

SINGLE CELL AND SINGLE MOLECULE
TECHNIQUES FOR THE ANALYSIS OF THE
EPIGENOME

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Christopher Benjamin Wallin

February 2016

© 2016 Christopher Benjamin Wallin

ALL RIGHTS RESERVED

SINGLE CELL AND SINGLE MOLECULE TECHNIQUES FOR THE
ANALYSIS OF THE EPIGENOME
Christopher Benjamin Wallin, Ph.D.
Cornell University 2016

Epigenetic regulation is a critical biological process for the health and development of a cell. Epigenetic regulation is facilitated by covalent modifications to the underlying DNA and chromatin proteins. A fundamental understanding of these epigenetic modifications and their associated interactions at the molecular scale is necessary to explain phenomena including cellular identity, stem cell plasticity, and neoplastic transformation. It is widely known that abnormal epigenetic profiles have been linked to many diseases, most notably cancer. While the field of epigenetics has progressed rapidly with conventional techniques, significant advances remain to be made with respect to combinatoric analysis of epigenetic marks and single cell epigenetics. Therefore, in this dissertation, I will discuss our development of devices and methodologies to address these pertinent issues.

First, we designed a preparatory polydimethylsiloxane (PDMS) microdevice for the extraction, purification, and stretching of human chromosomal DNA and chromatin from small cell populations down to a single cell. The valveless device captures cells by size exclusion within the micropillars, entraps the DNA or chromatin in the micropillars after cell lysis, purifies away the cellular debris, and fluorescently labels the DNA and/or chromatin all within a single reaction chamber. With the device, we achieve nearly 100% extraction efficiency of the DNA. The device is also used for in-channel immunostaining of chromatin

followed by downstream single molecule chromatin analysis in nanochannels (SCAN).

Second, using multi-color, time-correlated single molecule measurements in nanochannels, simultaneous coincidence detection of 2 epigenetic marks is demonstrated. Coincidence detection of 3 epigenetic marks is also established using a pulsed interleaved excitation scheme. With these two promising results, genome-wide quantification of epigenetic marks was pursued. Unfortunately, quantitative SCAN never materialized. Reasons for this, including poor signal to background, are explained in detail.

Third, development of mobility-SCAN, an analytical technique for measuring and analyzing single molecules based on their fluorescent signature and their electrophoretic mobility in nanochannels is described. We use the technique to differentiate biomolecules from complex mixtures and derive parameters such as diffusion coefficients and effective charges. Finally, the device is used to detect binding interactions of various complexes similar to affinity capillary electrophoresis, but on a single molecule level.

Fourth, we conclude by briefly discussing SCAN-sort, a technique to sort individual chromatin molecules based on their fluorescent emissions for further downstream analysis such as DNA sequencing. We demonstrate a 2-fold enrichment of chromatin from sorting and discuss possible system modifications for better performance in the future.

BIOGRAPHICAL SKETCH

Christopher was born and raised in Fergus Falls, MN. He attended Fergus Falls Public Schools and graduated in 2004. After graduation, he enrolled at North Dakota State University (NDSU) in Fargo, ND. In 2009, he graduated summa cum laude with a Bachelor of Science in Physics, Mathematics, and Economics along with a Bachelor of Science in Electrical Engineering. Christopher enrolled in the Applied Physics graduate program at Cornell University in 2010 where he studied under Professor Harold Craighead. After graduate school, Christopher plans to do postdoctoral research at the National Institute of Standards and Technology (NIST) in Gaithersburg, MD.

To my family.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Professor Harold Craighead, for giving me the opportunity to work in his lab over the course of my graduate studies. The scientific discussions and helpful suggestions that I have had throughout the years with Harold have made this dissertation possible. Through working in his lab, I have learned to be a creative, critical, and independent thinker which will be undoubtedly helpful in my continued professional development. For that, I thank you.

I would also like to thank my committee members, Professor Paul Soloway and Professor Brian Kirby, for their guidance and support throughout my time at Cornell.

I have had the privilege to work alongside many members of the Craighead lab: Juraj Topolancik, Jaime Benitez, David Latulippe, Christopher Kelly, Aline Cerf, Thomas Alava, Vivek Adiga, Kylan Szeto, Rob Barton, Ben Cipriany, Harvey Tian, Sarah Reinholt, Macey Ruble, Jia-Wei Yeh, Rahul Pathak, Chris Martin, Daniel Joe, Paul Zhu, and Aniket Kakatkar. I have had great experiences inside and outside the lab with these people and many have become good friends. I would like to especially thank Juraj Topolancik and Jaime Benitez for their support and mentorship when I first joined the lab. I also want to specifically thank Ben Cipriany and Kylan Szeto for their initial instruction and discussions regarding SCAN.

With the collaboration with the Soloway lab, I was fortunate to work alongside many talented people: Patrick Murphy, Jim Hagarman, Byung-Ryool (Yul) Hyun, John McElwee, and Jim Putnam. Their expertise in molecular biology, genetics, and epigenetics was sincerely appreciated and vital to much of this work. Further, the Soloway lab was always inviting to their activities outside of

lab. I will always remember the frisbee golf, bowling, golf, and other outings which we had throughout the years.

I would like to acknowledge the facilities and funding sources that have supported my graduate work and the completion of this dissertation. The Cornell NanoScale Facility (CNF) and the Nanobiotechnology Center (NBTC) which are funded by the National Science Foundation (NSF) have been invaluable user facilities where all of the devices were fabricated. Additionally, funding for this graduate work was generously supported by the National Cancer Institute (NCI) and the National Institutes of Health (NIH).

I would like to thank all the staff of the CNF facility. I would especially like to thank Rob Ilic, who has been a great friend and mentor throughout these years. I'm greatly looking forward to working alongside him at NIST.

And finally, to my family and friends, thank you for your encouragement since the beginning. None of this would have been possible without your love and support.

TABLE OF CONTENTS

Biographical Sketch	iii
Dedication	iv
Acknowledgements	v
Table of Contents	vii
List of Tables	x
List of Figures	xi
1 Introduction	1
1.1 Organization of this Dissertation	5
2 Microfluidic Extraction and Analysis of Human Chromosomal DNA and Chromatin from Single Cells	10
2.1 Introduction	10
2.2 Materials and Methods	13
2.2.1 Device Design and Fabrication	13
2.2.2 Microscopy and Experimental Setup	15
2.2.3 Cell Culture	17
2.2.4 Cell Capture and Lysis	17
2.2.5 DNA Purification and Visualization	18
2.2.6 Chromatin Purification, Visualization, and Antibody Labeling	18
2.2.7 DNA Release and Off-Chip Fluorescence Analysis	19
2.2.8 Single Cell Real-Time PCR Analysis	20
2.3 Results and Discussion	21
2.3.1 Off-Chip Quantification	23
2.3.2 Microfluidic Device to Nanochannel Coupling for Single Cell Analysis and SCAN	27
2.3.3 Microfluidic Chromatin Isolation From Small Cell Populations	30
2.3.4 Antibody Labeling of Chromatin in the Micropillar Device with SCAN Coupling	34
2.4 Conclusions	36
3 Single Molecule Epigenetic Mark Detection and Quantification in Nanofluidic Channels	42
3.1 Introduction	42
3.1.1 SCAN Overview	43
3.2 Materials and Methods	44
3.2.1 Device Fabrication	44
3.2.2 Optical Setup	45
3.2.3 Chromatin Extraction	47
3.2.4 Sucrose Gradient Purification of Chromatin	48

3.2.5	Methyl Binding Domain Synthesis and Labeling	49
3.2.6	Antibody labeling	49
3.2.7	Binding Reactions	49
3.2.8	Size Exclusion Chromatography	50
3.2.9	Nanochannel SCAN Preparation	51
3.3	Results and Discussion	51
3.3.1	2-Color Epigenetic Mark Detection	51
3.3.2	3-Color Epigenetic Mark Detection	57
3.3.3	Quantification of Epigenetic Marks	71
3.3.4	Theoretical Considerations for Quantitative SCAN	73
3.3.5	Signal Processing using Filters	82
3.3.6	Chromatin Sample Preparation Adjustments for Quantitative SCAN	87
3.3.7	Low Coincidence Rates of SCAN	90
3.4	Conclusions	97
4	Single Molecule Mobility Studies in Nanochannels	103
4.1	Introduction	103
4.2	Theoretical Considerations	107
4.2.1	Electrokinetic Phenomena in Nanochannels	108
4.2.2	Migration Time, Theoretical Plates, and Peak Broadening .	114
4.3	Materials and Methods	123
4.3.1	Device Fabrication	123
4.3.2	Device Layout	125
4.3.3	Nanochannel Voltage and the Equivalent Circuit	131
4.3.4	Optical System for Mobility Measurements	133
4.3.5	Sample Preparations	137
4.3.6	Measurement of Electroosmosis	141
4.4	Results and Discussion	143
4.4.1	Single Molecule Mobility Detection	143
4.4.2	Voltage Dependence on Migration Time	147
4.4.3	Length Dependence on Migration Time	149
4.4.4	Theoretical Plates of Alexa Fluor 647 and DNA by Varying Voltage	150
4.4.5	Theoretical Plates of Alexa Fluor 647 and DNA by Varying Detection Length	153
4.4.6	DNA Mobility in Nanochannels	155
4.4.7	Single Dye Mobility Resolution	160
4.4.8	Chromatin Mobility Measurements	165
4.4.9	Mobility Based Identification of 6 Biomolecules	172
4.4.10	Affinity Electrophoresis in Nanochannels	176
4.4.11	Affinity Electrophoresis of DNA-MBD1 Interactions	177
4.4.12	Affinity Electrophoresis of Chromatin- α -H3 Antibody Interactions	180

4.4.13	Affinity Electrophoresis of BSA-Primary Ab-Secondary Ab Interactions	183
4.5	Conclusions	187
5	Single Molecule Sorting of DNA and Chromatin in Nanofluidic Chan- nels	193
5.1	Introduction	193
5.2	Materials and Methods	195
5.2.1	Device Fabrication and Layout	195
5.2.2	Optical and Electrical Setup	195
5.2.3	Sample Preparation	198
5.3	Results and Discussion	198
5.3.1	Intensity Based Sorting of DNA	198
5.3.2	Sorting of Chromatin	201
5.3.3	False Positives with Sorting	203
5.3.4	Channel Backflow and Leakage	204
5.4	Conclusions	207
6	Conclusions and Future Prospectives	210

LIST OF TABLES

3.1	Antagonism of MBD1 and H3K27me3 for MF and MF60.1 cell lines	71
4.1	U-shaped nanochannel dimensions	127
4.2	Input/output channels and reservoir dimensions	127
4.3	Straight nanochannel dimensions	129
4.4	Filter dimensions for three filters. The filter width gives the total width available for fluid occupation across the entire reservoir. . .	130
4.5	DNA/Chromatin Peak Fitting Parameters with Gaussian fits. . .	168
4.6	DNA/chromatin peak fitting parameters with 2D Gaussian fits. .	169
4.7	DNA subpeaks fitting parameters with 2D Gaussian fits.	171
4.8	Biomolecules used for mobility based species identification using complex mixture experiment and their relevant physical parameters.	173
4.9	Probability of detection. The table should be read as the probability of detection for a molecule in the left-hand column with regards to a binary decision involving a molecule in the top row.	176
4.10	Summary of DNA/MBD fitting parameters with Gaussian fits .	178
4.11	Summary of Chromatin/ α -H3 fitting parameters with Gaussian fits	181

LIST OF FIGURES

2.1	Schematic and micrograph of the microfluidic device. The channel consists of a random array of micropillars for cell capture and DNA extraction. The non-random pillars are for structural support. The scale bar is 100 μm . Figure reproduced with permission from [1].	14
2.2	PDMS microchannel fabrication process. 1. Spin Microdeposit S1813 photoresist onto a SOI wafer. 2. Pattern resist using contact lithography. 3. Etch the silicon 20 μm to the oxide layer. 4. Deposit a monolayer of FOTS. 5. Pour PDMS onto the silicon mold. 6. Remove cured PDMS. 7. Bond PDMS to a fused silica wafer.	16
2.3	C_p standard curve derived from known amounts of DNA and a no-template control. Figure reproduced with permission from [1].	21
2.4	Schematic representation of the device operation. (A) A cell is captured under flow via size exclusion in the micropillars. (B) Cell lysis is performed with a ionic surfactant (1% SDS). Cell debris, nucleosomes, and other proteins are washed away leaving entangled chromosomal DNA. (C) DNA is labeled with intercalating dye. (D) The DNA is released using restriction endonucleases. Figure reproduced with permission from [1].	22
2.5	(A) A single M0-91 cell trapped in the random pillar array. (B) Detailed fluorescent image of the DNA strands held by the micropillars under flow. (C) Stretched DNA from the single cell shown in (A). The extension of the DNA was more than 10 mm downstream. Figure reproduced with permission from [1].	24
2.6	(A) Bright field image of 70 cells captured with the microdevice. (B) Fluorescent image of the entangled DNA following lysis. (C) Time sequence showing the digestion of the DNA. DNA was fully released after 2 minutes of digestion. (D) Amount of DNA extracted from different number of cells. The linear best fit of the data shown by the solid line gives a slope of $6.7 \pm 0.2 \text{ pg}/\text{cell}$. Figure reproduced with permission from [1].	26
2.7	Real-time PCR amplification curves. Figure reproduced with permission from [1].	27
2.8	Initial device design for microfluidic/nanochannel coupling. . .	28
2.9	Single cell micropillar device coupled to SCAN. (a) Single cell trapped within the micropillar array. (b) SCAN time trace from the single cell sample.	30

2.10	Schematic representation of chromatin purification using the microfluidic device. (A) A single cell is captured in the micropillars by size exclusion. (B) The cell is lysed using a nonionic surfactant (Triton X-100). The chromatin unwraps around the pillars and the cell debris is washed away. (C) The chromatin is intercalated and immunostained with antibodies targeting histone modifications. (D) The chromatin is released with micrococcal nuclease.	31
2.11	(a) Bright field micrograph showing 18 HeLa-GFP cells captured in the micropillar array. (b) Fluorescence image of 18 HeLa-GFP cells from (a). The GFP is localized within the nucleus prior to lysis. (c) Elongated chromatin after lysis under flow. (d) Coincidence offset histogram for GFP and TOTO-3. The coincidence peak signified that the chromatin from the microdevice remained intact for downstream analysis such as with SCAN.	33
2.12	Morphological differences between elongated DNA and chromatin from M0-91 cells. Both were stained with PicoGreen. (a) DNA (b) Chromatin	34
2.13	(a) Fluorescent micrograph showing H2B-GFP. (b) Fluorescence image of the Alexa Fluor 647 α -H3. (c) Image overlay with GFP (blue), Alexa Fluor 647 α -H3 (red), and their colocalization (purple). (d) Time trace of H2B-GFP and Alexa Fluor 647 α -H3 showing coincident molecules with SCAN.	35
3.1	(A) Two color optical setup. (B) Three color optical setup. A confocal aperture (not shown) was placed between the quad-band bandpass filter and the 488 nm laser line emission filter. Dotted lines represent pulsed signals for pulsed interleaved excitation. A confocal aperture (not shown) was placed between the quad-band bandpass filter and the 405 nm laser line emission filter.	46
3.2	Epigenetic mark specificity using SCAN. (A) 2-channel time trace showing fluorescent emission from GFP and α -H3 molecules. Coincident events signifying antibody-chromatin complexes are highlighted with the blue shading. (B) Coincidence delay histogram illustrating simultaneous detection of GFP and α -H3 antibody. (C+D) YOYO-1 labeled wild type ES cell chromatin displaying (C) coincidence with α -H3 and (D) no coincidence with pre-immune mouse serum serving as a negative control. (E+F) α -H3 (E) coincidence with YOYO-1 labeled wild type ES cell chromatin and (F) reduced coincidence with H3K27me3 depleted Eed-/- ES cell chromatin labeled with YOYO-1. (G+H) MBD1, specific for mC, exhibiting (G) coincidence with YOYO-1 labeled wild type ES cell chromatin and (H) no coincidence with mC deficient DNMT TKO ES cell chromatin labeled with YOYO-1. Reproduced with permission from [14].	54

3.3	Epigenetic coincident mark detection using SCAN. (A) Chromatin bound α -H3 and α -H2B coincidence. (B) α -H3 and α -H2B coincidence for a negative control with no chromatin. (C) Histone modification coincidence detection using α -H3K27me3 and α -H2B. (D) Histone modification and methylation coincidence detection using α -H3K9me3 and MBD1. Reproduced with permission from [14].	56
3.4	Excitation and emission spectra for Alexa Fluor 405 (blue) and YOYO-1 intercalating dye (green). The boxes with slanted lines represent emission filter passbands (452/45 and 525/40). Lines refer to 405 nm and 488 nm laser lines. Crosstalk from Alexa Fluor 405 into the 525/40 passband can be observed. Note that spectra are normalized to their respective peaks and don't represent absolute quantities for direct comparison between dyes. Data from Thermo Fisher Fluorescence SpectraViewer.	62
3.5	Raw time trace demonstrating pulsing. The upper trace shows the 452/45 channel with Alexa Fluor 405 being excited. The lower trace shows the 525/40 channel with YOYO-1 being excited.	63
3.6	Corrected signals after downsampling and interpolating. The upper trace shows the 452/45 channel with Alexa Fluor 405. The lower trace shows the 525/40 channel with YOYO-1. The minor signal distortions are caused by the low pass interpolation filter.	64
3.7	Size exclusion chromatogram for YOYO-1 labeled chromatin (green), α -H3 (blue/Alexa Fluor 405), and α -H2B (red/Alexa Fluor 647). The fraction used for SCAN was 17.	65
3.8	(a) Coincidence of α -H3 and chromatin. (b) Coincidence of α -H2B and chromatin. (c) 3-color coincidence of α -H2B, α -H3, and chromatin.	67
3.9	Coincidence from mouse fibroblast chromatin (a) Coincidence of α -H3K27me3 and chromatin. (b) Coincidence of MBD1 and chromatin. (c) 3-color coincidence of α -H3K27me3, MBD1, and chromatin.	69
3.10	Coincidence from mouse fibroblast 60.1 chromatin (a) Coincidence of α -H3K27me3 and chromatin. (b) Coincidence of MBD1 and chromatin. (c) 3-color coincidence of α -H3K27me3, MBD1, and chromatin.	70
3.11	PCH for YOYO-1 labeled chromatin using Gamma-Poisson distribution. The noise fit starts to diverge from the data at SNR ≈ 10 or counts ≈ 20 suggesting a reasonable thresholding value. Adjusted $R^2 = 0.9998$	75

3.12	(a) Illustration showing the distribution overlaps for chromatin samples. Depending on the chromatin length, the desired population distribution may be shifted farther left than shown. (b) Burst count distribution for YOYO-1 labeled di- and trinucleosome chromatin. Signal can't be separated from background. (c) Semilog plot of (b)	78
3.13	Percent coincidence as YOYO-1 thresholding was varied from a lower bound of SNR=5 which corresponded to a thresholding value of ≈ 25 counts. The horizontal line is an artifact of the analysis and doesn't represent a leveling off percent coincidence.	79
3.14	Alexa Fluor 647 burst count distribution. Thresholding was set at SNR=10.	81
3.15	Underestimation of molecule counts as signal distributions tended towards the noise. The red line shows the ideal linear behavior	82
3.16	Noise reduction via wavelets. (a) Raw time trace (b) Wavelet denoised time trace using a symlet4 wavelet with a decomposition level of 5 and soft thresholding	87
3.17	Agarose gel showing sucrose gradient purified chromatin. Samples were generally taken from the far right lane consisting of di- and trinucleosomes.	88
3.18	Electrophoretogram of 525/40 channel with a HeLa-GFP sample. The right peak represents chromatin in the form of GFP. The left peak is unknown very low intensity contaminants. The thresholding was set at SNR=10. As the SNR increases, the left peak vanishes quickly relative to the right peak.	91
3.19	Monte Carlo simulations showing underreported coincidence. Coincidence molecules/minute: 100 (red), 250 (blue), 500 (purple), 1000 (green). Total simulation time: 30 minutes. (a) Unbound Species B = 1000 molecules/min (b) Unbound Species B = 2000 molecules/min (c) Unbound Species B = 5000 molecules/min	96
4.1	Schematic illustrating the operating principle of mobility-SCAN. Single molecule electrophoretic mobilities are determined by measuring the migration time between two observation volumes separated by a detection length, L_d	104
4.2	U-shaped nanochannel for single molecule mobility measurements. The red dots illustrate a possible beam spot location. By translating along the horizontal axis, measurements at different channel lengths can be obtained. The scale bar is $25 \mu m$	126
4.3	$50 \mu m$ nanochannel for single molecule mobility measurements. The red dots illustrate a the typical beam spot location. The scale bar is $10 \mu m$	129

4.4	Filter Array. The scale bar is $10 \mu\text{m}$	130
4.5	Optical setup for 2-color single molecule mobility experiments. Quarter wave plates in excitation paths were omitted for clarity.	135
4.6	Fiber number vs. coupling factor for 488 nm (Green) and 637 nm (Red) lasers. Larger coupling observed in the 637 nm laser can be attributed to a larger spot size. Error bars are smaller than the symbols (maximum error= ± 0.1 dB).	137
4.7	Chemical structures of the ATTO dyes used for EOF measurement. Note that the carboxylic acid functional groups will be deprotonated at neutral pH since they have $\text{pK}_a \approx 4.8$. (a) ATTO 565 with amine modification. (b) ATTO 647N with a carboxylic acid modification.	142
4.8	(a) 2 channel time traces for Alexa Fluor 647. Bursts are detected in Channel 1 and subsequently detected in Channel 2. (B) Electrophoretograms of Alexa Fluor 647 for 200 V (black), 100 V (blue), 50 V (red), and 25 V (green). The increased noise of the longer migration time peaks is an artifact of the binning of the histogram as their bins had fewer counts relative to their faster counterparts due to peak broadening. (c) 2 channel time traces for TOTO-3 labeled 900 bp DNA. (D) Electrophoretograms of the 900 bp DNA for 200 V (black), 100 V (blue), 50 V (red), and 25 V (green).	145
4.9	Normalized cross correlation of Channel 1 and 2 signals from TOTO-3 labeled 900 bp DNA experiment. V_a was varied as 200 V (black), 100 V (blue), 50 V (red), and 25 V (green)	146
4.10	V_a^{-1} vs. migration time. (a) Alexa Fluor 647. Error bars are smaller than the symbols (± 0.8 ms for maximum standard error of the mean). (b) 900 bp DNA fragments labeled with TOTO-3 intercalating dye. Error bars are smaller than the symbols (± 0.1 ms for maximum standard error of the mean)	148
4.11	L_d vs. migration time. (a) Alexa Flour 647 dye. Error bars are smaller than the symbols (± 0.02 ms for maximum standard error of the mean). (b) 900 bp DNA fragments labeled with TOTO-3 intercalating dye. Error bars are smaller than the symbols (± 0.007 ms for maximum standard error of the mean).	150
4.12	(a) V_a vs. theoretical plates for Alexa Fluor 647. The fitted curve is given by $N = \beta V_{\text{Applied}} + \alpha$ where $\beta = 29.2 \pm 0.8 \text{ V}^{-1}$ and $\alpha_2 = -60 \pm 70$. (b) V_a vs. theoretical plates for 900 bp DNA fragments labeled with TOTO-3 intercalating dye. The fitted curve is given by $N = L_d^2 / (\beta_2 V_a + \beta_1 + \beta_0 / V_a)$ where $\beta_2 = (2.4 \pm 0.5) \times 10^{-2} \text{ V}^{-1} \cdot \mu\text{m}^2$, $\beta_1 = (2.0 \pm 0.1) \mu\text{m}^2$, and $\beta_0 = (3.119 \pm 0.004) \times 10^2 \text{ V} \cdot \mu\text{m}^2$ with an adjusted $R^2 = 0.9391$	153

4.13	(a) L_d vs. theoretical plates for Alexa Fluor 647. The voltage was set at $V_{Applied} = 200$ V. The fitted curve is given by $N = \beta_{N,L}L_d + \alpha$ where $\beta_{N,L} = 10.6 \pm 0.2 \mu\text{m}^{-1}$ and $\alpha = -40 \pm 20$ with $R^2 = 0.99237$. (b) L_d vs. theoretical plates for 900 bp DNA fragments labeled with TOTO-3 intercalating dye. The least-squares fit gave $\beta_2 = (6.0 \pm 0.3) \times 10^{-5}$, $\beta_1 = (1.3 \pm 0.2) \mu\text{m}$, and $\beta_0 = (8.37 \pm 0.5) \times 10^{-1} \mu\text{m}^2$ with an adjusted $R^2 = 0.9974$.	155
4.14	(a) Burst count distribution for a 1 kb DNA ladder with an additional 250 bp fragment added. (b) Burst counts vs. DNA length. (c) Ratio of the standard deviation to the mean counts vs. DNA length. The resolution improves with length.	157
4.15	(a) Burst count distribution showing 1 kb, 5 kb, 10 kb, 20 kb, and 48.5 kb DNA fragment peaks. A significant amount of noise in the range of 300 counts and less existed due to short fragments of DNA in solution. SNR=10. (b) Experimental electrophoretogram for λ (black), 20 kb (blue), 10 kb (red), and 1 kb (green) DNA fragments. 5kb was not plotted for readability. (c) Electrophoretic mobility (red, left scale bar) and migration time (black, right scale bar) vs. DNA length	159
4.16	(a) Migration time distribution of WGA with exponential modified Gaussian fits for each distribution. These results are consistent with a Poisson distributed degree of labeling with a mean of $\mu = 1.93$. (b) Migration time/mean burst counts contour plot. The three distributions highlighted in red correspond to the three peaks observed in (a) signifying single dye mobility resolution.	162
4.17	Unnormalized Poisson distribution. The fit gave a degree of labeling of 1.93 ± 0.01 dyes/molecule.	163
4.18	Dye number vs. $1/\max(t_m)$. The linear relationship predicts the peak for 4 dyes should be located a 65 ± 4 ms.	165
4.19	(a) Migration time distributions of TOTO-3 labeled trinucleosome HeLa-GFP chromatin. TOTO-3 (red) has a bimodal distribution signifying a faster DNA band and a slower chromatin band. GFP (green) only has a single peak which is colocalized with the second TOTO-3 peak. The red chromatin peak is skew slightly which is most likely due to partly chromatin dissociation. The plots were normalized to the peak of the red chromatin peak to keep the ratio between green and red events constant. (b) Contour plot of TOTO-3/GFP coincidence and TOTO-3 transit time showing that coincident chromatin molecules correspond to the second peak only. A 2D Gaussian fit of the observed peak gave at $\bar{t}_{m,2Dx} = 12.3 \pm 0.2$ ms.	167

4.20	(a) Normalized contour plot of TOTO-3 migration time vs. mean burst intensity. The DNA and chromatin peaks are clearly visible. (b) Zoomed in normalized DNA peak showing two subpopulations. We hypothesize that the two peaks correspond to di- and trinucleosomes. (c) Zoomed in normalized chromatin peak showing two subpopulations. We conjecture that the dominant peak corresponding to the slowest migration represents fully chromatinized chromatin while the other faster peak is composed of partially dissociated chromatin.	170
4.21	Electrophoretograms of TOTO-3 labeled trinucleosome chromatin with crosslinking (black) and no crosslinking (red).	172
4.22	Electrophoretograms for the 6 biomolecules over varying detector lengths. Positive time represents flow from the anode to the cathode and negative time represents flow from cathode to anode. (a) $L_d = 40\mu\text{m}$ (b) $L_d = 240\mu\text{m}$ (c) $L_d = 570\mu\text{m}$. (d) Resolution of DNA/BSA (red) and BSA/SBA (black).	175
4.23	Electrophoretograms of a MBD/DNA mixture. (a) Electrophoretograms of Alexa Fluor 647 labeled MBD (red) and YOYO-1 labeled DNA (green). Overlapping peaks at 3.8 ms suggested a binding induced mobility shift of MBD1 while the red secondary represented unbound MBD. DNA only had 1 measurable peak consisting of both unbound and bound molecules. (b) Contour plot illustrating that the 3.8 ms peak consisted of MBD/DNA bound complexes.	179
4.24	(a) Electrophoretograms of the α -H3 antibody (red) and the YOYO-1 (green). Two peaks are shown for the YOYO-1 which coincide to DNA and chromatin bands respectively. The α -H3 peak also has two bands but of opposite direction of migration. Binding causes a reversal of flow direction for the α -H3. Positive time represents flow from the anode to cathode. (b) Contour plot of α -H3 migration time against α -H3 YOYO-1 coincidence time. A peak is observed centered at $\bar{t}_x = 9.67 \pm 0.07$ ms signifying that YOYO-1/ α -H3 coincidences corresponds to the chromatin peak. (c) Zoomed in contour plot of b.	182
4.25	Migration times for the single molecule complexes. Boxes represent $\bar{t}_m \pm \sigma_m$ for each band.	184

4.26	Electrophoretograms of a BSA/ α -BSA/secondary Ab mixture. Positive time corresponds to anode to cathode flow. (a) Electrophoretogram of Alexa Fluor 647 labeled BSA. The fastest peak corresponds to free BSA. The slower peak is a combination of BSA/ α -BSA/secondary Ab complexes and BSA/ α -BSA complexes. (b) Electrophoretogram of Alexa Fluor 488 labeled secondary Ab. The positive time peak corresponds to BSA/ α -BSA/secondary Ab complexes. The negative time peak corresponds to α -BSA/secondary Ab complexes and free secondary Ab. (c) Contour plot illustrating that BSA/ α -BSA/secondary Ab complexes are indeed coincident.	186
4.27	Electrophoretograms of control experiments. (a) BSA (red) and BSA/ α -BSA (black) (b) Secondary Ab.	187
5.1	Bright field image of straight channel sorting device. The input (bottom), sort (top), and default (left) are all shown. The angle between the default and sort is 45°.	196
5.2	Optical and electrical setup for sorting. Real-time sorting was achieved by signal processing with a FPGA	197
5.3	Time traces for intensity-based sorting. The sorting threshold was 400 kcps. All molecules were identified correctly by the FPGA. Over the 1.0 s interval, we had 1 false negative and 1 false positive count due to backflow.	200
5.4	Time traces for chromatin sorting. 6 coincident molecules were identified over the 200 ms time frame. All 6 appear to be sorted. The sort at $t \approx 0.11$ s had false positives due to either with backflow, leakage, sort pulse molecule overlap, or a combination of all three.	202
5.5	False positive probabilities as a function of input molecules per minute and sorting pulse width time.	204
5.6	COMSOL simulations showing leakage and backflow. Input channel=left, default channel=top, sort channel=bottom. (a) Half voltage condition: $V_{\text{default}} = V_{\text{sort}}/2$. (b) Backflow condition: $V_{\text{default}} < V_{\text{sort}}/2$. (c) Leakage condition: $V_{\text{default}} > V_{\text{sort}}/2$	205
5.7	Schematic representation of the 5 port device to eliminate backflow and leakage with COMSOL coloring showing voltage levels within the device. (red = high voltage, blue = low voltage)	206
5.8	Circuit schematic for voltage control of 1 port of the 5 port sorter	207

CHAPTER 1

INTRODUCTION

Since the completion of the Human Genome Project in the early 2000's, the progress of biological fields relating DNA sequencing technology has continued to rapidly proceed as the costs associated with sequencing have steadily declined. One such field, epigenetics, focuses not on the DNA "blueprint" which is already largely known, but on the mechanisms through which the genetic blueprint is regulated and expressed in both space and time. Regulation of gene expression is largely dependent on the dynamic interplay of the underlying DNA sequence and chromatin, a genome-organizing platform consisting of histone proteins. These histone proteins form an octomeric histone core (nucleosome) which is the fundamental repeating unit of the chromatin structure. Regulatory information is encoded with these nucleosomes in the form of nucleosome remodeling, histone variants, and histone modifications on their N-terminus tail structure. These histone modifications form the basis for the hypothesized histone code. A large number of histone modifications are known to exist including methylation (mono-, di-, trimethylation), acetylation, phosphorylation, sumoylation, and ubiquitylation [1]. Histone modifications act to spatially modulate chromatin structure and recruit various "reader" proteins [2]. Broadly speaking, epigenetics tries to understand observed phenotypic variation caused not by changes in the DNA sequence, but instead generated by alterations to chromatin. These alterations may be brought about by the previously mentioned histone modifications or by chemical modifications to the nucleotides themselves, most notably 5-methylcytosine (5mC). 5mC acts to provide sites for various methyl binding domain proteins to anchor resulting in a transformed chromatin structure and reduced gene expression [3].

Epigenetic mechanisms are most likely fundamental to understanding biological questions regarding cellular differentiation, stem cell plasticity, aging, cancer, and more [4–11]. Cancer, perhaps the most researched epigenetic related disease, has been clearly linked to epigenetic aberrations occurring with the onset of tumorigenesis [12]. Moreover, significant interest from the medical community regarding the link between cancer and epigenetics has resulted in epigenetic therapeutic drugs and positive clinical outcomes. Current drugs on the market utilize histone deacetylase (HDAC) inhibitors or DNA methyltransferases (DNMTs) and their inhibitors [13, 14].

The present state of understanding with regards to epigenetics has relied almost completely on two techniques: chromatin immunoprecipitation followed by sequencing (ChIP-seq) and bisulfite sequencing (BS-seq) [15]. ChIP involves the selective immunoprecipitation of chromatin fragments using antibodies targeting the histone modifications or chromatin binding proteins of interest followed by DNA sequencing. Bisulfite sequencing is the application of bisulfite to DNA followed by sequencing to determine the methylation profile. Bisulfite treatment of DNA results in unmethylated cytosine converting to uracil but does not alter 5mC.

Conventional ChIP-seq is an ensemble based method requiring a large amount of starting material. Most ChIP-seq protocols require 10^6 to 10^7 cells for adequate results [16, 17]. Inefficient antibody pull-down, dissociation of the DNA-protein complexes, and downstream processing losses all influence the efficiency and sensitivity of the assay. Recently, specialized microfluidic platforms have been developed to decrease the losses associated ChIP and analyze samples from lower cell counts [18, 19]. These devices have generally

been made using soft lithography techniques with PDMS and have complex integration of valve-actuated control which has limited their use beyond proof-of-principle. Further, these protocols were demonstrated using locus-specific qPCR which limits the obtained information compared to high-throughput sequencing. More recently, methodologies have been developed to enable ChIP-seq on as few as 100 cells enabling ChIP-seq to be performed on low abundance cells [20]. Additionally, their protocol did not rely on potentially biasing preamplification. Most recently, the ultimate resolution for ChIP-seq was achieved by performing ChIP-seq on single cells with a methodology involving droplet microfluidics [21]. Still, obtaining ChIP-seq information from low cell counts remains a challenging task.

Unlike ChIP-seq, advances in BS-seq technology, namely in the realm of protocols and not specialized devices, have given single cell resolution to BS-seq. It had been previously thought that bisulphite based methods were incompatible with single cell studies as significant DNA degradation and loss occurs during the bisulfite conversion, which deaminates unmodified cytosines and converts them to uracil [22]. However, by eliminating multiple purification steps and performing all reactions in a single tube, researchers have demonstrated single cell reduced representation bisulfite sequencing (scRRBS) on single mouse embryonic stem cells [23]. They recovered about 1 million unique CpG sites with bisulfite conversion rate of greater than 99%. Further, approximately half of the CpG sites were sequenced with a deep enough coverage for reliable quantification. Very recently, another advancement was demonstrated as genome-wide BS-seq was accomplished for the first time by using a modified post-bisulfite adaptor tagging protocol [24]. Using this method, researchers demonstrated a 48.4% CpG coverage across the genome.

As discussed above, both ChIP-seq and BS-seq have already reached single-cell capabilities. What has not been demonstrated, however, is the capability of assaying multiple epigenetic marks for small batches of material. The ability to examine multiple epigenetic marks simultaneously, with high resolution, allows for the study of their interdependency which has been hypothesized in the histone code and has been demonstrated in literature [15, 17, 25–27].

Unfortunately, standard techniques for multiple epigenetic mark interrogation such as re-ChIP or ChIP-BS-seq rely on simple sequential application of their respective methods. This implies significant losses due to the multiplicative effect of serial processing. Studies investigating more than two marks have proven essentially infeasible due to low yields. Moreover, chromatin concentrations after the 2nd re-ChIP are typically so low that, based on the dissociation constants of antibodies, inefficient chromatin-antibody binding occurs at best. Additionally, while microfluidic based platforms have performed ChIP on low cell counts, to the best of our knowledge, re-ChIP has never been demonstrated using these microfluidic devices. Thus, the ability to study the coordination of multiple epigenetic marks for small samples, ideally single cells, remains a challenge.

With that said, the goal of my graduate studies was to develop technologies to address the limitations of the traditional methods. We were motivated to provide micro- and nanoscale technological solutions to the epigenetic community. The technologies and methods described within this dissertation are therefore intended to address the issues hindering epigenetic analysis from small cell populations or for multiple epigenetic marks.

1.1 Organization of this Dissertation

We have briefly discussed the numerous reasons why researchers are investigating the epigenome in this post-genomic era. After succinctly discussing the current epigenetic methods, we motivated the notion that microstructures were necessary for epigenetic analysis on small amounts of material due to the losses with the standard techniques. In the remaining pages of this dissertation, I will discuss the my efforts regarding developing techniques using micro- and nanoscale devices for the analysis of the epigenome.

Chapter 2 describes a preparatory microdevice for the extraction, purification, and stretching of chromosomal DNA and chromatin from small cell populations down to a single cell. With the microfluidic device, we are able to effectively capture, purify, and fluorescently label the DNA and/or chromatin. We demonstrate nearly 100% extraction efficiency of the DNA using the device. Further, we show in-channel immunostaining of chromatin followed by downstream single molecule analysis of the sample.

Chapter 3 focuses on quantification of epigenetic marks using a single molecule method called single-molecule chromatin analysis in nanochannels (SCAN). Using multi-color, time-correlated single molecule measurements in nanochannels, we aim to quantify the abundance of epigenetic marks on a genome-wide scale. We show coincidence detection of 2 epigenetic marks demonstrating that SCAN can possibly be used to quantify marks similar to mass spectrometry or ChIP. Further, we modify our detection scheme to enable coincidence detection of three epigenetic marks signaling the potential capabilities of the system beyond what ChIP can achieve. We conclude Chapter 3 by

examining the issues regarding difficulties with quantifying epigenetic marks using SCAN.

Chapter 4 describes the development of mobility-SCAN, an analytical technique for measuring and analyzing single molecules based on their fluorescent signature and their electrophoretic mobility. Similar to capillary electrophoresis but the single molecule scale, we use the technique to differentiate biomolecules from complex mixtures and derive parameters such as diffusion coefficients and effective charges. Further, we demonstrate, at a single molecule level, binding induced mobility shifts of protein-DNA, protein-chromatin, and protein-protein interactions. We use mobility-SCAN to discriminate chromatin complexes and show the utility of the method as a diagnostic and analytical tool for quantitative SCAN and SCAN-sort.

Chapter 5 briefly discusses the current state of SCAN-sort, a technique to sort individual molecules based on their fluorescent emissions for further downstream analysis such as DNA sequencing. We show our best results using chromatin which was over a 2-fold enrichment and discuss possible system modifications for better performance.

Chapter 6 provides relevant discussions regarding single cell and single molecule techniques, especially with regards to the methods described in the previous chapters.

BIBLIOGRAPHY

- ¹C. D. Allis, T. Jenuwein, D. Reinberg, and M.-L. Caparros, *Epigenetics* (Cold Spring Harbor Laboratory Press Cold Spring Harbor, NY: 2007).
- ²D. J. Patel and Z. Wang, “Readout of epigenetic modifications”, Annual review of biochemistry **82**, 81–118 (2013).
- ³M. Fatemi and P. A. Wade, “Mbd family proteins: reading the epigenetic code”, Journal of cell science **119**, 3033–3037 (2006).
- ⁴T. Chen and S. Y. Dent, “Chromatin modifiers and remodelers: regulators of cellular differentiation”, Nature Reviews Genetics **15**, 93–106 (2014).
- ⁵D. G. Tang, “Understanding cancer stem cell heterogeneity and plasticity”, Cell research **22**, 457–472 (2012).
- ⁶M. F. Fraga and M. Esteller, “Epigenetics and aging: the targets and the marks”, Trends in Genetics **23**, 413–418 (2007).
- ⁷V. Calvanese, E. Lara, A. Kahn, and M. F. Fraga, “The role of epigenetics in aging and age-related diseases”, Ageing research reviews **8**, 268–276 (2009).
- ⁸A. Brunet and S. L. Berger, “Epigenetics of aging and aging-related disease”, The Journals of Gerontology Series A: Biological Sciences and Medical Sciences **69**, S17–S20 (2014).
- ⁹C. Plass, S. M. Pfister, A. M. Lindroth, O. Bogatyrova, R. Claus, and P. Lichter, “Mutations in regulators of the epigenome and their connections to global chromatin patterns in cancer”, Nature Reviews Genetics **14**, 765–780 (2013).
- ¹⁰M. T. Villanueva, “Epigenetics: chromatin marks the spot”, Nature Reviews Cancer **15**, 196–197 (2015).

- ¹¹M. A. Dawson and T. Kouzarides, “Cancer epigenetics: from mechanism to therapy”, *Cell* **150**, 12–27 (2012).
- ¹²M. Esteller, “Epigenetics in cancer”, *New England Journal of Medicine* **358**, 1148–1159 (2008).
- ¹³R. M. Campbell and P. J. Tummino, “Cancer epigenetics drug discovery and development: the challenge of hitting the mark”, *The Journal of clinical investigation* **124**, 64–69 (2014).
- ¹⁴A. Nebbioso, V. Carafa, R. Benedetti, and L. Altucci, “Trials with epigenetic drugs an update”, *Molecular oncology* **6**, 657–682 (2012).
- ¹⁵B.-R. Hyun, J. L. McElwee, and P. D. Soloway, “Single molecule and single cell epigenomics”, *Methods* **72**, 41–50 (2015).
- ¹⁶T. I. Lee, S. E. Johnstone, and R. A. Young, “Chromatin immunoprecipitation and microarray-based analysis of protein location”, *Nature protocols* **1**, 729–748 (2006).
- ¹⁷B. E. Bernstein, T. S. Mikkelsen, X. Xie, M. Kamal, D. J. Huebert, J. Cuff, B. Fry, A. Meissner, M. Wernig, and K. Plath, “A bivalent chromatin structure marks key developmental genes in embryonic stem cells”, *Cell* **125**, 315–326 (2006).
- ¹⁸A. R. Wu, J. B. Hiatt, R. Lu, J. L. Attema, N. A. Lobo, I. L. Weissman, M. F. Clarke, and S. R. Quake, “Automated microfluidic chromatin immunoprecipitation from 2,000 cells”, *Lab on a Chip* **9**, 1365–1370 (2009).
- ¹⁹T. Geng, N. Bao, M. D. Litt, T. G. Glaros, L. Li, and C. Lu, “Histone modification analysis by chromatin immunoprecipitation from a low number of cells on a microfluidic platform”, *Lab on a Chip* **11**, 2842–2848 (2011).
- ²⁰Z. Cao, C. Chen, B. He, K. Tan, and C. Lu, “A microfluidic device for epigenomic profiling using 100 cells”, *Nature methods* (2015).

- ²¹A. Rotem, O. Ram, N. Shores, R. A. Sperling, A. Goren, D. A. Weitz, and B. E. Bernstein, “Single-cell chip-seq reveals cell subpopulations defined by chromatin state”, *Nat Biotech* **33**, 1165–1172 (2015).
- ²²N. Plongthongkum, D. H. Diep, and K. Zhang, “Advances in the profiling of dna modifications: cytosine methylation and beyond”, *Nature Reviews Genetics* **15**, 647–661 (2014).
- ²³H. Guo, P. Zhu, F. Guo, X. Li, X. Wu, X. Fan, L. Wen, and F. Tang, “Profiling dna methylome landscapes of mammalian cells with single-cell reduced-representation bisulfite sequencing”, *Nature protocols* **10**, 645–659 (2015).
- ²⁴S. A. Smallwood, H. J. Lee, C. Angermueller, F. Krueger, H. Saadeh, J. Peat, S. R. Andrews, O. Stegle, W. Reik, and G. Kelsey, “Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity”, *Nature methods* **11**, 817–820 (2014).
- ²⁵B. D. Strahl and C. D. Allis, “The language of covalent histone modifications”, *Nature* **403**, 41–45 (2000).
- ²⁶R. Metivier, G. Penot, M. R. Hubner, G. Reid, H. Brand, M. Kos, and F. Gannon, “Estrogen receptor-alpha directs ordered, cyclical, and combinatorial recruitment of cofactors on a natural target promoter”, *Cell* **115**, 751–763 (2003).
- ²⁷E. Soutoglou and I. Talianidis, “Coordination of pic assembly and chromatin remodeling during differentiation-induced gene activation”, *Science* **295**, 1901–1904 (2002).

CHAPTER 2

MICROFLUIDIC EXTRACTION AND ANALYSIS OF HUMAN CHROMOSOMAL DNA AND CHROMATIN FROM SINGLE CELLS

2.1 Introduction

¹Genome-wide analysis of single cells is important in life science research and modern medicine in applications ranging from cancer diagnosis to understanding tissue development [2, 3]. Microfluidic devices have been explored as a promising platform for single cell studies, providing superior handling of minute sample and reagent volumes in engineered microstructures [4–9]. Microfluidic devices operate in an enclosed system, thereby reducing the chance for human error and cross contamination. Furthermore, the low reagent volumes needed give microfluidic devices a distinct advantage in lowering both the time and cost of operation. Isolation of nucleic acids from biological samples is an essential step for every type of genetic analysis. While numerous extraction methods have been explored [10–15], it remains rather challenging to isolate and analyze genomic DNA from small cell populations and single cells. Traditional microfluidic platforms utilize solid phase extraction (SPE) [15], a method dependent upon the binding of DNA to solid phase matrices such as silica [15] or functionalized magnetic microparticles [16] for extraction of nucleic acids from cell lysates. Various surface treatment protocols have been developed to enhance the affinity of the surface of microdevices to nucleic acids [13–15]. However, this binding affinity is extremely sensitive to factors such as pH, temperature, and buffer compositions which often require dynamic con-

¹Sections 2.1-2.3.1 were adopted with permission from [1].

trol in order to minimize DNA losses. Additionally, other components of the cellular lysate may potentially inhibit the intended binding reactions. For example, negatively charged proteins in the cell lysate can decrease the efficiency of extraction by interacting with positively charged surfaces within the device. Even when the binding chemistry is optimized, it is difficult to ensure that all of the DNA fragments are adsorbed on the solid phase matrix and that the whole genome is represented in purified extracts. An appreciable fraction of genomic DNA is often lost during the purification process when the cell debris is washed away. Additional DNA losses can be caused by incomplete elution. As a result, the current state-of-the-art microfluidic devices for DNA separation from cell lysates exhibit rather modest extraction efficiencies of 60-90% [11]. This level of extraction is sufficient for genetic analysis of large cellular populations as multiple copies of every gene are present within the extract, which statistically guarantees complete genome representation, but such collection losses are unacceptable when single-copy genes in a single cell are being investigated. Even with improved yield, it is fundamentally difficult to apply SPE techniques to single cell studies due to manufacturing complexities and technical challenges associated with manipulation of small volumes of fluids. Current microfluidic devices with such capabilities rely on complicated fluidic networks of channels, valves, and reaction chambers (or droplets) for storing, transporting, and mixing picoliter volumes of sample and reagents [6, 7]. Fundamentally different approaches to the isolation of genomic DNA for characterization of single cells should be explored.

This work describes a simple, valveless, two-port microfluidic device capable of highly-efficient isolation and fluorescent analysis of DNA or chromatin from single cells by trapping and elongating long strands of human genomic

DNA or chromatin in two-dimensional arrays of micropillars. The flow of DNA/chromatin through arrays of obstacles involves collisions which impede its propagation and can result in immobilization if the fragment size is sufficiently large [17]. Although DNA transport in obstacle arrays have been studied extensively for fragment size separation and elongation [18–20], entrapment with hydrodynamic flow is yet to be used to extract and purify human chromosomal DNA from cell lysates. This approach is fundamentally different from the conventional microchip-based SPE as well as from physical filtration through nanopores, which can cause DNA shearing and often results in device failure due to clogging issues [21]. In the microfluidic device presented here, long strands of human chromosomal DNA or chromatin released from cells through chemical lysis are looped around PDMS micropillars and are physically retained while the remaining cellular contents are washed away under hydrodynamic flow. In this manner, large genomic DNA or chromatin is separated from a multitude of cellular debris such as proteins and lipid membrane fragments as well as from much smaller mitochondrial DNA and RNA. Random DNA fragmentation is minimized by operating at low flow rates. This physical extraction method provides several unique capabilities that make it attractive for genetic and epigenetic analysis of DNA and chromatin contents of single cells. The suspended DNA or chromatin can also be released from the microarrays by enzymatic digestion with restriction endonucleases and subsequently collected for downstream analysis of the digestion product in nanofluidic channels by DNA fragment sizing [22] or multicolor fluorescence [23, 24]. Genome-wide analysis of DNA or chromatin contents of single cells could thus be performed one molecule at a time, without the need for amplification. To demonstrate the applicability and utility of the presented microfluidic device regarding single cell

analysis, the device was used to extract, purify, and elongate DNA and chromatin from single human cells.

2.2 Materials and Methods

2.2.1 Device Design and Fabrication

The device design is centered around the micropillar array which is the essential element of the entire device. Figure 2.1 shows the device schematic and photomicrograph of a section of the fabricated microchannels supporting the random array of PDMS micropillars. The microchannels are 240 μm wide, 20 μm deep and up to 70 mm long. The array of microposts (5 μm wide and 20 μm tall) was designed with a gradient in spacing to create a solid obstacle for cell capture. The average gap between the microposts varies continuously from 15 μm to 2 μm along the channel. Though intended primarily for single cell analysis, the device can hold up to approximately 20 cells with a broad size distribution without clogging the channel. The extraction method can be scaled up to perform extraction of larger quantities from thousands of cells by broadening the microchannels. The 500 μm -long array of micropillars also provides a solid phase matrix for immobilization of the chromosomal DNA. Random placement of micropillars maximizes the number of collisions with solid surfaces, thereby preventing DNA or chromatin strands from easily slipping through the structure. Linear arrays of micropillars spanning the channel width, spaced by 100 μm throughout the microchannel, were also incorporated to provide support and thus minimize shear-induced damage of the hundreds of megabases long

DNA strands. To avoid strong electrostatic adsorption of DNA to both the channel walls and the micropillars, the microchannels were initially primed with a 1% (wt) solution of polyvinylpyrrolidone (PVP) (Fluka; St. Louis, MO) and 1% (wt) bovine serum albumin (BSA) (Sigma-Aldrich) in tris-EDTA-saline (TES) buffer (10 mM Tris-HCl, 20 mM EDTA, 100 mM NaCl) for 5 hours. The coated microchannels were rinsed with TES buffer for 1 hour at a rate of 30 nL/min in room temperature prior to experimentation.

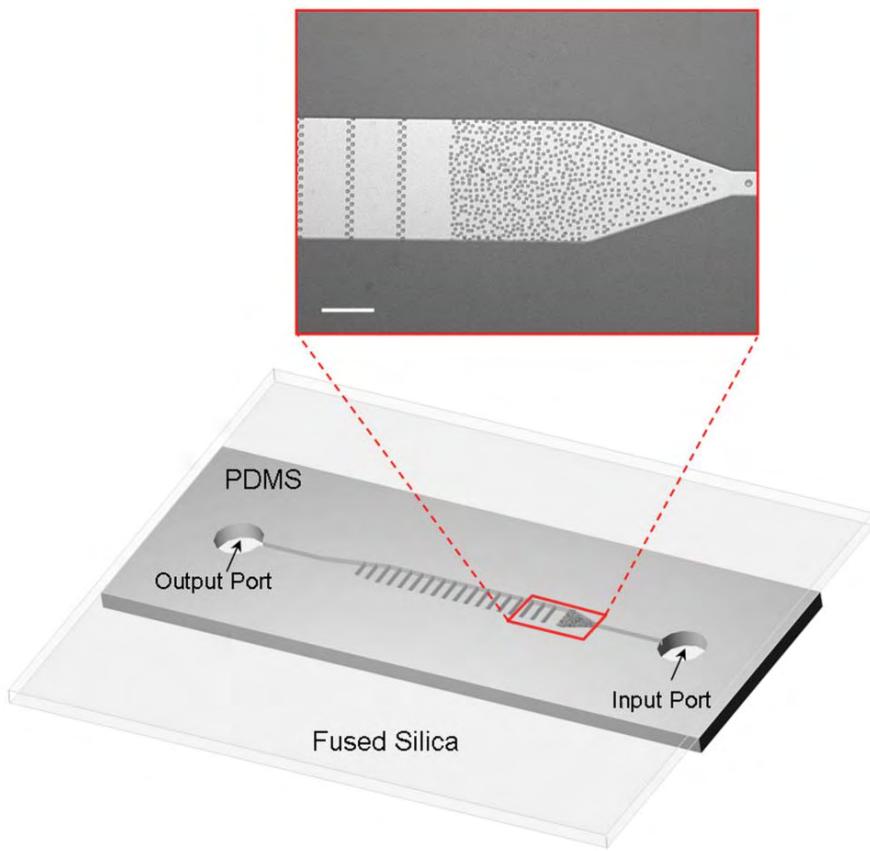


Figure 2.1: Schematic and micrograph of the microfluidic device. The channel consists of a random array of micropillars for cell capture and DNA extraction. The non-random pillars are for structural support. The scale bar is 100 μ m. Figure reproduced with permission from [1].

To fabricate the master molds, Microposit S1813 photoresist (Shipley; Marlborough, MA) was spun on silicon on insulator (SOI) wafers (Ultrasil; Hayward,

CA) and exposed by UV contact lithography (EV 620, EVG Group; Albany, NY). The exposed resist was developed in 726MIF developer (Microchemicals) and the pattern was transferred into the 20 μm -thick top silicon layer by Bosch process in a Unaxis SLR 770 deep reactive ion etching system (Unaxis USA Inc.; St. Petersburg, FL). A monolayer of (1H,1H,2H,2H-Perfluorooctyl)Trichlorosilane (FOTS) was deposited on the etched wafers in a MVD 100 molecular wafer deposition system (Applied Microstructures; San Jose, CA) to prevent sticking of the PDMS to the mold. Sylgard 184 (Dow Corning; Midland, MI) PDMS base resin was mixed with the curing agent at a 10 : 1 ratio, degassed under vacuum at room temperature, poured onto the master, and cured for 45 minutes at 150 °C. The elastomer casting was then peeled off the mold and access holes to the input and outputs of the microchannels were created with a 1.5 mm biopsy punch (Sklar Instruments; West Chester, PA). To complete channel fabrication, the patterned PDMS was treated with oxygen plasma for 1 minute and bonded to a 170- μm thick fused silica wafer (Mark Optics; Santa Ana, CA). A summary of the fabrication process is given in Figure 2.2.

2.2.2 Microscopy and Experimental Setup

The silica wafer containing the microfluidic devices for DNA/chromatin extraction was placed on a custom fabricated heating stage which was mounted on an Olympus IX70 inverted microscope (Olympus; Center Valley, PA). Stage temperature was controlled with a Model 1146D heated/refrigerated recirculator (VWR; Randor, PA). The microscope used for sample visualization was equipped with a 10x objective (Plan APO, N.A. 0.45), a 20x objective (Plan APO, N.A. 0.60), a 60x water immersion objective (Olympus UPlanSApo 60x /1.2W),

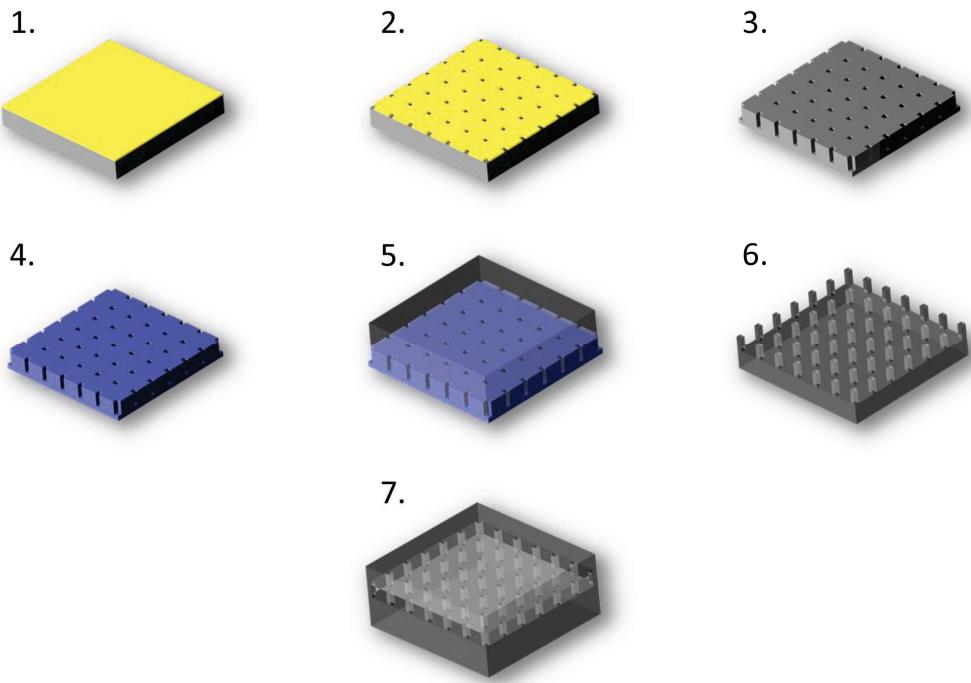


Figure 2.2: PDMS microchannel fabrication process. 1. Spin Microdeposit S1813 photoresist onto a SOI wafer. 2. Pattern resist using contact lithography. 3. Etch the silicon $20\ \mu\text{m}$ to the oxide layer. 4. Deposit a monolayer of FOTS. 5. Pour PDMS onto the silicon mold. 6. Remove cured PDMS. 7. Bond PDMS to a fused silica wafer.

an X-Cite 120 mercury light source (Lumen Dynamics; Mississauga, ON); a 485/535 nm filter cube (Chroma; Bellows Falls, VT), and a Cascade II charge-coupled device (CCD) camera (Photometrics; Tucson, AZ). A $100\ \mu\text{L}$ gas-tight syringe (Hamilton; Reno, NV) was connected to the microchannel via PEEK polymer tubing (I.D. = 254m, O.D. = 1.58mm) which was inserted into the output of the microdevice. Cells and reagents were delivered into the microdevice with a PHD 2000 pump (Harvard Apparatus; Holliston, MA) at a flow rate of $30\ \text{nL}/\text{min}$.

2.2.3 Cell Culture

M0-91 hematopoietic stem cells infected with myeloid leukemia were cultured in Dulbeccos Modified Eagle medium (DMEM) (Invitrogen) at 37 C and 5% CO₂ in a T25 flask. DMEM cell culture media was supplemented with DMEM High Glucose 1 (Gibco 11965), 100 non-essential amino acids (Gibco 11140), sodium pyruvate, 100mM (100) (Gibco 11360), 100 penicillin streptomycin (Gibco 15140), 1M HEPES buffer solution, (Gibco 15630), 2-mercaptoethanol (1:100 dilution with water) (Sigma 63689), and 10% (wt) fetal bovine serum (Atlanta Biologicals S11150). Cells were split every three days into equal volumes of fresh media.

HeLa cells expressing green fluorescent protein (GFP) on histone H2B were provided by Geoffrey M. Wahl of The Salk Institute of Biological Studies, USA. The HeLa-GFP cells were cultured under the same conditions as the M0-91 cells.

2.2.4 Cell Capture and Lysis

A solution containing M0-91 cells or HeLa-GFP cells suspended in original culturing media was injected into the input port of the microdevice and drawn under a constant flow into a random array of micropillars in which the cells become immobilized. The pressure driven flow was interrupted when a single cell was captured in the microarray and the input port was repeatedly rinsed to remove any remaining cells. The microchannel with a single immobilized cell was then rinsed with a TES buffer to remove residues of the cell culture growth medium. The M0-91 cells were lysed with a solution containing 1% (w/v) sodium dodecyl sulfate (SDS) in the same buffer. The HeLa cells

were lysed 1xTE buffer with 500 mM NaCl and 1% Triton-X 100 (v/v) to preserve the chromatin structure. The long strands of the released chromosomal DNA/chromatin became entangled in the microchannel under hydrodynamic flow while the unwanted components of the cell lysate such as lipid membrane fragments, non-genomic DNA, and RNA were washed away.

2.2.5 DNA Purification and Visualization

The immobilized genomic DNA was rinsed and purified by flowing a buffer containing proteinase K (Qiagen; Valencia, CA) through the microchannels to remove any remaining histone proteins bound to the stretched DNA strands. Thorough removal of the cellular debris and the lysis agent is often desirable for sample preparation to prevent interference with downstream processes such as polymerase chain reaction (PCR) amplification or single-molecule fluorescence analysis. The trapped chromosomal DNA was fluorescently labeled using the DNA intercalating fluorescent dye PicoGreen (Invitrogen; Carlsbad, CA) and visualized with the fluorescence microscopy setup described in Section 2.2.2. Background fluorescence inside the microchannel was reduced by washing off the unbound dye surrounding the entrapped chromosomal DNA.

2.2.6 Chromatin Purification, Visualization, and Antibody Labeling

The micropillar entrapped chromatin was rinsed and purified using 1xTE buffer with 500 mM NaCl to stabilize the chromatin. Chromatin was visualized by the

GFP on histone H2B and generally not intercalated in the channel but instead after collection. For immunostaining, Alexa Fluor 647 labeled α -H3 antibodies (Active-Motif, 39763) at a concentration of 10 μ g/mL in PBS buffer were added to the channel at a rate of 30 nL/min for 1 hour. The antibodies were crosslinked to the chromatin using a 0.75% solution of formaldehyde for 10 minutes. This crosslinking condition crosslinked the antibodies to some degree while not quenching the GFP fluorescence completely.

2.2.7 DNA Release and Off-Chip Fluorescence Analysis

To determine the extraction efficiency of the microfluidic device, the entrapped DNA strands were fragmented, collected, and quantified by fluorescence analysis using the Quant-iT PicoGreen dsDNA assay (Invitrogen). Fluorescently labeled genomic DNA suspended in the microchannel was released from the device by enzymatic digestion with BamHI (Invitrogen) restriction endonuclease. To optimize the digestion process, the microchannels were heated to 37 °C prior to the introduction of restriction enzymes. The fragmentation process was monitored in real time by observing the PicoGreen fluorescence with a CCD camera. Since the total DNA content of a single human diploid cell (\approx 6.6 pg [25]) is not sufficient for reliable off-chip fluorospectrometric quantification, the analysis was performed with larger quantities of M0-91 cells loaded into the obstacle array. The DNA trapped in the microchannel was eluted at 100 nL/min flow into less than 200 nL of 1x Digestion Buffer K (Invitrogen). Sample dilutions were controlled simply by adjusting the flow rates. To facilitate off-chip sample manipulation, the released DNA was diluted further by flowing an additional 20 μ L of TE buffer through the microdevice. The total extract was col-

lected in the polymer tubing that connects the syringe to the output channel. Once filled with fragmented DNA, the tubing was disconnected from the device and its contents were injected into a 0.2 ml Eppendorf tube for off-chip quantification. Thus, purified DNA from the known number of cells was suspended in 20 μ L of TE buffer. Solutions containing DNA extracts obtained from the microchannels were diluted with an equal amount of PicoGreen and their fluorescence intensity was measured using a NanoDrop 3300 fluorospectrometer (Thermo Scientific; Wilmington, DE). The instrument only requires a small amount of sample per measurement (1 to 2.5 μ L), thereby allowing multiple measurements ($N = 10$) to be taken for each extract. The fluorospectrometer was calibrated with bacteriophage T4 DNA standard (Wako Chemicals; Richmond, VA) and the total mass of extracted DNA was calculated from the fluorescent intensity signal.

2.2.8 Single Cell Real-Time PCR Analysis

Real-time PCR using a whole genome amplification kit (Single Cell WGA kit, NEB, E2620S/L) was used to quantify the extraction efficiency of the device for a single cell since fluorospectrometric quantification wasn't possible for such little amount of DNA. By locating the crossing point (C_p) of the qPCR, the initial quantity of input DNA was determined by comparing to a standard curve. The C_p standard curve was constructed using a batch sample of M0-91 cells. The cells were lysed using 1% SDS and treated with proteinase K. The M0-91 DNA was then purified by phenol-chloroform extraction and its concentration was determined using the PicoGreen assay. The M0-91 DNA was then diluted to 3.3, 6.6, and 9.9 μ g in a total volume of 10 μ L. The single cell whole genome am-

plification kit previously described was utilized to amplify the bulk M0-91 DNA for the C_p standard curve and the single cell's DNA which had been extracted using the methods described in Sections 2.2.4 and 2.2.7. The amplification was tracked by quantifying the fluorescence intensity of SYBR Green I (Invitrogen). C_p was calculated by finding the second derivative maximum in the amplification curve.

Figure 2.3 shows the C_p standard curve generated by the procedure described above. The linear least-squares regression gave a slope of $-(0.8 \pm 0.1) C_p/\text{pg}$, an intercept of $14.1 \pm 0.8 C_p$, and $R^2 = 0.988$.

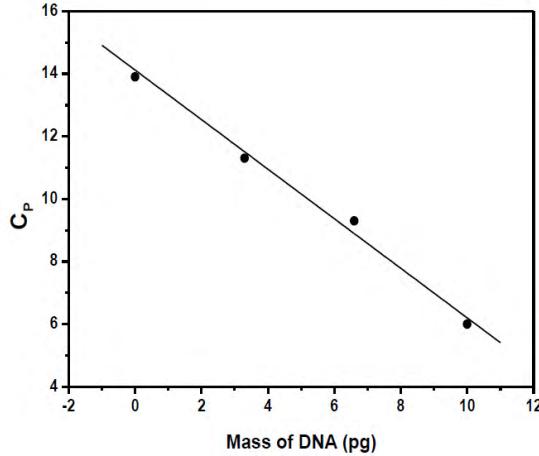


Figure 2.3: C_p standard curve derived from known amounts of DNA and a no-template control. Figure reproduced with permission from [1].

2.3 Results and Discussion

The microfluidic device described is designed to capture individual cells in an array of micropillars, perform lysis, extraction, purification, and then linearize the released human chromosomal DNA or chromatin. The rationale behind using micropillar obstacles is that chromosomal DNA is significantly larger than

the remaining cell contents and can be therefore separated from the lysate by size. Because DNA is delivered into the microdevice device in a living cell and the extraction is performed at very low flow rates, DNA fragmentation throughout the process is minimized. Figure 2.4 displays a schematic of the operation

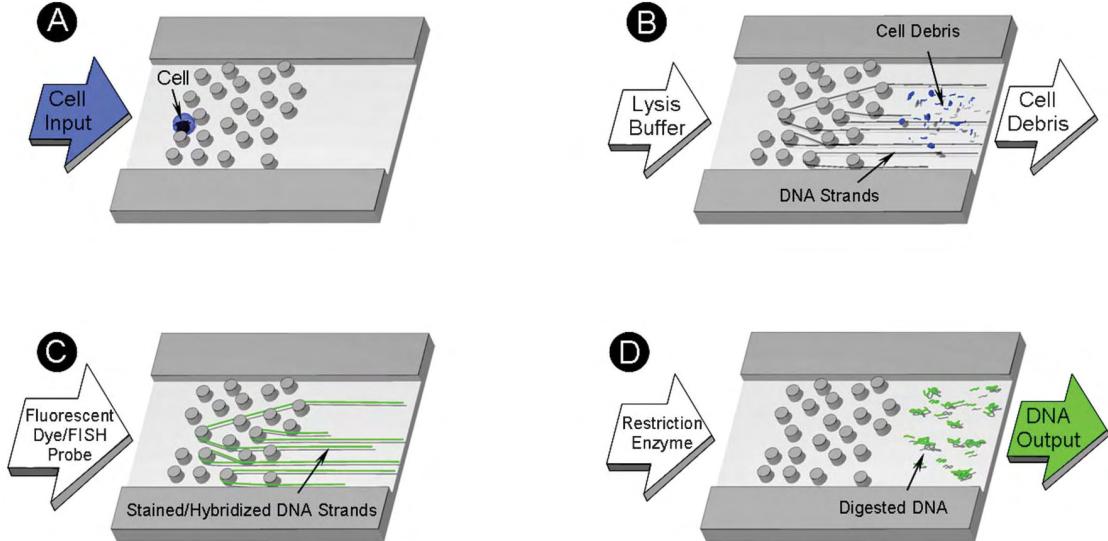


Figure 2.4: Schematic representation of the device operation. (A) A cell is captured under flow via size exclusion in the micropillars. (B) Cell lysis is performed with a ionic surfactant (1% SDS). Cell debris, nucleosomes, and other proteins are washed away leaving entangled chromosomal DNA. (C) DNA is labeled with intercalating dye. (D) The DNA is released using restriction endonucleases. Figure reproduced with permission from [1].

principles for the device for DNA. Briefly, an individual cell is delivered into the micropillar array by hydrodynamic flow. The captured cell is chemically lysed with 1% SDS, which simultaneously dissolves the plasma membrane and the nuclear membrane, thereby releasing the DNA into the lysis solution. The surfactant also gradually destroys the higher-order chromatin structure by denaturing and stripping off the histone proteins. The hydrodynamically driven DNA collides immediately with the micropillars and loops around them in a rope-over-pulley fashion. Dense arrays of randomly spaced micropillar obstacles were chosen for single-cell studies to maximize the number of collisions

encountered by the unwrapping DNA and to fan out the individual strands of chromosomal DNA as they unwind. This helps to separate and spread individual strands of DNA in the microchannel. DNA migration under hydrodynamic flow is inhibited by weak steric interactions with the non-functionalized micropillars and intermolecular interactions with other strands of chromosomal DNA.

Figure 2.5 shows a fluorescent image of DNA strands released from a single M0-91 cell. The DNA strands extend approximately 10 mm into the channel, which is shorter than the length of chromosomal DNA in B-form which ranges from 17 mm (chromosome 21) to 85 mm (chromosome 1) [25]. From the detailed fluorescence image shown in Figure 2.5B, we can see that the DNA strands were multiply folded onto themselves limiting their observed stretched length.

2.3.1 Off-Chip Quantification

To determine the extraction efficiency of the microfluidic device, multiple modified microdevices were loaded with up to 83 M0-91 cells as shown in Figure 2.6A. The modified microchannel design featured a semi-ordered array with gradually decreasing spacing to handle higher cell counts relative to the random array device. We used multiple cells because the small amount of genomic DNA contained in a single cell (6.6 pg per each human diploid cell [25]) was insufficient for reliable fluorospectrometric quantification. We also expected the error bars to be quite large for lower cell counts as the signal to noise ratio declined.

After device loading, the chromosomal DNA was isolated and labeled fol-

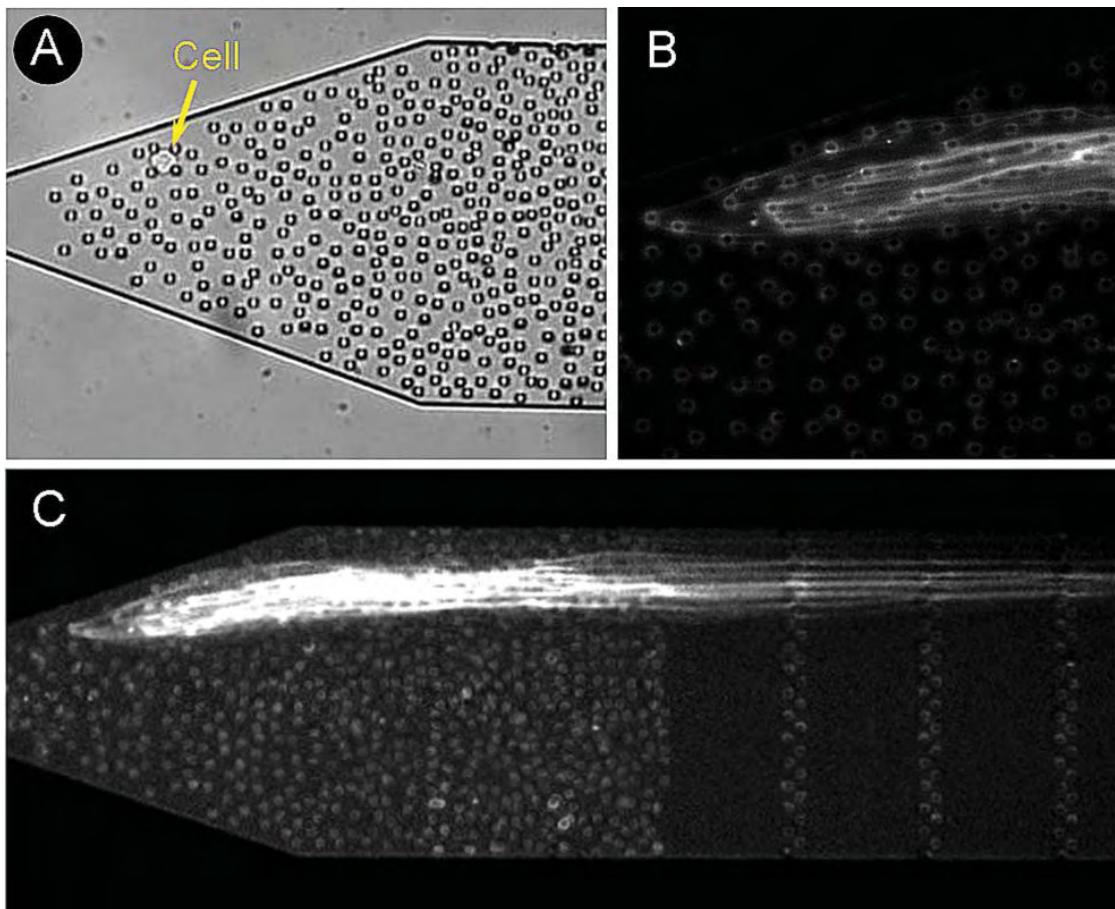


Figure 2.5: (A) A single M0-91 cell trapped in the random pillar array. (B) Detailed fluorescent image of the DNA strands held by the micropillars under flow. (C) Stretched DNA from the single cell shown in (A). The extension of the DNA was more than 10 mm downstream. Figure reproduced with permission from [1].

lowing the single cell procedures described in Sections 2.2.4 and 2.2.5. The purified DNA was released from the devices via enzymatic digestion described in Section 2.2.7. As illustrated in the sequence of images in Figure 2.6C, all DNA was released from the microchannel into the collection reservoir under 100 nL/min flow within 2 minutes.

Figure 2.6D shows the amounts of DNA extracted from six devices loaded with different numbers of M0-91 cells. The fluorescence intensity of the extracts measured with the NanoDrop 3300 fluorospectrometer was directly pro-

portional to the number of cells. The total mass of extracted DNA was calculated from the fluorescent intensity signals according to the calibration curve prepared with the bacteriophage T4 DNA standards. The solid line in Figure 2.6D gives the linear least-squares fit for the data. The curve gave a slope of 6.7 ± 0.2 pg/cell which is in excellent agreement with the expected amounts indicating that only negligible quantities of DNA are lost during the extraction process. The microfluidic device extracts essentially 100% of genomic DNA from small cell populations, outperforming both macroscopic and microchip-based SPE methods. The large vertical error bars reflect the inherent variation of the fluorospectrometer instrument. The uncertainty in cell counts is caused by the presence of dead and dying cells and other debris in the growth medium which makes cell identification and counting in micropillar arrays difficult. Only healthy, intact cells were considered in the calculation of the extraction efficiency because the DNA from dead or dying cells is already fragmented and the strands are too short to get captured in the microarray by looping around micropillars [26]. In a control experiment, no DNA strands were observed in the obstacle array when a population of dead M0-91 cells were immobilized and lysed in the microchannel.

Using the fluorospectrometer, we were unable to quantify a single cell's DNA due to the limitations of the measurement. To confirm that we were extracting close 100% of a single cell's DNA, we instead used real-time PCR. A single cell was trapped in the device and its DNA contents were extracted for off-chip analysis following the procedures outlined in Sections 2.2.4 and 2.2.7. The extracted DNA was amplified using WGA and quantified using real-time PCR as described in Section 2.2.8. The C_p value of the amplification was used to calculate the input quantity of DNA from a C_p standard curve. The amount

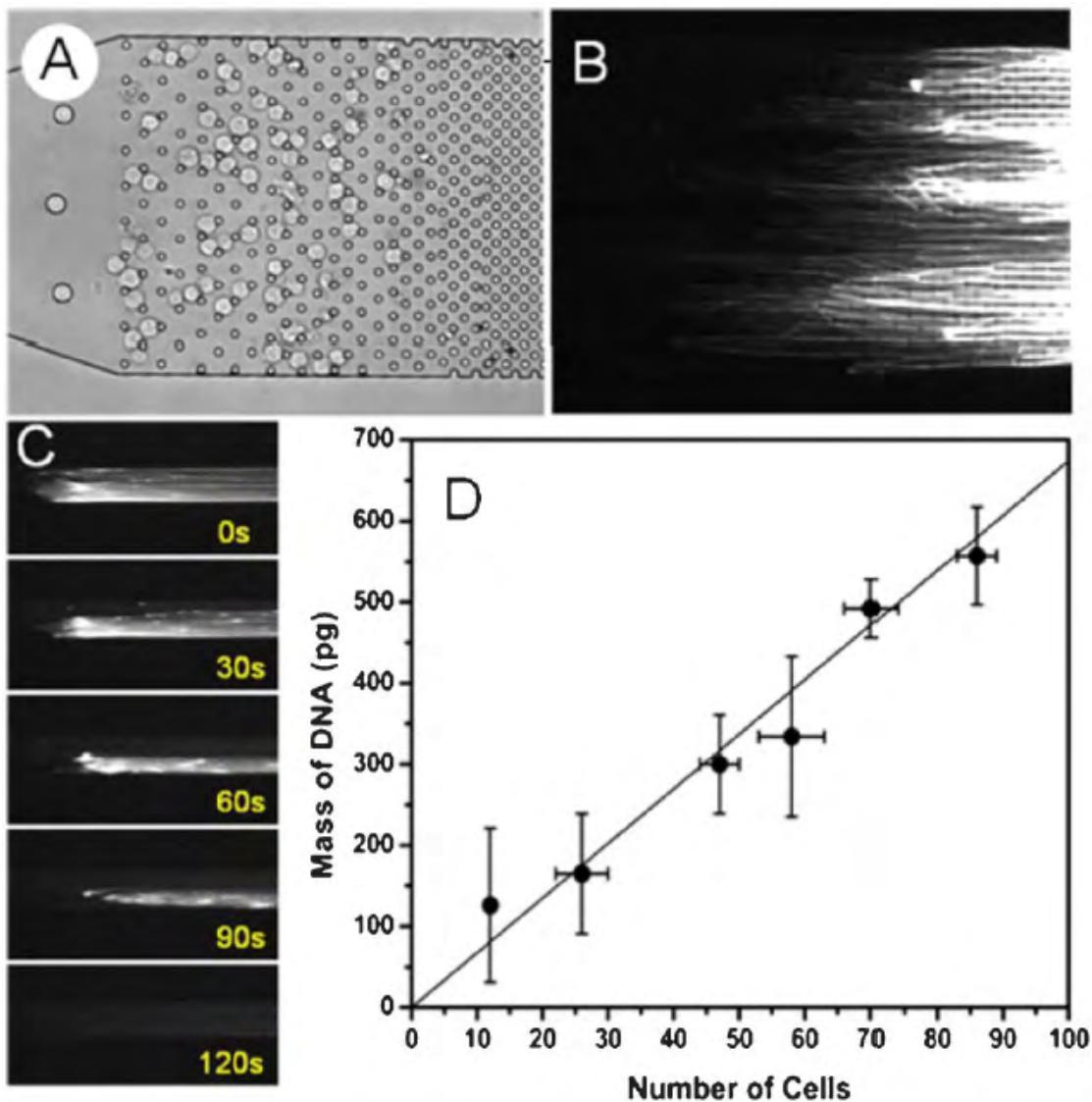


Figure 2.6: (A) Bright field image of 70 cells captured with the microdevice. (B) Fluorescent image of the entangled DNA following lysis. (C) Time sequence showing the digestion of the DNA. DNA was fully released after 2 minutes of digestion. (D) Amount of DNA extracted from different number of cells. The linear best fit of the data shown by the solid line gives a slope of 6.7 ± 0.2 pg/cell. Figure reproduced with permission from [1].

of DNA extracted from a single cell was found to be 8 ± 1 pg. Figure 2.7 shows the amplification curves and their associated C_p values which correspond to the inflection point. The slight discrepancy between the measured amount and accepted value of 6.6 pg could be due to lack of cell cycle control, differences in

amplification efficiencies, or perhaps contamination.

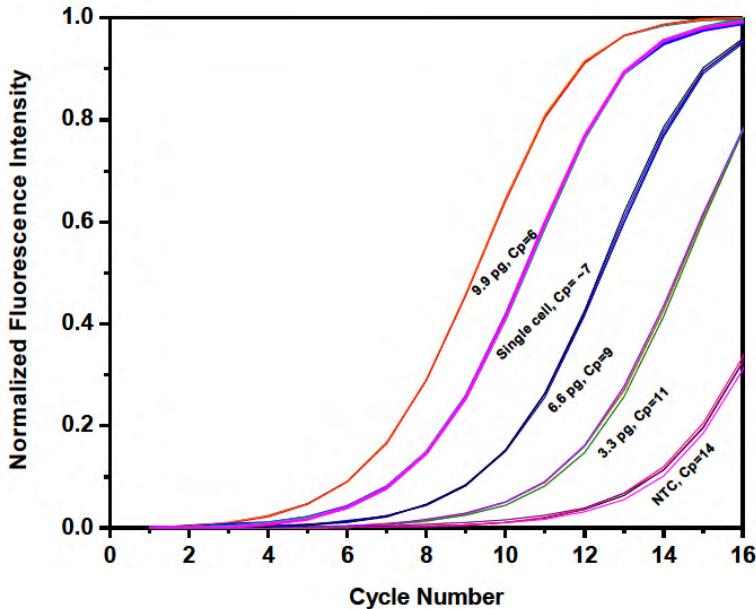


Figure 2.7: Real-time PCR amplification curves. Figure reproduced with permission from [1].

2.3.2 Microfluidic Device to Nanochannel Coupling for Single Cell Analysis and SCAN

Motivation for the micropillar device was for a preparatory tool for ensuing epigenetic analysis using single-molecule chromatin analysis in nanochannels (SCAN) first described by Cipriany et al [23]. The microfluidic device was to serve to isolate the chromatin/DNA from small cell populations, ideally single cells, and provide a purified immunostained sample for SCAN.

Initially, we conceived a device design illustrated in Figure 2.8. Devices were fabricated following the PDMS microfluidics process given in Section 2.2 and the nanochannel fabrication method given in Section 4.3.1. Unfortunately, the

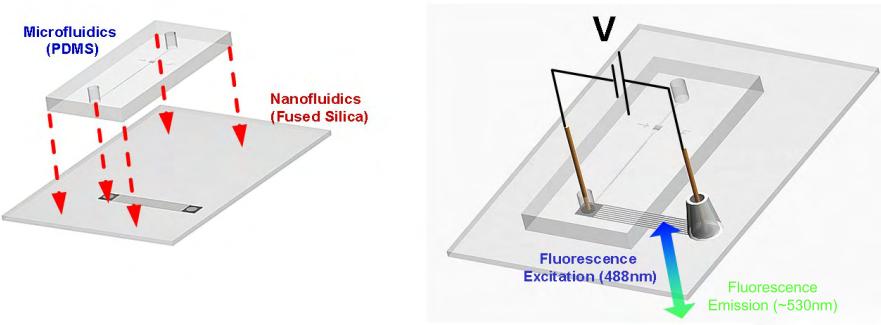


Figure 2.8: Initial device design for microfluidic/nanochannel coupling.

design suffered from many setbacks. First, the microchannels were prone to failure due to clogging, air bubbles, and delamination. Failure of the microchannel meant that the associated nanochannels became useless. With only nine devices per wafer, we were simply burning through too many devices to be sustainable with the configuration. Second, to ease the process of cell loading and buffer exchange, we operated our device by withdrawing fluid using a syringe pump. The input reservoir was open to atmosphere while the output reservoir was connected to a syringe via PEEK tubing. This created inherent problems in the coupling since the PEEK tubing would need to be removed in order to attach the SCAN electrodes. More advanced designs were made by depositing gold on the nanochannel wafer for embedded electrodes to eliminate the removal of the tubing. However, the main problem was that we were pulling the microfluidic debris (e.g. cell lysate, buffer contaminants) over the nanochannel input port and fouling the nanochannel in the process. In order to address this issue, we had two options: 1. Infuse the fluids in combination with creating a device two output channels, one for waste and one for the clean sample for SCAN, along with embedded PDMS valves for both output channels. 2. Extract DNA/chromatin into a microcentrifuge tube using procedures already developed and described in the previous pages of this manuscript and transfer it to

the SCAN device using a micropipette. We chose the second option since it conserved SCAN devices and needed limited development.

In order to process samples derived from the micropillar device with SCAN, we needed to modify the DNA extraction protocol given in Section 2.2. The protocol was altered to ensure that the resulting sample was in a buffer composed at least 95% standard SCAN buffer (1x TE, 0.5% PVP, 0.1% Triton-X 100). After lysis, the DNA was washed with 5 μ L of standard SCAN buffer to create a SCAN compatible fluid plug ahead of the DNA sample. The DNA was then digested with the restriction enzyme of choice in 0.5 μ L of the restriction enzyme buffer. Generally, the DNA was not intercalated in the channel due to reduced digestion efficiency caused by intercalation. Another plug of SCAN buffer (15 μ L) was created following the digest and was withdrawn into the PEEK tubing. The PEEK tubing was then removed from the output port of the microdevice and the fluid in the tubing, which contained the digested DNA with approximately 20 μ L of SCAN buffer, was infused into a microcentrifuge tube. The sample was then labeled using YOYO-1 and transferred to the input of the SCAN device.

To demonstrate the potential power of the micropillar/nanochannel device coupling, we aimed to detect DNA from a single cell using SCAN. We captured a single cell in a modified micropillar array as shown in Figure 2.9a and the cell was processed using the procedure described above. SCAN produced fluorescent bursts proportional to the DNA lengths. Figure 2.9b shows a 1 minute time trace of the single cell SCAN with many smaller bursts interspersed with large bursts. The burst count distribution gave an exponential distribution similar to Figure 3.12b. This was to be expected as the contour lengths for the DNA should

follow an exponential distribution assuming random cutting by the restriction enzyme. Informally, we can think of the cut sites as uniformly distributed along a large piece of DNA similar to the uniformly distributed event times in a Poisson process. The distance between cut sites is analogous to waiting times in a Poisson process which are well known to be exponentially distributed.

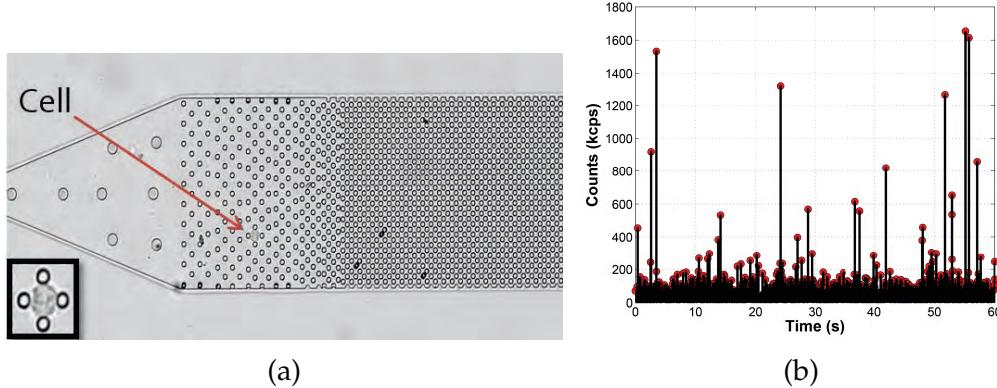


Figure 2.9: Single cell micropillar device coupled to SCAN. (a) Single cell trapped within the micropillar array. (b) SCAN time trace from the single cell sample.

2.3.3 Microfluidic Chromatin Isolation From Small Cell Populations

Conventional chromatin preps for ChIP experiments normally require approximately $10^6 - 10^7$ cells [27, 28]. The feasibility of harvesting cell counts on this order vary depending on the cell or tissue. In some cases, we simply do not have enough cells to analyze the epigenome. Some medically important cells that fall within this category include circulating tumor cells, embryonic stem cells, and various tissue samples.

With that said, we desired to modify our microfluidic system and protocol

to be able to extract, purify, and label chromatin efficiently from small cell populations with the ability to analyze their contents with downstream processes including SCAN. Figure 2.10 illustrates the process which we envisioned which was generally analogous to the previously outlined DNA technique except for the chemical treatments. Cells were to be trapped in the micropillar array under hydrodynamic flow, gently lysed so that the chromatin wouldn't dissociate, labeled with antibodies targeting epigenetic marks, crosslinked, and enzymatically digested using micrococcal nuclease for downstream analysis.

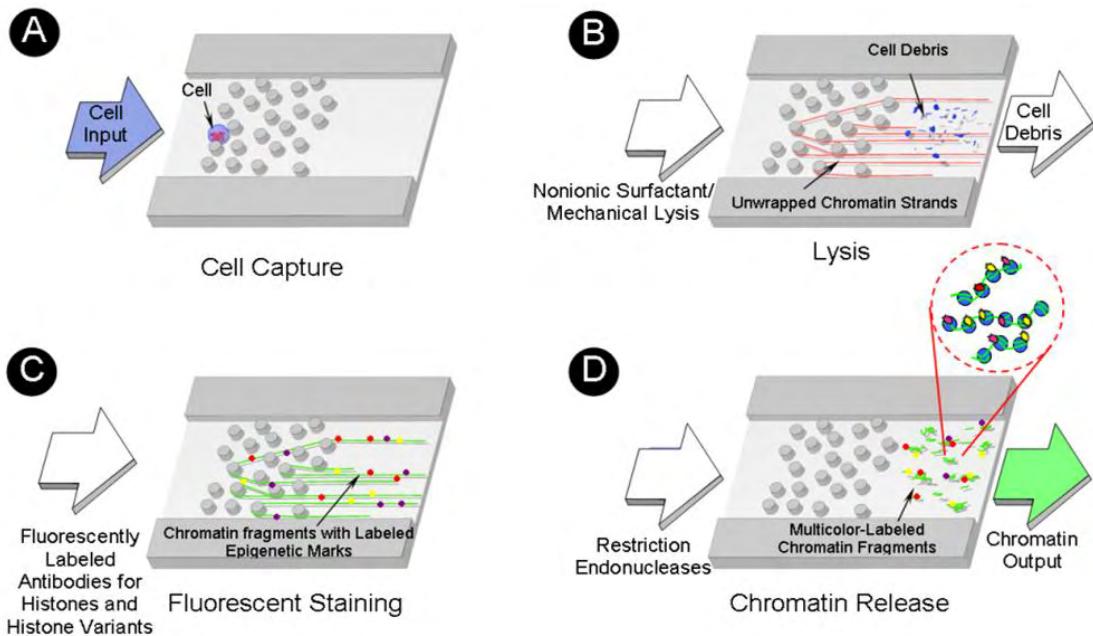


Figure 2.10: Schematic representation of chromatin purification using the microfluidic device. (A) A single cell is captured in the micropillars by size exclusion. (B) The cell is lysed using a nonionic surfactant (Triton X-100). The chromatin unwraps around the pillars and the cell debris is washed away. (C) The chromatin is intercalated and immunostained with antibodies targeting histone modifications. (D) The chromatin is released with micrococcal nuclease.

For our initial tests, we used HeLa-GFP cells which expressed GFP on the H2B histone. We used GFP as an indicator of chromatin. HeLa-GFP cells were flowed into our PDMS microdevice and entrapped within the pillars as shown

in Figure 2.11a. Figure 2.11b shows a dark field fluorescence image of the same cells with GFP emission localized within the nucleus. The HeLa-GFP cells were then chemically lysed using a 1xTE buffer with 500 mM NaCl and 1% Triton X-100, a non-ionic surfactant. The high salt stabilized the chromatin while the Triton X-100 perturbed and broke both the cell and nuclear membranes. Similar to chromosomal DNA, the chromatin became entangled in the pillar array and elongated under flow while the cell debris was washed away. The post lysis immobilized chromatin can be seen in Figure 2.11c. Following a buffer wash, the chromatin was digested using micrococcal nuclease and collected in a microcentrifuge tube. The 20 μ L collected sample was stained with TOTO-3 in the tube and transferred to an input SCAN port. Staining wasn't performed in the channel because previous results demonstrated reduced digestion efficiency caused by intercalation. The sample was then processed using 2-color SCAN. We observed a peak above background in the coincidence offset histogram as shown in Figure 2.11d. This implied that chromatin derived from the microdevice remains intact for downstream analysis and opened up the possibility of analyzing chromatin from a single cell.

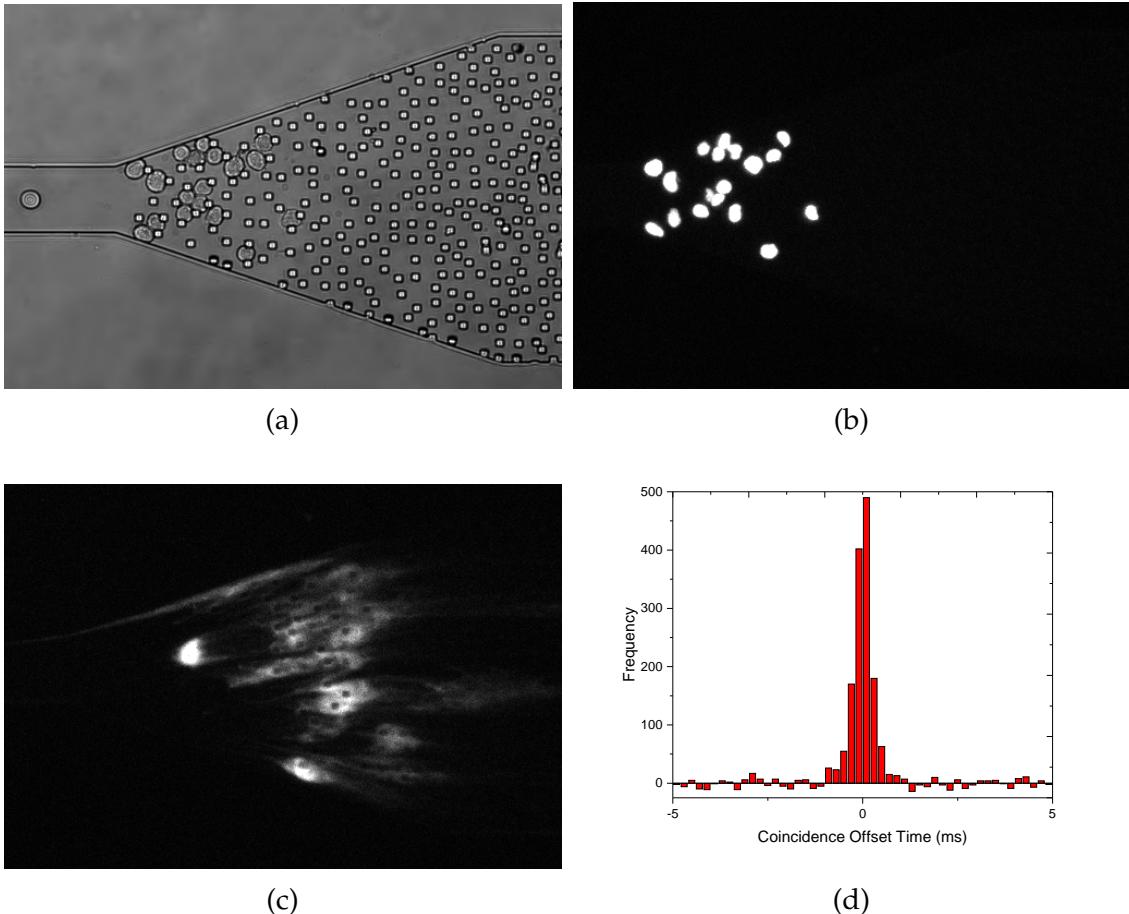


Figure 2.11: (a) Bright field micrograph showing 18 HeLa-GFP cells captured in the micropillar array. (b) Fluorescence image of 18 HeLa-GFP cells from (a). The GFP is localized within the nucleus prior to lysis. (c) Elongated chromatin after lysis under flow. (d) Coincidence offset histogram for GFP and TOTO-3. The coincidence peak signified that the chromatin from the microdevice remained intact for downstream analysis such as with SCAN.

Experiments not involving GFP did not require additional antibodies to conclude the existence of chromatin in the channel. We observed distinct morphological differences between DNA and chromatin when stretched. The DNA elongated in a highly linear manner whereas the chromatin structure was networked and cobwebbed. Figure 2.12 shows stretched DNA and chromatin contrasting the two configurations.

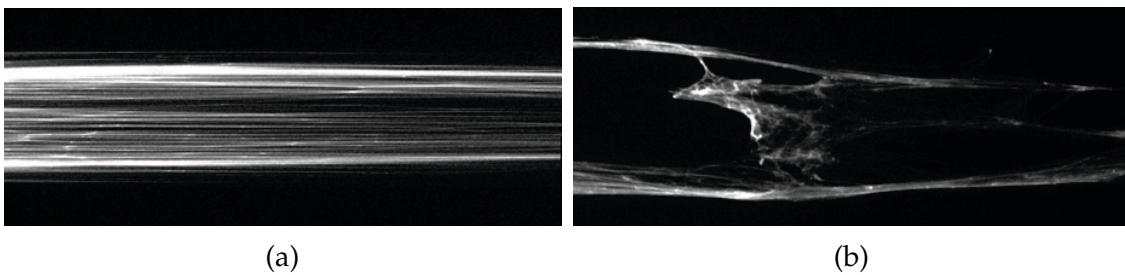
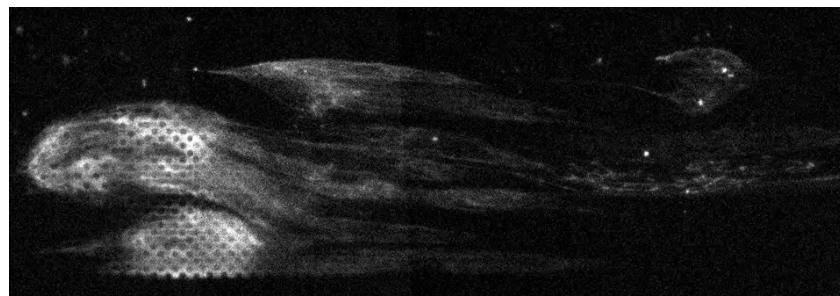


Figure 2.12: Morphological differences between elongated DNA and chromatin from M0-91 cells. Both were stained with PicoGreen. (a) DNA (b) Chromatin

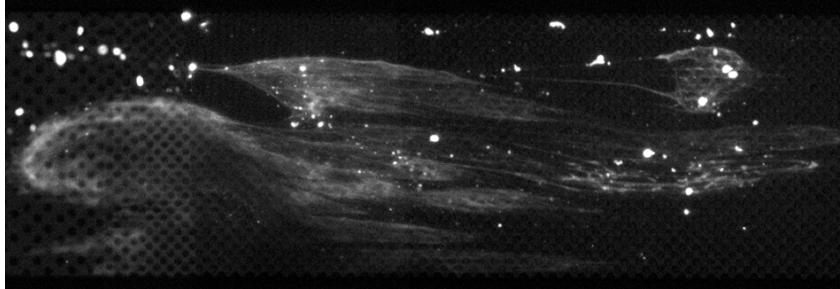
2.3.4 Antibody Labeling of Chromatin in the Micropillar Device with SCAN Coupling

The micropillar device's capability to entrap DNA and chromatin provides an excellent platform for performing binding reactions with epigenetic probes. The device has a significant advantage in that buffer exchanges can be easily achieved and excess reagents can be freely washed away without additional processing steps which may cause sample loss.

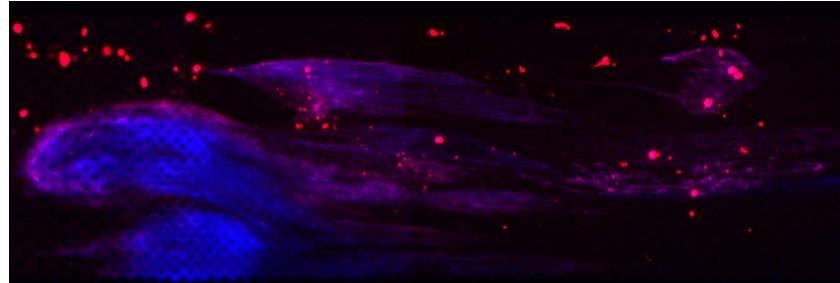
We trapped 60 cells in the micropillar array and lysed them with the 1xTE chromatin lysis buffer containing high salt and Triton X-100 described prior. We subsequently flowed Alexa Fluor 647 labeled α -H3 antibody for approximately 1 hour to immunostain the chromatin in the microchannel. Due to the densely packed nature of the chromatin, much of the flow was diverted around the chromatin. The result of this flow pattern can be seen by the relatively underlabeled areas of the chromatin in Figure 2.13c. After labeling, a 0.75% solution of formaldehyde was applied under flow to crosslink the antibodies to the chromatin. The chromatin was released by MNase digestion and analyzed with SCAN. H2B-GFP and α -H3 coincidence was observed as shown in Figure 2.13d.



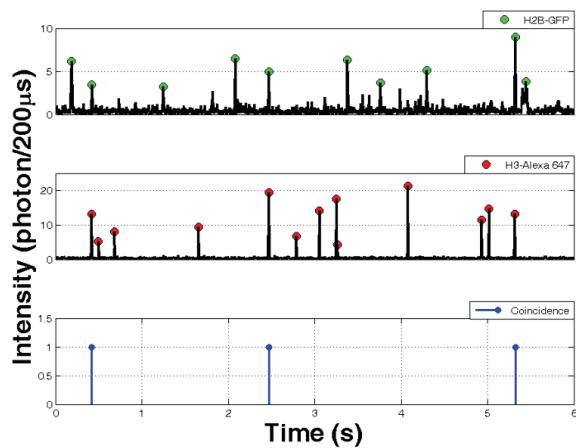
(a)



(b)



(c)



(d)

Figure 2.13: (a) Fluorescent micrograph showing H2B-GFP. (b) Fluorescence image of the Alexa Fluor 647 α -H3. (c) Image overlay with GFP (blue), Alexa Fluor 647 α -H3 (red), and their colocalization (purple). (d) Time trace of H2B-GFP and Alexa Fluor 647 α -H3 showing coincident molecules with SCAN.

As already mentioned, the underlabeling of the chromatin caused by flow diversion could be a limitation of the technique. In order to mitigate this problem, cells should be well separated from each other prior to lysis and flow rates should be minimized during immunostaining to enable sufficient time for diffusion to occur. Other Craighead members have had success with similar chromatin immunostaining experiments by following these principles.

2.4 Conclusions

The micropillar device proved to be an extremely effective method to trap, purify, and label genomic DNA and chromatin. The device relied on the physical capture and entanglement of DNA and chromatin unlike traditional methods which depend largely on surface functionalization or electrostatic interactions. The trapping method produced mostly unbroken DNA/chromatin fragments which could be easily labeled due to their predominant linear elongation. The chromatin fiber, while not completely linearized, was elongated to a degree for reasonable immunostaining. We believe that the large labeled fragments would be ideal for downstream optical mapping of epigenetic marks.

We showed that the microfluidic device was highly efficient in its DNA capture. We extracted nearly 100% of the available DNA down to a single cell. Based on the same capturing mechanism, we believe that we captured nearly 100% of chromatin, although the degree of dissociation is debatable. While we expected some degradation of chromatin in the elongation, the viability of the chromatin proved to be acceptable over time scales of a few hours as measured by GFP fluorescence and the ability to immunostain the sample.

The device has always been envisioned as a integrated preparatory device for further downstream analysis for small cell populations such as circulating tumor cells or stem cells. We were able to couple the device to SCAN and observed coincidence detection of both GFP/TOTO-3 and GFP/ α -H3. This demonstrated that chromatin from a small number of cells processed with the microfluidic device remained viable and detectable when performing downstream operations on them such as SCAN. While still challenging, by using the micropillar device in conjunction with a novel epigenetic analysis technique such as SCAN, we could one day open the door to analyzing the epigenome on a single cell level.

BIBLIOGRAPHY

- ¹J. J. Benitez, J. Topolancik, H. C. Tian, C. B. Wallin, D. R. Latulippe, K. Szeto, P. J. Murphy, B. R. Cipriany, S. L. Levy, and P. D. Soloway, "Microfluidic extraction, stretching and analysis of human chromosomal dna from single cells", *Lab on a Chip* **12**, 4848–4854 (2012).
- ²N. Navin, J. Kendall, J. Troge, P. Andrews, L. Rodgers, J. McIndoo, K. Cook, A. Stepansky, D. Levy, and D. Esposito, "Tumour evolution inferred by single-cell sequencing", *Nature* **472**, 90–94 (2011).
- ³K. Appasani and R. K. Appasani, *Stem cells and regenerative medicine: from molecular embryology to tissue engineering* (Springer Science and Business Media, 2010).
- ⁴R. N. Zare and S. Kim, "Microfluidic platforms for single-cell analysis", *Annual review of biomedical engineering* **12**, 187–201 (2010).
- ⁵A. R. Wheeler, W. R. Thronset, R. J. Whelan, A. M. Leach, R. N. Zare, Y. H. Liao, K. Farrell, I. D. Manger, and A. Daridon, "Microfluidic device for single-cell analysis", *Analytical chemistry* **75**, 3581–3586 (2003).
- ⁶A. K. White, M. VanInsberghe, I. Petriv, M. Hamidi, D. Sikorski, M. A. Marra, J. Piret, S. Aparicio, and C. L. Hansen, "High-throughput microfluidic single-cell rt-qpcr", *Proceedings of the National Academy of Sciences* **108**, 13999–14004 (2011).
- ⁷J. Clausell-Tormos, D. Lieber, J.-C. Baret, A. El-Harrak, O. J. Miller, L. Frenz, J. Blouwolff, K. J. Humphry, S. Köster, and H. Duan, "Droplet-based microfluidic platforms for the encapsulation and screening of mammalian cells and multicellular organisms", *Chemistry and biology* **15**, 427–437 (2008).

- ⁸G. Ocvirk, H. Salimi-moosavi, R. J. Szarka, E. Arriaga, P. E. Andersson, R. Smith, N. J. Dovichi, and D. J. Harrison, "Beta-galactosidase assays of single-cell lysates on a microchip: a complementary method for enzymatic analysis of single cells", *Proceedings of the Ieee* **92**, 115–125 (2004).
- ⁹Y. Marcy, T. Ishoey, R. S. Lasken, T. B. Stockwell, B. P. Walenz, A. L. Halpern, K. Y. Beeson, S. M. Goldberg, and S. R. Quake, "Nanoliter reactors improve multiple displacement amplification of genomes from single cells", *PLoS Genet* **3**, 1702–1708 (2007).
- ¹⁰M. B. Johns and J. E. Paulus-Thomas, "Purification of human genomic dna from whole blood using sodium perchlorate in place of phenol", *Analytical biochemistry* **180**, 276–278 (1989).
- ¹¹J. Kim, M. Johnson, P. Hill, and B. K. Gale, "Microfluidic sample preparation: cell lysis and nucleic acid purification", *Integrative Biology* **1**, 574–586 (2009).
- ¹²K. A. Wolfe, M. C. Breadmore, J. P. Ferrance, M. E. Power, J. F. Conroy, P. M. Norris, and J. P. Landers, "Toward a microchip-based solid-phase extraction method for isolation of nucleic acids", *Electrophoresis* **23**, 727–733 (2002).
- ¹³C. R. Reedy, C. W. Price, J. Sniegowski, J. P. Ferrance, M. Begley, and J. P. Landers, "Solid phase extraction of dna from biological samples in a post-based, high surface area poly (methyl methacrylate)(pmma) microdevice", *Lab on a Chip* **11**, 1603–1611 (2011).
- ¹⁴C. W. Price, D. C. Leslie, and J. P. Landers, "Nucleic acid extraction techniques and application to the microchip", *Lab on a Chip* **9**, 2484–2494 (2009).
- ¹⁵R. Boom, C. Sol, M. Salimans, C. Jansen, P. Wertheim-van Dillen, and J. Van der Noordaa, "Rapid and simple method for purification of nucleic acids", *Journal of clinical microbiology* **28**, 495–503 (1990).

- ¹⁶M. A. Gijs, "Magnetic bead handling on-chip: new opportunities for analytical applications", *Microfluidics and Nanofluidics* **1**, 22–40 (2004).
- ¹⁷P.-G. de Gennes, "Reptation of a polymer chain in the presence of fixed obstacles", *The journal of chemical physics* **55**, 572 (1971).
- ¹⁸N. P. Teclemariam, V. A. Beck, E. S. Shaqfeh, and S. J. Muller, "Dynamics of dna polymers in post arrays: comparison of single molecule experiments and simulations", *Macromolecules* **40**, 3848–3859 (2007).
- ¹⁹W. Volkmuth and R. Austin, "Dna electrophoresis in microlithographic arrays", *Nature* **358**, 600–602 (1992).
- ²⁰S. Turner, A. Perez, A. Lopez, and H. Craighead, "Monolithic nanofluid sieving structures for dna manipulation", *Journal of Vacuum Science and Technology B* **16**, 3835–3840 (1998).
- ²¹H. H. Heng, B. Spyropoulos, and P. B. Moens, "Fish technology in chromosome and genome research", *Bioessays* **19**, 75–84 (1997).
- ²²M. Foquet, J. Korlach, W. Zipfel, W. W. Webb, and H. G. Craighead, "Dna fragment sizing by single molecule detection in submicrometer-sized closed fluidic channels", *Analytical Chemistry* **74**, 1415–1422 (2002).
- ²³B. R. Cipriany, R. Zhao, P. J. Murphy, S. L. Levy, C. P. Tan, H. G. Craighead, and P. D. Soloway, "Single molecule epigenetic analysis in a nanofluidic channel", *Analytical chemistry* **82**, 2480–2487 (2010).
- ²⁴P. J. Murphy, B. R. Cipriany, C. B. Wallin, C. Y. Ju, K. Szeto, J. A. Haggarman, J. J. Benitez, H. G. Craighead, and P. D. Soloway, "Single-molecule analysis of combinatorial epigenomic states in normal and tumor cells", *Proceedings of the National Academy of Sciences* **110**, 7772–7777 (2013).

- ²⁵B. Alberts, D. Bray, K. Hopkin, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, *Essential cell biology* (Garland Science, 2013).
- ²⁶J. H. Zhang and X. Ming, "Dna fragmentation in apoptosis", *Cell research* **10**, 205–211 (2000).
- ²⁷T. I. Lee, S. E. Johnstone, and R. A. Young, "Chromatin immunoprecipitation and microarray-based analysis of protein location", *Nature protocols* **1**, 729–748 (2006).
- ²⁸B. E. Bernstein, T. S. Mikkelsen, X. Xie, M. Kamal, D. J. Huebert, J. Cuff, B. Fry, A. Meissner, M. Wernig, and K. Plath, "A bivalent chromatin structure marks key developmental genes in embryonic stem cells", *Cell* **125**, 315–326 (2006).

CHAPTER 3

SINGLE MOLECULE EPIGENETIC MARK DETECTION AND QUANTIFICATION IN NANOFUIDIC CHANNELS

3.1 Introduction

Epigenetic regulation is a critical biological process for the health and development of a cell. Epigenetic regulation is facilitated by covalent modifications to the underlying DNA and chromatin proteins. A fundamental understanding of these epigenetic modifications and their associated interactions at the molecular scale is necessary to explain phenomena including cellular identity, stem cell plasticity, and neoplastic transformation. In fact, it is widely known that abnormal epigenetic profiles have been linked to many diseases including psychiatric disorders, Alzheimer's, and most notably cancer [1–5]. While the field of epigenetics has progressed rapidly with conventional techniques, significant advances remain to be made with respect to combinatoric analysis of epigenetic marks and single cell epigenetics.

Limitations to Current Epigenomic Techniques

Previously, we gave a thorough discussion in Chapter 1 on the limitations of current epigenomic techniques. To briefly recapitulate, the two standard epigenetic profiling methods are bisulfite sequencing (BS-seq), which determines the 5mC content of a sample, and chromatin immunoprecipitation followed by sequencing (ChIP-seq), which identifies the genomic locations of the histone modifications and chromatin binding proteins [6]. While these techniques afford the

highest resolution possible, single base pair for BS-seq and single nucleosome for ChIP-seq, they are limited in their ability to analyze small amounts of material or process combinations of epigenetic marks beyond two. Therefore, the aim of SCAN was to have a technology to address the limitations of traditional methods. Specifically, SCAN was designed to

1. Probe combinations of epigenetic marks from biologically relevant samples.
2. Analyze materials down to the single cell regime.

In the remaining portion of this chapter, we describe the development of quantitative SCAN, a methodology to quantitatively assess the genome-wide level of epigenetic marks.

3.1.1 SCAN Overview

Single-molecule Chromatin Analysis at the Nanoscale (SCAN), a single-molecule technique to probe epigenetic marks, was first described by Cipriany et al [7]. SCAN uses laser-induced fluorescence to examine individual molecules flowing through a nanochannel. Through the nanochannel confinement, individual chromatin states can be determined as background is reduced and single molecule occupancy within the observation volume can be obtained. SCAN is a high-throughput method whose throughput is dependent on signal to noise (SNR) considerations and maintaining single molecule occupancy of the inspection volume. Typical throughputs are on the order of 10^3 molecules/minute.

Previously, nanofluidic platforms have been used to size femtogram quantities of DNA, quantum dot-organic dye conjugates, PCR amplification products and more [8–12]. Specifically regarding chromatin, however, Cipriany et al was the first to demonstrate that chromatin was capable of flowing through the nanochannels with concentrations less than 1 nM. They did this using HeLa-GFP chromatin to show coincidence of TOTO-3 intercalator and GFP which was expressed on the H2B histone. They extended their analysis to methylated DNA bound with methyl binding domain 1 (MBD1). By varying the ratio of methylated DNA to total DNA, they showed that coincidence varied linearly as the ratio increased. This experiment, however, showed the linearity of the SCAN detection with regards to ratios of reagents and did not demonstrate that they could quantify epigenetic marks. Nonetheless, it was with these results that we proceeded with optimism to try quantify multiple epigenetic marks on a genome-wide basis which is the subject of the remaining portion of this chapter.

3.2 Materials and Methods

3.2.1 Device Fabrication

The nanofluidic channels were fabricated on 4" UV fused silica wafers (Mark Optics). The GCA Autostep 200 stepper was used to pattern the nanochannels with a width of 500 nm. The channels were etched using reactive ion etching (Oxford 80) to a depth of approximately 250 nm. Prior to drilling access holes, the wafer was protected by spinning S1818 photoresist (Shipley) to a thickness of 2 μm on each side. The holes were then drilled using a focused jet of alu-

mina particles. The wafer was cleaned using 1165 Microposit Remover (Shipley) and hot Piranha solution ($3\text{ H}_2\text{SO}_4:\text{1 H}_2\text{O}_2$). The wafer was further cleaned using the standard RCA clean [13] and then touch bounded to a $170\text{ }\mu\text{m}$ cover-slip wafer. The wafers were permanently bonded by annealing at $1050\text{ }^\circ\text{C}$ for 6 hours. Ports were bonded onto the wafer using optical-grade epoxy (Norland Products). Further information on device fabrication can be found in Section 4.3.1.

3.2.2 Optical Setup

A fused silica wafer containing the nanofluidic channels was placed on an inverted microscope (IX-71, Olympus) with a laser side port for excitation. For 2 color experiments, a 488 nm laser (Sapphire, Coherent) was operated at $500\text{ }\mu\text{W}$ and a 639 nm laser (Cube, Coherent) was operated at 1 mW. For 3 color experiments, an additional 405 nm laser (Cube, Coherent) was digitally modulated at 10 kHz and outputted an average power of $300\text{ }\mu\text{W}$. The lasers were overlapped using free-space optics and reflected into a $60\times$, 1.2 NA water immersion objective (UPlanSAPO, Olympus) and focused on the nanochannels. Confocal filtering was performed using a $50\text{ }\mu\text{m}$ diameter pin-hole located 102 mm outside the left side port of the microscope. Fluorescence signals were then split chromatically using dichroic filters (Semrock) and bandpass filtered using emission filters (Semrock FF01-452/45, FF02-525/40, FF01-680/42). Fluorescence was gathered using $100\text{ }\mu\text{m}$ diameter core multimode optical fiber (OZ Optics). Photons were detected using a single-photon avalanche photodiode (APD) (SPCM-AQRH-14, Perkin Elmer). 2 color signals from the APDs were recorded with a digital correlator (Flex02-01D, Correlator.com) operating at a

sampling frequency of 100 kHz. 3 color signals were recorded with a multiple-event time digitizer (P7888, ComTec) operating at 20 kHz. The 405 nm laser and the digitizer operated in synchrony with the same clock source.

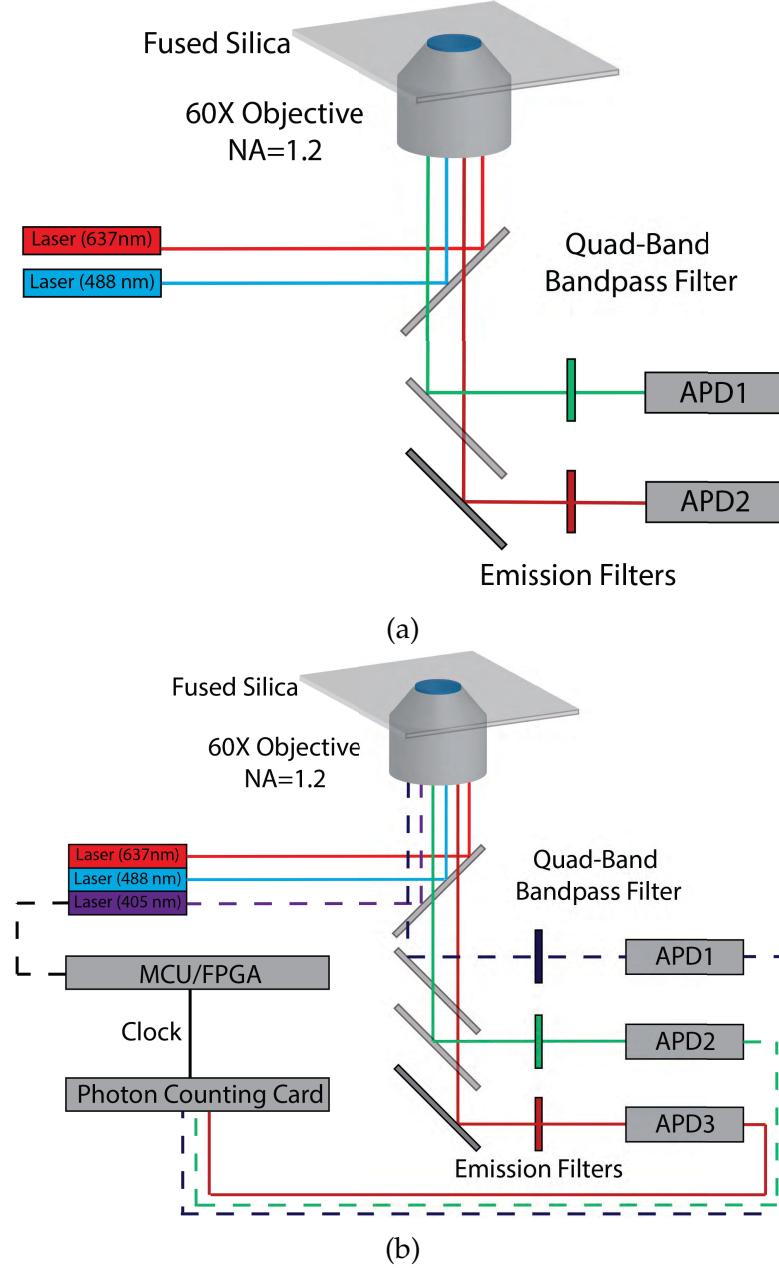


Figure 3.1: (A) Two color optical setup. (B) Three color optical setup. A confocal aperture (not shown) was placed between the quad-band bandpass filter and the 488 nm laser line emission filter. Dotted lines represent pulsed signals for pulsed interleaved excitation. A confocal aperture (not shown) was placed between the quad-band bandpass filter and the 405 nm laser line emission filter.

3.2.3 Chromatin Extraction

Chromatin extractions were performed by members of the Soloway lab. HeLa cells expressing green fluorescent protein on histone H2B were provided by Geoffrey M. Wahl of The Salk Institute of Biological Studies, USA.

Cells were cultured in Dulbecco's Modified Eagle's Medium with 5% fetal calf serum. Cells were removed from two 15 cm plates by scraping once a density of 1×10^6 cells/mL was reached. They were subsequently washed with 1xRSB (10 mM Tris pH 7.6, 15 mM NaCl, 1.5 mM MgCl₂), and resuspended in 5 mL 1xRSB buffer with 1% Triton-X 100. Homogenization followed and was performed using a dounce homogenizer with a loose pestle. Nuclei were recovered via 4000g centrifugation and resuspended in 1.5 mL of Buffer A (15 mM Tris pH 7.6, 15 mM NaCl, 60 mM KCL, 0.34 M sucrose, 0.5 mM spermidine, 0.15 mM spermine, 0.25 mM PMSF, 0.1% M CaCl₂ and 0.1% β -mercaptoethanol). Micrococcal nuclease (MNase) was added at a concentration of 266 gel-units/mL. 250 μ L aliquots of sample solution were removed at different time intervals ranging from 0 to 90 minutes depending on the application, followed by halting the micrococcal digest reaction by adding 5 μ L of 0.5 M EDTA. The resulting digests were centrifuged at 10,000g. The pellet was resuspended in 450 μ L 10 mM EDTA, 50 μ L, 0.5 M NaCl which solubilized the chromatin. DNA was then extracted from the solution for analysis of chromatin size and to confirm chromatin integrity which would be visible by a banding pattern in the ensuing gel. 60 μ L of chromatin was dissociated by adding 24 μ L of water, 6 μ L 10% SDS and 24 μ L 0.5 M NaCl. From here, the chromatin derived DNA samples were phenol-chloroform extracted with a 1:1 mixture of phenol-chloroform and run on a 1.5% agarose gel. Concentrations were determined by absorption measure-

ment using a spectrophotometer (Nanodrop 1000).

3.2.4 Sucrose Gradient Purification of Chromatin

Chromatin sucrose gradient purifications were performed by members of the Soloway lab using a 5-20% (w/v) linear sucrose gradient with 10 mM Tris-HCL pH 6.85 and 5 mM EDTA.

Briefly, a 5-20% (w/v) linear sucrose gradient was prepared by first adding 5% sucrose solution to the bottom of a centrifuge tube, followed by a 20% solution to layer, both having 10 mM Tris-HCL pH 6.85 and 5 mM EDTA as their buffer. A linear gradient was constructed using the Gradient Master (BioComp) gradient forming instrument following standard protocol given by the manufacturer. After formation of the gradient, 600 μ L was removed from the top of the solution in the tube to provide adequate volume for the sample to be added. 600 μ L of 1 mg/mL chromatin was loaded to the top of the tube from the side as to not disturb the gradient. The sample was centrifuged for 18 hours at 30,000 rpm (160,000 $\times g$) at 4°C using a SW 40 rotor (Beckman Coulter) to achieve differential centrifugation of the chromatin.

Following centrifugation, a hole was punctured in the bottom of the sample tube via a needle which connected to a pump. The sample was pumped out of the bottom of the tube at 1.5 mL/min through an absorption measurement apparatus, subsequently fractionated, and collected into microcentrifuge tubes.

3.2.5 Methyl Binding Domain Synthesis and Labeling

Monomeric MBD1 purification and labeling were performed by members of the Soloway lab. MBD1 was cloned into the pET-30b plasmid provided by Adrian P. Bird at The Wellcome Trust Centre for Cell Biology, University of Edinburgh, UK.

In brief, IPTG induced BL21(DE3) cultures were lysed and the His-tagged MBD1 was purified via affinity chromatography using Ni-NTA agarose (Qiagen), denaturation, and on-column renaturation cycles to refold the MBD1. Once refolded, the MBD was eluted off the column via a 5 mL PBS with 250 mM imidazole. Further dialysis processing produced a MBD1 protein free of imidazole in PBS solely. Labeling was achieved using the Microscale Protein Labeling Kit (Thermo Fisher A30006, A30009).

3.2.6 Antibody labeling

Antibodies unless otherwise noted were purchased from Active Motif. Labeling was achieved by using antibody labeling kits for 100 μ g of antibody. (Thermo Fisher A-20181, A-20184, A-20186) and following the standard protocol provided. Degree of labeling was determined by absorption.

3.2.7 Binding Reactions

Prior to performing the binding reaction, each constituent was blocked with 100 μ g/mL BSA for 1 hour at 4 °C in PBS and 0.1% Triton X-100. Binding reactions

typically had 500 μM of each antibody and/or MBD1. 500 μM was identified as the binding concentration of choice due to nonspecific binding at higher concentrations. However, the lack of a large molar excess of antibody likely cause some degree of precipitin complex formation. Chromatin was bound at a concentration of 5-30 ng/ μL depending on the experiment and the average length of chromatin used. Following an overnight binding reaction at 4 °C, the complexes were crosslinked using 0.75% formaldehyde for 15 minutes. The crosslinking reaction was quenched using glycine at 200 mM and diluted into the standard SCAN buffer (1xTE, 0.5% PVP, and 0.1% Triton X-100)

3.2.8 Size Exclusion Chromatography

The size-exclusion chromatography (SEC) methods were principally developed by Jim Hagerman of the Soloway Lab, but performed by both Jim and myself.

Size exclusion chromatography was performed using a Kontes Flex gravity column (Kimble Chase, 420401-0715) packed with Sephadryl S-300 (GE, 17-0599-10) with 1xTE as the running buffer. After binding, crosslinking, and quenching following the binding reaction protocol given in Section 3.2.7, the sample was injected into the top of the column and run at 4 °C under the influence of gravity. Fractions were collected into microwell plates and analyzed on a plate reader. After the desired fraction was determined, it was diluted to approximately 200 pM with standard SCAN buffer and analyzed using SCAN.

3.2.9 Nanochannel SCAN Preparation

Prior to running SCAN, the wafer containing the nanochannels was mounted on the microscope and filled with standard SCAN running buffer. The standard SCAN running buffer composed of 1xTE (10 mM Tris and 1 mM EDTA), 0.5% polyvinylpyrrolidone (PVP), and 0.1% Triton X-100 was loaded into the input reservoir only. The solution was pulled into the nanochannels via capillary forces. The device became sufficiently wetted after approximately 5 minutes. After visually ensuring no air bubbles were within the nanochannel, the output reservoir was also loaded. Both reservoirs were then washed with SCAN buffer multiple times to remove any contaminants from the reservoirs. After washing, samples at a concentration of approximately 250-500 pM were loaded into the input port and clean gold-plated electrodes were placed in both reservoirs. Electrokinetic flow was established by applying voltage to the electrodes. After approximately 5-20 minutes, the samples reached the inspection volume and SCAN data was taken using the multiple-event time digitizer or digital correlator described in Section 3.2.2.

3.3 Results and Discussion

3.3.1 2-Color Epigenetic Mark Detection

Earlier results from Cipriany et al [7] showed that SCAN could be utilized to detect individual chromatin molecules and DNA-MBD1 complexes in a rapid manner. Following this development, efforts were put forth to demonstrate 2-

color coincidence detection using chromatin and antibodies specific for relevant epigenetic marks. The following 2-color epigenetic mark experiments were led by Patrick Murphy from the Soloway Lab with myself contributing to device fabrication, optical instrumentation setup, and software/algorithm development for coincidence detection. Further information on the 2-color experiments can be found in [14].

First, as the most basic experiment applicable to the task at hand, we analyzed using SCAN a mixture of HeLa-GFP chromatin and a histone H3 specific antibody labeled with Alexa Fluor 647. Results from this experiment gave a peak above background centered around zero in the coincidence offset histogram proving that antibody-chromatin complexes can be detected using SCAN as shown in Figure 3.2. Surprisingly, the coincidence rate between α -H3 and H2B-GFP was only approximately 5% of the total GFP (chromatin) observed. This low coincidence rate was the first indication that there was something fundamentally wrong with SCAN’s detection, binding chemistry, and/or samples. While these low coincidence rates were typically attributed to chromatin degradation early on, this α -H3 and H2B-GFP experiment only probes intact chromatin, which suggests that chromatin degradation was not the main determinant of low coincidence. Also, we should note that bare H2B-GFP absorbs to the nanochannel surface due to its excessive positive charge so that the surplus of GFP signal can not be attributed to free flowing H2B-GFP.

While using HeLa-GFP as a chromatin source is tremendously useful for troubleshooting and validating certain aspects of SCAN, it is, in fact, a contrived system which would not available when we processed real-world samples. Therefore, we next prepared wild type ES cell chromatin labeled with

YOYO-1 intercalating dye and mixed it with α -H3 as a real-world analog to our previous HeLa-GFP system. Again, we observed YOYO-1/ α -H3 coincidence and tested for non-specific binding using a pre-immune mouse serum as shown in Figures 3.2 C,D. Additional testing of binding and non-specific binding using α -H3 and pre-immune mouse serum found that the crosslinked binding efficiency plateaued around a formaldehyde concentration of 0.75% and a 15 minute incubation time with lower concentrations and/or reaction times having a detrimental effect.

Further experiments were performed to show specificity with regards to antibody recognizing H3K27me3 and mC recognizing MBD1. For H3K27me3, two different chromatin sources were used: wild type ES cells and Eed -/- ES cells with negligible H3K27me3 modifications due to a mutation to a subunit of PRC2, the complex responsible for H3K27 methylation. Preferential binding occurred for α -H3K27me3 with regards to the wild type ES cell chromatin relative to Eed -/- ES cell chromatin as illustrated in Figures 3.2 E,F. Lastly, Figures 3.2 G,H validates via SCAN MBD1's specificity towards mC using wild type ES cell chromatin and DNMT TKO chromatin which has minimal, if any, mC. During the course of the MBD1 experiments and by varying the binding concentrations, it was observed that small amounts of non-specific binding did occur for MBD1 reaction concentrations greater than 625 nM which suggested limiting binding concentrations. However, as monomeric MBD1 has a poor binding affinity with a dissociation constant of $K_d = 30 \mu\text{M}$ [15], this would have an adverse effect on the true positive detection rate (MBD1/Chromatin complexes). This, perhaps, shows the potential for the inherent physical limitations of the probe causing uncertainty in absolute epigenetic mark quantification.

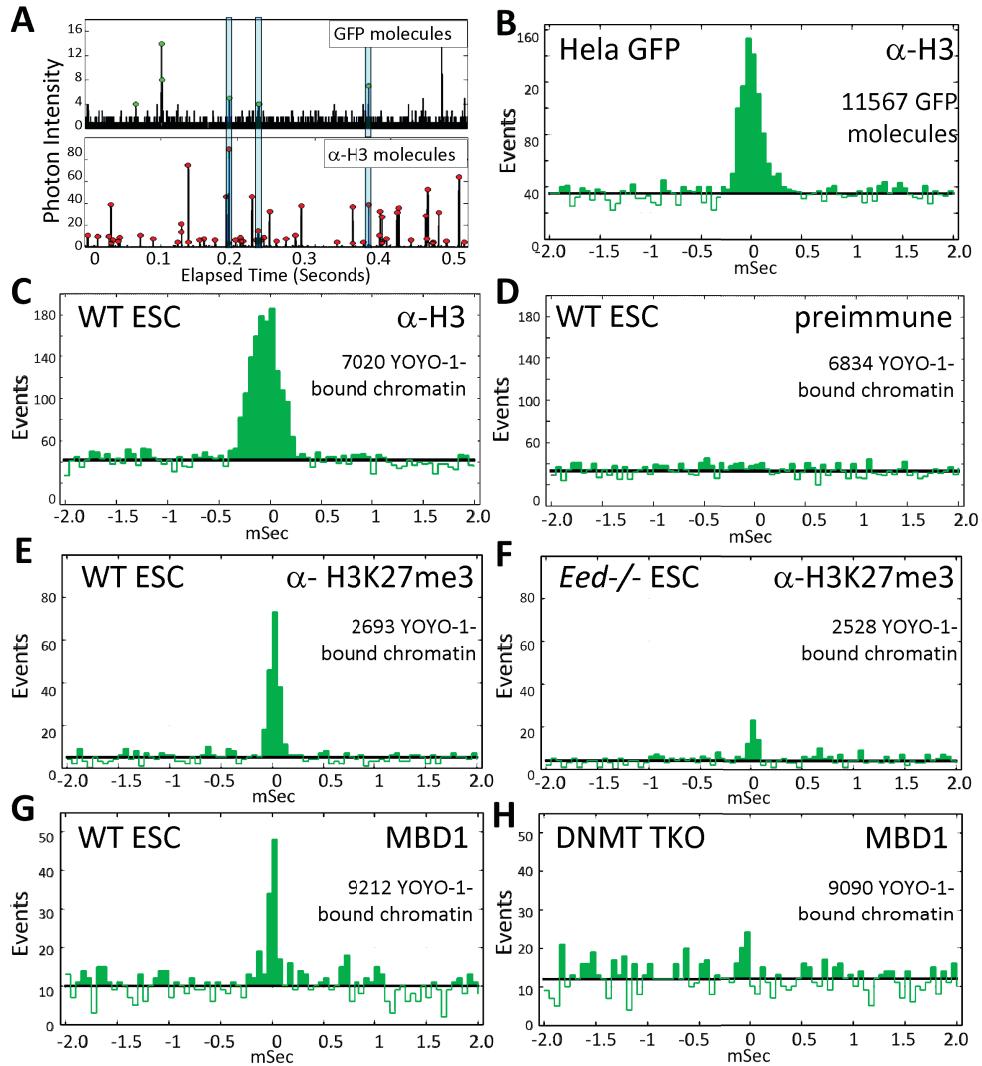


Figure 3.2: Epigenetic mark specificity using SCAN. (A) 2-channel time trace showing fluorescent emission from GFP and α -H3 molecules. Coincident events signifying antibody-chromatin complexes are highlighted with the blue shading. (B) Coincidence delay histogram illustrating simultaneous detection of GFP and α -H3 antibody. (C+D) YOYO-1 labeled wild type ES cell chromatin displaying (C) coincidence with α -H3 and (D) no coincidence with pre-immune mouse serum serving as a negative control. (E+F) α -H3 (E) coincidence with YOYO-1 labeled wild type ES cell chromatin and (F) reduced coincidence with H3K27me3 depleted $Eed^{-/-}$ ES cell chromatin labeled with YOYO-1. (G+H) MBD1, specific for mC, exhibiting (G) coincidence with YOYO-1 labeled wild type ES cell chromatin and (H) no coincidence with mC deficient DNMT TKO ES cell chromatin labeled with YOYO-1. Reproduced with permission from [14].

As noted prior, a considerable advantage of SCAN relative to other techniques is its capability to detect combinations of epigenetic marks simultaneously. To demonstrate this potential, we first bound two histone targeting antibodies (α -H2B and α -H3) with distinct labels to the wild type ES cell chromatin and processed it with SCAN. Figure 3.3A shows the coincidence observed which was less than 1% of the total number of α -H3 antibodies detected. While this low coincidence rate was at least partially due to the molar excess of antibody used in the binding, it does highlight a major issue with SCAN in that most of the detected molecules especially regarding antibodies are uninformative and unbound which ultimately limits throughput. Moreover, if we are achieving less than 1% coincidence with α -H2B and α -H3, we should expect far less coincidence for antibodies targeting histone modifications. As a negative control against α -H2B and α -H3 antibody aggregation, we performed the same experiment minus the chromatin as shown in Figure 3.3B. Further experiments were performed to show coincidence of histone modifications (α -H3K27me3) with α -H2B (Figure 3.3C) and also coincidence of histone modifications with DNA methylation (Figure 3.3D).

While the discussed two color experiments showed early promise for SCAN itself, they also highlighted certain aspects of SCAN that would continually hinder its capabilities such as meagerly low coincidence rates.

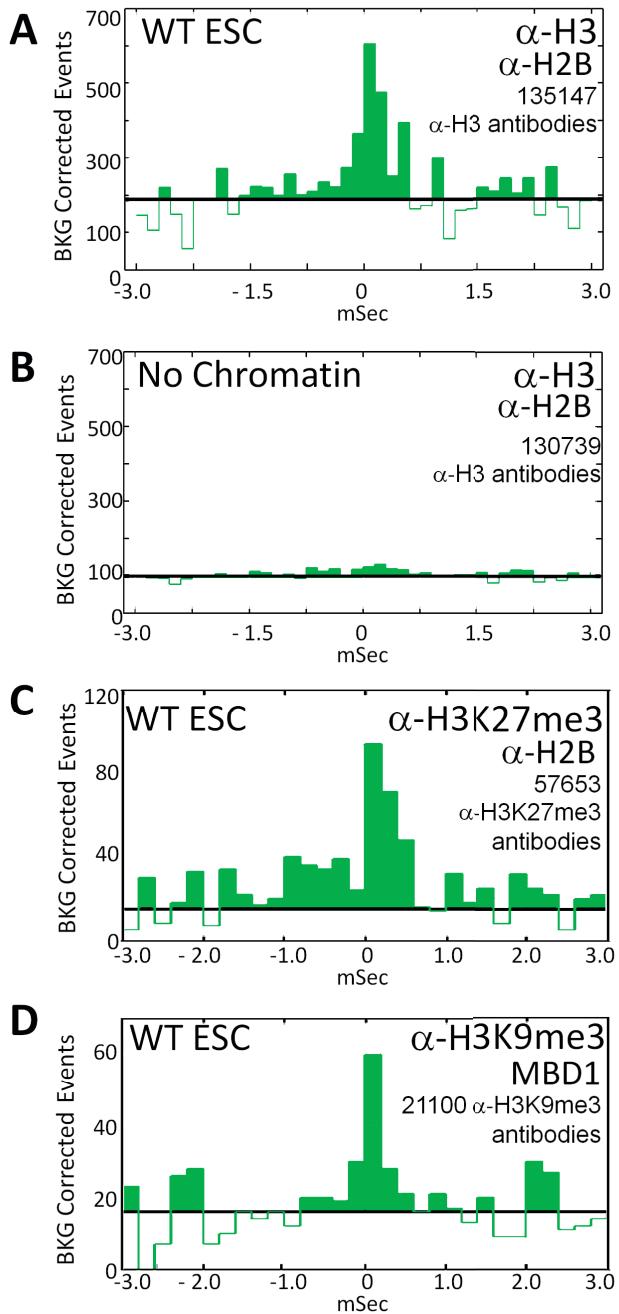


Figure 3.3: Epigenetic coincident mark detection using SCAN. (A) Chromatin bound $\alpha\text{-H3}$ and $\alpha\text{-H2B}$ coincidence. (B) $\alpha\text{-H3}$ and $\alpha\text{-H2B}$ coincidence for a negative control with no chromatin. (C) Histone modification coincidence detection using $\alpha\text{-H3K27me3}$ and $\alpha\text{-H2B}$. (D) Histone modification and methylation coincidence detection using $\alpha\text{-H3K9me3}$ and MBD1. Reproduced with permission from [14].

3.3.2 3-Color Epigenetic Mark Detection

Previous 2-color results implemented a standard 2-color confocal optical setup. The dyes being excited by the 488 nm laser and the 637 nm were adequately separated spectrally so that crosstalk was essentially not detectable for the desired molecules; large aggregates of undesired material created crosstalk which could be removed via post processing algorithms. The addition of a third spectrally distinct label was needed for two reasons:

1. Process 3 epigenetic marks simultaneously.
2. Reduce the number of 2-color experiments.

It is quite apparent why quantifying 3 epigenetic marks simultaneously would be appealing as standard ChIP protocols can not address more than 2 epigenetic marks. Reducing the number of experiments and increasing the accuracy of 2-color results was another reason for the development of the 3-color system as the 3-color system would allow for the number of experiments to be reduced by half. For a 2-color relative quantification of epigenetic marks, four independent experiments typically had to be performed sequentially. First, chromatin for which DNA was marked by intercalating dye and a histone recognizing antibody was analyzed using SCAN to determine the relative amount of chromatin to free DNA in the test sample. This was followed by a SCAN experiment analyzing the epigenetic mark of interest and the chromatin/DNA. This was then repeated using a control sample which we believed would control for variables including crosslinking effectiveness and efficacy of antibody binding. We previously quantified this operation as the normalized abundance

of epigenetic marks given by

$$NAE = \frac{\left[\left(\frac{B_e}{I} \right) / \left(\frac{B_2}{I} \right) \right]_T}{\left[\left(\frac{B_2}{I} \right) / \left(\frac{B_E}{I} \right) \right]_C} \quad (3.1)$$

where B_e = count of molecules bound to epigenetic probe and intercalator, I = count of intercalator bound molecules for the given experiment, B_2 = count of molecules bound to both intercalator and histone recognizing antibody such as α -H2B or α -H3, T = test cells, and C = control cells.

Choice of dyes used in the 3-color system was not arbitrary due to the constraints of the optical setup, namely the excitation lasers and emission filters. Additionally, familiarity with YOYO-1 and Alexa Fluor 647 implied that we were not going to change those parameters when switching to the 3-color system. Therefore, we were left with the choice of the third dye being either Alexa Fluor 405 or Alexa Fluor 568, both of which would have crosstalk into adjacent emission channels.

Generally, fluorescence crosstalk or bleed-through, is addressed through a process called linear spectral unmixing [16, 17]. The objective for linear spectral unmixing is to determine the relative proportion of the signals coming from each individual fluorophore and hence, reduce or eliminate crosstalk. This is done by deconvolving the intensity information from the fluorescence using reference spectra and a lambda stack. A lambda stack is generated by imaging over narrowbands across the spectrum giving spectrophotometry measurements for each individual detector (e.g. pixel, PMT, APD). The procedure can be summarized mathematically as

$$I(\lambda) = \sum_i c_i R_i(\lambda) \quad (3.2)$$

where i represents a dye, c_i is the relative concentration the dye, and $R_i(\lambda)$ is the reference spectrum of the i th dye. By inverting Equation (3.2), we can obtain the relative concentrations for each of dyes for each individual detector. In flow cytometry which has similarities to SCAN to some degree, lambda stacks are typically generated using a diffraction grating coupled to a CCD camera due to their small region of interest.

However, for low signal counts or signals with low SNR, linear unmixing is known to be unsatisfactory as significant errors can occur due to relative high count variability caused by counting statistics. Linear spectral unmixing is a spectrum averaged deconvolution scheme which in no way addresses count fluctuations. Ideally, fluctuations as a relative proportion of the mean should decrease as $N^{-1/2}$ due to Poisson statistics, where N is the number of counts.

Nevertheless, to determine the suitability linear unmixing would be for SCAN, we performed an unmixing of Alexa Fluor 405 into a 525/40 emission channel (data not shown). This unmixing was rather crude in the sense that the spectral bands for the unmixing were the relatively wideband bandpass emission filters. Prior to unmixing, a comparable number of false coincident molecules to total molecules were observed due to crosstalk. After unmixing, the coincidence rate dropped to approximately 25% of that observed before unmixing, which is still much too high for accurate SCAN coincidence analysis. Moreover, due to the inhomogeneity of a chromatin sample, we should expect greater sample-to-sample variability in false coincidence caused by the averaging nature of the method relative to homogeneous Alexa Fluor test sample.

Due to the infeasibility of the linear unmixing scheme for SCAN, we instead concentrated our efforts towards a time-division multiplexing strategy. Pulsed

laser techniques for fluorescence analysis have been used previously, most notably for single species Foerster Resonance Energy Transfer (spFRET) [18–20]. Termed pulsed interleaved excitation (PIE), alternating laser pulses are interleaved on a nanosecond time scale and synchronized to a time correlated single photon counting (TCSPC) device. Quasi-simultaneous recording of the temporal behavior of the sample molecule for each laser is achieved since the pulsing and TCSPC sample frequencies are both much higher than the underlying physical dynamics. Fluorescence crosstalk is effectively eliminated with this method as the time-division creates, for all practical purposes, two independent single dye experiments.

PIE has found great utility in FRET experiments since high FRET efficiency requires significant overlap of donor emission and accepter excitation spectra. When the donor and accepter are alternately excited using PIE, the donor, FRET, and accepter emission can all be independently verified. Essentially, PIE accurately differentiates FRET emission from accepter emission which proves the existence of the FRET and allows FRET detection for complexes with very low FRET efficiency.

SCAN-PIE had a number of different factors from traditional PIE. First, the synchronized clock from a field programmable gate array (FPGA) development board was run at 20 kHz instead of tens of megahertz typically seen with PIE. The period of the pulses was at least a full order of magnitude smaller than the average duration of a molecule inside the inspection volume ensuring adequate molecule sampling by all laser lines. Secondly, we reduced the number of pulsing lasers to one (405 nm) while interleaving the emission channels using post-processing techniques explained later. This was born out of necessity

and simplicity as we lacked, at the time, enough pulsing lasers with the correct emission lines for more intricate pulsing.

For initial 3-color experiments, we used Alexa Fluor 405, YOYO-1, and Alexa Fluor 647 as our dyes of choice. Crosstalk between Alexa Fluor 405 and/or YOYO-1 into the 680/42 channel was negligible due to the spectral separation whereas heavy crosstalk existed for Alexa Fluor 405 (emission: $\lambda_{Max} = 421$ nm), YOYO-1 (emission: $\lambda_{Max} = 509$ nm), and their associated emission channels. Therefore, a pulsing scheme was implemented as follows

1. With the 405 nm and 488 lasers both on, sample from the 452/45 channel only giving the Alexa Fluor 405 signal. YOYO-1 emission will not be detected in the 452/45 channel.
2. With the 405 nm off and the 488 nm laser on, sample the 525/40 channel giving the YOYO-1 signal. Alexa Fluor 405 will not be excited by the 488 nm laser.

For further understanding of the system design, the spectra of the Alexa Fluor 405 and YOYO-1 dyes along with the relevant emission filters are shown in Figure 3.4.

Synchronization of the laser pulsing and the time correlated single photon counting data acquisition (TCSPC DAQ) module was necessary so that each data bin could be mapped to the corresponding laser pulse. This synchronization was accomplished by generating two signals derived from a common clock signal using a FPGA development board. A 50% duty cycle 10 kHz square wave was inputed to the 405 nm laser to achieve pulsing and a 50 ns TCSPC triggering pulse signal operating at 20 kHz gave temporal data locations. Figure 3.5 shows

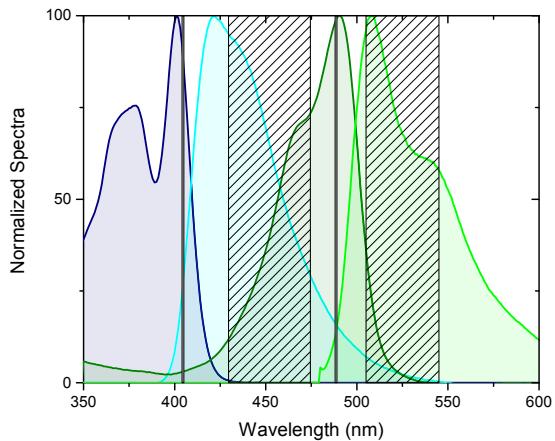


Figure 3.4: Excitation and emission spectra for Alexa Fluor 405 (blue) and YOYO-1 intercalating dye (green). The boxes with slanted lines represent emission filter passbands (452/45 and 525/40). Lines refer to 405 nm and 488 nm laser lines. Crosstalk from Alexa Fluor 405 into the 525/40 passband can be observed. Note that spectra are normalized to their respective peaks and don't represent absolute quantities for direct comparison between dyes. Data from Thermo Fisher Fluorescence SpectraViewer.

a raw time trace of a coincident molecule labeled with Alexa Fluor 405 (top) and YOYO-1 (bottom). Notice that the Alexa Fluor 405 trace goes to zero for every other $50\ \mu\text{s}$ data bin demonstrating the pulsing of the 405 nm laser. Being the 488 nm laser is on continuously, the YOYO-1 emission exhibits no pulsing behavior. However, depending on the molecule, some 525/40 emissions may exhibit oscillatory behavior due to crosstalk. Crosstalk of the Alexa Fluor 405 was estimated to be approximately 8% of the 452/45 signal.

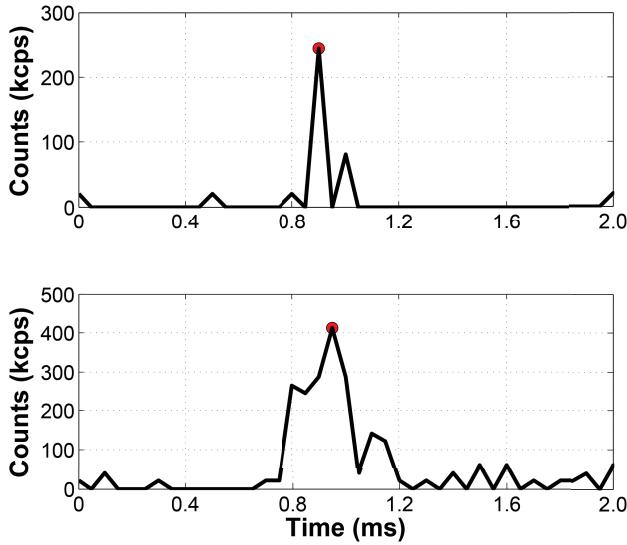


Figure 3.5: Raw time trace demonstrating pulsing. The upper trace shows the 452/45 channel with Alexa Fluor 405 being excited. The lower trace shows the 525/40 channel with YOYO-1 being excited.

We performed three operations on the pulsed raw time traces. First, we downsampled both the 452/45 and the 525/25 channel time traces to decrease the sampling rate by a factor of two and remove the unwanted parts of the signals. The phase of the downsampling was not arbitrary since we wanted to keep the 452/45 signal when the 405 nm laser was on and the 525/40 signal when the 405 nm laser was off. Second, we used an interpolation filter to upsample the signals back to the original sampling frequency for ease of downstream data analysis and to maintain proper timing between all three signals. Lastly, we delayed the 525/40 signal by one bin since the downsampling and upsampling operations advanced the signal one bin relative to the 452/45 signal. Figure 3.6 shows the result of the signal processing with the original signals shown in Figure 3.5. Slight signal distortions (e.g. signals drop below 0 counts) are observed in the resulting time traces due to the low-pass interpolation filter. However, these distortions cause little or no problems to the downstream SCAN analy-

sis and the benefits of the SCAN-PIE approach, namely removing crosstalk, far outweigh any chance for errors caused by signal distortions.

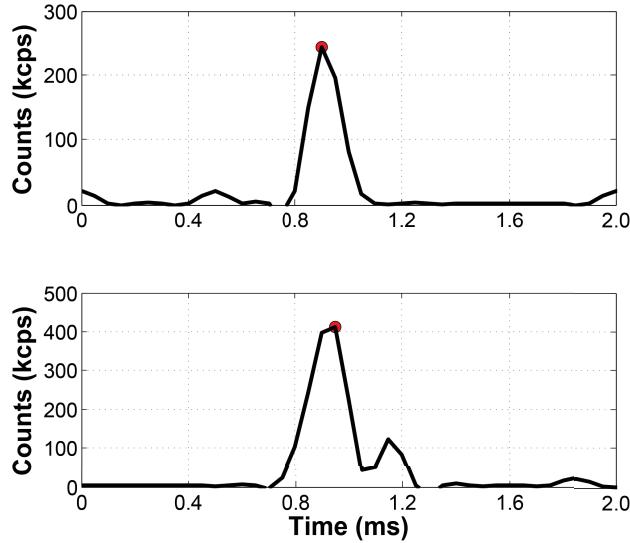


Figure 3.6: Corrected signals after downsampling and interpolating. The upper trace shows the 452/45 channel with Alexa Fluor 405. The lower trace shows the 525/40 channel with YOYO-1. The minor signal distortions are caused by the low pass interpolation filter.

Size Exclusion Chromatography for Free Antibody Removal

Prior to running 3-color SCAN, excess free antibodies were removed from the samples using size exclusion chromatography (SEC) as described in Section 3.2.8. Since binding reactions were performed with a high molar excess (10-50x) of antibodies to ensure adequate binding, we wanted to remove the unbound antibodies after crosslinking in order to have a high concentration of informative molecules and maintain throughput. Having excess antibodies created multiple problems for SCAN as it gave coincidence false positives, raised background which causes uncertainty in the coincidence rate, and it limited throughput for SCAN-Sort.

Figure 3.7 shows the result of a SEC run after analyzing the collected fractions on a plate reader. Chromatin (green) exhibited a broad peak relative to the free antibodies (blue, red) which enabled the removal of the free antibodies from the chromatin sample. We typically chose fractions which eluted just before the free antibodies such as fraction 17.

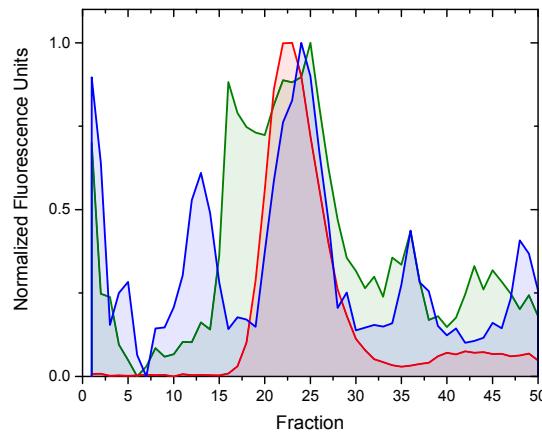


Figure 3.7: Size exclusion chromatogram for YOYO-1 labeled chromatin (green), α -H3 (blue/Alexa Fluor 405), and α -H2B (red/Alexa Fluor 647). The fraction used for SCAN was 17.

Gravity SEC was not without its drawbacks as we found that it generally lack reproducibility and robustness for the low concentration chromatin samples. SEC also diluted the sample down to at least the 1 nM range which caused problems with buffer exchanges for SCAN and plate reader detection. Recently, we have tried SEC using a HPLC system with limited success.

3-Color SCAN

Initial experiments for 3-color SCAN consisted of using YOYO-1 labeled chromatin and two histone targeting antibodies as they maximized the chances

for 3-color coincidence detection. We bound α -H3/Alexa Fluor 405 and α -H2B/Alexa Fluor 647 antibodies to chromatin fragments of approximately 1 kb in length, crosslinked, and ran the sample through the size exclusion column for free antibody removal prior to analysis using 3-Color SCAN. We detected α -H3/chromatin coincidence, α -H2B/chromatin coincidence and 3-color coincidence using SCAN. We identified roughly 1900 α -H3/chromatin coincident events giving a coincidence rate of about 4% with regards to total number of YOYO-1 events. A coincidence offset histogram depicting the coincidence peak above background is presented in Figure 3.8a. A similar coincidence rate was observed for α -H2B/chromatin with approximately 1900 detected coincident events as shown in Figure 3.8b. We observed about 550 three color coincident molecules as shown in Figure 3.8c. This represented approximately 1% on the total number of DNA observed. Based on these estimates of two and three coincidence rates, we have a positive covariance which implies a mutual positive dependence on the coincidence rates for α -H2B/chromatin and α -H3/chromatin. We expected this behavior. However, we should note that ideally the 2-color coincidence rates should be equal to the 3-color coincidence rate for α -H3 and α -H2B as each nucleosome contains two H3 histones and two H2B histones. Being that the 3 color coincident rate was only around 29% of the 2-color coincident rate, this suggests we had some issues with signal detection and binding chemistries.

3-Color Simultaneous Detection of Epigenetic Marks

While using α -H3 and α -H2B antibodies for a basic proof of principle demonstration of 3-color detection, we envisioned using 3-color SCAN to detect bona

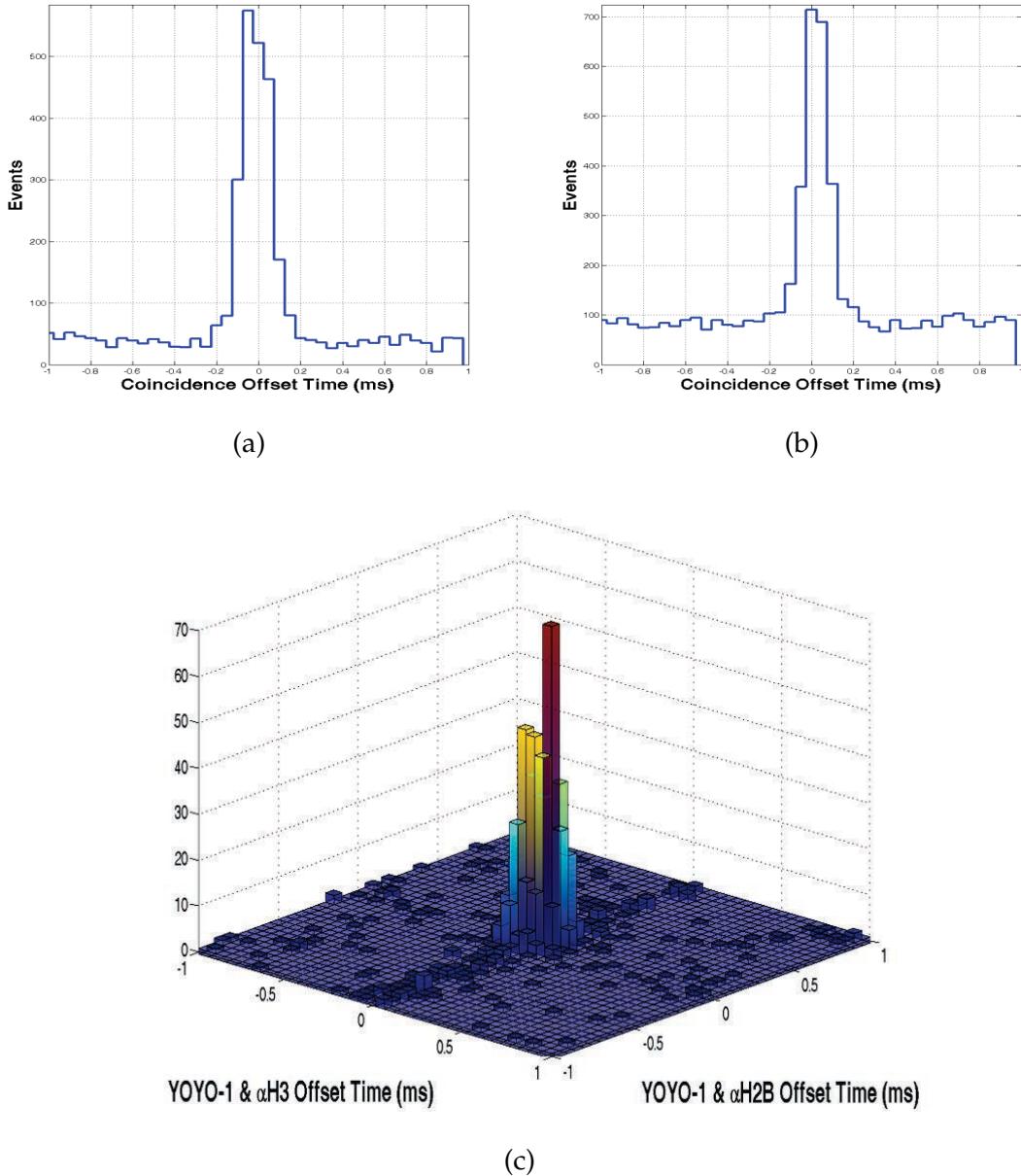


Figure 3.8: (a) Coincidence of α -H3 and chromatin. (b) Coincidence of α -H2B and chromatin. (c) 3-color coincidence of α -H2B, α -H3, and chromatin.

fide epigenetic marks. For this, we chose to investigate H3K27me3 and 5mC using 3-color SCAN. Previous investigations of these two marks using ChIP-BS-seq found that 5mC prevents or antagonizes H3K27me3 in the genome [21, 22]. Moreover, coordination of these marks have been found to change in cancers such as prostate cancer where an enrichment of DNA methylation and deple-

tion of H3K27me3 occurs in CpG islands and transcription start sites [23]. Since DNA methylation and H3K27me3 patterns change with tumorigenesis, we wanted to determine if we could measure these changes using primary mouse fibroblast (MF) and immortalized mouse fibroblast 60.1 cells. We bound α -H3K27me3 and MBD to MF chromatin and analyzed it using 3-color SCAN. The results, shown in Figure 3.9, show coincidence detection of the 2 probes with chromatin. We detected approximately 3900 coincident H3K27me3/chromatin events for a coincidence rate of 6% as shown in Figure 3.9a. Additionally, we detected 1000 coincident MBD/chromatin complexes as depicted in Figure 3.9b. Finally, we observed approximately 250 coincident 3-color events for a 3-color coincidence rate of roughly 0.4%.

We analyzed MF60.1 chromatin using the same procedure as used with the MF chromatin. Again, we observed 3-color coincidence as illustrated in Figure 3.10. We identified 5800 coincident H3K27me3/chromatin molecules, 1200 MBD1/chromatin molecules, and 400 coincident 3-color molecules which correspond to coincidence rates of 10%, 2%, and 0.7% respectively.

As will be explained in Section 3.3.4, the coincidence rates, especially those involving intercalators should be interpreted with some skepticism and understanding of the variability attributed to different threshold levels for molecule discrimination. The thresholds for the 3-color experiments were set at a SNR=10, which is an arbitrary yet consistent metric. With that said, we will proceed in the following analysis to demonstrate the potential power of 3-color SCAN. Since the analysis will use relative quantities and consistent thresholding, errors caused by thresholding will be assumed to cancel out and therefore be mitigated.

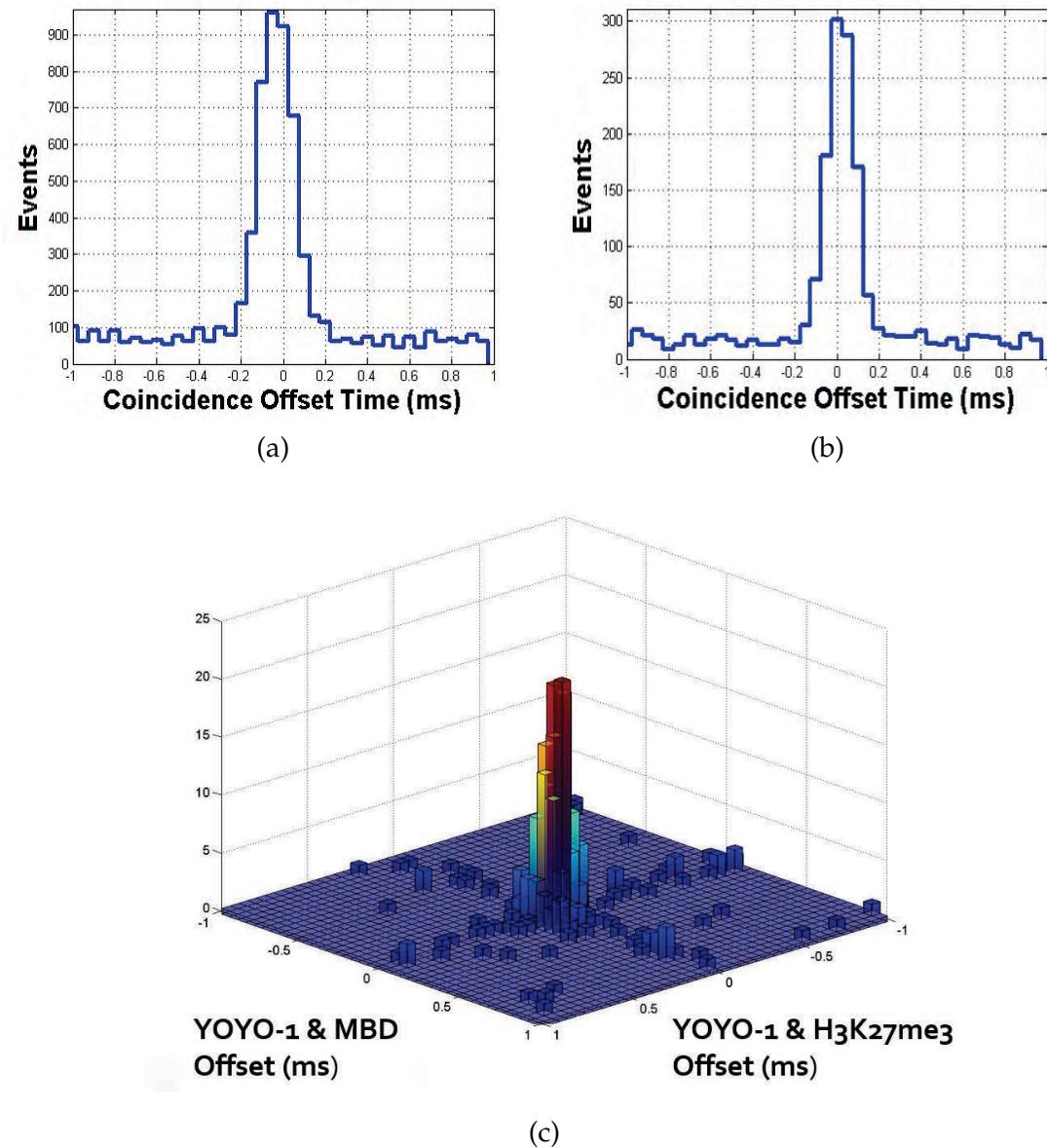


Figure 3.9: Coincidence from mouse fibroblast chromatin (a) Coincidence of α -H3K27me3 and chromatin. (b) Coincidence of MBD1 and chromatin. (c) 3-color coincidence of α -H3K27me3, MBD1, and chromatin.

All else being equal, the data suggests higher levels of 5mC, H3K27me3, and colocalized 5mC/H3K27me3 exist in the immortalized MF60.1 cell line relative to the MF cell line. However, this could be due to such factors as the MF60.1 having a greater percentage of chromatin relative to free DNA in the sample. To

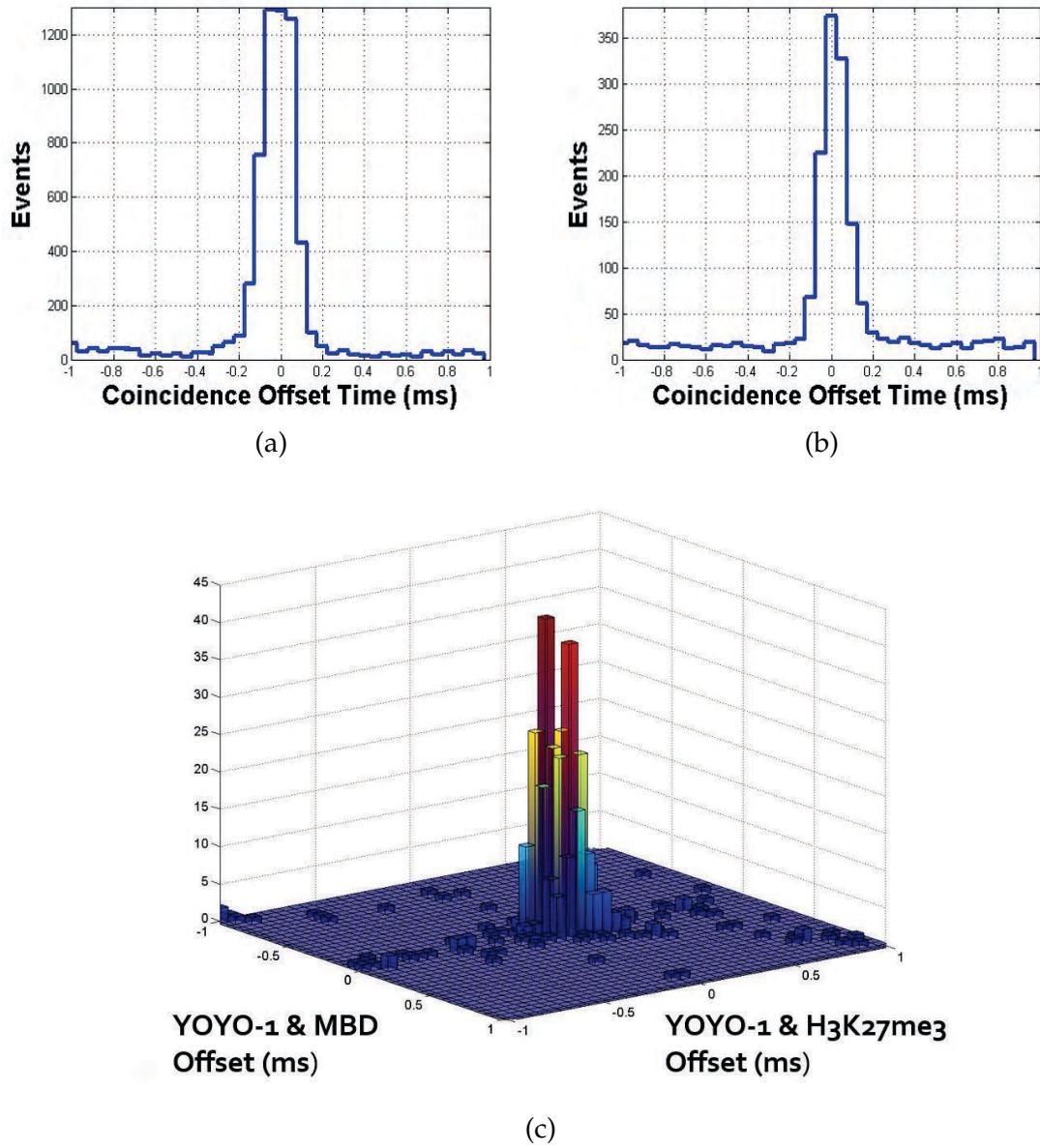


Figure 3.10: Coincidence from mouse fibroblast 60.1 chromatin (a) Coincidence of α -H3K27me3 and chromatin. (b) Coincidence of MBD1 and chromatin. (c) 3-color coincidence of α -H3K27me3, MBD1, and chromatin.

normalize these effects out, we looked at the ratio of 3-color coincident events to H3K27me3 coincident events along with the ratio of 3-color coincident events to MBD1 coincident events. We found slight increases in these ratios for the MF60.1 suggesting the possibility that antagonism decreases slightly for the im-

mortalized cell line. However, these increases were most likely within the error of the measurement.

	3-Color Coincidence Per Coincident H3K27me3	3-Color Coincidence Per Coincident MBD1
MF	6%	25%
MF60.1	7%	33%

Table 3.1: Antagonism of MBD1 and H3K27me3 for MF and MF60.1 cell lines

3.3.3 Quantification of Epigenetic Marks

Previously in Sections 3.3.1 and 3.3.2, we reported our ability to detect multiple epigenetic marks using SCAN. Logically progressing from this result, we set out to quantify the overall abundance of epigenetics marks using SCAN.

The power to quantify epigenetic marks, as noted before, is of great interest to researchers as deviations from normalcy can indicate cancer or other disease states. Standard ensemble based methods to address global levels of histone modifications are ChIP-seq and mass spectrometry, with ChIP giving underlying sequence information. Unlike ChIP which is limited to querying two histone modifications, mass spectrometry can report higher order combinations if the modifications are near one another on their histone tails. Quantitative LC-MS gives no sequence information however.

Analytical quantitative SCAN would give more information than ChIP and LC-MS methods on a genome wide level as it wouldn't be limited in spatial resolution or the number of combinations of epigenetic marks investigated simultaneously. Resolution could be adjusted by changing the lengths of the in-

put chromatin which could be done, at least on a macroscale level, by using differential centrifugation. Additionally, quantitative SCAN could possibly address larger combinations of epigenetic marks and their associated phenotypic variations based on their interplay. The number of simultaneous marks probed would be limited by the spectral bandwidth of the SCAN optics, the amount photons collected per dye, and the degree of spectral overlap between adjacent dye emissions. At the present time, we could envision at least four epigenetic marks being investigated in the future.

Compared to detection, quantification of epigenetic marks using SCAN requires much more information about the sample and the underlying distributions of the signal and background. One of quantitative SCAN's alluring features is the assumed simplicity of its operation; it is just a counting exercise as single molecule methods including SCAN provide the ultimate resolution - a single entity. However, if the signal isn't well separated from the background, then quantitative SCAN becomes much more difficult, if not impossible, to produce accurate results pertaining to the sample.

In the following, we will discuss how we analyzed SCAN results, illustrate the technical challenges of quantitative SCAN, and give some insights into how to address some of these issues. It will be necessary to keep in mind during the discussion that fluorescent bursts with SCAN can be classified into three distinct categories:

1. Noise inherent to the SCAN system including the fluorescence of the substrate, buffer, and detection optics, detector dark counts, SPAD afterpulsing, and Raman scattering.
2. Bursts due to contaminants and non-desired molecules (e.g. short frag-

ments of DNA). These bursts are typically low intensity yet above the inherent noise level.

3. Bursts due to sample molecules.

3.3.4 Theoretical Considerations for Quantitative SCAN

Appropriate statistical modeling of the underlying physical process is necessary to determine the signal and noise distributions associated with SCAN. Unlike a coherent source which is described by Poisson statistics and constant power output, partially coherent light such laser induced fluorescence exhibits broadening due to additional sources of randomness. These sources include fluctuations in the number of dyes in the inspection volume, spatial variations of laser excitation due to the nonuniform point spread function (PSF), and mechanical drift of the stage causing focusing variability.

Fundamentally, in the semiclassical approximation, the photon number statistics are governed by Mandel's formula [24, 25] given by

$$p_{PCH}(k) = \int_0^{\infty} \frac{(\eta_I I_D)^k \exp(-\eta_I I_D)}{k!} p(I_D) dI_D \quad (3.3)$$

where η_I is the detector efficiency and I_D is the intensity at the detector. Equation (3.3) illustrates that partially coherent light consists of a weighted Poisson distribution with the weights being determined the probability of intensity fluctuations at the detector. The effect of these fluctuations is to broaden the photon counting histogram (PCH) such that it becomes super-Poissonian with its variance greater than its mean. In general, these variations in fluorescence intensity are caused by spatial inhomogeneity of the point spread function and fluctuations in the number of dyes within the observation volume.

A rigorous derivation for an expression $p_{PCH}(k)$ containing experimental parameters relevant to SCAN can be performed following logic similar to Chan et al [25]. Assuming the Poisson probability, $P_\#(N)$, of observing N molecules in the inspection volume each with n number of dyes per molecule with a distribution given by $p_{dye}(n)$, we find that the photon number distribution for SCAN would be given by

$$p_{PCH}(k; \epsilon) = \sum_{N=0}^{\infty} \sum_{n=0}^{\infty} p^{(N)}(k; V_0, n\epsilon) p_{dye}(n) p_\#(N) \quad (3.4)$$

where

$$p^{(N)}(k; V_0, n\epsilon) = \overbrace{(p^{(1)} * \cdots * p^{(1)})}^{N \text{ times}}(k; V_0, n\epsilon) \quad (3.5)$$

and

$$p^{(1)}(k; V_0, n\epsilon) = \frac{1}{V_0} \int_{V_0} \frac{(n\epsilon \overline{\text{PSF}}(\vec{r}))^k \exp(-n\epsilon \overline{\text{PSF}}(\vec{r}))}{k!} d\vec{r} \quad (3.6)$$

Equation (3.6) gives the photon number distribution for a single molecule with n dyes each with a molecular brightness of ϵ . $\overline{\text{PSF}}(\vec{r})$ represents the normalized point spread function for the system. It was assumed that the molecule could be located with equal probability throughout the inspection volume V_0 . $p^{(N)}(k; V_0, n\epsilon)$ gives the photon number distribution for N molecules in the observation volume. The convolutions result from the fact that the probability distribution of a sum of statistically independent variables (each molecule) is given by the convolution of each individual probability distribution. More information on the derivation can be found in [25].

While the above is rigorous in scope, it is quite intractable for applications such as with SCAN. Looking back to Equation (3.3), if we make an assumption regarding the $p(I_D)$ with a sufficient number of fitting parameters, then we should have a tractable, potentially usefully model. Henceforth, we will assume that $p(I_D)$ can be modeled as a Gamma distribution which implies that $p_{PCH}(k)$

follows a Gamma-Poisson distribution (negative binomial distribution). We use this result as a model for noise in SCAN.

SCAN Noise Analysis

First, the fundamental noise in the SCAN system was characterized through the use of the photon counting histogram (PCH). Since most data bins were absent of bursts either caused by contaminants or the real signal, the PCH's behavior was dominated by the system noise. Thus, the experimental parameters of the noise such as the mean and standard deviation were obtained through a parametric fit of the underlying Gamma-Poisson distribution as shown in Figure 3.11

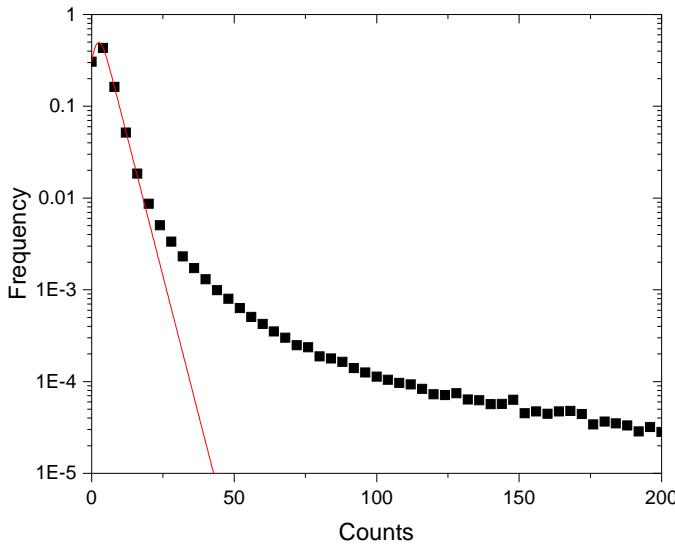


Figure 3.11: PCH for YOYO-1 labeled chromatin using Gamma-Poisson distribution. The noise fit starts to diverge from the data at $\text{SNR} \approx 10$ or counts ≈ 20 suggesting a reasonable thresholding value. Adjusted $R^2 = 0.9998$.

Single molecule identification was performed using varying thresholding al-

gorithms depending on the application. Once the noise properties were determined using the PCH, the thresholding was done by either specifying a desired SNR, a desired number of false-alarm molecules using the Neyman-Pearson rule, or by setting the threshold at some number of standard deviations above the mean noise level. As SNR is conventionally defined as the average signal power to the average noise power, $\text{SNR} = \mathbf{E}(s(t)^2)/\mathbf{E}(n(t)^2)$, we therefore determined the number of counts for thresholding by

$$N = \sqrt{\text{SNR}(\sigma_n^2 + \bar{n}^2)} \quad (3.7)$$

where the standard deviation, σ_n , and the mean, \bar{n} , were both estimated from the PCH noise fitting.

In another thresholding scheme, we trivially applied the Karlin-Rubin theorem coupled with the Neyman-Pearson lemma to provide a threshold value to minimize the probability of a miss while keeping the number of false positives below some user defined molecule count. Since we had an estimate for the noise distribution, $f(N|H_0)$, we were able compute the probability of a false positive which was given by

$$P_F = \int_{N_0}^{\infty} f(N|H_0) dN \quad (3.8)$$

where N_0 was an adjustable thresholding parameter. The thresholding was adjusted until the number of bins per minutes multiplied by P_F equaled a user defined false positive count per minute. Unfortunately, this approach to thresholding wasn't used as prevalently as others since it didn't take into consideration false positives from contaminants which contributed significantly to the overall count.

Once the molecules were determined, additional event filtering could be performed by using metrics such as total burst counts, burst duration, or mean

burst counts. Total burst counts and burst duration are, however, not well defined as they are dependent on the threshold. For example, burst duration was measured as the threshold-to-threshold intersection time difference. Depending on the application, mean burst counts gave more reliable data as the thresholding dependence somewhat mitigated. Case in point, using mean burst counts we were able to differentiate single dyes on individual biomolecules (data not shown).

Signal Distribution

Quantitative SCAN obviously requires sufficient information about the signal distribution in order to accurately estimate the number of legitimate molecules observed. Unfortunately, with the chromatin samples processed, we were never able to come to a reliable, accurate, consistent, and reproducible means to achieve this. Signal distribution separation was never achieved due to the heterogeneity of the sample, low chromatin burst counts due to inefficient intercalation, and excessive background from contamination and short fragments of DNA/chromatin. The heterogeneous sample combined with Poissonian dye fluctuations contributed to excessive broadening of the distribution as even uniform fragment lengths of 500 bp DNA have burst count distributions with standard deviations 15% of their mean. Figures 3.12b and 3.12c shows a typical burst area distribution observed with chromatin processed with SCAN. The distribution has no discernible features differentiating the signal from background as it is essentially a single decay due to factors described above. Another factor for the decaying distribution, at least in part, can be credited to the resulting chromatin length statistics of the MNase digestion.

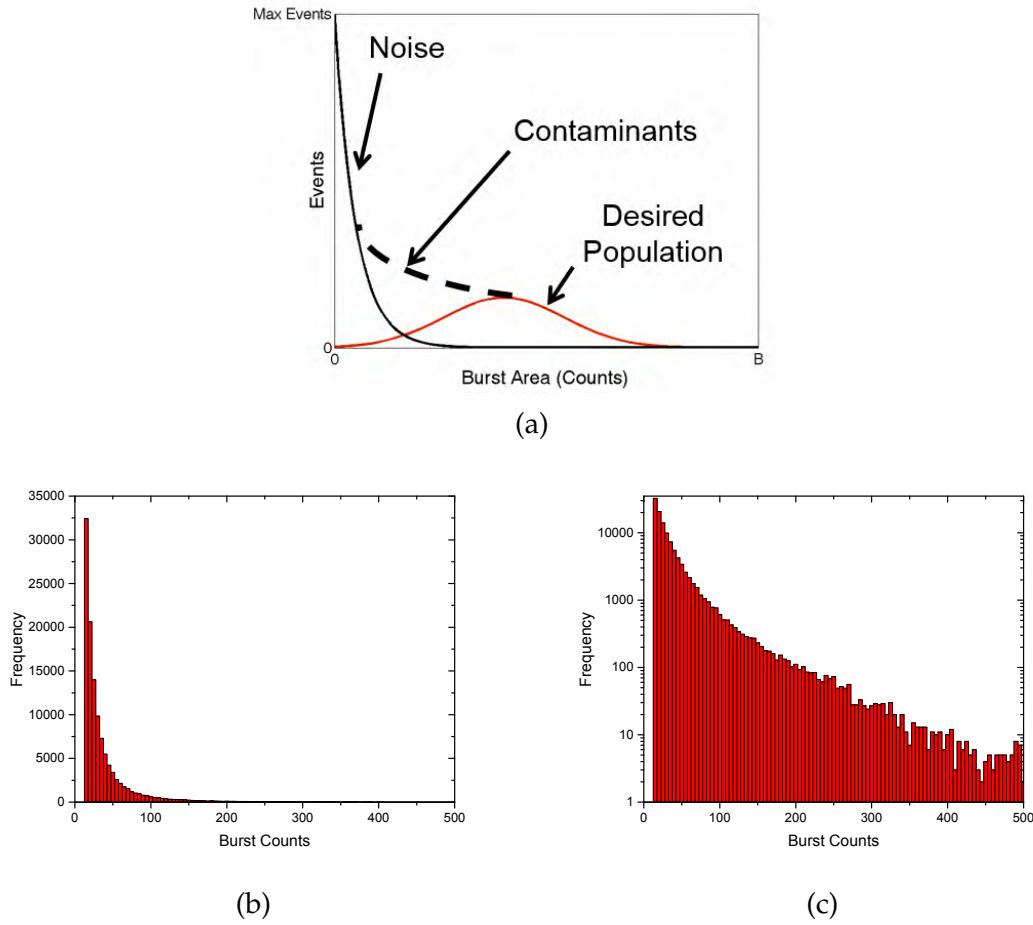


Figure 3.12: (a) Illustration showing the distribution overlaps for chromatin samples. Depending on the chromatin length, the desired population distribution may be shifted farther left than shown. (b) Burst count distribution for YOYO-1 labeled di- and trinucleosome chromatin. Signal can't be separated from background. (c) Semilog plot of (b).

Analytical SCAN focuses on determining the ratio of epigenetic marks observed to the total number of chromatin molecules observed. If we don't have an accurate and definitive measure of the number of chromatin molecules, then the utility of quantitative SCAN is nonexistent. To illustrate this point, we investigated how the percent coincidence varied as the intercalator thresholding for molecule identification was varied. We took YOYO-1 labeled chromatin with an average length of approximately 1 kb and an Alexa Fluor 647 α -H3 antibody

for our model system. After performing the binding reaction α -H3 at a fifty-fold molar excess, we processed the sample using SCAN. With a threshold held for the Alexa Fluor 647 at SNR=10, we analyzed the percent coincidence from a starting intercalator threshold of SNR=5. As shown in Figure 3.13, the percent coincidence monotonically increased as the intercalator thresholding increased. Due to this monotonicity, there is no justifiable reason to pick a given percent coincidence value over another.

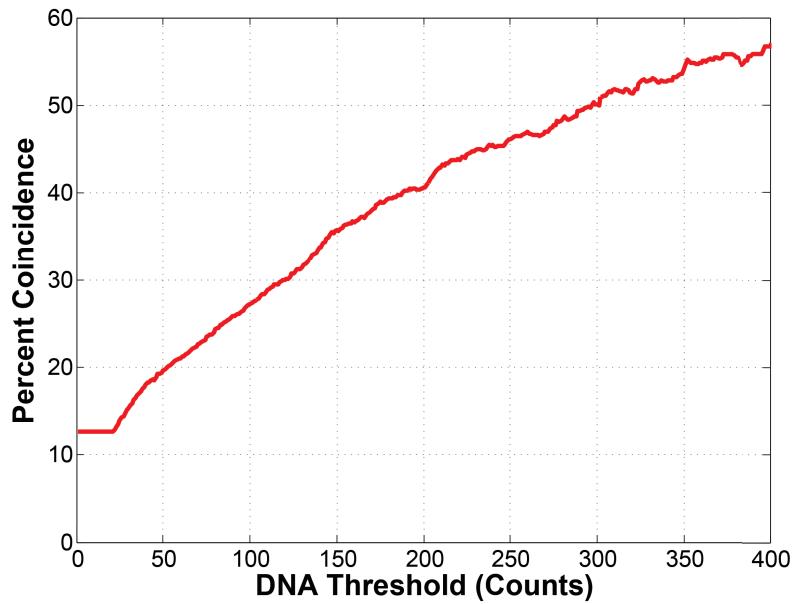


Figure 3.13: Percent coincidence as YOYO-1 thresholding was varied from a lower bound of SNR=5 which corresponded to a thresholding value of ≈ 25 counts. The horizontal line is an artifact of the analysis and doesn't represent a leveling off percent coincidence.

There are two main reasons for this bias. First, larger and brighter chromatin molecules have more binding sites per molecule implying that they would be more biased to have coincidence. Assuming we only have chromatin, the coincidence probability would have a skewed distribution towards longer chromatin fragments. Second, chromatin samples, especially prior to the implementation of the sucrose gradient purification, were excessively noisy as they

contained large amounts of small DNA/chromatin fragments or other fluorescence remnants from the chromatin preparation. These contaminants were generally of low brightness which overestimated the total number of observed DNA molecules.

Quantification of Alexa Fluor 647

As shown above, the inability to quantify chromatin samples using SCAN was due to having no reasonable or accurate estimate of the signal distribution from the generated data. However, this result cannot be broadly applied to all samples run by SCAN. To validate this assertion, we analyzed the smallest quanta of fluorescence, a single Alexa Fluor 647 dye, using SCAN. As illustrated in Figure 3.14, single dye burst count distributions are readily distinguished from noise which is below approximately 10 burst counts when analyzed at 1320 V/cm. The signal distribution was fit using a Gamma-Poisson distribution with a mean of 60.7 ± 0.2 counts and adjusted $R^2 = 0.97$. From the fit, we estimated the number of signal events as 4173 ± 68 molecules over a 4 minutes SCAN run. Since the signal was well-separated from the noise, we could have also simply counted the number of events above 10 burst counts which gave 4220 molecules which was within the fit's error. The distribution could have been also well-represented by log-normal distribution as suggested by [26, 27]. In any case, these results show that SCAN can detect and quantify single fluorophores.

Using Alexa Fluor 647 as a model, we set out to determine the possible molecule count errors associated with signal estimation when signals tended towards the noise. Since burst count distributions can be shifted towards zero by simply increasing the applied voltage, we could easily adjust the distributions

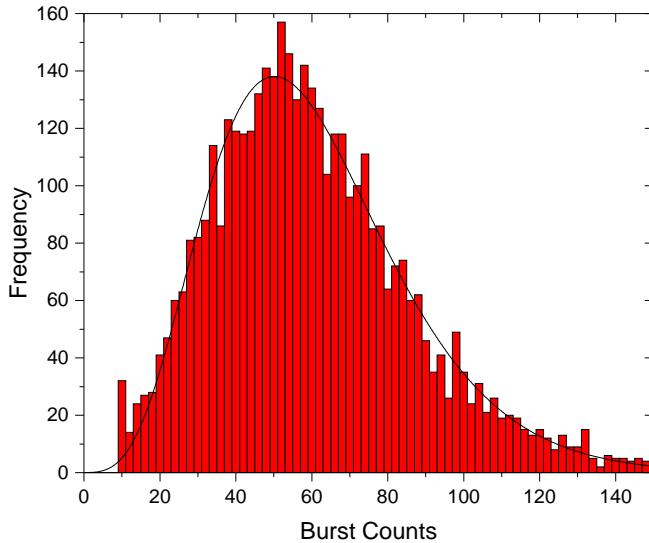


Figure 3.14: Alexa Fluor 647 burst count distribution. Thresholding was set at SNR=10.

and observe the effects. Further, as shown in Chapter 4, SCAN behaves linearly with regards to molecule speed as a function of voltage over standard operating voltages. Using this logic, if the voltage is doubled, then we should analyze twice as many molecules per unit time. Figure 3.15 gives the results from the experiment when the applied voltage was swept from $V_a = 10$ V to $V_a = 100$ V and the burst count distributions were fitted accordingly. Low applied voltages had burst count distributions which were easily distinguishable from the noise as Figure 3.14 illustrates for $V_a = 20$ V. However, as the distributions shifted closer to zero, deviations from the expected molecule counts grew larger. These results suggest that a necessary but not sufficient condition for accurate quantitative SCAN is that signal distributions be substantially resolved from the noise with a mean well above the zero lower bound. Reliance on modeling and fitting for burst count distributions may cause significant underestimation of the true molecule counts.

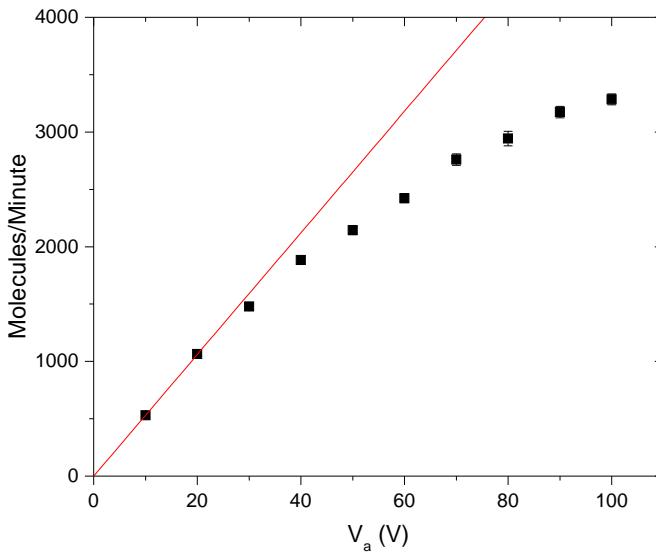


Figure 3.15: Underestimation of molecule counts as signal distributions tended towards the noise. The red line shows the ideal linear behavior.

3.3.5 Signal Processing using Filters

Prior to data analysis, rebinning and filtering were typically performed on the raw signals with the goal to increase the SNR and separate the signal from noise. Depending on the DAQ device, we either binned the raw data at 100 kHz or 20 kHz. This bin width was generally too short and the signals were oversampled. The raw data was rebinned such that the 3-4 bin widths equaled the average burst duration. By reducing the bandwidth through rebinning, we reduced the noise while still satisfying the Nyquist criterion. Since rebinning is essentially an unnormalized central moving average filter followed by a downsample, the high frequency noise was filtered. After rebinning, the signals were processed further using either matched filtering, Lee filtering, or wavelet denoising which are briefly described below.

Matched Filtering, Lee Filtering, and Wavelet Denoising Additional signal processing, smoothing, and filtering was generally performed on the single molecule signals after rebinning to minimize molecular detection error rates caused by high frequency fluctuations and false negatives.

In an ideal system without noise, molecular signals would be a smooth Gaussian due to the convolution of the Gaussian excitation beam with the traveling fluorescent dye. Unfortunately, Poissonian photocount fluctuations, dye number variations inside observation volume, and photophysical effects, namely photobleaching, all give the potential for single molecules to be identified as multiple molecules using standard burst detection algorithms based on thresholding. The result is skewed burst size distributions in which there are significant numbers of apparent low burst size molecules and unidentified larger molecules.

One method to curtail these undesired overcounting effects is to utilize a burst time duration threshold operation which causes two adjoining bursts to be classified as a single burst if the waiting time is below the time threshold. For standard experimental conditions, the burst time duration threshold should preferably be 2 to 3 standard deviations above the mean burst duration. The other method involves signal smoothing using various filters described below.

Almost all SCAN experiments with the exception of pure DNA experiments with analytes greater than 500 bp are performed in a regime with low SNR and SBR. This implies that false positives and negatives are a real concern as there is significant overlap of the underlying signal and noise distributions. Maximizing SNR by discriminating noise from signal using filtering will shift these distributions away from one another and reduce the overcounting of noise and

the undercounting of true molecules.

For additive noise, a matched filter is the optimal linear filter with respect to maximizing the signal to noise (SNR) for a given known signal pulse. The assumed ideal signal for our system representing a molecule of interest is a Gaussian with a width that can be estimated using autocorrelation of the signal. Arguments for the shape of the signal can be related to fluorescence correlation spectroscopy (FCS) theory which dictates that the autocorrelation under flow is governed by [28]

$$G(\tau) = \frac{1}{\langle N \rangle} (1 + (\tau/\tau_D))^{-1} (1 + a^{-2}(\tau/\tau_D))^{-1/2} \exp(-(\tau/\tau_v^2)(1 + (\tau/\tau_D))) \quad (3.9)$$

where τ_D is the characteristic diffusion dwell time, τ_v is the characteristic flow time, $\langle N \rangle$ is the average number of particles in the observation volume, and a is a shape parameter that is the ratio between the longitudinal and lateral axis. For the relevant time scales of SCAN, $\tau \gg \tau_D$ which implies that the autocorrelation and the signal pulses themselves will be Gaussian with a width dependent on τ_v

The impulse response of the matched filtering operation involves convolving the signal with a conjugated time-reversed version of the desired signal. In frequency space, the impulse response is given by [29]

$$H_{matched}(f) = C_f \frac{S^*(f)e^{-j\omega t_s}}{N(f)} \quad (3.10)$$

where $S(f)$ is the Fourier transform of the signal, $N(f)$ is the noise spectral density, t_s is an arbitrary delay, and C_f is a normalization constant. While the delay isn't needed for post-processed SCAN, real-time implementation of the matched filter for SCAN-Sort needs a delay so that it is physically realizable. Equation (3.10) illustrates the key idea of the matched filter in that it accentuates frequencies in the signal spectrum and diminishes frequencies dominated by noise.

Taking the inverse Fourier transform of Equation (3.10), assuming white noise ($N(f)=\text{const.}$), and noting that we are dealing with real signals, we arrive at the time domain form:

$$h_{\text{matched}} = C_t s(t_s - t) \quad (3.11)$$

Thus, our optimal matched filter is a Gaussian where its amplitude is defined by C_t , normalization constant and its width is estimated using the autocorrelation.

Another useful filter used in previous single molecule studies is the Lee filter [30]. This nonlinear filter works as a moving average filter along with a term which effectively modulates the relative amount of filtering based on the local variance and a user-defined parameter. In low variance situations such as observing background or dim molecules, the Lee filter works as a moving average filter. In high variance situations near molecules of interest, the averaging component is suppressed and more of the original signal is passed. The Lee filter can be mathematically represented by

$$\tilde{n}_k = \bar{n}_k + (n_k - \bar{n}_k) \frac{\sigma_k^2}{\sigma_k^2 + \sigma_0^2} \quad (3.12)$$

where the moving average component is given by

$$\bar{n}_k = \frac{1}{(2m+1)} \sum_{i=-m}^m n_{k+i} \quad (3.13)$$

and the variance is described by

$$\sigma_k^2 = \frac{1}{(2m+1)} \sum_{i=-m}^m (n_{k+i} - \bar{n}_{k+i})^2 \quad (3.14)$$

Generally, the filter width, $2m + 1$, should be comparable to the burst duration of the molecules and σ_0 should be of the order of the mean burst height.

Finally, the last signal processing operation used was wavelet denoising which we will discuss very briefly. Further information on wavelets can be

found in [31, 32]. Wavelets, with their multi-resolution capabilities, are ideal for processing SCAN data since we typically want filter out low amplitude, high frequency oscillations (noise) while maintaining high amplitude, high frequency bursts (molecules). Wavelet denoising involves taking the discrete wavelet transform (DWT), applying a threshold to the resulting coefficients, and taking the inverse discrete wavelet transform (IDWT) to reconstruct the denoised data. The DWT generally concentrates signals into a few large magnitude coefficients creating a sparse representation of the signal. The remaining coefficients, which are small, are mostly due to noise. By either hard or soft thresholding the coefficients, we can remove the noise while maintaining the large signal coefficients.

The DWT uses iterated low and high pass filter banks which are quadrature mirror filters of one another to create the details and approximations of the transform. Each level (iteration) decomposes the previous level's approximation coefficients into high and low frequency partitions via the high pass and the low pass filters defined by the particular wavelet of choice. These high and low frequency partitions are that level's detail and approximation coefficients for further decomposition at the next level. After a set number of decomposition levels which is dependent on the SNR of the original signal, the thresholding is applied and the IDWT is performed by using synthesis filter banks.

Figure 3.16 shows the original time trace SCAN signal and a wavelet denoised version. While wavelet denoising removes much of the noise, it also distorts the signal as all filtering operations do to some extent.

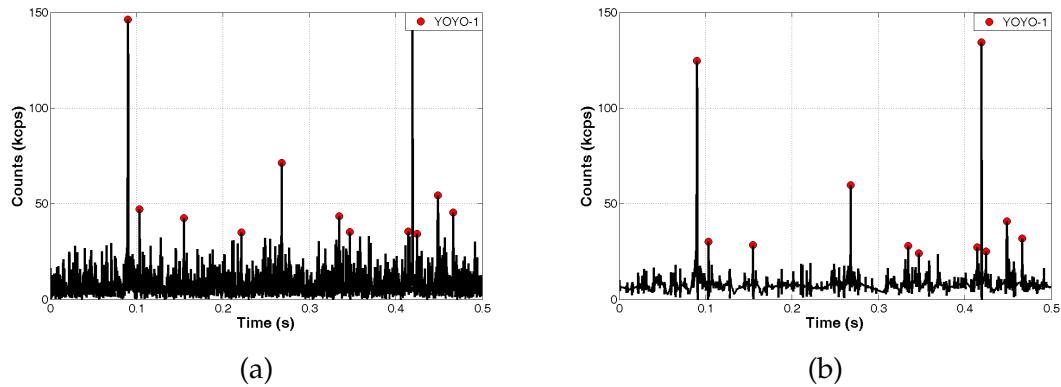


Figure 3.16: Noise reduction via wavelets. (a) Raw time trace (b) Wavelet denoised time trace using a symlet4 wavelet with a decomposition level of 5 and soft thresholding

3.3.6 Chromatin Sample Preparation Adjustments for Quantitative SCAN

We attempted to modify our input SCAN samples using two different approaches so that a distinguishable signal distribution could be observed. One approach, a sucrose gradient purification of the chromatin, focused on removing short fragments of DNA and chromatin along with making our sample more homogeneous. The second approach tried to append Alexa Fluor 488 labeled nucleoside triphophates to the chromatin molecules to vastly improve the signal.

First, the chromatin preps were processed using a sucrose gradient to fractionate the chromatin by equilibrium sedimentation. Mononucleosome, dinucleosome, trinucleosome and larger complexes were separated as illustrated in Figure 3.17. The banding pattern resulting from micrococcal digest of the chromatin indicates its viability along with the number of nucleosomes associated with it. Samples analyzed with SCAN were typically taken from frac-

tions containing trinucleosomes or larger such as the far right chromatin lane in the gel to maximize the intercalator signal. Using these trinucleosome sam-

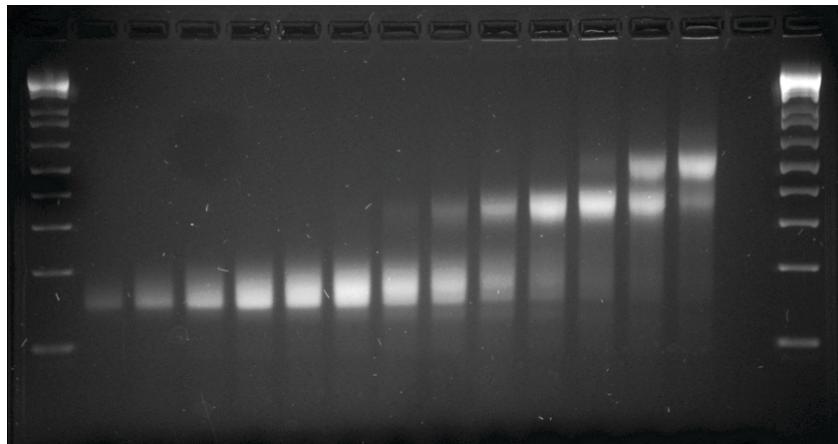


Figure 3.17: Agarose gel showing sucrose gradient purified chromatin. Samples were generally taken from the far right lane consisting of di- and trinucleosomes.

ples, we were still not able to distinguish the signal from background. Now, consider the far right chromatin band of the gel. Since the gel's band intensity proportional to its mass, the dinucleosome and trinucleosome bands have approximately equal molar concentrations. SCAN deals with samples on a per molecule basis so that we were processing approximately equal dinucleosomes and trinucleosomes. Additionally, the gel band has shorter than dinucleosome complexes which could be significant on a per molecule basis. Notice that the dinucleosomes have a length slightly less than 400 bp. Considering that 200 bp is the approximate lower bound for resolution of purified DNA under the best circumstances and that chromatin is at least 33% more dim than DNA, there is no question that quantifying di- and trinucleosomes was not obtainable using these methods.

In order to reliably separate signal from background for quantification, we believe significantly larger sucrose gradient purified chromatin fragments

would be needed. An estimate based on our experience would be upwards of 1.5 kb for the mean of the distribution. This would come at a considerable cost to the resolution of the technique since each fragment would have substantially more nucleosomes.

In order to maintain resolution, we tried a second method achieve a brighter signal by adding Alexa Fluor 488 conjugated dUTPs to the end of the chromatin molecules. There were two inherent benefits to this approach. For one, the Alexa Fluor dyes are generally brighter than intercalators by any order of magnitude. Secondly, our sample would be more homogeneously labeled instead of having a chromatin length dependent brightness. Since the chromatin length dependence would be eliminated, we would theoretically be able to quantify mononucleosome chromatin fragments using SCAN.

Terminal deoxynucleotidyl transferase (TdT), typically catalyze these reactions to the 3' end of DNA with protruding, recessed, or blunt ends. After considerable effort by members of the Soloway lab, we were unable to affix the dUTPs to the ends of the chromatin perhaps due to the nucleosome interfering with the TdT's activity and the method was abandoned.

In light of the recent report demonstrating single cell ChIP-seq [33], we now have reason to believe that DNA fragments can be ligated onto the ends of the mono-, di-, and tricnuleosome fragments. They used a DNA end repair kit for blunt ending the DNA in conjunction with T4 DNA ligase to ligate DNA bar-codes onto individual chromatin fragments. By affixing fluorescently labeled oligonucleotides to the mononucleosome chromatin ends by following the same protocol and removing the unreacted oligonucleotides and other short DNA by size exclusion, we would have a mono-dispersed, bright sample for quantitative

SCAN analysis.

3.3.7 Low Coincidence Rates of SCAN

To close out this chapter, we will discuss reasons for the low coincidence rates that we have continually observed with SCAN. There are many reasons for the low observed coincidence rates including:

1. The sample contains degraded chromatin in the form of free DNA.
2. Contaminants.
3. Binding concentrations imply precipitin reactions or other chemistry problems.
4. Antibody degrees of labeling.
5. Background coincidence correction.

First, we know from mobility-SCAN experiments described in Chapter 4 that there is a potential for a large amount of sample to be degraded chromatin. Based on experience from mobility experiments, at least 50 % of the sample will be degraded chromatin if no spin columns are applied in the sample prep. Chromatin rich samples should be promoted by using 100 kDa size exclusion spin columns after crosslinking to remove free DNA. Still, coincidence rate corrections for chromatin degradation should be done by analyzing the sample using mobility SCAN to determine the ratio of DNA to chromatin.

Second, based on mobility SCAN data and burst count distributions, we know that the chromatin samples are generally contaminated with very dim

fluorescent particulates that are nearly neutral and flow with the electroosmotic flow. While this example is for illustration only as the flow can be made unidirectional by only loading one port with the sample, it does suggest that contaminants could be an issue causing over counting of molecules. Figure 3.18 shows an electrophoretogram of HeLa-GFP trinucleosome chromatin for the 525/40 emission channel. The right peak corresponds to intact chromatin based on the overlap of the TOTO-3 electrophoretogram (data not shown). The left peak is low intensity contaminants of unknown origin, but not DNA or GFP. We know its not DNA as the flow is in the incorrect direction and free H2B-GFP doesn't flow due to its absorption to silica caused by its large positive charge. The contamination is most pronounced with chromatin samples and is generally nonexistent for commercially purchased reagents such as DNA and other labeled proteins.

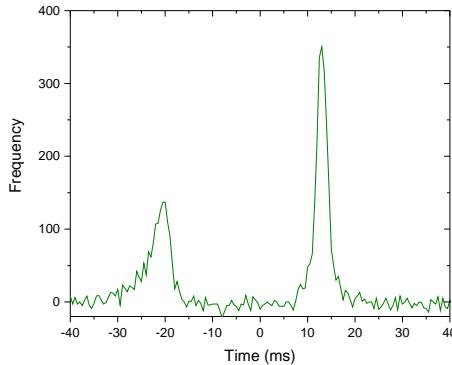


Figure 3.18: Electrophoretogram of 525/40 channel with a HeLa-GFP sample. The right peak represents chromatin in the form of GFP. The left peak is unknown very low intensity contaminants. The thresholding was set at SNR=10. As the SNR increases, the left peak vanishes quickly relative to the right peak.

Third, the binding reactions were executed at perhaps at non-optimal concentrations creating aggregation and poor binding depending on the K_D of the

antibodies used. Large scale fluorescent aggregates especially when analyzing chromatin samples were observed with SCAN which caused channel clogging and presumably low coincidence rates. The root cause of these aggregates was a precipitin reaction in which the multivalent antibodies bound to two different antigens and produced a lattice of antibody-antigen complexes. With precipitin reactions, it is well known that maximum precipitation occurs at the zone of equivalence where antibody and antigen concentrations are approximately equal. For sake of argument, consider a binding reaction of consisting of chromatin and a histone targeting antibody. For our binding reactions, we typically used 500 nM of antibody as nonspecific binding was thought to be a problem at higher concentrations. Additionally, it wasn't unusual to use 20-30 ng/ μ L of chromatin in the binding reactions which would put an average nucleosome concentration of 200-300 nM, independent of the chromatin size. Now, each nucleosome has two histone proteins of a given type which implies the effective antigen concentration was actually closer to 500-600 nM implying that we were binding near the zone of equivalence. Aggregation of this type, especially regarding histone targeting antibodies, could be one reason why the histone targeting antibody coincident rates were comparable to coincident rates observed using histone modification targeting antibodies.

While many antibodies have K_D of 10^{-9} or better, a reasonable fraction have K_D values much worse. Abcam estimates that over 15% of mouse antibodies have K_D values at 10^{-7} or worse. With these K_D values and binding at a 50 fold molar excess at 500 nM, we would have over 16% of the epitopes free of bound antibody creating a situation of underreported coincidence. In either case, better binding conditions probably would have occurred by using a larger molar excess of antibodies which would limit precipitin and shift the equilibrium to-

wards binding.

Fourth, many of the experiments were performed with probes which were underlabeled. For example, consider an antibody whose degree of labeling is two. Due to the Poisson statistics of the labeling process, 14% of the antibodies would have no label at all and hence contribute to lowering the coincidence rates. Practically, we should probably work with probes that have a degree of labeling of 3 or better as it would increase coincidence and signal. A degree of labeling of 3 or better ensures that over 95% of the probes have labels. Any binding degradation caused by dye loading would have few negative effects on the resulting SCAN measurements.

Fifth, simple coincidence background subtraction causes errors upwards of 20% which will be explained in detail below.

Background Coincidence Correction for SCAN

Background coincidence will occur in SCAN coincidence measurements due to uncorrelated events in the associated fluorescent channels. Correcting for this chance coincidence is not as simple background subtraction due to subtleties in the coincidence algorithms. Here, we will derive formulas relating to background coincidence, examine the nuances in the algorithms, and provide Monte Carlo simulations demonstrating these outcomes.

Consider a 2-color SCAN experiment with single molecule event rates given by r_1 and r_2 . Since we are dealing with a Poisson process, the arrival times will follow a uniform distribution due to the independent and stationary nature of the process. This implies that underlying background coincidence should

be uniform too. Now, we will derive the magnitude of the background as a function of r_1 and r_2 . Suppose an event occurs in channel 1. The properties of the Poisson process imply that the waiting time between adjacent events will be exponentially distributed. The probability without an event in channel 2 in the time interval T after the channel 1 event is $\exp(-r_2 T)$. Additionally, the probability that a channel 2 event will occur in the next infinitesimal time step dT is $r_2 dT$ which gives a differential probability of $r_2 \exp(-r_2 T) dT$. The differential rate of chance coincidence is the rate of the channel 1 weighted by the differential probability. Thus, the chance coincidence rate is simply then $dr_{chance} = r_1 r_2 \exp(-r_2 T) dT$. For $r_2 T \ll 1$, the exponential is essentially unity. Thus, we find the chance coincidence rate given by

$$r_{chance} = r_1 r_2 \Delta T \quad (3.15)$$

where ΔT is the time resolution (bin size) of the measurement. Eq. (3.15) illustrates that the background coincidence rate from uncorrelated inputs is proportional to both the singles coincidence rates r_1 and r_2 along with the coincidence resolving time. Naively, the true number of detected coincident events should be able to be determined by simply subtracting this background and integrating the area under the coincidence peak. However, this is not the case.

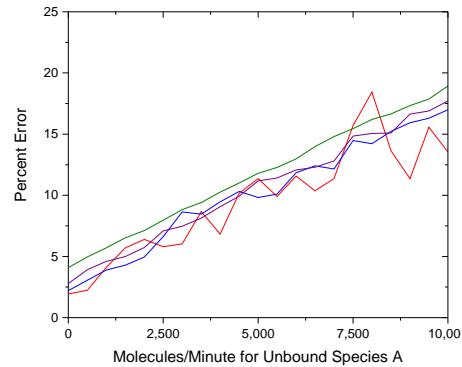
Consider analyzing two species, A and B , and their chemically crosslinked product AB using SCAN. First, only A and B are analyzed and are diluted in such a way that they don't preferential bind to each other and remain uncorrelated in the experiment. The resulting analysis would give no coincidence peak and a random background coincidence level given by Eq. (3.15). Ideally, if the crosslinked product were to be added to the sample, a Gaussian peak would emerge above background with its area being equal to the number of crosslinked AB molecules. In reality, due to the unbiased discrimination of coin-

cident molecules, the measured number of coincident molecules would be less than the number of AB molecules added to the system.

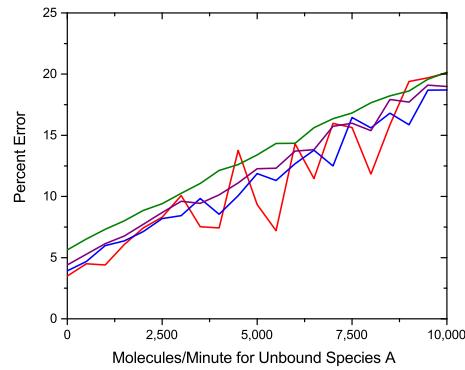
To understand this, assume that an AB coincident molecule is in the observation volume at time t_0 . For the sake of argument, assume we have n number of A molecules located at times $t_i = t_0 + \tau_i$ for $i = 1, 2, \dots, n$. Each A molecule will have some coincident offset time τ_i associated with it due to the AB molecule at $t = t_0$. The existence of these A molecules will elevate the overall coincidence background by a factor proportional to r_A . It follows that each coincident molecule will contribute less than one coincident event to the background subtracted coincidence. As the arrival rates of noncoincidence molecules increases, the coincidence rate error associated with simple background subtraction will become larger. In order to determine how large the errors could be with SCAN, we performed Monte Carlo simulations.

Monte Carlo Simulation

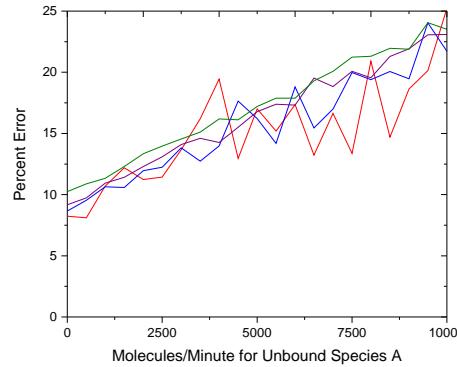
We developed a Monte Carlo simulation to model the underlying system using MATLAB. We randomly generated arrival times based on a uniform distribution with a time resolution of 1 ms. The input parameters for the simulations included the number of A , B , and AB molecules. We assumed realistic values for the simulation and demonstrated that coincidence rates due to this background effect can decrease detected coincidence by as much as 20% relative to the true coincidence as shown in Fig 3.19



(a)



(b)



(c)

Figure 3.19: Monte Carlo simulations showing underreported coincidence. Coincidence molecules/minute: 100 (red), 250 (blue), 500 (purple), 1000 (green). Total simulation time: 30 minutes. (a) Unbound Species B = 1000 molecules/min (b) Unbound Species B = 2000 molecules/min (c) Unbound Species B = 5000 molecules/min

3.4 Conclusions

Over the course of the chapter, we have developed techniques to detect epigenetic marks bound to individual chromatin fragments. We were able to show 2 and 3 color coincidence of these marks. 3-color detection required a pulsed interleaved excitation scheme to reduce crosstalk to negligible levels. However, we were never able to quantify the marks across the genome as the signal to background was poor and coincidence rates were below expected levels.

In order to quantify epigenetic marks at a resolution of mono-, di-, and trinucleosomes, sample preparation needs to be advanced as we simply cannot differentiate signal from background. This issue stems from two sources, the poor intercalation of the chromatin and the heterogeneity of the sample. We envision addressing the poor signal problem by ligating fluorescent oligonucleotides to the sucrose purified chromatin. Recent confirmation of the ability to ligate DNA onto chromatin fragments was provided by Rotem et al [33] giving optimism to the technique. After ligation, unreacted oligonucleotides, short fragments of DNA, and other contaminants would have to be removed. This would most likely be done via size exclusion methods.

The signal to background issue wasn't the only issue suffered by quantitative SCAN as low coincidence rates were also continuously observed. We outlined several causes for the low coincidence rates including degraded chromatin, sample contamination, background binding coincidence correction factors, and non-optimal binding conditions. Many of these issues revolve around sample preparation and the ability to remove excess antibodies from solution to which we haven't found a reliable solution.

Furthermore, due all these dynamic variables, the possibility of absolute quantification of epigenetic marks may have too many errors associated with it to ever be reliable or reproducible. If absolute quantification of epigenetic marks is ever going to be a reality, we would have to confirm our results to conventional ChIP-seq data. Only then could we be confident in our method.

BIBLIOGRAPHY

- ¹H. Sun, P. J. Kennedy, and E. J. Nestler, “Epigenetics of the depressed brain: role of histone acetylation and methylation”, *Neuropsychopharmacology* **38**, 124–137 (2013).
- ²J. Lord and C. Cruchaga, “The epigenetic landscape of alzheimer’s disease”, *Nat Neurosci* **17**, 1138–1140 (2014).
- ³C. Plass, S. M. Pfister, A. M. Lindroth, O. Bogatyrova, R. Claus, and P. Lichter, “Mutations in regulators of the epigenome and their connections to global chromatin patterns in cancer”, *Nature Reviews Genetics* **14**, 765–780 (2013).
- ⁴M. T. Villanueva, “Epigenetics: chromatin marks the spot”, *Nature Reviews Cancer* **15**, 196–197 (2015).
- ⁵M. A. Dawson and T. Kouzarides, “Cancer epigenetics: from mechanism to therapy”, *Cell* **150**, 12–27 (2012).
- ⁶B.-R. Hyun, J. L. McElwee, and P. D. Soloway, “Single molecule and single cell epigenomics”, *Methods* **72**, 41–50 (2015).
- ⁷B. R. Cipriany, R. Zhao, P. J. Murphy, S. L. Levy, C. P. Tan, H. G. Craighead, and P. D. Soloway, “Single molecule epigenetic analysis in a nanofluidic channel”, *Analytical chemistry* **82**, 2480–2487 (2010).
- ⁸M. Foquet, J. Korlach, W. Zipfel, W. W. Webb, and H. G. Craighead, “Dna fragment sizing by single molecule detection in submicrometer-sized closed fluidic channels”, *Analytical Chemistry* **74**, 1415–1422 (2002).
- ⁹M. Foquet, J. Korlach, W. R. Zipfel, W. W. Webb, and H. G. Craighead, “Focal volume confinement by submicrometer-sized fluidic channels”, *Analytical chemistry* **76**, 1618–1626 (2004).

- ¹⁰S. M. Stavis, J. B. Edel, Y. Li, K. T. Samiee, D. Luo, and H. G. Craighead, "Single-molecule mobility and spectral measurements in submicrometer fluidic channels", *Journal of Applied Physics* **98**, 044903 (2005).
- ¹¹S. M. Stavis, J. B. Edel, K. T. Samiee, and H. G. Craighead, "Single molecule studies of quantum dot conjugates in a submicrometer fluidic channel", *Lab on a Chip* **5**, 337–343 (2005).
- ¹²C. H. Reccius, S. M. Stavis, J. T. Mannion, L. P. Walker, and H. Craighead, "Conformation, length, and speed measurements of electrostatically stretched dna in nanochannels", *Biophysical Journal* **95**, 273–286 (2008).
- ¹³W. Kern, "Cleaning solutions based on hydrogen peroxide for use in silicon semiconductor technology", *RCA review* **31**, 187–206 (1970).
- ¹⁴P. J. Murphy, B. R. Cipriany, C. B. Wallin, C. Y. Ju, K. Szeto, J. A. Haggarman, J. J. Benitez, H. G. Craighead, and P. D. Soloway, "Single-molecule analysis of combinatorial epigenomic states in normal and tumor cells", *Proceedings of the National Academy of Sciences* **110**, 7772–7777 (2013).
- ¹⁵H. F. Jørgensen, K. Adie, P. Chaubert, and A. P. Bird, "Engineering a high-affinity methyl-cpg-binding protein", *Nucleic acids research* **34**, e96–e96 (2006).
- ¹⁶N. Keshava and J. F. Mustard, "Spectral unmixing", *Signal Processing Magazine, IEEE* **19**, 44–57 (2002).
- ¹⁷N. Keshava, "A survey of spectral unmixing algorithms", *Lincoln Laboratory Journal* **14**, 55–78 (2003).
- ¹⁸B. K. Muller, E. Zaychikov, C. Bräuchle, and D. C. Lamb, "Pulsed interleaved excitation", *Biophysical journal* **89**, 3508–3522 (2005).

- ¹⁹S. Rüttinger, R. Macdonald, B. Krämer, F. Koberling, M. Roos, and E. Hildt, "Accurate single-pair förster resonant energy transfer through combination of pulsed interleaved excitation, time correlated single-photon counting, and fluorescence correlation spectroscopy", *Journal of biomedical optics* **11** (2006).
- ²⁰S. Fore, Y. Yuen, L. Hesselink, and T. Huser, "Pulsed-interleaved excitation fret measurements on single duplex dna molecules inside c-shaped nanoapertures", *Nano letters* **7**, 1749–1756 (2007).
- ²¹A. M. Lindroth, Y. J. Park, C. M. McLean, G. A. Dokshin, J. M. Persson, H. Herman, D. Pasini, X. Miro, M. E. Donohoe, and J. T. Lee, "Antagonism between dna and h3k27 methylation at the imprinted rasgrf1 locus", *PLoS Genet* **4**, e1000145–e1000145 (2008).
- ²²J. A. Hagarman, M. P. Motley, K. Kristjansdottir, and P. D. Soloway, "Coordinate regulation of dna methylation and h3k27me3 in mouse embryonic stem cells", *PLoS One* **8**, e53880 (2013).
- ²³E. N. Gal-Yam, G. Egger, L. Iniguez, H. Holster, S. Einarsson, X. Zhang, J. C. Lin, G. Liang, P. A. Jones, and A. Tanay, "Frequent switching of polycomb repressive marks and dna hypermethylation in the pc3 prostate cancer cell line", *Proceedings of the National Academy of Sciences* **105**, 12979–12984 (2008).
- ²⁴L. Mandel, "Fluctuations of photon beams: the distribution of the photoelectrons", *Proceedings of the Physical Society* **74**, 233 (1959).
- ²⁵E. Y. Chan, N. M. Goncalves, R. A. Haeusler, A. J. Hatch, J. W. Larson, A. M. Maletta, G. R. Yantz, E. D. Carstea, M. Fuchs, and G. G. Wong, "Dna mapping using microfluidic stretching and single-molecule detection of fluorescent site-specific tags", *Genome research* **14**, 1137–1146 (2004).

- ²⁶L. L. Kish, J. Kameoka, C. G. Granqvist, and L. B. Kish, “Log-normal distribution of single molecule fluorescence bursts in micro/nano-fluidic channels”, Applied Physics Letters **99**, 143121 (2011).
- ²⁷S. A. Mutch, B. S. Fujimoto, C. L. Kuyper, J. S. Kuo, S. M. Bajjalieh, and D. T. Chiu, “Deconvolving single-molecule intensity distributions for quantitative microscopy measurements”, Biophysical journal **92**, 2926–2943 (2007).
- ²⁸R. Rigler and E. S. Elson, *Fluorescence correlation spectroscopy: theory and applications*, Vol. 65 (Springer Science and Business Media, 2012).
- ²⁹S. Haykin, *Communication systems* (John Wiley and Sons, 2008).
- ³⁰J. Enderlein, D. L. Robbins, W. P. Ambrose, P. M. Goodwin, and R. A. Keller, “The statistics of single molecule detection: an overview”, Bioimaging **5**, 88–98 (1997).
- ³¹M. Vetterli and J. Kovacevic, *Wavelets and subband coding* (Prentice-hall, 1995).
- ³²S. Mallat, *A wavelet tour of signal processing* (Academic press, 1999).
- ³³A. Rotem, O. Ram, N. Shoresh, R. A. Sperling, A. Goren, D. A. Weitz, and B. E. Bernstein, “Single-cell chip-seq reveals cell subpopulations defined by chromatin state”, Nat Biotech **33**, 1165–1172 (2015).

CHAPTER 4

SINGLE MOLECULE MOBILITY STUDIES IN NANOCHANNELS

4.1 Introduction

Biomolecules are frequently identified and separated using techniques based on differences in mobilities. One such method, capillary electrophoresis, is a technique in which analytes are electrokinetically driven and efficiently separated in a $20 - 200 \mu\text{m}$ i.d. narrow-bore capillaries. In recent decades as fabrication capabilities have improved, microchip capillary electrophoresis systems have been developed in large numbers due to their ease of fabrication, their massively parallel functionality, and their potential applications in the research, industrial, and medical fields [1–4]. Microchip capillary electrophoresis systems can obtain high resolution separations as the only fundamental source to band broadening comes from longitudinal diffusion. Additionally, with standard microfabrication techniques, detectors are readily integrated on-chip providing benefits such as low cost, portability, and fast analysis [5].

Standard microchip capillary electrophoresis channels are typically limited by detection sensitivity and clogging to a depth of $10 \mu\text{m}$ or more [4]. However, by shrinking the microchip capillary electrophoresis channel diameter by two orders of magnitude and with a suitable detector such as a single photon avalanche photodiode (APD), single molecule detection can be achieved at reasonable concentrations. Within the last decade, researchers have developed techniques to analyze single molecules traveling within these fused silica nanochannels based on their fluorescent signatures. The researchers have used these nanochannels to analyze DNA, chromatin, quantum dots, and PCR prod-

ucts on a single molecule level [6–12]. Moreover, keeping with the spirit of the its larger cousin, microchip capillary electrophoresis, they have also used the nanochannels to measure migration times of nucleic acid engineered fluorescent labels between two lasers located $7\ \mu\text{m}$ apart [13].

Here, we describe device and methodology which rapidly analyzes single molecule mobilities in a fused silica nanochannel device. The technique relies on the detection of laser induced fluorescence from labeled biomolecules at two adjacent locations within a nanochannel. The nanochannel confinement allows for single molecule detection as the molecules, moving under electrokinetic flow, are interrogated sequentially and their migration times are recorded. Figure 4.1 shows an illustration of the operation.

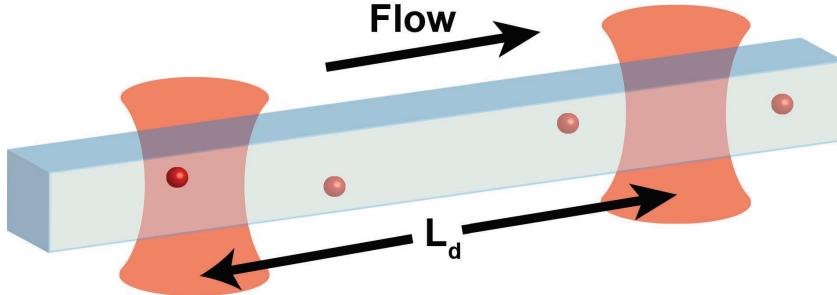


Figure 4.1: Schematic illustrating the operating principle of mobility-SCAN. Single molecule electrophoretic mobilities are determined by measuring the migration time between two observation volumes separated by a detection length, L_d .

Currently, our system allows for 2-color mobility detection for adjustable detection lengths up to $570\ \mu\text{m}$ and migration times nearing 1 second. The combination of 2-color spectral and mobility data provides a plethora of new information as conditional relationships can be investigated leading to the ability to analyze complex mixtures. Furthermore, the adjustable nature of the detection length allows for maximal mobility resolution as there was an observed detection length trade-off between ideal diffusion limited behavior and non-ideal

broadening caused by wall interactions and other phenomena. The efficiency of the technique as measured by theoretical plates was on par with microchip capillary electrophoresis with the efficiency being effectively limited the magnitude of the applied voltage and by dielectric breakdown due to the device design. We achieved theoretical plates over 12,000 for DNA and demonstrated the capability of detecting single dye variations for certain near-neutral molecules giving the ultimate resolution of mobility detection with fluorophores.

The system was shown to be capable of analyzing a diverse set of molecules including DNA, chromatin, dyes, and proteins, including some slightly cationic proteins which was somewhat surprising given the negatively charged silanol groups on the fused silica. Further, the technique was used to determine physical parameters such as electrophoretic mobilities, diffusion coefficients, effective molecular charge, and the Stokes radius using single molecule ensemble averaging.

A key feature of the device proved to be its ability to differentiate molecules based on mobility. In one demonstration of the device's potential, we took 6 different biomolecules (DNA, proteins, and dye) and were able to identify 83% of the molecules correctly based on mobility alone within the complex mixture. For many of the species, we had complete separation in mobility distributions. By identifying molecules based on mobility instead of color, we would enable a reduction of the number of distinct fluorophores needed for experiments and free up precious emission spectra for other labeled biomolecules.

Finally, the device was used to detect binding interactions of various complexes similar to affinity capillary electrophoresis, but on a single molecule level. We envision that binding induced mobility shifts could be used to detect un-

labeled antigens on a single molecule level, perhaps allowing for detection of scarce antigens not detected by other means.

Although we have discussed the benefits and results to the mobility-SCAN approach above, we would like to end Section 4.1 by explicitly putting forth certain questions relevant to SCAN to which were only answered after mobility measurements. They include:

1. What is the relative amount of chromatin to free DNA in our samples?
2. Is chromatin significantly underlabeled compared to DNA of the same size, presumably due to nucleosomes interfering with intercalation?
3. Why haven't we observed an intensity banding pattern for chromatin using SCAN?
4. Can we use mobility based methods to detect binding of epigenetic marks?
5. How contaminated are our samples? Are we counting molecules which aren't desired?
6. Can we detect unlabeled biomolecules using mobility shifts detected by SCAN?
7. Why has it been more difficult to control chromatin molecules using SCAN-sort relative to DNA?
8. Why have coincidence rates been subject to the direction of flow?
9. Can we determine diffusion coefficients, effective molecular charge, zeta potentials, electrophoretic mobilities and electroosmotic mobilities on a single molecule level and are they consistent with bulk measurements?

Answers to these questions and more will be elucidated in the following results.

4.2 Theoretical Considerations

Fundamentally, our mobility experiments consist of measuring transit times of single molecules undergoing drift and a random walk and producing ensemble based distributions to infer molecule identities. Ideally, the dynamics of each individual molecule can be modeled using the Fokker-Planck stochastic differential equation given by

$$dX_t = (\mu_{eo} + \mu_{ep})dt + \sigma_D dW_t \quad (4.1)$$

where W_t is a Wiener process (Brownian motion). The Fokker-Planck equation describes the location of a particle at X_t undergoing drift and Brownian motion with a diffusion coefficient $D = \sigma_D^2/2$. After application of Ito's lemma, we find the Fokker-Planck equation for probability is given by

$$\frac{\partial p}{\partial t} = D \frac{\partial^2 p}{\partial x^2} - (\mu_{eo} + \mu_{ep}) \frac{\partial p}{\partial x} \quad (4.2)$$

which is a standard drift-diffusion equation. The corresponding Green's function to Equation (4.2) is

$$p(x, t) = \frac{1}{\sqrt{4\pi Dt}} \exp\left[\frac{[x - (\mu_{eo} + \mu_{ep})t]^2}{4Dt}\right] \quad (4.3)$$

which is the distribution of the migration times of the molecules. In some respects with our bottom-up approach, we are, in essence, trying to experimentally confirm the mathematical models of Fokker-Planck. Before we can do this, however, we must first discuss certain electrokinetic phenomena in the nanochannels and also some concepts from separation science.

4.2.1 Electrokinetic Phenomena in Nanochannels

The single molecule transport in our channels predominately depends on two electrokinetic phenomena: electroosmosis and electrophoresis. Electroosmosis is the motion of fluid in the presence of an external electric field. Electrophoresis is the movement of particles relative to a fluid due to an external electric field. Fundamentally, electroosmosis and electrophoresis are the same physical phenomena differing only by the frame of reference. Typically, we describe electroosmosis in a frame relative to a motionless solid body. As an experimenter, however, we measure the motion of particles relative to a solid body with the fluid moving. Both processes are dependent on the presence and relative size of the electric double layers (EDL) which, along with the surface charge density on the given solid, couples the applied electric field to the particle motion and fluid flow.

Electrical Double Layer

In general, any charged surface whose boundary is an electrolyte solution will create an EDL in the solution localized at the boundary. These double layers consist of enrichment of oppositely charged counterions and a depletion of like-charged coions relative to bulk. Since the EDL has a non-zero charge density, a double layer potential exists with a characteristic decay length governed by the Debye length, λ . The Debye length is a characteristic of the solution and is given by

$$\lambda = \sqrt{\frac{\varepsilon RT}{2F^2 I_c}} \quad (4.4)$$

where I_c is the ionic strength of the bulk solution given by

$$I_c = \frac{1}{2} \sum_i z_i^2 C_{i\infty} \quad (4.5)$$

The Debye length appears naturally in the Gouy-Chapman model of the charges in the diffuse EDL as will be shown subsequently.

At equilibrium and under the assumption of the ions being point charges, the ion concentrations in the EDL obey a Boltzmann distribution given by

$$C_i = C_{i\infty} \exp\left(\frac{-z_i F \Phi}{RT}\right) \quad (4.6)$$

Combining the Boltzmann relation in Eq. (4.6) and the Poisson's equation gives the Poisson-Boltzmann equation

$$\nabla^2 \Phi = -\frac{F}{\varepsilon} \sum_i z_i C_{i\infty} \exp\left(-\frac{z_i F \Phi}{RT}\right) \quad (4.7)$$

which governs the mean field structure of the diffuse EDL. The nonlinearity of Eq. (3.17) prompts the expansion of the exponential term which leaves

$$\nabla^2 \Phi = -\frac{F}{\varepsilon} \sum_i z_i C_{i\infty} \left(1 - \frac{z_i F \Phi}{RT} + \dots\right) \quad (4.8)$$

Since the bulk solution is assumed to be electroneutral, Eq. (4.8) simplifies to

$$\nabla^2 \Phi = \frac{\Phi}{\lambda^2} \quad (4.9)$$

which is valid for $|z_i F \Phi / RT| \ll 1$. This approximation is denoted as the Debye-Huckel approximation. Solving Equation (4.9) for the double layer potential for a prescribed geometry gives decay profile dependent on λ .

As we have briefly discussed, the Debye length is the characteristic length scale relevant to our nanochannels and their behavior. If the Debye length is comparable to the critical dimension of the nanochannels, then more complexity is involved as the EDL becomes part of the bulk flow. For our system, the

standard SCAN running buffer consists of 10 mM Tris, 1 mM EDTA, 0.5% (w/v) PVP, and 0.1% (v/v) Triton X-100. PVP is a neutral, yet polar polymer and Triton X-100 is a nonionic surfactant; both don't contribute to the ionic strength of the solution. Tris, as a buffering agent, has $pK_a = 8.07$ at 25°C which is in the neighborhood of the pH of the running buffer. Using the Henderson-Hasselbalch equation, we calculate that $\approx 27\%$ of the Tris consists of its deprotonated conjugate base. EDTA has pK_a values at 2.0, 2.7, 6.2, and 10.3 and is essentially trivalent at neutral pH. Using Eq. (4.4) and Eq. (4.5), we calculate that $\lambda = 2.5$ nm for the running buffer. This is 100 times smaller than the critical dimension of our nanofluidic channels which implies that our nanochannel device behaves as a microscale device in that the EDL can be separated from the bulk flow by boundary-layer theory.

Electroosmosis

While the EDL is small compared to the channel dimensions, it still plays an important role as it generates electroosmotic flow (EOF) in the device. The ions in the EDL are mobile and it is this mobility that generates electroosmosis. In the EDL, the momentum equation for the fluid is given by

$$0 = -\nabla p + \eta \nabla^2 \vec{u} + \rho_e \vec{E} \quad (4.10)$$

Using Poisson's equation, we have

$$0 = -\nabla p + \eta \nabla^2 \vec{u} + \epsilon \nabla^2 \Phi \nabla \Phi \quad (4.11)$$

The potential can be decomposed into an equilibrium potential, Φ_{eq} , governed by the Poisson-Boltzmann equation and a nonequilibrium perturbation poten-

tial, Φ_{ne} , due to EDL distortion as

$$\Phi = \Phi_{eq} + \Phi_{ne} \quad (4.12)$$

For an undistorted EDL, Poisson's equation

$$\nabla^2 \Phi = -\frac{(\rho_e)_{eq}}{\epsilon} = \nabla^2 \Phi_{eq} \quad (4.13)$$

which implies that

$$\nabla^2 \Phi_{ne} = 0. \quad (4.14)$$

For an inert planar surface with $u_x = u_x(y)$ and a constant tangential applied electric field E_x , we have a flow with the double layer being undistorted due to the symmetries of the geometry and flow. It is apparent from Eq. (4.14) that up to an arbitrary constant that $\Phi_{ne} = -E_x y$. With the assumption that flow is pressure free, the momentum equation in the x -direction

$$\frac{d^2 u_x}{dy^2} = \frac{\epsilon E_x}{\eta} \frac{d^2 \Phi_{eq}}{dy^2} \quad (4.15)$$

subject to the boundary conditions

$$u_x(0) = 0, \quad \Phi(0) = \zeta, \quad \frac{du_x}{dy}(\infty) = 0, \quad \frac{d\Phi_{eq}}{dy}(\infty) = 0 \quad (4.16)$$

After a trivial integration of Eq. (4.15) and application of the boundary conditions (4.16), we arrive at

$$u_x(y) = -\frac{\epsilon E_x}{\eta} (\zeta - \Phi_{eq}(y)) \quad (4.17)$$

where $\Phi(y)$ is determined by solving some form of the Poisson-Boltzmann equation (linear, nonlinear, or modified) depending on the desired accuracy needed for the model. Using the Debye-Huckel approximation to maintain an analytically tractable equation, the flow profile is given by

$$u_x(y) = -\frac{\epsilon E_x \zeta}{\eta} (1 - \exp(-y/\lambda)) \quad (4.18)$$

As Eq. (4.19) illustrates, EOF has essentially constant flow profile beyond a couple Debye lengths away from the wall as the EDL decays into the bulk solution. While this derivation used the Debye-Hückel approximation, more accurate models using the nonlinear Poisson-Boltzmann equation or modifications thereof generate slightly sharper decay profiles for the double layer potential. In any case, for the experiments discussed in this manuscript where the thin EDL limit is always applicable, the flow can be considered as a plug flow with a velocity given by

$$u_x = -\frac{\varepsilon \zeta E_x}{\eta} \quad (4.19)$$

Furthermore, the velocity in the bulk of an arbitrary shaped channel will follow the same expression (4.19) provided the EDL is small compared to the transverse channel dimensions and the curvature of the walls. The flat flow profile enhances the mobility resolution of our devices relative to pressure driven Poiseuille flow since it doesn't exhibit Taylor dispersion.

Electrophoresis

Electrophoresis is the motion of charged in an electrolyte solution due to an applied electric field. At a fundamental physical level, electrophoresis and electroosmosis are the same phenomena, only differing by the choice of reference frame. Derived by Smoluchowski, the electrophoretic velocity of a particle is given by

$$u_x = \frac{\varepsilon \zeta E_x}{\eta} \quad (4.20)$$

which is applicable for particle shape as long as the EDL is sufficiently thin relative to the radius of curvature of the particle. If the EDL isn't thin, then the electric field can no longer be assumed uniform along the normal coordi-

nate causing a decreased mobility. In the thick EDL limit, the electrophoretic velocity is two-thirds that of Equation (4.20). Between the two extremes, the electrophoretic velocity is monotonic and described by Henry's function.

While expressing the electrophoretic mobility in terms of the ζ -potential is theoretically compelling, we also would like to relate the mobility through an effective charge and the Stokes radius of the particle to understand intuitively the relationships of molecule size and charge to the electrophoretic mobility. Ignoring the effect of counterions in the EDL, at steady state in a Stokes flow, the hydrodynamic drag force will be balanced by the Coulomb force on the particle. Therefore, we have for a spherical biomolecule of radius R

$$u_x = \frac{qE_x}{6\pi\eta R} \quad (4.21)$$

Equation 4.22 serves as a first approximation to the electrophoretic mobility as the effects of the EDL and the shape of the particle are not taken into account. The effects of the counterions tend to slow the particle down as they move in the opposite direction of the particle causing electrophoretic retardation. A more detailed analysis by Henry [14] gives

$$u_x = \frac{qE_x}{4\pi\eta R(1 + \kappa R)} H(\kappa R) \quad (4.22)$$

where κ is the inverse Debye length and H is the previously noted Henry's function which monotonically varies from $2/3$ to 1 . For small globular proteins in biocompatible buffers such as PBS, the value of κR is typically between 0.1 and 10 . Henry's function varies rapidly in this range implying that the electrophoretic mobility will be strongly dependent on electrolyte concentrations. Due to this and other difficulties, electrophoretic mobilities of complex molecules such as proteins are generally determined experimentally and related to semi-empirical formulas dependent on their size and shape [15]. The shape

of the molecule plays an important role in the mobility as globular molecules will migrate faster than elongated, fibrous molecules.

4.2.2 Migration Time, Theoretical Plates, and Peak Broadening

Fundamentally, single molecule mobility experiments simply involve measuring individual molecule's migration time between two inspection volumes and creating inferences regarding such things as molecule identity based on ensemble behavior. Factors that influence the average migration time and the migration time dispersion will be the subject of the following section.

Migration Time The migration time taken by a molecule to traverse a distance L_d between detection points is given by

$$t_m = \frac{L_d}{u_{eo} + u_{ep}} = \frac{L_d}{(\mu_{eo} + \mu_{ep})|\vec{E}|}, \quad (4.23)$$

where $|\vec{E}|$ is the magnitude of the electric field in the channel. Since the electric field is simply the channel voltage divided by its length, $|\vec{E}| = V_c/L_c$, and the channel voltage is related to the applied voltage by $V_c = \alpha V_a$ where $0 < \alpha < 1$ is a numerical factor dependent solely on the geometry of the device, we have

$$t_m = \frac{L_d L_c}{(\mu_{eo} + \mu_{ep})\alpha V_a} \quad (4.24)$$

The magnitude of α for our $50 \mu\text{m}$ device was $\alpha = 0.33$ and our long U-shaped device had $\alpha = 0.908$. Derivation of these prefactors are given in Section 4.3.3. From Equation (4.24), we see that mean electrophoretic mobility can be determined by ensemble averaging the single molecule measurements with respect

to the inverse migration times:

$$\bar{\mu}_{ep} = \frac{L_d L_c}{\alpha V_a} \left(\frac{1}{t_m} \right) - \bar{\mu}_{eo} \quad (4.25)$$

The mean electroosmotic mobility can be determined by using a neutral tracer dye as explained in Section 4.3.6.

Due to the functional form of Equation (4.24), the electrophoretic mobility measurement resolution is nonlinear. By resolution, we are not talking about resolving two different bands, but instead on $\Delta t_m / \Delta \mu_{ep}$ holding all other experimental parameters constant. These two concepts of resolution are, however, intertwined and dependent on one another. Partial differentiation of Equation (4.24) with respect to μ_{ep} , we find

$$\Delta t_m = - \frac{L_d L_c}{(\mu_{eo} + \mu_{ep})^2 \alpha V_a} \Delta \mu_{ep} \quad (4.26)$$

Therefore, all else being equal, very small differences in electrophoretic mobility can be resolved when $\mu_{eo} + \mu_{ep}$ are equal and opposite. The obvious drawback would be extremely long migration times demonstrating the tradeoff between mobility resolution and migration times.

We should also consider the effect of charge addition to a molecule's mobility for two cases. For the first case, consider $|\mu_{eo}| \gg |\mu_{ep}|$ which implies that the molecule is near its isoelectric point. The resolution of the measurement is then

$$\Delta t_m \approx - \frac{L_d L_c}{(\mu_{eo}^2 + 2\mu_{eo}\mu_{ep})\alpha V_a} \Delta \mu_{ep} = - \frac{L_d L_c}{(6\pi\eta R\mu_{eo} + 2q)\mu_{eo}\alpha V_a} \Delta q \quad (4.27)$$

Thus, the maximum resolution will occur when R and q are both small, all else equal. We achieved single dye mobility resolution using the nearly neutral wheat germ agglutinin by following this principle as we will show in Section 4.4.7

Similarly, consider the case when $|\mu_{ep}| \gg |\mu_{eo}|$ which is reasonable in some cases where surface coatings suppress the EOF. Using Equations (4.22), (4.26) and assuming R is constant, we find

$$\Delta t_m = -\frac{6\pi\eta RL_d L_c}{q^2 a V_a} \Delta q \quad (4.28)$$

Again, resolution will be best for molecules of less charge. However, for molecules of little charge, the assumption that $|\mu_{ep}| \gg |\mu_{eo}|$ would probably be invalid.

Individual molecule migration times are typically compiled into a electrophoretogram displaying the ensemble migration time distribution. As noted earlier, the functional form of the ensemble migration time distribution takes the form

$$p(t_m) = \frac{1}{\sqrt{2\pi}\sigma_T} \exp\left[\frac{[L_d - (\bar{u}_{eo} + \bar{u}_{ep})t_m]^2}{2\sigma_T^2}\right] \quad (4.29)$$

Since molecules can not have a migration time less than zero, migration time distributions will be slightly asymmetric for short migration times or large band variance σ_T as Equation (4.29) illustrates. An analytical estimation of the band asymmetry can be determined if the system is diffusion limited. Inserting $\sigma_T^2 = 2Dt_m$ into Equation (4.29) and performing elementary calculations for peak maximum and average gives

$$t_{m,max} = \frac{L_d}{\bar{u}Pe} \left(\sqrt{1 + Pe^2} - 1 \right) \quad (4.30)$$

and

$$\bar{t}_m = \frac{L_d}{\bar{u}Pe} \left(1 + \frac{2}{Pe} \right) \quad (4.31)$$

where $Pe = \bar{u}L_d/D$ is the Peclet number giving a relative measure of advection to diffusion and $\bar{u} = \bar{u}_{eo} + \bar{u}_{ep}$. As an example, consider now a dye which typically has a diffusion coefficient of around $4 \times 10^{-6} \text{ cm}^2/\text{s}$. We are considering

a dye since they will be the molecule with the largest diffusion coefficient and hence, the smallest Peclet number used in SCAN. Using standard SCAN parameters, we find a Peclet number of approximately 150. Plugging into Equations (4.30) and (4.31), we estimate a 2% deviation from peak maximum for the peak average. Thus, we find that peak asymmetry shouldn't be significant issue for diffusion limited system. However, when other factors such as wall interactions increase σ_T beyond that of diffusion, band asymmetries will become highly noticeable in the migration time distribution.

Theoretical Plates and Peak Broadening The efficiency of the single molecule mobility measurement in distinguishing different species can be defined using the concept of theoretical plates analogous to chromatographic separations. The concept of theoretical plates is defined as

$$N = \frac{L_d^2}{\sigma_T^2}, \quad (4.32)$$

where σ_T^2 is the total variance of the band and L_d is the length between molecule detection spots. Using this definition, we can see that separation efficiency increases as the number of theoretical plates increase due to reductions in total variance. In essence, theoretical plates is an indirect measure of the total variance in the zone length. Experimentally, the number of theoretical plates can be determined by the equivalent temporal relationship given by

$$N = 5.545 \left(\frac{t_m}{W_{FWHM}} \right)^2 \quad (4.33)$$

where W_{FWHM} is the band width at the full width at half maximum and the numerical prefactor stems from the relationship between σ and FWHM for a Gaussian function.

Fluctuations about the mean migration time given in Equation 4.24 will occur due to diffusion and non-ideal experimental factors causing band broadening. The total variance for a peak can be generally modeled by

$$\sigma_T^2 = \sigma_D^2 + \sigma_{det}^2 + \sigma_{bin}^2 + \sigma_{molecule}^2 + \sigma_{wall}^2, \quad (4.34)$$

where account for diffusion, detector variance, data binning quantization, inherent molecule mobility variances caused by things such as dye labeling fluctuations, and wall interactions respectively. Other dispersion mechanism typically seen in capillary electrophoresis such as injection variance, thermal Joule heating dispersion, EOF variation, and electromigration dispersion [16] are nonexistent or negligible for mobility-SCAN.

Diffusion broadening is well-known to be given by

$$\sigma_D^2 = 2Dt_m, \quad (4.35)$$

where t_m is given in Equation (4.24) and D is the diffusion coefficient of the molecule. Therefore, in a diffusion limited system, the number of theoretical plates is given by

$$N_D = \frac{(\mu_{ep} + \mu_{eo})\alpha V_a L_d}{2DL_c}, \quad (4.36)$$

Equation (4.36) illustrates that separation efficiency varies linearly with applied voltage and the distance between detection spots and thus illustrating why capillary electrophoresis is operated at high voltages and long detection lengths. Moreover, if diffusion dominates the peak broadening, then the diffusion coefficient can be experimentally determined by a straightforward mobility measurement.

The detector variance accounts for the beam widths of the detection scheme. For a square detector window of width, W , the single detector variance is given

by [17]

$$\sigma_{det}^2 = \frac{W^2}{12} \quad (4.37)$$

Since we have two detectors, the total contribution is $\sigma_{det}^2 = W^2/6$. Over the course of the experiments, we have typically found the detector variance to be small or negligible.

The data binning quantization variance accounts for the dispersion caused by time-binning the data. The variance is given by

$$\sigma_{bin}^2 = \frac{1}{12} \left(\frac{L_d}{t_m} \right)^2 \Delta t^2 \quad (4.38)$$

where Δt is the bin width. Equation (4.38) can be written in terms of experimental parameters using Equation (4.24) as

$$\sigma_{bin}^2 = \frac{1}{12} \left(\frac{(\mu_{eo} + \mu_{ep})\alpha V_a}{L_c} \right)^2 \Delta t^2 \quad (4.39)$$

The binning variance will obviously become more influential on the total variance as the migration time decreases. This gives another reason to use longer detection lengths.

The molecular dispersion accounts for the intrinsic band broadening due to fluctuations caused by things such as dye labeling or other charge related effects. For example, if we were processing SCAN data on a molecule whose functional group's pK_a equaled the pH, then we would observe band broadening caused by the single molecules being in various charge states. The dispersion caused by this effect should be proportional to L_d^2 and independent of the applied voltage.

Finally, because of the high surface-to-volume ratio of the nanodevices, surface interactions are expected to play a dominate role especially electrostatically

when the biomolecules are nearly neutral or positively charged relative to the negatively charged fused silica. Derived from Minarik et al [18], the contribution to band broadening for adsorption/desorption events in a capillary is given by

$$\sigma_{int}^2 = \left(\mu_{ep} + \mu_{eo} \right) \left(\frac{K^2}{D} \frac{r_c}{r_c + 2K} + \frac{4K}{(r_c + 2K)k_d} \right) \frac{\alpha V_a L_d}{L_c} \quad (4.40)$$

where $K = k_a/k_d$ and r_c is the radius of the capillary. For a given electric field, these surface interactions cause a broadening that is proportional to the applied voltage and also the separation distance between detectors. The nonlinear nature of the adsorption/desorption process not only shifts the peak centers, but can also cause significant band distortion with tailing towards greater migration times.

As previously mentioned, the total variance lacks terms traditionally seen in capillary electrophoresis including injection plug variation, thermal dispersion, and electromigration effects. The lack of a sample plug with regards to the single molecule measurement negates the peak broadening by this mechanism. Further, thermal effects are negligible in our system due to low currents inherent in devices which have resistances of greater than $1\text{ G}\Omega$. Our devices operate at currents typically less than 50 nA which implies a heat production of approximately $2\text{ }\mu\text{W}$ or less which can be readily dissipated through the device substrate. In capillary electrophoresis, the heat production causes band broadening as the electrophoretic and electroosmotic mobilities are temperature dependent via viscosity. The temperature within a capillary can be both spatially and temporally dependent. Spatial temperature gradients will exist across the radius of the capillary due to heat dissipation at the capillary surface. These spatial gradients modify the characteristic flat flow profile of the EOF to a more parabolic shape and generate Taylor dispersion. Temporal temperature

variations may exist due to processing samples prior to reaching a steady state. The amount of heat generated per unit time per unit length of capillary is

$$\frac{dE_g}{dt} = \frac{\sigma A V_c^2}{L_c^2} \quad (4.41)$$

which is a direct application of Ohm's law and implying lower cross-sectional area such as nanochannels would be beneficial regarding mitigating heat production. Experimentally, if heat dissipation was an issue, single molecule migrations times would be a non-linear function of voltage which has never been observed with our system. Interestingly, these non-ideal effects generally contribute to such an extent that capillary zone electrophoresis can never be considered diffusion-limited in many cases [17]. In contrast, with mobility-SCAN, we do have a system that has the capabilities for diffusion-limited migrations with which we can determine the diffusion coefficient of the migrating species.

Diffusion Coefficient Determination by Varying the Applied Voltage From Equation (4.24) a linear regression of the migration time as a function of V_a^{-1} gives

$$\beta_{m,V} = \frac{L_d L_c}{(\mu_{eo} + \mu_{ep})\alpha} \quad (4.42)$$

with $t_m = \beta_{m,V} V_a^{-1}$ and $\beta_{V,m}$ is a fitting parameter. Therefore, after a simple rearrangement, we have

$$\mu_{eo} + \mu_{ep} = \frac{L_d L_c}{\alpha \beta_{m,V}} \quad (4.43)$$

Likewise, a theoretical plate fitting parameter can be defined using Equation (4.36):

$$\beta_{N,V} = \frac{(\mu_{ep} + \mu_{eo})\alpha L_d}{2 D L_c} \quad (4.44)$$

Inserting Equation (4.47) into Equation (4.48) and rearranging we find that the diffusion coefficient for a diffusion limited system is given by

$$D_V = \frac{L_d^2}{2\beta_{m,V}\beta_{N,V}} \quad (4.45)$$

Diffusion Coefficient Determination by Varying the Detection Length From Equation (4.24) a linear regression of the migration time as a function of L_d gives

$$\beta_{m,L} = \frac{L_c}{(\mu_{eo} + \mu_{ep})\alpha V_a} \quad (4.46)$$

with $t_m = \beta_{m,L}L_d$ and $\beta_{m,L}$ is a fitting parameter. Rearranging, we have

$$\mu_{eo} + \mu_{ep} = \frac{L_d L_c}{\beta_{m,L} \alpha V_a} \quad (4.47)$$

A theoretical plate fitting parameter can be specified using Equation (4.36):

$$\beta_{N,L} = \frac{(\mu_{ep} + \mu_{eo})\alpha V_a}{2DL_c} \quad (4.48)$$

Inserting Equation (4.47) into Equation (4.48) and rearranging we find that the diffusion coefficient for a diffusion limited system is given by

$$D_L = \frac{1}{2\beta_{m,L}\beta_{N,L}} \quad (4.49)$$

Effective Charge of the Molecule through the Nernst-Einstein Relation The molecule's charge can be deduced via the Nernst-Einstein relation

$$z = \frac{RT\mu_{ep}}{FD}, \quad (4.50)$$

where R is the ideal gas constant, T is the temperature, and F is Faraday's constant. Using mobility-SCAN, we can experimental determine both the diffusion coefficient and the electrophoretic mobility which implies that we know

the effective charge state. Note that unlike the diffusion coefficient measurement which only needs a direct measurement of migration times for the analyte of interest, the molecule's charge requires μ_{ep} which implies that an EOF measurement would also need to be taken. Failure of the Nernst-Einstein relation occurs, most notably for DNA and other polyelectrolytes where the point charge approximation breaks down due to the free-draining nature of the molecule, and thus careful application of Equation 4.50 is needed.

4.3 Materials and Methods

4.3.1 Device Fabrication

The nanochannels were made in a 4" UV fused silica wafer (Mark Optics; Santa Ana, CA) with a thickness of 500 μm . To start the process, DUV42P ARC anti-reflective coating (Brewer Science; Rolla, MO) was spun at 2500 rpm for 60 seconds giving an ARC thickness of 60 nm and followed by a 205 °C proximity bake for 60 seconds. The UV210-0.6 DUV positive photoresist (Microchem; Westborough, MA) was then spun at 2000 rpm for 60 seconds and baked on a proximity hotplate for 60 seconds at 135 °C. The resulting resist thickness was approximately 650 nm. The wafers were subsequently exposed (30 mJ/cm²) using the ASML 300C DUV stepper and post exposure baked for 90 seconds on a proximity hotplate.

The ASML 300C DUV stepper gave better reproducibility and allowed for smaller feature sizes relative to the Autostep 200 i-line stepper which was used by previous generations of the Craighead lab to manufacture devices. The crit-

ical dimension employed in the devices was 250 nm which occurred in the upstream filters. Having the filter size smaller than the channel dimensions greatly reduced clogging, especially in the long mobility channels.

The exposed resist was developed using AZ 726MIF developer (MicroChemicals; Ulm, Germany). Prior to etching, the wafers were descummed using 150 W O₂ plasma. The channels were etched by an Oxford 80 reactive ion etcher in two steps. First, the ARC was removed using an Ar/O₂ etching chemistry. The fused silica was then etched to approximately 300 nm using CHF₃. The CHF₃ etch chemistry was necessary for the thin DUV resist as it provided a high selectivity. However, the inadequately low F/C ratio caused polymerization of the photoresist which was consequently removed by a barrel ashler.

Prior to sandblasting the holes that connect the channels to the reservoirs, the wafers were protected by spinning MICROPOSIT S1818 resist (Dow; Midland, MI) on both side of the wafer to a thickness of roughly 2 μ m. The resist coated wafers were then subjected to sandblasting in the predefined reservoir areas. Before bonding, the wafers were cleaned of sandblasting debris and other contaminants using the following chemicals: IPA, MICROPOSIT Remover 1165 (Dow; Midland, MI), NMP (1-Methyl-2-pyrrolidone), and piranha (H₂SO₄/H₂O₂ in 3:1 mixture). The final processing step before the wafers were bonded was the RCA clean:

1. RCA-1: NH₄OH/H₂O₂/H₂O - Organic residue removal
2. RCA-2: HCl/H₂O₂/H₂O - Metallic ion removal

Following the RCA processing, the cleaned wafers were directly bonded to 170 μ m UV fused silica wafer (Mark Optics). Barring excessive contamination,

the two glass wafers formed a semi-permanent bond through van der Waals' and hydrogen bonding. A permanent covalent fusion bond was obtained by annealing the wafers at 1085 °C for 6 hours. Following annealing, fluid reservoir ports were attached to the wafer using NOA63 optical adhesive (Norland Products).

4.3.2 Device Layout

Two different device designs were used for the biomolecule mobility measurements both of which were ultimately constrained by the emission optics. One design, which used a U-shaped channel architecture, was configured to maximize the separation efficiency and allowed for mobility measurements at different lengths on a continuum. The other layout was a 50 μm straight nanochannel device. The straight channel device had a much lower failure rate due to clogging. Clogging was caused by contamination and debris from manufacturing and sample absorption, both of which are dependent on the nanochannel length.

Additionally, for each device, filter arrays were implemented upstream and downstream of the nanochannel to filter out aggregates, large contaminants, and debris from entering the channels. While channel clogging was still an issue, especially for the long U-shaped devices, it was mitigated with the filters.

U-Shaped Channel Device

The U-shaped channel device consisted of 5 parallel U-shaped nanochannels connected in series with $10\ \mu m$ wide input/output channels which were themselves connected to loading reservoirs. A 10x optical micrograph of the device is shown in Figure 4.2.

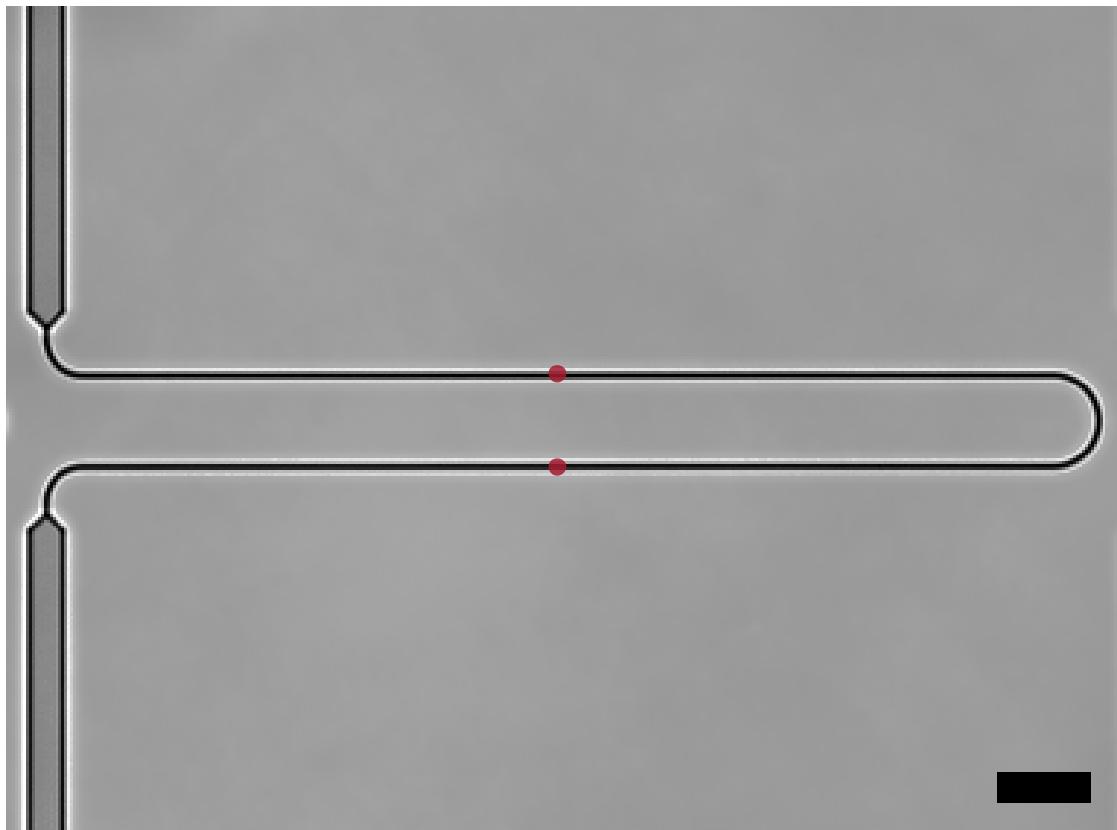


Figure 4.2: U-shaped nanochannel for single molecule mobility measurements. The red dots illustrate a possible beam spot location. By translating along the horizontal axis, measurements at different channel lengths can be obtained. The scale bar is $25\ \mu m$.

The U-shaped channel layout facilitated mobility measurements at various lengths from $L_d = 40\ \mu m$ to $L_d = 570\ \mu m$ by simply translating a motorized microscope stage. The minimum length corresponded to a measurement around

a semicircle of radius $r = 12.5 \mu\text{m}$. Measurement locations were determined to within $\pm 0.5 \mu\text{m}$ by using the electrical readout from the stage and translating from known locations such as the nanochannel endpoints or the semicircle endpoints. The geometrical parameters of the nanochannel are given in Table 4.1.

Parameter	Variable	Value
Channel length	L_c	$590 \mu\text{m}$
Channel width	W_c	500 nm
Channel height	H_c	250 nm
Detection length	L_d	$40\text{-}570 \mu\text{m}$
Distance between parallel channel segments	d_{\parallel}	$25 \mu\text{m}$

Table 4.1: U-shaped nanochannel dimensions

As noted before, the device also consisted of loading reservoirs for connecting sample ports and I/O channels, both embedded in the fused silica. The $10 \mu\text{m}$ wide I/O channels can be seen on the far left of Figure 4.2. Their dimension are given in Table 4.2.

Parameter	Variable	Value
I/O channel length	$L_{c,10\mu\text{m}}$	$925 \mu\text{m}$
I/O channel width	$W_{c,10\mu\text{m}}$	$10.0 \mu\text{m}$
I/O channel height	$H_{c,10\mu\text{m}}$	250 nm
Reservoir length	L_r	5.00 mm
Reservoir width	W_r	1.50 mm
Reservoir height	H_r	250 nm

Table 4.2: Input/output channels and reservoir dimensions

Voltage Constraints of the U-Shaped Channel Device The geometry of the channel and the dielectric strength of fused silica puts an upper bound on the

applied voltage which, in turn, limits the efficiency of the device. Due to the pre-defined distance of $1500\text{ }\mu\text{m}$ between the output optical fibers, the separation between the parallel nanochannel segments of the U-shape was constrained to $25\text{ }\mu\text{m}$ using a $60X$ objective. This subsequently put a limit on the maximum voltage that could be applied to the device before dielectric breakdown occurred. Typical values for the fused silica's dielectric strength are $10\text{-}40\text{ V}/\mu\text{m}$. For a $25\text{ }\mu\text{m}$ gap, this corresponds to a lower bound of 250 V . Experimentally, we determined that the devices largely remained operational when conducted at $V_a \leq 225\text{ V}$. For $V_a > 225\text{V}$, device failure caused by dielectric breakdown almost surely occurred. Thus, with an added margin of safety, all ensuing mobility measurements were limited to $V_a \leq 200\text{ V}$ for the U-shaped devices.

Straight Channel Device

The straight channel device consisted of 20 parallel nanochannels connected in series with $10\text{ }\mu\text{m}$ wide input/output channels and loading reservoirs. Other than the nanochannel geometry and the addition of 15 extra channels, all other dimensions were the same as the U-shaped device. An optical micrograph of the nanochannel is shown in Figure 4.3.

Table 4.3 give the nanochannel dimensions of the straight device. Note that the straight device is over 11 times shorter than the U-shaped channels.

Filter Arrays For each device, filter arrays were implemented upstream and downstream of the nanochannels to filter out aggregates, large contaminants, and debris from entering the channels. Prior to this implementation, channels would clog with great frequency especially when processing chromatin sam-

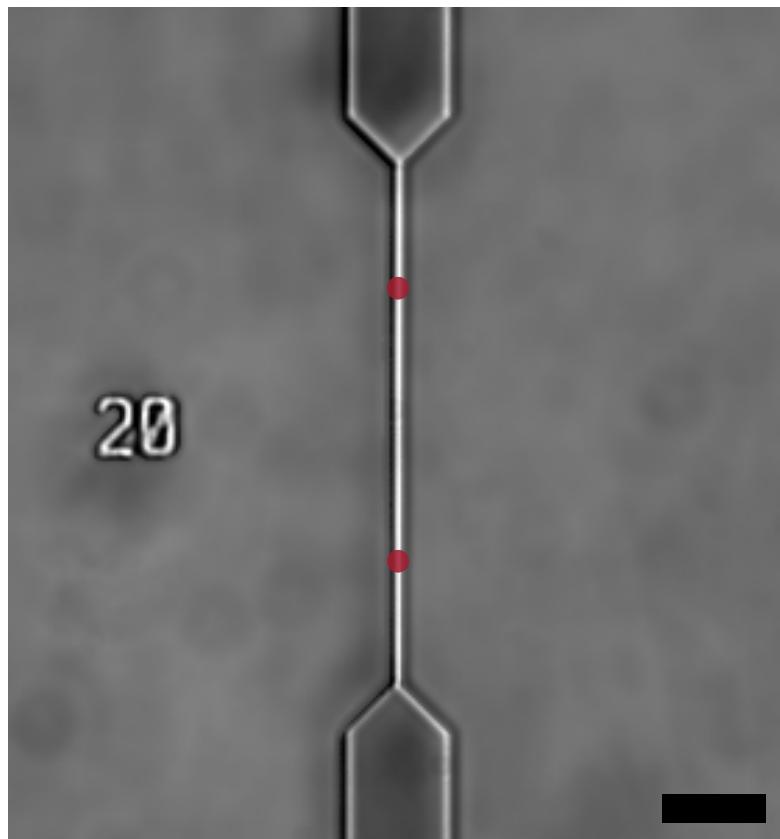


Figure 4.3: $50\text{ }\mu\text{m}$ nanochannel for single molecule mobility measurements. The red dots illustrate a the typical beam spot location. The scale bar is $10\text{ }\mu\text{m}$.

Parameter	Variable	Value
Channel length	L_c	$50\text{ }\mu\text{m}$
Channel width	W_c	500 nm
Channel height	H_c	250 nm
Detection length	L_d	$25\text{ }\mu\text{m}$

Table 4.3: Straight nanochannel dimensions

ples. For example, some SCAN-sort experiments, which only had one channel, were limited to 15 minutes due to clogging.

The filter arrays consisted of three linear filters with decreased spacing (void volumes) as one progressed towards the nanochannel. The array spanned the loading reservoirs. The spacings were $5\text{ }\mu\text{m}$, $1\text{ }\mu\text{m}$, and $0.5\text{ }\mu\text{m}$ respectively.

Figure 4.4 shows the filter array embedded within the device with Table 4.4 giving the relevant dimensions.

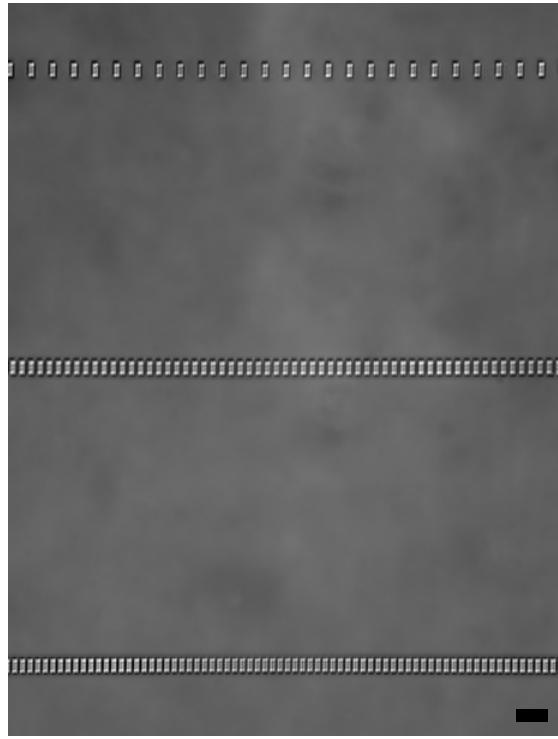


Figure 4.4: Filter Array. The scale bar is $10 \mu\text{m}$.

Parameter	Variable	Value
Filter length	L_{fil}	$5.0 \mu\text{m}$
Filter height	H_{fil}	250 nm
$5 \mu\text{m}$ filter width	$W_{fil,5\mu\text{m}}$	1.07 mm
$1 \mu\text{m}$ filter width	$W_{fil,1\mu\text{m}}$	$500 \mu\text{m}$
$0.5 \mu\text{m}$ filter width	$W_{fil,0.5\mu\text{m}}$	$300 \mu\text{m}$

Table 4.4: Filter dimensions for three filters. The filter width gives the total width available for fluid occupation across the entire reservoir.

While channel clogging was still in issue, especially for the long U-shaped devices, it was lessened with these filters arrays.

4.3.3 Nanochannel Voltage and the Equivalent Circuit

The system forms an equivalent circuit for which each individual component can be modeled as a resistor with a resistance given by

$$R = \frac{L}{\sigma A} \quad (4.51)$$

where L is the length of the element, σ is the fluid conductivity, and A is the cross-sectional area. Strictly speaking, this model holds only for systems with no pressure gradients and vanishing thin electric double layers. However, these conditions are sufficiently satisfied for our system to make it a very good approximation.

The resistance of the device is therefore given by the sum of the individual elements:

$$R_{device} = 2R_r + 2R_{fil,5\mu m} + 2R_{fil,1\mu m} + 2R_{fil,0.5\mu m} + 2R_{c,10\mu m} + 2R_{taper} + R_c \quad (4.52)$$

where the factor of 2 comes from the upstream and downstream of the sections of the device as the devices have a line of symmetry equidistant from the cathode and anode. Note that many of the terms in Eq. (4.52) are equivalent resistances accounting for the parallel nature of the device which follow the general form

$$R_{\parallel} = \frac{L_{single}}{N\sigma A_{single}} \quad (4.53)$$

where N is the number of parallel elements.

The resistance of the tapered section, which connect the nanochannels to the $10 \mu m$ wide channels, can be easily calculated using

$$R_{taper} = \int_0^L \frac{dl}{\sigma A(l)} \quad (4.54)$$

where $A(l)$ is a linear function of l which accounts for the taper. For our devices with a taper/ $\mu\text{m} = 1.9$, the trivial integration leads to a resistance given by

$$R_{taper} = \frac{1.57}{\sigma N_c H_c} \quad (4.55)$$

where $N_c = 20$, the number of parallel channels.

The voltage across the nanochannel, V_c , can be related to the applied voltage, V_a , via voltage division by the expression

$$\frac{V_c}{V_a} = \frac{R_c}{R_{device}} \quad (4.56)$$

With the assumption height uniformity across the device which is adequate for our purposes, we arrive at the following expression for the voltage ratio:

$$\frac{V_c}{V_a} = \frac{\frac{L_c}{W_c}}{\frac{2L_r}{W_r} + \frac{2L_{fil}}{W_{fil,5\mu\text{m}}} + \frac{2L_{fil}}{W_{fil,1\mu\text{m}}} + \frac{2L_{fil}}{W_{fil,0.5\mu\text{m}}} + \frac{2L_{c,10\mu\text{m}}}{W_{c,10\mu\text{m}}} + \frac{L_c}{W_c}} \quad (4.57)$$

Equation (4.57) illustrates the voltage across the nanochannel is dependent on the applied voltage and geometrical factors of the device. Using the U-shaped parameters given in Tables 4.1, 4.2, and 4.4, we find for the U-shaped device that

$$\frac{V_c}{V_a} = 0.908 \pm 0.003 \quad (4.58)$$

The large percentage of applied voltage across the U-shaped nanochannels is a result of its large resistance relative to the other resistances in the device. However, due to its length, the electric field within the U-shaped nanochannel remains comparatively small.

The straight nanochannels have a considerably lower voltage across them for a given V_a as their portion of the overall resistance is lower. Putting the geometrical parameters of the straight channels given in Tables 4.2, 4.3, and 4.4

into Eq. (4.57), we find that

$$\frac{V_c}{V_a} = 0.33 \pm 0.02 \quad (4.59)$$

The uncertainty in the voltage ratios given in (4.58) and (4.59) is dominated by reservoir hole placement with an estimated deviation of ± 1 mm per hole. Reservoir hole placement could be improved in the future by implementing a CO₂ laser cutting system for better precision control and reproducibility.

4.3.4 Optical System for Mobility Measurements

The optical system was designed to generate 2-color mobility measurements; thus enabling the detection of two different fluorophores while simultaneously measuring the mobilities of the various biomolecules to which the fluorophores are attached.

Coherent Obis 488 LX-50 (488 nm) and Coherent Cube 640-40C (640 nm) lasers were overlapped and subsequently split using Pellicle beamsplitters to create up to three equally spaced focused beam spots in a linear array. The distance between adjacent spots was 12.5 μm and using the two outer spots gave a separation distance of 25 μm . The 25 μm spot distance wasn't a design limitation since nanochannels perpendicular to the linear beam array could be fabricated as described in Section 4.3.2 with the U-shaped device.

The Pellicle beamsplitters were used to generate three focused beams which effectively eliminated ghosting and shifting of the beam caused by the reflections from the front and back surfaces typical of plate beamsplitters. However, Pellicle beamsplitters still have their drawbacks. For example, their transmis-

ision coefficient is wavelength and polarization dependent which caused difficulties in getting the input powers approximately equal for all laser spots. This problem was circumvented by inserting quarter waveplates into the beam paths.

Using Pellicle beamsplitters becomes especially cumbersome when adding more than three spots. If more spots were needed, a more robust and elegant design would involve diffractive beamsplitters. Commercial diffractive beam-splitters are widely available in varying configurations.

The emission optics consisted of a quad-band filter (Chroma ZET405/488/561/635M) to eliminate scattered laser light, a 580 nm edge BrightLine single-edge dichroic beamsplitter (Semrock FF580-FDi01), two emission filters (Semrock BrightLine FF02-525/40-25, Semrock BrightLine FF01-685/40-25), and two linear fiber optic arrays (Oz Optics VGA-8-250-0-X-10.3-3.8-2.03-QM-UVVIS-105/125-3-1-1-0-SP). The fiber arrays consisted of 8 individual multimode fibers spaced 250 μm apart in V-Groove assemblies with 105/125 μm core and cladding diameters.

No pinhole aperture exists in the setup since the 105 μm core effectively acts as a confocal aperture. The pinhole aperture diameter is typically designed to be approximately 1 Airy unit which optimally balances the trade-off between collection efficiency, axial resolution, and signal-to-background levels. An Airy unit corresponds to the first Airy disc minimum given by

$$d_{\text{Airy}} = \frac{1.22\lambda M}{NA} \quad (4.60)$$

For our system, one Airy unit corresponds to $d_{\text{Airy}} = 37 \mu\text{m}$ at $\lambda = 600 \text{ nm}$. Thus, our system's apertures are approximately three times larger than the optimal Airy unit. As a consequence, background levels are larger due to the approxi-

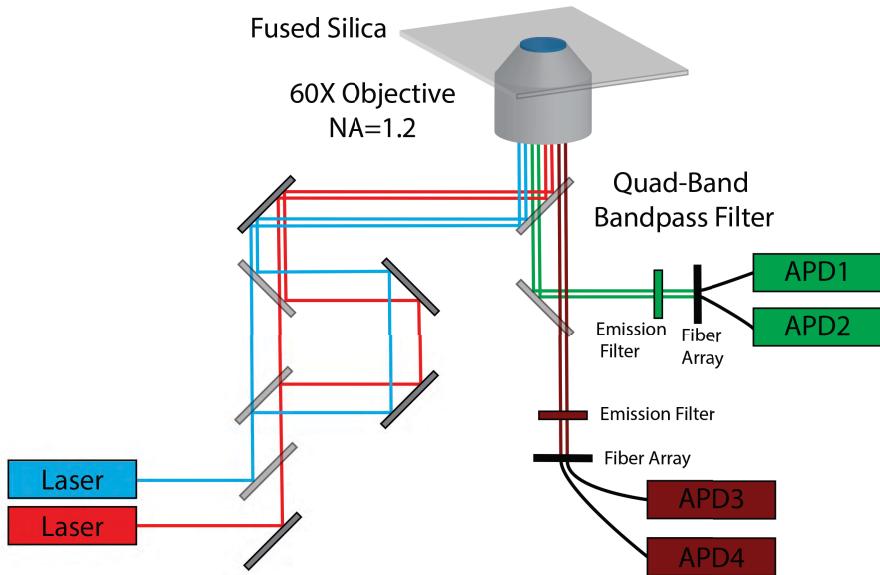


Figure 4.5: Optical setup for 2-color single molecule mobility experiments. Quarter wave plates in excitation paths were omitted for clarity.

mate linear growth for large pinholes in optical section thickness which is given by

$$dz \approx \sqrt{\left(\frac{\lambda n}{NA^2}\right)^2 + \left(\frac{\sqrt{2}n \cdot PH}{NA \cdot M}\right)^2} \quad (4.61)$$

where PH is the pinhole diameter. Using Equation (4.61), we find $dz \approx 2.8 \mu\text{m}$ where as the optimal, but not diffraction limited condition, gives $dz_{\text{AiryUnit}} \approx 1.2 \mu\text{m}$. However, since our devices confine the fluidics and the fluorescent molecules to a region approximately 250 nm thick, the additional optical section thickness only contributes background autofluorescence originating from the UV fused silica. Hence, the larger apertures are of little detriment to the detection capabilities of the system since the UV fused silica has very low autofluorescence.

Optical Crosstalk

Crosstalk between output fibers was a concern due to its capability to create false positives and to raise background levels especially in low signal to noise environments such as single dye or chromatin mobility measurements. At its worst, with complete coupling, the crosstalk would raise the effective concentration of detected molecules in each signal channel by a factor of 2. To address this concern, we measured the fluorescent emission into neighboring fibers with the laser centered over a given fiber. Fluorescence was generated using autofluorescent plastic slides (Chroma). The measurement served as an upper bound for the crosstalk as actual fluorescent molecules are confined to the 500 nm wide channel which is less than the laser beam width for either the 488 nm or 637 nm laser. The coupling factor is defined as

$$C_{n,1} = 10 \log \left(\frac{P_n}{P_1} \right) \quad (4.62)$$

where n is the fiber number. Physically, the distance between adjacent fibers is 250 nm and the distance between fibers 1 and 8 is $d_{1,8} = 1750$ nm. The results are shown in Figure 4.6.

The sharper coupling factor decay for the 488 nm laser line can be attributed to a smaller beam spot. The 488 nm laser had a FWHM of $1.78 \pm 0.05 \mu\text{m}$ which gives a beam width of $1.51 \pm 0.05 \mu\text{m}$. The 637 nm laser had a FWHM of $2.93 \pm 0.03 \mu\text{m}$ which corresponds to a beam width of $2.49 \pm 0.02 \mu\text{m}$. Beam spot measurements were made using a EMCCD camera (Andor). The spots weren't diffraction limited as larger beam spots gave better beam uniformity over the inspection volume in the nanofluidic channel.

Nearest-neighbor isolation was at least 10 dB for both wavelengths. Almost

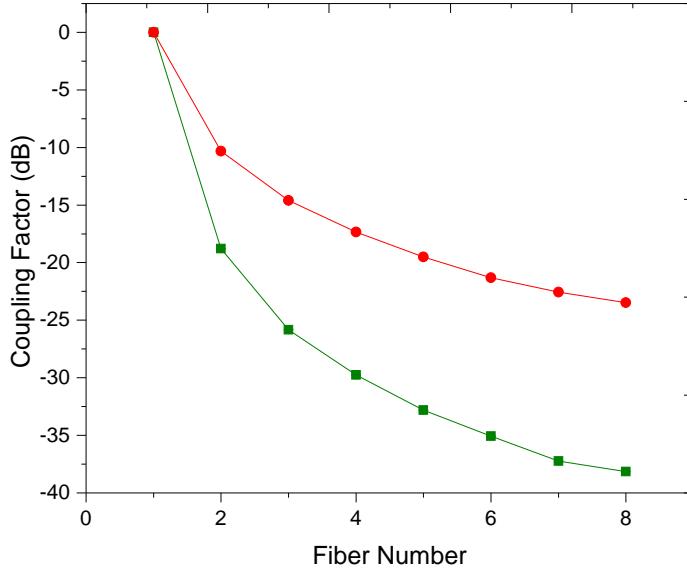


Figure 4.6: Fiber number vs. coupling factor for 488 nm (Green) and 637 nm (Red) lasers. Larger coupling observed in the 637 nm laser can be attributed to a larger spot size. Error bars are smaller than the symbols (maximum error=±0.1 dB).

all mobility measurements were taken at the 1-7 fiber configuration which offered over 20 dB of isolation and resulted in 25 μm of beam spot distance separation using a 60X objective. A few mobility measurements were made using the 1-4 fiber configuration which gave over 15 dB of isolation. This amount of isolation was experimentally sufficient in that no false coincidence resulting from crosstalk was ever observed.

4.3.5 Sample Preparations

Standard SCAN running buffer used in all experiments consisted of 1xTE, 0.5% (w/v) polyvinylpyrrolidone (PVP) at an average molecular weight of 40 kDa, and 0.1% (v/v) Triton X-100. The pH was adjusted to 7.5. Multiple experiments

utilized the Atto dye for EOF calibration. ATTO dyes were diluted into the standard SCAN running buffer at a concentration of approximately 200 pM.

Alexa Fluor 647 Experiments Alexa Fluor 647 carboxylic acid, tris(triethylammonium) salt (Thermo Fisher) was diluted into standard SCAN running buffer at a concentration of approximately 200 pM prior to performing SCAN.

Wheat Germ Agglutinin Experiments for Single Dye Mobility Resolution Wheat Germ Agglutinin, Alexa Fluor 647 Conjugate (Thermo Fisher) was diluted into standard SCAN running buffer at a concentration of approximately 200 pM prior to performing SCAN.

DNA Mobility Experiments A solution of equal molar concentrations λ , 20 kb, 10 kb, 5 kb, and 1 kb dsDNA fragments was prepared using the standard SCAN running buffer. The solution had an overall DNA concentration of 1 ng/ μ L which reduced DNA aggregation during nucleic acid staining. YOYO-1 intercalating dye was added to the solution at a 1 dye : 5 bp ratio ensuring adequate intercalation without DNA fragmentation caused by higher dye ratios. Samples were left to stain overnight at 4 °C. Prior to SCAN, the sample was diluted to approximately 200 pM.

Chromatin Experiments HeLa-GFP chromatin was isolated by members of the Soloway Lab using the chromatin extraction protocol given in Section 3.2.3. Further processing was performed the Soloway Lab to segregate mono/di/trinucleosome and larger chromatin via differential centrifugation

using a sucrose gradient as explained in 3.2.4. Buffer exchange and purification of trinucleosome chromatin was performed using an Amicon spin column (Millipore) with a nominal molecular weight limit of 50 kDa for 15 minutes at 5000g using PBS. A 50 kDa column was used instead of a 100 kDa used normally for chromatin purification as to keep free DNA in the sample to ensure both free DNA and chromatin distributions. For experiments involving crosslinking, the chromatin was crosslinked using a 1% solution of biological grade formaldehyde for 10 minutes followed by an addition of equal molar glycine to quench the reaction. After this, samples were diluted to 1 ng/ μ L to reduce aggregation during intercalation in the standard SCAN running buffer. TOTO-3 at a 1:5 dye to bp ratio was then reacted with the chromatin and left overnight at 4°. Prior to running SCAN, the sample was diluted further to approximately 200 pM and added to both ports of the SCAN device.

Mobility Based Identification of 6 Biomolecules TOTO-3 labeled 1 kb DNA (Thermo Fisher, SM1671) fragments; bovine serum albumin (BSA), Alexa Fluor 647 conjugate (Thermo Fisher, A34785); lectin soybean agglutinin (SBA) from Glycine max (soybean), Alexa Fluor 647 conjugate (Thermo Fisher, L32463); mouse IgG1, Alexa Fluor 647 conjugate (Thermo Fisher, MG121), ATTO 647N (ATTO-TEC GmbH) antibodies, and mouse IgG2b (Active Motif, 39763) labeled with Alexa Fluor 647 antibodies were each diluted separately into standard SCAN running buffer at a concentration of approximately 200 pM prior to performing SCAN.

Prior to dilution, the 1 kb DNA (Thermo Fisher, SM1671) was stained with TOTO-3 (Thermo Fisher, T3604). The TOTO-3 intercalating dye was added to a 1 ng/ μ L solution of the 1 kb DNA at a 1 dye : 5 bp ratio ensuring adequate

intercalation without DNA fragmentation caused by higher dye ratios. DNA samples were left to stain overnight at 4 °C.

The SCAN experiment processed the samples sequentially with the same nanochannel in the order given above. Between each sample, the nanochannels were flushed with standard SCAN buffer for at least 30 minutes and checked for the remnants of the previous sample. Additionally, ATTO 647N was analyzed between each sample to control for EOF time variant fluctuations.

Chromatin-Antibody Experiments Prior to performing the chromatin/ α -H3 binding reaction, each constituent was blocked with 1xBSA for 1 hour at 4 °C in PBS and 0.1 % Triton X-100. After blocking, the trinucleosome chromatin at a concentration of 20 ng/ μ L was then mixed with 500 nM α -H3 antibody and left to bind overnight at 4 °C. The complexes were crosslinked using 0.75% formaldehyde for 15 minutes. The crosslinking reaction was quenched using glycine at 200 mM and diluted into the standard SCAN buffer at a concentration of 200 pM.

BSA-Primary Ab-Secondary Ab Experiments Bovine serum albumin (BSA) Alexa Fluor 647 conjugate (Thermo Fisher, A34785), an α -BSA antibody (Thermo Fisher, A11133, unlabeled); and an α -Rabbit IgG (H+L) secondary antibody with Alexa Fluor 488 (Thermo Fisher, A11008) were mixed at equal molar in PBS. The binding reaction consisted of 13 nM of each component which translates to 2 μ g/mL for the antibodies. Following the overnight binding reaction at 4 °C, the complexes were crosslinked using 1.0% formaldehyde for 10 minutes. The crosslinking reaction was quenched using glycine at 200 mM. The sample was diluted then into the standard SCAN buffer at a concentration of approxi-

mately 200 pM and analyzed using SCAN.

For BSA control experiment, we diluted the BSA into the standard SCAN running buffer at a concentration of approximately 200 pM prior to performing SCAN. Likewise, for the secondary antibody control experiment, we diluted the secondary antibody into the standard SCAN running buffer at a concentration of approximately 200 pM prior to performing SCAN. For the BSA-primary antibody control experiment, the binding was performed at 13 nM BSA and a 500-fold molar excess (6.5 μ M) of antibody to push the equilibrium to near complete binding and reduce the chance for precipitin. Crosslinking and other steps were the same as described above. For the primary antibody-secondary antibody control experiment, the binding was performed at 13 nM for each component in PBS. Crosslinking and other steps were the same as described above.

4.3.6 Measurement of Electroosmosis

The direct method to measure the EOF of our system involves using a neutral tracer dye to measure the transit time between laser inspection volumes. Eligible fluorophores need to have enough brightness to enable single molecule measurements along with neutrality at the measured pH. Additionally, the dyes must be at least slightly hydrophilic to ensure solubility in the running buffers.

We have typically used Alexa Fluor (Thermo Fisher) dyes for the majority of our experiments due to their brightness, stability, and availability. Alexa Fluor dyes are synthesized from coumarin, rhodamine, xanthene, and cyanine dyes by adding sulfonic acid moieties. The sulfonation produces increased brightness due to ring sulfonation and enhanced hydrophilicity for the fluorophores.

Unfortunately, it also adds negative charge leaving all Alexa Fluor dyes as anionic molecules. Other popular single molecule fluorophores such as cyanine dyes can be synthesized with a carboxylic acids as their functional group giving them a neutral structure. However, their solubility is poor in aqueous solutions and require an organic phase (e.g. DMF, DMSO) to make the fluorophores soluble.

Instead, we turned to using ATTO (ATTO-TEC GmbH) fluorescent dyes to measure the EOF in our system. Specifically, ATTO 647N with a carboxylic acid modification and an amine modified ATTO 565 were used as tracer dyes as shown in Figure 4.7.

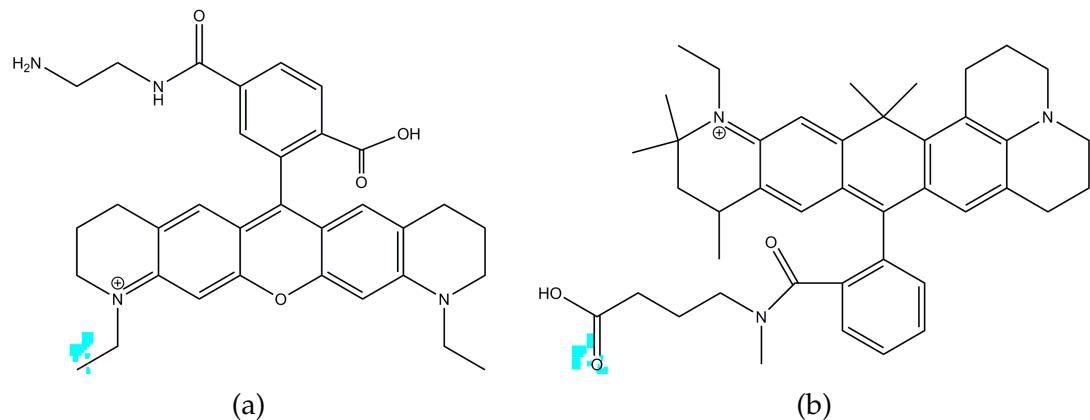


Figure 4.7: Chemical structures of the ATTO dyes used for EOF measurement. Note that the carboxylic acid functional groups will be deprotonated at neutral pH since they have $\text{pK}_a \approx 4.8$. (a) ATTO 565 with amine modification. (b) ATTO 647N with a carboxylic acid modification.

For most experiments, ATTO 647N was typically used to measure EOF since the optics were already setup for the other SCAN experiments. When using ATTO 647N, the other analytes were eliminated from the nanochannel by flowing fresh running buffer for approximately 30 minutes or until no sample was observed. Then, ATTO 647N was loaded into the reservoirs and subsequently analyzed. ATTO 565 eliminated the need for the washing step, but it required

a more complex optical setup. In the end, having two spectrally distinct EOF tracer dyes gave flexibility when processing samples.

4.4 Results and Discussion

4.4.1 Single Molecule Mobility Detection

As the fluorescently labeled biomolecules traversed the nanochannel device, they encountered focused lasers defining two 150 aL observation volumes which generated sequential laser-induced fluorescent burst signals. Figures 4.8a and 4.8c show time traces for Alexa Fluor 647 and TOTO-3 labeled 900 bp DNA respectively. For these time traces, the curved U shaped device was utilized with a nanochannel distance of 40 μm between beam spots and operated at electric field strength of $|\vec{E}| = 3.08 \text{ kV/cm}$. Molecules shown as red dots in the time trace were determined using a total burst count thresholding set for the minimum overlap between the noise distribution and the signal distribution for burst counts. Total molecule counts were determined by fitting a log-normal distribution to the burst area distribution. Both Alexa Fluor 647 and DNA essentially had a one-to-one correspondence between Channel 1 and Channel 2 with total molecule counts differing by no more than 2% for Alexa Fluor 647 and 0.5% for DNA.

A simple investigation of the time traces indicated that some Alexa dyes underwent photochemical destruction since a few dyes were only observed in Channel 1. While Figure 4.8a was chosen to illustrate this fact, it overestimates the actual effect due to counting statistics as the total photobleached molecule

were less than 2% when averaged over a sufficient number of molecules as noted before.

Alexa dyes also exhibited dispersion of transit times which are easily noticeable in the time trace by observing relative timing between burst events in each channel. Two bursts in channel 1 will be closer or farther apart in channel 2 signaling band broadening. DNA's band broadening isn't as apparent due to its reduced broadening and the resolution of the time traces.

From these time traces ensemble based electrophoretograms were constructed by finding the time differences between all molecules in channel 1 and channel 2 and producing a histogram of the result. Peak centers were defined by the centroid of the burst and not the peak as bursts generally exhibited a slight asymmetry due to photobleaching effects. Being able to differentiate migration peaks above background generally required lower concentrations relative to SCAN using a single beam spot, and the differentiation is both a function of the concentration of biomolecule and the degree of peak broadening. With sharp peaks such as with DNA, the necessity to dilute further than single spot SCAN is negligible. A criterion for guaranteeing adequate peak discrimination is to dilute the sample so that the average waiting time between events is large compared to the migration time of the molecule. In practice, we diluted to around 10 pM for some samples.

Figures 4.8b and 4.8d are electrophoretograms of Alexa Fluor 647 and TOTO-3 labeled 900 bp DNA for $L_d = 570 \mu\text{m}$ and varying voltages showing differences in migration time and peak variances. The curves are $V_a = 200 \text{ V}$ (black), $V_a = 100 \text{ V}$ (blue), $V_a = 50 \text{ V}$ (red), and $V_a = 25 \text{ V}$ (green) which corresponded to electric field strengths ranging from $|\vec{E}| = 3.08 \text{ kV/cm}$ to $|\vec{E}| = 0.385 \text{ kV/cm}$.

Peak broadening was readily obvious in both electrophoretograms as migration time increased with Alexa Fluor's dispersion being markedly greater than DNA's. Being that Alexa Fluor 647 has a diffusion coefficient two orders of magnitude larger than the DNA, this was expected.

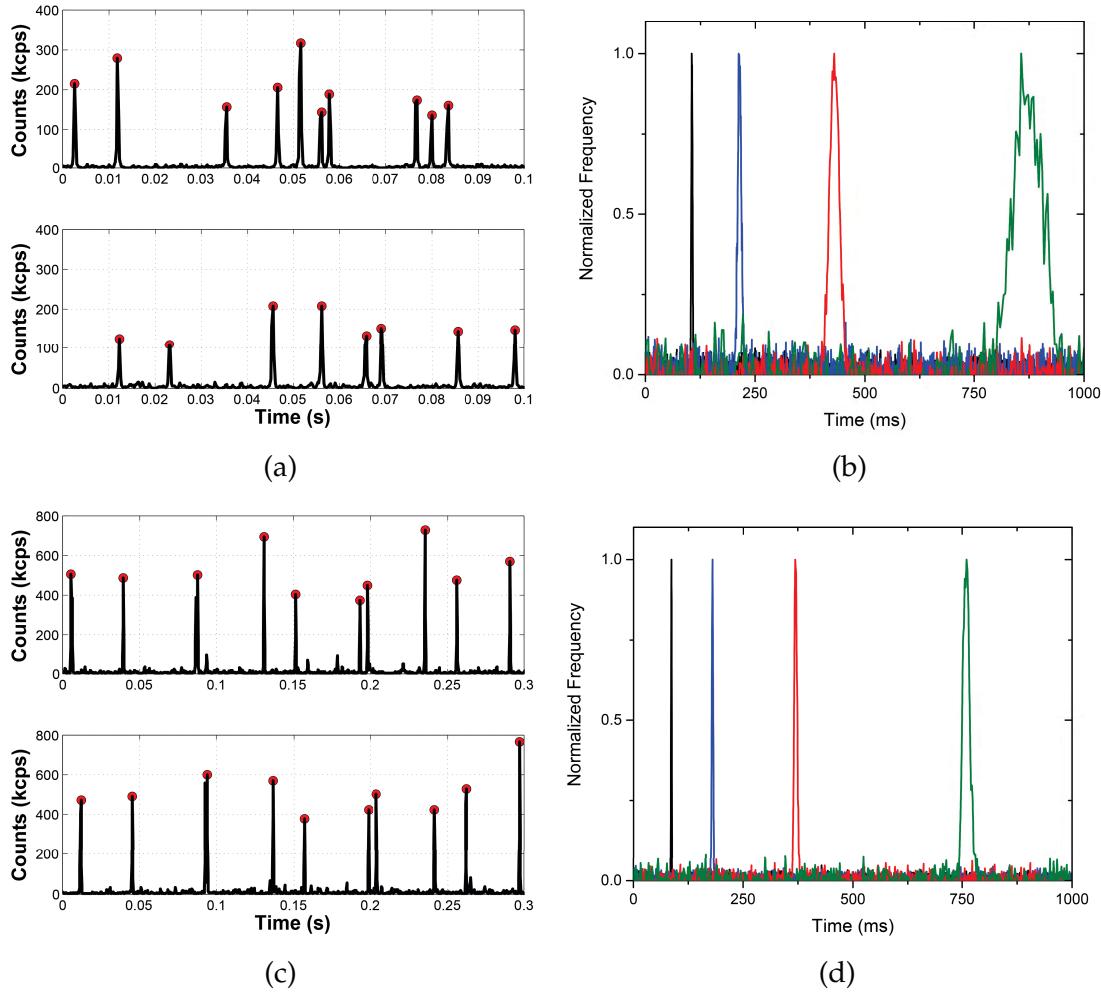


Figure 4.8: (a) 2 channel time traces for Alexa Fluor 647. Bursts are detected in Channel 1 and subsequently detected in Channel 2. (B) Electrophoretograms of Alexa Fluor 647 for 200 V (black), 100 V (blue), 50 V (red), and 25 V (green). The increased noise of the longer migration time peaks is an artifact of the binning of the histogram as their bins had fewer counts relative to their faster counterparts due to peak broadening. (c) 2 channel time traces for TOTO-3 labeled 900 bp DNA. (D) Electrophoretograms of the 900 bp DNA for 200 V (black), 100 V (blue), 50 V (red), and 25 V (green).

Migration Time Determination via Cross-Correlation

The electrophoretograms shown in Figures 4.8d and 4.8b were determined by locating all the molecules and compiling their migration times between observation spots. Similar migration time information could be obtained by calculating the cross-correlation between Channels 1 and 2. In fact, real-time monitoring of the mobilities was achieved in this manner during experiment. The unnormalized unbiased cross correlation of the time trace sequences for Channel 1 and Channel 2 was computed as

$$R_{12}(m) = \frac{1}{N-m} \sum_{n=0}^{N-m-1} x_{n+m} y_n, \quad m \geq 0 \quad (4.63)$$

where x_i and y_i are the i th terms for Channel 1 and 2 respectively and N is the number of terms in each sequence. Figure 4.9 shows the normalized cross-correlation between Channels 1 and 2 for the 900 bp DNA experiment with $L_d = 570 \mu\text{m}$. Notice the resemblance of Figure 4.9 to Figure 4.8d

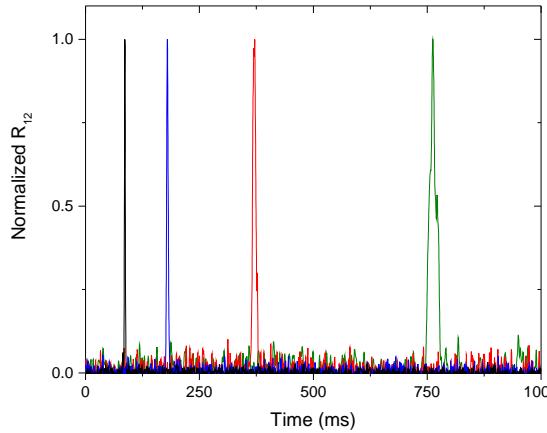


Figure 4.9: Normalized cross correlation of Channel 1 and 2 signals from TOTO-3 labeled 900 bp DNA experiment. V_a was varied as 200 V (black), 100 V (blue), 50 V (red), and 25 V (green)

4.4.2 Voltage Dependence on Migration Time

To characterize the device's dependence on voltage, we again chose to work with TOTO-3 labeled DNA and Alexa Fluor 647 since they have become ubiquitous to single molecule studies, are generally well-behaved in the nanochannels, and have orders of magnitude differences in size and diffusion coefficients. Earlier results in [13] using nucleic acid engineered fluorescent labels gave a linear relationship for molecular speed as a function of applied electric field as predicted by electrophoretic theory. Confirming this result served two purposes as it showed the validity of the result for nanochannels and it provided relevant fitting parameters in order to determine mobilities and diffusion coefficients for the investigated biomolecules which hasn't been demonstrated. Moreover, Stavis et al. had peaks which exhibited significant asymmetries presumably due to wall interactions which contributed to added uncertainties in their measurement.

For both the Alexa Fluor 647 and DNA experiments, a voltage sweep was performed between $V_a = 25$ V and $V_a = 200$ V at a nanochannel distance of $L_d = 570 \mu\text{m}$. This corresponds to a electric field sweep from $|\vec{E}| = 0.385 \text{ kV/cm}$ to $|\vec{E}| = 3.08 \text{ kV/cm}$. The resulting peaks showed little asymmetry as Figures 4.8b and 4.8d illustrate. This facilitated the use of a symmetric Gaussian for the peak fits. Linear regressions were performed on the resulting peak centers which theory predicts should be linear as a function of V_a^{-1} . The resulting regression curve for Alexa Fluor 647 shown in Figure 4.10a was given by $t_m = \beta V_a^{-1} + \alpha$ with $\beta = 21910 \pm 40 \text{ V}\cdot\text{ms}$, $\alpha = -5.1 \pm 0.7 \text{ ms}$, and adjusted $R^2 = 0.99998$. Likewise, the resulting regression curve for 900 bp DNA shown in Figure 4.10b was given by $t_m = \beta V_a^{-1} + \alpha$ with $\beta = 19250 \pm 60 \text{ V}\cdot\text{ms}$, $\alpha = -12 \pm 1 \text{ ms}$, and adjusted $R^2 = 0.99994$.

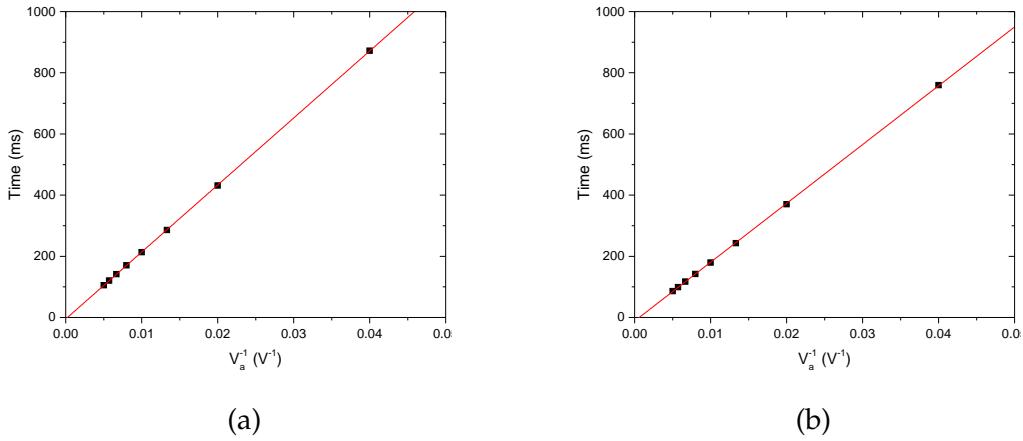


Figure 4.10: V_a^{-1} vs. migration time. (a) Alexa Fluor 647. Error bars are smaller than the symbols (± 0.8 ms for maximum standard error of the mean). (b) 900 bp DNA fragments labeled with TOTO-3 intercalating dye. Error bars are smaller than the symbols (± 0.1 ms for maximum standard error of the mean)

Electrophoretic mobilities were also calculated after determining the EOF using Atto 647N tracer dye. Analysis of the EOF gave $\mu_{eo} = (12.7 \pm 0.1) \times 10^{-9} \text{ m}^2\text{V}^{-1}\text{s}^{-1}$ for the Alexa Fluor 647 experiment and $\mu_{eo} = (8.3 \pm 0.1) \times 10^{-9} \text{ m}^2\text{V}^{-1}\text{s}^{-1}$ for DNA. These EOF's are constant with [19].

The strong linear voltage dependence on migration times also signals that Joule heating had a negligible effect on the system. Effects of Joule heating would be observed in the form of decreased migration times for higher voltages as temperatures increased. The lack of observable Joule heating can be attributed to the high surface-to-volume ratio of the device. The high surface-to-volume ratio implies that the device operates at low currents and has high heat transfer relative to traditional capillaries as the heat production scales with the cross-sectional area of the channel.

4.4.3 Length Dependence on Migration Time

Since the velocity of a particle is determined by the electric field and the mobility solely depends on buffer composition, channel wall chemistry, and inherent physical properties of the particle, the migration time should scale with the length between inspection volumes. We tested this preposition over L_d ranging from 40 μm to 570 μm using our U-shaped devices. The U-shaped devices allowed to probe length scales up to and perhaps beyond the practical limit of the device by simply translating the microscope stage along with probing length 75 times larger than previously measured.

We confirmed the linearity of the device over its entire length using $V_a = 200$ V corresponding to $|\vec{E}| = 3.08 \text{ kV/cm}$. Moreover, the linearity proved that the curvature in the channel had a negligible effect on migration times implying that the racetrack effect which plagued microchannel separations with curved channels was undetectable for our devices. Peaks were well-fitted using Gaussians due to their absence of asymmetry. Alexa Fluor 647 gave a linear regression of $t = \beta L_d + \alpha$ where $\beta = 0.1838 \pm 0.0008 \mu\text{m}^{-1}\cdot\text{ms}$ and $\alpha = -0.7 \pm 0.1 \text{ ms}$ with adjusted $R^2 = 0.9999$. This is shown in Figure 4.11a. DNA gave a linear regression of $t = \beta L_d + \alpha$ where $\beta = 0.1478 \pm 0.0004 \mu\text{m}^{-1}\cdot\text{ms}$ and $\alpha = 0 \pm 0.1 \text{ ms}$ with adjusted $R^2 = 0.99998$ which is shown in Figure 4.11b.

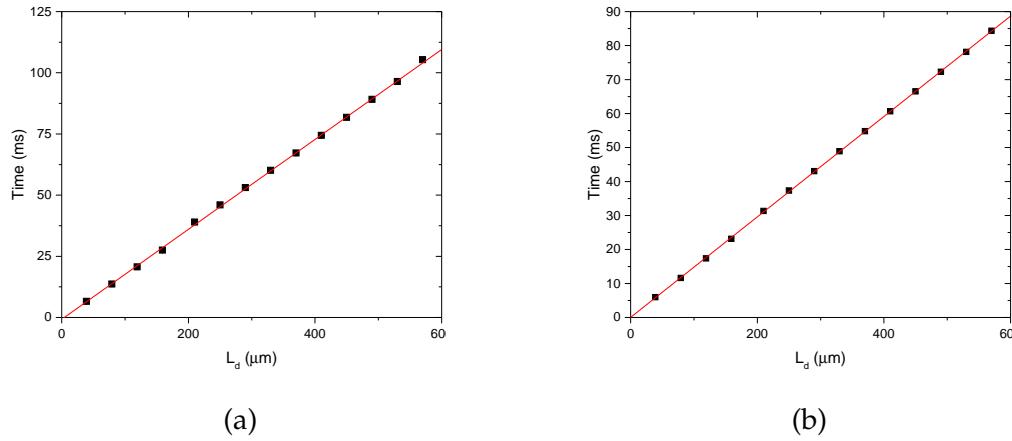


Figure 4.11: L_d vs. migration time. (a) Alexa Flour 647 dye. Error bars are smaller than the symbols (± 0.02 ms for maximum standard error of the mean). (b) 900 bp DNA fragments labeled with TOTO-3 intercalating dye. Error bars are smaller than the symbols (± 0.007 ms for maximum standard error of the mean).

4.4.4 Theoretical Plates of Alexa Fluor 647 and DNA by Varying Voltage

As noted previously, theoretical plates give a measure of the peak broadening and ultimately the resolution for which we can identify individual molecular species based on mobility alone. Therefore, it seems prudent to characterize the effect of certain experimental parameters have on the separation of the underlying migration time distributions.

Considering a Stokes sphere of radius R , it is well known from the Nernst-Einstein relation and Stokes' drag that

$$D = \frac{k_B T}{6\pi\eta R} \quad (4.64)$$

Since a dye will almost certainly have the smallest Stokes radius for any molecule used in SCAN and hence, the largest diffusion coefficient, we first

investigated its dispersion which presumably would be diffusion limited.

Using the Alexa Fluor 647 migration time voltage sweep measurements taken at $L_d = 570 \mu\text{m}$ from Section 4.4.2, we computed the number of plates using Equation (4.33). To recapitulate from earlier, in a diffusion limited system, the dependence on the theoretical plates should be linear in the applied voltage, V_a , as given in Equation (4.36). We indeed found a linear relationship between the number of plates and applied voltage as shown in Figure 4.12a. The fitted curve is given by $N = \beta V_a + \alpha$ where $\beta = 29.2 \pm 0.8 \text{ V}^{-1}$, $\alpha_2 = -60 \pm 70$ with adjusted $R^2 = 0.9952$.

Since Alexa Fluor 647's dispersion was diffusion limited, its diffusion coefficient was calculated using Equation (4.45) and the results from Section 4.4.2. We determined the Alexa Fluor 647's diffusion coefficient to be

$$D_{AF647} = (2.5 \pm 0.1) \times 10^{-10} \text{ m}^2\text{s}^{-1} \quad (4.65)$$

which is in reasonable agreement to that found by fluorescence correlation spectroscopy [20]. Using the D_{AF647} , the Nernst Einstein relation (4.50), and μ_{ep} from Section 4.4.2, we found that the effective charge of the Alexa Fluor 647 was $q = (-2.999 \pm 0.005)e$ at $T = 298 \text{ K}$. This value of q is in excellent agreement with the charge expected from the chemical structure, namely $q = -3e$.

We also investigated DNA broadening as a function of applied voltage. Figure 4.12b shows the number of plates vs. V_a for YOYO-1 labeled 900 bp DNA. Unlike the Alexa Fluor 647 which had a linear relationship between theoretical plates and V_a , we found the DNA data to be well modeled by

$$N = \frac{L_d^2}{\beta_2 V_a + \beta_1 + \beta_0/V_a} \quad (4.66)$$

The least-squares fit gave $\beta_2 = (2.4 \pm 0.5) \times 10^{-2} \text{ V}^{-1} \cdot \mu\text{m}^2$, $\beta_1 = (2.0 \pm 0.1) \mu\text{m}^2$,

and $\beta_0 = (3.119 \pm 0.004) \times 10^2 \text{ V}\cdot\mu\text{m}^2$ with an adjusted $R^2 = 0.9391$. The β_2 term gives a wall interaction broadening suggested by Equation (4.40). The β_1 term we again attribute to wall interactions. Other voltage independent dispersion terms such as detector variance and binning time quantization are at least two orders of magnitude too small. We assume the β_0 term is related to diffusion through the equation

$$\beta_0 = \frac{2DL_dL_c}{\mu_{ep} + \mu_{eo}} \quad (4.67)$$

After manipulation of Equation (4.24) with Equation (4.67), we find the DNA diffusion coefficient given by

$$D_{DNA} = \frac{\beta_0}{2\beta_m} \quad (4.68)$$

Using Equation (4.68), we calculated the diffusion coefficient for 900 bp DNA given by

$$D_{DNA} = (8.10 \pm 0.03) \times 10^{-12} \text{ m}^2\text{s}^{-1} \quad (4.69)$$

The experimental DNA diffusion coefficient is well within the typical range found by [21]

$$D \approx (2 - 5 \times 10^{-10})N_{bp}^{-0.57} \text{ m}^2\text{s}^{-1} \quad (4.70)$$

DNA gave the highest observed theoretical plate total for the device. While still 1-2 orders of magnitude smaller than capillary electrophoresis, this plate number is consistent with the fact that we operated our device 1-2 orders of magnitude lower in voltage compared to capillary electrophoresis and still very impressive for single molecule methods.

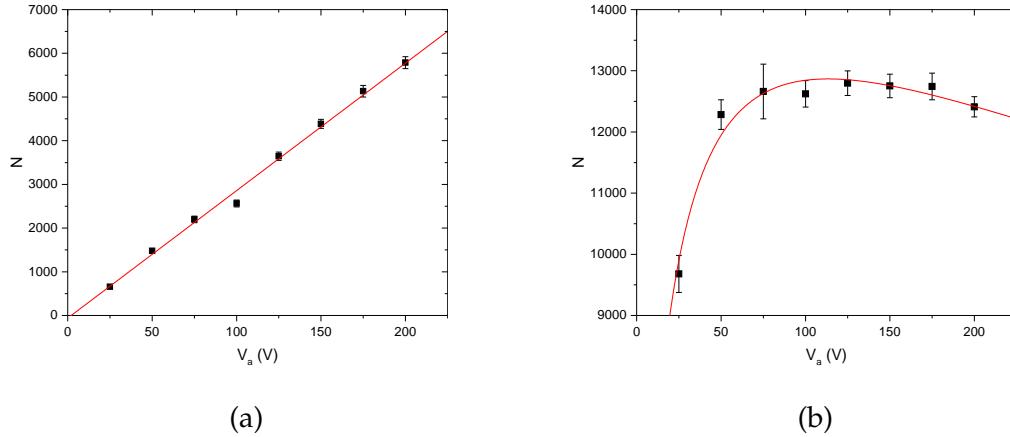


Figure 4.12: (a) V_a vs. theoretical plates for Alexa Fluor 647. The fitted curve is given by $N = \beta V_{\text{Applied}} + \alpha$ where $\beta = 29.2 \pm 0.8 \text{ V}^{-1}$ and $\alpha = -60 \pm 70$. (b) V_a vs. theoretical plates for 900 bp DNA fragments labeled with TOTO-3 intercalating dye. The fitted curve is given by $N = L_d^2 / (\beta_2 V_a + \beta_1 + \beta_0 / V_a)$ where $\beta_2 = (2.4 \pm 0.5) \times 10^{-2} \text{ V}^{-1} \cdot \mu\text{m}^2$, $\beta_1 = (2.0 \pm 0.1) \mu\text{m}^2$, and $\beta_0 = (3.119 \pm 0.004) \times 10^2 \text{ V} \cdot \mu\text{m}^2$ with an adjusted $R^2 = 0.9391$.

4.4.5 Theoretical Plates of Alexa Fluor 647 and DNA by Varying

Detection Length

In a diffusion limited system, the theoretical plates should be proportional to the detection length. Since we already knew Alexa Fluor 647 was diffusion limited from the voltage varying theoretical plate measurements, we expected a linear behavior when we adjusted the detection length. Using the data gathered in Section 4.4.3, we computed the theoretical plates from $L_d = 40 \mu\text{m}$ to $L_d = 570 \mu\text{m}$. Figure 4.13a shows the results which are well described by a line. The fitted curve is given by $N = \beta_{N,L} L_d + \alpha$ where $\beta_{N,L} = 10.6 \pm 0.2 \mu\text{m}^{-1}$ and $\alpha = -40 \pm 20$ with $R^2 = 0.99237$.

By using the detection length variation data, we were able to provide an orthogonal approach to check our methodology. Using Equation (4.49) to calculate

the diffusion coefficient in terms of $\beta_{m,L}$ and $\beta_{N,L}$, we found

$$D_{AF647,L} = (2.55 \pm 0.05) \times 10^{-10} \text{ m}^2\text{s}^{-1} \quad (4.71)$$

which is in excellent agreement with the diffusion coefficient found through voltage variation.

We also investigated DNA broadening as a function of detection length. Figure 4.13b shows the number of plates vs. V_a for YOYO-1 labeled 900 bp DNA. Unlike the Alexa Fluor 647 which had a linear relationship between theoretical plates and V_a , we found the DNA data to be well modeled by

$$N = \frac{L_d^2}{\beta_2 L_d^2 + \beta_1 L_d + \beta_0} \quad (4.72)$$

The least-squares fit gave $\beta_2 = (6.0 \pm 0.3) \times 10^{-5}$, $\beta_1 = (1.3 \pm 0.2) \mu\text{m}$, and $\beta_0 = (8.37 \pm 0.5) \times 10^{-1} \mu\text{m}^2$ with an adjusted $R^2 = 0.9974$. Unlike for voltage variation, the diffusion coefficient was not readily attainable from β_1 for detection length variation. If one assumes β_1 is only due to diffusion, then the DNA diffusion coefficient was overestimated by three orders magnitude. This overestimation was caused by β_1 being not only due to diffusion, but having contributions from wall interactions through Equation (4.40).

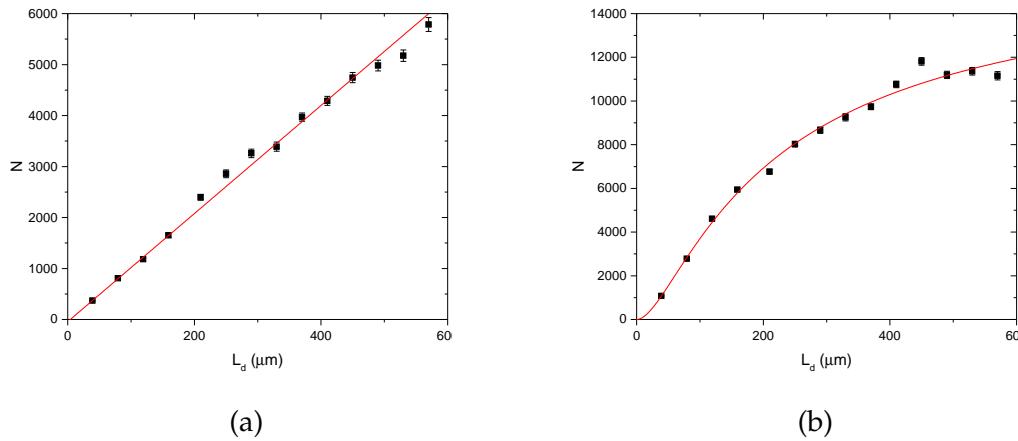


Figure 4.13: (a) L_d vs. theoretical plates for Alexa Fluor 647. The voltage was set at $V_{Applied} = 200$ V. The fitted curve is given by $N = \beta_{N,L} L_d + \alpha$ where $\beta_{N,L} = 10.6 \pm 0.2 \mu\text{m}^{-1}$ and $\alpha = -40 \pm 20$ with $R^2 = 0.99237$. (b) L_d vs. theoretical plates for 900 bp DNA fragments labeled with TOTO-3 intercalating dye. The least-squares fit gave $\beta_2 = (6.0 \pm 0.3) \times 10^{-5}$, $\beta_1 = (1.3 \pm 0.2) \mu\text{m}$, and $\beta_0 = (8.37 \pm 0.5) \times 10^{-1} \mu\text{m}^2$ with an adjusted $R^2 = 0.9974$.

4.4.6 DNA Mobility in Nanochannels

It is well-known that the free-solution electrophoretic mobility of DNA is essentially constant due its free-draining nature [21]. Still, we were interested in using mobility-SCAN to see if DNA lengths could be differentiated via mobility. In order to measure DNA lengths, we first needed to characterize our microscope setup and show that different DNA sizes could be distinguished by burst counts.

Burst Count Analysis for DNA Size Identification

Burst count analysis in submicrometer channels has been previously demonstrated using varying known lengths of DNA up to 48.5 kb λ DNA. Reccius

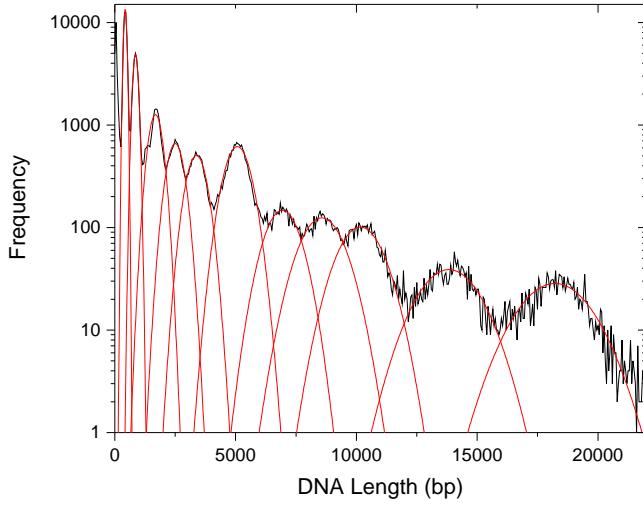
et al [12] were able to differentiate 125 bp DNA from noise and clearly resolve HindIII digested λ DNA based on photon burst histogram analysis. Reccius et al obtained 0.63 counts/bp. Earlier results from Foquet et al [6] were able to discriminate signal from background down to 564 bp and achieve 2.498 counts/bp but were unable to distinguish 130 bp from background. The ability to separate signal from background in single molecule experiments such as these especially rely on effective background rejection by reducing sample contaminants and background from the device which apparently limited Foquet et al.

Both Reccius et al and Foquet et al achieved signal strength orders of magnitude of 1 counts/bp. Experimental variability in the number of detected counts/bp is a function of many parameters including, but not limited to, molecule speed, laser power, choice of dye, emission optics efficiencies, and detector capabilities. Since we were using a different optical setup from the previous generation of Craighead members, it seemed prudent to test our detection efficiency relative to theirs and hopefully achieve comparable or better results.

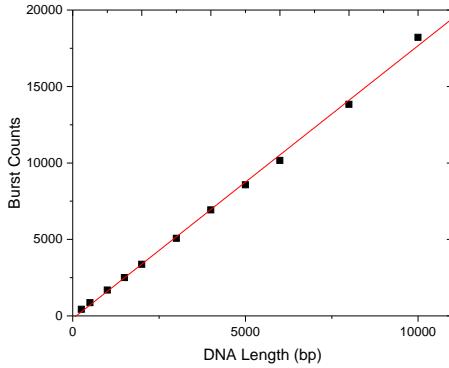
A 1 kb DNA ladder with an added 250 bp fragment was analyzed using SCAN with the resulting photon burst histogram shown in Figure 4.14a. The peaks correspond to DNA lengths of 0.25, 0.5, 1.0, 1.5, 2.0, 3.0, 4.0, 5.0, 6.0, 8.0, and 10.0 kb. Peaks were fit with Gaussian functions to determine their mean burst counts. Figure 4.14b shows the mean burst counts plotted against DNA length. The linear fit gave a slope of 1.79 ± 0.03 counts/bp which is comparable to both Reccius et al and Foquet et al.

As shown in Figure 4.14c, the ratio of the standard deviation to the mean decayed as $N_{bp}^{-0.17}$ implying that resolution improved with increasing DNA length unlike gel electrophoresis. The variability of DNA bursts counts were depen-

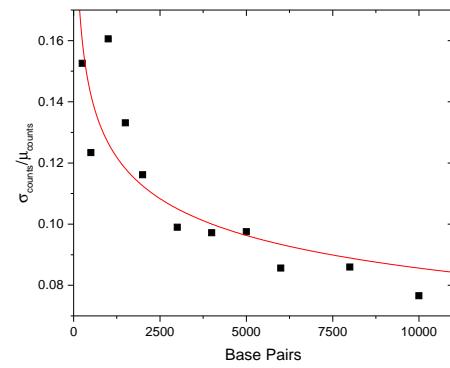
dent on photon counting statistics weighted by Poisson dye number fluctuations. Theoretically, this variability would be super-Poissonian which would imply that the ratio of the standard deviation to mean would scale as a power law and would decay slower than a Poisson distribution. This is exactly what was found experimentally.



(a)



(b)



(c)

Figure 4.14: (a) Burst count distribution for a 1 kb DNA ladder with an additional 250 bp fragment added. (b) Burst counts vs. DNA length. (c) Ratio of the standard deviation to the mean counts vs. DNA length. The resolution improves with length.

We then analyzed the mobility of a sample containing 1.0, 5.0, 10.0, 20.0, and 48.5 kb fragments with our straight nanochannels at $|\vec{E}| = 0.99$ kV/cm. Figure 4.15a illustrates that we were easily able to identify the fragments with excellent resolution based on burst counts. Since each fragment size burst count distribution did not overlap with adjacent distributions, we were able to compute migration time distributions for each DNA length in a straightforward manner. Figure 4.15b shows the result of this computation. Considering that bulk DNA has a mobility that is essentially independent of length, we were pleasantly surprised at the separation between different migration time distributions. For example, the mean migration times between 1.0 kb and 48.5 kb fragments differed by 8%.

Atto 647N was processed after recording the DNA sample using the same channel and input voltage. The resulting peak was fit using an exponentially modified Gaussian (EMG) distribution giving a mean migration time of $\bar{t}_m = 32.7 \pm 0.3$ ms. The electroosmotic mobility was found using Equation (4.24) to be $\mu_{eo} = (7.8 \pm 0.5) \times 10^{-9}$ m²V⁻¹s⁻¹. This result is consistent with other reports using glass microchips and capillaries for the given PVP concentration [19]. Finally, we computed the electrophoretic mobility using Equation (4.24) which is shown in Figure 4.15c. The obtained values for electrophoretic mobilities gave $\mu_{ep} \approx 2.6 \times 10^{-8}$ m²V⁻¹s⁻¹ which are consistent with bulk solution mobilities found in literature [21].

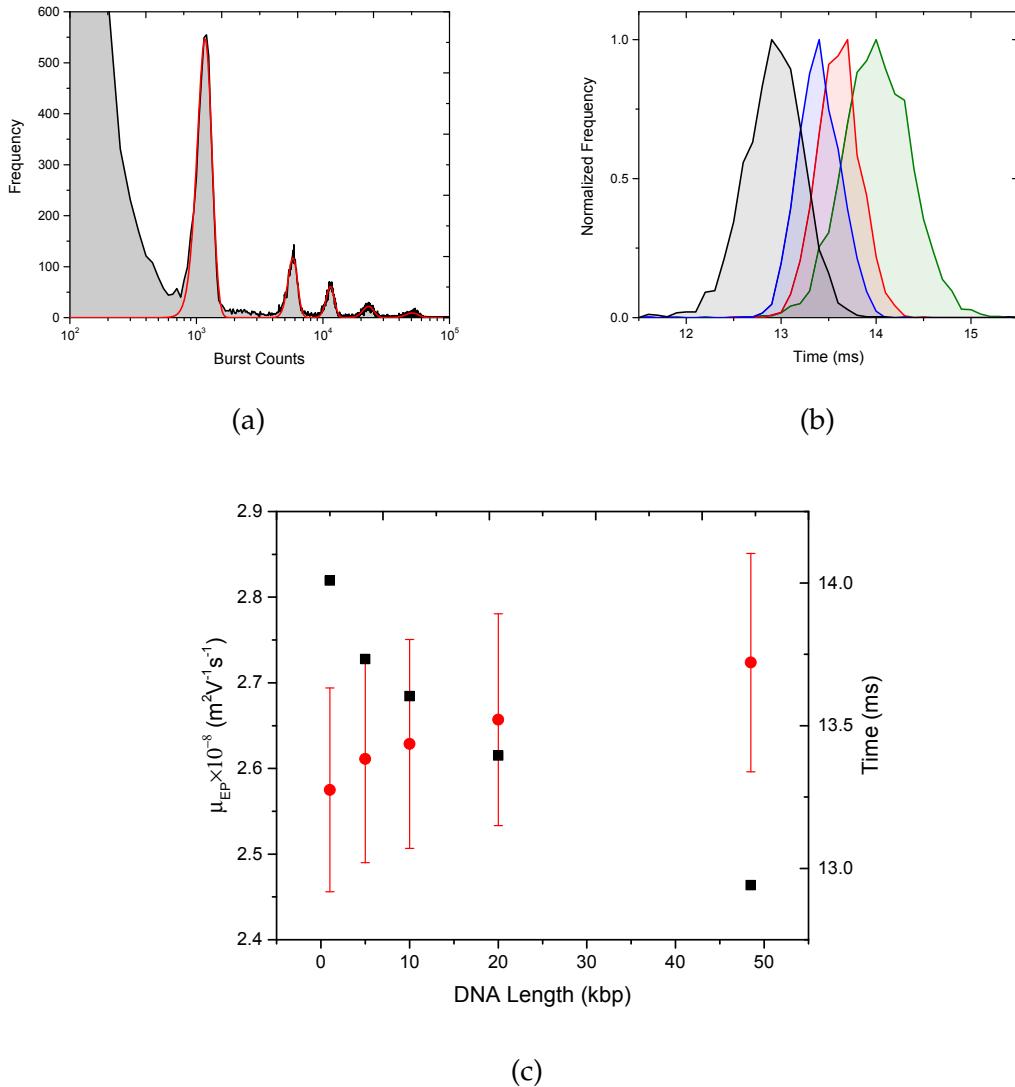


Figure 4.15: (a) Burst count distribution showing 1 kb, 5 kb, 10 kb, 20 kb, and 48.5 kb DNA fragment peaks. A significant amount of noise in the range of 300 counts and less existed due to short fragments of DNA in solution. SNR=10. (b) Experimental electrophoretogram for λ (black), 20 kb (blue), 10 kb (red), and 1 kb (green) DNA fragments. 5kb was not plotted for readability. (c) Electrophoretic mobility (red, left scale bar) and migration time (black, right scale bar) vs. DNA length

4.4.7 Single Dye Mobility Resolution

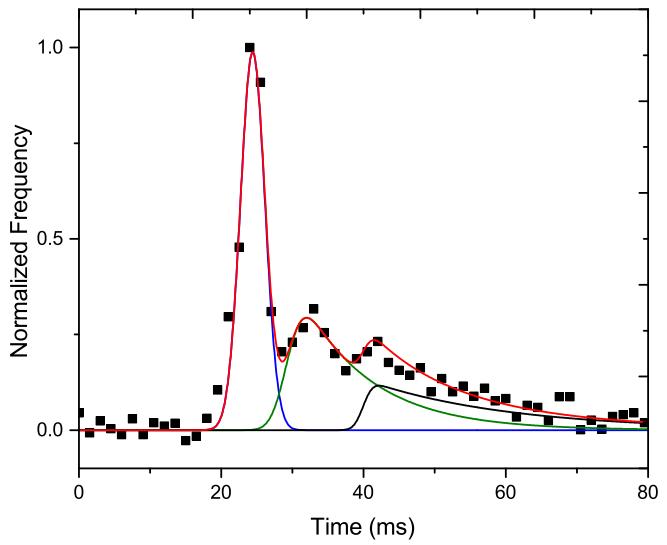
Previously, we noted in Section 4.2.2 that for $|\mu_{eo}| \gg |\mu_{ep}|$, maximum mobility resolution in time would occur for small near-neutral molecules. For the single dye resolution mobility experiments, we chose to work with wheat germ agglutinin (WGA) since it was near neutral with a $pI \approx 9$ [22]. Further, since WGA is extensively used in cell biology for fluorescence imaging of membranes, it was already commercially available with conjugated Alexa Fluor 647. Wheat germ agglutinin (WGA) is a family of plant lectins aimed at defending wheat against insects, bacteria, and yeast. As its name suggests, WGA agglutinates cells by binding to N-acetylglucosamine and sialic acid residues which are found on many membrane glycoproteins. WGA exists as a heterodimer with a molecular weight of ranging from 36-43 kDa [23].

As we noted before, WGA has a $pI \approx 9$. While isoelectric points give the pH of where the biomolecule is neutral, it gives no information of the magnitude of charge on the molecule when deviating from this point. Therefore, using the sequences of the 3 variants of WGA, we estimated the average charge of the WGA variants at a pH of 7.5 to be $q \approx -1.4e$ which doesn't agree with literature with regards to the isoelectric point. However, deviations from the estimated charge from experiments exist as the algorithm calculated the charge by assuming all protein residues had the same pKa as their isolated counterparts which wouldn't be absolutely valid for folded proteins. This calculation suggested to us that WGA would be nearly neutral at pH of 7.5. The near neutrality with a relatively small protein gave us the best shot to observe mobility differences with a single dye resolution.

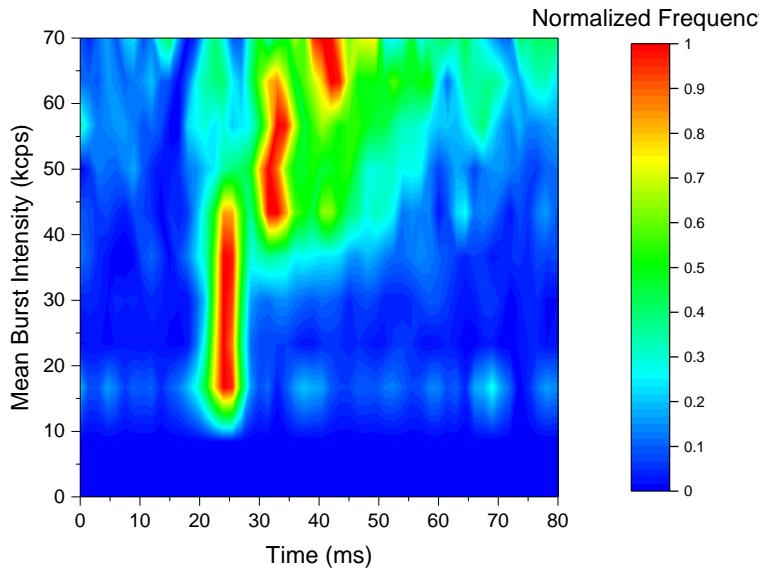
Alexa Fluor 647 labeled WGA was analyzed using SCAN with our U-shaped

channels at $L_d = 90 \mu\text{m}$ and $|\vec{E}| = 3080 \text{ V/cm}$. We recorded data at other detector-to-detector channel lengths but found $L_d = 90 \mu\text{m}$ gave us the best resolution for the measurement as it balanced the peak broadening from wall interactions from the other length dependent resolution gains. As shown in Figure 4.16a, we observed three distinct peaks in the electrophoretogram which we attributed to 1, 2, and 3 dyes per molecule. Since the WGA was near neutral, flow was from cathode to anode. Due to this flow, as the number of negatively charged dyes increased, the migration times increased as well. The peaks were fit using an exponentially modified Gaussian (EMG) function due to the significant positive skewness especially for the slower migration peaks. The mean of the EMG fits were found to be $\bar{t}_{m1} = 24.4 \text{ ms}$, $\bar{t}_{m2} = 39.1 \text{ ms}$, $\bar{t}_{m3} = 60.0 \text{ ms}$ whereas peak maximums were located at $24.4 \pm 0.1 \text{ ms}$, $33.0 \pm 0.5 \text{ ms}$, and $41.4 \pm 0.4 \text{ ms}$ respectively.

In order to conclusively demonstrate that the different peaks corresponded to varying number of dyes, we analyzed the mean burst counts in conjunction with the migration times. We used mean burst counts instead of total burst counts, which is dependent on migration time, as our metric to remain unbiased. The result of the analysis is shown in Figure 4.16b. Three different distributions were observed with increasing mean burst counts as migration time increases signaling additional dyes per molecule and single dye mobility resolution.



(a)



(b)

Figure 4.16: (a) Migration time distribution of WGA with exponential modified Gaussian fits for each distribution. These results are consistent with a Poisson distributed degree of labeling with a mean of $\mu = 1.93$. (b) Migration time/mean burst counts contour plot. The three distributions highlighted in red correspond to the three peaks observed in (a) signifying single dye mobility resolution.

Due to the Poisson statistics regarding dye labeling, we should observe that behavior when analyzing the area under the electrophoretogram curves shown in Figure 4.16a. We indeed had a Poissonian relationship between areas signaling again that the mobility shifts were caused by single dye differences. This can be seen in Figure 4.17. The fit gave a degree of labeling of 1.93 ± 0.01 dyes/molecule. This degree of labeling was also confirmed using two alternative sources. Using spectrophotometric measurements, we found a degree of labeling of 2 dyes/molecule. The certificate of analysis from Thermo Fisher likewise gave a degree of labeling of 2 dyes/molecule.

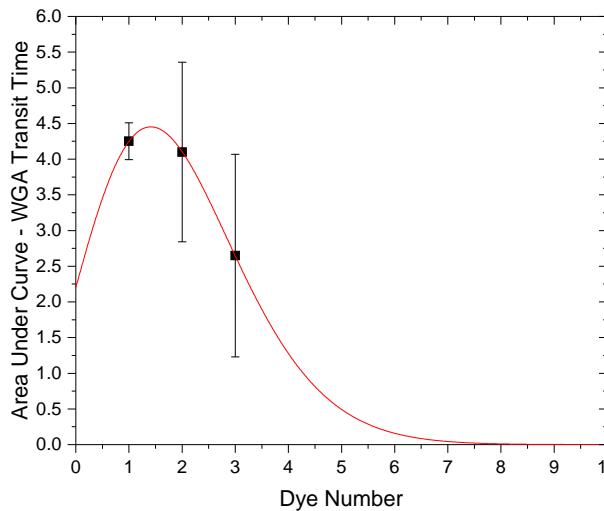


Figure 4.17: Unnormalized Poisson distribution. The fit gave a degree of labeling of 1.93 ± 0.01 dyes/molecule.

For an EOF of $\mu_{eo} = 8 \times 10^{-9} \text{ m}^2 \text{V}^{-1} \text{s}^{-1}$ typically observed in our nanochannels with our standard SCAN buffer, we would observe a mean migration time of approximately 37.5 ms. This suggests that the WGA-dye complex changed from a positively charged molecule to a negatively charged molecule as its dye load increases from one to three.

After closer inspection of Figure 4.16a, we saw a potential fourth peak centered at $t_m = 68$ ms. Unfortunately, due to the low counts and the noise associated with it, no distribution was observed in the migration time-mean burst intensity contour plots. However, in order to suggest that this was due to WGA molecules with 4 dyes, we will make two arguments. First, the RMS noise is $N_{RMS} = 0.02$ whereas the peak is 0.05 above the fitted EMG giving us a 2.5σ event. Secondly, by modeling the system using the Stokes law approximation, we will show that the fourth peak should be near the migration time observed.

The migration time is given by

$$t_m = \frac{L_d L_c}{\left(\mu_{EO} + \frac{q}{6\pi\eta R}\right)\alpha V_a} \quad (4.73)$$

where we have taken as convention migration from cathode to anode as positive. After rearrangement of Equation (4.73), we find

$$\frac{1}{t_r} = \frac{\alpha V_a}{L_d L_c} \left(\mu_{EO} + \frac{q}{6\pi\eta R} \right) \quad (4.74)$$

With the assumption that additional dye loading doesn't modify the Stokes radius, we have

$$\frac{1}{t_r} = \frac{\alpha V_a}{L_d L_c} \left(\mu_{EO} + \frac{q_{WGA} + nq_{Dye}}{6\pi\eta R} \right) \quad (4.75)$$

which factors as

$$\frac{1}{t_r} = \frac{q_{dye}\alpha V_a}{6\pi\eta R L_d L_c} n + \frac{\alpha V}{L_d L_c} \left(\mu_{EO} + \frac{q_{WGA}}{6\pi\eta R} \right) \quad (4.76)$$

Hence, if our crude model is correct, then plotting the inverse migration time with respect to dye number should give a linear relationship. Due to the significant tailing in the EMG distributions, the peaks were Gaussian fitted to determine the peak maximum locations. They were found to be at 24.4 ± 0.1 ms, 33.0 ± 0.5 ms, and 41.4 ± 0.4 ms. Figure 4.18 shows the resulting linear fit

using $1/\max(t_m) = \beta n + \alpha$ with $\beta = -8.5 \pm 0.6 \text{ s}^{-1}$ and $\alpha = 49 \pm 1 \text{ s}^{-1}$. Using these fitting parameters, the predicted peak for WGA molecules with 4 dyes should be located a $65 \pm 4 \text{ ms}$. The observed peak is located within 5% of the predicted peak and falls within the uncertainty of the fit and measurement.

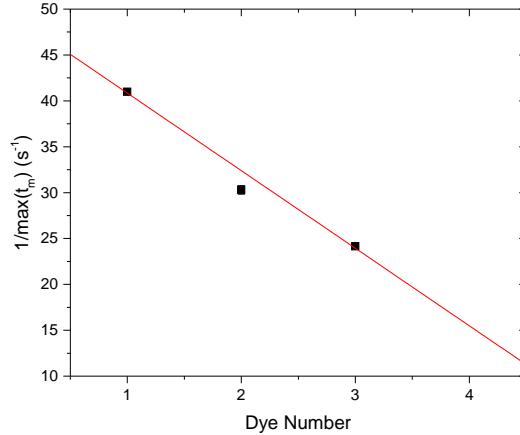


Figure 4.18: Dye number vs. $1/\max(t_m)$. The linear relationship predicts the peak for 4 dyes should be located a $65 \pm 4 \text{ ms}$.

4.4.8 Chromatin Mobility Measurements

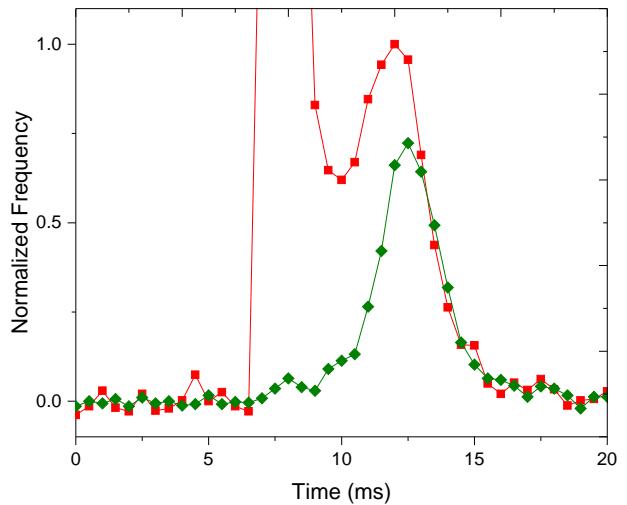
Due to its phosphate backbone, DNA is a highly negatively charged polyelectrolyte. Each phosphate group contributes one negative charge (two charges per bp) at physiological pH [24]. On the other hand, nucleosomes consist of 160-220 positive charges depending mainly on the basic amino acids. Structurally as a group, the nucleosome core particle (NCP) consists of a negatively charged core of approximately $-240e$ and eight relatively free and flexible histone tails with a positive charge of roughly $90e$. All told, on average, the NCP has a net charge $-150e$ [25]. The inter-nucleosome spacing varies from 10-70 bp of linker DNA depending on the cell type and chromosomal loci [26]. Therefore, taking

the median of 40 bp of linker DNA for analysis, the associated NCP and linker DNA complex has an approximate 40% reduction in charge relative to its free DNA counterparts from $-380e$ to $-230e$. This analysis suggests that free DNA and chromatin could be separately identified on a single molecule level using mobility differentiation.

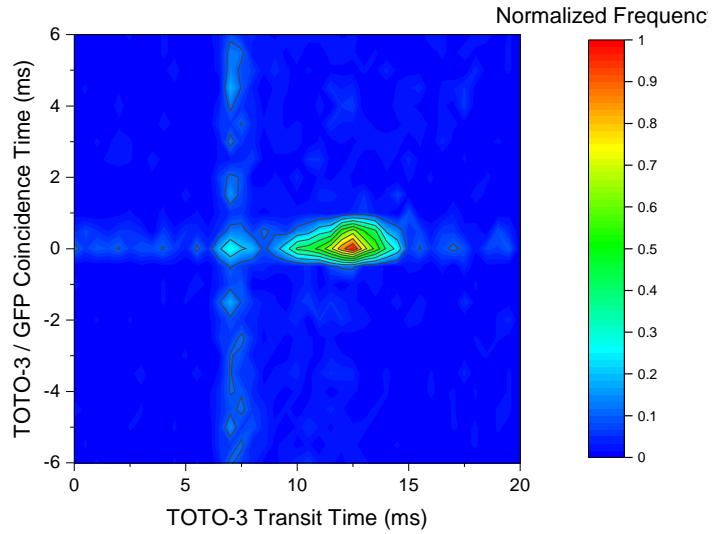
In order to demonstrate that chromatin and DNA exhibit different mobilities which can be distinguished using SCAN, we utilized HeLa-GFP which has a H2B-GFP transgene as the sample of choice due to its ability to identify chromatin from free DNA by the GFP signal and not having to rely on additional complications caused by the introduction of anti-histone antibodies. Initial studies used the 50 μm straight devices as they were less prone to clogging. We performed the experiments at 1.32 kV/cm for electric field strength.

Figure 4.19a illustrates a bimodal distribution from the TOTO-3 channel (red) which gave a faster DNA band and a slower chromatin band as expected. The GFP (green) exhibited a single band which coincided with the second TOTO-3 peak and signaled intact chromatin. However, the TOTO-3 chromatin peak center was shifted slightly relative to the GFP peak. This was most likely due to partial chromatin dissociation which caused GFP counts to drop below the SNR=10 threshold. Laser misalignment was calculated using incoming and outgoing coincidence measurements and estimated at most to be a 0.5 bin width (250 ms). The Gaussian peak fitting parameters are given in Table 4.5.

Being able to identify chromatin from DNA based on mobility would be not only beneficial for quantitative SCAN but also for SCAN-sort. We found that DNA had a probability of detection of 99.3% and chromatin had a probability of detection of 97.1% based on mobility alone.



(a)



(b)

Figure 4.19: (a) Migration time distributions of TOTO-3 labeled trinucleosome HeLa-GFP chromatin. TOTO-3 (red) has a bimodal distribution signifying a faster DNA band and a slower chromatin band. GFP (green) only has a single peak which is colocalized with the second TOTO-3 peak. The red chromatin peak is skew slightly which is most likely due to partly chromatin dissociation. The plots were normalized to the peak of the red chromatin peak to keep the ratio between green and red events constant. (b) Contour plot of TOTO-3/GFP coincidence and TOTO-3 transit time showing that coincident chromatin molecules correspond to the second peak only. A 2D Gaussian fit of the observed peak gave at $\bar{t}_{m,2Dx} = 12.3 \pm 0.2$ ms.

Parameter	Value (ms)	Standard Error (ms)
$\bar{t}_{m,r1}$	7.69	0.01
$\sigma_{m,r1}$	0.46	0.01
$\bar{t}_{m,r2}$	11.70	0.04
$\sigma_{m,r2}$	1.52	0.05
$\bar{t}_{m,g2}$	12.56	0.02
$\sigma_{m,g2}$	1.17	0.02
$\bar{t}_{m,2Dx}$	12.25	0.02
$\sigma_{m,2Dx}$	1.45	0.02

Table 4.5: DNA/Chromatin Peak Fitting Parameters with Gaussian fits.

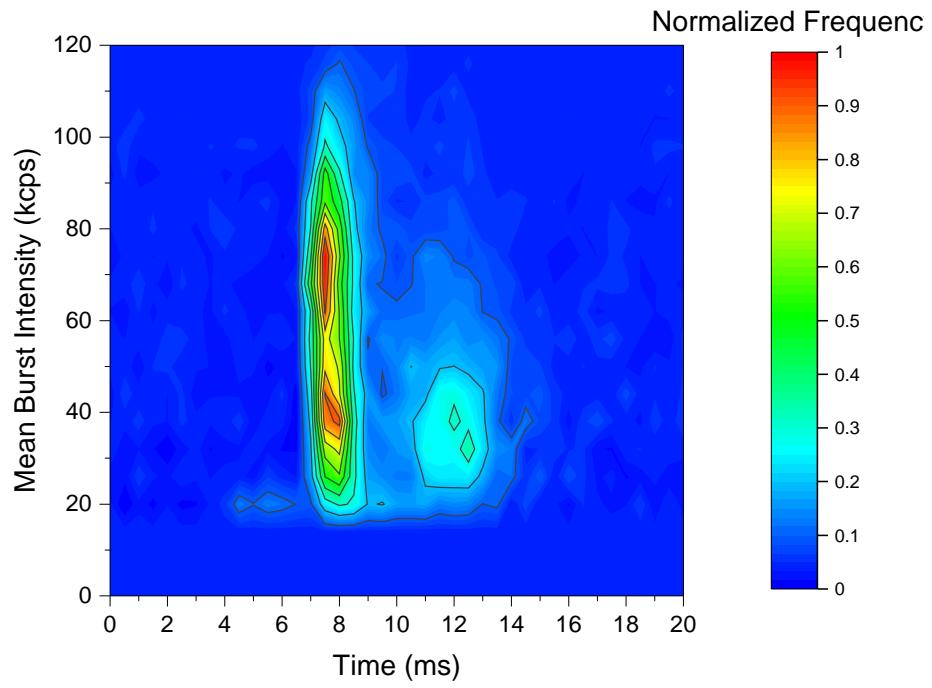
As previously noted, having both mobility and burst intensity data substantially increases the available information since conditional probabilities and statements can be investigated. We used this technique to show that chromatin is under-intercalated relative to DNA of the same length. Using the data previously shown in Figure 4.19, we computed both migration times and mean burst intensity of each molecule. Figure 4.20a shows the contour plot of the result which illustrates that DNA is much brighter than chromatin and that both the DNA and chromatin peaks have subpopulations within them. We fit the chromatin and DNA peaks using 2D Gaussian functions with the results shown in Table 4.6. The chromatin peak was 33% more dim than DNA based on mean burst intensity which proves that chromatin is not intercalated as efficiently as DNA of the same size.

Parameter	Value	Standard Error
$x_{c,DNA}$	7.69 ms	0.01 ms
$\sigma_{xc,DNA}$	0.48 ms	0.02 ms
$y_{c,DNA}$	56.6 kcps	0.8 kcps
$\sigma_{yc,DNA}$	26.6 kcps	0.9 kcps
θ_{DNA}	0.008	0.002
$x_{c,Chromatin}$	11.81 ms	0.04 ms
$\sigma_{xc,Chromatin}$	1.33 ms	0.05 ms
$y_{c,Chromatin}$	38.0 kcps	0.4 kcps
$\sigma_{yc,Chromatin}$	15.0 kcps	0.4 kcps
$\theta_{Chromatin}$	6.281	0.008

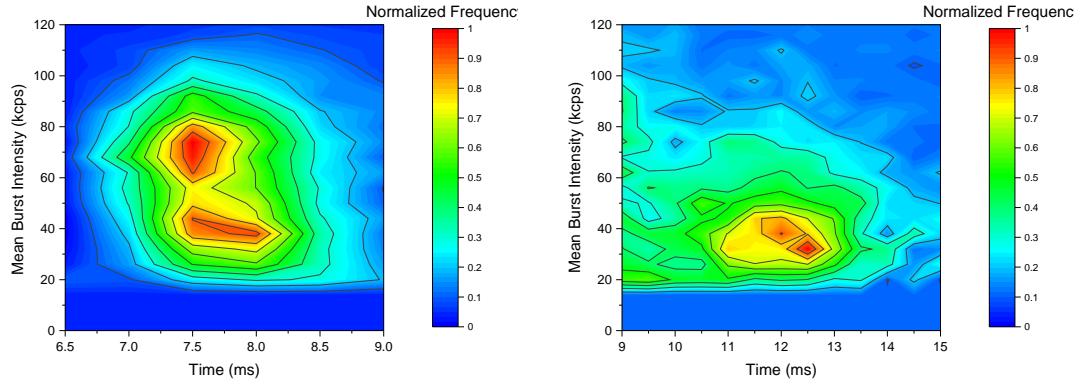
Table 4.6: DNA/chromatin peak fitting parameters with 2D Gaussian fits.

As shown in Figures 4.20b and 4.20c, both the DNA peak and chromatin peak had subpeaks within their structure. We will first consider the DNA subpeaks which were fitted using 2D Gaussians with fitting parameters given in Table 4.7. We hypothesize that the two peaks correspond to di- and trinucleosomes. Based on the differences in DNA length of di- and trinucleosomes, we expect the mean burst intensity of trinucleosomes to be approximately 1.6 times greater than dinucleosomes. The fits of the two peaks give a value of 1.76 which is a deviation of less than 10%. Further, the samples contained near equal molar concentrations of di- and trinucleosomes. This is reflected in Figure 4.20b and the far right gel lane in Figure 3.17 of which the sample was derived.

The chromatin band also had two peaks within it. Since the chromatin has multiple lengths and attached nucleosomes, the analysis is less straightforward compared to DNA. With the faster peak centered at 12 ms being the brighter of the two, this may suggest that its population consisted of chromatin that was not fully chromatinized as some nucleosomes may have dissociated.



(a)



(b)

(c)

Figure 4.20: (a) Normalized contour plot of TOTO-3 migration time vs. mean burst intensity. The DNA and chromatin peaks are clearly visible. (b) Zoomed in normalized DNA peak showing two subpopulations. We hypothesize that the two peaks correspond to di- and trinucleosomes. (c) Zoomed in normalized chromatin peak showing two subpopulations. We conjecture that the dominant peak corresponding to the slowest migration represents fully chromatinized chromatin while the other faster peak is composed of partially dissociated chromatin.

Parameter	Value	Standard Error
$x_{c,DNA1}$	7.76 ms	0.01 ms
$\sigma_{xc,DNA1}$	0.49 ms	0.02 ms
$y_{c,DNA1}$	42.3 kcps	0.6 kcps
$\sigma_{yc,DNA1}$	15.6 kcps	0.7 kcps
θ_{DNA1}	0.008	0.002
$x_{c,DNA2}$	7.62 ms	0.01 ms
$\sigma_{xc,DNA2}$	0.45 ms	0.02 ms
$y_{c,DNA2}$	74.6 kcps	0.8 kcps
$\sigma_{yc,DNA2}$	19.2 kcps	0.9 kcps
θ_{DNA2}	0	0.001

Table 4.7: DNA subpeaks fitting parameters with 2D Gaussian fits.

Chromatin Crosslinking While crosslinking is necessary to maintain bound antibody-chromatin complexes, it is also necessary to stabilize the chromatin itself. To show the effects of crosslinking, we compared crosslinked chromatin to non-crosslinked chromatin taken from the same trinucleosome chromatin source as shown in Figure 4.21. The crosslinked chromatin was crosslinked for 15 minutes with 1.0% formaldehyde. Both samples were labeled with TOTO-3 and processed with mobility-SCAN. Of the detected molecules at an SNR=10, the crosslinked chromatin sample consisted of 56% chromatin. For the non-crosslinked chromatin sample, the amount of chromatinized material dropped to 34% indicating dissociation occurred due to concentration or ionic strength effects.

Moreover, the crosslinking itself caused a mobility shift to occur most prominently in the chromatinized material. The chromatin peak centers shifted from 11.9 ± 0.4 ms to 10.83 ± 0.03 ms, a difference of just over 9%. We have observed crosslinked mediated mobility shifts with other protein too. We hypothesize the shift is due to the crosslinking of positive side chain amino acids such as lysine or arginine which have ammonium cations. Crosslinking of these amino acids

with formaldahyde decreases the net charge of the crosslinked complex [27].

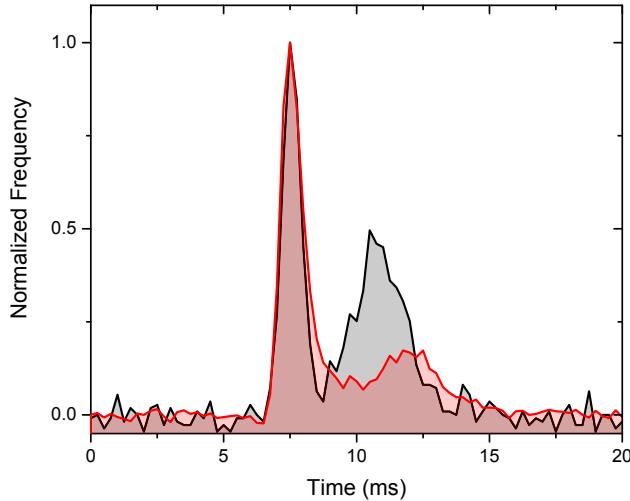


Figure 4.21: Electrophoretograms of TOTO-3 labeled trinucleosome chromatin with crosslinking (black) and no crosslinking (red).

4.4.9 Mobility Based Identification of 6 Biomolecules

We have shown that molecules can be identified by their electrophoretic mobility instead of their fluorescent signature. Using the U shaped nanochannels, we wanted to further demonstrate the potential of SCAN mobility based methods for identifying various molecules using a complex mixture. We also wanted to experimentally test the degree of resolution changes between different bands due to increases in channel length. For these experiments, we used 6 different molecules outlined in Table 4.8.

We analyzed each individual species separately using SCAN for this proof of principle experiment. We did this using the same channel by sequential processing each biomolecule. Channels were washed between samples with standard

Biomolecule	M_w (kDa)	pI	Degree of Labeling
DNA (1kb)	660	4.3	1:5 dye to bp
Bovine Serum Albumin (BSA)	66.5	4.7	4
Soybean Agglutinin (SBA)	120	6.0	2
Mouse IgG1	150-170	7.3±1.2	-
ATTO 647N	0.746	-	-
Mouse IgG2b	150-170	7.3±1.2	2

Table 4.8: Biomolecules used for mobility based species identification using complex mixture experiment and their relevant physical parameters.

SCAN buffer until no prior sample could be detected. Further, any possibility of the EOF changing over time was controlled for by observing ATTO 647N migration times between runs.

We did this sequential processing of individual molecules for three main reasons. First, overlaps in migration time distributions could possibly create minor ambiguities in analysis and we wanted to provide conclusive evidence for peak separation. Second, processing each material separately allowed for normalization of each peak and removed any difficulties regarding creating a mixture of approximately equal concentrations. Third, if mobility based methods were going to be used to analyze real world samples, then prior knowledge of their migration time distributions would have to be known. We essentially performed this experiment to discover their independent migration time distributions.

The biomolecules were analyzed at three different locations to explore the effect of channel length on peak separation. These locations corresponded to a nanochannel distance of $40\ \mu\text{m}$, $240\ \mu\text{m}$, and $570\ \mu\text{m}$. With the detection lengths varying over an order of magnitude, we ensured that any effects associated with channel length would be observed since the resolution should increase rather

slowly as $L_d^{1/2}$ for a diffusion limited regime. However, with the long channel lengths and a diverse population of molecules, we expected resolution to increase more slowly, if at all, due to predominant peak tailing with some analytes.

As shown in Figures 4.22a, 4.22b, and 4.22c, each individual peak was well resolved for all detection lengths. Four of the molecules flowed from the anode towards the cathode: DNA (black), BSA (orange), SBA(blue), (red) Mouse IgG1. Mouse IgG2b (purple) and the neutral ATTO 647N (green) flowed in the opposite direction. Notice that their migration times were correlated to their respective isoelectric points. Nonetheless, this correlation would generally not be true for all instances as pI only gives the pH of molecule neutrality and gives no indication of the absolute magnitude of the charge away from the pI. If one desired, the charge for each molecule could have been roughly estimated using known pK_a values of the isolated residues and functional groups since the protein sequence or molecular structure for each biomolecule is widely known.

Visual inspection of the electrophoretograms gave limited indication of any resolution benefits to longer channels. To confirm this, we chose anode-to-cathode flowing DNA, BSA, and SBA to calculate their associated resolutions between one another. We neglected Mouse IgG1 in the calculations due to its large peak asymmetries. Defining the resolution as

$$R = \frac{t_{m2} - t_{m1}}{0.5(W_1 + W_2)} \quad (4.77)$$

where the baseline bolus width was taken as $W_i = 4\sigma$ for $i = 1, 2$, we found some evidence for a positive correlation between channel length and resolution as shown in Figure 4.22d. The resolution of the BSA/SBA increased by 16% from $L_d = 40\mu\text{m}$ to $L_d = 570\mu\text{m}$ whereas the DNA/BSA resolution peaked at the

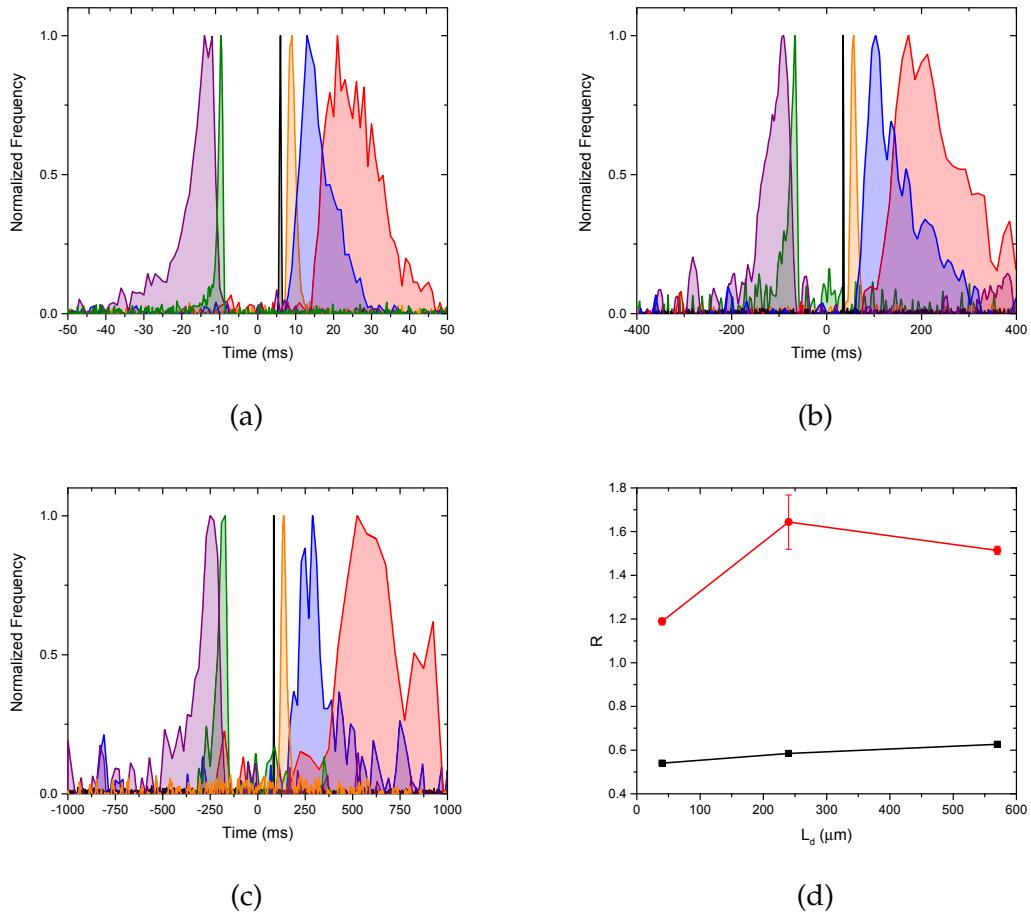


Figure 4.22: Electrophoretograms for the 6 biomolecules over varying detector lengths. Positive time represents flow from the anode to the cathode and negative time represents flow from cathode to anode. (a) $L_d = 40\mu\text{m}$ (b) $L_d = 240\mu\text{m}$ (c) $L_d = 570\mu\text{m}$. (d) Resolution of DNA/BSA (red) and BSA/SBA (black).

intermediate channel length with a 38% increase from $L_d = 40\mu\text{m}$ to $L_d = 240\mu\text{m}$.

Since we envision using mobility to differentiate single molecules instead of using distinctly colored fluorophores, we wanted to determine how well would could do this with our 6 molecule mixture with $L_d = 240\mu\text{m}$. We computed the probability of detection for all molecules relative to some other molecule using maximum a posteriori (MAP) decision criteria which is given by [28]

$$\frac{p_{Y|H_1}(y|H_1)}{p_{Y|H_0}(y|H_0)} \stackrel{H_1}{>} \frac{P_0}{P_1} \quad (4.78)$$

where $p_{Y|H_i}(y|H_i)$ are the probability density functions of the Y corresponding to hypothesis $i = 0, 1$ in a binary detection scheme. The MAP decision rule minimizes the probability of error in the binary decision. After computing the likelihood ratios for all combinations of migration time distributions and assuming $P_0 = P_1$ implying equal molecule concentrations, we compiled the probabilities of detection as shown in Table 4.9. Most peaks were well resolved with near 100% detection probability for all but two cases involving SBA/IgG1 and ATTO 647N/IgG2b owing to the large amount tailing of SBA into the IgG1 distribution and the ATTO 647N into IgG2b.

	DNA	BSA	SBA	IgG1	ATTO 647N	IgG2b
DNA	-	100	100	100	100	100
BSA	99.9	-	99.5	100	100	100
SBA	100	98.5	-	58.8	100	100
IgG1	100	99.9	83.9	-	100	100
ATTO 647N	100	100	100	100	-	73.0
IgG2b	100	100	100	100	82.4	-

Table 4.9: Probability of detection. The table should be read as the probability of detection for a molecule in the left-hand column with regards to a binary decision involving a molecule in the top row.

4.4.10 Affinity Electrophoresis in Nanochannels

Affinity electrophoresis techniques are based on the measurement of changes in electrophoretic mobilities through molecule interactions or complex formation. One of the most popular assays of this type is the electrophoretic mobility shift assay (EMSA) which detects protein-DNA, protein-RNA, and protein-protein interactions via binding-induced mobility shifts visualized in a gel medium. Similarly, affinity capillary electrophoresis addresses the same interactions but

uses the high separation efficiency of capillary electrophoresis. These methods are also capable of quantitatively measuring binding constants and the stoichiometry for the interactions.

Motivated by affinity capillary electrophoresis, we wanted to determine if our SCAN mobility system was able to detect mobility shifts upon binding of different analytes especially with regards to chromatin interactions with various epigenetic probes. In the following, we will describe our results pertaining to DNA-MBD shifts, chromatin-antibody shifts, and a more complex mixture containing BSA, α -BSA antibody, and a secondary antibody.

4.4.11 Affinity Electrophoresis of DNA-MBD1 Interactions

Methyl-CpG-binding domain protein 1 (MBD1) binds specifically to methylated DNA which causes repression of gene expression at the loci. DNA methylation, the most well studied epigenetic mark, can therefore be analyzed using SCAN via fluorescently labeled MBD1 proteins.

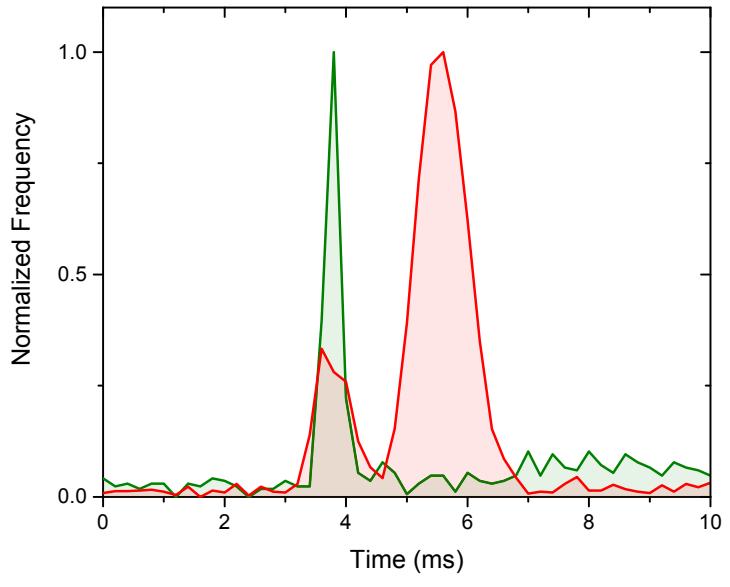
Due to the DNA's exceedingly charged state (2 negative charges per bp at physiological pH) and its large size as measured in metrics such as radius of gyration relative to proteins, we expected little to any shift in the DNA's electrophoretic mobility upon protein binding. Due to this, we figured that if we had any chance of detecting mobility shift due to binding, it would be with DNA and a basic protein such as MBD1. Therefore, we took a high mass DNA ladder, methylated it with CpG methyltransferase (M.SssI), bound the Alexa Fluor 647 MBD1 probe to it, and ran it on SCAN.

While we didn't observe any discernible binding shift with the DNA, we did find that the DNA-MBD1 complex was shifted considerably relative to unbound MBD1 giving a bimodal distribution for the Alexa Fluor 647 channel (red) as shown in Figure 4.23a. To further prove that the overlapping DNA-MBD1 peak was indeed due to binding, we analyzed the input DNA-MBD1 coincidence in conjunction with the MBD1 migration time and found that DNA-MBD1 complexes were only associated with the fastest peak as shown in Figure 4.23b.

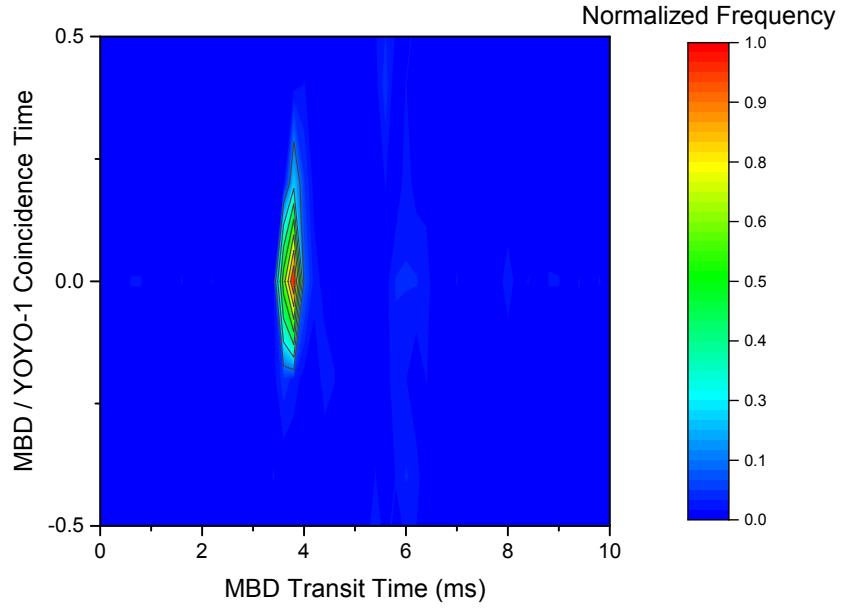
The resulting electrophoretograms in Figure 4.23 were fitted using Gaussian functions. The single DNA peak gave fitting parameters of $\bar{t}_g = 3.775 \pm 0.004$ ms and $\sigma_g = 0.122 \pm 0.003$ ms while the two MBD peaks gave $\bar{t}_{r1} = 3.76 \pm 0.01$ ms, $\sigma_{r1} = 0.28 \pm 0.01$ ms, $\bar{t}_{r2} = 5.574 \pm 0.004$ ms, and $\sigma_{r2} = 0.411 \pm 0.004$ ms. The overlapping MBD and DNA peaks were within 1% of each other with regard to peak center. A two dimensional Gaussian fit of the contour data had $\bar{t}_x = 3.740 \pm 0.002$ ms and $\bar{t}_y = 0.0270 \pm 0.004$. A summary is given in Table 4.10.

Parameter	Value (ms)	Standard Error (ms)
\bar{t}_{r1}	3.76	0.01
σ_{r1}	0.28	0.01
\bar{t}_{r2}	5.574	0.004
σ_{r2}	0.411	0.004
\bar{t}_g	3.775	0.004
σ_g	0.122	0.003
$\bar{t}_{x,2D}$	3.740	0.002
$\bar{t}_{y,2D}$	0.0270	0.004

Table 4.10: Summary of DNA/MBD fitting parameters with Gaussian fits



(a)



(b)

Figure 4.23: Electrophoretograms of a MBD/DNA mixture. (a) Electrophoretograms of Alexa Fluor 647 labeled MBD (red) and YOYO-1 labeled DNA (green). Overlapping peaks at 3.8 ms suggested a binding induced mobility shift of MBD1 while the red secondary represented unbound MBD. DNA only had 1 measurable peak consisting of both unbound and bound molecules. (b) Contour plot illustrating that the 3.8 ms peak consisted of MBD/DNA bound complexes.

It was somewhat unfortunate that the DNA's mobility didn't shift upon MBD1 binding as that would have allowed for the elimination of MBD1 being labeled and would have provided quasi-label free detection of epigenetic marks. Further, it would have freed up the emission spectra for another labeled protein. However, the DNA-MBD1 mobility shift experiment was important in the development of mobility-SCAN as it was our first demonstration of these binding-induced changes on a single molecule level.

4.4.12 Affinity Electrophoresis of Chromatin- α -H3 Antibody Interactions

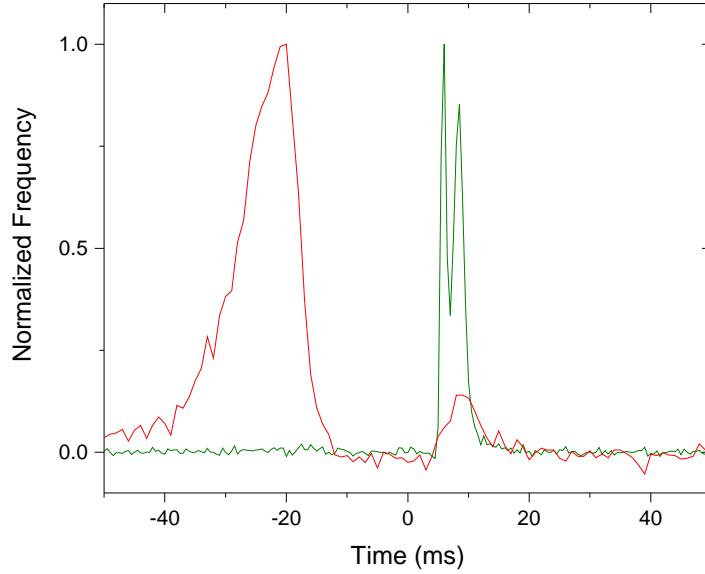
SCAN relies heavily on antibodies for detection of epigenetic marks. We saw in the mobility based separation of 6 biomolecules in Section 4.4.9 that we can have antibodies flow in either direction using the standard SCAN buffer at pH=7.5. We therefore investigated if we could use a binding-induced flow reversal to filter out unbound antibodies which have been plaguing SCAN since inception. We took an Alexa Fluor 647 labeled α -H3 antibody which we knew flowed from cathode to anode and bound it to YOYO-1 labeled trinucleosome chromatin using a 10-fold molar excess of antibody. We then loaded both the cathode and anode ports with sample and processed the sample using SCAN. We observed two peaks in both the chromatin and antibody crosscorrelations with the antibody crosscorrelation suggesting flow reversal. After analyzing the single molecule data, we found that the antibody indeed reversed flow. As exhibited in Figure 4.24a, the YOYO-1 chromatin signal (green) gave two peaks characteristic of free DNA and chromatin in its migration time distribution. The Alexa Fluor 647 sig-

nal representing the antibody gave two oppositely flowing peaks with one peak colocalized with the chromatin peak.

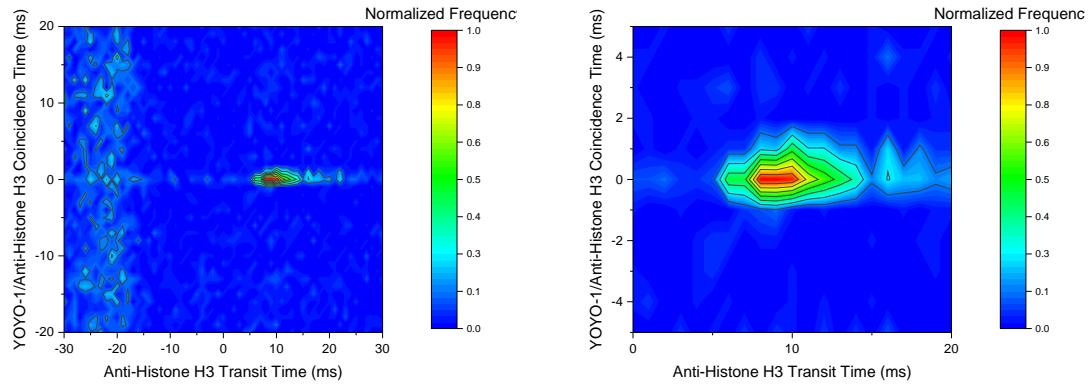
The peaks for the YOYO-1 distribution were fit using Gaussians and found to have centers located at $\bar{t}_{g1} = 5.876 \pm 0.006$ ms and $\bar{t}_{g2} = 8.34 \pm 0.01$ ms with standard deviations of $\sigma_{g1} = 0.421 \pm 0.007$ ms and $\sigma_{g2} = 0.93 \pm 0.01$ ms. The colocalized peak of the antibody was found to have a center at $\bar{t}_r = 9.3 \pm 0.1$ ms with a standard deviation of $\sigma_r = 2.5 \pm 0.1$ ms. The colocalized antibody peak was broader than the chromatin peak, but essentially located only under the slower chromatin. The chromatin peak didn't widen accordingly because the relative proportion of bound antibody-chromatin complexes to total chromatin was low. To show that the antibody only bound to the chromatin molecules, we looked at coincidence YOYO-1/Alexa Fluor 647 molecule and mapped them to YOYO-1's migration time. The results given in Figures 4.24b and 4.24c show a single peak centered at $\bar{t}_x = 9.67 \pm 0.07$ ms and $\bar{t}_y = 0.26 \pm 0.04$. A summary is given in Table 4.11.

Parameter	Value (ms)	Standard Error (ms)
\bar{t}_{g1}	5.876	0.006
σ_{g1}	0.421	0.007
\bar{t}_{g2}	8.34	0.01
σ_{g2}	0.93	0.01
\bar{t}_r	9.3	0.1
σ_r	2.5	0.1
$\bar{t}_{x,2D}$	9.67	0.07
$\bar{t}_{y,2D}$	0.26	0.04

Table 4.11: Summary of Chromatin/ α -H3 fitting parameters with Gaussian fits



(a)



(b)

(c)

Figure 4.24: (a) Electrophoretograms of the α -H3 antibody (red) and the YOYO-1 (green). Two peaks are shown for the YOYO-1 which coincide to DNA and chromatin bands respectively. The α -H3 peak also has two bands but of opposite direction of migration. Binding causes a reversal of flow direction for the α -H3. Positive time represents flow from the anode to cathode. (b) Contour plot of α -H3 migration time against α -H3 YOYO-1 coincidence time. A peak is observed centered at $\bar{t}_x = 9.67 \pm 0.07$ ms signifying that YOYO-1/ α -H3 coincidences corresponds to the chromatin peak. (c) Zoomed in contour plot of b.

4.4.13 Affinity Electrophoresis of BSA-Primary Ab-Secondary Ab Interactions

Using SCAN to detect unlabeled biomolecules in solution would be a powerful tool for diagnostic purposes. Since the biomolecules are unlabeled, a fluorescent labeled antibody or other reagent specific for the biomolecule of interest would need to bind to the molecule and cause a binding-induced mobility shift which could be detected using SCAN. In order to demonstrate the potential of this method, we used BSA labeled with Alexa Fluor 647, an unlabeled α -BSA primary antibody, and a Alexa Fluor 488 labeled secondary antibody in our proof-of-principle experiments. While it would have been insightful to have the primary antibody labeled, we neglected to do so because our optical setup for mobility measurements was limited to only two spectrally distinct dyes. It didn't hinder the resulting analysis and the relative mobility of the primary antibody was inferred from the experiments. Plus, in actual real-world experiments which may more often than not employ secondary antibodies in their detection to minimize labeling and reagent costs. Another benefit for using a primary and secondary antibody for detection lies in the fact that larger mobility shifts can be obtained as our data shows.

Due to the complexity of the mixture and lack of label on the primary antibody, we needed to perform the actual BSA/primary Ab/secondary Ab experiment along with control experiments involving BSA only, secondary Ab only, BSA/ primary Ab, and primary Ab/ secondary Ab. The central result demonstrating that we detected mobility shifts of all the molecules implying the possibility of using SCAN for unlabeled molecule detection is given in Figure 4.25. We deduced from the data that BSA was the most anionic, followed by the

primary antibody, and then the secondary antibody even though the primary antibody wasn't detectable using fluorescence as it was unlabeled.

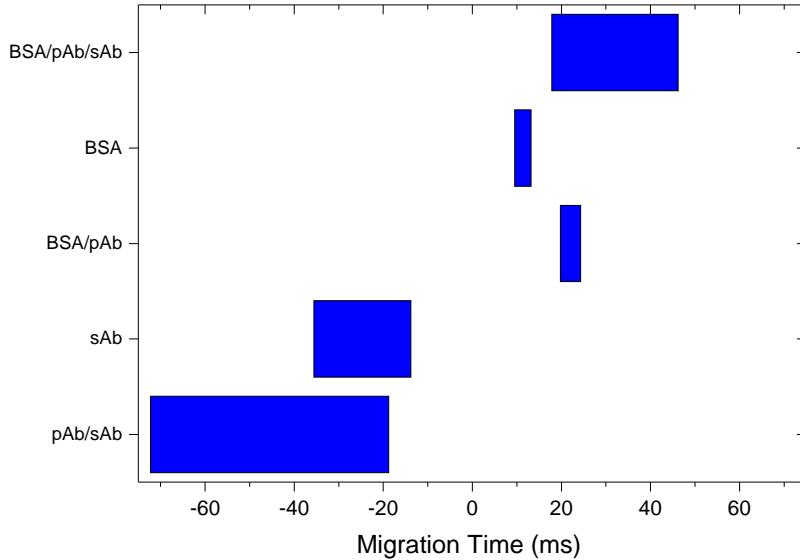


Figure 4.25: Migration times for the single molecule complexes. Boxes represent $\bar{t}_m \pm \sigma_m$ for each band.

The electrophoretograms for the main experiment consisting of BSA, α -BSA, and the secondary antibody are shown in Figure 4.26. After performing the binding reaction and crosslinking reactions, the sample was processed using SCAN. BSA labeled with Alexa Fluor 647 exhibited two distinct peaks in its migration distribution. The fastest peak was from completely unbound BSA as confirmed with a control experiment with only BSA shown in Figure 4.27a (red). The BSA control experiment gave a mean migration time of $\bar{t}_{BSAcon} = 9.70 \pm 0.05$ ms whereas the peak center of the fastest peak BSA in the main experiment was found to be $\bar{t}_{r1} = 9.68 \pm 0.06$. A second BSA control experiment with the unlabeled primary α -BSA antibody was performed to determine its influence. The α -BSA antibody was bound at a 500 fold molar excess to push the equilibrium to

complete binding along with reducing the chance to produce precipitates due to a precipitin reaction. Since the α -BSA antibody was unlabeled, we didn't need to be concerned for its molar excess as they were undetectable in SCAN. Relative to BSA, the BSA/ α -BSA complex was shifted towards a longer migration time as shown in Figure 4.27a (black). Due to the high molar excess used in the binding, essentially no free BSA was observed. The BSA/ α -BSA complex gave a migration time distribution described by an EMG with a mean migration of $\bar{t}_{BSApAb} = 22.06 \pm 0.04$ ms and a standard deviation of $\sigma_{BSApAb} = 2.26 \pm 0.08$ ms.

The main experiment gave a second peak with regards to the Alexa Fluor 647 signal with a mean of $\bar{t}_{r2} = 26.8 \pm 0.5$ ms and a standard deviation of $\sigma_{r2} = 7.2 \pm 0.5$ ms when fitted to an EMG. The slower second peak was composed of both BSA/ α -BSA complexes and BSA/ α -BSA/secondary Ab complexes. Confirmation of the BSA/ α -BSA complexes within the second peak stemmed from the already mentioned BSA/ α -BSA control experiment. The existence of BSA/ α -BSA/secondary Ab complexes was established by looking at the Alexa Fluor 488 secondary antibody signal. The Alexa Fluor 488 secondary antibody signal gave two peaks in its electrophoretogram as shown in Figure 4.26b. One was colocalized with the slower peak of the Alexa Fluor 647 suggesting binding while the other was flowing in the opposite direction. Verification of BSA/ α -BSA/secondary Ab complexes was achieved by looking at input coincidence of Alexa Fluor 488 and 647 and correlating them to the secondary antibody transit time as illustrated in 4.26c. Moreover, we inferred that the Alexa Fluor 488 peak corresponding to cathode to anode flow consisted of both free secondary antibody and primary Ab/ secondary Ab complexes by the large amount of tailing in the peak relative to a secondary Ab control given in Figure 4.27b.

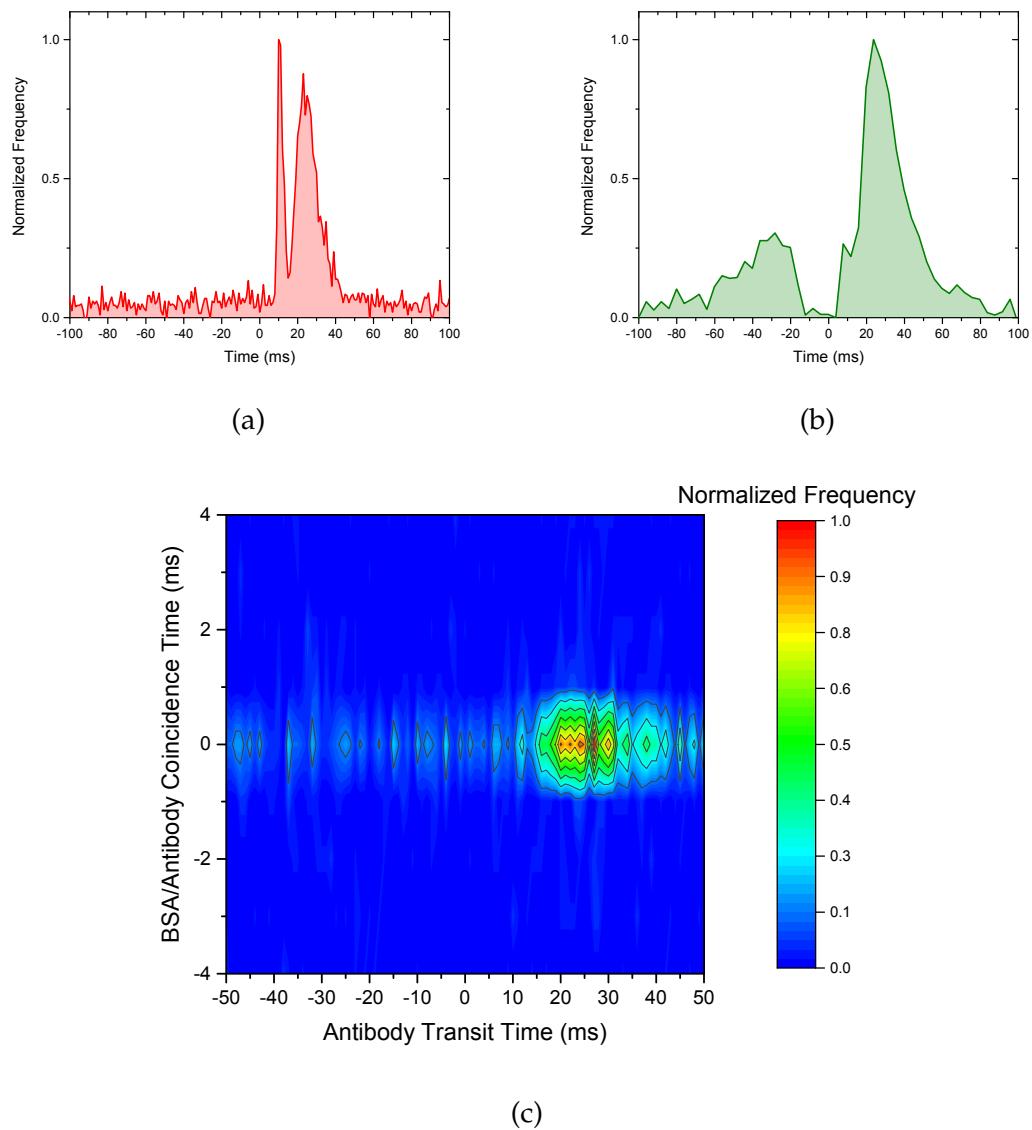


Figure 4.26: Electrophoretograms of a BSA/ α -BSA/secondary Ab mixture. Positive time corresponds to anode to cathode flow. (a) Electrophoretogram of Alexa Fluor 647 labeled BSA. The fastest peak corresponds to free BSA. The slower peak is a combination of BSA/ α -BSA/secondary Ab complexes and BSA/ α -BSA complexes. (b) Electrophoretogram of Alexa Fluor 488 labeled secondary Ab. The positive time peak corresponds to BSA/ α -BSA/secondary Ab complexes. The negative time peak corresponds to α -BSA/secondary Ab complexes and free secondary Ab. (c) Contour plot illustrating that BSA/ α -BSA/secondary Ab complexes are indeed coincident.

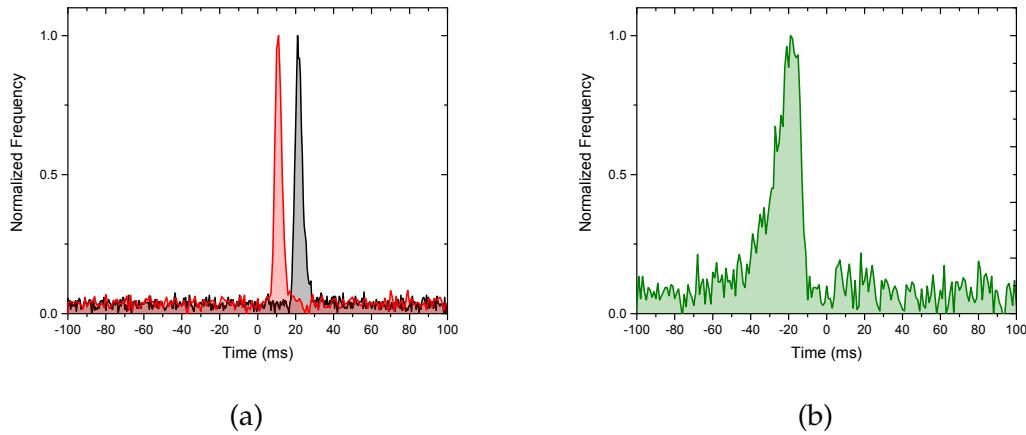


Figure 4.27: Electrophoretograms of control experiments. (a) BSA (red) and BSA/α-BSA (black) (b) Secondary Ab.

4.5 Conclusions

We presented a device and technique, mobility-SCAN, which rapidly evaluated single molecule mobilities in a fused silica nanochannels as they traveled between two inspection volumes. The single molecule mobilities were generated without physical separation typically seen in microchip capillary electrophoresis. Instead, we measured each molecule's migration time individually and constructed the distributions through the compilation of many such trajectories. With our system, 2-color detection was utilized for both spectral and mobility measurements giving single molecule information not previously available to just spectral measurements. Mobility-SCAN proved capable of measuring physical parameters such as electrophoretic mobilities, diffusion coefficients, effective molecular charge, and the Stokes radius using single molecule ensemble averaging. All single molecule mobility parameter results were comparable to those obtained by other methods such as FCS or capillary electrophoresis. We investigated the maximum practical limits for channel lengths and voltages for mobility-SCAN and produced theoretical plates above 12,000 for DNA. While

we altered the detection lengths and voltages for various measurements, we didn't change the pH or the surface chemistries of the device. Depending on the application and molecules involved, these two parameters give additional freedom to the experimenter for better distribution separations. Using near-neutral molecules, we demonstrated the capability of detecting single dye variations by mobility differences alone, proving, in some sense, that we reached a fundamental limit of the technique.

A major element of the device's utility proved to be its ability to differentiate DNA, chromatin, proteins, and dye based only on mobility. We took 6 distinct biomolecules (DNA, proteins, and dye) and were able to determine 83% of the molecules correctly based on mobility alone within the complex mixture. We had complete segregation for many of the species. By distinguishing molecules based on mobility instead their fluorescent emission spectra, we could diminish the number of distinct fluorophores needed for experiments. Finally, the system was used to investigate binding interactions of various complexes similar to affinity capillary electrophoresis, but on a single molecule level. We envision that binding induced mobility shifts could be used to detect unlabeled antigens on a single molecule level, perhaps allowing for detection of scarce antigens not detected by other means.

BIBLIOGRAPHY

- ¹C. L. Colyer, T. Tang, N. Chiem, and D. J. Harrison, "Clinical potential of microchip capillary electrophoresis systems", *Electrophoresis* **18**, 1733–1741 (1997).
- ²N. A. Lacher, K. E. Garrison, R. S. Martin, and S. M. Lunte, "Microchip capillary electrophoresis/electrochemistry", *Electrophoresis* **22**, 2526–2536 (2001).
- ³D. Milanova, R. D. Chambers, S. S. Bahga, and J. G. Santiago, "Electrophoretic mobility measurements of fluorescent dyes using on-chip capillary electrophoresis", *Electrophoresis* **32**, 3286–3294 (2011).
- ⁴V. Dolnik, S. Liu, and S. Jovanovich, "Capillary electrophoresis on microchip", *Electrophoresis* **21**, 41–54 (2000).
- ⁵M. Hashimoto, K. Tsukagoshi, R. Nakajima, K. Kondo, and A. Arai, "Microchip capillary electrophoresis using on-line chemiluminescence detection", *Journal of chromatography A* **867**, 271–279 (2000).
- ⁶M. Foquet, J. Korlach, W. Zipfel, W. W. Webb, and H. G. Craighead, "Dna fragment sizing by single molecule detection in submicrometer-sized closed fluidic channels", *Analytical Chemistry* **74**, 1415–1422 (2002).
- ⁷M. Foquet, J. Korlach, W. R. Zipfel, W. W. Webb, and H. G. Craighead, "Focal volume confinement by submicrometer-sized fluidic channels", *Analytical chemistry* **76**, 1618–1626 (2004).
- ⁸B. R. Cipriany, R. Zhao, P. J. Murphy, S. L. Levy, C. P. Tan, H. G. Craighead, and P. D. Soloway, "Single molecule epigenetic analysis in a nanofluidic channel", *Analytical chemistry* **82**, 2480–2487 (2010).

⁹B. R. Cipriany, P. J. Murphy, J. A. Hagarman, A. Cerf, D. Latulippe, S. L. Levy, J. J. Benítez, C. P. Tan, J. Topolancik, and P. D. Soloway, "Real-time analysis and selection of methylated dna by fluorescence-activated single molecule sorting in a nanofluidic channel", Proceedings of the National Academy of Sciences **109**, 8477–8482 (2012).

¹⁰S. M. Stavis, J. B. Edel, K. T. Samiee, and H. G. Craighead, "Single molecule studies of quantum dot conjugates in a submicrometer fluidic channel", Lab on a Chip **5**, 337–343 (2005).

¹¹S. M. Stavis, S. C. Corgié, B. R. Cipriany, H. G. Craighead, and L. P. Walker, "Single molecule analysis of bacterial polymerase chain reaction products in submicrometer fluidic channels", Biomicrofluidics **1**, 034105 (2007).

¹²C. H. Reccius, S. M. Stavis, J. T. Mannion, L. P. Walker, and H. Craighead, "Conformation, length, and speed measurements of electrostatically stretched dna in nanochannels", Biophysical Journal **95**, 273–286 (2008).

¹³S. M. Stavis, J. B. Edel, Y. Li, K. T. Samiee, D. Luo, and H. G. Craighead, "Single-molecule mobility and spectral measurements in submicrometer fluidic channels", Journal of Applied Physics **98**, 044903 (2005).

¹⁴D. Henry, "The cataphoresis of suspended particles. part i. the equation of cataphoresis", in Proceedings of the royal society of london a: mathematical, physical and engineering sciences, Vol. 133 (), pp. 106–129.

¹⁵P. D. Grossman and J. C. Colburn, *Capillary electrophoresis: theory and practice* (Academic Press, 2012).

¹⁶M. L. Marina, A. Ríos, and M. Valcárcel, *Analysis and detection by capillary electrophoresis*, Vol. 45 (Newnes, 2005).

- ¹⁷X. Huang, W. F. Coleman, and R. N. Zare, "Analysis of factors causing peak broadening in capillary zone electrophoresis", *Journal of Chromatography A* **480**, 95–110 (1989).
- ¹⁸M. Minarik, B. Gas, A. Rizzi, and E. Kenndler, "Plate height contribution from wall adsorption in capillary zone electrophoresis of proteins", *JOURNAL OF CAPILLARY ELECTROPHORESIS* **2**, 89–96 (1995).
- ¹⁹D. Milanova, R. D. Chambers, S. S. Bahga, and J. G. Santiago, "Effect of pvp on the electroosmotic mobility of wet-etched glass microchannels", *Electrophoresis* **33**, 3259–3262 (2012).
- ²⁰A. Loman, C. Müller, F. Koberling, W. Richtering, and J. Enderlein, "Absolute and precise measurements of the diffusion of small fluorescent dye molecules across the visible spectrum", in 14th international workshop on single molecule spectroscopy and ultrasensitive analysis in life science, poster, berlin, germany () .
- ²¹B. J. Kirby, *Micro-and nanoscale fluid mechanics: transport in microfluidic devices* (Cambridge University Press, 2010).
- ²²P. D. Rye and N. Bovin, "Carbohydrate affinity page for the study of carbohydrate-binding proteins", *BioTechniques* **25**, 146–151 (1998).
- ²³A. Allen, A. Neuberger, and N. Sharon, "The purification, composition and specificity of wheat-germ agglutinin", *Biochem. j* **131**, 155–162 (1973).
- ²⁴G. M. Blackburn, *Nucleic acids in chemistry and biology* (Royal Society of Chemistry, 2006).
- ²⁵T. M. Nordlund, *Quantitative understanding of biosystems: an introduction to biophysics* (CRC Press, 2011).

²⁶S. A. Grigoryev, “Nucleosome spacing and chromatin higher-order folding”, *Nucleus* **3**, 493–499 (2012).

²⁷G. T. Hermanson, *Bioconjugate techniques* (Academic press, 2013).

²⁸H. L. Van Trees, *Detection, estimation, and modulation theory* (John Wiley and Sons, 2004).

CHAPTER 5

**SINGLE MOLECULE SORTING OF DNA AND CHROMATIN IN
NANOFLUIDIC CHANNELS**

5.1 Introduction

We have emphasized throughout this dissertation that novel methods need to be developed to analyze multiple epigenetic marks from low cell counts. We have focused on single molecule techniques to achieve this, signaling a possible paradigm shift away from ChIP-based methods. Results thus far have concentrated on steady state operations of SCAN and have focused on quantifying epigenetic marks without possibility of obtaining the underlying sequence information. For sequence information to be obtained, sample partitioning based on the chromatin's histone modifications and subsequent recovery are needed. The sample partitioning can be achieved using a nanofluidic fluorescence activated single molecule sorter which can electrokinetically sort individual molecules based on their fluorescent identity in real-time.

Previous results from Cipriany et al [1] have already demonstrated that SCAN-sort could potentially be used as a single molecule sorter to separate molecules with the desired epigenetic marks for downstream sequencing. Using MBD1 and a mixture of methylated pML4.2 and unmethylated pUC19 DNA, they reported a 3.5 ± 0.3 enrichment of pML4.2 after sorting for coincidence of MBD1-pML4.2 complexes at approximately 300 molecule/minute input rate.

Here, we present our results regarding the further development of SCAN-

sort. We designed a new sorting device shown in Figure 5.1 which shortened the input and output channels compared to Cipriany et al. These modifications allowed for sorting at higher concentrations and lowered false positives rates as trigger times could be reduced by 75%. Sorting of DNA based on intensity was optimized by minimizing false positives using the new devices at a rate of 2000 molecules/minute. The sorting operation produced a sorted sample that was composed of over 98% of the desired DNA molecules. We further extended the sorting to a chromatin sample and sorted based coincidence of GFP and TOTO-3 at approximately 5000 molecules/minute. We estimated that the chromatin was enriched by over 2-fold by sorting, which provided optimism that SCAN-sort could be used to map epigenetic marks.

The sorting device, however, suffered from intrinsic instabilities caused by the device geometry itself. While stable of short periods of time, device operation suffered from drift which caused backflow or leakage molecules and contributed to sorting errors. To fix this problem, we designed, but never implemented, a five-channel design which, in theory, should solve both backflow and leakage.

While analogous in spirit to fluorescence-activated cell sorting (FACS), single molecule sorting differs from FACS in practice by many factors including actuation mechanisms, throughput, and fluorescence signal levels. Fluorescence signal levels are proportional to the antigen densities on the cell membrane. Typical cells have antigen densities of 10^4 - 10^5 antigens/cell implying that each cell has orders of magnitude more dyes for signal discrimination compared to single molecules which may only have one dye [2]. This directly translates into throughput capabilities as FACS systems can process up to 300,000 cells per

minute while our single molecule sorter can achieve around 3,000 molecules per minute implying that massive parallelization would be needed to achieve relevant molecule counts.

5.2 Materials and Methods

5.2.1 Device Fabrication and Layout

The sorting devices were fabricated following the procedure given in Section 4.3.1. Instead of the “*m*” design shown in Cipriany et al, we opted for a straight design which minimized channel length. The previous devices used by Cipriany et al suffered from long sorting actuation times which reduced molecule throughput and created sorting errors. Moreover, they easily clogged due to their length. Figure 5.1 shows the device layout. The length of the channels were 10 μm for the input, sort, and default channels, each with a width of 500 nm.

5.2.2 Optical and Electrical Setup

The optical setup for SCAN-sort was identical to the mobility-SCAN optics given in Section 4.3.4. We either used two or three beam spots depending on the device. For the straight sorting device shown in Figure 5.1, only two adjacent spots were used to monitor the input and sort while three beam spots gave the opportunity to monitor the input, sort, and default channels if a “*m*” shaped device was used as was done in Cipriany et al. The input beam spot was placed

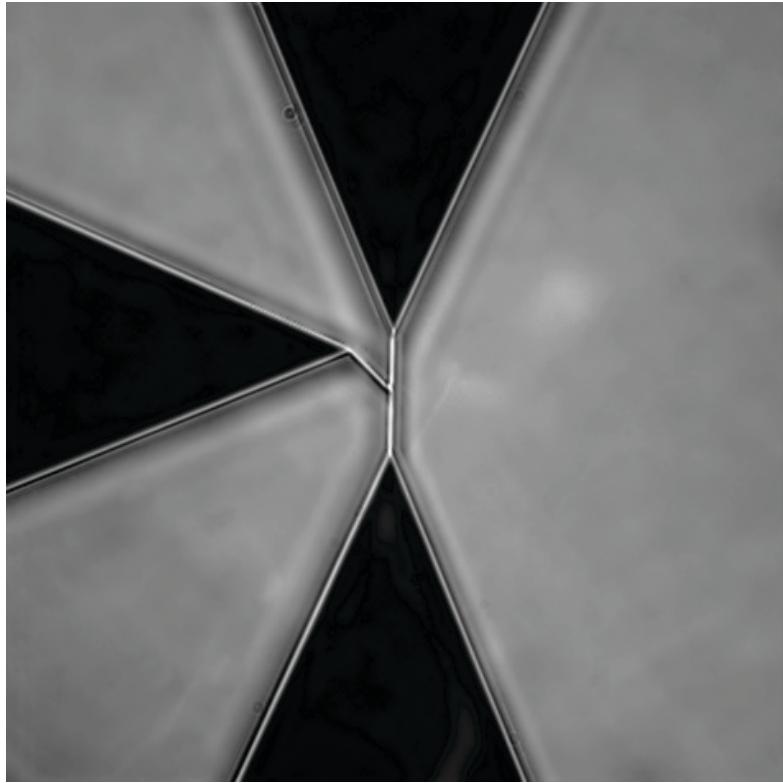


Figure 5.1: Bright field image of straight channel sorting device. The input (bottom), sort (top), and default (left) are all shown. The angle between the default and sort is 45° .

a few microns before the bifurcation to allow for the inherent delay caused by the FIR filtering of the field programmable gate array (FPGA).

Real-time photon counting and signal processing was performed on a Altera DE2-115 Development Board with a Cyclone IV FPGA. Signals from the photon counting module (FastComTec P7888) were sampled every $50\ \mu s$ by the FPGA. Cyclone IV processed the incoming data stream by performing FIR Gaussian filtering, burst detection, and coincidence detection. Depending on the experiment, triggering events were called if the burst height exceeded a certain threshold or if coincidence was detected. If a triggering event was determined, then a sort trigger pulse of a predetermined time duration was applied to solid state relays which actuated the voltage switching in the device. Post-experiment data

analysis using MATLAB was used to ensure proper triggering of the FPGA.

Electrical actuation of the device was achieved by using solid state relays as noted previously. The outputs of the relays were connected to the two output ports of the sorting device. At any given instant, only one relay would be on. One output port would be at high voltage while the other output port would be floating. The voltage of the low voltage port was adjusted via a potentiometer between the two output ports in a Δ network of resistors spanning the three ports (input, sort, and default). Setting the optimal voltage on the low voltage port was non-trivial and required constant monitoring during SCAN-sort operation.

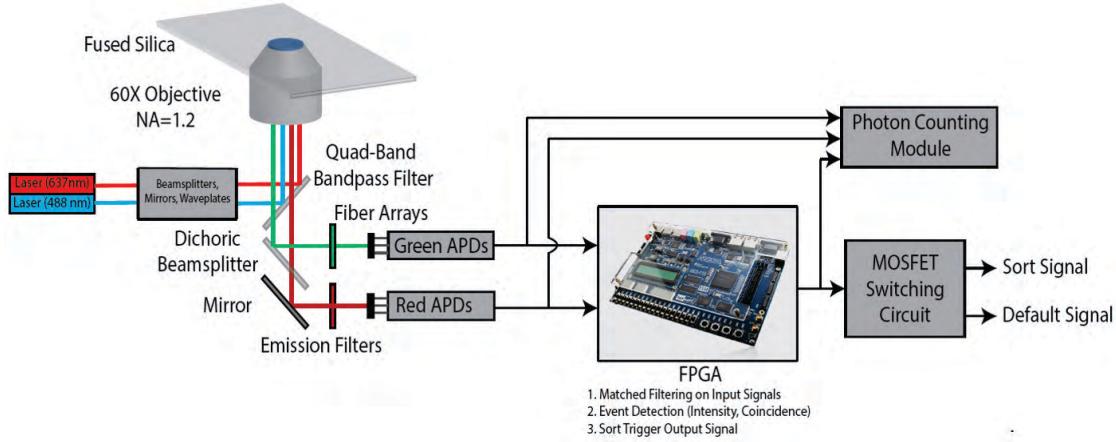


Figure 5.2: Optical and electrical setup for sorting. Real-time sorting was achieved by signal processing with a FPGA

5.2.3 Sample Preparation

DNA Experiments

A solution of equal molar concentrations 1 kb and 0.5 kb dsDNA fragments was prepared using the standard SCAN running buffer. The solution had an overall DNA concentration of 1 ng/ μ L which reduced DNA aggregation during nucleic acid staining. YOYO-1 intercalating dye was added to the solution at a 1 dye : 5 bp ratio ensuring adequate intercalation without DNA fragmentation caused by higher dye ratios. Samples were left to stain overnight at 4 °C. Prior to SCAN-sort, the sample was diluted to approximately 200 pM.

Chromatin Experiments

HeLa-GFP chromatin was isolated by members of the Soloway Lab using the chromatin extraction protocol given in Section 3.2.3. TOTO-3 at a 1 dye : 5 bp ratio was reacted with chromatin at a concentration of 30 ng/ μ L and left overnight at 4°. Prior to running SCAN, the sample was diluted to approximately 200 pM and added to input port the SCAN device

5.3 Results and Discussion

5.3.1 Intensity Based Sorting of DNA

For initial testing of the straight sorting device, we chose DNA to sort due to its brightness and uniform mobility. A 100 pM mixture of approximately equal mo-

lar 0.5 kb and 1.0 kb fragments was loaded into the input reservoir of the straight sorting device. Since our detection optics were arranged linearly, we chose to monitor the input and the sort channels of the device. With the on-voltage set to 5 V, we adjusted the off-voltage via a potentiometer such that the false positives fell below the detectable level which would give an estimated precision ($TP/(TP+FP)$) greater than 98%. The detectable limit was estimated by calculating the standard deviation of the Poisson fluctuations of the migration time background bin counts for a given average background. The off-voltage was set at, on average, $V_{off} = 3.30$ V. However, over the course of the experiment, the device behavior was slightly time dependent and the amount of backflow would increase over time. We therefore had to continually monitor the experiment and adjust the off voltage accordingly. At the beginning of the experiment, we operated at $V_{off} = 3.26$ V and we operated at $V_{off} = 3.33$ V at the end, 50 minutes later. The time-dependent nature of the sorting was also observed by Cipriany et al.

Figure 5.3 shows a 1.0 s time trace of the sorting operation. The trigger was a signal from the FPGA showing that the FPGA recognized that the molecule at the input satisfied the sorting threshold of 350 kcps. Once triggered, the voltage switched for a 5 ms duration. An example of backflow in the sort channel is shown at $t \approx 0.57$ s. Notice that its burst duration of the backflow is much longer than the remaining molecules due to the magnitude of the electric field being much smaller in the off-channel. The time trace also has a false negative at $t \approx 0.15$ s where a molecule above threshold was detected by the FPGA but not subsequently observed in the sort channel.

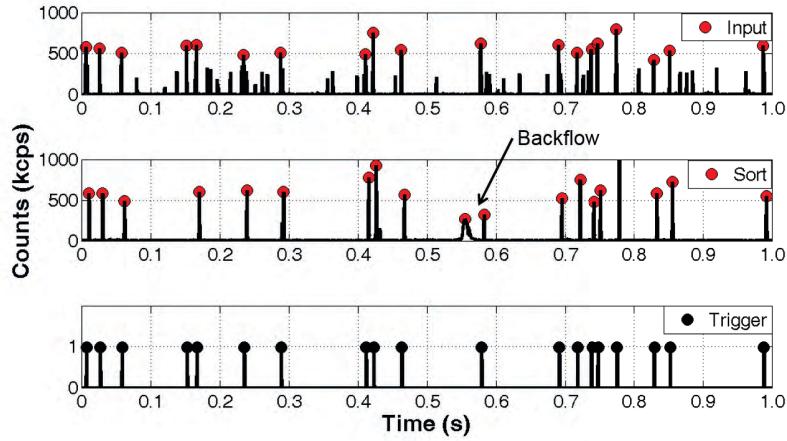


Figure 5.3: Time traces for intensity-based sorting. The sorting threshold was 400 kcps. All molecules were identified correctly by the FPGA. Over the 1.0 s interval, we had 1 false negative and 1 false positive count due to backflow.

With lowering the false positives to below the detectable limit, we, in turn, caused an increase in backflow and a decreased sensitivity (true positive rate). Molecular backflow was differentiated by its large burst duration which could be easily discerned from forward flowing molecules. We estimated that approximately 19% of the detected molecules in the sort were actually molecules flowing backward which reduced sorted throughput. Neglecting backflow, we had a true positive rate ($TPR = TP / (TP + FN) = TP / P$) of roughly 60%.

We found that the trade-off between backflow and leakage (FP) was inherent to the bifurcated device geometry. Leakage caused large increases in false positives as undesired molecules leaked into the sort channel when the default channel was at high voltage. It also caused false negatives as desired molecules entered the default channel when the sort channel was at high voltage. Backflow, on the other hand, effectively removed already sorted molecules, reducing true positives, and increasing false negatives.

These initial experiments demonstrated the precision necessary for sorting. For example, by changing the off-voltage by as little as 3% (10 mV), we went from a high amount of leakage with approximately 50% of the sort molecules being leakage ($FP \approx TP$), to an ideal condition with a slight amount of backflow with good leakage rejection. This slim margin of error suggested that if device operation was going to be robust in the future, a new channel design and better real-time voltage manipulation was probably going to be needed.

5.3.2 Sorting of Chromatin

After optimizing the sorting conditions with DNA, we moved on to sort chromatin for its potential epigenetic profiling utility. Chromatin, though, has additional challenges regarding signal detection which have been described throughout this dissertation. Additionally, chromatin samples contain molecules with two different mobilities with relatively broad distributions which add complications to the dynamics of the sorting operation.

Using HeLa-GFP as the sample of choice, we sorted molecules based on the coincidence condition. Figure 5.4 gives a time trace generated during the experiment. Six coincident molecules were observed over the time span and were sorted correctly. The sort at $t \approx 0.11$ s had false positives due to either with backflow, leakage, sorting pulse molecule overlap, or a combination of all three. Also, the last trigger consisted of two coincident sorting events. Since the two molecules were within the 20 ms sort pulse duration of one another, only one trigger signal was generated at falling edge of the sorting pulse. After approximately twenty minutes of sorting, the device started fouling as aggregates built

up at the entrance to the nanochannel input. Prior to this fouling, the device operated reasonably given the drawbacks of the leakage/backflow trade-off.

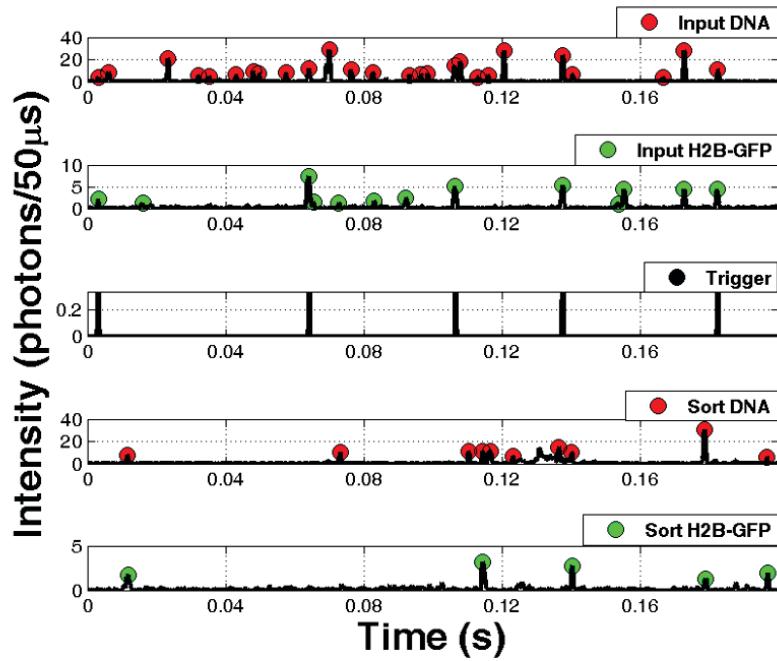


Figure 5.4: Time traces for chromatin sorting. 6 coincident molecules were identified over the 200 ms time frame. All 6 appear to be sorted. The sort at $t \approx 0.11$ s had false positives due to either with backflow, leakage, sort pulse molecule overlap, or a combination of all three.

If we were to process real-world chromatin samples, we would be interested in the enrichment of the sorting. Enrichment is the important sorting metric as false positives would be mapped randomly on the genome whereas true positives would be mapped to sparse histone modification locations. Simply enriching the sample such that the histone modifications are represented at least twice in the sorted output would be sufficient, theoretically, to create a signal above background in the sequence data.

For the chromatin sorting, we calculated the enrichment by

$$\text{Enrichment} = \frac{N_{S,c}/N_S}{N_{I,c}/N_I} \quad (5.1)$$

where $N_{S,c}$ was the total number of coincident molecule in the sort, N_S was the total TOTO-3 molecules detected in the sort, $N_{I,c}$ was the total number of coincident molecule in the input, N_I was the total TOTO-3 molecules detected in the input. We estimated that the sorting achieved at least a 2-fold enrichment of chromatin to total TOTO-3 molecules. The combinations of leakage molecules, backflow, and signal dropping below background all contributed to difficulty in the analysis of the data. However, their contributions would have increased N_S and decreased $N_{S,c}$ and thus we most likely underestimated the actual enrichment.

5.3.3 False Positives with Sorting

The lack of considerable enrichment with the chromatin stemmed, at least partially, from the sorting pulse being too long which enabled ensuing molecules coming down the nanochannel to be sorted with the desired original molecule. Figure 5.5 shows the probability of the false positives occurring as a function of number of input molecules per minute (concentration) and the sorting pulse width, T_{on} . The plot was derived using Poisson statistics for molecule occupation within the sorting volume. For the chromatin sorting experiment in Section 5.3.2, we were processing roughly 5000 molecules per minute and had a T_{on} of 20 ms giving a false positive probability close to 30% (data not shown). For the DNA sorting experiment in Section 5.3.1, we were running at a rate of about 700 molecules/minute and a $T_{on} = 5$ ms which gave a much lower probability ($\approx 2.5\%$) of having false positives.

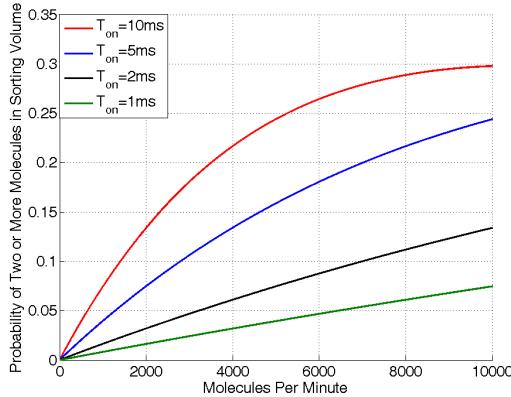


Figure 5.5: False positive probabilities as a function of input molecules per minute and sorting pulse width time.

Due to signal detection issues, chromatin may never be able to be analyzed such the $T_{on} < 5$ ms, even under ideal circumstances. Thus, sorting throughput on a per-channel basis is limited. To increase throughput, massive parallelization would have to be achieved which brings about its own challenges that will be briefly discussed in Section 5.4.

5.3.4 Channel Backflow and Leakage

As already mentioned and demonstrated, the single molecule sorter has a trade-off between backflow and channel leakage (false negatives). Backflow is undesirable as it transfers already sorted molecules to the default and it complicates data analysis. Likewise, channel leakage contributes to missed molecules which are extremely valuable especially for rare epigenetic marks and false positives. Moreover, channel leakage is a function of the size of the molecules themselves as it is dependent on the location in flow of the molecule. With large DNA molecules whose projected volume takes up much of the cross-sectional area of

the channel, the dependence is negligible. For smaller molecules such as short chromatin fragments, the chance of leakage will depend on the path of the individual molecules.

Figure 5.6 demonstrates using COMSOL simulations both backflow (Figure 5.6b) and channel leakage (Figure 5.6c) with the half voltage condition shown in (Figure 5.6a). The simulations solve Laplace's equation with the appropriate boundary conditions. Notice that leakage will be path dependent.

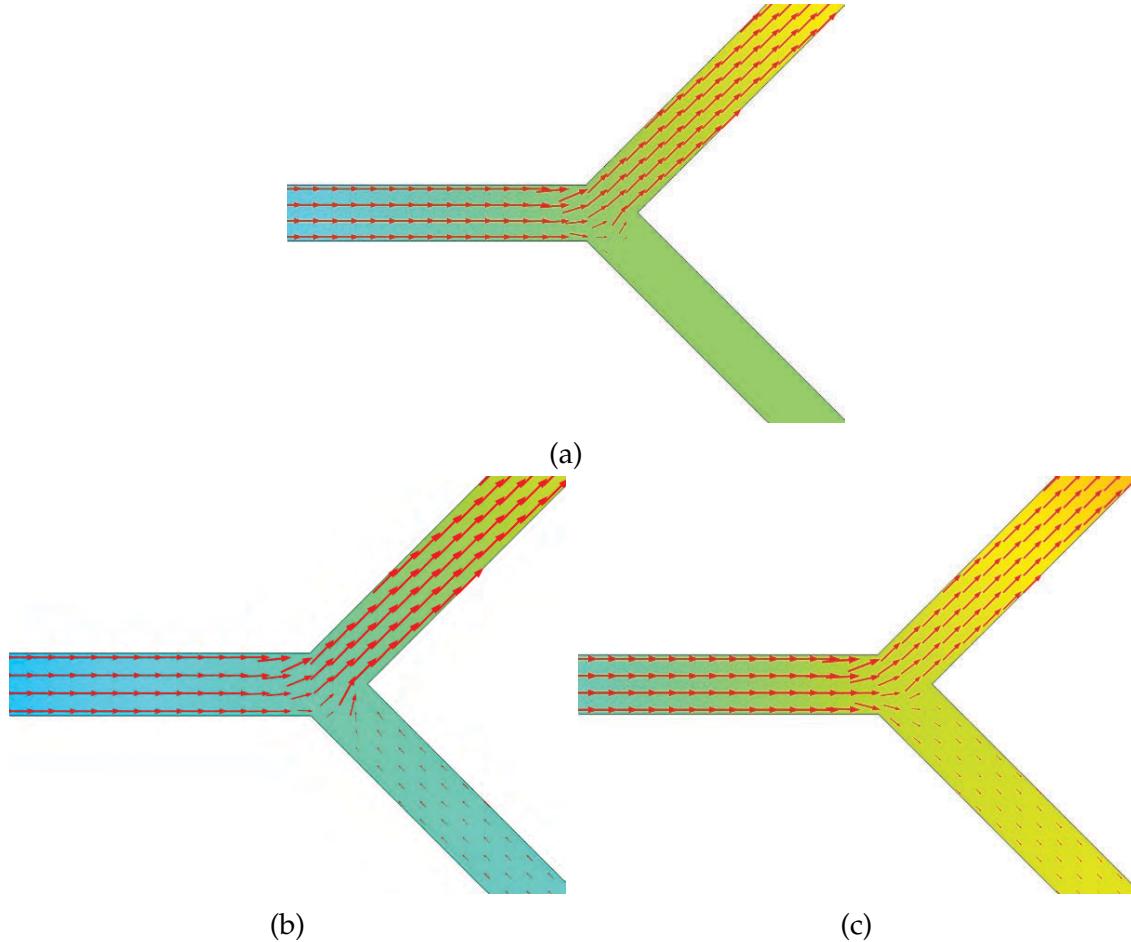


Figure 5.6: COMSOL simulations showing leakage and backflow. Input channel=left, default channel=top, sort channel=bottom. (a) Half voltage condition: $V_{\text{default}} = V_{\text{sort}}/2$. (b) Backflow condition: $V_{\text{default}} < V_{\text{sort}}/2$. (c) Leakage condition: $V_{\text{default}} > V_{\text{sort}}/2$

Using a three port device, we can't remove both backflow and leakage as there is only one boundary condition that is adjustable. To eliminate backflow and leakage, we designed a new five port device outlined in Figure 5.7. By switching the voltages between V_{LowOne} and V_{LowTwo} along with $V_{HighOne}$ and $V_{HighTwo}$, leakage and backflow are both removed. The obvious drawback of this design is the additional electrical components needed to operate the sorter. An initial design schematic of the needed components is given in Figure 5.8. Voltages were to be controlled by linear voltage regulators which were themselves manipulated by digital potentiometers. A N-channel power MOSFET switching circuit comprises the remaining portion of the circuit. All of the ports except the input port would have had independent voltage and switching control. The five port design was never fabricated and the PCB was never manufactured. If the PCB was manufactured today, we would most likely use digitally programmable regulators controlled by the FPGA or other microcontroller instead of using digital potentiometers and linear regulators, but the overall idea is the same.

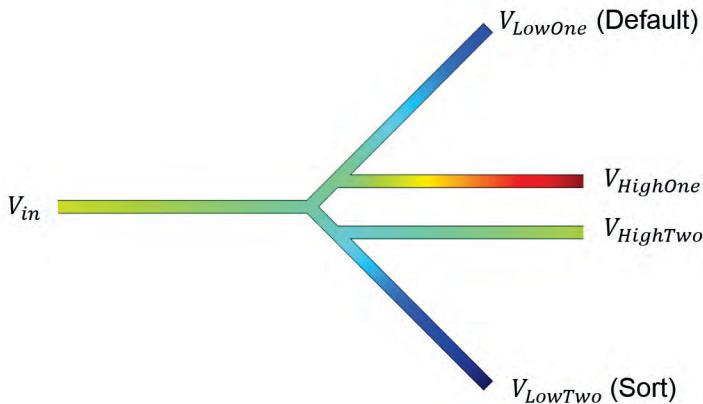


Figure 5.7: Schematic representation of the 5 port device to eliminate backflow and leakage with COMSOL coloring showing voltage levels within the device. (red = high voltage, blue = low voltage)

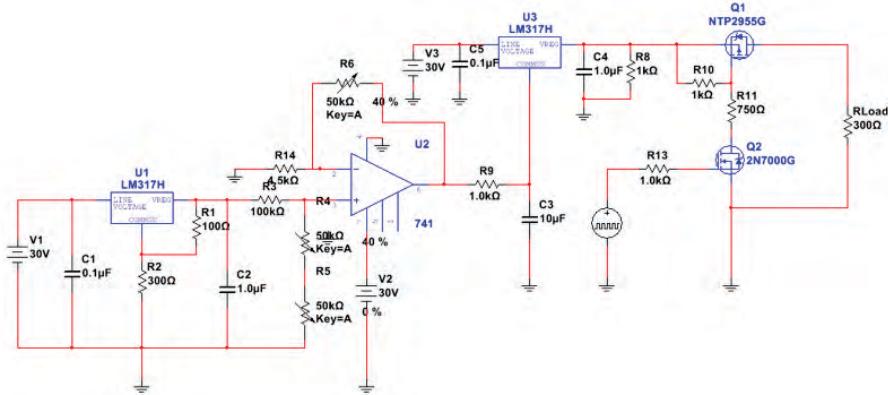


Figure 5.8: Circuit schematic for voltage control of 1 port of the 5 port sorter

5.4 Conclusions

With a shortened sorting device, we sorted DNA based on intensity at a rate of 2000 molecules/minute. The sorting procedure produced a sorted sample that was composed of over 98% of the wanted DNA molecules. Chromatin was also sorted based on coincidence of GFP and TOTO-3 at approximately 5000 molecules/minute. We calculated that chromatin was enriched by over 2-fold by sorting, which provides hope that SCAN-sort could be used to sort epigenetic marks in the future.

However, SCAN-sort has many challenges ahead if it wants to be a viable technique for mapping epigenetic marks. First, SCAN-sort will need to be parallelized as single SCAN-sort devices clog too frequently and provide throughput that is much too low. For a single SCAN-sort device, we should expect molecule counts per minute around 3000. Assuming that we are running mononucleosomes with SCAN, it would take approximately 5 days to interrogate a human genome worth of material. If we ran 100 devices in parallel, it would take approximately 1 hour. However, parallelization brings a whole host of challenges regarding device integration, EMCCD signal detection, sorting actuation, and

real-time signal processing, feedback, and calibration for each device. Real-time monitoring of each individual sorter via hardware and software with appropriate feedback loops for things such as voltage set points would need to be implemented. Second, the problems associated with backflow, leakage molecules, and stability will need to be addressed by using the 5 port devices described in Chapter 5 or some other means. Challenges certainly will arise when testing the 5 port device. Third, typical coincidence rates for histone modifications have generally been below 1%. These low coincidence rates further suggest that high throughput parallelization is necessary and possible investigations into binding chemistries are needed. Fourth, the removal of free antibodies from solution remains an unresolved problem which would needed to be addressed.

BIBLIOGRAPHY

¹B. R. Cipriany, P. J. Murphy, J. A. Hagarman, A. Cerf, D. Latulippe, S. L. Levy, J. J. Benítez, C. P. Tan, J. Topolancik, and P. D. Soloway, "Real-time analysis and selection of methylated dna by fluorescence-activated single molecule sorting in a nanofluidic channel", *Proceedings of the National Academy of Sciences* **109**, 8477–8482 (2012).

²D. P. Clark and N. J. Pazdernik, *Biotechnology: applying the genetic revolution* (Newnes, 2015).

CHAPTER 6

CONCLUSIONS AND FUTURE PROSPECTIVES

This dissertation described our results pertaining to the development of micro- and nanoscale devices and methods for epigenetic analysis for selected cells.

In Chapter 2, we designed a novel PDMS microfluidic device based on micropillar entrapment for the extraction, isolation, and recovery of human chromosomal DNA and chromatin from small cell populations. Using the device, we were able to recover chromatin and DNA from single cells for downstream fluorospectrometric measurements, qPCR, and SCAN measurements. The recovery was near 100% for DNA and sufficiently high for chromatin. The device also allowed for in-channel immunostaining and DNA labeling followed by buffer exchange.

Although operated as a stand-alone device, we envision this device to be the first stage in a multi-stage device for epigenetic profiling. However, further modifications to the device in the form of microfluidic valving would most likely be needed. Valving would be necessary for stage isolation, buffer exchanges, contamination reduction, and for overall better control of flow and pressures throughout the multistage device. We also foresee a massive parallelization of the device to improve single cell throughput. In the past, single cell parallelization was limited by the loading mechanism or lack thereof. Single cell capture within the micropillars relied on cells being sufficiently dilute. After one cell became trapped, the extra cells would be removed from the input reservoir. Recently, Craighead group members have implemented single cell capture cups within the device to actively capture single cells. After capture, the cells are lysed and the DNA/chromatin becomes entangled within micropillar

system as before. With the capture cup technology, chromatin from hundreds of isolated single cells could be processed in parallel, simultaneously.

In Chapter 3, we demonstrated the ability to detect coincident epigenetic marks on a single molecule level in a rapid fashion. Since combinations of epigenetic marks act in concert, researchers are interested detection of multiple epigenetic features. Due to the spectral overlap of dyes and the low SNR of the measurement, we implemented a time-division multiplexing scheme similar to pulsed interleaved detection for 3-color detection of epigenetic marks.

Unfortunately, the ability of SCAN to accurately quantify epigenetic marks on a genome-wide level was never demonstrated due to significant overlap between signal and background. With that said, the future prospects of using SCAN for high resolution epigenetic quantification are dependent almost exclusively on sample preparation improvements. The complete reliance on the inefficient intercalation of chromatin for signal detection coupled with heterogeneous samples containing multiple lengths of DNA and chromatin produced SCAN data which gave ambiguous results regarding molecule counts and coincidence rates. To increase the signal relative to background, we attempted to add fluorescent nucleotides to the chromatin fragments via terminal transferase with no success. A recent paper by Rotem et al [1], however, demonstrated ligation of DNA barcodes to mono-, di-, and trinucleosomes using T4 DNA ligase (NEB Quick Ligation Kit) and a DNA end repair kit suggesting that we could ligate fluorescent DNA oligonucleotides onto the chromatin fragments for signal enhancement and signal separation from background. Assuming that this ligation could be reproduced in our lab, we would have the capabilities to quantify multiple epigenetic marks at a mononucleosome resolution. Only then

would we be able to compare our data to ChIP and become confident in our methodology.

In Chapter 4, we established mobility-SCAN, a technique to measure single molecule mobilities. As a diagnostic and analysis tool, mobility-SCAN answered many questions regarding things such as chromatin/DNA ratios, chromatin intercalation, and chromatin/antibody binding. Moreover, we measured single molecule electrophoretic mobilities, diffusion coefficients, effective molecular charges and more with the technique. We also demonstrated molecule identification solely based on mobility by identifying 6 distinct biomolecules. We presented single molecule affinity electrophoresis by showing binding induced mobility shifts of various antibodies and other proteins. We believe that the affinity based binding shifts could be utilized for biomolecule detection of some unlabeled antigen. We believe that mobility-SCAN is fundamental to moving forward with any SCAN method, SCAN-sort or quantitative SCAN.

Finally, in Chapter 5, we discussed SCAN-sort which involves single molecule electrokinetic sorting for enrichment in a bifurcated nanochannel. With large fragments of DNA, we were able to minimize the false positives to less than 2% while maintaining a true positive rate of approximately 60%. Additionally, we sorted chromatin based on TOTO-3/GFP coincidence and were able to enrich by over 2-fold.

SCAN-sort has many challenges ahead if it wants to be a viable technique for mapping epigenetic marks. First, SCAN-sort will need to be parallelized as a single SCAN-sort device provides throughput that is much too low. For a single SCAN-sort device, we should expect molecule counts per minute around

3000. Assuming that we are running mononucleosomes with SCAN, it would take approximately 5 days to interrogate a human genomes worth of material. If we ran 100 devices in parallel, it would take approximately 1 hour. However, parallelization brings a whole host of challenges regarding device integration, EMCCD signal detection, sorting actuation, and real-time signal processing, feedback, and calibration for each device. Real-time monitoring of each individual sorter via hardware and software with appropriate feedback loops for things such as voltage set points would need to be implemented. Second, we envision addressing the problems associated with backflow, leakage molecules, and stability by using the 5 port devices described in Chapter 5. Challenges certainly will arise when testing the 5 port device. Third, typical coincidence rates for histone modifications have generally been below 1%. These low coincidence rates further suggest that high throughput parallelization is necessary. Fourth, the removal of free antibodies from solution remains an unresolved problem which would be needed to be addressed.

Micro- and nanoscale device structures will continue to form, at least in part, the foundation to which advancement will made in single cell epigenetics. With the continued growth of the epigenetics field and the recent advances in therapeutic drugs inviting research dollars from well-financed biotechnology and drug companies, it will be interesting to observe the progression of the epigenetics field and its interplay with method and device development for the years to come. All I hope is that some of the methods and results described within this dissertation will be of use for those future developers.

BIBLIOGRAPHY

¹A. Rotem, O. Ram, N. Shoresh, R. A. Sperling, A. Goren, D. A. Weitz, and B. E. Bernstein, “Single-cell chip-seq reveals cell subpopulations defined by chromatin state”, *Nat Biotech* **33**, 1165–1172 (2015).