

Innovations in Droplet Microfluidics for Enzyme Evolution

by

Daniel A. Holland-Moritz

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemistry)
in the University of Michigan
2020

Doctoral Committee:

Professor Robert T. Kennedy, Chair
Professor Ryan C. Bailey
Professor Katsuo Kurabayashi
Dr. Jeffrey Moore, Merck & Co.
Professor Alison Narayan

Daniel A. Holland-Moritz

dholla@umich.edu

ORCID iD: [0000-0003-4634-6122](https://orcid.org/0000-0003-4634-6122)

© Daniel A. Holland-Moritz 2020

Dedication

To my Father, Mother, Sister and Brother, who have been here with me every step of the way. You are my biggest role models, my closest friends; without your infinite well of support and love, I would not be here today. You taught me how to be a really useful engine.

Acknowledgements

Thank you to my research advisor, Dr. Robert Kennedy. Without your support, mentorship, and advice none of this would have been possible. Thank you for guiding me all these years, even from across the country as I worked for two of these years from “Kennedy Lab East,” at Merck’s Rahway NJ site. Thank you to all of my committee members for their support and advice each step along this journey, and a very special thank you to Dr. Jeffrey Moore, who made the Michigan-Merck collaboration possible and served as a second PhD advisor and mentor during my time in New Jersey. Thank you Dr. Shuwen Sun, for being a fantastic role model and office mate throughout my time at Merck, and yet another scientific mentor and advisor while we worked together in New Jersey.

Thank you to my MANY collaborators at Merck and Michigan. Without the support of Mike Wismer, Dr. Kaori Hiraogi, and Dr. Wai-Ling Cheung-Lee at Merck, none of the critical engineering, molecular biology, or analytical chemistry of this project could have happened. Similarly, it would be remiss not to thank my lab mates and collaborators at Michigan, especially Emory Payne. You all taught me so much, and I am deeply grateful for your patience and instruction over the past two years. You gave meaning to the term “collaboration” and you taught me to trust and learn from the expertise of the scientists around me. I am a better researcher for it.

Finally, thank you to all of my friends, for their unwavering support through all the ups and downs of the past years. Stephanie Chun, William Walker, and Kevon Stanford, you have always been there for me, and the memories we made together will last a lifetime. Jeremy

Kallick and Isabel Colon-Bernal - thank you for your unerring loyalty and for always telling it how it is. Thank you Alex Sun, for your patient edits and encouragement during the thesis writing process and my difficult final months. You guys weathered this with me, and I'm proud to count you among my friends.

Table of Contents

Dedication	ii
Acknowledgements.....	iii
List of Tables	vii
List of Figures	viii
List of Abbreviations	xi
Abstract	xiv
Chapter 1: Introduction.....	1
1.1 High-Throughput Screening Technologies.....	1
1.2 The Interplay of Analytical Throughput and Sample Introduction	1
1.3 Microfluidics for High Throughput Sample Generation and Manipulation	3
1.4 Analytical Challenges for Microdroplet Interrogation	7
1.5 Direct MS Assessment in Droplet Microfluidics.....	9
1.6 High Throughput Screening for Directed Evolution	12
1.7 Dissertation Overview	16
Chapter 2: Mass Activated Droplet Sorting (MADS) Enables High Throughput Screening of Enzymatic Reactions at Nanoliter Scale	19
2.1 Introduction.....	19
2.2 Experimental - Mass Activated Droplet Sorting.....	22
2.3 Mass Activated Droplet Sorting - System Overview.....	30
2.4 Experimental - MADS in Complex Matrices and Reactions.....	36
2.5 Reading Enzyme Activity in ivTT Reaction Mix	41
2.6 Proof of Concept: Enzyme Screening.....	48
2.7 Conclusions.....	52
2.8 Supplemental Information – Transaminase Sequence	53

Chapter 3: Chapter 3 Integrated Nanoliter Sample Processing and Mass Spectrometry for the Expression and Selection of a Transaminase Library	56
3.1 Introduction.....	56
3.2 Experimental Methods	60
3.3 Results: Implementation of an in-Droplet Protein Expression Workflow	74
3.4 Initial Efforts to Screen a WT Transaminase Library.....	90
3.5 Rational for Screening a Degenerate Transaminase	94
3.6 Challenges and Prospects for the MADS system	104
3.7 Conclusions/Impact.....	108
Chapter 4: Insights and Future Directions	110
4.1 Challenges and Opportunities for MADS and Droplet MS	110
4.2 Alternate Applications for the MADS Platform	128
4.3 Improving the throughput and sensitivity of droplet MS.....	130
4.4 Future Directions for Droplet Infusion	131
4.5 Conclusions.....	135
References	136

List of Tables

Table 2-1: Sorting results for MADS proof of concept, with user-defined thresholding	34
Table 2-2: Sorting results for MADS proof of concept, with activity-based thresholding.....	51
Table 3-1: PCR Protocols	62
Table 3-2: Targeted amino acids of the Tier 1 SSM library.	63
Table 3-3: Primers used for the PCR protocols described in this work.....	64
Table 3-4: Wildtype Tier 1 library Sequencing	84
Table 3-5: S17P Tier 1 library sequencing	98
Table 3-6: Single droplet sequencing of S17P library amplified in droplets.....	99
Table 3-7: Droplet sequencing of S17P library blank reveals source of WT DNA	100

List of Figures

Figure 1-1: Chemical structure of PFPE-PEG-PFPE surfactants	5
Figure 1-2: Unit operations in droplet microfluidics	6
Figure 1-3: Schematic of droplet mass spectrometry	11
Figure 1-4: The molecular biology workflow of directed evolution	15
Figure 2-1: SU-8 patterning for soft lithography.....	23
Figure 2-2: Storage, incubation, and reinjection strategy for nanoliter scale samples.	25
Figure 2-3: Typical trace from direct infusion of oil-segmented droplets.....	26
Figure 2-4: Full schematic and detail of the MADS device	28
Figure 2-5: Operating principle for signal alignment between on-chip imaging and MS	32
Figure 2-6: Decision tree detailing simple threshold-based droplet sorting.....	33
Figure 2-7: Imaged results from a representative droplet sorting run	35
Figure 2-8: Schematic and micrograph of a droplet-pairing reagent addition device	39
Figure 2-9:One-pot ivTT expression of transaminase analyzed in bulk and droplets	42
Figure 2-10: Concentration dependent signal intensity for droplets of ivTT with amine 1 ...	44
Figure 2-11: Demonstration of accurate assessment of mixed droplet concentrations	45
Figure 2-12: Demonstration of complete in-droplet expression and activity of ATA117.....	46
Figure 2-13: Decision tree detailing activity-based thresholding in droplet sorting	47
Figure 2-14: Representative signal trace for activity based thresholding in MADS	50
Figure 2-15: Droplet expression and sorting results.....	52
Figure 3-1: Mechanism for the ATA117-mediated transformation of an amine.....	59

Figure 3-2: Structure of the ATA117 dimer	63
Figure 3-3: Schematic of the fluorescent activated droplet sorting device.....	67
Figure 3-4: Experimental data and strategies for PCR amplification in droplets	76
Figure 3-5: Strategies for detection of PCR amplification in droplets.	78
Figure 3-6:Flourescent images of digital PCR dilutions	80
Figure 3-7: Bulk ivTT reactions reveal enzyme activity as a drop in amine signal	81
Figure 3-8: Experimental workflow and data for single gene droplet expression	83
Figure 3-9: Fluorescence activated droplet sorting enriches DNA containing droplets.....	87
Figure 3-10: Single droplet sequencing data for amplified DNA in droplets.....	89
Figure 3-11: Z factor of the separation between WT and inactive droplet samples.	92
Figure 3-12: Initial Screening results from the assessment of a transaminase library.....	93
Figure 3-13: Assessment of the deleterious mutation S17P in bulk and in droplets	95
Figure 3-14: Improved conditions for screening differences between WT and S17P.....	97
Figure 3-15: Screening results from the assessment the S17P transaminase library	101
Figure 3-16: Rescreen data from the re-expression of hits from the S17P library	103
Figure 3-17: Time course read of wildtype ATA117 in droplets	107
Figure 4-1: Variability in droplet signal with increasing concentration	111
Figure 4-2: Linear increase in droplet signal variation with mean signal intensity.....	113
Figure 4-3: Linear increase in droplet signal variation with mean signal intensity.....	114
Figure 4-4: Instrumental signal variation predicts the spread of droplet data	115
Figure 4-5: Comparison of analyte retention within droplets.....	117
Figure 4-6: The benefit of a quenched enzyme reaction.....	120
Figure 4-7: K-channel reagent addition enables scalable addition to all droplets	122

Figure 4-8: Drift in K-channel operation at nanoliter scales	124
Figure 4-9: Sampling frequency on an Agilent 6120B monitoring 4 ions	126
Figure 4-10: Magnetic sorting of 65 pl droplets	127
Figure 4-11: Magnetic biasing to recapture beads in split 100 pl droplets.....	129
Figure 4-12: Peristaltic sampling of microfluidic droplets from a well plate.....	132
Figure 4-13: Vacuum driven infusion of samples onto a mass spectrometer	135

List of Abbreviations

AADS	Absorbance Activated Droplet Sorting
AC	Alternating Current
ADC	Analog to Digital
APD	Avalanche Photodiode
ATA-117	Transaminase ATA-117
CCD	Charge Coupled Device
CCFL	Cold Cathode Fluorescent Lamp
CE	Capillary Electrophoresis
DART-MS	Direct Analysis in Real Time Mass Spectrometry
DCC	DNA Clean and Concentrator
ddPCR	Digital Droplet Polymerase Chain Reaction
DE	Directed Evolution
DEP	Dielectrophoresis
DNA	Deoxyribonucleic Acid
dNTP	Deoxyribonucleotide Triphosphate
dsDNA	Double Stranded DNA
EGFR	Epidermal Growth Receptor Factor
ESI-MS	Electrospray Ionization Mass Spectrometry
FACS	Fluorescence Activated Cell Sorting

FAD	Flavin Adenine Dinucleotide
FADS	Fluorescence Activated Cell Sorting
FITC	Fluorescein Isothiocyanate
FRET	Förster Resonance Energy Transfer
GOI	Gene of Interest
HPLC	High Performance Liquid Chromatography
HTS	High Throughput Screening
ID	Inner Diameter
IR	Infrared
ivTT	<i>In Vitro</i> Transcription and Translation
LC	Liquid Chromatography
LED	Light Emitting Diode
MADS	Mass Activated Droplet Sorting
MALDI	Matrix-Assisted Laser Desorption Ionization
MISER	Multipl Injections in a Single Experimental Run
MS	Mass Spectrometry
MWP	Multiwell Plate
NEB	New England Biolabs
nESI-MS	Nano-Electrospray Ionization Mass Spectrometry
NMR	Nuclear Magnetic Resonance
OD	Outer Diameter
PC	Personal Computer
PDMS	Polydimethyl Siloxane

PFO	Perfluorooctanol
PFPE-PEG-PFPE	Perfluoropolyether-Polyethylene Glycol (copolymer)
PLP	Pyridoxal Phosphate
PMP	Pyridoxamine Phosphate
PMT	Photomultiplier Tube
PVC	Polyvinyl Chloride
RSD	Relative Standard Deviation
RTOS	Real Time Operating System
SDHA	Succinate Dehydrogenase
SOE	Splicing by Overhang Extension
SPE	Solid Phase Extraction
SSM	Single Site Mutation
TRIS	Tris(Hydroxymethyl)aminomthane
USB	Universal Serial Bus
UV	Ultraviolet
UV-Vis	Ultraviolet/Visible light
WT	Wildtype

Abstract

High throughput screening (HTS) is a critical technology in discovery and process chemistry. Typically, HTS has relied on robotic handling of well plates of samples to rapidly perform and assess reactions, but novel technologies that take advantage of microfluidic sample processing hold potential as alternatives to typical well plate based assays. Droplet based microfluidics has opened the door to the execution of millions of reactions on the nanoliter scale but has lagged in application to high throughput screening due to limitations in the options for analysis of nanoliter samples, which has largely been performed using optical detection. The work described here addresses this analytical challenge, detailing the design and development of an integrated microfluidic system for mass activated droplet sorting (MADS). The MADS system uses electrospray ionization mass spectrometry (ESI-MS) to analyze the contents of nanoliter scale reactions, and microfluidics to both perform these reactions at nanoliter scale and sort and recover the droplets in which they were performed. MADS expands the toolbox for droplet detection and recovery, broadening the applicability of droplet sorting to protein engineering, drug discovery, and diagnostic workflows.

To demonstrate the utility of the MADS system, it is applied to the process of Directed Evolution. Directed evolution of biocatalytic enzymes enables the design of enzymes that perform new catalytic functions, but requires screening thousands of modified enzymes to identify mutations favorable to new catalytic functions. In the pursuit of applying MADS sorting to biocatalytic reactions in droplets, this work details the development of the microfluidic

devices and molecular biology workflow for the expression of a library of enzyme variants *in vitro*, each in its own 30 nL droplet. Digital droplet polymerase chain reaction (ddPCR) is applied to amplify individual DNA copies in microdroplets, overcoming statistical limitations in droplet loading using a high throughput fluorescence activated droplet sorting (FADS) system to enrich for individual droplets containing amplified DNA. PCR amplified droplet samples are paired with droplets of *in vitro* transcription and translation (ivTT) matrix to create a final library of droplets expressing enzymes *in vitro*. These samples are screened using MADS to demonstrate the capability of the system to enrich the best performing enzymes from a pool of library variants.

As proof of concept, a deleterious mutation is introduced into an existing transaminase enzyme and this enzyme is then used as the backbone for a 2000 variant enzyme library. Proof of concept is demonstrated by enrichment of the original wildtype enzyme from this library, and recovery of the original enzyme activity. The work completed demonstrates the capability of the MADS system to screen tens of thousands of enzyme variants, using just a few milliliters of reagent and at significantly improved throughputs relative to traditional screening.

Chapter 1: Introduction

1.1 High-Throughput Screening Technologies

High-throughput screening (HTS) is a core technology in the pharmaceutical industry and a critical component of drug discovery and development pipelines. The label of “High-Throughput” is generally used to describe techniques which allow the interrogation of 10^3 - 10^6 samples per day¹, but the terminology tends to vary between applications depending on the complexity of the assay and the analytical readout used to assess the result of the experiment. Typically, high throughput screening is carried out in multi-well plates (MWP), which were introduced in the 1950s² and rose in popularity in the 1970s³ as tools to improve the efficiency of laboratory experimentation. These plates have become standardized in size, allowing experiments to be carried out in 96, 384, and 1536-well formats by robotic systems as well as by hand³. In these formats, reactions are carried out in volumes ranging from 5 to ~300 microliters, with each reaction taking place in a single well.

1.2 The Interplay of Analytical Throughput and Sample Introduction

In a typical HTS assay, the most significant bottlenecks determining assay throughput are analytical throughput and sample processing throughput. Information rich analytical methodologies, such as liquid chromatography (LC), mass spectrometry (MS), nuclear magnetic resonance (NMR) spectroscopy, and infrared (IR) spectroscopy, have analytical speeds ranging from 30s seconds to 30 min per sample. As such, these technologies are generally incapable of screening more than a few hundred to a few thousand samples per day, and do not lend

themselves to rapid screening. As such, most high throughput screens have been carried out using photometric spectroscopy, in particular absorbance and fluorescence. Absorbance and fluorescence spectroscopy are comparably information-light relative to their counterparts in the analytical toolbox, but the high sensitivity and rapid response times of photodetectors give the technique a distinct edge when applied to high throughput analytical assessment. For example, photomultiplier tubes (PMTs) possess pulse pair resolutions on the order of 10^{-8} s,⁴ which readily accommodates analytical throughput on the order of one hundred thousand (10^5) of samples per second in flow based formats like flow cytometry.⁵ The introduction of charged coupled devices (CCD) for plate imaging has enabled high density plates to be assessed in a single image, allowing luminescence, fluorescence and scintillation to be assessed across hundreds of wells with a single imaging device.⁶ Given a rapid and sensitive mode of detection, the throughput limit is governed by the rate at which samples may be introduced to the analytical detector.

Scanning a MWP with the appropriate optical configuration allows modern CCD plate readers to analyze a 384 well plate in under a minute, allowing for throughputs ranging from 1-10 samples per second. At these speeds, tens of thousands of samples may be analyzed per day, but this still lags far behind the theoretical limit of photometric assessment. The limitation of MWP technology for screening throughput stems from the physical size, separation, and positioning of samples in space, which determines the speed with which robotic and imaging systems can access these samples. However, faster sample introduction to the detector is possible in flow based setups, where serial samples are passed by the detector. A particularly good example of this is fluorescent activated cell sorting (FACS), which was developed in the 1970s and combines the liquid handling technology behind ink jet printing⁷ with the rapid analytical assessment of fluorescent detection. Using a piezoelectric resonator to vibrate a capillary of

continuously flowing cell suspension, sample is broken into droplets containing single cells, and individual droplets are sorted based on photometric readouts.⁸ FACS systems are capable of analyzing and sorting individual cells at throughputs of 10^5 to 10^6 samples per second. This innovation demonstrates that given a high analytical acquisition rate, the limit to throughput is determined by sample handling and introduction speeds. In FACS, continuous flow of samples to a detector allows sample interrogation to proceed up to the throughput limit of detection.

1.3 Microfluidics for High Throughput Sample Generation and Manipulation

The past four decades have seen the rise of automated liquid handling robotics capable of addressing each well in a MWP, dispensing volumes ranging from the nanoliter (Caliper Life Science SciClone pin tool) to the microliter range. While this strategy reduces the bottleneck of sample generation and allows thousands of well plate reactions to be created and run in concert, it simply cannot generate samples at throughputs that approach FACS analytical speeds. However, nanoliter scale sample handling is regularly performed in capillary flow-based systems, such as those used to manipulate liquid in chromatography systems. These systems use micron scale (microfluidic) channels to mix and transport small volumes with pinpoint accuracy. Unfortunately, this form of microfluidic sample handling has been difficult to adapt to HTS given the difficulty of maintaining a series of discrete samples in continuous flow. Switch valve configurations such as those used in standard chromatography injection systems allow discrete sample plugs to be serially injected and cleared from a capillary flow path but these components require large dead volumes to counter carryover and Taylor dispersion^{9,10} between samples.

Segmented flow has been used in microfluidics to address the challenges of single phase microfluidic systems with regards to sample separation, surface fouling, and inefficient mixing¹¹⁻¹³. Simple flow-focusing and T-junction microchannel geometries¹³⁻¹⁵ allow immiscible liquids

to be brought together in capillary flow to form dispersed droplets of one phase “carried” by the other. By using an oil carrier phase that is sufficiently omniphobic with respect to analytes in the aqueous phase, each droplet may be treated as an isolated sample. Typically, fluorocarbon oils such as perfluorodecalin are used to provide this physical and chemical barrier. This form of microfluidic sample handling is known as “droplet microfluidics.”

The highly uniform generation of droplets in microfluidics makes this technology a potentially attractive alternative to MWPs, as each microdroplet acts as its own separate reactor (Figure 1-1 A). Performing chemical reactions in discrete droplet volumes allows for rapid in-line analysis, much like cell analysis in FACS systems.¹⁴ This presents an attractive opportunity to scale complex liquid handling to speeds that match the true throughput limit of photometric assessment.

To perform discrete chemistry with droplets, they must be stored, manipulated, and analyzed. While the simplest way to store droplets separately is done in-line *via* micro-channels or capillaries, the back pressure generated by fluidic resistance within such systems can be problematic. High throughput screening operations require thousands of samples to be manipulated in series, so the use of amphipathic surfactants that prevent droplet coalescence have become commonplace and allow droplets to be stored in bulk.^{16,17} Surfactants such as the triblock copolymer perfluoropolyether-polyethylene glycol (PFPE-PEG-PFPE, Figure 1-1) mimic the function of phospholipid membranes in biological systems and prevent droplet merging when in contact.¹⁷

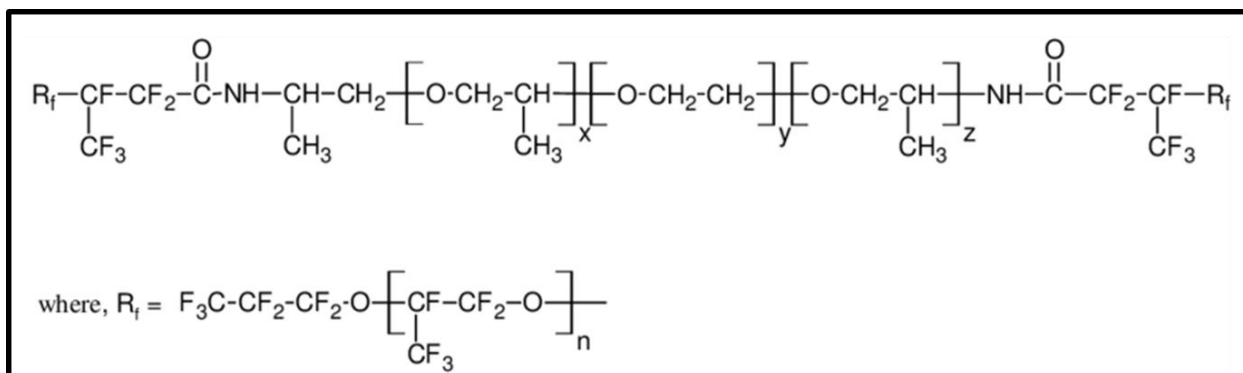


Figure 1-1: Chemical structure of PFPE-PEG-PFPE surfactants

Droplet microfluidics enables many of the core functionalities of microplate sample manipulation. Advances in microfabrication¹⁸ have enabled the creation of complex capillary geometries in monolithic chips for performing secondary operations on droplets. These have been designed to address many of the core functionalities of well plate screening.¹⁴ In well plates, robotic pipetting systems enable the addition of reagent to individual reactions. In droplet microfluidics, reagent addition is most commonly achieved by direct injection¹⁹ or pairwise merging²⁰ of droplets (Figure 1-2 B). When placed in series, direct injection allows multiple reagents to be added to each droplet in rapid succession. These functions often exceed the throughput of robotic handling by an order of magnitude or more.

In MWP format, small volumes of sample may be recovered for analysis, further reactions, or long-term storage. In droplet format, sampling can be performed by splitting droplets in flow using microchannel bifurcations^{21,22} (Figure 1-2 C) electrostatic forces,²³ or acoustics waves.^{24,25} Droplet splitting may also be used to perform sample exchange and solid phase extractions, allowing selective retention of targeted analytes in droplets via bead based capture and wash workflows.²⁶⁻²⁸

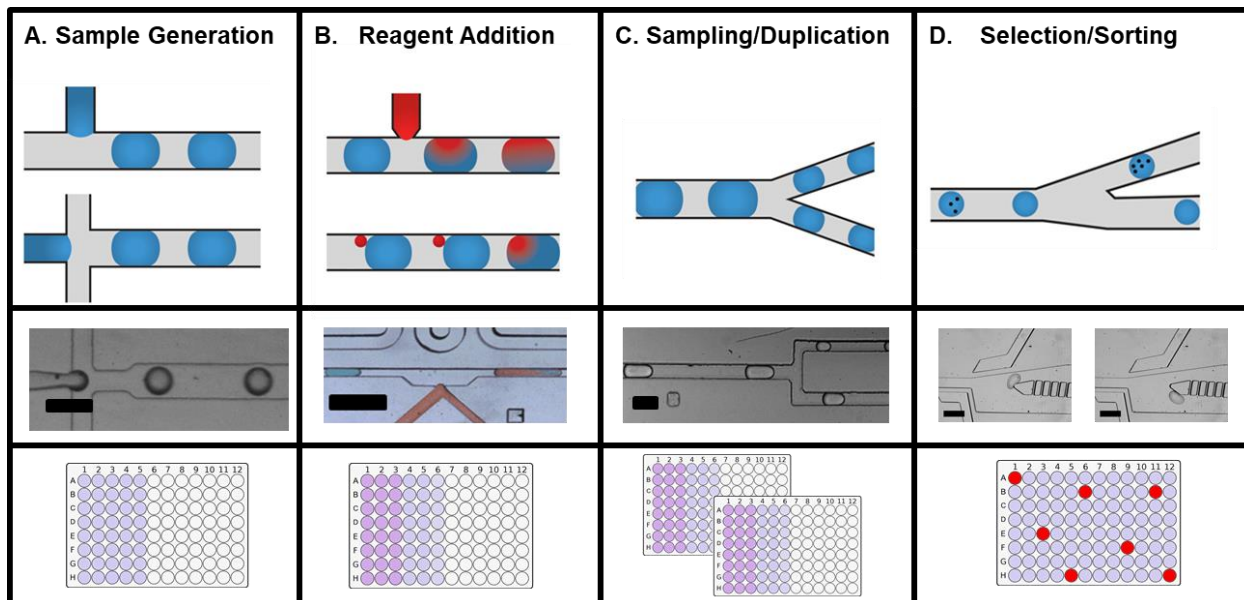


Figure 1-2: Unit operations in droplet microfluidics

Droplet microfluidics enables many of the core functionalities of microplate sample manipulation to be replicated in flow. A) Droplets are generated using flow focusing microstructures, creating digital separation of samples in distinct droplets (scale bar 100 μm). B) Droplet content may be modified through the addition of reagent to distinct droplets, either by droplet pairing or picoinjection (scale bar 300 μm). C) Reactions may be sampled in flow using droplet splitting architectures to break droplets into two daughter droplets, creating sample replicates (scale bar 100 μm). D) Droplets may be selected and recovered for further analytical assessment or statistical enrichment using active and passive sorting techniques like dielectrophoresis and magnetophoresis (Scale bar 500 μm).

The confinement of samples in a well plate facilitates easy sample recovery and secondary analysis, as well as clear spatial encoding of droplet identities. Droplets must be recovered following primary analytical readout if secondary analysis is necessary. The most common strategy for doing so is dielectrophoresis (DEP) where a droplet may be deflected in an electric field to exit a device via a selected channel (Figure 1-2 D).^{29,30} This sorting strategy operates in a similar manner and throughput to the electrostatic sorting of droplets in FACS, allowing droplets to be sorted at throughputs up to 10^5 samples per second.³¹ Fluorescence activated and absorbance activated³² droplet screening (FADS and AADS) are robust droplet screening techniques employed across a wide range of applications.^{29,32-40} While spatial information is difficult to maintain using emulsion based sorting and recovery methods, it is

certainly possible,⁴¹ and droplet microfluidics allows sample handling to begin to approach the throughput limits of photometric detection.

1.4 Analytical Challenges for Microdroplet Interrogation

Both droplet microfluidics and traditional MWP screening methodologies are limited in utility when coupled with the low information density of photometric detection. The analytical information obtained is simply an analog output on a single channel (photometric intensity) and gives little information on the underlying chemistry or biology within the sample without careful linkage through assay development. Only reactions that engender a distinct change in fluorescence or absorbance may be assessed using MWP readers, and few reactions exhibit these properties natively. As a result, significant research and development has been devoted to the discovery and use of fluorescent markers as indicators of chemical or biological activity. Advances in fluorescence polarization (FP), Förster resonance energy transfer (FRET) and fluorescence quenching have broadened the scope of reactions that may be probed using fluorescence,⁴² and have allowed spectroscopic techniques to maintain their dominance in the field of HTS. Unfortunately, these techniques still carry inherent disadvantages, in that they rarely probe the reaction of interest directly and suffer from significant risk of off target effects.⁴³ False positives may be caused by unintentional quenching or auto-fluorescence, off target effects (such as inhibition or activation of secondary reporter enzymes that produce a visible signal), and secondary or unforeseen chemical reactivity or affinity within the assay. Assay development must be undertaken with care, and often industrially relevant reactions prove difficult or impossible to link to optical outputs. In these instances, more information dense analytical assessment becomes the dominant driver of assay design and development, and HTS becomes less feasible.

When information content and targeted compound identification and quantification of small molecule analytes is paramount to analytical assessment, liquid chromatography and mass spectrometry dominate the laboratory toolbox. Historically, these techniques have not been accessible for high throughput screening applications. LC coupled to ultraviolet (UV) and MS detection systems have been adapted to MWP sample format, but the throughput of these systems is defined by the throughput of chromatographic separations and switch valve auto sampling. Chromatographic separations on analytical columns typically take a minimum of several minutes to run and require skilled method development and operation to maximize throughput, while switch valve based sample handling systems require significant dead volume (in the microliter range) and a serial load-inject workflow that limits the frequency with which samples may be introduced to analytical assessment. This bottleneck has been a significant area of research for analytical chemists. Recent developments, such as the use of multiple injections in a single experimental run (MISER), have enabled LC separations to be carried out in less than one minute per sample, which is a hugely enabling improvement relative to the multiple-minute separations of the past and represent a major step forward in throughput.⁴⁴

One potential work-around for high information content analytical assessment lies in direct analysis of samples without the use of chromatographic separation. Mass spectrometry is a potential candidate for this kind of analysis, due to the ability of MS to rapidly separate and detect multiple analytes in the gas phase. Mass spectrometers are comprised of three core parts: an ion source, a mass analyzer, and a detector. The ion source ionizes the analytes in the interrogated sample, ejecting them as charged molecules in the gas phase. The mass analyzer uses carefully shaped and focused electric fields to select and separate these gaseous ions based on their mass and charge. The wide range of mass analyzers available in standard MS

instruments⁴⁵ enable high resolution separations of these complex mixtures of ions, as well as the selection of a single ions of interest from a mixture. These separated ions are then sent to the detector to give in-depth chemical information on the mass of the analytes within the complex mixture.

Direct mass spectrometry has historically been limited in throughput by sample cleanup and sample introduction throughputs. Sample cleanup must typically be performed to minimize interference from ionization suppressing salts and buffers, but developments such as the Agilent Rapidfire, which can incorporate solid phase extraction (SPE) into the sampling workflow and rapidly load and inject samples from a MWP, have enabled faster sample injection to MS analysis.⁴⁶ More recently, acoustic sample injection (ECHO-MS) has been demonstrated and commercialized for direct MS analysis off of a wellplate,⁴⁷ and direct analysis in real time (DART-MS) have enabled more rapid sample analysis by MS without significant sample pre-processing.⁴⁸ These technologies have enabled MS analysis at speeds ranging from 10 s per sample to 0.3 s/sample, and have demonstrated the potential for coupling more complex analytical assessment to existing MWP technologies.

1.5 Direct MS Assessment in Droplet Microfluidics

Analytical assessment in droplet microfluidics has historically been limited in scope. The relative ease of adapting photometric detection to droplet analysis has made it the predominant approach in droplet microfluidic development, similar to MWP-based HTS. Microscopic sample sizes in droplet systems do not lend themselves to more data rich analytical platforms, largely due to the limited volume available for analysis. The need for greater analytical depth is well recognized within the microfluidics research community, and recent research has laid the groundwork for an expanded toolbox of analytical assessment in nanoliter

to picoliter droplet volumes.^{49,50} While chromatography with segmented flow at nanoliter to picoliter scales has been more difficult to achieve, recent work from our group has shown facile coupling of microchip capillary electrophoresis (CE) to droplet introduction, which enables individual separations to be performed at sub-second time scales.⁵¹ In addition to CE with fluorescent detection, Raman spectroscopy has been applied to single cell detection in droplets as a detection method that is fast enough to trigger dielectrophoretic sorting of droplets in flow.⁵²⁻⁵⁴ Raman spectroscopy allows label-free detection of analyte in droplets and in cells;⁵⁵ however, signal error can arise from interference from the oil phase and optical distortion in the droplets. Additionally, Raman spectroscopy suffers from low sensitivity without the use of surface enhancement or extended analytical timescales, which makes the technique more difficult to apply towards complex matrices in which non-specific molecular adsorption to the necessary gold and silver nanoparticles can limit analytical utility.^{49,53,54} Absorption spectroscopy has been applied in droplet microfluidic workflows in several formats as well,^{32,56} and electrochemical detection in droplets has been performed as well.⁵⁷

Mass spectrometry has also been explored as an analytical method in droplet microfluidics.⁵⁸⁻⁶¹ MS is particularly attractive as an approach to droplet analysis for several key reasons. First, MS is commonly coupled to continuous flow separation methods, as is exemplified in LC-MS systems. Ionization techniques, in particular electrospray ionization (ESI), may be easily integrated into the fluid path of a microfluidic system and are capable of detecting the contents of droplets as small as 60 pL in volume.⁶¹ After ionization, analytes in a mass spectrometer may be separated by the mass analyzer, allowing specific analytes to be detected based on their unique mass signature (Figure 1-3). While a wide variety of mass analyzers may be used for selective ion selection and gas-phase mobility separation in MS, many

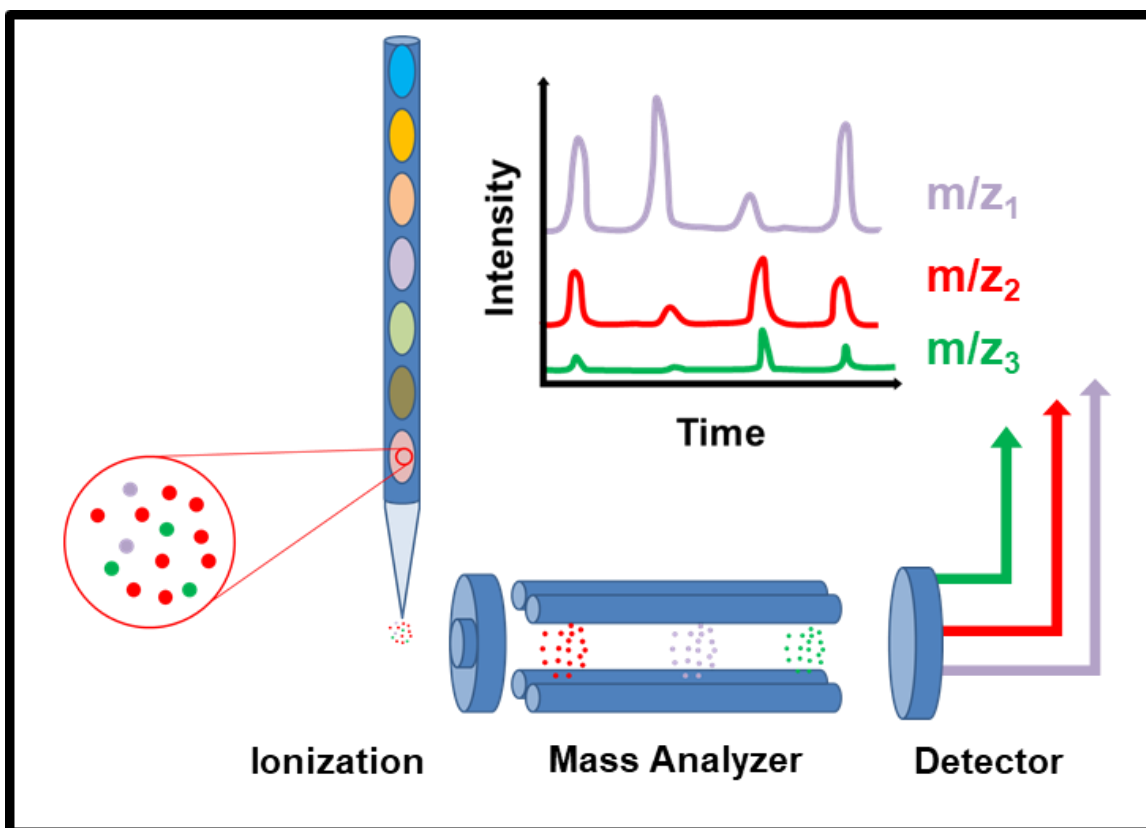


Figure 1-3: Schematic of droplet mass spectrometry

Droplet Mass spectrometry via electrospray ionization allows individual droplets to be sprayed by mass spectrometry as they are infused through a capillary to the ionization source. As each droplet flows to the tip of the electrospray emitter, the liquid within is nebulized and the contents are ionized by the high voltage electric field at the tip of the emitter. As the oil in between the samples exits the tip, electrospray ceases and the dielectric oil is left un sprayed. Propelled by focused electric fields, the ionized analytes within the droplet enter the mass spectrometer's mass analyzer, where the ionized materials are selected and isolated based on their mass to charge ratio and sent to a detector. The detector outputs an analog signal of ion intensity which, depending on the selection mode of the mass analyzer, may correspond to a single selected mass of interest (single ion monitoring) or a spectrum of intensities across a scanned mass range.

operate on timescales in the millisecond range,⁶² creating the potential for rapid analytical assessment that pairs well with droplets in flow. Finally, the end point detection of MS utilizes electron multipliers, which operate in a manner analogous to PMTs. The response time of these detectors are comparable with PMTs, and as a result, throughput limitation in droplet MS arises from a combination of sample introduction and scan rate. With proper optimization, as many as

30 droplet samples have been detected per second using mass spectrometry.⁶³

Despite the potential for MS to be adapted to high throughput screening in droplet flow paradigms, state-of-the-art systems have largely been limited to a few hundred samples per screen.^{58,59} The label free nature of MS detection, coupled with its compatibility to flow and segmented flow based systems, make it an attractive and practical analytical strategy in the development of microfluidic systems for high throughput screening.

1.6 High Throughput Screening for Directed Evolution

The use of fluorescent detection for high throughput screening has seen widespread use in the areas of drug discovery, probing biological interactions, and cytometry. These fields lend themselves to interrogation by photometric detection, because the output phenomena being probed is often a macromolecular response – i.e. a protein interaction that is blocked, a binding process that is inhibited, or a change in cell viability. These phenomena may often be observed by using a fluorescent indicator to signal a macromolecular response of interest. For example, the binding of two proteins may be probed for disruption by using a FRET donor or acceptor on each binding unit.

Unfortunately, systems and reactions of interest in HTS do not always lend themselves to interrogation by fluorescent tagging. Synthetic chemistry and biocatalytic processes involving the production of a small molecule or metabolite are particularly difficult to track with fluorescence. Tracking catalytic activity is difficult because placing a bulky fluorophore on a small molecule substrate can alter the chemical properties of these compounds and in turn, impact the way they interact with biological targets of interest.⁴³ Even if initial screens are possible using creative assay development, hits must often be validated with secondary, more robust analytical methods, and false positives or negatives are a consistent challenge.

Biocatalytic transformations are of growing import in the field of synthetic chemistry.⁶⁴ Reactions catalyzed by enzymes in biological systems tend to be highly chemoselective, regioselective, and enantioselective.⁶⁵ As such, enzymes have been shown to enable synthetic transformations that are often inaccessible or prohibitively costly when performed using transition metal catalysis. Notably, enzymes can perform challenging chemical syntheses under comparatively mild conditions, without the use of toxic solvents, high temperatures, or carefully controlled reactors.^{66,67} Millions of years of evolution have built these macromolecular structures into nature's chemical factories. Given its expedited access to enantiopure compounds, as well as sustainable, process-friendly reaction conditions, biocatalysis has emerged as a highly enabling synthetic strategy across the pharmaceutical and agrochemical industries.

The greatest drawback to using biocatalysis for chemical synthesis is the process by which one goes about identifying optimal enzyme biocatalysts. Enzymes have undergone millions of years of evolution to perform a specific task; they are rarely very active outside their native substrate scope. However, in order for the natural world to contend with evolving synthetic needs, it has developed ways to rapidly modify its catalysts to fit new and prescient activities. Enzymes have been significantly re-programmed to produce novel catalytic outputs through the process of evolution. The synthetic needs of modern chemists operate at timescales vastly smaller than the millions of years of evolution required for an organism to restructure an enzyme to achieve a new biological purpose. In order to adapt biocatalytic enzymes to prescient needs, modern biocatalytic chemists must find ways to rapidly alter the synthetic capabilities of enzymes. This has engendered the field of directed enzyme evolution (DE).

Directed evolution is the mutation and selection of enzymes for the purpose of creating and identifying new variants capable of performing previously inaccessible chemistry, either

through rational design or untargeted screening (Figure 1-4). The field was pioneered by Frances Arnold and her group at the California Institute of Technology and was recognized with the Nobel Prize in chemistry in 2018.⁶⁸ By screening thousands of enzyme variants, researchers have created catalysts capable of a remarkable array of transformations.^{64-66,69} These evolved biocatalysts are becoming common for use in the industrial production of chemicals ranging from biofuels to pharmaceuticals.⁷⁰

Unfortunately, the process of directed evolution requires laborious/extended screening campaigns to identify enzymes that are active against non-native substrates. The typical enzyme evolution workflow involves the expression of a library of enzymes in host cells, followed by the separation of each enzyme variant into its own individual well on a plate, induced protein expression for each captured colony, and lysis to release the expressed protein into solution. The lysate is then screened for activity on the small molecule of interest. In industry, a typical round of evolution involves assessment of anywhere between 10^3 and 10^5 enzyme variants. Additionally, an enzyme may undergo more than ten rounds of selection and mutation before a sufficiently stable, active, and efficient variant is identified. While industrial HTS facilities have the capability to screen libraries of such magnitude, conventional photometric methods of analysis are incompatible with biocatalytic transformations with regard to analytical depth. Chemoenzymatic processes typically occur on small molecule substrates and catalytic activity is intrinsically tied to the structure of this substrate. As a result, fluorescent detection becomes more challenging to implement, and direct detection by LC-MS or LC-UV is used for reaction analysis. This dependence on LC-based analysis significantly extends the timeline of biocatalytic method development and severely limits the number of novel enzymes produced by directed evolution.

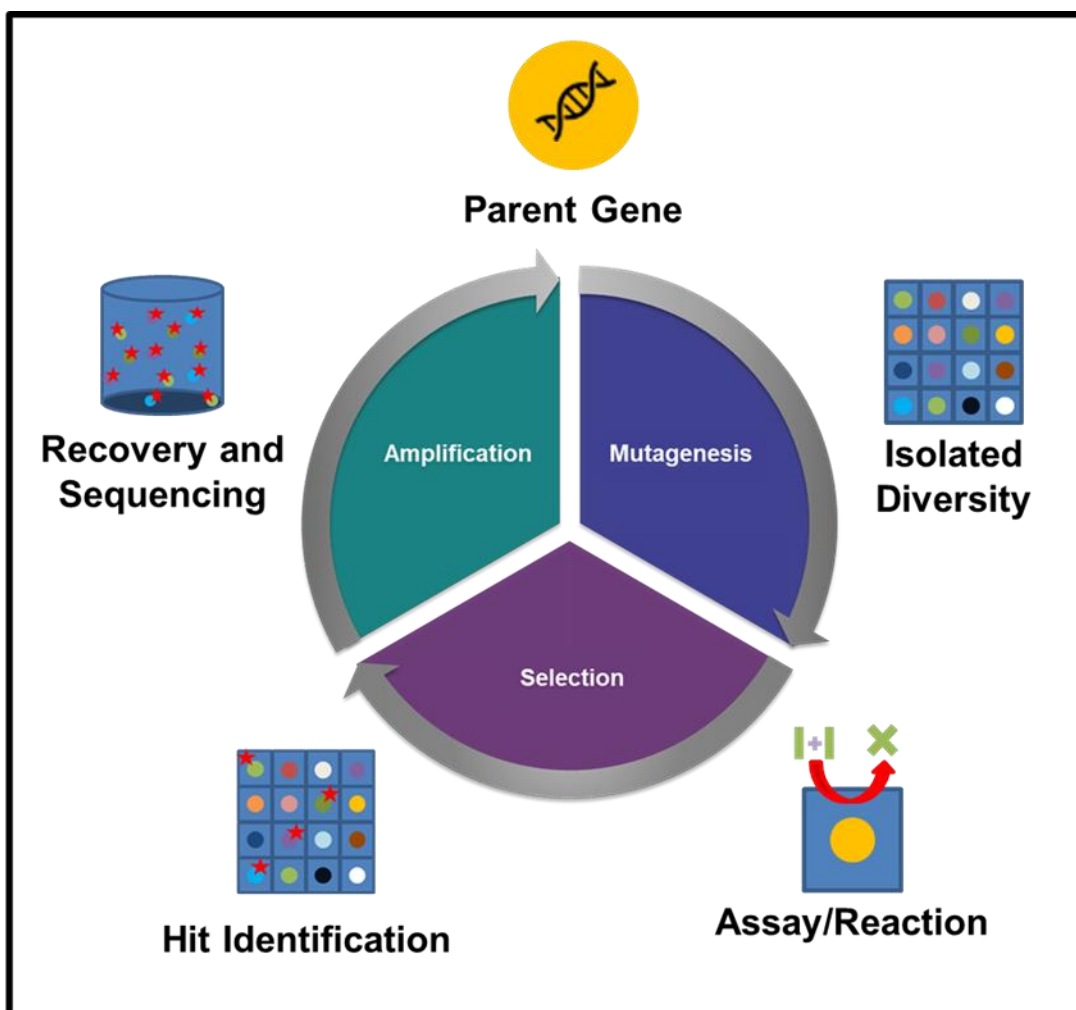


Figure 1-4: The molecular biology workflow of directed evolution

The core workflow of Directed Evolution follows the same cycle as natural selection and evolution. A parent gene is chosen as the basis for a library of enzyme variants. This gene is mutated to create the library, and each gene is transformed into cells and expressed in cell culture. Individual cells (and therefore enzyme variants) are separated into individual wells on a plate. The cells are cultured and lysed to release protein into solution, and then assayed for activity on the target compound. Hits are analytically assessed and identified, and the genetic information coding for those hits becomes the basis for the next round of parent genes.

For cases in which photometric detection may be applied to detect enzyme activity, rapid droplet microfluidics-based photometric screening methods have enabled millions of enzymes to be screened and sorted in record time.^{32,37,40,71-76} These methods have been groundbreaking for the field of biocatalysis and highlight the feasibility of screening for and identifying a new enzyme within a week's timeline. However, the severe limitations of photometric detection for

detecting catalytic outputs have stymied any effort to adopt these techniques on a broader scale. Mass spectrometry-based detection of microfluidic droplet contents represents a potentially revolutionary development for advancing HTS of enzymes. This would enable the rapid generation of droplets that contain individual enzymes, as well the ability to test sample activity in a label-free manner using MS. With droplet MS, 10^5 - 10^6 samples could be assessed in a single day, allowing rounds of evolution to be carried out in record time, regardless of the substrate.

1.7 Dissertation Overview

Early experiments with MWP based droplet MS screening indicated that enzyme activity tracked well with MS signal in direct droplet infusion experiments.⁷⁷ Many of these early experiments relied on MWP formats for reaction incubation as well as subsequent MS readout. As droplet microfluidics-based sample generation and manipulation became rapidly more versatile, novel HTS systems that demonstrate the feasibility of rapid creation and *in vitro* expression of enzyme libraries in droplets have been developed.^{40,78,79} In this context, the aim of this thesis work is to demonstrate facile integration of droplet MS into a droplet sorting device, while enabling reliable creation and MS assessment of droplets containing expressed enzyme. More significantly, this work aims to integrate these two systems for use in the complete workflow of directed enzyme evolution.

In Chapter 2, this work details the development and completion of first component system of these aims: mass activated droplet sorting (MADS)⁸⁰. This is a microfluidic system for high-throughput sorting of nanoliter droplets based on direct detection using electrospray ionization mass spectrometry (Figure 1-5). In this integrated system, droplets are sampled and

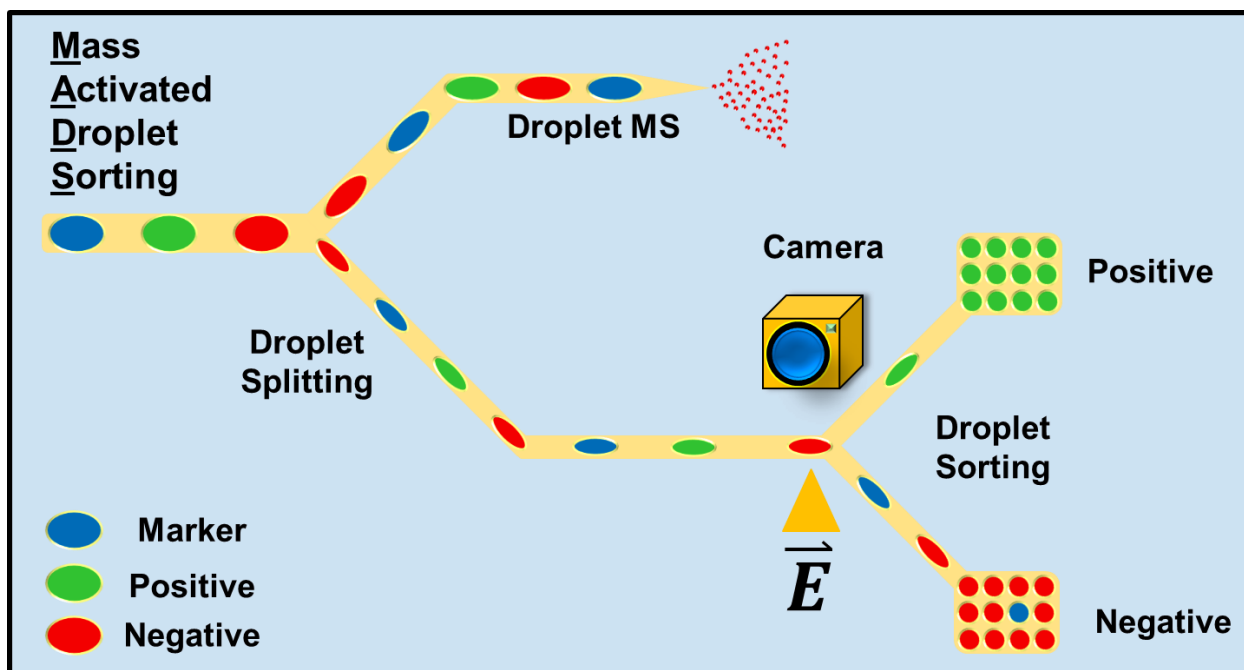


Figure 1-5: Schematic of the MADS system.

Mass activated droplet sorting (MADS) begins with the infusion of droplets onto the microfluidic device. On the MADS chip, droplets are re-spaced with carrier oil and split into two portions. The larger of the two portions exits the device and is sprayed at the mass spectrometer. The smaller of the two portions remains on the microfluidic chip where it is imaged by a camera and then sorted based off the alignment of the MS and camera signals and the MS signal relative to the selection threshold. Marker droplets infused into the device give a shared signal at both the camera and the mass spectrometer that enable accurate alignment of the two concurrent data streams.

analyzed by ESI-MS with a portion retained and sorted based on the MS result. We demonstrate a throughput of 0.7 samples/s at 98% accuracy and use the system to screen ~15,000 samples in 6 h. The hardware, software, and engineering supporting the microfluidic devices at the core of our system is also described in depth. In Chapter 3, application of the MADS system is demonstrated for directed evolution high-throughput screening. A microfluidics based workflow for performing *in vitro* expression of large enzyme libraries in droplet samples is detailed. Droplet polymerase chain reaction⁸¹ followed by fluorescent activated droplet sorting is performed to create and enrich a library of droplets containing isogenic DNA for the expression

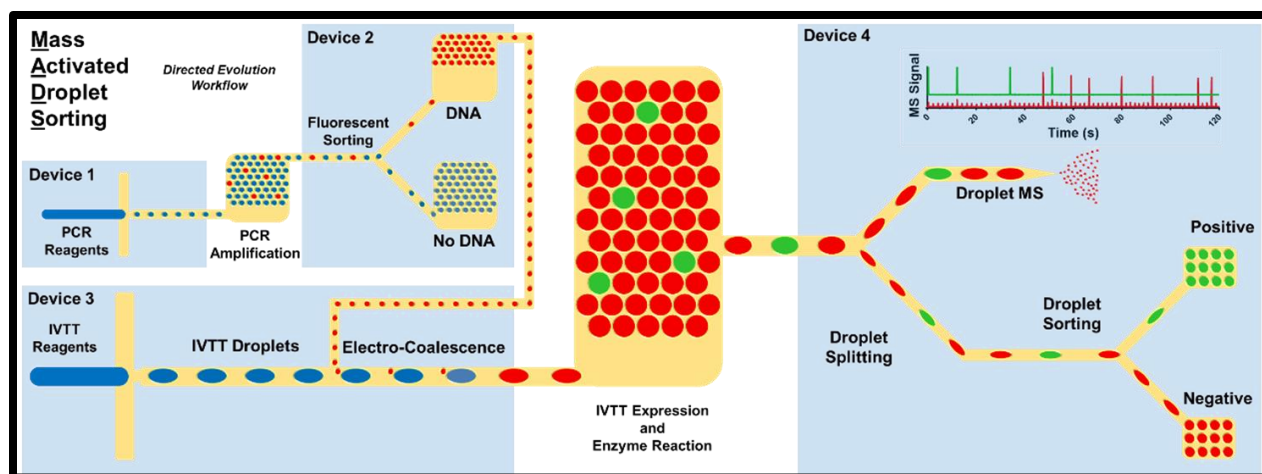


Figure 1-6: Schematic of the MADS workflow for directed evolution

The microfluidic workflow proposed in this thesis for the Directed Evolution of Enzymes in nanoliter droplets. The work begins with the production of single droplets of DNA with unique sequences coding for enzyme variants using digital PCR (Device 1). These droplets are then sorted by their fluorescence (Device 2) before being paired with *in vitro* expression reagents to express a unique protein in each droplet (Device 3). After expression and incubation with reagents of interest, individual droplets are split, with half of the droplet running to the mass spectrometer for analysis and half of the droplet running to a microfluidic sorting device for recovery (Device 4). Droplets are sorted based on the signal from the mass spectrometer, recovered, and analyzed for improved enzyme activity.

of a transaminase single site mutation (SSM) library. Following FADS enrichment, we use droplet pairing strategies to combine DNA bearing droplets with *in vitro* expression materials to express each droplet of amplified DNA as a unique transaminase enzyme, captured within a single 30 nL droplet. MADS is used to identify and enrich for winners performing better than a positive control parent enzyme (Figure 1-6).

In Chapter 4 the critical challenges to the implementation of the MADS system are detailed, and future areas of development for droplet mass spectrometry and mass activated droplet sorting are highlighted. Key challenges such as inter-droplet chemical transfer, reaction quenching in droplet systems, and the analytical limitations of direct mass spectrometry of microfluidic samples are discussed, and areas of further development potential for droplet mass spectrometry and mass activated droplet sorting are detailed.

Chapter 2: Mass Activated Droplet Sorting (MADS) Enables High Throughput Screening of Enzymatic Reactions at Nanoliter Scale

Adapted with permission from Holland-Moritz, et al. *Angew. Chem. Int. Ed.* **2020**, 59, 4470.

Copyright © Wiley

2.1 Introduction

Droplet microfluidics enables experiments to be performed at nanoliter to femtoliter scale, increasing throughput and decreasing unit costs of chemical and biological experimentation.^{77,82} A decade of research in the field has demonstrated the utility of droplet systems for a range of applications, including single cell gene expression profiling,^{75,83,84} small molecule screening,^{85,86} and diagnostics.^{87,88} Dielectrophoretic droplet sorting³⁰ has made possible the rapid recovery of selected samples for analysis.^{29,73,74,89} The ability of microfluidic systems to create, assay, and sort microscale samples is attractive in applications where sample preparation and analysis are bottlenecks, e.g. directed evolution of enzymes.

Active sorting of microfluidic droplets largely relies on optical detection.^{49,90} Fluorescence activated droplet sorting is most frequently utilized because of its high speed and sensitivity. FADS has found use in ultrahigh throughput screening for directed evolution.^{32,34,37,40,76,91} Screening large libraries is often the rate-limiting step in biocatalyst development, where thousands of enzyme variants must be tested for catalytic activity,⁷⁷ and plate based screens are time and resource intensive. FADS has enabled screening of libraries containing millions of variants in a few hours.

Fluorescence detection requires a reporter molecule, a condition that is difficult to meet in many applications.^{92,93} Fluorescent indicators must be carefully selected to ensure that they are retained within droplets,⁹⁴⁻⁹⁶ do not interfere with the process being investigated, and provide a readout dependent only upon the assay of interest. These limitations have restricted the use of droplet assays for high throughput biocatalyst screening in industry, where most target analytes are small molecule pharmaceuticals that are difficult to adapt to fluorescent assays. For example, in developing a transaminase for the production of the pharmaceutical sitagliptin, Savile et. al⁶⁶ screened nearly 36,000 variants of a transaminase. High performance liquid chromatography (HPLC)-MS was used to guide the selection of variants because the transformation (ketone to chiral amine) produced no significant change in the optical properties of the substrate and product.

To address the limits of fluorescence based screens, recent work has expanded the analytical techniques that can be applied to the active sorting of droplets. UV/visible light (UV-Vis) absorbance,^{32,56} droplet imaging,^{97,98} and Raman spectroscopy^{52,54} have been employed as alternatives to fluorescence detection in droplets. Although these methods have widened the range of analytical techniques available for droplet sorting, they too have significant limitations in their implementation. Raman spectroscopy allows label-free detection of the target analyte but the signal is negatively affected by interference from the oil phase and optical distortion in the droplets. Raman additionally suffers low sensitivity without the use of surface enhancement, which makes the technique more difficult to apply in complex matrices where non-specific molecular adsorption to the necessary metal nanostructures can limit analytical utility.^{49,53,54}

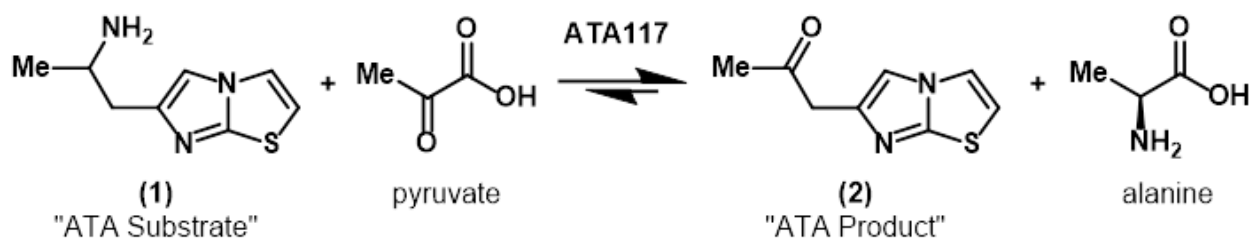
Absorbance spectroscopy offers another unlabeled approach to droplet detection. However, it is both less specific and less sensitive than FADS, necessitating the use of secondary

reporters for detection when the target analyte absorbs weakly and when the reaction does not result in a sufficient change in absorbance.³² Droplet imaging techniques are limited to visible characteristics such as size, color, or particle content, none of which are direct methods for observing the chemical content of a droplet sample.

Analysis and sorting of droplets using mass spectrometry would be a valuable addition to the methods for droplet sorting. MS offers nearly universal label-free detection with high sensitivity and selectivity, as well as the flexibility for multiplexing. Recent work with ESI^{58-61,77,99} and matrix-assisted laser desorption/ionization (MALDI) MS¹⁰⁰⁻¹⁰² has shown that MS can be used for droplet analysis; however, sample destruction by ESI has been an obstacle to coupling it to microfluidic droplet sorting.

Mass Activated Droplet Sorting (MADS) is a technique that couples droplet ESI-MS to DEP sorting. MADS is capable of sorting thousands of nanoliter droplets at ~0.7 samples/s based on their MS signal with up to 98% accuracy. The method relies on a programmable sorting algorithm that enables MS-based sample identification and dynamic thresholding.¹⁰³

The utility of this system is first demonstrated by sorting a pool of droplets based on 1-(pyridin-3-yl)ethan-1-amine (“pyridinyl amine”) concentration. Following this, we screen samples for activity of the transaminase ATA-117 after *in vitro* expression in droplets. We enrich droplets based on their conversion of the non-native substrate 1-(imidazo[2,1-b]thiazol-6-yl)propan-2-amine (“ATA Substrate”, Scheme 2-1). This screen would not be feasible by either FADS or AADS methodology, illustrating the potential of MADS to screen reactions that are not accessible by other methods.



Scheme 2-1: ATA117 conversion of 1-(imidazo[2,1-b]thiazol-6-yl) propan-2-amine

The reaction for the transformation of the non-native ATA Substrate into its corresponding ATA product ketone is catalyzed by ATA 117.

2.2 Experimental - Mass Activated Droplet Sorting

In order to perform Mass Spectrometry based sorting of microfluidic droplets, we had to design and build a microfluidic system capable of both reading and sorting nanoliter scale droplets on a single device. To achieve this, we built the microfluidic devices, optimized their operation, and integrated the software control systems that would make and execute microfluidic sorting decisions.

2.2.1 Microfluidic Device Fabrication

In order to perform microfluidic functions, we designed and built a series of microfluidic devices to perform both droplet mass spectrometry, and reagent addition to microfluidic droplets. Microfluidic devices were fabricated using standard soft lithography techniques.¹⁰⁴ Devices were designed using AutoCAD software (Autodesk) and printed onto mylar film masks at 25000-50000 dpi (Fineline Imaging). Using these masks, SU-8 2075 (Microchem Corp) was patterned (Figure 2-1) to a depth of 200 μm on a 7.62 cm diameter silicon wafer (University Wafer). Poly dimethyl siloxane (PDMS) (Curbell Plastics) was prepared in a 1:10 ratio of activator to monomer, poured over the wafers and cured 1 h at 70 $^{\circ}\text{C}$ before it was removed from

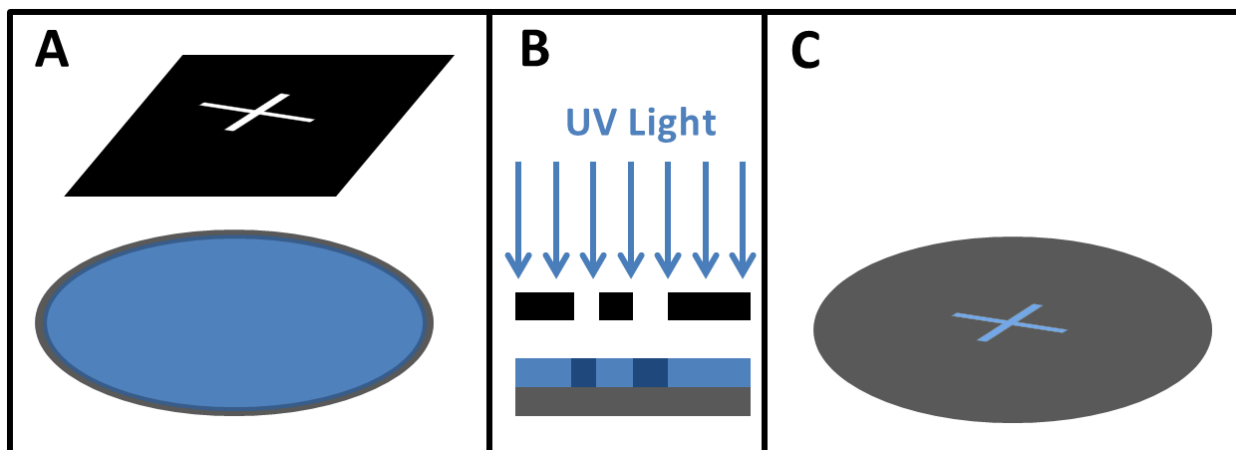


Figure 2-1: SU-8 patterning for soft lithography.

Fabrication of master molds for microfluidic devices was performed using established techniques for SU-8 patterning on silicon wafers. (A) SU-8 2075 was coated on the surface of a silicon wafer after application to the surface by spinning the wafer at 1200 RPM. (B) The SU-8 on the surface was exposed to UV light through a Mylar mask and baked on a hotplate to cure the polymer exposed to the light. (C) Uncured SU-8 was washed away with solvent, leaving behind a positive image of the mask design on the surface of the wafer. This positive image was then used to mold microfluidic devices in PDMS.

the mold and hard baked for 1 hour at 120°C. Punched ports were created in the cured PDMS using a sharpened 20 or 22 ga syringe needle (BD) and the punched holes were cleaned by passing a short length of fused silica capillary (360 µm OD, 20 µm ID, Molex) through the hole to clear any particulate left behind by the punch.

Devices were cleaned with Scotch tape (3M) and bonded to 7.62 cm long (3 x 1 in) glass microscope slides (Corning) after exposing both the slide and the molded PDMS to oxygen plasma for 15 s each using a corona discharge wand (Electrotechnics). Bonded devices were placed in an oven to cure at 80 °C. Within 10 min of bonding, chips were flushed with a derivatization solution of trichloro(1H,1H,2H,2H-perfluorooctyl)silane (Sigma-Aldrich) dissolved 2% by volume in perfluorodecalin (95%, Sigma-Aldrich). Flushed devices were baked overnight at 70 °C.

2.2.2 Fluid Flow Control

Both Syringe pumps and pressure pumps were used to drive flow in microfluidic devices. When using syringe pumps to drive fluid flow, Hamilton Gastight glass syringes were connected to Teflon (PFA) tubing using 250 μm ID PEEK unions (Valco). 30 ga Teflon tubing was sheathed in 1/16" outer diameter (OD), 1/32" inner diameter (ID) Teflon tubing (Cole Parmer) to allow the 1/16" Peek union ferrule to grip the smaller diameter tubing. These syringes were driven using a Harvard Apparatus PHD syringe pump. When using gas pressure to drive fluid flows onto a device, sealed vials (Chemglass Lifesciences) were pressurized using an Elveflow OB1-Mk3 pressure pump. Tygon tubing running from the pressure pump was connected to syringe needles via luer-lock connectors and these were inserted into the vial septa. 30 ga thin wall Teflon tubing (Cole Parmer) inserted through the same vial septa allowed fluids to exit the vials and flow onto the microfluidic devices when the vial headspace was pressurized.

2.2.3 Droplet Storage, Incubation, and Re-injection

Droplets were collected in a custom built storage chamber made from an 8 cm Pasteur pipette with the tapered tip removed (Figure 2-2). Silicone or rubber stoppers (00, Fisher) were cut with biopsy punches (6 mm, Fisher) to cap the ends and access holes were cut to allow the insertion of 30 ga Teflon tubing through the caps. During chamber filling, tubing from the droplet generation or reagent addition device was inserted into the bottom of the chamber, allowing generated droplets to flow directly into the chamber. A second segment of tubing inserted into the bottom of the chamber allowed excess oil to be drained off. During chamber emptying, this second segment was tied off and the first was used to pump oil into the chamber. With no available escape through the bottom of the chamber, droplets flow out through the tubing at the top of the chamber.

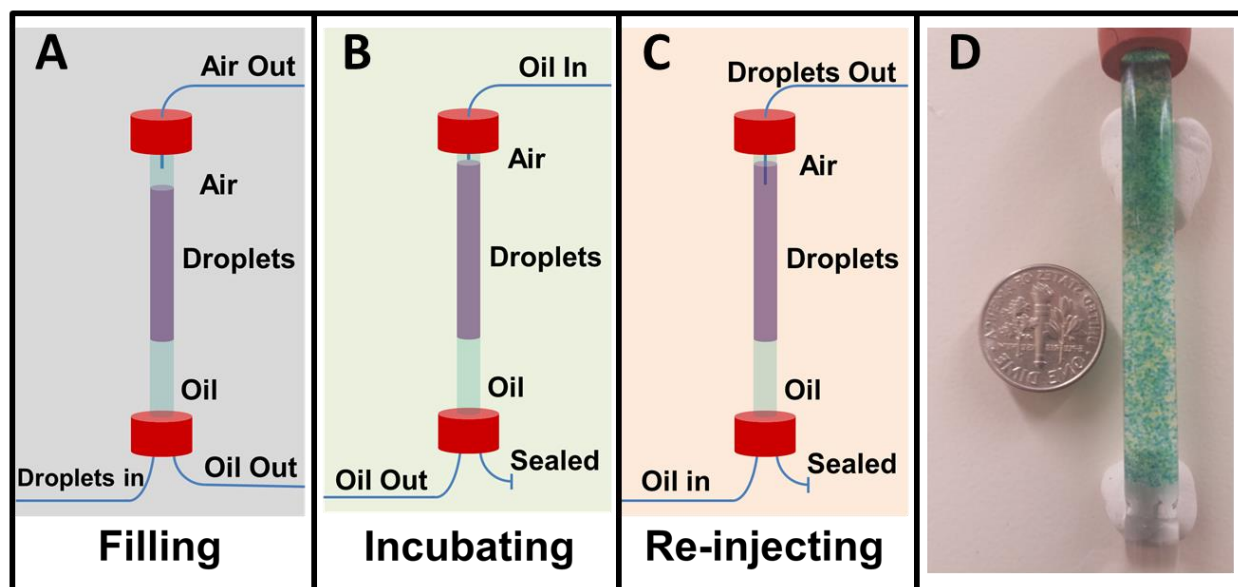


Figure 2-2: Storage, incubation, and reinjection strategy for nanoliter scale samples.

Storage and incubation strategy for surfactant stabilized microfluidic droplets. (A) During generation, droplets flow from the outlet of the droplet generator and into the bottom of the storage device. The less dense aqueous droplets float to the top of the chamber, allowing excess oil to be drained through a drain in the bottom. (B) For extended incubation periods, droplets are stored in the glass chamber, but oil is circulated using a peristaltic pump. Oil flows in through the inlet at the top of the chamber, while an equivalent volume is removed through the outlet at the bottom. (C) During droplet re-injection for both reagent addition and MADS protocols, oil is pumped into the chamber through the bottom of the storage device, and droplets are pumped out through tubing inserted into the top. (D) An image of 30 nL droplets stored within this microfluidic droplet storage device, next to a US dime for scale. Individual droplets may be seen as specks of blue in the device.

During extended incubation (>3 h), oil was circulated through the container by drawing oil from the bottom of the chamber and simultaneously pumping replacement oil in through the upper tubing. Oil was circulated using a Masterflex L/S peristaltic pump (Cole Parmer).

2.2.4 Device Imaging

During device operation, droplet size, frequency, color, and contents were monitored using one of two cameras. For high speed monitoring of droplet formation, reagent addition, and high frequency events, a Phantom Miro C110 high speed camera was connected to a Nikon TS2-FL microscope. During MS-directed droplet sorting, a Cognex 7600C image sensor connected

to the same microscope was used to monitor the droplets exiting the device to track and count droplets prior to sorting.

2.2.5 Mass Spectrometer Operating Conditions

Droplet analysis was performed using an Agilent single quadrupole mass spectrometer (6120B). Droplets were directly infused from the sorting device into the mass spectrometer through a 360 μm OD Teflon capillary directly inserted into a commercial Agilent CE-MS source.⁷⁷ Parallel sheath flow composed of deionized water was driven by a Harvard Apparatus PHD syringe pump and a 10 mL Hamilton gastight syringe at 30 $\mu\text{L}/\text{min}$. Capillary voltage was 3 kV, with 350 $^{\circ}\text{C}$ drying gas set at 30 L/min and a source pressure of 10 psi. Single ion monitoring was used to track the small molecules (Table S1) pyridinyl amine (m/z 123.0), ATA substrate amine (m/z 182.0), carnitine (m/z 162.3), neostigmine (m/z 223.0), and chlorocholine (m/z 122.3). Peak height (MS dwell time) was set to 0.005s. A sample MS trace from a continuous stream of droplets can be seen in Figure 2-3.

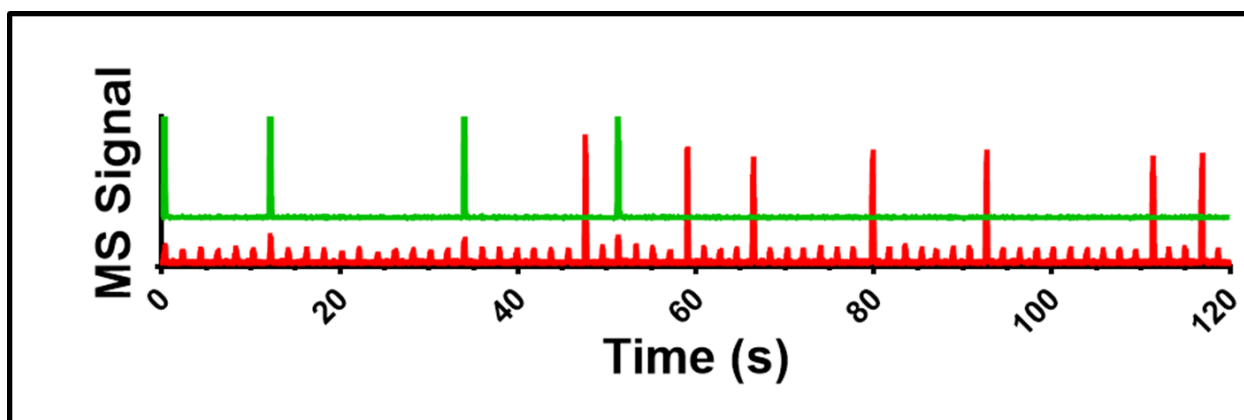


Figure 2-3: Typical trace from direct infusion of oil-segmented droplets

A sample trace of droplets infused onto a mass spectrometer while monitoring amine **1** (green trace) and pyridinyl amine (red trace) within the droplet. As each droplet is nebulized off of the tip of the electrospray needle, the analytes within ionize and may be read by the mass spectrometer. Droplets are interspersed with an inert oil carrying phase which does not ionize during electrospray. When the spacing oil is nebulized, no signal is observed on the mass spectrometer. Thus, each individual droplet appears as a peak on the two traces, with the peak intensity corresponding to the amount of the small molecule contained within the sample droplet.

2.2.6 Device Operation: Mass Activated Droplet Sorting

Flow in the MADS chip is driven by a combination of syringe pumps and gas pressure. A 10 mL Hamilton syringe filled with perfluorinated oil respaces droplets as they enter the device while a 1 mL syringe drives oil into the storage chamber and droplets onto the device. Both syringes are driven by a single syringe pump, programmed to push the 1mL syringe at 1.5 $\mu\text{L}/\text{min}$. Thus, the net flow onto the device is 16.5 $\mu\text{L}/\text{min}$, Droplets entering the device are spaced with oil at a 1:10 ratio.

Due to the differences in back pressure between the MS and sorting exits of the device, the sorting outlet vials and the sorting region are held at elevated pressure. To do this, a pressurized vial drives oil into an inlet in the device located at the end of the delay line but prior to the sorting junction (Figure 2-4). The sorting bifurcation flows into two vials that are pressurized equally using a T-split line from the gas pressure pump. The difference in pressure between the oil vial and the collection vials dictates the velocity of droplet flow through the sorting region, and may be used to accelerate droplets as they enter the sorting region. Typically, this pressure drop is held at 0.2 psi (1.4 kPa).

The applied pressures at the sorting region of the device must be slowly ramped as the device begins to operate and the two exit paths are filled with droplets. This ensures that both the on-chip delay line and the connected capillary maintain approximately equal back pressures as they fill with samples. Once both exits are filled and flowing at the same rate, the applied back pressure can be maintained indefinitely and the device will operate with minimal user input for as much as 6 h. We found the optimal operating back pressures for our system with a ~ 60 cm

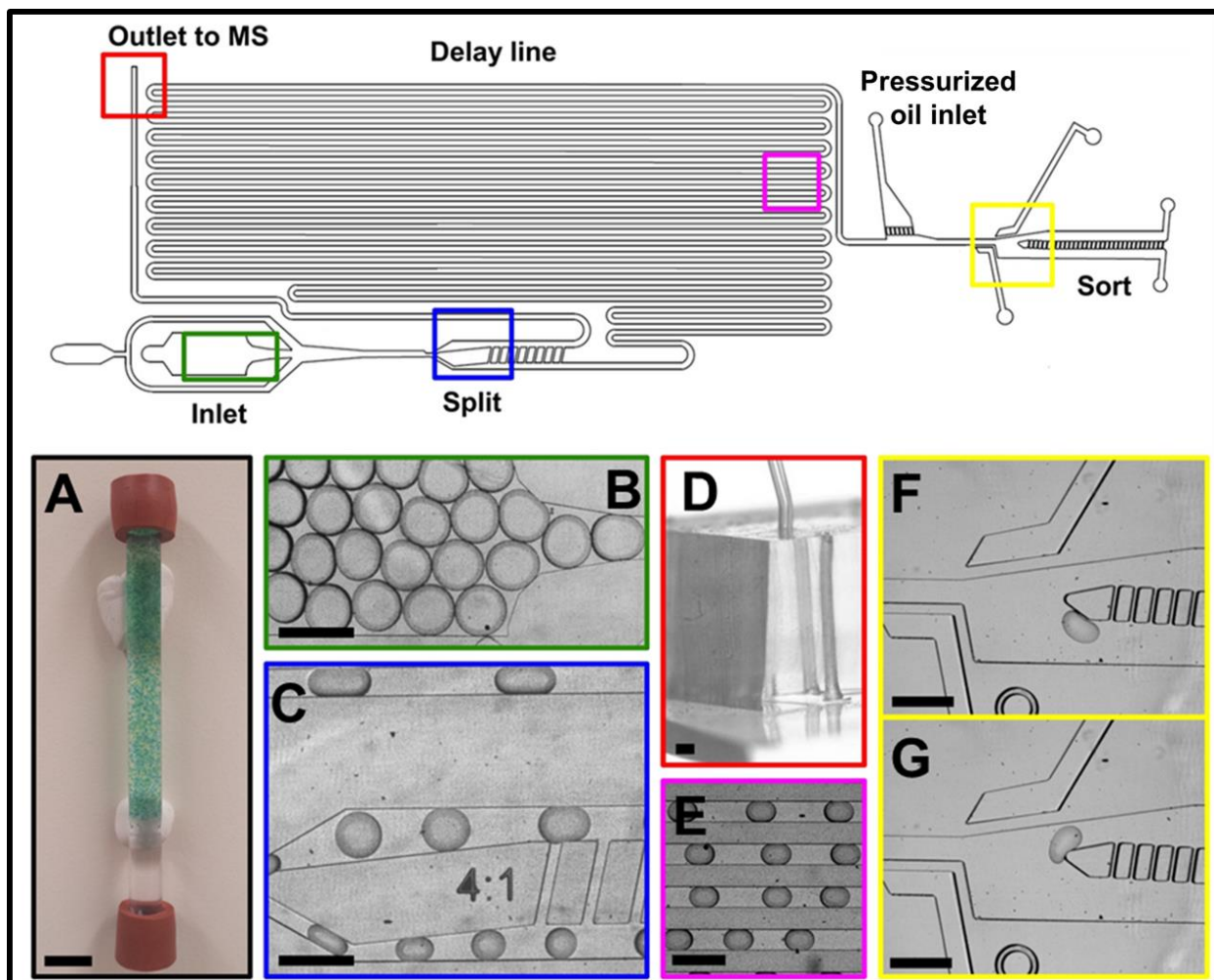


Figure 2-4: Full schematic and detail of the MADS device

(A) ~25 nL droplets are stored in bulk prior to being pumped onto the device. (B) As they enter the microfluidic device, (C) droplets are split into two daughter droplets. (D) The larger of the two halves leaves the device via an inserted Teflon capillary while (E) the smaller of the two remains on chip and is pumped into an extended delay line. After the larger droplet is sprayed, a 1.5 kV alternating current (AC) pulse at the sorting junction (F, G) directs droplets into one of two outlets. A constant flow rate through the sorting region is maintained by keeping a constant pressure difference between a source vial of oil connected to the pressurized oil inlet and the outlets of the device. Scale bar in (A) is 1 cm, all others are 500 μm .

transfer capillary to be 1.2-1.6 psi (8.3-11 kPa) at the oil inlet, but this will vary depending on the input flow rates, exact capillary lengths, droplet composition, and the MS source pressure.

2.2.7 DEP Sorting and Imaging Analysis

In simple sorting experiments of analyte in water, peak threshold voltages for sorting were set based off of the experimentally observed peak intensities from the first few droplets. If a sprayed droplet was determined to meet the sorting criteria, a high voltage alternating current (AC) pulse applied to the upper salt water electrode in the sorting region would pull it into the positive outlet; otherwise the droplet would be hydrodynamically directed to waste (Figure 2-4 F). Collected pools of droplets were imaged on a Nikon TS2-FL with a Phantom Miro C110, and analyzed using imageJ software to determine sorting accuracy and enrichment. Positive “Hit” droplets could be confirmed via fluorescent imaging of the hit droplets, which contained flavin adenine dinucleotide (FAD) to set them apart from the non-fluorescent marker and blank droplets. Excitation wavelength for FAD imaging was 470 nm (blue light emitting diode, LED) and a 525 nm Nikon fluorescein isothiocyanate (FITC) filter cube was used to filter the emission.

2.2.8 Integrated Electronics

Droplet signals from the electrospray of infused droplets are acquired from a four channel DAC card (Agilent G1960-67039) installed in the mass spectrometer. The frequency response of the DAC card was improved by removing the two 1 μ F capacitors from the lowpass filter stage for each 10x channel. The 10x channels are used to provide better signal to noise performance. A standalone four channel programmable gain Bessel lowpass filter (Alligator USBPGF-S1/LX) is used to adjust signal gain and frequency response prior to acquisition. A Sigilent SDS 1104x-E oscilloscope was used to monitor and record signal outputs in real time. On the microscope, a machine vision camera with open collector logic outputs is used to visually identify droplets (Cognex 7600C).

A STM32F407 ARM Cortex M4 microcontroller board (Mikroelektronika 1685) runs custom software to synchronize the outputs from the mass spectrometer and the machine vision camera. The mass spectrometer outputs are wired into four 12-bit analog to digital (ADC) inputs, and the open collector camera outputs are wired into two digital inputs with internal pull-ups. A digital output is used to control the DEP sorting voltage. This signal is buffered by a driver integrated circuit (Infineon IR2125), which selectively applies 5V power to a 1.5 kV AC 30 kHz cold cathode fluorescent lamp (CCFL) inverter board (JKL BXA-601). The outputs of the CCFL inverter are wired in parallel and connected through two 100k resistors in series to the sorting junction electrodes.

The microcontroller runs software written in C using Eclipse and ChiBIOS RTOS libraries. The microcontroller board is interfaced via universal serial bus (USB) to control software written in Labview on the host computer (PC).

2.3 Mass Activated Droplet Sorting - System Overview

The destruction of sample during ESI renders impossible the direct sorting of material that has been analyzed by MS. This challenge was resolved by asymmetrically splitting the droplets²² and performing analysis and sorting on the two different portions. In this approach, ~25 nL droplets are pumped from a storage chamber (Figure 2-2) and onto the chip (Figure 2-4), where they are split into two daughter droplets. The larger of the two daughter droplets flows into PFA capillary that is mated to a single quadrupole mass spectrometer via a sheath-flow ESI-MS source, where it is directly analyzed without the need for oil removal (Figure 2-3).^{77,105} The smaller of the two daughter droplets travels into an on-chip delay line to allow time for its sister droplet to reach the mass spectrometer. The smaller sister droplet may be deflected into the appropriate exit using dielectrophoresis once the larger daughter droplet has been analyzed and a

sorting decision made. For testing, the accuracy of the system is determined by collecting and imaging the sorted droplets after each experiment.

2.3.1 Sorting Strategy

In principle, accurate droplet sorting could be achieved either by counting the samples detected by MS and aligning that count to the samples at the DEP junction, or by setting a time delay between the signal on the MS and the DEP pulse that matches the delay between electrospray and sorting. In practice, neither of these strategies is sufficiently stable. The time delay approach is problematic because slight differences in droplet size, spacing and velocity will tend to result in sorting errors. A simple counting approach is problematic because a single miscounted event can result in a frame shift where every subsequent sample is sorted incorrectly.

We used a modified counting approach in which “marker” droplets were randomly mixed with the sample droplets so that the system could proofread its counting (Figure 2-5). To match the droplet entering the DEP sorting junction with its corresponding mass spectrometer signature, a camera monitors droplets as they enter the sorting region. The camera has image processing capabilities that allow it to trigger a digital signal based on droplet color. The marker droplets contain a unique mass analyte that is monitored by the mass spectrometer and a colored dye that is recognized by the camera. The signal given by marker droplets is used to synchronize the mass spectrometer data with the sorted droplet stream.

The microcontroller software driving the sorting decisions uses a real-time operating system (RTOS) that runs several processes in parallel. The first process acquires analog input signals from a modified digital to analog converting board installed in the mass spectrometer. A peak detection algorithm identifies peaks on each input signal and records the maximum value for each peak. The magnitudes of these values are then used to classify the peaks and determine

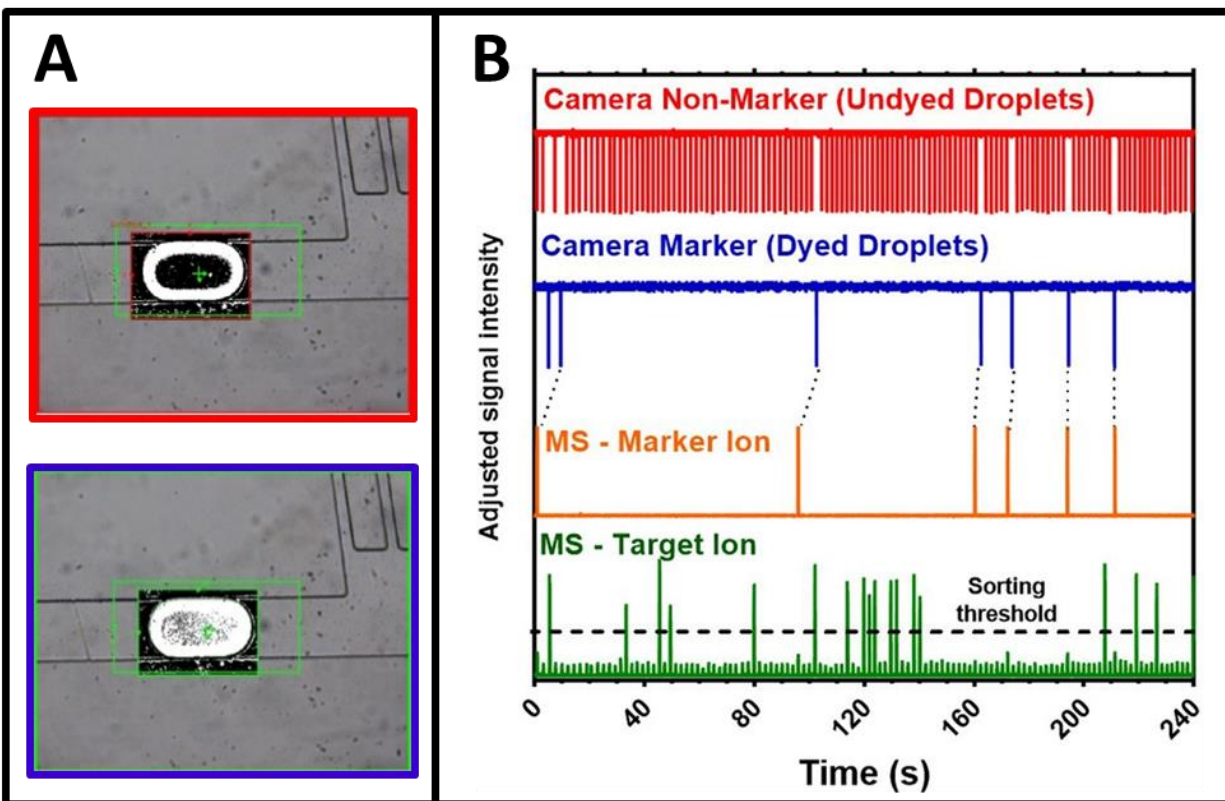


Figure 2-5: Operating principle for signal alignment between on-chip imaging and MS

(A) A camera with on-board image-recognition software capable of recognizing the difference between uncolored and colored droplets may be used to track droplets within the MADS system. The addition of a small molecule (either amine 1 or carnitine) to a droplet dyed blue with food dye allows a distinct signal to be read in each split droplet as it passes the mass spectrometer and as it passes the camera. The alignment of these two signals (B) allows the MADS system accurately track droplets as they exit either side of the microfluidic device used to analyze and sort them.

which are markers, and which will be targeted for sorting. This data is stored in a queue of virtual droplets. The second process simultaneously monitors the digital output from the camera, classifying the droplets as markers or non-markers based on the output signal. The third process handles communication with the host PC, allowing the user to set sorting parameters and read real-time sorting statistics.

The system synchronizes the mass spectrometer and camera droplet streams by monitoring the intervals between marker droplets in each. It starts operation with all droplets

directed to waste. As droplets start flowing past the camera junction, the system counts the number of non-marker droplets between marker droplets. Once the system detects an interval at the camera that matches an interval of virtual droplets stored in the queue, it enters the synchronized state. In this state, the stored droplet information is used to make sorting decisions about the droplets detected at the camera. The system will continue to monitor the intervals between marker droplets, and if the interval between marker droplets ever differs from the interval in the queue by more than three samples, the system enters the non-synchronized state. It then attempts to re-synchronize, starting with the next interval. The system is thereby self-correcting and capable of responding to anomalous events such as merged or split droplets that can cause miscounting. A schematic of this basic decision tree is shown in Figure 2-6.

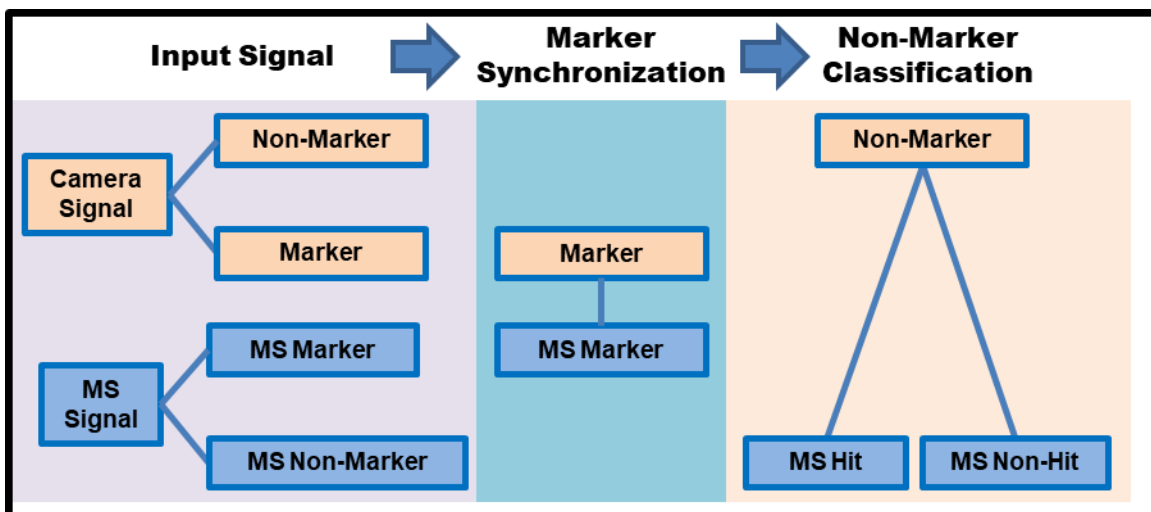


Figure 2-6: Decision tree detailing simple threshold-based droplet sorting

The MADS operating system under basic threshold operating mode. Input signals are sent to the microcontroller controlling MADS sorting from both the camera at the sorting region and the mass spectrometer. The alignment of the camera and the mass spectrometer is performed using marker droplets, allowing the synchronization of signals and enabling accurate sample tracking. Droplet “Hit” thresholding is based on a user defined threshold set at the beginning of operation, and hit droplets are identified from among the non-marker droplets based on signal intensity above this threshold.

In most experiments we aimed for a marker frequency of about 20%, because the frequent re-alignment (an average of one in five droplets) provides a regular check for the system to ensure it continues to sort accurately. However, the frequency of these markers is flexible and we have observed successful sorting with as low as 7% markers. At higher marker frequencies, a minimum alignment interval length may be set below which the microcontroller will not attempt to align. This reduces the chance of alignment to duplicate intervals, which become more common with increased marker frequency.

2.3.2 Sorting Accuracy

To assess the efficacy of the MADS device and its supporting software, we generated, mixed, and sorted three types of ~25 nL droplets. The bulk (~70-80%) of the droplets contained 50 μ M pyridinyl amine. Approximately 10% of the droplets contained 500 μ M pyridinyl amine (a tenfold increase in target signal) and FAD.

Experiment	Before Sorting		Marker droplets (%)	After Sorting	
	Samples Screened	Hit droplets in initial pool (%)		Hit droplets in positive pool (%)	Hit droplets in negative pool (%)
Experiment A	4159	4.7	7.2	88.5	0.9
Experiment B	4444	7.4	11.8	91.7	3.1
Experiment C	3625	9.8	18.8	91.6	4.1
Experiment D	3660	10.8	20.1	96.0	2.9
Experiment E	4001	11.0	20.1	98.7	1.5

Table 2-1: Sorting results for MADS proof of concept, with user-defined thresholding
 Sorting results from experimental runs of the mass activated droplet sorting device. Sorted droplets contained two levels of pyridinyl amine, 50 μ M and 500 μ M. In these experiments, target droplets with 500 μ M pyridinyl amine are enriched 9-10 fold higher in the final pool relative to the starting pool

These latter droplets served as high-signal targets for sorting. Although not used to make sorting decisions, the FAD is both visibly yellow and fluorescent, allowing sorting accuracy to be evaluated and confirmed by imaging after each experiment. The final 10-20% of the samples contained 500 μM ATA Substrate as a marker ion and blue food dye. These served as markers for sorting alignment.

The mixed population of droplets was pumped onto a sorting device where samples were split and sorted based on the signal from the pyridinyl amine. The threshold for sorting was manually set based on the signal intensity observed for the low and high concentration of pyridinyl amine. Sorted droplets were collected and imaged to analyze the accuracy of the sorting system. Table 2-1 summarizes five representative sorting experiments, and the results of

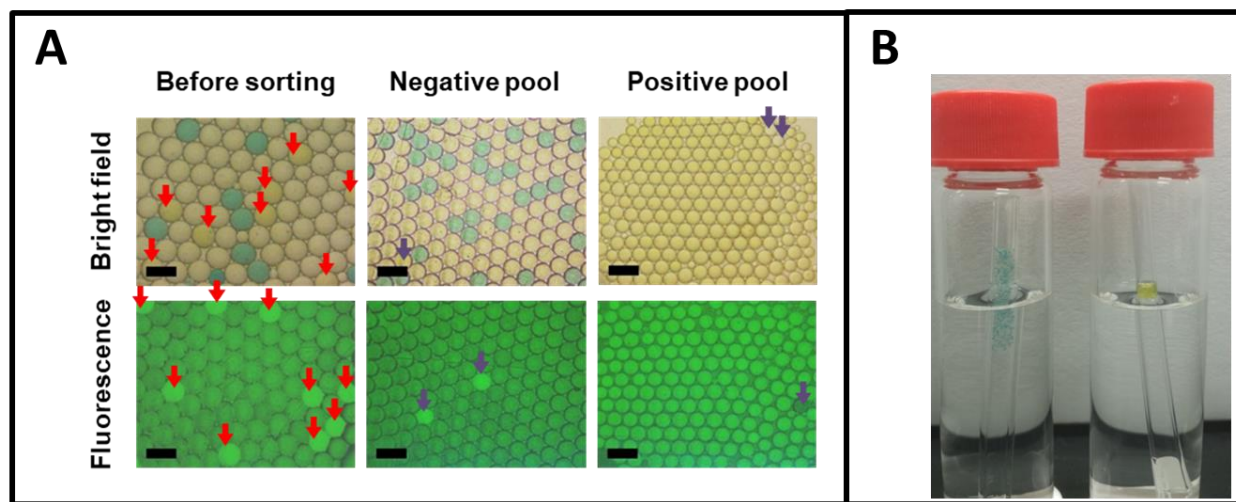


Figure 2-7: Imaged results from a representative droplet sorting run

(A) Brightfield and fluorescent Images of droplets before and after sorting in experiment E in table 2-1. Prior to sorting, approximately 11% of droplets contain the high concentration of pyridinyl amine along with the FAD that makes them visibly yellow. After sorting, more than 98% of the collected droplets exhibit the color and fluorescence associated with the target population. Scale bars are 500 μm . (B) Successful sorting is also clear when viewing the collected negative and positive pools of droplets macroscopically. Collected droplets are visibly separated into a yellow target pool and a blue and clear pool of mixed marker and non-target droplets.

droplet sorting in one of these experiments are shown in Figure 2-7. In all of these experiments, false positive rates in the collected droplets were <12%, and false negative rates were <5% .

We hypothesize that the majority of false negatives are collected during periods in which the microcontroller is in the non-synchronized state, directing all samples to waste, and that the majority of false positives occur when the microcontroller directing the sorting operation becomes misaligned. The microcontroller will synchronize as soon as it recognizes a marker to marker interval on the camera that matches an interval stored in the queue of droplet data from the MS signal. If it aligns to an incorrect interval, it will sort incorrectly until it recognizes a subsequent marker signal that does not match the expected interval, enters the non-synchronized state and attempts to realign. A minimum sync interval (in this work set to 5) may be chosen to avoid misalignment, as longer intervals will be more infrequent, and less likely to result in a misaligned interval. We observed that misalignment and non-synchronized periods were often preceded by false sorting events. Additionally, MADS operation is sensitive to miscounted events once it is aligned and sorting accurately. If the microcontroller counts an unexpectedly large or small number of droplets in a given interval during aligned operation, it will attempt to re-establish alignment. A “sync error limit” may be set to modify the stringency with which alignment signals must match after the MADS microcontroller is synced. In the case of this work, we set the sync error limit to 3 droplets, meaning that after three extra, or three too few droplets within an alignment interval, the microcontroller drops alignment and attempts to re-align.

2.4 Experimental - MADS in Complex Matrices and Reactions

Our results in the initial investigation of sorting efficiency demonstrated that MADS was capable of sorting mixtures of droplet samples based on their chemical contents with a high

degree of accuracy. However, these early tests were conducted with samples containing binary levels of the analyte of interest. We recognized that, in a typical screening scenario, these samples would neither be as simple as analyte-in-water, nor would hits be as discrete (a tenfold difference in concentration).

To test the feasibility of the MADS device for use in a directed evolution workflow, we aimed to sort a model library of droplet samples that contained *in vitro* expressed wildtype (WT) transaminase ATA117, from *Arthrobacter* sp. KNK168¹⁰⁶ (See 2.6, Transaminase Sequence). This transaminase is the same enzyme starting point that was previously evolved to produce the sitagliptin transaminase used in the production of the pharmaceutical Januvia.⁶⁶ The WT enzyme is fairly promiscuous, and can convert the non-native ATA substrate amine **1** that we used as a marker ion in our early experiments into ATA product ketone **2** (1-(imidazo[2,1-b]thiazol-6-yl)propan-2-one) when the reaction is run in the thermodynamically favored direction (Scheme 2-1). The transformation results in a net mass shift of 1 amu, and none of the substrates or products of the reaction are distinguishable optically in the reaction mix without chromatography.

2.4.1 Bulk ivTT Reactions

Microliter scale ivTT assays were carried out to assess the activity profile of enzymes expressed using commercial ivTT reagents combined with the substrates and cofactors of the enzyme targets. 25 μ L reactions were prepared by adding 100ng of purified plasmid DNA coding for the transaminase enzyme of interest to 25 μ L reactions of ivTT from New England Biolab's (NEB) PURExpress® In Vitro Protein Synthesis Kit. 10 μ L and 7.5 μ L respectively of solutions A and B (a proprietary mix of purified components of E.Coli protein translation) were mixed with 1.0 μ L of the RNase inhibitor murine (40000 units/mL, NEB) and 1.0 μ L of 2.5 mM Pyridoxal phosphate (PLP) in 10 mM Tris(Hydroxymethyl)aminomethane (TRIS) buffer (pH

7.5). DNA or PCR product and reaction substrates Pyruvate and 1-(imidazo[2,1-b]thiazol-6-yl) propan-2-amine (dissolved in 10mM TRIS buffer, pH 7.5) were added to the desired concentrations and 10 mM TRIS added to bring the reactions to volume. For time-course studies, this reaction mix was prepared in larger scale and 25 μ L samples were removed and quenched by freezing. Reactions were monitored over an 24 hour time period, with samples taken every hour for the first 4 h, and then at 6 h, 8h, and 24 h.

2.4.2 Plug-flow Mass Spectrometry for Assessment of Bulk ivTT Reactions

Bulk reaction analysis was performed using plug-flow infusion onto an Agilent single quadrupole mass spectrometer (6120B). Samples were directly infused into the mass spectrometer through a 360 μ m OD, 150 μ m ID Teflon capillary directly inserted into a commercial Agilent CE-MS source⁷⁷. Parallel sheath flow composed of deionized water was driven by a Harvard Apparatus PHD syringe pump and a 10 mL Hamilton gastight syringe at 50 μ L/min. Capillary voltage was 3 kV, with 350 °C drying gas set at 30 L/min and a source pressure of 10 psi. Single ion monitoring was used to track the small molecules of interest in the reaction solution. Peak height (MS dwell time) was set to 0.015s.

Sample plugs for infusion onto the MS were generated using a 500 μ L Hamilton gas tight syringe filled with Novec7500 oil and connected to 60 cm of 360 μ m OD, 150 μ m ID Teflon capillary. The syringe was placed into a Harvard Apparatus syringe pump set to withdraw mode. The pump was set to withdraw at 50 μ L/min and the Teflon capillary tubing inserted into a vial of oil. Once liquid began to flow into the tubing, it was serially dipped into the sample of interest and then the vial of oil, dwelling in each for 5-10s. This created a series of discreet sample plugs, which were then infused at 50 μ L/min into the MS for analysis.

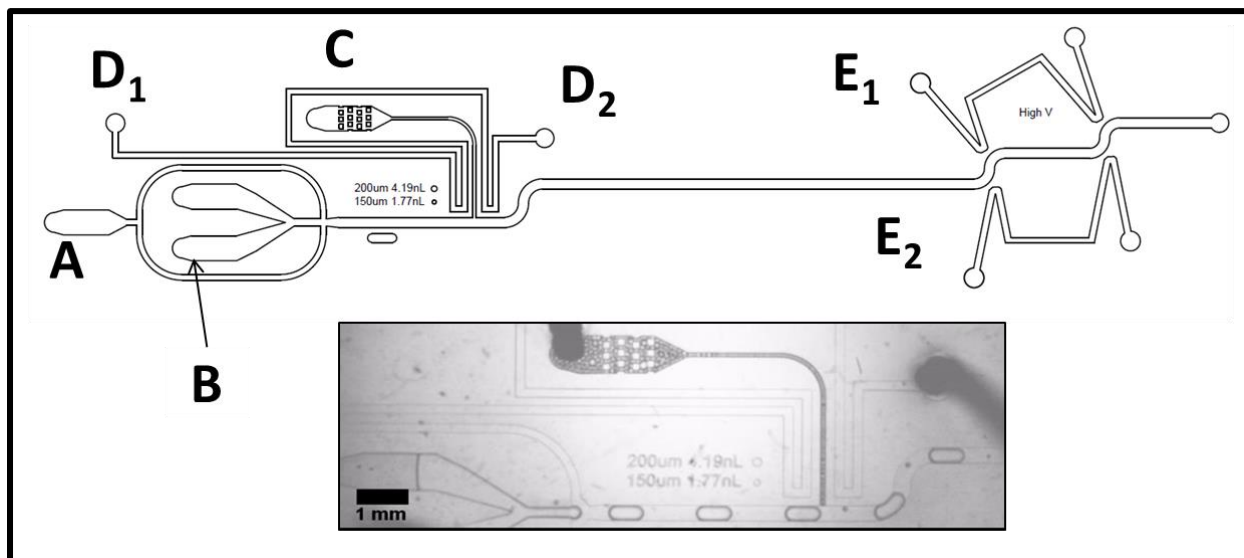


Figure 2-8: Schematic and micrograph of a droplet-pairing reagent addition device

A droplet pairing reagent addition device is shown with important regions labeled. An oil inlet (A) and an aqueous solution inlet (B) combine aqueous sample with fluorinated oil to produce droplets. These inlets are driven by a pressure pump which pressurizes vials of the respective liquids and drives them onto the chip through inserted tubing. A syringe pump drives 1 nL droplet reinjection by pumping oil into a glass storage chamber filled with the droplets, which are then driven into inlet (C) at a regular flow rate. Moat electrodes (D) are filled with salt water and electronically connected to ground to prevent any on-chip electric fields from merging these small 1 nL droplets as they run onto the Small droplets and large droplets pair up in the horizontal region between D₂ and E₁ (inset micrograph) before being merged in the high voltage electric field applied across channels E₁ and E₂.

2.4.3 Device Operation: Reagent Addition

Reagent addition is used to produce sample droplets with complex contents by mixing two different droplets together on a microfluidic device. We used this technique to express transaminase DNA *in vitro* within microfluidic droplets by adding 1 nL droplets of transaminase DNA to droplets of ivTT. This was performed on a droplet pairing device (Figure 2-8) and was driven using a combination of gas pressure and syringe pumps to drive fluid flow. To ensure 1 nL droplets flowed on to the device at a steady rate, a 500 μ L Hamilton gastight syringe was used to flow perfluorinated oil at 0.5-1.0 μ L/min into the bottom of the droplet storage chamber, driving droplets out through tubing inserted into the top. Meanwhile, the gas pressure pump was

used to pressurize two vials to 2 psi (13.8 kPa). One vial was used to hold the perfluorinated carrier phase for droplet generation while the other holds the aqueous sample used for droplet generation. The pressure on the vial of aqueous material was adjusted to create droplets of a pre-determined (30 nL) size at the flow focusing region of the device, while the syringe pump flow rate was adjusted to match the frequency of 1 nL droplet introduction to the frequency of droplet generation. Paired droplets flow downstream past two salt water¹⁰⁷ channels where an applied 1.5 kV AC field merges the two droplets. Exiting droplets flow directly into a second droplet storage chamber. This allowed droplets to be transferred from device to device without ever necessitating the use of a pipette which could shear large 30 nL samples.

2.4.4 ivTT Reaction Conditions for Signal Calibration

To make each dilution of ATA substrate amine in ivTT for calibration experiments, 250 μ L of cell free expression buffer was created from 100 μ L and 75 μ L respectively of solutions A and B (a proprietary mix of purified components of E.Coli protein translation), 10 μ L of the RNase inhibitor murine (40000 units/mL, NEB), 10 μ L of 2.5 mM Pyridoxal phosphate (PLP) in 10 mM TRIS (pH 7.5), and 1-(imidazo[2,1-b]thiazol-6-yl) propan-2-amine in nuclease free water such that seven final samples were created at 5 mM, 1 mM, 500 μ M, 250 μ M, 100 μ M, 50 μ M, and 0 μ M. Every other solution received 1 μ l of McCormick blue food dye to aid with visualization.

2.4.5 ivTT Reaction Conditions for Reagent Addition

To make the stock solutions of ivTT to which DNA was added in our enzyme assays, 1 mL of cell free expression buffer was created from 400 μ L and 300 μ L respectively of solutions A and B, 40 μ L of RNase inhibitor, 40 μ L of 2.5 mM PLP in 10 mM TRIS, 20 μ L of nuclease free water, and 200 μ L 5 mM of 1-(imidazo[2,1-b]thiazol-6-yl) propan-2-amine and 5 mM

pyruvic acid in 10 mM TRIS (pH 7.5). Final samples contained 1 mM amine **1**, 100 μ M PLP cofactor, and 1 mM pyruvate. An identical 250 μ L solution of expression buffer was created for the marker droplets with the addition of 2.5 μ L blue McCormick food dye and 2.5 μ L of 50 mM L-carnitine in water to the 50 μ L 5 mM amine **1** and pyruvic acid added.

1 nL droplets containing DNA (50 ng/ μ L final), FAD (5 mM final) and neostigmine (25 mM final) in water were formed using a commercial flow-focusing device (BioRad DG8). Upon addition to droplets of ivTT, these droplets express transaminase and are both visibly yellow and fluorescent. These small DNA droplets were pooled and mixed with blank water droplets and droplets containing 50 ng/ μ L WT plasmid and 50 mM chlorocholine. ivTT droplets that received these secondary DNA-containing droplets show activity, but not fluorescence and were used as a positive control signal for thresholding.

The combined pool of mixed 1 nL droplets contained all three varieties of droplet at a 1:1:6 ratio with the blanks making up the bulk. This mixed emulsion was reinjected into a reagent addition device that segmented ivTT solution into \sim 30 nL volumes before pairing these larger volumes with a small droplet and merging them in a 1.5 kV AC electric field (Figure 2-8)

2.5 Reading Enzyme Activity in ivTT Reaction Mix

In order to verify that we could observe the enzyme activity assay in droplets, the ivTT reaction was first performed in bulk and analyzed in microfluidic plug-flow format on the mass spectrometer. 100 ng/ μ L DNA was added to ivTT reaction mix at a 25:1 volume ratio and kept at room temperature on the benchtop. The reaction was quenched by freezing 20ul samples at regular time intervals. Droplet plugs were then formed from these samples in capillary using a syringe pump in withdraw mode as described in 2.4.2.^{22,59} Both the substrate and the product of the reaction could be monitored in this format. The time course experiment shows that the ATA

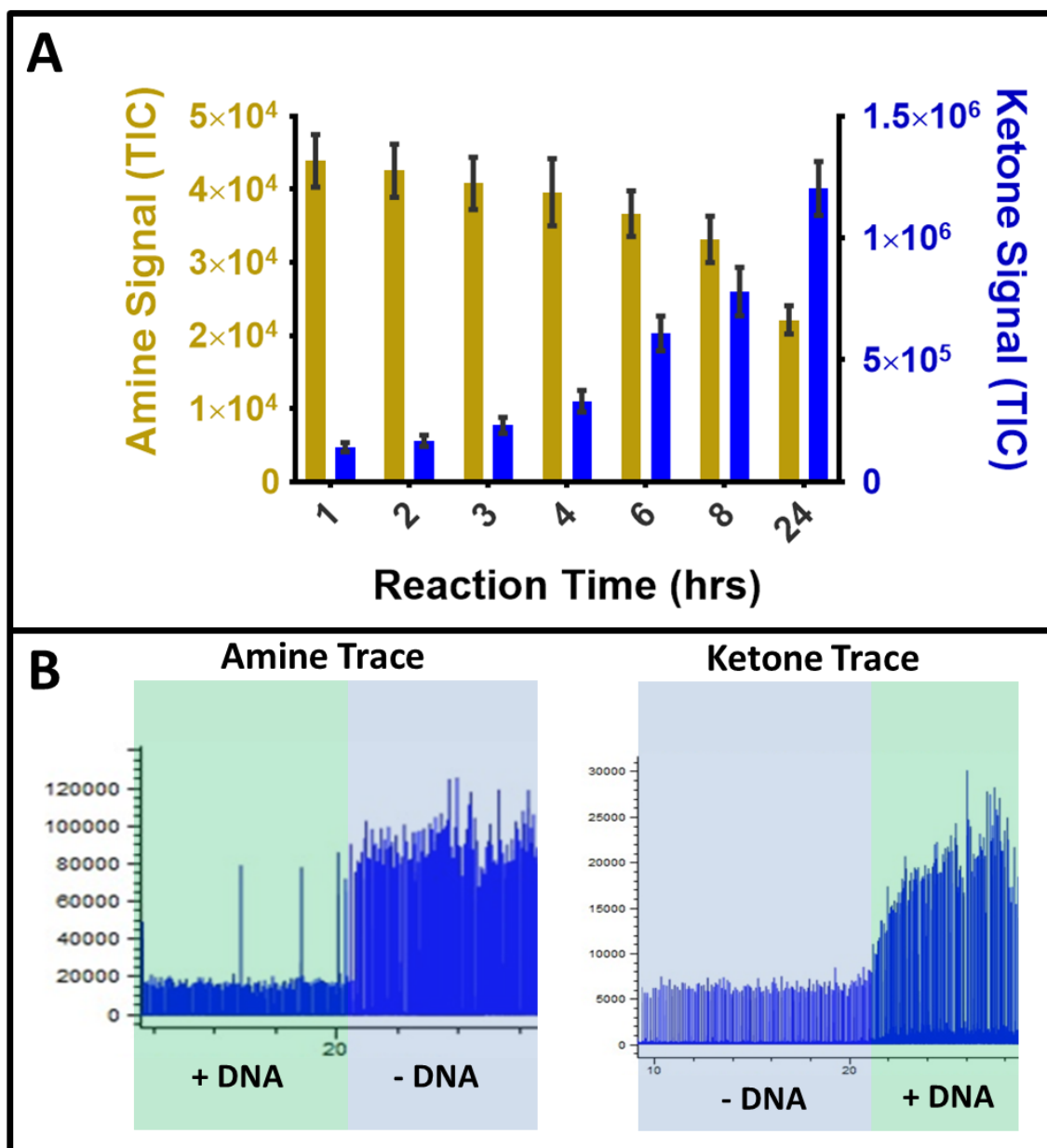


Figure 2-9: One-pot ivTT expression of transaminase analyzed in bulk and droplets

(A) *in-vitro* expression of the transaminase enzyme was performed by adding DNA to ivTT solution along with amine substrate for the transamination. This reaction was sampled periodically to determine if enzyme expression could be observed as a drop in the amine concentration and a corresponding rise in the formation of the ketone product. Plug-flow analysis of the transaminase reaction performed in bulk shows that the amine 1 signal decreases as the ketone 2 signal rises, demonstrating that the transamination occurs as DNA is expressed within the reaction. (B) When the fully reacted (24 h after DNA addition) and the unreacted bulk ivTT was then segmented into 30 nL droplets and infused using the MADS device to read samples, both the decrease in the amine signal and the increase in the ketone signal within 30 nL samples could still be observed. However, in this format it is evident that the ketone product is not as cleanly retained within the droplet samples as the Amine substrate is.

substrate concentration begins to decrease noticeably in the first few hours, and that it continues to drop over the course of a full day (Figure 2-9).

Following this confirmation, 30 nL droplets were produced from the bulk reaction mixes. The negative control reaction mixture that received no DNA and the reaction mixture that incubated for 24 h were compared by infusing these droplet samples onto the MADS device and reading the infused 30 nL droplets. Figure 2-9 B shows the ion traces observed by monitoring first amine **1** and then ketone **2** during the transition between the two samples. In this experiment, the transition between active and inactive enzyme samples is distinct when monitoring the amine substrate, but when monitoring the ketone product, the change appeared more gradual. This suggested that ketone **2** was moving between samples. This type of molecular transfer has been reported previously,^{94,95,108-110} and we were concerned that this would ultimately make active and inactive samples difficult to distinguish. Given this data, we elected to monitor reaction progress by observing the reduction of the ATA substrate (amine **1**), rather than the increase in the ketone **2** product.

2.5.1 MS Dynamic Range in Biological Matrix

To assess the ability of droplet MS to quantitatively distinguish between varied concentrations of amine **1** in complex ivTT matrices, droplets were made from ivTT doped with amine **1** at concentrations from 50 μ M to 5 mM. These bulk solutions were then segmented into 30 nL droplets and collected in layers in the same storage chamber. The droplet samples were pumped onto the MADS device one concentration at a time by drawing the chamber outlet tubing through each layer of stored droplets (Figure 2-10). Signal intensities for each sprayed concentration show clear step-change transitions between each layer of stored samples, with signal RSDs ranging from 6% to 9%. Signals were extracted for 225 peaks at each

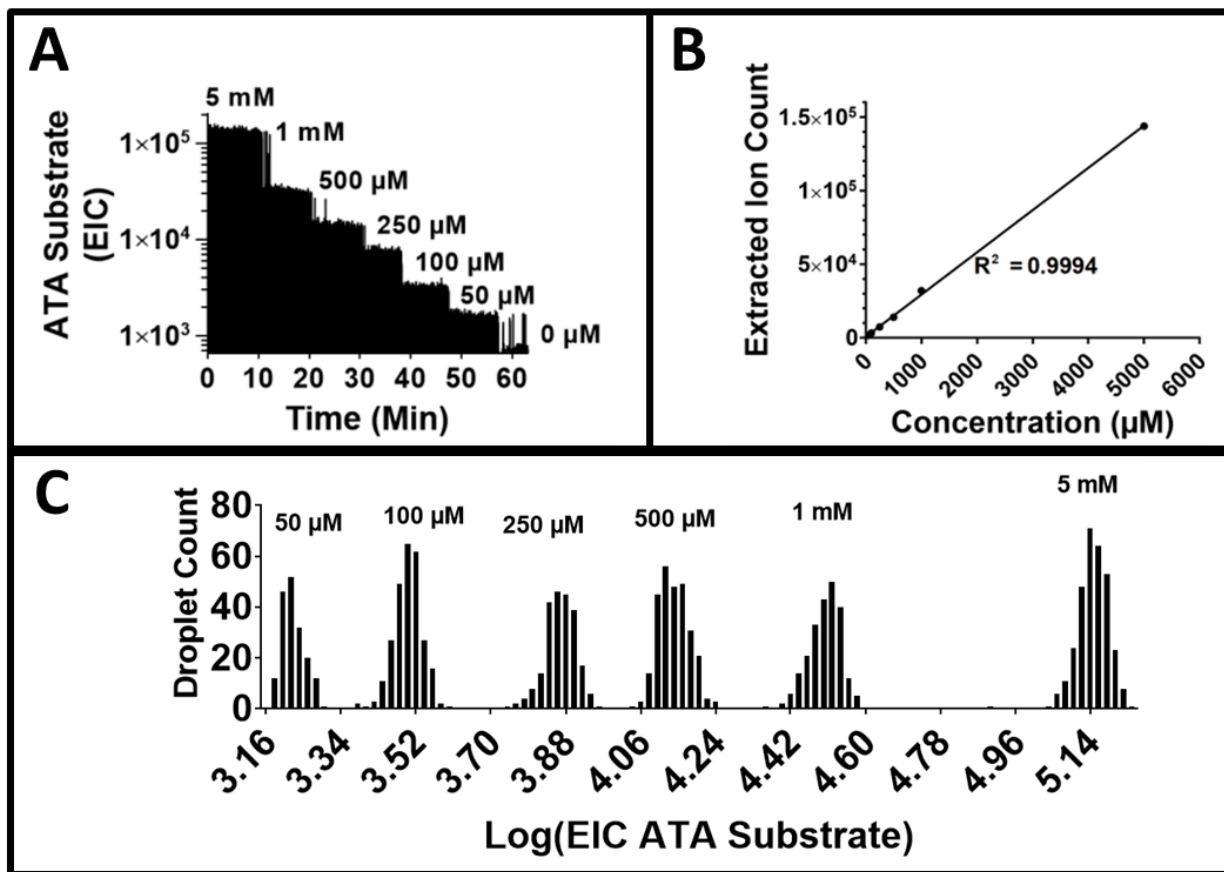


Figure 2-10: Concentration dependent signal intensity for droplets of ivTT with amine 1

(A) Six concentrations of amine **1** in droplets of ivTT are sprayed sequentially. Clear step changes are visible as each population is sprayed and (B) the averaged extracted ion count (EIC) from 225 peaks in each population show a linear response in signal with increasing substrate. Error bars show standard error of the mean peak height, which at these scales are all smaller than marker size. (C) The six populations of droplets are all clearly distinct and baseline resolved when unmixed, with none showing significantly wider spread relative to the log of signal intensity.

concentration, and plotted, showing a linear response to amine **1** concentration with a limit of detection of 30 μM. The logarithmic transformation of all extracted peak values was then taken, and the frequency distribution of each signal intensity range plotted, revealing clear clusters of sample intensities corresponding to the 6 sample concentrations.

To demonstrate the capability of the MS to distinguish between these droplets in a mixture, the droplets were mixed and sprayed after an hour of incubation. The raw trace and

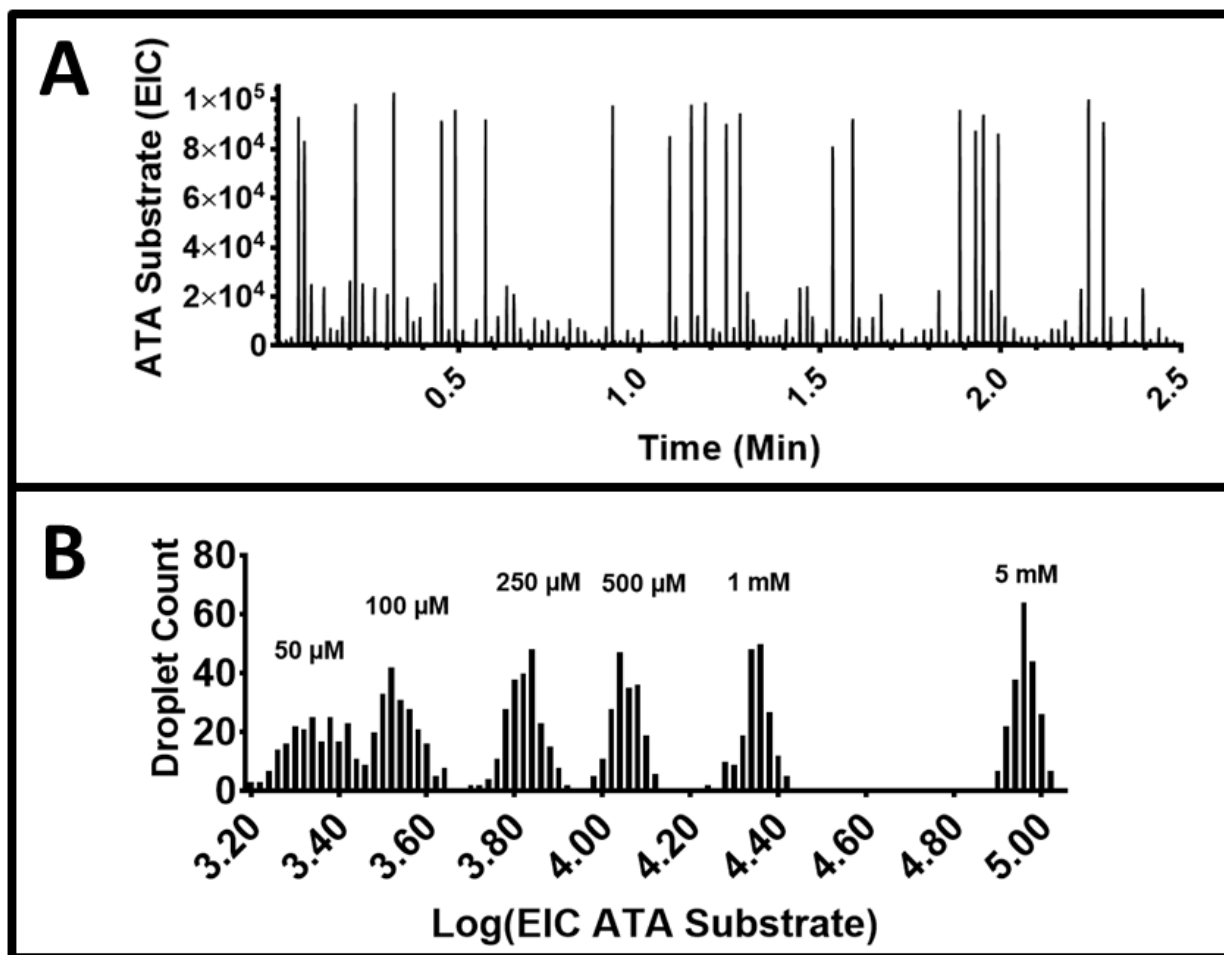


Figure 2-11: Demonstration of accurate assessment of mixed droplet concentrations

(A) A mixed population of droplets containing 6 concentrations of amine **1** in ivTT are sprayed after an hour of incubation. (B) Extracted peak heights from this mixture of droplets reveals the 6 component populations within the mixture. These histograms of droplet intensity appear less resolved from one another than their unmixed counterparts in Figure 2-11, likely due to some low-level transfer of amine **1** between droplets with higher and lower concentration.

binned ion intensity for the mixed droplets are shown in Figure 2-11. The histogram of droplet signal intensities shows resolution of each of the six underlying concentrations of ATA Substrate amine in the droplets. These histograms exhibit some drift towards the center when compared to those generated from the sequential spray experiment, possibly due to a small degree of substrate transfer between droplets. Nevertheless, this data may be used to calculate the Z' factor for a potential screen of samples starting at 1 mM and running to 95% conversion (50 μ M), giving a Z' of 0.757. This result demonstrates the ability of the MS to identify a wide range of analyte

concentrations in droplets containing complex, practical sample, and the potential to sort these populations based on MS signal.

2.5.2 Droplet MS Drift in Dynamic Reaction Screening

Signal drift can affect sorting accuracy if a fixed MS signal threshold is used as the sorting criteria. In our early experiments with DNA bearing ivTT droplets, we observed two forms of MS signal drift. First, enzymatic activity in droplets is not quenched at the initiation of analysis. Figure 2-9 illustrates the change over 24 h in substrate and product in a bulk reaction of expression and enzymatic turnover. The same occurs in droplets and can be seen as an increasing difference in target signal between inactive and active droplet signals over several hours of a screen (Figure 2-12).

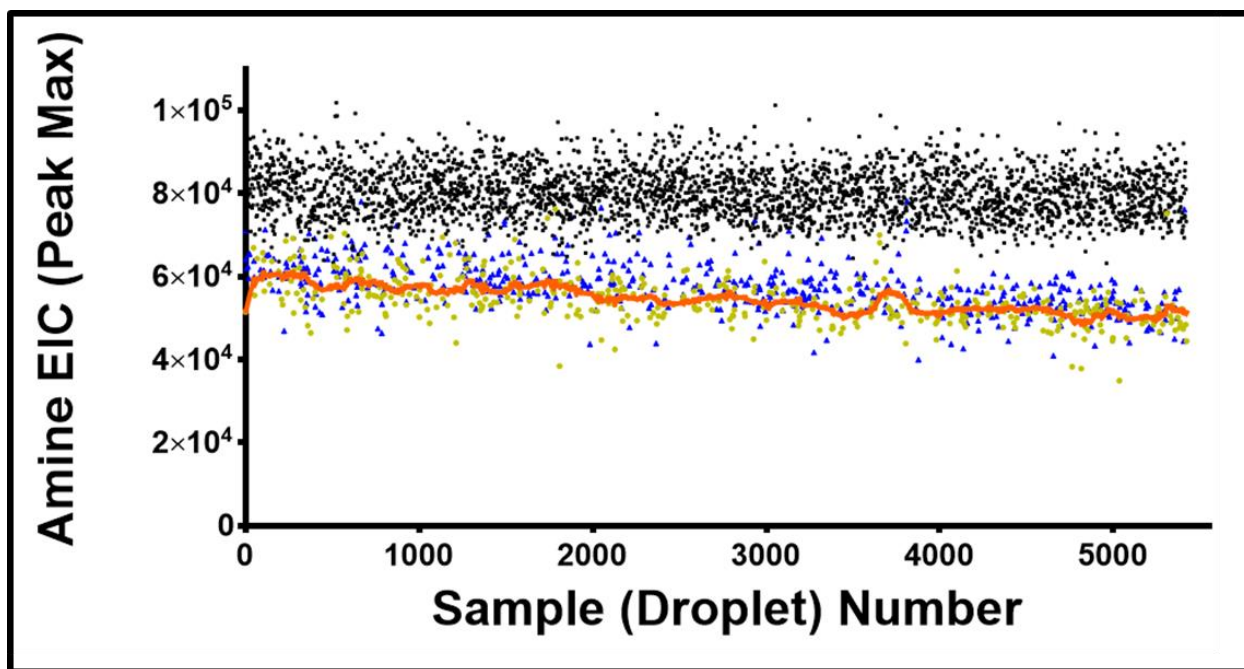


Figure 2-12: Demonstration of complete in-droplet expression and activity of ATA117

A 3 h trace of amine 1 signal in droplets expressing transaminase. A gradual increase in the separation between inactive (black circles) and active droplets containing DNA (blue and yellow circles) demonstrates the effect of an unquenched enzyme reaction in droplets. In this plot, active samples are divided into two populations: positive control (yellow circles) and model library (blue circles). The positive control signal can be averaged by the adaptive sorting algorithm (orange trace), and this moving average may be used to set sorting criteria for the model library.

Second, we also observed a more gradual drift in the maximum MS intensity of the ATA Substrate in our system, which declined by 15-20% over 6 h. Significantly, the signal for the marker ions remained stable during this same period, suggesting that the signal loss was specific to the ATA Substrate. We theorized that it stemmed from the slow transfer of ATA Substrate from the inactive droplets in the system to those expressing active protein. These two mechanisms for change in signal over time indicate the need to account for this signal drift during long analysis.

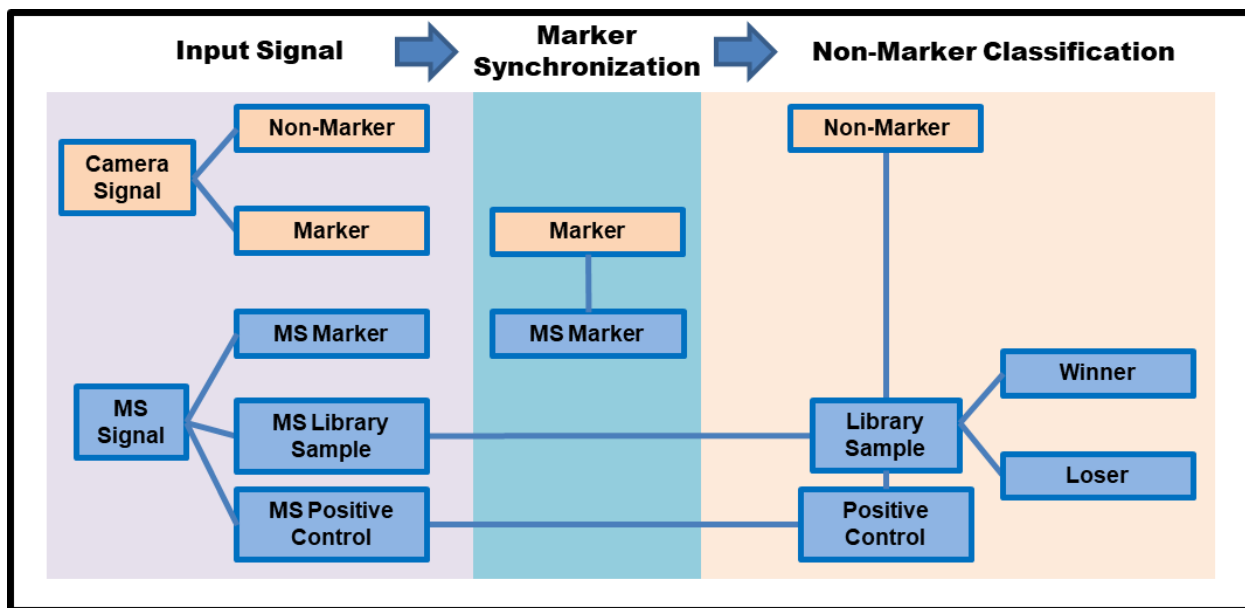


Figure 2-13: Decision tree detailing activity-based thresholding in droplet sorting

The MADS operating system under adaptive thresholding mode. Input signals are sent to the microcontroller controlling MADS sorting from both the camera at the sorting region and the mass spectrometer. The use of three separate MS channels for droplet identification allows for the categorization of droplets as alignment markers, library samples, and positive control samples. The alignment of the camera and the mass spectrometer is performed using marker droplets, allowing the synchronization of signals and enabling accurate sample tracking. Droplet “Hit” thresholding is based on the average signal from positive control samples rather than a user defined threshold set at the beginning of operation, and is only applied to the sorting of droplets categorized as sample droplets. This enables the system to adapt to changing concentrations of sample within the droplets over time.

To maintain accuracy over time, we developed an “adaptive thresholding” technique to change the threshold for selection as signal drift occurred (Figure 2-13). In this modified algorithm, hits are defined in comparison to the running average (16 samples) of a set of positive control signals mixed into the screened population. These positive control samples can be marked with a unique ion that distinguishes them from the sample population, allowing the algorithm to avoid sorting them as hits. We used the small molecules chlorocholine and neostigmine to mark the positive control samples and the model library samples respectively; these quaternary amines are robustly contained within droplets and readily ionize under ESI conditions.

In the adaptive mode of operation, the program will only attempt to collect samples marked by the presence of neostigmine. The threshold defaults to the average signal of the samples containing chlorocholine (Figure 2-12, orange line), but may be adjusted to be more or less stringent using a multiplier, termed the “sorting ratio”. Increasing the sorting ratio raises the threshold for a hit above the average positive signal, and allows a larger portion of the model library to be targeted.

2.6 Proof of Concept: Enzyme Screening

To demonstrate MADS on enzyme samples expressed in droplets, we created and screened a model library of transaminase enzymes by *in vitro* expression. For this experiment, three types of sample droplet were formed through the electrocoalescence of 1 nL droplets to ~25 nL droplets of *ivTT* reaction mix and ATA Substrate.¹¹¹ When the added 1 nL droplet contained WT DNA, neostigmine, and FAD, a visibly yellow, fluorescent droplet expressing transaminase was produced. These model library samples were targeted for sorting, and the added FAD made it possible to confirm the accuracy of the screen using fluorescent imaging. When the 1 nL

addition contained WT DNA and chlorocholine, an active positive control droplet was produced. 1 nL water droplets made up the bulk of the small droplets, and the addition of these produced inactive ivTT samples. Marker droplets for synchronization were generated from ivTT doped with carnitine and blue food dye and were mixed into the final pool of droplet samples.

The microcontroller utilized the running average of the amine **1** substrate signal in positive control droplets containing chlorocholine to adjust the sorting threshold. The neostigmine channel was used to identify model library members that would be considered for sorting. The carnitine channel was used to identify marker droplets.

Because all model library droplets were expressing the same WT DNA as the positive control droplets, the default threshold produced by the average wildtype signal only targeted approximately 50% of the active droplets. The sorting ratio was therefore set above 1 in experiments that aimed to recapture all of the model library droplets from the mixed pool. Droplets that showed amine **1** signal below this threshold were targeted for sorting, and those that did not were rejected and allowed to flow to waste. In a screening scenario, this cutoff will likely be set to more stringent values to select only for high performance hits.

After 3-4 h incubation to give the enzyme time to react, the droplets were reinjected and sorted, with samples taken and imaged at 1 or 2 h intervals. A representative trace of the mass signal for the droplets is shown in Figure 2-14. Model library droplets show a reduction in ATA Substrate, a high neostigmine signal, and are fluorescent. Positive control droplets show a similar reduction in ATA Substrate but display high chlorocholine signal, and are therefore not targeted for sorting. Marker droplets and droplets that did not receive an addition containing DNA remain unchanged in their ATA Substrate amine concentration.

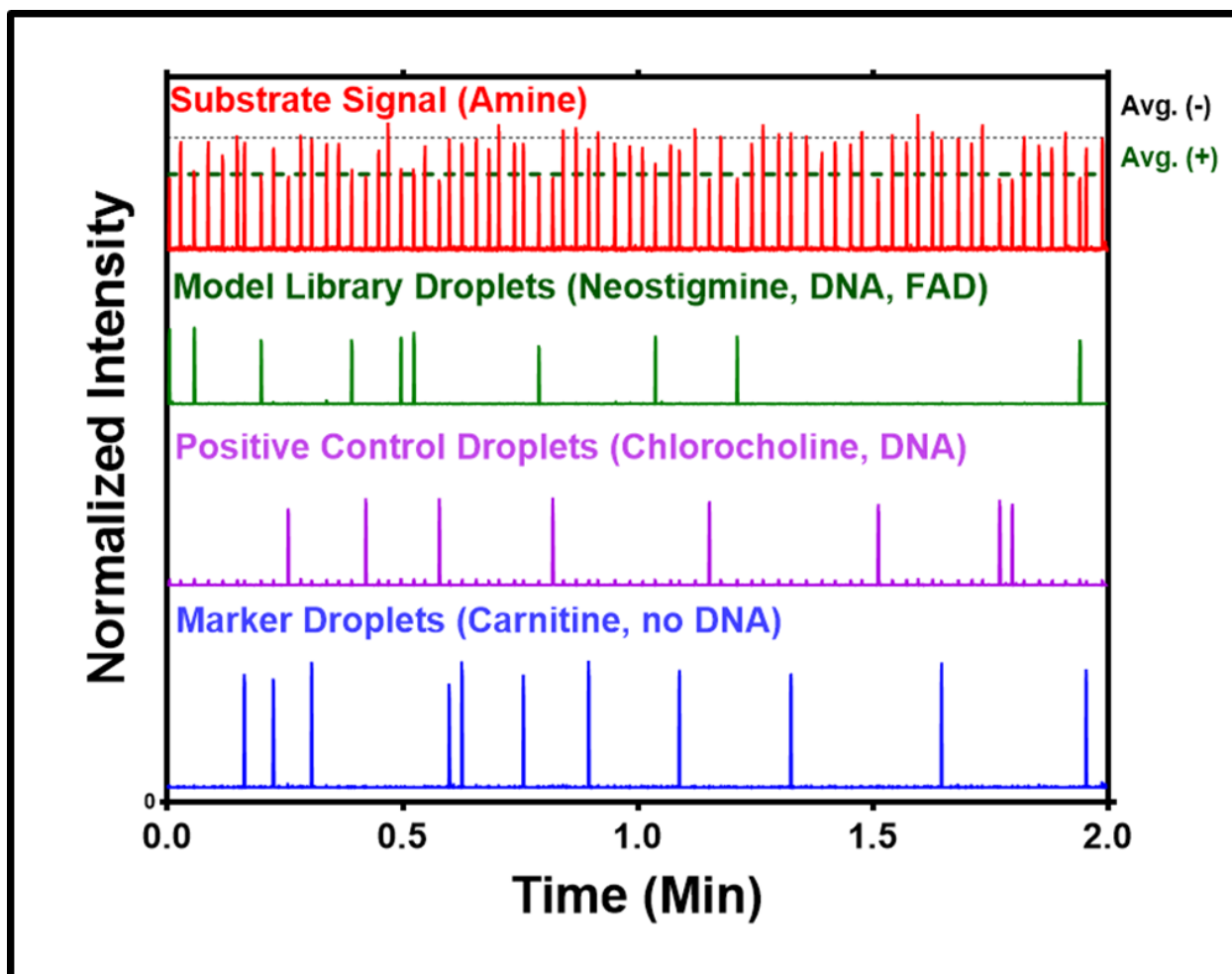


Figure 2-14: Representative signal trace for activity based thresholding in MADS
 A mixture of 1 nL droplets that occasionally contained plasmid DNA with MS traceable analytes are added to 30 nL droplets to create a mixed pool of large droplets where some express active protein. The droplets are sprayed, and the small molecule tracers neostigmine and chlorocholine allow the MADS microcontroller to identify those that have received and expressed this DNA. A modified sorting algorithm reads the resultant trace and uses the average amine signal from the droplets containing chlorocholine (purple trace) to set the sorting threshold for those containing neostigmine (green trace). Average signals for the positive control, shown as Avg.(+), and average signals from negative samples, shown as Avg.(-), are marked to highlight how all droplets containing these two marker signals exhibit the expected drop in Amine signal.

In all experiments with a sorting ratio above 1.3, the collected droplet pools showed enrichment above 90%, from a starting occupancy of approximately 10%. Collected droplets were imaged and analyzed, and Table 2-2 shows the results from 5 experiments with a sorting ratio of 1.45 (targeting all droplets with wildtype activity and neostigmine), where 98% of the

Experiment	Before Sorting			After Sorting	
	Samples Screened	Hit droplets in initial pool (%)	Marker droplets (%)	Hit droplets in positive pool (%)	Hit droplets in negative pool (%)
Experiment A	4292	9.9	18.8	93.0	3.1
Experiment B	2340	11.1	10.1	91.0	5.3
Experiment C	2637	9.8	17.2	97.5	3.1
Experiment D	2562	10.5	20.1	94.5	4.8
Experiment E	2067	10.3	20.4	90.5	3.5

Table 2-2: Sorting results for MADs proof of concept, with activity-based thresholding

Sorting results from experimental runs of the mass activated droplet sorting device after ivTT expression of DNA in droplets. Sorted droplets contained either no DNA and therefore no active enzyme to consume the loaded amine **1**, or DNA and ivTT expressed active enzyme. Half of those containing DNA received chlorocholine with DNA during reagent addition, while half received neostigmine and FAD. Those containing neostigmine were sorted based on the average signal of those containing chlorocholine. Target droplets are enriched 8-9 fold higher in the final pool relative to the starting pool.

droplets dosed with DNA, FAD and Neostigmine were targeted for sorting. Collected droplets were imaged after sorting, and confirmed as winners using the fluorescence of the FAD to evaluate accurate sorting (Figure 2-15). Accuracy was found to be on par with our previous experiments using aqueous solutions of known analyte concentration.

The results of this experiment show the capability of the system to perform multiplex analysis in a complex reaction mixture and use information from an ongoing chemical reaction to accurately sort a target population.

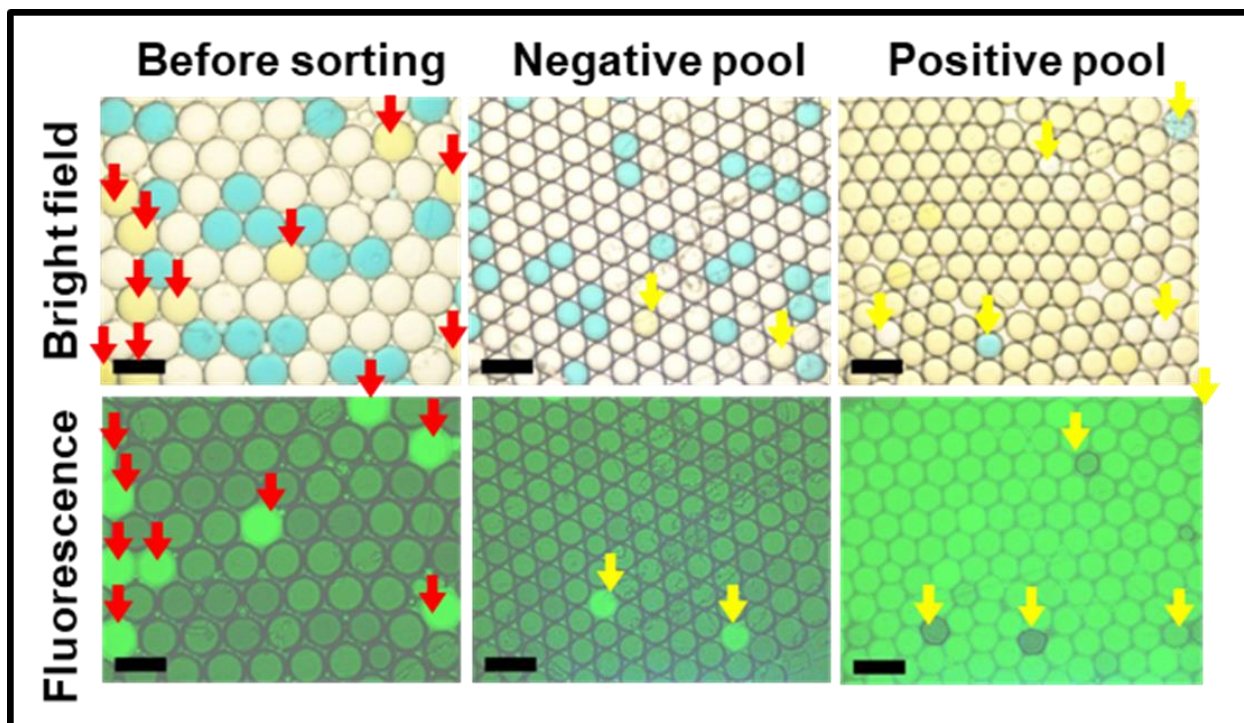


Figure 2-15: Droplet expression and sorting results

(A) Brightfield and fluorescent Images of droplets before and after sorting in a representative experiment from table 2-2. Droplets expressing transaminase DNA (yellow) are sorted based on the presence of active enzyme using the adaptive thresholding technique. These are highlighted with red arrows prior to sorting. False positives and negatives in the collected pools are highlighted with yellow arrows in the images of droplets after sorting. Scale bars are 500 μm .

2.7 Conclusions

Although FADS has been the most utilized approach for sorting droplet samples since its introduction⁷³ and has enabled sorting at throughputs unachievable in conventional systems, indirect assays often lead to off-target selection.⁹² A screen that directly interrogates the analyte of interest is far preferable. With careful assay development, both FADS and AADS have been used to screen industrially relevant enzymes,^{32,93} but the vast majority of potential targets will not be amenable to such reporter assays.

Here, we have demonstrated an adaptive, self-correcting MADS system with the ability to screen complex biological reactions. This work links a versatile, label-free analytical tool to

microfluidics workflows and significantly broadens the applicability of these miniaturized systems.

The inherent label free nature of MS makes MADS a valuable tool for probing chemistries that cannot be easily adapted to FADS, broadening detection to include most molecules that are readily ionizable with ESI. While it does not achieve the full information density of HPLC-MS analysis, the resolution of direct MS allows numerous analytes to be simultaneously monitored, creating the potential for selective, multifaceted probes of activity. For screening chemical reactions, the ability to detect changes in mass is more broadly applicable than changes in spectra. Although MADS sorting rates are 1000-fold slower than those demonstrated with FADS, they are still 100-fold faster than industry standard HPLC-MS methods. The higher detection limits of MADS as demonstrated are on par with absorbance screens. This should be sufficient for most applications, but could be improved by adapting the method to a triple quadrupole mass spectrometer if desired.

2.8 Supplemental Information – Transaminase Sequence

```
tcgcgcgtttcggatgatgacggtgaaaacctctgacacatgcagctcccggagacggtcacagcttgtctgtaagcggatgccg
ggagcagacaagcccgtcagggcgcgtcagcgggtgtggcgggtgtcggggctggcttaactatgcggcatcagagcagattgtactg
agagtgcacatgatgcgggtgtgaaataccgcacagatgcgtaaggagaaaataccgcatcaggcgcattcgcattcaggctgcgcaac
tgttgggaagggcgatcgggtgcgggcctcttcgctattaccagctggcgaaagggggatgtgctgcaaggcgattaagttgggtaacg
ccagggtttcccagtcacgacgttgtaaacgacggccagtgaattgacgcgtattgggatcaaggaatggtgcatgcaaggagctccgt
agatagtcaccaggcacagcgaattaatacactcactataggagaccacaacccgcatatacggcgggacacacacaaggagac
catATGAAACAGAACAAAGAACAGATCACCAAACAGAACATCAACGCGAGCTCTGCA
TTCAGCGCAGATACTAGCGAGATTGTTTACACGCACGATACGGGTCTGGACTACATT
```

ACCTATAGCGACTATGAACTGGACCCGGCGAATCCGCTGGCTGGCGGCGCGGCCTG
GATCGAGGGTGCATTCGTCCC GCCGTCCGAGGCGCGTATTTCTATCTTTGACCAGGG
TTATCTGCACAGCGACGTGACCTATACCGTGTTCCACGTCTGGAACGGTAATGCCTT
TCGTCTGGATGATCATATTGAGCGTCTGTTCTCTAACGCAGAGAGCATGCGCATTAT
CCCTCCGCTGACCCAAGATGAAGTGAAAGAAATTGCGCTGGAAC TGGTGGCCAAGA
CCGAGTTGCGTGAGGCTTTCGTTAGCGTTAGCATCACCCGCGGTTACAGCAGCACGC
CGGGTGAGCGTGACATCACGAAGCACCGTCCACAAGTATAACATGTACGCAGTGCCG
TATCAGTGGATCGTGCCATTTGATCGCATTTCGTGATGGTGTCCACGCGATGGTTGCA
CAGTCCGTGCGCCGCACCCCTCGCAGCAGCATTGACCCGCAAGTCAAAA ACTTCCA
GTGGGGTGACTTGATCCGTGCGGTGCAAGAAACCCATGATCGTGGTTTTGAGGCGCC
GCTGCTTCTGGACGGTGACGGCCTGCTGGCGGAAGGCTCGGGCTTCAATGTCGTCGT
TATCAAGGATGGCGTTGTGCGCAGCCCGGGTCGTGCCGCGCTGCCGGGCATCACGC
GTAAAACCGTCCTGGAAATTGCTGAGAGCCTGGGCCATGAAGCTATTTTGGCGGAC
ATTACTCTGGCGGAACTGCTGGATGCGGATGAAGTTCTGGGTTGCACGACCGCCGGT
GGCGTTTGGCCGTTTGTAGCGTCGACGGTAATCCGATCAGCGACGGTGTCCCGGGC
CCGATCACGCAGTCAATTATCCGTCGTTACTGGGAACTGAACGTTGAGTCTAGCTCC
TTGCTGACCCCGGTGCAGTACTAATAActcgaggatccggctgctaacaagcccgaaggaagctgagttggc
tgctgccaccgctgagcaataactagcataacccttggggcctctaaacgggtcttgaggggttttggctgaaaggaggaactatatccgg
atatccacaggacgtctgtcatgcactagtctgtcgtctctgtagcggcgattaagcatcccaatggcgcgccgagcttgctcgagcatg
gtcatagctgttctgtgtgaaattgttatccgctcacaattccacacaacatacagccggaagcataaagtgtaaagcctgggggtgcctaa
tgagtgagctaaactacattaattgcgttgcgctcactgcccgtttccagtcgggaaacctgtcgtgccagctgcattaatgaatcgccaac
gcgcggggagaggcggttgcgtattgggcgctctccgcttctcgtcactgactcgtcgcgctcggctgttcggctgcggcgagcgg
atcagctcactcaaagcggtataacggttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaagg

ccaggaaccgtaaaaaggccgctgtgctggcggtttccataggtccgccccctgacgagcatcacaaaaatcgacgctcaagtcagag
gtggcgaaacccgacaggactataaagataccaggcggttccccctggaagctccctcgtgctctctctgtccgacctgccgcttaccg
gatacctgtccgctttctcccttcgggaagcgtggcgctttctcatagctcacgctgtaggtatctcagttcgggtgtaggtcgttcgctcaag
ctgggctgtgtgcacgaacccccgttcagcccagccgctgcgcttatccgtaactatcgtcttgagtccaacccggtaagacacgactt
atgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcgggtctacagagttctgaagtgggtggcctaactac
ggctacactagaagaacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaaaa
accaccgctggtagcgggtggtttttgttgcaagcagcagattacgcgcagaaaaaaggatctcaagaagatccttgatctttctacgg
ggctgacgctcagtggaacgaaaactcacgtaagggttttggtcatgagattatcaaaaaggatcttcactagatccttttaattaaaa
tgaagtttaaatcaatctaaagtatatatgagtaaacttggtctgacagttagaaaaactcatcgagcatcaaatgaaactgcaatttattcatat
caggattatcaataccatattttgaaaaagccgtttctgtaatgaaggagaaaactcaccgagcagttccataggatggcaagatcctggta
tcggctctgcgattccgactcgtccaacatcaatacaacctattaatttcccctcgtcaaaaaataaggttatcaagtgagaaatcacatgagtgga
cgactgaatccgggtgagaatggcaaaagttatgcatttcttccagactgttcaacaggccagccattacgctcgtcatcaaaatcactcgc
atcaaccaaaccgttattcattcgtgattgcgctgagcgagacgaaatacgcgatcgtgttaaaaggacaattacaacaggaatcgaat
gcaaccggcgcaggaactgccagcgcatacaaatatttccactgaatcaggatattcttctaataacctggaatgctgtttcccagggat
cgcagtggtgagtaacctgcatcatcaggagtacggataaaatgcttgatggcggaaaggcataaattccgtcagccagtttagtctga
ccatctcatctgtaacatcattggcaacgctacctttgccatgttccagaaacaactctggcgcacgggcttccatacaatc gatagattgct
gcacctgattgcccacattatcgcgagccatttataccatataaatcagcatccatggtggaatttaacgcggcctagagcaagacgttt
cccgtgaaataggctcactcttcttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatattgaaatgatttagaa
aaataaacaataggggttccgcgcacatttccccgaaaagtccacctgacgtctaagaaaccattattatcatgacattaacctataaaaaat
aggcgtatcacgaggcccttctc

Chapter 3: Chapter 3 Integrated Nanoliter Sample Processing and Mass Spectrometry for the Expression and Selection of a Transaminase Library

3.1 Introduction

Enzyme-mediated catalysis is a powerful tool in the arsenal of modern synthetic chemistry. The enhanced selectivity, efficiency, and safety of biocatalytic transformations make enzyme catalysis an attractive alternative to transition metal-based synthetic methods.^{64,66,67,69,70,112} The application of chemoenzymatic and multi-enzymatic processes in pharmaceutical synthesis has driven both the rapid preparation of lead compound libraries and the sustainable production of active pharmaceutical ingredients.⁶⁶ For example, Merck's design of an innovative biocatalytic cascade for the manufacture of islatravir⁶⁹ highlights the industrial utility of enzymes as robust catalysts for accelerating pharmaceutical syntheses, and the potential to use re-purposed enzyme cascades to affect complex, multi-step transformations on non-native chemical substrates. However, in order to optimize enzymes for enabling this stepwise transformation of a novel substrate, intensive engineering and screening programs must be undertaken to identify biocatalysts with optimized activity on non-native building blocks for each step along the way. In this process of directed evolution, tens of thousands of unique enzymes must be expressed and assayed to select the top candidates with the desired properties.

In the search for novel active enzymes for biocatalytic applications, the use of robust, high-throughput screening and detection methods is of paramount importance.^{71,92,113} Current strategies rely on a combination of chromatographic separation and mass spectrometry or UV detection, processes that possess analytical throughputs of several minutes per sample. Assay development and setup also contribute to limited throughputs; library preparation, cell growth,

protein induction, cell lysis, and chemical assays must all be carried out stepwise in well plate format. With evolution programs screening thousands of enzymes per round, current directed evolution methods that are assessed chromatographically necessitate weeks of screening work per round, which translates to a timescale of months to years needed to develop a robust enzyme catalyst.

Since sample production and screening present critical bottlenecks to protein engineering efforts, the development of rapid analytical strategies and sample processing workflows is of utmost import. Flow cytometry⁵ and microfluidic droplet sorting systems^{29,32,34,37,38,40,73,75,89,114,115} have enabled rapid assessment of enzyme libraries reaching 10⁶ members, but have relied almost exclusively on photometric detection and fluorescent labeling, which are often challenging to implement in small molecule transformations. Systems that combine nanoliter droplet sampling with direct mass spectrometry have been explored as label free analytical methods to both reduce sample consumption and improve throughput for protein engineering, synthetic method development, and drug discovery workflows.^{47,77,100,101,116} We have recently demonstrated the coupling of droplet microfluidics with electrospray MS for the label-free assessment of up to 2,500 nanoliter samples per hour using mass activated droplet sorting (MADS).⁸⁰

3.1.1 Transaminases as Biocatalytic Targets

In the previous chapter, nanoliter droplet mass spectrometry and microfluidic droplet sorting were combined to demonstrate mass activated droplet sorting (MADS) of 30 nL microfluidic droplets. *in vitro* expression of a transaminase enzyme was performed within 30 nL droplets using a reagent addition device, and assessed these samples with the MADS platform to demonstrate its analytical and microfluidic sorting capabilities with a biological reaction in a

droplet. Transaminase enzymes are valuable biosynthetic tools for the production of enantiomerically pure chiral amines that are often used as intermediates in the synthesis of complex, pharmaceutically important organic molecules.¹¹⁷ Transaminases catalyze the exchange of an amine and a ketone functional group between two molecules with these moieties, a catalytic process that enables cells to form amino acids from the corresponding keto-acid *in vivo*. Transaminases that accept aliphatic ketones and amines as their substrates are attractive targets for directed evolution efforts.¹¹⁸

Most transaminases perform the synthesis of enantiomerically pure amines through the use of the coenzyme pyridoxal-5'-phosphate (PLP). In the transaminase active site, PLP readily forms a Schiff base with amines, which allows it to be covalently bound to lysine in the active site of transaminase enzymes (Figure 3-1). Nucleophilic addition of the amine substrate onto this PLP-bound imine displaces the lysine and forms an imine intermediate with the substrate amine, which then undergoes proton rearrangement and hydrolysis to release the ketone product, leaving the enzyme-associated PLP aminated as pyridoxamine phosphate (PMP). In the second half of this “ping pong” reaction, the PLP cofactor is regenerated from PMP. Here, a ketone donor undergoes attack by PMP and subsequent amination to generate and release an amine. Enantiospecificity in this reaction stems from favored coordination of one enantiomer over the other during the protonation of the imine by the catalytic lysine in the active site.

Transaminases have seen notable adaptation to industrial processes and the development of higher throughput methods for their evolution is of particular interest. In this work we aimed to further utilize microfluidic droplet mass spectrometry and establish a workflow that would (i) reduce reagent consumption and automate reaction setup by expressing a library of ~2000

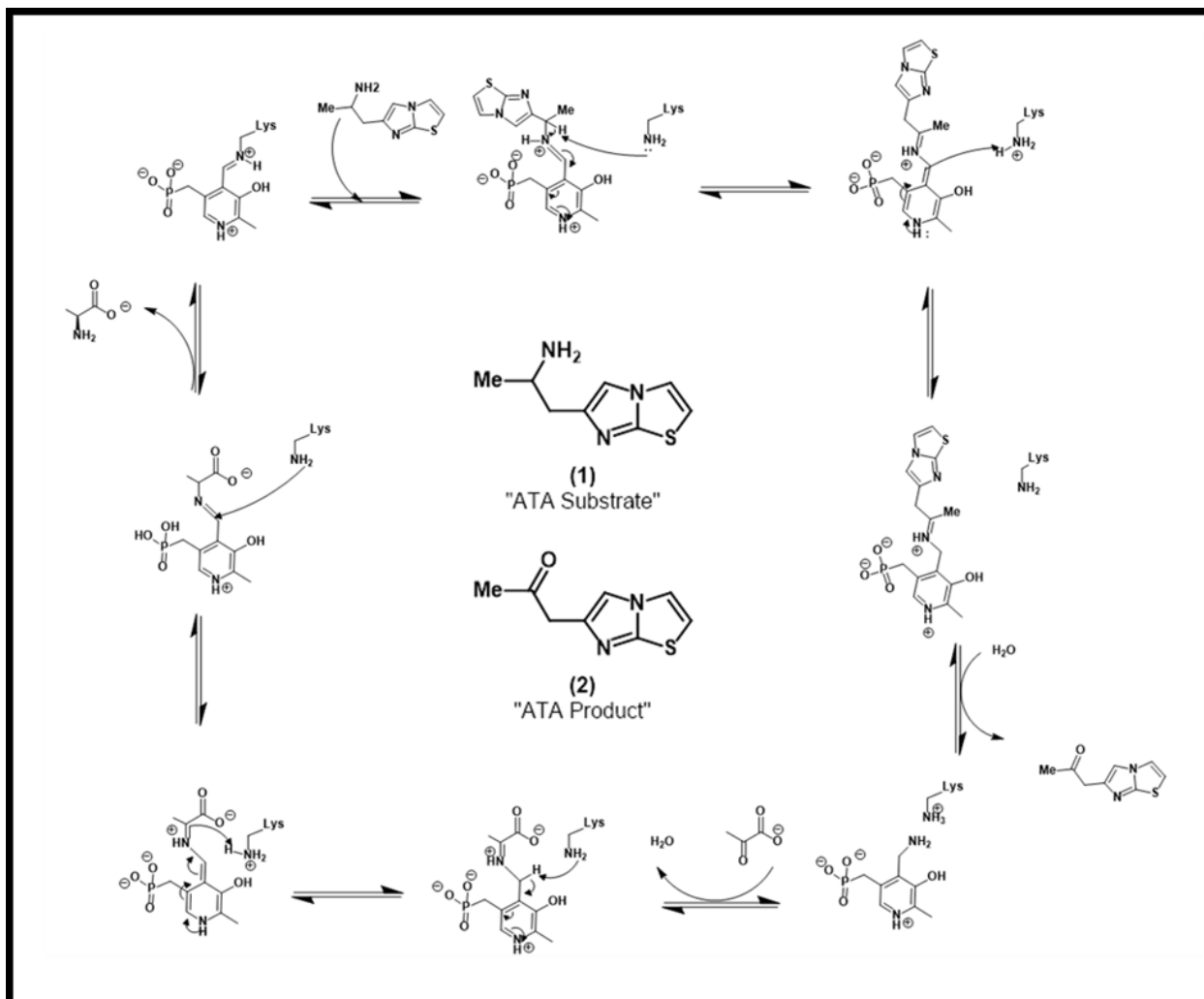


Figure 3-1: Mechanism for the ATA117-mediated transformation of an amine

Enzyme mechanism for the transformation of non-native substrate 1 into product 2 by the wildtype transaminase ATA117. This ping-pong two substrate reaction is catalyzed by the cofactor PLP and simultaneously converts pyruvate to alanine. Here it is shown using 1-(imidazo[2,1-b]thiazol-6-yl)propan-2-amine, or the “ATA Substrate” pictured here as the amine donor in the transformation of Pyruvate to Alanine.

different mutations on a parent transaminase enzyme expressed in individual microdroplets, (ii) analyze the relative activity of the expressed enzymes using droplet MS, and (iii) accurately recover and identify potential catalysts that outperformed the parent enzyme from which they were derived and to which they were compared.

3.2 Experimental Methods

Protein engineering and directed evolution is carried out in a stepwise process in which a library of enzymes is expressed, isolated, and assayed to determine the individual activity of each library member. The best variants are then sequenced, enabling the identification of specific mutations that improve the catalytic efficiency, selectivity, and specificity of the enzyme target. These mutations may then be combined or further iterated upon to drive the enzyme target towards the ideal enzyme for the transformation at hand. In order to express and screen a library of enzymes in microdroplets, we developed a microfluidics based workflow to perform each of the core steps of the protein engineering and directed evolution screening workflows within microdroplets.

3.2.1 Bulk ivTT Reactions

Microliter scale in vitro transcription and translation assays were carried out to assess the activity of enzymes expressed using commercial ivTT reagents combined with the substrates and cofactors of the enzyme targets. 25 μ L reactions were prepared by adding 100 ng of purified linear DNA coding for the enzyme of interest to 25 μ L reactions of ivTT from New England Biolab's PURExpress® In Vitro Protein Synthesis Kit. 10 μ L and 7.5 μ L respectively of solutions A and B (a proprietary mix of purified components of E.Coli protein translation) were mixed with 1.0 μ L of the RNase inhibitor murine (40000 units/mL, NEB) and 1.0 μ L of 2.5 mM Pyridoxal phosphate in 10 mM TRIS (pH 7.5). DNA or PCR product and reaction substrates Pyruvate and 1-(imidazo[2,1-b]thiazol-6-yl) propan-2-amine (dissolved in 10mM TRIS buffer, pH 7.5) were added to the desired concentrations and 10 mM TRIS added to bring the reactions to volume. For time-course studies, this reaction mix was prepared in larger scale and 25 μ L

samples were removed and quenched by the addition of 1 μL of 10% formic acid. Reactions were monitored over an 8 hour time period, with samples taken every hour.

3.2.2 Plug-flow Mass Spectrometry for Assessment of Bulk ivTT Reactions

Bulk reaction analysis was performed using plug-flow infusion onto an Agilent single quadrupole mass spectrometer (6120B). Samples were directly infused into the mass spectrometer through a 360 μm OD, 150 μm ID Teflon capillary directly inserted into a commercial Agilent CE-MS source⁷⁷. Parallel sheath flow composed of deionized water was driven by a Harvard Apparatus PHD syringe pump and a 10 mL Hamilton gastight syringe at 50 $\mu\text{L}/\text{min}$. Capillary voltage was 3 kV, with 350 $^{\circ}\text{C}$ drying gas set at 30 L/min and a source pressure of 10 psi. Single ion monitoring was used to track the small molecules of interest in the reaction solution. Peak height (MS dwell time) was set to 0.015s.

Sample plugs for infusion onto the MS were generated using a 500 μL Hamilton gas tight syringe filled with Novec7500 oil and connected to 60 cm of 360 μm OD, 150 μm ID Teflon capillary. The syringe was placed into a Harvard Apparatus syringe pump set to withdraw mode. The pump was set to withdraw at 50 $\mu\text{L}/\text{min}$ and the Teflon capillary tubing inserted into a vial of oil. Once liquid began to flow into the tubing, it was serially dipped into the sample of interest and then the vial of oil, dwelling in each for 5-10s. This created a series of discreet sample plugs, which were then infused at 50 $\mu\text{L}/\text{min}$ into the MS for analysis.

3.2.3 PCR Amplification of Transaminase DNA

DNA sequences were amplified from template plasmid or linear DNA using New England Biolab's Q5® High-Fidelity 2X Master Mix. Each 20 μL PCR reaction was created with 10 μL of master mix (containing polymerase, dNTPs, and the necessary buffer components for PCR), forward and reverse primers (100 nM final concentration in the reaction), and DNA

Method Name	Method {(temp,time) repeat#}	Purpose
Q5_Droplet	(98,1:00){(98,0:10)(65,0:20)(72,0:55)40x}(72,2:00)(20,hold)	Digital droplet PCR
Q5_Bulk	(98,0:30){(98,0:10)(65,0:20)(72,0:55)29x}(72,2:00)(4,hold)	Linear DNA generation, DNA recovery
Colony_PCR	(94,5:00){(94,0:15)(60,0:20)(68,1:30)29x}(68,5:00)(12,hold)	Colony sequencing
Q5_SSM	(98,1:00){(98,0:10)(65,0:20)(72,0:45)24x}(72,2:00)(4,hold)	Library fragment generation
SOE_PCR	(98,1:00){(98,0:10)(65,0:20)(72,0:45)24x}(72,2:00)(4,hold)	SOE PCR for library assembly
Q5_BB	(98,0:30){(98,0:10)(65,0:20)(72,1:30)29x}(72,2:00)(4,hold)	Vector backbone PCR

Table 3-1: PCR Protocols

PCR recipes used for the amplification, assembly, and sequencing of DNA in the MADS directed evolution workflow.

template (ranging in concentration from 1 pg/μL to 1 ng/μL in the final volume). These reactions were amplified according to the protocol “Q5_Bulk” in Table 3-1. PCR Product was purified using ZYMO Research DNA Clean and Concentrator (DCC) columns, according to the manufacturer’s protocol. Final eluted DNA concentration was quantified using a NanodropTM Spectrophotometer.

3.2.4 Single Site Mutation (SSM) Library Creation

pUC57 vector containing the wildtype transaminase and a gene for kanamycin resistance were provided by our collaborators at Merck. In order to create a library of single site enzyme variants, primers with NNK mutations (coding for every available amino acid) at each amino acid position for the wildtype transaminase were designed using Mutation Maker software and ordered from Integrated DNA Technologies. These mutations were distributed into three primer libraries consisting of positions at the core of the protein (Figure 3-2), highly accessible surface residues (tier 2), and remaining surface residues (tier 3). Tier 1, consisting of 96 selected mutations, was used for all experiments described in this work (Table 3-2).

For SSM fragment generation, 500 nM forward and reverse primers coding for a single site mutation are paired with 500 nM reverse and forward primers (respectively, in two separate

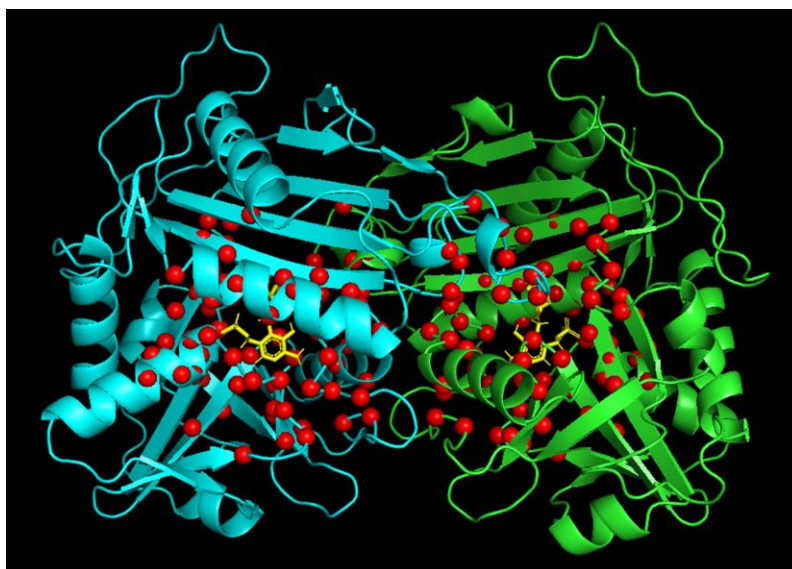


Figure 3-2: Structure of the ATA117 dimer

Structure of the ATA117 dimer with PLP and amine 1 bound, and the core amino acids targeted by the Tier 1 SSM library highlighted with red spheres. 96 sites within the core of the protein were targeted for mutagenesis

SSM Tier 1 Sites					
K2X	A18X	H87X	W172X	I212X	P260X
Q3X	F19X	R95X	I173X	L227X	G261X
N4X	S20X	D98X	V174X	L228X	I262X
K5X	A21X	H99X	R193X	A236X	T263X
E6X	D22X	R102X	T194X	E237X	R264X
Q7X	T23X	L103X	D200X	G238X	K265X
I8X	S24X	F138X	P201X	S239X	T266X
T9X	F72X	V139X	Q202X	G240X	V267X
K10X	Q74X	S140X	V203X	F241X	G296X
Q11X	L77X	S149X	K204X	N242X	C297X
N12X	H78X	T150X	N205X	V243X	T298X
I13X	T82X	P151X	F206X	V244X	T299X
N14X	Y83X	G152X	Q207X	P254X	A300X
A15X	T84X	E153X	W208X	A257X	G301X
S16X	V85X	Y170X	D210X	A258X	G302X
S17X	F86X	Q171X	L211X	L259X	V303X

Table 3-2: Targeted amino acids of the Tier 1 SSM library.

reactions) bracketing the gene of interest in a 20 μ L PCR reaction containing 10 μ L of NEB Q5 2x master mix and 5 ng of template plasmid. PCR amplification (Following protocol Q5_SSM, Table 3-1) in the presence of template plasmid generates fragments running in opposite directions from the mutation site to the end of the genes of interests. Primers NEST_FW and NEST_RV (Table 3-3) were used to generate forward and reverse fragments of the gene of interest, including the T7 promoted ribosome binding domain and the T7 terminator bracketing the gene. 2.5 μ L of both the forward and the reverse fragment product were combined, and these fragment mixtures were then treated with 2 μ L of ExoSAP-IT (thermoFisher) and 0.5 μ L of Dpn1 (Promega) and incubated at 37 degrees for an hour to remove remaining plasmid template after the PCR reaction. 7.5 μ L reactions were diluted 20-fold in ultrapure water after treatment.

After fragment treatment and dilution, 5 μ L of fragment mix was added to 12.5 μ L of Q5

Primer Name	Sequence	Purpose
NEST_FW	GAAAGGGGGATGTGCTGCAAG	Linear Fragment Generation; Library Building
NEST_RV	CCGGCTCGTATGTTGTGTGGA	Linear Fragment Generation; Library Building
NEST_FW_RV	CTTGCAGCACATCCCCCTTTC	Vector assembly for library
NEST_RV_FW	TCCACACAACATACGAGCCGG	Vector assembly for library
EXT_FW	CCCAGTCACGACGTTGTAAAACG	Digital droplet PCR
EXT_RV	GGCTCGAGCATGGTCATAGCTG	Digital droplet PCR
M13_For_Long	GTTGTAAAACGACGGCCAGTG	ivTT DNA Recovery Single droplet PCR
pUC_S7_Rev	CAGAGACGACAGACTAGTGC	ivTT DNA Recovery Single droplet PCR
M13_For_Long_RV	CACTGGCCGTCGTTTTACAAC	Vector backbone assembly for recovered DNA
pUC_S7_Rev_FW	GCACTAGTCTGTCGTCTCTG	Vector backbone assembly for recovered DNA
T7	TAATACGACTCACTATAGGG	Sanger Sequencing
T7 Term	GCTAGTTATTGCTCAGCGG	Sanger Sequencing

Table 3-3: Primers used for the PCR protocols described in this work.

master mix, along with 1.25 μ L of forward and reverse NEST primers at 10 mM, with final reaction volume brought to 25 μ L. Splicing by overhang extension (SOE) PCR was performed by amplifying these mixtures using the SOE_PCR protocol (Table 3-1). 10 μ L of each SOE PCR reaction was pooled to give a mixture of all library members in a single solution, and the mixture was run on a 1% agarose gel for fragment isolation. The band corresponding to the gene of interest (GOI) was cut from the gel and purified using a ZYMO Research gel extraction and purification kit and quantified using a nanodrop spectrophotometer.

Purified linear sequences were assembled into pUC57 backbone via Gibson assembly¹¹⁹ prior to transformation into NEB 10 β *E. coli* cells (See 2.8 Supplemental Information). Linear vector backbone was generated in 20 μ L Q5 PCR reactions from the template plasmid (5 ng/rxn) using primers NEST_FW_RV and NEST_RV_FW (0.1 mM final) and PCR protocol Q5_BB. PCR Product was treated with the addition of 2 μ L of DpnI and incubated at 37 $^{\circ}$ C for 1hr and then purified using a ZYMO DCC kit, and quantified on a nanodrop. 50 ng of pUC57 backbone and 50 ng of library fragment were combined with 5 μ L of NEBuilder[®] Hifi DNA assembly master mix in a 10 μ L reaction and incubated at 50 $^{\circ}$ C for 15 min. Reactions were diluted 3-fold and 1 μ L was then directly added to 50 μ L of NEB 10 β electrocompetent cells and electroporated using a Bio-Rad MicroPulser.

After transformation and plating, single cell colonies were picked for library sequence confirmation and high-density plated cells were scraped from plates to miniprep plasmid stock solutions. Plasmid miniprep was carried out using ZYMO Research's plasmid miniprep kit. Library fragments for digital PCR were generated using 20 μ L PCR reactions containing 10 μ L Q5 2x master mix, 100 nM NEST_FW and NEST_RV, and 1 ng of plasmid, amplified using the

Q5_Bulk protocol. Reaction product was purified on a ZYMO DCC column and eluted in nuclease free water.

3.2.5 Digital PCR and Droplet DNA Dilutions

Purified DNA was diluted serially in nuclease free water at concentrations ranging from 1 ng/ μ L to 1 fg/ μ L. 20 μ L PCR reactions were created using 25 μ L of BioRad Supermix for probes (No dUTP), 75 μ L of NEB Q5[®] HiFi PCR 2x Master mix, 76 μ L of SYBR Green [invitrogen, 10,000x in DMSO] diluted 2000 fold in nuclease free water (5x final concentration), 2 μ L each of forward and reverse primers (10 μ M), 10 μ L of 50 mM Neostigmine or 200 mM chlorocholine in nuclease free water, and 10 μ L of diluted DNA. 20 μ L PCR mix samples were segmented into droplets in Novec 7500 fluorinated oil (3M) with 2% (w/w) surfactant (008, RAN biotechnologies) using a commercial droplet generator system (BioRad QX200). Droplets were thermocycled using the protocol “Q5_Droplet” in Table 3-1, and the resultant emulsion imaged on a Nikon TS2 fluorescent microscope under 470 nm LED excitation and 534/55 nm filtering to determine droplet occupancy.

Based on a rough molecular weight of 1100 kDa and a droplet volume of 1.0 nL, a stock DNA concentration of 72 fg/ μ L should give an average of 1 DNA molecule per droplet. Distribution of DNA in droplets will follow a Poisson distribution with $\lambda=1$, and at this concentration, 59.3% of droplets will be empty, 31.0% will contain 1 strand of DNA, 8.1% will contain 2 strands and the remainder will contain 3 or more. Typical running dilutions for DNA added to digital PCR reactions ranged from 10 fg/ μ L to 100 fg/ μ L depending on the template material molecular weight, purity and concentration read on the Nanodrop[™] spectrophotometer (Thermo Scientific), and fresh dilution series were performed each time digital PCR was performed. DNA dilution series were made from frozen stocks diluted to 1 ng/ μ L.

3.2.6 PCR Droplet Fluorescence Activated Droplet Sorting

After digital PCR amplification, PCR droplets were transferred with a pipette into a custom built storage chamber made from an 8-12 cm long segment of NMR tube. Silicone stoppers (00, Fisher) were cut with biopsy punches (4 mm, Fisher) to cap the ends and access holes were cut with 22 ga needles to allow the insertion of 30 ga Teflon tubing through the caps (See Figure 2-2).

Droplets were sorted based on their relative fluorescent intensity to isolate DNA-bearing samples. A 500 μL syringe pumped oil into the storage device at 2-3 $\mu\text{L}/\text{min}$ to drive droplets from the storage device onto a simple Y-junction sorting device (Figure 3-3). A second 10 mL

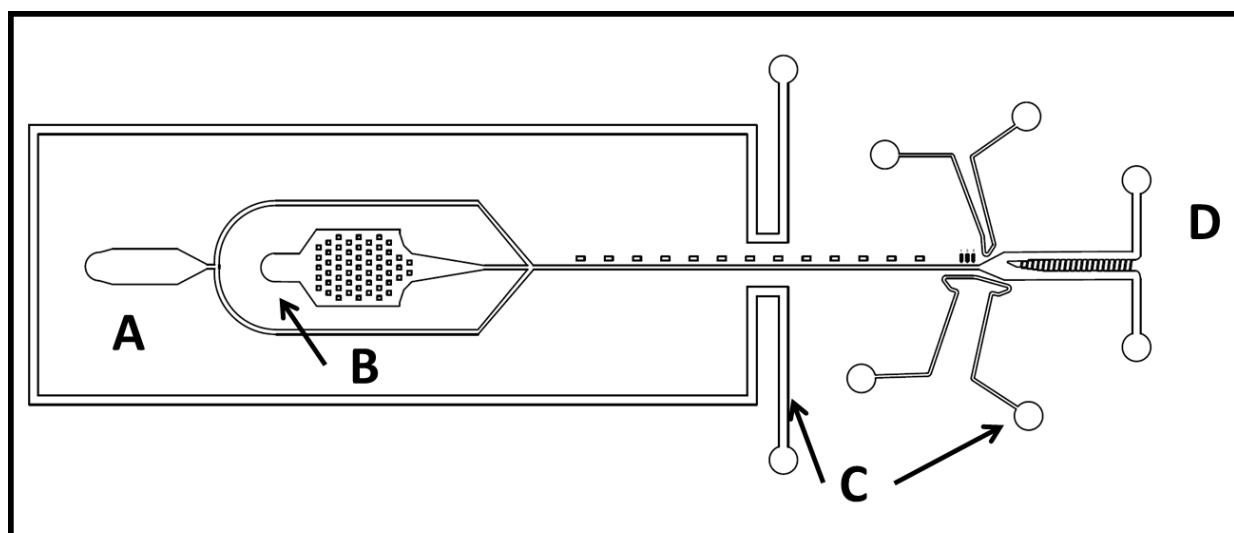


Figure 3-3: Schematic of the fluorescent activated droplet sorting device

The microfluidic device for fluorescent activated droplet sorting (FADS) is detailed. These devices are fabricated to 100 μm depth, with standard channel widths of 100 μm in all regions except the inlets and the sorting region. At the sorting region, the channel flares outward to 650 μm and a series of pillars centered initially 250 μm from the upper edge create an asymmetric division for dielectrophoretic sorting. During operation, oil is infused at inlet (A) to re-space small, 1 nL droplets re-injected into inlet (B). These droplets flow past an interrogation point right before the sorting region, where salt water electrodes allow the application of high voltage AC fields across the sorting region while maintaining ground potential around the droplet inlet to enable dielectrophoretic sorting of the re-injected droplets. Droplets exit the device through one of two outlets (D) after being selected and sorted.

syringe infused oil onto the same device at 8-10 $\mu\text{L}/\text{min}$ to respace these droplets as they entered. The device was set on a Nikon TS2 florescent microscope with a 10x objective aimed at the sorting junction. To detect the fluorescence of these droplets, we attached an Avalanche Photodetector (APD) from Thorlabs (410A2) to the eyepiece of our microscope using a c-mount Nikon 38mm eyepiece adaptor (LMscope). This APD outputs an analog signal corresponding to the light intensity that strikes the detector. The existing microcontroller software was modified to operate in “FADS mode”. The analog voltage readout from droplet signals was sent through a 12-bit ADC input to a STM32F407 ARM Cortex M4 microcontroller board (Mikroelektronika 1685) running custom software. The microcontroller software used an RTOS that acquired analog input signals from the APD and used a peak detection algorithm to identify droplets as signals above a user-defined threshold input on custom Labview software. Peak voltage thresholds were set manually based on the observed readout from droplets passing the detector window. The excitation wavelength for FADS was a 470 nm LED, and a 525 nm Nikon FITC filter cube was used to filter the emission.

A digital output was used to control the DEP sorting voltage. This signal is buffered by a driver integrated circuit (Infineon IR2125), which selectively applies 5V power to a 1.5 kV AC 30 kHz CCFL inverter board (JKL BXA-601). The outputs of the CCFL inverter were wired in parallel and connected through two 100k resistors in series to the sorting junction electrodes via a 3 M salt water filled syringe. The microcontroller runs software written in C using Eclipse and ChiBiOS RTOS libraries and was interfaced via USB to control software written in Labview on the host PC. When a peak was registered above the threshold intensity, the DEP sorting voltage was triggered, pulling the detected droplet into the upper outlet channel. Sorted droplets were

collected in 1.5 mL Eppendorf ® tubes and transferred with a pipette into a second glass storage tube for reagent addition.

3.2.7 Single PCR Droplet Sequencing

To sequence the contents of PCR amplified droplets, single droplets were captured in PCR tubes during the FADS sorting process. This was achieved by moving the FADS outlet tubing rapidly between PCR tubes as the sorting occurred. To ensure capture of single droplets, the flow rate of the FADS sorting was reduced from 2 $\mu\text{l}/\text{min}$ to 0.5 $\mu\text{l}/\text{min}$. With droplets of approximately 1 nl and a droplet DNA occupancy of approximately 1 in 10, fluorescent droplets are sorted and exit the outlet tubing approximately once per second. The relatively low frequency of PCR amplified droplets, combined with the minimal dwell time in each PCR tube enabled the capture of single exiting droplets in single PCR tubes. Tubes containing single droplets were identified after deposition using fluorescence imaging to confirm single occupancy. 20 μl of PCR master mix (prepared as described in 3.2.3) was then added to each tube identified as containing a single fluorescent droplet. The single droplets were merged with this PCR mix through the addition of 1 μL of perfluorooctanol (PFO) to the PCR tube. PCR reactions were amplified using M13_Long_For and pUC_57_Rev primers and Q5_Bulk PCR protocol before being sent to Genewiz (South Plainfield, NJ) for Sanger sequencing.

3.2.8 Preparation of ivTT for in-Droplet Expression

In-Droplet enzyme reactions were run by expressing transaminase DNA using New England Biolab's PureExpress (ivTT). 1 mL of cell free expression buffer was created from 400 μL and 300 μL respectively of ivTT solutions A and B, 40 μL of RNase inhibitor (Murine, NEB), 40 μL of 2.5 mM PLP in 10 mM TRIS buffer, 20 μL of nuclease free water, and 100 μL each of 10, 50, or 100 mM of 1-(imidazo[2,1-b]thiazol-6-yl) propan-2-amine and equimolar

pyruvic acid in 10 mM TRIS (pH 7.5). Final samples contained 1, 5, or 10 mM substrate amine, 100 μ M PLP, and 1, 5, or 10 mM pyruvate. An identical 250 μ L solution of expression buffer was created for the marker droplets used to track samples in the MADS workflow⁸⁰ with the addition of 2.5 μ L of 50 mM Carnitine and 2.5 μ L of McCormick blue food dye.

3.2.9 Addition of Digital PCR Droplets to ivTT Droplets

Reagent addition was driven using a combination of gas pressure and syringe pumps to drive fluid flow on the droplet pairing device²⁰ (Figure 2-8). A 500 μ L Hamilton gastight syringe drove oil at 1.0-1.5 μ L/min into the droplet storage chamber, driving droplets onto the device. The gas pressure pump (Elveflow OB1-MK2) was used to pressurize two vials to 3 psi (20.6 kPa). One vial held the perfluorinated carrier phase for ivTT droplet generation while the other held a 1.5 μ L tube of ivTT solution. The pressure on the vials created 30 nL droplets while the syringe introduced 1 nL droplet downstream at the frequency of droplet generation. Paired droplets flowed past two channels containing 3 M salt water solution¹⁰⁷ connected to a custom built high voltage inverter (CXA-L0512-NJL; TDK-Lamda) supplied with 12VDC and modulated by a potentiometer. High voltage applied to syringes connected to these salt water channels caused surfactant-stabilized droplets to merge as they passed the on-board electrodes. Exiting droplets flowed directly into a second, larger droplet storage chamber made from a Pasteur pipette and capped with Silicone plugs (Figure 2-2).

3.2.10 Droplet Storage and MADS Operation

After DNA addition, droplets were incubated in the larger storage chamber for between 4 and 18 h in circulating perfluorinated oil¹²⁰ at room temperature. Marker droplets made from dyed ivTT were added at the outset of incubation such that approximately 20% of incubated droplets contained the marker material.

After incubation, droplets were infused from the storage chamber and onto the MADS device. A 1 mL syringe filled with carrier oil was connected to the storage chamber and a 10 mL syringe containing carrier oil was connected to the MADS device. Oil infusion from the 1 mL syringe drove droplets from the storage chamber onto the MADS device where oil from the 10 mL syringe respaced them and drove flow on the MADS device. Using an Elveflow OB1-MK3 pressure pump, the outlet vials of the MADS device were pressurized to 0.8 PSI and the pre-sorting oil inlet flow was driven using an oil-filled vial pressurized to 1.2 PSI to balance the back pressure of the droplets running to the mass spectrometer. To ensure equal flow to the mass spectrometer and the sorting region on the microfluidic device as the system filled with samples, these pressures were ramped manually over the first 5-10 min of operation, in increments of 0.2-0.4 PSI from starting conditions of 0.2 PSI and 0.0 PSI respectively.

During screening, the mass spectrometer was set to monitor four ion channels: Amine 1- $M/z=182$, Carnitine- $M/z=162.3$, Neostigmine- $M/z=223$, and chlorocholine- $M/z=122.3$. Amine 1 was monitored to determine enzyme turnover in droplets, Carnitine was monitored to track droplet signals from dyed markers, Chlorocholine was monitored to track positive control droplets and determine baseline selection criteria, and Neostigmine was monitored to identify droplets containing library DNA.

3.2.11 DNA Recovery from ivTT

Following MADS sample sorting, collected hit droplets were pooled in a PCR tube and diluted with the addition of 20 μL 5mM Hepes Buffered Saline (HBS, 5mM HEPES, 20mM NaCl, pH 7.3) and 2 μL perfluorooctanol. The resulting dilution was then filtered on a ZYMO spin column using the “PCR product” protocol and eluted in 10 μL of nuclease-free water. This was used as template in a subsequent PCR containing 1 μL of recovered DNA, 10 μL Q5 2X

master mix, 8.6 μL of nuclease free water, and 0.2 μL each of forward and reverse primer (M13_Long_For, pUC_S7_Rev, 10 μM). This PCR reaction was carried out following the Q5_Bulk protocol in Table 3-1, and purified on a ZYMO DCC column.

3.2.12 DNA Sequencing

Recovered DNA fragments from ivTT were assembled into pUC57 vector using Gibson assembly to vector backbone amplified using primers M13_For_Long_RV and pUC_S7_Rev_FW, as previously described (3.2.4). After assembly, the vector was transformed into *E. coli* (NEB 10 β) with electroporation, and cells were plated and allowed to grow on solid media overnight. After colony growth, colonies were selected and diluted in 100 μL of liquid broth. 2 μL of the resulting broth were then used as template in a 25 μL PCR reaction containing 12.5 μL OneTaq Hot Start 2x Master Mix (NEB), and 200 nM NEST forward and reverse primers following the “Colony PCR” protocol. The PCR product of this reaction was sent to Genewiz for Sanger sequencing using T7 For and T7Term primers (Table 3-3).

3.2.13 Device Fabrication

Microfluidic devices were fabricated using standard soft lithography techniques¹⁰⁴. Devices were designed using AutoCAD software (Autodesk) and printed onto mylar film masks at 25000-50000 dpi (Fineline Imaging). Using these masks, SU-8 2075 (Microchem Corp) was patterned to a depth of 200 μm on a 7.62 cm (3 in) diameter silicon wafer (University Wafer). PDMS (RTV 615, Curbell Plastics) was prepared in a 1:10 ratio of activator to monomer, poured over the wafers and cured 1 h at 70 $^{\circ}\text{C}$ before it was removed from the mold and hard baked for 1 hour at 120 $^{\circ}\text{C}$. Punched ports were created in the cured PDMS using a sharpened 20 or 22 ga syringe needle (BD) and the punched holes were cleaned by passing a short length of fused silica

capillary (360 μm OD, 20 μm ID, Molex) through the hole to clear any particulate left behind by the punch.

Devices were cleaned with Scotch tape (3M) and bonded to 7.62 cm long (3 x 1 in) glass microscope slides (Corning) after exposing both the slide and the molded PDMS to oxygen plasma using a Tergeo plasma cleaner operating using the “PDMS Bonding” default protocol. Bonded devices were placed in an oven to cure overnight at 80 °C.

3.2.14 Data Analysis

Raw signal data from droplet MS screens was exported to .CSV files after screening, with one CSV produced for each monitored ion. These files were collated into a single spreadsheet listing the timestamp and peak intensity of each of the 4 channels. Peak heights for each droplet were identified by reporting maximum values above a user-defined threshold, and droplets were then automatically categorized based on their intensities using an excel spreadsheet. Amine peak values classified as belonging to blank, library, marker, and control populations were then extracted and plotted using GraphPad Prism.

While this data plotted does not directly represent the data observed by the microcontroller sorting the droplets in the MADS system, it gives an accurate representation of the droplet signals observed by the microcontroller. The microcontroller receives an analog trace directly from the mass spectrometer, and makes its determination on what constitutes a peak and how that peak is categorized based on the minimum peak height thresholds set by the operator during the sorting process. Data plotted after a screen is reported by the chemstation software as an ion count signal, rather than an analog voltage on the mass spectrometer. This data is processed similarly in excel, with peak detection performed by defining peaks based on ion count intensity above a set threshold in the amine channel, and then reporting the maximum

values in each channel corresponding to the identified peak on the amine channel. Plotted data represents a reasonable representation of the data observed by the MADS hardware, but because one peak detection threshold is set based on the analog signal from the MS and the other based off of the ion count data stream, the magnitude of this threshold is slightly different. Ultimately, the impact is that it is possible for these two processes to mismatch on the decision to categorize a peak as a droplet if a signal falls below the threshold of one but not the other, but practically speaking this is exceedingly unlikely, because the signal to noise in the peak identification channel – that of the amine– is sufficient to ensure that the threshold set for peak identification is reached by all droplets, and is well above any baseline signal on that channel.

Due to the slow transfer of substrate between active and inactive droplets, droplet populations monitored over several hours exhibited signal drop that followed the activity of the enzyme expressed within active droplets. This decay was fit with quadratic functions to describe the spread of population data as the enzyme turned over substrate to product. 95% confidence intervals for this fit were calculated using graphpad software, and 95% prediction intervals were calculated and plotted as well. These best-fit functions were used to assess the sorting metrics chosen for each population, analyze the sorted samples, and confirm the confidence with which hits were identified.

3.3 Results: Implementation of an in-Droplet Protein Expression Workflow

Previous examples of enzyme evolution in droplet microfluidics have relied heavily on protein expression *in vivo*, which limits the breadth of droplet screening to cell permeable compounds and non-cytotoxic proteins^{32,73,76,89,91,114,121}. In our earlier demonstration of in-droplet transaminase activity monitoring, we were able to observe successful conversion of the non-native substrate amine **1** into ketone **2** by wildtype ATA117 expressed in droplets using in-

vitro methods⁸⁰. Given our ability to observe a distinct difference in amine substrate signal between droplets and bulk reactions that had received a dose of purified DNA and those that had not, it was now important to verify that this change in amine signal could be produced through in-droplet expression, and that this expression could be produced via the addition of unpurified PCR product in the droplet form.

3.3.1 Design and Optimization of PCR Amplification in Droplets

Similar to single cell protein expression and single cell encapsulation in microfluidic droplets, *in vitro* protein expression and evolution in droplets requires the delivery of single enzyme variant DNA to each microfluidic sample to produce unique proteins in each droplet. Stochastic loading of dilute solutions of DNA into droplets would enable single-gene, single droplet capture of DNA, but to generate sufficient concentrations of genetic material for *in vitro* protein expression, PCR must be used to amplify this isogenic material prior to ivTT expression^{40,78}. This means that to perform in-droplet expression, we needed to combine single droplets of PCR product into single droplets of ivTT to observe enzyme activity.

PCR is a well-established protocol for the amplification of DNA, but the degree of amplification, purity of product, and concentration of the DNA produced are intrinsically linked to the template concentration, number of cycles performed, and polymerase enzymes used in the reaction mix. We needed to design a PCR protocol that enabled us to a) amplify single genes in single droplets without significant side products, b) track single gene amplification within each individual droplet to gauge the appropriate concentration of template to add, and c) track the ultimate addition of DNA to the ivTT solution using an added small molecule.

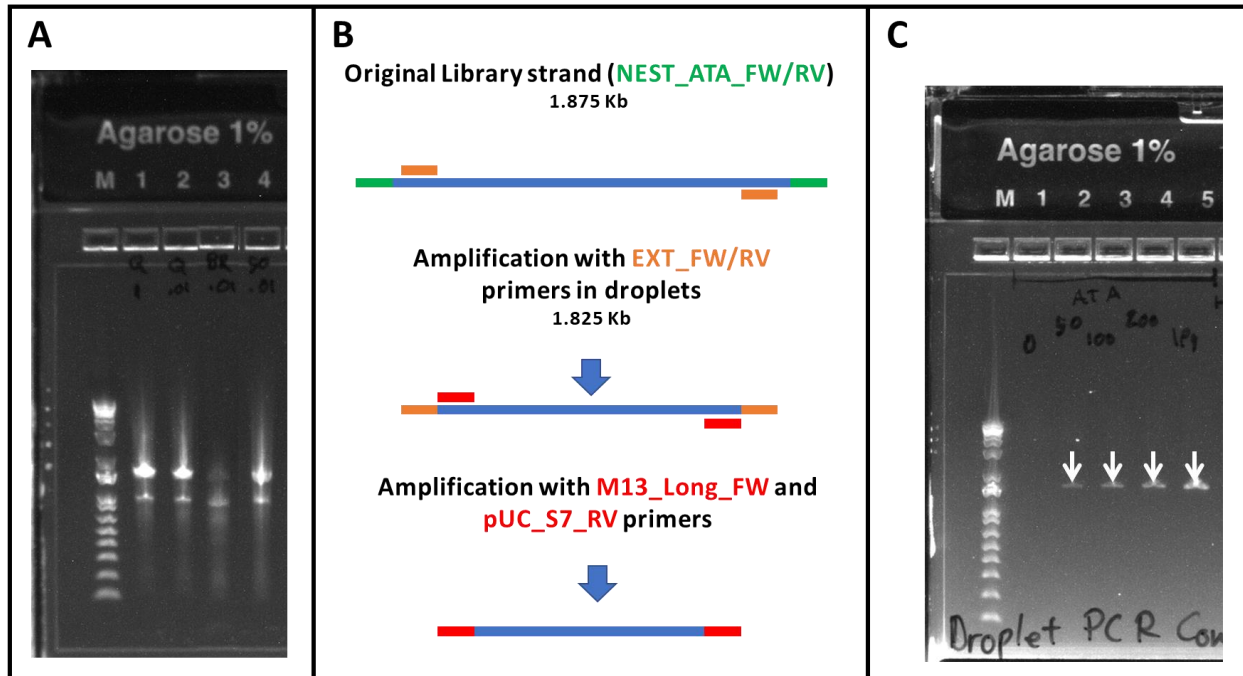


Figure 3-4: Experimental data and strategies for PCR amplification in droplets

(A) PCR efficiency was lower when using PCR mix for digital qPCR than it was when using PCR mix for high fidelity amplification. Q5 master mix (lanes 1 and 2) gives far more efficient DNA amplification when compared to the master mix made for digital PCR by BioRad (lane 3). However, droplets made from Q5 master mix did not survive thermocycling. Mixing the two commercial master mix solutions (lane 4) resulted in droplets that were both stable and gave high concentration product. (B) Side products formed during PCR amplification of linear template led to the development of a “nested” primer strategy for PCR amplification. Linear fragment containing the gene of interest is formed from plasmid template using NEST primers. This template is distributed into droplets where it is amplified by EXT primers for expression. The EXT amplified fragment is then recovered and amplified for sequencing using M13 and pUC_S7 primers. (C) EXT primers amplifying NEST template in droplets gives clear, clean, single-product signal around 1500 kb (marked with white arrows). A dilution of 0, 50, 100, 200, and 1000 fg per reaction shows increasing product formation with increasing template loading.

PCR amplification in droplet format proved to be a challenge to the implementation of in-droplet expression. In particular, droplet stability during PCR thermocycling proved problematic, as was the appearance of secondary PCR products in the PCR reaction. Initially, we found that PCR droplets did not survive thermocycling when the PCR reaction mixture was made solely using NEB’s Q5 High Fidelity 2x Master Mix, but they appeared to survive when made using BioRad ddPCR Supermix for Probes (no dUTP). We then found that the BioRad

master mix appeared to give less efficient PCR at the low template concentrations necessary for digital PCR to achieve single-droplet template loading. By combining these two solutions, we found we could achieve both droplet stability, and successful amplification (Figure 3-4 A). During these investigations, we noted the formation of multiple PCR products, especially when the primers used to amplify the linear sequence of DNA were the same as the ones used to build the linear fragment initially. We found that we could eliminate this by utilizing “nested” primers for our gene of interest which successively moved inward on the template strand with each round of DNA amplification. Nested primer sets (which were reportedly used in previous droplet PCR protocols⁴⁰) reliably gave clean product when successive PCR was performed on linear template (Figure 3-4 C). The first set of primers, NEST_FW and NEST_RV, bracketed the gene of interest 223 and 295 bases from the start and stop codon of the transaminase gene on our pUC-57_ATA117 construct. These were used for the library build, and formed the initial linear gene fragment that would be purified and dispersed into droplets. The second set of primers, EXT_FW and EXT_RV, laid 25-35 bases internal to the NEST primers. These were used for the in-droplet PCR amplification, and would be what was expressed in the ivTT reaction. A third set of primers, M13_Long_FW and pUC_S7_Rev, could then be used to recover DNA from the ivTT reactions and regenerate the vector for cloning and sequencing hit sequences.

3.3.2 Detection of PCR Amplification in Droplets

To fluorescently track in-droplet PCR and template distribution, two approaches were explored. Initially, quantitative PCR (qPCR) probes, which use the 3' → 5' exonuclease activity of Taq polymerase to degrade a sequence-specific oligomeric probe and release a quenched fluorescent signal, were used to detect PCR amplification. However, these probes tended to disrupt full-length gene amplification in the droplet format, and PCR efficiency was low in

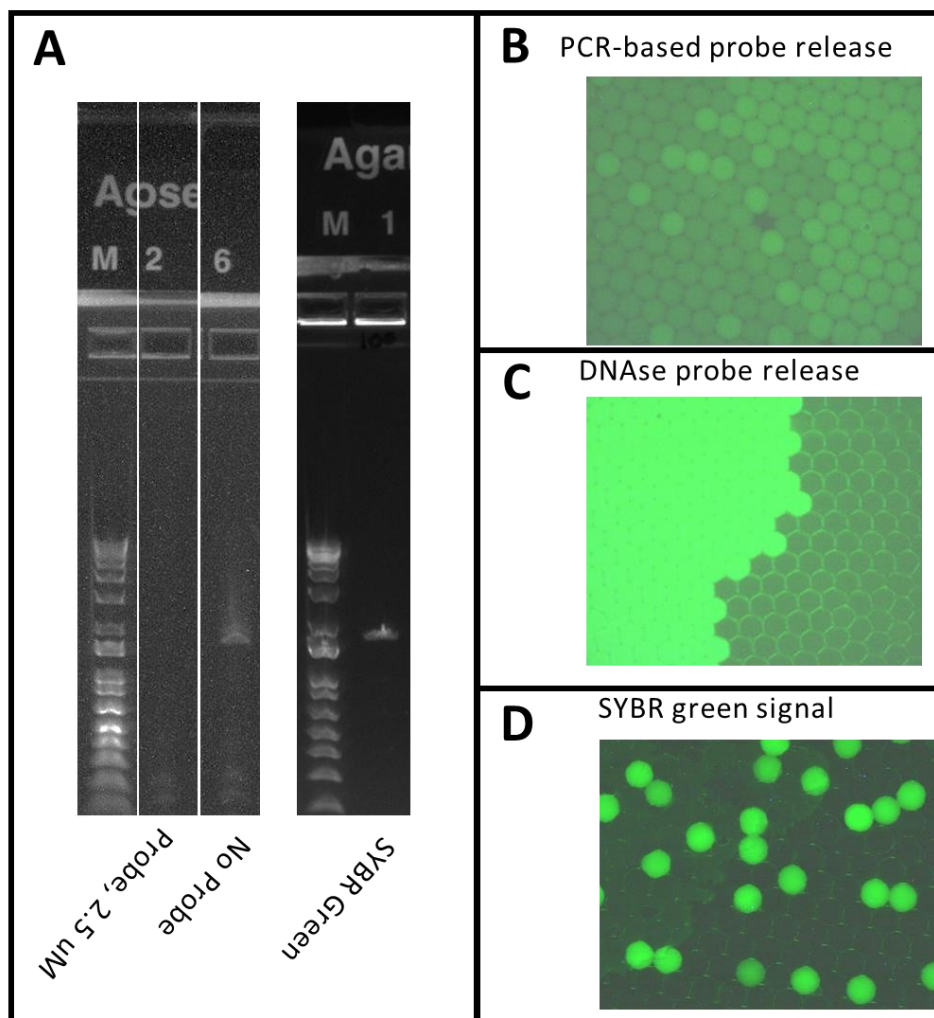


Figure 3-5: Strategies for detection of PCR amplification in droplets.

(A) In early investigations, we found that the presence of qPCR probe in the PCR reaction greatly disrupted PCR amplification, but that SYBR Green dye, a small molecule that is fluorescent in the presence of double stranded DNA, did not. (B) While probe-based detection gave positive signal for DNA amplification, the signal to noise was weak relative to what could be achieved with full probe cleavage by DNase (C), and the signal to noise of SYBR Green in PCR amplification (D) is higher, making DNA bearing droplets more easy to detect.

reactions containing this probe material (Figure 3-5 A). Furthermore, the release of the fluorophore within these droplets was far less efficient than expected, and the background fluorescence of unreleased probe was relatively high, likely due to the fact that the Q5 polymerase that made up 75% of the reaction mix, does not contain 3' → 5' exonuclease activity (Figure 3-5 B). SYBRTM green I, an intercalating DNA dye, proved to be a better sensor for

DNA replication. This dye is fluorescent only in the presence of double stranded DNA (dsDNA), so upon amplification, PCR product solution turns brightly fluorescent (Figure 3-5 D). The small molecule neostigmine was also added to the PCR reaction mix, which allowed the mass spectrometer to track addition of PCR mix to the microfluidic droplets of ivTT (final concentration 2.5 mM).

Preliminary calculations that assumed an average per-base molecular weight of 330 g/mol and a DNA strand length of 1500 base pairs were used to estimate how much DNA was necessary to load into the PCR solution to get microfluidic droplets (assuming a 1 nL total volume) where approximately one in ten contained DNA ($\lambda=0.10^{122}$ in Poisson distribution statistics). This calculation showed that DNA would need to be diluted to low fg levels to achieve single droplet occupancy. While this gave a reliable estimate for how much it was necessary to dilute template DNA in PCR reaction mix, a dilution series of the template DNA in PCR master mix was performed each time digital PCR experiments were run, to determine the appropriate dilution for any given droplet PCR run. This is because the concentration of the DNA stocks measured after purification are measured at significantly higher concentration than the dilution necessary to achieve single droplet loading in droplets (in the ng/ μ l range), and the dilution necessary to bring each stock to single occupancy concentrations tended to vary between 10 and 100fg, likely due to the limitations in the spectrophotometric read of the starting DNA concentration and DNA loss to the surface of the pipettes and tubes with which stocks were diluted.

Typically, we aimed for an encapsulation ratio (λ) of $\lambda=0.10^{122}$, or 0.10 DNA molecules per droplet. At this dilution, 90% of droplets should receive no DNA, and of those that do, 90% should receive only a single copy of DNA, with the remaining 1% of droplets getting 2 or more.

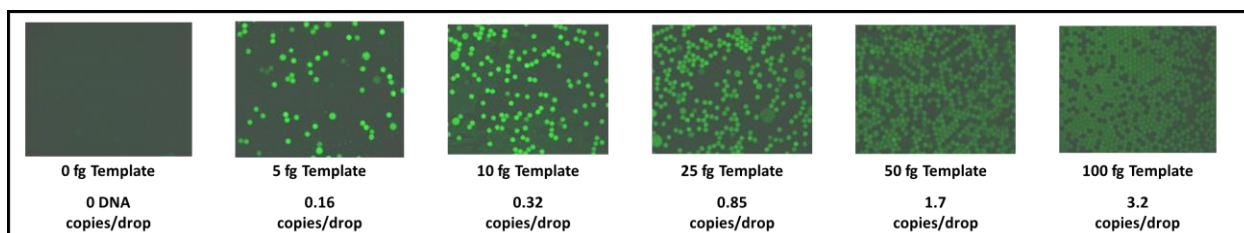


Figure 3-6: Fluorescent images of digital PCR dilutions

Digital PCR at a range of template concentrations reveals the stochastic loading of droplets with PCR template prior to PCR. At low template concentrations, single template strands of DNA are captured in single droplets and amplified to give a high concentration of DNA, all stemming from a single parent sequence.

Figure 3-6 shows a representative dilution series. Typically, a dilution series such as this was used to set the concentration for digital PCR, with the aim being to select the concentration closest to the desired 10% occupancy. PCR protocol Q5_Droplet was used to amplify the droplets during digital PCR.

3.3.3 Expression of PCR Reaction Product by ivTT

To test the ability of our ivTT reaction mix to express unpurified PCR-amplified DNA and exhibit a drop in the signal of amine **1**, a set of bulk 25 μ L ivTT reactions with varied template loading were performed at microliter scale. 1 μ L of water, purified T7 promoted epidermal growth factor receptor protein (EGFR) template (125 ng/ μ L), and de-emulsified ddPCR master mix amplified in droplets with 0, 10 fg, 100 fg, 1 pg, or 2 pg of linear template (per 20 μ L reaction) was added to each. After 6 h at room temperature, the reactions were analyzed via direct mass spectrometry of infused plugs (see 3.5.12), monitoring the mass of the substrate amine (**1**) (Figure 3-7). The addition of droplet PCR-amplified DNA to the ivTT reactions resulted in a decrease in amine signal ($p < 0.001$) for all samples receiving more than 10 fg/reaction of DNA amplified in droplets. The addition of PCR mix that did not have template and PCR mix with very low template concentrations (10 fg/reaction) allowed confirmation that this change did not stem from the addition of PCR reaction mix alone, and that a threshold

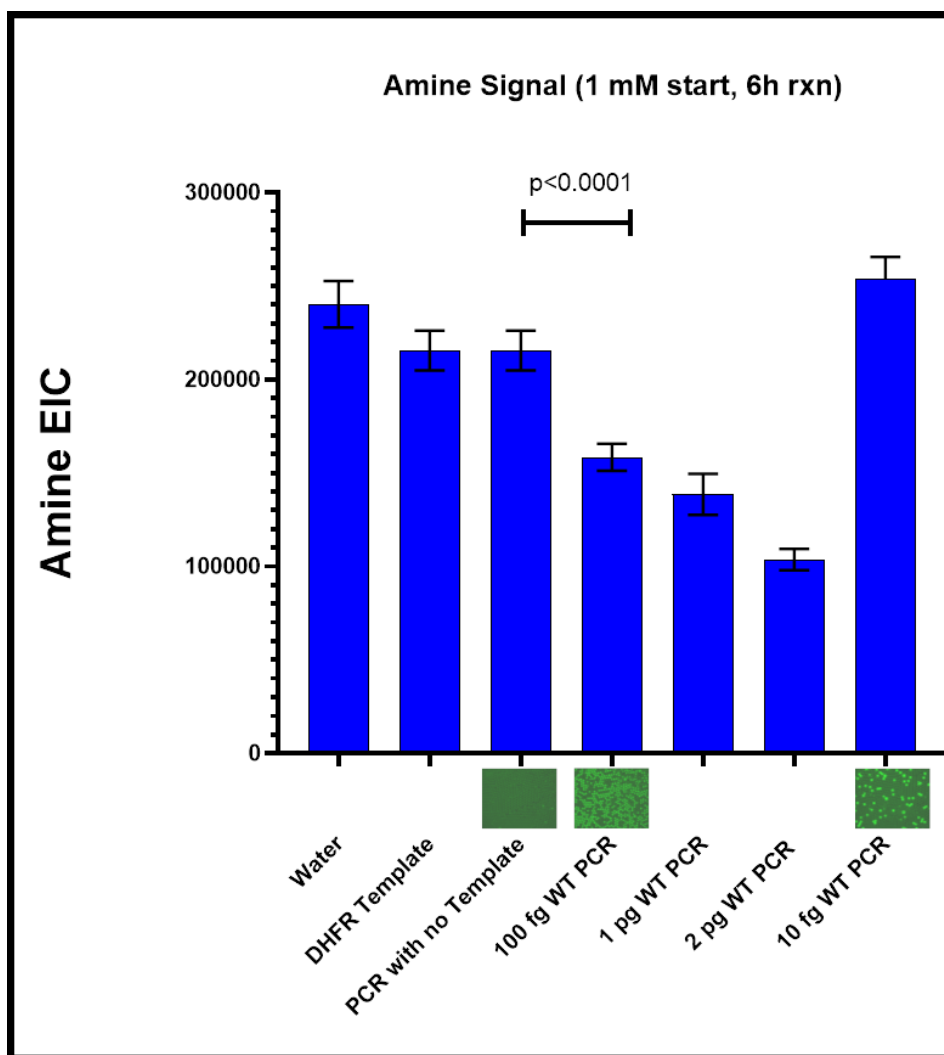


Figure 3-7: Bulk ivTT reactions reveal enzyme activity as a drop in amine signal

Amine signal read by direct MS of droplets of ivTT reaction solution. No expression, and expression of an inactive protein (columns 1-3) result in a high concentration of amine substrate remaining. Expression of the transaminase ATA117 gene in bulk results in a distinct decrease in signal. The extent to which this decrease occurs corresponds to the quantity of transaminase DNA added. When low template concentration of 10 fg/reaction is used, ~90% of droplets are empty, and the dilution of template that occurs when the droplet PCR emulsion is broken is sufficient to result in insufficient DNA expression for activity to be observed via loss of amine substrate. Images inset below the bars reveal the occupancy of PCR amplified droplets when fewer than 100% of droplets receive template DNA.

amount of DNA was necessary to express enough protein to observe amine signal decrease. The expression of EGFR DNA in the same reaction conditions (EGFR protein has no transamination

activity) confirmed that this drop in substrate concentration stemmed from the presence of transaminase DNA being expressed, not merely DNA expression.

3.3.4 In-Droplet Expression of PCR Reaction Product by ivTT

Confirmation that protein could be expressed in single droplets from digitally amplified droplets of DNA was undertaken by confirming transaminase expression after reagent addition of digital PCR material to ivTT. Diluted NEST-amplified ATA117 template was added to PCR master mix and thermocycled so that each amplified droplet would contain (and therefore induce expression of) the same template DNA. Using a dilution of 5 fg of template in 20 μ L of PCR master mix containing 2.5 mM neostigmine, NEST_ATA117 was amplified in ddPCR and added to 30 nL droplets of ivTT. After incubating 6 h, these droplets were infused onto the mass spectrometer using the MADS device.

Figure 3-8 shows both the experimental workflow and the resulting populations of droplets and their substrate amine signals. Transaminase expression could occur only when ddPCR samples containing amplified DNA were added to droplets of ivTT, and the addition of ddPCR material was traceable independent of protein expression in the form of neostigmine in the added PCR material. As before, DNA expression and subsequent transamination could be observed using MS monitoring of amine **1**, while successful addition of PCR materials could be observed by monitoring the neostigmine from the PCR solution. If the proportion of droplets displaying a change in amine **1** signal matched the proportion showing fluorescent labeling after digital PCR, this would suggest successful expression from singly amplified droplets of DNA. As expected, a reduction in in amine **1** signal was observed when the added ddPCR droplet contained amplified and expressed ATA117, and the frequency of this result closely matched the frequency of fluorescent droplets in the ddPCR reaction (Figure 3-8 B). Confirmation of this

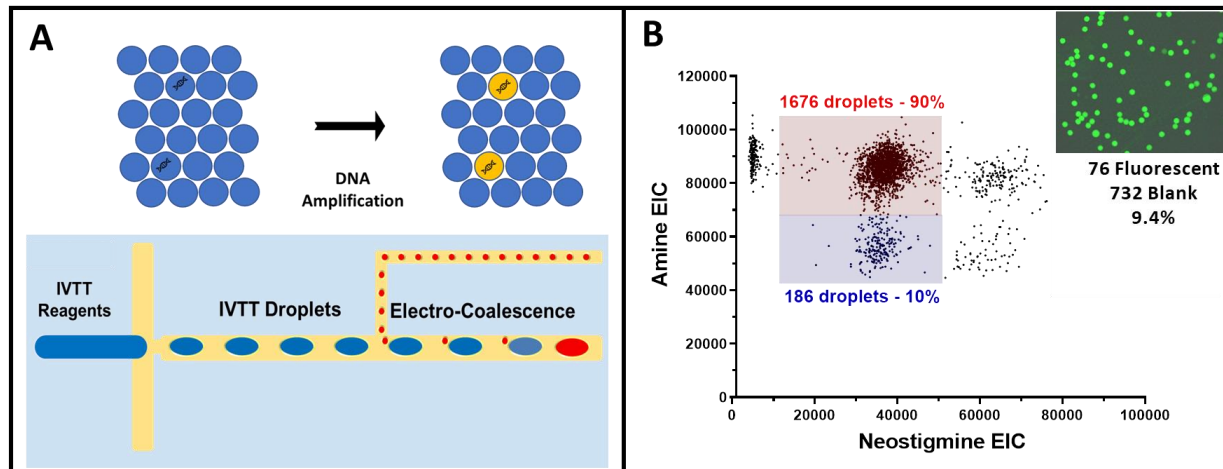


Figure 3-8: Experimental workflow and data for single gene droplet expression

(A) The experimental workflow for determining if single droplet expression is possible from PCR amplified DNA in droplets is shown. DNA is diluted to 5-10 fg per 20 μ l PCR reaction and neostigmine is added to the mixture before it is broken into 1 nl droplets. After PCR thermocycling, amplified DNA will only appear in droplets that contained single template strands of DNA prior to amplification. These droplets are then added to droplets of ivTT with 1 mM of the substrate amine **1**, and the signal of the droplets after 4 h of expression is read via droplet MS. The small molecule neostigmine in the PCR reaction allows the addition of single droplets of DNA to droplets of ivTT to be tracked via mass spectrometry. (B) When the neostigmine and amine signals in each droplet are plotted against one another, distinct populations of droplets may be identified. Droplets that receive no addition through the reagent addition process show low neostigmine signal, as there is no neostigmine in the ivTT and it only can be added through reagent addition. Because these droplets have undergone no addition, they all have high amine signal, corresponding to no enzyme activity. The population that received amplified DNA droplets from the reagent addition process show a reduction in the amine signal and an elevated neostigmine signal (blue region), while those that received empty droplets without DNA amplified within show an elevated neostigmine signal but no reduction in the amine signal. The ratio of the droplets containing DNA as observed in digital PCR fluorescent imaging (inset) closely matches the proportion showing activity within the expressed droplet pool. Addition of two droplets of DNA to one of ivTT may also be observed in the two clusters of droplet signals to the right of the red and blue highlighted clusters, showing that neostigmine concentration scales with addition volume. This also explains the scatter along the x axis between clusters – reagent addition is not 100% consistent and some droplets will receive more or less than a single volumetric addition, either due to size variation within the DNA droplets, or due to pairing of more than one droplet to each ivTT sample.

result indicated that ivTT expression of isogenic DNA was indeed occurring in droplet samples, and that this translated directly into a reduction in the amine concentration.

3.3.5 SSM Library Generation

To screen for transaminase enzymes capable of outperforming the wildtype enzyme in the transformation of amine **1**, a site saturation mutagenesis (SSM) library of transaminase variants (see section 3.2.4) was built and screened for activity. This library of enzyme gene variants was cloned into the plasmid pUC57 and evaluated for the appropriate diversity (Table 3-4) through *E. coli* expression and Sanger sequencing. 69% of the sequenced colonies displayed expected library SSM mutations, and 6% carried the parent (WT) gene, with the remainder showing more than one mutation or frameshifts. This library of gene fragments encompassing the first tier (Table 3-2, representing 1,919 enzyme variants from saturation mutagenesis of 96 positions in the enzyme core) was then used as template in droplet PCR for single droplet library amplification.

Tier 1 WT SSM	% (16 samples)	# Colonies
Wildtype	6%	1
Mixed Population	12%	2
Frameshift	6%	1
Single SSM Mutation	69%	11

Table 3-4: Wildtype Tier 1 library Sequencing

Library quality control data for the SSM build off of WT transaminase shows the majority of sequences read show expected, single-site mutations on the DNA sequence, all at targeted SSM locations.

3.3.6 Fluorescent Activated Droplet Sorting to Enrich for DNA Expression

Droplet expression data from early work on the expression of PCR amplified ATA117 DNA highlighted a key challenge in the execution of single droplet protein library expression.

Due to the initial distribution of low copy number DNA into PCR droplets, the majority of PCR amplified droplets will contain no DNA after amplification, and will therefore demonstrate no activity upon addition to ivTT solution. To maximize MADS sample processing rates, enrichment of library DNA in the DNA bearing ddPCR droplets was necessary prior to ivTT expression. A microfluidic device for fluorescent activated droplet sorting and used FADS enrichment to isolate droplets with DNA after PCR was therefore developed to sort the samples based on SYBR green fluorescence.

PCR amplified droplets containing WT library DNA were sorted using this FADS setup (Figure 3-3). Without sorting, only 14% of these droplets are fluorescent and have the potential of containing the DNA for an active (or inactive) variant. While many of these droplets should contain DNA coding for active mutations, the nature of an SSM is such that some of these droplets will code for deactivating mutations, and unlike the PCR experiment in Figure 3-8, fewer than 14% of these samples should be active when expressed. These droplets were sorted by their fluorescence, and imaged before and after sorting to verify accurate FADS enrichment. Bulk enrichment of the DNA in sorted droplets was also confirmed via gel electrophoresis (Figure 3-9 B). Fluorescent images show enrichment to 97% fluorescent from 14%, and a significant increase in signal intensity on the gel after sorting (Figure 3-9 B, lane 2).

The expression of unenriched library DNA should show far fewer active samples than that of the FADS enriched library DNA, and that the number of active samples in the unenriched pool should be less than 14%, due to the nature of the SSM library, which will exhibit both deleterious and advantageous mutations. In the instance shown in Figure 3-9 C and D, only 7.5% of expressed library droplets initially showed signal within the 95% prediction interval of the WT activity distribution. This prediction interval is an attractive estimate for activity within

library samples, because it represents the activity distribution of the unmodified, wildtype enzyme.

Given that only 14% of droplets showed fluorescence prior to sorting, but 97% showed fluorescence after, a commensurate increase in the number of signals showing enzyme activity should be observed after sorting. Following FADS enrichment, 60% of droplets in the library appeared within the 95% prediction interval of the wildtype positive control, an 8-fold increase that closely aligns to the fold-increase in fluorescent droplets (Figure 3-9 E and F). This result demonstrated the effect of pre-sorting DNA droplets; doing so minimized the number of empty droplets processed by the MADS system.

The utility of eliminating these empty samples is best understood through an assessment of the time and cost of the alternative to fluorescent sorting. Without sorting, ivTT samples would receive DNA for expression 14% of the time, meaning it takes approximately seven times as much time to read a single sample, and seven times as much ivTT. Fluorescent sorting (at current rates) for this work took approximately 16 h per screen, due to the low per-droplet volume (1 nL), relatively low speeds (50 droplets/s) and the high emulsion volume (~100 μ L) necessary to run the reagent addition device. This meant that a screen over an 8 hour day with FADS sorting took three times as long as one without, but covered 7-8 times as many samples and saved 7-8 times the volume of ivTT, leading to a reduction of a week of screening time and a cost savings of several hundred percent. Given the expense of the ivTT material used in this screening workflow, this is a significant factor in the decision to pursue pre-enrichment of droplets with DNA.

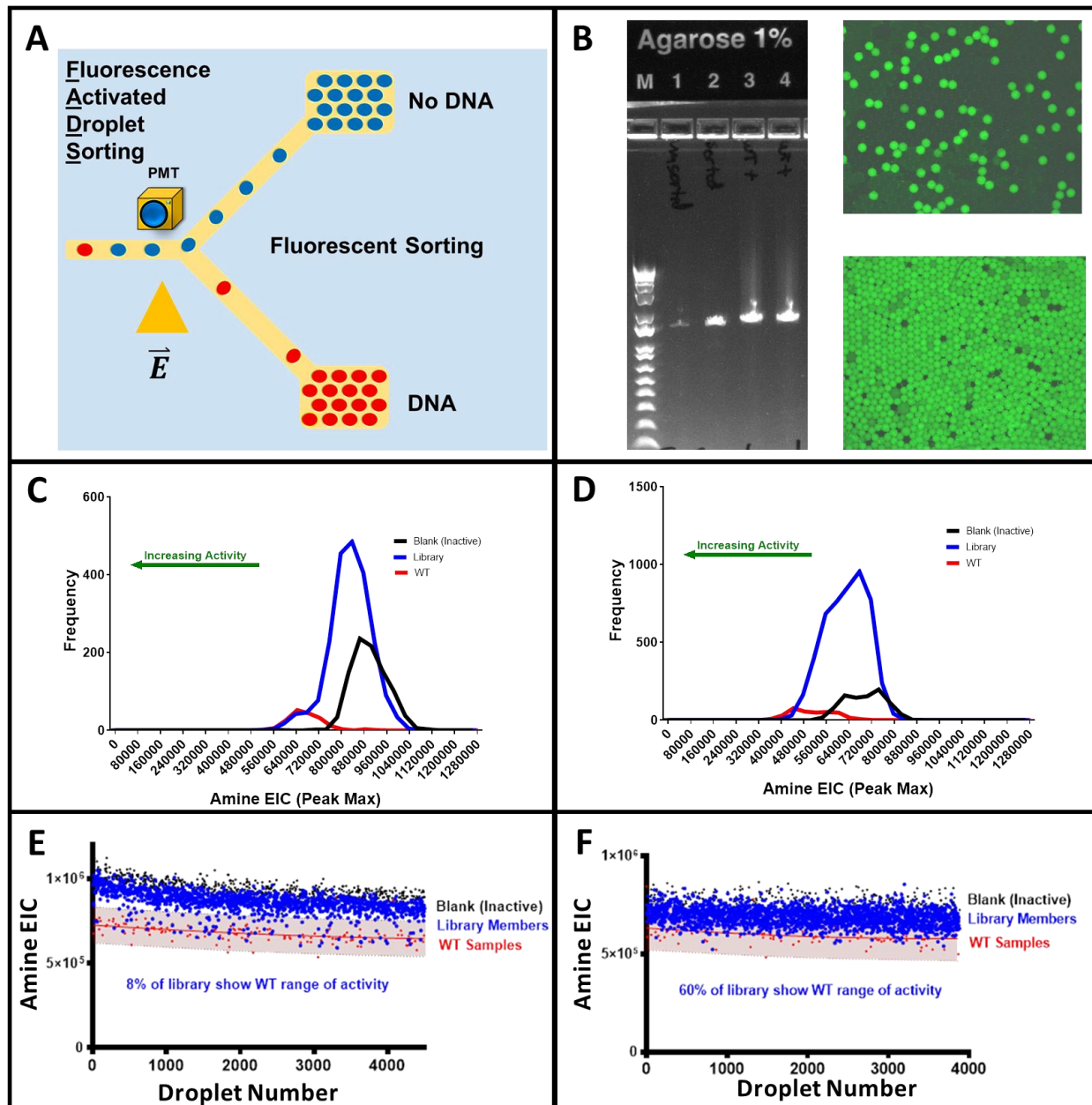


Figure 3-9: Fluorescence activated droplet sorting enriches DNA containing droplets

(A) Representation of Fluorescence Activated Droplet Sorting (FADS). This technique enables the enrichment of DNA containing droplets prior to reagent addition in the droplet format. A microfluidic device allows 100-125 μm , ~ 1 nl droplets to be reinjected at 1-2 $\mu\text{l}/\text{min}$ into a 100 μm channel and separated by re-spacing continuous phase. As these droplets flow past an illumination and detection point, they are interrogated serially by an avalanche photodiode, which is used to trigger a sorting pulse on downstream electrodes. (B) Imaging before and after sorting reveals enrichment for droplets containing DNA from 14% to 97% fluorescent, and gel electrophoresis confirms the higher intensity of signal (and therefore enriched DNA concentration) in sorted droplets. Panels C, D, E, and F show the distribution of droplet activity (measured by the reduction of amine) in a library of samples prior to fluorescent sorting of the expressed WT library (C/E) and after sorting (D/F). Both reactions were performed with 10 mM

Amine and read starting at 4 h after reagent addition. 95% prediction intervals in E and F mark the region expected to contain 95% of signals with wildtype activity. Fluorescent enrichment increases the percentage of library samples that fall within this plotted region from 7.5% in (E) to 60% in (F). This demonstrates that pre-enrichment with fluorescent sorting enriches the number of active samples read within a single library screen by nearly 8-fold.

3.3.7 Single Droplet Sequencing Enables Sequence Confirmation

To test the contents of PCR amplified droplet, single droplets were captured in PCR tubes during the FADS process (Figure 3.3.6). These single droplets were then merged with 20uL of PCR master mix and underwent PCR amplification (Figure 3-10), followed by Sanger sequencing. 73 droplets were read this way, with 55 giving full reads. Ideally, the sequence distribution read this way would closely match that of the original library material, with around 70% of the reads carrying a single site mutation. DNA loading into the droplets will follow a Poisson distribution, and the statistics for single gene encapsulation are such that the number of strands captured in each droplet can be modeled by the equation:

$$P(X) = \frac{\lambda^X e^{-\lambda}}{X!}$$

The probability of a droplet containing X DNA sequences is given by the term λ , which measures the number of DNA strands available in the original solution, per droplet volume (copies/drop). With 14% of droplets showing fluorescence, it is possible to back-calculate that this λ value was approximately 0.15 DNA sequences per droplet, which closely matches the observed number from a dilution of template to approximately 5 fg/reaction (0.16) (Figure 3-6). This value of λ may then be used to calculate the fraction of the fluorescent droplets with DNA that will contain a single strand, in this case 92% of our samples. Given the distribution of wildtype, SSM, and truncated sequences observed in the original library sequencing, single gene encapsulation should occur approximately 64% of the time (70% library multiplied by 92%

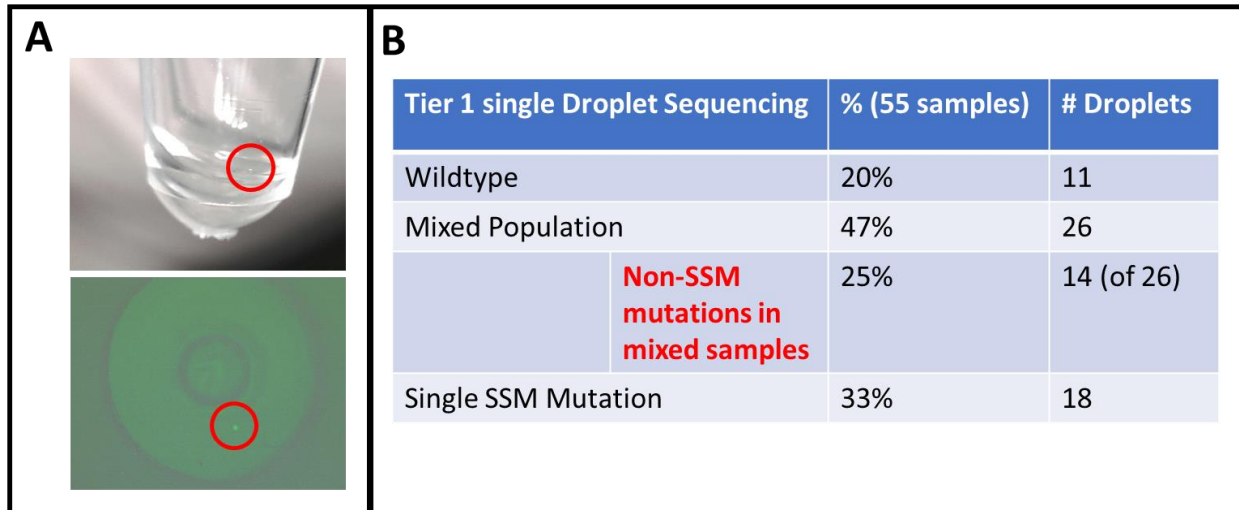


Figure 3-10: Single droplet sequencing data for amplified DNA in droplets

(A) Fluorescent activated droplet sorting enables the capture of single droplets in PCR tubes. Brightfield and fluorescent images show these single droplets in PCR tubes, which have been highlighted with red circles. These single 1 nL droplets may be combined with PCR reagents and amplified, enabling single droplet Sanger sequencing. The table in (B) summarizes the results of single droplet sequencing, and reveals that after encapsulation and amplification, non SSM mutations (in red) are present in the DNA that were not present prior to amplification. This tells us that while single droplet PCR is amplifying single gene variants within 1 nL droplets, the process is not producing isogenic droplets with perfect fidelity. Ongoing work seeks to resolve this.

single capture), and wild type DNA approximately 8% of the time, with the remainder carrying mixed materials.

Single droplet sequencing results (Figure 3-10) revealed that among the 55 full reads, 20% (11 reads) showed the parent DNA. 33% (18 reads) showed library members with single site mutations. 47% (26 reads) show multiple mutations or mixed sequence reads at a single mutation site, a phenomenon that could be caused by multiple gene encapsulation, PCR droplet merging, or mutation during amplification.

The high frequency of mixed mutations was concerning, given the statistics of droplet encapsulation and library composition previously gathered. Further investigation of the 26 sequences containing multiple or mixed mutations reveals that 14 contained mutations not

explained by the NNK library build, which only targeted 96 sites using an NNK degenerate codon system that should ensure that the final nucleotide of any mutant codon is a G or T. Only 2 (4% of the total reads) sequences contained clear evidence of multiple NNK codons in a single read, and 9 showed evidence of mixed sequence at a single mutation site. Taken together, this evidence suggested that (1) the digital PCR process was introducing extra mutations into samples that appeared to have started as single NNK variants in single droplets, and (2) single gene encapsulation and amplification was occurring in the droplet context. Investigation into the observed mutation phenomena described here is ongoing, and ultimately an important challenge to resolve. Among other things, SYBR green degradation products have been shown to be potential inhibitors of PCR¹²³, so further work optimizing the droplet reaction conditions and SYBR green concentrations could help resolve these issues. Nonetheless, single droplet sequencing provides an (as yet unreported) opportunity to gain insight into the genetic contents of droplets amplified in digital PCR, and this work enabled verification that the library distribution in droplet PCR produced isogenic amplified DNA for expression in droplets of ivTT. Even with the appearance of these non-SSM reads, 80% of reads showed non-WT DNA amplified within the droplets, and none showed more than 2 mutations on the WT sequence. Single droplet expression therefore appears to be capable of producing single protein expression in a microfluidic droplet.

3.4 Initial Efforts to Screen a WT Transaminase Library

With a procedure in place for the complete *in vitro* expression and activity assay of transaminase enzymes, we considered our potential ability to screen a library of enzymes using this system. Analytical data from previous work with the MADS platform revealed that our Z' factor, which measures the ratio between the separation of the variance in two populations and

their respective means¹²⁴, was negative when measured between the wildtype enzyme and the blank samples at the beginning of the reaction, though it did increase as a function of reaction time (Figure 3-11). Typically, a strong screen will have a Z factor of 0.5 or better between positive controls (hits) and negative controls. A negative Z' means that the 99% confidence intervals of the two populations overlap and distinguishing the two populations with 99% confidence is not possible. Because positive enzyme variants (better than WT) will show greater separation from this baseline, and because the WT activity is a metric of minimal activity, not maximal for the transamination this suggested the MADS system would be capable of identifying WT activity or better. Previous droplet calibrations (Figure 2-10) demonstrated that the mass spectrometer had the necessary dynamic range to see samples outperforming the wildtype enzyme, and the Z score of a reduction from 1 mM amine **1** in ivTT to 50 μ M in droplets was 0.757⁸⁰, suggesting that the screen contained the necessary sensitivity to see something better than wildtype, even if the difference between wildtype and blanks was not fully resolved.

A library of transaminase variants built from the wildtype enzyme template was expressed and screened using the MADS workflow. The linear library template material (NEST primer amplified) was distributed into droplets for ddPCR, amplified, and sorted according to our FACS protocol. 1 pg of the parent transaminase DNA was also amplified in PCR droplets containing 10 mM chlorocholine rather than neostigmine, and the droplets of parent DNA and sorted library DNA were mixed at a 1:8 ratio to create an addition pool coding for both the parent enzyme 12.5% of the time (marked with chlorocholine) and the library enzyme 87.5% of the time (marked with neostigmine). A reaction solution containing ivTT, 1 mM Amine substrate and 1 mM Pyruvate, and 0.1 mM PLP was prepared and using the reagent addition

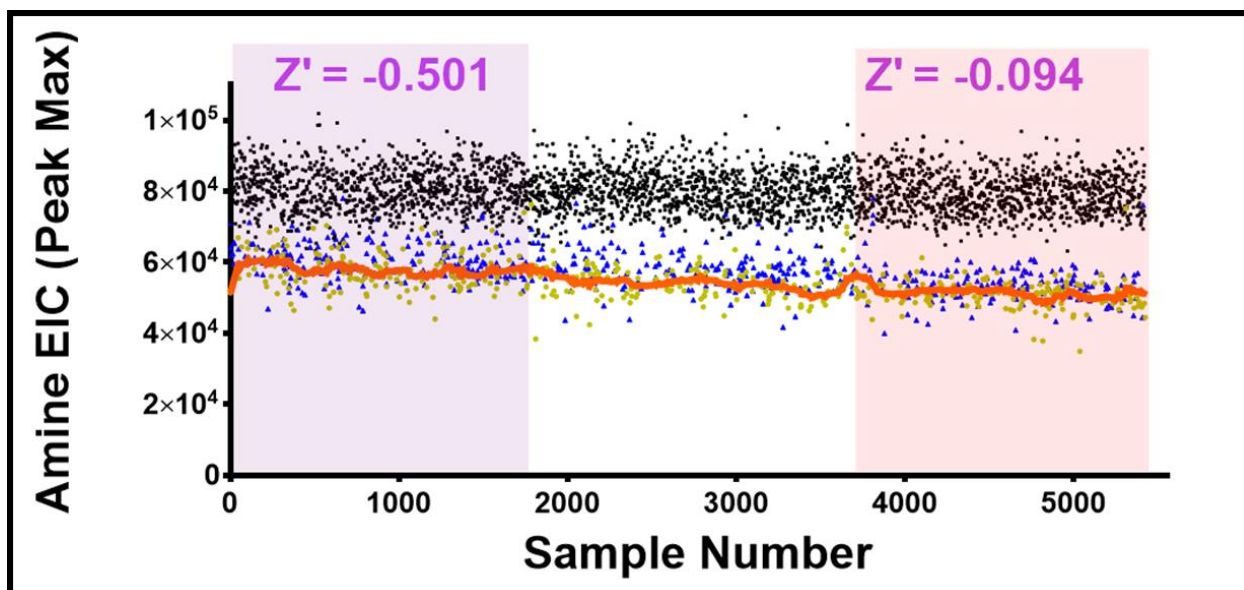


Figure 3-11: Z factor of the separation between WT and inactive droplet samples.

Raw data from figure 2-12 is re-analyzed to assess the separation between the wildtype transaminase and the blank samples at the 1 mM screening conditions read after 4h incubation. Here, yellow and blue data points express the wildtype transaminase enzyme, and the separation between these samples and the samples not expressing active enzyme is given as Z factor for two regions of the plot. This value measures how fully separated the two populations are relative to their analytical spread. Typical screening conditions are optimal when the Z factor separating negative and positive controls is greater than 0.5, and the low Z factor between the WT and negative samples tells us that the WT enzyme is not separated enough from the inactive samples during the reaction for something with WT activity to be captured without the risk of capturing false positives. Given that we were screening for enzymes more active than the WT parent, any such enzyme would be further separated from the inactive samples, and an initial screen of the WT library at these conditions was warranted.

protocol described in chapter 2.4.3, the droplets of DNA were paired and merged with droplets of ivTT. Collected droplets were then incubated 4 h and read on the MADS device.

Data from the first attempt at library expression and assay is shown in Figure 3-12. The raw signal from the amine trace was extracted for each droplet peak (See data analysis, 3.2.14), with droplet identity assigned based on the signal from neostigmine (library members) and chlorocholine (WT parent DNA). Although the wildtype enzyme clearly appears distinct from

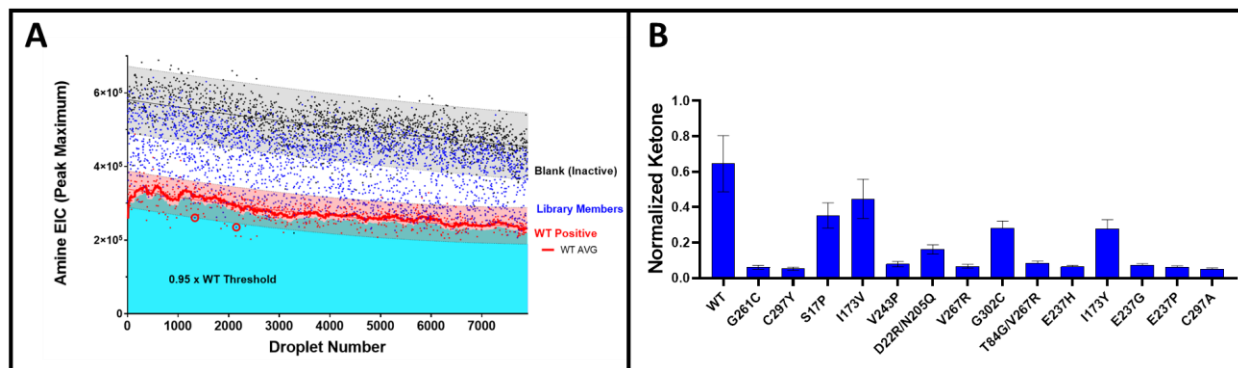


Figure 3-12: Initial Screening results from the assessment of a transaminase library

(A) Plotted amine signals for each of the approximately 8000 samples screened in the initial test of the MADS system for directed evolution. Wildtype positive control samples are depicted in red, library samples in blue, and samples that had neither chlorocholine nor neostigmine (blank) in black. 95% prediction intervals, calculated from the quadratic fit of the blank and WT control populations, are calculated and displayed on the plot as shaded regions in the color of the sample data points. These provide a rough metric of the data spread of these two samples. Only two samples appear below the 95% prediction interval of the wildtype samples, suggesting that very little, if anything in this screen is outperforming the activity range that we would expect from the wildtype samples. The red trace on this plot marks the running average of the wildtype samples in the data set, and a sorting ratio of 0.95 times the positive control average was used to collect anything below the light blue shaded region. These samples were recovered and the DNA sequenced. (B) Fourteen sequences recovered from these samples were tested in bulk ivTT reactions run for 4 h after the addition of 100 ng of DNA per 25 μ L ivTT reaction. Each sample was then infused as plugs onto the mass spectrometer and the amine concentration read. The signal intensity averages of these plugs is shown, with error bars showing the standard deviation in signal intensity (normalized to the wildtype control) for each plug (N>200 for all samples). None of the recovered sequences appeared to outperform the WT enzyme.

the blank droplet samples and almost 2500 unique library samples were observed, only 2 of these samples were below the 95% prediction interval for wildtype signals (Figure 3-12 A). During the screen of these library samples, we collected all samples that came out below 0.95x the positive control signal calculated from the running average of the last 16 wildtype signals (red line, Figure 3-12 A). These were transformed into *E. coli* and colonies picked for sequencing. Sequencing revealed 89 different variants in 141 reads, suggesting that we had not particularly reduced the library diversity in this experiment. 15 variants appeared in more than one read,

including the parent enzyme, which was in 36 colonies sequenced (25%). This was only slightly higher than the percentages observed in the single droplet sequencing (20%).

Given the lack of convincing hit variants both in the plotted droplet data and in the sequence reads, 15 unique reads (including WT) that appeared multiple times in this sequence pool were assessed for activity. None of the 14 selected variants produced more ketone than the WT enzyme when the reaction was run in bulk in a format emulating the conditions in the droplet: 1 mM amine and pyruvate in ivTT, with the read taken 4 h after the addition of 100 ng of DNA per 25 μ L (Figure 3-10 B).

3.5 Rational for Screening a Degenerate Transaminase

Results from the initial SSM screen led us to consider the question: if we have two unique enzyme variants in our microfluidic system, can we differentiate their activity in droplet format? Our initial attempt at screening had no true positive control towards which we could screen to confirm activity based selection. To ensure that enzyme variants of differing activity could be accurately differentiated in this system, we tested the capability of the MADS electrospray platform to distinguish the wildtype enzyme from an enzyme variant containing a deleterious mutation. We chose the mutation S17P, which was one of the 14 re-tested variants from our screen of the wildtype library, and showed approximately 50% reduced activity for the catalytic transformation (Figure 3-12).

We began by observing the reaction of the S17P enzyme and the WT enzyme in bulk as a function of time. If the S17P enzyme was indeed the weaker enzyme, we expected it would produce product more slowly, and reduce substrate concentration slower, than the wildtype enzyme. Data from running the two reactions in bulk using previously established methods

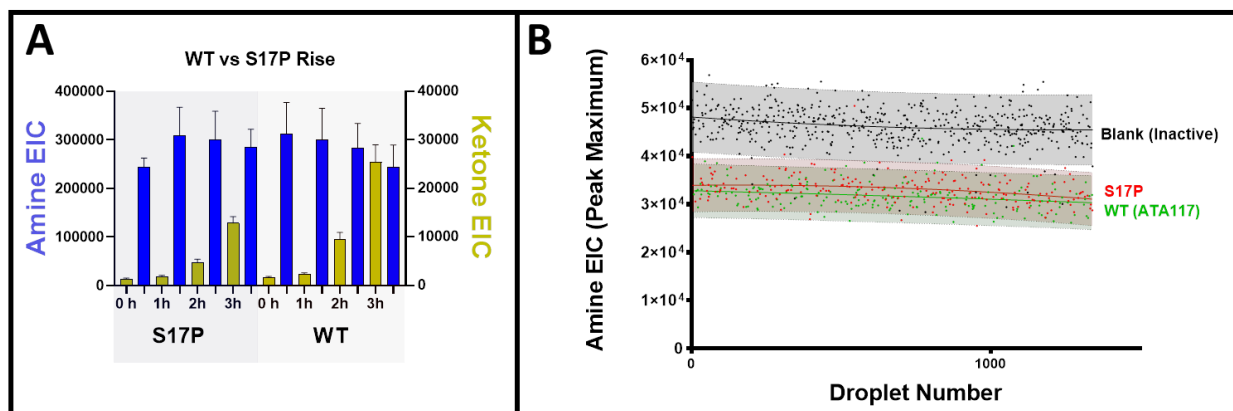


Figure 3-13: Assessment of the deleterious mutation S17P in bulk and in droplets

(A) Bulk reactions of the S17P and the wildtype transaminase variants were performed and quenched to reveal the slower product production in the S17P enzyme variant. However, while product formation rates are clearly different in these two enzymes, the relative error in substrate signal made distinguishing the two by amine **1** concentration problematic. These bulk reactions made it clear that the current reaction conditions and measurement of the substrate loss was not viable for distinguishing the two enzyme variants. (B) When the two enzyme variants were co-expressed in droplets and screened at 1 mM substrate amine after 4 h incubation, we observed similar results. Both variants show similar amine **1** levels and their 95% prediction intervals largely overlap, making amine concentration a poor metric for telling the two populations apart. We hypothesized that one potential challenge to distinguishing these two variants was the relative concentration of enzyme and substrate in the reactions.

showed that this was indeed the case (Figure 3-13A), but that at the concentrations of the previously performed reactions (1 mM amine) it was difficult to observe a difference between the two when monitoring the amine signal drop. This was also the case when the reaction was performed in the droplet format using reagent addition to droplets of ivTT (Figure 3-13B).

We hypothesized that we might be able to better distinguish the two by optimizing the reaction conditions. Typically, when biocatalysts are screened, it is necessary to optimize substrate concentration, reaction time, and enzyme concentration to ensure that a difference may be observed between two differentially active enzymes. It's common to screen reaction conditions for the optimal enzyme or substrate loading that allows enzymes to be easily distinguished. At low enzyme concentrations where substrate availability is far greater than either enzyme can fully utilize, a faster enzyme will demonstrate higher conversion of the

available substrate than its weaker counterpart because it will be capable of converting the available material at a higher rate. If enzyme concentration is too high or substrate concentration is too low in a reversible reaction, both enzymes will rapidly reach the reaction equilibrium, and will no longer be possible to distinguish. Because ivTT is being used to express droplets of DNA in the MADS system, droplet samples' protein concentration is not easily diluted for optimized screening conditions. However, substrate concentration may be increased along with reaction time in order to maximize the separation between the S17P and the WT enzymes. 1 M, 2.5 mM, 5 mM, and 10 mM amine concentrations were run and assessed both 4 h after reagent addition and 18 h after reagent addition. It was determined that at a 10 mM Amine with the overnight reaction incubation, enough difference between the activity of the two enzyme variants existed to better resolve the two populations (Figure 3-14) when monitoring substrate loss.

To investigate how fully resolved the S17P and the WT populations were with this 10 mM, overnight reaction, and see if it was possible to screen a library of S17P variants for WT level activity, we calculated the Z' for a screen at these new conditions, where the wildtype is the positive control. With the higher amine and longer incubation, the Z' of the separation between the WT enzyme and the negative control samples was 0.53 (Figure 3-14), significantly improved relative to the 1 mM, 4 h reaction conditions used originally ($Z'=-0.17$). A hit with wildtype level activity from a library of samples whose activity falls between the negative and positive control should be completely distinct from the negative control population, with more than 6σ separation between the averages of the positive and negative samples.

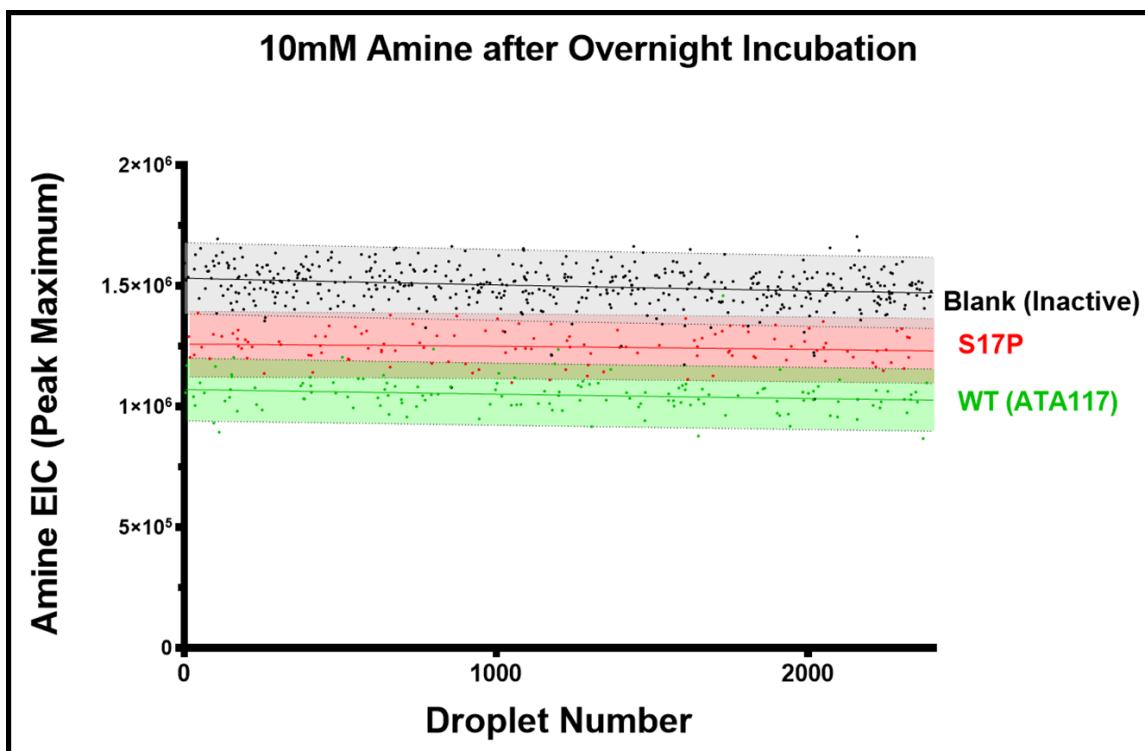


Figure 3-14: Improved conditions for screening differences between WT and S17P

Difference in activity read in droplet samples expressing the S17P enzyme vs the WT enzyme after 18 h incubation at 10 mM amine 1 in the reaction. The difference in enzyme activity read by the MADS system was tested at 1 mM, 2.5 mM, 5 mM, and 10mM amine and pyruvate, reading the reactions after 4 h and 18 h incubation. We determined that running the reaction for 18h at 10 mM amine enabled us to more clearly distinguish the activity of the S17P and wildtype enzymes. At these conditions, the two populations still overlap, but the Z' between the wildtype enzyme and the blank samples is 0.53, while the Z' between the blank signal and the S17P signal is -0.54. At this level of separation, WT level activity is clearly differentiated from blank samples but S17P is not. This distinction should be clear enough to screen a population of samples whose most active members show WT level or greater. 95% prediction intervals are plotted for all three populations, to highlight that there is still some overlap between the WT and S17P population activities.

3.5.1 Construction of a Degenerate Transaminase Library

With a procedure in place for the *in vitro* expression and activity assay of transaminase enzymes in droplets and the confirmed ability to better distinguish two enzyme variants, we next aimed to show the MADS system's ability to enrich a population of enzymes for those outperforming their peers. We now had two enzymes with a single site mutation difference that

performed at two distinct levels and enzymes with the greater of the two activities could be completely distinguished (with more than 99% confidence) from the negative.

With this in mind, we ordered a commercially built (Genscript) single site mutation library using the weaker S17P enzyme variant as the backbone. This library consisted of the same 96 targeted sites screened during our initial screen of the WT enzyme, and was made up of 1,919 unique sequences. We assembled this library into *E.coli* for quality control and sequencing, and found it contained approximately 15% parent (S17P) sequence, and 67% SSM variants (Table 3-5). This was on par with the in-house built library built from the wildtype ATA117 backbone. We performed digital PCR of the sample, as previously described.

Tier 1 S17P SSM	% (85 samples)	# Colonies
Wildtype	15%	13
Mixed Population	18%	15
Single SSM Mutation	67%	57

Table 3-5: S17P Tier 1 library sequencing

Library quality control data for the SSM build off of the S17P backbone transaminase shows the majority of sequences read show expected single-site mutations on the DNA sequence, all at targeted SSM locations. The number of single mutation sequences read in colony based sequencing was on-par with that read on the in-house built SSM on the wildtype backbone

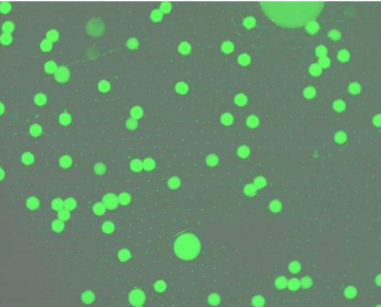
Tier 1 (S17P) single Droplet Sequencing, 100 fg		% (15 samples)	# Droplets	Fluorescent Image
Round 1 Backbone (S17P)		20%	3	
Mixed Population		60%	9	
	Non-SSM mutations in mixed samples	60%	9	
Single SSM Mutation		13%	2	
WT SSM		6%	1	

Table 3-6: Single droplet sequencing of S17P library amplified in droplets

Single droplet sequencing of 15 captured droplets from the unsorted S17P SSM amplified in droplets revealed a high rate of excess mutations on PCR amplified droplets, as had been observed previously. The vast majority of these samples had received a single template from the SSM library build, but the template received additional mutations during amplification. One of the 15 samples showed the P17S reversion to wildtype, which we believe stemmed from the presence of background WT SSM DNA in the PCR mix.

3.5.2 Single Droplet Sequencing Reveals DNA Droplet Composition

Fluorescent sorting of the PCR amplified SSM library enabled us to perform sequence reads on single, fluorescently labeled amplified droplets after digital PCR. In the case of the droplet sequences read prior to the screen of the S17P SSM library, we read 15 sequences at the dilution used for the library screening. Of those 15 sequences, 3 showed the library backbone, an observation that closely matched the read of the library prior to encapsulation and PCR (Table 3-6). 9 of these sequences had multiple mutations in the read, while only 2 had the desired single mutation to the S17P backbone. However, every single mixed sequence showed a combination of a single SSM codon change and a single non-SSM site mutation – a phenomenon which we had previously observed in the WT SSM. This result suggested that 11 of the 15 had initially received library DNA, but that many of these experienced non-SSM mutation during the digital PCR process. The 15th sequence contained an SSM mutation built off of the wildtype backbone. Investigation of the blank control samples made in tandem (Table 3-7) revealed the likely source of this WT-A257G material to be background WT SSM material. In 14 reads of fluorescent

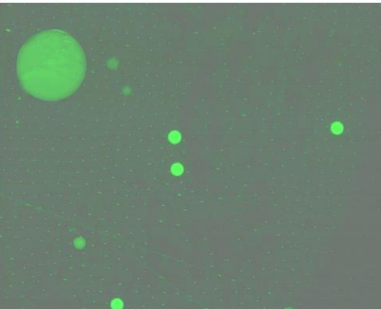
Tier 1 (S17P) single Droplet Sequencing, 0 fg (background)		% (14 samples)	# Droplets	Fluorescent Image
Round 1 Backbone (S17P)		0%	0	
Mixed Population		0%	0	
Single SSM Mutation		0%	0	
WT SSM		100%	14	
	Non-SSM mutations in mixed samples	57%	8	

Table 3-7: Droplet sequencing of S17P library blank reveals source of WT DNA

Single droplet sequencing of 14 captured droplets from the blank PCR reaction amplified in droplets revealed the presence of background WT SSM material contaminating the PCR reaction. All of these samples were WT SSM members. Like the other single droplet sequencing results, these droplets exhibited a high rate of untargeted mutation.

droplets collected from PCR samples to which no DNA template had been added, all showed the wildtype sequence at position 17, and 50% showed WT-A257G. This discovery suggested that a) our droplet PCR samples had background signal from WT SSM DNA mixed in, and b) these PCR reactions contained 5-7% WT SSM members, and ~95% S17P SSM material.

3.5.3 Single Droplet Expression and Sorting

We fluorescently sorted the S17P SSM DNA amplified in droplets and performed single droplet expression of the S17P SSM DNA, using tandem expression of S17P isogenic DNA amplified in droplets (1 pg per 20 μ L reactions) as a positive control. Figure 3-15 shows the distribution of samples in the population from 7 h of read time, during which more than 14,000 droplet samples were interrogated. As expected, the library samples in the screen ranged broadly in amine signal, from the activity of the inactive samples to activity surpassing the active “S17P” parent samples.

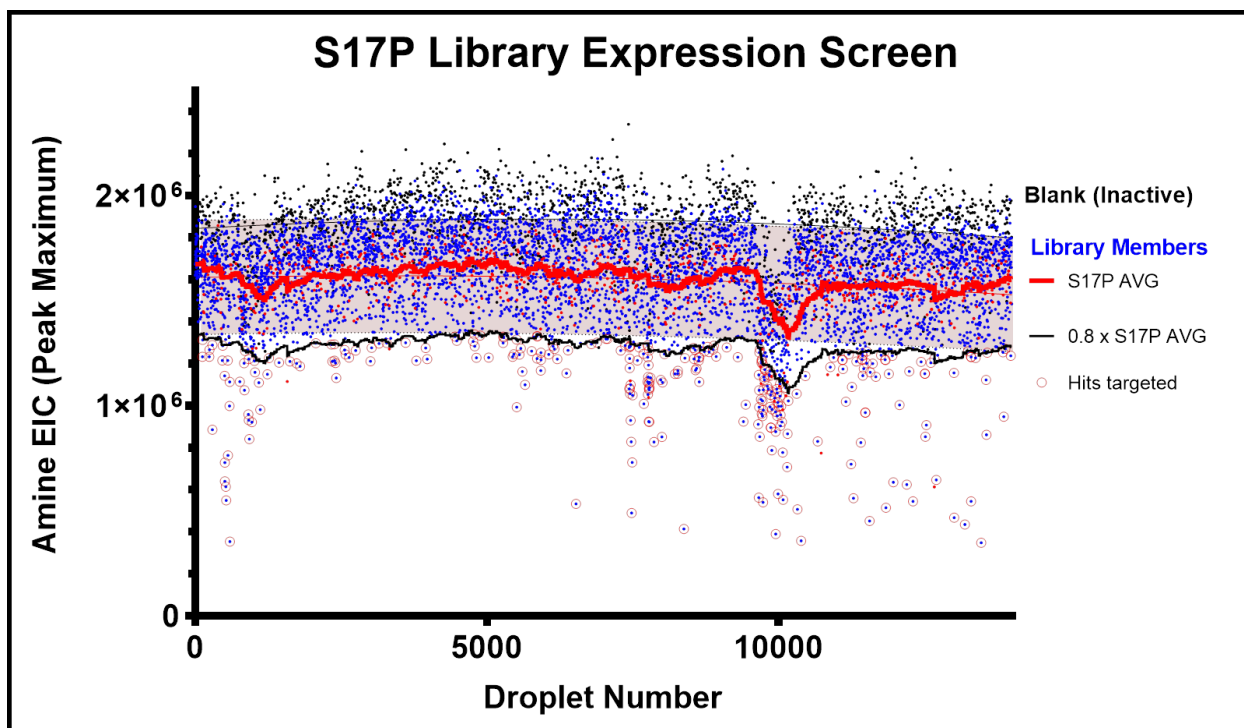


Figure 3-15: Screening results from the assessment the S17P transaminase library
 Screening data from the selection of the library built off of the S17P backbone. The running average of the S17P parent enzyme samples is shown in the red trace, with individual samples plotted in red. Library samples expressing DNA variants built from the S17P backbone are shown in blue. The shaded region of the plot represents the 95% prediction interval for wildtype samples, calculated from the quadratic fit of the WT sample data. Unlike the initial wildtype SSM screen, these reaction conditions and library members reveal samples clearly outside and below the 95% prediction interval window, suggesting the presence of enzymes expressed that are better than the S17P parent. The black trace follows the selection criteria applied to this library, which was any signal that appeared lower than 80% of the intensity of the average S17P signal. These samples are highlighted above with circles drawn around them, and represent approximately 5% of the expressed library samples.

During the screen, we used the running standard deviation from the S17P backbone average, which was actively calculated and reported by the MADS software, to set a selection threshold of 0.8x the S17P average, as this appeared to be approximately 2 standard deviations from the mean of the parent backbone sample. This threshold, denoted as a black trace on figure 3-15, consistently falls below the 95% prediction interval calculated off of the parent population (red shaded band) and served as a selection threshold stringent enough to capture enzymes

outperforming the parent S17P enzyme. The hits whose signal appeared below this level were collected into a single pool of droplets, merged, and the DNA recovered. Over the course of the screen, 286 of 5823 library members were identified as hits below this threshold, representing approximately 5% of the original population. This matches closely to the expected number of WT sequences (6%) in the pool based off of initial single droplet sequencing results on the library.

3.5.4 Assessment of Sorting Results

Droplet sorting results were assessed using both sequencing and re-encapsulation and re-screening the hit sequences. For sequencing, hit DNA was assembled into vector pUC57 and transformed into *E. coli* for picking and sequencing. Notably, 28% of all sequences in the collected pool of DNA exhibit the mutation P17S, a reversion that should have recovered the activity of the original wildtype enzyme. 52% of these P17S reversions bore the single site mutation A257G, the same mutation observed in 50% (7/14) of the single sequencing reads from the background signal. This represents approximately 5-fold enrichment from the starting distribution observed by single droplet sequencing.

Additionally, when the recovered population of samples was transformed into *E. coli*, the resultant DNA could be put back through the MADS workflow to assess the activity distribution of the population after enrichment. This material, when FADS sorted and re-expressed in the droplet format, exhibited a distinct shift in the population distribution of activities (Figure 3-16). Unlike the original droplet population, the sorted DNA shows a bi-modal distribution of activity with a shift in the frequency of more active samples, suggesting two core activity profiles within the population – that of the original and more active ATA117 enzyme, and that of the weaker S17P. This suggests enrichment occurred for the more active form of the transaminase gene

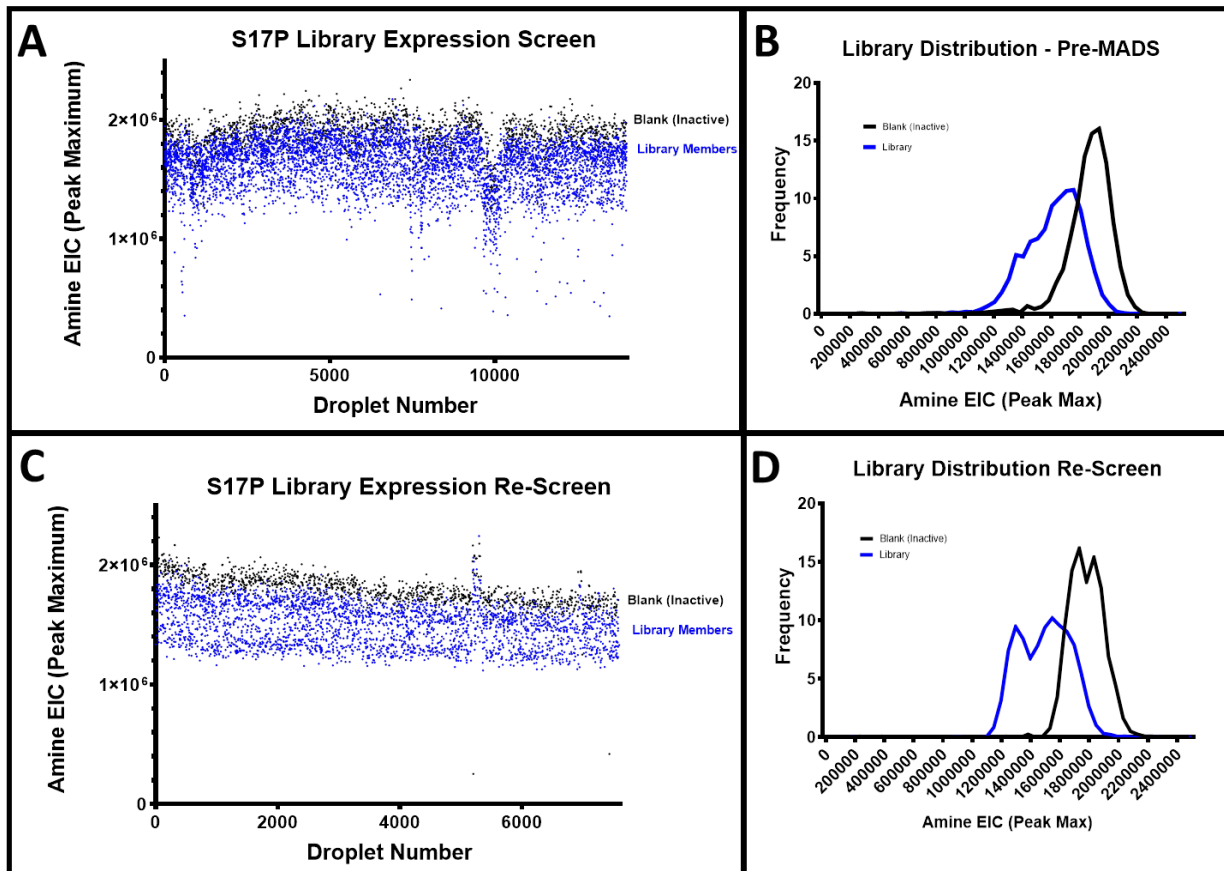


Figure 3-16: Rescreen data from the re-expression of hits from the S17P library

Library activity distribution before (A, B) and after (C, D) MADS enrichment. MADS enrichment of the samples in the upper half of the figure results in a library of samples that reveal a bimodal distribution of activity in the re-screen. This is additionally supported by sequencing data showing a higher number of wildtype mutations in the second library than in the first.

from the original droplet population, but that this enrichment was imperfect – we still captured less active variants in the hit population, likely due to the relatively wide spread of the droplet signals. The enrichment seen in the re-expression of the enzyme samples was further supported by sequencing data showing that more than 25% of the samples contained the wildtype sequence after MADS sorting.

3.6 Challenges and Prospects for the MADS system

The results described above suggest that MADS is capable of enriching a population of enzymes for samples outperforming the parent enzyme, but there remain challenges to the full implementation of this work, and opportunities for further improvement to the workflow and system throughput. In particular, single droplet sequencing has revealed that both fidelity and sample purity during digital DNA amplification was a challenge in this work, and there are additional challenges associated with the assessment of the substrate amine, rather than the product formation in the reaction. Ongoing work with the enzyme assay conditions suggests that reaction timing, substrate concentration, and reaction conditions may all be optimized to yield better MADS sorting conditions with this transaminase system. In the following section, we outline the key areas of improvement and ongoing work to address the difficulties that we have encountered thus far.

3.6.1 Single Droplet Sequences: Multiple Mutations and Background Signal

One clear concern stemming from the single droplet sequencing of the S17P library is the introduction of unexpected mutations to the library sequences between library quality control testing and reagent addition of PCR amplified DNA droplets to droplets of ivTT. Nine of fifteen sequences read in the pre-screen single droplet sequencing (60%), and 82 of 354 (28%) read in the post-screen DNA sequencing exhibited the presence of mutations that were not present in the original library pool. There are several possible causes for this change that will need to be investigated, as the unintentional addition of diversity in the library pool could prove problematic when screening for specific mutations. The first possible source of mutation in the droplets is the polymerase used to amplify the samples. Current PCR protocol utilizes a mixture of a high fidelity polymerase (Q5) and BioRad's proprietary polymerase for digital PCR. Polymerases for

digital qPCR need to have 3' → 5' exonuclease activity in order to degrade detection probes, and there is no requirement that these polymerases maintain high fidelity because their main purpose in these reaction mixtures is to produce as much product for detection as possible. If this polymerase is one of low fidelity, it could be introducing mutations into the sequences in the droplets during the PCR process. Notably, the vast majority of the unexpected/untargeted mutations observed in this work are single nucleotide mutations, which may be a result of promiscuity in the polymerase mixture. Ongoing investigation into the deactivation of this polymerase seeks to eliminate this potential variable from the droplet PCR process.

Another possible source of these mutations in the digital PCR process is the presence of SYBR green in the PCR reaction mix. Commercial SYBR Green is sold at “10000x” concentration, which appears to be approximately 10 mg/ml, a common dilution of stock DNA dye solutions like ethidium bromide¹²⁵. The final SYBR Green concentration used in this work is a dilution in the PCR master mix to 2.5x (0.2 mg/ml), and it is possible that at this concentration, the intercalation of dye into DNA as it is replicated can prove mutagenic. With single droplet sequencing possible, it would be valuable to investigate various SYBR Green dilutions for mutagenic effects, as high concentrations of the dye and its degradation byproducts have been linked to PCR failure in the past¹²³.

A third potential culprit lies in the number of cycles of PCR performed during digital amplification. High numbers of cycles have been associated with the introduction of side-products in PCR, and the digital protocol utilizes 40 rounds of amplification to generate DNA in the droplets. While gel separations of the PCR amplified DNA in droplets do not show significant secondary product bands, It would be valuable to investigate this further to determine if fewer cycles result in a lower frequency of the unintended mutations in the droplet sequences.

3.6.2 Enzyme Assay Optimization

As can be seen in the section 3.5, the key challenge to using MADS to sort a library of transaminases lay in the identification of ideal screening and assay conditions. Transaminases catalyze a reversible reaction, which means that even when running in the thermodynamic direction (as we do here) there is an equilibrium concentration past which the enzyme cannot convert more substrate into product. In typical biocatalytic reaction optimization, additional enzymes^{66,117} or extraction steps may be used to remove products from a reaction and continue to drive product formation forward, but in the microfluidic context explored in this work, such strategies are not implemented. This means that the reaction will run until it reaches this equilibrium, so the assay must be built such that it takes the encapsulated enzyme a significant amount of time to do so, giving plenty of opportunity for a faster enzyme to distinguish itself from the slower. One possible solution lies in the addition of secondary enzymes that remove product in a secondary reaction, to drive the first reaction forward.

3.6.2 Effect of Time on Unquenched Reactions

Reactions performed through the addition of DNA to droplets of ivTT are not quenched. This presents a challenge, in part because it means that reaction reads are dynamically changing as the screen progresses (Figure 3-11, 3-12). While we have partially addressed this by adding a running reaction positive control to select samples of interest, doing so has introduced its own set of challenges. We have seen this in particular when reading the reaction in droplets over extended time periods. Figure 3-17 shows the changing concentration of droplet samples over the course of a 9 h read starting within one hour of DNA addition to droplets of ivTT containing 10mM substrate. Several hours of reaction time are necessary to see a difference between the positive DNA expressing samples, and the negative samples, and that the ability of the MS to

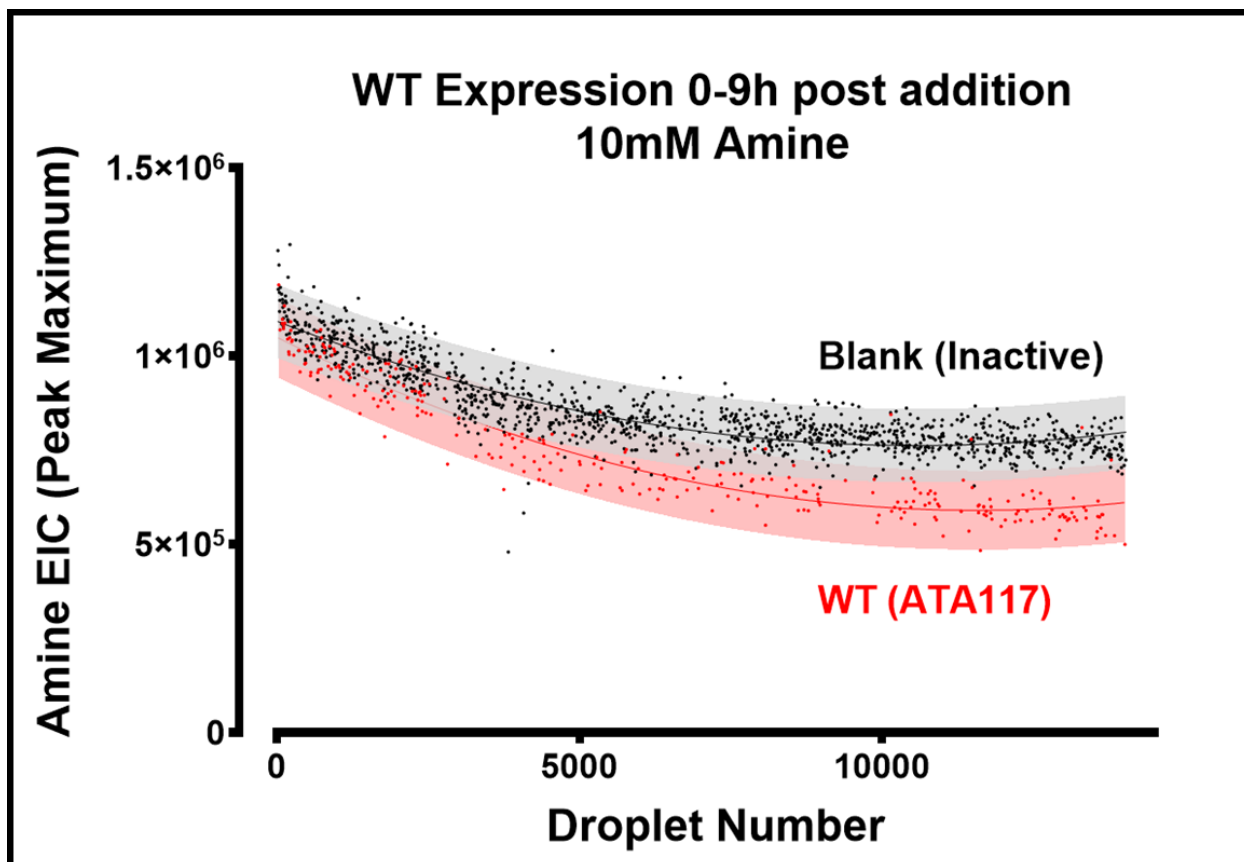


Figure 3-17: Time course read of wildtype ATA117 in droplets

A 9-hour run of the MADS system at 10 mM amine monitoring the activity of the wildtype ATA117 enzyme reveals the difficulty of screening an active enzyme reaction in microfluidic droplets. As time progresses, the WT enzyme is expressed and begins to reduce substrate concentration within the sample droplets. Screening an unquenched reaction such as this requires screening with a dynamic threshold for hits, and the screen becomes more robust as time progresses, making the selection criteria applied to the later portion of droplets inherently more robust than that applied to the earlier portion. By quenching the reaction, all droplets would be screened under the same conditions with the same criteria, regardless of when they are assessed.

distinguish these changes will increase as time progresses. Quenching the reaction could allow us more freedom to screen all samples under the same conditions, and would reduce the need to use a dynamic thresholding technique.

3.6.3 Reaction Monitoring and Chemical Crosstalk

One of the largest challenges to the success of this work has been the movement of small molecule substrates between microfluidic droplet samples. The rapid loss of distinct ketone

product signal in the droplets has led us to monitor this reaction by substrate loss, but this presents a significant challenge analytically. The choice to monitor amine in the reaction means that the difference between a parent enzyme that converts 5% of the substrate and an evolved variant that doubles this activity is only 5% of the signal read. Given typical signal RSDs of 5-10% on the droplet MS, this is a significant challenge, because it requires us to find an enzyme and screening conditions in which the wildtype enzyme may convert enough substrate to be visible as distinct relative to the blank samples, but which leaves plenty of conversion (dynamic range) possible before a better enzyme reaches the equilibrium condition of the reaction.

This is further complicated by chemical cross talk, which appears to happen with the substrate amine as well as the product ketone, albeit at a much slower rate. This leakage feeds more substrate into active droplets from surrounding samples as the reaction runs, reducing the magnitude of separation between positive and negative samples. A system in which a product that does not transfer between droplets is monitored would avoid many of these issues; a 2-fold increase in activity when monitoring production of a product is a 100% increase in signal, rather than a few percent decrease, even at relatively low conversion rates. Monitoring product formation would more cleanly separate distinct enzymes from their parent sequence, and enable more accurate assessment of the relative activities of different enzymes.

3.7 Conclusions/Impact

Directed evolution is a labor, time, and resource intensive proposition, with the largest bottlenecks being the generation of libraries of enzyme samples, and the efficient screening of these populations. Microfluidics offers the opportunity to rapidly generate large enzyme libraries in droplet format, and MADS enables label free screening of these libraries at unprecedented throughput. In this demonstration we showed that a single researcher operating

the MADS system is capable of expressing and screening more than 2000 enzyme variants for rapid discovery, all in the course of a few days. Including the library generation, sample preparation, screening, and sequencing, a full round of expression and selection may be carried out in just one week's work. Typical throughputs for directed evolution require a team of scientists to spend weeks of screening to achieve similar results.

From the initiation of protein expression to the final MS assay, the full microfluidic workflow for directed evolution using MADS takes only 24-30 h of lab time for a single researcher. In a conventional round of enzyme evolution, a single SSM screen can take a team of 4 researchers 3-4 weeks to complete, while the MADS platform allows a single researcher to perform the same work in the course of a week. Per-sample cost in MADS is reduced to \$0.03 per 30 nL sample. If these reactions were performed in well plates at 25 μ l volume (the manufacturer's recommendation for ivTT reactions) a screen of 2000 enzyme variants would require 50 ml of ivTT and cost more than \$45,000.

High throughput mass spectrometry coupled with microfluidic sample processing enables rapid experimentation in protein engineering workflows. This is an area of research that has traditionally relied heavily on time intensive molecular biology and LC based analytical workflows, which has limited the utility of the technique and slowed widespread adoption of biocatalytic synthetic approaches. The time and cost savings of performing directed evolution in droplet samples offers the prospect of significantly more expansive enzyme screening campaigns across a much broader spectrum of potential targets.

Chapter 4: Insights and Future Directions

4.1 Challenges and Opportunities for MADS and Droplet MS

The work described in the preceding sections of this thesis has covered a broad range of microfluidic techniques and devices for the production, manipulation, and analytical assessment of microfluidic droplets and their contents. While chapters 2 and 3 describe key milestones delivered from the design and preliminary application of the MADS system, this chapter seeks to (1) assess and address the key challenges and ongoing investigation in the implementation of the MADS system for the directed evolution of a transaminase, and (2) highlight the future of droplet based microfluidic systems, with an emphasis on future areas of exploration.

4.1.1 Distinguishing populations of droplets

There are two factors that play into the robustness of a screen and the ability of a screening technique to make an accurate determination of the validity of a hit, and these are intrinsic to the way we calculate the reliability of a screen with Z' factor. The first is the dynamic range of the screen – the distance between positive and negative signals, often expressed by the difference in signal means. The second is the spread of the signals at any given concentration – the sample to sample variability. The first of these may be improved by changing the concentrations of reagents loaded into the reaction, increasing the incubation time, or shifting the reaction equilibrium to drive the reaction to completion, but the second is a bit more challenging to address.

While the calibration of the MS signal presented in chapter 2 showed that the mean response of the mass spectrometer to Amine 1 in ivTT was linear with respect to the amine signal, it also revealed that the relative variance of the signal increased with the increase in analyte concentration. This is particularly evident when the 6 populations plotted in Figure 2-10 are plotted on a linear scale, rather than a log scale (Figure 4-1).

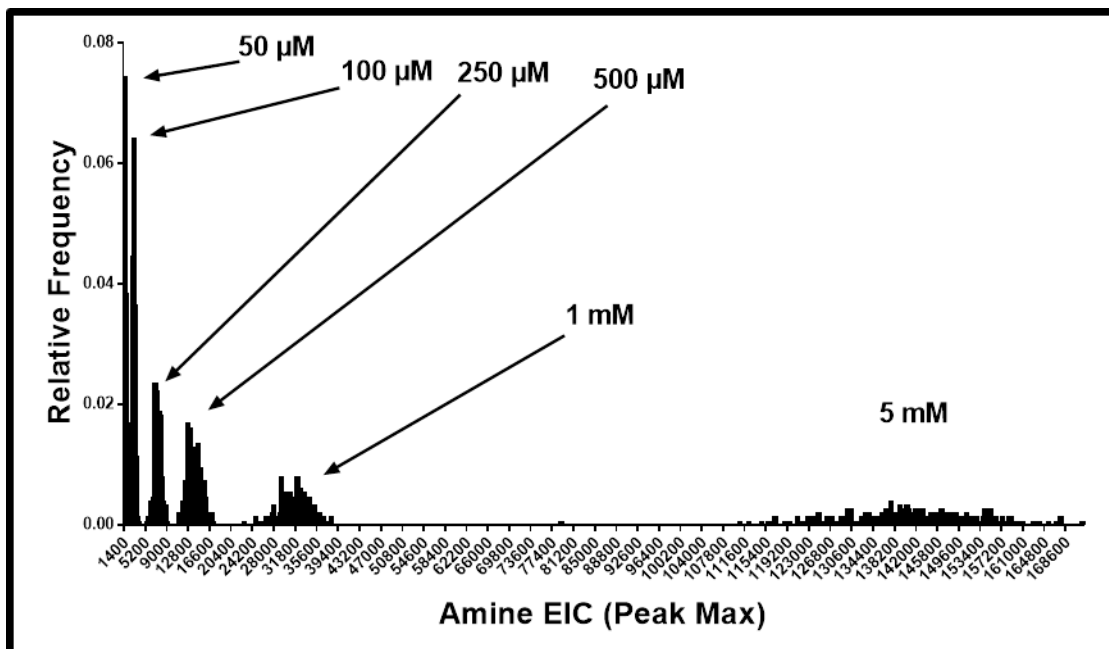


Figure 4-1: Variability in droplet signal with increasing concentration

The raw distribution of signal intensity for 225 droplets of ivTT doped with 6 concentrations of amine shows that the signal variation increases as the signal intensity increases.

If the variance was similar at all concentrations, the relative standard deviation (RSD) of the signal would decrease with increasing signal intensity (concentration) and the minimum concentration change necessary to distinguish two populations of droplets would be constant regardless of signal intensity. This does not appear to be the case. This means that depending on the intensity of the signal, the magnitude of change in average signal necessary to distinguish a second population of samples will be different.

To explore this phenomenon further, it is valuable to consider (a) how we might define full resolution of two sample populations and (b) how the signal variance changes as a function of sample concentration and signal intensity. To address the first of these, we can use the analytical description of resolution in chromatography as our guide. In chromatography, peak resolution is defined as

$$R = \frac{2(\mu_1 - \mu_2)}{W_1 + W_2}$$

where μ describes the mean value of a peak (usually expressed in retention time) and W describes the spread of that peak, as described by 2 full standard deviations in either direction from the mean. A resolution of 1, therefore, is achieved when the separation between the means of two Gaussian peaks is equal to two standard deviations from the center of the first peak plus two standard deviations from the second. In other words, when the 95% confidence interval of two peaks lie edge to edge, the two distributions are considered fully resolved ($R=1$). If we use these criteria to evaluate populations of sprayed droplets, we can calculate the necessary concentration change for full 4σ separation by considering the mean and standard deviation of our samples from the calibration data from chapter 2.

In Figure 4-2, we plot the mean and upper and lower 2σ values for the extracted ion counts reported for 225 peaks at 6 different concentrations. The increase in upper and lower bounds for the Gaussian distributions of these signals is linear at these concentration ranges, and can be fit to a line that describes the relative increase in standard deviation (2σ) expected from an increase in sample concentration. Because this signal variation scales linearly with signal intensity, it's possible to calculate the change in concentration (C) necessary to fully resolve two Gaussian distributions whose 95% confidence intervals are described by these linear functions.

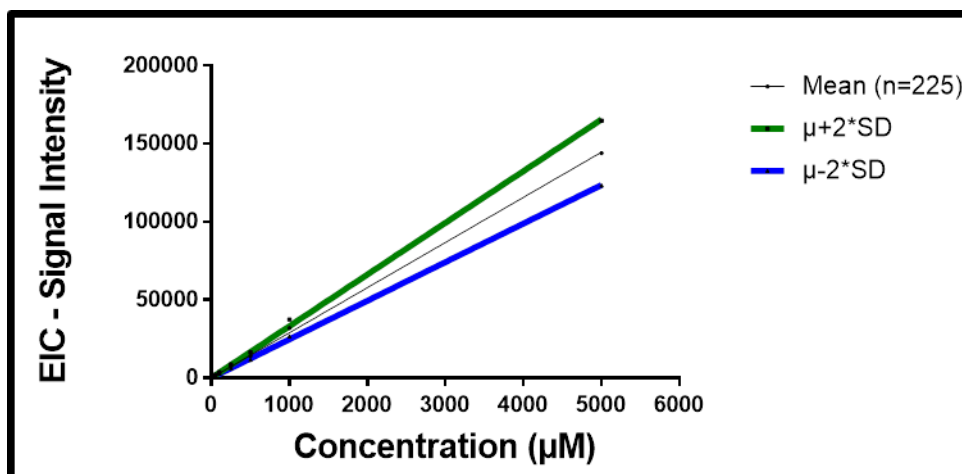


Figure 4-2: Linear increase in droplet signal variation with mean signal intensity

The standard deviation of signal from the droplet samples increases linearly with increasing concentration of amine in the sprayed droplets. This linear trend can be used to predict the variability of a population of droplet samples given a sample concentration or mean signal intensity.

For two droplet distributions to be fully resolved, the lower bound of the 95% confidence interval of one must be equal to the upper bound of the other. Because the lower and upper bounds of the confidence intervals can be described by linear equations, we can define the point where the two theoretical distributions would reach a resolution of 1, $Y_{Resolution}$, as a function of both the line describing the upper bound of signal variation and the line describing the lower bound of variation such that

$$m_{upper}(C_2) = Y_{Resolution} = m_{lower}(C_1)$$

where C_1 is the concentration of the population of droplets with mean signal μ_1 and C_2 is the concentration of the population with mean signal μ_2 . Here, m_{lower} describes the slope of the linear fit of the lower bound of the 95% confidence intervals of MS signal distribution and m_{upper} describes the slope of the linear fit of the upper bound of the 95% confidence interval (Figure 4-3). Rearrangement of this equation gives us the relationship between the slope of the lines describing instrumental variation and the minimum concentration difference between two fully resolved sample distributions, such that

$$\frac{m_{upper}}{m_{lower}} = \frac{(C_1)}{(C_2)}$$

For the calibration performed in chapter 2 (Figure 2-10) and the linear fits shown in Figure 4-2, this ratio is 1.34, and tells us that for a given concentration in a droplet to be fully distinguishable from another, the second population must have 1.34 times the analyte concentration of the first.

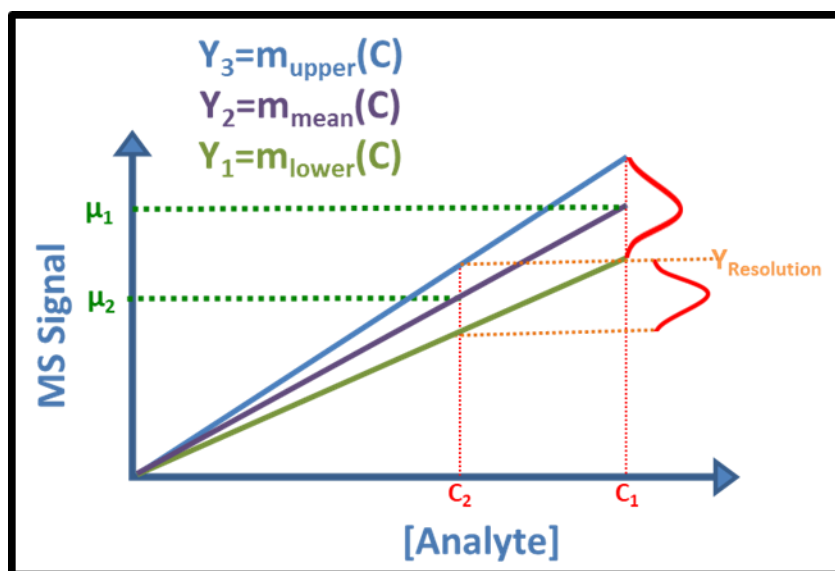


Figure 4-3: Linear increase in droplet signal variation with mean signal intensity

An annotated re-creation of the plot in figure 3 shows how the linear fit of the 2σ upper and lower bound on the calibrated droplet concentrations may be used to predict the minimum concentration difference necessary to create two populations of droplets whose distribution of signals are baseline resolved. The difference between C_1 and C_2 may be calculated by finding these values at $Y=Y_{Resolution}$ for the linear fit of the upper and lower bound of the 95% confidence intervals of the calibrated droplet concentrations. This helps conceptualize that the data distribution is going to scale linearly with concentration and that the difference in concentration necessary to resolve two populations of droplets is concentration dependent.

4.1.2 Signal Variation in Expressed Enzyme Samples

We can use the linear relationships between signal variation and the signal intensity calculated using the calibration data in Figure 2-10 to predict the signal variation for a signal with given mean intensity μ . Comparing the distribution calculated using this calibration data to

real-worked data collected while reading enzyme expressed in microfluidic droplets allows us to compare the signal variation observed during the enzyme assay to the expected signal variation from the instrument and determine if there is a significant difference between the two. When this is done for the first 100 samples expressing enzyme DNA and the first 100 blank samples of the data from Figure 2-12, the calibration-predicted signal distribution very closely matches and predicts the observed distribution (Figure 4-4). A t-test comparison of the observed and the calibration based predication of signal distribution reveals no significant difference ($P > .05$) between the distribution of signal intensity seen in the actual enzyme expressed in droplet samples, and the predicted instrumental variation at that same signal intensity.

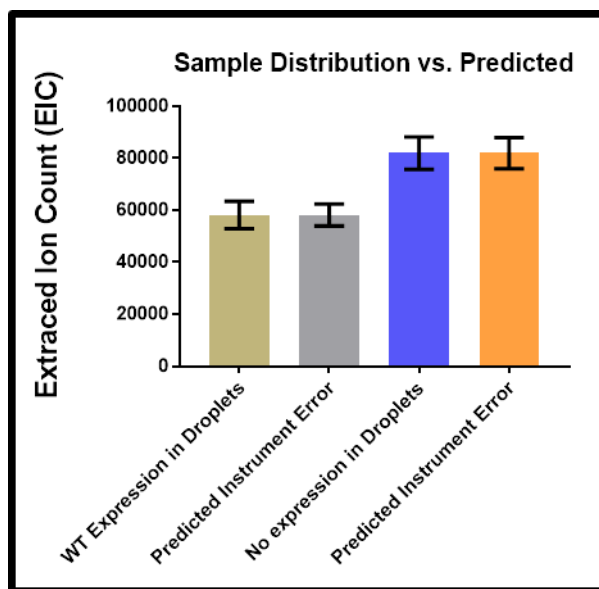


Figure 4-4: Instrumental signal variation predicts the spread of droplet data

When the calibration data from the ivTT calibration is used to predict the distribution of signal intensities at the mean intensity of the first 100 samples from the ivTT expression experiment shown on the right, the predicted data distribution closely resembles that of the observed data, suggesting that the variation seen during expression is largely predicted by instrumental variation.

This result suggests that the observed variation in droplet signal is largely explained by instrumental variation, rather than concentration variation within the population of droplets expressing transaminase. If the expression of transaminase was increasing the variability of the

droplet signals, we would expect the signal distribution to be significantly different, but the t-test tells us that our null hypothesis – that the two distributions are equivalent – cannot be rejected.

4.1.3 Chemical Transport between Droplets

Chemical transport between microfluidic droplets is perhaps the single largest hurdle to the field of droplet microfluidics. Fluorinated oils and surfactants were initially introduced as omniphobic continuous phases to better contain droplet analytes^{95,126} but researchers have observed chemical crosstalk between stabilized droplet samples, especially when stored in bulk. As the transfer of droplet analytes appears to be largely confined to small molecule analytes, this presents a significant barrier to the implementation of systems such as MADS for HTS of small molecule production.

In order to better understand the factors that give rise to droplet crosstalk, several groups have investigated molecular retention in droplet emulsions using fluorescent detection^{94,95,109,110,127-129} Buffer pH, fluorophore solubility, analyte hydrophobicity, and micellar transport have all been proposed as contributors to fluorophore loss both into the surrounding oil phase and into neighboring droplet samples. However, no definitive mechanism of molecular transport is widely agreed upon. Solutions to this issue are still in their infancy, but a number of potential options have been proposed and appear to improve retention in some contexts.

The addition of bovine serum albumin or sugars to droplet solutions has been demonstrated to increase fluorophore containment in a few examples,^{95,128} likely through the improvement of fluorophore affinity for the droplet phase and the disruption of fluorophore affinity for the surfactants used to maintain droplet separation. The modification of fluorophores with permanently charged groups, as well as the use of analytes with permanent charge, appears to discourage chemical loss between samples.¹²⁹ Indeed, the three analytes used to monitor and

identify droplet populations in the MADS system, neostigmine, carnitine, and chlorocholine, were all selected because they are quaternary amines, with permanent positive charge. This permanent charge made them easy to detect, and no chemical transfer was observed between them, despite extended incubation and operation times with droplets in close contact (Figure 4-5). Nanoparticle and dendritic¹³⁰ surfactants have also been shown to reduce micelle formation and discourage surfactant mediated transport, leaving carrier phase solubility as the only driving force for chemical loss^{108,109}

Ultimately, significant investigative work still needs to be done in order to rigorously define and address the true driving mechanisms behind chemical transport between droplets. Fortunately, droplet mass spectrometry provides a unique opportunity to investigate molecular transport in droplet systems. Continuous infusion of droplets stored in bulk has enabled us to

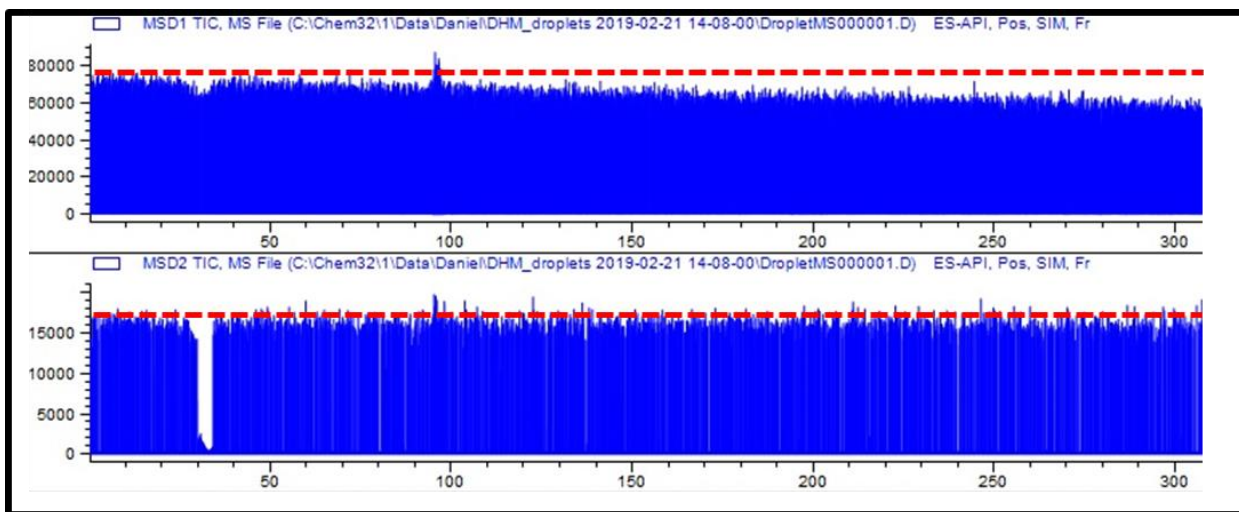


Figure 4-5: Comparison of analyte retention within droplets

Raw MS traces are shown for a 5-hour run of droplet mass spectrometry reveals the gradual drop in signal of amine **1** (upper trace) while the signal of chlorocholine (lower trace) remains at a consistent intensity. In this experiment, only 20% of samples contained chlorocholine, and more than 20% of samples contained amine **1** with no active enzyme, but all amine signals dropped over the 5 hour run, suggesting the leakage of amine from high concentration droplets to low concentration droplets containing active and expressed transaminase

monitor reaction progress over time, and we've observed the leakage of small molecules in these systems over the course of extended experiments as well (Figure 4-5). Prior to the advent of this label free assessment technique, most of the analytes investigated for retention in droplet samples were fluorescent materials that could be tracked with optical detection. Mass spectrometry has allowed us to track the movement of materials across droplets, and offers a unique opportunity to characterize the nature of chemical transfer, and ultimately develop better, more broadly useable surfactant materials. This will be critical to the potential for systems such as MADS to be broadly useable across applications and chemical systems

4.1.4 Quenching Reactions in a Microfluidic System

In typical enzyme screening workflows, reactions are quenched in order to provide an identical basis of comparison for all samples assessed. This typically requires the addition of reagent to rapidly denature proteins in solution and stop the activity of the enzyme. In our microfluidic workflow, we opted not to perform this step and instead assessed reactions as they ran in real time. This creates two potential issues – if some proteins are expressed more rapidly than others, they may appear to be more active in this system simply due to this, and because the reaction is running in real time, selection thresholds must adapt as the screen is running to compensate for the expected activity of the average sample.

The first of these two issues is one that is common to protein engineering workflows performed in cells. Expression hits that are more easily produced by cells can appear to be more active than their less easily expressed counterparts. Fortunately, ivTT is unlikely to have this issue, because the expression rate of DNA *in vitro* should be relatively similar across all variant DNA where the difference between two DNA variants is just a single codon, and cytotoxicity is a non-issue. Variability in the start of the expression process also appears to have a relatively

minor effect on signal spread in the droplets. Even when expressing the same DNA across all droplets (such as was done in figure 2-12) where each droplet received DNA at a different time over the course of reagent addition, the signal variability does not seem to be greater than what would be expected from instrumental variation alone.

The second of these challenges is perhaps more concerning - in an unquenched reaction, hit identification is complicated by the activity of the samples throughout the screen. If we choose to assess all samples based on the same thresholding criteria (as was done in the original testing of the MADS system), samples that are read later in the screen have more time to react, and therefore are more likely to achieve the criteria necessary for selection and recovery. This means that we risk type 2 errors (collecting false negatives) for the first half of the screen, and we risk type 1 errors (collecting false positives) in selection for the second half of the screen (Figure 4-6 A). In the case of the dynamic thresholding used in MADS, we chose to select samples based off of a moving threshold (determined from the parent enzyme activity) to minimize the chance that we would collect a false positive or negative sample. This significantly reduces the amount of screening time during which we risk a false selection, but does not fully resolve the issue because the threshold activity will eventually reach the maximum activity of the screen, after which it cannot differentiate a hit. In this setup, the dynamic range of the reaction (the distance between a positive and a negative sample) changes as a function of time. Initially, the dynamic range is small and the activity difference between a parent enzyme and an improved variant is low. With time, the activity of the improved variant becomes distinct from the parent, but as activity slows and the enzyme reaches static equilibrium in the reaction, the parent enzyme catches up and the distance between the parent and the faster enzyme falls to zero (Figure 4-6 B). This means that

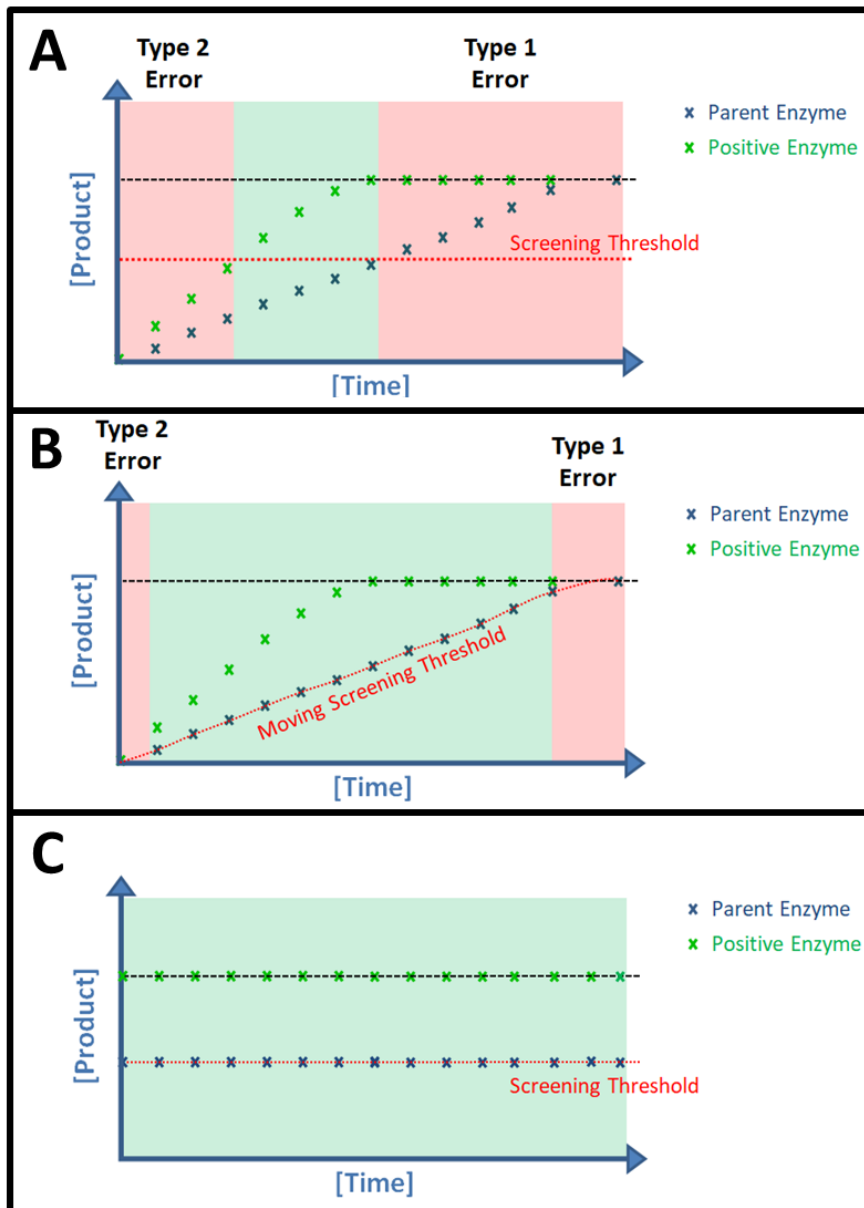


Figure 4-6: The benefit of a quenched enzyme reaction

Theoretical plots of enzyme activity in unquenched reactions reveal the limitations of (A) static thresholding for activity selection in an unquenched reaction, (B) adjusted thresholding for activity selection in an unquenched reaction (as performed in the MADS system), and (C) quenched reaction screening with static thresholding. In a quenched reaction, the difference between a positive and a negative variant is always the same, and selection may be undertaken without the risk that changing enzyme activity will affect the screening results. In all three plots, ideal screening conditions are highlighted by green shading, revealing that only the quenched reaction condition enables error-free screening for the entirety of the screening period. Type 2 errors (false negatives) and type 1 errors (false positives) are far more likely in unquenched reaction screening.

the confidence with which hits are chosen will be highest in the midpoint of the reaction, and will fall off preceding and following this point.

In a quenched reaction, the risk of a false positive or negative selection is independent of the screening time, and the dynamic range is static. By quenching the reaction at the time when a positive and a negative control sample are most distinct, we maximize the selection window's range, and we enable the reaction to be screened in perpetuity. Figure 4-6 gives a theoretical comparison of all three selection methods. Due to the additional complexity of quenching microfluidic reactions in droplets, we chose to use the second method for our screen, but the relatively high variation (RSD was between 6% and 9% for the calibration in Figure 2-10) in signal for droplets of the same concentration limited the screening window to the middle period of the reaction. Under this mode of operation, false negative and false positive selection became more likely early and late (respectively) in the screen. Given the necessity of a screening period of several hours to achieve the library coverage desired in a full SSM screen (10^3 enzyme variants with at least 3 replicates), the MADS system currently benefits from screening slower reactions and wider dynamic ranges.

One solution to this challenge is to perform reagent addition a second time in the microfluidic workflow, after the addition of DNA to individual droplets, and after the expression of that DNA. This would be possible to implement using picoinjection¹³¹ (or K-channel picoinjection¹³²) reagent addition, which we pursued early in this project as a potential method of addition to microfluidic droplets. In pico-injection based reagent addition, droplets are re-injected and flowed past a perpendicular channel filled with aqueous reagent connected to a

pressurized vial. By closely matching the applied pressure in this reagent stream to the Laplace pressure of the oil and water interface of the injection channel and the cross flow, the interface

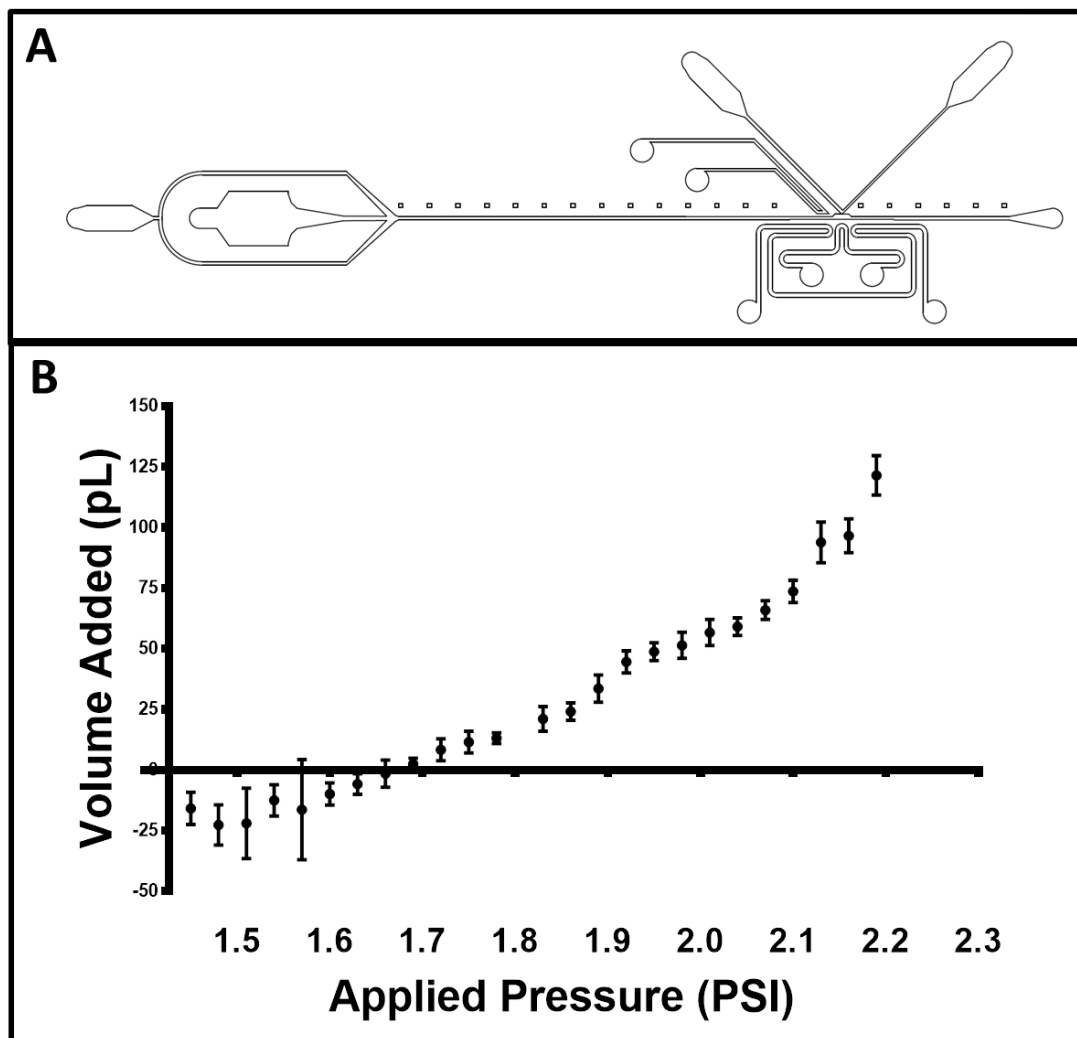


Figure 4-7: K-channel reagent addition enables scalable addition to all droplets

K-channel architecture¹³² on a microfluidic device may be used to add reagent from a continuous stream to a series of re-injected droplets as they pass the picoinjection region on the device. In (A), a schematic of a device for the processing of 160 pl droplets is shown. The device is fabricated to 50 μm depth with 50 μm channels and a 10 μm interface between the reagent cross flow (V-shaped channel) and the droplet stream (horizontal channel). (B) Data from experimentation with the device shows that as the pressure applied to the reagent cross flow rises, the volume added to the droplet rises. Such a system would allow the chemical quenching of in-droplet reactions after expression and activity assay, and may represent an opportunity to improve the quality of assays performed in microdroplets.

may be maintained without producing any droplets of reagent in the cross flow. When aqueous droplets flow past this architecture, a high voltage electric field applied to parallel electrodes on the device destabilizes the surfactant which otherwise prevents the droplet from merging with the added reagent, and allows the injector to add reagent to the droplet. We initially implemented such a system for reagent addition during the early work on the MADS project, using the K-channel architecture to add up to 130 pl to droplets of approximately 160 pl in a device designed with 50 μm square channels (Figure 4-7).

One of the challenges to implementation in this sort of system is the stability of the oil/water interface in a pico-injection system. Because the pressure of the system fluctuates as the droplets flow through the channel,¹³³ the interface between the oil and the aqueous solution in a picoinjector must move to adjust. To minimize this movement and avoid allowing the solution to bulge too far out into the cross flow and create satellite droplets in the microfluidic system, the Laplace pressure differential across this interface needs to be high. Because Laplace pressure of the oil/water interface scales inversely with the radius of curvature of that interface, the connection between the two channels must be kept as small as possible to ensure as small and highly curved of an interface as possible. While this is relatively easy to do at smaller fabrication scales, as the channel size increases to accommodate larger droplet volumes, the stability of the picoinjector decreases, and its sensitivity to minor changes in pressure on the reagent vial increases. This means that minor changes in applied pressure, liquid height in the vial, and vial height relative to the device can have large scale effects on the consistency of the reagent addition. Figure 4-8 shows a series of images of the output of a scaled up reagent addition device operating over a 25 minute period. This device scaled channel width to 100 μm

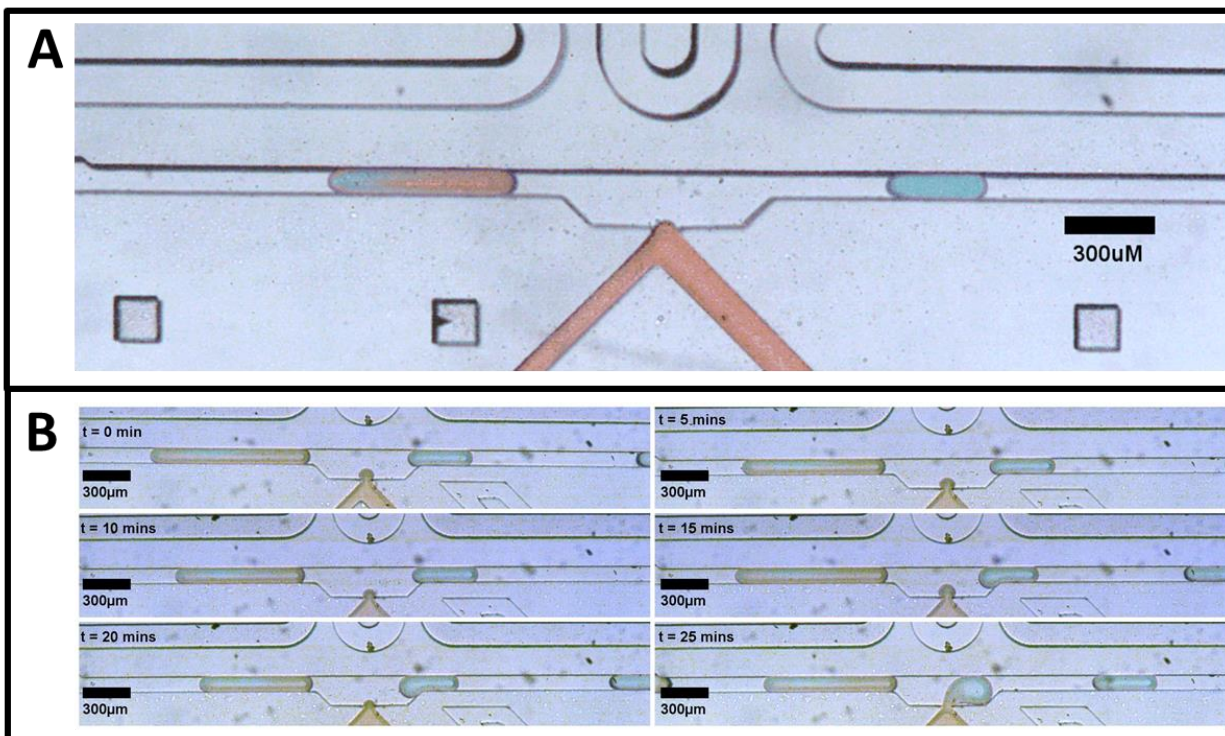


Figure 4-8: Drift in K-channel operation at nanoliter scales

(A) K channel geometry is used to add 3 nL to a 3 nL droplet in continuous flow. This scaled device is capable of processing droplets that are on the same volumetric scale as the samples screened in MADS, and holds the potential for use as a reaction quenching mechanism for assays carried out in droplets. However, in order to fully implement such a system, it will be first critical to address the drift in added volume observed over extended operation of the device (B). This variation is likely due to the large (100 μm) interface between the cross-flow and the droplet injection channel, which is sensitive to relatively small pressure changes in the system during operation.

and height to 100 μm. This 400% increase in channel cross sectional area was sufficient to take 3 nL droplets and more than double their volume (up to 8nL) but the channel expansion made stable operation of the picoinjection platform much more challenging. The longer the device operated, the more the vial holding the injected liquid drained. As this occurred, the effective pressure on the injected reagent drops and the volume injected fell in tandem. This may be possible to buffer against by using high resistance (back-pressure) fluidic lines or larger volume source vials, but more investigation and optimization will be necessary for the implementation of a picoinjection device for larger, 30 nL reaction droplets. Alternatively, digital PCR, fluorescent

sorting, and ivTT expression could be scaled down to pl volumes, and the resulting library could be quenched at picoliter scales. The resulting material could then be diluted in larger droplets for analytical assessment.

4.1.5 Data density and MS Cycle Time

Another key challenge in the MADS system is the data acquisition rate of the mass spectrometer. Under current operating conditions, the minimum dwell time for each ion channel on the mass spectrometer is 0.005 s. This means that the instrument takes 200 reads per second, divided over 4 channels for a total of 50 data points per channel per second. At current flow rates and oil spacing, droplets are sprayed every 1.4 s, and spaced with oil at a 10:1 ratio. This means that droplets dwell at the tip for 0.13 s, allowing approximately 25 points to be taken across the four MS channels. This means that at the most, each channel gets approximately 6 data points across the droplet (Figure 4-9). The microcontroller running the MADS system extracts the maximum data point read during the course of a single peak and utilizes this data to make a sorting decision, but this means that the droplet MS is sensitive to variations in signal intensity that occur over the course of those 6 samples.

Faster sampling rates in the mass spectrometer may help combat some of the variability across droplet reads, especially with the implementation of average peak height detection and reporting, rather than maximum peak height detection. Slowing the flow rate, and reducing the spacing of droplets in the microfluidic system may also help this, as doing so would allow the droplet volume to be detected over a longer period of time, and provide greater numbers of samples across a droplet. Improvement in the instrument sampling and scanning rate is certainly feasible. Recently, Waters collaborated with Kempa et al⁶³ to read as many as 30 droplets per second on a mass spectrometer. Ultimately, improvements in instrumentation and MADS

technique should be capable of yielding reliable, data-rich reads of nanoliter and picoliter scale droplets, enabling MADS with greater reliability.

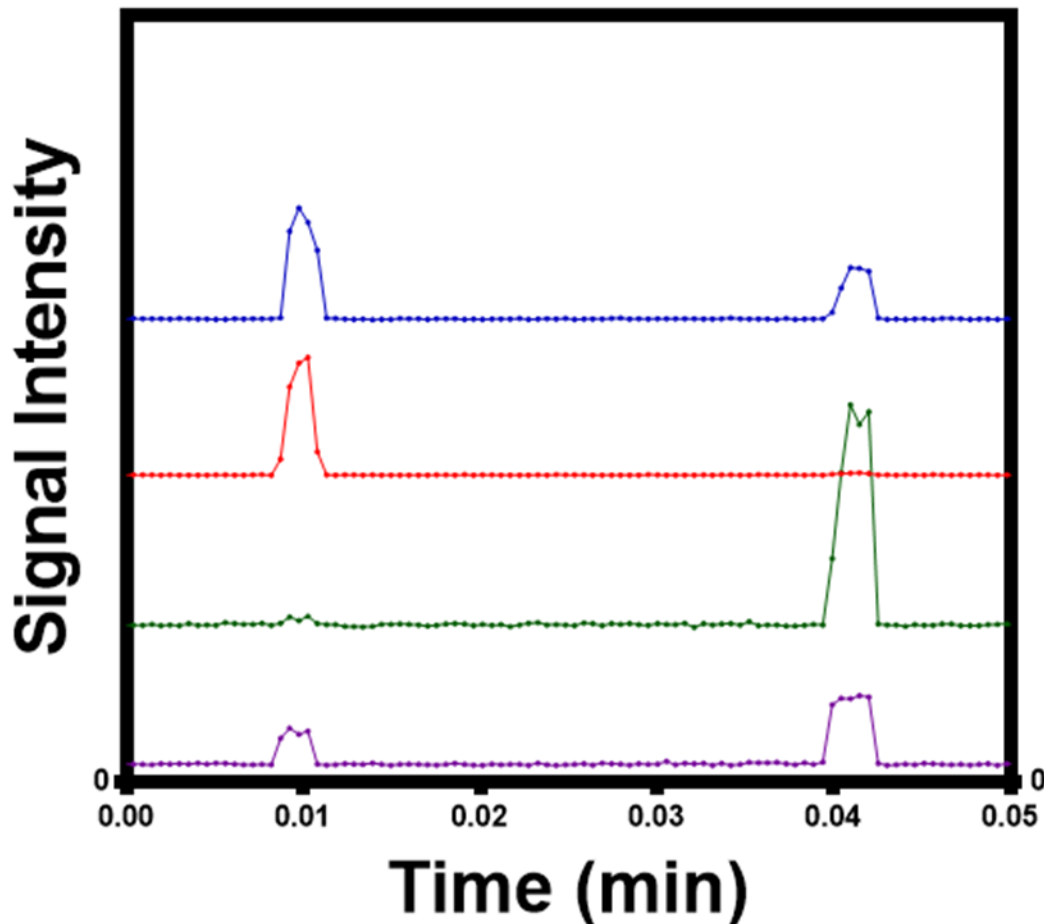


Figure 4-9: Sampling frequency on an Agilent 6120B monitoring 4 ions

A sample trace from 30 nl MADS droplet screening shows two droplets read by the mass spectrometer. Individual signal reads are marked to highlight the relatively low number of reads collected across individual droplet samples. In this case, only 5-6 reads are taken across a droplet. This limits the analytical robustness of MADS. Faster sampling mass spectrometers and the use of average peak height rather than peak maxima to assess droplet concentration may improve the robustness of the MADS technology.

4.1.6 DNA Loading

One ongoing challenge to the MADS system is the isogenic loading of droplets with DNA for expression. Early on in our work, we proposed the use of one-bead-one-gene libraries,

and we optimized microfluidic devices for rapidly sorting droplets for those with encapsulated magnetic beads (Figure 4-10). This device allowed us to generate and sort 50 μm (65 pl) droplets containing 10 μm magnetic beads (Spherotech) at greater than 500 Hz, rapidly enriching for singly-encapsulated beads in droplets. The original intention was to functionalize these beads with DNA and then pair them with ivTT for expression, but we found the digital PCR process to be more amenable to early implementation, and ultimately did not pursue this system. However, DNA functionalized beads used in ivTT expression have recently been reported for the droplet based expression of an oxidoreductase,¹³⁴ and have been demonstrated in 30 nL droplets.

The use of microbeads for enzyme expression could be a significant improvement relative to single droplets of amplified DNA for several reasons. First, the microbeads cannot merge and combine with one another like droplets may, reducing the potential for singly captured genes ending up mixing prior to addition to ivTT. Additionally, microbeads are easier to capture, and could be separated out into individual wells or locations after sorting, maintaining

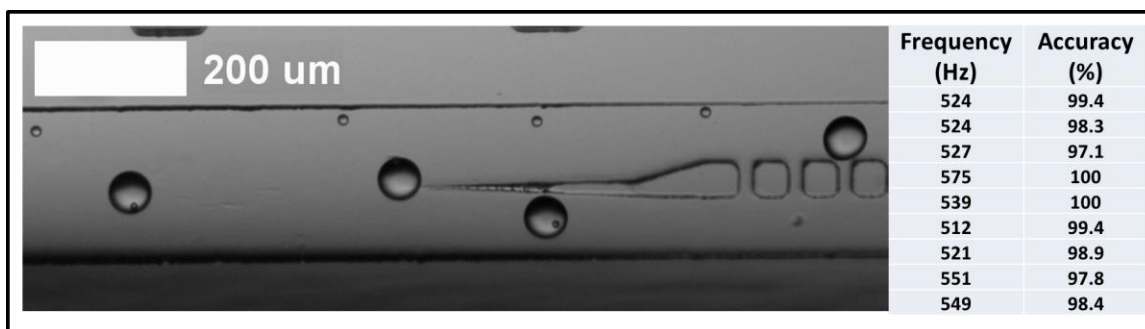


Figure 4-10: Magnetic sorting of 65 pl droplets

A micrograph and data from 9 experimental runs of a magnetic sorting device for 65 pl droplets containing 10 μm magnetic beads. Upstream, droplets are generated from an aqueous suspension of magnetic beads in a 60 μm x 60 μm flow focusing cross (SU-8 patterning height 50 μm). Droplets flow into a 200 μm sorting channel located 200 μm from a permanent N42 neodymium magnet. A tapered, offset pillar array enables droplets pulled slightly (~ 10 μm) off center by the magnetic field to be sorted into one of two exit channels, enabling rapid enrichment of beads in droplets.

the digital character of the hit material, even after de-emulsification. There remains significant potential in the development of the MADS system around bead-based gene libraries, and much of the groundwork for this has previously been laid.

4.2 Alternate Applications for the MADS Platform

MADS screening offers an opportunity for rapid, label free screening at the nanoliter scale. In this thesis work, we have applied this system to performing directed evolution; however, we envision the MADS platform to be amenable to a broader range of screening methods as well. Recently, FADS based screening has been implemented in the high throughput screening of DNA encoded, bead bound libraries of small molecules. DNA encoded libraries are gaining popularity in discovery chemistry as potential solutions to scale down the size and increase the throughput and the chemical space accessible in compound library screens.^{135,136} Using split-pool synthesis¹³⁷ techniques coupled with DNA oligo construction on the same scaffold, a library of small molecules may be constructed such that each unique chemical is associated with a genetic sequence. In a recent report, Cochrane and coworkers created a DNA encoded library of small molecules on solid support (microbeads) to screen for enzyme inhibitors⁸⁵. The resultant beads contained both a DNA oligo coding for the synthetic split and pool sequence, as well as the bound small molecule. Beads were encapsulated into droplets of enzyme and substrate, and upon UV irradiation, the bead-bound small molecule could be released from the bead's surface and into the droplet within. If the activity of the probed enzyme was reduced by the released small molecule, it could be detected by a fluorescent sorting device as a decrease in the relative fluorescence of the droplet, leading to capture of the droplet. Following sorting, beads were sequenced to reveal the small molecule library members responsible for this inhibition.

One potentially attractive direction for the MADS platform is the performance of similar small molecule screens for discovery chemistry. MADS provides the distinct advantage of being able to probe virtually any reaction that may be detectable by MS, eliminating the need for a fluorescent product. Upon generating a DNA encoded library on small magnetic beads, similar small molecule dosing could easily be achieved for the droplets entering the MADS device.

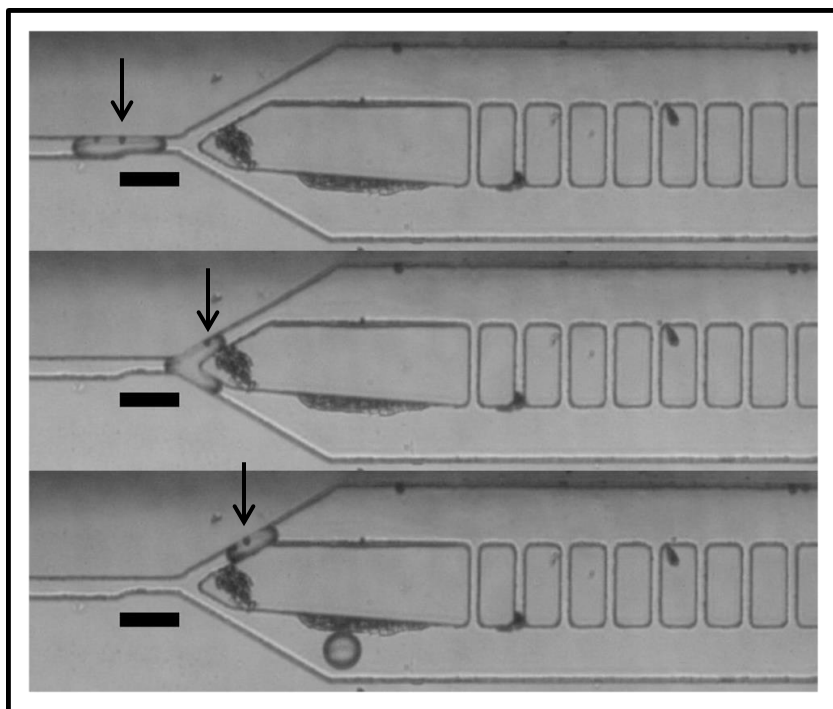


Figure 4-11: Magnetic biasing to recapture beads in split 100 pL droplets

This series of images shows a droplet splitting device dividing a droplet containing a 10 μm magnetic bead (highlighted by black arrows). Because the split is performed in close proximity (200 μm away) to a permanent magnet, the bead is biased into the upper channel every time. This technique could be used to recover a bead containing a DNA barcode from individual droplets.

These droplets could easily be sorted based on bead content using our existing magnetic bead-based droplet sorter. The bead can subsequently be recovered using magnetic field assisted pull down during the splitting step of MADS (Figure 4-11). This could assure that the bead would not be lost to the mass spectrometer, and hit beads could be recovered and sequenced as described

by Chochrane and coworkers. A MADS system such as this could potentially screen more than 100,000 library members in a week, at the cost of a few milliliters of reagent and a few milligrams of magnetic beads. Such a system would be easily scaled to screen the same library of bead bound compounds against hundreds of targets, with no modification to the system apart from the input reaction mixtures containing the target enzyme system.

4.3 Improving the throughput and sensitivity of droplet MS

NanoESI-MS (nESI-MS) would be a valuable tool to apply to the MADS system. Recent work from our group has demonstrated nESI-MS of droplets in the picoliter range, and at these low flow rates and microscopic spray tip dimensions, droplet electrospray has proven to be far more sensitive, achieving low micromolar to nanomolar limits of detection.⁶¹ In addition, nESI-MS has been demonstrated at throughputs up to 10 samples per second and could prove readily amenable to higher throughput screening.

In order to incorporate nESI-MS with MADS, device scale down might be necessary to produce smaller sample droplets that may be rapidly introduced into the MS spray tip. At smaller droplet scales, more droplets may be infused at the same flow rates, and nanoliter per minute flow rate ranges may still be used to sample multiple droplets per second. To perform MADS at these smaller scales, each of the devices for the MADS unit operation would require scale down to dimensions in the 25-50 μm range (channel width) from their current dimensions to accommodate the smaller droplet sizes. Fortunately, many of the MADS component devices were in fact initially developed for this scale of droplet manipulation, and these existing designs could be easily applied to testing nESI-MADS. MADS at picoliter scale is certainly within reach in the near term; indeed, the company Sphere Fluidics, based in the UK, is currently in the process of developing one such system.

4.4 Future Directions for Droplet Infusion

The MADS workflow provides many key advantages in terms of reduced sample usage and improved assay throughput, but there will inevitably be some reactions that remain better suited to well plate formatted screens. Complex biochemical and organic synthesis reactions with numerous reagent addition steps are difficult to perform in droplets, as each microfluidic unit operation introduces added variability in droplet size, contents, and stability. As such, automated workflows that require minimal manual manipulations appear to be better suited to these droplet strategies. Moreover, in several applications, there is significant value in using droplet analysis for rapid assessment of well plate samples, as demonstrated by Diefenbach et al.⁷⁷ In such instances, continual sampling of droplets from a well plate into a mass spectrometer is of significant interest. To enable continuous sampling, there are two strategies that we have preliminarily investigated and which would benefit from dedicated efforts for implementation.

4.4.1 Droplet Infusion Using Peristaltic Pumping

From a setup standpoint, one of the simplest ways to achieve continuous infusion into a mass spectrometer involves using a flow-through pumping system, such as a peristaltic pump. This would allow samples to be simultaneously pulled from a well plate and pushed onto a mass spectrometer in continuous flow, without the need to disconnect and reconnect tubing from sampling to infusion (Figure 4-12). This was preliminarily investigated using a Gilson Minipuls peristaltic pump to infuse samples onto a microchip capillary electrophoresis setup. In this setup, a Gilson Minipuls pump operating at 1 RPM was used to withdraw ~20 nL samples of 1 μ M “SDHA,” a short peptide derived from succinate dehydrogenase (5FAMGGQSLK(succ)FGKG).⁵¹ These droplets were pulled from a wellplate into 360 μ M OD, 150 μ M ID teflon capillary, which was in turn connected to the inlet of the 250 μ M ID PVC

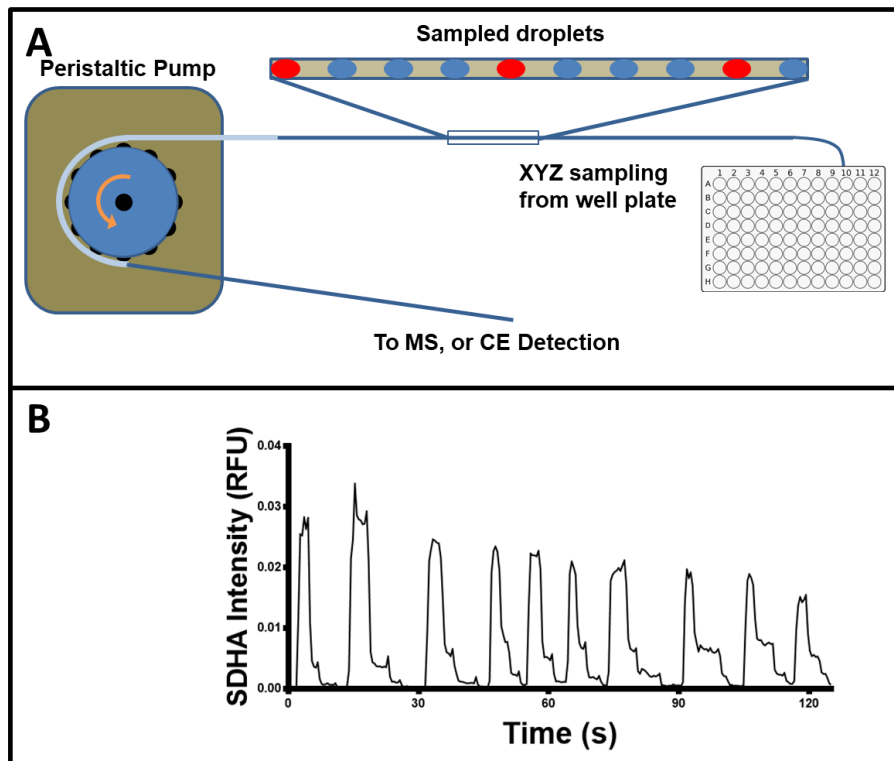


Figure 4-12: Peristaltic sampling of microfluidic droplets from a well plate

(A) The experimental setup for a preliminary test of peristaltic sampling from a well plate. A peristaltic pump was used to pull nanoliter samples from a well plate containing blank and fluorescent aqueous samples onto a Teflon capillary. These samples are drawn through the flexible tubing of the peristaltic pump, and then exported off of this tubing into Teflon capillary. This capillary may then be set up to transfer the droplet samples into a device for analytical assessment. In the case of the proof of concept work shown here (B), droplets containing a 1 μM solution of the peptide 5FAMGGQSLK(succ)FGKG were introduced into a microchip CE system.⁵¹ Each sample droplet was followed by four wash droplets, and 5 samples were read each minute using LIF detection of the peptide after a 0.25s separation on the microchip CE device. Peristaltic sampling offers the opportunity to perform continuous flow infusion of samples from source to detection.

tubing fitted to the peristaltic pump. As the pump operated, an XYZ positioner delivered the trailing end of the capillary into wells on a plate of samples covered with a layer of perfluorodecalin oil. Droplets were pulled up from the plate as the positioner moved from well to well, separated by carrier oil. These droplets were then transferred through the pump tubing, and out into another short length of Teflon capillary running to a microfabricated CE chip. Preliminary data suggested that samples could be infused and cleared from the system every 15s,

suggesting the potential to use the same system to infuse droplets into a mass spectrometer instead.

The largest hurdles to the implementation of a peristaltic pump sampling system are likely to arise from complexities in the surface chemistry of the peristaltic pump tubing and the pulsatile nature of peristaltic pump flow. Peristaltic pumps rely on deformable tubing in the rotor region to drive flow through the system. This tubing is often made of materials like polyvinyl chloride (PVC) that can withstand repeated deformation without wear, and these materials are more susceptible to analyte retention and sample-to-sample carryover than fluorinated Teflon tubing. However, carryover can be minimized through the use of periodic wash droplets, and the simplicity of such a setup might outweigh the drawbacks of potential sample carryover.

4.4.2 Continuous Droplet Infusion Using Pressure or Vacuum

Another possible alternate method for rapid, continual sample introduction to the MS is to use pressure or vacuum driven flow to continually infuse samples onto the mass spectrometer. Pressure driven flow requires the pressurization of a sealed container holding the sample, which could prove challenging to implement safely. However, by pressurizing the headspace above a well plate and running sampling capillary from this pressurized well plate the MS source, continual flow of droplets to the MS could perhaps be achieved. This sort of setup might be possible using existing technologies. For instance, by setting up an XYZ droplet sampling system in a glove box at elevated pressures and running the sampling line through a side of the glove box, continuous positive pressure could be applied to drive flow from a well plate of samples to analytical assessment.

Perhaps more simply, we can take advantage of the reduced pressures at the mass spectrometer source to draw in samples in for analysis. This has recently been demonstrated in preliminary work with an Agilent 6120B mass spectrometer using a CE-MS source connected to 60 cm of 360/150 teflon capillary. Using the same plug flow mass spectrometry conditions described in 3.2.3, plugs of enzyme reaction were infused into the mass spectrometer for analysis using nothing more than the vacuum draw of the source region itself. 32 samples were infused into the mass spectrometer in 5 min by manually dipping the trailing end of the Teflon capillary connected to this CE-MS source into wells on a wellplate. By alternating between wells of oil and wells of sample, we were able to avoid sample to sample contact in the capillary, clearing out each subsequent sample with a plug of oil (Figure 4-13). This technique would be simple to implement for rapid assessment of nanoliter scale plugs from a multiwell plate, and could easily prove to be the most straightforward method for sample delivery for high throughput mass spectrometry.

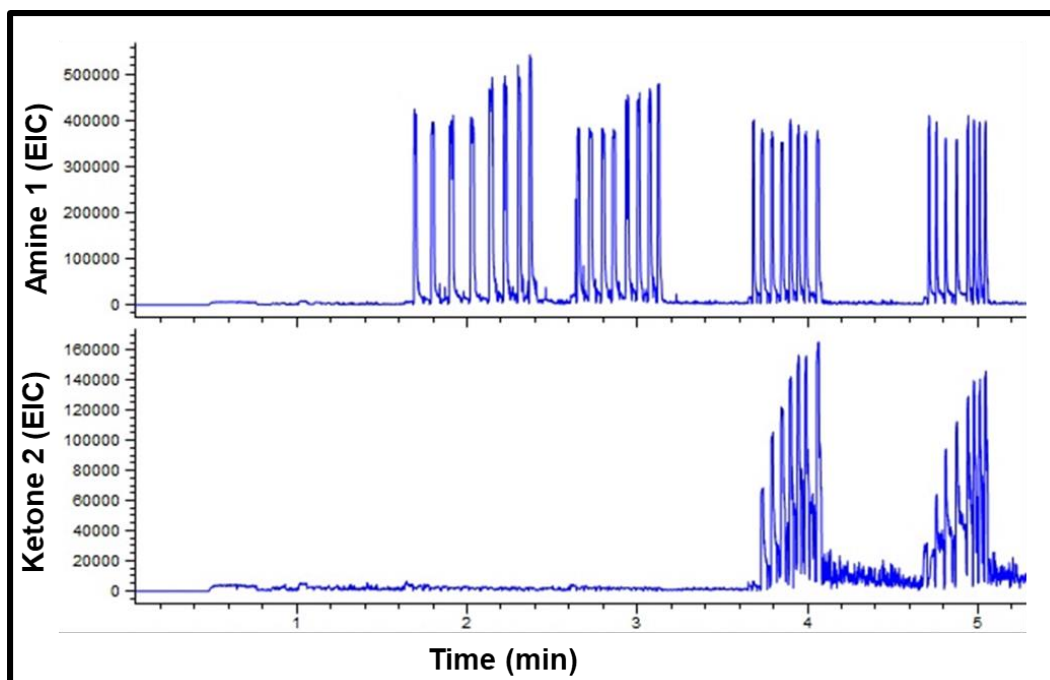


Figure 4-13: Vacuum driven infusion of samples onto a mass spectrometer

Facile sampling to a mass spectrometer is achieved on nanoliter scale plugs of ivTT reaction mix using only the negative pressure in the source of an Agilent 6120B to drive flow. Here, one enzyme variant is assessed in each cluster of samples (4 reactions, assessed at 1 h time points for 8 h) for the transamination of amine. The first two samples show no activity for the transamination, but the second two show rising signal from the ketone product of the reaction.

4.5 Conclusions

High throughput screening drives discovery in the field of chemistry. Droplet microfluidics offers a unique opportunity for laboratory miniaturization and automation in ultra-rapid screening workflows, but has largely failed to deliver on its promise over the last decade, largely due to limitations of analytical assessment in droplet systems. With the expansion of microfluidic analysis into label free detection via mass spectrometry, droplet microfluidics is no longer confined to detection on fluorogenic assays, and is now capable of application across a broad range of systems. We've demonstrated one such application – mass spectrometry based sorting of microfluidic droplets for directed evolution – and we anticipate continued development in the years to come as the versatility of sub microliter analysis via droplet mass spectrometry continues to mature.

References

- (1) Mayr, L.; Bojanic, D. *Current Opinion in Pharmacology* **2009**, *9*, 580-588.
- (2) Takatsy, G. *Acta Microbiol Acad Sci Hung* **1955**, *3*, 191-202.
- (3) In *Pharmaceutical Sciences Encyclopedia*, pp 1-44.
- (4) k.k., H. P. In *Basics and Applications*, World Technical Writing, I., Ed.; Hamamatsu Photonics, 2007, p 323.
- (5) Leary, J. F. *Cytometry Part A* **2005**, *67A*, 76-85.
- (6) *Drug Discovery Today* **1999**, *4*, 401 - 410.
- (7) SWEET, R. *Review of Scientific Instruments* **1965**, *36*, 131-&.
- (8) BONNER, W.; SWEET, R.; HULETT, H.; HERZENBERG, L. *Review of Scientific Instruments* **1972**, *43*, 404-+.
- (9) Aris, R. a. T. G. I. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* **1956**, *235*, 67-77.
- (10) Taylor, G. I. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* **1953**, *219*, 186-203.
- (11) Chiu, D.; Lorenz, R.; Jeffries, G. *Analytical Chemistry* **2009**, *81*, 5111-5118.
- (12) Baroud, C.; Gallaire, F.; Dangla, R. *Lab on a Chip* **2010**, *10*, 2032-2045.
- (13) Thorsen, T.; Roberts, R.; Arnold, F.; Quake, S. *Physical Review Letters* **2001**, *86*, 4163-4166.
- (14) Shang, L.; Cheng, Y.; Zhao, Y. *Chemical Reviews* **2017**, *117*, 7964-8040.
- (15) Garstecki, P.; Gitlin, I.; DiLuzio, W.; Whitesides, G. M.; Kumacheva, E.; Stone, H. A. *Applied Physics Letters* **2004**, *85*, 2649-2651.
- (16) Gruner, P.; Riechers, B.; Orellana, L.; Brosseau, Q.; Maes, F.; Beneyton, T.; Pekin, D.; Baret, J. *Current Opinion in Colloid & Interface Science* **2015**, *20*, 183-191.
- (17) Baret, J. *Lab on a Chip* **2012**, *12*, 422-433.
- (18) Xia, Y.; Whitesides, G. M. *Angewandte Chemie International Edition* **1998**, *37*, 550-575.
- (19) O'Donovan, B.; Tran, T.; Sciambi, A.; Abate, A. *Jove-Journal of Visualized Experiments* **2014**.
- (20) Lee, M.; Collins, J.; Aubrecht, D.; Sperling, R.; Solomon, L.; Ha, J.; Yi, G.; Weitz, D.; Manoharan, V. *Lab on a Chip* **2014**, *14*, 509-513.
- (21) Link, D.; Anna, S.; Weitz, D.; Stone, H. *Physical Review Letters* **2004**, *92*.
- (22) Nie, J.; Kennedy, R. *Analytical Chemistry* **2010**, *82*, 7852-7856.
- (23) de Ruiter, R.; Pit, A.; de Oliveira, V.; Duits, M.; van den Ende, D.; Mugele, F. *Lab on a Chip* **2014**, *14*, 883-891.
- (24) Jung, J.; Destgeer, G.; Ha, B.; Park, J.; Sung, H. *Lab on a Chip* **2016**, *16*, 3235-3243.
- (25) Schmid, L.; Weitz, D. A.; Franke, T. *Lab on a Chip* **2014**, *14*, 3710-3718.
- (26) Doonan, S.; Lin, M.; Bailey, R. *Lab on a Chip* **2019**, *19*, 1589-1598.
- (27) Xu, C.; Xie, T. *Industrial & Engineering Chemistry Research* **2017**, *56*, 7593-7622.
- (28) Lee, H.; Xu, L.; Oh, K. *Biomicrofluidics* **2014**, *8*.

- (29) Baret, J.; Miller, O.; Taly, V.; Ryckelynck, M.; El-Harrak, A.; Frenz, L.; Rick, C.; Samuels, M.; Hutchison, J.; Agresti, J.; Link, D.; Weitz, D.; Griffiths, A. *Lab on a Chip* **2009**, *9*, 1850-1858.
- (30) Ahn, K.; Kerbage, C.; Hunt, T.; Westervelt, R.; Link, D.; Weitz, D. *Applied Physics Letters* **2006**, *88*.
- (31) Sciambi, A.; Abate, A. *Lab on a Chip* **2015**, *15*, 47-51.
- (32) Gielen, F.; Hours, R.; Emond, S.; Fischlechner, M.; Schell, U.; Hollfelder, F. *Proceedings of the National Academy of Sciences of the United States of America* **2016**, *113*, E7383-E7389.
- (33) Hasan, S.; Geissler, D.; Wink, K.; Hagen, A.; Heiland, J.; Belder, D. *Lab on a Chip* **2019**, *19*, 403-409.
- (34) Caen, O.; Schutz, S.; Jammalamadaka, M.; Vrignon, J.; Nizard, P.; Schneider, T.; Baret, J.; Taly, V. *Microsystems & Nanoengineering* **2018**, *4*.
- (35) Qiao, Y.; Zha, X.; Zhu, J.; Tu, R.; Dong, L.; Wang, L.; Dong, Z.; Wan, Q.; Du, W. *Lab on a Chip* **2018**, *18*, 190-196.
- (36) Obexer, R.; Pott, M.; Zeymer, C.; Griffiths, A.; Hilvert, D. *Protein Engineering Design & Selection* **2017**, *30*, 531-531.
- (37) Obexer, R.; Pott, M.; Zeymer, C.; Griffiths, A.; Hilvert, D. *Protein Engineering Design & Selection* **2016**, *29*, 355-365.
- (38) Beneyton, T.; Coldren, F.; Baret, J.; Griffiths, A.; Taly, V. *Analyst* **2014**, *139*, 3314-3323.
- (39) Eastburn, D.; Sciambi, A.; Abate, A. *Nucleic Acids Research* **2014**, *42*.
- (40) Fallah-Araghi, A.; Baret, J.; Ryckelynck, M.; Griffiths, A. *Lab on a Chip* **2012**, *12*, 882-891.
- (41) Cole, R.; Tang, S.; Siltanen, C.; Shahi, P.; Zhang, J.; Poust, S.; Gartner, Z.; Abate, A. *Proceedings of the National Academy of Sciences of the United States of America* **2017**, *114*, 8728-8733.
- (42) Lerner, E.; Cordes, T.; Ingargiola, A.; Alhadid, Y.; Chung, S.; Michalet, X.; Weiss, S. *Science* **2018**, *359*, 288+.
- (43) Zi-Yi, Y.; He, J.-H.; Lu, A.-P.; Hou, T.-J.; Cao, D.-S. *Drug Discovery Today* **2020**, *25*, 657 - 667.
- (44) Welch, C.; Gong, X.; Schafer, W.; Pratt, E.; Brkovic, T.; Pirzada, Z.; Cuff, J.; Kosjek, B. *Tetrahedron-Asymmetry* **2010**, *21*, 1674-1681.
- (45) Haag, A. M. In *Modern Proteomics – Sample Preparation, Analysis and Practical Applications, 2016*; Springer, Cham, 2016.
- (46) www.agilent.com, 2020.
- (47) Sinclair, I.; Stearns, R.; Pringle, S.; Wingfield, J.; Datwani, S.; Hall, E.; Ghislain, L.; Majlof, L.; Bachman, M. *Jala* **2016**, *21*, 19-26.
- (48) Cody, R.; Laramée, J.; Durst, H. *Analytical Chemistry* **2005**, *77*, 2297-2302.
- (49) Zhu, Y.; Fang, Q. *Analytica Chimica Acta* **2013**, *787*, 24-35.
- (50) Kalantarifard, A.; Saateh, A.; Elbuken, C. *Chemosensors* **2018**, *6*, 23.
- (51) Guetschow, E. D.; University of Michigan. Library. Deep Blue., 2016.
- (52) Wang, X.; Ren, L.; Su, Y.; Ji, Y.; Liu, Y.; Li, C.; Li, X.; Zhang, Y.; Wang, W.; Hu, Q.; Han, D.; Xu, J.; Ma, B. *Analytical Chemistry* **2017**, *89*, 12569-12577.
- (53) Jahn, I.; Zukovskaja, O.; Zheng, X.; Weber, K.; Bocklitz, T.; Cialla-May, D.; Popp, J. *Analyst* **2017**, *142*, 1022-1047.
- (54) Syme, C.; Martino, C.; Yusvana, R.; Sirimuthu, N.; Cooper, J. *Analytical Chemistry* **2012**, *84*, 1491-1495.

- (55) Wang, X.; Xin, Y.; Ren, L.; Sun, Z.; Zhu, P.; Ji, Y.; Li, C.; Xu, J.; Ma, B. *Science Advances* **2020**, *6*, eabb3521.
- (56) Maceiczky, R.; Hess, D.; Chiu, F.; Stavrakis, S.; deMello, A. *Lab on a Chip* **2017**, *17*, 3654-3663.
- (57) *Analytica Chimica Acta* **2013**, *796*, 68 - 74.
- (58) Sun, S.; Kennedy, R. *Analytical Chemistry* **2014**, *86*, 9309-9314.
- (59) Sun, S.; Slaney, T.; Kennedy, R. *Analytical Chemistry* **2012**, *84*, 5794-5800.
- (60) Smith, C.; Li, X.; Mize, T.; Sharpe, T.; Graziani, E.; Abell, C.; Huck, W. *Analytical Chemistry* **2013**, *85*, 3812-3816.
- (61) Steyer, D. J.; Kennedy, R. T. *Analytical Chemistry* **2019**, *91*, 6645-6651.
- (62) Kirk, A. T.; Bohnhorst, A.; Raddatz, C.-R.; Allers, M.; Zimmermann, S. *Analytical and Bioanalytical Chemistry* **2019**, *411*, 6229-6246.
- (63) Kempa, E.; Smith, C.; Li, K.; Bellina, B.; Richardson, K.; Pringle, S.; Barran, P. **2020**.
- (64) Galanie, S.; Entwistle, D.; Lalonde, J. *Natural Product Reports* **2020**, *37*, 1122-1143.
- (65) *Bioorganic & Medicinal Chemistry* **2018**, *26*, 1241 - 1251.
- (66) Savile, C.; Janey, J.; Mundorff, E.; Moore, J.; Tam, S.; Jarvis, W.; Colbeck, J.; Krebber, A.; Fleitz, F.; Brands, J.; Devine, P.; Huisman, G.; Hughes, G. *Science* **2010**, *329*, 305-309.
- (67) Sheldon, R. A.; Woodley, J. M. *Chemical Reviews* **2018**, *118*, 801-838.
- (68) Arnold, F. H. *Angewandte Chemie International Edition* **2019**, *58*, 14420-14426.
- (69) Huffman, M. A.; Fryszkowska, A.; Alvizo, O.; Borra-Garske, M.; Campos, K. R.; Canada, K. A.; Devine, P. N.; Duan, D.; Forstater, J. H.; Grosser, S. T.; Halsey, H. M.; Hughes, G. J.; Jo, J.; Joyce, L. A.; Kolev, J. N.; Liang, J.; Maloney, K. M.; Mann, B. F.; Marshall, N. M.; McLaughlin, M., et al. *Science* **2019**, *366*, 1255-1259.
- (70) Adams, J. P.; Brown, M. J. B.; Diaz-Rodriguez, A.; Lloyd, R. C.; Roiban, G.-D. *Advanced Synthesis & Catalysis* **2019**, *361*, 2421-2432.
- (71) Markel, U.; Essani, K. D.; Besirlioglu, V.; Schiffels, J.; Streit, W. R.; Schwaneberg, U. *Chemical Society Reviews* **2020**, *49*, 233-262.
- (72) Ma, F.; Chung, M.; Yao, Y.; Nidetz, R.; Lee, L.; Liu, A.; Feng, Y.; Kurabayashi, K.; Yang, G. *Nature Communications* **2018**, *9*.
- (73) Sjostrom, S.; Bai, Y.; Huang, M.; Liu, Z.; Nielsen, J.; Joensson, H.; Svahn, H. *Lab on a Chip* **2014**, *14*, 806-813.
- (74) Beneyton, T.; Wijaya, I.; Postros, P.; Najah, M.; Leblond, P.; Couvent, A.; Mayot, E.; Griffiths, A.; Drevelle, A. *Scientific Reports* **2016**, *6*.
- (75) Zhu, X.; Shi, X.; Wang, S.; Chu, J.; Zhu, W.; Ye, B.; Zuo, P.; Wang, Y. *Rsc Advances* **2019**, *9*, 4507-4513.
- (76) Colin, P.; Kintses, B.; Gielen, F.; Miton, C.; Fischer, G.; Mohamed, M.; Hyvonen, M.; Morgavi, D.; Janssen, D.; Hollfelder, F. *Nature Communications* **2015**, *6*.
- (77) Diefenbach, X.; Farasat, I.; Guetschow, E.; Welch, C.; Kennedy, R.; Sun, S.; Moore, J. *Acs Omega* **2018**, *3*, 1498-1508.
- (78) Mazutis, L.; Baret, J.; Treacy, P.; Skhiri, Y.; Araghi, A.; Ryckelynck, M.; Taly, V.; Griffiths, A. *Lab on a Chip* **2009**, *9*, 2902-2908.
- (79) Mazutis, L.; Araghi, A. F.; Miller, O. J.; Baret, J.-C.; Frenz, L.; Janoshazi, A.; Taly, V.; Miller, B. J.; Hutchison, J. B.; Link, D.; Griffiths, A. D.; Ryckelynck, M. **2009**.
- (80) Holland-Moritz, D. A.; Wismer, M. K.; Mann, B. F.; Farasat, I.; Devine, P.; Guetschow, E. D.; Mangion, I.; Welch, C. J.; Moore, J. C.; Sun, S.; Kennedy, R. T. *Angewandte Chemie International Edition* **2020**, *59*, 4470-4477.

- (81) Vogelstein, B.; Kinzler, K. W. *Proceedings of the National Academy of Sciences* **1999**, *96*, 9236.
- (82) Agresti, J.; Antipov, E.; Abate, A.; Ahn, K.; Rowat, A.; Baret, J.; Marquez, M.; Klibanov, A.; Griffiths, A.; Weitz, D. *Proceedings of the National Academy of Sciences of the United States of America* **2010**, *107*, 6550-6550.
- (83) Klein, A.; Mazutis, L.; Akartuna, I.; Tallapragada, N.; Veres, A.; Li, V.; Peshkin, L.; Weitz, D.; Kirschner, M. *Cell* **2015**, *161*, 1187-1201.
- (84) Macosko, E.; Basu, A.; Satija, R.; Nemesh, J.; Shekhar, K.; Goldman, M.; Tirosh, I.; Bialas, A.; Kamitaki, N.; Martersteck, E.; Trombetta, J.; Weitz, D.; Sanes, J.; Shalek, A.; Regev, A.; McCarroll, S. *Cell* **2015**, *161*, 1202-1214.
- (85) Cochrane, W.; Malone, M.; Dane, V.; Cavett, V.; Satz, A.; Paegel, B. *Acs Combinatorial Science* **2019**, *21*, 425-435.
- (86) MacConnell, A.; Price, A.; Paegel, B. *Acs Combinatorial Science* **2017**, *19*, 181-192.
- (87) Tao, Y.; Rotem, A.; Zhang, H.; Chang, C.; Basu, A.; Kolawole, A.; Koehler, S.; Ren, Y.; Lin, J.; Pipas, J.; Feldman, A.; Wobus, C.; Weitz, D. *Lab on a Chip* **2015**, *15*, 3934-3940.
- (88) DeKosky, B.; Kojima, T.; Rodin, A.; Charab, W.; Ippolito, G.; Ellington, A.; Georgiou, G. *Nature Medicine* **2015**, *21*, 86-91.
- (89) Najah, M.; Calbrix, R.; Mahendra-Wijaya, I.; Beneyton, T.; Griffiths, A.; Drevelle, A. *Chemistry & Biology* **2014**, *21*, 1722-1732.
- (90) Tran, T.; Lan, F.; Thompson, C.; Abate, A. *Journal of Physics D-Applied Physics* **2013**, *46*.
- (91) Kintsjes, B.; Hein, C.; Mohamed, M.; Fischlechner, M.; Courtois, F.; Leine, C.; Hollfelder, F. *Chemistry & Biology* **2012**, *19*, 1001-1009.
- (92) Farinas, E.; Bulter, T.; Arnold, F. *Current Opinion in Biotechnology* **2001**, *12*, 545-551.
- (93) Debon, A.; Pott, M.; Obexer, R.; Green, A.; Friedrich, L.; Griffiths, A.; Hilvert, D. *Nature Catalysis* **2019**, *2*, 740-747.
- (94) Chen, Y.; Gani, A.; Tang, S. *Lab on a Chip* **2012**, *12*, 5093-5103.
- (95) Courtois, F.; Olguin, L.; Whyte, G.; Theberge, A.; Huck, W.; Hollfelder, F.; Abell, C. *Analytical Chemistry* **2009**, *81*, 3008-3016.
- (96) Skhiri, Y.; Gruner, P.; Semin, B.; Brosseau, Q.; Pekin, D.; Mazutis, L.; Goust, V.; Kleinschmidt, F.; El Harrak, A.; Hutchison, J.; Mayot, E.; Bartolo, J.; Griffiths, A.; Taly, V.; Baret, J. *Soft Matter* **2012**, *8*, 10618-10627.
- (97) Girault, M.; Kim, H.; Arakawa, H.; Matsuura, K.; Odaka, M.; Hattori, A.; Terazono, H.; Yasuda, K. *Scientific Reports* **2017**, *7*.
- (98) Zang, E.; Brandes, S.; Tovar, M.; Martin, K.; Mech, F.; Horbert, P.; Henkel, T.; Figge, M.; Roth, M. *Lab on a Chip* **2013**, *13*, 3707-3713.
- (99) Wink, K.; Mahler, L.; Beulig, J.; Piendl, S.; Roth, M.; Belder, D. *Analytical and Bioanalytical Chemistry* **2018**, *410*, 7679-7687.
- (100) Kuster, S.; Fagerer, S.; Verboket, P.; Eyer, K.; Jefimovs, K.; Zenobi, R.; Dittrich, P. *Analytical Chemistry* **2013**, *85*, 1285-1289.
- (101) Kuster, S.; Pabst, M.; Jefimovs, K.; Zenobi, R.; Dittrich, P. *Analytical Chemistry* **2014**, *86*, 4848-4855.
- (102) Hatakeyama, T.; Chen, D.; Ismagilov, R. *Journal of the American Chemical Society* **2006**, *128*, 2518-2519.
- (103) Lee, H.; Crane, M.; Zhang, Y.; Lu, H. *Integrative Biology* **2013**, *5*, 372-380.
- (104) Becker, H.; Gartner, C. *Analytical and Bioanalytical Chemistry* **2008**, *390*, 89-111.

- (105) Pei, J.; Li, Q.; Lee, M.; Valaskovic, G.; Kennedy, R. *Analytical Chemistry* **2009**, *81*, 6558-6561.
- (106) Iwasaki, A.; Yamada, Y.; Ikenaka, Y.; Hasegawa, J. *Biotechnology Letters* **2003**, *25*, 1843-1846.
- (107) Sciambi, A.; Abate, A. *Lab on a Chip* **2014**, *14*, 2605-2609.
- (108) Pan, M.; Rosenfeld, L.; Kim, M.; Xu, M.; Lin, E.; Derda, R.; Tang, S. *Acs Applied Materials & Interfaces* **2014**, *6*, 21446-21453.
- (109) Pan, M.; Lyu, F.; Tang, S. *Analytical Chemistry* **2015**, *87*, 7938-7943.
- (110) Etienne, G.; Vian, A.; Biocanin, M.; Deplancke, B.; Amstad, E. *Lab on a Chip* **2018**, *18*, 3903-3912.
- (111) Ahn, K.; Agresti, J.; Chong, H.; Marquez, M.; Weitz, D. *Applied Physics Letters* **2006**, *88*.
- (112) Sheldon, R. A.; Brady, D.; Bode, M. L. *Chemical Science* **2020**, *11*, 2587-2605.
- (113) Packer, M.; Liu, D. *Nature Reviews Genetics* **2015**, *16*, 379-394.
- (114) Mazutis, L.; Gilbert, J.; Ung, W.; Weitz, D.; Griffiths, A.; Heyman, J. *Nature Protocols* **2013**, *8*, 870-891.
- (115) Bai, Y.; Weibull, E.; Joensson, H.; Andersson-Svahn, H. *Sensors and Actuators B-Chemical* **2014**, *194*, 249-254.
- (116) Sinclair, I.; Bachman, M.; Addison, D.; Rohman, M.; Murray, D. C.; Davies, G.; Mouchet, E.; Tonge, M. E.; Stearns, R. G.; Ghislain, L.; Datwani, S. S.; Majlof, L.; Hall, E.; Jones, G. R.; Hoyes, E.; Olechno, J.; Ellson, R. N.; Barran, P. E.; Pringle, S. D.; Morris, M. R., et al. *Analytical Chemistry* **2019**, *91*, 3790-3794.
- (117) Stewart, J. D. *Current Opinion in Chemical Biology* **2001**, *5*, 120-129.
- (118) Shin, J. S.; Kim, B. G. *Biotechnol Bioeng* **1999**, *65*, 206-211.
- (119) Gibson, D. G.; Young, L.; Chuang, R.-Y.; Venter, J. C.; Hutchison, C. A.; Smith, H. O. *Nature Methods* **2009**, *6*, 343-345.
- (120) Mahler, L.; Tovar, M.; Weber, T.; Brandes, S.; Rudolph, M.; Ehgartner, J.; Mayr, T.; Figge, M.; Roth, M.; Zang, E. *Rsc Advances* **2015**, *5*, 101871-101878.
- (121) Baret, J.; Beck, Y.; Billas-Massobrio, I.; Moras, D.; Griffiths, A. *Chemistry & Biology* **2010**, *17*, 528-536.
- (122) Collins, D. J.; Neild, A.; deMello, A.; Liu, A.-Q.; Ai, Y. *Lab on a Chip* **2015**, *15*, 3439-3459.
- (123) Karsai, A.; Müller, S.; Platz, S.; Hauser, M.-T. *BioTechniques* **2002**, *32*, 790-796.
- (124) Zhang, J. H.; Chung, T. D.; Oldenburg, K. R. *J Biomol Screen* **1999**, *4*, 67-73.
- (125) Zipper, H.; Brunner, H.; Bernhagen, J.; Vitzthum, F. *Nucleic Acids Res* **2004**, *32*, e103.
- (126) Roach, L. S.; Song, H.; Ismagilov, R. F. *Analytical Chemistry* **2005**, *77*, 785-796.
- (127) Janiesch, J.-W.; Weiss, M.; Kannenberg, G.; Hannabuss, J.; Surrey, T.; Platzman, I.; Spatz, J. P. *Analytical Chemistry* **2015**, *87*, 2063-2067.
- (128) Sandoz, P.; Chung, A.; Weaver, W.; Di Carlo, D. *Langmuir* **2014**, *30*, 6637-6643.
- (129) Najah, M.; Mayot, E.; Mahendra-Wijaya, I.; Griffiths, A.; Ladame, S.; Drevelle, A. *Analytical Chemistry* **2013**, *85*, 9807-9814.
- (130) Chowdhury, M. S.; Zheng, W.; Kumari, S.; Heyman, J.; Zhang, X.; Dey, P.; Weitz, D. A.; Haag, R. *Nat Commun* **2019**, *10*, 4546.
- (131) Abate, A.; Hung, T.; Mary, P.; Agresti, J.; Weitz, D. *Proceedings of the National Academy of Sciences of the United States of America* **2010**, *107*, 19163-19166.
- (132) Doonan, S. R.; Bailey, R. C. *Analytical Chemistry* **2017**, *89*, 4091-4099.

- (133) Vanapalli, S. A.; Banpurkar, A. G.; van den Ende, D.; Duits, M. H.; Mugele, F. *Lab Chip* **2009**, *9*, 982-990.
- (134) Goto, H.; Kanai, Y.; Yotsui, A.; Shimokihara, S.; Shitara, S.; Oyobiki, R.; Fujiwara, K.; Watanabe, T.; Einaga, Y.; Matsumoto, Y.; Miki, N.; Doi, N. *Lab on a Chip* **2020**, *20*, 852-861.
- (135) Favalli, N.; Bassi, G.; Scheuermann, J.; Neri, D. *FEBS Lett* **2018**, *592*, 2168-2180.
- (136) Dickson, P.; Kodadek, T. *Organic & Biomolecular Chemistry* **2019**, *17*, 4676-4688.
- (137) Houghten, R. A.; Pinilla, C.; Blondelle, S. E.; Appel, J. R.; Dooley, C. T.; Cuervo, J. H. *Nature* **1991**, *354*, 84-86.