Nanofluidic Device for Single-Molecule RNA Sequencing

By

Chad M. Vietz

Submitted to the graduate degree program in Department of Chemistry and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Master of Science.

Chair, Steven A. Soper, Ph. D.

Michael A. Johnson, Ph. D.

Meredith D. Hartley, Ph. D.

Date Defended: 22 November, 2022

The thesis Committee for Chad M. Vietz certifies that this is the approved version of the following thesis

Nanofluidic Device for Single-Molecule RNA Sequencing

Chair, Steven A. Soper, Ph. D.

Date Approved: 22 November, 2022

Abstract

Single-molecule sequencing (SMS) technologies, specifically the Oxford Nanopore Technology (ONT), offer a number of advantages over next generation sequencing (NGS) technologies such as the ability to perform direct RNA sequencing negating the need for amplification, long reads, label-free electrical readout, and a simpler workflow. However, the ONT is not without challenges, namely high cost and high mass requirements (~1 µg) even though only a single molecule is sequenced. We are developing a SMS nanofluidic device (X-ToF) that is fabricated in a thermoplastic via nano-injection molding. This device uses an immobilized enzyme, such as XRN1, to processively clip through a single stranded RNA molecule base-by-base and detects and identifies single ribonucleotides using their molecular-dependent time-of-flight (ToF; *i.e.*, electrophoretic mobility). X-ToF has the ability to significantly lower sequencing cost, reduce library preparation time, and significantly improve the call accuracy. The X-ToF chip consists of a network of nanofluidic channels and within this network is placed a solid-phase bioreactor where XRN1 is covalently immobilized for processively cleaving RNA. Following the clipping of RNA, detection of single ribonucleotide monophosphates (rNMPs) is achieved by electrokinetically shuttling cleaved bases through a nanochannel containing two in-plane nanopores. The ToF of each rNMP is determined via the time it takes the rNMPs to pass through the two in-plane nanopores on either end of a nanochannel flight tube. Here, we report work toward the development of this technology focusing on: (1) The fabrication of an immobilized nanoscale enzymatic reactor (INER); and (2) the modification of nanofluidic structures to facilitate the entry of RNA through in-plane nanopores. XRN1 immobilization to PMMA/COC devices was accomplished using EDC/NHS coupling chemistry. Fluorescently-labeled ssRNA molecules were electrokinetically translocated through the *X-ToF* nanochannels with real-time single-molecule tracking using a fluorescence microscope. We observed electrokinetically translocating RNA in

real-time and its association to immobilized XRN1 within the nanoreactor. Digestion of RNA was accomplished by supplying XRN1's cofactor (Mg^{+2}) and observing the fluorescence decay rate of a single RNA molecule. From this data, the clipping rate was found to be 23 ± 3 nt s^{-1}. Furthermore, by using fluorescently-labeled ssRNA and DNA, the behavior of these biomolecules was studied as they were electrokinetically driven through in-plane nanopores. Observations indicated the failure of these molecules to pass through the in-plane nanopores. A new device was fabricated that contained larger in-plane nanopores and nanostructures that were less abrupt in their depth dimensions. Results from these nanofluidic modifications are presented herein.

Acknowledgements

I would like to extend my gratitude to my parents, Steven and Roxann Vietz for all of the support and guidance they have provided me throughout my life. It is because of your continued hard work and commitment to providing a bright future for your children that I was able to pursue the opportunities that ultimately led me here. I will be forever thankful to both of you for instilling in me the importance of working hard, doing my best, and taking a long-term perspective on life. I cannot thank you both enough.

Next, I would like to thank my wife, Elizabeth Vietz. You have been a solid and unwavering pillar of support during my time in graduate school. Through the ups and downs, you supported me, celebrated with me, and never once doubted my abilities. I could not have done this without you by my side cheering me on through it all.

I would like to express my sincere thanks to my research advisor, Professor Steven Soper for his support while I completed my research work under his guidance. During my time in Dr. Soper's research group I was able to work in a wide range of areas such as biology, engineering, and chemistry. I was given the opportunity to attend conferences, give a high number of oral presentations, and independently work on research projects. These opportunities helped me develop critical thinking, problem solving, and data interpretation skills. I would also like to thank my two other research advisor committee members: Dr. Meredith Hartley and Dr. Michael Johnson. Thank you for your time and helpful suggestions as I completed my research. I am deeply appreciative of you both.

I would also like to acknowledge a number of past and present members from the Soper research group, including Dr. Maggie Witek, Dr. Matt Hupert, Dr. Uditha Athapattu, Dr. Anishkumar Manoharan, Dr. Farhad Shiri, Katie Childers, Ian Freed, Dr. Harshani Wijerathne, Maximillian Chibuike, Khurshed Akabirov, Sachindra Gamage, Hanna Nguyen, Dami Oluwadamilola, Indu Anushika, and Emma Goodrich. I had a lot of good times and laughs working with a number of you and wish you all the best. I would also like to extend my gratitude to the collaborators I worked with: Collin McKinney, Matt Verber, Dr. Junseo Choi, Dr. Sunggook Park, and Dr. Adam Hall. I would also like to thank Avery Meadows, the graduate program coordinator for KU’s chemistry department for her help navigating the process for completing my degree.

Table of Contents

Abbreviations and Acronyms

SMS	Single-molecule sequencing
NGS	Next generation sequencing
RNA	Ribonucleic acid
PCR	Polymerase chain reaction
XToF	Exonuclease time-of-flight
XRN1	Exoribonuclease 1
ToF	Time-of-Flight
rNMP	Ribonucleotide monophosphate
INER	Immobilized nanoscale enzymatic reactor
PMMA	Polymethyl methacrylate
COC	Cyclic olefin copolymer
EDC	3-(3-dimethylaminopropyl) carbodiimide
NHS	N-hydroxysuccinimide
ssRNA	Single stranded RNA
DNA	Deoxyribonucleic acid
RNA-seq	RNA sequencing
SBS	Sequencing by synthesis
SMRT	Single-molecule real time sequencing
c-DNA	Complementary DNA
ddNTP	Dideoxynucleotide triphosphate
SNV	Single nucleotide variants

dNTP	Deoxynucleotide triphosphate
CCD	Charged coupled device
ssDNA	single stranded DNA
ATP	Adenosine 5'-triphosphate
PNPase	Polynucleotide phosphorylase
RPS	Resistive pulse sensing
T_g	Glass transition temperature
FIB	Focused ion beam
PET	Polyethylene terephthalate
NIL	Nano-imprint lithography
PUA	Polyurethane acrylate
MeOH	Methanol
Cas9	CRISPR Associated Protein 9
NA	Numerical aperture
EMCCD	Electron multiplying charge coupled device
SEM	Scanning electron microscope
AFM	Atomic force microscope
ETA	Ethanolamine
EOF	Electroosmotic flow

Chapter 1. Overview of technologies for RNA sequencing

1.1 Introduction

1.1.1 RNA-seq introduction

The first generation of RNA sequencing (RNA-seq) was powered by Sanger sequencing, which is based upon synthesizing a complementary DNA (cDNA) template and terminating the synthesis of this template with a labelled dideoxynucleotide (ddNTP). This is followed by gel electrophoresis to sort fragments according to their size and followed by detection. Unfortunately, this did not provide high throughput sequencing, which eventually resulted in RNA-seq using a Next Generation Sequencing (NGS) approach that resulted in taking a huge leap forward in 2005[1] when the first high-throughput sequencing platform was released. Soon after, multiple other methods were developed.[2] These NGS sequencers offered a much higher throughput and lower cost than what was achievable with Sanger sequencing. From these NGS technologies, Illumina has emerged as a clear leader. To this point, 95% of the published RNA-seq data on the Short Read Archive has been produced using Illumina's short-read sequencing technology.[3] While the NGS sequencing technology produced significantly higher throughput and lower cost, they were not without their disadvantages. These include: (1) lengthy library preparation, which adds significant cost due to labor and involves one or more steps of PCR amplification, which can result in the loss of epigenetic information[4] and amplification errors;[5] (2) high cost to initially purchase a NGS machine and the continued need for kits or reagents, which adds significant cost to running sequencing reactions; and (3) the inability to directly sequence RNA without converting to cDNA, which can cause the loss of important RNA modification information such as epitranscriptome information.[6,7] The disadvantages of NGS eventually spurred research to develop the third generation of sequencing or single-molecule-sequencing (SMS) technologies, the first of which became commercially available in 2010.[8] Since then, ONT has become a leader in this area with their nanopore-based sequencing that produces an electrical readout.[9] ONT and other SMS technologies offer advantages compared to NGS such as: (1) The ability to directly sequence RNA;[10] (2) produce simple electrical readouts which do not require fluorescence labeling;[9] and (3) reduced cost for initial instrument purchase as well as for continual sequencing runs. Beyond these immediate advantages of SMS technologies, there is great potential for these third-generation technologies to further simplify library preparation, greatly increase read lengths, and reduce the amount of input sample needed to perform a successful sequencing reaction thus eliminating the need for PCR and opening the door to new biomarkers that do not produce as much sample as needed to complete a NGS reaction (approximately 30 ng).

Due to the continued technological improvements, the market for technology pertaining to RNA sequencing has exploded in recent years. This growth in market share is attributed to a growing amount of research harnessing RNA sequencing to answer biological questions about both humans and animals. Market growth is also being driven by this technology's application in the clinical setting to empower precision medicine.[11] Currently, NGS technologies primarily dominate the landscape of RNA-seq. However, due to their current and ever improving advantages, SMS technologies will inevitably capture a share of the available market. The increase in the overall sequencing market remains to be driven in part by a large increase in clinical applications.[12,2] These clinical applications are primarily focused on treating infectious diseases and cancer by enabling the detection of biomarkers and therapeutic targets such as gene fusions, deletions, insertions, and single nucleotide variants (SNVs), which can be detected at both the RNA or DNA level aiding clinicians in both diagnosing a patient and determining the best course of therapeutic treatment.[12,13]

In this first chapter, we review considerations and strategies for sequencing RNA along with past and present sequencing technologies, including first generation technologies, NGS, and concluding with third generation technologies. Table 1.1 compares and summarizes these different RNA sequencing platforms. These technologies and their underlying mechanisms will be examined specifically in the light of their applications for RNA sequencing.

Platform	Company	Read length(bases)	Run time	Volume per run	Cost	Template preparation	Sequencing chemistry
The first-generation sequencing							
Sanger	Life sciences	800 bp	2 h	1 read	$2400 per million bases	Bacterial cloning	Dideoxynucleo-sides terminator
The next-generation sequencing							
Roche 454 pyrosequencing	454 Life sciences	700 bp	< 24 h	0.7 Gb	$10 per million bases	Emulsion PCR	Sequencing by synthesis, pyrosequencing
Illumina HiSeq	Illumina	100 bp	3–10 days	120–1500 Gb	$0.02—$0.07 per million bases	Bridge PCR	Reversible termi-nator sequenc-ing
Illumina MiSeq	Illumina	100 bp	1–2 days	0.3–15 Gb	$0.13 per million bases	Bridge PCR	Reversible termi-nator sequenc-ing
SOLiD	Applied biosys-tems instru-ments (ABI)	50–75 bp	7–14 days	30 Gb	$0.13 per million bases	Emulsion PCR	Sequencing by ligation
DNA nanoball sequencing	Complete genom-ics	440–500 bp	9 days	20–60 Gb	$4400/genome	Rolling circle replication	Hybridization and ligation
Ion torrent	454 Life sciences	200–500 bp	4–5 h	660 Mb; 11 Mb	$300 to $750 per run	Emulsion PCR	Sequencing by synthesis
The third-generation sequencing							
SMRT	Pacific biosciences	> 900 bp	1-2 h	0.5–1 Gb	$2 per million bases	No need	Sequencing by synthesis
Helicos sequenc-ing	Helicos bio-sciences	25–60 bp	8 days	21–35 Gb	$0.01 per million bases	No need	Hybridization and synthesis
Nanopore sequencing	Oxford nanopore technologies	Up to 98 kb	48/72 h	Up to 30 Gb	< $1 per million bases	No need	Nanopore

Table 1.1 A comparison of different platforms for RNA sequencing arranged by generation of sequencing. (Reproduced from reference [2]).

1.1.2 Consideration and strategies for sequencing RNA

1.1.2.1 Reverse transcription

Reverse transcription is the process of converting RNA to cDNA through the use of a reverse transcription enzyme (see Figure 1.1) This process is the first step in any RNA sequencing endeavor for first and NGS techniques as they require double or single stranded DNA inputs and are unable to sequence RNA directly.[14] Reverse transcription is completed using ssRNA, several enzymes, primers, and dNTPs. First, the ssRNA molecules and primers are heated to eliminate secondary structures of the RNA and then snap cooled on ice. During the cooling step, the primers anneal to the RNA. Following this, a reaction buffer, the reverse transcriptase enzyme, and dNTPs are added. These reagents are then heated to allow for the reverse transcriptase to operate and synthesize a cDNA strand by adding one dNTP molecule at a time starting at the primer with the reaction template by the RNA molecule. The dNTP added is always complementary to the base in the RNA strand, thus preserving the sequence of the RNA when converting it to cDNA. However, once the first cDNA strand has been created, the original RNA strand must be replaced with DNA. This is accomplished by nicking the RNA with an enzyme, such as RNase H.[15] This enzyme will remove some of the RNA nucleotides causing gaps in the strand which have an available 3’ -OH sights. DNA polymerase I can then be used to add dNTPs starting at this 3’ -OH sight. This enzyme will continue to add complementary dNTPs and remove the remaining RNA in the process due to its 5’ to 3’ exonuclease activity. DNA ligase is then used to join the newly created DNA fragments by catalyzing the formation of phosphodiester bonds. Finally, this results in double stranded cDNA.[16,15,17]

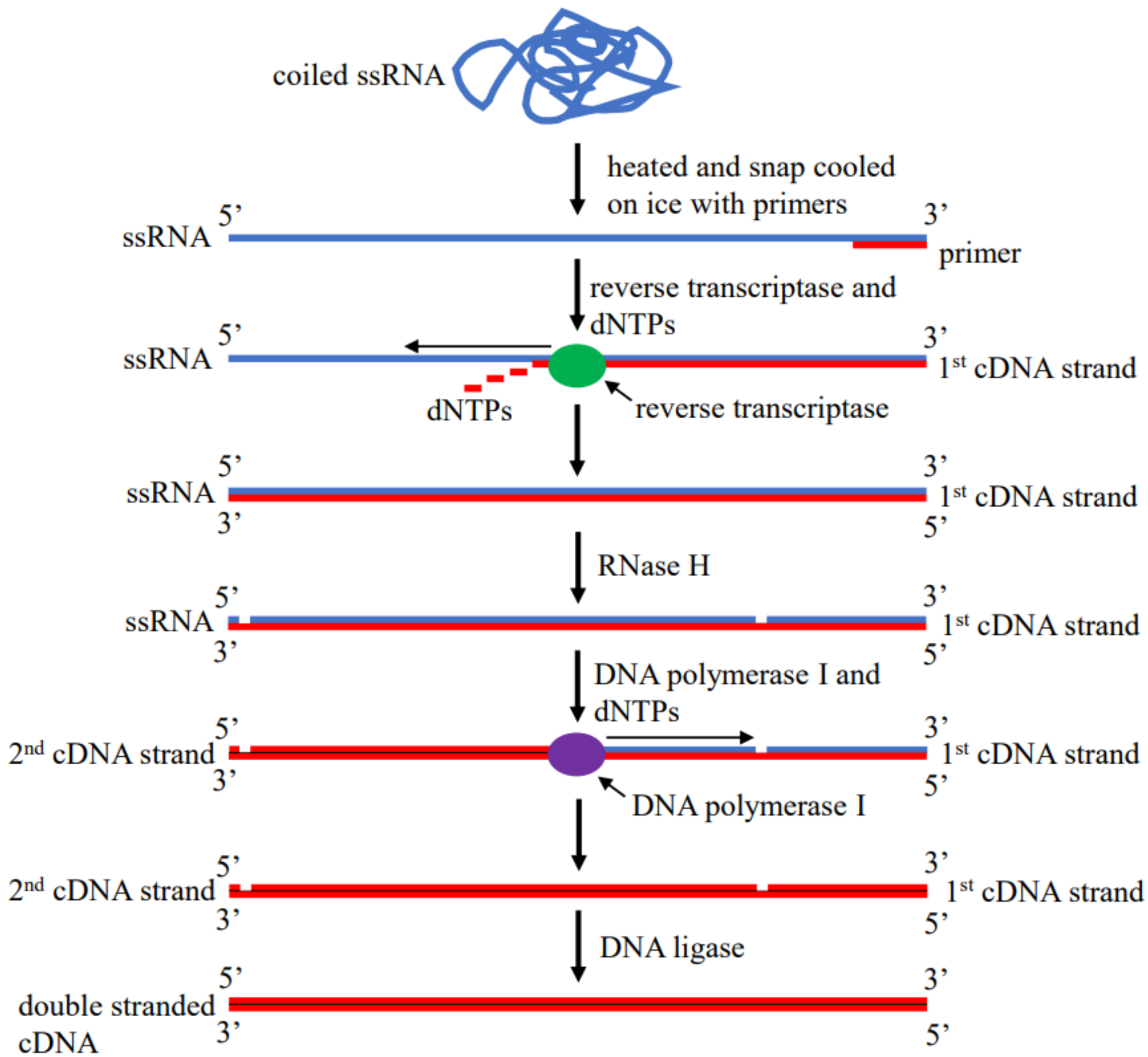

Figure 1.1 The process of reverse transcription to generate double stranded cDNA from ssRNA.

1.1.2.2 Strand specific RNA-seq

In the classic reverse transcription reaction used to produce double stranded cDNA from RNA, information about the orientation of the original RNA strand is lost. This information is vital when interrogating the transcriptome. Further, understanding the transcriptome allows one to record species of the transcript, investigate transcriptional gene structure, determine post-transcriptional modifications, and measure the expression level of transcripts.[18] Researchers have discovered a way to remedy this issue using strand specific or stranded RNA-seq. This is a method for producing cDNA from RNA while keeping information about the original orientation of the RNA in place.[14] This process is accomplished by distinguishing between the first and second strands of cDNA. The methods used can sorted into three categories: (1) Directly sequence the first cDNA strands;[19] (2) ligate adaptors in predetermined locations to either the RNA or to the first strand of cDNA;[20] and (3) chemically alter the second strand of cDNA with uracil so that it can be degraded and eliminated altogether in the follow library preparation steps.[21] The inclusion of a few extra library preparation steps allows for the orientation of the original RNA strand to be preserved.[22]

1.1.2.3 RNA-seq targeted approaches

Multiple routes of targeted RNA-seq may be taken to lower the overall cost required to gain insights from RNA-seq results and increase the number of samples that may be analyzed at once. This is especially applicable when harnessing RNA-seq to inform clinical diagnosis and treatment decisions.[14] Common targeted approaches include using (1) sequence-specific DNA with complementary targeting sites to cDNA[23], (2) double stranded cDNA reverse transcribed from poly(A) RNA to enrich designated cDNA[24], and (3) selection of RNA, DNA, or cDNA via surfaces containing target-specific oligonucleotides[14]. Using available targeted RNA-seq methods, fusion transcripts, allele-specific expression, mutations, and RNA-editing events in transcripts may all be selected for.[24,23,25]

1.2 First generation sequencing

Sanger sequencing was developed in 1977 and was based on a method that caused chain termination.[26] Sanger sequencing offered a straight forward method to sequencing DNA or cDNA and is still highly regarded as the gold standard in many applications. This technology was invented as a DNA sequencing method but can be used to sequence RNA by first converting RNA to cDNA through reverse transcription.[14] After double stranded cDNA is generated from RNA, Sanger sequencing may be completed. This method starts with four different reaction mixtures made of a ssDNA template, primers, DNA Polymerase, dNTPs (dATP, dTTP, dGTP, and dCTP), and a comparatively low concentration of one of the four ddNTPs (ddATP, ddTTP, ddGTP, or ddCTP), which are radioactively- or fluorescently-labeled. In the Sanger reactions, the primers first bind to the ssDNA, followed by DNA polymerase attachment, which begins adding complementary dNTPs to the ssDNA target. However, when a ddNTP is added to the growing strand, the polymerase activity would be terminated. This is due to the fact that for the DNA polymerase to add another dNTP, there needs to be an available 3' -OH group present on the last nucleotide that was added. If this 3' -OH group is not present, as is the case with ddNTP molecules, the strand lengthening reaction is terminated.[27] Because there is a much higher concentration of dNTPs available than ddNTPs, the ddNTPs have a low statistical probability of being incorporated into the DNA fragment during a complementary base addition. This results in different length fragments of DNA due to termination of the chain at different points for different DNA polymerase reactions. Further, the four reactions will each contain fragments that terminate with the ddNTP that was included in each of 4 reactions (ddATP, ddTTP, ddGTP, or ddCTP). In classical Sanger sequencing, these varying length fragments are then ordered in four separate columns (one for each reaction mixture) by their size using polyacrylamide gel electrophoresis with imaging being completed by detecting the radioactive label on the ddNTPs using autoradiography. The sequence of the DNA can then be determined by the size order of the fragments and what column they appear in.[26]

Since the development of classical Sanger sequencing, modern Sanger sequencing has been developed by updating the types of labels used on the ddNTPs, the fragment sorting method, and the detection method of the labeled ddNTPs. Fluorescently-labeled ddNTPs are used now, which allows the user to utilize a single reaction instead of four separate reactions.[28] Capillary electrophoresis is used as the size sorting technique with the detection of the ddNTPs typically achieved using fluorescence detection.[29,28]

Sanger sequencing offered, for the first time, a method to sequence DNA. However, Sanger sequencing was high cost with a low throughput and had large mass requirements. These drawbacks eventually led to the evolution of next generation of sequencing platforms, which are known as NGS.[30]

1.3 Next generation sequencing (NGS)

1.3.1 Introduction

The first NGS sequencing method, pyrosequencing, emerged in 2005.[1] Since this time, NGS has provided significant leaps in RNA-seq technology. NGS reduced the mass requirements from first generation technologies down to approximately 30 ng, increased the throughput capabilities, and greatly reduced the overall cost of sequencing. To this point, NGS platforms now allow for the data-generation phase of sequencing the human genome to be accomplished in under 2 weeks with a cost of ~$5,000.[31] NGS technology has been making its way into the clinic due to its ability to inform both diagnostic and treatment decisions by providing an in-depth view of the transcriptome, detecting RNA transcript variation, SNVs, insertions, deletions, splice variants, and chimeric gene fusions.[32,12] NGS also provides a wide array of available commercial sequencing platforms, which utilize different mechanisms (see Figure 1.2) of sequencing such as Illumina's reversible dye terminator platform, 454 Life Science's sequencing by synthesis platform, and the SOLiD sequencing by ligation method.[2] It is important to note that to achieve RNA-seq on these platforms, like Sanger sequencing, RNA must first be converted to cDNA via reverse transcription.

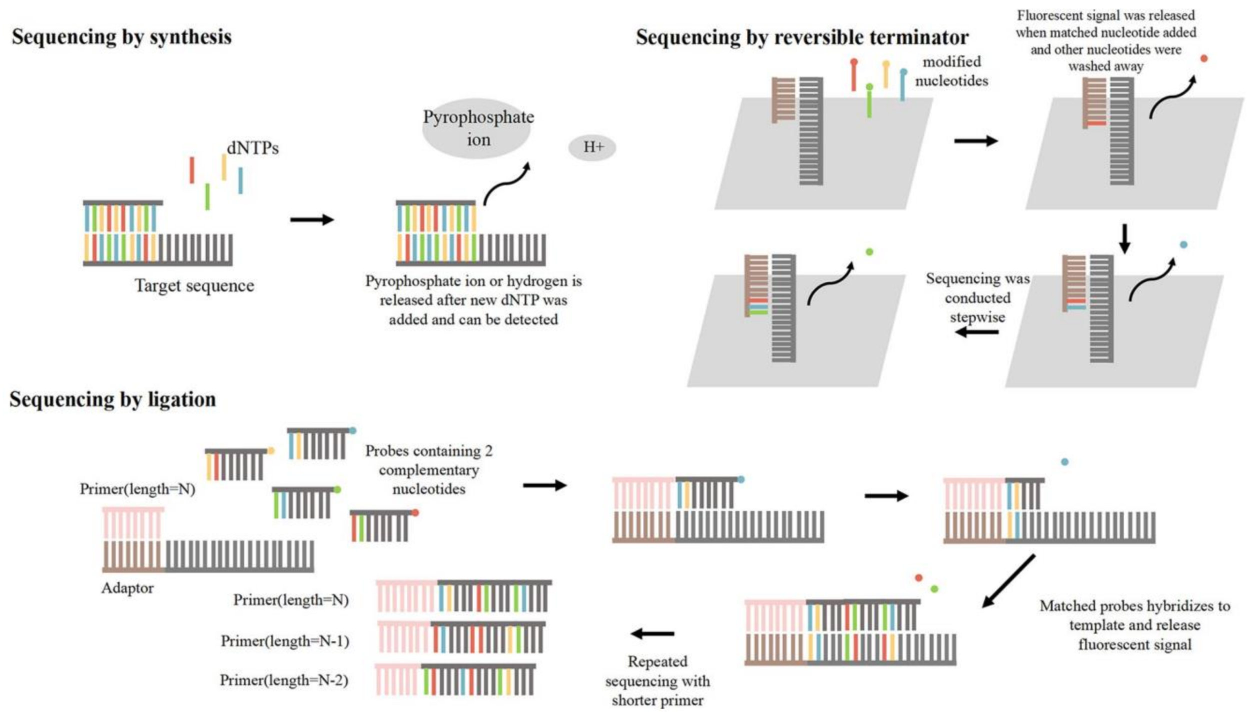

Figure 1.2 Three well established types of NGS methods, sequencing by synthesis, sequencing by ligation, and sequencing by reversible terminator. (Reproduced from reference [2]).

1.3.2 Methods of NGS

1.3.2.1 Introduction

NGS sequencing platforms make up the vast majority of sequencing machines used today. Their improvement on Sanger sequencing to provide paralleled sequencing on a massive scale precipitously dropped the cost and time required for sequencing in spite of the fact that the read lengths can be up to 10-fold less in length (see Table 1.1). As NGS platforms entered the sequencing landscape, this prompted the development of different methods to accomplish sequencing. In this section we will cover the hallmark methods that left their footprint on the sequencing landscape, several of which continue to be used to this day.

1.3.2.2 Pyrosequencing

Pyrosequencing is accomplished by detecting the addition of dNTPs through the production of light. This sequencing method is initiated by fragmenting and attaching primers via ligation to the DNA of interest. These individual fragments are then attached to beads and amplified via emulsion PCR. Equimolar concentrations of beads housing sequences that are complementary to the library adapter sequence, DNA fragments, DNA polymerase, and PCR reagents are mixed.[33] Oil micelles are formed in an emulsion, which house a 1:1 ratio of beads and DNA fragments.[31] The thermocycling step of PCR is carried out to produce copies of the DNA fragments. These beads are then isolated and held in a well located in a flow cell to keep them stationery and ease delivery of reagents. Primers are added to the ssDNA fragments on the beads followed by supplying one dNTP at a time to the flow cell. If the dNTP added is complementary to the sequence, DNA polymerase will add it to the strand. During this base addition process, pyrophosphate is released. The pyrophosphate, with adenylyl sulfate is converted to ATP via ATP sulfurylase. Luciferase then uses ATP with luciferin and oxygen as a substrate to produce light. This light is proportional to the number of dNTPs incorporated. The light is then detected with a CCD. By recording what wells in the flow cell produce light when each dNTP is provided, the sequence of each fragment can be determined.[1,34] Common errors arise from nucleotide repeats of more than five due to nonlinear production of light.[35]

1.3.2.3 Reversible dye terminator sequencing

The reversible dye terminator sequencing method of NGS (see Figure 1.3) begins with fragmentation of the DNA, enzymatic trimming, adenylation of the fragment ends, and ligation of adapters to first prepare a DNA library to be sequenced. The concentration of the library is measured and diluted to ensure adequate spacing between the clusters, which are generated in a following step. The fragments are hybridized via complementary adapter sequences onto the surface of the flow cell, which is made of eight microfluidic channels housed in glass. Solid-phase PCR or bridge amplification is used to create clusters for the following sequencing reactions. A chemical treatment step to release fragment ends housing the same adapter is completed. This resulting end is then equipped with a primer that houses a free 3’ -OH group. This primer is then used as the building block for extension in following sequencing reactions. During each reversible dye terminating reaction step, all four nucleotides are supplied in the presence of DNA polymerase, each with a different fluorescent label.[36] Only single nucleotide additions are possible due to a blocking group being housed at the 3’ -OH location of the ribose. Each step is made up of the following separate events: (i) DNA polymerase adds a nucleotide; (ii) unused nucleotides are removed; (iii) imaging of the flow cell is completed to record fluorescent signals; and (iv) the dye and the blocking group are chemically removed so that the reaction steps can then be repeated up to 150 times. After these reactions, second read preparations occur so that reads can be made from the opposite side of each cluster. Synthesized strands are removed via denaturization followed by bridge amplification to regenerate the clusters. The fragment clusters are primed with a reverse primer and sequencing steps are repeated.

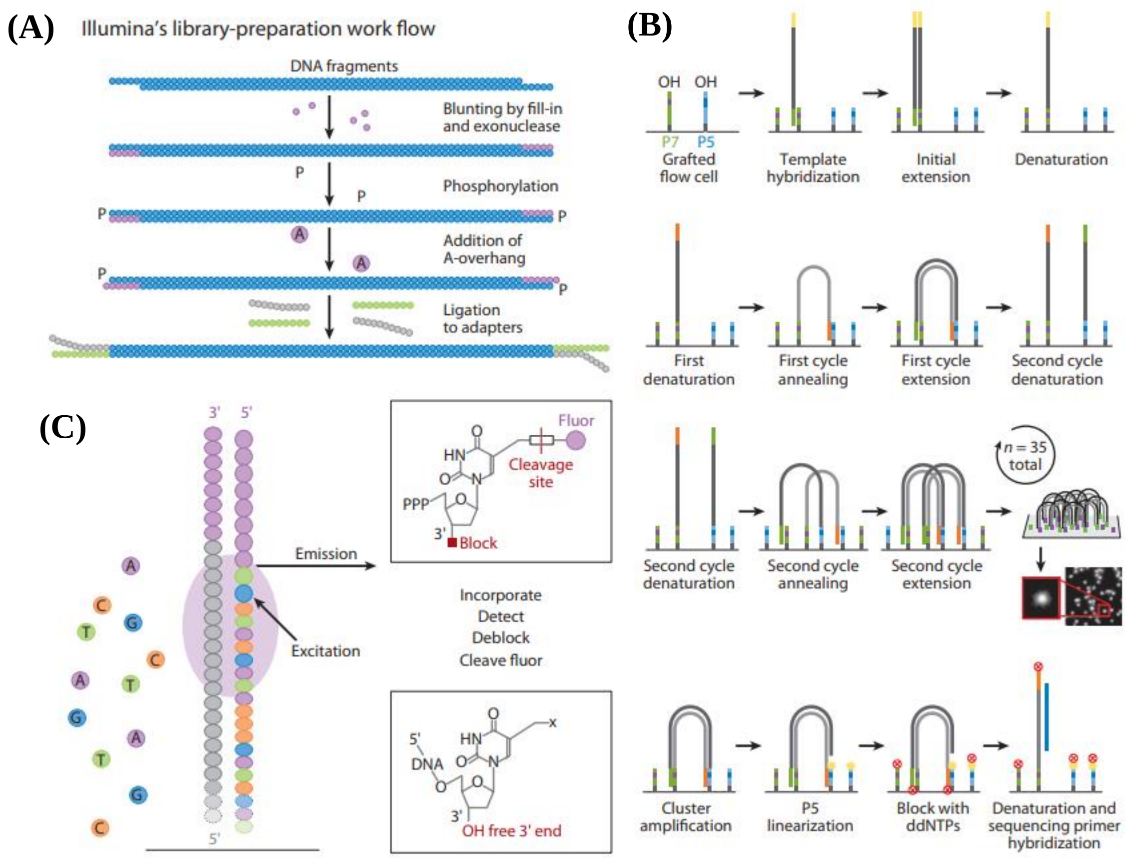

Figure 1.3 Reversible dye terminator method of NGS. (A) Formation of library from DNA sample. (B) Process of bridge amplification to generate clusters for sequencing. (C) Reversible dye terminator sequencing procedure. (Used with permission of Annual Reviews, Inc., from Next-Generation Sequencing Platforms, Elaine R. Mardis, 6, copyright 2013; permission conveyed through Copyright Clearance Center, Inc.) (Reference [31]).

Errors observed from this method occur primarily from substitution errors, incomplete deblocking from previous cycles or lack of blocking group on the incorporated nucleotide, and incomplete fluorescence cleavage. This method enjoys a high level of throughput due to performing up to 150 bp reads and using increasingly more concentrated library solutions to boost cluster density on the flow cell.[31,37]

1.3.2.4 Sequencing-by-synthesis (SBS)

SBS methods accomplish sequencing by taking advantage of the process that occurs when nucleotides are added one by one to ssDNA of interest.[37] One technology that exploits this process is Ion Torrent. Ion Torrent accomplishes sequencing by monitoring hydrogen ion release as a nucleotide is incorporated into a ssDNA. (see Figure 1.4)

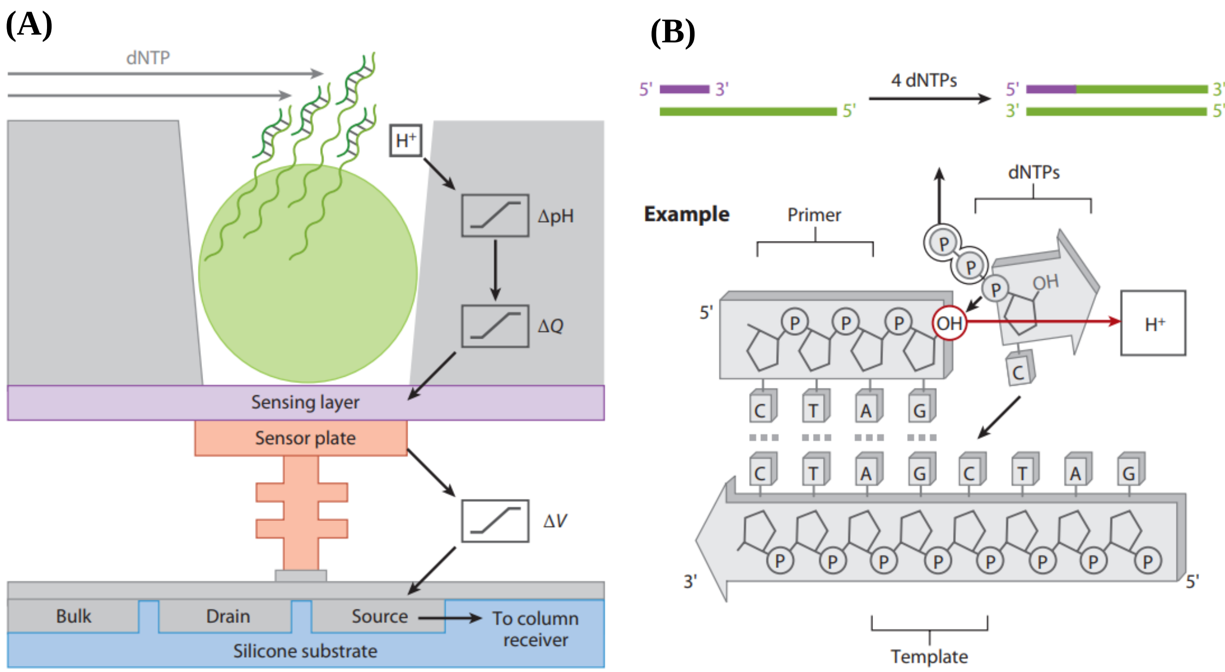

Figure 1.4 Ion Torrent chip and sequencing method **(A)** A single well from the Ion Torrent Ion Chip. Beads with the DNA of interest are housed in individual wells and the change in pH is monitored as dNTPs are incorporated. **(B)** Method of sensing the pH change as dNTPs are added into the DNA strand of interest located on the surface of the beads. (Used with permission of Annual Reviews, Inc., from Next-Generation Sequencing Platforms, Elaine R. Mardis, 6, copyright 2013; permission conveyed through Copyright Clearance Center, Inc.) (Reference [31]).

This is done in a massively parallel fashion so that a large amount of DNA can be sequenced quickly. The process starts with fragmentation of the DNA library to generate short fragments, followed by adaptor ligation. These individual fragments are then attached to beads and amplified via emulsion PCR. The beads are then isolated and dispersed across an ion chip. This chip contains millions of wells, and each bead inserts itself into an unoccupied well. After the beads have been dispersed, each well is flooded with one of the four dNTPs. If the dNTP flooded into the well is complementary to the current strand and incorporated by DNA polymerase into the ssDNA fragment on the bead, a hydrogen ion will be released. In order to exploit this process to obtain sequence information, the pH in the well is monitored. Therefore, if a change in the pH is detected when a dNTP is supplied, a base is called. Further, the amount of pH change is monitored to determine if only one base should be called or if more should be. This pH change is converted to voltage and recorded by the sequencing platform. This technology has the advantage of not requiring fluorescence labelling and instead providing data via an electrical readout.[38]

Errors in sequencing from this method typically occur due to insertions, deletions, and nonlinear electric response when five or more nucleotides are incorporated during one dNTP supply step. The Ion Torrent platform only requires approximately 2 hours to complete a sequencing run and has the potential to enable sequencing of the human genome in a few hours.[37]

1.3.2.5 Sequencing by ligation

The method of sequencing by ligation entered the NGS industry in 2005 with a proof-of-concept demonstration[39] and was successfully commercialized in 2007 with the release of the SOLiD NGS platform.[40] It offered massive parallel sequencing by utilizing DNA ligase to repair breaks in DNA fragments attached to the surface of beads.[40] To begin the process of sequencing by ligation, the sample library is fragmented, and primer ligation is completed. These fragments are then attached to complementary primers on beads and amplified by emulsion PCR. The beads are isolated and attached to the surface of a glass slide. Following this, the beads are exposed to primers followed by a solution of probes, each of which have a different fluorescent tag. The primers serve to attach to the ssDNA on the beads and provide a 3’ -OH group for the probes to build off of. The probes are made up of three different sections. The first two nucleotides in the probe are complementary to the sequence, the next three bases are degenerate nucleotides, and the final three nucleotides are universal nucleotides meaning they can interact with any other nucleotide. A fluorophore is attached at the end of this last group. Once a complementary probe is added onto the existing ssDNA, DNA ligase completes the connection, non-ligated probes are washed away, and the fluorescent signal is recorded.[41] The three bases on the end of the freshly added probe are removed through cleavage of the bond between the fifth and sixth base. A new complementary fluorescent probe will be attached to the existing strand and this process will repeat. Following this, the newly fabricated fragment is removed, and the entire process is repeated a number of times, each time using primers that are one nucleotide shorter with each processive round. (see Figure 1.5) This allows for interrogation the entire sequence to be determined with full coverage.

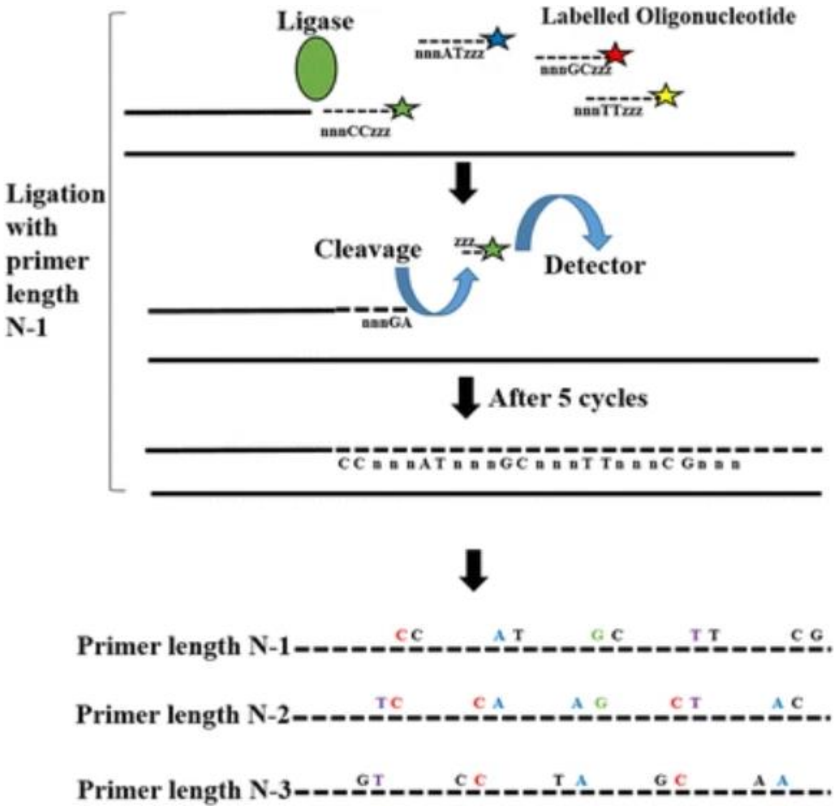

Figure 1.5 Sequencing by ligation method. A primer is added to a ssDNA molecule and fluorescently labeled probes are added by DNA ligase. The probes are then detected, and cleavage occurs paving the way for the next probe to be added. This process is repeated with primers of shorter lengths. (Reprinted by permission from [Springer Nature Customer Service Centre GmbH] : [Springer Nature] [Indian Journal of Microbiology] [42] (High Throughput Sequencing: An Overview of Sequencing Chemistry, Sheetal Ambardar et al), [COPYRIGHT] (2016)).

The method of sequencing by ligation boasts an extremely low error rate due to repeated coverage of the sequence with each round of sequencing using a shorter and shorter primer. Errors that do occur are typically due to substitutions.[40]

1.3.3 Challenges of NGS

NGS technologies suffer from issues that reside with sequencing fragments amplified via polymerase reactions. These include errors such as incorrectly copied bases, biased amplification of certain sequence areas, and loss of DNA methylation information.[31] NGS relies on short reads, which can be troublesome when encountering sequences with extensive repeats. When looking specifically at RNA, NGS lacks the ability to directly sequence RNA. Like Sanger sequencing, RNA must first undergo reverse transcription to create cDNA for sequencing. This can be particularly troublesome considering the extensive number of epitranscriptomic modifications (>100) that can occur in RNA and the role they play in biology and diseases;[43] this information is typically lost during the reverse transcription and polymerase reactions. These downsides call for a method of sequencing that is able to directly sequence RNA and integrate individual molecules to obtain sequence information.

1.4 Third generation sequencing

1.4.1 Introduction

Third generation sequencing or SMS provides much longer reads[44] compared to the short reads provided by NGS. SMS addresses the limitations associated with NGS such as simpler library preparation and the ability to probe single molecules as well as in some cases forgoing the need for reverse transcription and polymerase reactions due to the fact that SMS can read an RNA sequence directly. There have been two major advantages of using SMS specifically for RNA-seq, however. First, as previously mentioned SMS allows for long reads, which allows for improved characterization of transcripts of interest.[9] Second, this generation offered methods for the first time that are able to directly sequence RNA without first transcribing it to cDNA.[6,7]

1.4.2 Sequencing-by-synthesis (SBS)

1.4.2.1 Direct RNA sequencing

In 2009, a method for directly sequencing RNA was developed by Helicos Biosciences that did not require RNA to be converted to cDNA.[6] This method is based upon sequencing-by-synthesis, harnessing the detection of fluorescent reporters and attachment of RNA via a poly(A) tail to a glass surface containing millions of poly(T) tails. (see Figure 1.6)

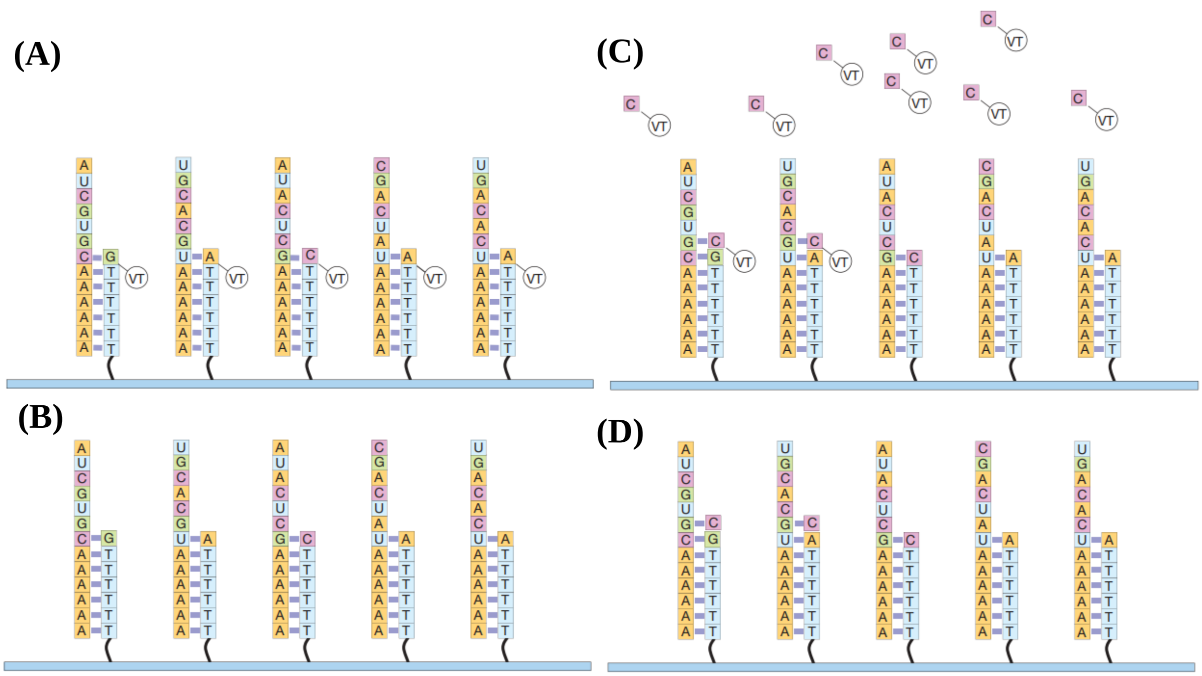

Figure 1.6 Process of direct RNA sequencing. **(A)** RNA is immobilized on surface and followed by a fill and lock step to correct any misalignments that may occur and ensure sequencing does not occur in the poly(A) tail. **(B)** the dye-nucleotide linker is cleaved for the next incorporation of nucleotides. **(C)** One type of VT nucleotide is added followed by an imaging step to record the sequence of each template. **(D)** The dye is cleaved, preparing the RNA template for the next nucleotide addition. (Adapted by permission from [Springer Nature Customer Service Centre GmbH]: [Springer Nature] [Nature] [6] (Direct RNA Sequencing, Fatih Ozsolak et al), [COPYRIGHT] (2009)).

The first step of this process is to add a poly(A) tail to the target ssRNA molecules. This is achieved using poly(A) polymerase I (PAPI) and abruptly adding 3'deoxyATP to the reaction. This facilitates generation of a poly(A) tail of approximately 150 nucleotides with a 3' blocking group at the end. The blocking group is needed to ensure that nucleotides are not added to the poly(A) tail during sequencing. The poly(A) tail on the ssRNA is hybridized to poly(T) tails on the sequencing surface. Following hybridization dTTP and a polymerase enzyme are supplied with A, C, and G virtual terminator (VT) nucleotides to lock the RNA into position. These VT nucleotides contain a blocking group on at the 3' group to prevent additional nucleotide incorporations and a fluorescent reporter, both of which can be chemically cleaved.[45] Images are taken to record the fluorescent signals and then the blocking group on the nucleotide and the fluorescent label are chemically removed. A, T, C, and G VT nucleotides are then supplied in an alternating fashion. In between each new nucleotide supply step, a washing stage is completed, images of the florescent signals are captured, and the blocking group and florescent label are cleaved. This process is repeated a large number of times to achieve sequencing of the RNA. Errors experienced from this method (approximately 4%) originate from insertions, deletions, substitutions, and incomplete rinsing of nucleotides between sequencing steps.[6,46]

1.4.2.2 Single-molecule real-time sequencing (SMRT)

SMRT, a method developed by Pacific Biosciences[47] offers the ability to perform long reads with real time monitoring of the sequencing process by detecting fluorescently labeled nucleotides as they are incorporated into a DNA strand by a polymerase. The library is prepared for sequencing by attaching primers onto the DNA. These primers provide a location for DNA polymerase to bind and begin adding complementary nucleotides to the existing strand. The sample is loaded onto a SMRT chip composed of SiO_2 with a thin aluminum layer on top. This chip contains zero-mode waveguides (ZMW). These ZMW are cylindrical cavities in the chip surface which are approximately 100 nm deep and 50 nm in diameter.[48] A single polymerase molecule is attached to the bottom of a ZMW which is glass.[49] Due to this, one DNA strand containing the hairpin adaptors, and primers will be house in one ZMW. The four types of nucleotides are added to the SMRT chip at once. These can be added at the same time because each of the four nucleotides contain a different fluorescent label.[47] DNA polymerase begins adding complementary nucleotides to the DNA strand one at a time. The detection of which nucleotide is being incorporated into the strand by the polymerase is achieved by taking advantage of the ZMW architecture. The ZMW is illuminated by a laser, but due to the ZMW's size, the laser only penetrates into the very bottom of the cavity. This creates a very small detection volume at the surface of the glass where the polymerase is attached.[48] All four different nucleotides diffuse in and out of this volume very quickly during the reaction, however when the DNA polymerase adds a complementary nucleotide, it is held in the detection volume for a much longer time. This produces a higher fluorescent signal for this nucleotide type than the other three and a base is called. As the DNA polymerase moves on to incorporate the next nucleotide, it cleaves the fluorescent label off, which is present on the phosphate chain, rather than at the base of the nucleotide which allows this sequencing process to be repeated.[47,50,42]

1.4.3 Nanopore sequencing

1.4.3.1 Strand sequencing

Strand sequencing is the method of passing a single strand of RNA or DNA through a nanopore while applying a voltage and monitoring the resulting current across the nanopore. As the strand passes through the pore's small geometry, it causes a disruption to the ionic current, which can be measured as a change in current based on the nucleotide that is currently residing in the pore.[9],[51] Motor enzymes are used to slow down the translocation of the single strand to achieve better sequence identification.[52]

Oxford Nanopore developed a nanopore based strand sequencing method for sequencing RNA and DNA that uses an array of solid state nanopores. Since this technology's debut on the commercial market in 2014[9] Oxford Nanopore has continue to improve and develop its technology. Significant developments include improving the base calling accuracy from initial moderate accuracy in the 80 - 90% range[9] for single reads up to approximately 99% through consensus reads[53],[54] and demonstrating the ability to directly sequence RNA on their platform.[7]

This sequencing is accomplished by translocating a strand of DNA through a biological nanopore while applying a voltage and monitoring the resulting current. To begin, the library of interest is ligated with an adapter sequence that contains a motor enzyme. It is then added to a flow cell which contains 512 channels with four biological nanopores in each channel, yielding a total of 2048 nanopores. Each channel contains a separate electrode which is individually controlled and monitored using an application-specific integration circuit (ASIC). One of the four nanopores in each channel are selected for sequencing by the software.[55,56,9] DNA and the processive motor enzyme move to the membrane surface and interact with the nanopore. The motor enzyme then aids in unzipping the double stranded DNA and translocating one strand through the nanopore while the other strand is not translocated.[9] As this strand passes through biological nanopore's small geometry, which for α-hemolysin is 5 nm long with a diameter ranging from ~1.4 to ~2.4 nm[56], it will displace a number of ions currently occupying the nanopore based on the nucleotide currently residing in the nanopore and the measured current will drop proportional to the ions displaced.[57,58] Therefore, by recording the current trace as the strand is translocated through the pore, the sequence of the strand can be deduced. Once the entire strand has passed through the nanopore, the motor enzyme departs and the pore is available for the process to be repeated.[9]

The accuracy of Oxford Nanopore's technology is steadily being pushed higher with the development of new base calling algorithms, such as the Guppy and Bonito base callers that can achieve DNA sequencing accuracies of over 99%[53,54]. These basecallers use machine learning and are trained on a known data set to learn what signal amplitudes correspond to different nucleotides. These basecallers can also be trained on different data sets depending on the sequencing application. If certain epigenetic modifications are expected to be observed in a sample, the basecaller can first be trained on a data set which contains these modifications.[59]

For the first time in 2018, Oxford Nanopore's technology was used to directly sequence RNA without using reverse transcription or any form of amplification.[7] The ability to directly sequence RNA without first converting it to cDNA removes reverse transcription biases, retains epitranscriptomic information, and retains RNA base modifications vital for investigating the transcriptome.[60,61] Direct RNA sequencing was accomplished on this platform by first annealing and ligating an oligo(dT) containing duplex adaptors. This is followed by another ligation step used to attach of sequencing adaptors which contain a motor protein used for aiding translocation of the RNA through the nanopore array.[7] (see Figure 1.7).

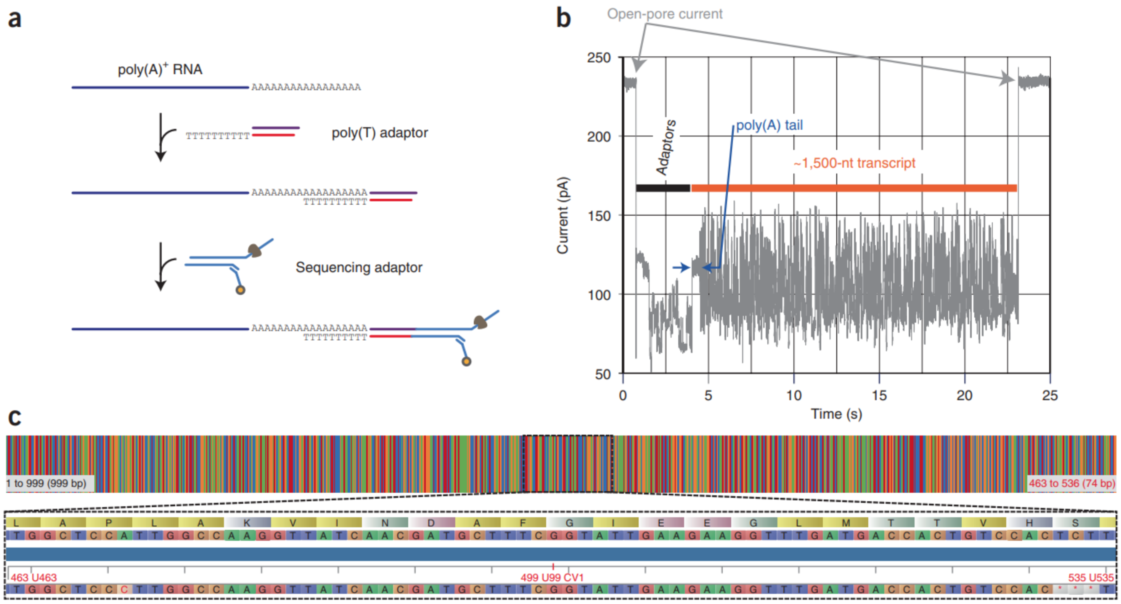

Figure 1.7 Direct RNA sequencing process and example data acquired. **(A)** Library preparation of RNA consisting of addition of an oligo(dT) containing duplex adaptors and sequencing adaptors containing a motor protein. **(B)** Open-pore current observed during the sequencing experiment. **(C)** Comparison of a Saccharomyces cerevisiae S228C read to the transcriptome. (Reprinted by permission from [Springer Nature Customer Service Centre GmbH]: [Springer Nature] [Nature Methods] [7] (Highly parallel direct RNA sequencing on an array of nanopores, Daniel R Garalde et al), [COPYRIGHT] (2018)).

From this point, the RNA is directly sequenced from the 3' to the 5' direction. This approach generated RNA read lengths as long as 10 kb, but with the most common read length being around 1000 bp long.[7,62]

1.4.3.2 Exo-sequencing

In contrast to strand sequencing, exo-sequencing is a method that first removes individual nucleotides from the intact DNA/RNA strand using a processive exonuclease enzyme before passing them one-by-one through a nanopore for detection and identification based on the current drop experienced as they disrupt the ionic current in the nanopore as they pass through.[63] This has been thoroughly demonstrated as a method that can determine the sequence of RNA using an enzyme such as PNPase and biological nanopores.[64,63] Issues with this method occur primarily due to diffusion after enzymatic digestion, which results in low injection efficiencies of the nucleotides through the nanopore and a possible loss of order of the nucleotides before they reach the nanopore.[65]

Exoribonuclease 1 (XRN1) was investigated and found to be a promising enzyme that could be exploited for an exo-sequencing application.[66] XRN1 processively digests RNA in the 5’ to 3’ direction.[67] It was demonstrated that: (i) XRN1 could be covalently immobilized onto a solid support while retaining its ability to digest RNA; (ii) XRN1 could digest through methylated RNA sequences sites and secondary structures; and (iii) XRN1 possessed a clipping rate of approximately 26 nt s^{-1} with a processivity of >10.5 kb at 25°C.[66]

1.5 References

(1) Margulies, M.; Egholm, M.; Altman, W. E.; Attiya, S.; Bader, J. S.; Bemben, L. A.; Berka, J.; Braverman, M. S.; Chen, Y. J.; Chen, Z.; et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **2005**, *437* (7057), 376-380. DOI: 10.1038/nature03959 From NLM Medline.

(2) Hong, M.; Tao, S.; Zhang, L.; Diao, L. T.; Huang, X.; Huang, S.; Xie, S. J.; Xiao, Z. D.; Zhang, H. RNA sequencing: new technologies and applications in cancer research. *J Hematol Oncol* **2020**, *13* (1), 166. DOI: 10.1186/s13045-020-01005-x License at https://creativecommons.org/licenses/by/4.0/legalcode From NLM Medline.

(3) Kodama, Y.; Shumway, M.; Leinonen, R.; International Nucleotide Sequence Database, C. The Sequence Read Archive: explosive growth of sequencing data. *Nucleic Acids Res* **2012**, *40* (Database issue), D54-56. DOI: 10.1093/nar/gkr854 From NLM Medline.

(4) Kristensen, L. S.; Hansen, L. L. PCR-Based Methods for Detecting Single-Locus DNA Methylation Biomarkers in Cancer Diagnostics, Prognostics, and Response to Treatment. *Clinical Chemistry* **2009**, *55* (8), 1471-1483. DOI: 10.1373/clinchem.2008.121962 (acccessed 9/6/2022).

(5) Sharifian, H. Errors induced during PCR amplification. Swiss Federal Institute of Technology, Department of Computer Science, 2010.

(6) Ozsolak, F.; Platt, A. R.; Jones, D. R.; Reifenberger, J. G.; Sass, L. E.; McInerney, P.; Thompson, J. F.; Bowers, J.; Jarosz, M.; Milos, P. M. Direct RNA sequencing. *Nature* **2009**, *461* (7265), 814-818. DOI: 10.1038/nature08390 From NLM Medline.

(7) Garalde, D. R.; Snell, E. A.; Jachimowicz, D.; Sipos, B.; Lloyd, J. H.; Bruce, M.; Pantic, N.; Admassu, T.; James, P.; Warland, A.; et al. Highly parallel direct RNA sequencing on an array of nanopores. *Nat Methods* **2018**, *15* (3), 201-206. DOI: 10.1038/nmeth.4577 From NLM Medline.

(8) Milos, P. M. Helicos single molecule sequencing: unique capabilities and importance for molecular diagnostics. *Genome Biology* **2010**, *11* (1), I14. DOI: 10.1186/gb-2010-11-s1-i14.

(9) Jain, M.; Olsen, H. E.; Paten, B.; Akeson, M. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biology* **2016**, *17* (1), 239. DOI: 10.1186/s13059-016-1103-0.

(10) Leger, A.; Amaral, P. P.; Pandolfini, L.; Capitanchik, C.; Capraro, F.; Miano, V.; Migliori, V.; Toolan-Kerr, P.; Sideri, T.; Enright, A. J.; et al. RNA modifications detection by comparative Nanopore direct RNA sequencing. *Nature Communications* **2021**, *12* (1), 7198. DOI: 10.1038/s41467-021-27393-3.

(11) *NGS-Based RNA-Sequencing Market 2022-Growth Strategy, Top Trends And Business Opportunity*. 2022. https://www.comserveonline.com/news-releases/ngs-based-rna-sequencing-market-2022-growth-strategy-top-trends-and-business-opportunity/10135183 (accessed 2022 October, 16).

(12) Byron, S. A.; Van Keuren-Jensen, K. R.; Engelthaler, D. M.; Carpten, J. D.; Craig, D. W. Translating RNA sequencing into clinical diagnostics: opportunities and challenges. *Nature Reviews Genetics* **2016**, *17* (5), 257-271. DOI: 10.1038/nrg.2016.10.

(13) Wu, H.; Li, X.; Li, H. Gene fusions and chimeric RNAs, and their implications in cancer. *Genes Dis* **2019**, *6* (4), 385-390. DOI: 10.1016/j.gendis.2019.08.002 From NLM.

(14) Ozsolak, F.; Milos, P. M. RNA sequencing: advances, challenges and opportunities. *Nature reviews genetics* **2011**, *12* (2), 87-98.

(15) Schultz, S. J.; Champoux, J. J. RNase H activity: structure, specificity, and function in reverse transcription. *Virus Res* **2008**, *134* (1-2), 86-103. DOI: 10.1016/j.virusres.2007.12.007 From NLM Medline.

(16) Rio, C. D. A., M.; Hannon, J. G.; Nilsen, W. T. *RNA: A Laboratory Manual*; Cold Spring Harbor Laboratory Press, 2011.

(17) Eckert, K. A.; Kunkel, T. A. DNA polymerase fidelity and the polymerase chain reaction. *PCR Methods Appl* **1991**, *1* (1), 17-24. DOI: 10.1101/gr.1.1.17 From NLM.

(18) Wang, Z.; Gerstein, M.; Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics* **2009**, *10* (1), 57-63. DOI: 10.1038/nrg2484.

(19) Ozsolak, F.; Goren, A.; Gymrek, M.; Guttman, M.; Regev, A.; Bernstein, B. E.; Milos, P. M. Digital transcriptome profiling from attomole-level RNA samples. *Genome Res* **2010**, *20* (4), 519-525. DOI: 10.1101/gr.102129.109 From NLM.

(20) Mamanova, L.; Andrews, R. M.; James, K. D.; Sheridan, E. M.; Ellis, P. D.; Langford, C. F.; Ost, T. W. B.; Collins, J. E.; Turner, D. J. FRT-seq: amplification-free, strand-specific transcriptome sequencing. *Nature Methods* **2010**, *7* (2), 130-132. DOI: 10.1038/nmeth.1417.

(21) Borodina, T.; Adjaye, J.; Sultan, M. Chapter five - A Strand-Specific Library Preparation Protocol for RNA Sequencing. In *Methods in Enzymology*, Jameson, D., Verma, M., Westerhoff, H. V. Eds.; Vol. 500; Academic Press, 2011; pp 79-98.

(22) Levin, J. Z.; Yassour, M.; Adiconis, X.; Nusbaum, C.; Thompson, D. A.; Friedman, N.; Gnirke, A.; Regev, A. Comprehensive comparative analysis of strand-specific RNA sequencing methods. *Nature Methods* **2010**, *7* (9), 709-715. DOI: 10.1038/nmeth.1491.

(23) Li, J. B.; Levanon, E. Y.; Yoon, J. K.; Aach, J.; Xie, B.; Leproust, E.; Zhang, K.; Gao, Y.; Church, G. M. Genome-wide identification of human RNA editing sites by parallel DNA capturing and sequencing. *Science* **2009**, *324* (5931), 1210-1213. DOI: 10.1126/science.1170995 From NLM Medline.

(24) Levin, J. Z.; Berger, M. F.; Adiconis, X.; Rogov, P.; Melnikov, A.; Fennell, T.; Nusbaum, C.; Garraway, L. A.; Gnirke, A. Targeted next-generation sequencing of a cancer transcriptome enhances detection of sequence variants and novel fusion transcripts. *Genome Biol* **2009**, *10* (10), R115. DOI: 10.1186/gb-2009-10-10-r115 From NLM Medline.

(25) Zhang, K.; Li, J. B.; Gao, Y.; Egli, D.; Xie, B.; Deng, J.; Li, Z.; Lee, J. H.; Aach, J.; Leproust, E. M.; et al. Digital RNA allelotyping reveals tissue-specific and allele-specific gene expression in human. *Nat Methods* **2009**, *6* (8), 613-618. DOI: 10.1038/nmeth.1357 From NLM Medline.

(26) Sanger, F.; Nicklen, S.; Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* **1977**, *74* (12), 5463-5467. DOI: 10.1073/pnas.74.12.5463 From NLM Medline.

(27) Chidgeavadze, Z. G.; Beabealashvilli, R. S.; Atrazhev, A. M.; Kukhanova, M. K.; Azhayev, A. V.; Krayevsky, A. A. 2',3'-Dideoxy-3' aminonucleoside 5'-triphosphates are the terminators of DNA synthesis catalyzed by DNA polymerases. *Nucleic Acids Res* **1984**, *12* (3), 1671-1686. DOI: 10.1093/nar/12.3.1671 From NLM.

(28) Prober, J. M.; Trainor, G. L.; Dam, R. J.; Hobbs, F. W.; Robertson, C. W.; Zagursky, R. J.; Cocuzza, A. J.; Jensen, M. A.; Baumeister, K. A System for Rapid DNA Sequencing with Fluorescent Chain-Terminating Dideoxynucleotides. *Science* **1987**, *238* (4825), 336-341. DOI: doi:10.1126/science.2443975.

(29) Cohen, A. S.; Najarian, D. R.; Paulus, A.; Guttman, A.; Smith, J. A.; Karger, B. L. Rapid separation and purification of oligonucleotides by high-performance capillary gel electrophoresis. *Proceedings of the National Academy of Sciences* **1988**, *85* (24), 9660-9663. DOI: doi:10.1073/pnas.85.24.9660.

(30) Fakruddin, M.; Chowdhury, A. Pyrosequencing an alternative to traditional Sanger sequencing. *American Journal of Biochemistry and Biotechnology* **2012**, *8* (1), 14-20.

(31) Mardis, E. R. Next-generation sequencing platforms. *Annu Rev Anal Chem (Palo Alto Calif)* **2013**, *6*, 287-303. DOI: 10.1146/annurev-anchem-062012-092628 From NLM Medline.

(32) Mortazavi, A.; Williams, B. A.; McCue, K.; Schaeffer, L.; Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* **2008**, *5* (7), 621-628. DOI: 10.1038/nmeth.1226 From NLM Medline.

(33) Dressman, D.; Yan, H.; Traverso, G.; Kinzler, K. W.; Vogelstein, B. Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. *Proceedings of the National Academy of Sciences* **2003**, *100* (15), 8817-8822. DOI: doi:10.1073/pnas.1133470100.

(34) Rothberg, J. M.; Leamon, J. H. The development and impact of 454 sequencing. *Nature Biotechnology* **2008**, *26* (10), 1117-1124. DOI: 10.1038/nbt1485.

(35) Ronaghi, M.; Uhlén, M.; Nyrén, P. A Sequencing Method Based on Real-Time Pyrophosphate. *Science* **1998**, *281* (5375), 363-365. DOI: doi:10.1126/science.281.5375.363.

(36) Guo, J.; Xu, N.; Li, Z.; Zhang, S.; Wu, J.; Kim, D. H.; Sano Marma, M.; Meng, Q.; Cao, H.; Li, X.; et al. Four-color DNA sequencing with 3'-O-modified nucleotide reversible terminators and chemically cleavable fluorescent dideoxynucleotides. *Proc Natl Acad Sci U S A* **2008**, *105* (27), 9145-9150. DOI: 10.1073/pnas.0804023105 From NLM Medline.

(37) Stranneheim, H.; Lundeberg, J. Stepping stones in DNA sequencing. *Biotechnol J* **2012**, *7* (9), 1063-1073. DOI: 10.1002/biot.201200153 From NLM Medline.

(38) Rothberg, J. M.; Hinz, W.; Rearick, T. M.; Schultz, J.; Mileski, W.; Davey, M.; Leamon, J. H.; Johnson, K.; Milgrew, M. J.; Edwards, M.; et al. An integrated semiconductor device enabling non-optical genome sequencing. *Nature* **2011**, *475* (7356), 348-352. DOI: 10.1038/nature10242 From NLM Medline.

(39) Shendure, J.; Porreca, G. J.; Reppas, N. B.; Lin, X.; McCutcheon, J. P.; Rosenbaum, A. M.; Wang, M. D.; Zhang, K.; Mitra, R. D.; Church, G. M. Accurate Multiplex Polony Sequencing of an Evolved Bacterial Genome. *Science* **2005**, *309* (5741), 1728-1732. DOI: doi:10.1126/science.1117389.

(40) Metzker, M. L. Sequencing technologies — the next generation. *Nature Reviews Genetics* **2010**, *11* (1), 31-46. DOI: 10.1038/nrg2626.

(41) Landegren, U.; Kaiser, R.; Sanders, J.; Hood, L. A ligase-mediated gene detection technique. *Science* **1988**, *241* (4869), 1077-1080. DOI: 10.1126/science.3413476 From NLM.

(42) Ambardar, S.; Gupta, R.; Trakroo, D.; Lal, R.; Vakhlu, J. High Throughput Sequencing: An Overview of Sequencing Chemistry. *Indian Journal of Microbiology* **2016**, *56* (4), 394-404. DOI: 10.1007/s12088-016-0606-4.

(43) Barbieri, I.; Kouzarides, T. Role of RNA modifications in cancer. *Nature Reviews Cancer* **2020**, *20* (6), 303-322. DOI: 10.1038/s41568-020-0253-2.

(44) Bleidorn, C. Third generation sequencing: technology and its potential impact on evolutionary biodiversity research. *Systematics and Biodiversity* **2016**, *14* (1), 1-8. DOI: 10.1080/14772000.2015.1099575.

(45) Bowers, J.; Mitchell, J.; Beer, E.; Buzby, P. R.; Causey, M.; Efcavitch, J. W.; Jarosz, M.; Krzymanska-Olejnik, E.; Kung, L.; Lipson, D. Virtual terminator nucleotides for next-generation DNA sequencing. *Nature methods* **2009**, *6* (8), 593-595.

(46) Ozsolak, F.; Milos, P. M. Single-molecule direct RNA sequencing without cDNA synthesis. *WIREs RNA* **2011**, *2* (4), 565-570. DOI: https://doi.org/10.1002/wrna.84.

(47) Eid, J.; Fehr, A.; Gray, J.; Luong, K.; Lyle, J.; Otto, G.; Peluso, P.; Rank, D.; Baybayan, P.; Bettman, B. Real-time DNA sequencing from single polymerase molecules. *Science* **2009**, *323* (5910), 133-138.

(48) Levene, M. J.; Korlach, J.; Turner, S. W.; Foquet, M.; Craighead, H. G.; Webb, W. W. Zero-Mode Waveguides for Single-Molecule Analysis at High Concentrations. *Science* **2003**, *299* (5607), 682-686. DOI: doi:10.1126/science.1079700.

(49) Korlach, J.; Marks, P. J.; Cicero, R. L.; Gray, J. J.; Murphy, D. L.; Roitman, D. B.; Pham, T. T.; Otto, G. A.; Foquet, M.; Turner, S. W. Selective aluminum passivation for targeted

immobilization of single DNA polymerase molecules in zero-mode waveguide nanostructures. *Proceedings of the National Academy of Sciences* **2008**, *105* (4), 1176-1181.

(50) Xu, M.; Fujita, D.; Hanagata, N. Perspectives and challenges of emerging single-molecule DNA sequencing technologies. *Small* **2009**, *5* (23), 2638-2649.

(51) Cockroft, S. L.; Chu, J.; Amorin, M.; Ghadiri, M. R. A single-molecule nanopore device detects DNA polymerase activity with single-nucleotide resolution. *Journal of the American Chemical Society* **2008**, *130* (3), 818-820.

(52) Spies, M. *DNA helicases and DNA motor proteins*; Springer Science & Business Media, 2012.

(53) Wick, R. R.; Judd, L. M.; Holt, K. E. Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome biology* **2019**, *20* (1), 1-10.

(54) Xie, S.; Leung, A. W.-S.; Zheng, Z.; Zhang, D.; Xiao, C.; Luo, R.; Luo, M.; Zhang, S. Applications and potentials of nanopore sequencing in the (epi)genome and (epi)transcriptome era. *The Innovation* **2021**, *2* (4), 100153. DOI: https://doi.org/10.1016/j.xinn.2021.100153.

(55) Wang, Y.; Zhao, Y.; Bollas, A.; Wang, Y.; Au, K. F. Nanopore sequencing technology, bioinformatics and applications. *Nature Biotechnology* **2021**, *39* (11), 1348-1365. DOI: 10.1038/s41587-021-01108-x.

(56) Deamer, D.; Akeson, M.; Branton, D. Three decades of nanopore sequencing. *Nat Biotechnol* **2016**, *34* (5), 518-524. DOI: 10.1038/nbt.3423 From NLM.

(57) Smeets, R. M.; Keyser, U. F.; Krapf, D.; Wu, M.-Y.; Dekker, N. H.; Dekker, C. Salt dependence of ion transport and DNA translocation through solid-state nanopores. *Nano letters* **2006**, *6* (1), 89-95.

(58) Parkin, W. M.; Drndić, M. Signal and noise in FET-nanopore devices. *ACS sensors* **2018**, *3* (2), 313-319.

(59) Vereecke, N.; Bokma, J.; Haesebrouck, F.; Nauwynck, H.; Boyen, F.; Pardon, B.; Theuns, S. High quality genome assemblies of Mycoplasma bovis using a taxon-specific Bonito basecaller for MinION and Flongle long-read nanopore sequencing. *BMC Bioinformatics* **2020**, *21* (1), 517. DOI: 10.1186/s12859-020-03856-0.

(60) Schwartz, S.; Motorin, Y. Next-generation sequencing technologies for detection of modified nucleotides in RNAs. *RNA biology* **2017**, *14* (9), 1124-1137.

(61) Jenjaroenpun, P.; Wongsurawat, T.; Wadley, T. D.; Wassenaar, Trudy M.; Liu, J.; Dai, Q.; Wanchai, V.; Akel, N. S.; Jamshidi-Parsian, A.; Franco, A. T.; et al. Decoding the epitranscriptional landscape from native RNA sequences. *Nucleic Acids Research* **2020**, *49* (2), e7-e7. DOI: 10.1093/nar/gkaa620 (acccessed 9/5/2022).

(62) Workman, R. E.; Tang, A. D.; Tang, P. S.; Jain, M.; Tyson, J. R.; Razaghi, R.; Zuzarte, P. C.; Gilpatrick, T.; Payne, A.; Quick, J. Nanopore native RNA sequencing of a human poly (A) transcriptome. *Nature methods* **2019**, *16* (12), 1297-1305.

(63) Ayub, M.; Hardwick, S. W.; Luisi, B. F.; Bayley, H. Nanopore-based identification of individual nucleotides for direct RNA sequencing. *Nano letters* **2013**, *13* (12), 6144-6150.

(64) Astier, Y.; Braha, O.; Bayley, H. Toward single molecule DNA sequencing: direct identification of ribonucleoside and deoxyribonucleoside 5 ‘-monophosphates by using an engineered protein nanopore equipped with a molecular adapter. *Journal of the American Chemical Society* **2006**, *128* (5), 1705-1710.

(65) Reiner, J. E.; Balijepalli, A.; Robertson, J. W.; Drown, B. S.; Burden, D. L.; Kasianowicz, J. J. The effects of diffusion on an exonuclease/nanopore-based DNA sequencing engine. *The Journal of chemical physics* **2012**, *137* (21), 214903.

(66) Athapattu, U. S.; Amarasekara, C. A.; Immel, J. R.; Bloom, S.; Barany, F.; Nagel, A. C.; Soper, S. A. Solid-phase XRN1 reactions for RNA cleavage: application in single-molecule sequencing. *Nucleic Acids Res* **2021**, *49* (7), e41. DOI: 10.1093/nar/gkab001 License at https://creativecommons.org/licenses/by/4.0/legalcode.

(67) Nagarajan, V. K.; Jones, C. I.; Newbury, S. F.; Green, P. J. XRN 5′→3′ exoribonucleases: Structure, mechanisms and functions. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **2013**, *1829* (6), 590-603. DOI: https://doi.org/10.1016/j.bbagrm.2013.03.005.

Chapter 2. Operation of an Immobilized Nanoscale Enzymatic Reactor (INER) and Fluidic Structures to Facilitate Entry of Single RNA/DNA Molecules into *X-ToF*

2.1 Introduction

2.1.1 Characteristics of the *X-ToF* SMS

We are developing an innovative nanofluidic device for single-molecule sequencing (SMS) of both DNA and RNA and uses an immobilized processive exonuclease enzyme to sequentially cleave single nucleotides from an intact DNA or RNA molecules and identifies them based on their time-of-flight (i.e., electrophoretic mobility), *X-ToF* (see Figure 2.1). In the case of RNA sequencing to fully explain device operation, *X-ToF* is based on processive riboexonuclease digestion of single RNA molecules that sequentially generates ribonucleic monophosphates (rNMPs); the processivity of the enzyme determines the effective RNA read length. This is followed by the detection and identification of single rNMPs via their molecular-dependent time-of-flight (ToF), which is deduced as they pass through two in-plane nanopores located in series and separated by a nanoscale flight tube that acts as an electrophoresis column. The *X-ToF* device was designed to avoid diffusion issues that are known to accompany exo-sequencing;[1] this is accomplished by placing the exonuclease within a nanofluidic network and operating under high Peclet numbers to avoid diffusional misordering and increase loading efficiency of single rNMPs. Due to *X-ToF's* use of exo-sequencing and ToF detection, it has been coined *X-ToF*. *X-ToF's* revolutionary design and operational characteristics is poised to significantly improve upon NGS and other SMS technologies by: (i) Multi-parameter identification (ToF, dwell time, peak amplitude) of individual rNMPs (or dNMPs for DNA) to provide high read accuracy, even for single reads; (ii) the ability to read epigenetic or epitranscriptomic information directly with high accuracy because of lack of need for PCR and reverse transcription; (iii) fabrication via nano-injection molding of the device, which will provide a production route conducive to high-scale production at low-cost; (iv) interrogating single molecules with long reads (>50 Kb); (v) ability to process small amounts of sample (<5 ng) even without PCR amplification; (vi) simple library preparation; (viii) ability to integrate other microfluidic devices to *X-ToF* to provide full automation of sample preparation; and (iv) the ability to fine tune conditions (pore sizes, applied voltage, pH of buffer, Mg^{2+} conc., *etc*.) to improve read accuracy.

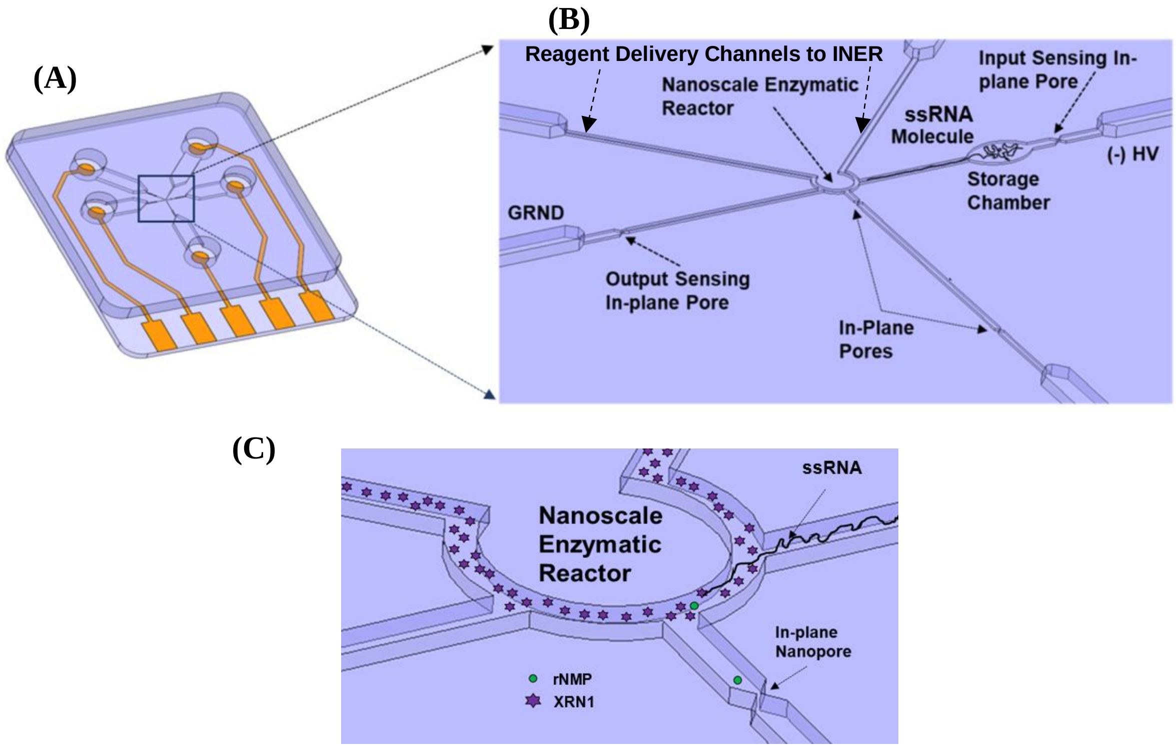

Figure 2.1. Overview of *X-ToF* featuring three different views (Figure adapted with permission from reference [2]). **(A)** *X-ToF* is fabricated in a thermoplastic using a replication process and is thermally fusion bonded to a cover plate containing electrodes for simple connection and control of process steps. **(B)** Entire nanochannel network showing the four in-plane nanopores used for sensing the input, association of RNA to the immobilized enzyme, and detecting the ToF of single rNMPs following enzymatic cleavage. **(C)** View of the INER with immobilized XRN1 processive enzyme for digestion of a ssRNA releasing rNMPs for detection and identification.

2.1.2 Operation of the *X-ToF* SMS

X-ToF is composed of three functional units: (i) An input/output network equipped with two in-plane pores to transduce the entry of RNA/DNA into the INER; (ii) an INER comprised of two feed channels for the functionalization of the INER, which consists of a surface-immobilized enzyme; and (iii) the ToF channel containing two in-plane pores that flank either end of the ToF channel to allow for single-molecule electrophoresis (see Figure 2.1(B)). The *X-ToF* nanofluidic network is interfaced to access microchannels via tapered funnels to allow for introduction of targets from the macro-world. The access microchannels are also connected to reservoirs for sample and reagent input as well as application of the required voltage sequence for SMS.

In a sequencing run for ssRNA, a single ssRNA is electrokinetically driven through an input tapered micro-funnel, then translocates through the first in-plane nanopore and is shuttled into the INER. For RNA, the INER will possess XRN1 or another riboexonuclease that is covalently immobilized to its surface.[3] The 5’ end of the ssRNA will associate with the XRN1 enzyme by entering its active site and be held in place via electrostatic interactions.[4] If the association does not occur, then a second resistive pulse signal (RPS) will be generated at the second in-plane pore located in the exit channel, but absent if the RNA/XRN1 association occurs. The cofactor for the XRN1 enzyme, Mg^{2+}, is supplied using the two upper channels connected to the INER. Mg^{2+} is driven electrokinetically into the INER, thereby successfully delivering the cofactor initiating enzymatic activity. Following delivery of the Mg^{2+} cofactor, processive digestion of the ssRNA target will occur (see Figure 2.1(C)). Based on our results, surface immobilized XRN1 will cleave a single rNMP from the strand approximately every 26 s.[3] The enzymatically released rNMP is then electrokinetically driven through a first and second in-plane nanopore that flank the ToF nanochannel. The ToF of any particular rNMP is the amount of time it takes it to pass between the first and second in-plane nanopores and provide the signals for measuring the ToF elucidated via RPS.[5,6] The ToF for each rNMP (A, U, G, C) will be different due to their molecular-dependent electrophoretic mobilities.[7] By comparing the measured ToF of any rNMP to the known ToF for each rNMP, bases can be called generating the RNA sequence.

There are a total of four in-plane nanopores housed within the fluidic network of the *X-ToF* device. Two are used for the ToF detection and two are used to monitor the location of ssRNA as it enters the nanochannel network of *X-ToF*. One in-plane nanopore is located after the input micro-funnel as noted above, but before the INER, while the other is located after the INER. While monitoring the RPS signals across this nanochannel containing the two in-plane nanopores, we secure information as to the success of a RNA/XRN1 association event to make sure that only ONE ssRNA molecule is positioned within the INER and available for sequencing. If multiple molecules are present within the INER and due to the non-synchronicity of the enzymes' activity, the sequencing data acquired will be meaningless. Conversely, if there are no ssRNAs within the INER, then no sequencing data would be generated.

As noted above, the two in-plane nanopores in the ToF nanochannel are poised for the detection and identification of the rNMPs sequentially generated by XRN1's activity on the input ssRNA. As a rNMP passes through the first in-plane nanopore, an RPS signal will occur followed by a second RPS signal as the same rNMP molecule passes through the second in-plane nanopore. The time difference between the first and second RPS signal is recorded as the ToF for the rNMP and used to make a base call.

2.1.3 Fabrication of the *X-ToF* SMS

The *X-ToF* device can be fabricated in a thermoplastic, such as polymethyl methacrylate (PMMA) or any other thermoplastic. The choice of a thermoplastic is predicated on the fact that plastics can be produced in a high production mode via replication,[8] such as nanoimprint lithography (NIL) or nano-injection molding (nIM).[9] Following replication of the fluidic circuit, a cyclic olefin copolymer (COC) 8007 cover plate is used to enclose the fluidic network via thermal fusion bonding.[10] The choice of COC 8007 is that it has a lower glass transition temperature (T_g) compared to PMMA. This combination allows for thermal fusion bonding to be completed at a temperature that ensures no significant deformation of the underlying nanofluidic channels to occur due to the higher T_g of PMMA with respect to COC 8007 (process yield rate >90%).[10] Fabrication of *X-ToF* starts by creating the desired microchannels in Si via photolithography. The micro-funnels and nanochannels are then fabricated into the Si wafer via focused ion beam (FIB) milling. From this Si master, a resin stamp may be fabricated by using a variety of UV curable resins (see Figure 2.2A).[11] This resin stamp in turn contains the reverse tone of fluidic structures as present on the Si master. This resin stamp is then used to thermally imprint these fluidic structures into the PMMA polymer using either thermal NIL or nano-injection molding (see Figure 2.2B). Following the replication step, the COC 8007 cover plate is added via thermofusion bonding after an oxygen plasma treatment to facilitate this process.[10, 12]

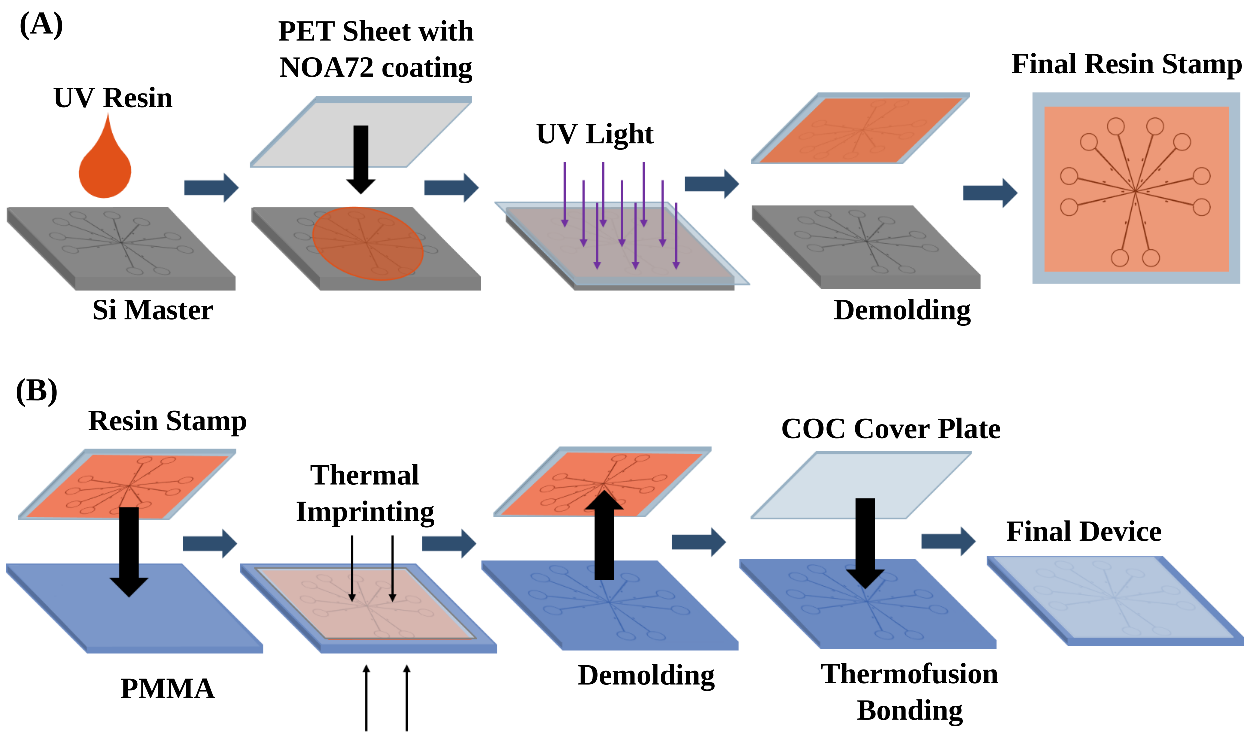

Figure 2.2. Experimental process of fabricating the *X-ToF* device starting with the Si master. **(A)** Resin stamps are fabricated from polyurethane acrylate (PUA) UV curable resin and a PET sheet with an adhesive coating. **(B)** Resin stamps are used to imprint structures into polymethyl methacrylate (PMMA) thermoplastic using thermal imprinting. The stamp is demolded from the resulting device and a COC cover plate is added to yield the final device.

2.2 INER with XRN1 for digestion of single RNA molecules

Immobilized microscale reactors (IMERs) and INERs (Immobilized Nanoscale Enzymatic Reactors) are two subgroups of solid phase bioreactors.[13] IMERs contain structures, which are on the microscale, while INERs contain structures on the nanoscale. This difference in scale leads to a much higher confinement in INERs compared to IMERs, which can reduce the effects of diffusion and produce higher reaction rates for the immobilized enzyme.[14] Previously, Nyote *et al.* fabricated a thermoplastic IMER using λ-Exo immobilized via EDC/NHS coupling chemistry onto the surface of PMMA to investigate its digestion of dsDNA. Digestion studies were first carried out by pumping a solution of dsDNA through the IMER and detecting the fraction of undigested dsDNA using a spectrofluorimeter. Following this, digestion was investigated in real time by digesting a fluorescently-stained dsDNA in an IMER while observing the process using an epifluorescence microscope. It was found that complete digestion occurred using the IMER and exhibited a higher digestion efficiency than in free solution reactions. Further, the real time observation of the enzymatic reaction revealed the digestion rate was $1.0 \times 10^3 \pm 100$ nt/s and the processivity was >40 kbp at 37°C. This higher processivity was attributed to an increase in the stability of the enzyme when immobilized.[13] Athapattu *et al.* similarly used thermoplastic IMERs to investigate XRN1's digestion of RNA. Immobilization of the XRN1 enzyme was again achieved using EDC/NHS coupling chemistry. A solution of RNA was pumped through an IMER and the undigested RNA fraction was detected using a spectrofluorimeter. It was found that 87.6 ±2.8% of RNA was digested by the IMER while a lower fraction at (78.3 ±4.4%) was digested when performed in free solution. This was followed by real time observation of XRN1's digestion of fluorescently-stained RNA in an IMER using an epifluorescence microscope. From these results, it was determined that XRN1 exhibited a clipping rate of 26 ±5 nt/s and had an apparent processivity of >10.5 kb at 25° when immobilized.[3]

Here we focused on the work to develop an INER, which includes selectively functionalizing the reactor's surface with XRN1 enzyme and demonstrating its ability to digest RNA in nano-confinement. For INERs, we will be interested in scaling effects. For example, most IMERs (in terms of the reaction kinetics are diffusion-limited. However, this may not be the case for INERs because diffusional distances are significantly reduced and thus, leading to higher substrate/enzyme encounter rates providing higher turnover rates for INERs compared to IMERS.

Indeed, Bohn and coworkers found that reaction rates for INERs using horseradish peroxidase were ~10-fold higher than IMERs.[14] This will be mitigated by the fact that in this work we are dealing with a processive enzyme (XRN1), which encounters its substrate, forms a complex with the substrate, and then due to its motor activity, will move along the substrate carrying out the enzymatic reaction. In the case of XRN1, this consists of clipping a terminal ribonucleotide from the intact RNA molecule in the 5' → 3' direction.[15] As a final note, we will be working in a domain for our INERs in which the product of the inverse Debye length κ (~0.8 nm^{-1}) and the channel diameter (~250 nm) is >1, which will reduce concentration polarization that can exclude co-ions (exclusion of negatively charged substrates from the INER with negatively charged surface).

2.3 Methods

2.3.1 Device fabrication

Fabrication of *X-ToF* was achieved using resin stamps made from a Si master, nano-imprint lithography (NIL) to imprint the structures into PMMA, and the addition of a COC cover plate using thermofusion bonding following oxygen plasma treatment (see Figure 2.2). The oxygen plasma treatment was performed to generate carboxylic acid groups[16,12] to make the thermoplastic surfaces more hydrophilic. This facilitated the thermal fusion bonding process[17] and provided chemical groups on the INER that can later be used for covalent attachment of XRN1.[3] This device fabrication was followed by multiple reaction steps to achieve enzyme immobilization thereby completing fabrication of the INER.

Resin stamps were fabricated by dispensing 70 µL of polyurethane acrylate (PUA) resin onto the surface of the Si master. A PET sheet with a NOA72 adhesive coating was placed on top of the resin and gently pressed down to eliminate trapped air bubbles. The resin stamp was then exposed to UV/O_3 (3 min, 19.5 mW/cm^2 measured at 254 nm) to cure the resin. The resin stamp was carefully demolded from the Si master and the edges were trimmed. The fabricated resin stamp was then used to transfer structures to PMMA using thermal NIL. The resin stamp was placed on top of PMMA polymer, and thermal imprinting was carried out (145°C, 300 psi, 5 min). After thermal imprinting, structures were successfully transferred from the resin stamp to the PMMA polymer with high replication fidelity. These structures were covered with a protective sheet of plastic while the edges of the devices were trimmed, and reservoirs drilled using a 1 mm drill bit. To complete the *X-ToF* device fabrication, a COC 8007 cover plate was added to enclose the fluidic channels fabricated in the PMMA substrate. The cover plate and PMMA device were exposed to O_2 plasma (50 mW, 20.5 sccm, 1 min) to modify the surface, facilitate fusion bonding, and generate chemical carboxylic acid groups[16] for enzyme immobilization. The cover plate and PMMA device were bonded together using NIL (70°C, 110 psi, 15 min).

2.3.2 Enzyme immobilization

Attachment of XRN1 to the INER within the *X-ToF* device was achieved using EDC/NHS coupling chemistry.[3] To ensure delivery of reagents to the INER in the *X-ToF* device, a waveform generator was used to apply a bias voltage across the INER. This was done by immersing electrodes into reservoirs of the microchannels that connect to the INER (Figure 2.1). A schematic for this immobilization reaction can be found in Figure 2.3. Both the COC 8007 cover plate and PMMA device were activated with O_2 plasma before bonding with the NIL. This generated carboxyl groups on the thermoplastic surface.[16] After bonding, a solution of 50% MeOH was filled into the microchannels for 5 min. The MeOH solution was then removed using a vacuum pump and replaced with 1× NEBuffer 3 (100 mM NaCl, 50 mM Tris-HCl, 1 mM DTT, pH = 7.9). This buffer was removed from two microchannels connected to two nanochannels, which were used to deliver reagents to the INER (Figure 2.1). These microchannels were then filled with 0.1 M MES buffer containing 10 mg of EDC and 1 mg of NHS. The waveform generator was used to apply a biased voltage of 5 V across these microchannels and the INER for 30 min. This served to drive the solution of EDC and NHS through the nanoreactor to ensure that these reagents reached and reacted with the carboxylic acid groups on the thermoplastic surfaces. This solution was removed and replaced with a 737 nM XRN1 enzyme solution. The waveform generator was again used with a biased voltage of 5 V to drive the enzyme solution through the INER for 30 min. The waveform generator was removed and the device was left for an additional 90 min to ensure enzyme immobilization.

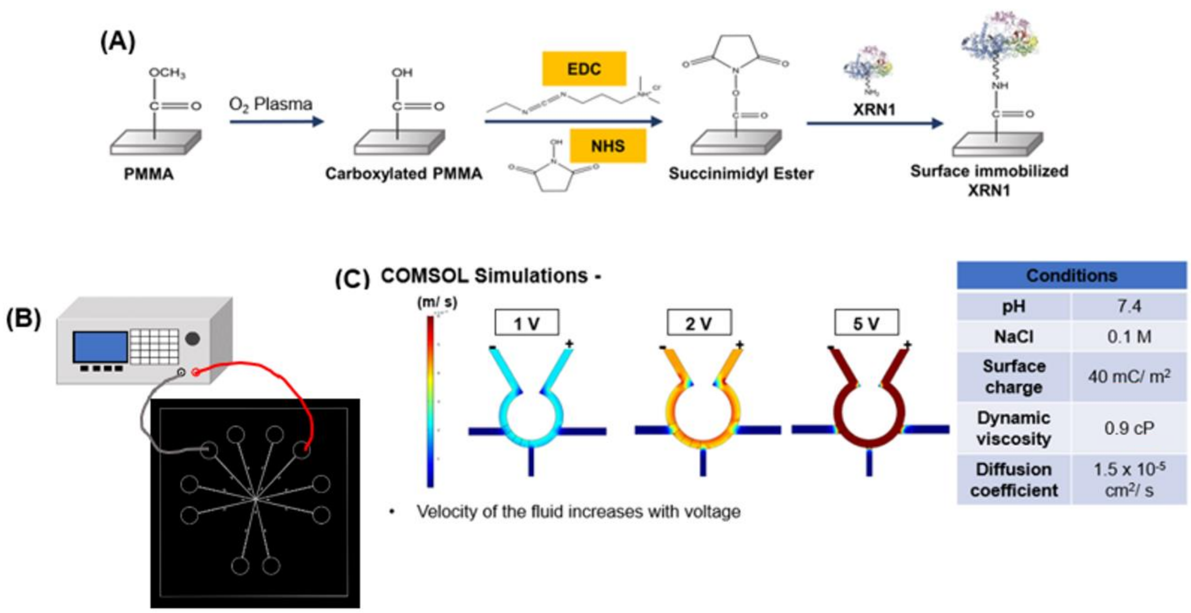

Figure 2.3. Immobilization of XRN1 onto the INER of the *X-ToF* device. **(A)** The XRN1 enzyme was immobilized onto the PMMA surface using EDC/NHS coupling chemistry. Oxygen plasma treatment was performed before bonding the cover plate to the device. EDC/NHS and later XRN1 were filled into the reagent delivery channels of the INER to deliver necessary reagents. (Figure adapted from reference [3]). **(B)** Electrodes were connected to a waveform generator and immersed

into the microchannels that connect to the INER. A voltage was applied across the INER to drive reagents through the INER. **(C)** COMSOL simulations to investigate the behavior of solutions as they electrokinetically travel through the INER. It was observed that as the voltage was increased, the speed of the solution moving through the INER increased (adapted from reference [18]).

During XRN1 immobilization, electrodes from a waveform generator were immersed into the reagent delivery reservoirs, which connected to either side of the INER, to electrokinetically deliver reagents to the INER. This strategy had been investigated previously using COMSOL simulations by Dr. Swarnagowri Vaidyanathan (see Figure 2.3(C)). In these simulations, a voltage was applied across the nanoreactor in the *X-ToF* device and the fluid dynamics of the solution observed. It was found that as the applied voltage increased, the speed of the solution passing through the INER increased. It was also found that the solution was confined to the INER when driven through the nanochannels electrokinetically and did not leak significantly into adjoining nanochannels. This was important to ensure that the enzyme was only immobilized only within the INER. To further ensure that reagents were being delivered through the INER, the conductance was checked after emptying and refilling microchannels with a new solution. This was done to confirm that solution was filling from the microchannels and into the INER and that there were no air bubbles in the channels that would disrupt an electrical connection. If devices displayed little or no conductance after multiple attempts to fill the devices, they were discarded.

2.4 Results and discussion

2.4.1 Site-directed immobilization onto INER nanopillar of X-ToF (Data and experimental results reproduced from reference [2] with permission).

Before the enzyme immobilization reaction was undertaken and to ensure the flow of the solution around the INER nanopillar could be restricted electrokinetically to only the INER, we carried out experiments using a 20 nM Rhodamine B tracer dye solution. The Rhodamine B solution was introduced into one reagent delivery microchannel of *X-ToF* (see Figure 2.1(B)). Then, a 5 V potential was applied across the channels to electrically drive the dye from one delivery channel to another with the intervening element being the INER. As can be seen from the fluorescence image, the direction of the flow could be controlled by adjusting the applied potential. Moreover, we carried out COMSOL simulations to further demonstrate the ability to electrokinetically control the solution to selectively immobilize biomolecules around the bioreactor. Increasing the voltage increased the velocity of the solution (see Figure 2.3(C)). At all voltages tested, minimum diffusion of the solution to the adjoining nanochannels of the INER was seen.

Assembled devices were primed with 50% MeOH for 5 min followed by the introduction of 1× PBS buffer. After 15 min, the PBS buffer in one of the enzyme input/output reservoirs was replaced with a solution containing a 50 nM Cy3-labeled oligonucleotide reporter and 20 mg/mL EDC in PBS buffer. After introduction of the Cy3-oligonucleotide, a DC voltage was applied across the INER reagent deliver channel reservoirs to electrokinetically drive the dye-labeled oligonucleotide molecules around the nanopillar of the INER. After 15 min, the dye solution was completely removed using a vacuum and washed with 0.05% Tween 20 followed by 50% MeOH to remove non-specifically bound molecules. After washing, the devices were imaged using an epifluorescence microscopy (λ_{ex} = 532 nm).

To make sure non-specific adsorption of the Cy3-labeled oligonucleotide was negligible, we first used a 50 nM solution of Cy3 oligonucleotide that was introduced into the *X-ToF* device without EDC (see Figure 2.4(A)). A driving potential of 5 V was applied across the reagent delivery channels and fluorescence images of the *X-ToF* device were collected at different times. After 15 min, the oligonucleotide solution was removed and the device was washed with 0.05%

tween 20 and 50% MeOH. The fluorescence intensity of the device at different positions were measured (see Figure 2.4(A)). The fluorescence intensity at each position was similar to the background intensity indicating that non-specific adsorption was negligible.

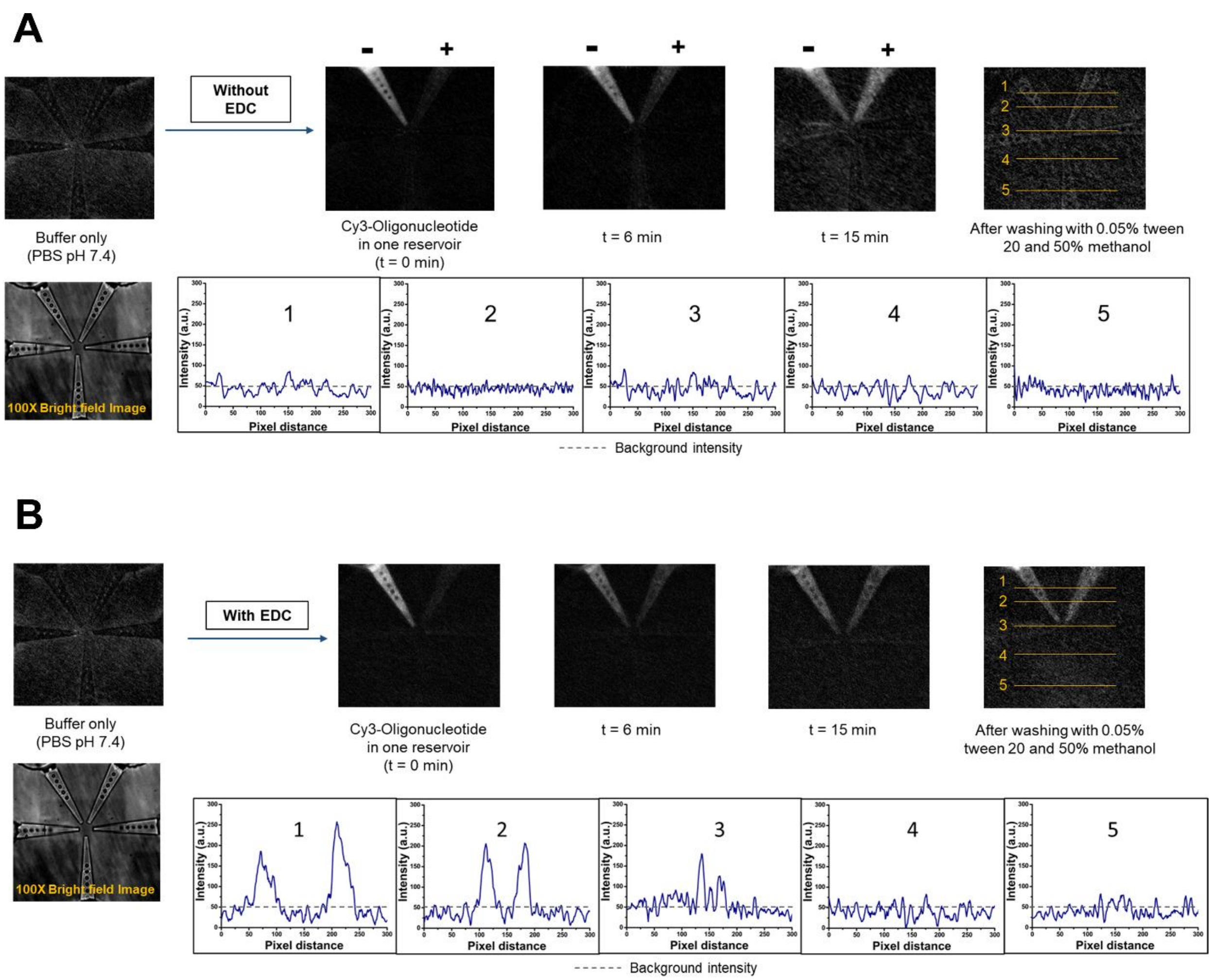

Figure 2.4. (A) Non-specific adsorption of Cy3-labeled oligonucleotides to a non-functionalized containing INER. (B) Covalent attachment of Cy3-labeled oligonucleotides using a one-step EDC immobilization reaction. Figure reproduced with permission from reference [2]

Next, the Cy3-labeled oligonucleotide was introduced into the INER, but this time in the presence of 20 mg/ mL EDC followed by 5 V potential to drive the solution through the INER (see

Figure 2.4(B)). After 15 min, the solution was removed and washed with 0.05% Tween 20. The fluorescence intensity was measured at different positions of the device. As can be seen from Figure 2.4(B), the fluorescence intensity was above background in the reagent delivery channels (position 1 and 2) and around the INER's pillar (position 3). This indicated that electrokinetic control of the flow of solution through the INER could selectively immobilize the enzyme. Any diffusional effects into channels where the reaction is not slated to occur can be mitigated by using electroosmotic or electrophoretic forces to reduce this diffusional artifact by applying voltages in the input/output channels and the flight tube.

2.4.1.1 Validation of successful XRN1 immobilization via single-molecule fluorescence imaging

INERs were fabricated using the previously described process. The purpose of these experiments was to ensure enzyme immobilization was successful and to study the activity of XRN1 in the INER with nanoscale confinement. The enzyme activity was monitored by observing fluorescently labeled RNA in an INER using single-molecule fluorescence microscopy. A comparison was made of RNA behavior in an INER with XRN1 covalently immobilized to that which did not have immobilized enzyme. This was followed by a study to complex RNA with immobilized XRN1 in the INER to observe its potential digestion by XRN1 in the INER. Several key metrics were acquired such as the XRN1 clipping rate.

2.4.1.2 INER digestion of Single RNA Molecules

The immediate goal was to compare INERs that had undergone XRN1 immobilization against INERs that had no XRN1. These devices are referred to as XRN1 positive INERs and XRN1 negative INERs. The dynamics of RNA travel and digestion were studied in each group using single-molecule fluorescence microscopy. The general experimental setup can be found in Figure 2.5.

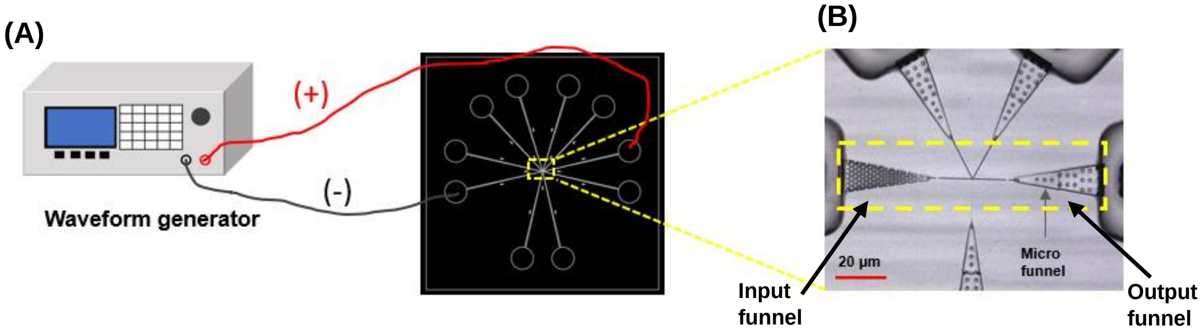

Figure 2.5. (A) A waveform generator was connected across the INER by immersing electrodes into input/output channel reservoirs on either side. Cas9 RNA labelled with SYTO82 dye was added to the microchannel reservoir, which consisted of the input channel containing an in-plane pore and connected to the INER. 1-2 V DC was used to drive the RNA through the INER and to the output channel, which also contained an in-plane pore. **(B)** Zoomed image of the INER. The RNA in these experiments was driven through the input/output channels of *X-ToF*. The input channel with a tapered funnel has approximately 80 pillars to facilitate unwinding of the RNA and entropically ease its entry into the device.

After fabrication, *X-ToF* devices were filled with buffer (100 mM NaCl, 50 mM Tris-HCl, 1 mM DTT, pH = 7.9; this is called NEBuffer 3 and is used for XRN1 enzymatic reactions). Cas9 RNA (4,245 nt) was labelled with SYTO82 (RNA staining dye) and added to access microchannel that connects to the input tapered funnel. Electrodes were connected to a waveform generator and immersed in reservoirs on either side of the INER. The entire device assembly was mounted onto the stage of an epifluorescence microscope, which consisted of a NIKON TE 2000 microscope, 100X/1.4 NA oil-immersion objective, and an Andow iXon3 EMCCD camera. The laser source was a Coherent Nd:YAG; λ_{ex} = 532 nm; P = 0.01-5 W; 2.2 mm beam diameter.

With the application of the DC driving voltage, Cas9 RNA was electrokinetically driven through the *X-ToF* device at 1 to 2 V DC. Images were collected using the EMCCD camera with

Metamorph advanced v7.5.6.0 software. ImageJ software was later used to subtract the background from recorded videos for further imaging processing. Fluorescence intensity data was processed using Origin 2021b SR2 version 9.85.212 software with a 50% percentile filter applied to reduce noise.

The travel dynamics of RNA was observed as it moved through the nanochannels and to the INER. Initial testing demonstrated that RNA could translocate from the input microfunnel through the nanochannels, past the INER, and to the output microfunnel. It was hypothesized that when XRN1 was present within the INER, the translocating RNA would enter the immobilized enzyme's active site through electrostatic interactions and effectively complex with the enzyme causing it to be held in place.[15] These experiments were performed using NEBuffer 3 without Mg^{2+} needed to initiate digestion, but the RNA could still enter XRN1s active site and complex with the enzyme.[3]

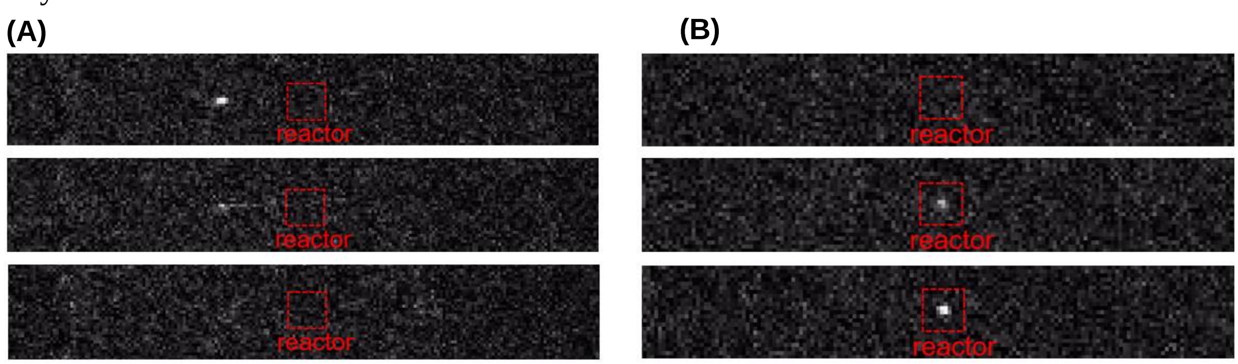

Figure 2.6. (A) SYTO82 labelled Cas9 RNA was observed in an XRN1 negative INERs that had not undergone the enzyme immobilization process. This RNA was electrokinetically driven from the input microfunnel of the *X-ToF*, past the INER, and through the output microfunnel. No RNA remained at the INER, but instead moved past and exited the INER. **(B)** SYTO82 labelled RNA was observed in an INER containing covalently immobilized XRN1. A SYTO82 labelled RNA was electrokinetically driven through the input microfunnel and to the INER. It was observed that

the RNA became stationary at the INER, complexing with XRN1. This resulted in a bright fluorescent stationary signal being observed within the INER.

INERs that did not contain XRN1 it was noted that several SYTO82 labelled RNA molecules could be electrokinetically driven into the input nanochannels and past the INER. Representative images from one of these translocation events is shown in Figure 2.6(A). This RNA translocated past the INER without remaining stationary. INERs that had undergone XRN1 immobilization were tested as well. In this case, RNA would move into the INER and become stationary, producing a fluorescent area of high intensity. Images from these experiments are shown in Figure 2.6(B).

After initial characterization of XRN1 positive and negative INERs, we sought to determine if digestion of the complexed RNA could occur in the INER. To rule out photobleaching, a negative control was performed to determine the photostability of SYTO82 labelled Cas9 RNA. An *X-ToF* device was filled with NEBuffer 3 that did not contain Mg^{2+}. SYTO82 labelled RNA was added to the same microchannel reservoir. The RNA solution contained 4% β-mercaptoethanol to act as an oxygen scavenger and to minimize photobleaching.[19,20] The waveform generator was used to drive RNA to the INER where it complexed and was held in place by XRN1. Then, the waveform generator was turned off and the fluorescence intensity of the RNA was observed. It was found that RNA did not exhibit signs of photobleaching for 13 min, at which point the experiment was ended due to the expected digestion time of Cas9 RNA by XRN1 to be <13 min. A graph of florescence intensity and representative images from the experiment are displayed in Figure 2.7. As can be seen from this experiment, the fluorescence signal was stable and showed no signs of fluorescence loss arising from photobleaching. This is consistent with the literature, which reports SYTO82 exhibits low amounts of photobleaching.[21]

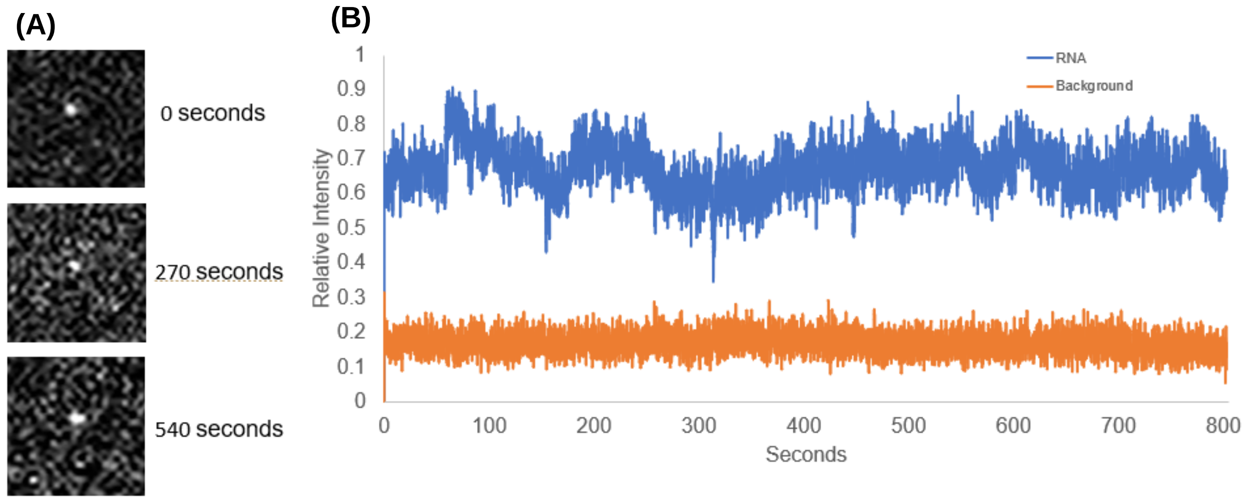

Figure 2.7. SYTO82 labelled Cas9 RNA was complexed with immobilized XRN1 on the INER and observed photostability. This experiment was conducted for 800 s, at which point it was terminated. **(A)** Still images of SYTO82 labelled Cas9 RNA within the INER and captured at 0, 270, and 540 s during the experiment. **(B)** The relative fluorescence intensity of Cas9 RNA and the background was plotted against time, showing that the RNA did not undergo photobleaching during the experiment. (Reproduced with permission from reference [2]).

After ruling out any photobleaching, the experiment was repeated with a NEBuffer 3 (100 mM NaCl, 50 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM DTT, pH = 7.9) that did contain $MgCl_2$. This provided XRN1 with the cofactor needed to facilitate digestion of the RNA. It was expected that if the RNA was digested, the fluorescence intensity would decay. From this decay, we could calculate the experimental clipping rate and compare it to previously published clipping rates. This experiment was performed and proceeded as hypothesized. The RNA was driven to the INER where it complexed with the XRN1 becoming stationary. The fluorescence intensity slowly decreased as the RNA was digested. Representative images from the experiment are displayed in Figure 2.8(A) with a graph of the fluorescence intensity being displayed in Figure 2.8(B).

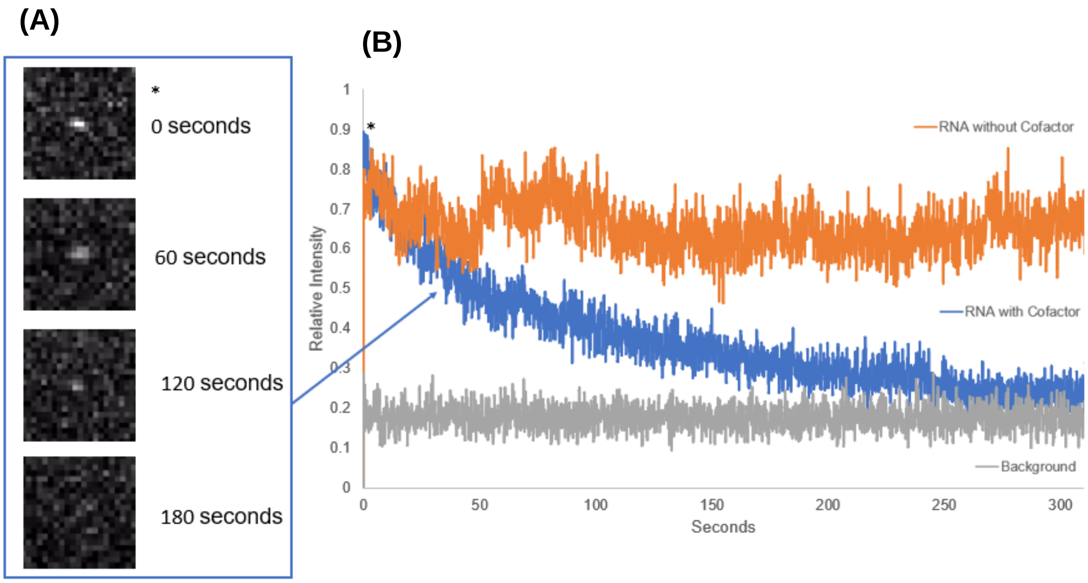

Figure 2.8. SYTO82 labelled Cas9 RNA was complexed with XRN1 enzyme immobilized on the nanoreactor in the presence of XRN1's cofactor. (Reproduced with permission from reference [2]). **(A)** Images of SYTO82 labelled Cas9 RNA stationary on the nanoreactor were captured at 0, 60, 120, and 180 seconds during the digestion experiment. **(B)** The relative fluorescence intensity of the RNA during the previous control experiment and the digestion experiment were plotted with the background intensity against time, showing that RNA was digested by XRN1 when $MgCl_2$ was added into the buffer and that this loss in fluorescence was not due to photobleaching.

From the shape of the decay curve and the initial fluorescence intensity observed, it was determined that a single RNA was complexed with the immobilized XRN1 enzyme within the INER and had undergone digestion. Using the decay in fluorescence intensity, the clipping rate of XRN1 enzyme in nano-confinement was calculated. The clipping rate was 23 ±3 nt s^{-1} (n=3). This did not differ significantly from the clipping rate previously reported from our group of 26 ±5 nt

s^{-1} of XRN1 using an IMER.[3] From these experiments, it was concluded that an INER utilizing XRN1 for the digestion of RNA could be used to first complex with an intact RNA molecule with the RNA translocation process controlled electrokinetically and in the absence of Mg^{+2}, no enzymatic cleavage would occur. Then, when Mg^{+2} is introduced into the INER, the enzymatic reaction could be initiated indicating the start of a sequencing reaction. This is an important result because during an *X-ToF* run, the RNA loading step is decoupled from the sequencing reaction meaning that for devices having many *X-ToF* sensors, all can be loaded with RNA followed by initiating the sequencing process in unison.

2.4.2 Development of nanostructures and operational parameters to facilitate RNA entry into the *X-ToF* device

The *X-ToF* device used for initial testing did not contain in-plane nanopores. These initial tests included the ability to electrokinetically drive RNA through the device and toward the immobilized XRN1 and initiate enzyme activity with the addition of Mg^{+2} using input and output channels that were 250 nm in depth and width, which is much larger than the persistence length of the ssRNA, which is ~9 nm under the salt conditions employed herein. These experiments were completed before adding in-plane nanopores due to the increased complexity of the nanofluidic design after their addition that can arise from co-ion exclusion within these in-plane pores resulting from slight overlap in the electrical double layer. The in-plane nanopores also present challenges when attempting to translocate RNA through them due to RNA's ability to form stable secondary structures[22] or resist translocation due to exclusion effects from the EOF.[23] While these issues are resolvable, initial experiments were focused on device design and the enzymology. After these initial studies, the next step was to fabricate a new Si master for the *X-ToF* device that contained a total of four in-plane pores. These in-plane nanopores were included so that RPS could be performed to sense RNA and rNMPs. RPS is the method of detecting an analyte as it passes through a voltage biased nanopore. During this process, ions from a buffer move through the nanopore and allow for electrical current to be conducted as a bias voltage is applied. When an analyte, such as RNA, enters a nanopore of comparable size it will displace a fraction of these ions. This displacement of ions causes an increase in the pore resistance and thus, a drop in the measured current.[24] This allows for the detection of an analyte as it passes through a nanopore. With this next design of the *X-ToF* device, the final proposed design could be further tested in terms of RPS of input RNA to optimize the loading efficiency of RNA into the device. After initial fabrication of these devices, the immediate goal was to translocate RNA from the input micro-funnel, through the input nanochannels, past the INER, and out the output nanochannel. Fluorescence microscopy was used to observe single fluorescently-labeled RNA molecules as they translocated through the device to determine the loading efficiency, which has a net effect on the amount of input RNA that is required to initiate a sequencing run.

2.4.2.1 Fabrication of *X-ToF* with in-plane pores

A total of four in-plane nanopores were added into the *X-ToF* design for RNA sensing and ToF detection. One in-plane nanopore was added in the input nanochannel between the microfunnel and the entropic trap, one in-plane nanopore in the output nanochannel, and two in-plane nanopores in the ToF nanochannel. A diagram of the *X-ToF* device with these four in-plane nanopores can be seen in Figure 2.9. The addition of these in-plane nanopores allowed for RNA and rNMP detection via RPS.

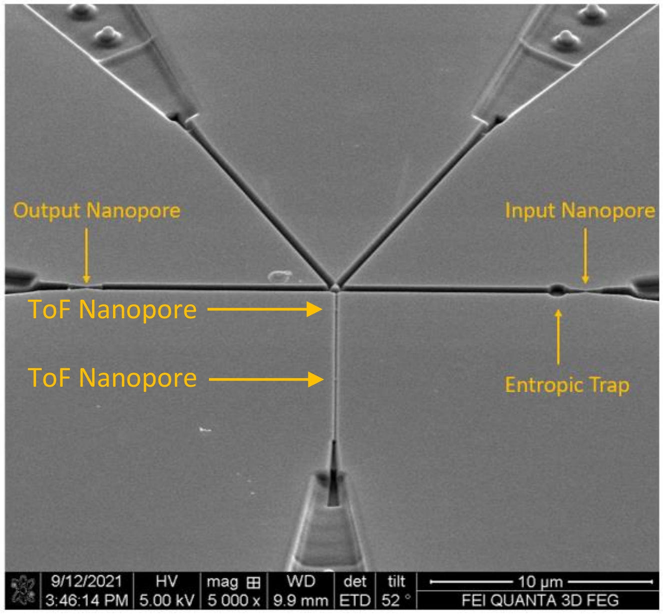

Figure 2.9. SEM image of a Si master of the *X-ToF* device with different components labeled for clarity (SEM image courtesy of Dr. Junseo Choi).

A combination of photolithography and FIB milling was used to create the Si master.[25] Photolithography was used to create the microscale structures in the Si master. This included the reservoirs and microchannels. Bitmaps of the microfunnels and nanochannels were used to direct the FIB milling (Courtesy of Dr. Junseo Choi) of the nanostructures such as the nanochannels, in-plane nanopores, entropic trap, and INER. A scanning electron microscope (SEM) image captured of the fabricated nanochannel network in the Si master is displayed as Figure 2.9. In this iteration, the input and output nanopores were ~35 nm (effective width and depth), the first ToF nanopore was 19 nm, and the second ToF nanopore was 15 nm.

Thermoplastic devices were fabricated using the same workflow as detailed in section 2.1.3. Briefly, resin stamps were fabricated using a PUA resin stamp to copy the structures from the Si master. Nanoimprint lithography (NIL) was then used to transfer the structures from the resin stamps to a PMMA thermoplastic. The microchannel reservoirs were drilled, and a COC 8007 cover plate was added to the PMMA device using thermal fusion bonding after O_2 plasma treatment.

2.4.2.2 RNA and DNA observations at in-plane pores

Experiments utilizing epifluorescence microscopy were used to test the new *X-ToF* design. Thermoplastic devices were fabricated in PMMA and treated with ethanolamine (ETA). Treatment with ETA was completed to suppress the surface charge thereby mitigating any issues such as high EOF or co-ion exclusion.[23] This was accomplished by using EDC/NHS coupling chemistry. A solution of 30 mM ETA containing 10 mg EDC and 1 mg NHS was prepared in 0.1 M MES buffer with a final volume of 500 µL. After fabrication, this solution was filled into the device and allowed to react for 30 min. Following this, the device was emptied and refilled with 1× NEBuffer 3. The device was mounted on the stage of the epifluorescence microscope and SYTO82 labelled RNA was added to the microchannel reservoir, which connected to the input nanochannel. Electrodes from a waveform generator were immersed into the reservoir containing the RNA and a reservoir connected to the output nanochannel. The RNA was fluorescently tracked while being electrokinetically driven through the input nanochannel and toward the in-plane nanopore.

It was observed that RNA could be driven through the input microfunnel and to the in-plane nanopore and entropic trap structures using 1-2 V DC. However, when the RNA reached the entropic trap, it would not translocate through toward the INER. This entropic trap was included in the original design of the *X-ToF* device for controlled storage and ejection of biomolecules.[26] Representative images from these experiments are presented in Figure 2.10 along with SEM and atomic force microscope (AFM) images detailing the nanofluidic structures.

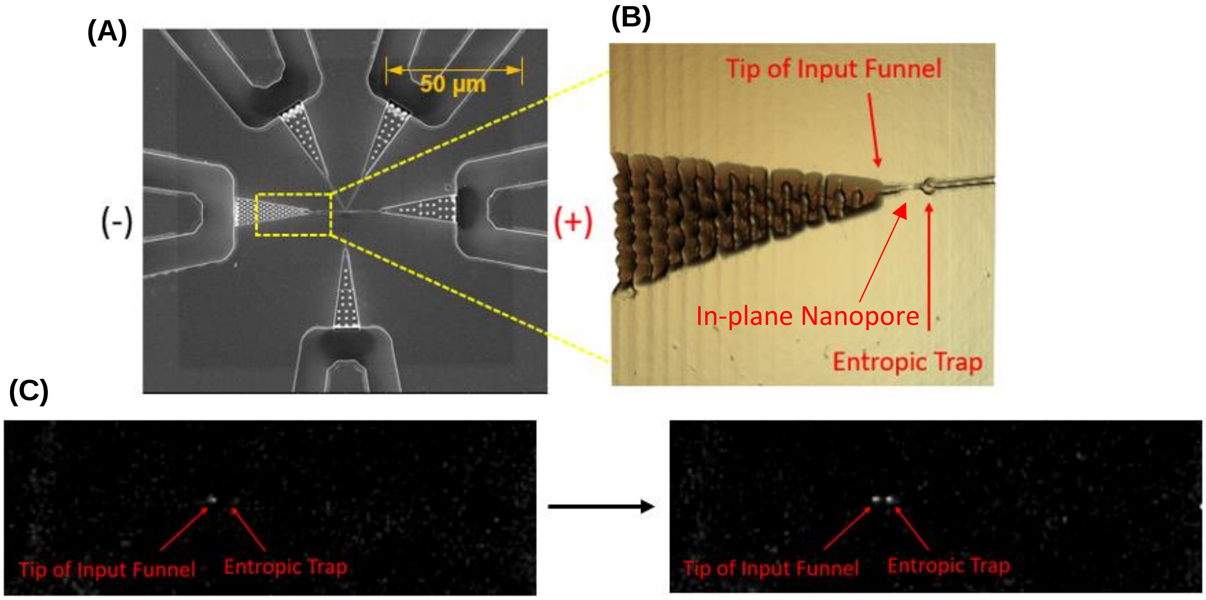

Figure 2.10. (A) SEM of the *X-ToF* device containing nanostructures. A yellow box surrounds the input funnel, in-plane nanopore, and entropic trap (Figure adapted with permission from reference [2]) (SEM image courtesy of Dr. Junseo Choi). **(B)** 3D generated AFM image showing features highlighted within the yellow box in **(A)**. The tip of the input funnel and entropic trap with the in-plane nanopore in between them are labelled for clarity. **(C)** Still frames captured from fluorescence microscopy observations. Fluorescently labelled Cas9 RNA was electrokinetically driven through the input funnel and channel. In the left image, RNA can be seen at the tip of the input funnel. The image on the right shows RNA partially proceeding through the in-plane nanopore but not completely translocating through it. As shown, RNA was observed on multiple occasions to partially translocate through the in-plane nanopore and entropic trap structures, but this translocation process would not go to completion.

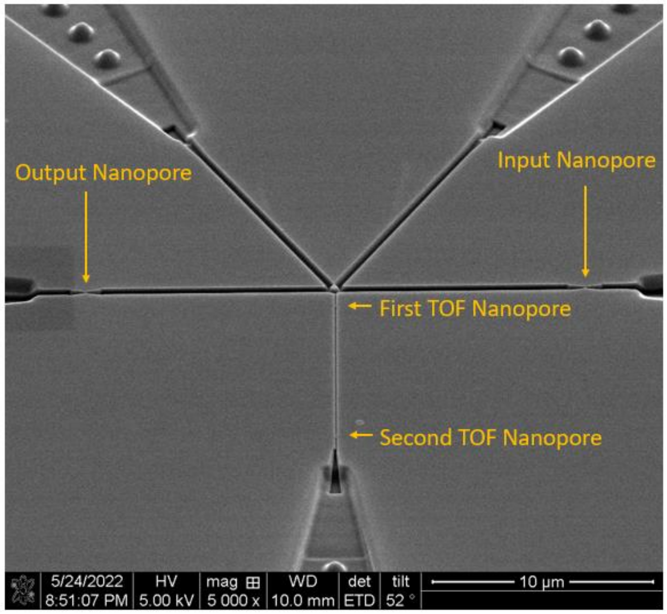

Figure 2.11. SEM image of *X-ToF* design containing in-plane nanopores, but no entropic trap (SEM image courtesy of Dr. Junseo Choi).

RNA could be driven through the input funnel and to the in-plane nanopore. However, when the RNA reached the in-plane nanopore and entropic trap structures, it would not translocate through. At this point, it was hypothesized that the translocation event was unsuccessful due to: (1) The electrophoretic force exerted on the RNA significantly dropping as it reached the entropic trap due to the increased size of the entropic trap compared to the in-plane nanopore and nanochannel; and (2) the RNA coiling to form stable secondary structures.[22] Based on this hypothesis, a new Si master of *X-ToF* design was fabricated without the entropic trap. In this iteration, the input nanopore was found to be 25 nm, the output nanopore was 23 nm, and both the first and second ToF nanopores were 20 nm in effective size. A SEM image of this Si master is displayed in Figure 2.11.

Devices were fabricated in PMMA with a COC cover plate. The devices were treated with ETA to reduce the surface charge and the device was filled with NEBuffer 3. In these experiments, it was continually observed that RNA would travel through the input funnel to the entrance of the in-plane nanopore but would not translocate through the in-plane nanopore (see Figure 2.12).

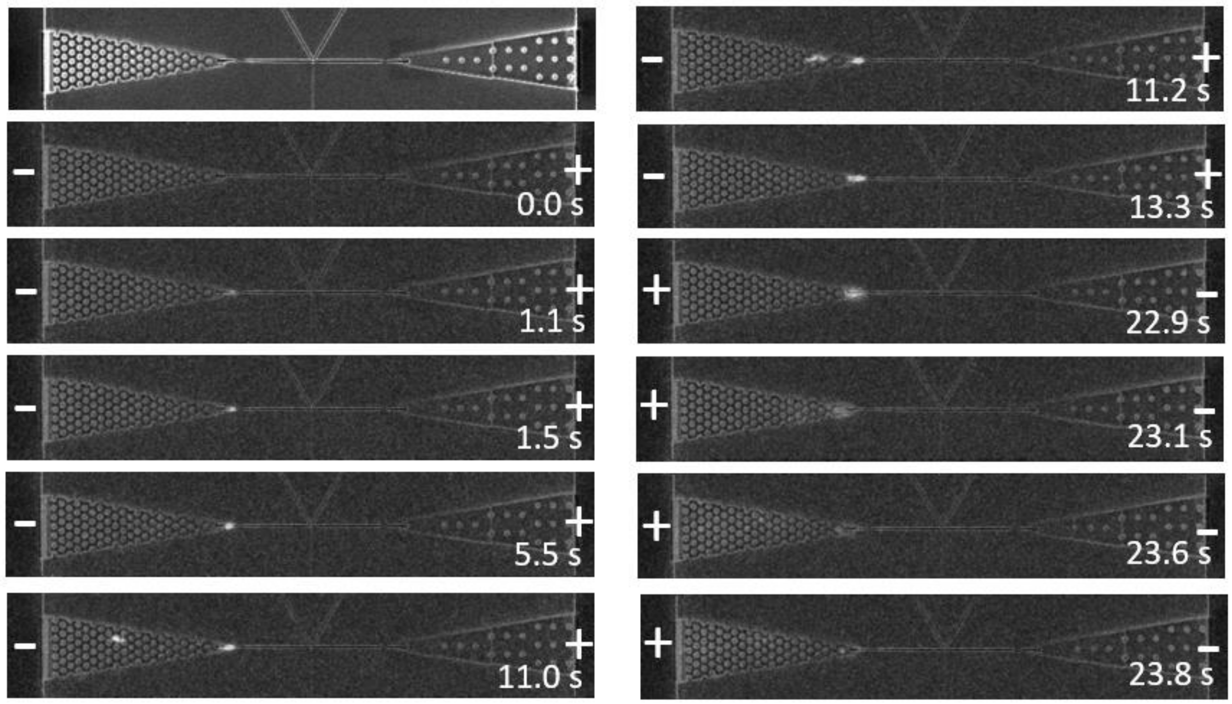

Figure 2.12. Frames from the fluorescence tracking of single RNA molecules electrokinetically traveling through the *X-ToF* design containing in-plane nanopores, but no entropic trap. A semi-transparent image of the *X-ToF* device has been laid over the top of these frames to clarify where the RNA is positioned within each frame. The RNA can be observed traveling to the tip of the input funnel from times 0.0 s until 13.3 s. Starting at 22.9 s, the polarity of the applied voltage was reversed. The RNA could then be seen exiting the tip of the input funnel and traveling back toward the access microchannel.

Following these observations, experiments were completed using a simpler dual in-nanopore device (DNP) that has been reported previously.[23] This device contained two in-plane nanopores in series with an approximate size of 30 nm that are separated by a nanochannel that is

5 μm in length and 50 nm in depth and width.[23] The buffer chosen for this experiment was 0.5× Tris borate EDTA (TBE), which was based on previous results where DNA was successfully driven into and stretched in nanochannels ranging from 300 x 200 nm down to 35 x 35 nm (depth and width).[27] Lambda DNA stained with POPO-3 was also used as the biomolecule of choice to test successful entry into and through the in-plane nanopores. Results from these experiments demonstrated that λ-DNA also encountered difficulties in terms of translocation when reaching the in-plane nanopores. Representative images from this experiment are displayed in Figure 2.13.

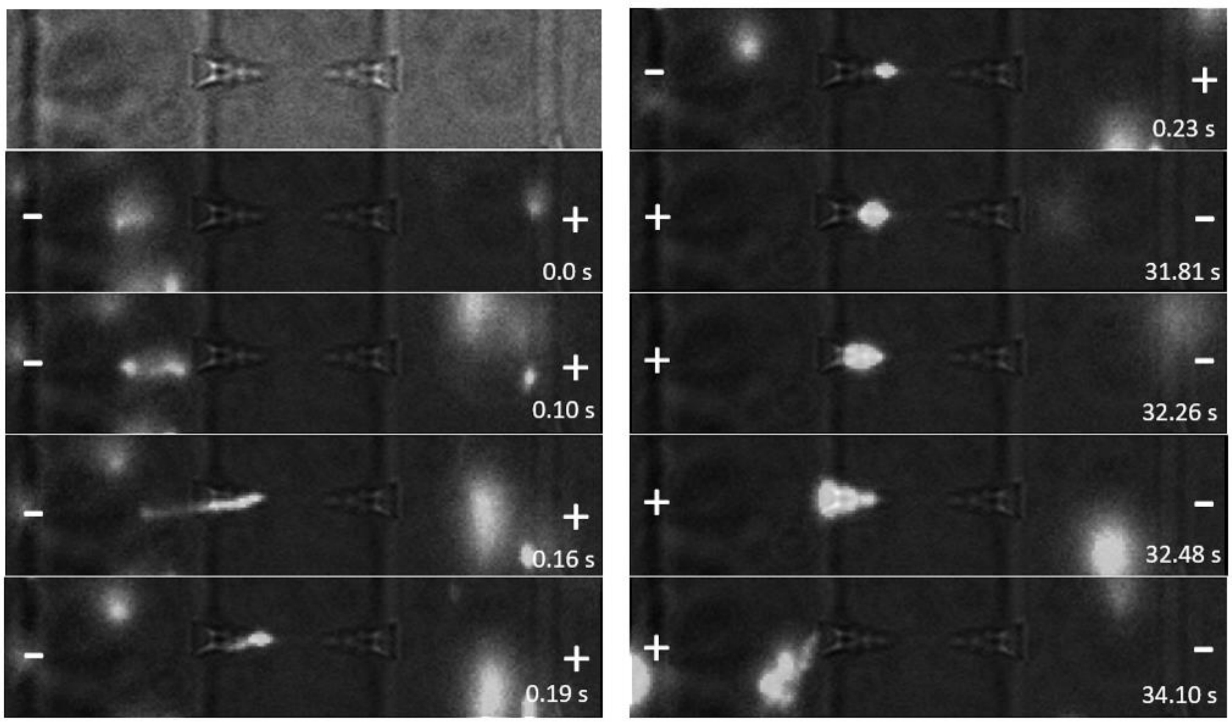

Figure 2.13. Images of fluorescence from single molecule tracking of labeled DNA with a semi-transparent image of the DNP device laid over the fluorescence image to demonstrate where DNA was located during each frame of the imaging. In these frames, DNA can be seen stretching and being driven toward the input funnel. At 0.16 s, it can be seen passing through the input funnel and coiling at the entrance to the first in-plane nanopore. The polarity of the voltage being applied was then reversed for ~32 s into and the DNA could be seen backing out and exiting the input funnel.

This experiment demonstrated that λ-DNA was not entering the in-plane nanopore due to coiling at the end of the input funnel. We speculated that RNA or DNA was not translocating through the in-plane nanopores as a result of coiling of RNA and DNA at the in-plane nanopore due to entropic effects arising from an inconsistent depth of the input nanochannel and the diameter of the in-plane nanopores being too narrow to allow for partially coiled RNA or DNA to translocate through. As a result of these observations, we made further changes in the device design.

2.4.2.3 Revised nanostructures to facilitate RNA/DNA entry through in-plane nanopores

In order to facilitate translocation efficiency, the next iteration of the *X-ToF* device possessed input and output nanopore sizes targeted to be 50 nm. It was hypothesized that an in-plane nanopore size of 50 nm would be large enough to facilitate the translocation of Cas9 RNA, which is 4,245 bp and has a radius of gyration of approximately 26 nm,[28] but also provide good SNR on the RPS signal. If smaller RNAs are being used that are below the detection limit of the 50 nm pore size, the in-plane nanopore can be tuned to the desired size by increasing the pressure during thermal fusion bonding; pores sizes can be reduced by 25%.[23] The input and output in-plane nanopores were also moved 2.5 µm away from the input funnel for two reasons: (i) The RNA would be confined in the 2.5 µm length of nanochannel before reaching the in-plane nanopore, which will slightly stretch the molecule allowing easier translocation through the in-plane pore. The contour length of Cas9 RNA is approximately 1.3 µm. Therefore, the RNA would be completely contained within the nanochannel even if stretched to its full contour length. This confinement allows the biomolecule to be stretched from its initial randomly coiled state to a more non-coiled state. Previously, Uba *et al.* demonstrated the stretching of dsDNA confined in thermoplastic nanochannels. It was shown that as the nanochannel width and depth were reduced, the confined dsDNA molecule would stretch more reaching nearly its full contour length in a 35 × 35 nm nanochannel.[10] (ii) Allow determination via fluorescence observations if the RNA would enter the nanochannel and travel to the in-plane nanopore or simply remain at the tip of the input funnel based on spatial considerations. This was challenging to accomplish in the previous design due to the in-plane nanopore being in the nanochannel near the tip of the input funnel. An SEM image of the next iteration of this *X-ToF* device is displayed in Figure 2.14.

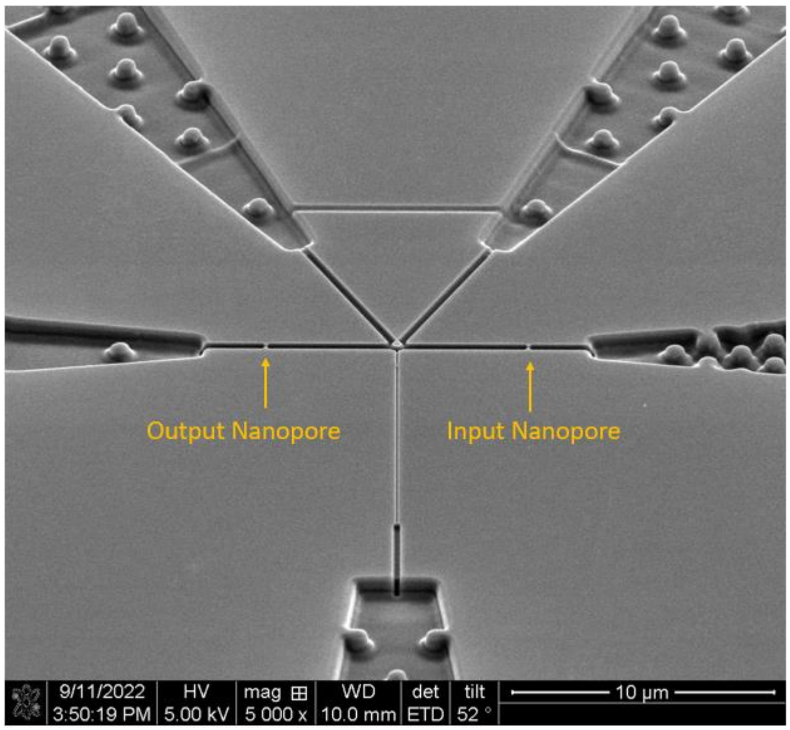

Figure 2.14. SEM image of *X-ToF* design containing approximately 50 nm input and output in-plane nanopores with 2.5 µm spacing between the microfunnel and in-plane nanopores (SEMs courtesy of Dr. Junseo Choi).

This iteration of the *X-ToF* device was evaluated using single-molecule fluorescence microscopy. PMMA devices were fabricated from the Si master using UV resin stamps and NIL to thermally imprint structures into the substrate followed by thermal fusion bonding a COC 8007 cover plate to enclose the fluidic network. Following fabrication and assembly, devices were treated with 30 mM ETA for 30 min to reduce the surface charge of the thermoplastic.[23] The fluid in the device was removed and replaced with NEBuffer 3. Finally, the SYTO82 labeled Cas9 RNA was added to the device and electrokinetically driven through the nanochannel network by applying a bias voltage. In the ETA treated devices, it was observed that the Cas9 RNA would travel to the end of the input funnel but would not enter into the nanochannel and travel to the first in-plane nanopore (see Figure 2.15).

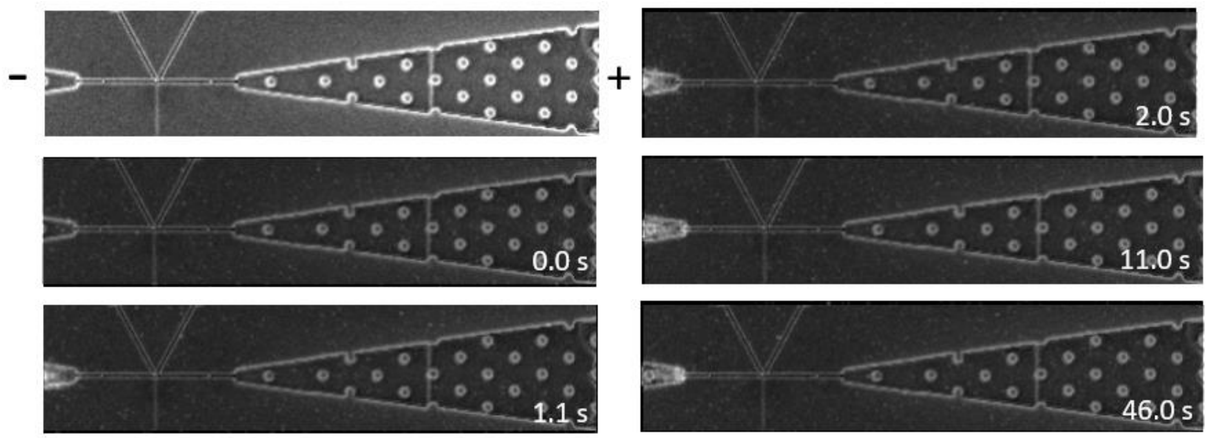

Figure 2.15. Fluorescent images of Cas9 RNA in an ETA-treated PMMA device. Top left image is of the device and has been made semi-transparent and overlaid over fluorescent frames from in the following 5 images. The fluorescently-labeled RNA can be seen entering the tip of the input funnel but not moving past the input funnel and into the nanochannel.

Following this experiment, another PMMA device was used, but it was not treated with ETA. When RNA was added to this device and driven through the input funnel and into the nanochannels, it could clearly be seen entering into the nanochannel and past the input funnel and up to the first input in-plane nanopore. Images from this experiment are displayed in Figure 2.16.

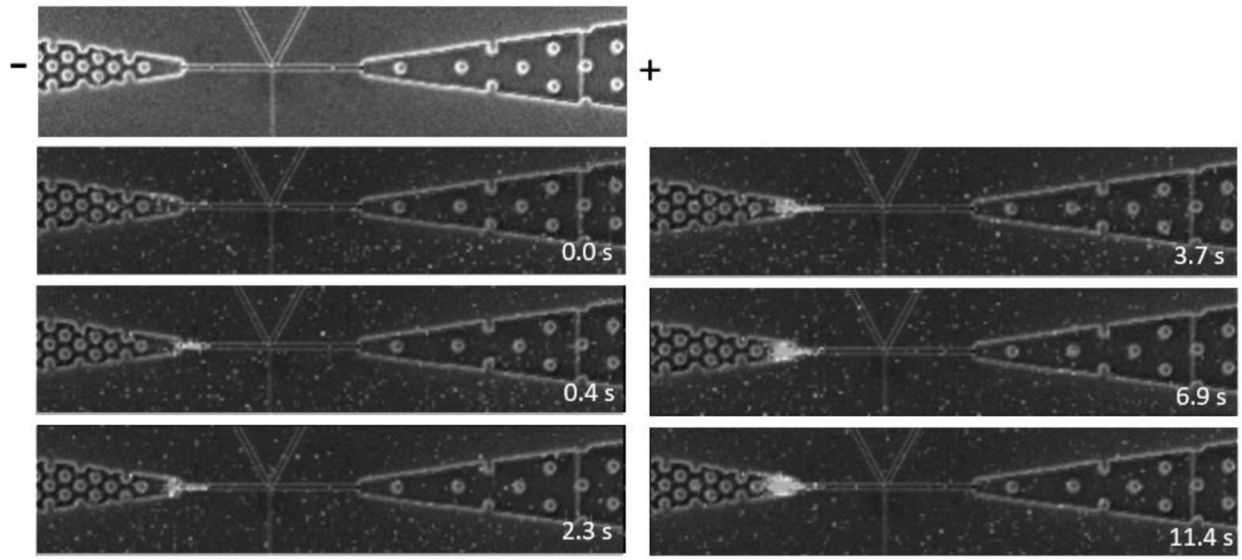

Figure 2.16. Fluorescence microscopy images of SYTO82-labeled Cas9 RNA in a PMMA device that was not treated with ETA following O_2 plasma activation and then, thermally fusion bonded to a COC 8007 cover plate. Top left image is of the device that has been made semi-transparent and overlaid over remaining images showing RNA translocation through the fluidic network. The fluorescently labeled RNA can be seen entering the tip of the input funnel and moving into the nanochannel and up to the first in-plane nanopore.

From these experiments, it was concluded that treating the PMMA device with ETA prevented RNA from entering into the nanochannels of the device. A comparison of the tip of the input funnel during each experiment clearly showed differences in the RNA's ability to enter the nanochannel (see Figure 2.17). This observed difference could be due to ETA deforming the nanostructures due to swelling or not properly defiling from the channel causing blockage.

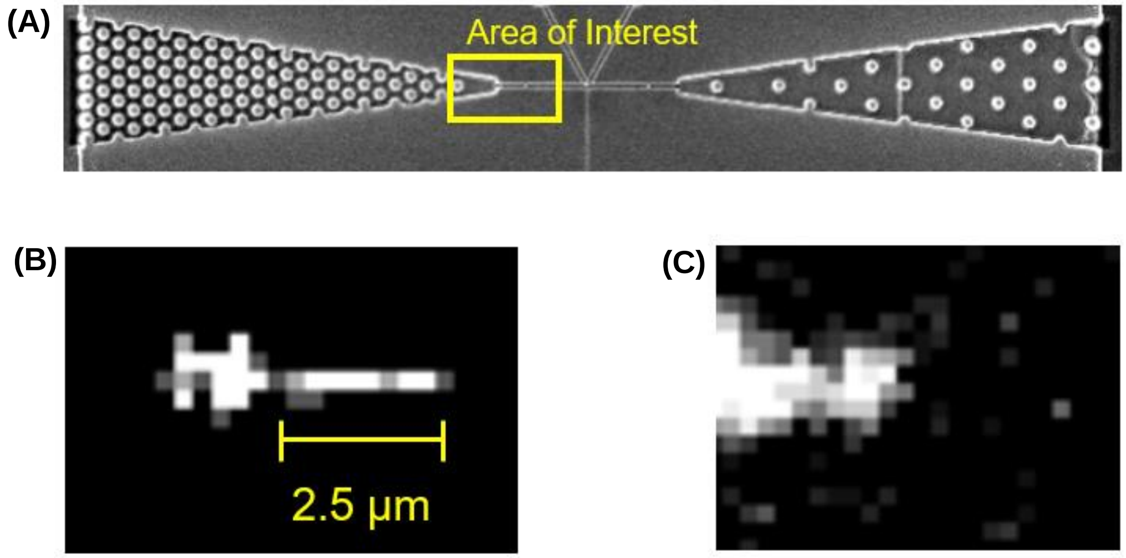

Figure 2.17. Comparison of RNA entry into the nanochannel adjacent to the input funnel in *X-ToF* PMMA device with and without ETA treatment. **(A**) Image of the device with a yellow box around the region of interest, ROI. **(B)** Image of the ROI for an untreated *X-ToF* PMMA device. Labeled RNA can be seen at the tip of the input funnel and traveling into the nanochannel and up to the input in-plane nanopore located 2.5 µm into the nanochannel from the input funnel tip. **(C)** ROI of an ETA treated PMMA device showing the translocation of Cas9 RNA. The labeled RNA can be seen in the tip of the input funnel but was not observed entering the nanochannel.

In the ETA non-treated PMMA device, the RNA could be driven to the first in-plane nanopore, but did not appear to translocate through it. This may have occurred due to co-ion exclusion effects[23] between the surface of the PMMA device and the RNA both of which are negatively charged, which can lead to a low success rate of RNA translocation. Further, the frame rate of the epifluorescence microscope employed for this experiment was operating at 16 fps. Therefore, between each image there is 62.5 ms. If the RNA would translocate through the nanochannels at a speed faster than this, it could be missed by the camera. Based on these results, RPS was employed for the next step due to the limited frame rate readout of the EMCCD.

RPS experiments in COP devices were employed for the next step. Initially PMMA devices were utilized for RPS experiments, but no RPS events were observed. In these experiments, 1 M KCl was added to the NEBuffer 3. This high salt concentration served to provide charge shielding around the negatively charged RNA and COP device surface.[29] Charge shielding was necessary due to the RNA and the walls of the PMMA device both carrying a negative charge causing co-ion exclusion, which resulted in the probability of successful translocation to be very low. Addition of KCl mitigated these charge-based issues by shielding the charge on the wall of the device and the RNA. In addition, this high salt concentration will reduce the thickness of the electrical double layer.

X-ToF devices were imprinted in COP and bonded with a COC 8007 cover plate. Following bonding, NEBuffer 3 containing 1 M KCl was filled into each device. Non-labeled Cas9 RNA with SYTO82 was added to the access microchannel connected to the input funnel. Electrodes were dipped into the microchannel reservoirs on either side of the nanochannel containing the input and output in-plane nanopores. Then, 2.5 V was applied to drive the Cas9 RNA through the in-plane nanopores. A patch clamp system was used to supply this voltage and record the resulting current transient events. Data was recorded using a 10 kHz low pass filter and a 400 Hz high pass filter. Resistive pulse events were observed during this experiment clearly demonstrating that the Cas9 RNA was successfully passing through the in-plane nanopores. Results are displayed in Figure 2.18 where a total of 106 RPS events were detected.

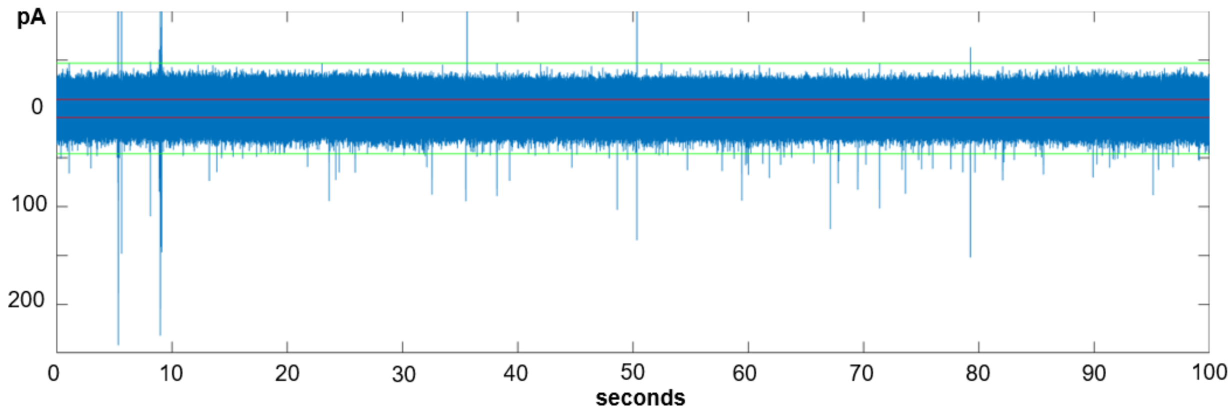

Figure 2.18. RPS generated from translocating Cas9 RNA through the input and output in-plane nanopores of the COP *X-ToF* device. In this experiment, NEBuffer 3 containing 1 M KCl was used with no modification of the surfaces with ETA. Experimental current trace in which drops in current can be observed as RNA passes through the in-plane nanopore.

The results in Figure 2.18 demonstrated that when 1 M KCl was added to the running buffer, the probability of RNA translocating through the in-plane nanopores increased precipitously due to charge shielding by the high salt concentration. To further investigate these results, similar experiments were conducted with KCl concentrations of 0, 250, 500, 750, and 1000 mM addition to NEBuffer 3. This provided insight into the ideal KCl concentration to use to maximize the probability of successful RNA translocations. To complete these experiments, devices were once again fabricated in COP and fusion bonded with a COC 8007 cover plate. Following bonding, the relevant buffer was filled into the entire device. A solution of the buffer containing 1 nM Cas9 RNA was then used to replace the buffer solution within the access microchannel. Electrodes were placed into the microchannel reservoirs on either side of the nanochannel containing the input and output in-plane nanopores and 2.5 V was applied. Experiments were conducted for 100 s each and the current traces were recorded. Following each experiment, the buffer in the device was replaced with the next investigated buffer and the experiment was repeated. This process was completed for each of the five KCl concentrations in different COP devices. It was found that RNA translocation increased in the presence of 1,000 mM KCl, but translocation events still occurred at 0 mM KCl, but at a lower frequency. Representative current traces from each KCl concentration are displayed in Figure 2.19.

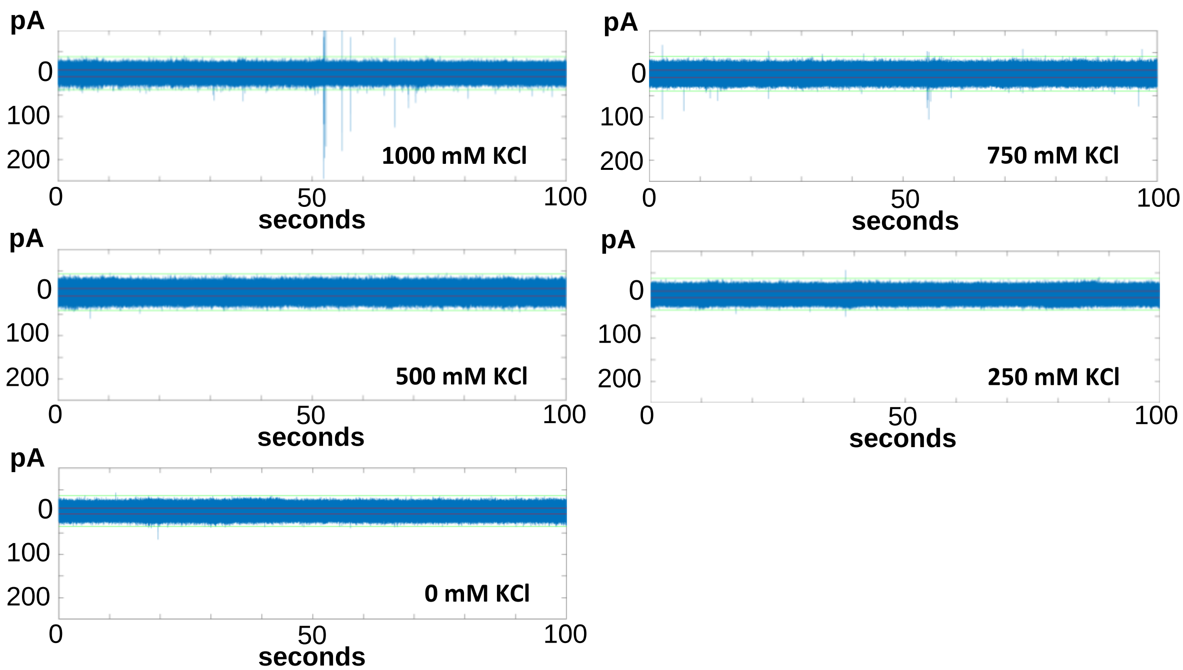

Figure 2.19. Representative current traces from each of the KCl concentrations tested. Event frequency and amplitude was observed to be the highest in the presence of 1,000 mM KCl. The red lines mark the RMS noise of the open pore current while the green lines define the threshold that a drop in current much reach to be defined as an event.

Events from these current traces were identified by applying a threshold that was 5x the RMS noise of the open pore current. These event numbers were used to calculate the average as well as investigate the inter-device variability through the standard deviation and RSD percentage of event number between devices. Variation between devices is most likely due to the stochastic nature of RPS, slight differences in surface activation, and individual device fabrication. A bar graph and table displaying the statistical results from these experiments are displayed in Figure 2.20 and Table 2.1.

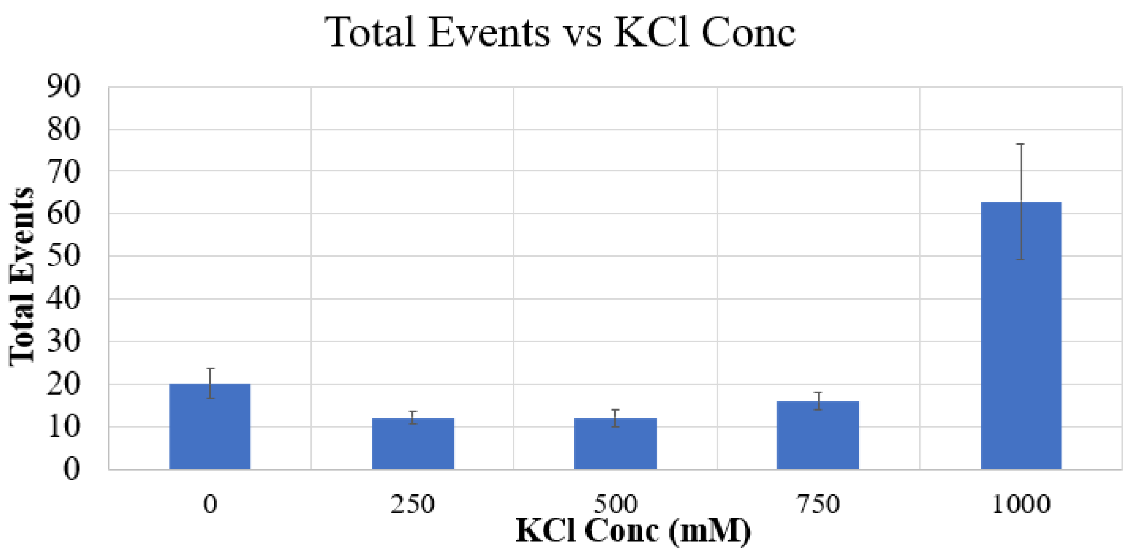

Figure 2.20. Graph displaying the total average events plotted against KCl concentration with standard deviation bars. This data was collected using COP/COC 8007 devices that were O_2 plasma activated, but no ETA-modification was done on the devices.

KCl Conc (mM)	Avg	Std. Dev.	RSD
1000	63	27	43 %
750	16	4	25 %
500	12	4	33 %
250	12	3	25 %
0	20	7	35 %

Table 2.1. Total average events, standard deviation, and RSD for RPS data secured using an *X-ToF* device consisting of COP/COC 8007 and different KCl concentrations added to NEBuffer 3.

To maximize the likelihood of RNA successfully translocating through the in-plane nanopore, one could employee 1,000 mM KCl in the running buffer of choice. With this in mind, even if lower amounts of KCl are used, RNA can still be successfully driven through the in-plane nanopore and into the INER of *X-ToF*. The lower probability of RNA entering through the in-plane nanopore at lower KCl concentrations would result in higher mass requirements for the *X-ToF* device. However, here it has been demonstrated that a solution of 1 nM Cas9 RNA with no KCl is sufficient to successfully inject RNA past the in-plane nanopore and into the INER. Approximately 3 µL of this solution is needed to fill the access microchannel of the device. This would put the mass requirements of this device at ~8 ng in its current iteration. This mass requirement could be further reduced by decreasing the size of the access microchannel or adding KCl into the buffer. The choice of what buffer used will be based upon what KCl concentration XRN1 can function under. It is possible that high salt concentrations will disrupt the electrostatic interactions[4,15] in the active site of this enzyme causing sequencing failure.

2.5 Conclusions and future directions

We presented results demonstrating the successful generation of an INER containing XRN1 in the *X-ToF* device and successful injection of RNA through in-plane nanopores of the *X-ToF* device. Areas requiring further investigation to achieve the ultimate goal of single-molecule sequencing include XRN1's ability to digest RNA in the presence of KCl, hardware and software developments, and the experimental design to demonstrate proof-of-concept sequencing.

XRN1's digestion efficiency of RNA in the presence of increasing KCl concentrations is vital. High salt concentrations have the potential to disrupt the electrostatic interactions essential for the enzyme's digestion process.[4,15] The digestion efficiency results will play a role in determining what buffer is used in the *X-ToF* device. The buffer used must balance the ability of XRN1 to efficiently digest RNA as well as efficiently inject RNA into the *X-ToF's* in-plane nanopores leading to the INER while still allowing for sufficient detection of the rNMPs. XRN1's digestion capabilities in the presence of KCl can be tested by free solution and solid phase reactions using IMERs. Solid phase and free solution reactions may yield varying results as XRN1's enzymatic efficiency has been shown to be improved when immobilized onto a solid support.[3]

Investigation of XRN1's ability to capture RNA as it passes through the INER can be probed using RPS. After device fabrication and XRN1 immobilization within the INER, RNA can be translocated through the input and output in-plane nanopores. Due to an in-plane nanopore being present before and after the INER, if only a single resistive pulse is observed, it can be determined that the RNA molecule was captured by the immobilized XRN1. On the other hand, if two resistive pulse events occur it can be surmised that the XRN1 was unsuccessful in capturing the RNA molecule (see Figure 2.21).

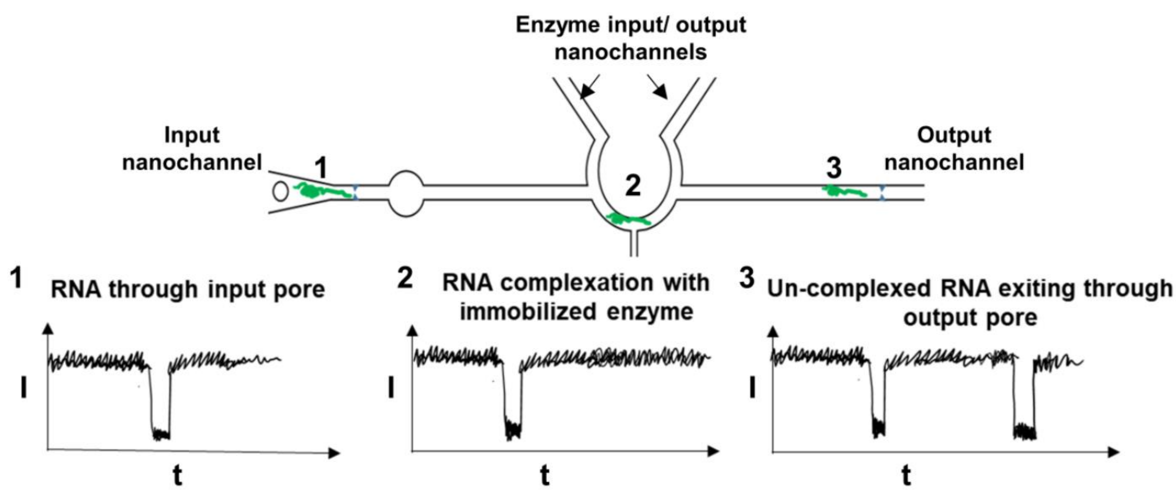

Figure 2.21. Diagram displaying input and output in-plane nanopores for the detection of entering and exiting RNA. (Reproduced with permission from reference [2]).

This simple procedure will allow us to rapidly screen different conditions for optimizing the capture rate, such as optimizing the salt concentrations, bias voltage, and pore size with respect to the input molecule (RNA length, and ss- versus dsDNA as well as length).

A proof-of-concept sequencing experiment using the *X-ToF* device should be performed next. After initial device fabrication and enzyme immobilization, the sequencing reaction will be performed in two steps using a two channel TIA empowered by automated voltage control. First, RNA will be driven into the INER by applying a voltage only across the input/output nanochannels. When a single peak is recorded, indicating the RNA has been captured by the enzyme, the applied voltage will be switched off. A voltage will then be applied across the ToF nanochannel to drive Mg^{2+} into the INER to initiate digestion and also drive the released rNMPs through the two in-plane nanopores to identify their ToF for base calling purposes. This second step, which involves the delivery of the cofactor to the enzyme and movement of the rNMPs through the in-plane nanopores for ToF detection, may be achieved in two ways. The first option is to rely upon the electrophoretic mobilities of the Mg^{2+} and rNMPs. The second option is to rely upon electroosmotic flow to transport both the cofactor and rNMPs. Figure 2.22 demonstrates how delivery of the cofactor and rNMP detection in each method would be achieved.

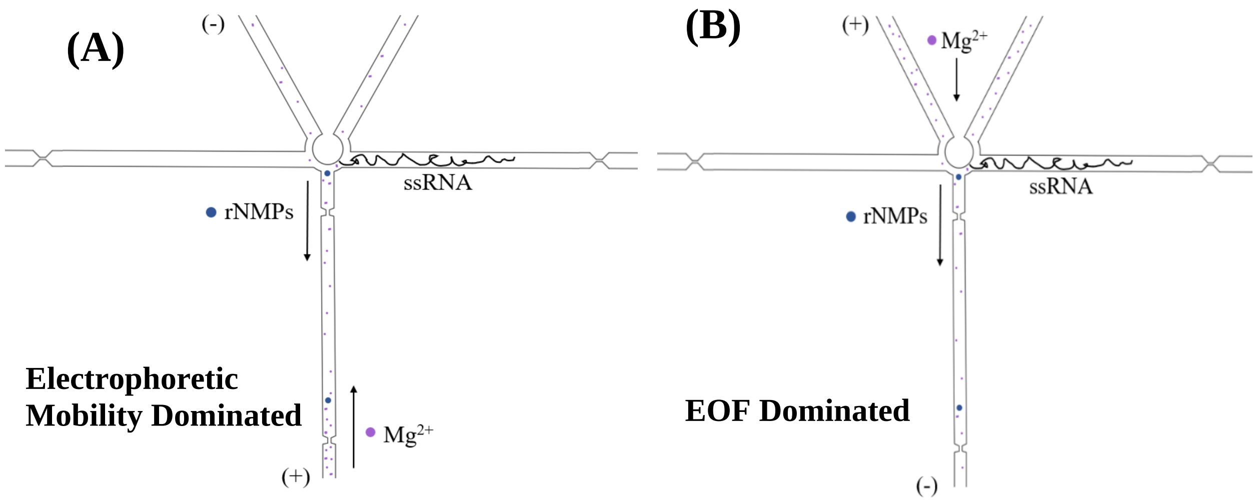

Figure 2.22. Diagram of the *X-ToF* device showing electrophoretic mobility and EOF dominated flow. **(A)** Electrophoretic mobility dominates causing rNMPs and Mg^{2+} to travel in opposite directions. **(B)** EOF dominates causing rNMPs and Mg^{2+} to travel in the same direction.

As stated previously, the voltage control for the proof-of-concept experiment will be automated. This is vital to ensure only a single RNA molecule is injected into the device for sequencing. This automated control will work by first applying a designated voltage across the input/output channels by channel one of the TIA while simultaneously monitoring the current. Electrical connections to the *X-ToF* device will be made using zebra connectors attached to thin film Ag/AgCl electrodes, which will be present on the cover plate of the device. When a single resistive pulse event is generated that is above a designated threshold, the applied voltage is switched off. A second designated voltage will then be applied across channel two of the TIA. The electrodes for channel two will be connected across the nanoscale enzymatic reactor and the ToF nanochannel. The automated control of the two channel TIA will allow for high temporal control of applied voltages to ensure only a single RNA molecule is injected into the *X-ToF* at a time.

Following initial development of the *X-ToF* sensor and completion of proof-of-concept, multiple *X-ToF* sensors will be fabricated on a thermoplastic substrate. The TIA's for each individual *X-ToF* sensor will be miniaturized to fit directly on top of each sensor. By massively increasing the sensors employed at one time, the sequencing throughput for this platform will also increase greatly which is needed to be a competitive platform in the current sequencing market.

2.6 References

(1) Reiner, J. E.; Balijepalli, A.; Robertson, J. W.; Drown, B. S.; Burden, D. L.; Kasianowicz, J. J. The effects of diffusion on an exonuclease/nanopore-based DNA sequencing engine. *The Journal of chemical physics* **2012**, *137* (21), 214903.

(2) Athapattu, U. S. Development of Nanoscale Technologies for Applications in Single-Molecule Sequencing. University of Kansas, 2021.

(3) Athapattu, U. S.; Amarasekara, C. A.; Immel, J. R.; Bloom, S.; Barany, F.; Nagel, A. C.; Soper, S. A. Solid-phase XRN1 reactions for RNA cleavage: application in single-molecule sequencing. *Nucleic Acids Res* **2021**, *49* (7), e41. DOI: 10.1093/nar/gkab001 License at https://creativecommons.org/licenses/by/4.0/legalcode.

(4) Jinek, M.; Coyle, Scott M.; Doudna, Jennifer A. Coupled 5′ Nucleotide Recognition and Processivity in Xrn1-Mediated mRNA Decay. *Molecular Cell* **2011**, *41* (5), 600-608. DOI: https://doi.org/10.1016/j.molcel.2011.02.004.

(5) O'Neil, C.; Amarasekara, C. A.; Weerakoon-Ratnayake, K. M.; Gross, B.; Jia, Z.; Singh, V.; Park, S.; Soper, S. A. Electrokinetic transport properties of deoxynucleotide monophosphates (dNMPs) through thermoplastic nanochannels. *Anal Chim Acta* **2018**, *1027*, 67-75.

(6) Choi, J.; Jia, Z.; Riahipour, R.; McKinney, C. J.; Amarasekara, C. A.; Weerakoon-Ratnayake, K. M.; Soper, S. A.; Park, S. Label-Free Identification of Single Mononucleotides by Nanoscale Electrophoresis. *Small* **2021**, *17* (42), 2102567. DOI: https://doi.org/10.1002/smll.202102567.

(7) Amarasekara, C. A.; Rathnayaka, C.; Athapattu, U. S.; Zhang, L.; Choi, J.; Park, S.; Nagel, A. C.; Soper, S. A. Electrokinetic identification of ribonucleotide monophosphates (rNMPs) using thermoplastic nanochannels. *Journal of Chromatography A* **2021**, *1638*, 461892.

(8) Kim, Y.; Song, J.; Lee, Y.; Cho, S.; Kim, S.; Lee, S.-R.; Park, S.; Shin, Y.; Jeon, N. L. High-throughput injection molded microfluidic device for single-cell analysis of spatiotemporal dynamics. *Lab on a Chip* **2021**, *21* (16), 3150-3158.

(9) Rathnayaka, C.; Amarasekara, C. A.; Akabirov, K.; Murphy, M. C.; Park, S.; Witek, M. A.; Soper, S. A. Nanofluidic devices for the separation of biomolecules. *Journal of Chromatography A* **2022**, *1683*, 463539. DOI: https://doi.org/10.1016/j.chroma.2022.463539.

(10) Uba, F. I.; Hu, B.; Weerakoon-Ratnayake, K.; Oliver-Calixte, N.; Soper, S. A. High process yield rates of thermoplastic nanofluidic devices using a hybrid thermal assembly technique. *Lab on a Chip* **2015**, *15* (4), 1038-1049.

(11) Jia, Z.; Choi, J.; Park, S. Selection of UV-resins for nanostructured molds for thermal-NIL. *Nanotechnology* **2018**, *29* (36), 365302.

(12) Uba, F. I.; Pullagurla, S. R.; Sirasunthorn, N.; Wu, J.; Park, S.; Chantiwas, R.; Cho, Y.-K.; Shin, H.; Soper, S. A. Surface charge, electroosmotic flow and DNA extension in chemically modified thermoplastic nanoslits and nanochannels. *Analyst* **2015**, *140* (1), 113-126.

(13) Oliver-Calixte, N. J.; Uba, F. I.; Battle, K. N.; Weerakoon-Ratnayake, K. M.; Soper, S. A. Immobilization of Lambda Exonuclease onto Polymer Micropillar Arrays for the Solid-Phase Digestion of dsDNAs. *Analytical Chemistry* **2014**, *86* (9), 4447-4454. DOI: 10.1021/ac5002965.

(14) Wang, Z.; King, T. L.; Branagan, S. P.; Bohn, P. W. Enzymatic activity of surface-immobilized horseradish peroxidase confined to micrometer- to nanometer-scale structures in nanocapillary array membranes. *Analyst* **2009**, *134*, 851-859.

(15) Chang, J. H.; Xiang, S.; Xiang, K.; Manley, J. L.; Tong, L. Structural and biochemical studies of the 5'-->3' exoribonuclease Xrn1. *Nat Struct Mol Biol* **2011**, *18* (3), 270-276. DOI: 10.1038/nsmb.1984.

(16) Amarasekara, C. A.; Athapattu, U. S.; Rathnayaka, C.; Choi, J.; Park, S.; Soper, S. A. Open-tubular nanoelectrochromatography (OT-NEC): gel-free separation of single stranded DNAs (ssDNAs) in thermoplastic nanochannels. *Electrophoresis* **2020**, *41* (18-19), 1627-1640. DOI: 10.1002/elps.202000109.

(17) O'Neil, C. E.; Taylor, S.; Ratnayake, K.; Pullagurla, S.; Singh, V.; Soper, S. A. Characterization of activated cyclic olefin copolymer: effects of ethylene/norbornene content on the physiochemical properties. *Analyst* **2016**, *141* (24), 6521-6532. DOI: 10.1039/c6an01448h.

(18) Vaidyanathan, S. Single Molecule Particle Analysis using Nanotechnology. University of Kansas, 2021.

(19) Andrew, T. L.; Swager, T. M. Reduced photobleaching of conjugated polymer films through small molecule additives. *Macromolecules* **2008**, *41* (22), 8306-8308.

(20) Chen, H.; Gai, H.; Yeung, E. S. Inhibition of photobleaching and blue shift in quantum dots. *Chemical communications* **2009**, (13), 1676-1678.

(21) Wlodkowic, D.; Skommer, J.; Darzynkiewicz, Z. SYTO probes in the cytometry of tumor cell death. *Cytometry A* **2008**, *73* (6), 496-507. DOI: 10.1002/cyto.a.20535.

(22) Doty, P.; Boedtker, H.; Fresco, J.; Haselkorn, R.; Litt, M. Secondary structure in ribonucleic acids. *Proceedings of the National Academy of Sciences* **1959**, *45* (4), 482-499.

(23) Athapattu, U. S.; Rathnayaka, C.; Vaidyanathan, S.; Gamage, S. S.; Choi, J.; Riahipour, R.; Manoharan, A.; Hall, A. R.; Park, S.; Soper, S. A. Tailoring thermoplastic in-plane nanopore size by thermal fusion bonding for the analysis of single molecules. *ACS sensors* **2021**, *6* (8), 3133-3143.

(24) Bayley, H.; Martin, C. R. Resistive-Pulse SensingFrom Microbes to Molecules. *Chemical Reviews* **2000**, *100* (7), 2575-2594. DOI: 10.1021/cr980099g.

(25) Wu, J.; Chantiwas, R.; Amirsadeghi, A.; Soper, S. A.; Park, S. Complete plastic nanofluidic devices for DNA analysis via direct imprinting with polymer stamps. *Lab Chip* **2011**, *11* (17), 2984-2989. DOI: 10.1039/c1lc20294d.

(26) Vaidyanathan, S.; Weerakoon-Ratnayake, K. M.; Uba, F. I.; Hu, B.; Kaufman, D.; Choi, J.; Park, S.; Soper, S. A. Thermoplastic nanofluidic devices for identifying Abasic sites in single DNA molecules. *Lab on a Chip* **2021**, *21* (8), 1579-1589.

(27) Uba, F. I.; Hu, B.; Weerakoon-Ratnayake, K.; Oliver-Calixte, N.; Soper, S. A. High process yield rates of thermoplastic nanofluidic devices using a hybrid thermal assembly technique. *Lab Chip* **2015**, *15* (4), 1038-1049. DOI: 10.1039/c4lc01254b From NLM.

(28) Weerakoon-Ratnayake, K. M.; O'Neil, C. E.; Uba, F. I.; Soper, S. A. Thermoplastic nanofluidic devices for biomedical applications. *Lab on a Chip* **2017**, *17* (3), 362-381.

(29) Peng, R.; Tang, X. S.; Li, D. Detection of Individual Molecules and Ions by Carbon Nanotube-Based Differential Resistive Pulse Sensor. *Small* **2018**, *14* (15), 1800013. DOI: https://doi.org/10.1002/smll.201800013.