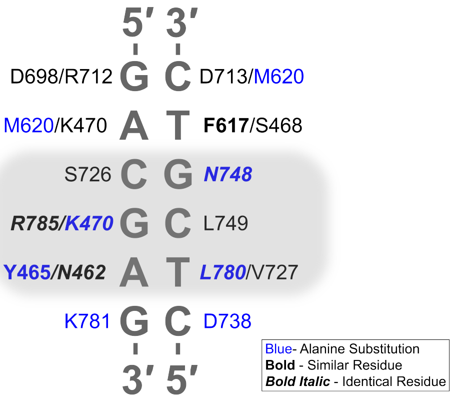


Figure 3.5: Predicted interactions between DnmA and its target sequence 5′–GACGAG–3′. Residues that are biochemically similar to aligned residues in the homologous MTase MmeI are in bold face while residues that are identical are in bold italic face. Residues chosen for alanine substitutions are in blue font color. The grey shading surrounding 5′–CGA–3′ represent the shared nucleotides between the cognate recognition sites for DnmA and MmeI.

ant of DnmA (DnmA[6AA*]) which has substitutions at key residues we predict are involved in 5′–GACGAG–3′ recognition (Figure 3.5). Further, we generated a catalytically inactive DnmA variant by introducing an alanine at position 465, replacing a tyrosine needed for stabilizing base–flipping during the methyl transfer reaction (DnmA[Y465A]), reviewed in [139]). We have previously shown this substitution renders DnmA inactive *in vivo* and *in vitro* [63]. *In vitro* analysis of unmethylated dsDNA binding by the DnmA variants showed a decrease in estimated affinity to DNA, with the most severe effect in DnmA[6AA*] which had a 12–fold greater EC50 (458.3 ± 185.2 nM) compared to WT DnmA while DnmA[Y465A] had a 4–fold greater EC50 (164.2 ± 26.8 nM) (Figure 3.3D–F).

We also introduced the DnmA variants fused to PAmCherry into the cell and checked for stability and functionality *in vivo*. The DnmA variants were not degraded *in vivo*, demonstrated by intact DnmA–PAmCherry fusions in Western blot analysis (Figure 3.1A). Importantly, the variants were unable to complement reporter activity in a Δ*dnmA* background, indicating both DnmA[6AA*] and DnmA[Y465A] are inactive *in vivo* (Figure 3.1B). Single–molecule tracking data and normalized localization density maps were generated for these two variants. The diffusion coefficient distributions for the two variants are lower than those of WT DnmA–PAmCherry (Figure 3.4A). The two variants demonstrated a decreased ability to bind DNA in vitro, yet *in vivo* we observed an increase in the weight fraction of slow–moving molecules (Y465A – 70%, 6AA* – 60%, Figure 3.4G) compared to WT DnmA (50%, Figure 3.4G). Strikingly, DnmA[6AA*] also has a strong negative correlation with DnaX (−0.26) while DnmA[Y465A] has a correlation similar to WT DnmA (0.40 vs 0.37, Figure 3.3I,J). These data suggest DnmA[Y465A] is still able to scan and search DNA for available substrate but is unable to catalyze methylation because of its inability to stabilize the flipped base, whereas the DnmA[6AA*] variant is unable to scan and search DNA, relegating it to positions outside of the nucleoid region. Taken together, our results indicate that, regardless of substrate or variant, the mobility of DnmA is slower under these conditions, and that DnmA localization is primarily influenced by

DNA binding rather than by active methylation.

3.2.4 DnmA is Part of a Conserved Gene Cluster with YeeB and YeeC

Our *in vivo* single-molecule results suggest that DnmA, in part, co-localizes with the replisome to fully methylate the *B. subtilis* chromosome as replication occurs, raising questions about the function of m6A in *B. subtilis*. We have previously shown that m6A regulates the transcription of a subset of genes and that there is no difference in transformation efficiency in cells lacking m6A under the conditions tested [63]. However, we had not tested a role for m6A in protection from bacteriophage predation. In prior work, we showed that m6A functions in the Gram-positive pathogen Streptococcus pyogenes both in the regulation of gene expression and as part of a functioning RM system, supporting the idea that DnmA can play a role in restriction modification as part of the putative operon consisting of *dnmA*, *yeeB*, and *yeeC* genes [138]. YeeB has a C-terminal Superfamily II DNA/RNA helicase domain like those found in restriction endonucleases, while YeeC has a C-terminal T5 orf172-domain, a largely uncharacterized domain that is predicted to have multiple functions involving DNA binding [146]. In a bioinformatic survey, Makarova et al. identified YeeB and YeeC homologs as putative anti-phage genes often found in a type of genomic island termed defense islands, suggesting that the *dnmA* operon could be involved in phage defense [147]. The *dnmA* gene is also adjacent to two genes involved in DNA mobility (*yefB* and *yefC*) and to two putative toxin-antitoxin systems (*yeeD-yezA* and *yezG-yeeF*), while the whole region from *yefB* to *yeeF* is in a local GC-minimum compared to the surrounding genome (Figure 3.6A). Together, these findings suggest *dnmA*, *yeeB*, and *yeeC* were horizontally acquired and could represent a phage defense island [147, 148].

Given the information above, we asked if the *dnmA-yeeB-yeeC* gene cluster is conserved in other microorganisms and adjacent to genes with defense-associated protein families. We analyzed the genomic neighborhoods surrounding homologs of DnmA (10 genes upstream and 10 genes downstream) and scored the number of genes with predicted defense associated protein families (see 3.4). Neighborhoods harboring DnmA had, on average, 1.8 ± 1.2 genes with defense-associated protein families, while randomly selected regions of similar size had 0.56 ± 0.67 genes with defense-associated protein families (Figure 3.6B). The most common protein families adjacent to *dnmA* were homologous to *yeeB* (ATP-dependent helicase/Superfamily II DNA or RNA helicase protein families) and to *yeeC* (T5 orf172-domains containing protein/GIY-YIG nuclease protein families) (Figure 3.6C). In addition, these protein families were found at the 1st and 2nd position downstream of *dnmA*, respectively, indicating that the operon structure in these organisms is the same as the gene organization found in *B. subtilis* (Figure 3.6C). The Uncharacterized Protein Family, which likely represents multiple protein functions, is found throughout the neighborhood upstream or downstream of *dnmA*. This family could represent another member of the *dnmA-yeeB-yeeC* locus in some bacteria, however these genes are uncharacterized, with no known function, making their level of functional conservation unclear.

3.2.5 The DnmA Recognition Motif is Found in Bacteriophage Genomes

The fact that YeeB and YeeC co-occur with DnmA in a conserved cluster and that YeeB and YeeC have putative anti-phage activities suggests that the *dnmA-yeeB-yeeC* gene cluster functions as a restriction modification system. One anti-restriction strategy by bacteriophage is the avoidance of a given restriction site within their genome, a phenomenon often observed for Type II RM systems composed of one MTase and one REase [149]. If the *dnmA-yeeB-yeeC* gene cluster functions as an RM system, then one prediction is that the DnmA recognition motif would be under-enriched in bacteriophage genome sequences. We tested this hypothesis by comparing the observed number of recognition motifs to the expected number of recognition motifs in a sample of bacteriophage genomes, using observed-expected(O/E) ratios of 0.72 and 1.30 as thresholds for under and over-enrichment, respectively [149]. As a control, we measured the O.E. ratio of the recognition

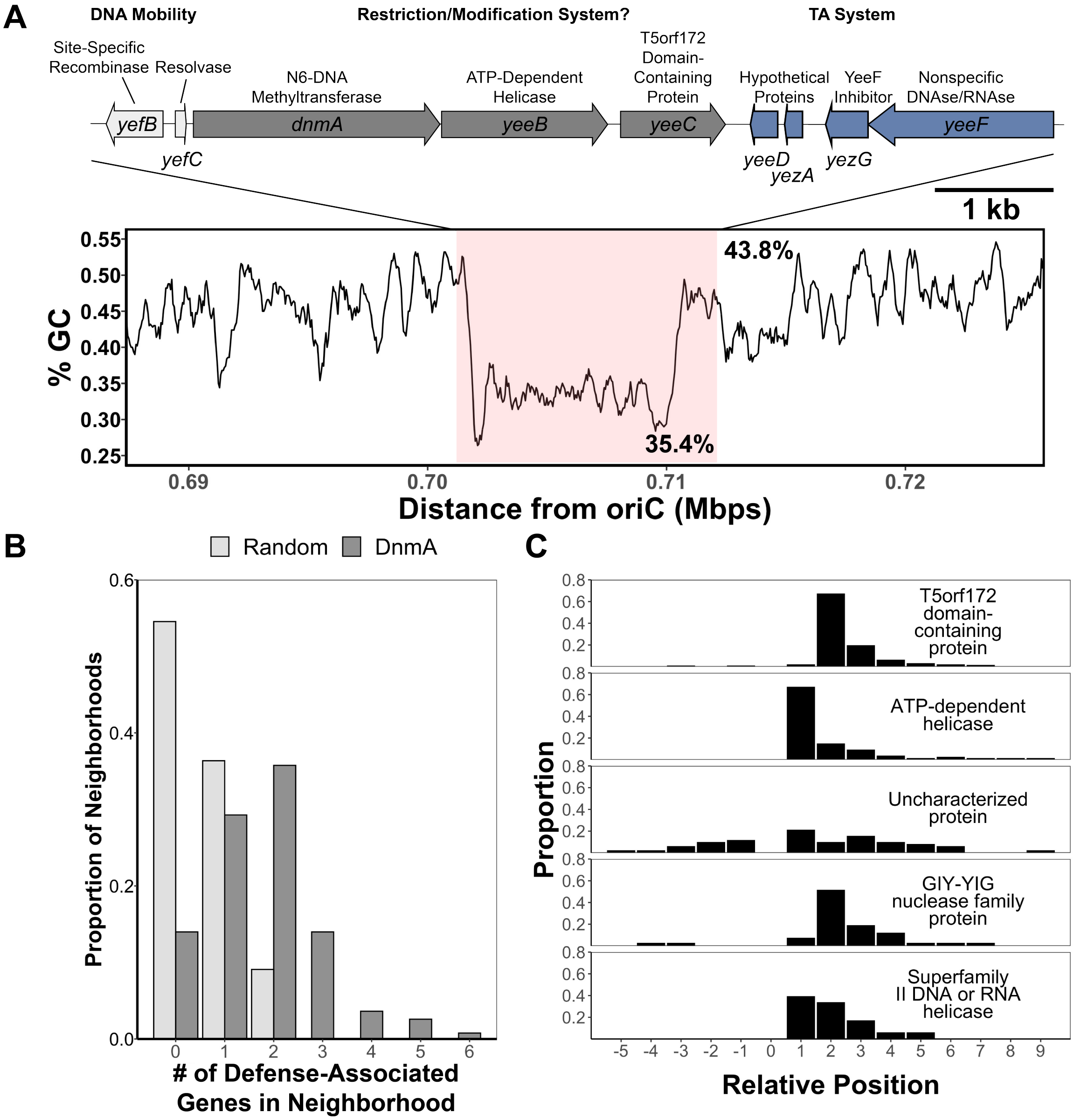


Figure 3.6: (A) (Top) Genome architecture of the locus surrounding *dnmA* in *B. subtilis* PY79. (Bottom) Percent GC content of the *B. subtilis* PY79 genome approximately 15 kb upstream and downstream of the *dnmA* locus. The GC content from *yefB* to *yeeF* is highlighted in pink to emphasize the local minimum. The mean percent GC inside the pink box is 35.4%, and 43.8% is the mean percent GC of the genome. (B) The proportion of genome neighborhoods with a given number of defense-associated protein families. Light gray, the distribution from randomly sampled genomic neighborhoods; dark gray, the distribution from neighborhoods surrounding DnmA homologs. (C) The relative positions of the top five most frequent neighboring defense-associated protein families. 0 indicates the position of *dnmA*, positive integers indicate positions downstream (3′) of *dnmA*, and negative integers indicate positions upstream (5′) of *dnmA*.

sequence for the Type II 5-methylcytosine MTase BsuMM (5′-CTCGAG-3′), which is part of an active Type II RM system found in *B. subtilis* PY79 [150]. In genomes with at least 5 expected motifs, the BsuMM motif has a mean O/E ratio of 0.43 (Figure 3.7). Furthermore, 62.2% of the analyzed genomes have an O/E ratio below the threshold of 0.72, indicating the BsuMM motif is under-enriched in bacteriophage genomes. We repeated the same analysis with the DnmA recognition motif 5′-GACGAG-3′ and a mock recognition motif with the same GC content as the DnmA recognition motif (5′-CTGCTC-3′). In contrast to the BsuMM motif, the DnmA and mock DnmA motifs have O/E ratios of 0.97 and 0.99, respectively. Additionally, they have a lower percentage of genomes with an O/E ratio below the 0.72 threshold (DnmA motif 6.0% and mock DnmA motif 2.6%, Figure 3.7). Together, these data demonstrate that the DnmA motif is naïve to the selective pressure observed with the BsuMM motif from an active RM system. Thus, if the DnmA-YeeB-YeeC gene cluster acts to restrict phage infection or amplification, the mechanism must be distinct from canonical RM systems such as BsuMM-BsuMR (27).

3.2.6 The *dnmA-yeeB-yeeC* Locus Does Not Influence *B. subtilis* Susceptibility to Bacillusphage Nf, Bacillusphage SBS-ΦJ, or Bacillus Virus $\Phi 29$

Though the DnmA recognition site in bacteriophage genomes is not under-enriched, the conservation of both gene arrangement and orientation suggests there is a selective advantage to maintaining *dnmA*, *yeeB*, and *yeeC*, such as limiting bacteriophage infection. We created single gene deletions to directly test the hypothesis that lack of *yeeB* and *yeeC* will result in increased susceptibility to phage infection. Phage were chosen based on the enrichment and total number of DnmA sites within their respective genomes including: Bacillusphage Nf (0 sites), Bacillus virus Φ29 (3 sites, under-enriched), and Bacillusphage $SBS - \Phi J$ (44 sites, no enrichment). In the absence of phage, all strains grew similarly, indicating single-gene deletions of *yeeB* and *yeeC* are not deleterious for growth (Figure 3.8A). Regardless of strain, phage addition at T_0 caused clearing of the culture within two

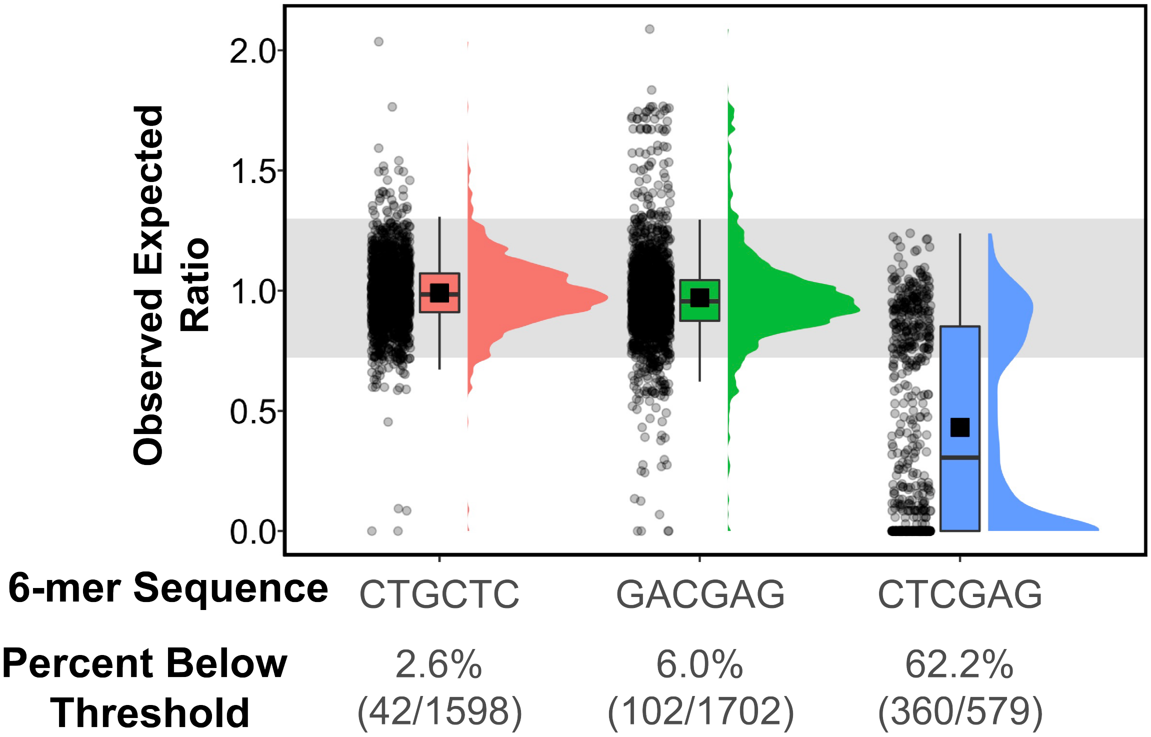


Figure 3.7: Genomes of bacteriophages from the Herelleviridae, Siphovirdae, Myoviridae, and Podoviridae families were analyzed for the presence of the DnmA motif (5′-GACGAG-3′), a mock DnmA motif (5′ − CTGCTC − 3′), and the active RM MTase BsuMM (5′-CTCGAG-3′). Each observed number of motifs was normalized to the expected number of motifs as determined by the compositional bias method [149]. Points above or below the shaded region were considered over- or under-enriched (see 3.4). The number of genomes with an O/E ratio below the threshold of 0.72 (x-axis, numbers Below Threshold") are provided with percentages in parentheses. Each point represents an individual data point, boxplots represent the interquartile range of the data, and the density plots represent the distributions of the data values. Data plotted are for genomes containing at least 5 expected motifs, thus n = 1,598 (CTGCTC), 1,702 (GACGAG), and 579 (CTCGAG).

hours (Figure 3.8B-D). Single gene deletions did not alter phage production either, as the efficiency of plaquing (EOP) was similar between all strains and phages tested (Figure 3.8E-G). Since $\Delta yeeb$ and $\Delta yeec$ backgrounds had similar susceptibility to phage infection and EOP, these data indicate that the *dnmA-yeeB-yeeC* gene cluster is dispensable for protection against bacteriophage infection under the conditions tested here. Given that our results show 1) no evidence of under-enrichment of the DnmA site in phage genomes; 2) no difference in phage predation when comparing deletions of *dnmA*, *yeeB*, and *yeeC* to WT; and 3) no effect of *dnmA* on DNA uptake during natural transformation [63], we suggest that the *dnmA-yeeB-yeeC* cluster does not function as an RM or antiphage system. Instead, we suggest that DnmA is functionally an orphan MTase from a nonfunctional relic of an RM system.

3.3 Discussion

Genes encoding RM systems are found in many bacterial species, yet the functionality of most of these systems remains unknown (http://rebase.neb.com/rebase/) [134]. In *B. subtilis*, the DNA methyltransferase DnmA was previously identified and characterized as a MTase that controls gene expression [63]. Here, we explore how substrate composition and key amino acid residues in DnmA influences kinetics and to expand the biological role of DnmA, we used single-molecule, bioinformatic, and genetic approaches to study DnmA function and regulation of dynamic movement and localization. Our *in vitro* and *in vivo* analyses of DnmA show that disrupting DNA binding either by manipulating DNA substrate availability or DNA binding residues influence DnmA-DNA interactions *in vitro* and DnmA localization *in vivo*. We show that DnmA is coincident with two genes with putative restriction functions, however, our data support the conclusion that DnmA does not participate as an anti-phage system in *B. subtilis*.

We characterized the mobility and localization of DnmA *in vivo* through single - molecule tracking analyses, one of a handful of studies utilizing this technology to bet-

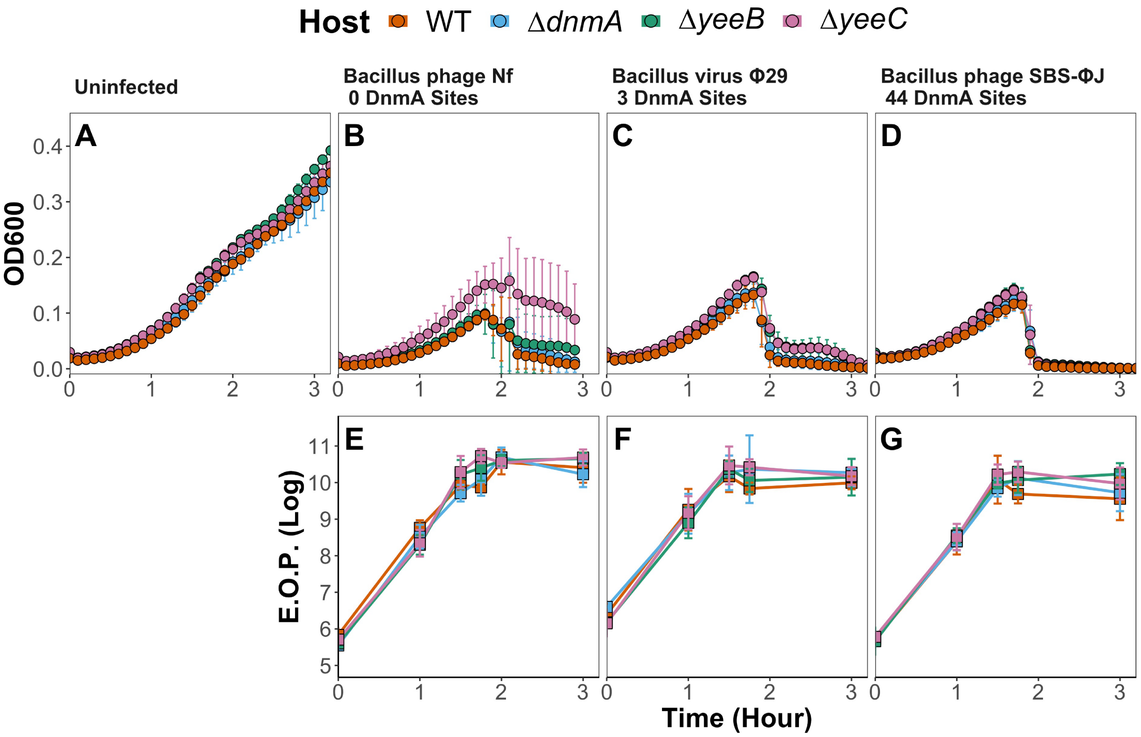


Figure 3.8: (A to D) Growth curves of uninfected *B. subtilis* (A) and cells infected with: (B) Bacillus phage Nf, (C) Bacillus virus Φ29 , and (D) Bacillus phage *SBS* − ΦJ . *B. subtilis* strains are differentiated by color. Cultures were pregrown, and phage addition (MOI of 0.1) occurred at time 0. Growth was monitored by Optical Density (OD) measurements every 5 min for 3 h. Each point is the mean of 4 to 6 biological replicates, and error bars indicate standard deviation. (E to G) Separately, EOP was monitored over the same timescale after phage addition. Squares represent the mean EOP value, and error bars denotes the standard deviaton.

ter understand prokaryotic DNA methylation *in vivo* [151]. In unperturbed cells, DnmA is found throughout the center of the cell, likely interacting with the nucleoid and has a positive correlation with the position of the replisome (Figure 3.2). Negri and colleagues analyzed the mobility and localization of the DNA MTase M.Csp231I, which functions in an active RM system [151]. Similar to our findings, M.Csp231I localizes throughout the nucleoid with a high probability of localizing near the mid and quarter cell positions, suggesting a common DNA searching mechanism among DNA MTases in bacteria [151].

Single-molecule studies of DNA-binding enzymes in *E. coli* have suggested that the slower-moving enzyme molecules are involved in catalytic functions [152]. However, due to the essential nature of the enzymes, catalytically inactive versions were not studied. Here, the use of the inactive DnmA variant DnmA[Y465A] allowed us to assess how catalysis influences DNA mobility and localization *in vivo*. Interestingly, DnmA[Y465A] and WT DnmA have similar percentages of slow-moving molecules, indicating that the slower moving molecules are not necessarily enzymes involved in active catalysis (Figure 3.1). Additionally, although we reasoned that disrupting DNA interactions in DnmA, either by amino acid substitution or manipulating available substrate pools, would result in a larger population of fast-moving molecules compared to WT DnmA, we instead found that mobility remains largely unchanged compared to WT DnmA in unperturbed conditions except for upon HPUra treatment. Thus, our results highlight the importance of targeted amino acid substitutions and other approaches to better explain single molecule results of catalytic enzymes.

Under high R-loop conditions (Δ*rnhC*), the localization pattern of DnmA switches from a concentration near the mid-cell and co-localization with the replisome to DnmA being negatively correlated with the replisome position (Figure 3.2). Since DnmA binds RNA-DNA hybrids poorly *in vitro* (Figure 3.3) and DNA binding is necessary for DNA methylation [140], our data suggest specific DNA interactions are necessary for proper DnmA localization and correlation with DnaX. This conclusion is supported by the even stronger negative DnaX correlation in DnmA[6AA*], which lacks the ability to recognize the DnmA recognition sequence *in vitro* (Figure 3.3). While the mechanism for DnmA[6AA*] repositioning is not clear, protein sequestration and localization are used in bacteria to regulate enzymatic activity. In C. crescentus, the cell cycle regulating DNA MTase CcrM is inhibited by polar sequestration [153]. While we do not observe strict polar DnmA localization under R-loop stress or in DnmA[6AA*], it is tempting to speculate that MTase repositioning in the cell represents a broad mechanism to negatively regulate DNA methylation and epigenetic gene expression in bacteria.

Morgan et al. found that *B. subtilis* DnmA (previously YeeA) is homologous to the Type IIL MTase-REase protein MmeI, which has MTase and REase domains in a single polypeptide [144]. The authors noted that DnmA did not encode a REase motif but was adjacent to YeeB and YeeC homologs. We expanded on this finding to include DnmA homologs from various species and found that genomes encoding DnmA likely encode two genes with helicase and nuclease functions (putative YeeB and YeeC homologs, respectively) within a 20-gene neighborhood, demonstrating that gene synteny and architecture are conserved (Figure 3.6B). The putative recombinase genes *yefB* and *yefC* and the toxin-antitoxin pair *yeeF* and *yezG*, however, are not adjacent to DnmA at a high enough frequency for identification in our analysis. This result suggests that these genes represent *B. subtilis*-specific gene acquisitions. In *S. pneumoniae*, genes encoding the MTase specificity subunits, which direct the MTase to a given sequence, are subject to phase-variation through recombination, resulting in heterogenous methylation patterns in the genome [154]. Thus, it is possible the adjacent recombinase genes may play a similar role in *B. subtilis.* In our previous characterization of DnmA, however, we observed homogenous methylation patterns under standard growth conditions [63]. Additionally, we did not identify any sequence signatures suggestive of site-specific recombination flanking the low-GC region in the genome, such as inverted or direct repeats. The yeeF gene has an N-terminal LXG domain which allows for secretion through the Type 7 Secretion System (T7SS) encoded by the distally located genes yukEDCB-yueBCD [155]. The C-terminal domain of YeeF encodes nuclease activity that is inactivated by the neighboring antitoxin YezG [156]. Thus, our data suggests that this region represents a defunct mobile genetic element that is maintained through a selective benefit of DnmA and/or the antitoxin YezG.

The conservation and putative functions of *yeeB* and *yeeC* suggest a conserved function. We assessed the anti-phage activity of DnmA, YeeB, and YeeC by testing whether single-deletion mutants had any effect on host survival and/or bacteriophage amplification. Despite using bacteriophages with a range of DnmA motifs in their genome, the single-deletion mutants had no effect on bacteriophage-mediated host killing or production ((Figure 3.8), leading us to conclude that DnmA is part of a remnant of a non-functional RM system. This observation is important because of the pervasive occurrence of MTases and DNA methylation in the domain Bacteria [134].

DnmA-YeeB-YeeC homologs in the marine microorganism *Vibrio crassostreae* were identified in a recent study [157]. Deletion mutations of *dnmA* and *yeeB* caused an increase in bacteriophage sensitivity to some subclades of bacteriophage, while having no effect when other subclades were used [157]. An amino acid alignment of DnmA, YeeB, and YeeC from *B. subtilis* and V. crassostreae shows that all three proteins share high sequence homology in putative active site domains (Figure 3.9). However, YeeB and YeeC from *B. subtilis* are missing several amino acid-long stretches in the C-terminal domain. Therefore, it is possible that YeeB and YeeC in *B. subtilis* are missing critical residues necessary for antiphage function. Our data suggest that numerous bacterial MTases detected in the bacterial methylome also originate from defunct phage defense systems, similar to *dnmA* in *B. subtilis* [134]. These defunct defense systems could have maintained an active MTase either for epigenetic control or due to the presence of a toxin-antitoxin system that selects for the acquired region while losing restriction activity.

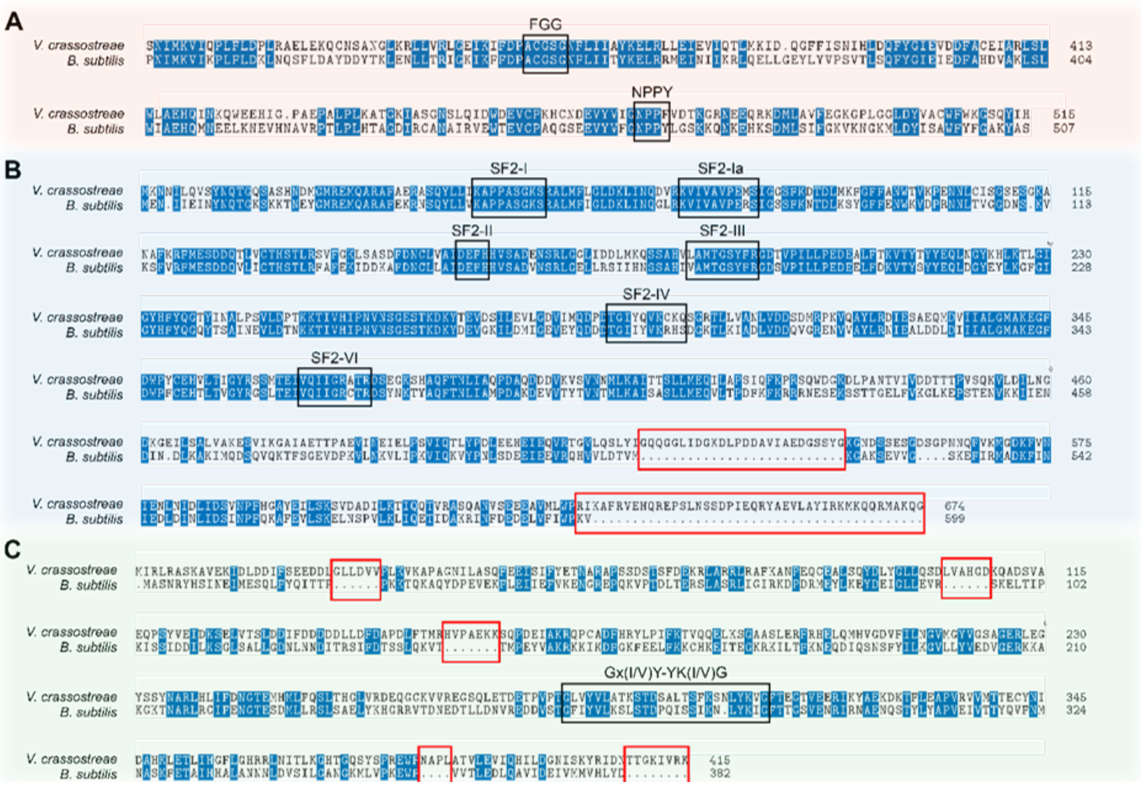


Figure 3.9: Identical residues for each alignment are in blue. Amino acids missing in *B. subtilis* homologs are highlighted with red boxes. Motifs involved in protein function are highlighted with black boxes with the motif name printer above the box. (A) Alignment of DnmA amino acid sequences from 298-515 for V. crassostreae and 289-507 for *B. subtilis*. Boxes highlight the residues found in the two conserved active sites FGG and NPPY. DnmAVcr is 57% similar to DnmABsu (B) Alignment of amino acid sequences of YeeB homologs. Boxes highlight conserved motifs found in Domain 1 and Domain 2, respectively, of superfamily 2 (SF2) helicases. YeeBVcr is 56.76% similar to YeeBBsu. (C) Alignment of amino acid sequences of YeeC homologs. The black box highlights conserved residues found in GIY-YIG nuclease family protein/T5orf172-domain containing proteins. YeeCVcr is 32.96% similar to YeeCBsu.

3.4 Materials and Methods

3.4.1 Cloning and Strain Construction

The plasmids and strains used in this Chapter are from the Simmons Lab at the University of Michigan. Please refer to corresponding manuscript of this chapter for details [158].

3.4.2 Electrophoretic Mobility Shift Assay

Production of m6A in oNLF001 was carried out by Integrated DNA Technologies (IDT) and was determined to be 98% pure by electrospray ionization mass spectrometry (IDT). For annealing, solutions of the unmethylated probe (oligos oTMN67/oTMN68), methylated probe (oNLF001/oTMN68), and the RNA-DNA hybrid (oTMN67/oJRR271) were mixed at a final concentration of 50 nM and incubated at room temperature overnight, covered from light. Purified ScoC was mixed in a binding reaction consisting of 5X EMSA reaction buffer (500 mM Tris-HCl pH 8, 1.25 M NaCl, 10% glycerol v/v) and 5 nM (final concentration) annealed oligos. Reactions were incubated for 30 minutes at 25 $°C$. Afterwards, 8 μL of the mixture were loaded onto and resolved via pre-run 6% Native-PAGE which was performed covered from light and on ice for 60 minutes at 100V in 1X TBE. The samples were visualized with the LI-COR Odyssey imager. Intensities of the shifted and unshifted bands were quantified using Fiji image software using the Gels feature [159]. The fraction bound was calculated by first subtracting the background signal (region of gel with no band) from the intensity measurement of each band. The intensity of the bound substrate was divided by the sum of intensities of the bound and unbound substrate, yielding the fraction bound. Fraction bound data was modeled using the four-parameter log-logistic function in the drc package for R and the effective concentration for half maximal binding (EC50) was measured for each replicate [160].

3.4.3 Flow Cytometry

Strains of interest were struck out on LB agar plates and incubated 16 hours overnight at 30°C. The next day, 6 isolated colonies were inoculated in 250 µL LB in wells of a 96-well plate and grown at 37°C in an orbital shaker at 250 RPM until early exponential phase. Cultures were then moved to microcentrifuge tubes and diluted 1:1 with 200 µL sterile 1X PBS and single cell fluorescence was measured using an AttuneTM NxT Acoustic Focusing Cytometer (ThermoFisher Scientific). Fluorescence data was acquired from 200,000 cells with the following settings: flow rate, 25 µl/min; FSC voltage, 200; SSC voltage, 250; BL1 voltage, 250.

3.4.4 Live-Cell Single-Molecule Imaging

B. subtilis strains expressing DnmA-PAmCherry (PY79 and Δ*rnhC* PY79) and DnmA variants (DnmA[Y465A]-PAmCherry and DnmA-6AA*-PAmCherry) were grown overnight on LB-agar plates at 37 °C. The cells were washed from the plate with filtered S750 minimal media and inoculated in filtered S750 minimal media at OD600 0.1, followed by growth with shaking at 200 RPM at 30 °C for 4 h until reaching OD600 0.5 - 0.6 (S750 minimal media - 1X S750 salts [10X S750 salts: 0.5 M MOPS, pH 7.4, 100 mM Ammonium Sulfate, 50 mM Potassium Phosphate Monobasic, filter sterilzed], 1X Metals [100X metals: 0.2 M $MgCl_2$, 70 mM $CaCl_2$, 5 mM $MnCl_2$, 0.1 mM $ZnCl_2$, 100 µg/mL Thiamine HCl, 2 mM HCl, 0.5 mM FeCl3 (added last to prevent precipitation), filter sterilized)], 1% glucose, 0.1% glutamate, 40 µg/mL tryptophan, 40 µg/mL phenylalanine). Experiments in 6-(p-hydroxyphenyIazo)-uracil (HPUra) were done by adding HPUra at a final concentration of 162 µM to the culture immediately before imaging. Coverslips were cleaned via argon plasma etching (PE-50, Plasma Etch) for 30 minutes and 2% agarose pads were prepared with freshly made, filtered S750 media to reduce background fluorescent signals. Cells were pipetted onto agarose pads and sandwiched between coverslips for imaging. Once prepared, the sample was mounted on a wide-field inverted microscope (Olympus

IX71, Melville, NY) for single-molecule imaging.

Prior to imaging, the cells and background were photobleached with a 561-nm laser (Sapphire 561-50, Coherent, Bloomfield, CT) for two minutes at a power density of 630 W/cm2. Single DnmA-PAmCherry molecules were photoactivated with 400-ms pulses of 405-nm laser (Cube 405-100, Coherent, Bloomfield, CT) with a power density of 21.6 W/cm2 at the start of the imaging and after photo-bleaching. Photo-activated DnmA-PAmCherry molecules were imaged with a 561-nm laser with a power density of 69.2 W/cm2 and appropriate dichroic and long-pass filters. Fluorescence was collected via a 1.40 NA 100× oil-immersion phase-contrast objective and detected with a 512 × 512 pixel electron multiplying charge-coupled device camera (Photometrics, Acton, MA). Images were recorded with 40-ms exposure time.

3.4.5 Single-Molecule Detection, Tracking, and Analysis

Phase-contrast images were used to provide a reference mask for single-molecule detection and fitting within cell boundaries. Single-molecule fitting was done via the Single-Molecule Localization by Local Background Subtraction (SMALL-LABS) algorithm [12]. The fit positions were connected into trajectories using the Hungarian algorithm [65].

The diffusion coefficients for each trajectories is fitted through [100]: $MSD = 4D\tau + 2\sigma^2$ Where MSD is the squared displacement. τ is the time lag and σ is the localization precision. The normalized heatmaps in Figure 3.2E-J include the positions of all single molecules in all cells under each condition. First, the cell outlines were determined from segmentation of the phase contrast images, then the Feret properties of each cell were calculated (Matlab function bwferet) to determine the long and short axis of each cell. The single-molecule localizations of DnmA in each cell were projected onto the corresponding cell's long and short axes to acquire the relative position of that molecule in the cell. Based on assuming the cells are symmetric along long and short axes, the 2D relative position of each single-molecule were symmetrized along the two axes.

The curve fitting for the histogram of diffusion coefficients in Figure 3.2D and Figure 3.4A depict the single-component Gaussian fitting and the logarithm of single-trajectory diffusion coefficients are weighted based on track length. Figure 3.4B-F depicts the 2-component Gaussian fitting of the logarithm of single-trajectory diffusion coefficients in Figure 3.4A. The Spot-On algorithm was applied to fit the probability density function of single-molecule displacements to a 2-state model and a 3-state model to get the weight fraction of each component for WT DnmA [18]. For Spot-On analysis of the other datasets, the fitted diffusion coefficient range for each state is fixed within the confidence interval of the corresponding state's WT DnmA diffusion coefficient value to enable a direct comparison of weight fraction of each state between different datasets.

3.4.6 Gene Neighborhood Analysis, Phage Propagation and Infection

The genetic anlysis and phage assays used in this Chapter are from the Simmons Lab at the University of Michigan. Please refer to corresponding manuscript of this chapter for details [158].

CHAPTER IV

HP1 Oligomerization Compensates for Low-Affinity H3K9me Recognition and Provides a Tunable Mechanism for Heterochromatin-Specific Localization

The work presented in this chapter was previously published in
Science Advances.

Biswas, S.*, Chen, Z.*, Karslake, J.D.*, Farhat, A.*, Ames, A., Raiymbek, G.,
Freddolino, P.L., Biteen, J.S. and Ragunathan, K.
HP1 oligomerization compensates for low-affinity H3K9me recognition and provides a
tunable mechanism for heterochromatin-specific localization.
Science Advances, **8**(27), eabk0793 (2022). DOI: 10.1126/sciadv.abk0793

In this work, I participated in the conceptualization of the project. I performed the single-molecule imaging experiments and dynamics analysis together with Saikat Biswas and Josh Karslake. I designed and implemented the spatial analysis, plotted the analysis of the single-molecule dynamics, and validated the dynamics detection range. *: equal contribution.

4.1 Introduction

Despite having identical genomes, eukaryotic cells can establish distinct phenotypic states that remain stable and heritable throughout their lifetimes [161]. In the context of a multicellular organism, the persistence of epigenetic states is vital to establish and maintain distinct cellular lineages [162]. This process of phenotypic diversification depends, in part, on the posttranslational modifications of DNA packaging proteins called histones [163,164]. Proteins that can "read, write, and erase" histone modifications interact weakly and transiently with their histone substrates. Nevertheless, dynamic, low-affinity interactions between histone modifiers and their substrates can encode stable memories of gene expression that can be inherited following DNA replication and cell division [165].

H3K9me is a conserved epigenetic modification that is associated with transcriptional silencing and heterochromatin formation [54]. Heterochromatin establishment is critical for chromosome segregation, sister chromatid cohesion, transposon silencing, and maintaining lineage-specific patterns of gene expression [67]. These diverse cellular functions associated with heterochromatin are critically dependent on an evolutionarily conserved HP1 family of proteins that recognize and bind to H3K9me nucleosomes [73]. HP1 proteins have a distinct architecture that consists of two conserved structural domains: (i) an N-terminal chromodomain (CD) that recognizes H3K9me nucleosomes and (ii) a C-terminal chromoshadow domain (CSD) that promotes dimerization (Figure 4.1A) [166]. The HP1 CD domain binds to H3K9me peptides with low micromolar affinity (1 to 10 μM) [167,168]. The HP1 CSD domain promotes protein-protein and protein-nucleosome interactions and oligomerizes to form higher-order, phase-separated HP1-containing chromatin complexes that exhibit liquid-like properties [79, 169, 170]. In addition, HP1 proteins are posttranslationally modified, and these modifications affect both nucleosome binding and HP1 protein–protein interactions [171,172]. A flexible and unstructured hinge region connects the Swi6 CD and CSD domains. The hinge region binds to nucleic acids without any sequence specificity [76]. At present, it is unclear how the competing demands of hinge-mediated nucleic acid binding, CD-dependent H3K9me recognition, and CSD-mediated oligomerization and nucleosome interactions influence HP1 enrichment at specific locations in the genome.

In fission yeast *Schizosaccharomyces pombe (S. pombe)*, Histone H3 lysine 9 methylation (H3K9me) is enriched at sites of constitutive heterochromatin, which includes the pericentromeric repeat sequences (dg and dh), the telomeres (tlh), and the mating-type locus (mat) [71]. A conserved SET domain–containing methyltransferase, Clr4, is the sole enzyme that catalyzes H3K9me in *S. pombe* [72, 173]. The major *S. pombe* HP1 homolog, Swi6, senses the resulting epigenetic landscape and binds to H3K9me chromatin with low affinity but high specificity [174]. Swi6 is an archetypal member of the HP1 family of proteins [74]. Its ability to simultaneously recognize H3K9me and oligomerize enables linear spreading across broad segments of the chromosome encompassing several hundred kilobases of DNA [175]. Heterochromatin spreading subsequently leads to the silencing of genes that are distal from heterochromatin nucleation centers.

Only 2% of nucleosomes in the *S. pombe* genome are marked with H3K9me [176]. Given the limiting amount of substrate, most Swi6 molecules (80 to 90%) are located elsewhere in the genome, potentially engaged in promiscuous DNA- and chromatin-dependent interactions [177, 178]. Under conditions of acute heterochromatin misregulation, these promiscuous Swi6-chromatin interactions lead to epimutations that alter cellular fitness [179]. Overexpressing Swi6 enhances epigenetic silencing of a reporter gene at sites of constitutive heterochromatin [180]. Therefore, Swi6 functions as a dose-sensitive heterochromatin-associated protein. Altering the fractional occupancy of Swi6 at sites of H3K9me is likely to have a profound impact on transcriptional silencing, heterochromatin stability, and epigenetic inheritance.

On the basis of fluorescence recovery after photobleaching (FRAP) measurements, the turnover rates of Swi6 and other HP1 homologs from sites of heterochromatin range from a few hundred milliseconds to several seconds [181, 182]. Point mutations that impair CD binding to H3K9me nucleosomes or CSD-mediated dimerization abolish Swi6 binding to nucleosomes. Hence, the rapid turnover of HP1 proteins from heterochromatin involves both (i) CD-dependent binding and unbinding of Swi6 from H3K9me nucleosomes and (ii) CSD-dependent association or dissociation of Swi6 oligomers. CSD-dependent oligomerization drives the formation of heterochromatin condensates, but we do not understand whether such interactions inhibit or enhance H3K9me recognition *in vivo*. In addition, reconstitution studies show that an increase in the on-rate for DNA binding enhances the interaction between HP1 proteins and nucleosomes [75]. However, given a genome that is replete with nucleic acids (DNA and RNA), it is equally likely that the on-rate associated with nucleic acid binding prevents specific Swi6 binding to H3K9me and could titrate Swi6 away from sites of H3K9me. In essence, we lack a fundamental understanding of the extent of coupling between CSD-dependent oligomerization, CD-dependent H3K9me recognition, and hinge-dependent nucleic acid binding in the context of the fission yeast nucleus. In this study, we use single-particle tracking photoactivated localization microscopy to measure the *in vivo* binding dynamics of Swi6 in real time as it samples the fission yeast nucleus [1, 11]. *In vitro* binding measurements have thus far served as the gold standard to measure interactions between chromatin readers and modified histone peptides or recombinant nucleosome substrates [73, 74, 79, 82, 183–185]. However, these studies are typically carried out under dilute, noncompetitive solution conditions, which do not reflect the chromatin environment that Swi6 encounters in the nucleus. Rather, here, we analyze individual Swi6 molecule trajectories with high spatial and temporal resolution in living cells. Our studies determined the precise biochemical attributes of Swi6 that give rise to distinct mobility states. Our measurements enabled us to engineer precise degrees of multivalency within Swi6 that entirely circumvent the need for CSD-dependent oligomerization while suppressing the inhibitory effects of nucleic acid binding. We find that the simultaneous engagement of at least four H3K9me CD domains is both necessary and sufficient for the heterochromatin-specific targeting of Swi6, while nucleic acid binding competes with oligomerization. Overall, our results demonstrate that the evolutionarily conserved phenomenon of HP1 oligomerization may represent a tunable mechanism that compensates for weak, low-affinity H3K9me recognition that outcompetes promiscuous nucleic acid binding.

4.2 Results

4.2.1 The single-molecule dynamics of Swi6 in live *S. pombe* indicate a heterogeneous environment

Because Swi6 molecules within the fission yeast nucleus undergo binding and unbinding events in a complex environment, we used single–molecule tracking to measure their heterogeneous dynamics. For instance, a Swi6 molecule bound to an H3K9me nucleosome is, on average, likely to exhibit slower, more confined motion compared to rapidly diffusing proteins. Hence, the biochemical properties of Swi6 will directly influence its mobility within the fission yeast nucleus. We transformed *S. pombe* cells with PAmCherry fused to the N terminus of Swi6 [3]. This fusion protein replaces the WT endogenous Swi6 gene and serves as the sole source of Swi6 protein in fission yeast cells (Fig. 4.1B). To test the functionality of PAmCherry–Swi6 in heterochromatin assembly, we used strains where a *ura4+* reporter is inserted within the outermost pericentromeric repeats (*otr1R*) (39). We concluded that PAmCherry–Swi6 is functional since cells expressing the fusion protein exhibit reduced growth on -URA medium consistent with *ura4+* silencing (Fig. 4.1B). In contrast, cells lacking the H3K9 methyltransferase Clr4 ($clr4\Delta$) exhibit no growth inhibition when plated on –URA–containing medium, consistent with $ura4$+ expression and the loss of H3K9me-dependent epigenetic silencing (Fig. 4.1B). In addition, chromatin immunoprecipitation–quantitative polymerase chain reaction (ChIP-qPCR) measurements show that H3K9me2 and Swi6 binding to sites of constitutive heterochromatin is preserved in cells expressing PAmCherry-Swi6 (fig. 4.2, A and B).

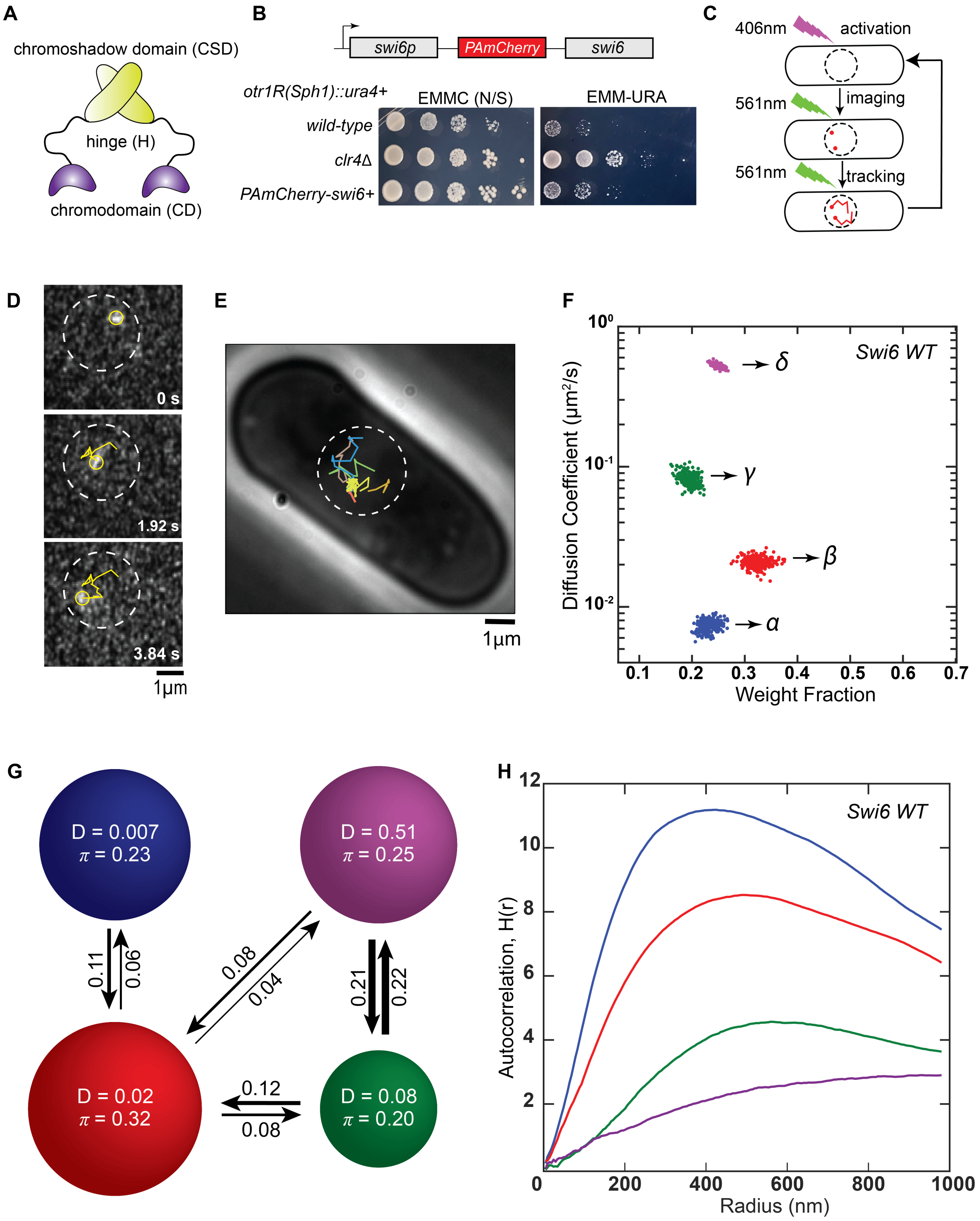


Figure 4.1: (A) Each Swi6 domain has a distinct biochemical role. CD: H3K9me recognition; H: nucleic acid binding; CSD: protein oligomerization. **(B)** Top: pamCherry-swi6 expressed from the endogenous *swi6+* promoter. Bottom: Silencing assay using a *ura4+* reporter inserted at the pericentromeric repeats (*otr1R*). Caption continues

Figure 4.1: (C) Single-molecule experiment workflow. PAmCherry-Swi6 molecules are photoactivated (406 nm, 1.50 to $4.50W/cm^2$, 200 ms), then imaged, and tracked until photobleaching (561 nm, $71.00W/cm^2$, 25 frames/s). The cycle is repeated 10 to 20 times per cell. (D) Image of a single photoactivated PAmCherry-Swi6 molecule at three time points. Yellow circle, molecule position; yellow line, Swi6-PAmCherry trajectory from photoactivation until the current frame; white circle, approximate nucleus position. **(E)** Representative single-molecule trajectories in a live *S. pombe* cell. Each trajectory is acquired after a new photoactivation cycle. **(F)** SMAUG identifies four distinct mobility states, (α, β, γ, and δ) for PAmCherry-Swi6 in WT cells. Each point is the average single-molecule diffusion coefficient, D, of Swi6 molecules in that state at a saved iteration of the Bayesian algorithm after convergence. Dataset: 10,095 steps from 1491 trajectories. **(G)** Average probabilities (arrows) of a PAmCherry-Swi6 molecule transitioning between the mobility states (circles) from (F). Each circle area is proportional to the weight fraction, π; D is in square micrometers per second. Low-frequency transition probabilities (below 0.04) are excluded. **(H)** Ripley's analysis shows higher autocorrelation for PAmCherry-Swi6 in the slower (blue and red) states compared to $H(r) \leq 2$ for the faster (green and purple) states.

We used single-molecule microscopy to investigate the dynamics of Swi6 molecules with high spatial (20 to 40 nm) and temporal (40 ms) resolution. We photoactivated zero to two PAmCherry-Swi6 fusion proteins per activation with 406-nm light. Next, we imaged the photoactivated Swi6 molecules with 561-nm laser excitation until photobleaching, obtained a 5- to 15-step trajectory based on localizing molecules at 40-ms intervals, and then repeated the photoactivation/imaging cycle with another PAmCherry-Swi6 molecule 10 times per cell across several individual cells (Fig. 4.1, C and D). Our experimentally measured localization accuracy is about 36 nm (fig. 4.2C). We observed both stationary and fast-moving molecules (Fig. 4.1E). We hypothesized that each type of motion, which we term a "mobility state," corresponds to a distinct biochemical property of Swi6 in the cell (e.g., bound versus unbound) and that molecules can transition between the different mobility states during a single trajectory. Thus, rather than assign a single diffusion coefficient to each single-molecule trajectory, we analyzed our data using single-molecule analysis by unsupervised Gibbs (SMAUG) sampling algorithm (4.4) [31]. SMAUG estimates the biophysical descriptors of a system by embedding a Gibbs sampler in a Markov Chain Monte Carlo framework. This nonparametric Bayesian analysis approach deter-

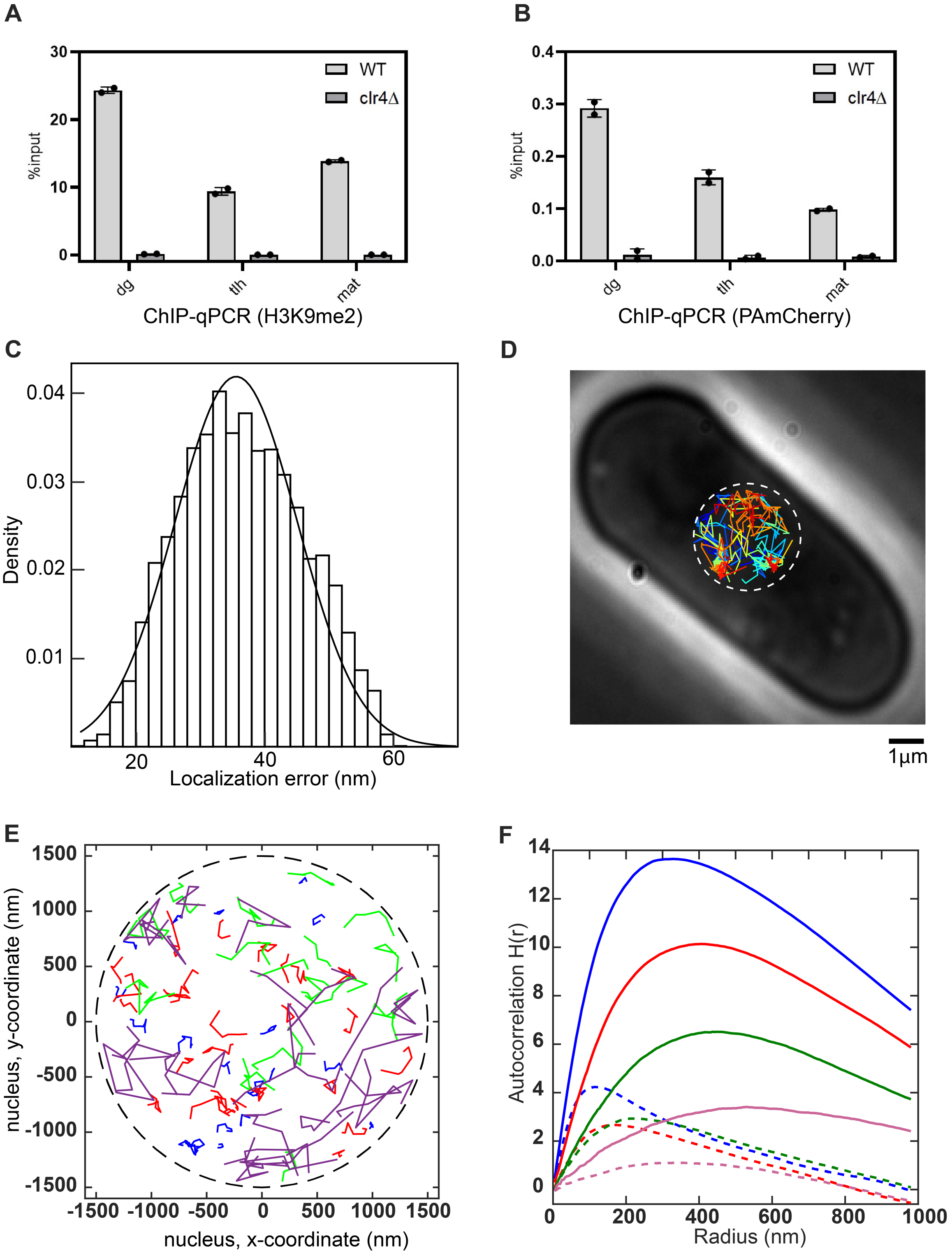

A
30
20
10
0
%input
WT
clr4Δ
dg
tlh
mat
ChIP-qPCR (H3K9me2)
B
0.4
0.3
0.2
0.1
0.0
%input
WT
clr4Δ
dg
tlh
mat
ChIP-qPCR (PAmCherry)
C
0.04
0.03
0.02
0.01
Density
20
40
60
Localization error (nm)
D
1μm
E
1500
1000
500
0
-500
-1000
-1500
nucleus, y-coordinate (nm)
-1500 -1000 -500 0 500 1000 1500
nucleus, x-coordinate (nm)
F
14
12
10
8
6
4
2
0
Autocorrelation H(r)
0 200 400 600 800 1000
Radius (nm)

Figure 4.2

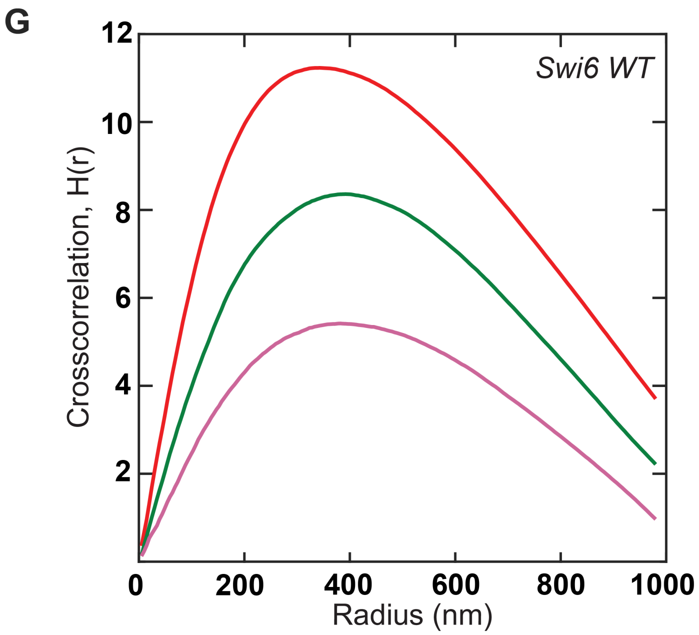


Figure 4.2: (A). ChIP-qPCR measurements of H3K9me2 levels at sites of heterochromatin formation (dg pericentromeric repeats, tlh telomere and mat mating type locus) (N = 2), for Swi6 WT and *clr*4Δ cells. Error bars represent standard deviation. **(B)**. ChIP-qPCR measurements of PAmCherry-Swi6 levels at sites of heterochromatin formation (dg pericentromeric repeats, tlh telomere and mat mating type locus) (N = 2), for Swi6 WT and *clr*4Δ cells. Error bars represent standard deviation. **(C)**. Density histogram of single-molecule localization fit error from 20 individual measurements and the plotted curve is the t-distribution fit. **(D)**. A collection of thirty-five single-particle trajectories in live *S. pombe* cells. Colors represent different single-particle trajectories that are acquired following sequential photoactivation cycles. Clusters of molecules appear at sites that correspond to constitutive heterochromatin. The heterogeneous tracks correspond to Swi6 molecules that exhibit different mobility states. **(E)**. Randomly simulated trajectories for the four wild-type Swi6 mobility states. The nucleus is approximated to be a circle with radius 1.5μm as indicated by the black dashed circle. Blue, red, green and purple simulated trajectories correspond to α, β, γ and δ mobility states of wild-type Swi6. **(F)**. Ripley's H function plot for experimental trajectories and simulated trajectories of wild-type Swi6. Blue, red, green, and purple lines correspond to α, β, γ and δ states, respectively. The solid line corresponds to experimentally acquired trajectories without normalization, and the dashed lines correspond to simulated trajectories shown in B, **(G)**. Ripley's $H(r)$ function plot for cross-correlation between β, γ and δ state in relation to the α state as indicated by the red, green, and purple plots, respectively. Simulated trajectories exhibit negligible cross-correlation values, and therefore the cross-correlation plots shown here do require any additional normalization.

mines the most likely number of mobility states and the average diffusion coefficient of single molecules in each state, the occupancy of each state, and the probability of transitioning between different mobility states between subsequent 40-ms frames. For each strain, we sampled over a thousand trajectories of PAmCherry-Swi6 within the nucleus of *S. pombe* cells (fig. 4.2D). SMAUG analyzed the 10,000 steps from these trajectories in aggregate. For cells expressing fusions of PAmCherry-Swi6, the algorithm converged to four mobility states and estimated the diffusion coefficient, D, and the fraction of molecules in each state, π, for each iteration (each dot in Fig. 4.1F is the assignment from one iteration of the SMAUG analysis). We also compared our results from SMAUG to other single-particle tracking algorithms such as Spot-On and vbSPT (4.4). Our comparisons with different single-molecule tracking methods reveal that for our single-particle tracking datasets, SMAUG and other analysis methods are consistent with respect to the number of states and the estimated diffusion coefficients (D_avg) for each state.

We refer to the four mobility states as α, β, γ, and δ in order of increasing Davg (Fig. 4.1F). The clusters of points correspond to estimates of D and π for each of the four mobility states, and the spread in the clusters indicates the inferential uncertainty. The slowest mobility state comprises 23% of the Swi6 molecules; its average diffusion coefficient ($D_{avg\alpha} = 0.007 \pm 0.001 \mu m^2/s$) is close to the localization precision of the microscope, indicating no measurable motion for these molecules (error bars indicate the 95% credible interval). Given that Swi6 forms discrete foci at sites of constitutive heterochromatin (centromeres, telomeres, and the mating-type locus), the α mobility state likely corresponds to Swi6 molecules that are stably bound at these constitutive sites, which are enriched with H3K9me. The fastest diffusion coefficient that we estimated is $D_{avg\delta} = 0.51 \pm 0.03 \mu m^2/s$. We tested the extent to which our experimental temporal resolution affects the estimation of D_avg and found that using a shorter (20-ms) exposure time results in larger localization errors due to the limited brightness of PAmCherry, although it does not affect the weight fraction of molecules in the fast-diffusing δ state. Because of this reduced signal-to-noise ratio, 40 ms was chosen as an optimal temporal resolution. Although the 40-ms time resolution of our measurements is likely to result in uncertainty and potential underestimation of the D_avg associated with the fast state (δ state), our measurements are appropriate to quantify the slower-diffusing α, β, and γ populations.

We also measured how often Swi6 molecules in one mobility state transition to another mobility state. We found that Swi6 molecules are most likely to transition between adjacent rather than nonadjacent mobility states. The transition probabilities reveal that a distinct hierarchy of biochemical interactions dictates how Swi6 interacts with H3K9me sites in the fission yeast genome (Fig. 4.1G). Notably, only molecules in the β intermediate-mobility state transition with high probability to the slow-mobility α state, which (as noted above) likely corresponds to Swi6 molecules stably bound at sites of H3K9me.

We used a spatial autocorrelation analysis to measure inhomogeneities in Swi6 diffusion (Fig. 4.1H). The Ripley H function, $H(r)$, measures deviations from spatial homogeneity for a set of points and quantifies the correlation as a function of the search radius, r [186]. We calculated $H(r)$ values for the positions of molecules in each mobility state from the single-molecule tracking dataset. To eliminate the bias that would come from the spatial correlation between steps along the same trajectory, we compared the experimental observations to a null model by randomly simulating confined diffusion trajectories for each of the four mobility states (4.4). We represented the spatial autocorrelation values associated with each state to an $H(r)$ function normalized using simulated trajectories (fig. 4.2, E and F). The real trajectories show substantially higher-magnitude and longer distance correlations than a realistic simulated dataset. We also observed low autocorrelation values for the δ and γ states (purple and green) and high autocorrelation values for the β and α states (blue and red) (Fig. 4.1F). The cross-correlation between the β, γ, and δ mobility states and the slow-moving α state further supports our model of spatial confinement: Molecules in the β state are most likely to be proximal to H3K9me-bound molecules in the α state (fig. 4.2G). In summary, the combination of transition plots and spatial homogeneity maps suggests that the β and γ intermediate-mobility states represent biochemical intermediates that sequester Swi6 molecules from transitioning between the H3K9me-bound α state and the fast-diffusing, chromatin-unbound δ state.

4.2.2 The slowest moving Swi6 molecules correspond to foci that are typically observed at sites of constitutive heterochromatin

Following the baseline characterization of the different mobility states associated with Swi6 and their relative patterns of spatial confinement, we used fission yeast mutants to assign individual biochemical properties to each mobility state. As a first approximation, we hypothesized that the major features likely to affect Swi6 binding within the nucleus are (i) CD-dependent H3K9me recognition, (ii) hinge-mediated nucleic acid binding, and (iii) CSD-mediated oligomerization [73].

We deleted the sole *S. pombe* H3K9 methyltransferase, Clr4, to determine how the four mobility states associated with Swi6 diffusion respond to the genome-wide loss of H3K9 di- and trimethylation. The mobility of PAmCherry-Swi6 is substantially different in (H3K9me0) *clr*4Δ cells compared to *clr*+ cells. Most prominently, most Swi6 molecules in *clr*4Δ cells move rapidly and show no subnuclear patterns of spatial confinement compared to WT cells. The slow-mobility α state is absent, and the fraction of molecules in the intermediate-mobility β state decreases twofold (Fig. 4.3A). In contrast, the weight fractions of the γ and δ states substantially increase in H3K9me0 *clr*4Δ cells, suggesting that these mobility states are exclusively H3K9me independent (Fig. 4.3A). The estimated transition probabilities between states show that in the absence of the H3K9me-dependent low-mobility α state, PAmCherry-Swi6 molecules predominantly reside in and transition between the γ and δ fast-mobility states with only rare transitions to the β state (Fig. 4.3B). Hence, both the α and β states depend on Swi6 binding to H3K9me nucleosomes, with α occurring exclusively in *clr4+* cells and β representing a mixture of H3K9me-dependent and H3K9me-independent substates (present in *clr*+ and *clr*4Δ cells).

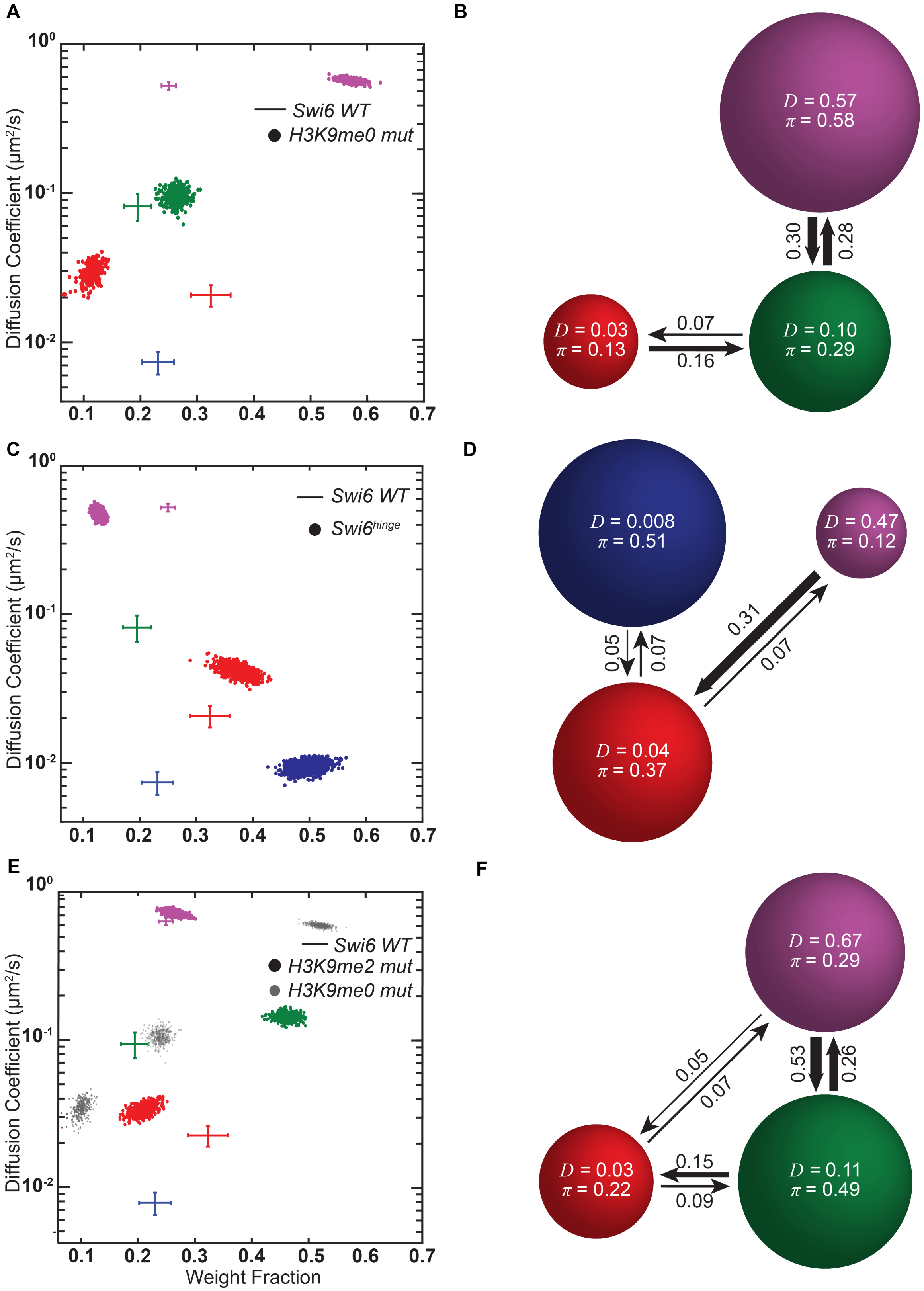

A
B
C
D
E
F
Diffusion Coefficient (μm²/s)
10⁰
10⁻¹
10⁻²
0.1
0.2
0.3
0.4
0.5
0.6
0.7
Weight Fraction
Swi6 WT
H3K9me0 mut
Swi6hinge
H3K9me2 mut
D = 0.57
π = 0.58
0.30
0.28
D = 0.10
π = 0.29
0.07
0.16
D = 0.03
π = 0.13
D = 0.008
π = 0.51
D = 0.47
π = 0.12
0.05
0.07
0.31
0.07
D = 0.04
π = 0.37
D = 0.67
π = 0.29
0.53
0.26
0.05
0.07
D = 0.03
π = 0.22
0.15
0.09
D = 0.11
π = 0.49

Figure 4.3

Figure 4.3: **(A)** SMAUG identifies three distinct mobility states for PAmCherry-Swi6 in H3K9me0 mut (*clr*4Δ) cells; the α mobility state is absent. Dataset: 10,432 steps from 2463 trajectories. **(B)** Average probabilities (arrows) of a PAmCherry-Swi6 molecule in H3K9me0 mut cells transitioning between the mobility states (circles) from **(A)**. **(C)** SMAUG identifies three distinct mobility states for PAmCherry-Swi6^{hinge}(*swi6* KR25A) molecules; the γ mobility state is absent. Dataset: 12,788 steps from 1210 trajectories. **(D)** Average probabilities (arrows) of a PAmCherry-Swi6^{hinge}molecule transitioning between the mobility states (circles) from **(C)**. **(E)** SMAUG identifies three distinct mobility states for PAmCherry-Swi6 in H3K9me2 mut (*mst2* F449Y mutant) cells; the α mobility state is absent. Dataset: 14,837 steps from 2308 trajectories. H3K9me0 mut clusters from **(A)** are provided as a reference (gray circles). **(F)** Average probabilities (arrows) of a PAmCherry-Swi6 molecule in H3K9me2 mut cells transitioning between the mobility states (circles) from **(E)**. Each point in **(A)**, **(C)**, and **(E)** is the average single-molecule diffusion coefficient, D, of Swi6 molecules in that state at a saved iteration of the Bayesian algorithm after convergence; the average and SD of the WT Swi6 clusters (blue α, red β, green γ, and purple δ, respectively, from Fig. 4.1F) are provided as a reference (crosshairs). Each circle area in **(B)**, **(D)**, and **(E)** is proportional to the weight fraction, π; D is in square micrometers per second; low-frequency transition probabilities below 0.04 are excluded.

A tryptophan to alanine substitution (W104A) within the Swi6 CD attenuates H3K9me binding approximately 100-fold (Swi6 CD^{mut}) [168, 187]. We expressed PAmCherry–Swi6 CD^{mut} in *S. pombe* cells. Our SMAUG analysis identified three mobility states for PAmCherry – Swi6 CD^{mut} (fig. 4.4A), and the distribution of mobility states also showed depletion of the α state similar to the depletion of the α state of Swi6 in H3K9me0 *clr*4Δ cells (c.f. Fig. 4.3A). A leucine to glutamate substitution (L315E) within the Swi6 CSD disrupts dimerization and higher-order oligomerization (Swi6 CSD^{mut}). We expressed PAmCherry-Swi6 CSD^{mut}) in *S. pombe* cells. Our SMAUG analysis identified three mobility states for PAmCherry-Swi6 CSD^{mut}(fig. 4.4B). Notably, the low-mobility α state is absent in PAmCherry-Swi6 CSD cells. We also mutated two residues within the Swi6 CD domain (R93A K94A), which disrupts CD-CD–dependent Swi6 oligomerization [187], and measured the mobility states associated with PAmCherry-Swi6 Loop-X. We were unable to detect substantive changes in the fraction of molecules in the α state upon introducing alanine substitutions with the ARK loop (fig. 4.4C). This could be because CD-CD interactions have a more subtle effect on Swi6 oligomerization *in vivo* and play a minor role relative to CSD-dependent interactions. It is likely that additional destabilizing mutations within the CD-CD interface might be needed to further disrupt CD-CD–dependent Swi6 oligomerization *in vivo*. On the basis of our measurements, we conclude that the low-mobility α state depends on the H3K9me substrate, on CD-mediated H3K9me recognition, and on CSD-mediated dimerization.

4.2.3 Transient nucleic acid binding leads to an intermediate apparent diffusion coefficient for Swi6

HP1 proteins have a variable-length hinge region that connects the H3K9me recognition CD domain and the CSD oligomerization domain [75]. In the case of Swi6, the hinge region has 25 lysine and arginine residues that modulate the interaction between Swi6 and nucleic acids (DNA and RNA). We replaced all 25 lysine and arginine residues with alanine (Swi6^{hinge}) and imaged the mobility patterns of PAmCherry-$Swi6^{hinge}$ [177]. Neutralizing the net positive charge within the hinge region results in fewer fast-diffusing molecules and a substantial increase in the proportion of low-mobility Swi6^{hinge}molecules relative to the WT protein. These qualitative observations were consistent with our quantitative analysis by SMAUG. We detected three mobility states (as opposed to four in the case of the WT Swi6 protein), a substantial increase in the populations of the H3K9me-dependent α slow-mobility and β intermediate-mobility states, and a concomitant decrease in the population of the δ fast-diffusing state (Fig. 4.3C). Notably, the γ intermediate-mobility state is absent in the case of PAmCherry-Swi6^{hinge}(Fig. 4.3C). Hence, we infer that the γ intermediate-mobility state corresponds to Swi6 bound to nucleic acids (DNA or RNA) via its hinge region. Furthermore, the twofold increase in the weight fraction of the slow α state suggests that, in the absence of DNA or RNA binding, PAmCherry - $Swi6^{hinge}$ molecules preferentially interact with H3K9me chromatin. Last, we noticed that the diffusion coefficient of the β mobility state exhibits a twofold increase, suggesting that the loss of nucleic acid binding additionally destabilizes this intermediate.

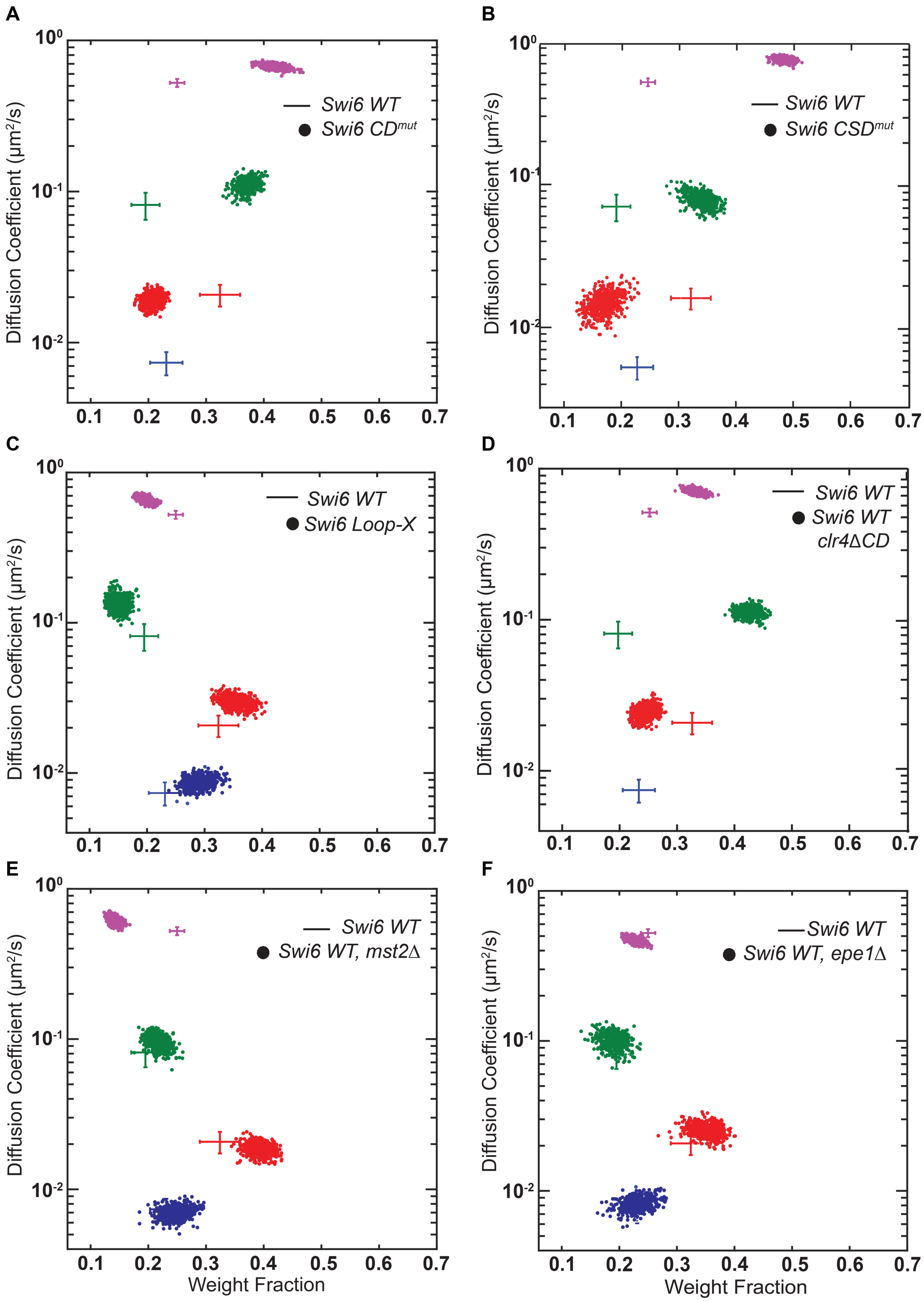

A
Swi6 WT
● Swi6 CD^{mut}
B
Swi6 WT
● Swi6 CSD^{mut}
C
Swi6 WT
● Swi6 Loop-X
D
Swi6 WT
● Swi6 WT clr4ΔCD
E
Swi6 WT
● Swi6 WT, mst2Δ
F
Swi6 WT
● Swi6 WT, epe1Δ
Diffusion Coefficient (μm²/s)
10^{0}
10^{-1}
10^{-2}
0.1
0.2
0.3
0.4
0.5
0.6
0.7
Weight Fraction

Figure 4.4

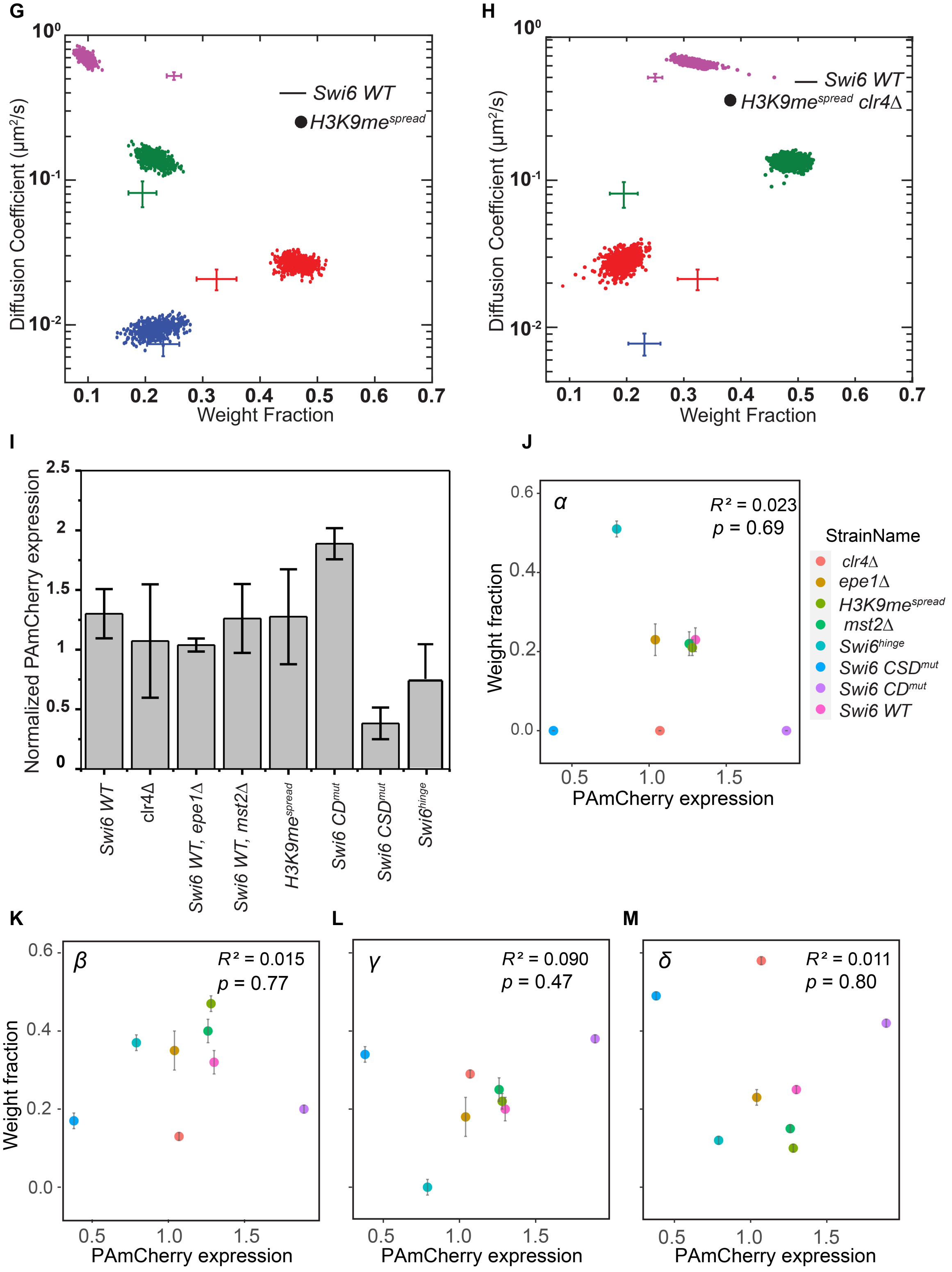

G
10⁰
10⁻¹
10⁻²
Diffusion Coefficient (µm²/s)
0.1 0.2 0.3 0.4 0.5 0.6 0.7
Weight Fraction
— Swi6 WT
● H3K9me^spread
H
10⁰
10⁻¹
10⁻²
Diffusion Coefficient (µm²/s)
0.1 0.2 0.3 0.4 0.5 0.6 0.7
Weight Fraction
— Swi6 WT
● H3K9me^spread clr4Δ
I
Normalized PAmCherry expression
0 0.5 1 1.5 2 2.5
Swi6 WT
clr4Δ
Swi6 WT, epe1Δ
Swi6 WT, mst2Δ
H3K9me^spread
Swi6 CD^mut
Swi6 CSD^mut
Swi6^hinge
J
α
R² = 0.023
p = 0.69
Weight fraction
0.0 0.2 0.4 0.6
0.5 1.0 1.5
PAmCherry expression
StrainName
clr4Δ
epe1Δ
H3K9me^spread
mst2Δ
Swi6^hinge
Swi6 CSD^mut
Swi6 CD^mut
Swi6 WT
K
β
R² = 0.015
p = 0.77
Weight fraction
0.0 0.2 0.4 0.6
0.5 1.0 1.5
PAmCherry expression
L
γ
R² = 0.090
p = 0.47
0.5 1.0 1.5
PAmCherry expression
M
δ
R² = 0.011
p = 0.80
0.5 1.0 1.5
PAmCherry expression

Figure 4.4

Figure 4.4: (A). Average single-molecule diffusion coefficients and weight fraction estimates for PAmCherry-Swi6 CD^{mut} expressing cells (*swi6 W104A*, which has a tryptophan to alanine mutation within the Swi6 chromodomain that disrupts H3K9me recognition and binding). SMAUG identifies three distinct mobility states, (red β, green γ, and purple δ, respectively) for PAmCherry-Swi6 in CD^{mut} cells. Each point represents the average single-molecule diffusion coefficient vs. weight fraction of PAmCherry-Swi6 molecules in each distinct mobility state at each saved iteration of the Bayesian algorithm after convergence. The dataset contains 10075 steps from 1624 trajectories. The wild-type Swi6 clusters (Figure4.1F) are provided for reference (cross hairs).**(B)**. Average single-molecule diffusion coefficients and weight fraction estimates for PAmCherry-Swi6 CSD^{mut} expressing cells (*swi6 L315E*, which has a leucine to glutamate mutation within the Swi6 chromoshadow that disrupts H3K9me recognition and binding). SMAUG identifies three distinct mobility states, (red β, green γ, and purple δ, respectively) for PAmCherry-Swi6 in CSD^{mut} cells. Each point represents the average single-molecule diffusion coefficient vs. weight fraction of PAmCherry-Swi6 molecules in each distinct mobility state at each saved iteration of the Bayesian algorithm after convergence. The dataset contains 11206 steps from 1625 trajectories. The wild-type Swi6 clusters (Figure 4.1F) are provided for reference (cross hairs).**(C)**. Average single-molecule diffusion coefficients and weight fraction estimates for cells expressing PAmCherry-Swi6-LoopX. SMAUG identifies four distinct mobility states, (blue α, red β, green γ, and purple δ, respectively) for PAmCherry-Swi6 in Swi6-LoopX cells. Each point represents the average single-molecule diffusion coefficient vs. weight fraction of PAmCherry-Swi6 molecules in each distinct mobility state at each saved iteration of the Bayesian algorithm after convergence. The dataset contains 12132 steps from 1413 trajectories The WT clusters (Figure 4.1F) are provided for reference (cross hairs).**(D)**. Average single-molecule diffusion coefficients and weight fraction estimates for PAmCherry-Swi6 molecules expressed in chromodomain deleted *clr4* ($clr4\Delta CD$) cells. SMAUG identifies three distinct mobility states, (red β, green γ, and purple δ, respectively) for PAmCherry-Swi6 in $clr4\Delta CD$ mut cells. Each point represents the average single-molecule diffusion coefficient vs. weight fraction of PAmCherry-Swi6 molecules in each distinct mobility state at each saved iteration of the Bayesian algorithm after convergence. The dataset contains 10205 steps from 2551 trajectories. Specifically, the α mobility state is absent in $clr4\Delta CD$ cells. The average and standard deviation of the WT Swi6 clusters (Figure 4.1F) are provided as a reference (cross hairs).**(E)**. Average single-molecule diffusion coefficients and weight fraction estimates for PAmCherry-Swi6 molecules expressed in $mst2\Delta$ cells. SMAUG identifies four distinct mobility states, (blue α, red β, green γ, and purple δ, respectively) for PAmCherry-Swi6 in $mst2\Delta$ cells. Each point represents the average single-molecule diffusion coefficient vs. weight fraction of PAmCherry-Swi6 molecules in each distinct mobility state at each saved iteration of the Bayesian algorithm after convergence. The dataset contains 11341 steps from 1619 trajectories. The wild-type Swi6 mobility states (Figure 4.1F) are provided as a reference (cross hairs).

Figure 4.4: (F). Average single-molecule diffusion coefficients and weight fraction estimates for PAmCherry-Swi6 molecules expressed in *epe*1Δ cells. SMAUG identifies four distinct mobility states, (blue α, red β, green γ, and purple δ, respectively) for PAmCherry-Swi6 in *epe*1Δ cells. Each point represents the average single-molecule diffusion coefficient vs. weight fraction of PAmCherry-Swi6 molecules in each distinct mobility state at each saved iteration of the Bayesian algorithm after convergence. The dataset contains 12790 steps from 2095 trajectories. The wild-type Swi6 mobility states (Figure 4.1F) are provided as a reference (cross hairs).**(G)**. Average single-molecule diffusion coefficients and weight fraction estimates for PAmCherry-Swi6 molecules in H3K9mespread (*mst*2Δ, *epe*1Δ cells. SMAUG identifies four distinct mobility states, (blue α, red β, green γ, and purple δ, respectively) for PAmCherry-Swi6 in H3K9mespread cells. Each point represents the average single-molecule diffusion coefficient vs. weight fraction of PAmCherry-Swi6 molecules in each distinct mobility state at each saved iteration of the Bayesian algorithm after convergence. The dataset contains 11425 steps from 1287 trajectories. The wild-type Swi6 diffusion coefficients (Figure 4.1F) are provided as a reference (cross hairs).**(H)**. Average single-molecule diffusion coefficients and weight fraction estimates for PAmCherry-Swi6 molecules expressed in chromodomain deleted *clr*4 (*clr*4Δ*CD*) cells. SMAUG identifies three distinct mobility states, (red β, green γ, and purple δ, respectively) for PAmCherry-Swi6 in *clr*4Δ*CD* mut cells. Each point represents the average single-molecule diffusion coefficient vs. weight fraction of PAmCherry-Swi6 molecules in each distinct mobility state at each saved iteration of the Bayesian algorithm after convergence. The dataset contains 10205 steps from 2551 trajectories. Specifically, the α mobility state is absent in *clr*4Δ*CD* cells. The average and standard deviation of the wild-type Swi6 clusters (Figure4.1F) are provided as a reference (cross hairs).**(I)**. The level of PAmCherry-Swi6 was examined by western blot using an mCherry antibody for Swi6 WT, *clr*4Δ Swi6 WT, *epe*1Δ Swi6 WT, *mst*2Δ, H3K9mespread, Swi6 CD^{mut}, Swi6 CSD^{mut} and Swi6-hinge cells (N= 2). Error bars represent standard deviation.**(J)**. The scatter plot of normalized PAmCherry expression level vs weight fraction of state α. Each color denotes a strain. Error bar indicates the estimated standard deviation for state fraction. Strains that have no state α in their SMAUG analysis are scattered with weight fraction zero. The R^2 and p-value note in text are for the linear regression between expression level and weight fraction.**(K)**. The scatter plot of normalized PAmCherry expression level vs weight fraction of state β. Each color denotes a strain. Error bar indicates the estimated standard deviation for state fraction. Strains that have no state corresponding state in their SMAUG analysis are scattered with weight fraction zero. The R^2 and p-value note in text are for the linear regression between expression level and weight fraction of the state.**(L)**. The scatter plot of normalized PAmCherry expression level vs weight fraction of state γ. Each color denotes a strain. Error bar indicates the estimated standard deviation for state fraction. Strains that have no state corresponding state in their SMAUG analysis are scattered with weight fraction zero. The R^2 and p-value note in text are for the linear regression between expression level and weight fraction of the state.**(M)**. The scatter plot of normalized PAmCherry expression level vs weight fraction of state δ. Each color denotes a strain. Error bar indicates the estimated standard deviation for state fraction. Strains that have no state corresponding state in their SMAUG analysis are scattered with weight fraction zero. The R^2 and p-value note in text are for the linear regression between expression level and weight fraction of the state.

On the basis of the assignment of transition probabilities between the mobility states, we identified two critical features of Swi6 dynamics in the hinge mutant: (i) PAmCherry-$Swi6^{hinge}$ molecules in the slow-mobility α state transition less often to the remaining β and δ mobility states, and (ii) the probability of cross-transitions between the fast-diffusing δ state and the intermediate-mobility β state increases (Fig. 4.3D). Although negligible in the case of the WT PAmCherry-Swi6 protein, the cross-transitions (δ state to β state) become prominent in the case of PAmCherry-$Swi6^{hinge}$ mutant. Hence, by eliminating nucleic acid binding, we observed a substantial increase in H3K9me-dependent and H3K9me-independent chromatin association. Overall, these measurements strongly suggest that nucleic acid binding interactions compete with H3K9me localization and could potentially promote Swi6 unbinding from chromatin.

4.2.4 Weak chromatin interactions also result in an intermediate apparent diffusion coefficient for Swi6

Deleting Clr4 reduces but does not fully eliminate the population of the β intermediate-mobility state (Fig. 4.3A). Therefore, the β state must have both H3K9me-dependent and H3K9me-independent components. We hypothesized that the β intermediate-mobility state likely represents the transient sampling of chromatin by Swi6 (H3K9me or H3K9me0) before stable binding at sites of H3K9me (α state). Swi6 binds to H3K9me3 chromatin with higher affinity compared to H3K9me1/2. To eliminate the high-affinity Swi6 binding state and exclusively interrogate transient chromatin interactions, we replaced the H3K9 methyltransferase Clr4 with a mutant methyltransferase (Clr4 F449Y, referred to here as the Clr4 H3K9me2 mutant) that catalyzes H3K9 mono- and dimethylation (H3K9me1/2) but is unable to catalyze trimethylation (H3K9me3) due to a mutation within the catalytic SET domain [176]. We verified the expression of the Clr4 mutant protein in fission yeast cells using a myc epitope tag.

Following single-particle tracking measurements of Swi6, we found that cells expressing the Clr4 H3K9me2 mutant exhibit only three mobility states (Fig. 4.3E), having lost the α state. These results are consistent with our expectations and previously published data where selectively eliminating H3K9me3 interrupts stable Swi6 association at sites of heterochromatin formation [176]. Notably, we observed a twofold increase in the weight fraction of molecules residing in the β intermediate-mobility state relative to H3K9me0 cells, suggesting that H3K9me2 is sufficient to drive an increase in the occupancy of the Swi6 chromatin-bound population (Fig. 4.3E). Mapping the transition probabilities of Swi6 molecules between the remaining three mobility states further supports our conclusions. We observed an increase in the transition probability between the fast-mobility δ and γ mobility states, which is negligible or absent in H3K9me0 cells (Fig. 4.3F). Therefore, H3K9me2 enhances Swi6 chromatin binding but is incapable of driving Swi6 occupancy to the exclusively H3K9me3-dependent α state. We also tested how Swi6 mobility states change upon expression of a Clr4 CD–deficient protein, which leads to a substantial reduction in H3K9me3 levels and an increase in H3K9me2 levels [176]. The mobility states of Swi6 measured in cells expressing the Clr4 ΔCD mutant notably resemble those in cells expressing Clr4 F449Y, further supporting our observations that H3K9me2 is sufficient to promote an increase in Swi6 chromatin occupancy (β state) (fig. 4.4D). Hence, we conclude that the β intermediate corresponds to a chromatin sampling state consisting of weak and unstable interactions between Swi6 and H3K9me or H3K9me0 nucleosomes and is an on-pathway intermediate to the lowest-mobility α state.

Simultaneously, we also tested how heterochromatin misregulation affects Swi6 dynamics. We deleted proteins that are involved in maintaining proper heterochromatin boundaries: (i) Epe1, a putative H3K9 demethylase that erases H3K9me, and (ii) Mst2, an H3K14 acetyltransferase that acetylates histones and promotes active transcription [80, 81, 188]. We deleted either *epe1* or *mst2* individually in cells expressing PAmCherry-Swi6 (fig. 4.4, E and F). We observed relatively few changes in the weight fractions of the different Swi6 mobility states in these individual mutants. However, simultaneously deleting both *epe1* and *mst2* leads to a more marked rearrangement of the Swi6 mobility states (fig. 4.4G). We refer to this double mutant as an $H3K9me^{spreading}$ mutant. Unlike the H3K9me2 mutant, the fraction of molecules in the β intermediate-mobility state increases nearly twofold in the $H3K9me^{spreading}$ mutant, which also coincides with a near-complete depletion of Swi6 molecules from the unbound δ state (fig. 4.4G). As expected, deleting *clr4* in this $H3K9me^{spreading}$ mutant ($H3K9me^{spreading}$) $Clr4\Delta$) collapses the β intermediate-mobility state from 48 to 14%, similar to what we observed in $epe1 + mst2 + Clr4\Delta$ cells (Fig. 4.3A and fig. 4.4H). Hence, transient chromatin interactions increase in the case of $H3K9me^{spreading}$ mutants, suggesting that heterochromatin misregulation leads to an increase in the chromatin-bound fraction of Swi6. Through measurements of PAmCherry-Swi6 expression levels across different mutant backgrounds, we determined that there is no correlation between protein expression and mobility state occupancy (fig. 4.4, I to M).

4.2.5 Fine-grained kinetic modeling transitions reveal H3K9me binding specificity *in vivo*

Our fission yeast mutants enabled us to assign biochemical properties to each of our experimentally measured mobility states (Fig. 4.5A). The state-to-state transitions inferred by SMAUG (Fig. 4.1F) estimate the probability with which a molecule of Swi6 assigned to one state during one 40-ms imaging frame will be assigned to some other state in the next frame (40 ms later). To infer the direct biochemical processes underlying the observed Swi6 dynamics, we implemented a high–temporal resolution model using a Bayesian synthetic likelihood (BSL) approach (see 4.4 for details) to determine the most consistent set of fine-scale chemical rate constants [189]. Note that the model we have used is a statistical inference algorithm, not a generative model: As opposed to being trained using our experimental data to simulate other data, the model infers rate constants that can best describe the experimental results (as well as our uncertainties regarding those quantities).

We applied the BSL inference algorithm to the WT Swi6 data (shown in Fig. 4.1F) to

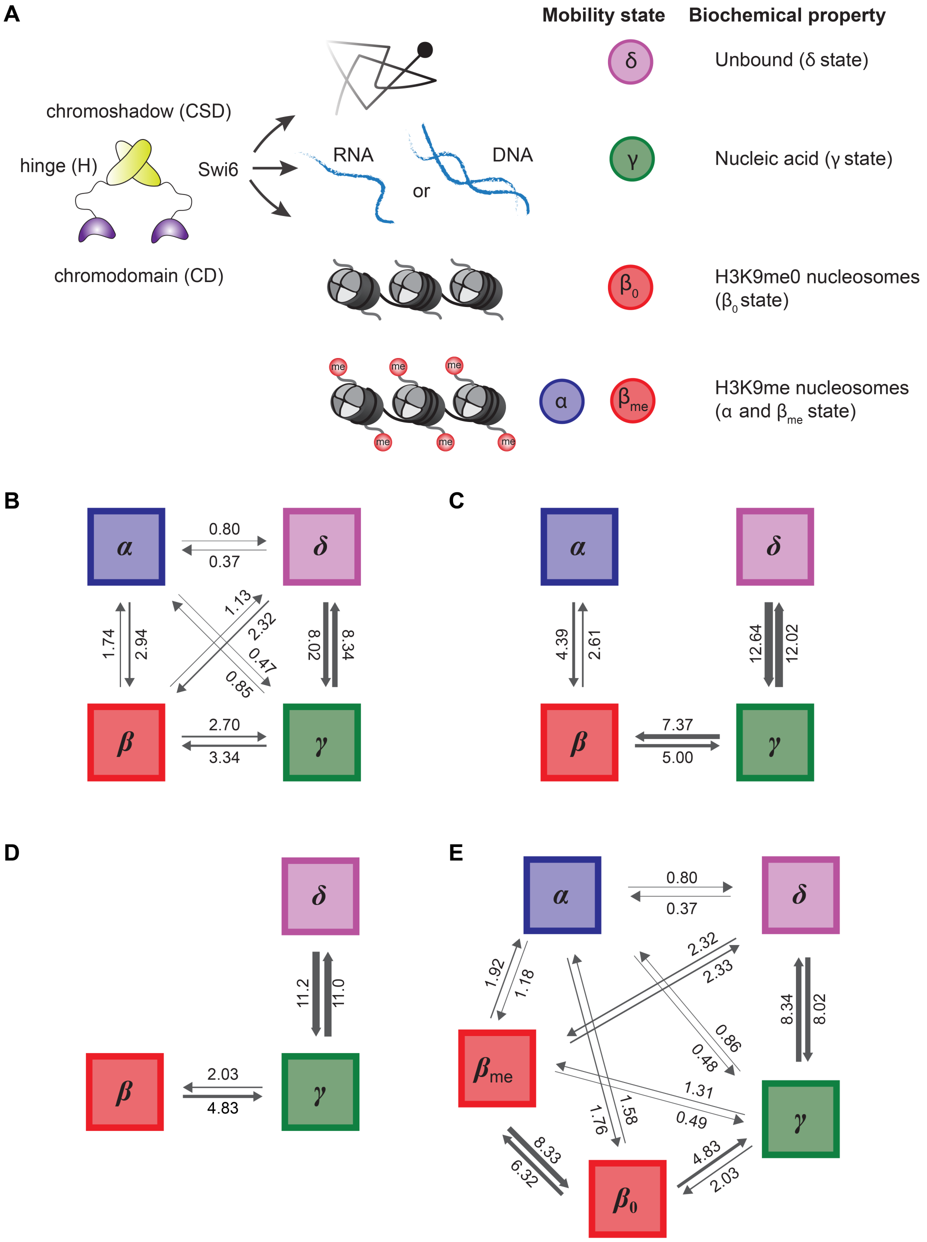


Figure 4.5: (A) Schematic showing the inferred biochemical nature of each state considered in our fine-grained chemical simulations. **(B)** Inferred rate constants for cells expressing WT Swi6, in units of 1/s, for transitions of Swi6 assuming that there are four biochemical states of Swi6 and that any state can chemically transition to any other. **(C)** Inferred rate constants (in units of 1/s) for the same situation as in **(B)** but assuming that states can only transition to the adjacent mobility states. **(D)** Inferred rate constants for *clr*4Δ cells, in units of 1/s, for the Swi6 transitions, assuming a three-state model as inferred by SMAUG. **(E)** Inferred rate constants, in units of 1/s, assuming five biochemical states of Swi6 in the WT cells. This model assumes that the β state corresponds to two chemical states: one with Swi6 bound to fully methylated H3K9me and one with Swi6 bound to unmethylated H3K9me; the latter is assigned parameters based on the *clr*4Δ simulation shown in **(B)**.

simulate transition rates between the different mobility states (α, β, γ, and δ). We compared the complete simulated posterior distributions (violin plots in fig.4.6A) to the naïve transition rates obtained by assuming that the SMAUG transitions were exclusively single chemical transitions (dashed lines in fig. 4.6A). When we simulate the experiment using the inferred chemical rate constants from BSL, the simulations agree qualitatively with the direct analysis of the experimental results, demonstrating that our SMAUG analysis indeed captures the relevant time scales underlying Swi6 interstate transitions. Formal model comparison (4.4) indicates that the rates inferred via BSL show a favorable value of the Bayesian information criterion ($\Delta BIC = -5.4 \times 10^{-3}$) relative to the naïve rates. Hence, a detailed consideration of the chemical kinetics underlying the observed Swi6 transitions yields more quantitatively accurate information. Next, we compared our rate constant inferences between the models where (i) Swi6 transitions directly between nonadjacent mobility states (a dense model; Fig.4.5B) and (ii) Swi6 transitions between adjacent mobility states (a sparse model; Fig. 4.5C). Since we have full posterior distributions for our inferences, we switch to the widely applicable information criterion (WAIC) for these comparisons. We found that the dense model is favored ($\Delta WAIC = -3.7 \times 10^{4}$), indicating that transitions between both adjacent and nonadjacent states are possible within a single (40-ms) experimental observation. We also confirmed that our algorithm inferred the absolute quantities of our rate constants, not only relative quantities (4.4).

We next examined the effects of eliminating H3K9me (replicating the scenario observed in $clr4\Delta$ cells) on interstate transitions. Consistent with the experimental data from Fig. 4.3B, our simulations revealed that the transition rates between the γ and δ states do not appreciably change in $clr4\Delta$ cells relative to WT cells (Fig. 4.5D). However, the transitions involving the β state are significantly altered in $clr4\Delta$ cells, with far more rapid transitions from the β to the γ state, indicating that the β state is less stable (Fig. 4.5D): When we did a separate BSL inference on the $clr4\Delta$ experiments, the new inference agreed better with the experiments than the WT inference rates without the α state

($\Delta WAIC = -8.9 \times 10^4$). As previously noted, our single-molecule tracking experiments led us to infer that the β state consists of H3K9me-dependent and H3K9me-independent components, with only the H3K9me-independent component present in $clr4\Delta$ cells.

To deconvolute the two β state components, we performed simulations in WT cells under the assumption that the β state observed in $clr4\Delta$ cells solely represents the H3K9me0 β state. Denoting the proposed substates of β as β_0 (H3K9me0) and β_{me} (H3K9me1/2/3), we repeated our BSL inference on the WT Swi6 data using a five-state model of the system. We constrained parameters involving transitions between the β0 and γ states and the β0 and δ states to match those for the $clr4\Delta$ cells while using the WT rates for the transitions between the γ and δ states. The resulting rate constants agree substantially better with the experimental data than did the original four-state model ($\Delta WAIC = -4.7 \times 10^4$ relative to the four-state model) (Fig. 4.5E). On the basis of these results, we conclude that the β0 state is an unstable intermediate, with high rates of $\beta_0 \rightarrow \gamma$ transitions (H3K9me0 to nucleic acid binding) and $\beta_0 \rightarrow \beta_{me}$ transitions (H3K9me0 to H3K9me binding). Hence, despite the relative abundance of H3K9me0 chromatin in *S. pombe* cells, the fast dissociation of Swi6 from the β_0 state ensures that Swi6 spends little time bound nonspecifically to H3K9me0 chromatin.

Last, on the basis of estimates that 1 to 2.5% of the *S. pombe* genome consists of H3K9me3 nucleosomes, we found that the preference of Swi6 to bind H3K9me versus H3K9me0 chromatin (ratio of rate constants between β_0 and β_{me}) is 94-fold (24). While our inference regarding the equilibrium constant is sensitive to the proportion of H3K9, which is trimethylated, the most favorable plausible equilibrium for Swi6 still favors βme more than 35-fold (fig. 4.6B). Since the low-mobility α state depends on H3K9me recognition and oligomerization (fig. 4.4, A and B), we propose that CSD oligomerization amplifies the ability of Swi6 to discriminate between H3K9me0 and H3K9me chromatin in the nucleus. In summary, our analysis using the BSL-based model achieves three important goals: (i) validates the time resolution we use in our single-molecule imaging study, (ii) identifies rate constants for transitions between mobility states with virtually identical diffusion coefficients (β_0 and β_{me3})), and (iii) provides a direct measure of the H3K9me3 recognition specificity in cells.

4.2.6 Increased CD valency compensates the disruption of Swi6 oligomerization

To uncouple H3K9me recognition from CSD-dependent oligomerization, we replaced the Swi6 CSD oligomerization domain with a glutathione S-transferase (GST) tag (fig. 4.8A) [190]. Overall, the GST fusion construct is expected to maintain Swi6 dimerization while eliminating higher-order CSD-mediated oligomerization. We refer to this hybrid protein construct as Swi6^{1XCD}-GST since the newly engineered protein has only one intact CD. GST homodimerization of the Swi6^{1XCD}-GST results in a complex that precisely consists of two CDs. We expressed PAmCherry-Swi6^{1XCD}-GST in fission yeast cells, which lack an endogenous copy of Swi6 ($swi6\Delta$). Following a strong 406-nm activation pulse, we imaged the ensemble of PAmCherry-Swi6^{1XCD}-GST molecules in the *S. pombe* nucleus. We observed a diffuse distribution of PAmCherry-Swi6^{1XCD}-GST proteins within the nucleus in contrast to WT Swi6, which exhibits prominent clusters (fig. 4.8A). These observations differ from previous studies of the mammalian HP1 isoform, HP1β, in which case replacing the CSD domain with GST did not affect HP1β localization [183]. However, unlike Swi6, HP1β exhibits a reduced oligomerization capacity and fails to form condensates *in vitro* [183, 191].

Next, we added a second CD domain to the existing PAmCherry-Swi6^{1XCD}-GST construct to generate PAmCherry-Swi6^{2XCD}-GST. GST homodimerization of the Swi6^{2XCD}-GST results in an engineered Swi6 protein that has a precise twofold increase in CD valency relative to WT Swi6 or PAmCherry-Swi6^{1XCD}-GST (Fig. 4.7A). We used high-intensity 406-nm illumination to activate the ensemble of PAmCherry-Swi6^{2XCD}-GST molecules and acquired z sections by imaging at 561 nm (Fig. 4.7C). We observed prominent foci of PAmCherry-Swi6^{2XCD}-GST molecules that qualitatively resemble WT Swi6

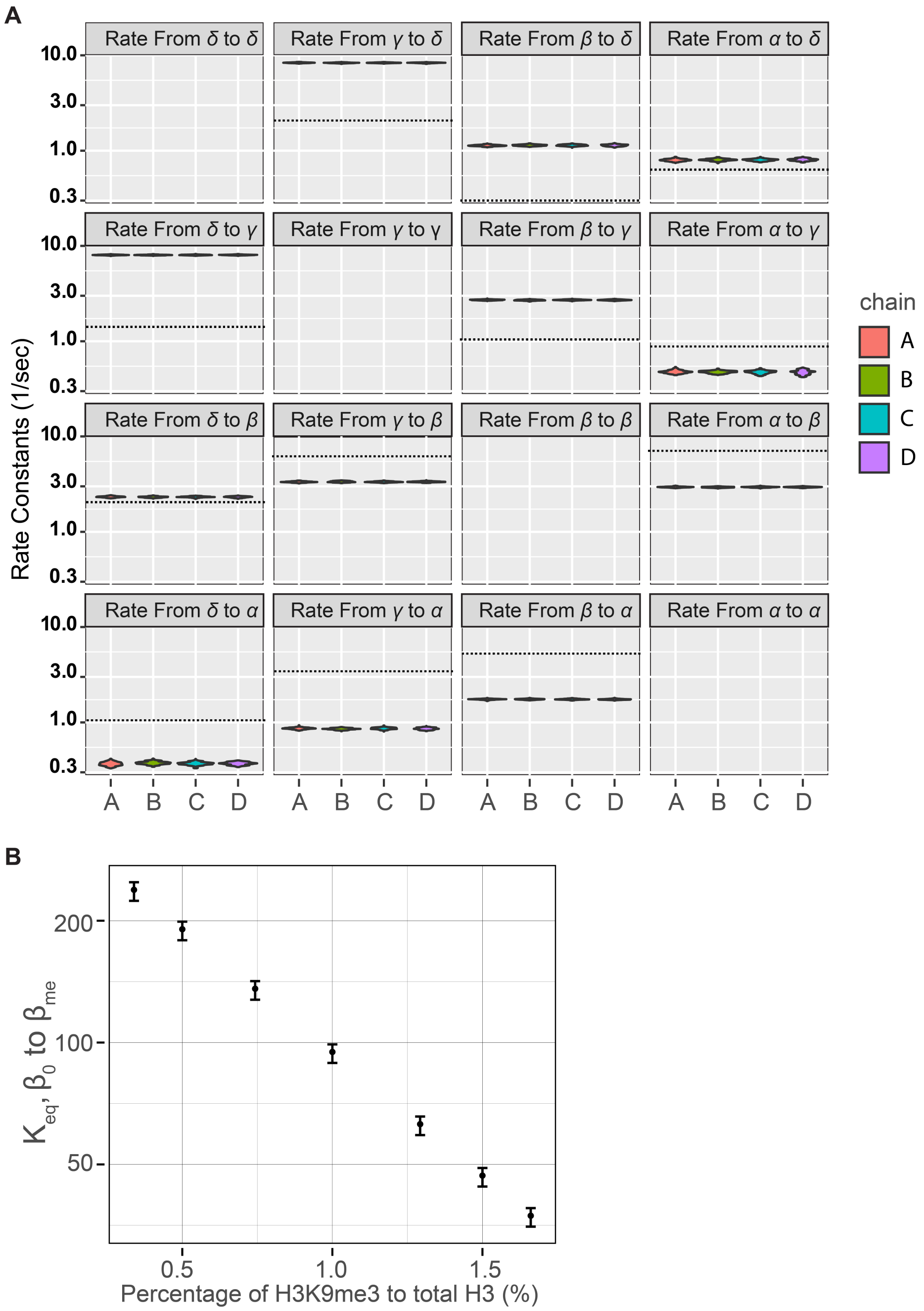


Figure 4.6: (A). The posterior distributions of the Swi6 transition rate constants for the four state wild type transitions from four independent Monte Carlo chains (referred to as A-D). The dashed lines represent the rate constants that would be deduced by transforming the SMAUG transition probabilities directly into rate constants assuming that each transition represents a single chemical transformation over the experiment.**(B)**. We calculated the posterior means for the equilibrium constant from β0 to βme given the proportion of H3K9 which is trimethylated based on quantitative mass-spectrometry measurements of total histones (1%) and H3K9me3 measurements based on Swi6 IPs (2.5%). We report equilibrium constants under a wide range of potential values for the trimethylation rate to illustrate the sensitivity of our inference to the H3K9me3 fraction. Error bars indicate the 95% credible intervals for the equilibrium constants.

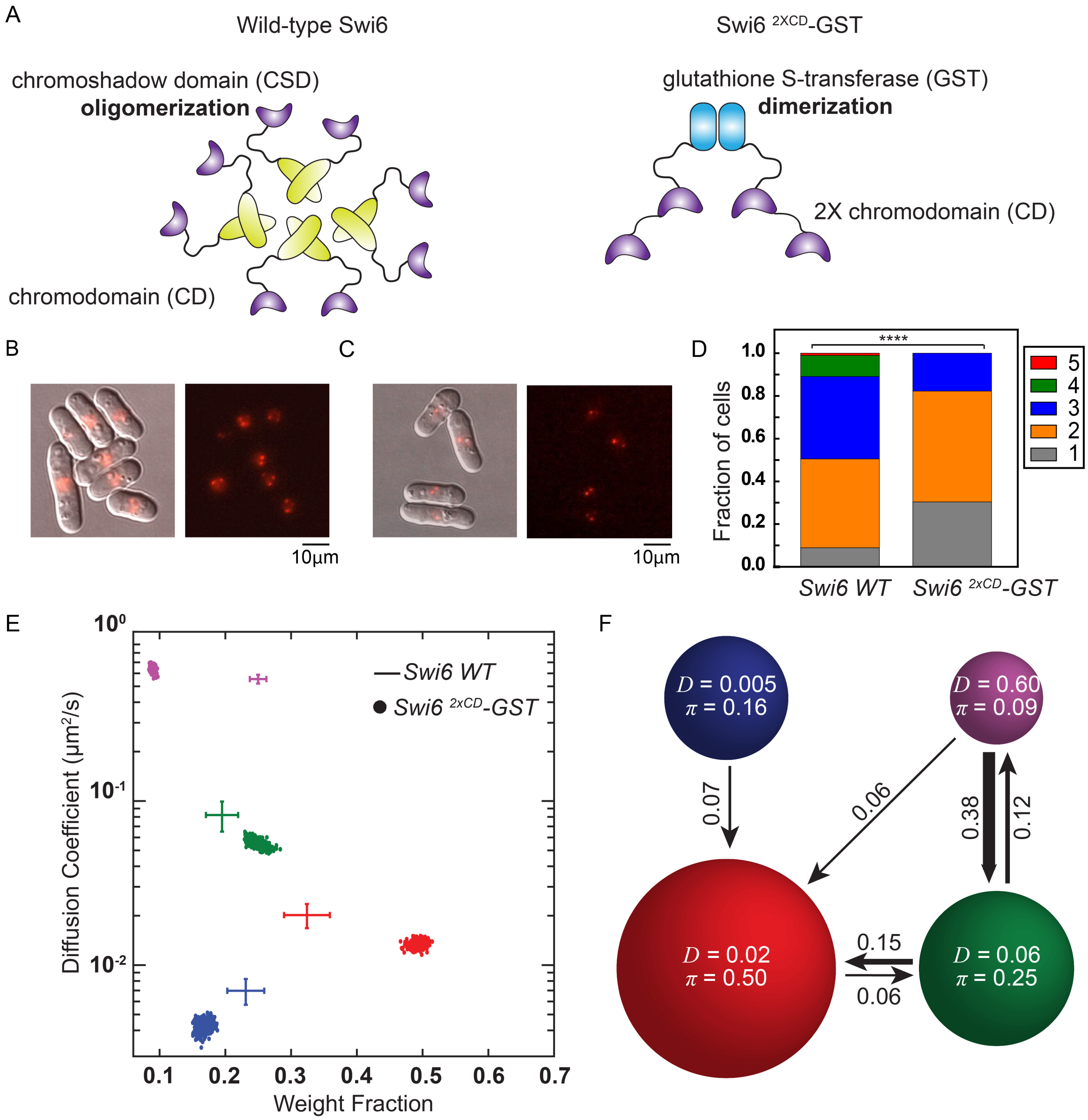


Figure 4.7: (A) CSD-mediated oligomerization of Swi6 (left) and GST-mediated dimerization of an engineered Swi6^{2XCD}-GST mutant with two tandem CDs. **(B)** Overlaid differential interference contrast and epifluorescence images (left) and epifluorescence images alone (right) of PAmCherry-Swi6 simultaneously activated using high-power 405-nm excitation and imaged using 561-nm excitation. The images are a maximum intensity projection of a Z-stack consisting of 13 images acquired at 250-nm z-axis intervals.**(C)** Overlaid differential interference contrast and epifluorescence images (left) and epifluorescence images alone (right) of PAmCherry-Swi6^{2XCD}-GST simultaneously activated using high-power 405-nm excitation and imaged using 561-nm excitation. The images are a maximum intensity projection of a Z-stack consisting of 13 images acquired at 250-nm z-axis intervals.

Figure 4.7: (D) Distribution of the number of photoactivated PAmCherry. ****p = 0.0005, Pearson's chi-squared test. **(E)** SMAUG identifies four distinct mobility states for PAmCherry-Swi6^{2XCD}-GST molecules (dots); the average and SD of the WT Swi6 clusters from Fig. 4.1F in the same four states are provided as a reference (crosshairs). Each point is the average single-molecule diffusion coefficient, D, of molecules in that state at a saved iteration of the Bayesian algorithm after convergence. Dataset: 42,382 steps from 5182 trajectories. **(F)** Average probabilities (arrows) of a PAmCherry-Swi6^{2XCD}-GST molecule transitioning between the mobility states (circles) from **(E)**.

foci in cells (Fig. 4.7B). We measured expression levels of Swi6^{1XCD} and Swi6^{2XCD}-GST proteins and found no correlation between their expression levels and mobility state occupancy (fig. 4.8, B and C). We measured H3K9me2 levels and the binding of our PAmCherry fusion proteins at sites of constitutive heterochromatin (pericentromeric repeats, mating-type locus, and telomeres) using ChIP-qPCR (fig. 4.8, D and E). H3K9me2 levels are intact at dg and tel1 in cells expressing PAmCherry-Swi6^{2XCD}-GST. H3K9me2 is absent at the mat locus, consistent with previous studies that have shown that deleting Swi6 sensitizes the mating-type locus to the complete loss of H3K9me (28). Furthermore, using ChIP-qPCR, we confirmed that PAmCherry-Swi6^{2XCD}-GST binds to chromatin at dg and tel1 in a manner that is similar to that of WT PAmCherry-Swi6 (fig. 4.8E).

Comparing the distribution of foci numbers per cell reveals a skew in the distribution with a more significant proportion of cells that exhibit three to five foci in the case of WT Swi6 compared to the 2XCD-GST fusion construct (Fig. 4.7D). To assess the localization and silencing capability of Swi6^{2XCDT}-GS in the absence of any protein-mediated dimerization, we introduced seven mutations within the GST dimer interface to attenuate the high-affinity GST dimerization [192] to create PAmCherry-Swi6^{2XCD}-GST mutant. The majority of cells exhibit zero or one cluster (71%) compared to the dimerization-competent GST allele (fig. 4.8, F and I). Hence, GST-mediated dimerization of two tandem CDs is necessary for H3K9me recognition, and mutations that affect dimerization reduce CD-mediated localization. In addition, we introduced CD binding mutations (W104A) to either one (PAmCherry-Swi6$^{2XCD-mut1}$-GST) or both CDs (PAmCherry-Swi6$^{2XCD-mut1/2}$)

in the context of the 2XCD-GST fusion construct. Imaging cells where one CD was mutated resulted in 66% of cells having no clusters, whereas mutating both CDs resulted in a complete loss of PAmCherry foci (fig. 4.8, G to I). Hence, the second tandem CD partially contributes to PAmCherry-Swi6$^{2XCD-GST}$ localization even when one CD has been mutated.

Using glutaraldehyde-based cross-linking, we set up reactions where we added increasing amounts of recombinant 3XFLAG epitope–tagged WT Swi6, Swi6 L315E, and Swi6^{2XCD}-GST. We confirmed that the ^{2XCD}-GST fusion construct fails to form higher-order oligomers unlike WT Swi6 (fig. 4.8J). Last, we tested whether the expression of the 2XCD-GST fusion construct in cells that lack Swi6 (*swi*6Δ) restores epigenetic silencing using an ade6+ reporter inserted at the pericentromeric repeats (*otr1R*) where we observe PAmCherry-Swi6^{2XCD}-GST localization. Unlike WT Swi6–expressing cells, which appear red due to silencing of the ade6+ reporter, PAmCherry-Swi62XCD-GST–expressing cells exhibit a loss of red pigmentation consistent with the absence of silencing (fig. 4.8K). Hence, restoring CD-dependent protein localization is insufficient to rescue epigenetic silencing, confirming that additional Swi6 CSD–mediated protein-protein interactions are required for this process.

We mapped the mobility states associated with PAmCherry-Swi6^{2XCD}-GST in *S. pombe*. Despite the differences in the overall number of foci, the twofold increase in CD valency completely circumvents the need for higher-order oligomerization: Swi6^{2XCD}-GST fully restores the localization of Swi6 to sites of heterochromatin at levels that rival those of WT Swi6 (Fig. 4.7E). We measured a slow-mobility state (α state) population of 20%, similar to that of the WT Swi6 protein (Fig. 4.1F). Besides, there is a substantial increase in the β intermediate sampling state, indicating that PAmCherry-Swi6^{2XCD}-GST exhibits increased chromatin association. We confirmed that the localization of PAmCherry-Swi6^{2XCD}-GST within the genome depends exclusively on H3K9me by deleting Clr4 (PAmCherry-Swi6^{2XCD}-GST *clr*4Δ). PAmCherry-Swi6^{2XCD}-GST *clr*4Δ cells exhibit a complete loss of the slow-

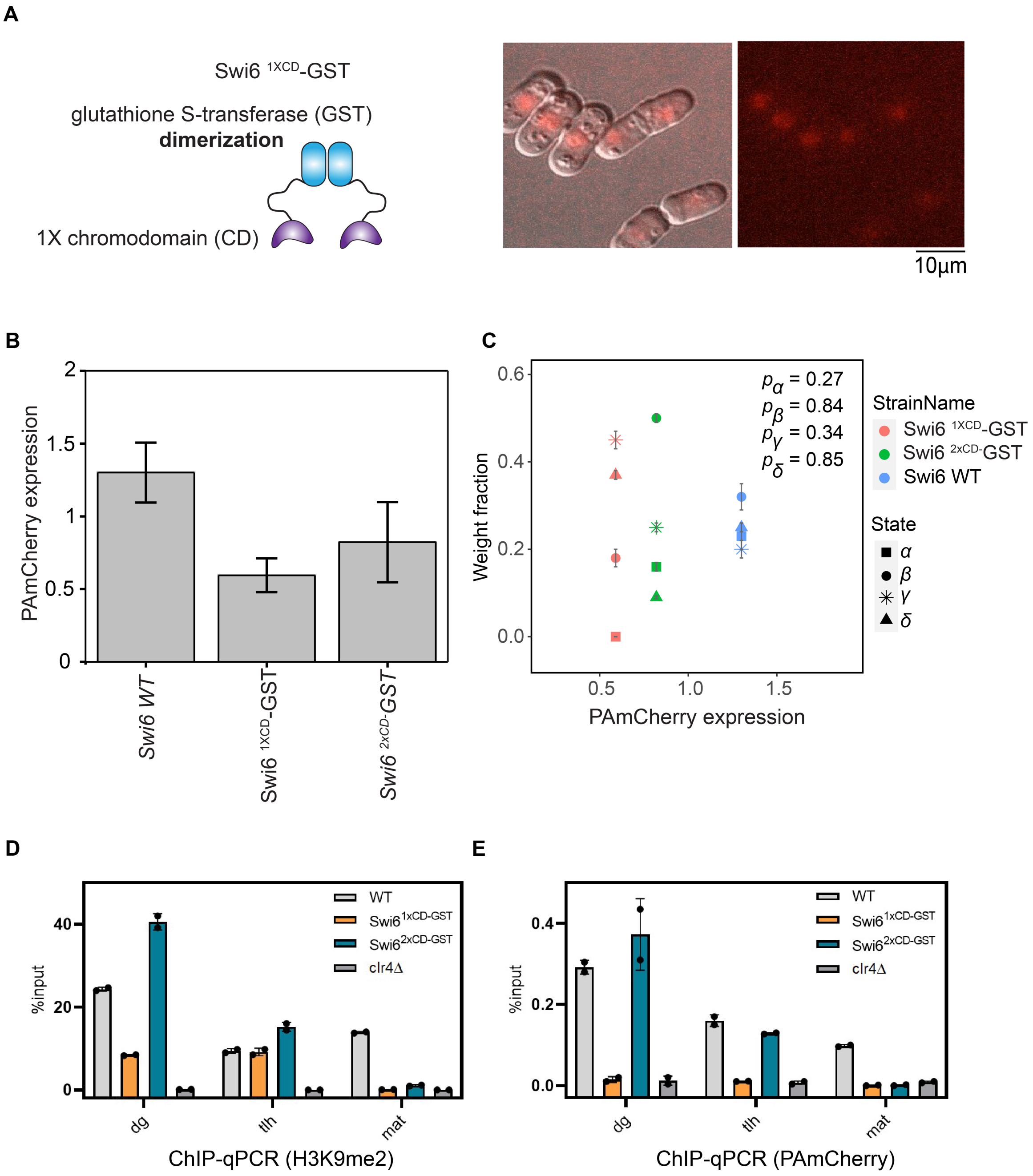

A
Swi6 1XCD-GST
glutathione S-transferase (GST)
dimerization
1X chromodomain (CD)
10µm
B
PAmCherry expression
2
1.5
1
0.5
0
Swi6 WT
Swi6 1XCD-GST
Swi6 2xCD-GST
C
Weight fraction
0.6
0.4
0.2
0.0
0.5
1.0
1.5
PAmCherry expression
p_α = 0.27
p_β = 0.84
p_γ = 0.34
p_δ = 0.85
StrainName
Swi6 1XCD-GST
Swi6 2xCD-GST
Swi6 WT
State
α
β
γ
δ
D
%input
40
20
0
WT
Swi6 1xCD-GST
Swi6 2xCD-GST
clr4Δ
dg
tlh
mat
ChIP-qPCR (H3K9me2)
E
%input
0.4
0.2
0.0
WT
Swi6 1xCD-GST
Swi6 2xCD-GST
clr4Δ
dg
tlh
mat
ChIP-qPCR (PAmCherry)

Figure 4.8

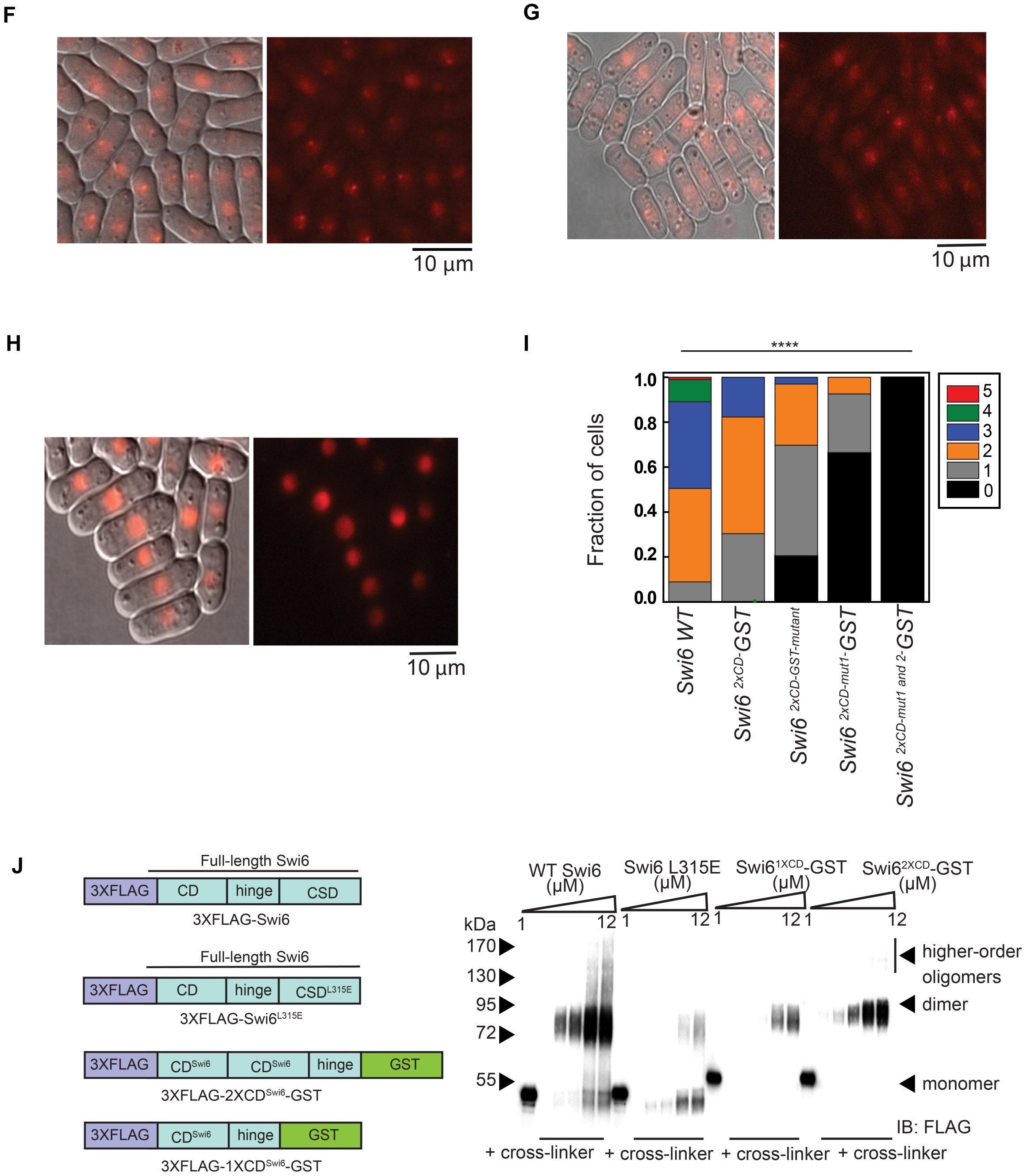

F
10 µm
G
10 µm
H
10 µm
I
****
Fraction of cells
1.0
0.8
0.6
0.4
0.2
0.0
5
4
3
2
1
0
Swi6 WT
Swi6 2xCD-GST
Swi6 2xCD-GST-mutant
Swi6 2xCD-mut1-GST
Swi6 2xCD-mut1 and 2-GST
J
Full-length Swi6
3XFLAG
CD
hinge
CSD
3XFLAG-Swi6
Full-length Swi6
3XFLAG
CD
hinge
CSDL315E
3XFLAG-Swi6L315E
3XFLAG
CDSwi6
CDSwi6
hinge
GST
3XFLAG-2XCDSwi6-GST
3XFLAG
CDSwi6
hinge
GST
3XFLAG-1XCDSwi6-GST
WT Swi6 (µM)
Swi6 L315E (µM)
Swi61XCD-GST (µM)
Swi62XCD-GST (µM)
kDa
1
12
1
12
1
12
1
12
170
130
95
72
55
higher-order oligomers
dimer
monomer
IB: FLAG
+ cross-linker
+ cross-linker
+ cross-linker
+ cross-linker

Figure 4.8

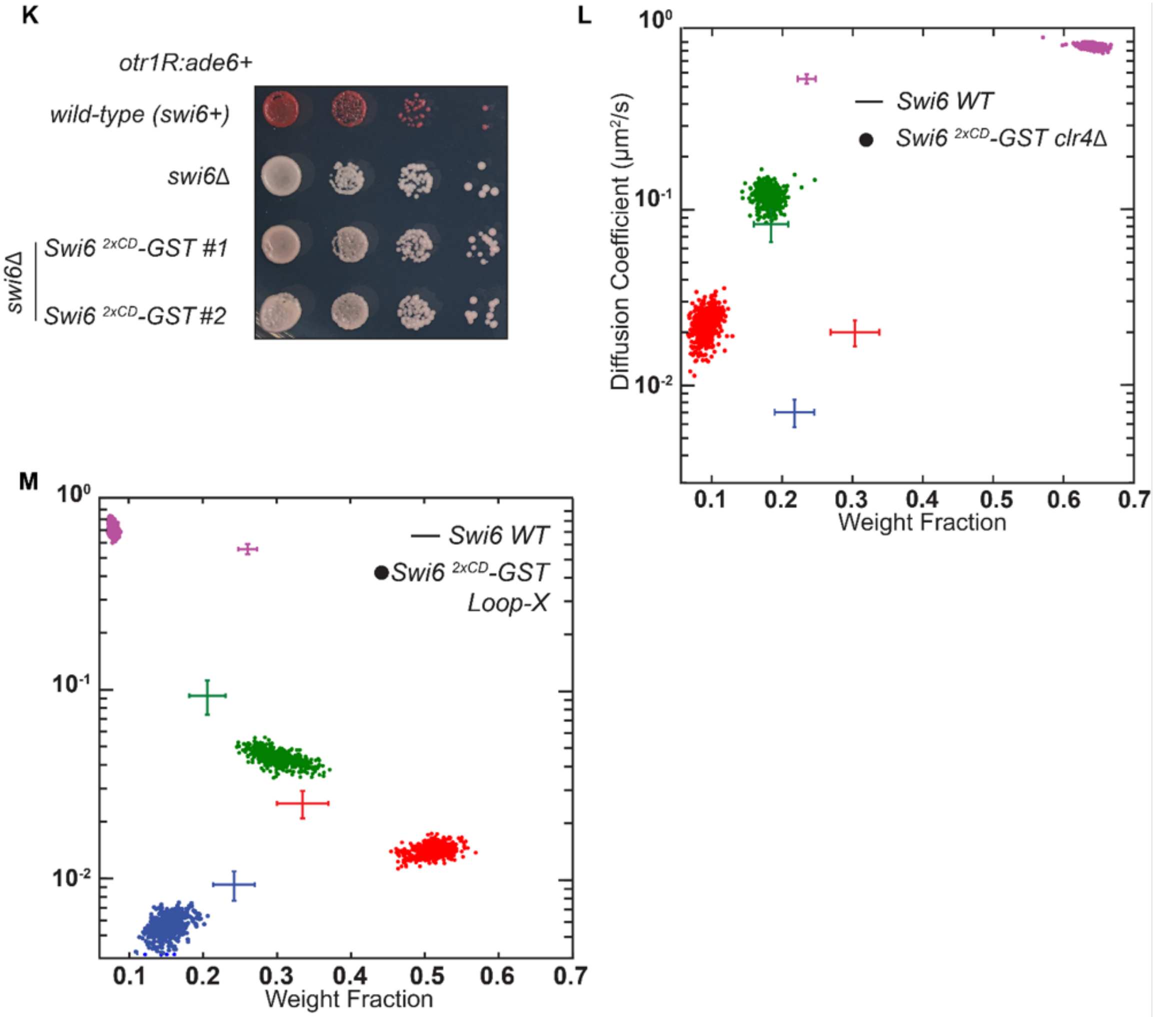


Figure 4.8: (A). Schematic representation of an engineered, dimerization competent Swi6^{1XCD}-GST protein (left panel). Overlay of differential interference contrast and epi-fluorescence image of collection of PAmCherry-Swi6^{1XCD}-GST molecules which are simultaneously activated using a high 405 nm excitation pulse and imaged using 561 nm excitation (left panel). Epi-fluorescence image of collection of PAmCherry-Swi6^{1XCD}-GST molecules activated with high 405 nm excitation and imaged with 561 nm (right panel). The images are a maximum intensity projection of a Z-stack consisting of 13 images acquired at 250 nm z-axis intervals (scale bar 10 µm).**(B)**. The level of PAmCherry-Swi6 was examined by western blot for Swi6 WT, Swi6^{1XCD}-GST and Swi6 ^{2XCD}-GST (N= 2). Error bars represent standard deviation. **(C)**. The scatter plot of normalized PAmCherry expression level vs weight fraction for Swi6 WT, Swi6^{1XCD}-GST and Swi6 ^{2XCD}-GST. Each color denotes a strain and each symbol denotes a state. Error bar indicates the estimated standard deviation for state fraction. Strains that have no state corresponding state in their SMAUG analysis are scattered with weight fraction 0. The p-value note in text are for the linear regression between expression level and weight fraction of each state.

Figure 4.8: (D). ChIP-qPCR measurements of H3K9me2 levels at sites of heterochromatin formation (dg pericentromeric repeats, tlh telomere and mat mating type locus) (N = 2), for Swi6 WT, $clr4\Delta$, Swi6^{1XCD}-GST and Swi6^{2XCD}-GST cells. Error bars represent standard deviation. **(E)**. ChIP-qPCR measurements of PAmCherry-Swi6 levels at sites of heterochromatin formation (dg pericentromeric repeats, tlh telomere and mat mating type locus) (N = 2), for Swi6 WT, $clr4\Delta$, Swi6^{1XCD}-GST and Swi6^{2XCD}-GST cells. Error bars represent standard deviation. **F**. Overlay of differential interference contrast and epi-fluorescence images of collection of PAmCherry-Swi6^{2XCD}-GST-multant molecules that are simultaneously activated using a high 405 nm excitation power and imaged using 561 nm excitation (left panel). Epi-fluorescence image of collection of PAmCherry-Swi6^{2XCD}-GST-multant molecules simultaneously activated with high 405 nm excitation and imaged with 561 nm (right panel). The images are a maximum intensity projection of a Z-stack consisting of 13 images acquired at 250 nm z-axis intervals (scale bar 10 µm).**(G)**. Overlay of differential interference contrast and epi-fluorescence images of collection of PAmCherry-Swi6$^{2XCD-mut1}$-GST molecules that are simultaneously activated using a high 405 nm excitation power and imaged using 561 nm excitation (left panel). Epi-fluorescence image of collection of PAmCherry-Swi6$^{2XCD-mut1}$-GST molecules simultaneously activated with high 405 nm excitation and imaged with 561 nm (right panel). The images are a maximum intensity projection of a Z-stack consisting of 13 images acquired at 250 nm z-axis intervals (scale bar 10 µm).**(H)**. Overlay of differential interference contrast and epi-fluorescence images of collection of PAmCherry-Swi6$^{2XCD-mut1/2}$-GST molecules that are simultaneously activated using a high 405 nm excitation power and imaged using 561 nm excitation (left panel). Epi-fluorescence image of collection of PAmCherry-Swi6$^{2XCD-mut1/2}$-GST molecules simultaneously activated with high 405 nm excitation and imaged with 561 nm (right panel). The images are a maximum intensity projection of a Z-stack consisting of 13 images acquired at 250 nm z-axis intervals (scale bar 10 µm).**(I)**. Distribution of the number of photoactivated PAmCherry clusters in PAmCherry-Swi6, PAmCherry-Swi6^{2XCD}-GST, PAmCherry-Swi6$^{2XCD-GSTmut}$, PAmCherry-Swi6$^{2XCD-mut1}$-GST and PAmCherry-Swi6$^{2XCD-mut1/2}$-GST expressing cells (****, p-value= 0.0005, Pearson's Chi-squared test). Cells expressing wild-type Swi6 exhibit more considerable variance in the number of clusters per cell.**(J)**. Swi6^{1XCD}-GST and Swi6^{2XCD}-GST is primarily a dimer and, unlike WT Swi6, does not show evidence of further oligomerization. Left: Schematic representation of recombinant FLAG-Swi6 constructs used to observe oligomerization. Right: WT Swi6 (left), Swi6 L315E (middle), Swi6^{1XCD}-GST, and Swi6^{2XCD}-GST were crosslinked using 0.01% glutaraldehyde. Non-crosslinked input controls (1 µM) and crosslinked protein samples (1, 2, 5, 7.5, and 12 µM) were separated using SDS-PAGE and detected by anti-FLAG western blot.

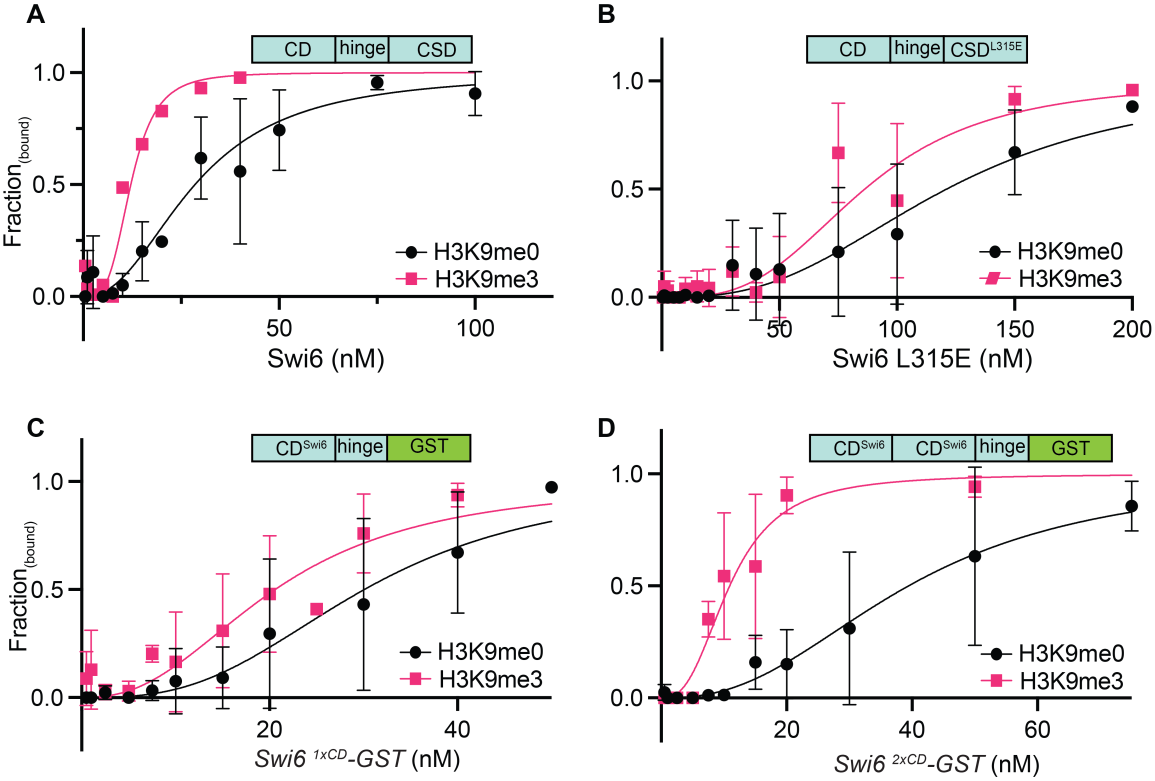
Figure 4.8: (K). A color-based assay to detect heterochromatin establishment at the silent mating type locus (*otr1Rp::ade6+*). ade6+ silencing results in the appearance of red colonies. Loss of ade6+ leads to the appearance of white colonies. Colonies represent ten-fold dilutions of wild-type (*swi6+*), $swi6\Delta$ and PAmCherry-Swi6^{2XCD}-GST $swi6\Delta$ cells. **(L)**. Average single-molecule diffusion coefficients and weight fraction estimates for PAmCherry-Swi6^{2XCD}-GST expressed in H3K9me0 ($clr4\Delta$) cells. SMAUG identifies three distinct mobility states, (red β, green γ, and purple δ, respectively) for PAmCherry-Swi6 in Swi6^{1XCD}-GST cells. Each point represents the average single-molecule diffusion coefficient vs. weight fraction of PAmCherry-Swi6 molecules in each distinct mobility state at each saved iteration of the Bayesian algorithm after convergence. The dataset contains 8265 steps from 2817 trajectories. The WT clusters (Figure 4.1F) are provided for reference (cross hairs).**(M)**. Average single-molecule diffusion coefficients and weight fraction estimates for cells expressing PAmCherry-Swi6$^{2XCD-GST}$-LoopX. SMAUG identifies four distinct mobility states, (blue α, red β, green γ, and purple δ, respectively) for PAmCherry-Swi6 in Swi6$^{2XCD-GST}$-LoopX cells. Each point represents the average single-molecule diffusion coefficient vs. weight fraction of PAmCherry-Swi6 molecules in each distinct mobility state at each saved iteration of the Bayesian algorithm after convergence. The dataset contains 22432 steps from 3138 trajectories The WT clusters (Figure 4.1F) are provided for reference (cross hairs).

mobility α state and a concomitant decrease in the β intermediate-mobility state from 50 to 12% (fig. 4.8L). We also introduced previously characterized Swi6 CD Loop-X mutations to suppress Swi6 CD domain–dependent oligomerization, which could confound our interpretations [187]. We determined that there are no quantitative differences in the mobility states associated with PAmCherry-Swi6^{2XCD}-GST Loop-X versus PAmCherry-Swi6^{2XCD}-GST constructs without the CD Loop-X mutation (fig. 4.8M). Hence, the recovery of the slow-mobility state in the case of PAmCherry-Swi6^{2XCD}-GST depends solely on CD-dependent H3K9me recognition. Since binding occurs in the context of a chimeric protein that acts as a dimer and fails to form higher-order oligomers, we concluded that four CD domains are necessary and sufficient for Swi6 to localize at sites of H3K9me.

Last, we inferred transition probabilities between the different mobility states in the case of PAmCherry-Swi6^{2XCD}-GST (Fig. 4.7F). Notably, we observed that molecules rarely exchange between the H3K9me-dependent α and β states, unlike what we detect in the case of the oligomerization-competent, WT Swi6 protein. The forward and reverse transition probabilities between the α and β states decrease approximately fourfold, suggesting that the PAmCherry-Swi6^{2XCD}-GST protein is less dynamic with regard to its chromatin-associated states. Thus, although engineered multivalent CD domains are sufficient to achieve target search *in vivo*, PAmCherry-Swi6^{2XCD}-GST molecules are less dynamic and exhibit fewer binding and unbinding transitions from chromatin. To understand the molecular basis for the enhanced binding of the Swi6^{2XCD}-GST fusion construct, we performed nucleosome binding assays. Using H3K9me0 and H3K9me3 nucleosomes, we measured the apparent binding affinity and binding specificity for recombinant WT Swi6, Swi6 CSD^{mut}, Swi6^{1XCD}-GST, and Swi6^{2XCD}-GST. Although Swi6^{2XCD}-GST binds with similar affinity to H3K9me3 nucleosomes as WT Swi6, the fusion protein exhibits a twofold increase in H3K9me3 nucleosome binding specificity (fig. 4.9).

4.2.7 Oligomerization directly competes with nucleic acid binding to promote Swi6 localization at sites of heterochromatin formation

Since an engineered multivalent CD–containing construct (Swi6^{2XCD}-GST) is less dynamic, we hypothesized that the association and dissociation of Swi6 oligomers regulate H3K9me-dependent binding and turnover from heterochromatin. We tested our hypothesis by achieving oligomerization-dependent Swi6 localization that is independent of H3K9me recognition. We made strains that express combinations of WT Swi6 and a Swi6 CD^{mut} protein that has an inactive CD domain but retains an intact CSD oligomerization domain (Fig. 4.10A). We coexpressed mNeonGreen-Swi6 protein in cells that also express PAmCherry-Swi6 CD^{mut}. Since mNeonGreen and PAmCherry emissions are spectrally distinct, we used our photoactivation approach to image single PAmCherry-Swi6 CD^{mut} molecules (red channel) after verifying the presence of discrete mNeonGreen-Swi6 foci at sites of constitutive heterochromatin (green channel). We previously showed that the lack of CD-mediated H3K9me recognition eliminates the low-mobility α state (fig. 4.4A). In contrast, our SMAUG analysis of strains coexpressing WT Swi6 and Swi6 CD^{mut} re-



	$K_{d, apparent}$ (nM)		
	H3K9me0	H3K9me3	Specificity
Swi6 WT	28.5 ± 1.83	12.1 ± 0.62	2.4
Swi6 CSD^{mut}	121.1 ± 9.86	86.3 ± 6.04	1.4
Swi6 ^{1xCD}-GST	30.1 ± 2.58	20.6 ± 1.61	1.5
Swi6 ^{2xCD}-GST	39.4 ± 3.56	10.9 ± 0.84	3.6

Figure 4.9: Quantification of two gel shift experiments using H3K9 (black circle) and H3K9me3 (pink rectangle) to determine $K_{1/2}$ and specificity ($K_{1/2}$ H3K9/$K_{1/2}$ H3K9me3) for **(A)** Swi6 WT, **(B)** Swi6 CSD^{mut}, **(C)** Swi6^{1xCD}-GST and **(D)** Swi6^{2xCD}-GST. A schematic of recombinant protein construct is included above each binding curve. Summary table of apparent $K_{1/2}$ and binding specificity for H3K9me3 mononucleosomes relative to H3K9me0 mononucleosomes for the different Swi6 and GST fusion proteins.

vealed that 5% of PAmCherry-Swi6 CD^{mut} proteins now reside in the slow-mobility α state (Fig. 4.10B). Therefore, CSD-mediated oligomerization alone can drive Swi6 localization independently of H3K9me recognition. We cannot eliminate the possibility of heterodimers consisting of WT and Swi6 CD^{mut} proteins from our measurements. However, it is unlikely that heterodimers consisting of a single, functional CD domain will contribute significantly to the observed localization patterns since at least four CD domains are required for H3K9me binding (Fig. 4.7E). Consistent with the fraction being small, PAmCherry foci are not visible in our epifluorescence data where we imaged an ensemble of PAmCherry-Swi6 CD^{mut} molecules, although mNeonGreen-Swi6 foci are intact (fig. 4.11A).

To confirm that the recovery of the α slow-mobility state is due to CSD-dependent Swi6 interactions, we coexpressed a Swi6 CSD mutant (PAmCherry-Swi6 CSD^{mut}), which is unable to oligomerize but has an intact CD domain, together with the WT mNeonGreen-Swi6 protein. The coexpression of WT Swi6 protein fails to restore any measurable occupancy of PAmCherry-Swi6 CSD^{mut} protein in the low-mobility α state (Fig. 4.10C). Suppressing nucleic acid binding results in increased occupancy of Swi6 molecules in the α and β states (Fig. 4.3C). We hypothesized that eliminating nucleic acid binding might enhance oligomerization-mediated recruitment of PAmCherry-Swi6 CD^{mut}. Therefore, we coexpressed PAmCherry-Swi6^{hinge} CD^{mut} proteins in cells that also express mNeonGreen-Swi6. Notably, we observed a substantial increase in the occupancy of PAmCherry-Swi6^{hinge} CD^{mut} molecules in the α mobility state (approximately 20%) (Fig. 4.10D). PAmCherry foci could be easily visualized in our epifluorescence data where we simultaneously imaged an ensemble of PAmCherry-Swi6^{hinge} CD^{mut} molecules and mNeonGreen-Swi6 (fig. 4.11B). Therefore, our results reveal that nucleic acid binding and Swi6 oligomerization are in direct competition. Disrupting nucleic acid binding promotes a CD-independent mode of Swi6 binding at sites of H3K9me with the CSD domain having a causal role in driving interactions between the differently labeled Swi6 molecules.

We performed a biochemical assay to directly measure the competition between nucleic acid binding, oligomerization, and H3K9me recognition using recombinant Swi6 proteins and multivalent H3K9me chromatin derived from fission yeast cell extracts as substrates. We incubated recombinant WT 3XFLAG Swi6 or 3XFLAG Swi6 L315E protein with fission yeast cell extracts. We then pulled down Swi6-bound chromatin in the WT and mutant protein context (Fig. 4.10E). In the absence of DNA, both WT Swi6 and Swi6 L315E pull down a similar amount of H3K9me chromatin as detected using an H3K9me3-specific antibody (Fig. 4.10F). Our pull-downs of H3K9me chromatin using recombinant Swi6 proteins are specific since extracts prepared using $clr4\Delta$ cells failed to recover any chromatin. We added an exogenous 1.6-kb DNA fragment and observed that the amount of H3K9me chromatin that we recovered substantially decreased in the case of Swi6 L315E (CSD^{mut}) but remains unchanged in the case of WT Swi6 (Fig. 4.10G). Hence, our observations suggest that CSD-dependent interactions that promote Swi6 oligomerization can indeed resist the effects of promiscuous nucleic acid binding.

4.3 Discussion

Our results reveal the molecular basis for how Swi6 identifies sites of H3K9me within the complex and crowded chromatin landscape of the *S. pombe* nucleus. Despite only 2% of chromatin being marked with H3K9me, Swi6 readily discriminates between modified H3K9me chromatin and unmodified H3K9me0 chromatin; we found that Swi6 binds *in vivo* to H3K9me nucleosomes with 94-fold specificity. Our numbers most closely resemble *in vitro* measurements of Swi6 binding to H3K9me3 peptides as opposed to nucleosomes [74]. Hence, modified H3K9me histone tails are the primary specificity determinants of Swi6 binding in the nucleus. The reduced specificity observed in *in vitro* studies is likely due to nucleic acid binding, which leads to Swi6-nucleosome interactions that are independent of the histone tails being modified [185]. In contrast, our *in vivo* studies reveal that nucleic acids, given their large excess in a native chromatin context, promote

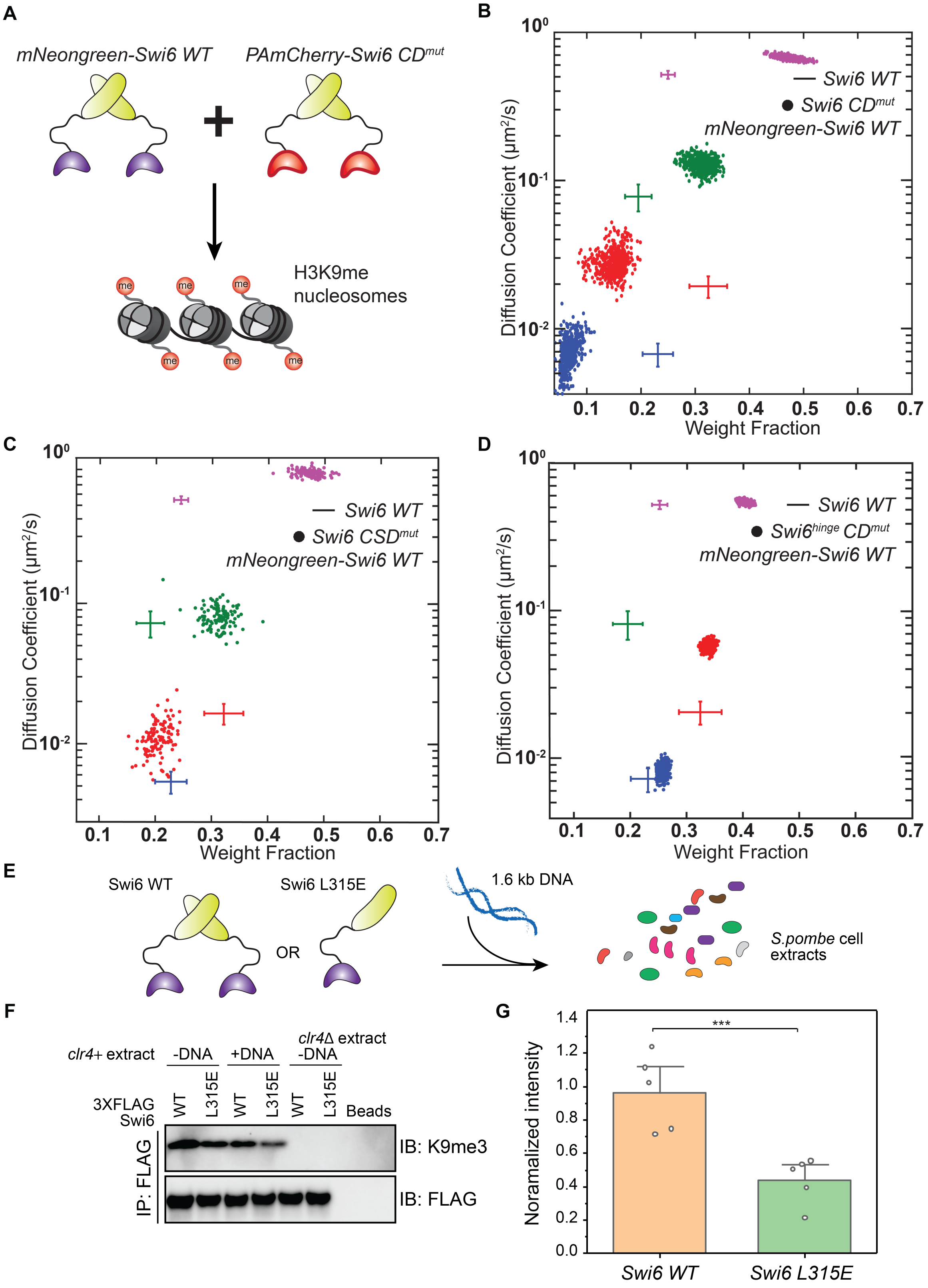

A
mNeongreen-Swi6 WT
PAmCherry-Swi6 CD^{mut}
H3K9me
nucleosomes
me
B
Diffusion Coefficient (μm^2/s)
Weight Fraction
10^0
10^{-1}
10^{-2}
0.1 0.2 0.3 0.4 0.5 0.6 0.7
— Swi6 WT
● Swi6 CD^{mut}
mNeongreen-Swi6 WT
C
Diffusion Coefficient (μm^2/s)
Weight Fraction
— Swi6 WT
● Swi6 CSD^{mut}
mNeongreen-Swi6 WT
D
Diffusion Coefficient (μm^2/s)
Weight Fraction
— Swi6 WT
● Swi6^{hinge} CD^{mut}
mNeongreen-Swi6 WT
E
Swi6 WT
Swi6 L315E
OR
1.6 kb DNA
S.pombe cell extracts
F
clr4+ extract
-DNA
+DNA
clr4Δ extract -DNA
3XFLAG Swi6
WT
L315E
WT
L315E
WT
L315E
Beads
IP: FLAG
IB: K9me3
IB: FLAG
G
Noramalized intensity
0.0 0.2 0.4 0.6 0.8 1.0 1.2 1.4
***
Swi6 WT
Swi6 L315E

Figure 4.10

Figure 4.10: **(A)** PAmCherry-Swi6 CD^{mut} and mNeonGreen-Swi6 (WT) proteins are coexpressed; mobility states are measured using single-molecule tracking of PAmCherry-Swi6 CD^{mut}. **(B)** SMAUG identifies four distinct mobility states for PAmCherry-Swi6 CD^{mut} (*swi6* W104A) in cells coexpressing mNeonGreen-Swi6. Dataset: 13,194 steps from 1900 trajectories. **(C)** SMAUG identifies three distinct mobility states for PAmCherry-Swi6 CSD^{mut} (*swi6 L315E*) in cells coexpressing mNeonGreen-Swi6; the α mobility state is absent. Dataset: 3200 steps from 1270 trajectories. **(D)** SMAUG identifies three distinct mobility states for PAmCherry- Swi6^{hinge}CD^{mut} (*swi6 KR25A W104A*) in cells coexpressing mNeonGreen-Swi6; the γ mobility state is absent. Dataset: 15,462 steps from 3250 trajectories. Each point in (B) to (D) is the average single-molecule diffusion coefficient, D, of Swi6 molecules in that state at a saved iteration of the Bayesian algorithm after convergence; the average and SD of the WT Swi6 clusters (blue α, red β, green γ, and purple δ, respectively, from Fig. 4.1F) are provided as a reference (crosshairs). **(E)** Schematic of the competition between Swi6 oligomerization and nucleic acid binding with 3XFLAG-Swi6 or 3XFLAG-Swi6 L315. **(F)** FLAG IP assay to detect histone H3 and histone H3K9me3 bound to 3XFLAG-Swi6 or 3XFLAG-Swi6 L315E using an H3K9me3 antibody in extracts from WT (*clr4+*) or H3K9me0 (*clr*4Δ) cells. **(G)** Mean intensity of H3K9me3 histones detected upon addition of DNA relative to no DNA. Error bars: SD (N = 5, ***P = 0.009, Wilcoxon rank sum test).

Swi6 unbinding by directly competing with oligomerization. Swi6 dimers are incapable of staying bound at sites of H3K9me, which, in turn, places a unique emphasis on the coordination between oligomerization and H3K9me recognition (Fig. 4.12). We propose that Swi6 oligomerization stabilizes higher-order molecular configurations consisting of at least four CDs to promote cooperative and multivalent H3K9me recognition and binding.

Unlike earlier FRAP measurements, our model-independent superresolution assessment of Swi6 diffusion identifies four distinct mobility states [178, 181, 182]. Using fission yeast mutants, we have validated the biochemical attributes associated with each mobility state. Note that in our mutants, the chromatin environment and sequence composition of Swi6 significantly change. Hence, our assignments of mobility states are based on the order of magnitude of the D value only, and the specific value of D for the same assigned state might differ between different strains; for example, $D_{avg\delta WT} = 0.51\mu m^2/s$ is lower than $D_{avg\delta H3K9me2} = 0.67\mu m^2/s$. Therefore, in our comparisons between SMAUG results

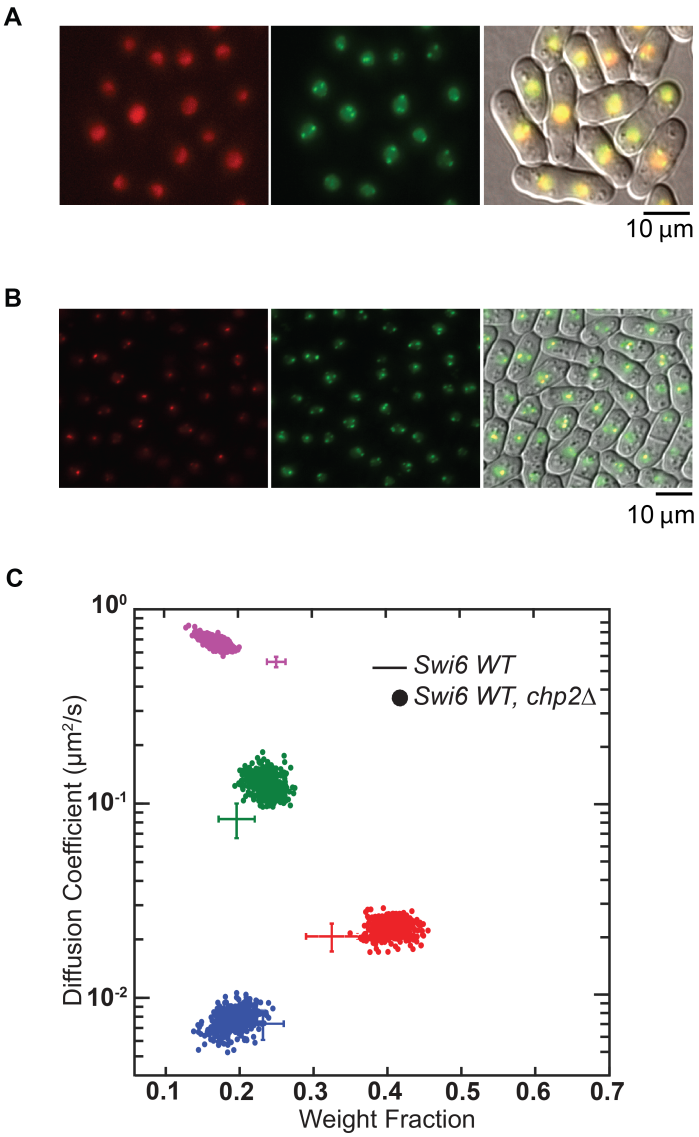


Figure 4.11: (A). Live cell imaging of PAmCherry-Swi6CD^{mut} and mNeongreen-Swi6 WT. The three images correspond to PAmCherry-Swi6CD^{mut} (561 nm excitation after 406 nm activation), mNeongreen-Swi6 WT (514 nm excitation), and overlay of the emission channels with DIC image.**(B)**. Live cell imaging of Swi6^{hinge}CD^{mut} mNeongreen-Swi6 WT. Three images correspond to 561 nm excitation after 406 nm activation, 514 nm excitation, and overlay of the two emission channels with DIC image. mNeonGreen-Swi6 and PAmCherry- Swi6 form co-localized foci in green and red emission channels with DIC image, respectively. **(C)**. Average single-molecule diffusion coefficients and weight fraction estimates for PAmCherry-Swi6 molecules expressed in *chp*2Δ cells. SMAUG identifies four distinct mobility states, (blue α, red β, green γ, and purple δ, respectively) for PAmCherry-Swi6 in *chp*2Δ cells. Each point represents the average single-molecule diffusion coefficient vs. weight fraction of PAmCherry-Swi6 molecules in each distinct mobility state at each saved iteration of the Bayesian algorithm after convergence. The dataset contains 27033 steps from 2665 trajectories. The average and standard deviation of the wild-type Swi6 clusters (Figure 4.1F) are provided as a reference (cross hairs).

of different strains, we focus on the depletion or emergence of a mobility state and on the change of weight fraction for each mobility state, not on subtle differences in D for the state in different mutants.

The primary drivers of Swi6 mobility are nucleic acid binding and weak and strong H3K9me-dependent interactions. The transition probabilities reveal how each mobility state functions to sequester or titrate Swi6 molecules, suggesting that altering their relative occupancy ultimately affects the H3K9me-bound population of the protein. Most prominently, we find that nucleic acid binding titrates Swi6 away from sites of H3K9me, while neutralizing nucleic acid binding promotes stable interactions at sites of H3K9me (α state), and it also increases the overall chromatin-bound population of the protein (β state). We propose that a significant function of Swi6 oligomerization is to counterbalance inhibitory and titratable molecular interactions that would otherwise wholly suppress Swi6 localization in cells.

Our studies highlight how the high-resolution tracking of the *in vivo* dynamics of single molecules in cells can fully recapitulate all the biochemical features of proteins despite their heterogeneous dynamics in a native chromatin context. Our studies represent a vital step toward the ultimate goal of *in vivo* biochemistry, where the on and off rates of proteins and their substrates can be reliably and directly measured in their cellular environment. Last, we used our inferred transition rates, combined with known biochemical parameters, to infer the precise chemical rate constants governing the behavior of Swi6. The main caveats of these inferred rate constants are that they cannot capture spatial effects, and they model binding reactions as pseudo first order. While modeling spatial effects will require further study, we can infer true binding rate constants from our inferred pseudo first-order rates by assuming that Swi6 is the limiting reagent of its binding reactions. Furthermore, the unbinding reactions are truly first order, and thus, our inferred rates for these reactions are true rate constants.

Preserving the same degree of nucleic acid binding but preventing oligomerization

(Swi6^{1XCD}-GST) disrupts Swi6 localization at sites of H3K9me (fig. 4.8A). Simply adding a second CD restores H3K9me-specific localization. Although multivalency represents a longstanding principle for how HP1 proteins bind to chromatin, our results suggest well-defined stoichiometric configurations that enable stable and selective H3K9me binding. Our engineered constructs reveal that four tandem CDs are both necessary and sufficient for effective H3K9me-dependent localization in cells. Although recombinant Swi6 purified from *Escherichia coli* is predominantly a dimer (83%) *in vitro*, about 10% of Swi6 molecules form tetramers [74]. On the basis of our results, we hypothesize that oligomerization and subsequent phase separation might increase the local concentration of Swi6 molecules to shift their equilibrium distribution from dimers to tetramers. In this manner, the ability of HP1 proteins to form condensates could be vital to coordinate oligomerization and H3K9me recognition, as shown in our studies. The underlying mechanisms that promote Swi6 oligomerization in our experiments remain unclear since the Swi6 Loop-X mutant exhibits virtually identical mobility states relative to WT Swi6. It is possible that the Loop-X mutation has a subtle effect *in vivo*, and additional residues are needed to fully disrupt the CD-CD oligomerization interface. An alternative model that is also consistent with our data is that chromatin offers a multivalent surface not only for CD-H3K9me interactions but also for CSD-nucleosome interactions, leading to additional modes of Swi6 oligomerization [79]. We cannot differentiate between oligomers being stabilized before H3K9me binding and oligomers being stabilized after binding to sites of H3K9me. A model where Swi6 dimers “probe” chromatin before oligomerizing would be entirely consistent with our data. We would expect that such probing would lead to likely encounters between additional Swi6 molecules at sites enriched with H3K9me.

In the case of Swi6 and its mammalian homolog, HP1α, oligomerization is essential to promote the formation of chromatin condensates that exhibit liquid-like properties [79, 169, 193]. Our results suggest that the presence of low-affinity CD domains could be one reason why some classes of HP1 proteins are proficient in oligomerization. The ability of specific HP1 isoforms to oligomerize promotes the multivalent recognition of H3K9me nucleosomes since the dimeric state of a protein such as Swi6 is insufficient for stable heterochromatin binding. Our findings raise the question of why high-affinity CD domains are not more prevalent among HP1 proteins since this could represent the most straightforward solution to the localization question. It is noteworthy that an engineered version of Swi6 with two CDs and no oligomerization exhibits reduced binding and unbinding transitions to other intermediate states from the H3K9me-dependent low-mobility α state (Fig. 4.7F). Therefore, our engineered protein constructs, once engaged at sites of H3K9me, exhibit little to no protein turnover.

On the basis of our observations, we speculate that protein recruitment that is exclusively dependent on high-affinity CD binding lacks tunability. The lack of exchange and protein turnover from chromatin would also impede subsequent downstream, CD-dependent binding events. Most notably, in *S. pombe*, the H3K9 methyltransferase (Clr4) and the second HP1 protein (Chp2) both have CD domains that recognize and bind to H3K9me and are essential for epigenetic silencing [87, 194]. Deleting Chp2 leads to a decrease in the fraction of the fast-moving Swi6 molecules and a concomitant increase in the fraction of molecules in chromatin sampling β state (fig. 4.11C). Instead, we propose that weak oligomerization and protein turnover ensure a time-sharing approach that provides opportunities for regulatory inputs either via protein-protein interactions or post-translational modifications. The formation of heterochromatin condensates, in addition to serving as mechanisms that promote epigenetic silencing through physical changes to the genome, could be fundamentally involved in shifting the equilibrium states of Swi6 oligomerization to promote efficient and highly selective H3K9me target recognition in living cells [169, 170].

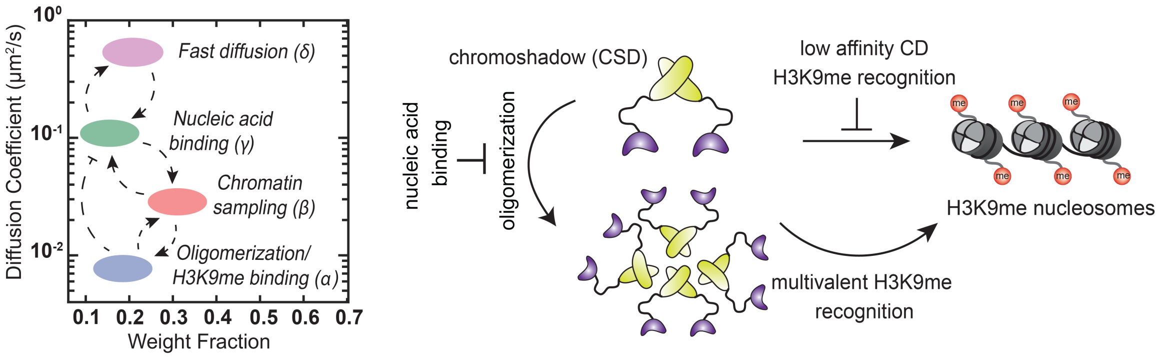


Figure 4.12: Although CDs have the requisite specificity to localize at sites of H3K9me, nucleic acid binding titrates proteins away from sites of heterochromatin formation. Oligomerization stabilizes higher-order configurations of the Swi6 CD domain to ensure rapid and efficient localization of Swi6 at sites of heterochromatin formation and outcompetes nucleic acid binding.

4.4 Materials and Methods

4.4.1 Plasmids and Strains

The plasmids and strains used in this Chapter are from the Ragunathan Lab at the University of Michigan. Please refer to corresponding manuscript of this chapter for details [106].

4.4.2 *S. pombe* live-cell imaging

Yeast strains containing a copy of PAmCherry-Swi6 or PAmCherry-Swi6 mutants under the control of the native Swi6 promoter were grown in standard complete YES media (US Biological, catalog no. Y2060) containing the full complement of yeast amino acids and incubated overnight at 32°C. The seed culture was diluted and incubated at 25°C with shaking to reach an optical density at 600 nm (OD_{600}) of 0.5. To maintain cells in an exponential phase and eliminate extranuclear vacuole formation, the culture was maintained at OD_{600} 0.5 for 2 days, with dilutions performed at 12-hour time intervals. To prepare agarose pads for imaging, cells were pipetted onto a pad of 2% agarose prepared in YES media, with 0.1 mM N-propyl gallate (Sigma-Aldrich, catalog no. P-3130) and 1% gelatin (Millipore, catalog no. 04055) as additives to reduce phototoxicity during imaging. *S. pombe* cells were imaged at room temperature (RT) with a 100× 1.40 numerical aperture (NA) oil-immersion objective in an Olympus IX-71 inverted microscope. First, the fluorescent background was decreased by exposure to 488-nm light (Coherent Sapphire, 200 W/cm^2 for 20 to 40 s). A 406-nm laser (Coherent Cube, 405-100; 102 W/cm^2) was used for photoactivation (200-ms activation time), and a 561-nm laser (Coherent Sapphire, 561-50; 163 W/cm^2) was used for imaging. Images were acquired at 40-ms exposure time per frame. The fluorescence emission was filtered with Semrock LL02-561-12.5 filter and Chroma ZT488/561rpc 488/561 dichroic to eliminate the 561-nm excitation source and imaged using a 512 × 512 pixel Photometrics Evolve EMCCD camera.

For the epifluorescence images in Fig. 4.7 and fig. 4.8, a 405-nm light-emitting diode (LED) light source (Lumencor SpectraX) at 25 mW/nm (100% power) was used to photoactivate cells, and a 561-nm LED was used to image them subsequently. Images were collected with 100-ms exposure time per frame with a 100× 1.45 NA oil-immersion objective using a Photometrics Prime95B sCMOS camera.

4.4.3 Single-molecule trajectory analysis with SMAUG algorithm

Recorded Swi6-PAmCherry single-molecule positions were detected and localized with two-dimensional Gaussian fitting with home-built MATLAB software as previously described and connected into trajectories using the Hungarian algorithm [17, 195]. These single-molecule trajectory datasets were analyzed by a nonparametric Bayesian framework to reveal heterogeneous dynamics [31]. This SMAUG algorithm uses nonparametric Bayesian statistics and Gibbs sampling to identify the number of distinct mobility states, n, in the single-molecule tracking dataset in an iterative manner. It also infers the parameter such as weight fraction, π_i, and effective diffusion coefficient, D_i, for each mobility state (i = $\ldots$,n), assuming a Brownian motion model. To ensure that even rare events would be captured, we collected more than 10,000 steps in our single-molecule tracking dataset for each measured strain, and we ran the algorithm over >10,000 iterations to achieve a thoroughly mixed state space. The state number and associated parameters were updated in each iteration of the SMAUG algorithm and saved after convergence. The final estimation (e.g., Fig. 4.1F) shows the data after convergence for iterations with the most frequent state number. Each mobility state, i, is assigned a distinct color, and for each saved iteration, the value of Di is plotted against the value of πi. The distributions of estimates over the iterations give the uncertainty in the determination of D_i. Furthermore, the transition probabilities (e.g., Fig. 4.1G) give the average probability of transitioning between states from one step to the next in any given trajectory. For static molecules from imaging fixed *S. pombe* cells, SMAUG converges to a single state with $D_{avg} = 0.007 \pm 0.001 \mu m^2/s$. The average localization error for single-molecule localizations in this fixed-cell imaging is 32.6 nm.

4.4.4 Clustering analysis for the Swi6 distributions

The spatial pattern (i.e., dispersed, clustered, or homogeneously distributed and at what scale) of each mobility state was investigated using Ripley's K function [186]

$$K(r) = \lambda^{-1} \sum_{i=1}^{n} \sum_{i \neq j} \frac{I(r_{ij} < r)}{n} \tag{4.1}$$

where r is the search radius, n is the number of points in the set, λ is the point density, and r_{ij} is the distance between the ith and jth point. $I(x)$ is an indicator function (1 when true and 0 when false). For convenience, we further normalized K(r) to attain Ripley's H function

$$H(r) = (\frac{K(r)}{\pi})^{1/2} - r \tag{4.2}$$

where $H(r) = 0$ for a random distribution, $H(r) > 0$ for a clustered distribution pattern, and $H(r) < 0$ for a dispersed pattern. The maximum of $H(r)$ approximately indicates the cluster size [196]. The cross-correlation between different states was studied with the same method. In all analysis, the nucleus was approximated as a circle to determine the area and perform edge correction [197]. We calculated $H(r)$ for each cell, and then we consolidated data from different cells into an overall $H(r)$ from the average across all cells weighted by the point density.

To eliminate effects from the intrinsic spatial correlation between steps that come from the same trajectories, we simulated diffusion trajectories with similar confined area size, average track length, and overall density as experimental trajectories by drawing step lengths from the step size distribution of the corresponding experiment steps. These trajectories are random in the initial position and step direction. We calculated a four-state $H(r)$ distribution for trajectories simulated corresponding to the Swi6 dataset (fig. 4.2C). To eliminate the contribution of the in-track autocorrelation of steps in $H(r)$, we subtracted $H(r)$ of the randomly simulated trajectories from $H(r)$ of the experimental data for each mobility state. The same $H(r)$ simulation and subtraction were carried out for all Ripley autocorrelation analyses.

4.4.5 Fine-grained chemical rate constant inference

The ine-grained chemical rate constant inference and strains used in this Chapter are from the freddolino Lab at the University of Michigan. Please refer to corresponding manuscript of this chapter for details [106].

4.4.6 *in vitro* Biochemisty assays

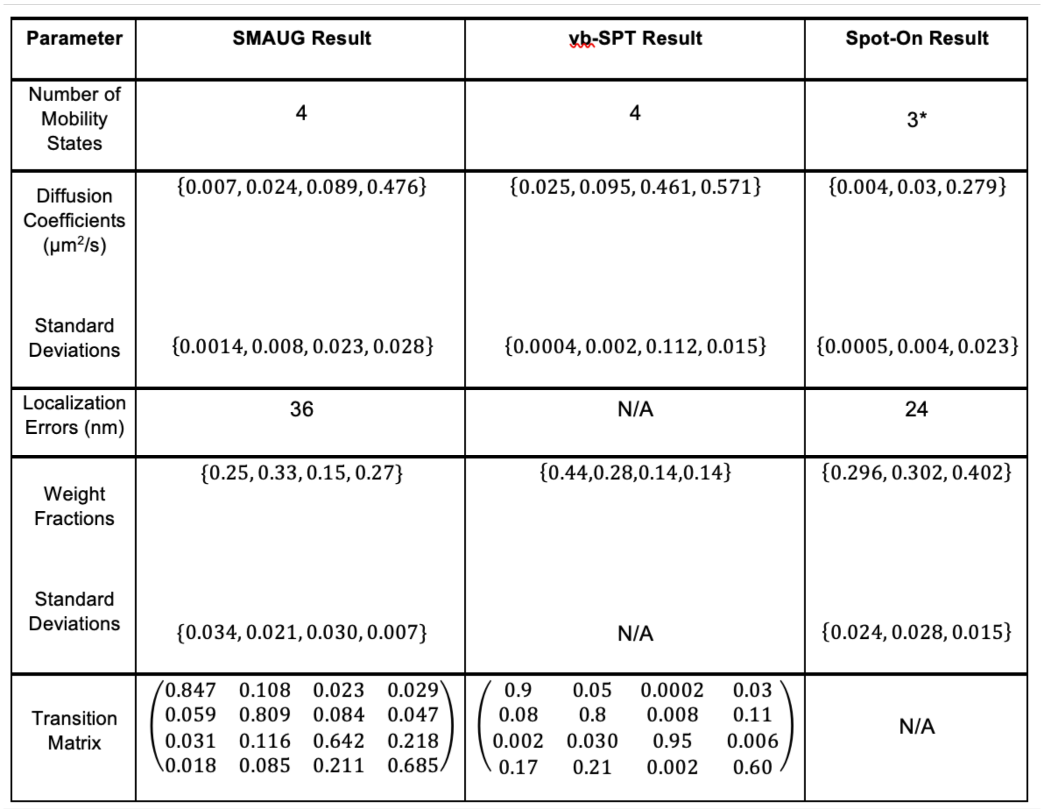
The Nucleosome electrophoretic mobility shift assays, Glutaraldehyde protein cross-linking oligomerization assay, Biochemical measurements of the competition between Swi6 oligomerization, nucleic acid binding, and H3K9me recognition and Chromatin immunoprecipitation are from the Ragunathan Lab at the University of Michigan. Please refer to corresponding manuscript of this chapter for details [106].

4.4.7 Comparison between SMAUG and other single particle tracking analysis tools

We analyzed our datasets with SMAUG as well as with two broadly used single-molecule tracking analysis tools: Spot-On and vbSPT [18, 29]. Spot-On is a probability-based kinetic modeling framework which is known to correct for multiple biases and for its user-friendly interface. Spot-On can estimate the diffusion coefficients and weight fraction for each state in the context of a model with a pre-determined number of diffusive states. However, the probability of transitioning between states cannot be acquired from this probability-based kinetic model fitting and presetting the number of states might introduce extra bias.

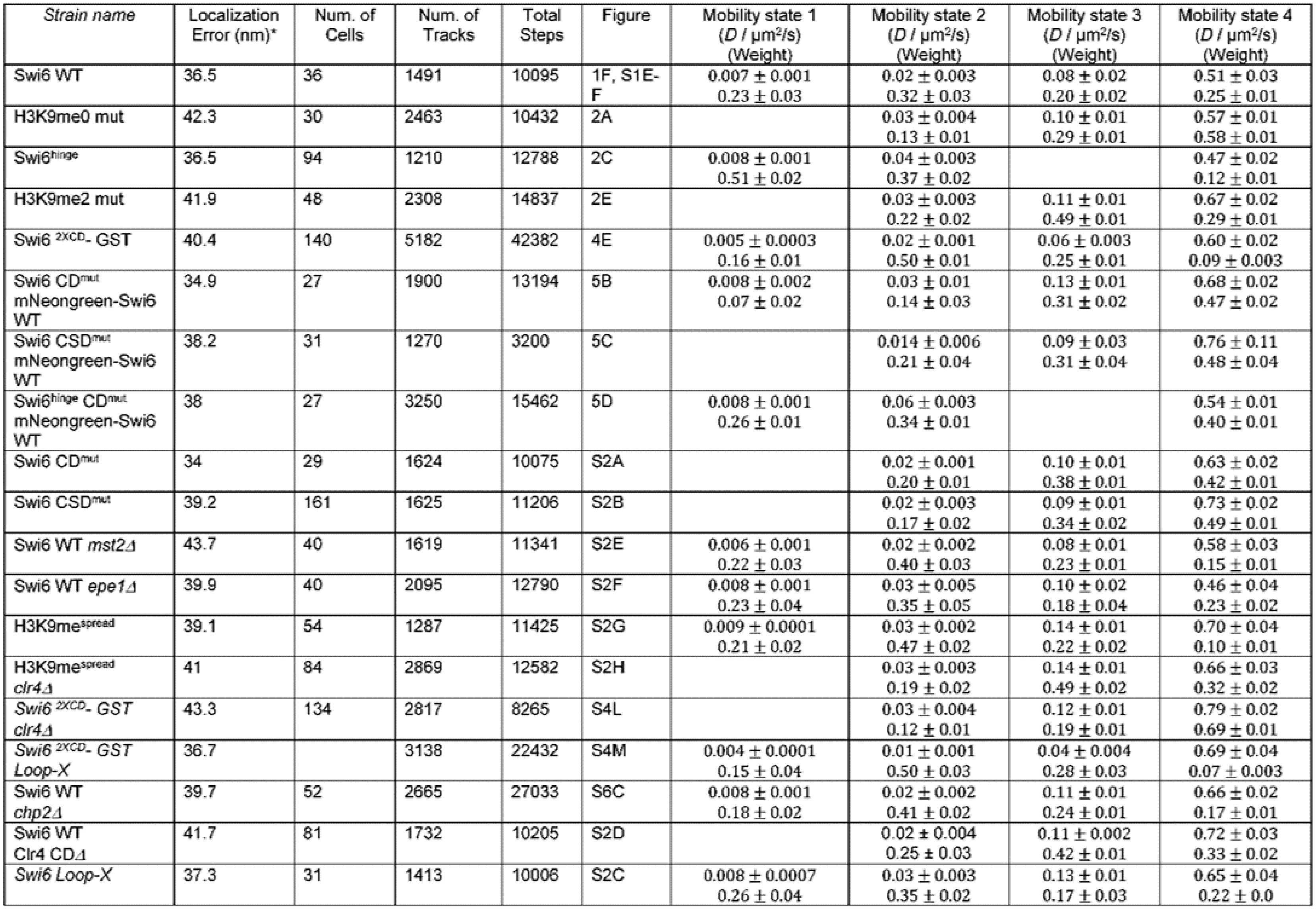
Like SMAUG, vbSPT analyzes single-molecule trajectory data within a variational Bayesian framework that can also identify the diffusive state number and estimate the diffusion coefficients, weight fractions, and transition probabilities. The major difference between SMAUG and vbSPT is that SMAUG uses a Dirichlet process-based nonparametric Bayesian framework to decide the most probable states number, while vbSPT fits the model under different state number and then uses a max-evidence criterion to select the number of states. A detailed comparison of the performance of SMAUG and vbSPT has been done previously [31]; in this case, SMAUG outperformed vbSPT in its ability to accurately estimate the parameters for a simulated dataset of trajectories containing a mixture of four diffusion coefficients. We analyzed our single-molecule trajectories of PAmCherry-Swi6 (Figure 4.1) with the three analysis methods. The nonparametric SMAUG algorithm converged to uncover four diffusive states (Figure 4.1F-G). We found that vbSPT also selected a 4-state model by its max-evidence criterion. Since Spot-On only includes a 2- or 3-state model, we fit our data to a 3-state model in this comparison. The three methods yield very similar results for the order of magnitude and the weight fractions of the slow, intermediate, and fast diffusive states, though there are some differences in the identification of each population.

For example, the two fastest terms in the vbSPT results have similar D values to the single fastest term in the SMAUG result, but the total occupancy of the two vbSPT fast terms (π = 0.28) is similar to the weight fraction of the fastest term in SMAUG (π = 0.27). Considering the large standard deviation for the fast term diffusion coefficient in vbSPT, vbSPT appears to have oversampled the fast term to produce two terms. In addition, both Spot-On and SMAUG estimate a very slow (D < 0.01 $\mu m^2/s$) state which is missing in the vbSPT results. Overall, this comparison of the PAmCherry-Swi6 results for different analysis methods leads to similar conclusions about the advantages of using the SMAUG algorithm [31] (1) Unlike Spot-on which need a predetermined number of states, SMAUG can objectively determine the number of diffusive states using nonparametric Bayesian statistics; and (2) SMAUG performs better than vbSPT in terms of fully resolving a mixture of diffusive states and their associated parameters (Table 4.1). Results for analysis of our set of single-molecule trajectories of PAmCherry- Swi6 in *S. pombe* (Figure 4.1) with the three different analysis methods.



Parameter	SMAUG Result	vb-SPT Result	Spot-On Result
Number of Mobility States	4	4	3*
Diffusion Coefficients ($\mu m^2/s$) Standard Deviations	$\{0.007, 0.024, 0.089, 0.476\}$ $\{0.0014, 0.008, 0.023, 0.028\}$	$\{0.025, 0.095, 0.461, 0.571\}$ $\{0.0004, 0.002, 0.112, 0.015\}$	$\{0.004, 0.03, 0.279\}$ $\{0.0005, 0.004, 0.023\}$
Localization Errors (nm)	36	N/A	24
Weight Fractions Standard Deviations	$\{0.25, 0.33, 0.15, 0.27\}$ $\{0.034, 0.021, 0.030, 0.007\}$	$\{0.44,0.28,0.14,0.14\}$ N/A	$\{0.296, 0.302, 0.402\}$ $\{0.024, 0.028, 0.015\}$
Transition Matrix	$\begin{pmatrix} 0.847 & 0.108 & 0.023 & 0.029 \\ 0.059 & 0.809 & 0.084 & 0.047 \\ 0.031 & 0.116 & 0.642 & 0.218 \\ 0.018 & 0.085 & 0.211 & 0.685 \end{pmatrix}$	$\begin{pmatrix} 0.9 & 0.05 & 0.0002 & 0.03 \\ 0.08 & 0.8 & 0.008 & 0.11 \\ 0.002 & 0.030 & 0.95 & 0.006 \\ 0.17 & 0.21 & 0.002 & 0.60 \end{pmatrix}$	N/A

Table 4.1: *Spot-On was fixed to 3 states due to limitations in the original code



Strain name	Localization Error (nm)*	Num. of Cells	Num. of Tracks	Total Steps	Figure	Mobility state 1 (*D* / μm^2/s) (Weight)	Mobility state 2 (*D* / μm^2/s) (Weight)	Mobility state 3 (*D* / μm^2/s) (Weight)	Mobility state 4 (*D* / μm^2/s) (Weight)
Swi6 WT	36.5	36	1491	10095	1F, S1E-F	0.007 ± 0.001 0.23 ± 0.03	0.02 ± 0.003 0.32 ± 0.03	0.08 ± 0.02 0.20 ± 0.02	0.51 ± 0.03 0.25 ± 0.01
H3K9me0 mut	42.3	30	2463	10432	2A		0.03 ± 0.004 0.13 ± 0.01	0.10 ± 0.01 0.29 ± 0.01	0.57 ± 0.01 0.58 ± 0.01
Swi6^{hinge}	36.5	94	1210	12788	2C	0.008 ± 0.001 0.51 ± 0.02	0.04 ± 0.003 0.37 ± 0.02		0.47 ± 0.02 0.12 ± 0.01
H3K9me2 mut	41.9	48	2308	14837	2E		0.03 ± 0.003 0.22 ± 0.02	0.11 ± 0.01 0.49 ± 0.01	0.67 ± 0.02 0.29 ± 0.01
Swi6 ^{2XCD}- GST	40.4	140	5182	42382	4E	0.005 ± 0.0003 0.16 ± 0.01	0.02 ± 0.001 0.50 ± 0.01	0.06 ± 0.003 0.25 ± 0.01	0.60 ± 0.02 0.09 ± 0.003
Swi6 CD^{mut} mNeongreen-Swi6 WT	34.9	27	1900	13194	5B	0.008 ± 0.002 0.07 ± 0.02	0.03 ± 0.01 0.14 ± 0.03	0.13 ± 0.01 0.31 ± 0.02	0.68 ± 0.02 0.47 ± 0.02
Swi6 CSD^{mut} mNeongreen-Swi6 WT	38.2	31	1270	3200	5C		0.014 ± 0.006 0.21 ± 0.04	0.09 ± 0.03 0.31 ± 0.04	0.76 ± 0.11 0.48 ± 0.04
Swi6^{hinge} CD^{mut} mNeongreen-Swi6 WT	38	27	3250	15462	5D	0.008 ± 0.001 0.26 ± 0.01	0.06 ± 0.003 0.34 ± 0.01		0.54 ± 0.01 0.40 ± 0.01
Swi6 CD^{mut}	34	29	1624	10075	S2A		0.02 ± 0.001 0.20 ± 0.01	0.10 ± 0.01 0.38 ± 0.01	0.63 ± 0.02 0.42 ± 0.01
Swi6 CSD^{mut}	39.2	161	1625	11206	S2B		0.02 ± 0.003 0.17 ± 0.02	0.09 ± 0.01 0.34 ± 0.02	0.73 ± 0.02 0.49 ± 0.01
Swi6 WT *mst2Δ*	43.7	40	1619	11341	S2E	0.006 ± 0.001 0.22 ± 0.03	0.02 ± 0.002 0.40 ± 0.03	0.08 ± 0.01 0.23 ± 0.01	0.58 ± 0.03 0.15 ± 0.01
Swi6 WT *epe1Δ*	39.9	40	2095	12790	S2F	0.008 ± 0.001 0.23 ± 0.04	0.03 ± 0.005 0.35 ± 0.05	0.10 ± 0.02 0.18 ± 0.04	0.46 ± 0.04 0.23 ± 0.02
H3K9me^{spread}	39.1	54	1287	11425	S2G	0.009 ± 0.0001 0.21 ± 0.02	0.03 ± 0.002 0.47 ± 0.02	0.14 ± 0.01 0.22 ± 0.02	0.70 ± 0.04 0.10 ± 0.01
H3K9me^{spread} *clr4Δ*	41	84	2869	12582	S2H		0.03 ± 0.003 0.19 ± 0.02	0.14 ± 0.01 0.49 ± 0.02	0.66 ± 0.03 0.32 ± 0.02
Swi6 ^{2XCD}- GST clr4Δ	43.3	134	2817	8265	S4L		0.03 ± 0.004 0.12 ± 0.01	0.12 ± 0.01 0.19 ± 0.01	0.79 ± 0.02 0.69 ± 0.01
Swi6 ^{2XCD}- GST Loop-X	36.7		3138	22432	S4M	0.004 ± 0.0001 0.15 ± 0.04	0.01 ± 0.001 0.50 ± 0.03	0.04 ± 0.004 0.28 ± 0.03	0.69 ± 0.04 0.07 ± 0.003
Swi6 WT *chp2Δ*	39.7	52	2665	27033	S6C	0.008 ± 0.001 0.18 ± 0.02	0.02 ± 0.002 0.41 ± 0.02	0.11 ± 0.01 0.24 ± 0.01	0.66 ± 0.02 0.17 ± 0.01
Swi6 WT Clr4 CDΔ	41.7	81	1732	10205	S2D		0.02 ± 0.004 0.25 ± 0.03	0.11 ± 0.002 0.42 ± 0.01	0.72 ± 0.03 0.33 ± 0.02
Swi6 Loop-X	37.3	31	1413	10006	S2C	0.008 ± 0.0007 0.26 ± 0.04	0.03 ± 0.003 0.35 ± 0.02	0.13 ± 0.01 0.17 ± 0.03	0.65 ± 0.04 0.22 ± 0.0

Table 4.2: All SMAUG results for used *S. pombe* strains

CHAPTER V

H3K9 methylation Enhances HP1-associated Epigenetic Silencing Complex Assembly and Suppresses Off-Chromatin Binding

The work presented in this chapter is in submission

Chen, Z., Seman, M., Farhat, A., Fyodorova, Y., Biswas, S., Levashkevich, A.,
Freddolino, P.L., Biteen, J.S. and Ragunathan, K.
H3K9 methylation enhances HP1-associated epigenetic silencing complex assembly and suppresses off-chromatin binding.
bioRxiv, 2023.03.08.531771. DOI: 10.1101/2023.03.08.531771

In this work, I conceptualized the proposed mechanism in the project and designed the needed experiments and strains. I performed single-molecule imaging experiments and dynamics analysis. I designed and implemented the single-molecule dynamics and spatial analysis. I plotted the single-molecule dynamics analysis, reconstructed localization images, and estimated the kinetics rate of target proteins.

5.1 Introduction

Genetically identical cells can exhibit different phenotypic characteristics due to the covalent modification of DNA packaging proteins called histones [161]. One modification, Histone H3 lysine 9 methylation (H3K9me), is enriched within non-transcribed regions of the genome, called heterochromatin [89]. Heterochromatin is important for maintaining the integrity of the genome, silencing repetitive DNA sequences, and maintaining cell identity [54]. The function of epigenetic modifications like H3K9 methylation relies on the actions of specific proteins called histone modifiers. These include "writer" proteins that add modifications to histones, "reader" proteins that recognize and bind to these modifications, and "eraser" proteins that remove these modifications [161]. Histone modifiers often form large multi-protein complexes with other accessory factors to regulate chromatin structure, genome organization, and transcription [54, 57].

H3K9 methylation acts as a binding platform for the recruitment of a conserved family of proteins called HP1 [89]. HP1 proteins play multiple roles in forming heterochromatin [168, 198]. This includes the recruitment of histone modifiers that catalyze H3K9 methylation deposition and spreading across large chromosomal regions, chromatin compaction through oligomerization, and epigenetic inheritance after DNA replication [73]. HP1 proteins recognize H3K9 methylation through a conserved domain called the chromodomain (CD) and interacts with its binding partners through a second protein domain called the chromoshadow domain (CSD) [168, 198]. Our current understanding is that H3K9 methylated chromatin simply acts as a scaffold that recruits HP1 and its partner proteins to silence transcription [199].

In the model organism *Schizosaccharomyces pombe (S. pombe)*, H3K9 methylation is enriched at the pericentromeric repeats, telomeres, and the mating type locus. The protein Clr4 is responsible for adding methyl groups to H3K9me, to create a binding site for HP1 proteins [200]. Two HP1 orthologs bind to H3K9me in *S. pombe*, Swi6, and Chp2 [201–203]. Despite their structural similarity and shared evolutionary origin, Swi6 and Chp2 are expressed at very different levels in the cell and have distinct roles in heterochromatin formation [87]. Swi6 is expressed at levels that are at least 100 times higher than Chp2 in cells [194]. *In vitro* studies of Swi6 and Chp2 capture key biochemical features associated with oligomerization and their interaction with mononucleosomes and oligonucleosomes [73, 82, 194]. Such studies have shown that Swi6 and Chp2 have similar tendencies to form dimers and oligomers [166], but Swi6 binds more strongly to nucleosomes (approximately 3-fold higher) compared to Chp2 [171]. However, it is unclear how these significant differences in expression levels and binding affinities between Swi6 and Chp2 extend to how both proteins interact with H3K9me in living cells.

Deletions of Swi6 and Chp2 have additive effects on epigenetic silencing. These observations suggest that Swi6 and Chp2 have distinct roles in establishing heterochromatin [82, 194]. The two HP1 proteins preferentially interact with different binding partners. Epe1 is a putative H3K9 demethylase that opposes H3K9 methylation [81] and interacts with Swi6 both *in vitro* and *in vivo* [58, 80]. On the other hand, the SHREC complex in *S. pombe*, which consists of two major chromatin modifying enzymes - Clr3, a histone deacetylase, and Mit1, a chromatin remodeler – preferentially forms complexes with Chp2 [81, 83, 86, 204]. Mit1 interacts with Chp2 both *in vitro* and *in vivo.* It is still unclear how Swi6 and Chp2 specifically and selectively recruit their respective binding partners to sites of heterochromatin formation. One possibility is that Swi6 and Chp2 first form a complex with their partner proteins (Epe1, Mit1, or Clr3) off-chromatin, then search the genome, and ultimately bind at sites that are enriched for H3K9 methylation. Cells lacking Clr4 and H3K9 methylation exhibit a significant loss in HP1-mediated protein interactions, suggesting that chromatin may play a causal role in enabling protein-protein interactions in living cells [171]. Hence, an alternative model is that HP1 proteins form complexes with their binding partners at sites of heterochromatin formation rather than off-chromatin.

Immunoprecipitation followed by mass spectrometry (IP-MS) is useful to detect protein-

protein interactions. Yet, the types of interactions they detect can vary depending on factors such as lysis conditions, salt concentrations, and protein abundance. As a result, these assays may not fully represent the range of interactions that occur between proteins in living cells. Detecting protein-protein interactions after lysing cells also removes proteins from their native, complex, and crowded chromatin environment. This leads to chromatin associated factors exhibiting divergent properties *in vitro* versus how they behave in cells. For example, in the case of Swi6, nucleosome binding *in vivo* is inhibited, not enhanced, by interactions with nucleic acids, unlike *in vitro* binding assays. This is because the large excess of DNA in the cell can displace Swi6 from its binding site and promote protein turnover at sites of heterochromatin formation [106]. Attempts to bridge the gap between *in vitro* and *in vivo* studies such as FRET and two color imaging measurements rely on protein-protein interactions that are infrequent and transient given the dynamic properties of chromatin binding proteins. Additionally, FRET poses critical methodological challenges due to its limited working distance (< 0 nm) and the rare chances of spontaneous interactions between labeled molecules [205].

Single-molecule microscopy of target protein-photoactivatable fluorescent protein fusions is a powerful tool to study protein dynamics *in vivo* [206, 207]. Live-cell imaging can access the interaction of histone modifiers with their chromatin substrates [106, 208, 209]. When combined with critical advances in statistical inference methods for analyzing high spatiotemporal resolution imaging data (Bayesian statistics applied to single-particle tracking) [31, 210], we can map the biophysical mobility states of proteins (as measured by their diffusion coefficients) to their biochemical properties in living cells [106]. Furthermore, analyzing the probabilities of transitioning between or dissociating from each detected mobility state can provide estimates of the biochemical properties for protein-protein and protein-chromatin interactions in cells [189].

Here, we use single-molecule tracking photoactivated localization microscopy to study the dynamics and interactions of the HP1 proteins—Swi6 and Chp2—and the proteins they form complexes with—Epe1, Mit1, and Clr3. Our goal here is to determine whether the mobility states of proteins can be used to infer how they form complexes within the context of native chromatin. Based on a combination of single particle tracking measurements and mathematical modeling, we propose a mechanism for how H3K9 methylation not just encourages specific complex formation between HP1 proteins and their interactors but also suppresses the propensity of heterochromatin-associated proteins to form off-chromatin complexes. As opposed to an inert platform or scaffold to direct HP1 binding, our study rebrands chromatin as an active participant in enhancing HP1 mediated complex formation in living cells.

5.2 Results

5.2.1 *S. pombe* HP1 orthologs, Swi6 and Chp2 exhibit distinct, non-overlapping biophysical states in living cells

We previously used single-molecule tracking to identify biophysical diffusive states that map to distinct biochemical properties of proteins in living cells [106]. We measured the *in vivo* dynamics of Swi6, one of two HP1 proteins in fission yeast. We have determined that Swi6 has four distinct mobility states each of which maps to a specific biochemical property in cells [106]. Here, we measured the mobility states of the second conserved HP1 protein, Chp2. We labeled the N-terminus of the endogenous copy of Chp2 with PAmCherry (PAmCherry-Chp2) but were unable to observe an appreciable number of photoactivation events for single particle tracking likely due to its low expression level. Instead, we inserted a second copy of an N-terminally labeled PAmCherry-Chp2 under the regulation of a thiamine-repressible promoter, *nmt1*, *nmt41* and *nmt81* (Figure 5.1A, B). We first ensured that nmt81-dependent expression of PAmCherry-Chp2 complements *chp2*Δ cells by measuring the silencing of a *ura4+* reporter inserted at the mat locus (*Kint2*). If *ura4+* is silenced, cells grow on 5-fluoroorotic acid (EMMC+FOA) containing media and

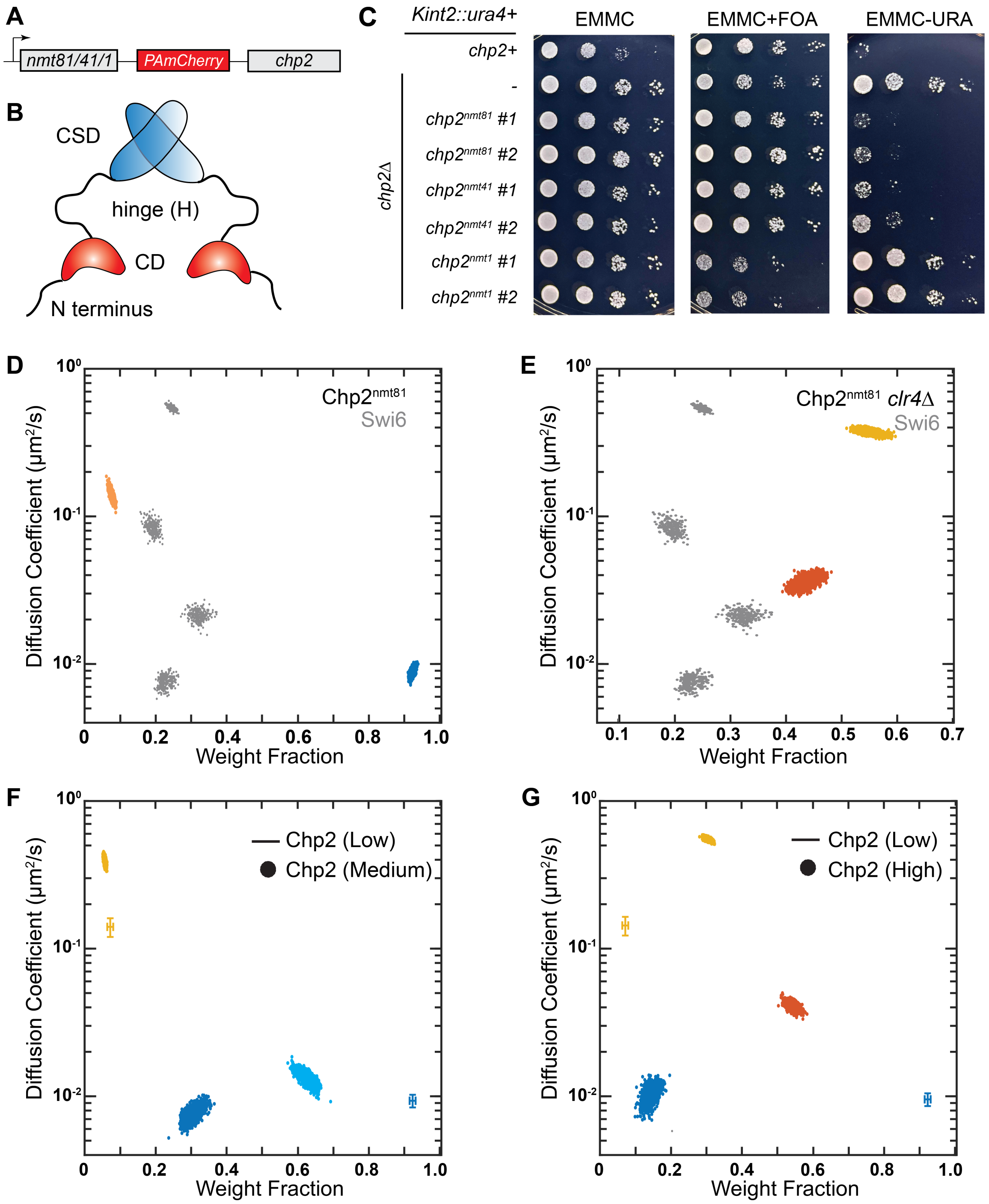

A
nmt81/41/1
PAmCherry
chp2
B
CSD
hinge (H)
CD
N terminus
C
Kint2::ura4+
EMMC
EMMC+FOA
EMMC-URA
chp2+
-
chp2^{nmt81} #1
chp2^{nmt81} #2
chp2^{nmt41} #1
chp2^{nmt41} #2
chp2^{nmt1} #1
chp2^{nmt1} #2
chp2Δ
D
Chp2^{nmt81}
Swi6
Diffusion Coefficient (μm^2/s)
10^0
10^{-1}
10^{-2}
0
0.2
0.4
0.6
0.8
1.0
Weight Fraction
E
Chp2^{nmt81} clr4Δ
Swi6
Diffusion Coefficient (μm^2/s)
10^0
10^{-1}
10^{-2}
0.1
0.2
0.3
0.4
0.5
0.6
0.7
Weight Fraction
F
Chp2 (Low)
Chp2 (Medium)
Diffusion Coefficient (μm^2/s)
10^0
10^{-1}
10^{-2}
0
0.2
0.4
0.6
0.8
1.0
Weight Fraction
G
Chp2 (Low)
Chp2 (High)
Diffusion Coefficient (μm^2/s)
10^0
10^{-1}
10^{-2}
0
0.2
0.4
0.6
0.8
1.0
Weight Fraction

Figure 5.1

Figure 5.1: A: Design of the PAmCherry-Chp2 construct: PAmCherry is fused to the N-terminus of Chp2 and expressed from ectopically using a series of inducible promoters- *nmt1*, *nmt41*, or *nmt81*. B: Schematic representation of Chp2 domains. CD: chromodomain (H3K9me recognition); H: hinge (nucleic acid binding); CSD: chromoshadow domain (dimerization interface). C: Silencing assay using an *ura4+* reporter inserted at the mat locus (*Kint2:ura4*). 10-fold serial dilutions of cells expressing Chp2 from different nmt promoter variants were plated on EMMC, EMMC+FOA and EMM-URA plates. D-E: NOBIAS identifies two distinct mobility states for PAmCherry-Chp2^{nmt81}. Each colored point is the average single-molecule diffusion coefficient of PAmCherry-Chp2 molecules in that state sampled from the posterior distribution of NOBIAS inference at a saved iteration after convergence in WT cells (D) and $clr4\Delta$ cells (E). Grey points are the previously reported PAmCherry-Swi6 single-molecule dynamics [106]. F-G: NOBIAS identifies multiple mobility states for PAmCherry-Chp2^{nmt41} (medium expression, F) and PAmCherry-Chp2^{nmt1} (high expression, G). Each colored point is the average single-molecule diffusion coefficient sampled from the posterior distribution for PAmCherry-Chp2 at the indicated expression level. Colored line crosses represent the data from PAmCherry-Chp2^{nmt81} (low expression; data in D).

fail to grow on media lacking uracil (EMM-URA). We noted that *nmt81*-PAmCherry-Chp2 is functional and successfully restores *ura4+* silencing in $chp2\Delta$ cells (Figure 5.1C).

Next, we tracked individual PAmCherry-Chp2 molecules expressed from the *nmt81* promoter in *S. pombe*. PAmCherry-Chp2 was briefly photoactivated with 405-nm laser light and imaged with 561-nm laser excitation light. We repeated this measurement until all PAmCherry-Chp2 molecules that can be activated were photobleached (see Methods). The activation-excitation-imaging cycle was repeated approximately $10-20$ times for each cell, and the single molecules were localized and tracked in the recorded fluorescence movies with the SMALL-LABS algorithm [12]. We model the motion of Chp2 molecules inside the *S. pombe* nucleus as a diffusive process and thus can assign diffusion coefficients to quantify the different mobility states associated with Chp2. We define a mobility state as a subpopulation of molecules with a distinct diffusion coefficient (D). In contrast to the four mobility states that we observed in the case of PAmCherry-Swi6 (a mixture of stationary and mobile molecules), nearly all PAmCherry-Chp2 proteins in *S. pombe* are stationary (Figure 5.2B). Hence, our live cell imaging data reveals a substantially different binding configuration between Swi6 and Chp2.

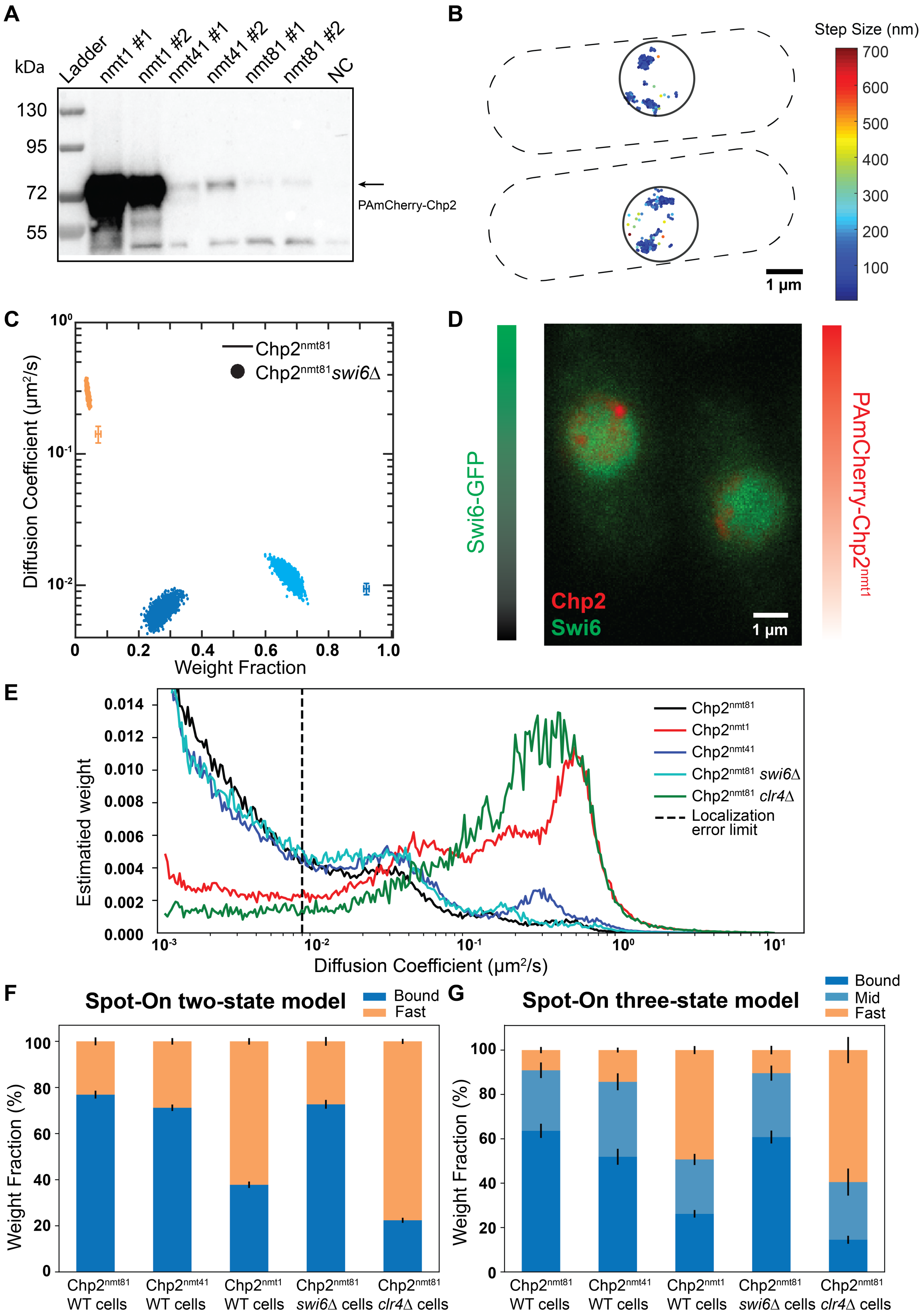

A
kDa
Ladder
nmt1 #1
nmt1 #2
nmt41 #1
nmt41 #2
nmt81 #1
nmt81 #2
NC
130
95
72
55
PAmCherry-Chp2
B
Step Size (nm)
700
600
500
400
300
200
100
1 μm
C
Chp2nmt81
Chp2nmt81swi6Δ
Diffusion Coefficient (μm²/s)
10⁰
10⁻¹
10⁻²
0
0.2
0.4
0.6
0.8
1.0
Weight Fraction
D
Swi6-GFP
Chp2
Swi6
1 μm
PAmCherry-Chp2nmt1
E
Estimatied weight
0.014
0.012
0.010
0.008
0.006
0.004
0.002
0.000
10⁻³
10⁻²
10⁻¹
10⁰
10¹
Diffusion Coefficient (μm²/s)
Chp2nmt81
Chp2nmt1
Chp2nmt41
Chp2nmt81 swi6Δ
Chp2nmt81 clr4Δ
Localization error limit
F
Spot-On two-state model
Bound
Fast
Weight Fraction (%)
100
80
60
40
20
0
Chp2nmt81 WT cells
Chp2nmt41 WT cells
Chp2nmt1 WT cells
Chp2nmt81 swi6Δ cells
Chp2nmt81 clr4Δ cells
G
Spot-On three-state model
Bound
Mid
Fast
Weight Fraction (%)
100
80
60
40
20
0
Chp2nmt81 WT cells
Chp2nmt41 WT cells
Chp2nmt1 WT cells
Chp2nmt81 swi6Δ cells
Chp2nmt81 clr4Δ cells

Figure 5.2

Figure 5.2: 1A The expression level of PAmCherry-Chp2$^{nmt1/nmt41/nmt81}$ is quantified by western blot using an mCherry antibody. The cross-reactivity of the mCherry antibody can specifically detect PAmCherry protein fusions. 1B Single-molecule step size map for PAmCherry-Chp2^{nmt81}. Dashed lines: approximate *S. pombe* cell outlines; solid circles: approximate nucleus borders. 1C: NOBIAS identifies distinct mobility states for PAmCherry-Chp2^{nmt81} in $swi6\Delta$ cells. Each colored point is the average single-molecule diffusion coefficient of molecules in that state sampled from the posterior distribution of NOBIAS inference at a saved iteration after convergence. The colored crosses show the data for PAmCherry-Chp2^{nmt81} in WT cells (Figure 5.1D). 1D: Two-color imaging of cells with Swi6-GFP expressed from the endogenous promoter and PAmCherry-Chp2^{nmt1}. Green colorbar: Swi6-GFP intensities; Red colorbar: reconstructed PAmCherry-Chp2 density map. Both color channels are normalized to the maximum pixel intensity. E: Posterior distribution of diffusion coefficients of single-molecule trajectory datasets inferred from DPSP analysis(2). Vertical dashed line: Lower bound of detectable diffusion coefficients given the experimental localization error. 1F-1G: Weight fractions of each mobility state for PAmCherry-Chp2^{nmt81} single-molecule trajectories inferred from Spot-On analysis(3) with a two-state model (F) and a three-state model (G).

To investigate any potential heterogeneity in the dynamics within the observed static molecules, we applied NOBIAS, a nonparametric Bayesian framework that can objectively determine the number of mobility states giving rise to a single-molecule tracking dataset [210]. We identified two mobility states associated with PAmCherry-Chp2: over 92% of the Chp2 molecules are in the low mobility state with an average diffusion coefficient, $D_{\alpha,Chp2}$ = 0.007 $\mu m^2/s$ (Figure 5.1D) and around 7.5% of Chp2 molecules in a fast mobility state with $D_{\delta,Chp2}$ = 0.13 $\mu m^2/s$. NOBIAS analysis also provides the probability of a molecule transitioning between two mobility states within its trajectory: Chp2 molecules in the fast mobility δ state are much more likely to transition to the slower state compared with the reverse transition (Figure 5.3A). These weight fractions and transition probabilities indicate that Chp2 molecules predominantly occupy the slow α mobility state and only a very small proportion of Chp2 molecules occupy the fast δ state. This slow Chp2 motion is very different from the motion of the second *S. pombe* HP1 protein, Swi6: similar Bayesian analysis using the SMAUG package found that Swi6 molecules are distributed across 4 distinct mobility states [106].

To determine if the dominant slow mobility state of Chp2 corresponds to H3K9me-bound Chp2, we deleted Clr4, the only H3K9 methyltransferase in *S. pombe* [72,173]. In a $clr4\Delta$ background, the slowest PAmCherry-Chp2 mobility state is completely absent (Figure 5.1E). PAmCherry-Chp2 molecules in $clr4\Delta$ cells switch over to the fast mobility state consistent with Chp2 proteins moving around the nucleus in an unconstrained manner (weight fraction= 56%, D_{fast} = 0.36 $\mu m^2/s$). In addition, we observed a new mobility state that we did not previously detect in *clr4+* cells (weight fraction=44%, D_{int} = 0.03 $\mu m^2/s$). The new mobility state most closely matches the chromatin sampling (β state) that we previously observed in the case of Swi6. Therefore, without H3K9 methylation, Chp2 exhibits a substantial degree of binding to unmethylated chromatin. In contrast, only 10% of Swi6 molecules are in a chromatin sampling configuration with >60% of Swi6 molecules exhibiting fast, unconstrained diffusion in $clr4\Delta$ cells.

The overarching goal of our studies is to measure the biochemical properties of proteins and how they form complexes in the context of living cells. The appearance of a new mobility state in $clr4\Delta$ cells led us to hypothesize that Chp2 protein molecules that dissociate from H3K9me engage in a substantial degree of promiscuous off-target interactions. Unlike an *in vitro* experiment, we cannot change concentrations of proteins incrementally to determine binding affinities and specificities between proteins and their cognate ligands. Instead, we used two additional *nmt* promoter variants (*nmt41* and *nmt1*) to alter the overall Chp2 levels in wild-type cells. We used western blots to quantify the differences in expression across the three promoters. The difference between Chp2 expression driven by *nmt41* and *nmt81* is approximately 50-fold (Figure 5.2A). In contrast, the expression level of Chp2 is at over 1000 fold higher when expressed from an *nmt1* promoter compared to *nmt81*-driven expression (Figure 5.2A). Hence, the promoter variants give us a substantial dynamic range in terms of Chp2 concentration to assess whether the mobility states we observed are in any way limited by substrate availability (H3K9 methylated nucleosomes in *clr4+* cells).

The dynamics of the medium expressed (50 fold higher) PAmCherry-Chp2^{nmt41} are slightly increased compared to low expression *nmt81* driven PAmCherry-Chp2. The slow diffusive state in the low expressing PAmCherry-Chp2^{nmt81} cells splits into two states in the medium expression PAmCherry-Chp2^{nmt41} cells with D_{slow1} =0.005 $\mu m^2/s$ and D_{slow2} =0.010 $\mu m^2/s$ (Figure 5.1F). In contrast, in the highly expressed PAmCherry-Chp2^{nmt1} (over 1000 fold compared to nmt81), we observe that only 15% of Chp2 is in the slow diffusive state; the remaining Chp2 molecules are in intermediate and fast states (Figure 5.1G). The new state we observed in the case of the high expression PAmCherry-Chp2 cells is comparable to Chp2 dynamics in a $clr4\Delta$ background- where Chp2 has no substrate. In this background, Chp2 molecules adopt a new mobility state that most closely resembles the chromatin sampling β state we observed previously in our Swi6 single particle measurements. Our *ura4+* reporter based silencing assays revealed that PAmCherry-Chp2 expressed from an *nmt41* promoter (PAmCherry-Chp2^{nmt41}) preserves *ura4+* reporter gene silencing whereas PAmCherry-Chp2 expressed from the high expression nmt1 promoter (PAmCherry-Chp2^{nmt1}) disrupts silencing (Figure 5.1C) [194]. This is possibly because Chp2 outcompetes other chromodomain containing proteins for a limiting amount of H3K9me substrate. Hence, maintaining the equilibrium of Chp2 in a low diffusion state (H3K9me dependent) preserves its heterochromatin-associated silencing functionality.

Swi6 and Chp2 both have a chromodomain that is responsible for H3K9me binding specificity (Figure 5.1B). We asked to what extent Swi6 competes with Chp2 to bind to H3K9 methylated nucleosomes. We imaged PAmCherry-Chp2 in cells lacking the major HP1 protein, Swi6 ($swi6\Delta$). Like in WT cells, the majority of Chp2 molecules in $swi6\Delta$ cells exhibit slow mobility. However, unlike WT cells, the slow population is split into two distinct slow mobility states with D_{slow1} =0.005 $\mu m^2/s$ and D_{slow2} =0.010 $\mu m^2/s$, with only a very small portion in the fast state (Figure 5.2C). Similar to WT cells, the fast Chp2 state is very unstable: there is a high probability of transitioning from the fast state to one of the faster slow states (Figure 5.4A). The appearance of a split new mobility state is likely because deleting Swi6 disrupts epigenetic silencing or because Swi6 makes unknown contributions to stabilizing Chp2 binding, although our results cannot distinguish between these two possibilities. Because nonparametric Bayesian approaches are known to have the potential for over-splitting [37], we validated the the existence of the two slower states of Chp2^{nmt81} in $swi6\Delta$ cells and in PAmCherry-Chp2^{nmt41} cells by analyzing all of our Chp2 datasets using two other different single molecule tracking methods: Dirichlet process mixture models for single-particle tracking (DPSP) [107] and Spot-On [18]. Both analysis methods capture a similar increase in dynamics and significant heterogeneity in the low mobility state tracks as NOBIAS (Figure 5.2).

Unlike the split in the mobility in Chp2 in $swi6\Delta$ cells, we observed a change in the weight fraction of molecules in the chromatin sampling β state upon imaging PAmCherry-Swi6 in $chp2\Delta$ cells [106]. Hence, the deletions of the individual HP1 proteins have very different impacts on protein dynamics. At the extreme limit, the *nmt1* promoter-driven expression of PAmCherry-Chp2 leads to the complete displacement of Swi6 from sites of heterochromatin formation. As expected, we observed labeled mNeongreen-Swi6 molecules uniformly distributed across the nucleus upon PAmCherry-Chp2^{nmt1} overexpression (Figure 5.2D).

5.2.2 Chp2 dissociates faster from the H3K9me site *in vivo* than *in vitro*

The preponderance of the stationary H3K9me-binding state for PAmCherry-Chp2^{nmt81} implies that our high-resolution single particle tracking measurements may overestimate Chp2 dissociation rates due to photobleaching. This parameter is crucial to determine Chp2 binding kinetics *in vivo* and determine the extent to which such measurements correlate with *in vitro* assays. We estimated the dissociation rate using two approaches: 1) single-molecule tracking followed by a Bayesian synthetic likelihood (BSL) simulation [189]. This simulation-based approach has the benefit of not being affected by experimental time-resolution limits. 2) single-molecule time-lapse imaging at different time intervals to ensure that photobleaching did not lead to an overestimation of the dissocia-

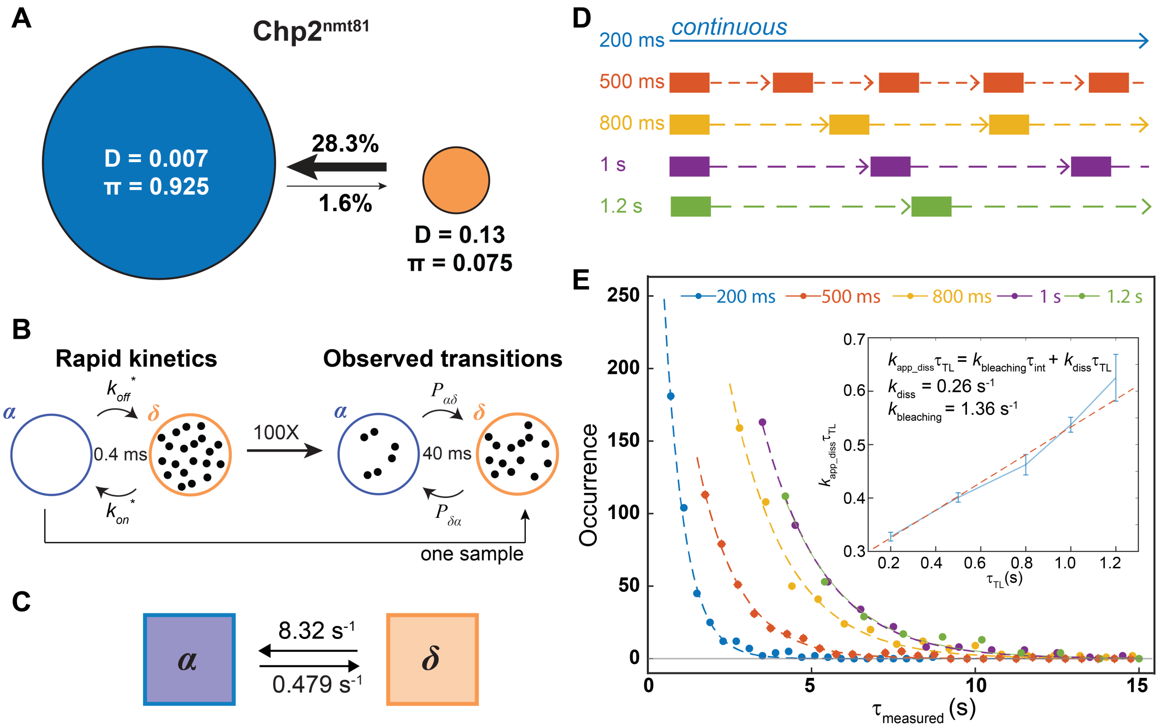


Figure 5.3: 2A: Inferred transition probabilities between the two mobility states of PAmCherry-Chp2^{nmt81} from single-molecule tracking (Figure 5.1D). Diffusion coefficient, D, in units of $\mu m^2/s$ and weight fraction, π, are indicated. Arrow widths are proportional to transition probability. 2B: Fine-grained chemical kinetic simulation with Bayesian Synthetic Likelihood algorithm. The reaction on/off rate is proposed and simulated at a 0.4-ms time interval to calculate the likelihood based on transition probabilities from A at the 40-ms experimental imaging time interval. 2C: Inferred rate constants for PAmCherry-Chp2^{nmt81}. 2D: Schematic for single-molecule time-lapse imaging. The time-lapse period, τ_{TL}, is the sum of the 200-ms integration time and the time delay. Five different time delays were used to access: τ_{TL} = 200, 500, 800, 1000, and 1200 ms. 2E: Dwell time distributions for PAmCherry-Chp2^{nmt81}. The distributions are shown with fits to an exponential decay. Insert: linear fit (red dashed line) of $k_{diss_{app}}\tau_{TL}$ vs. τ_{TL}, from which the dissociation rate constant, k_{diss}, and the photobleaching rate constant, $k_b leaching$, are obtained. Errors bars are the standard deviation of the exponential decay fitting.

tion rate [211].

To infer the rate constants of transitions based on the NOBIAS transition matrices, we used a BSL algorithm, which has previously been applied to assess Swi6 dynamics [106, 189]. At each step, we simulated the experimental outcome of the transitions with 0.4 ms time steps 2000 times for a set of rate constants (Figure 5.3B). BSL methods infer the most justifiable distribution of rate constants to estimate the value and uncertainty of the reaction rate. We applied the BSL method to analyze the output of the single-molecule tracking analysis and estimated that $k_{diss} = 0.479 \pm 0.005 s^{-1}$(Figure 5.3C). We also experimentally determined Chp2 residence times and dissociation rates using single-molecule time-lapse imaging (5.4). Based on single-molecule time-lapse imaging at five different time intervals (Figure 5.3D), we calculated a Chp2-H3K9me disassociation rate of $k_{diss} = 0.260 \pm 0.018 s^{-1}$ and an average dwell time of 3.85 s (Figure 5.3E). In contrast, time-lapse imaging of Swi6 gives $k_{diss} = 0.454 \pm 0.051 s^{-1}$ and and an average dwell time of 2.20 s (Figure 5.4C). Both experimental approaches (single-molecule tracking and single-molecule photobleaching) measured a dissociation rate more than 10-fold faster *in vivo* compared to previous *in vitro* measurements of Chp2 binding to H3K9me $9.6 \pm 0.60 \times 10^{-3} s^{-1}$ for me2 and $1.5 \pm 0.27 \times 10^{-2} s^{-1}$for me3) [194]. Furthermore, comparing the results for Chp2 to previously reported Swi6 dissociation results using the same BSL analysis of single-molecule tracking data(Biswas et al., 2022), we find that the Chp2 dissociation rate ($0.479 s^{-1}$) is lower than that of Swi6 ($1.27 s^{-1}$). The *in vivo* time-lapse measurement of Chp2 and Swi6 disassociation rates reveal that Chp2 remains bound to H3K9me for a longer time than Swi6 *in vivo* suggesting that in fact, Chp2 binds to H3K9me chromatin with higher affinity. Deleting Swi6 did not affect the residence time and dissociation rate of PAmCherry-Chp2^{nmt81} in $swi6\Delta$cells which revealed a $k_{diss} = 0.269 \pm 0.031 s^{-1}$ and and an average dwell time of 3.72 s (Figure 5.4D). The similarity in disassociation rates between PAmCherry-Chp2^{nmt81} and PAmCherry-Chp2^{nmt81} in $swi6\Delta$ cells indicates that although deleting Swi6 perturbs Chp2 dynamics, it does not affect the intrinsic affinity between Chp2 and H3K9me chromatin.

5.2.3 The anti-silencing factor Epe1 co-localizes with its HP1 binding partner primarily at sites of H3K9 methylation and exhibits limited off-chromatin dynamics

Having established the baseline dynamics of two major HP1 proteins in *S. pombe*-Swi6 and Chp2, we sought to determine how HP1 proteins interact with accessory factors to facilitate heterochromatin assembly. The putative H3K9me demethylase Epe1 is a major determinant of heterochromatin stability [80, 81, 88, 212]. Epe1 directly binds to Swi6 and this interaction is essential for Epe1 recruitment to sites of H3K9 methylation. Deleting Epe1 leads to both unregulated H3K9 methylation spreading and increased epigenetic inheritance [58,81]. We labeled Epe1 at the C-terminus with PAmCherry (Epe1-PAmCherry-). To confirm if Epe1 molecules successfully localize at heterochromatin sites, we labeled Swi6 with mNeonGreen (mNeonGreen-Swi6) in cells and imaged the emission in the green channel (488-nm excitation) alongside Epe1-PAmCherry in the red channel (561-nm excitation). Overlaying mNeonGreen images with Epe1-PAmCherry super-resolution images indicates that Epe1 foci form at the periphery of Swi6-heterochromatin foci (Figure 5.5A).

To identify the mobility states associated with Epe1, we tracked single Epe1-PAmCherry molecules and inferred the number of mobility states, the diffusion coefficients, and the weight fraction for each Epe1 state. Since the interaction between Epe1 and Swi6 is direct, we expected to observe four mobility states similar to what we previously observed with Swi6. In contrast, we found that Epe1 has only two mobility states and that the predominant slower state (weight fraction, π_{slow} 94%, $D_{slow,Epe1} = 0.008\ \mu m^2/s$) (Figure 5.5B). Only 6% of Epe1 are assigned to a faster state with $D_{fast,Epe1} = 0.22\ \mu m^2/s$. The transition probabilities indicate that transitioning from the fast state to the slow state is much more favored than the reverse transition (21% to 0.8%) (Figure 5.5F). These results suggest that in the presence of H3K9me, Epe1 preferentially remains in the H3K9me bound state

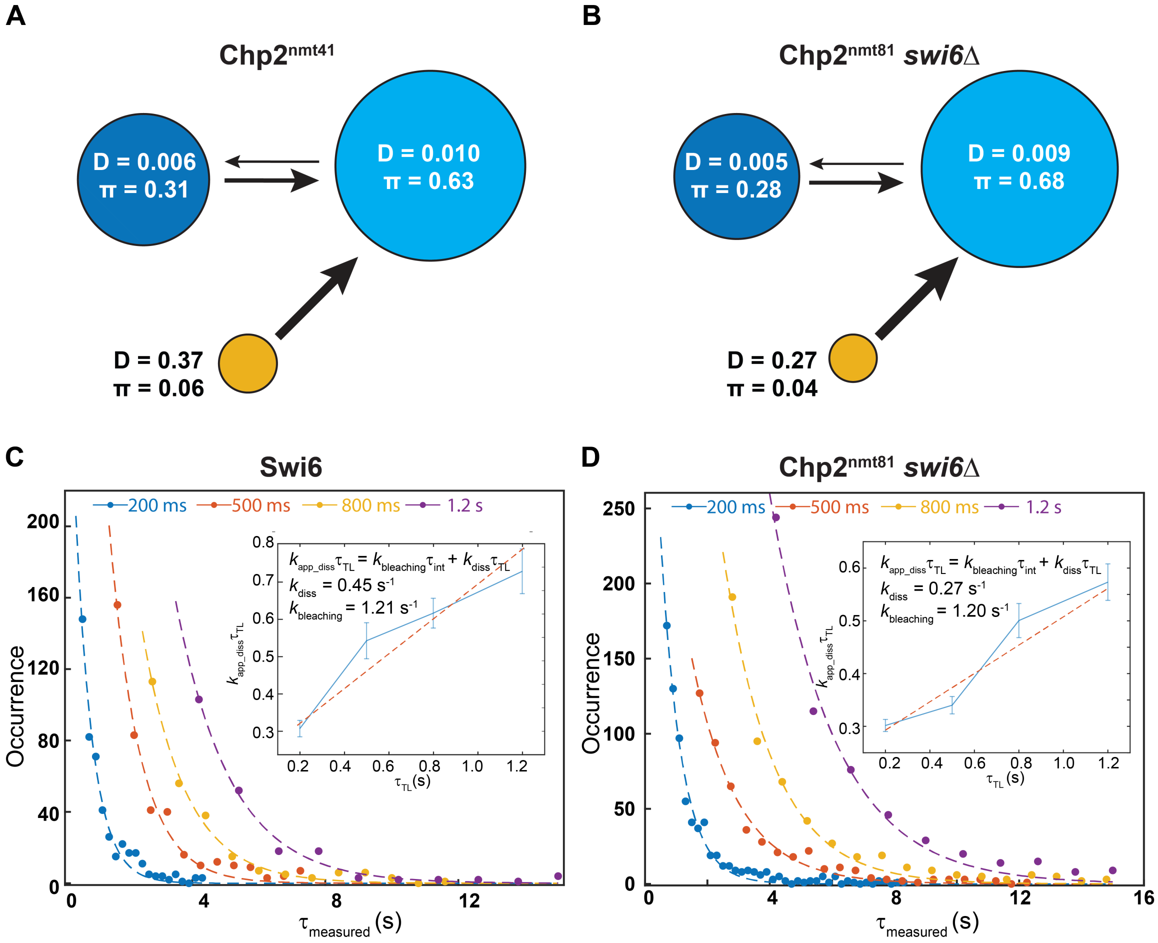


Figure 5.4: 2A-B Transition probabilities between the three mobility states of PAmCherry-Chp2^{nmt41} (A) and PAmCherry-Chp2^{nmt81} in *swi*6Δ cells (B) from NOBIAS. Diffusion coefficient, D, in units of $\mu m^2/s$ and weight fraction, π, are indicated. The arrow widths are proportional to the transition probabilities. 2C-D: Dwell time distributions for PAmCherry-Swi6 expressed under endogenous promoter (C) and PAmCherry-Chp2^{nmt81} (D) in *swi*6Δ cells. The distributions are shown with fits to an exponential decay. Insert: linear fit (red dashed line) of $k_{diss_{app}}\tau_{TL}$ versus τ_{TL}, from which the dissociation rate constant k_{diss} and the photobleaching rate constant $k_{bleaching}$ are obtained. Errors bars are the standard deviation of the exponential decay fitting.

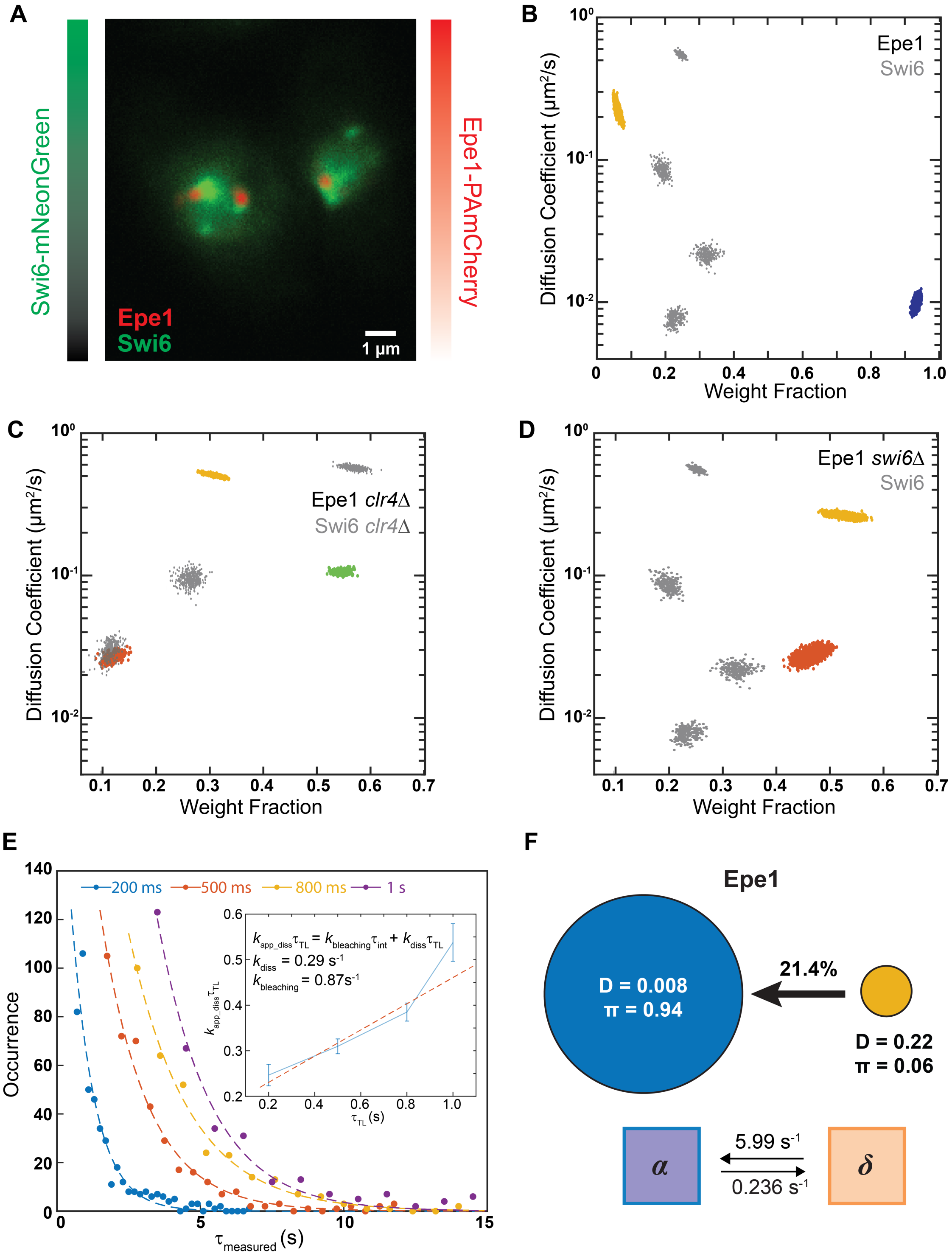

A
Swi6-mNeonGreen
Epe1-PAmCherry
Epe1
Swi6
1 µm
B
Epe1
Swi6
Diffusion Coefficient (µm²/s)
10⁰
10⁻¹
10⁻²
0
0.2
0.4
0.6
0.8
1.0
Weight Fraction
C
Epe1 clr4Δ
Swi6 clr4Δ
Diffusion Coefficient (µm²/s)
10⁰
10⁻¹
10⁻²
0.1
0.2
0.3
0.4
0.5
0.6
0.7
Weight Fraction
D
Epe1 swi6Δ
Swi6
Diffusion Coefficient (µm²/s)
10⁰
10⁻¹
10⁻²
0.1
0.2
0.3
0.4
0.5
0.6
0.7
Weight Fraction
E
200 ms
500 ms
800 ms
1 s
Occurrence
0
20
40
60
80
100
120
140
0
5
10
15
τmeasured (s)
k_app_diss τTL = k_bleaching τint + k_diss τTL
k_diss = 0.29 s⁻¹
k_bleaching = 0.87s⁻¹
k_app_diss τTL
0.2
0.3
0.4
0.5
0.6
0.2
0.4
0.6
0.8
1.0
τTL (s)
F
Epe1
D = 0.008
π = 0.94
21.4%
D = 0.22
π = 0.06
α
5.99 s⁻¹
0.236 s⁻¹
δ

Figure 5.5

Figure 5.5: 3A: Two-color imaging of cells expressing mNeongreen-Swi6 and Epe1-PAmCherry. Swi6 and Epe1 are expressed from their endogenous promoters. Green colorbar: Swi6-mNeonGreen intensities; Red colorbar: reconstructed Epe1-PAmCherry density map. Both color channels are normalized to the maximum pixel intensity. 3B-3D: NOBIAS identifies distinct mobility states for Epe1-PAmCherry. Each colored point is the average single-molecule diffusion coefficient of PAmCherry-Chp2^{nmt81} molecules in that state sampled from the posterior distribution of NOBIAS inference at a saved iteration after convergence in WT cells (B), $clr4\Delta$ cells (C), and $swi6\Delta$cells (D). Grey points are the previously reported PAmCherry-Swi6 single-molecule dynamics [106]. 3E: Dwell time distributions for Epe1-PAmCherry expressed under its endogenous promoter. The distributions are shown with fits to an exponential decay. Insert: linear fit (red dashed line) of $k_{diss_{app}}\tau_{TL}$ versus τ_{TL}, from which the dissociation rate constant, k_{diss}, and the photobleaching rate constant $k_{bleaching}$ are obtained. Errors bars are from standard deviation of exponential decay fitting. 3F: Top: Transition probabilities between the two mobility states of Epe1-PAmCherry from (B) NOBIAS analysis. Diffusion coefficient, D, in units of $\mu m^2/s$ and weight fraction, π, are indicated. Bottom: Inferred rate constants for Epe1-PAmCherry from the fine-grained chemical kinetic simulation.

presumably through its direct interaction with Swi6.

We were surprised to note that Epe1 and Swi6 exhibit a significant mismatch in mobility states except for the fact that they localize as expected at sites of heterochromatin formation. To determine the role that H3K9 methylation might play in promoting complex formation, we performed Epe1-PAmCherry single particle tracking measurements in $clr4\Delta$ cells. As expected, we noticed that the previously observed Epe1 foci in wild-type cells disappear and Epe1-PAmCherry molecules in $clr4\Delta$ cells exhibit a diffuse distribution throughout the nucleus (Figure 5.6C). Also, we observed a complete loss of the slowest state given that neither Swi6 nor Epe1 can localize to sites of heterochromatin in the absence of their cognate H3K9 methylation ligand (Figure 5.5C). Remarkably, we observed that Epe1 now exhibits three mobility states, and the diffusion coefficients of these states perfectly align with those of Swi6 (Figure 5.5C). Swi6, in the absence of H3K9 methylation, also exhibits three mobility states since the slowest state that depends on H3K9me binding is absent [106]. Hence, our results suggest that Epe1 and Swi6 indeed can also directly interact with each other to form off-chromatin complexes. However, the presence of H3K9me chromatin significantly shifts the equilibrium towards a chromatin-bound state:

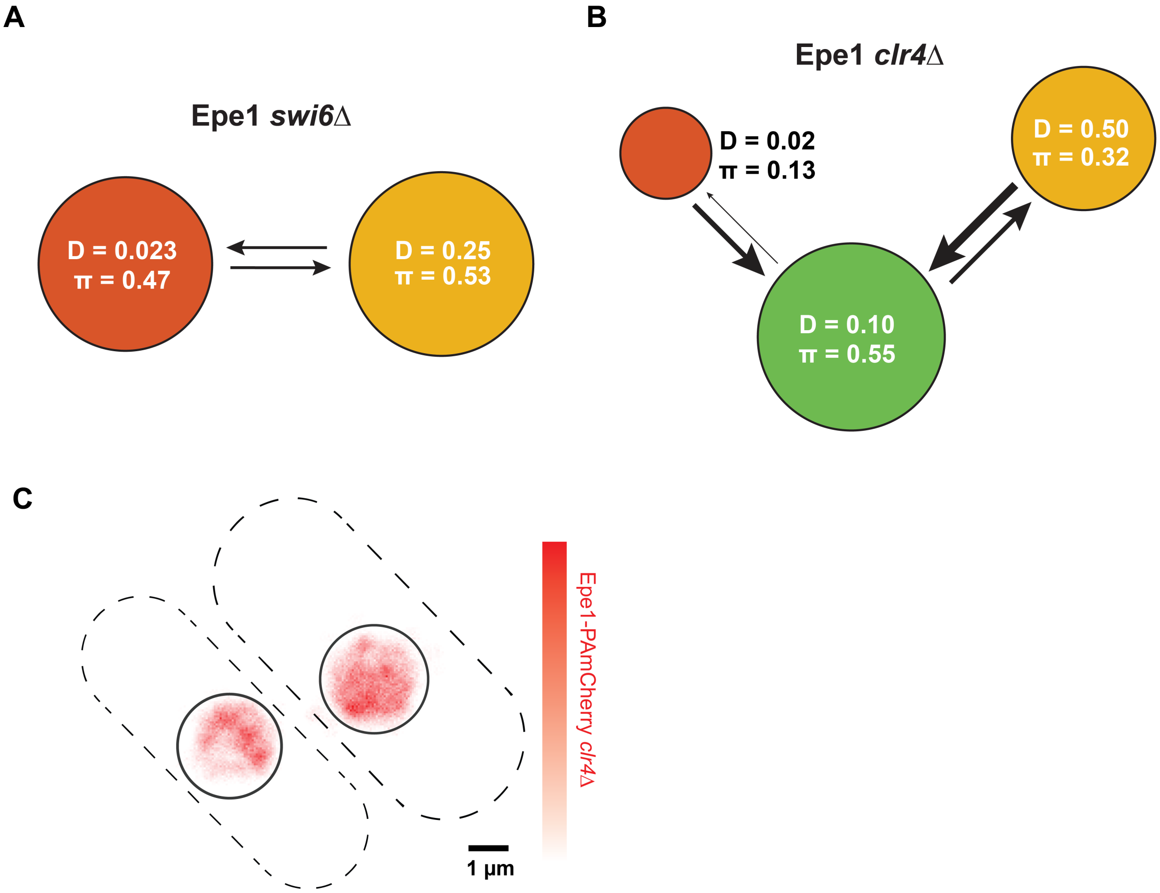


Figure 5.6: 3A-B Transition probabilities between the mobility states of Epe1-PAmCherry in *swi*6Δ cells (A) and in *clr*4Δ cells (B) from NOBIAS. The arrow widths are proportional to the transition probabilities. 3C: Reconstructed single-molecule fits density map for Epe1-PAmCherry in *clr*4Δ cells. Dashed lines: approximate *S. pombe* cell outlines; solid circles: approximate nucleus borders.

in the clr4+ case where H3K9me is present, the transition out of the fast state is 26 times higher than the transition into the fast state (Figure 5.5F) whereas this ratio decreases to 1.9 in $clr4\Delta$ cells (Figure 5.6B).

To validate that the recruitment of Epe1 to sites of H3K9 methylation (i.e., the slow state) is dependent on Swi6, we performed single particle tracking measurements of Epe1-PAmCherry in a $swi6\Delta$ background. As expected, we observed a complete loss of the slow state and the appearance of a new mobility state with a diffusion coefficient 3 times higher than the slowest state that we measured in WT cells. In addition, the weight fraction for the faster state, πfast, increases from 6% in the WT background to over 50% in $swi6\Delta$ (Figure 5.5D).

Finally, we performed time-lapse imaging to measure the Epe1 dissociation rate. We estimated that Epe1 dissociates from sites of heterochromatin formation at a rate that is $k_{diss} = 0.288 \pm 0.044s^{-1}$ according to single-molecule time-lapse imaging with four timefour-time intervals (Figure 5.5E). Consistently, BSL analysis of the single-molecule tracking transition matrix gave $k_{diss} = 0.236 \pm 0.003s^{-1}$ (Figure 5.5F) or a dwell time of 4.24s. These data suggest that Epe1 remains bound to heterochromatin for dwell times that are much longer than that of Swi6 despite Swi6 and Epe1 directly interacting with each other to form a complex. This suggests H3K9me or multivalency arising from Swi6 oligomerization might promote the stable association of Epe1 at sites of heterochromatin formation.

5.2.4 Histone remodeler Mit1 and histone deacetylase Clr3 assemble into SHREC complex only at heterochromatin

Given our observation that Epe1 and Swi6 preferentially form complexes at sites of H3K9 methylation and not off-chromatin, we wanted to determine the extent to which the principle of H3K9me-directed complex assembly might be generalizable to other HP1 protein complexes. The SHREC complex consists of a histone remodeler Mit1 and histone deacetylase (HDAC) Clr3 (Figure 5.7A) [83, 86]. Unlike Epe1, which critically depends on Swi6 for its recruitment to heterochromatin, proteins that are part of the SHREC complex preferentially form complexes with Chp2 [83, 87]. The C-terminus of Chp2 forms a complex with the N-terminus of Mit1 and their interactions have been characterized using X-ray crystallography [83]. This is further supported by studies of Swi6 purification followed by mass spectrometry in *chp*2Δ cells which reveals a precipitous loss of Mit1 from heterochromatin [171]. The recruitment of Clr3 is more complex and depends both on HP1-dependent and HP1-independent interactions [80, 83, 87].

We previously determined that Chp2 exhibits two distinct mobility states and hence we extended our studies to identify the mobility states associated with its primary interacting partners - Mit1 and Clr3. We fused PAmCherry to the N-terminus of Mit1 and Clr3 and expressed the two fusion proteins using a thiamine-repressible nmt81 promoter. We determined that PAmCherry-Mit1^{nmt81} preserved epigenetic silencing at the mat locus by using a *ura4+* based silencing assay (Figure 5.8A). As previously described, the establishment of *ura4+* silencing leads to growth in FOA (EMMC+FOA) containing media and the lack of growth in media without uracil (EMM-URA). Our single-molecule tracking data for PAmCherry-Mit1^{nmt81} and PAmCherry-Clr3^{nmt81} reveals that both proteins exhibit three mobility states (Figure 5.7B,D,E). The diffusion coefficients for Mit1 and Clr3 only match each other for the slowest states (D_{slow} = 0.005 $\mu m^2/s$) with comparable weight fractions. There are 28% of slow-state single-molecule steps for Mit1 and 25% for Clr3. Notably, the D_{slow} values for these two proteins are again at levels similar to what we have observed in the case of other heterochromatin-associated factors (D_{slow} of Swi6, Chp2, and Epe1). Reconstructed single-molecule fits density heatmap of Mit1 show that their high-density hotspots also exhibit spatial patterns that are similar to Chp2, Swi6, and Epe1 while Clr3 has a more widely dispersed pattern (Figure 5.8C).

We analyzed transition probabilities and calculated spatial autocorrelations for Mit1 and Clr3 based on our single molecule tracking data. We noticed that Clr3 has a higher

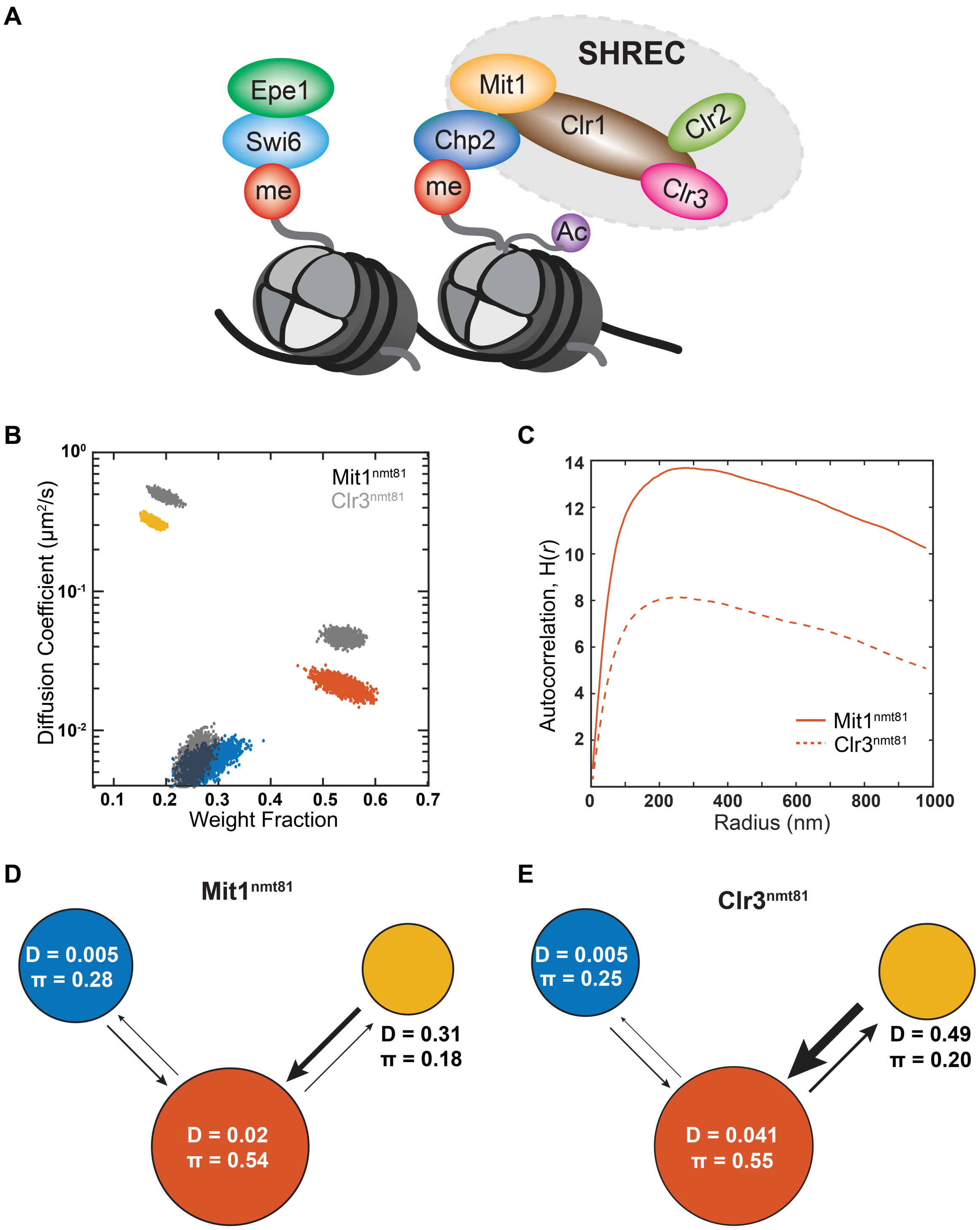
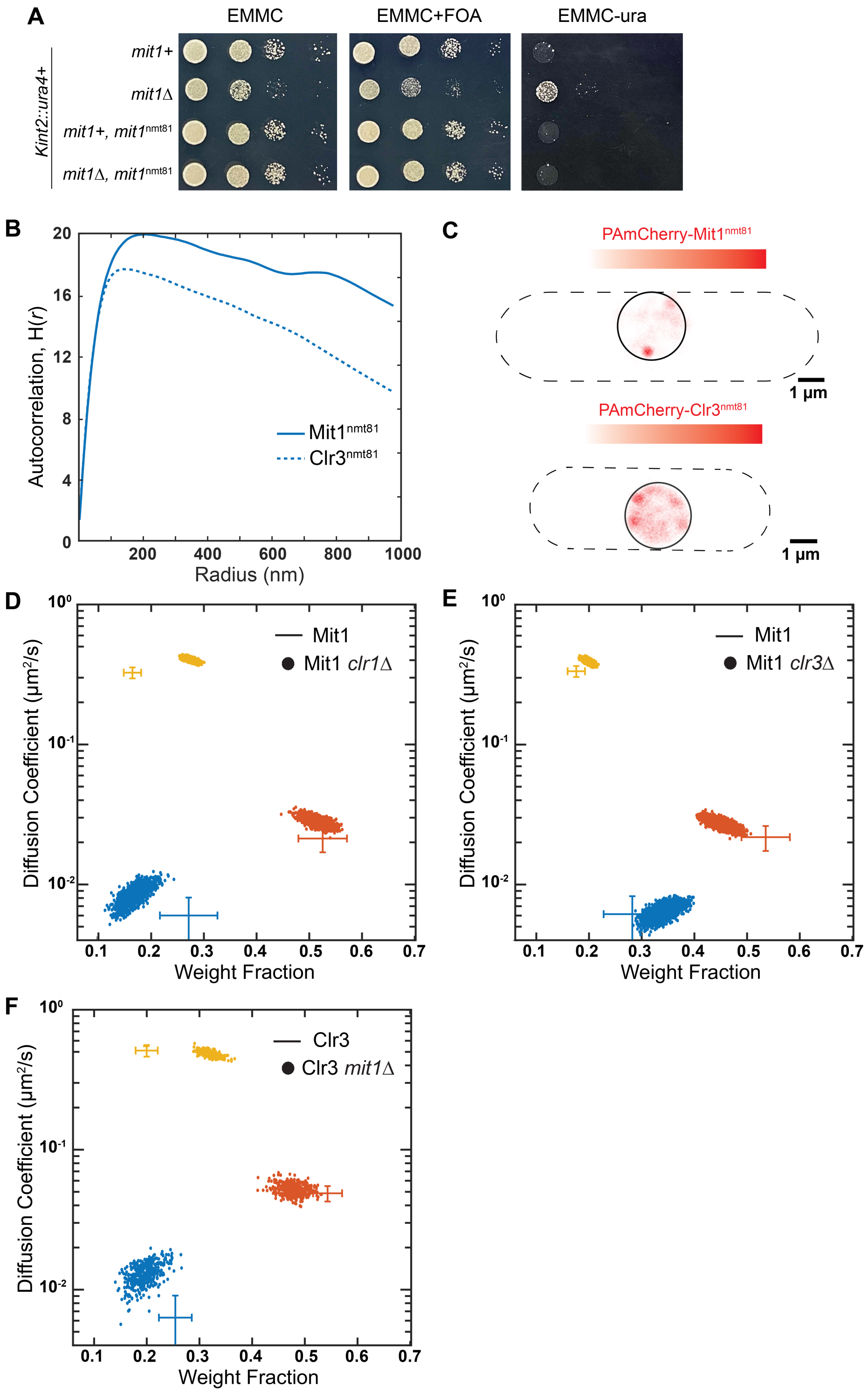

A
SHREC
Epe1
Swi6
me
Mit1
Clr1
Clr2
Chp2
me
Clr3
Ac
B
10^0
10^{-1}
10^{-2}
Diffusion Coefficient ($\mu m^2/s$)
Mit1^{nmt81}
Clr3^{nmt81}
0.1
0.2
0.3
0.4
0.5
0.6
0.7
Weight Fraction
C
14
12
10
8
6
4
2
Autocorrelation, H(r)
Mit1^{nmt81}
Clr3^{nmt81}
0
200
400
600
800
1000
Radius (nm)
D
Mit1^{nmt81}
D = 0.005
π = 0.28
D = 0.31
π = 0.18
D = 0.02
π = 0.54
E
Clr3^{nmt81}
D = 0.005
π = 0.25
D = 0.49
π = 0.20
D = 0.041
π = 0.55

Figure 5.7

Figure 5.7: 4A: Schematic of H3K9 methylated nucleosomes interacting with HP1 proteins and forming HP1 sub complexes. Swi6 binds to Epe1, and Chp2 interacts with the SHREC complex through Mit1. 4B: NOBIAS identifies three distinct mobility states for PAmCherry-Mit1^{nmt81} and PAmCherry-Clr3^{nmt81}. Each point is the average single-molecule diffusion coefficient of PAmCherry-Mit1^{nmt81} molecules (colored points) or PAmCherry-Clr3^{nmt81} molecules (grey points) in that state sampled from the posterior distribution of NOBIAS inference at a saved iteration after convergence. 4C: The intermediate state of Mit1^{nmt81} (solid line) has a higher Ripley's H(r) than the intermediate state of Clr3^{nmt81} (dashed line). Each autocorrelation plot is normalized with randomly simulated trajectories from the same state (5.4). 4D-E Transition probabilities between the three mobility states of PAmCherry-Mit1^{nmt81} (D) and PAmCherry-Clr3^{nmt81} (E) from NOBIAS. The arrow widths are proportional to the transition probabilities. Diffusion coefficient, D, in units of $\mu m^2/s$ and weight fraction, π, are indicated.

transition probability from the fast state to the intermediate state compared with Mit1 (Figure 5.7D,E). Spatial autocorrelation analysis is useful especially when combined with the state label NOBIAS provides for each step. We used Ripley's H function to determine the potential spatial overlap between Mit1 and Clr3 for different mobility states [186]. A higher H(r) value indicates a higher clustering level at searching radius r, and Mit1 has a higher H function value than Clr3 in the intermediate state at all searching radii (Figure 5.7C). In contrast, there is little difference between the H functions for the slowest states of Mit1 and Clr3, indicating that the clustering level of Mit1 and Clr3 slow state is similar (Figure 5.8B). In summary, the spatial auto-correlation analysis and single-molecule dynamic measurements for Mit1 and Clr3 suggest that the SHREC complex components preferentially co-localize on chromatin and is unlikely to form off-chromatin complexes in live *S. pombe* cells.

Whether Chp2 and Mit1 recruit the HDAC module Clr3, to sites of H3K9me or the two SHREC complex components are recruited to the H3K9me site independently remains an open question [83,87]. To test between these possibilities we deleted Clr1, a protein in the SHREC complex that is thought to link the remodeling and HDAC modules. We observed an increase in the fast state weight fraction and a decrease in the bound state weight fraction for Mit1 (Figure 5.8D), which further confirmed that Mit1's bound state depends



A
EMMC
EMMC+FOA
EMMC-ura
Kint2::ura4+
mit1+
mit1Δ
mit1+, mit1nmt81
mit1Δ, mit1nmt81
B
Autocorrelation, H(r)
0
4
8
12
16
20
200
400
600
800
1000
Radius (nm)
Mit1nmt81
Clr3nmt81
C
PAmCherry-Mit1nmt81
1 μm
PAmCherry-Clr3nmt81
1 μm
D
Diffusion Coefficient (μm²/s)
10⁰
10⁻¹
10⁻²
0.1
0.2
0.3
0.4
0.5
0.6
0.7
Weight Fraction
Mit1
Mit1 clr1Δ
E
Diffusion Coefficient (μm²/s)
10⁰
10⁻¹
10⁻²
0.1
0.2
0.3
0.4
0.5
0.6
0.7
Weight Fraction
Mit1
Mit1 clr3Δ
F
Diffusion Coefficient (μm²/s)
10⁰
10⁻¹
10⁻²
0.1
0.2
0.3
0.4
0.5
0.6
0.7
Weight Fraction
Clr3
Clr3 mit1Δ

Figure 5.8

Figure 5.8: 4A: Silencing assay using a *ura4+* reporter inserted at the mat locus (*Kint2:ura4*). 10-fold serial dilution of cells were plated on EMMC, EMMC+FOA and EMM-URA plates to determine if PAmCherry-Mit1^{nmt81} expression can establish heterochromatin in $mit1\Delta$ cells.4B: The slow state of Mit1 (solid line) has similar Ripley's H(r) value as the slow state of Clr3 (dashed line) in WT cells, indicating that both proteins are clustered in slow states. Each autocorrelation plot is normalized with randomly simulated trajectories from the same state (5.4).4C: Reconstructed single-molecule fits density map for PAmCherry-Mit1^{nmt81} (top) and PAmCherry-Clr3^{nmt81} (bottom) in WT cells. Dashed lines: approximate *S. pombe* cell outlines; solid circles: approximate nucleus borders. 4D-F: NOBIAS identifies distinct mobility states for PAmCherry-Mit1^{nmt81} in $clr1\Delta$ cells (D), and in $clr3\Delta$ cells (E). NOBIAS also identifies distinct mobility states for PAmCherry-Clr3^{nmt81} in $mit1\Delta$ cells (F). Each colored point is the average single-molecule diffusion coefficient of molecules in that state sampled from the posterior distribution of NOBIAS inference at a saved iteration after convergence. The colored crosses show the data from PAmCherry-Mit1^{nmt81} or PAmCherry-Clr3^{nmt81} molecules in WT cells (Figure 5.7B).

on heterochromatin as Clr1 also interacts with Chp2 [87]. We also acquired and analyzed single-molecule tracking data for PAmCherry-Mit1 in $clr3\Delta$ cells, in which the number of diffusive states remains to be 3, and there is little change in the corresponding D and weight fraction for each state (Figure 5.8E), which means the binding of the remodeler module (Mit1) does not depend on the HDAC module (Clr3). In contrast, we found that the slow state of PAmCherry-Clr3 in $mit1\Delta$ cells has a decreased weight fraction and an increase in the diffusion coefficient associated with the slow state (D_{slow} changes from 0.005 $\mu m^2/s$ to 0.010 $\mu m^2/s$) (Figure 5.8F). These results suggest that Clr3 binding might depend on the successful binding and recruitment of Mit1. Alternatively, the deletion of Mit1 could have a larger effect on heterochromatin stability but our imaging experiments cannot distinguish between these two possibilities.

5.2.5 SHREC complex dynamics is affected by H3K9 methylation

To test whether the slowest mobility state corresponding to Mit1 and Clr3 depends on H3K9 methylation, we performed single-molecule tracking measurements in $clr4\Delta$ cells. In $clr4\Delta$cells, Mit1 and Clr3 exhibit a substantial increase in the fastest mobility state (17.7% to 37.8% for Mit1 and 20.0% to 46.0% for Clr3) with a concomitant decrease in the slowest mobility state (28.4% to 10.7% for Mit1 25.5% to 13.0% for Clr3) (Figure 5.9A-B). However, unlike what we observed in the case of the two HP1 proteins—Swi6 and Chp2—or Epe1, the slowest mobility state is not fully eliminated for either PAmCherry-Mit1 or PAmCherry-Clr3 in *clr*4Δ cells. These results suggest that other mechanisms in addition to H3K9 methylation are responsible for the slow mobility state of Mit1 and Clr3 (although more than half of its -bound state is determined by H3K9 methylation). In the Ripley's H cluster analysis, for all steps of Mit1 and Clr3 in WT cells and *clr*4Δ cells, we notice a substantial decrease in H(*r*) value for both proteins in the absence of Clr4, consistent with reduced clustering. (Figure 5.9C). Reconstructed localization maps of PAmCherry-Mit1 and PAmCherry-Clr3 in *clr*4Δ cells also show an overall unclustered spatial pattern for both proteins (Figure 5.9D).

Next, we tested the extent to which the slow mobility state of Mit1 and Clr3 depends on the two HP1 proteins- Swi6 and Chp2. We acquired single-molecule tracking data of PAmCherry-Mit1 in *chp*2Δ, *swi*6Δ, and *chp*2Δ*swi*6Δ cells. We analyzed the Mit1 single particle trajectories associated with each dataset and inferred the number of diffusive states and associated D and W values (Figure 5.9E-F, 5.10A). We notice that Mit1 in *chp*2Δ cells exhibits a substantial decrease in the bound state weight fraction compared to WT cells, but this decrease is less than what we observed in the case of *clr*4Δ cells. In contrast, we observed a similar weight fraction for all 3 diffusive states of Mit1 in *swi*6Δ cells compared to Mit1 in WT cells. These results suggest and indeed confirm that Chp2 is the primary HP1 protein interacting with Mit1. In the absence of Chp2, Swi6 can play a compensatory role highlighting the potential for cross-talk and shared binding sites between the two proteins. Indeed, we observed that Mit1 dynamics in *chp*2Δ*swi*6Δ produced an additive effect resulting in a further decreased bound state weight fraction compared with only *chp*2Δ cells.

For the HDAC Clr3, we acquired single-molecule tracking data for PAmCherry-Clr3 in *chp*2Δ and *swi*6Δ cells. In the analysis of Clr3 in *chp*2Δ cells (Figure 5.10B), we no-

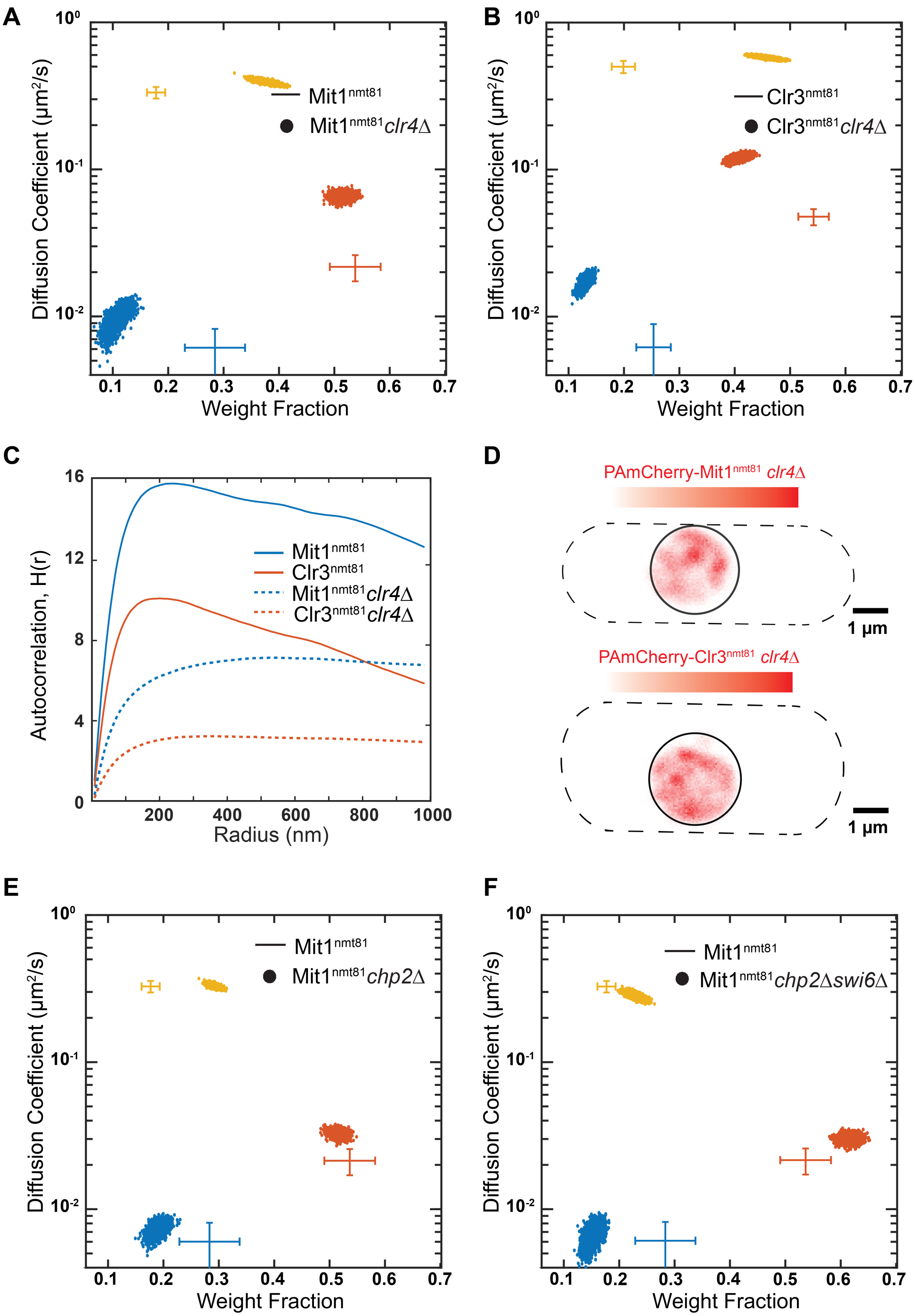

A
10^0
10^{-1}
10^{-2}
Diffusion Coefficient (μm^2/s)
Mit1^{nmt81}
Mit1^{nmt81}*clr4Δ*
0.1 0.2 0.3 0.4 0.5 0.6 0.7
Weight Fraction
B
10^0
10^{-1}
10^{-2}
Diffusion Coefficient (μm^2/s)
Clr3^{nmt81}
Clr3^{nmt81}*clr4Δ*
0.1 0.2 0.3 0.4 0.5 0.6 0.7
Weight Fraction
C
16
12
8
4
0
Autocorrelation, H(r)
Mit1^{nmt81}
Clr3^{nmt81}
Mit1^{nmt81}*clr4Δ*
Clr3^{nmt81}*clr4Δ*
200 400 600 800 1000
Radius (nm)
D
PAmCherry-Mit1^{nmt81} *clr4Δ*
1 μm
PAmCherry-Clr3^{nmt81} *clr4Δ*
1 μm
E
10^0
10^{-1}
10^{-2}
Diffusion Coefficient (μm^2/s)
Mit1^{nmt81}
Mit1^{nmt81}*chp2Δ*
0.1 0.2 0.3 0.4 0.5 0.6 0.7
Weight Fraction
F
10^0
10^{-1}
10^{-2}
Diffusion Coefficient (μm^2/s)
Mit1^{nmt81}
Mit1^{nmt81}*chp2Δswi6Δ*
0.1 0.2 0.3 0.4 0.5 0.6 0.7
Weight Fraction

Figure 5.9

Figure 5.9: 5A-B: NOBIAS identifies three distinct mobility states for PAmCherry-Mit1^{nmt81} (A) and PAmCherry-Clr3^{nmt81} (B) in $clr4\Delta$ cells. Each colored point is the average single-molecule diffusion coefficient of molecules in that state sampled from the posterior distribution of NOBIAS inference at a saved iteration after convergence. The colored crosses show the data for PAmCherry-Mit1^{nmt81} and PAmCherry-Clr3^{nmt81} (Figure 5.7B).5C: Ripley's H analysis for steps from all states for Mit1^{nmt81} and Clr3^{nmt81} in WT cells and $clr4\Delta$ cells. The Mit1^{nmt81} and Clr3^{nmt81} (dashed blue and orange line) in $clr4\Delta$ cells has lower Ripley's H(r) value than Mit1^{nmt81} and Clr3^{nmt81} (solid blue and orange line) in WT cells. 5D: Reconstructed single-molecule density map for PAmCherry-Mit1^{nmt81} (top) and PAmCherry-Clr3^{nmt81} (bottom) in $clr4\Delta$ cells. Dashed lines: approximate *S. pombe* cell outlines; solid circles: approximate nucleus borders. 5E-F: NOBIAS identifies three distinct mobility states for PAmCherry-Mit1^{nmt81} in $chp2\Delta$ cells (E) and $chp2\Delta swi6\Delta$ cells (F). Each colored point is the average single-molecule diffusion coefficient of molecules in that state sampled from the posterior distribution of NOBIAS inference at a saved iteration after convergence. The colored crosses show the data for PAmCherry-Mit1^{nmt81} in WT cells (Figure 5.7B).

ticed the same decrease in the bound state as for Mit1 in $chp2\Delta$ cells, which supports the hypothesis of Chp2-mediated HDAC recruitment. Interestingly, in $swi6\Delta$ cells (Figure 5.10C), we observe not only a decreased weight fraction for the Clr3 bound state, but also an increased D_{slow} from 0.005 $\mu m^2/s$ to 0.010 $\mu m^2/s$. Our data shows that the stable nucleosome-bound state of the HDAC component Clr3, requires H3K9me, Chp2, and Mit1. We thus infer that the two modules of the SHREC complex only co-localize in presence of H3K9me and HP1 proteins at heterochromatin sites, given the substantial differences in recruitment behavior between the remodeler component and the HDAC component of the SHREC complex.

5.3 Discussion

We have used single-particle tracking approaches to investigate how heterochromatin-associated factors form complexes in living fission yeast cells. Our observations of the properties of heterochromatin-associated proteins in cells deviate in important and substantive ways from *in vitro* studies. Previous studies have shown that Swi6 binds to nucleosomes with a 3-fold higher affinity than Chp2 [82]. In contrast, our data based on

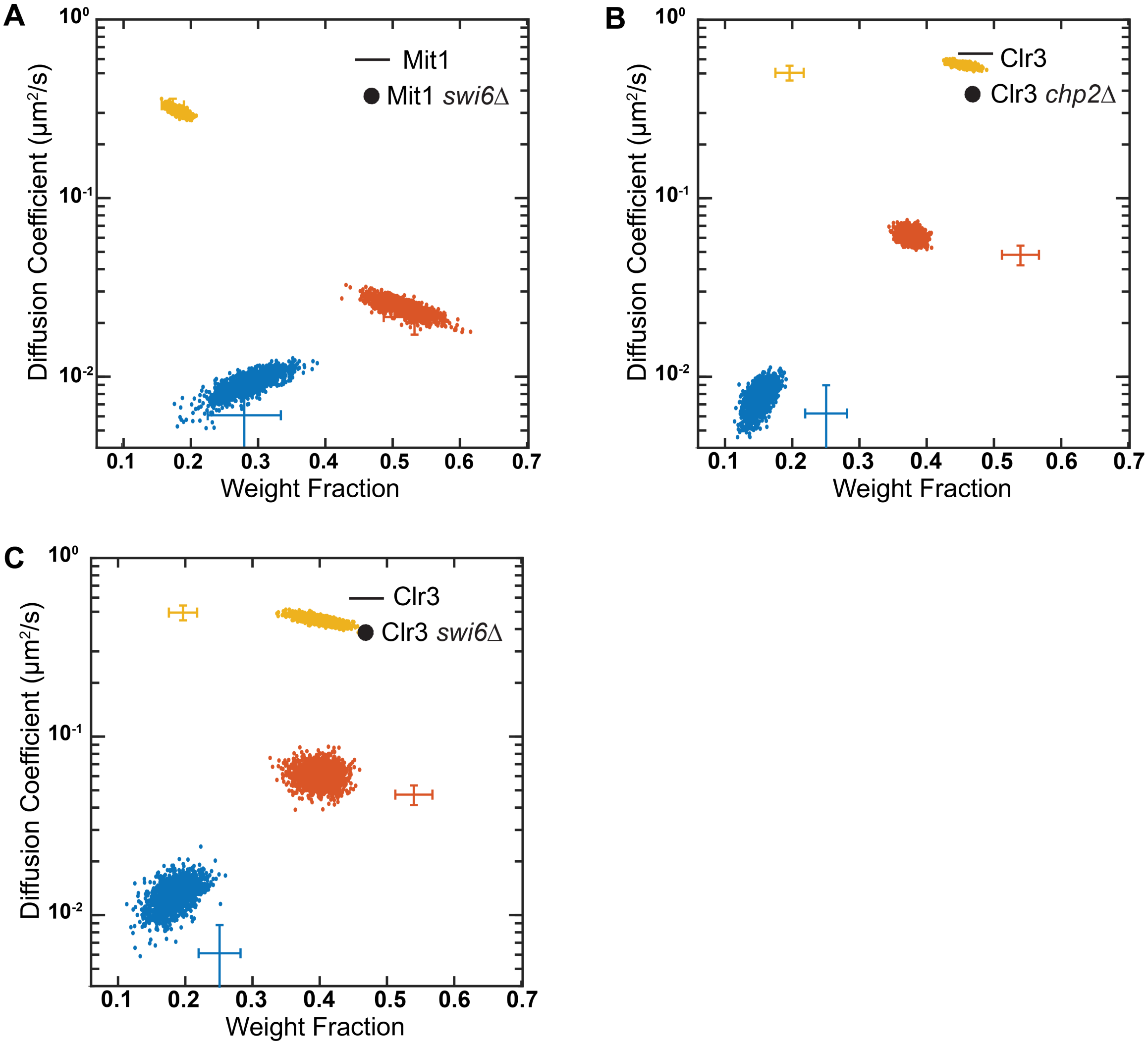


Figure 5.10: NOBIAS identifies distinct mobility states for PAmCherry-Mit1^{nmt81} in *swi6*Δ cells (A). NOBIAS also identifies distinct mobility states for PAmCherry-Clr3^{nmt81} in *chp2*Δ cells (B) and in *swi6*Δ cells (C). Each colored point is the average single-molecule diffusion coefficient of molecules in that state sampled from the posterior distribution of NOBIAS inference at a saved iteration after convergence. The colored crosses show the data from PAmCherry-Mit1^{nmt81} or PAmCherry-Clr3^{nmt81} molecules in WT cells (Figure 5.7B).

1) the weight fractions of molecules in the H3K9 methylation-dependent slow mobility state, 2) the transition rates of molecules between the free and bound states, and 3) time-lapse imaging to measure koff demonstrates that the majority of Chp2 molecules are in the H3K9 methylation-bound state and Chp2 binds with higher affinity to H3K9me chromatin. By varying Chp2 protein expression levels, we also reveal how Chp2 binds with exquisite specificity to H3K9me chromatin when expressed in limiting (and physiologically relevant) amounts. Hence, despite the two HP1 proteins having very similar domains, their different amino acid compositions, especially within the nucleic acid binding hinge domain, likely leads to different biochemical interactions in cells. These results might explain why Chp2 is not easily displaced by Swi6 despite the levels of Chp2 protein being 100-fold lower than that of Swi6 in cells.

In our earlier work, we noted that deleting Chp2 had little effect on the overall dynamics of Swi6 with a slight increase in the chromatin binding (β state) [106]. We concluded that this limited dependence was likely because H3K9me is not substrate-limiting in cells. Yet, deleting Swi6 led to increased dynamics and the emergence of a new diffusive state for Chp2. The increased dynamics and the new Chp2-associated mobility state might appear either because 1) Swi6 oligomerization may stabilize Chp2 binding at sites of heterochromatin formation. This would also explain the differences between H3K9me binding we observed in the *in vitro* and *in vivo* data for Swi6 relative to Chp2; or because 2) Swi6 is not directly involved in Chp2-H3K9me binding, but the loss of Swi6 results in an overall reduction in heterochromatin stability, making Chp2 less bound.

The binding properties of the two HP1 proteins- Swi6 and Chp2, serve as an important point of departure for our measurements on heterochromatin complex assembly in living cells. Epe1, a major anti-silencing factor that interacts with Swi6, exhibits only two mobility states suggesting that Epe1 interacts with Swi6 exclusively at sites of H3K9 methylation. These studies are consistent with our earlier observations showing that the addition of an H3K9 methylated peptide to *in vitro* binding assays dramatically increased

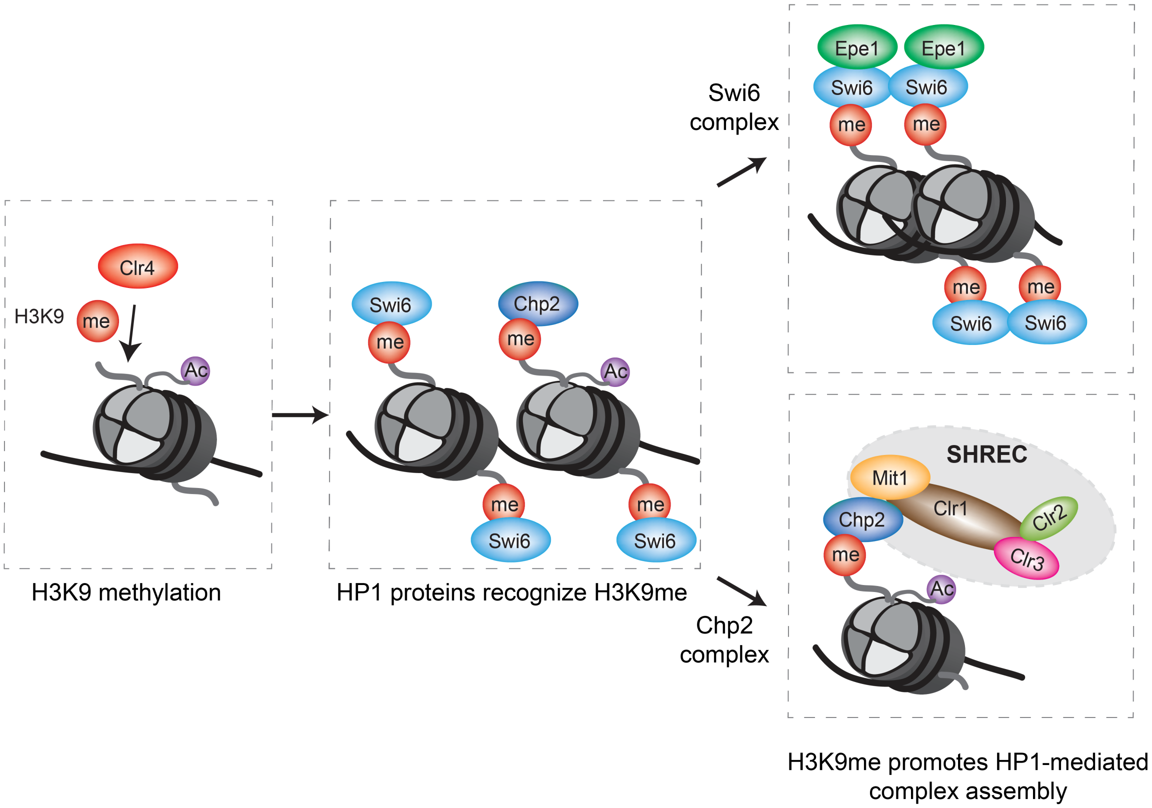


Figure 5.11: Left to right: The sequence specific recruitment of Clr4 promotes H3K9 methylation deposition at sites of heterochromatin; HP1 proteins Swi6 and Chp2 recognize H3K9me with millisecond scale kinetics; H3K9me enhances Swi6 and Chp2 dependent protein complex assembly at sites of heterochromatin. The increased likelihood of complex formation at sites of H3K9me attenuates other possible off-chromatin interactions in the case of Swi6 and Chp2 binding partners (Epe1, Mit1 and Clr3).

the extent of binding between Epe1 and Swi6 [212]. Swi6 IP-MS studies also show that Epe1 interacts with Swi6 in *clr4+* but not *clr4*Δ or H3K9R mutants [212]. Indeed, our studies precisely define that it is in fact the presence of H3K9 methylation itself that attenuates other non-productive Epe1-Swi6 interaction states. Deleting Clr4 leads to Epe1 exhibiting three mobility states, the diffusion coefficients of which perfectly align with that of Swi6 in *clr4*Δ cells. Hence, Swi6 and Epe1 likely form off-chromatin complexes and bind directly to each other. However, the presence of H3K9 methylation dramatically shifts the equilibrium populations toward a chromatin-bound state.

We tested whether the principle of H3K9 methylation enhancing complex formation could be extended to other proteins such as the chromatin remodeler, Mit1, and the histone deacetylase, Clr3, both of which form complexes with the second HP1 protein, Chp2. Unlike Chp2, which has only two mobility states, Mit1 and Clr3 exhibit three mobility states. Both Mit1 and Clr3 exhibit mobility states with different diffusion coefficients and spatial autocorrelation functions except for the slow state which we attribute to an H3K9 methylation bound fraction. These results suggest that Mit1 and Clr3, which are components of the SHREC complex, co-localize only at sites of H3K9 methylation. Our results are consistent with recent structural work on SHREC complex proteins highlighting the special role that Chp2 has in recruiting Mit1 to heterochromatin [83, 204]. These results suggest alternative modes of SHREC complex component recruitment which eventually lead to the co-localization of the remodeler and deacetylase modules, binding as independent components. For example, Mit1 can be recruited via HP1-dependent interactions, CD domain-dependent nucleic acid interactions like DNA binding proteins at the telomeres [86,213]. Clr3, the HDAC module, interacts with Clr2 which has an MBD domain that binds to nucleic acids and also directly interacts with DNA binding proteins such as Atf1 and Pcr1 [83]. Hence, the availability of different binding partners that associate with the SHREC complex could lead to a low mobility state of Mit1 and Clr3 even in the absence of H3K9 methylation.

Chromatin is largely thought to be merely a scaffold that recruits histone-binding proteins to particular locations in the genome [214]. Our single-molecule imaging measurements of heterochromatin proteins and their binding partners reveal a vital role for H3K9 methylation as an enhancer of complex formation in living cells. Although the proteins whose properties we measured directly bind to each other and form pairwise interactions *in vitro*, we observed little off-heterochromatin co-localization when H3K9 methylation is present. Our results reveal the dramatic shift in the equilibrium binding states induced by the presence of H3K9 methylation. Our results have important implications for the reconstitution and structural biology of heterochromatin-associated factors. Specifically, our results emphasize the need to explicitly include H3K9 methylated chromatin substrates when describing models of how heterochromatin-associated factors form complexes both *in vitro* and in cells given its role in enhancing complex formation. Although the mechanisms of such enhancement are not well understood, it is likely due to protein conformational changes that are triggered by nucleosome binding and H3K9 methylation that switches heterochromatin associated proteins from a low-affinity to a high affinity interaction state [212].

5.4 Methods and Materials

5.4.1 Plasmids and Strains

The plasmids and strains used in this Chapter are from the Ragunathan Lab at the Brandeis University. Please refer to corresponding manuscript of this chapter for details.

5.4.2 Funcational assay and western gels

The Funcational assay and western gels used in this Chapter are from the Ragunathan Lab at Brandeis University. Please refer to corresponding manuscript of this chapter for details.

5.4.3 *S. pombe* live-cell imaging

Yeast strains were grown in standard complete YES media (US Biological, catalog no. Y2060) containing the full complement of yeast amino acids and incubated overnight at 32°C. For PAmCherry-Epe1 strains and Epe1 mutants under the control of the native Epe1 promoter, the seed culture was diluted and incubated at 25°C with shaking to reach an optical density at 600 nm (OD_{600}) of 0.5. For strains with the nmt1, nmt41, or nmt81 promoter, the seed culture was diluted into EMMC media (FORMEDIUM, cat. PMD0402) containing the full complement of yeast amino acids and incubated at 25°C with shaking to reach an optical density at 600 nm (OD_{600}) of 0.5. To maintain cells in an exponential phase and eliminate extranuclear vacuole formation, the culture was maintained at OD_{600} 0.5 for 2 days, with dilutions performed at 12-hour time intervals(24 hour time interval for EMM media culture. Cells were pipetted onto a pad of 2% agarose prepared in EMM media and each agarose pad sample was imaged for less than 1 hour. *S. pombe* cells were imaged at room temperature with a 100× 1.40 numerical aperture (NA) oil-immersion objective in an Olympus IX-71 inverted microscope. First, the fluorescent background was decreased by exposure to 488-nm light (Coherent Sapphire, 377 W/cm^2 for 20 to 40 s). A 406-nm laser (Coherent Cube, 405-100; 1-5 W/cm^2) was used for photoactivation (200-ms activation time), and a 561-nm laser (Coherent Sapphire, 561-50; 70.7 W/cm^2) was used for imaging. Images were acquired at 40-ms exposure time per frame. The fluorescence emission was filtered to eliminate the 561-nm excitation source and imaged using a 512 × 512-pixel Photometrics Evolve EMCCD camera.

5.4.4 Single-molecule trajectory analysis

Recorded PAmCherry single-molecule positions were localized and tracked with SMALL-LABS software [12]. A mask of the nucleus of each cell was determined based on auto-florescence outside the nucleus in the 488nm bleaching step. Only the signal within the nucleus mask was analyzed. Single-molecule trajectory datasets were analyzed by a non-parametric Bayesian framework NOBIAS to infer the number of mobility states, the parameter for each state, and the transition between states [210]. More than 1000 trajectories for each SPT dataset are put in the framework for robust analysis and to eliminate rare events. Reported parameters for each state are the posterior mean after the number of mobility states stabilize, and reported uncertainty is the standard deviation from the posterior distribution. Some datasets were also analyzed with two publicly available SPT analysis software DPSP [107] and Spot-On [18]. In DPSP analysis the chosen range of diffusion coefficients was 10^{-3} - 10 $\mu m^2/s$. In Spot-On analysis, the number of components is set to 2 and 3 separately.

5.4.5 Fine-grained chemical rate constant inference

The ine-grained chemical rate constant inference and strains used in this Chapter are from the freddolino Lab at the University of Michigan. Please refer to corresponding manuscript of this chapter for details [106].

5.4.6 Clustering analysis for the Swi6 distributions

The spatial pattern of each mobility state was investigated using the Ripley's K function [186]

$$K(r) = \lambda^{-1} \sum_{i=1}^{n} \sum_{i \neq j} \frac{I(r_{ij} < r)}{n} \tag{5.1}$$

where r is the search radius, n is the number of points in the set, λ is the point density, and r_{ij} is the distance between the ith and jth point. $I(x)$ is an indicator function (1 when true and 0 when false). For convenience, we further normalized K(r) to attain Ripley's H function

$$H(r) = (\frac{K(r)}{\pi})^{1/2} - r \tag{5.2}$$

where $H(r) = 0$ for a random distribution, $H(r) > 0$ for a clustered distribution pattern, and $H(r) < 0$ for a dispersed pattern. In the analysis, the nucleus was approximated as a circle to determine the area and perform edge correction [197]. We calculated $H(r)$ for each cell, then we calculated an overall $H(r)$ from the average of all cells weighted by the fits density.

To eliminate effects from the spatial correlation between single-molecule steps from the same trajectories, we simulated diffusion trajectories with similar confined area size, average track length, and overall density as experimental trajectories by drawing step lengths from the step size distribution of the corresponding experiment steps. This normalization is reported in previous work [106].

5.4.7 Reconstructed single-molecule heatmap

For each cell, the nucleus and cell outlines were obtained from the fluorescence image of the nucleus and the phase-contrast image of the cell; these outlines were then approximated by a circle and a rectangle with circular caps, respectively. Every frame was analyzed by SMALL-LABS to identify single molecules, and the position and frame number of each single molecule was saved. To generate the reconstructed single-molecule heatmap for the cell, the pixel intensities after subtraction of the fitted offset in the appropriate diffraction-limited region about each single molecule were summed and the sum of all well-fit molecules was globally normalized.

5.4.8 Single-molecule time-lapse imaging

We model the binding of Chp2 and H3K9me or Swi6 to Epe1 as a direct two-component association/disassociation reaction:

$$AB \leftrightarrow A + B$$

The measured residence time of each PAmCherry-Chp2 or PAmCherry-Epe1 molecule is estimated from the lifetime of the stationary fluorescence signal. $k_{app_{d}iss}$ is acquired by fitting the probability distribution function, P, of the measured residence times, τmeasured, to a single exponential decay function:

$$P = exp(-k_{diss_{a}pp}\tau_{measured})$$

.

The measured apparent disassociation rate, $k_{app_{d}iss}$, consists of the true disassociation rate, k_{diss}, and the photobleaching rate of the PAmCherry label, $k_{bleaching}$; we separated these contributions by collecting data at multiple delay times to measure the photobleaching rate. For static molecules, we introduced a dark period with each time interval that we kept the integration time, τ_{int}, the same and introduced different lengths of dark delay times, τ_{delay}. In this way, the contribution of photobleaching was kept the same for different total time intervals, $\tau_{TL} = \tau_{int} + \tau_{delay}$. We measured the residence time $\tau_{measured} = (n - 1)\tau_{TL}$ by counting the total number of sequential frames, *n*, in which the molecule was detected. Finally, the true disassociation rate, k_{diss}, was estimated from a linear regression of the two-term relationship [211]:

$$k_{diss_app}\tau_{TL} = \tau_{TL}k_{diss} + \tau_{int}k_{bleaching}$$

This linear regression also took the uncertainty of each data point from the exponential fitting into consideration and gives the final fitted slope k_{diss} and its uncertainty.

CHAPTER VI

Conclusions

6.1 Introduction

Single-molecule tracking (SPT) is a high spatiotemporal resolution experimental tool to investigate the motion of target molecules in living systems, and nonparametric Bayesian statistics-based analysis objectively determines the number of components within the heterogeneous dynamics. Together, these advances in experimental and computational approaches make *in vivo* understanding of dynamics in living biological systems possible. In this dissertation, the epigenetic modification systems in prokaryotes and eukaryotes systems were further understood by the SPT experiment of epigenetic modification proteins and the nonparametric Bayesian statistical analysis of the SPT datasets. In this final chapter, I will summarize the conclusions from the previous chapters in the dissertation and discuss promising future directions for applying SPT to study epigenetic modifications.

6.2 Nonparametric Bayesian statistics and anomalous diffusion for SPT

In Chapter II, I developed the Bayesian SPT analysis framework called NOBIAS which combined nonparametric Bayesian and machine learning classification of anomalous diffusion. Given a single-molecule trajectories dataset, NOBIAS first uses an HDP-HMM framework to objectively determine the number of mobility states in the SPT datasets, then estimates the parameters associated with each state such as weight fraction, diffusion coefficients, and the transition matrix between states. Each single-molecule step is assigned its diffusive state label, and track segments with the same state label are classified into anomalous diffusion models with a pre-trained recurrent neural network. Simulated SPT datasets with different data qualities are applied to evaluate the performance of NOBIAS in different potential scenarios. Then NOBIAS was also applied to an experimental dataset to find the potential subdiffusion of SusG membrane protein in the human gut microbiome *Bacteroides thetaiotaomicron.* Compared with other earlier nonparametric methods which also can infer the number of mobility states from the dataset itself, for example, vbSPT and SMAUG [29,31], NOBIAS shows higher computational efficiency and stability for validating simulated datasets with ground truth. On top of that, NOBIAS also enables additional features for considering asymmetric Brownian motion and anomalous diffusion model classification.

Despite the strengths, NOBIAS is also limited by how it is designed as well. First, NOBIAS treats single-molecule steps as the basic unit of analysis and relies on assigning state labels for each step and counting transitions within the track. For very short trajectories with lengths of 2-4 steps, NOBIAS does not perform well. As a great complement, Hecker et. al. develop single-trajectory-based Bayesian frameworks DPSP and state array which are designed to deal with short tracks [107]. For the performance of both modules of NOBIAS, the longer trajectories provide better stability, yet the RNN module has a much higher requirement of track segment length as the trained neural network only stably predicts diffusion type with at least 20 steps [45]. For live cell imaging to acquire such long tracks is challenging due to z-axis detection depth and photobleaching, this limitation makes the RNN module less applicable for many datasets. Lastly, the design of the HDP-HMM module is based on the Gaussian emission function, which means two diffusive states with similar apparent diffusion coefficients would be assigned into one state with great possibility, which could lead to confusion in the RNN classification module.

A promising future direction to further improve the SPT dataset analysis is to combine the single-molecule localization and tracking part together with the trajectory analysis part together to provide more information for the statistical learning of tracking analysis. Almost all current SPT trajectories analysis take the fitted and tracked trajectories as input directly instead of analyzing the raw fluorescence movie. The trajectories input is favored for its low dimension and high computational efficiency. With the improvement of computational resources in scientific research, especially the usage of the graphical processing unit (GPU) for image processing, it is promising to use the high-dimension raw fluorescence images for diffusion and dynamics analysis directly. Previous work has applied the Bayesian framework or neural network method for single-molecule localization. I have preliminarily started the motion analysis directly from fluorescence images in collaboration with Prof. Jonathan Terhorst. In the new model, we still apply a Bayesian framework with Markov Chain Monte Carlo (MCMC) approaches, but now instead of the 1D state of step transition we monitor, we directly observe the transition between 2D images. Based on fluorescence intensity distribution in each image, the previous image, and the current 4D transition matrix, the probability of localization in each frame is determined. In this model, the localization probability 2D matrix will inform the position of the molecules in each frame, and the 4D transition matrix that projects the previous frame to the next frame contains information about the motion properties. By putting different prior on the transition matrix this model can be adapted into any motion type with Markovian property, for example, directed motion, Brownian motion, and confined Brownian motion. For anomalous diffusion though, as these diffusion models contain memories from previous frames and do not follow Markovian property, a higher dimensional transition matrix could be needed to capture them, but the computational complexity could exponentially increase. A further statistical model based on the anomalous diffusion model with proper approximation could reduce the complexity and make this movie-based SPT method applicable to anomalous diffusion.

6.3 DNA methylation and RNA-DNA hybrid regulate DNA methyltransferase localization

In Chapter III, I performed single-molecule tracking and localization for DNA methyltransferase DnmA in *Bacillus subtilis*, in WT cells and mutants including DNA replication arresting RNA-DNA hybrid cleavage disability, and two DnmA binding mutants. The two color imaging of DnmA and the replisome component DnaX suggest unmethylated DNA drive DnmA positively colocalizes with nucleoid and replisome, but also explore other regions of the cells. The decrease of correlation between DnmA and DnaX under treatment of DNA replication inhibitor further validates that colocalization is driven by newly replicated unmethylated DNA. The negative correlation between DnmA and DnaX under increased RNA-DNA hybrid and DnmA sequence recognition mutant 6AA* suggest that DNA interactions are necessary for proper DnmA localization and correlation with DnaX. The dynamics analysis only shows a subtle decrease in dynamics under these disturbances compared with WT DnmA, suggesting that DnmA dynamics only partially reflect its functional status. These single-molecule results together with other biochemical and bioinformatics assays conclude that DnmA is part of a remnant of a nonfunctional restriction-modification system.

Bacillus subtilis has been widely studied as a model gram-positive bacteria. DnmA is the only DNA methyltransferase in *B. subtilis*, and the single-molecule tracking and localization experiments of DnmA inform us greatly about how the nucleic acid components in the live cell could change important modification processes. RnhB and RnhC are two RNase H proteins in *B. subtilis* that are responsible for the cleavage of DNA-RNA hybrids such as R-loops, complementary RNA-DNA, ribopatches, and Okazaki fragments [138, 215]. As was shown in Chapter III, Δ*rnhC* will increase the R-loop and significantly affect the stability of the genome. To understand the role RnhB and RnhC play in the cleavage of different RNA-DNA hybrid, *in vivo* dynamics and localization information of these two RNase H would help provide crucial information that *in vitro* assay may neglect. How the RNA-DNA hybrid cleavage is carried out together with other modifications such as DNA replication, repair, and RNA transcription should be investigated through a comparison of single-molecule dynamics and localizations after disturbance in these processes. The development in multi-color imaging and correlation analysis in multiple fluorescence channels could provide a more powerful imaging tool to understand RNA-DNA hybrid cleavage in *Bacillus subtilis*.

6.4 Oligomerization of HP1 protein provides a tunable heterochromatin localization mechanism

In Chapter IV, I investigated the single-molecule dynamics of an HP1 protein Swi6 in the yeast model system *Schizosaccharomyces pombe*. With the previously invented Bayesian SPT analysis framework SMAUG, the Swi6 *in vivo* dynamics display 4 mobility states. With the design of mutant strains for comparison, these 4 biophysical mobility states are mapped into their biochemical meanings with the following correspondences: a high mobility state close to free motion in cells (δ), a nucleic acid-binding state (γ), a chromatin-sampling state that involves weak binding to H3K9me0 and H3K9me3 nucleosomes (β), and a stable H3K9me3-dependent bound state (α). Swi6 recognizes the H3K9me substrate with low binding affinity but high specificity, with the SMAUG analysis of Swi6 dynamics under different oligomerization states, it is demonstrated that Swi6 uses its oligomerization through the CSD domain to multivalently increase the concentration of its H3K9me recognition CD domain. Loss of oligomerization results in a deficiency in the H3K9me binding state of Swi6, and engineered 2×CD on top of the loss of oligomerization can recover the Swi6 H3K9me3 binding state. In the application of the SMAUG framework, the transition between states and the assignment of mobility state labels at the single-step level is acquired in addition to the diffusion coefficients and weight fractions for each state. The transition rate between states are further input to the BSL model for the extraction of the transition chemical rate, and the assigning of state label enable the mapping of mobility states to their cellular spatial distribution. The single-molecule spatial localization maps of Swi6 mobility states are analyzed through Ripley's K clustering method to reveal that the clustering level for these four states decreases with the increase of dynamics, which further validates their biochemical meanings.

Swi6's nucleic acid binding state represents the hinge region's nucleic acid binding affinity, validated by the Swi6^{hinge} mutant. Further, single-molecule tracking experiments and FLAG IP assay demonstrate that the nucleic acid binding state is directly complete with oligomerization. Consequence of nucleic acid binding promotes release from the H3K9me while oligomerization promotes H3K9me binding and heterochromatin formation. The competition and balance between oligomerization and nucleic acid binding establish a tunable heterochromatin localization mechanism for Swi6's H3K9me binding. The formation of heterochromatin could be involved in the equilibrium state of Swi6's oligomerization to promote its H3K9me binding. Besides the biological interpretation and insights that are further understood regarding Swi6's H3K9me recognition mechanism, this chapter displays that the methodology combining *in vivo* SPT and nonparametric Bayesian statistics could put a step further to the goal of live cell quantitative biochemistry. The methodology can be adapted for similar microbe and mammalian cell systems to study various systems and proteins of interest.

In the future, further engineering of Swi6 could inform us even more regarding the H3K9me epigenetic modification systems in *S. pombe*. HP1α proteins are reported to form biocondensates by liquid-liquid phase separation *in vitro* and *in vivo* [79, 169, 170]. However, full physical characterization and how the potential phase separation enables HP1 protein's regulatory function remains unknown. An allele of HP1α Swi6—*swi6-sm1* displays disruption of the condensation-like domain of Swi6 with normal CD-H3K9me binding and CSD dimerization [216]. Single-molecule tracking of Swi6-sm1 could reveal the dynamics changes in this allele and could further characterize the potential biological condensation mechanism for HP1 proteins. In addition to Swi6-sm1, a point mutation at 278 threonine to Lysine (T278K) shows complete loss of epigenetic maintenance even when epe1 is deleted, while Tyrosine (T278Y) or Phenylalanine (T278F) shows enhanced maintenance. It has been shown that Swi6-Epe1 interaction is completely disrupted under these T278 mutants. The *in vivo* dynamics of these mutants together with other *in vitro* biochemistry assays could further reveal how Swi6 and Epe1 enable the establishment and maintenance of epigenetic memory.

6.5 H3K9 methylation reinforces heterochromatin-specific complex assembly

In Chapter V, I present a mechanism where histone modification protein complexes assembly at heterochromatin site regulated by H3K9me and two HP1 proteins in *S. pombe.* Following Chapter IV, I further use single-molecule tracking and a new Bayesian SPT analysis framework NOBIAS to characterize the other HP1 protein Chp2 *in vivo*. Chp2 displays a very static dynamics pattern with more than 93% of steps assigned to a bound state and a static spatial pattern with clear foci within the nucleus. The single-molecule time-lapse imaging of Chp2 and Bayesian synthetic likelihood simulation based on SPT transition from NOBIAS together quantify the dissociation rate of Chp2 *in vivo* is faster than reported *in vitro* rate, yet Chp2's H3K9me bound is more stable than Swi6. The distinct expression level and H3K9me binding affinity for Swi6 and Chp2 reveal a competitive and cooperative relationship between these two HP1 proteins.

HP1 proteins recognize H3K9mes and recruit other proteins to heterochromatin. Epe1 is a putative demethylase and anti-silencing factor which prevent the spreading of heterochromatin [80, 81]. Epe1 is shown to bind to Swi6 *in vitro*, and I have shown with single-molecule tracking of Epe1 that more than 92% of Epe1 is in static mobility and colocalize with Swi6 and heterochromatin region. With the disruption of H3K9me and Swi6, we demonstrate that H3K9me attenuates off-heterochromatin Swi6-Epe1 interactions that only H3K9me bound Swi6 recruits and binds to Epe1 in the live cell. The other HP1 protein Chp2 is known to recruit the SHREC complex which consists of histone remodeler Mit1 and H3K14 deacetylase Clr3 [83, 86]. The single-molecule imaging of Mit1 and Clr3 under different H3K9me and HP1 proteins disruption shows that SHREC complex only assembly at the heterochromatin site, and H3K9me is necessary for Chp2 to associate with SHREC complex. Using the Epe1-Swi6 and SHREC-Chp2 interaction as two examples, I propose a general mechanism for HP1 protein complexes, where heterochromatin is a necessary substrate and location for these protein complexes' assembly. Despite *in vitro* assays and structural biology evidence have demonstrated how these proteins can pairwise interact and assemble into complexes without the presence of histone or heterochromatin, *in vivo* data show that H3K9 methylation reinforces heterochromatin specific HP1 protein complex assembly and attenuates promiscuous protein interactions outside.

In the future, multiple directions within this chapter should be further explored. The first is to examine if this heterochromatin-mediated complex assembly mechanism extends to more HP1 protein complexes. Chp2's major role in *S. pombe* is to interact with the SHREC complex, yet Swi6 has much broader functions and various proteins are bound to Swi6, for example, CHD family chromatin remodeler Hrp3, meiotic cohesin protection protein Sgo1, histone chaperone FACT complex component Spt16 [87, 216]. Single-molecule dynamics of these proteins *in vivo* and under disruption of H3K9me and Swi6 could reveal the role H3K9me and heterochromatin plays in their assembly. The other promising follow-up of this project is to look more into the SHREC complex. Data in Chapter V supports the model where the SHREC complex is recruited to Chp2 through

Chp2-Mit1-Clr1-Clr2/Clr3 order. However, it is also shown that the bound state of Mit1 and Clr3 decreases but remains with the deletion of Chp2 or the entire H3K9me. Mit1 and Clr3 have other pathways to be recruited to chromatin and retain a bound state [83]. Preliminary data has shown that double deletion of Clr4 and an H3K4 methyltransferase Set1 will result in the complete loss of the Mit1 bound state. This result indicates that Mit1 could have an H3K4 methylation-related function and recruitment pathway. Besides, the role of Clr1 and Clr2 plays in the recruitment and assembly *in vivo* remains unclear. Clr1 is also known to bind to Chp2 and structurally connects Mit1 and Clr3 in the complex, and Clr2 can bind to nuclei acid and potentially has sequence-specific recognition of DNA which could recruit Clr3 to target DNA sequence [204]. Advances in multicolor imaging could further help the understanding of the SHREC complex. The further study of Mit1 and Clr3 recruitment mechanisms could be widely applied to other proteins that have multiple recruitment pathways to the chromatin site.

6.6 Overarching Conclusions

In this dissertation, I have further extended single-molecule tracking into nonparametric, asymmetric, and anomalous diffusion to understand the heterogeneous dynamics of target molecules in living systems. The advantage of nonparametric Bayesian statistics analysis for *in vivo* SPT datasets is that it can objectively determine the number of mobility states for the biomolecule of interest. Under the assumption that each distinct mobility state corresponds to a different biochemical function, this dissertation presents a methodology that quantifies *in vivo* biochemistry by combinding *in vivo* SPT and nonparametric Bayesian statistics analysis. For complicated systems like the fission yeast epigenetic modification systems, previous SPT analysis methods could very easily overlook the potential biochemical meanings of the intermediate states. Nonparametric methods like SMAUG and NOBIAS enable a much more decisive comparison between the dynamics of target proteins under different conditions by monitoring the disappearance or emergence of a biochemical state rather than being limited to only an overall increase or decrease in dynamics. The diffusivity of proteins is set as a benchmark in live cells to compare the function and existing form of the same proteins and also between different proteins. The single-step level diffusive state label assignment also opens up the road to linking dynamics with the spatial distribution and would be promising if combined with two-color imaging. The transition matrix from the Bayesian analysis could also be used to infer the chemical rate as reported in Chapters IV and V. This methodology only requires trajectories as input and is widely applicable to other biological systems and other experimental setups, not limited to microbes and sptPALM. However, this methodology also has limitations. The first limitation comes from the assumption that distinct mobility states corresponding to different biochemical functions could be false for some molecules. However, within our resolution, two states with the same mobility could carry out different biochemical functions, and proteins engaged in a single biochemical function could have a widely distributed mobility. The second limitation comes along with the discretely inferred the number of the mobility state. In this discretized interpretation, a two-state result would be completely different from a three-state one, but the dynamics between the two datasets could be similar. In other words, the difference between 2.49 versus 2.51 could be exaggerated to 2 versus 3 in this approach. More often in these cases, the framework itself will not be consistent under multiple parallel runs. Knowledge and experiences with Bayesian statistics are strongly recommended to prevent overinterpretation of the results.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] H. H. Tuson, J. S. Biteen, "Unveiling the Inner Workings of Live Bacteria Using Super-Resolution Microscopy", *Analytical Chemistry* **87**, 42 (2015). DOI: 10.1021/ac5041346.

[2] Z. Liu, L. Lavis, E. Betzig, "Imaging Live-Cell Dynamics and Structure at the Single-Molecule Level", *Molecular Cell* **58**, 644 (2015). DOI: 10.1016/j.molcel.2015.02.033.

[3] F. V. Subach, *et al.*, "Photoactivatable mCherry for high-resolution two-color fluorescence microscopy", *Nature Methods* **6**, 153 (2009). DOI: 10.1038/nmeth.1298.

[4] J. B. Grimm, *et al.*, "Bright photoactivatable fluorophores for single-molecule imaging", *Nature Methods* **13**, 985 (2016). DOI: 10.1038/nmeth.4034.

[5] J. B. Grimm, *et al.*, "A general method to improve fluorophores for live-cell and single-molecule microscopy", *Nature Methods* **12**, 244 (2015). DOI: 10.1038/nmeth.3256.

[6] G. V. Los, *et al.*, "HaloTag: A Novel Protein Labeling Technology for Cell Imaging and Protein Analysis", *ACS Chemical Biology* **3**, 373 (2008). DOI: 10.1021/cb800025k.

[7] W. E. Moerner, L. Kador, "Optical detection and spectroscopy of single molecules in a solid", *Physical Review Letters* **62**, 2535 (1989). DOI: 10.1103/PhysRevLett.62.2535.

[8] S. W. Hell, J. Wichmann, "Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy", *Optics Letters* **19**, 780 (1994). DOI: 10.1364/OL.19.000780.

[9] E. Betzig, *et al.*, "Imaging Intracellular Fluorescent Proteins at Nanometer Resolution", *Science* **313**, 1642 (2006). DOI: 10.1126/science.1127344.

[10] M. J. Rust, M. Bates, X. Zhuang, "Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM)", *Nature Methods* **3**, 793 (2006). DOI: 10.1038/nmeth929.

[11] S. Manley, *et al.*, "High-density mapping of single-molecule trajectories with photoactivated localization microscopy", *Nature Methods* **5**, 155 (2008). DOI: 10.1038/nmeth.1176.

[12] B. P. Isaacoff, Y. Li, S. A. Lee, J. S. Biteen, "SMALL-LABS: Measuring Single-Molecule Intensity and Position in Obscuring Backgrounds", *Biophysical Journal* **116**, 975 (2019). DOI: 10.1016/j.bpj.2019.02.006.

[13] M. Saxton, “Single-particle tracking: the distribution of diffusion coefficients”, *Biophysical Journal* **72**, 1744 (1997). DOI: 10.1016/S0006-3495(97)78820-9.

[14] H. Qian, M. Sheetz, E. Elson, “Single particle tracking. Analysis of diffusion and flow in two-dimensional systems”, *Biophysical Journal* **60**, 910 (1991). DOI: 10.1016/S0006-3495(91)82125-7.

[15] X. Michalet, A. J. Berglund, “Optimal diffusion coefficient estimation in single-particle tracking”, *Physical Review E* **85**, 061916 (2012). DOI: 10.1103/PhysRevE.85.061916.

[16] D. Mazza, A. Abernathy, N. Golob, T. Morisaki, J. G. McNally, “A benchmark for chromatin binding measurements in live cells”, *Nucleic Acids Research* **40**, e119 (2012). DOI: 10.1093/nar/gks701.

[17] D. J. Rowland, J. S. Biteen, “Measuring molecular motions inside single cells with improved analysis of single-particle trajectories”, *Chemical Physics Letters* **674**, 173 (2017). DOI: 10.1016/j.cplett.2017.02.052.

[18] A. S. Hansen, *et al.*, “Robust model-based analysis of single-particle tracking experiments with Spot-On”, *eLife* **7**, e33125 (2018). DOI: 10.7554/eLife.33125.

[19] B. Hebert, S. Costantino, P. W. Wiseman, “Spatiotemporal Image Correlation Spectroscopy (STICS) Theory, Verification, and Application to Protein Velocity Mapping in Living CHO Cells”, *Biophysical Journal* **88**, 3601 (2005). DOI: 10.1529/biophysj.104.054874.

[20] K. Bacia, S. A. Kim, P. Schwille, “Fluorescence cross-correlation spectroscopy in living cells”, *Nature Methods* **3**, 83 (2006). DOI: 10.1038/nmeth822.

[21] L. Xiang, K. Chen, R. Yan, W. Li, K. Xu, “Single-molecule displacement mapping unveils nanoscale heterogeneities in intracellular diffusivity”, *Nature Methods* **17**, 524 (2020). DOI: 10.1038/s41592-020-0793-0.

[22] M. E. Beheiry, M. Dahan, J.-B. Masson, “InferenceMAP: mapping of single-molecule dynamics with Bayesian inference”, *Nature Methods* **12**, 594 (2015). DOI: 10.1038/nmeth.3441.

[23] E. P. Perillo, *et al.*, “Deep and high-resolution three-dimensional tracking of single particles using nonlinear and multiplexed illumination”, *Nature Communications* **6**, 7874 (2015). DOI: 10.1038/ncomms8874.

[24] F. Balzarotti, *et al.*, “Nanometer resolution imaging and tracking of fluorescent molecules with minimal photon fluxes”, *Science* **355**, 606 (2017). DOI: 10.1126/science.aak9913.

[25] J. Cnossen, *et al.*, “Localization microscopy at doubled precision with patterned illumination”, *Nature Methods* **17**, 59 (2020). DOI: 10.1038/s41592-019-0657-7.

[26] S. Hou, J. Exell, K. Welsher, "Real-time 3D single molecule tracking", *Nature Communications* **11**, 3607 (2020). DOI: 10.1038/s41467-020-17444-6.

[27] P. Jouchet, *et al.*, "Nanometric axial localization of single fluorescent molecules with modulated excitation", *Nature Photonics* **15**, 297 (2021). DOI: 10.1038/s41566-020-00749-9.

[28] L. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition", *Proceedings of the IEEE* **77**, 257 (1989). DOI: 10.1109/5.18626.

[29] F. Persson, M. Lindén, C. Unoson, J. Elf, "Extracting intracellular diffusive states and transition rates from single-molecule tracking data", *Nature Methods* **10**, 265 (2013). DOI: 10.1038/nmeth.2367.

[30] N. Monnier, *et al.*, "Inferring transient particle transport dynamics in live cells", *Nature Methods* **12**, 838 (2015). DOI: 10.1038/nmeth.3483.

[31] J. D. Karslake, *et al.*, "SMAUG: Analyzing single-molecule tracks with nonparametric Bayesian statistics", *Methods* **193**, 16 (2021). DOI: 10.1016/j.ymeth.2020.03.008.

[32] M. J. Johnson, A. S. Willsky, "Bayesian Nonparametric Hidden Semi-Markov Models", *Journal of Machine Learning Research* **14**, 673 (2013).

[33] T. S. Ferguson, "A Bayesian Analysis of Some Nonparametric Problems", *The Annals of Statistics* **1**, 209 (1973).

[34] Y. W. Teh, M. I. Jordan, M. J. Beal, D. M. Blei, "Hierarchical Dirichlet Processes", *Journal of the American Statistical Association* **101**, 1566 (2006). DOI: 10.1198/016214506000000302.

[35] C. P. Robert, *The Bayesian Choice*, Springer Texts in Statistics (New York, NY, 1994).

[36] W. R. Gilks, S. Richardson, D. Spiegelhalter, *Markov Chain Monte Carlo in Practice* (1995).

[37] E. B. Fox, E. B. Sudderth, M. I. Jordan, A. S. Willsky, *Proceedings of the 25th international conference on Machine learning - ICML '08* (Helsinki, Finland, 2008), pp. 312–319.

[38] R. Metzler, J.-H. Jeon, A. G. Cherstvy, E. Barkai, "Anomalous diffusion models and their properties: non-stationarity, non-ergodicity, and ageing at the centenary of single particle tracking", *Phys. Chem. Chem. Phys.* **16**, 24128 (2014). DOI: 10.1039/C4CP03465A.

[39] J.-H. Jeon, M. Javanainen, H. Martinez-Seara, R. Metzler, I. Vattulainen, "Protein Crowding in Lipid Bilayers Gives Rise to Non-Gaussian Anomalous Lateral Diffusion of Phospholipids and Proteins", *Physical Review X* **6**, 021006 (2016). DOI: 10.1103/PhysRevX.6.021006.

[40] A. Caspi, R. Granek, M. Elbaum, “Diffusion and directed motion in cellular transport”, *Physical Review E* **66**, 011916 (2002). DOI: 10.1103/PhysRevE.66.011916.

[41] M. Bauer, R. Metzler, “In Vivo Facilitated Diffusion Model”, *PLoS ONE* **8**, e53956 (2013). DOI: 10.1371/journal.pone.0053956.

[42] C. Manzo, M. F. Garcia-Parajo, “A review of progress in single particle tracking: from methods to biophysical insights”, *Reports on Progress in Physics* **78**, 124601 (2015). DOI: 10.1088/0034-4885/78/12/124601.

[43] S. Bo, F. Schmidt, R. Eichhorn, G. Volpe, “Measurement of anomalous diffusion using recurrent neural networks”, *Physical Review E* **100**, 010102 (2019). DOI: 10.1103/PhysRevE.100.010102.

[44] N. Granik, *et al.*, “Single-Particle Diffusion Characterization by Deep Learning”, *Biophysical Journal* **117**, 185 (2019). DOI: 10.1016/j.bpj.2019.06.015.

[45] A. Argun, G. Volpe, S. Bo, “Classification, inference and segmentation of anomalous diffusion with recurrent neural networks”, *Journal of Physics A: Mathematical and Theoretical* (2021). DOI: 10.1088/1751-8121/ac070a.

[46] G. Muñoz-Gil, *et al.*, “AnDi: The Anomalous Diffusion Challenge”, *Emerging Topics in Artificial Intelligence 2020* p. 44 (2020). DOI: 10.1117/12.2567914.

[47] G. Muñoz-Gil, *et al.*, “Objective comparison of methods to decode anomalous diffusion”, *arXiv:2105.06766 [cond-mat, physics:physics, q-bio]* (2021).

[48] D. Moazed, “Mechanisms for the Inheritance of Chromatin States”, *Cell* **146**, 510 (2011). DOI: 10.1016/j.cell.2011.07.013.

[49] G. Felsenfeld, “A Brief History of Epigenetics”, *Cold Spring Harbor Perspectives in Biology* **6**, a018200 (2014). DOI: 10.1101/cshperspect.a018200.

[50] F. Mohn, D. Schübeler, “Genetics and epigenetics: stability and plasticity during cellular differentiation”, *Trends in Genetics* **25**, 129 (2009). DOI: 10.1016/j.tig.2008.12.005.

[51] J. L. Miller, P. A. Grant, *Epigenetics: Development and Disease*, T. K. Kundu, T. K. Kundu, eds. (Dordrecht, 2013), vol. 61, pp. 289–317.

[52] D. Dussoix, W. Arber, “Host specificity of DNA produced by Escherichia coli”, *Journal of Molecular Biology* **11**, 238 (1965). DOI: 10.1016/S0022-2836(65)80054-7.

[53] K. Vasu, V. Nagaraja, “Diverse Functions of Restriction-Modification Systems in Addition to Cellular Defense”, *Microbiology and Molecular Biology Reviews* **77**, 53 (2013). DOI: 10.1128/MMBR.00044-12.

[54] S. I. S. Grewal, D. Moazed, “Heterochromatin and Epigenetic Control of Gene Expression”, *Science* **301**, 798 (2003). DOI: 10.1126/science.1086887.

[55] K. Hyun, J. Jeon, K. Park, J. Kim, “Writing, erasing and reading histone lysine methylations”, *Experimental & Molecular Medicine* **49**, e324 (2017). DOI: 10.1038/emm.2017.11.

[56] M. R. Motamedi, *et al.*, “Two RNAi Complexes, RITS and RDRC, Physically Interact and Localize to Noncoding Centromeric RNAs”, *Cell* **119**, 789 (2004). DOI: 10.1016/j.cell.2004.11.034.

[57] A. Verdel, *et al.*, “RNAi-Mediated Targeting of Heterochromatin by the RITS Complex”, *Science* **303**, 672 (2004). DOI: 10.1126/science.1093686.

[58] K. Ragunathan, G. Jih, D. Moazed, “Epigenetic inheritance uncoupled from sequence-specific recruitment”, *Science* **348**, 1258699 (2015). DOI: 10.1126/science.1258699.

[59] M. A. Sánchez-Romero, J. Casadesús, “The bacterial epigenome”, *Nature Reviews Microbiology* **18**, 7 (2020). DOI: 10.1038/s41579-019-0286-2.

[60] J. Casadesús, D. Low, “Epigenetic Gene Regulation in the Bacterial World”, *Microbiology and Molecular Biology Reviews* **70**, 830 (2006). DOI: 10.1128/MMBR.00016-06.

[61] A. Løbner-Olesen, O. Skovgaard, M. G. Marinus, “Dam methylation: coordinating cellular processes”, *Current Opinion in Microbiology* **8**, 154 (2005). DOI: 10.1016/j.mib.2005.02.009.

[62] A. K. Thakur, L. Movileanu, “Real-time measurement of protein–protein interactions at single-molecule resolution using a biological nanopore”, *Nature Biotechnology* **37**, 96 (2019). DOI: 10.1038/nbt.4316.

[63] T. M. Nye, *et al.*, “Methyltransferase DnmA is responsible for genome-wide N6-methyladenosine modifications at non-palindromic recognition sites in Bacillus subtilis”, *Nucleic Acids Research* **48**, 5332 (2020). DOI: 10.1093/nar/gkaa266.

[64] Y. Li, Z. Chen, L. A. Matthews, L. A. Simmons, J. S. Biteen, “Dynamic Exchange of Two Essential DNA Polymerases during Replication and after Fork Arrest”, *Biophysical Journal* **116**, 684 (2019). DOI: 10.1016/j.bpj.2019.01.008.

[65] Y. Liao, J. W. Schroeder, B. Gao, L. A. Simmons, J. S. Biteen, “Single-molecule motions and interactions in live cells reveal target search dynamics in mismatch repair”, *Proceedings of the National Academy of Sciences* **112** (2015). DOI: 10.1073/pnas.1507386112.

[66] Y. Liao, Y. Li, J. W. Schroeder, L. A. Simmons, J. S. Biteen, “Single-Molecule DNA Polymerase Dynamics at a Bacterial Replisome in Live Cells”, *Biophysical Journal* **111**, 2562 (2016). DOI: 10.1016/j.bpj.2016.11.006.

[67] D. Nicetto, K. S. Zaret, “Role of H3K9me3 heterochromatin in cell identity establishment and maintenance”, *Current Opinion in Genetics & Development* **55**, 1 (2019). DOI: 10.1016/j.gde.2019.04.013.

[68] E. L. Greer, Y. Shi, “Histone methylation: a dynamic mark in health, disease and inheritance”, *Nature Reviews Genetics* **13**, 343 (2012). DOI: 10.1038/nrg3173.

[69] K. Fukuda, *et al.*, “Regulation of mammalian 3D genome organization and histone H3K9 dimethylation by H3K9 methyltransferases”, *Communications Biology* **4**, 1 (2021). DOI: 10.1038/s42003-021-02089-y.

[70] Q. Bian, E. C. Anderson, Q. Yang, B. J. Meyer, “Histone H3K9 methylation promotes formation of genome compartments in Caenorhabditis elegans via chromosome compaction and perinuclear anchoring”, *Proceedings of the National Academy of Sciences* **117**, 11459 (2020). DOI: 10.1073/pnas.2002068117.

[71] R. C. Allshire, K. Ekwall, “Epigenetic Regulation of Chromatin States in *Schizosaccharomyces pombe*”, *Cold Spring Harbor Perspectives in Biology* **7**, a018770 (2015). DOI: 10.1101/cshperspect.a018770.

[72] K. Ekwall, *et al.*, “Mutations in the fission yeast silencing factors clr4+ and rik1+ disrupt the localisation of the chromo domain protein Swi6p and impair centromere function”, *Journal of Cell Science* **109**, 2637 (1996). DOI: 10.1242/jcs.109.11.2637.

[73] D. Canzio, A. Larson, G. J. Narlikar, “Mechanisms of functional promiscuity by HP1 proteins”, *Trends in Cell Biology* **24**, 377 (2014). DOI: 10.1016/j.tcb.2014.01.002.

[74] D. Canzio, *et al.*, “Chromodomain-Mediated Oligomerization of HP1 Suggests a Nucleosome-Bridging Mechanism for Heterochromatin Assembly”, *Molecular Cell* **41**, 67 (2011). DOI: 10.1016/j.molcel.2010.12.016.

[75] L. C. Bryan, *et al.*, “Single-molecule kinetic analysis of HP1-chromatin binding reveals a dynamic network of histone modification and DNA interactions”, *Nucleic Acids Research* **45**, 10504 (2017). DOI: 10.1093/nar/gkx697.

[76] J. F. Smothers, S. Henikoff, “The Hinge and Chromo Shadow Domain Impart Distinct Targeting of HP1-Like Proteins”, *Molecular and Cellular Biology* **21**, 2555 (2001). DOI: 10.1128/MCB.21.7.2555-2569.2001.

[77] L. Schmiedeberg, K. Weisshart, S. Diekmann, G. Meyer zu Hoerste, P. Hemmerich, “High- and Low-mobility Populations of HP1 in Heterochromatin of Mammalian Cells”, *Molecular Biology of the Cell* **15**, 2819 (2004). DOI: 10.1091/mbc.e03-11-0827.

[78] K. P. Müller, *et al.*, “Multiscale Analysis of Dynamics and Interactions of Heterochromatin Protein 1 by Fluorescence Fluctuation Microscopy”, *Biophysical Journal* **97**, 2876 (2009). DOI: 10.1016/j.bpj.2009.08.057.

[79] S. Sanulli, *et al.*, “HP1 reshapes nucleosome core to promote phase separation of heterochromatin”, *Nature* **575**, 390 (2019). DOI: 10.1038/s41586-019-1669-2.

[80] M. Zofall, S. I. Grewal, “Swi6/HP1 Recruits a JmjC Domain Protein to Facilitate Transcription of Heterochromatic Repeats”, *Molecular Cell* **22**, 681 (2006). DOI: 10.1016/j.molcel.2006.05.010.

[81] N. Ayoub, *et al.*, "A Novel jmjC Domain Protein Modulates Heterochromatization in Fission Yeast", *Molecular and Cellular Biology* **23**, 4356 (2003). DOI: 10.1128/MCB.23.12.4356-4370.2003.

[82] R. S. Isaac, *et al.*, "Biochemical Basis for Distinct Roles of the Heterochromatin Proteins Swi6 and Chp2", *Journal of Molecular Biology* **429**, 3666 (2017). DOI: 10.1016/j.jmb.2017.09.012.

[83] G. Job, *et al.*, "SHREC Silences Heterochromatin via Distinct Remodeling and Deacetylation Modules", *Molecular Cell* **62**, 207 (2016). DOI: 10.1016/j.molcel.2016.03.016.

[84] O. Aygün, S. Mehta, S. I. S. Grewal, "HDAC-mediated suppression of histone turnover promotes epigenetic stability of heterochromatin", *Nature Structural & Molecular Biology* **20**, 547 (2013). DOI: 10.1038/nsmb.2565.

[85] S. Bornelöv, *et al.*, "The Nucleosome Remodeling and Deacetylation Complex Modulates Chromatin Structure at Sites of Active Transcription to Fine-Tune Gene Expression", *Molecular Cell* **71**, 56 (2018). DOI: 10.1016/j.molcel.2018.06.003.

[86] T. Sugiyama, *et al.*, "SHREC, an Effector Complex for Heterochromatic Transcriptional Silencing", *Cell* **128**, 491 (2007). DOI: 10.1016/j.cell.2006.12.035.

[87] M. R. Motamedi, *et al.*, "HP1 Proteins Form Distinct Complexes and Mediate Heterochromatic Gene Silencing by Nonoverlapping Mechanisms", *Molecular Cell* **32**, 778 (2008). DOI: 10.1016/j.molcel.2008.10.026.

[88] K. Bao, C.-M. Shan, J. Moresco, J. Yates, S. Jia, "Anti-silencing factor Epe1 associates with SAGA to regulate transcription within heterochromatin", *Genes & Development* **33**, 116 (2019). DOI: 10.1101/gad.318030.118.

[89] S. I. S. Grewal, S. Jia, "Heterochromatin revisited", *Nature Reviews Genetics* **8**, 35 (2007). DOI: 10.1038/nrg2008.

[90] S. T. Hess, T. P. Girirajan, M. D. Mason, "Ultra-High Resolution Imaging by Fluorescence Photoactivation Localization Microscopy", *Biophysical Journal* **91**, 4258 (2006). DOI: 10.1529/biophysj.106.091116.

[91] A. Yildiz, "Myosin V Walks Hand-Over-Hand: Single Fluorophore Imaging with 1.5-nm Localization", *Science* **300**, 2061 (2003). DOI: 10.1126/science.1084398.

[92] J. Deich, E. M. Judd, H. H. McAdams, W. E. Moerner, "Visualization of the movement of single histidine kinase molecules in live Caulobacter cells", *Proceedings of the National Academy of Sciences* **101**, 15921 (2004). DOI: 10.1073/pnas.0404200101.

[93] S. Elmore, M. Müller, N. Vischer, T. Odijk, C. L. Woldringh, "Single-particle tracking of oriC-GFP fluorescent spots during chromosome segregation in Escherichia coli", *Journal of Structural Biology* **151**, 275 (2005). DOI: 10.1016/j.jsb.2005.06.004.

[94] I. Izeddin, *et al.*, “Single-molecule tracking in live cells reveals distinct target-search strategies of transcription factors in the nucleus”, *eLife* **3**, e02230 (2014). DOI: 10.7554/eLife.02230.

[95] A. Badrinarayanan, R. Reyes-Lamothe, S. Uphoff, M. C. Leake, D. J. Sherratt, “In Vivo Architecture and Action of Bacterial Structural Maintenance of Chromosome Proteins”, *Science* **338**, 528 (2012). DOI: 10.1126/science.1227126.

[96] H. Y. Park, *et al.*, “Visualization of Dynamics of Single Endogenous mRNA Labeled in Live Mouse”, *Science* **343**, 422 (2014). DOI: 10.1126/science.1239200.

[97] C. A. Bayas, *et al.*, “Spatial organization and dynamics of RNase E and ribosomes in Caulobacter crescentus”, *Proceedings of the National Academy of Sciences* **115**, E3712 (2018). DOI: 10.1073/pnas.1721648115.

[98] G. Schütz, H. Schindler, T. Schmidt, “Single-molecule microscopy on model membranes reveals anomalous diffusion”, *Biophysical Journal* **73**, 1073 (1997). DOI: 10.1016/S0006-3495(97)78139-6.

[99] A. Kusumi, Y. Sako, M. Yamamoto, “Confined lateral diffusion of membrane receptors as studied by single particle tracking (nanovid microscopy). Effects of calcium-induced differentiation in cultured epithelial cells”, *Biophysical Journal* **65**, 2021 (1993). DOI: 10.1016/S0006-3495(93)81253-0.

[100] A. J. Berglund, “Statistics of camera-based single-particle tracking”, *Physical Review E* **82**, 011917 (2010). DOI: 10.1103/PhysRevE.82.011917.

[101] H. Deschout, K. Neyts, K. Braeckmans, “The influence of movement on the localization precision of sub-resolution particles in fluorescence microscopy”, *Journal of Biophotonics* **5**, 97 (2012). DOI: 10.1002/jbio.201100078.

[102] M. Lindén, V. Ćurić, E. Amselem, J. Elf, “Pointwise error estimates in localization microscopy”, *Nature Communications* **8**, 15115 (2017). DOI: 10.1038/ncomms15115.

[103] J. Elf, G.-W. Li, X. S. Xie, “Probing Transcription Factor Dynamics at the Single-Molecule Level in a Living Cell”, *Science* **316**, 1191 (2007). DOI: 10.1126/science.1141967.

[104] A. S. Hansen, I. Pustova, C. Cattoglio, R. Tjian, X. Darzacq, “CTCF and cohesin regulate chromatin loop stability with distinct dynamics”, *eLife* **6**, e25776 (2017). DOI: 10.7554/eLife.25776.

[105] T. Sungkaworn, *et al.*, “Single-molecule imaging reveals receptor–G protein interactions at cell surface hot spots”, *Nature* **550**, 543 (2017). DOI: 10.1038/nature24264.

[106] S. Biswas, *et al.*, “HP1 oligomerization compensates for low-affinity H3K9me recognition and provides a tunable mechanism for heterochromatin-specific localization”, *Science Advances* **8**, eabk0793 (2022). DOI: 10.1126/sciadv.abk0793.

[107] A. Heckert, L. Dahal, R. Tijan, X. Darzacq, “Recovering mixtures of fast-diffusing states from short single-particle trajectories”, *eLife* **11**, e70169 (2022). DOI: 10.7554/eLife.70169.

[108] M. Bauer, R. Metzler, “Generalized Facilitated Diffusion Model for DNA-Binding Proteins with Search and Recognition States”, *Biophysical Journal* **102**, 2321 (2012). DOI: 10.1016/j.bpj.2012.04.008.

[109] A. Gentili, G. Volpe, “Characterization of anomalous diffusion classical statistics powered by deep learning (CONDOR)”, *Journal of Physics A: Mathematical and Theoretical* (2021). DOI: 10.1088/1751-8121/ac0c5d.

[110] J. Sethuraman, “A Constructive Definition of Dirichlet Priors”, *Statistica Sinica* **4**, 639 (1994).

[111] H. Ishwaran, M. Zarepour, “Dirichlet Prior Sieves in Finite Normal Mixtures”, *Statistica Sinica* **12**, 941 (2002).

[112] E. B. Fox, E. B. Sudderth, A. S. Willsky, *2007 10th International Conference on Information Fusion* (Quebec City, QC, Canada, 2007), pp. 1–8.

[113] J. Van Gael, Y. Saatci, Y. W. Teh, Z. Ghahramani, *Proceedings of the 25th international conference on Machine learning - ICML ’08* (Helsinki, Finland, 2008), pp. 1088–1095.

[114] A. Gelman, A. Gelman, eds., *Bayesian data analysis*, Texts in statistical science (Boca Raton, Fla, 2004), second edn.

[115] B. B. Mandelbrot, J. W. Van Ness, “Fractional Brownian Motions, Fractional Noises and Applications”, *SIAM Review* **10**, 422 (1968). DOI: 10.1137/1010093.

[116] H. Scher, E. W. Montroll, “Anomalous transit-time dispersion in amorphous solids”, *Physical Review B* **12**, 2455 (1975). DOI: 10.1103/PhysRevB.12.2455.

[117] J. Klafter, G. Zumofen, “Lévy statistics in a Hamiltonian system”, *Physical Review E* **49**, 4873 (1994). DOI: 10.1103/PhysRevE.49.4873.

[118] J.-H. Jeon, R. Metzler, “Fractional Brownian motion and motion governed by the fractional Langevin equation in confined geometries”, *Physical Review E* **81**, 021103 (2010). DOI: 10.1103/PhysRevE.81.021103.

[119] S. Hochreiter, J. Schmidhuber, “Long Short-Term Memory”, *Neural Computation* **9**, 1735 (1997). DOI: 10.1162/neco.1997.9.8.1735.

[120] K. S. Karunatilaka, E. A. Cameron, E. C. Martens, N. M. Koropatkin, J. S. Biteen, “Superresolution Imaging Captures Carbohydrate Utilization Dynamics in Human Gut Symbionts”, *mBio* **5** (2014). DOI: 10.1128/mBio.02172-14.

[121] E. C. Martens, H. C. Chiang, J. I. Gordon, “Mucosal Glycan Foraging Enhances Fitness and Transmission of a Saccharolytic Human Gut Bacterial Symbiont”, *Cell Host & Microbe* **4**, 447 (2008). DOI: 10.1016/j.chom.2008.09.007.

[122] H. H. Tuson, M. H. Foley, N. M. Koropatkin, J. S. Biteen, "The Starch Utilization System Assembles around Stationary Starch-Binding Proteins", *Biophysical Journal* **115**, 242 (2018). DOI: 10.1016/j.bpj.2017.12.015.

[123] A. Lepore, *et al.*, "Quantification of very low-abundant proteins in bacteria using the HaloTag and epi-fluorescence microscopy", *Scientific Reports* **9**, 7902 (2019). DOI: 10.1038/s41598-019-44278-0.

[124] R. W. Hamming, "Error Detecting and Error Correcting Codes", *Bell System Technical Journal* **29**, 147 (1950). DOI: 10.1002/j.1538-7305.1950.tb00463.x.

[125] N. M. Koropatkin, T. J. Smith, "SusG: A Unique Cell-Membrane-Associated α-Amylase from a Prominent Human Gut Symbiont Targets Complex Starch Molecules", *Structure* **18**, 200 (2010). DOI: 10.1016/j.str.2009.12.010.

[126] H. Shen, *et al.*, "Single Particle Tracking: From Theory to Biophysical Applications", *Chemical Reviews* **117**, 7331 (2017). DOI: 10.1021/acs.chemrev.6b00815.

[127] J. Elf, I. Barkefors, "Single-Molecule Kinetics in Living Cells", *Annual Review of Biochemistry* **88**, 635 (2019). DOI: 10.1146/annurev-biochem-013118-110801.

[128] A. Robson, K. Burrage, M. C. Leake, "Inferring diffusion in single live cells at the single-molecule level", *Philosophical Transactions of the Royal Society B: Biological Sciences* **368**, 20120029 (2013). DOI: 10.1098/rstb.2012.0029.

[129] S. Thapa, M. A. Lomholt, J. Krog, A. G. Cherstvy, R. Metzler, "Bayesian analysis of single-particle tracking data using the nested-sampling algorithm: maximum-likelihood model selection applied to stochastic-diffusivity data", *Physical Chemistry Chemical Physics* **20**, 29018 (2018). DOI: 10.1039/C8CP04043E.

[130] A. G. Cherstvy, S. Thapa, C. E. Wagner, R. Metzler, "Non-Gaussian, non-ergodic, and non-Fickian diffusion of tracers in mucin hydrogels", *Soft Matter* **15**, 2526 (2019). DOI: 10.1039/C8SM02096E.

[131] W. Arber, "Host-Controlled Modification of Bacteriophage", *Annual Review of Microbiology* **19**, 365 (1965). DOI: 10.1146/annurev.mi.19.100165.002053.

[132] W. Arber, D. Dussoix, "Host specificity of DNA produced by Escherichia coli", *Journal of Molecular Biology* **5**, 18 (1962). DOI: 10.1016/S0022-2836(62)80058-8.

[133] T. A. Bickle, D. H. Krüger, "Biology of DNA restriction", *Microbiological Reviews* **57**, 434 (1993). DOI: 10.1128/mr.57.2.434-450.1993.

[134] M. J. Blow, *et al.*, "The Epigenomic Landscape of Prokaryotes", *PLOS Genetics* **12**, e1005854 (2016). DOI: 10.1371/journal.pgen.1005854.

[135] D. Wion, J. Casadesús, "N6-methyl-adenine: an epigenetic signal for DNA–protein interactions", *Nature Reviews Microbiology* **4**, 183 (2006). DOI: 10.1038/nrmicro1350.

[136] X. Nou, *et al.*, “Regulation of pyelonephritis-associated pili phase-variation in Escherichia coli: binding of the Papl and the Lrp regulatory proteins is controlled by DNA methylation”, *Molecular Microbiology* **7**, 545 (1993). DOI: 10.1111/j.1365-2958.1993.tb01145.x.

[137] A. Anjum, *et al.*, “Phase variation of a Type IIG restriction-modification enzyme alters site-specific methylation patterns and gene expression in *Campylobacter jejuni* strain NCTC11168”, *Nucleic Acids Research* **44**, 4581 (2016). DOI: 10.1093/nar/gkw019.

[138] T. M. Nye, *et al.*, “DNA methylation from a Type I restriction modification system influences gene expression and virulence in Streptococcus pyogenes”, *PLOS Pathogens* **15**, e1007841 (2019). DOI: 10.1371/journal.ppat.1007841.

[139] S. Bheemanaik, Y. Reddy, D. Rao, “Structure, function and mechanism of exocyclic DNA methyltransferases”, *Biochemical Journal* **399**, 177 (2006). DOI: 10.1042/BJ20060854.

[140] C. B. Woodcock, A. B. Yakubov, N. O. Reich, “*Caulobacter crescentus* Cell Cycle-Regulated DNA Methyltransferase Uses a Novel Mechanism for Substrate Recognition”, *Biochemistry* **56**, 3913 (2017). DOI: 10.1021/acs.biochem.7b00378.

[141] K. S. Lang, *et al.*, “Replication-Transcription Conflicts Generate R-Loops that Orchestrate Bacterial Stress Survival and Pathogenesis”, *Cell* **170**, 787 (2017). DOI: 10.1016/j.cell.2017.07.044.

[142] J. S. Lenhart, *et al.*, “RecO and RecR Are Necessary for RecA Loading in Response to DNA Damage and Replication Fork Stress”, *Journal of Bacteriology* **196**, 2851 (2014). DOI: 10.1128/JB.01494-14.

[143] S. J. Callahan, *et al.*, “Structure of Type IIL Restriction-Modification Enzyme MmeI in Complex with DNA Has Implications for Engineering New Specificities”, *PLOS Biology* **14**, e1002442 (2016). DOI: 10.1371/journal.pbio.1002442.

[144] R. D. Morgan, E. A. Dwinell, T. K. Bhatia, E. M. Lang, Y. A. Luyten, “The MmeI family: type II restriction–modification enzymes that employ single-strand modification for host protection”, *Nucleic Acids Research* **37**, 5208 (2009). DOI: 10.1093/nar/gkp534.

[145] R. F. Albu, M. Zacharias, T. P. Jurkowski, A. Jeltsch, “DNA Interaction of the CcrM DNA Methyltransferase: A Mutational and Modeling Study”, *ChemBioChem* **13**, 1304 (2012). DOI: 10.1002/cbic.201200082.

[146] L. M. Iyer, E. V. Koonin, L. Aravind, “Extensive domain shuffling in transcription regulators of DNA viruses and implications for the origin of fungal APSES transcription factors”, *Genome Biology* **3**, research0012.1 (2002). DOI: 10.1186/gb-2002-3-3-research0012.

[147] K. S. Makarova, Y. I. Wolf, S. Snir, E. V. Koonin, "Defense Islands in Bacterial and Archaeal Genomes and Prediction of Novel Defense Systems", *Journal of Bacteriology* **193**, 6039 (2011). DOI: 10.1128/JB.05535-11.

[148] M. Juhas, *et al.*, "Genomic islands: tools of bacterial horizontal gene transfer and evolution", *FEMS Microbiology Reviews* **33**, 376 (2009). DOI: 10.1111/j.1574-6976.2008.00136.x.

[149] I. S. Rusinov, A. S. Ershova, A. S. Karyagina, S. A. Spirin, A. V. Alexeevski, "Avoidance of recognition sites of restriction-modification systems is a widespread but not universal anti-restriction strategy of prokaryotic viruses", *BMC Genomics* **19**, 885 (2018). DOI: 10.1186/s12864-018-5324-3.

[150] H. Ohshima, S. Matsuoka, K. Asai, Y. Sadaie, "Molecular Organization of Intrinsic Restriction and Modification Genes *Bsu* M of *Bacillus subtilis* Marburg", *Journal of Bacteriology* **184**, 381 (2002). DOI: 10.1128/JB.184.2.381-389.2002.

[151] A. Negri, *et al.*, "Regulator-dependent temporal dynamics of a restriction-modification system's gene expression upon entering new host cells: single-cell and population studies", *Nucleic Acids Research* **49**, 3826 (2021). DOI: 10.1093/nar/gkab183.

[152] M. Stracy, *et al.*, "Transient non-specific DNA binding dominates the target search of bacterial DNA-binding proteins", *Molecular Cell* **81**, 1499 (2021). DOI: 10.1016/j.molcel.2021.01.039.

[153] X. Zhou, J. Wang, J. Herrmann, W. E. Moerner, L. Shapiro, "Asymmetric division yields progeny cells with distinct modes of regulating cell cycle-dependent chromosome methylation", *Proceedings of the National Academy of Sciences* **116**, 15661 (2019). DOI: 10.1073/pnas.1906119116.

[154] J. Li, *et al.*, "Epigenetic Switch Driven by DNA Inversions Dictates Phase Variation in Streptococcus pneumoniae", *PLOS Pathogens* **12**, e1005762 (2016). DOI: 10.1371/journal.ppat.1005762.

[155] K. Kobayashi, "Diverse LXG toxin and antitoxin systems specifically mediate intraspecies competition in Bacillus subtilis biofilms", *PLOS Genetics* **17**, e1009682 (2021). DOI: 10.1371/journal.pgen.1009682.

[156] S. Kaundal, A. Deep, G. Kaur, K. G. Thakur, "Molecular and Biochemical Characterization of YeeF/YezG, a Polymorphic Toxin-Immunity Protein Pair From Bacillus subtilis", *Frontiers in Microbiology* **11**, 95 (2020). DOI: 10.3389/fmicb.2020.00095.

[157] D. Piel, *et al.*, "Genetic determinism of phage-bacteria coevolution in natural populations", *preprint*, Microbiology (2021).

[158] N. L. Fernandez, *et al.*, "DNA Methylation and RNA-DNA Hybrids Regulate the Single-Molecule Localization of a DNA Methyltransferase on the Bacterial Nucleoid", *mBio* pp. e03185–22 (2023). DOI: 10.1128/mbio.03185-22.

[159] J. Schindelin, *et al.*, "Fiji: an open-source platform for biological-image analysis", *Nature Methods* **9**, 676 (2012). DOI: 10.1038/nmeth.2019.

[160] C. Ritz, F. Baty, J. C. Streibig, D. Gerhard, "Dose-Response Analysis Using R", *PLOS ONE* **10**, e0146021 (2015). DOI: 10.1371/journal.pone.0146021.

[161] C. D. Allis, T. Jenuwein, "The molecular hallmarks of epigenetic control", *Nature Reviews Genetics* **17**, 487 (2016). DOI: 10.1038/nrg.2016.59.

[162] R. Bonasio, S. Tu, D. Reinberg, "Molecular Signals of Epigenetic States", *Science* **330**, 612 (2010). DOI: 10.1126/science.1191078.

[163] T. Jenuwein, C. D. Allis, "Translating the Histone Code", *Science* **293**, 1074 (2001). DOI: 10.1126/science.1063127.

[164] B. D. Strahl, C. D. Allis, "The language of covalent histone modifications", *Nature* **403**, 41 (2000). DOI: 10.1038/47412.

[165] K. R. Stewart-Morgan, N. Petryk, A. Groth, "Chromatin replication and epigenetic cell memory", *Nature Cell Biology* **22**, 361 (2020). DOI: 10.1038/s41556-020-0487-y.

[166] N. P. Cowieson, J. F. Partridge, R. C. Allshire, P. J. McLaughlin, "Dimerisation of a chromo shadow domain and distinctions from the chromodomain as revealed by structural analysis", *Current Biology* **10**, 517 (2000). DOI: 10.1016/S0960-9822(00)00467-X.

[167] R. M. Hughes, K. R. Wiggins, S. Khorasanizadeh, M. L. Waters, "Recognition of trimethyllysine by a chromodomain is not driven by the hydrophobic effect", *Proceedings of the National Academy of Sciences* **104**, 11184 (2007). DOI: 10.1073/pnas.0610850104.

[168] S. A. Jacobs, S. Khorasanizadeh, "Structure of HP1 Chromodomain Bound to a Lysine 9-Methylated Histone H3 Tail", *Science* **295**, 2080 (2002). DOI: 10.1126/science.1069473.

[169] A. G. Larson, *et al.*, "Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin", *Nature* **547**, 236 (2017). DOI: 10.1038/nature22822.

[170] A. R. Strom, *et al.*, "Phase separation drives heterochromatin domain formation", *Nature* **547**, 241 (2017). DOI: 10.1038/nature22989.

[171] N. Iglesias, *et al.*, "Native Chromatin Proteomics Reveals a Role for Specific Nucleoporins in Heterochromatin Organization and Maintenance", *Molecular Cell* **77**, 51 (2020). DOI: 10.1016/j.molcel.2019.10.018.

[172] G. LeRoy, *et al.*, "Heterochromatin Protein 1 Is Extensively Decorated with Histone Code-like Post-translational Modifications", *Molecular & Cellular Proteomics* **8**, 2432 (2009). DOI: 10.1074/mcp.M900160-MCP200.

[173] A. V. Ivanova, M. J. Bonaduce, S. V. Ivanov, A. J. S. Klar, "The chromo and SET domains of the Clr4 protein are essential for silencing in fission yeast", *Nature Genetics* **19**, 192 (1998). DOI: 10.1038/566.

[174] K. Ekwall, *et al.*, "The Chromodomain Protein Swi6: A Key Component at Fission Yeast Centromeres", *Science* **269**, 1429 (1995). DOI: 10.1126/science.7660126.

[175] S. Haldar, A. Saini, J. S. Nanda, S. Saini, J. Singh, "Role of Swi6/HP1 Self-association-mediated Recruitment of Clr4/Suv39 in Establishment and Maintenance of Heterochromatin in Fission Yeast", *Journal of Biological Chemistry* **286**, 9308 (2011). DOI: 10.1074/jbc.M110.143198.

[176] G. Jih, *et al.*, "Unique roles for histone H3K9me states in RNAi and heritable silencing of transcription", *Nature* **547**, 463 (2017). DOI: 10.1038/nature23267.

[177] C. Keller, *et al.*, "HP1Swi6 Mediates the Recognition and Destruction of Heterochromatic RNA Transcripts", *Molecular Cell* **47**, 215 (2012). DOI: 10.1016/j.molcel.2012.05.009.

[178] R. Stunnenberg, *et al.*, "H3K9 methylation extends across natural boundaries of heterochromatin in the absence of an <span style="font-variant:small-caps;">HP</span> 1 protein", *The EMBO Journal* **34**, 2789 (2015). DOI: 10.15252/embj.201591320.

[179] J. Wang, B. D. Reddy, S. Jia, "Rapid epigenetic adaptation to uncontrolled heterochromatin spreading", *eLife* **4**, e06179 (2015). DOI: 10.7554/eLife.06179.

[180] J.-i. Nakayama, A. J. Klar, S. I. Grewal, "A Chromodomain Protein, Swi6, Performs Imprinting Functions in Fission Yeast during Mitosis and Meiosis", *Cell* **101**, 307 (2000). DOI: 10.1016/S0092-8674(00)80840-5.

[181] T. Cheutin, S. A. Gorski, K. M. May, P. B. Singh, T. Misteli, "In Vivo Dynamics of Swi6 in Yeast: Evidence for a Stochastic Model of Heterochromatin", *Molecular and Cellular Biology* **24**, 3157 (2004). DOI: 10.1128/MCB.24.8.3157-3167.2004.

[182] T. Cheutin, *et al.*, "Maintenance of Stable Heterochromatin Domains by Dynamic HP1 Binding", *Science* **299**, 721 (2003). DOI: 10.1126/science.1078572.

[183] K. Hiragami-Hamada, *et al.*, "Dynamic and flexible H3K9me3 bridging via HP1β dimerization establishes a plastic state of condensed chromatin", *Nature Communications* **7**, 11310 (2016). DOI: 10.1038/ncomms11310.

[184] S. Kilic, A. L. Bachmann, L. C. Bryan, B. Fierz, "Multivalency governs HP1α association dynamics with the silent chromatin state", *Nature Communications* **6**, 7313 (2015). DOI: 10.1038/ncomms8313.

[185] G. Nishibuchi, *et al.*, "N-terminal phosphorylation of HP1α increases its nucleosome-binding specificity", *Nucleic Acids Research* **42**, 12498 (2014). DOI: 10.1093/nar/gku995.

[186] B. D. Ripley, “The second-order analysis of stationary point processes”, *Journal of Applied Probability* **13**, 255 (1976). DOI: 10.2307/3212829.

[187] D. Canzio, *et al.*, “A conformational switch in HP1 releases auto-inhibition to drive heterochromatin assembly”, *Nature* **496**, 377 (2013). DOI: 10.1038/nature12032.

[188] B. D. Reddy, *et al.*, “Elimination of a specific histone H3K14 acetyltransferase complex bypasses the RNAi pathway to regulate pericentric heterochromatin functions”, *Genes & Development* **25**, 214 (2011). DOI: 10.1101/gad.1993611.

[189] S. N. Wood, “Statistical inference for noisy nonlinear ecological dynamic systems”, *Nature* **466**, 1102 (2010). DOI: 10.1038/nature09319.

[190] Y. Maru, D. E. Afar, O. N. Witte, M. Shibuya, “The Dimerization Property of Glutathione -Transferase Partially Reactivates Bcr-Abl Lacking the Oligomerization Domain”, *Journal of Biological Chemistry* **271**, 15353 (1996). DOI: 10.1074/jbc.271.26.15353.

[191] M. M. Keenen, *et al.*, “HP1 proteins compact DNA into mechanically and positionally stable phase separated domains”, *eLife* **10**, e64563 (2021). DOI: 10.7554/eLife.64563.

[192] A.-M. Abdalla, C. M. Bruns, J. A. Tainer, B. Mannervik, G. Stenberg, “Design of a monomeric human glutathione transferase GSTP1, a structurally stable but catalytically inactive protein”, *Protein Engineering, Design and Selection* **15**, 827 (2002). DOI: 10.1093/protein/15.10.827.

[193] A. G. Larson, G. J. Narlikar, “The Role of Phase Separation in Heterochromatin Formation, Function, and Regulation”, *Biochemistry* **57**, 2540 (2018). DOI: 10.1021/acs.biochem.8b00401.

[194] M. Sadaie, *et al.*, “Balance between Distinct HP1 Family Proteins Controls Heterochromatin Assembly in Fission Yeast”, *Molecular and Cellular Biology* **28**, 6973 (2008). DOI: 10.1128/MCB.00791-08.

[195] J. Munkres, “Algorithms for the Assignment and Transportation Problems”, *Journal of the Society for Industrial and Applied Mathematics* **5**, 32 (1957). DOI: 10.1137/0105003.

[196] M. A. Kiskowski, J. F. Hancock, A. K. Kenworthy, “On the Use of Ripley’s K-Function and Its Derivatives to Analyze Domain Size”, *Biophysical Journal* **97**, 1095 (2009). DOI: 10.1016/j.bpj.2009.05.039.

[197] F. Goreaud, R. Pélissier, “On explicit formulas of edge effect correction for Ripley’s K -function”, *Journal of Vegetation Science* **10**, 433 (1999). DOI: 10.2307/3237072.

[198] S. A. Jacobs, *et al.*, “Specificity of the HP1 chromo domain for the methylated N-terminus of histone H3”, *The EMBO Journal* **20**, 5232 (2001). DOI: 10.1093/emboj/20.18.5232.

[199] S. Sanulli, J. D. Gross, G. J. Narlikar, "Biophysical Properties of HP1-Mediated Heterochromatin", *Cold Spring Harbor Symposia on Quantitative Biology* **84**, 217 (2019). DOI: 10.1101/sqb.2019.84.040360.

[200] S. Rea, *et al.*, "Regulation of chromatin structure by site-specific histone H3 methyltransferases", *Nature* **406**, 593 (2000). DOI: 10.1038/35020506.

[201] A. J. Bannister, *et al.*, "Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain", *Nature* **410**, 120 (2001). DOI: 10.1038/35065138.

[202] D. Halverson, G. Gutkin, L. Clarke, "A novel member of the Swi6p family of fission yeast chromo domain-containing proteins associates with the centromere in vivo and affects chromosome segregation", *Molecular Genetics and Genomics* **264**, 492 (2000). DOI: 10.1007/s004380000338.

[203] G. Thon, J. Verhein-Hansen, "Four Chromo-domain Proteins of Schizosaccharomyces pombe Differentially Repress Transcription at Various Chromosomal Locations", *Genetics* **155**, 551 (2000). DOI: 10.1093/genetics/155.2.551.

[204] K. Leopold, A. Stirpe, T. Schalch, "Transcriptional gene silencing requires dedicated interaction between HP1 protein Chp2 and chromatin remodeler Mit1", *Genes & Development* **33**, 565 (2019). DOI: 10.1101/gad.320440.118.

[205] D. W. Piston, G.-J. Kremers, "Fluorescent protein FRET: the good, the bad and the ugly", *Trends in Biochemical Sciences* **32**, 407 (2007). DOI: 10.1016/j.tibs.2007.08.003.

[206] A. Gahlmann, W. E. Moerner, "Exploring bacterial cell biology with single-molecule tracking and super-resolution imaging", *Nature Reviews Microbiology* **12**, 9 (2014). DOI: 10.1038/nrmicro3154.

[207] A. Kusumi, T. A. Tsunoyama, K. M. Hirosawa, R. S. Kasai, T. K. Fujiwara, "Tracking single molecules at work in living cells", *Nature Chemical Biology* **10**, 524 (2014). DOI: 10.1038/nchembio.1558.

[208] S. Basu, *et al.*, "Live-cell 3D single-molecule tracking reveals how NuRD modulates enhancer dynamics", *preprint* (2020).

[209] A. Ranjan, *et al.*, "Live-cell single particle imaging reveals the role of RNA polymerase II in histone H2A.Z eviction", *eLife* **9**, e55667 (2020). DOI: 10.7554/eLife.55667.

[210] Z. Chen, L. Geffroy, J. S. Biteen, "NOBIAS: Analyzing Anomalous Diffusion in Single-Molecule Tracks With Nonparametric Bayesian Inference", *Frontiers in Bioinformatics* **1**, 742073 (2021). DOI: 10.3389/fbinf.2021.742073.

[211] J. C. M. Gebhardt, *et al.*, "Single-molecule imaging of transcription factor binding to DNA in live mammalian cells", *Nature Methods* **10**, 421 (2013). DOI: 10.1038/nmeth.2411.

[212] G. Raiymbek, *et al.*, “An H3K9 methylation-dependent protein interaction regulates the non-enzymatic functions of a putative histone demethylase”, *eLife* **9**, e53155 (2020). DOI: 10.7554/eLife.53155.

[213] K. M. Creamer, *et al.*, “The Mi-2 Homolog Mit1 Actively Positions Nucleosomes within Heterochromatin To Suppress Transcription”, *Molecular and Cellular Biology* **34**, 2046 (2014). DOI: 10.1128/MCB.01609-13.

[214] S. D. Taverna, H. Li, A. J. Ruthenburg, C. D. Allis, D. J. Patel, “How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers”, *Nature Structural & Molecular Biology* **14**, 1025 (2007). DOI: 10.1038/nsmb1338.

[215] E. K. McLean, T. M. Nye, F. C. Lowder, L. A. Simmons, “The Impact of RNA-DNA Hybrids on Genome Integrity in Bacteria”, *Annual Review of Microbiology* **76**, 461 (2022). DOI: 10.1146/annurev-micro-102521-014450.

[216] Y. Yamagishi, T. Sakuno, M. Shimura, Y. Watanabe, “Heterochromatin links to centromeric protection by recruiting shugoshin”, *Nature* **455**, 251 (2008). DOI: 10.1038/nature07217.