NORTHWESTERN UNIVERSITY

Applications of SAMDI Mass Spectrometry: Shifting Limits in Directed Evolution and the Development of Spatiotemporal Sequencing

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Biomedical Engineering

By

Adam John Pluchinsky

EVANSTON, ILLINOIS

June 2022

Abstract

SAMDI-MS, which stands for self-assembled monolayers (SAMs) for matrix-assisted laser desorption ionization (MALDI) mass spectrometry (MS), is a powerful tool that has enabled the development of novel high-throughput screening and experimentation methods for decades. SAMDI-MS works by immobilizing analytes to functionalized SAMs prior to MS analysis and is capable of studying enzymatic and chemical reactions performed in solution or directly on the surface. In this dissertation, I adopt the SAMDI platform for the development and application of unique bioassays for use in two major fields of research, directed evolution and protein sequencing. In these fields, I use the technology to remove a most persistent bottleneck and add a new dimension of analysis, respectively. I then further build out the technique's flexibility to analyze molecules of interest.

In the first chapter, I harness SAMDI-MS's ability to rapidly screen thousands of cell-based reactions to develop a screening assay that is capable of screening reactions from libraries of enzyme variants significantly faster than state-of-the-art techniques. Suitable high-throughput and generalizable screening techniques are mandatory in directed evolution, as libraries often exceed several hundred enzyme variants; however, many directed evolution campaigns still rely on the use of low-throughput chromatography-based screening methods. Here, I present a high-throughput strategy for screening libraries of enzyme variants for improved activity. Unpurified reaction products are immobilized to a self-assembled monolayer and analyzed by mass spectrometry, allowing for direct evaluation of thousands of variants in under an hour. The method was demonstrated with libraries of randomly mutated cytochrome P411 variants to identify improved catalysts for a non-natural biochemical C–H alkylation reaction. This reaction was chosen because it is challenging to detect using traditional methods, demonstrating SAMDI-MS's

flexibility for a wide variety of reactions. The evolved catalyst may also find use in organic synthesis as its products are difficult to synthesize by chemical means otherwise. The technique may be tailored to evolve enzymatic activity for a variety of transformations where higher throughput is needed. It is with this research, in collaboration with the lab of Frances Arnold, the recipient of the 2018 Nobel Prize in chemistry, that we shift the attention of the field from this bottleneck to new challenges.

In the second chapter, I use a recently developed extension of the SAMDI technique, imaging SAMDI (iSAMDI), that enables high resolution microfluidic sequencing of surface-bound peptides that is capable of resolving amino acids of identical mass. Techniques that offer single-residue resolution of amino acids are important in for proteomic research, especially where the precise sequence of a protein is not already known in a database or where disease related mutations have altered the sequence of a protein. Here, I immobilize peptides to the floor of a microfluidic flow cell and use exopeptidases to generate peptide ladders that span the channel. I then use iSAMDI to provide a record of the peptide ladders. While the difference in mass between the ladders, read by MALDI-MS, can be used to reveal the individual amino acids in the peptide, the difference in specificity the exopeptidases exhibit for each amino acid can also be used to distinguish between residues by a simple visual analysis of images of their kinetic signatures provided by iSAMDI. This extension allows for the resolution of isobaric amino acids—information that could not be obtained by MALDI-MS alone and generally difficult to obtain using modern sequencing techniques. The iSAMDI assay is shown with a variety of peptides and exopeptidases.

Finally, I describe new developments and additional findings in immobilization techniques that may find use in future SAMDI-MS applications. I describe methods for capturing analytes to

the surfaces via the azide-alkyne “click” cycloaddition using both copper and copper-free systems with multiple strategies. Having a large toolbox of capture chemistries is important for selecting the best system to analyze samples.

Acknowledgments

The level of success of my Ph.D. would not have been possible without my advisor, Professor Milan Mrksich. Thank you. Your quest to make discoveries that drive science is the primary reason I chose to conduct my thesis work at Northwestern. With your support, I had the freedom to wrestle with both interesting and consequential scientific challenges and, as a result, provide herein new innovation to the scientific community. Because of your guidance and mentorship, I can now intuit solutions to problems in ways that I didn't and couldn't before. Your outlooks have contributed significantly to my matriculation as a scientist, developed my professionalism and allowed me to understand the importance of where I am, what I'm doing, the people I meet and where I decide to go.

Thank you to my committee, Dr. Evan Scott, Dr. Michael Jewett, and Dr. Neha Kamat for your academic support and pleasant conversation. I also recognize several of the many other professors that have inspired me during my time at Northwestern, including Dr. Eric Perreault, Dr. Matthew Tresch, Dr. Neil Kelleher, and Dr. Patrick Kiser. Thanks to Saman Shafaie for helping my research to run smoothly in IMSERC and to the National Science Foundation for providing funding for my research.

Within the scientific community at Northwestern, I am thankful to have met so many individuals that helped me to push my thesis work forward and heighten its impact. In my immediate research group, I thank Dr. Justin Modica for his wealth of knowledge and rigorous discourse, Dr. Kevin Metcalf for his mentorship and teachings in lab and in life, Dr. Patrick O'Kane for getting me up-to-speed with our lab's technology, Dr. Alexei Ten for helping me to innovate solutions to problems, Dr. Jennifer Grant for her methodological brilliance and for serving as a graduate student role model, Juliet Roll for being a wonderful lab partner, Dr. Jose-

Marc Techner for his support in the lab and friendship, Dr. Blaise Kimmel for his resourcefulness, my office desk neighbor Dr. Elamar Hakim Moully for helping me grow, and to Yael Mayer, Brianna Bullock, Deborah Cundiff, Nakendra Jackson and Christina Doelling for making me feel welcome in the department from day one, for keeping the lab and department up and running and for many lovely conversations.

I notably recognize and thank Dr. Jasmine Hershewe for sitting next to me in the front row of my very first class, for all of the support you provided in my research and for always being someone I could build a meaningful and long-lasting friendship with.

A large portion of my research was as a result of a collaboration external to the University, at The California Institute of Technology. I'd like to thank Dr. Frances Arnold and her group members, namely, Daniel Wackelin, Dr. Xiongyi Huang, Dr. Ruijie Chang and Dr. Sabine Brinkmann-Chen for believing in our vision, working with us to develop an impactful demonstration of technology and teaching me the value of working with others with different expertise.

Northwestern University is an excellent place to complete a Ph.D. outside of the lab, too. I acknowledge that my research significantly benefited from an institution with one of the strongest and most endowed research programs in the world; however, my work also found balance with the greater campus community. Here, I'd like to thank the Center for Leadership and Dr. Adam Goodman for supplementing my growth with interdisciplinary training that enhanced my ability to work with others by developing my understanding of my own leadership. I also thank the board members of two graduate student associations I served on, QPGSA and ADCA, for believing in my ability to bring my skills to help transform the organizations. Moreover, thank you to Vimal Bellamkonda for your friendship, and, together with Kirthana Sandepudi and Rishika Dugyala, for

being the best dance team captains I could ask for to keep me physically active in graduate school, and, together with Kripa Guha, Areesha Majeed, Swetha Marisetty, Manasa Pagadala, Braylan Saunders-Effort, Aditi Rathore, my dance partner Deepika Sriram and joint choreographer Tejas Sekhar, for making me feel welcomed into the Desi community forever. I also thank Elysse Longiotti who continues to go above and beyond her role as a career advisor to me and thank all of the professors in the Kellogg School of Management for letting me audit their courses.

Being in Chicago, I also made long lasting friendships and relationships outside of the University that further rounded out my career as a graduate student with additional friendship and love. Here, I thank Neil Ramos, Ryan Thomas Lay, Amy Jia, Joshua Hills, Clayton Rische, and Alexandra Kolberg for bringing joy and care to my graduate career. With them, I also want to thank my childhood and college friends, Briana Hyndman, Andrew Backhaus, Dr. Thomas Spanarkel, Arnesh Jajoo and Sydney Garay for supporting my studies and continuing to believe in me. I am also thankful for having had Michael Roach and Aaron Bernacho in my life during this time and for the memories we shared.

Lastly, I am thankful for my loving family who continue to be proud of my scholastic achievements.

Preface

With rigorous training in chemistry from Fordham University, a background heavy in engineering from Columbia University, and a desire to journey into the fields of molecular and synthetic biology, I chose to pursue my graduation education in order to engineer new and highly impactful tools as the intersection of these fields. The laboratory of Dr. Milan Mrksich, which conducts research at this interface, along with the collaborative approach to research that Northwestern University cites as gospel, provided the perfect environment to achieve this goal. As a result, I was not only able further innovate the technology in my home research facility, but I also had the opportunity to apply these innovations to solve significant challenges as a collaborator for another prominent research group at The California Institute of Technology – a group which I also showed significant interest in when applying to graduate school. I could not have achieved my current level of success without my advisor's continued enthusiasm and support for my potential, the University's spirit, and my drive to pursue things far greater than myself.

List of Abbreviations

SAM	Self-assembled monolayer
MALDI	Matrix-assisted laser desorption/ionization
TOF	Time-of-flight
MS	Mass spectrometry
MS/MS or **MS^2**	Tandem mass spectrometry
AUC	Area under the curve
SAMDI	Self assembled monolayers for MALDI
iSAMDI	Imaging SAMDI
TI-SAMDI	Traceless immobilization SAMDI
GCMS	Gas chromatography mass spectrometry
HPLC	High pressure liquid chromatography
SFC	Supercritical fluid chromatography
FACS	Fluorescence activated cell sorting
epPCR	Error-prone polymerase chain reaction
THAP	2,4,6-trihydroxyacetophenone
PEG	Polyethylene glycol
EDA	Ethyldiazoacetate
TTN	Total turnover
DBCO	Dibenzylcyclooctyne
CuAAC	Copper(I)-catalyzed azide-alkyne cycloaddition

"*Those are the discoveries that drive science.*"

Table of Contents

List of Figures

Chapter 1

SAMDI Mass Spectrometry

1.1 Design

For two decades, the Mrksich Group has harnessed the combinatorial power of self-assembled monolayers and mass spectrometry to enable the development of rapid, generalizable and creative biosensing and detection platforms that have advanced a myriad of fields. This section describes the technology, how it has been used in our group to extend the boundaries of research in several areas and sets the stage for the two fields of significance where I've applied the technology.

SAMDI-MS, which fully stands for self-assembled monolayers for matrix-assisted laser desorption/ionization (MALDI) mass spectrometry (MS), in the simplest technical description, is an analytical technique that aims to post-experimentally analyze samples that have been immobilized to an engineered surface. The self-assembled monolayers that make up these engineered surfaces offer a number of advantages that compliment MALDI instrumentation both in their design and readout.

At their foundation, the engineered surfaces used by our group make use of the semi-covalent bond formed between sulfur and gold (Figure 1.1). By coating the surface of an analysis plate with gold, a limitless number of sulfur-terminated self-assembled monolayers are able to be bound to the surface. This bond is particularly useful for several reasons. First, preparation is fairly simple and rapid in that it only requires a gold-evaporated surface to be soaked with sulfur-terminated molecules. Secondly, the strength of the gold-sulfur bond has been shown to be stable

under a wide variety of reaction conditions (including biological conditions) while being readily ablated during MALDI mass spectrometric analysis.

As we work our way up the self-assembled monolayers, our group and the work discussed herein makes use of alkane thiolate terminated monolayers that next comprise of repeating units of polyethylene glycol (PEG). The alkane groups have been shown to provide heat stability to the monolayer, where in their unterminated form have shown to be stable to upwards of 400K[1]; by terminating a proportionally larger number of these monolayers in PEG, the surface gains the ability to prevent non-specific adsorption of biological and other materials[2]. This is incredibly useful for experimentation involving whole cells and proteins, so that cellular debris and enzymes themselves are not left mixed with an analyte of interest during analysis in the mass spectrometer.

A final and key aspect of the technique is the control the user has over the monolayer's surface chemistry. By extending 10 to 20% of the SAMs beyond those of which are terminated with PEG, one can strategically functionalize the surface with chemical moieties for use in a variety of applications. In general, these chemistries can be used to selectively immobilize libraries of proteins, peptides, small molecules and other reaction constituents based on the assay being developed. Several ligation chemistries have been developed and used by our group for such purposes (Figure 1.1). These reactions, such the maleimide-thiol reaction, azide-alkyne reaction, Diels-Alder reaction, and amine-NHS-ester reaction include several click chemistries, several non-covalent reactions such as streptavidin-biotin binding, as well as others[3–6]. Additionally, the monolayer can be adjusted to contain a positively charged molecule in order to consolidate the readout to a desired adduct.

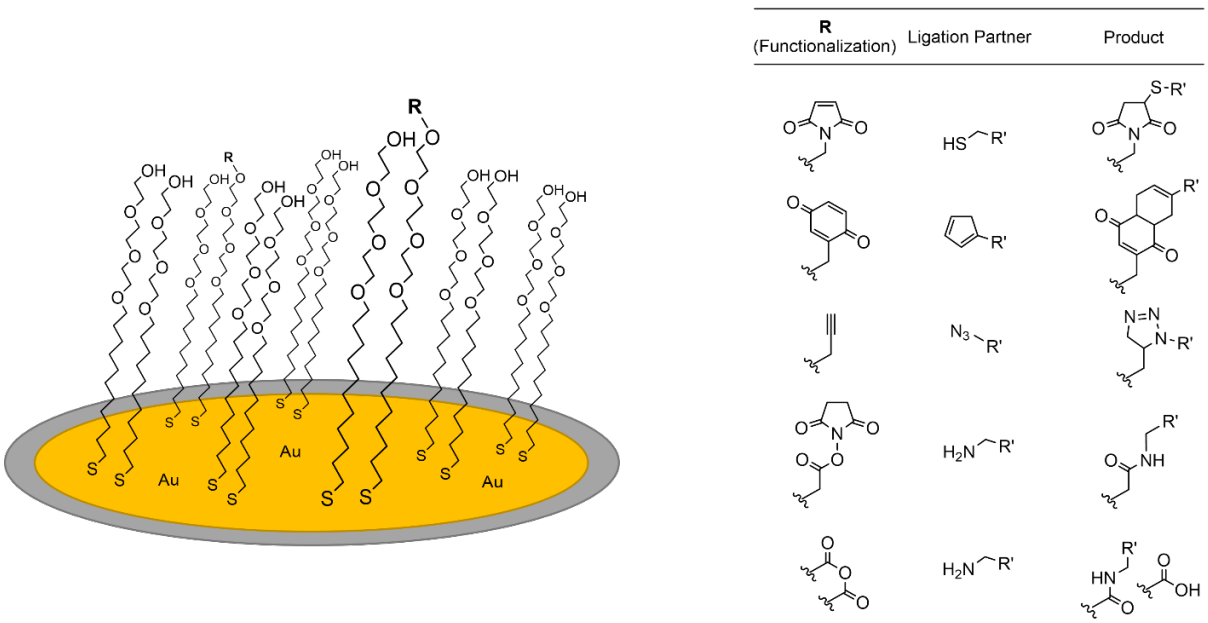


Figure 1.1. Surface Functionalization with Ligation Chemistries. Several reaction schemes are shown that may be used to functionalize the surface in order to further immobilize a substrate to the SAM. Surface functionality is typically chosen based on the chemistries the substrate presents.

The work presented in this thesis relies almost entirely upon the maleimide-thiol reaction. This chemistry is particularly useful as it does not require a catalyst, is biocompatible, highly specific and very fast. However, in some instances I make use of a diazirine-functionalized monolayer to highlight the surface's flexibility and also present a chapter with multiple strategies to expand SAMDI's toolbox with a dibenzylcyclooctyne (DBCO) terminated monolayer for copper-free click ligation to the surface.

1.2 Readout

To analyze the self-assembled monolayers, our group uses matrix-assisted laser desorption/ionization time-of-fight (MALDI-TOF) mass spectrometry (Figure 1.2). In recent

decades, MALDI-TOF has quickly grown as a popular method for the detection of small molecules, peptides, and other biomolecules as large as proteins[7–9]. To analyze samples with MALDI-TOF, a sample is mixed or coated with another compound known as matrix. The dried sample is then put into the instrument to be ablated and ionized by a laser. While not very well understood, it is thought that the matrix serves either to absorb the laser's energy and transfer it to the sample, or that the matrix removes energy from the sample. Importantly, this "soft ionization" method offered by MALDI-TOF causes little sample fragmentation, thus providing straightforward and rapid detection of molecules. The molecules then travel over a set distance, hit a detector, and are visualized based on their differences in travel time to the detector. From this information and the charge of the molecules being analyzed, the mass of molecules are derived and distributed across a spectral readout with their relative quantities represented as peaks of varying intensity. The method is not only helpful for detecting molecules, but also for identifying when a change in mass has occurred to molecules as well.

Coupled with MALDI mass spectrometry, the self-assembled monolayers produce spectra that are straightforward to interpret with generally no fragmentation of the monolayers occurring in the instrument other than breakage of the disulfide bond[10]. As with any sample, MALDI mass spectrometry is able to provide mass information on self-assembled monolayers and their tethered analyte[11]. Knowing the mass on the unreacted monolayer allows the user to subtract out the mass of the monolayer in order to understand the mass of the analyte of interest. This is a particularly rewarding feature for small molecules that would otherwise be convoluted in the lower mass region of spectral readout. Because all molecules have a mass, the technique is applicable to essentially all biochemical assays and can provide yields for reactions for as long as both the substrate and product can be observed. Even in some cases of constitutional isomers where reaction constituents

have the same mass but different structural configurations, it may be possible to distinguish between them with post-reaction processing.

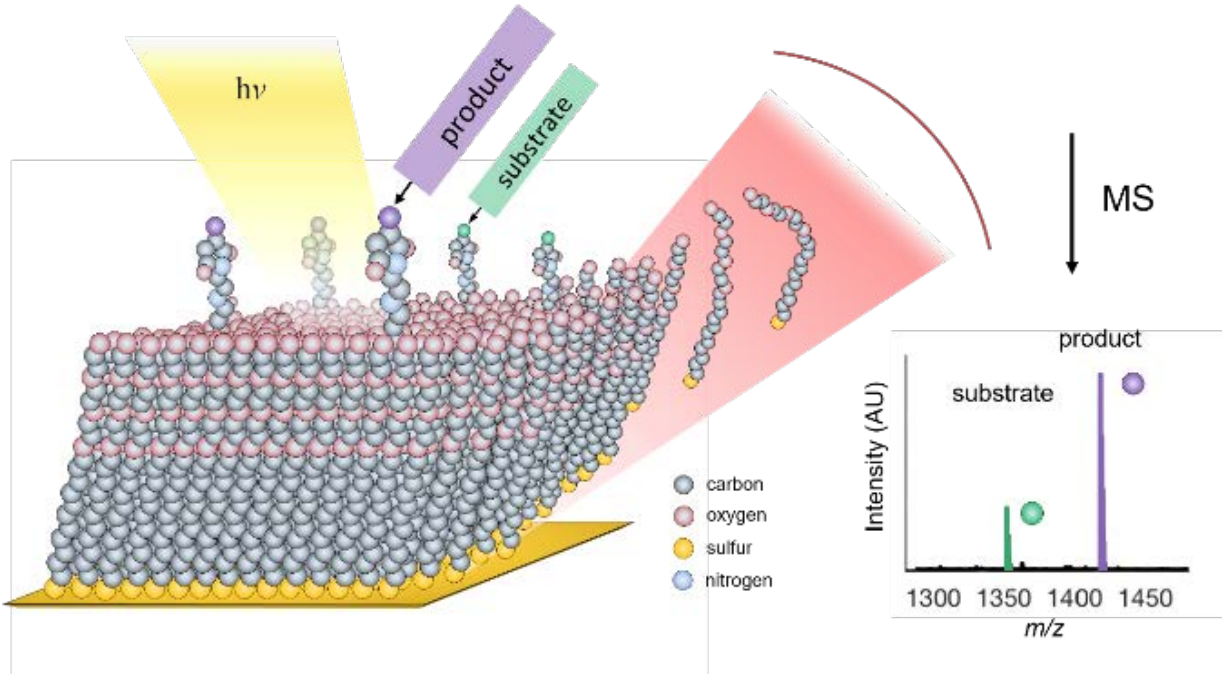


Figure 1.2. MALDI Mass Spectrometry of Self-Assembled Monolayers. Reaction products can be immobilized to self-assembled monolayers of alkane thiolates on gold. The laser source from the MALDI desorbs and ionizes the monolayers. The resulting mass spectra displays peaks that correspond to the mass of the immobilized analytes, where their relative intensities can be used to determine their quantities, and thus provide a yield for the reaction. Image adapted from Mrksich, M., Mass spectrometry of self-assembled monolayers: A new tool for molecular surface science. *ACS Nano* **2008**, 2, 7-8 (reference 10) to include all atoms of the self-assembled monolayer our group typically employs, including maleimide functionalization.

Another important feature of the SAMDI method is that it is very high-throughput compared to MALDI alone, especially in the case of complex samples that can convolute the signal. Like with MALDI, samples to be analyzed by SAMDI can be prepared in an array format,

and thus are compatible with other high throughput liquid handling technologies. Presently, our group most commonly analyzes samples in 384- and 1536- array formats, although 6144 and higher densities have been explored (Figure 1.3). The analysis plate itself only requires simple modification with gold using to create an array of spots. A surface of self-assembled monolayers is then added by simply soaking the plate in an ethanolic solution of pre-functionalized monolayers and is re-functionalized as necessary. After a sample of interest is immobilized to the monolayer, the plate is simply rinsed, and 2,4,6-trihydroxyacetophenone (THAP) matrix is applied to each spot and dried, as this matrix has shown to be well suited for detection of alkanethiolates. From there, the MALDI mass spectrometer is able to quickly acquire data for each spot, with the instrument requiring approximately 30minutes for a 384 plate, and 50 minutes for a 1536 plate[3]. Lastly, fast data processing can be performed using peak-finding software or computer programs like Profiler that can automatically detect and integrate mass spectra peaks. To demonstrate the throughput of the technique, early work in our group showed the ability of SAMDI-MS to analyze small molecule libraries containing hundreds of thousands of samples in a day[12]. Given the upper limit on throughput, the technique has found comfortable use is studies where thousands of samples require analysis. This includes recent work optimizing the traceless Petasis reaction as a result of screening more than 1800 unique reaction conditions[13], and identifying and characterizing four new glycosyltransferases as a result of screening nearly 15,000 reaction conditions[14].

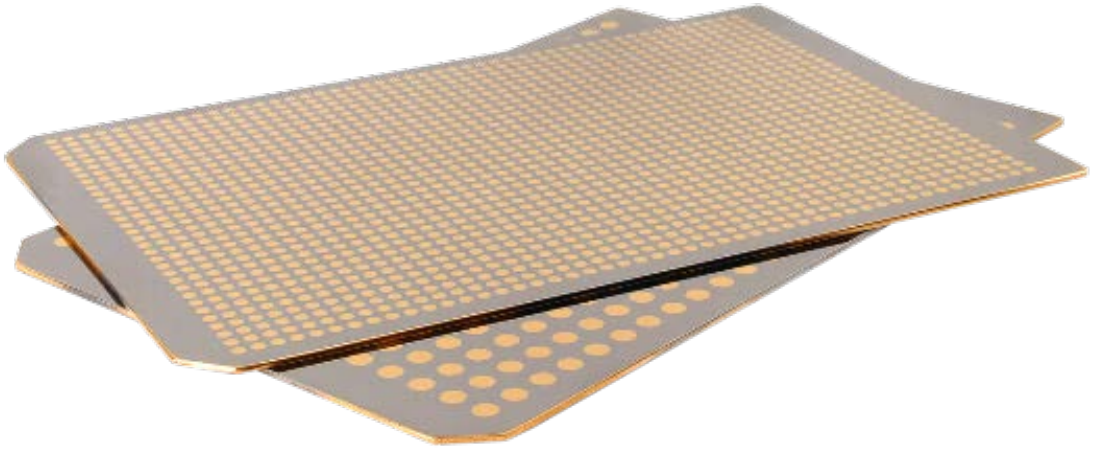


Figure 1.3. 1536 and 384 SAMDI Array Plates. Reprinted from *Methods in Enzymology*, *607*, Syzmczak, L. C.; Huang, C.-F.; Berns, E. J.; Mrksich, M., Combining SAMDI mass spectrometry and peptide arrays to profile phosphatase activities, 389-403, **2018**.

1.3 Applications

Early examples of SAMDI focused on characterizing enzymatic activity in biochemical systems where purified enzymes were allowed to act on immobilized substrates. While molecules are typically analyzed by MALDI after an in-solution reaction, one of the benefits of SAMDI is the ability analyze both in-solution reactions and those conducted directly on the surface, *in situ*. This is particularly helpful for providing a more accurate assessment of enzymatic activity when the enzyme's native substrate, such as a peptide, is naturally present at a biointerface (i.e. the tail of histone). For this reason, the platform has found extensive use in characterizing kinase, phosphatase, deaminase, glycosyl transferase, protease, hydrolase, acetyl transferase, and deacetylase activity on entire libraries of samples, without the need for a label or tag on the peptide[3,14–21]. Rather, peptides are immobilized to the surface via the Michael addition between

the maleimide present on the surface and the naturally occurring sulfur group present in the amino acid cysteine. The reactions performed *in situ*, typically recapitulating post-translational modifications (PTMs) that occur naturally in the cell, can be monitored by SAMDI. Toggling the amino acids along the peptide substrate that border the amino acid undergoing a reaction allows the user to understand entire specificity profiles of the enzyme being question. Beyond peptides, the method has also been shown to be able to characterize reactions on immobilized proteins. This strategy has also been useful for screening enzyme inhibitors for drug discovery including those for anthrax lethal factor, SIRT3, and isocitrate dehydrogenase 1[22–24].

Other applications of the technology make use of the "pull-down" format in addition to reactions that are conducted *in situ*. In this format, rather than immobilizing a species prior to any reaction, enzymatic and chemical reactions are carried out in-solution first, and then applied to the surface for detection and analysis. For example, making use of the alkyne-azide ligation strategy to immobilize analogs of tolbutamide, SAMDI has been used to characterize and measure drug metabolism kinetics of a cytochrome P450 in order to show its usefulness in screening for and identifying potential adverse drug-drug interactions[25]. This strategy has also found use in quantifying up to a dozen reaction intermediates simultaneously, and has also enabled the development of various high-throughput reaction discovery approaches[13,26,27].

Moving beyond these two formats, SAMDI has also been used to enable the development of new assays. For example, our group has grown and lysed cells directly on the surface, in order to quantify protein activity[21,28]. Building off of these studies, the technique has since been used to measure protein activity of single cells to better understand cellular heterogeneity in cancer[29] and has enabled the development of live cell analysis devices capable of monitoring the dynamics of cellular activity of cellular activity from small populations of cells in response to stimuli[30]. As a

screening tool for nanoparticle therapeutics, SAMDI rapidly measured the immune activities of spherical nucleic acids and underscored the importance of structure-activity relationships when developing new medicines[31]. Interestingly, the field of data storage has also benefited from SAMDI, with new methodology reported for storing information in molecules on the surface of an analysis plate[32]. Here, peptides of varying masses can represent information as bit, dependent on their presence/absence on the surface of self-assembled monolayers. Reactions on the surface have also been used to amplify and quantify low affinity protein-ligand interactions in high-throughput fashion[33]. From these examples, it can be argued that SAMDI has had and continues to be capable of providing substantial impact in more fields than most technologies due to its simplicity, speed, and versatility.

In this work, I bring SAMDI to two more important fields, namely directed evolution and protein sequencing, to show how the technology can enable new developments in these areas of research. In the next two chapters, 2 and 3, I introduce the field of directed evolution and show how SAMDI-MS can be used to develop catalysts both in record time and for reactions that could not be explored otherwise. In chapters 4 and 5, I introduce the field of protein sequencing and describe the development of a SAMDI-MS assay that enables high resolution microfluidic sequencing capable of resolving amino acids of identical mass. Lastly, in chapter 6, I introduce my own designs and methods for functionalizing the monolayer with strained alkynes that may find use in future studies.

Chapter 2

Directed Evolution

2.1 The Cycle of Directed Evolution: Mutagenesis and Screening

In Nature, the development of all complex catalysts and compounds has been driven by evolution. From microorganisms to fungi, plants and animals, all species adapt through the process of natural evolution when a micro-level change in heritable traits leads to a competitive advantage over generations[34]. In rapidly dividing bacteria subjected to a selection pressure, this process can happen rather quickly. In a common example of bacteria subjected to an antibiotic, a single bacterium with a mutation that confers resistance to the antibiotic will eventually become dominant across billions of cells[35]. Such mutations can occur in any number of the proteins that make up the bacterium, giving the bacterium a multitude of options for surviving using new proteins. Several examples include proteins that make up bacterial membranes mutated to selectively keep drugs from entering, displays of protein pumps that selectively remove antibiotics from the cell, catalysts that actively destroy the antibiotic, or the development altogether of new protein pathways that circumvent those the antibiotic set out to disrupt. Here, small changes in the genotype of microorganisms allow them to survive. For macroorganisms, these processes and their benefits are only seen over much greater time scales. One of most prominent examples in modern history is sickle cell anemia, where two mutations in the gene for haemoglobin, the protein which carries human blood, adversely changes the shape of the cells but, also helps to confer a level of resistance to malaria[36]. The origins of this genetic shift can be dated back to the year 5300 B.C., and today, the World Health Organization estimates that roughly 5% of the population have a blood cell mutation[37]. It is through this selection process that developed the tens of thousands of functions

we have today, but the scientific community realize that enzymes can do much more. Fortunately, Mother Nature has loaded organisms with catalysts for us to study, use and/or evolve ourselves.

As scientists, we love copying Nature. From developing reusable, high load and easy release materials that mimic the feet of geckos[38], to our attempts to recreate the brain using artificial intelligence[39], some of our best inventions are a result of biomimicry. Directed evolution is the attempt to take existing enzymes and recreate the process of evolution in the lab with the goal to develop new catalysts for industrial, research and medicinal applications by engineering enzymes with any intended function. Typically, enzymes found in nature are not in their best condition to perform off-target reactions, thus, methods must be used to express them recombinantly, increase their purity and productivity, and sometimes change their function altogether. In addition to these attributes, directed evolution efforts can be used to increase an enzyme's thermostability, substrate specificity, optimum working pH, and organic solvent resistance[40–45]. By mimicking Nature with directed evolution, scientists are able to quickly develop catalysts.

In its most general sense, the directed evolution paradigm consists of two major steps: subjecting an enzyme to iterative rounds of mutagenesis to create a library of variants, followed by a selection or screening method to extract variants that show improved function defined either by an increase in productivity, a change in phenotype and/or change in some other property. In the first step, the gene that encodes for an enzyme of interest, called a parent enzyme, is amplified and subjected to mutagenesis methodology that alters the nucleic acid sequence of the gene. The altered genes are then allowed to produce protein variants, some of which, depending on the mutation they receive, differ from the parent enzyme by one to several amino acid mutations. It is these changes

in amino acids that lead to difference in properties between a variant enzyme and its parent. In the second step, a selection or screen is applied to isolate the variants that have contain the new or changed property. In a selection, undesired variants are automatically eliminated whereas with a screen, each individual variant is evaluated for the desired property. While seemingly simple, a screening method is often responsible for sorting through anywhere from 10^3 to 10^{15} samples and distinguishing variants from one another based on the property being tested is not always straightforward.

A large driver behind the protein engineering technique is the pharmaceutical industry, where newer and more efficient catalysts can be used to either produce non-protein drugs by circumventing costly and time-consuming chemical synthesis reaction pathways in drug development, by replace other synthetic catalysts, or as use a protein therapeutics themselves[46,47]. For example, engineering of monoclonal antibodies can provide greater safety and efficacy to patients while exhibiting improved delivery and reduced immunogenicity. In 2016, the global protein engineering market size was valued at 823 million USD due to the broad interest among academic institutions and in industry to develop new catalysts to aid in drug discovery. Directed evolution, comprises roughly a third of this research (Figure 2.1) and its use is expected to surpass a market size of 1 billion in the next few years.

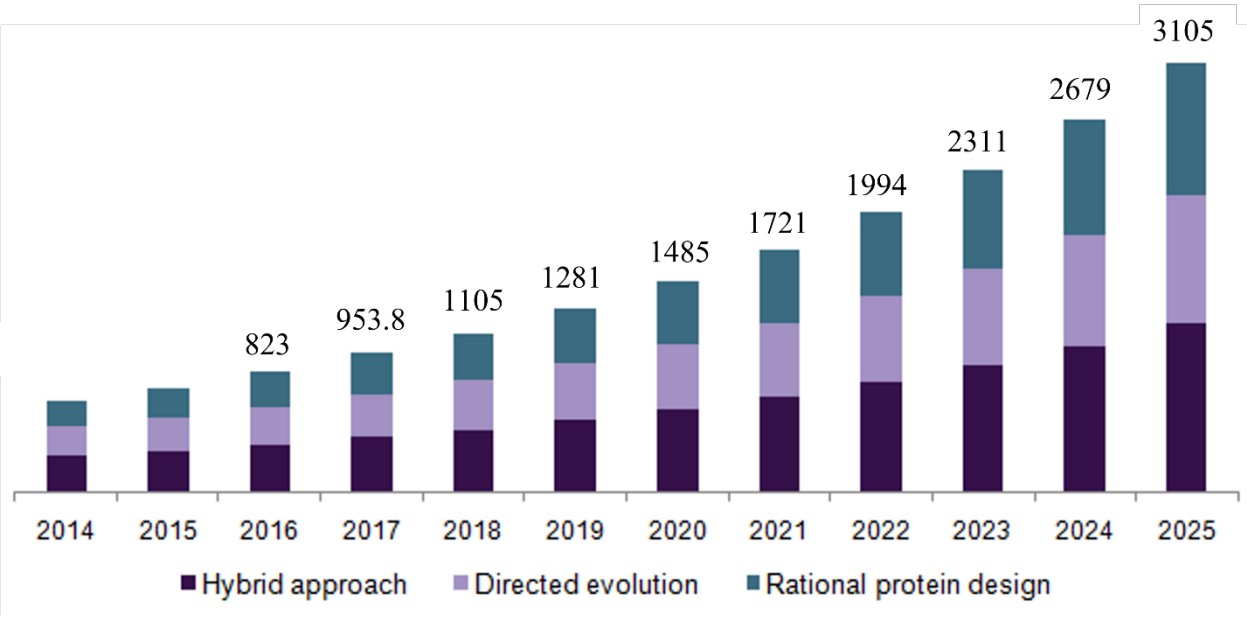


Figure 2.1. U.S. Protein Engineering Market by Technology, 2014 - 2025 (USD Million).

The global protein engineering market size was valued at 823 million in 2016. With an anticipated growth rate of 15.9% CAGR the market would exceed 3 billion USD in the next few years. Here directed evolution is represented as areas where rational protein design is not utilized or where both are used in combination. Data was obtained and forecasted from Grand View Research copyright 2017.

Today, over 90% of industrial enzymes are produced recombinantly (outside of its native host organism) for maximal purity and productivity[48]. For example, the Arnold group at Caltech has designed a new enzyme, namely the P411, from an existing P450 for in vivo cyclopropenation catalysis[49], which was later enabled to enantioselectively synthesize precursors to levomilnacipran[50] and ticagrelor[51] and pushed further to gram-scale synthesis[52]. Subsequent evolution showed the P411 could also catalyze C-H amination[53] and has recently displayed promiscuity toward highly-challenging carbene transfer reactions including C-H bond insertion

activity. This property could be extensively exploited for late-stage drug diversification with the help of an efficient screening method.

No directed evolution experiment is successful without sufficient genetic diversity and a high-throughput screen or selection method[54]. With modern advancements in molecular biology techniques, mutagenesis has been able to generate libraries on the order of 10^{4-6} when conducted *in vivo*[55] and 10^{15} different variants using *in vitro* compartmentalization techniques[56]. Moreover, gene amplification techniques are relatively simple using PCR. The most commonly used method for generating libraries is the error-prone polymerase chain reaction (epPCR), which randomly introduces mutations into the gene using a DNA polymerase that is reduced in fidelity. Arguably, this method is one that is closest to Nature's mechanisms, as new enzymes that find use are a result of random mutations occurring in the genome. This process is achieved in the laboratory by introducing alkylating agents to the PCR reaction. Typically, variants contained within the library are aimed to contain on to five base pair mutation[55]. A lower mutation rate is preferred so as to not have a higher likelihood of killing the enzymes activity with the introduction of too many mutations. Even with a low mutation rate, the number of variants contained in a library that are different from the parent and each other is astronomical. For example, with 20 canonical amino acids and a rate of one amino acid substitution per variant for a 100 amino acid protein, the number of different variants that could be generated would be 20x100=2000 variants, and for two substitutions, 40,000 variants. However, as more substitutions are introduced per variant, this number can swell to 20^{100} variants. Additionally, according to the Tawfik group, the frequency of mutations that are considered beneficial is typically very low (on the order of 0.1%) and more than 40% of mutations produce defunct enzymes[57]. To add more complexity, in some cases, two neutral or deleterious mutations on their own may be beneficial when combined. As a result, the burden

of successfully identifying an improved variant has shifted onto the screening method required to survey such large space.

The field of directed evolution has been consistently limited by our abilities to screen a library for desirable variants. In directed evolution experiments confined to microplates like in the studies mentioned above, conventional screening efforts usually rely on high-performance liquid chromatography (HPLC) and usually require a chromophore on the molecule of interest[58,59]. Such assays can only screen several hundred variants per round of evolution, thus limiting the rate of evolution and sequence space that can be explored[60,61]. It is no secret that the field of directed evolution could benefit immensely from a robust screening method, that is, one not only high-throughput in manner, but also easily transferable to measuring other catalytic activity.

2.2 Throughput Levels in Screening: A Persistent Bottleneck

Directed evolution not only relies on the generation of a library but also is dependent on exploring a maximum of combinatorial diversity[62]. In the studies mentioned above, the rate of product formation was screened using various methods of chromatography including gas chromatography (GC), supercritical fluid chromatography (SFC), high pressure liquid chromatography (HPLC), in combination with NMR and mass spectrometry. Enzyme reactions are first carried out by adding the reaction components and either whole cells, crude cell lysates, or the purified proteins to microtiter plates. The use of microwell plates remains the most commonly used platform and are superior to shake flasks due to their compatibility with liquid handling technologies, convenience and reliability[63,64]. The samples are then extracted or purified and individually sampled via chromatography. Although adequate, these screens are low throughput as each sample must be run separately and may require different combinations of these

methods. As a result, the sequence space explored is limited to under a thousand mutants per round of evolution.

To handle the screening of larger libraries, many techniques based on colorimetric assays have been developed using plate readers[54]. For example, the absorbance of 2,2—diphenyl-1-picrylhydrazyl was measured at 517nm in order to develop a simple and rapid pre-screening method for identifying fermentation conditions and gene combinations that result in the biosynthesis of monoterpenes but is dependent on the chemistry of the metabolites requiring a change in UV/Visible spectra to be identifiable[65]. Another example of success is where Blanch and Clark reported high throughput methods that couple in vitro expression with glycan hydrolysis for screening glycoside hydrolases in order to develop enzymes with higher activity and stability to be more suitable for industrial applications[66]. Microfluidics can also be adapted for ultrahigh-throughput measurement of the absorbance of enzymatic products. For example Hollfelder and coworkers improved the activity of a phenylalanine dehydrogenase toward its native substrate by measuring reduction of its cofactor[67]; however, the droplet-sorting infrastructure required here is not always available in most laboratories, and not all transformations have cofactors.

In other developments, fluorescence activated cell sorting (FACS) has been shown to handle libraries on the order of 10^7. For example, the Reetz group of Germany developed a method to isolate enantioselective hydrolases using fluorescent dyes and cell surface display[68]. Additionally, sialyltransferases have been evolved by monitoring the formation of sialosides in intact *E. Coli*[69]. Here, the authors selectively trapped fluorescently labeled products in the cells by using a carefully designed fluorescently labeled acceptor sugar. These techniques like many FACS-based techniques, however, are dependent on the generation of a tag that which is not always available for evolving many transformations[70,71]. It has also been shown that the presence

of a tag can lead to false positive results with one of the most common instances being the resveratrol conundrum[72]. Moreover, all of these methods mentioned above are typically useful for screening the specific applications they were designed for but may not be compatible with other reactions.

As a more general approach, assays based on high-throughput MALDI-MS have been looked upon to screen variants[73]. The tool is considered invaluable because it offers a label-free analysis of molecule as it only depends on changes in molecular weight alone. The technique has also been shown to rapidly inspect a large number of biological samples[74,75]. However, the purification steps prior to analysis sacrifices throughput for sensitivity. Modern high-throughput techniques that couple mass spectrometry (MS) to other devices are also either costly or not easily adoptable[76–78] and are often only amenable to only one type of transformation[79]. Moreover, while these methods may be used to screen analytes unbiasedly, a chromatographic purification step typically impedes throughput. This has prompted the development of alternative MS-based assays that rely on multiple capturing strategies to immobilize products prior to analysis[80]. A simpler and more general technique enabling high throughput identification of desired function from large libraries of variants created randomly would prove to be beneficial. In all, there still does not exist an adequate high-throughput platform capable of screening a wide variety of these biochemical transformations in an efficient manner, leaving a large portion of new pathways yet to be unlocked.

When a high-throughput screen is unavailable, our efforts turn either to reducing the size of the pool by rational design[81–83] or to modern bioassays utilizing affinity- and activity-based reporter systems or survival selection systems[84]. For example, small site-saturation mutagenesis libraries can be created and screened by HPLC. These approaches fail, however, when the relationship between structure and function is poorly understood[85] or from incompatibilities with

non-natural chemistry due to a greater diversity of building blocks and the chemical reactions involved,[46] respectively. When structural knowledge about the protein is not well documented making rational design impractical and/or attempts to mutate the active site do not achieve significant gains in activity, methods of random mutagenesis are turned to[81]. Such libraries often contain few beneficial mutations resulting in the need to screen hundreds to thousands of variants to find improvement, thus requiring extended instrument time for screening. While machine learning looks to be a promising aid in directed evolution, such models are fairly new and still require robust screening methods for data input[86]. Until the paradigm shifts, directed evolution will continue to command efficient high-throughput and versatile screens for its success. Therefore, a technique enabling the identification of desired function from large libraries of variants created randomly would prove to be beneficial.

Despite decades of achievements, the analysis of activity – which relies largely on methods of chromatography – remains the bottleneck in the method, lacking throughput, generalizability and an ability to handle complex solutions. For this reason, prominent scientists across the globe have been calling for and working to develop high-throughput screens. In a recent review, the Hilvert lab of Switzerland underscored the importance of high-throughput screening methods for investigating sequence-function relationships, calling the exploration of sequence space of enzyme catalysts a "numbers game[87]." Professors Schwaneberg of Germany and his peers regularly cite the time involved in conventional screening methods as "a major bottleneck" in the field and that major design considerations must be implemented in screening assays to ensure that they can produce highly significant results in a cost effective manner[88,89]. Professor Arnold of the United States and 2018 Nobel Laureate in Chemistry for the "Directed Evolution of Enzymes" and her group, in their most recent development of a high throughput screen for terpene synthase

cyclization activity, have written that while directed evolution is a reliable method for optimizing enzyme performance, “it requires an appropriate high-throughput assay for screening mutant libraries”[90]. Until this limitation in the paradigm shifts, directed evolution will continue to command efficient, high-throughput and versatile screens for its success.

In the next chapter, I describe my work at Northwestern University, in collaboration with the group of Frances Arnold at Caltech, to develop self-assembled monolayers for matrix-assisted laser desorption ionization (SAMDI) as a useful high-throughput screening tool that provides a suitable and more general platform to screen a wide variety of directed evolution campaigns in order to meet that demand.

Chapter 3

High Throughput Screening with SAMDI Mass Spectrometry for Directed Evolution

This work was done in collaboration with Dr. Frances Arnold and members of her lab at The California Institute of Technology. The research and all figures presented in this chapter have been published or are adapted from the publication: Pluchinsky, A. J.; Wackelin, D. J.; Huang, X.; Arnold, F. H.; Mrksich, M. High Throughput Screening with SAMDI Mass Spectrometry for Directed Evolution. J. Am. Chem. Soc. 2020, 142 (47), 19804–19808. https://doi.org/10.1021/jacs.0c07828.[91]

3.1 Introduction

Directed evolution represents a viable route to developing biocatalysts for synthetic organic chemistry,[92–95] including many non-natural transformations.[96–101] With substantial advances in our ability to generate genetic diversity[102] and prepare libraries exceeding thousands of variants,[59] the screening of activity remains a significant bottleneck for many reactions. Conventional screening efforts largely rely on optical methods, which oftentimes have a defined reaction scope[103,104] as they usually require a suitable chromophore on the molecule of interest[105–107] or are based on the detection of a coupled co-product[79]. While methods employing mass spectrometry (MS) have the advantage that they are label-free and therefore quite general, a chromatographic separation step is often necessary, which can limit throughput.[59,108–111] These methods limit both the rate at which evolution is performed and the sequence space explored, often with only several hundred variants screened per round of evolution in a suitable timeframe.[60,61] This constraint has prompted the

development of alternative MS-based assays;[77,80,88,106,112,113]however, prior to the research demonstrated in this thesis, no screening effort based on molecular mass alone had been developed to reach the throughput of conventional screening methods that rely on absorbance or fluorensence[111]. In this chapter, I describe the development and first application of self-assembled monolayers for matrix-assisted laser desorption ionization (SAMDI) that provides both a generalizable platform to enable screening in a variety of directed evolution campaigns and screening throughput on-par with that of plate readers, thus, eliminating the screening bottleneck in directed evolution.

In recent work, a cytochrome P411—originally derived from a cytochrome P450 by substituting the native cysteine axial ligand for serine—was further evolved to become the first iron-based biocatalyst to perform alkylation of sp^3 C–H bonds through carbene C–H insertion, providing an efficient biocatalytic route for this highly-challenging and valuable transformation.[59] However, evolving this catalyst to perform this transformation on a variety of substrates requires the use of chromatography to detect the reaction products, with benzylic and alpha-amino alkylation relying on high pressure liquid chromatography (HPLC) for their detection and alkylation of substrates containing allylic C–H bonds limited to even lower throughput methods of gas chromatography (GC). Further evolution of this variant would benefit from a higher throughput screen able to identify any of these transformations. Such a technique would also be designed to enable the identification of desired function from large libraries of variants created randomly.

3.2 Results

Leveraging SAMDI's abilities to assay enzyme activity[21,114] and rapidly analyze thousands of small molecule reactions directly from complex solutions,[12,115] we sought to continue evolving this catalyst for C–H insertion activity in high throughput. We chose an allylic substrate which was among the most challenging to detect[59] (Figure 1A); developing a screen for this reaction is difficult because the products are not easily ionizable, do not possess a significant chromophore or generate a fluorescent signal, and cannot be linked to the viability of the cell, coupled to a measurable co-product, or make use of a biological reporter system.

To evolve enzymes for this reaction, we generate libraries containing cytochrome P411 variants in well plates and allow the variants to catalyze the reaction on an acetate-protected substrate (Figure 1A). With our approach, we then use self-assembled monolayers to selectively immobilize the substrate and reaction product directly from cell suspensions (Figure 1B). Based on the chemistry available on the reaction products, we chose to engineer the surface to present maleimide groups against a background of tri(ethylene glycol) groups. We can treat the reaction products with acid to reveal the thiol, which allows immobilization to the monolayer via a Michael addition. We then use SAMDI MS to measure the masses of the analyte-alkanethiolate conjugates by matrix-assisted laser-desorption ionization mass spectrometry (MALDI-MS).[3,10] In this way, we need only identify the products by a corresponding change in mass and integrate the peaks of the substrate and product to provide a yield for the reaction (Figure 1C).

In this study, the peak corresponding to substrate capture was present at 1033 Da, and the product peak was shifted by +86 Da (Figure 1C, right). For each spectrum acquired, we calculated

relative product yields from the area under the curve (AUC) of each peak using $AUC_{product}$ / ($AUC_{substrate}$ + $AUC_{product}$). Each variant was screened in quadruplicate to acquire an average yield and account for variability in the deprotection and immobilization steps. We then normalized the values by the average value of parent activity on each respective plate to acquire a fold improvement. For each library, we generated heat maps to visualize the relative activities of each variant (Figure 1C, left). Variants were shaded by their fold improvement where orange represents increased activity and purple, decreased activity relative to parent-like activity (white). The most promising library members were run at analytical scale and validated using gas chromatography mass spectrometry (GCMS). We then selected the best variant to be the parent of the next round of evolution.

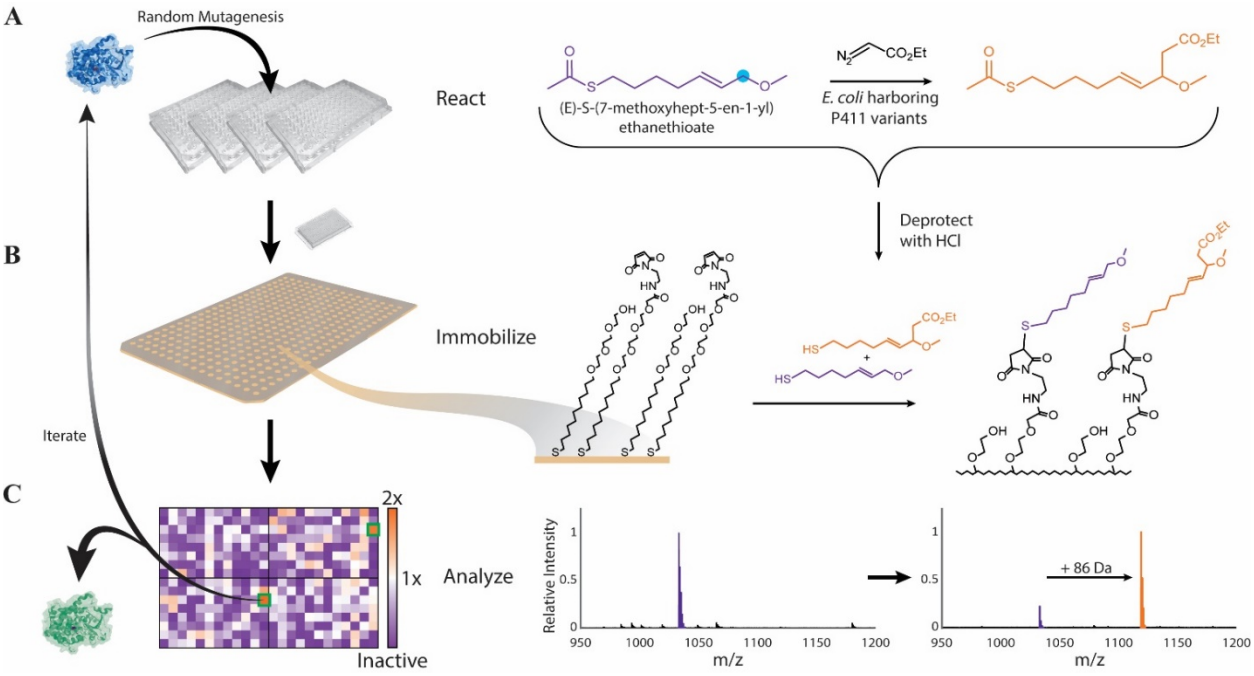


Figure 3.1. Use of the SAMDI screening assay in a cycle of directed evolution. (A) Libraries of cytochrome P411 are expressed in 96-well plates and allowed to react with the substrate (purple) and ethyl diazoacetate to form the ester product (orange). (B) Reaction products are deprotected and transferred directly to a SAMDI plate where they immobilize to a maleimide-presenting monolayer. (C) The array is analyzed by MS and results are displayed as a heat map where each variant is shaded by fold improvement measured over its parent.

To identify the best starting variant for the first round of evolution, we screened a diverse panel of heme proteins. Variants of cytochromes P411, cytochromes c, and globin homologues were screened, including variants evolved by Zhang et al. for C–H functionalization[59]. Reactions were carried out using (E)-S-(7-methoxyhept-5-en-1-yl) ethanethioate and ethyl diazoacetate (EDA) in *Escherichia coli* whole cells using the general procedure for small scale enzymatic reactions. Product formation for each variant was characterized and, while several variants

exhibited activity on the substrate, one variant was chosen that had the highest product formation and with one mutation (P74T) from P411-CHF identified by Zhang et al.[59] With P411-CHF(P74T) as the initial parent, we first used SAMDI to measure the retention of function of enzymes in libraries generated by error-prone PCR at various manganese chloride concentrations (Figure 3.2).

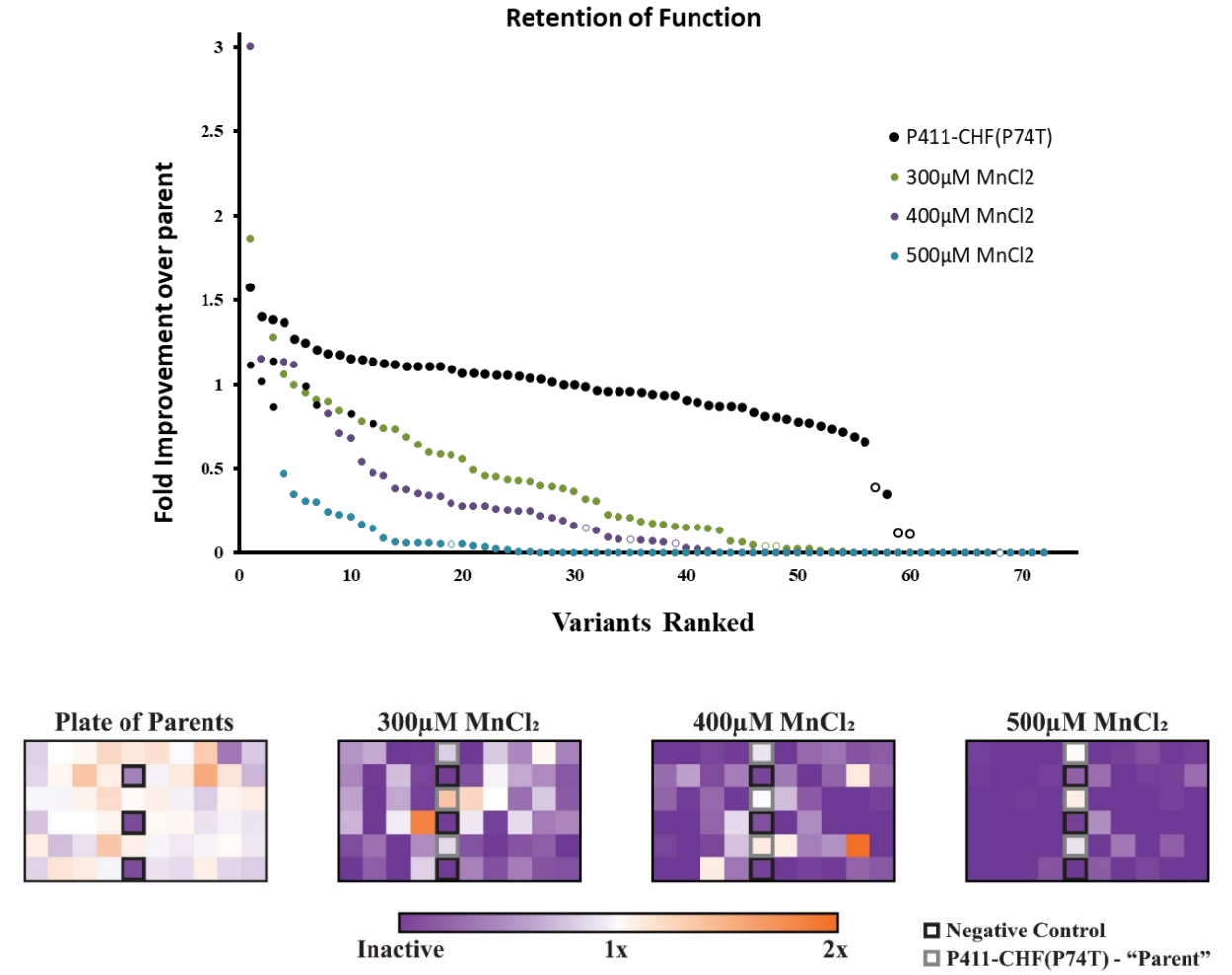


Figure 3.2. Results from 3 mutational loads compared to a plate of parent enzymes. Following the procedure in the experimental section, random libraries of P411-CHF(P74T) were generated using error-prone PCR and various concentrations of $MnCl_2$. A mutation rate using 400µM $MnCl_2$ was adequate to proceed with this enzyme. All parents are show in black.

From these data, we found that SAMDI was able to rank variants with a least-squares correlation of R^2=0.96 to data collected by GCMS (Figure 3.3). While both techniques identified one variant in this library as having potentially improved activity, further validation confirmed this hit to be a false positive (See Appendix Figure S1).

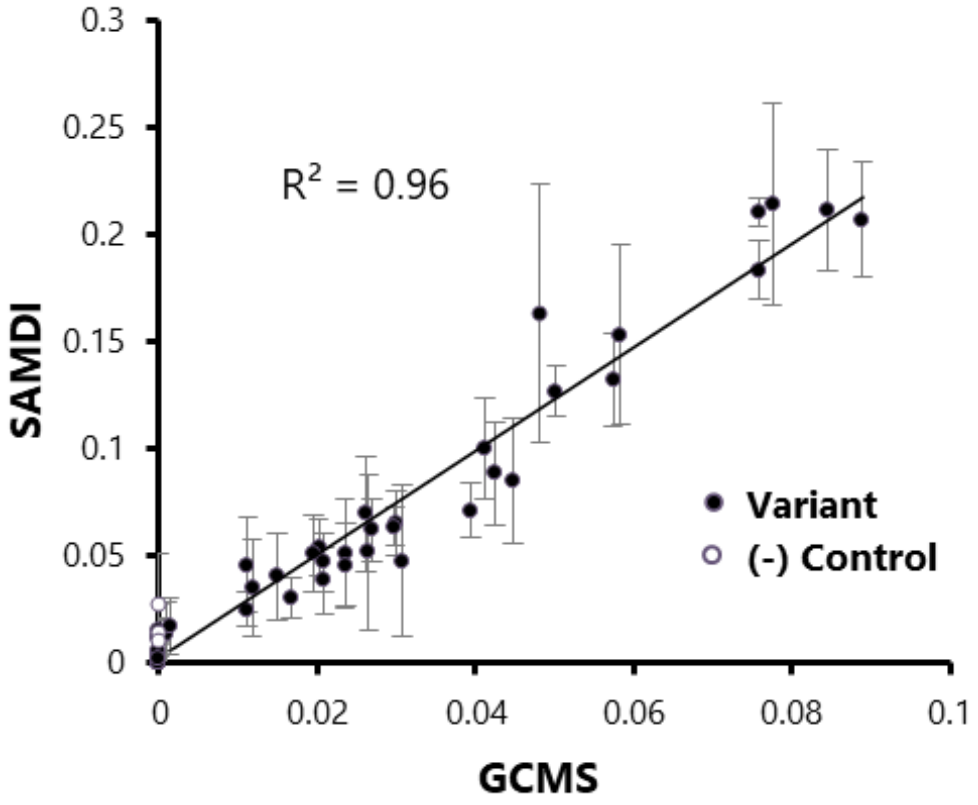


Figure 3.3. Scatterplot of screening results. Screening results were collected via GCMS and SAMDI from a library of 70 mutants. Values were calculated as a fraction of product over the total of the remaining starting material and product formed. Correlation was determined using least squares linear regression.

In order to identify biocatalysts with increased activity, we performed iterative rounds of mutagenesis and screening in whole *E. coli* cells. Because SAMDI can handle the large sequence space of random libraries, we opted to generate mutations throughout the entire gene using error-prone PCR. Over the course of three rounds of evolution, we acquired data for nearly 5,000 variants (Figure 3.4).

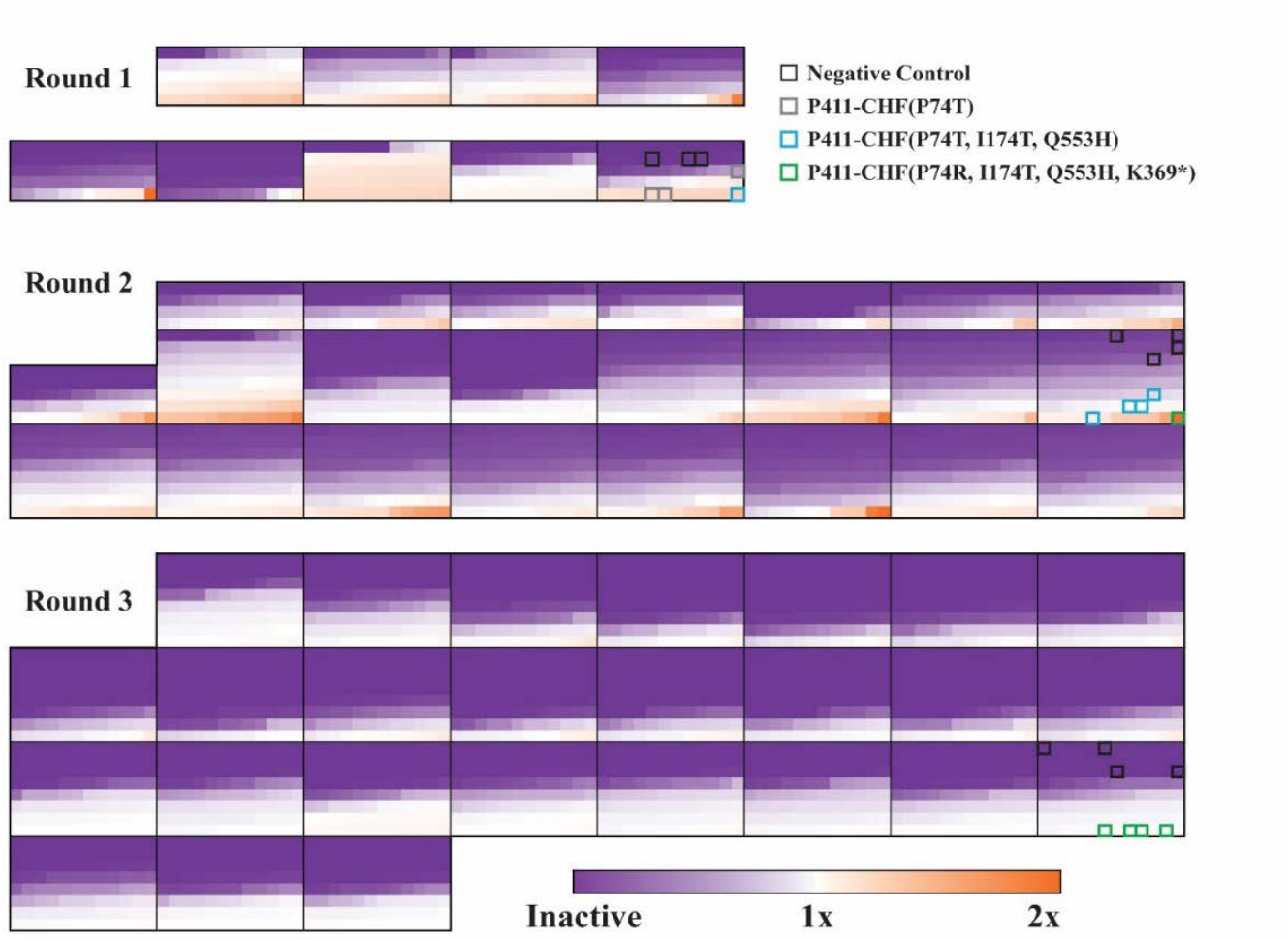


Figure 3.4. Heatmaps display the relative activity of 4,956 samples measured via SAMDI. Mutants are shaded and organized by relative fold improvement normalized by the average of parent controls on each respective 96-well plate. Visible differences in the amount of active versus inactive variants in each heat map can be attributed to the number of mutations in each variant.

Variants identified by SAMDI as having a fold improvement of more than two times the relative standard deviation on the average of parent activity from each library were selected to be validated by GCMS. The mutant confirmed to be the most active was then chosen for the next round of evolution. The best confirmed variant from each round (boxed in blue for round 1 and in green for round 2) were used as the parent in each subsequent round of evolution. In the third round of experiments, none of the variants were verified either by SAMDI or via GCMS to have significantly improved activity over the previous parent. In the third round, we did not identify a variant with significantly improved activity out of the 2,496 variants screened and decided to end the campaign. We chose the best variant from round 2 as the final variant.

Importantly, all of the data for these variants was acquired approximately 140-fold more rapidly than what would be expected with GCMS (Figure 3.5). Here, data generation for each round required only a few hours, reducing the total analysis time from 24 days (for 1 replicate) to 17 hours (for 4 replicates). In the third round we screened nearly 2,500 variants and did not find a significantly improved enzyme, suggesting that the enzyme may be approaching a local maximum in activity or may need stabilizing mutations before further activating mutations can be found. Experimental procedure may also need to be reworked to avoid possible limitations in dynamic range as enzymes improve. As we sought to evolve on this platform as a proof of principle and managed to do so, we decided to end the campaign.

	Round 1	Round 2	Round 3	Total
# of Variants	540	1,920	2,496	4,956
Time GCMS 7min/sample	63 hours	224 hours	291 hours	24 days
Time SAMDI 3sec/sample	27 mins	96 mins	125 mins	248 mins
x4 replicates	1.8 hours	6.4 hours	8.3 hours	~17 hours

Figure 3.5. Comparison of throughput and total screening effort of SAMDI to conventional screening methodology. Screening with SAMDI-MS significantly shortens the time required to screen in a directed evolution campaign compared to one where GCMS would be utilized.

To acquire the fold improvement of each parent over their predecessor, we ran the top variants from each round at analytical scale and measured their activities against their parents by GCMS. The final variant displayed a 2-fold improvement (1300 total turnovers (TTN)) (Figure 3.6).

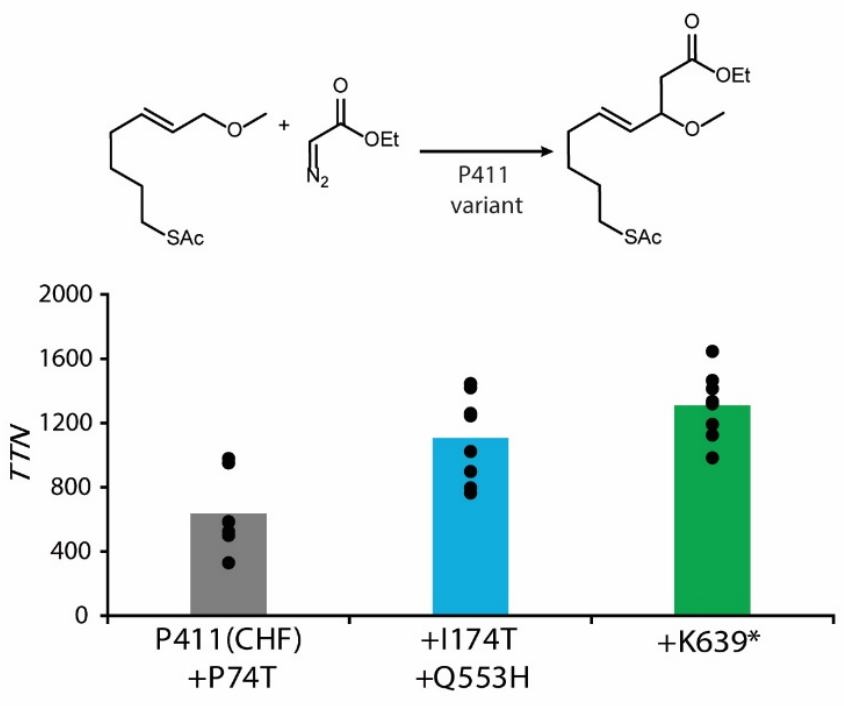


Figure 3.6. Characterization of parent lineage by GCMS. We used GCMS to characterize the hits and obtain the total turnover (TTN) for each variant. The evolutionary lineage of P411 for C–H alkylation is displayed. Bars represent mean yields (performed from two independent cell cultures, each used for duplicate reactions). Reaction conditions were as follows: cytochrome P411 in *E. coli* whole cells (optical density at 600 nm, OD600, of 1), 5 mM substrate, 5 mM ethyl diazoacetate, 5 vol% EtOH in M9-N buffer at room temperature under anaerobic conditions for 18 h. Asterisk represents the introduction of a stop codon.

Interestingly, none of the beneficial mutations were in the active site of the enzyme or at sites previously mutated in rational approaches (Figure 3.7). By not restricting the sequence space explored, we were able to identify potential allosteric effects and provide new sites that may be investigated in targeted evolution (See Appendix Figure S2).

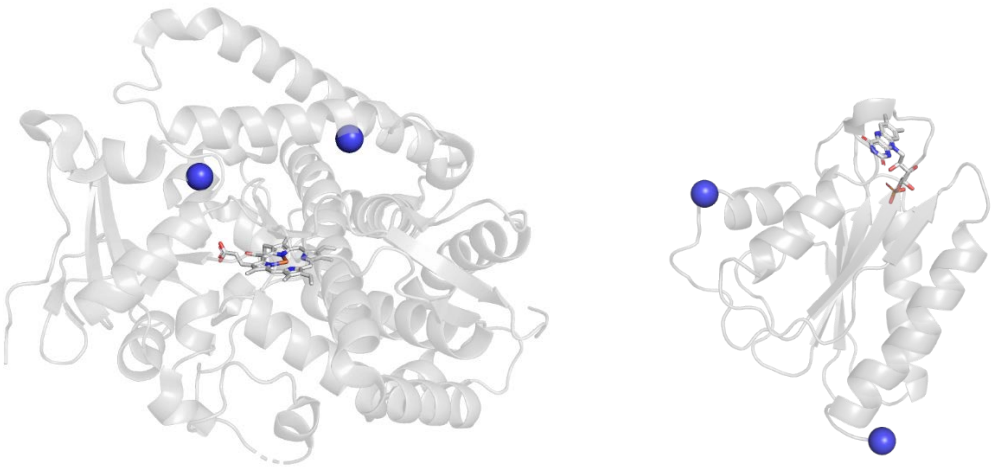


Figure 3.7. Structural visualization of amino acids mutated during evolution of P411-CHF to our final variant. The final variant was identified to contain 4 amino acid mutations from the initial parent with 2 in the heme domain and 2 in the FMN domain. The following mutations were accumulated (from left to right): In the heme domain: P74T, I174T. In the FMN domain: Both a Q553H mutation and a stop codon at K639. While the P74T mutation identified in the initial parent is located in the active site, none of the mutated residues identified by SAMDI are more than 15 Å from the iron atom in the center of the heme cofactor.

Note: The structure of the heme domain was modeled using the crystal structure of a related P411 variant (PDB 5UCW), which contains nine additional mutations. The structure of the FMN domain was modeled using the crystal structure from a related variant (PDB 1BVY).

To demonstrate the reproducibility of the SAMDI technique, we selected and scaled up one variant from the final round to be screened repeatedly with SAMDI. Here we found a standard deviation of 2.3% with a resolving power of 0.1 m/z (See Appendix Figure S3). The primary source of variability about the mean is likely due to application of matrix, which leads to modest differences in signal strength from spot to spot.[10] We also note that while the SAMDI technique is able to accurately quantitate the extent of each reaction,[116] we only required relative product yields to proceed with evolution, and thus, accurate yields were determined only for the variants validated by GCMS. Experimental reproducibility at-large was shown by inducing multiple colonies of the same clone for the initial variant. Here, we found a coefficient of variation (CV) of 14% (See Appendix Figure S4). A value of 15% or lower typically represents an acceptable assay as variability in the measurement is expected to exist within plates of the same enzyme, due to differences in enzyme expression. From this metric we estimate that SAMDI can reliably identify improvements greater than 3 times the CV (i.e. 1.42 fold or higher).

In this study, the use of a thiol-containing reactant to allowed for convenient immobilization of reaction products to the monolayer. The presence of a thioacetate is also high yielding and orthogonal to the molecule's existing moieties, and importantly, the acetate protecting group is unaffected by the transformation. While an unprotected thiol is chemically reactive and would interact with cellular milieu and the carbene, thioacetate protection masks the thiol and

allows for the reaction to proceed; however, molecules with these group would not be compatible with many other reactions. In the event that a thioacetate group may not be preferable on the substrate of interest, the monolayer may be modified with other functional groups to accommodate different chemical moieties. Hence, we first repeated the C–H insertion reaction demonstrated in this study, but we chose to immobilize via a different capture scheme via a 'traceless' immobilization method that did not rely on the presence of the thiol group. Here we found the technique was able to capture all four species that are both reaction products both in their free thiol and unprotected thioacetate forms (Figure 3.8) In this scheme, a monolayer presenting a diazirine group reveals a carbene on irradiation, which reacts non-selectively to immobilize nearly all molecules[26,117].

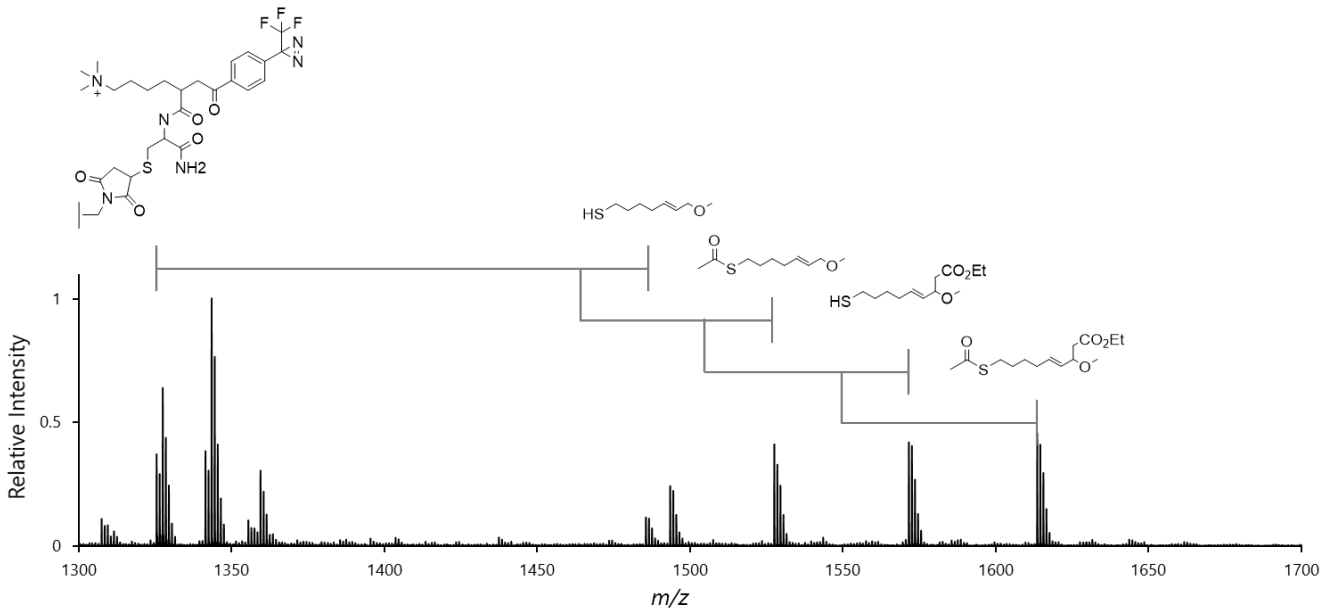
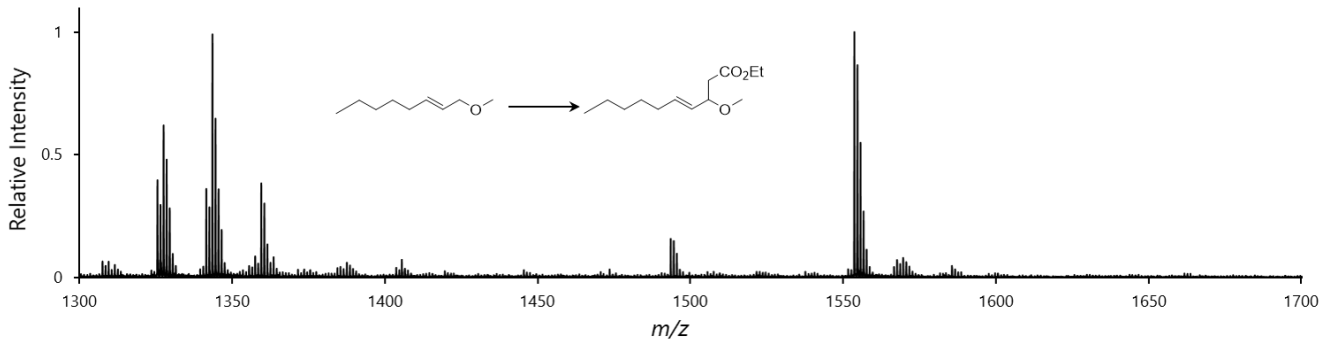


Figure 3.8. Spectra acquired via 'traceless' immobilization reveal capture of all four species. Using methods previously reported elsewhere[117], a diazirine-terminated monolayer was used to analyze the reaction products for the reaction from Figure S5, demonstrating that analysis of this reaction was not reliant on thiol-maleimide chemistry. Here, SAMDI-MS spectra displays

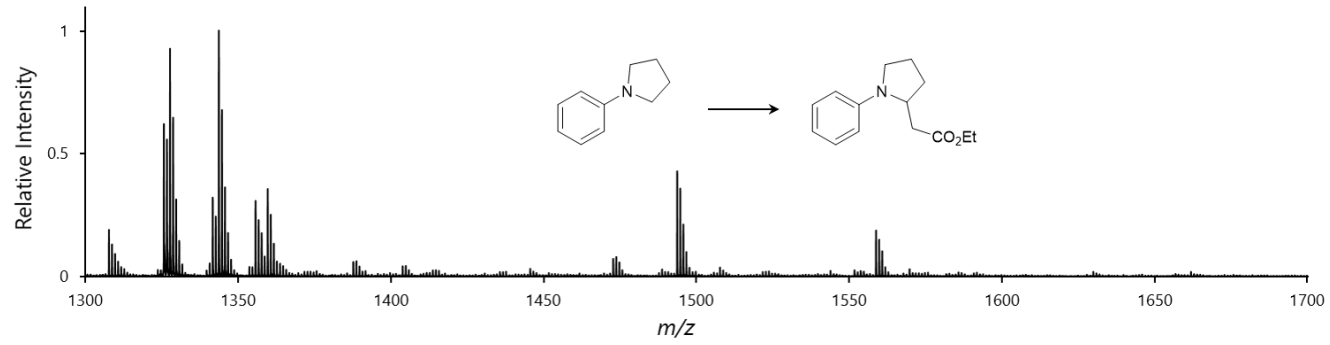
the diazirine monolayer at m/z 1325 (after loss of nitrogen and conversion to carbene) and clear peaks corresponding to both the deprotected and remaining acetate-protected reaction products. Crude lysates were briefly deprotected in 28mM HCl for 30minutes and quenched with saturated sodium bicarbonate in order to show the ability of this monolayer to immobilize all 4 species. No extraction or centrifugation was performed prior to analysis. The deprotected and acetate-protected substrate appear at m/z 1485 and m/z 1527, respectively, and their corresponding insertion product are seen + 86 Daltons at m/z 1571 and m/z 1613, respectively. Primary byproducts are seen due to reaction with water (m/z 1343) and 2,4,6-trihydroxyacetophenone (m/z 1493), the MALDI matrix.

We then performed the C–H insertion reactions for five additional substrates that lacked a thiol or thioacetate altogether and used the traceless immobilization scheme to analyze products, and in each case observed peaks in the mass spectra that corresponded to reactant and product (Figure 3.9). This example demonstrates that reactants need not be functionalized for immobilization, and in turn suggests this method will have a very broad relevance in directed evolution.

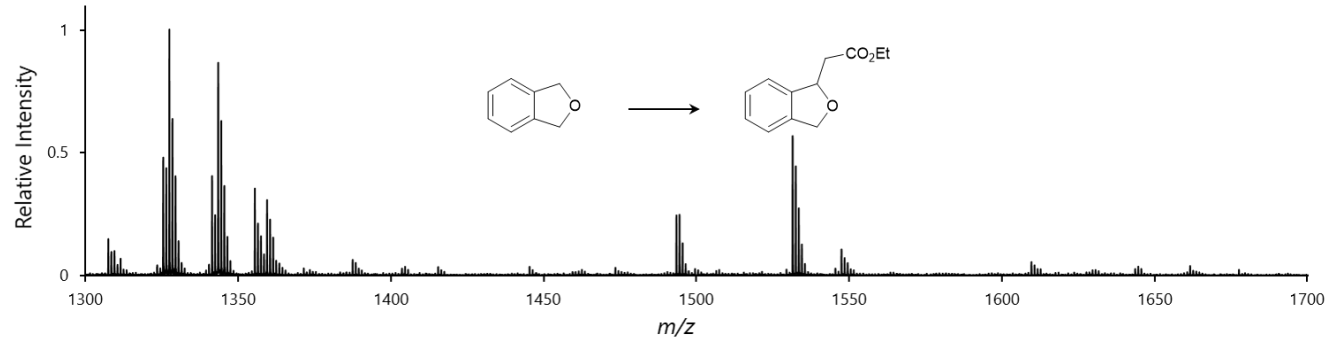
a



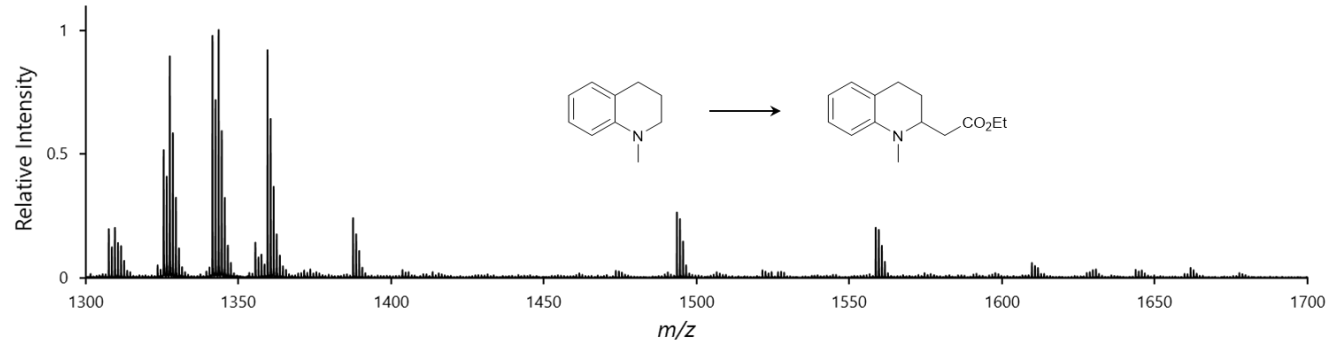
b


Relative Intensity
1
0.5
0
1300
1350
1400
1450
1500
1550
1600
1650
1700
m/z
N
N
CO2Et

c


Relative Intensity
1
0.5
0
1300
1350
1400
1450
1500
1550
1600
1650
1700
m/z
O
O
CO2Et

d


Relative Intensity
1
0.5
0
1300
1350
1400
1450
1500
1550
1600
1650
1700
m/z
N
N
CO2Et

e

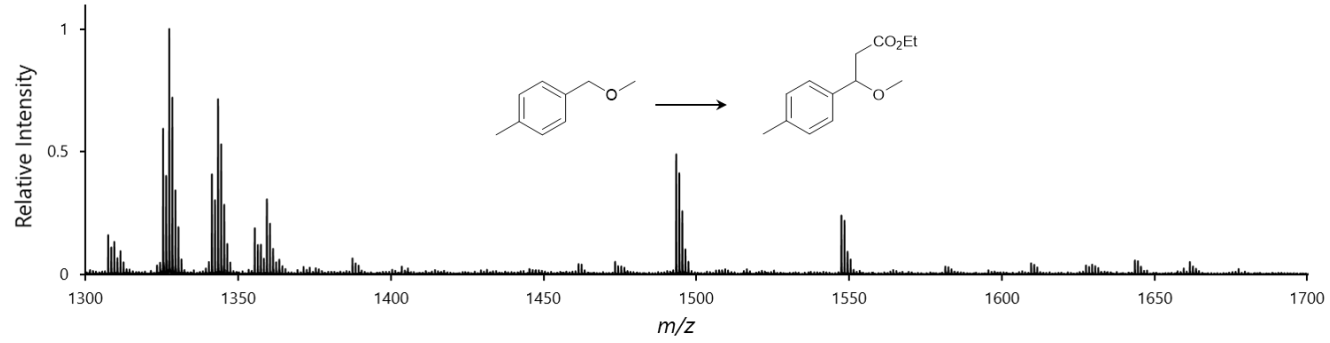

Relative Intensity
1
0.5
0
1300
1350
1400
1450
1500
1550
1600
1650
1700
m/z
O
O
CO2Et

Figure 3.9 (a-e): Spectra acquired via 'traceless' immobilization for an expanded substrate scope. SAMDI-MS data that demonstrates the technique's generalizability by its capacity to record product formation for a variety of reactions within the substrate scope of the enzyme. Because the expected reaction products do not contain functional groups that may be exploited for straight-forward immobilization, reactions were analyzed using a diazirine monolayer and conditions reported in Figure S7. The evolved P411 variant was used to catalyze the reaction of ethyl diazoacetate and **a.** (E)-1-Methoxyoct-2-ene, **b.** Ethyl 2-(1-phenylpyrrolidin-2-yl)acetate **c.** Ethyl 2-(1,3-dihydroisobenzofuran-1-yl)acetate **d.** Ethyl 2-(1-methyl-1,2,3,4-tetrahydroquinolin-2-yl)acetate and **e.** 1-(Methoxymethyl)-4-methylbenzene. The peaks corresponding to the substrate and product for these reactions are as follows: **a.** m/z 1467 and 1553, **b.** m/z 1472 and 1558, **c.** m/z 1445 and 1531, **d.** 1472 and 1558, and **e.** m/z 1461 and 1547, respectively. As expected, all product peaks are shifted +86 Da. Further knowledge of the differences in ionization and immobilization efficiency between the substrate and their product is required to quantitate yields; however, it is clear that the evolved P411 variant remains active for a number of substrates.

A range of non-natural enzymatic chemistries have recently been demonstrated using a cytochrome[59]. Such discoveries also led to the current research. However, in these studies, HPLC was used as the highest throughput option available, with a screen time of 3 minutes per variant at best. Moreover, relying on GCMS for those molecules that were difficult to detect otherwise required a minimum of 7 minutes per well. Throughout the course of this study, 22,944 spectra were generated and processed. With each plate requiring only 30 minutes per run (~4 seconds per sample), SAMDI collected data more than one hundred-fold faster than classic GCMS and approximately ten-fold faster than recent developments in the state-of-the-art thereof.[109,118] If

higher throughput is desired, the method may be accelerated by approximately two-fold by working with groups of sixteen 96-well plates in 50 minutes per run, with each sample requiring only 0.5 μL from each well.[115] To assess the entire screening effort at large, we first take into account the 30 minutes to deprotect, ~1.5 hours to immobilize and prepare each plate and 30 minutes to evaluate data for each group of plates. With each reaction deprotected and spotted in quadruplicate, the entire screening process can be reduced to a sum of under 48 working hours. While this represents the ideal case if plates were to be run concurrently, we note that directed evolution is an iterative cycle requiring a multitude of steps in between rounds of screening.

3.3 Conclusions

SAMDI-MS has been used extensively to profile enzymatic activity both in biochemical reactions and from complex lysates, while permitting the analysis of up to thousands of samples per hour and more than 30,000 experiments per day.[12,21,25,32,33,116] Hence, it is clear that the throughput in this study was not limited by the number of variants that could be screened.

The approach described here has the benefits that it is low cost, high-throughput, compatible with existing liquid handling technologies and all library diversification techniques performed in multi-well plates, requires no purification outside of simply rinsing the SAMDI plate post-immobilization, and may be applied to any reaction that produces a shift in mass.[115] While epPCR allowed SAMDI to find modest fold improvements in the present work, we expect that utilizing other diversification techniques will result in greater improvements.

While many developed screening assays are linked to the activity they seek to measure, SAMDI-MS can accommodate a wide variety of chemical transformations—without sacrificing throughput—as other immobilization strategies have been demonstrated and are readily available.[3,25,26,117,119] In this way, the assay is not limited to certain classes of reactions but can be adapted to many organic transformations. We note that this method cannot be applied to reactions where the product and substrate share the same mass—including stereoisomeric and tautomeric structures—and would in those cases require a second reaction step (that is selective for one of the molecules), tandem mass spectrometry, or a separation step. We also note that while the monolayers have excellent stability in aqueous environments at biological pH and room temperature, their use is not compatible with strong oxidizing, acidic, or basic reagents and temperatures exceeding 70 °C. For use in whole-cell reactions, however, their use is not limiting.

As directed evolution continues to add new chemistries to Nature's repertoire, generating small molecules with increasing complexity,[120] the need for high-throughput and generalizable screening tools is paramount. Here, I showed how SAMDI can be expanded as a new high-throughput screening platform that resolves the bottleneck in directed evolution for synthetic organic chemistry. This work demonstrated that access to a high-throughput screen not only allows for faster evolution, but also allows for evolution to be performed without limiting the sequence space that can be explored. Notably, the technique requires only a few microliters of whole cell lysates, without prior purification and without the requirement of a chromophore or fluorescent tag on the molecule of interest; however, it is the use of immobilization chemistry that distinguishes SAMDI's throughput and substantiates the method to be well suited for evaluating variants in applications of directed evolution. This platform enables directed evolution efforts to

evolve enzymes for improved activity and interrogate wider areas of protein space. By resolving this bottleneck, we anticipate that further use of this method will lead to exploring larger areas of chemical space in high throughput and help uncover unexpected solutions for creating better enzymes.

3.4 Methodology

General. Unless otherwise noted, all chemicals and reagents were obtained from commercial suppliers (Sigma-Aldrich, VWR, Alfa Aesar, Combi-Blocks) and used without further purification. Disulfides used to form self-assembled monolayers were purchased from ProChimia Surfaces. Silica gel chromatography was carried out using AMD Silica Gel 60, 230-400 mesh. ^{1}H and ^{13}C NMR spectra were recorded on a Varian Inova 300 MHz, Varian Inova 500 MHz, or Bruker Prodigy 400 MHz instrument in CDCl3 and are referenced to residual protio solvent signals. Data for ^{1}H NMR are reported as follows: chemical shift (δ ppm), multiplicity (s = singlet, d = doublet, t = triplet, q = quartet, p = pentet, sext = sextet, m = multiplet, dd = doublet of doublets, dt = doublet of triplets, ddd = doublet of doublet of doublets), coupling constant (Hz), integration. Sonication was performed using a Qsonica Q500 sonicator. High-resolution mass spectra were obtained at the California Institute of Technology Mass Spectral Facility. Synthetic reactions were monitored by thin layer chromatography (TLC, Merck 60 gel plates) using a UV-lamp or an appropriate TLC stain for visualization. *E. coli* cells were grown using Luria-Bertani medium (LB) or Hyperbroth (AthenaES) (HB) with 0.1 mg/mL ampicillin (LBamp or HBamp). Primer sequences are available upon request. T5 exonuclease, Phusion polymerase, and Taq ligase were purchased from New England Biolabs (NEB, Ipswich, MA). M9-N minimal medium (abbreviated as M9-N buffer; pH 7.4) was used as a buffering system for whole cells

and lysates, unless otherwise specified. M9-N buffer was used without a nitrogen source; it contains 47.7 mM Na_2HPO_4, 22.0 mM KH_2PO_4, 8.6 mM NaCl, 2.0 mM $MgSO_4$, and 0.1 mM $CaCl_2$.

E. coli cells were grown using Luria-Bertani medium (LB) or Hyperbroth (AthenaES) (HB) with 0.1 mg/mL ampicillin (LBamp or HBamp). Primer sequences are available upon request. T5 exonuclease, Phusion polymerase, and Taq ligase were purchased from New England Biolabs (NEB, Ipswich, MA). M9-N minimal medium (abbreviated as M9-N buffer; pH 7.4) was used as a buffering system for whole cells and lysates, unless otherwise specified. M9-N buffer was used without a nitrogen source; it contains 47.7 mM Na_2HPO_4, 22.0 mM KH_2PO_4, 8.6 mM NaCl, 2.0 mM $MgSO_4$, and 0.1 mM $CaCl_2$

Chromatography. Chiral GC was conducted using an Agilent 7820A instrument (FID) and an Agilent CycloSil-B column (30 m × 0.32 mm, 0.25 µm film). Gas chromatography-mass spectrometry (GCMS) analyses were carried out using a Shimadzu GCMS-QP2010SE system and J&W HP-5ms column (30 m × 0.25 mm, 0.25 µm film).

Cloning and error prone mutagenesis. pET22b(+) was used as a cloning and expression vector for all enzymes described in this study. Error prone mutagenesis was performed using Taq DNA Polymerase. The PCR products were digested with DpnI, gel purified, and ligated using Gibson Mix$^{TM\ 121}$. The ligation mixture was used to directly transform electrocompetent *E. coli* strain E. cloni BL21(DE3) cells (Lucigen).

Expression of P411 variants in 96-well plates. Single colonies from LBamp agar plates were picked using sterile toothpicks and cultured in deep-well 96-well plates containing 3 LBamp (300 µL/well) at 37 ºC, 220 rpm shaking, and 80% relative humidity overnight. After, HBamp (950 µL/well) in a deep-well 96-well plate was inoculated with an aliquot (50 µL/ well) of these overnight cultures and allowed to shake for 2.75 hours at 37 ºC, 220 rpm, and 80% relative humidity. The plates were cooled on ice for 30 minutes and the cultures were induced with 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and 1.0 mM 5-aminolevulinic acid (final concentrations). Expression was conducted at 20 °C, 150 rpm for 16–20 hours.

Reaction screening in 96-well plate format. E. coli (E. cloni BL21(DE3)) cells in deep-well 96-well plates were pelleted (5,000 × g, 3 min, RT) and resuspended in M9-N buffer (20 µL/well) by gentle vortexing. The 96-well plate was transferred into an anaerobic chamber. In the anaerobic chamber, reaction buffer (33 mM glucose in M9-N, 380 µL/well) was added, followed by alkane substrate (10 µL/well, 400 mM in EtOH) and ethyl diazoacetate (10 µL/well, 400 mM in EtOH). In some cases, the substrates and reaction buffer were mixed together prior to addition to the plate. The plate was sealed with an aluminum foil and shaken at room temperature, 500 rpm in the anaerobic chamber. After 24 hours, the seal was removed, and the reactions were worked up following the appropriate method below.

Product formation screening using GC-MS.

After 24 hours, a solution of 20 mM 1,3,5-trimethoxybenzene (internal standard) in a mixed solvent system (cyclohexane/ ethyl acetate = 1:1, 20 µL), followed by 580 µL pure solvent, was added. The plate was tightly sealed with a reusable silicone mat, vortexed (15 s × 3) and

centrifuged (3,000 × g, 5 min) to completely separate the organic and aqueous layers. The organic layers (200 µL/well) were transferred to 300 µL vial inserts, which were then placed in 2 mL vials and analyzed by GC.

Product formation screening by SAMDI Mass Spectrometry

After 24 hours, 190 µL crude reaction mixture was aliquoted to PCR plates and frozen on dry ice. Plates were shipped overnight on dry ice.

Preparation of SAMDI Array Plate **–** Gold-evaporated array plates were made in-house from outsourced polished stainless-steel plates (Ace Metal Crafts). Specifically, purchased steel plates with dimensions matching standard MALDI plates (12.3cm × 8cm) were washed with hexanes, ethanol and water. Using an electron beam evaporator (Thermionics Laboratory Inc., Hayward, CA), the plates were coated at 0.02nm/s with a 5nm layer of titanium followed by a 35nm layer of gold spots (2.8 mm diameter) at 0.05nm/s in a 384-array fashion. Before use, plates were stored under vacuum. At least two days prior to screening, the plates were soaked in a 0.5 mM ethanolic solution of alkyl disulfides (Prochimea Surfaces) at 4°C. The solution consisted of disulfide alkanes either symmetrically terminated with tri(ethylene glycol) groups or asymmetrically terminated with tri(ethylene glycol) and a maleimide functional group in a stoichiometric ratio of 2:3 expecting to yield a self-assembled monolayer of 20% maleimide surface density[32]. The plates were removed from solution, rinsed with ethanol and dried with $N_{2(g)}$ immediately before use.

Deprotection of Reaction Constituents **–** 25 µL of each reaction was aliquoted, organizing four 96-well plates into one 384 plate. Liquid handling was performed by a Tecan Freedom EVO 200 Robot. To each well, 10 µL of 100mM HCl was added using an automatic reagent dispenser (Multidrop Combi, Thermo Scientific). Plates were sealed with aluminum sealing tape, mixed briefly, and centrifuged at 5000g for 1 min. After 30 min, 3 equivalents of $NaHCO_3$ from saturated solution was added and the plates were sealed, mixed and centrifuged at 5000g for 1 min. The plates were either frozen or analyzed using the steps below.

Activity Assay by SAMDI Mass Spectrometry **–** Using a TECAN liquid-handling robot, 3 µL from each well were spotted to SAMDI plates. The plates were allowed to incubate for 1 hour at 37°C in a humidity chamber, providing sufficient time for immobilization of the thiolated reaction constituents to react with the maleimide. The plates were rinsed with 1% Alconox detergent, deionized ultrafiltered (DIUF) water and ethanol with sufficient pressure to remove all noticeable debris, dried with a stream of $N_{2(g)}$, and coated with matrix (2',4',6'-trihydroxyacetophenone (THAP) in acetone (18mg/ml). Each spot was analyzed in positive reflector mode by MALDI-TOF (AB Sciex 5800 series). Captured reaction constituents from whole cell reactions were identified by their mass shifts relative to the unreacted maleimide. Analysis of more than 5100 spectra collected was performed using the analysis tools built into the TOF/TOF™ Series Explorer™ Software, which works to integrate the area under the curve (AUC) of the peak corresponding to a specified mass/charge ratio. For every spectrum in this study, the following peaks were integrated: 1033 Da (thiolated substrate + monolayer), 1119 Da (thiolated C-H insertion product + monolayer). Peaks below a signal to noise threshold set at 20 were not accepted. Each mass spectra generated for each well was used to determine the relative

conversion for each variant by dividing the AUC of the product peak by the AUC of both the unreacted substrate and C-H insertion product peaks. The conversion for the three parent enzymes on each 96-well plate were averaged and all other values normalized to this value to acquire relative fold improvement for each variant. The variants with the highest fold improvement from each 96-well library selected for product formation screening using GC and GC-MS detailed above.

Expression of P411 variants. E. coli (E. cloni BL21(DE3)) cells carrying plasmid encoding the appropriate P411 variant were grown overnight in 5 mL LBamp. Preculture (5 mL) was used to inoculate 45 mL of HBamp in a 125 mL Erlenmeyer flask; this culture was incubated at 37 ºC, 240 rpm for 2.25 hours. The culture was then cooled on ice (20–30 min) and induced with 0.5 mM IPTG and 1.0 mM 5-aminolevulinic acid (final concentrations). Expression was conducted at 24 ºC, 140 rpm, for 20 hours. Following, *E. coli* cells were pelleted by centrifugation (2,600 × g, 10 min, 4 ºC or 3,000 × g, 5 min, 4 ºC). Media was removed and the resulting cell pellet was resuspended in M9-N buffer to OD600 = 60. An aliquot of this cell suspension (3 mL) was taken to determine P411 concentration using the hemochrome assay after lysis by sonication. When applicable, remaining cell suspension was further diluted with M9-N buffer to the OD600 used for the biotransformation and the concentration of P411 protein in the biotransformation was calculated accordingly.

Hemochrome assay for the determination of heme protein concentration. E. coli cells expressing heme protein and resuspended in M9-N buffer were lysed by sonication using a Qsonica Q500 sonicator equipped with a microtip (1 min, 1 second on, 1 second off, 35% amplitude); samples

were kept on wet ice for this process. The resulting lysed solution was centrifuged (20,000 × g, 10 min, 4 °C) to remove cell debris. The supernatant (clarified lysate) was separated from the pellet and kept on ice until use. In a falcon tube, a solution of 0.2 M NaOH, 40% (v/v) pyridine, 0.5 mM $K_3Fe(CN)_6$ was prepared (pyridine-NaOH-$K_3Fe(CN)_6$ solution). Separately, a solution of 0.5 M $Na_2S_2O_4$ (sodium dithionite) was prepared in 0.1 M NaOH. To an Eppendorf tube containing 500 µL of clarified lysate in M9-N buffer was added 500 µL of the pyridine-NaOH-$K_3Fe(CN)_6$ solution, mixed, and transferred to a cuvette; the UV-Vis spectrum of the oxidized Fe(III) state was recorded immediately. To the cuvette was then added 10 µL of the sodium dithionite solution. The cuvette was sealed with parafilm and the UV-Vis spectrum of the reduced Fe(II) state was recorded immediately. A cuvette containing 500 µL of M9-N, 100 µL 1 M NaOH, 200 µL pyridine, and 200 µL water (complete mixture without protein and $K_3Fe(CN)_6$) was used as a reference for all absorbance measurements. Concentrations of cytochromes P411 were determined using a published extinction coefficient for heme b, ε556(reduced)-540(oxidized) = 23.98 mM^{-1} cm^{-1} [122].

Biotransformations using whole E. coli cells. Suspensions of *E. coli* (E. cloni BL21(DE3)) cells expressing the appropriate heme protein variant in M9-N buffer (typically OD600 = 60) were degassed by bubbling with argon in sealed vials for at least 40 minutes; the cells were kept on ice during this time. Separately, a solution of D-glucose (250 mM in M9-N) was degassed by sparging with argon for at least 30 minutes. All solutions were then transferred into an anaerobic chamber for reaction set up. To a 2 mL vial were added D-glucose (40 µL of 250 mM stock solution in M9-N buffer), M9-N buffer (typically 343.33 µL), suspension of *E. coli* expressing P411 (OD600 = 1, 6.67 µL), alkane substrate (5 µL of 400 mM stock solution in EtOH), and diazo compound (5 µL of 400 mM stock solution in EtOH) in the listed order. Final reaction volume was 400 µL; final

concentrations were 5 mM alkane substrate, 5 mM diazo compound, and 30 mM D-glucose. Note: reaction performed with *E. coli* cells resuspended to OD600 = 1 indicates that 6.67 µL of OD600 = 60 cells were added, and likewise for other reaction OD600 descriptions. The vials were sealed and shaken at room temperature and 500 rpm for 18 hours in the anaerobic chamber. The reactions were worked up and analyzed by GC; the reaction workup procedures are outlined in detail in Section VII. The expression of heme protein was measured using the hemochrome assay, and the concentration of heme protein in the biotransformation was calculated accordingly.

Screening of enzymes for C–H alkylation activity

Testing heme proteins for reaction discovery. A composite plate of 92 heme proteins and their variants from Bacillus megaterium and Rhodothermus marinus were screened for formation of ethyl (E)-9-(acetylthio)-3-methoxynon-4-enoate from substrates (E)-S-(7-methoxyhept-5-en-1-yl) ethanethioate and ethyl diazoacetate. These proteins were cloned and used in other studies, including carbene Si–H insertion[123], alkene cyclopropanation[51,124], and alkane C-H functionalization[59]. Expression of these proteins followed the procedures as described by the prior studies and testing for initial activity was carried out with whole *E. coli* cells. The general procedure for reaction screening in 96-well plate format (Section I (E)) was employed and the reactions were analyzed by GC-MS. Note: Since this was an initial test, proper expression of the indicated proteins was not verified and experimental replicates were not performed. A variant closely related to P411-CHF from prior alkane C-H functionalization studies[59], P411-CHF +P74T, was identified as the most promising variant for evolution.

Directed evolution of C–H alkylation enzymes. Error-prone libraries were generated using *Taq* Polymerase and differing concentrations of manganese chloride, and screened in 96-well plates. Following the general screening in 96-well plate procedure, variants which exhibited higher formation of C–H alkylation product were identified. A summary of the beneficial mutation(s) is presented in Figure S2. The locations of the selected beneficial mutations are displayed on a structural model of the P411 enzyme.

Variants which were identified to show higher activity during screening were streaked out on LB_{amp} agar plates. A single colony was selected, sequenced, and the TTN measured for our product using whole *E. coli* cells overexpressing the desired protein (the cell density was normalized such that each reaction contained the same cell density).

Chapter 4

Protein Sequencing

4.1 Protein Sequencing: Digestion and Analysis Strategies

Novel peptide sequencing methodologies that provide more precise protein sequence information are necessary for identifying proteins and furthering our understanding of the proteome[125]. The development of such techniques is also important in biological research, especially where disease-associate proteins require single-residue resolution for their discovery. With significant advancements in mass spectrometry (MS) and bioinformatics, the generation of overlapping peptide fragments by endopeptidase digest (commonly by trypsin) and reading of individual amino acids by tandem MS have become essential for *de novo* sequencing; however, such methods struggle to identify proteins' N- and C-terminal residues—information that can be used to significantly enhance a peptide mass fingerprint for protein identification and that may uncover important mutations and/or post translational modifications in a protein. Moreover, tandem MS generally lacks the ability to distinguish between residues of identical and similar masses[126]. One approach based on tandem MS has been able to achieve isobaric distinction across canonical and non-canonical regimes, but still relies on complex analysis of fragmentation ions[127].

For single residue sequencing, Edman degradation via a protein's N-terminal still remains a widely-used and reliable tool[128,129]. In this approach, the amino-terminal residue is labeled with phenyl isothiocyanate under mild alkane conditions to form a derivative of the amino acid. Under acidic conditions, the amino acid is then cleaved off as a thiazoline. These individual derivatives are extracted sequentially and examined either by chromatography or electrophoresis. From the beginning, this technique was capable of identifying all of the canonical amino acids for peptides

of up to 50 residues long[128]. After roughly 70 years since its development, it still remains a gold standard for sequencing today. The technique, however, has its challenges. With cycle times requiring roughly an hour per amino acid, and analysis reliant on chromatography, the method is generally regarded as slow. Additionally, volume of sample required and sample loss are also concerns as proteins and peptides cannot be amplified like nucleic acids. Moreover, the technique requires a free N-terminus.

To identify amino acid sequences using tandem mass spectrometry, known as MS/MS or MS^2, molecules of interest are first separated via chromatography, ionized and separated again by their mass to charge ration (*m/z*). Particular ions are then selected and split into smaller fragments prior to a second analysis by mass spectrometry which separates these smaller fragments by their m/z ratio. A number of methods using liquid chromatography to separate components prior to analysis by MS have been developed in abundance to interface with mass spectrometers that have found significant value in the characterization of peptide mixtures[130–133]. Gas phase fragmentation methodologies have also been developed for use with tandem MS, including collision-induced dissociation, photodissociation, electron capture dissociation and electron transfer dissociation, to name a few[134–137]. While high energy collision-induced dissociation has achieved single-residue resolution, including residues of identical mass, namely leucine and isoleucine, fragmentation of backbones and sidechains produces complex tandem MS spectra[138–140]. While more modern techniques have been developed[127,141–143], no new and simpler methodology has been explored. Without specialized facilities, many groups are either turning to deep learning to interpret spectra[144] or still rely on Edman degradation.

Edman degradation chemistry has also found use in the first report of peptide ladder generating methodology, characterized by the analysis of the remaining peptides versus the

individually removed amino acid analogues during degradation[145]. In this technique, Chait et.al. originally described a new principle in protein sequencing that combines multiple steps of Edman degradation and capping "ladder generating" chemistry with a final, single step mass spectrometric readout of the amino acid sequence using MALDI-MS. Since then, a variety of other chemical and exopeptidase-based cleaving methodologies have been developed for generating peptide ladders, the majority of which rely on matrix-assisted laser desorption ionization mass spectrometry (MALDI-MS), notable for substantially simpler and higher throughput data analysis in comparison to tandem MS[146–155]. By nature of this analysis, the ability of these techniques to resolve isobaric residues remains absent. Thus, ladder-generating and MALDI-based sequencing methodology that can go beyond reading changes in mass alone would advance the current state of the field.

A variety of microfluidic devices and techniques with governing microfluidic principles have been developed to interface with a MALDI mass spectrometer to study peptides and proteins. For example, it was shown that after running either peptide or protein analytes through a sodium dodecyl sulfate-polyacrylamide gel after electrophoretic separation, the gel could be directly analyzed by mass spectrometry without the use of matrix[156,157]. Kawaura and coworkers developed an isoelectric focusing chip with removable resin tape was also designed to separate proteins prior to detection by mass spectrometry[158]. In this study, a serpentine channel 60 mm long, 10 μm deep and 400 μm wide was developed to create high resolution separation of proteins. After separation, the resin was removed, the samples were freeze dried and analyzed by MALDI, where spectra were generated at 500 μm increments across the channel. A 2D plot was then created to visualize the proteins based on their m/z and pI values[158]. It was also shown that plates for thin-layer chromatography with an attached layer of porous polymer monolith could be coupled directly with MALDI to determine the molecular weight of peptides and proteins[159]. In another approach, it was

shown that infrared MALDI could be coupled to thin layer chromatography separations to study proteins from cultured cells. Other studies used microfluidic devices to achieve spatial separation and identification of protein and peptides after tryptic digest[160–164]. In all of these studies, however, while sequence coverage was enough to identify the proteins, the methods either looked at properties of whole proteins or were not sufficient to provide single residue resolution of amino acids throughout the peptide fragments. In all, microfluidic devices offer an ability to perform multiple reactions, reduce reaction volume and separate reaction products – elements all useful for sequencing; however, their use in sequencing applications has been limited due to an incapability with reagents, such as those used in Edman degradation, necessary for singular amino acid removal.

Solid-phase proteomic techniques that rely on generating peptide fragments on beads using chemical digestion[148,150,165] or ion fragmentation,[166] have also been shown to be compatible with MALDI analysis. In a most recent study, Woodbury and coworkers showed that peptides ladders could be generated in gas phase ammonia. A library of peptides, which were synthesized on solid support beads and labeled at their N-terminal with a positive charge. High pressure ammonia gas was then shown to cleave the peptides at random locations along the peptide backbone and generate peptide ladders separated by single amino acids. It was then the N-terminal labeled fragments that were separated from remaining fragments on the beads and analyzed by MALDI-MS to uncover the sequence. It was shown, however, that cleavage between different amino acid pairs was sequence dependent and required fine control over exposure time to and pressure of the ammonia gas. Additionally, although this studied claimed to demonstrate using this information along with side-chain modifications to distinguish between residues of the same and similar mass, they focus on residues of similar mass such as lysine/glutamine, glutamine/glutamate and

asparagine/aspartate, they ignore addressing the isobaric residue pair, leucine/isoleucine. In another study, Wittman and colleagues generated peptide fragments directly with MALDI-MS[166]. This method made use of a photo linker and an optimized procedure for laser-induced cleavage from the beads, followed by Fourier transform ion cyclotron resonance and MALDI tandem MS. Distinguishing between isobaric residues using the generated fragmentation ions was not explored.

With the development of so many technologies, it is clear the precise determination of amino acid sequences remains an area of significance. However, unambiguously distinguishing between amino acids, like the pairs listed above, still presents a challenge for many methods. In the cases of the residues of similar masses, it was shown that pre-analysis modifications can be used to further separate them. This liberty is not available however for the hydrophobic and isobaric amino acids, leucine and isoleucine. We hypothesized that the SAMDI method, a hallmark technology of the Mrksich group, can be used to sequence peptides in only one degradation and one analysis step, while bringing significant advancements to the field. In the next chapter, I will present a microfluidic enzyme reactor that generates and separates peptide ladders. Similar to Edman degradation, we plan to sequence by individual cleavage of N-terminal residues without disruption to the peptide bonds between other amino acid residues. Instead of using chemical digest, however, we plan to us aminopeptidases that liberate amino acids individually from the N-terminus. By digesting with proteins, we reason this system can take advantage of the difference in exopeptidase specificity for individual amino acids. In this way, spatiotemporal data from biocatalytic degradation of peptides via aminopeptidases can be obtained with iSAMDI and used to both decode sequences and resolve residues of identical mass based on differences in catalytic activity and may be useful in uncovering unknown post-translational modifications. By being conducted in a microfluidic channel and with analysis by MALDI, our technique will provide

ladders of peptides that are also spatiotemporally resolved and offer higher resolution over state-of-the-art techniques, only limited by the instrument's laser's ablation diameter. The platform will also serve to profile the activity of aminopeptidases. In the next two sections, I will first discuss both the importance that spatiotemporal resolution has brought to fields of research and the previous uses of iSAMDI that we will build upon, respectively.

4.2 Spatiotemporal Control in Microfluidic Devices

Developing spatiotemporal resolution in microfluidic platforms have allowed for unprecedented advances in fields of research. These dynamics are particularly useful when added to devices meant to study biological processes to reveal additional information on the origin, differentiation, function, and fate decision of cells[167]. For example, several groups have shown the benefit of adding spatial and temporal components when investigating gene expression profiles in single cell transcriptomics. The DBiT-seq platform, developed by Liu and coworkers, uses parallel channels against a tissue slide to deliver DNA barcodes that facilitates spatial transcriptomic analysis at the cellular level and reveals linkages between gene expression profiles and spatial distributions[168]. Qiu and coworkers developed scNT-seq, a droplet microfluidic-based method that encapsulates and pairs single-cells to barcoded beads[169]. The platform was used to jointly profile old and newly transcribed RNAs to recreate RNA regulatory dynamics in heterogeneous cell populations, monitor gene regulatory network activities in response to cell stimuli and in stem cell transition states. Other studies have used microfluidic systems to recapitulate physiological systems. For example, Bhattacharjee and coworkers developed a microfluidic gradient chamber generator array with 1024 chambers to investigate spatiotemporally complex axon guidance and

growth dynamics[170]. These examples demonstrate ways in which spatiotemporal control can reveal additional pieces of information from a microfluidic system.

Microfluidic devices have also been used to spatially separate enzymatic reactions to recreate enzyme cascades and synthesize products. For example, Ono and coworkers synthesized tetrasaccharide with a microfluidic chip by binding glycosyl transferases to agarose beads and separating them in series[171]. Maier and coworkers achieved two-step sequential biotransformations using printed esterases, hydrogenases and decarboxylases in agarose-based inks[172]. More recently, Obst and coworkers were able to achieve enzymatic synthesis of sialic acids in a microfluidic device, while avoiding cross-inhibitions and incompatible reaction steps[173]. Here, three enzymes necessary for synthesis, but the substrate for the third enzyme was also an inhibitor to the first two enzymes. To get around this, the enzymes were immobilized in hydrogels and compartmentalized. Continuous flow of different substrates was introduced at different points along the enzyme cascade so as to not inhibit the activity of enzymes upstream. Production of sialic acids was only possible by controlling the spatiotemporal components of the systems.

Other than small molecules creating negative feedback, enzymes themselves, depending on the order they act on a single substrate, can also exhibit cross talk and affect the type and/or quantity of product produced. To this end, a number of different devices have been developed to perform multistep enzyme cascades[174–177]. A common example of this phenomenon exists in the world of gene expression and regulation. Here, post translational modifications are made to the DNA-winding histone proteins in order to turn genes on or off. These modifications themselves, however, have the ability to affect the activity of other post translationally modifying proteins to modify neighboring amino acids' PTMs. In our group, we have used SAMDI to show that the activity of KDAC8 on the lysine at position 12 in the histone 3 protein (H3K12) can be affected

by PTMs made at nearby distal sites[18]. We've also shown that citrullination of H3R8 by the p300 enzyme can affect the activity of acetylation at the nearby H3K14[15]. These crosstalk discoveries have also been made external to our group as well[178]. More recently, a microfluidic reactor was built, based on the principles of SAMDI, capable of not only identifying cross talk between enzymes, but also controlling the order of activity by separating the enzymes and controlling the timing of substrate interaction. Thus, the addition of a spatiotemporal component to analysis provides significant benefit to understanding the system being studied. Because of the impact this new technology stands to have and its use in later chapters of this dissertation, this technology and all of the findings from this study and other studies that utilize this reactor is to be described in the next chapter.

4.3 iSAMDI Mass Spectrometry: History, Design and Uses

With previous work in profiling the activities of post translationally modifying enzymes, and with an understanding that PTMs themselves can also have an effect on the activities of these enzymes in the form of cross talk, our group, led by the efforts of Dr. Grant, set out to develop a system, using the principles of SAMDI mass spectrometry, that could control the order and timing of enzymatic reactions taking place on self-assembled monolayers and quantify the reaction products. It was then from this technology, that we added the ability to perform biotransformations in solution and track the completion of reactions using self-assembled monolayers over a span of the channel. The system was then re-invented to conduct chemical reactions solely in solution and capture reaction products throughout the entire span of the channel. These surfaces could then be analyzed by imaging mass spectrometry by creating an image of reaction progress using the imaging capabilities of our most modern mass spectrometers.

To design this system capable of studying reactions with SAMDI in a flow cell, a glass slide, rather than a steel plate, is coated with gold and engineered with an appropriate functionalized self-assembled monolayer. A polydimethylsiloxane (PDMS) channel is then clamped over the surface, and the floor of the microfluidic channel is further modified by overflowing species. This technology would become known as Imaging of SAMDI-MS, or iSAMDI-MS.

As mentioned in the previous chapter, the first account of our group combining self-assembled monolayers and microfluidic was to create an enzyme reactor that provides spatiotemporal control over reaction products[179]. In this study, the reactor had two enzymes organized on the floor of the microfluidic channel such that they could encounter a substrate sequentially. These enzymes, peptidyl arginine deiminase type 1 (PAD1) and acetyltransferase p300/CBP associated factor (PCAF) are able to catalyze citrullination and acetylation reactions, respectively, and have been shown in our group to exhibit a cross talk where their activity depends on the order in which they act (as explained in 4.3). Here we know that PCAF is unable to acetylate a peptide that has first been citrullinated by PAD1, whereas acetylation had no effect on citrullination. By adding spatiotemporal control over the enzymes and their peptide substrates, the different mixtures of these reaction products can be better understood.

To adapt the microfluidic SAMDI system to study these enzymes' activities, a SAMDI flow cell was created that presented maleimide as the functional group, and the enzymes were tethered to the surface of self-assembled monolayers via a protein immobilization method where fusion proteins attach to the surface via covalent binding to a ligand presented on the surface. In this case, by immobilizing benzyl guanine and alkyl chlorides to the surface, both SnapTag fused to a PCAF and HaloTag fused to a PAD1 are able to covalently bind to these molecules,

respectively, and present their fused enzyme to the inner chamber of the flow cell for reactions to take place. During the course of the reaction of overflowing substrate, the self-assembled monolayer is also able to immobilize the substrate, intermediates and products at every step of the cascade, avoiding the need to capture and analyze each reaction product separately.

A few important findings can be taken away from this study. Primarily, the new SAMDI-based methodology was able to produce expected cross talk results between two enzymes that were consistent with literature. Here, spatiotemporally organizing the enzymes in different orders produced two different ratios of reaction products. Next, the convenience of the method was demonstrated in that reaction products were immobilized directly to the chip in which the reactions were taking place, providing quick and easy analysis. Here the yields analyzed on the reactor surface act as a good approximation of yields that would be collected in an effluent. Decreasing the substrate flow rate was also shown to increase the yield of acetylated and citrullinated products, which as important experimental condition for future studies. This was consistent with other studies also reporting similar findings where an increase in product formed as a result of lower flow rates; this has been surmised to be a result of mass transfer effects[180,181]. Importantly, it was shown that multiple enzymatic reactions could be carried out in a single experiment using a SAMDI-based microfluidic reactor and all reaction products, including intermediates, able to be measured easily by SAMDI, before and after each reaction. It is this one-surface approach to reaction and detection that makes the method desirable to study a variety of other multi-enzyme reactions.

Building off of the development of this enzyme reactor, Grant and coworkers set out to use the design frameworks to create a SAMDI-based microfluidic device capable of mapping reaction

progress throughout longer sections of the channel. The device would also be capable of acquiring enough data to determine an enzyme's kinetics in one experiment. In this study, the flow cell and analysis were enhanced in the following ways: First, the channel was adjusted to a 3D format in order to handle mixing of more reagents in larger combinations. Second, microfluidics were used to create gradients of substrate, and the enzyme was applied in solution to these gradients. While many of the principles of the previous channel remained the same, the monolayer was allowed to immobilize reactions product after mixing of the enzyme and its substrate. Lastly, imaging mass spectrometry was used to acquire time course data in order to predict the enzyme's kinetic Michaelis constant. This technique would become the first example of what is known as iSAMDI-MS.

The design of this reactor consisted of two layers sitting atop of the gold slide containing self-assembled monolayers. In the bottom layer, an enzymes substrate and buffer are introduced and mixed using controlled diffusive mixing to create gradients. After the gradients are established, the top layer introduces the enzyme to each of the gradients. As the enzyme catalyzes the reaction, the reaction products are able to immobilize to the floor of the channel. This produces sections of channels that can be imaged by a mass spectrometer. Each position on the floor of the channel corresponds to a distinct reaction time for each concentration, and the channels of different concentrations provide enough information to calculate K_M.

The enzyme used in this system was human glutathione reductase which acts to reduce glutathione. Interestingly, because the substrate only exists as a disulfide and the product, containing a free thiol, only the product is able to be immobilized to the maleimide monolayer. Because of this, the authors chose to flow through a constant concentration of radiolabeled product for use as an internal standard showing that both a substrate and product need not be require to

provide a yield for the reaction. Rather, when calculating kinetic constants for the enzyme, it is more useful to readout the change in concentration of product over time which is inversely related to the change in substrate. The capture of this molecule creates a spatial map of the reaction progress that can be imaged by the mass spectrometer.

To acquire an image of the reactions, MALDI imaging MS or MALDI IMS was used. This technique has recently emerged as a powerful tool for a wide use of functions including being capable of determining spatial distributions of a variety of molecules[182–187]. The technique works by creating an array of 2D pixels, each containing its own spectrum. From this data, an image can be created for any mass value within the spectra, and species with that mass value can be analyzed by visual inspection. By combining IMS with SAMDI, high throughput data acquisition is also made possible, with the present study acquiring data every 200 microns along each channel, and with some instruments being capable of achieving down to 1 micron in width[188]. At the resolution the authors chose, nearly 2600 data points were acquired across eight channels of different substrate concentrations. From these data, eight concentration vs time plots were generated, and the collection of these plots were used to calculate the K_M for the enzyme.

A few important findings can be taken away from this study. Importantly, this study showed that iSAMDI could be used to monitor biochemical reactions and quantitate enzyme kinetics and may enable high throughput enzyme activity studies for a broad range of enzymes with shorter assay times over conventional approaches. Moreover, all of the reactions required could be completed on one plate and required a significantly small amount of reagent on orders of a few hundred microliters. The device also circumvents the need for large numbers of integrated valves on a chip. Finally, the imaging component gives the research the ability to visually inspect the reaction, which may be helpful in subsequent studies.

Building yet again off of these two studies, Grant and coworkers then tested the ability of iSAMDI to study chemical reaction kinetics by measuring the rate constant of a reaction[189]. In this study, a simpler channel was used. The channel contained two inputs and one output that was used to mix two reagents. As the chemical reaction takes place, self-assembled monolayers covalently react with the reactions products to imprint a permanent spatiotemporal record of the reaction mixture as it flows through the microfluidic channel. As in the previous study, iSAMDI could then be used to generate mass intensity maps of immobilized molecules. Each pixel generated corresponded to a unique reaction time. Using the spectral data contained in each pixel, the extent of the reaction at each position along the channel could be calculated. It was also shown that it is necessary to model the dispersion fluid front characteristics in order to quantitate second order rate constants for the reaction.

The model reaction used in this system was one first described in our group and added to this publication to report it. Specifically, phenyl glyoxal was added to one end of the channel and a cysteine-containing peptide to the other. When the aldehyde reacts with an N-terminal cysteine of a peptide, the two molecules reconfigure and join in a reversible manner. The joined molecule can also undergo an acyl transfer that further reconfigures the molecule to present the thiol as free again. It is this molecule and the original cysteine-containing substrate that are able to be immobilized to the monolayer to quantify the extent of the reaction. Here, the substrate and product present on the surface were used to calculate yields throughout the channel. However, to calculate the results, the fluid front characteristics were also important.

To accurately predict the extent of the reactions, a determination of the fluid velocity was first used to precisely calculate between position on the surface and time. This was achieved using

an injection of fluorescein. Fluorescent intensities were plotted over time and a linear regression was used to obtain the velocity. Importantly, it was seen that the velocity found experimentally was higher than what would be calculated theoretically using the cross-sectional area of the channel and flow rate. This was to be believed as a result of clamping of the channel which can squeeze the cross-sectional area and reduce the volume of the overall channel. Importantly, this also showed that convective flow had a greater effect on the reaction rate over diffusion. An important finding in this study showed that pH has an effect on the reaction, and that the flow cell is able to accurately show this. The reaction was carried out at four different pH levels to which the reaction proceeded faster at higher pHs.

In each of the studies above, SAMDI was used in combination with flow cells to organize and characterize three different reactions: a two-step enzyme cascade, a one-step biotransformation, and a chemical reaction. I the first study, enzymes were bound to the surface, substrates were passed over them to react and subsequently captured by the monolayer at 3 distinct time points. In the second study, enzymes and a gradient of substrate were mixed in solution, and the reaction's product subsequently captured at the end of the channel. In the third study, chemical reactants were mixed in solution, and, again, the reaction products subsequently captured on the channel, but throughout the entire channel. In all three cases, the extent of the reaction was able to be recorded and useful information pulled from the spatial distribution and time course information. The use of iSAMDI made possible a high resolution and high throughput method for making these calculations by acquiring high densities of information across the areas in the channel analyzed by the instrument. These approaches were enabled by the use of self-assembled monolayers that were able to capture molecules directly on the chip in which the experiments took place and the SAMDI technique to analyze the monolayers.

In the next chapter of this dissertation, I present a SAMDI-based enzyme reactor that is capable of monitoring the progress of a cascade of biotransformations. In this research, one substrate, a peptide, is immobilized to the surface, and the enzyme is able to act on the peptide as it is flown over the surface. The same enzyme acts repeatedly on the substrate and every intermediate produced, where the number of intermediates possible is equal to the number of amino acids in the starting peptide substrate. After each experiment, a permanent record of the substrate, all of the intermediates, and spatiotemporal map of where each intermediate exists in the flow cell remains. With this information, both the sequence of amino acids can be deduced and the rates of all of the reactions of their removal can be monitored. The system takes advantage of microfluidic devices and iSAMDI's ability to monitor reactions over time for surface biotransformations.

Chapter 5

Using Microfluidics and Imaging SAMDI Mass Spectrometry for Spatiotemporal Peptide Laddering and Sequencing

This work was done in collaboration with Juliet Roll, M.S. The research and all figures presented in this chapter are unpublished and will be submitted as Pluchinsky, A. J.; Roll, J.; Mrksich, M. Using Microfluidics and Imaging SAMDI Mass Spectrometry for Spatiotemporal Peptide Laddering and Sequencing.

Microfluidic platforms have been developed to provide advanced methodology for biochemical assays by reducing material consumption and enabling spatiotemporal control over reaction products; however, their application as peptide sequencing platforms is largely underexplored due to the incompatibility of reagents required of single residue sequencing. In this chapter, I will demonstrate a simple microfluidic device (Figure 5.1) where peptides are bound to a surface of self-assembled monolayers at the bottom of the flow cell and subjected to time-dependent degradation by overflowing exopeptidases. The resulting ladder of peptides, each differing by one residue, span the channel. The surface is analyzed by imaging self-assembled monolayers for matrix-assisted laser desorption/ionization mass spectrometry (iSAMDI-MS). With this approach, peptide sequences are read directly from the resulting spectra and residues that cannot be distinguished by a change in mass alone are resolved by the difference in specificity the exopeptidase has for those residues. The method is demonstrated on the isobaric canonical amino acids, leucine and isoleucine. This article expands the use of microfluidic systems to sequencing surface-bound peptides and introduces a new dimension to peptide sequencing.

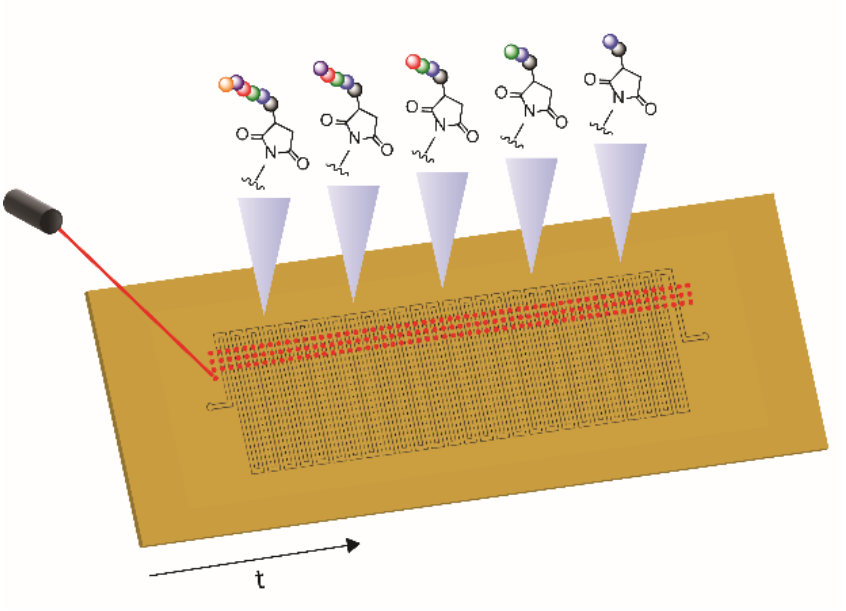


Figure 5.1. Cartoon of iSAMDI sequencing strategy. Peptide substrates, whose amino acids are represented as strings of colored balls, are immobilized to a gold-coated maleimide-functionalized surface. Ladders of peptides are generated throughout the surface and are scanned by MALDI mass spectrometry.

5.1 Introduction

A variety of methodologies exist for single residue peptide sequencing, which have found importance identifying novel proteins[190,191] and discovering important residue mutations and post translational modifications[146,192]. State-of-the-art methods of peptide sequencing at present include the decades old Edman degradation[128,129,193] and tandem mass spectrometry (MS2). Edman degradation is an automatable process that works by chemically removing residues from peptides bound to resin beads in cycles and analyzes the amino acids by chromatography, but the process is generally regarded as slow[194]. MS2 relies on pre-digests of protein, usually with an endopeptidase like trypsin. While faster, the technique necessitates solution-phase analysis and requires sophisticated instrumentation and analysis of complex fragmentation ions[127,195,196].

Moreover, modern MS2 techniques can struggle to provide full sequence coverage[197], and limited methods exist for resolving residues of identical mass[127,198].

In recent years, matrix-assisted laser/desorption ionization mass spectrometry (MALDI-MS) has become a popular tool for detecting biomolecules by their molecular mass and is a suitable technique for analyzing peptides from both chemical and enzymatic digestions[125,147,199–201]. By combining MALDI and Edman degradation chemistry, a higher throughput method for sequencing was developed that worked by analyzing the difference in mass between peptide backbones called peptide ladders, rather than individually cleaved residues[145,154]. Using chemical digestions, it was also shown that peptide ladders could be sequenced directly on solid surfaces for use in high throughput MALDI-based library screening[148,150,165,166]. Generating peptide ladders via enzymatic digestion has also been performed; however, digestion with exopeptidases is usually only performed at the C-terminus of peptides and is limited by the enzyme's specificity for different pairs of amino acids, the majority of which are incompletely understood[149,152–155,202–204]. Thus, a technique that can reveal the specificity profile of an exopeptidase to aid in sequencing at a peptide's N-terminus would be beneficial.

A variety of microfluidic and microfluidic-like devices have also been developed to interface with a MALDI mass spectrometer for proteomic analysis[160]. These approaches offer the advantages of being high throughput, easy to use and that they can separate proteins and peptides spatiotemporally; however, they either rely on the mass spectrometer to interpret fragmentation ions or identify proteins from whole peptides after tryptic digest, without sequencing individual amino acids[161–163,205]. In one study, a serpentine microfluidic channel atop a bed of resin was used to spatially separate proteins by both their mass and isoelectric point and a 2D map was generated by MALDI to visualize these properties; however, this study did not explore separating and

sequencing smaller protein components[158]. As such, the sequencing of peptides still relies heavily on low throughput chromatography and MS-based methods for separating and identifying amino acids.

In this article, we use exopeptidase digestion to generate surface bound peptide ladders in a microfluidic flow cell and use imaging self-assembled monolayers for matrix-assisted laser/desorption ionization mass spectrometry (iSAMDI-MS) to sequence peptides. Contrary to other sequencing methods facilitated by microfluidic columns that rely on chemical degradation and separation of products by chromatography[126,206], our device consists of a microfluidic channel designed to incorporate chemically defined self-assembled monolayers that can be directly analyzed by MALDI-MS. The monolayer is functionalized to covalently bind peptides, and exopeptidases are flown over the surface to generate peptide ladders on the solid support. Because it does not rely on Edman degradation chemistry, our device is able to make use of the most widely used material for microfluidic fabrication, poly(dimethylsiloxane) (PDMS), that is rarely used in sequencing applications[207]. In our group we recently developed similar microfluidic devices that are capable of providing spatiotemporal control over enzyme reaction products and showed that iSAMDI could be used to obtain high-density data sets to calculate relative concentrations of reaction products and generate a quantitative map of reaction progress[179,208]. In the examples presented here, we collected up and processed up to 34 000 spectra per experiment using no more than 200 µL of reagents and determined the sequence of peptides up to 6 residues long. Importantly, we show that a kinetic profile of each intermediate can be used to distinguish between isobaric residues in the sequence.

5.2 Results

Experimental Sequencing Strategy.

Our strategy for sequencing peptides first using a traditional SAMDI approach is shown in Figure 5.2A. A standard 384 gold-spotted SAMDI plate is engineered such that each spot contains a layer of functionalized self-assembled monolayers. The monolayers present a maleimide functional group which serve to immobilize cysteine (Cys, C) -terminated peptides via a Michael addition of the thiol group and the alkene of the maleimide. The tri(ethylene) glycol are present to prevent non-specific adsorption of enzymes. It has been shown previously that these monolayers are amenable to analysis by MALDI-MS as they have been used to analyze a broad range of enzymatic reactions on peptides conducted on the surface[14,19,21,114,209].

An exopeptidase with specificity for cleaving singular amino acids is then added to the surface at different concentrations or across different time intervals and allowed to act on the peptide substrate. As the exopeptidase and peptide react, amino acids are removed from the terminal end of the peptide, one-by-one. Positions where the enzyme has been spotted to the surface for a longer time intervals or where the enzyme was spotted in higher concentrations are expected to contain intermediates of peptides that contain fewer amino acids versus positions where the enzyme is spotted for shorter durations or in lower concentrations. Hence, this generates a ladder of peptide intermediates each differing by one amino acid residue. The reactions are quenched by rinsing the plate with water and ethanol and the monolayers are analyzed by mass spectrometry.

In the present example shown in Figure 5.2A , we use a surface presenting LAKAPSGC peptides. The amino acids leucine (Leu, L), lysine (Lys, K) and alanine (Ala, A) are represented in orange, red, and purple and green, respectively. Proline (Pro, P) is represented as blue. The C-

terminal Cys immobilizes to the surface via its free thiol leaving the N-terminal available to react. We then treated the surface with aminopeptidase I from streptomyces griseus at varying concentrations. In this example, the combination of exopeptidase and peptide were chosen because the specificity of the aminopeptidase I for Leu, Ala and Lys has been well reported, previously[210–213]. The reported activity of this enzyme differs across these residues over one hundred to nearly one thousand orders of magnitude and would thus provide a good sense of the range of concentrations of enzyme and time required for the enzyme to hydrolyze a variety of residues in situ with SAMDI.

A set of four spectra generated are shown in Figure 5.2B where the spectrum closest to the front was generated from a surface of peptide untreated by enzyme and the others generated from three selected concentrations of enzyme varying ten-fold from 1µg/mL to 100µg/mL with equal reaction times. The second spectrum from the front was generated from the experiment with the lowest concentration of enzyme and the fourth with the highest concentration of enzyme. In all cases, enzyme was added to the surface for 30 minutes.

The SAMDI-MS spectrum of the monolayer after the peptide was immobilized showed peaks at m/z 1618 corresponding to the maleimide-alkyldisulfide conjugate and peptide LAKAPSGC. Treatment of the monolayer with enzyme at 1µg/mL reveals a peak at m/z 1505 corresponding to the peptide intermediate AKAPSGC. Moreover, treatment of the monolayer with enzyme at 10µg/mL and 100µg/mL revealed peaks at m/z 1434 and 1306 corresponding to the KAPSGC and APSGC peptide intermediates, respectively. No additional peaks corresponding to any other peptide intermediates were observed rising downstream of the peak corresponding to APSGC.

Using the peaks observed across all spectra, the difference in m/z values between adjacent peaks can be used to identify the order of residues existing on the peptide. Here, a difference of 1618 and 1505 produces a mass difference of 113 Daltons, which corresponds to the monoisotopic mass of either the Leu or isoleucine (Ile, I) amino acid. Thus, when considering SAMDI data based on changes in mass alone, a distinction of either I or L cannot be made. Differences between 1505, 1434 and 1306 produces mass differences of 71 Daltons and 128 Daltons, corresponding to the monoisotopic masses of the Ala and Lys amino acids.

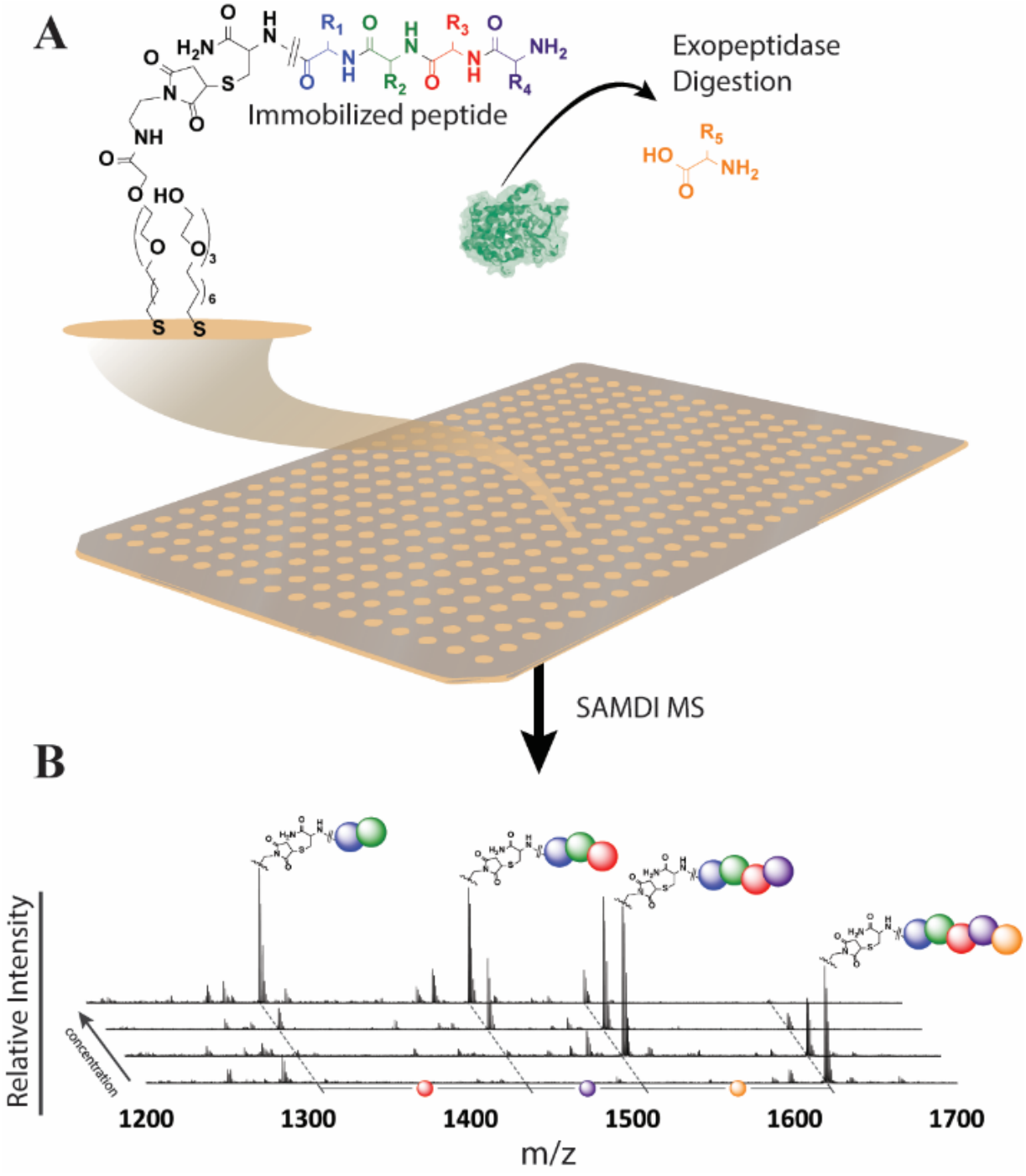


Figure 5.2. Strategy for sequencing peptides with SAMDI mass spectrometry. (i) A peptide is immobilized to a surface of self-assembled monolayers displaying maleimide via the free thiol of a N-terminal Cys residue. An exopeptidase is applied to the surface at specified time intervals or

at varying concentrations. (ii) Self-assembled monolayers are analyzed by MS and the results are displayed as a series of spectra that each capture a different intermediate peptide. The difference in m/z values between peaks corresponds to the mass of the removed residue between intermediates. Here, mass differences of 113, 71 and 128 Daltons are represented by orange, purple and red balls, and are identified as Leu/Ile, Ala, and Lys, respectively.

From this data, we surmised that we could design a microfluidic flow cell that makes use of this strategy to automate the ladder generating process and generate high resolution spatiotemporal data. Adopting such a device would allow us to make use of time, rather than differences in enzyme concentration, for generating peptide ladders. While high concentrations of enzyme are useful for quickly hydrolyzing amino acids, peptides of longer lengths would inevitably require longer reaction times. Whereas the concentration of enzyme can also hit a ceiling, the length of a microfluidic channel can easily be extended and, thus, the time over which it can be run. The device would also be one of few microfluidic devices that have ever been used for sequencing amino acids. Therefore, we decided to design a microfluidic device that uses the SAMDI strategy. Before designing a longer peptide for use in this system, however, we first opted to use SAMDI's profiling capabilities to uncover all of the activities of aminopeptidase I and identify an ideal peptide candidate.

Profiling Activities of Aminopeptidase I.

Having tested SAMDI's ability to monitor exopeptidase activities on surface-bound peptide substrates, we next describe an experiment to obtain the entire specificity profile of the aminopeptidase on a peptide array. We designed a 361-member array containing a library of X|ZPGSAC peptides, where X and Z are any of the canonical amino acids, excluding Cys. The array was designed with a Pro in the third position because the aminopeptidase I's inability to cleave residues adjacent to a Pro has been previously reported[210], thus, it is expected that only the terminal-most residue in each member of the library would be cleaved. In this way, we can determine which residues, either being cleaved or adjacent to the residue being cleaved, promotes or inhibits activity of the aminopeptidase I.

The array was designed using the 384 gold-spotted SAMDI plate shown in Figure 5.2. Similarly to the previous experiments, the spots were modified with a layer of self-assembled monolayers presenting maleimide functional group. The peptides were synthesized using standard protocols with FMOC-protected amino acids and transferred to the array plate of monolayers where each peptide was immobilized to the surface via a covalent bond between the peptide's C-terminal Cys and the maleimide functional group.

Three peptide arrays were created and treated with aminopeptidase I at a concentration of 1μg/mL for three separate durations of time, 1 hour, 3 days and 3 weeks. During these time intervals, the arrays were placed in a humidity chamber at 37 ℃. Arrays were then treated with THAP matrix and analyzed by MALDI to acquire spectra for each spot. The SAMDI-MS spectra revealed separate peaks corresponding to the substrate peptides and the product of the reactions. The conversion of the seven-member peptides to six-member peptides was characterized by

integration of the corresponding peaks and is given by Equation 1, where AUC refers to the area under the curve.

$$\text{Activity} = \text{AUC}_{\text{product}}/(\text{AUC}_{\text{substrate}} + \text{AUC}_{\text{product}}) \times 100\% \quad (1)$$

We note that the ionization efficiencies of the substrates and their products are not identical, and therefore the nominal conversions are not calibrated; however, the relative activities are useful for understanding the scope of substrates the aminopeptidase I is able to act on.

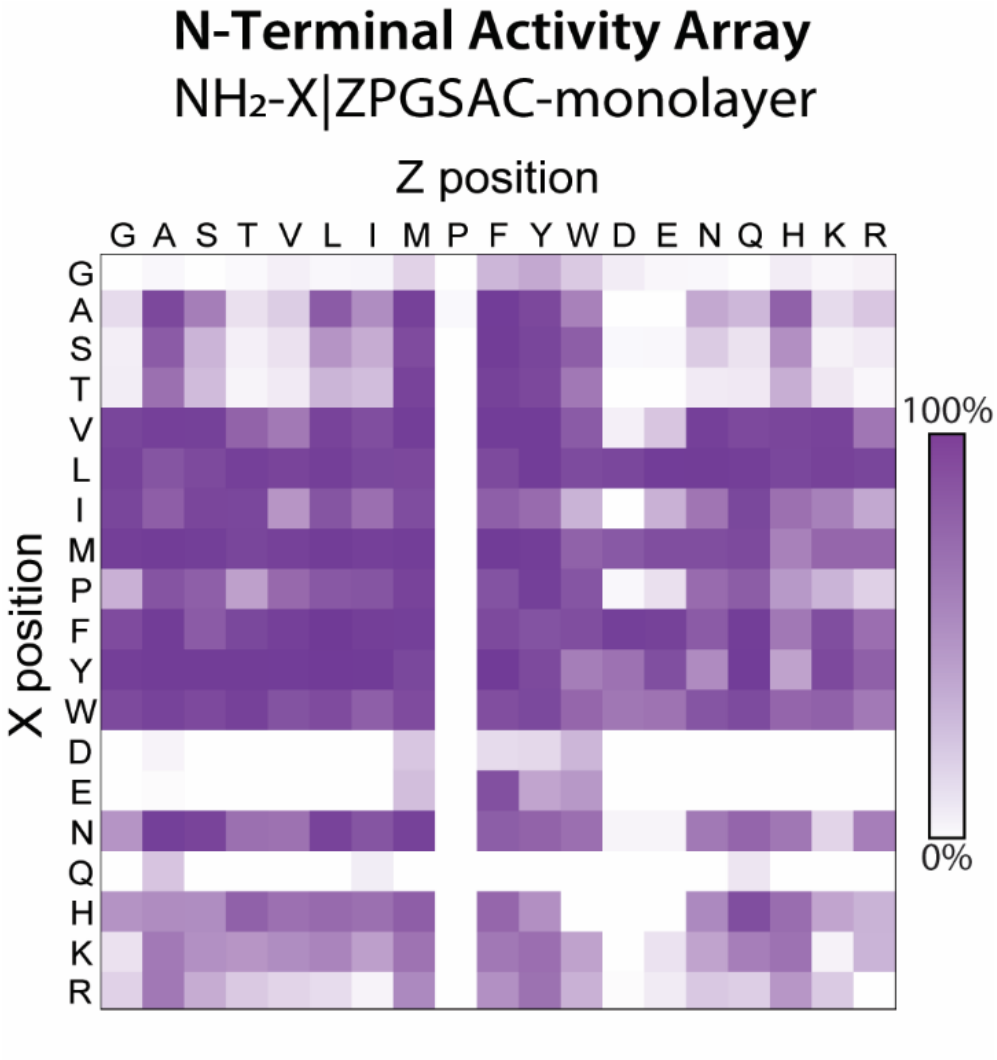


Figure 5.3. Profiling aminopeptidase I activity on a peptide array. Peptides of varying N-terminal residue pairs were immobilized onto SAMDI arrays. The arrays were treated with buffered aminopeptidase I for 3 weeks. The extent of cleavage of the N-terminal residue in the X position of each member in the array was measured using SAMDI-MS in which full cleavage in denoted by dark purple. Each square represents a peptide of sequence X|ZPGSAC and X and Z residues are denoted on the vertical and horizontal axes, respectively.

The activities for each peptide after 3 weeks are represented in a 19 × 19 heat map where every row defines the terminal amino acid to be cleaved in the X position and where every column defines the amino acid adjacent in the Z position (Figure 5.2). The percent cleavage is represented in a scale bar with purple corresponding to 100% activity and white corresponding to 0% activity. Results from treating the array for a shorter time intervals of 1 hour and 3 days can be found in Figure S# of the Supporting Information. Importantly, these data of relative activities reveal the sequencing capability of the aminopeptidase I and allowed us to strategically design peptides for our future experiments.

From these data, we see the entire specificity profile of the aminopeptidase I. Among the most consistent trend in our data is that the enzyme is highly specific for cleaving Leu regardless of the adjacent amino acid. This is consistent with literature, where the activity for Leu was determined to be 64.5 μmol . min-1 . mg enzyme-1, which was nearly an order of magnitude greater than the next most favorable residues, methionine (Met, M) and phenylalanine (Phe, F)[210]. From our array, we not only see that the enzyme readily exhibits hydrolysis of Met and Phe, but also that the enzyme struggles to cleave the negatively charged aspartic acid (Asp, D) and glutamic acid (Glu, E) for most adjacent residues. Interestingly, while Asp and Glu are not easily hydrolyzed, we reveal that these residues, as well as most of the others, are more efficiently cleaved when either Met or any aromatic residue is directly adjacent to the terminal residue, noted by the dark columns in the middle of the array.

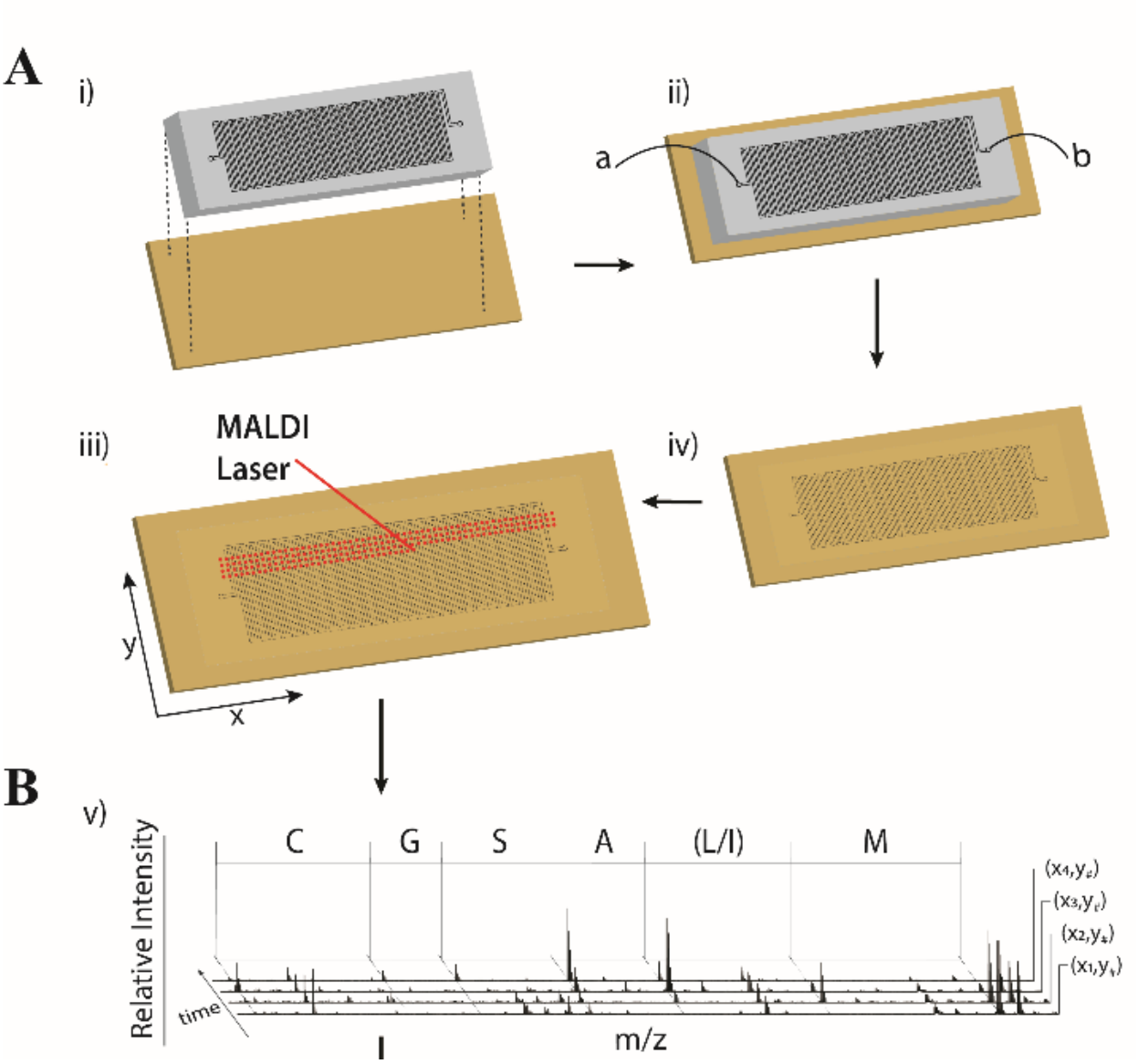
Previous studies of the specificity of this aminopeptidase found no hydrolysis of glycine (Gly, G) or Ala, when Pro was directly adjacent to these residues. Our profiling suggests that the enzyme exhibits no activity for any residue adjacent to Pro; however, at earlier time points (Figure S#, Supplemental Information), it is seen that the enzyme exhibits activity for Asp when adjacent

to Pro, both contrary to expectation. In previous work, high activity was observed for Phe, modest activity was seen for valine (Val, V), Pro and Lys, and minimal activity was reported for Gly and Ala. This is not only consistent with our findings, but we also show that the enzyme is readily able to hydrolyze the other aromatic residues, tyrosine (Tyr, Y) and tryptophan (Trp, W), regardless of the adjacent amino acid excluding Pro. We also newly find that the enzyme generally exhibits little hydrolysis for serine (Ser, S), threonine (Thr, T), and glutamine (Gln, Q), except when Ala is in the second position. Additionally, the enzyme exhibits modest activity for asparagine (Asn, N), histidine (His, H), arginine (Arg, R) and Ile. Interestingly, we also find that Leu, Met, Phe, Tyr, and Trp are the only residues are more readily hydrolyzed when either Glu or Asp are the second residue.

Microfluidic Device Design and Operation. Our device for sequencing peptides in higher resolution with a microfluidic flow cell is shown in Figure 5.4A. A poly(dimethyl siloxane) (PDMS) block is conveniently cast from 3D printed masters according to previous methods[179,189,208]. The channel was designed to be 500 µm wide, 500 µm tall and 886 mm in length. The PDMS mold is lightly clamped to a gold-coated glass slide where the floor of the flow cell consists of self-assembled monolayers functionalized with maleimide present at a density of 20% relative to the tri(ethylene glycol) background. The device has one input through which a peptide substrate, buffer, and enzyme enter and are allowed to travel through the length of the unidirectional channel. A detailed schematic of the workflow can be found in Figure S# of the Supporting Information. With peptide being the first sample to pass through the flow cell, the self-assembled monolayers capture and become saturated with the peptide by covalently reacting with the free thiol of a terminal cysteine. As an exopeptidase enzyme flows over the surface, peptides lose their terminal amino acids, one-by one. Peptides closest to the inlet are allowed to react with

the overflowing enzyme for longer durations and thus, more amino acids are removed from these peptides during the course of enzyme flow. Hence, the position along the channel corresponds to a different reaction time. The surface is back washed with water and ethanol right before the enzyme reaches the end of the channel in order to push the enzyme back out through the inlet and wash the channel such that peptides closest to the outlet do not receive enzyme treatment. In this way, the monolayer keeps a record of the original substrate and all intermediate lengths of peptides along the channel. The flow rate of enzyme through the channel can be adjusted to accommodate different lengths of peptide substrates. For longer peptides, the flow rate is reduced in order to increase the total reaction time, thus, allowing more amino acids to be cleaved. Removal of the PDMS top reveals the monolayer that contains a spatiotemporal record of all species present on the surface over varying distances. After application of matrix, the chip can then be analyzed by iSAMDI-MS.

In the present example, we synthesized MLASGC peptides to be presented on the surface. We designed this peptide candidate to be a good substrate for the enzyme based on data obtained from profiling the specificity of the enzyme, such that all residues could be cleaved during the experiment. Here, Met and Leu would be removed readily relative to an expected modest removal of Ala and slow removal of Ser and Gly to reveal Cys. Peptide was first allowed to flow through and allowed to immobilize to the surface. As previously described, Cys immobilizes to the surface allowing Met to be the first residue to react at the N-terminal. We then treated the surface with aminopeptidase I at a concentration of 6.3 μg/mL and at a flow rate such that the fluid front would require a total of 15 hours to reach the end of the channel. At the end of the experiment, the flow cell was disassembled, and the slide was washed with water and ethanol and dried. Matrix was applied to the slide using a matrix sprayer so that the slide could be analyzed by mass spectrometry.



(Figure 5.4 continued on next page)

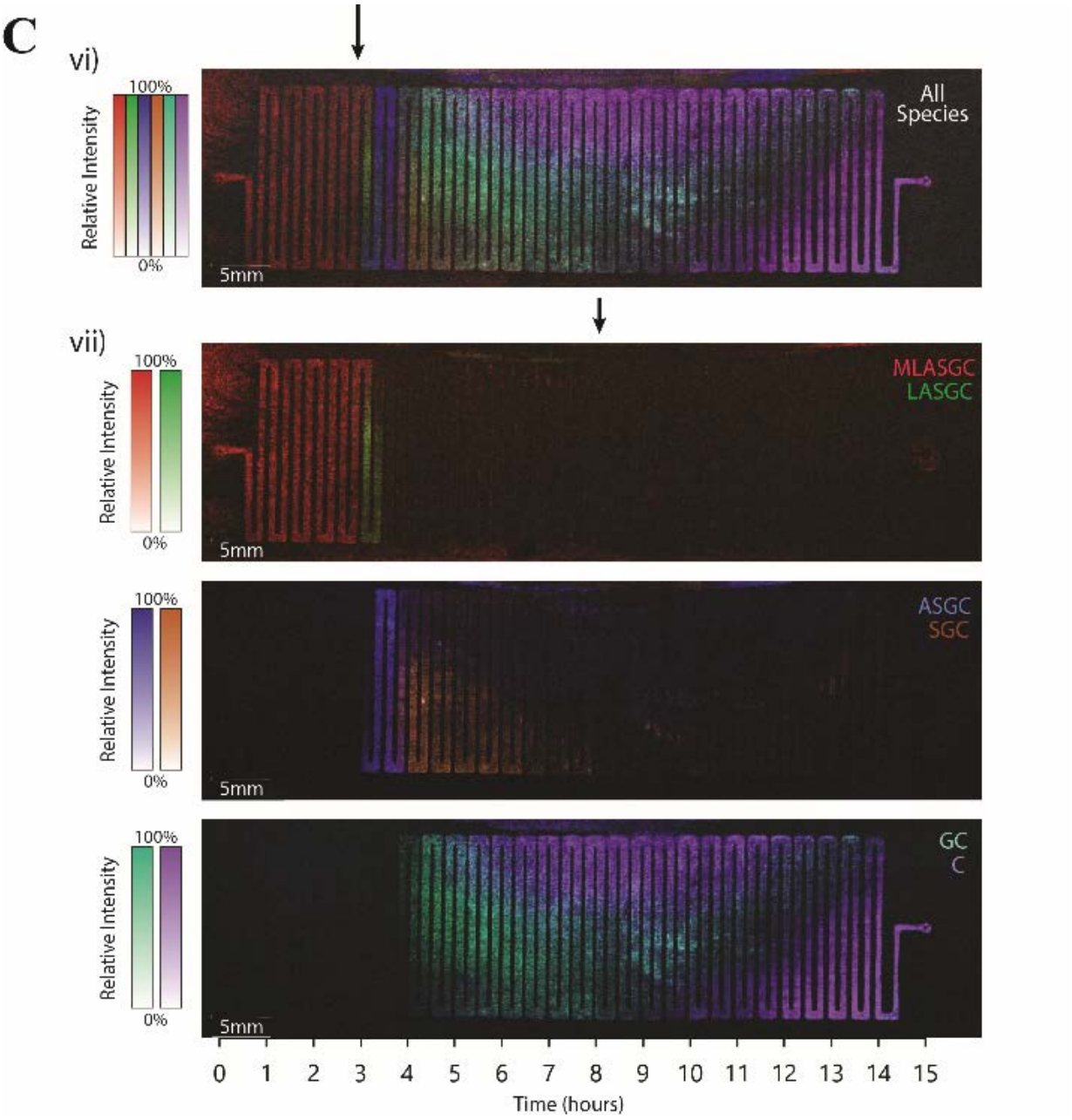


Figure 5.4. Design, operation and analysis of the microfluidic screening platform. (A) Design and operation of the microfluidic device. (i) A poly(dimethyl siloxane) (PDMS) structure containing 1mm2 channels is fixed to a gold-coated slide pre-functionalized with self-assembled monolayers to create a flow cell with a sinuous channel. (ii) A peptide substrate is flown through the flow cell from inlet ‘a’ to outlet ‘b”, during which the peptide undergoes immobilization to the monolayer. The substrate is followed by buffer and an exopeptidase. (iii) The PDMS top is

removed, and the slide is coated with matrix. (B) Data acquisition and visualization with iSAMDI. (iv) iSAMDI-MS is used over the entire region of the slide once below the channel. (v) The region scanned contains an array of pixels, where each pixel contains one spectrum. Spectra reveal different peaks of intermediates arising over different lengths along the channel. (vi) The starting substrate and each intermediate are displayed as a heatmap of various colors as a measure of relative intensity versus m/z, indicating their presence on the surface at different locations and for varying distances and peak intensities. (vii) To visualize different peptide intermediate species, heatmaps are separated into individual or groups of color channels.

Data Acquisition, Sequencing and Visualization.

Using imaging mass spectrometry via iSAMDI-MS, an array of pixels is collected where each pixel contains a unique SAMDI spectrum. Each spectrum reports the peptide species present on the surface at each respective pixel location along the channel as peaks of varying relative intensity. The collective of spectra reveals the m/z value for all intermediate peptides along the length of the channel. When considered together a ladder of intermediate peptides is revealed.

In this example, the sequence of amino acids was determined by both visual and computer inspection (see Supporting Information). In brief, the spectra throughout the channel are scanned to identify peaks that rise over time. All peaks are considered together, and the order of amino acids is determined. As previously shown, the distance between m/z values of these peaks can be tied to the mass of cleaved amino acids and the specific order in the full-length peptide substrate can be determined. In Figure 5.4B, we show four spectra selected from different points along the channel that collectively contain all of the peaks that were identified as peptide intermediates. The points were chosen at x values corresponding to the time range from 2 to 5 hours. The distances

between these peaks were calculated to be 131, 113, 71, 87, 57 and 103 Daltons, corresponding to the monoisotopic masses of Met, either Leu or Ile, Ala, Ser, Gly and Cys, respectively. This analysis, based on changes in mass alone, provides a sequence of M(L/I)ASGC.

With the iSAMDI-MS data, we generated mass intensity maps to visualize all species immobilized to the monolayer throughout the channel floor (Figure 5.4C). In these figures, time corresponds to the amount of time the enzyme was allowed to pass over a specific point along the surface. Therefore, the images are shown such that enzyme enters from the right (where the shortest peptides are expected) and flows to the left (where the longest peptides are expected). The colors correspond to different peptide intermediates on the surface and their brightness is representative of the relative intensity of the peak corresponding to that intermediate. Thus, the resulting data set serves as a spatial map of mass spectra containing different peptide intermediates and represents a kinetic profile for all of the reactions. Here, we immediately see that the time points we chose correspond to pixels with x values that are located at areas of overlapping colors that are consistent with the intermediate peptides found in the spectra peaks. From these data, we see that amino acids less readily hydrolyzed persist on the surface for longer durations along the channel, evident by the presence of peptides where such amino acids are in the terminal position. Comparing this visualization to data obtained from profiling the specificity of the aminopeptidase, we see that Met, either Leu or Ile, and Ala are rapidly hydrolyzed while Ser and Gly are less readily removed.

With the entire specificity profile of the enzyme revealed prior, we found that the aminopeptidase tended to exhibit less hydrolysis activity for Ile relative to Leu, regardless of the residue in the second position; the Leu-Ala and Ile-Ala pairs were no exception. For this reason, we surmised that this difference in activity could be extracted from the time course data and be

used to distinguish between these isobaric residues. To highlight this distinction, we selected both peptides LAPGSAC and IAPGSAC from our library to rigorously compare differences in the rate of hydrolysis the enzyme has for both Leu and Ile when adjacent to an Ala. Because we were only looking at the removal of one residue, we opted to manually generate time points by spotting protein to the surface at various time points over the course of 2 hours (Figure S# Supporting Information). From this experiment, we saw a clear preference for Leu over Ile, with less activity. For these experiments, substrates were present on a 20% surface of maleimide. While the surface density of maleimide and therefore peptide concentration can be varied, the concentration of the 2D surface is not straightforward to measure and remains significantly smaller than the concentration of enzyme. For this reason, quantifying the different enzyme kinetics for these substrates is not straightforward. Fortuitously, the iSAMDI images could potentially allow us to make the distinction based on simple visual pattern analysis alone.

To test iSAMDI's ability to sequence residues of identical mass, two peptides, MLASGC and MIASGC were synthesized and blindly immobilized to the bottom of two separate channels. Aminopeptidase I was flown through each of the channels simultaneously at a flow rate such that the fluid front would require a total of only 1 hour to reach the end of the channel. iSAMDI was then used to acquire pixels of data for the channel. Intensity values at m/z = 1321, which corresponds to both LASGC and IASGC intermediates, are shown for every pixel in the resulting heatmaps (Figure 5.5). By simple visual analysis, we see a difference in the persistence of these two peptide intermediates in the flow cell channel. In the top heatmap, the intermediate is shorter lived and has a less max intensity at its peak versus the bottom heatmap where the intermediate is longer lived and peaks at a higher intensity. Therefore, these patterns alone suggest that the enzyme

hydrolyzes the terminal amino acid more readily from the intermediate peptide in the top experiment, suggesting that the top heatmap originally contained peptide MLASGC.

We plotted the abundances of these two intermediates, which is a measure of the average pixel peak intensity of the intermediate over all of the species present on the surface for a specific time point along the channel. Each data set resulted in approximately 4000 individually addressable reactions that occurred in the microfluidic device. The curve in green, generated from data from the top panel, again shows that the intermediate is more quickly digested versus the species in gray. For this reason and with the knowledge that aminopeptidase I is more specific for hydrolyzing Leu, we were able to accurately label the green curve LASGC and the gray curve IASGC. Therefore, the original sequence that generated the top heatmap was MLASGC and the sequence that generated the bottom heat map was MIASGC.

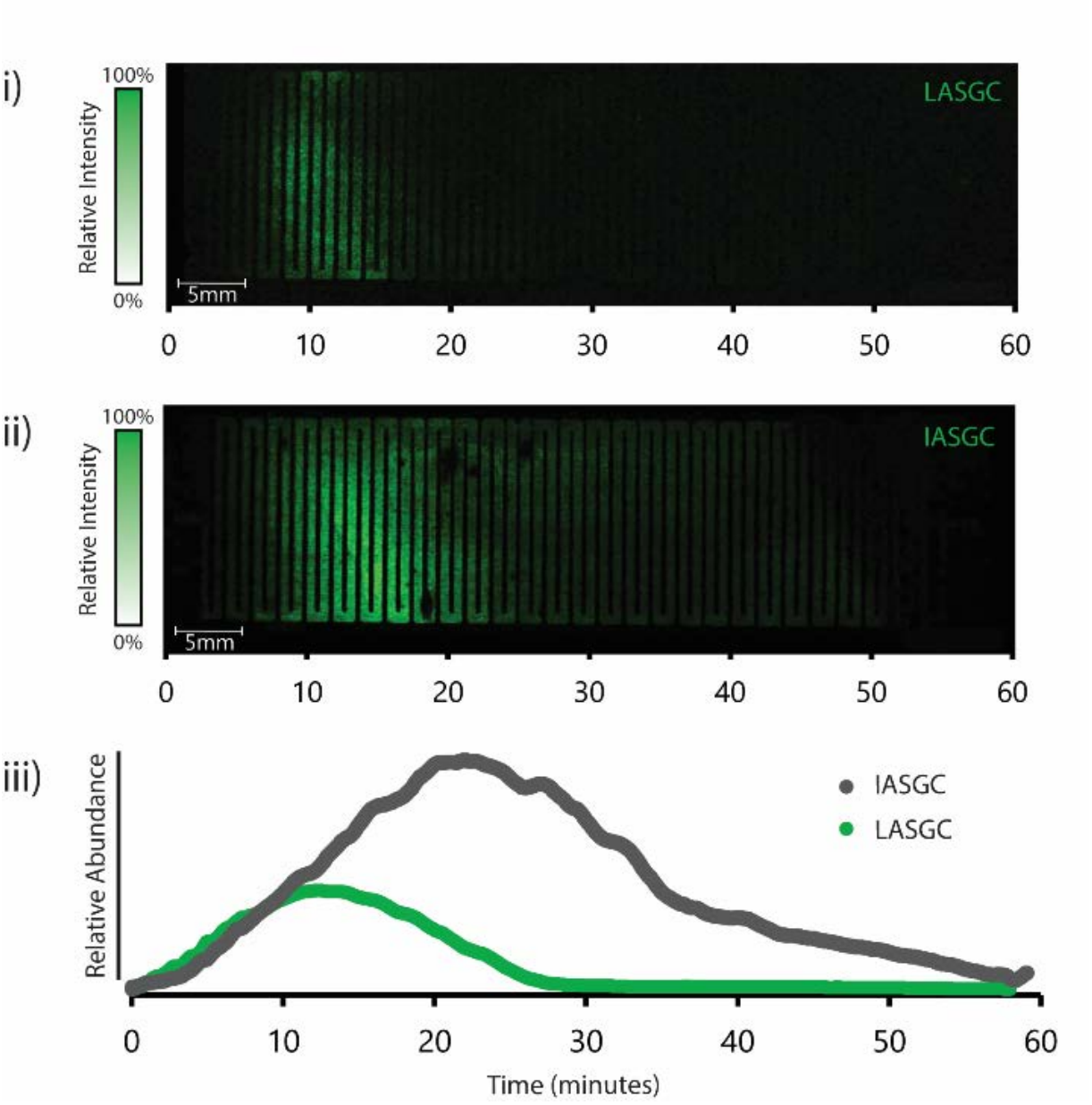


Figure 5.5. Time course of peptides intermediates. Heatmaps display the relative intensity for intermediate peptides LASGC (i) and IASGC (ii) after aminopeptidase I was flown over a surface of MLASGC and MIASGC peptides, respectively. (iii) The abundances of both LASGC and IASGC are displayed over time relative to all other species present.

5.3 Discussion

This article demonstrates how the SAMDI approach can be used with exopeptidases to sequence surface-bound peptides, and how iSAMDI can be integrated with a simple microfluidic device to acquire high resolution time points of peptide ladder intermediates. Importantly, iSAMDI is capable of providing these ladders at a temporal resolution with high enough density for identifying subtle differences in activity the enzyme has for each residue in a complex string or peptides, which has not been previously reported using other technologies. By performing the reactions in flow and on substrates immobilized to the floor of the flow cell, the temporal resolution is determined by the spatial resolution of the mass spectrometry and the flow rate. In our longest experiment, we imaged the surface at a pixel resolution of 100 μm, which corresponds to a temporal resolution of 6.1 seconds. Here, the instrument collected nearly 90,000 spectra, a third of which were used for analysis, in a matter of 2 hours. Such time points can be used to create an image representative of the kinetics of the enzyme for each residue. By resolving the relative kinetics of the enzyme for different residues, we are able to distinguish amino acids between peptide ladder intermediates and show that MALDI mass spectrometry is no longer reliant on changes in mass alone. Because our device only requires use of reagents in aqueous conditions, we were able to design our flow cell using PDMS, which is rarely used in proteomic analysis due to the surface properties of PDMS that leads to adsorption of proteins[160].

In this study, peptides were designed to contain pairs of amino acids within the substrate scope of the exopeptidase being used to digest the peptide. For this reason, sequencing is limited to coverage of a chosen peptidase; however, this may be rectified by creating a cocktail of peptidases whose collective activities hydrolyze all amino acids pairs[213–219]. The use of broadly specific carboxypeptidases in sequencing via the C-terminal of peptides has also been demonstrated and

may make good candidates for this method[220]. Recently, a 'Edmanase' enzyme was engineered by directed evolution to be broadly specific for us in high throughput peptide sequencing, and may negate the need for a cocktail of enzymes[221]. While it can be argued that chemical degradation methods are cheaper, abundant and more reliable than exopeptidase degradation, it is the difference in catalytic activity of the enzymes that substantiates the method for sequencing residues that cannot be seen by changes in mass alone.

The SAMDI platform has been successful in profiling the post translationally modifying activities for a variety of enzymes[14,19,21,114,209], with most recent work showing the strategy's usefulness for studying proteases by profiling the activity of an endopeptidase[16]. In this report, we extend the assay to profiling exopeptidases and show how an array of peptides can be used to uncover the specificity of exopeptidases for different amino acid combinations. Here, we chose to study the sequencing capability of an N-terminal peptidase, as these enzymes tend to be studied and used less for sequencing application compared to C-terminal peptidases. Profiling the specificity of aminopeptidases with SAMDI comes with the advantage that the method is label free and does not rely on tags that signal activity. This is in contrast to commonly based methods that use colorimetric tags which can affect an enzymes activity. For the enzyme we chose, its ability to cleave all canonical amino acid pairs was unknown, with studies only reporting the enzyme's ability to cut several residues from a 4-nitroanilide tag. [210,211] Our method was able to identify the activity of this protease for all native substrates of different pairs of amino acids and managed to identify unexpected pairs of amino acids that facilitate hydrolysis. While this array was useful for studying the specificity of aminopeptidase I for different pairs of amino acids, with Pro barring the enzyme's activity past the first residue and its activity independent of third residues, we note that other arrays would be required to survey the substrate scope of other exopeptidases.

We anticipate the platform will find use in uncovering the specificity profiles of other exopeptidases, including carboxypeptidases and ‘Edmanase’.

A corresponding limitation to this approach is that it requires peptides to contain a functional group at their terminus for immobilization. In this case, we took advantage of the chemistry native to cysteine by synthesizing our peptides to contain this residue at the C-terminal but recognize that peptides obtained via a protein digest would not have this advantage. Using other strategies for immobilizing peptides to solid supports, including ligation chemistries readily available and used by our group in the past[13,14,222–227], it may be possible to develop a generalized functionalization strategy for either C- or N-terminal residues. We also note that accurate quantitation of all peptide intermediates is not possible without prior calibration of the ionization efficiencies of every possible sequence combination as the removal of different residues can greatly affect ionization. Moreover, different sequences can favor different adduct peaks and tend to form proton, sodium and potassium adducts. The addition of more peaks can complicate accurate sequence determination but may be resolved by suppression of all but one adducts[228].

Another important caveat to our approach in that the resolving power of iSAMDI-MS is equivalent to that of MALDI-MS. The intensity of every peak is relative to the intensity of all peaks on the surface. Therefore, the ability to detect and quantify species is reduced as more species become present on the surface. This may be problematic peptides of longer lengths, especially for sequences where a majority of residues exhibit slow removal. We also anticipate that our assay could be complicated by the use of a cocktail of exopeptidases that may contribute to significant self-degradation and affect the time the experiments may be run. This may be avoided by engineering and expressing the enzymes to have tails that are resistant to hydrolysis.

5.4 Conclusion

In conclusion, we have developed a combined exopeptidase digest and iSAMDI-MS approach for sequencing peptides from peptide ladders generated in a microfluidic flow cell that is not reliant on changes in mass alone when distinguishing between residues. To our knowledge, this is the first example of integrating sequencing surface bound peptide sequencing, microfluidic flow cells and iSAMDI-MS, providing an alternative to methods using chemical digestion and chromatographic analysis. We expect that this high-resolution method will be useful for sequencing non-canonical, post translationally modified and other amino acids, especially modifications that are MS-labile that cannot be identified by changes in mass alone.

5.5 Methodology

Buffers. Peptide immobilization to surface was conducted in water (pH 6). All reactions and measurements were conducted in 20mM tricine, pH 8.0, with 0.05% bovine serum albumin (protein reconstitution buffer).

Self-Assembled Monolayer Preparation. Standard glass microscope slides were cleaned using ethanol and water in a sonication bath. An electron beam evaporator (Thermionics VE-100) was used to deposit 5 nm Ti (0.02 nm s−1) followed by 30 nm Au (0.05 nm s−1) at a pressure between 1 × 10−6 and 8 × 10−6 Torr. The slides were soaked overnight at 25 °C in an ethanolic solution (0.5 mM total disulfide concentration) having a 2:3 ratio of an asymmetric disulfide terminated with a maleimide group and tri(ethylene glycol) group to a symmetric disulfide terminated with tri(ethylene glycol) groups. The slides were then rinsed with ethanol and water.

Microfluidic Device Fabrication. Poly(dimethylsiloxane) (PDMS) masters were rendered on SolidWorks software. The master had one 800 μm inlet branching into a 500 μm width and 500 μm height channel. The files were converted to stl format and printed in a digital printing mode using a Stratasys Connex 350 3D printer in VeroWhite material (Stratasys Direct) with a glossy finish. The 3D printed masters were prepared for PDMS polymerization as previously described. PDMS prepolymer mixture was mixed in a 1:10 ratio (w/w curing agent to prepolymer), degassed in a vacuum desiccator for 15 min, and poured into the 3D printed master. The master containing PDMS was degassed in a vacuum desiccator for 45 min and placed in a 43 °C oven overnight. The PDMS blocks were then peeled off of the mold and treated in a 130 °C oven for 4.5 min. The 3D-printed molds were washed with water, dried and reused for additional PDMS curing cycles. A 0.8 mm biopsy punch was used to form the inlets of PDMS layer. The PDMS layer was then treated with 50 W air plasma for 35 s at 200 mTorr (Solarus Plasma Cleaner, Gatan, Inc.). The PDMS device was placed onto the Au slide functionalized with the self-assembled monolayer so that the bottom PDMS layer was in contact with the slide. The PDMS and Au slide assembly was held together using light pressure from an external clamp made from extruded polycarbonate secured with four screws at each corner. The clamp had two 1.5 mm diameter holes to match the location of the two device inlets. PTFE tubing (0.042″ outer diameter, ColeParmer) was inserted into the inlet and outlet via stainless steel catheter couplers (22 ga × 15 mm, Instech), and was primed with buffer using a syringe pump. Any remaining bubbles were removed by individually applying flow rates >500 μL min^{-1} to the syringe.

Microfluidic Device Operation. Peptide was introduced into the channel at a flow rate of 5 μL min^{-1} until peptide solution noticeably began to exit the outlet. Peptide was allowed to stay in the

channel for 50 minute to 1 hour. Buffer was introduced at the same flow rate to push out peptide solution. Aminopeptidase I enzyme was injected into the inlet at a concentration of 6.25 μg/mL for 15 hours at a flow rate of 8uL $hour^{-1}$ such that the enzyme would get close to but not reach the end of the channel. Immediately at the end of the enzymes flow time, the syringed was used to cuk up water through the outlet to reverse the flow of enzyme and clear the channel of enzyme in less than 3 seconds. The clamp was disassembled, the PDMS layers were peeled off of the chip, and the chip was rinsed with ethanol, water, and acetone. A solution of 2′,4′,6′-trihydroxyacetophenone monohydrate (THAP; ≥ 99.5%, Sigma-Aldrich) in 50% acetonitrile (25 mg/mL) was applied to the chip using a matrix sprayer. We found that the matrix crystallized more densely on the area of the monolayer in contact with the microfluidic channels, which allowed us to easily locate the entire sinuous channel using the video camera on the MALDI TOF instrument (RapifleX, Bruker Daltonics). iSAMDI-MS spectra were acquired in reflector positive mode using a mass range of 600−2000 m/z over a selected region of interest (ROI) that spanned the entire reach of the channel. The laser was operated at 200 Hz with 125 laser pulses applied per spot using the “medium” aperture setting.

Data and Image Analysis. The [M + Na+] + adducts of all peptide intermediates were determined both visually and computationally (MATLAB R2019a). Spectra were opened using FlexImaging software (Bruker Daltonics). Ion intensity maps were generated on FlexImaging software at 1452 ± 0.5, 1321 ± 0.5, 1208 ± 0.5, 1137 ± 0.5, 1055 ± 0.5, 993 ± 0.5, 873 ± 0.5 and 693.5 ± 0.5 Da, corresponding to the peptide intermediates MLASGC, LASGC, ASGC, SGC, GC, C - alkanedisulfide conjugates, and maleimide-alkanedisulfide conjugate and EG3, respectively. ROIs were selected to highlight all areas of the channel. Areas under the curve for all species’ color

channels were exported for all spectra in the ROIs. The abundances of each species were calculated computationally (MATLAB R2019a, see SI) for all spectra by taking the area under the curve of a species and dividing it by the total area under the curve for all species. Abundances were averaged across the cross section of the channel.

Chapter 6

Additional Surface Engineering and Modifications

6.1 Development of a Strained Alkyne Self-Assembled Monolayer

In this chapter, I report several methods I've explored for developing a surface of self-assembled monolayers presenting a strained alkyne. The surface is useful for immobilizing azide-containing molecules in a conjugation reaction known as copper-free click chemistry. I also discuss several benefits and drawbacks of each method based on empirical findings.

"Click" chemistry, in its most general definition, comprises a class of small molecule reactions that are typically employed to join two biomolecules, such as a molecule of interest to a probe, although the chemistry is useful in various applications. While not referring to a singular chemistry, the copper(I)-catalyzed azide-alkyne cycloaddition (CuAAC) has gained significant popularity over other ligation reactions and is more often used synonymously with the "click reaction." While considered biorthogonal in that the reaction's components are highly selective, chemically/biologically inert, exhibit fast kinetics, work in biological conditions (i.e. aqueous) and capable of being incorporated into biomolecules, the CuAAC click reaction is not often used in live-cell studies due to the toxicity of the Cu(I) ion. For this reason, a number of copper-free click reactions have been developed[229–236].

Outside of probing and joining biomolecules, click reactions are also very useful for selectively immobilizing samples to a surface of self-assembled monolayers for analysis by SAMDI-MS. Importantly, the high orthogonality of the reactions makes the chemistries particularly useful in complex solutions as it allows the surface to selectively extract samples even from environments like cell lysates. Or group has demonstrated that several of these

immobilization chemistries are useful in this way, with many studies, including my own making, use of the maleimide-thiol Michael addition, and in early work, the Diels-Alder reaction, to name a few (see Figure 1.1). Recently in 2016, our group first reported the use of the CuAAC reaction to immobilize molecules to a surface of self-assembled monolayers[25]. In this study, the maleimide surface was modified in only 10 minutes using 400 μM of propyne thiol in ethanol. Thus, the surface was modified to present an alkyne that would be able to immobilize azide-containing molecules. To capture the azide-tagged molecules, the authors required a common cocktail of reagents including 10mM $CuSO_4$ (for the Cu(I) ions), 40mM ascorbate (to reduce Cu(II) to Cu(I)), and 20mM THPTA (to stabilize Cu(I) and reduce side reactions) for 1 hour. While capable of adequately immobilizing samples to a surface to quantify reaction products, the reagents required for the CuAAC reaction are damaging to the surface of self-assembled monolayers, and are not forgiving if left on the surface for more than an hour. From my own experience, these reagents can also react with and degrade the product of a reaction, making it indistinguishable from its substrate in the mass spectrometer. Thus, an alternative approach for immobilize azide-containing molecules to the surface would be helpful for when the reagents and reaction products are not compatible and/or longer immobilization times are required but not possible when using CuAAC. In the following example, I present multiple strategies for modifying the surface to present a strained alkyne, specifically a dibenzylcyclooctyne also known as DBCO (pronounced dib'- co.).

In my first example of surface modification, I make use of the maleimide's reactivity with amines at higher pHs. Because of the reactivity of the strained alkyne, these molecules are typically not used as molecules co-functionalized with free thiols. For this reason, I opted to use molecule that contains a DBCO at one end and a free amine at the other end (Figure 6.1). In this approach, 10mM of DBCO-amine is added to a maleimide-functionalized self-assembled monolayer in

phosphate buffer at a pH of about 8.5. The 10% DMSO is present as it was used to dissolve the pure DBCO-amine solids. At this concentration, the surface is functionalized after only 15 minutes to produce a "DBCO monolayer". From there, a sample tagged-with an azide can be conjugated to the surface.

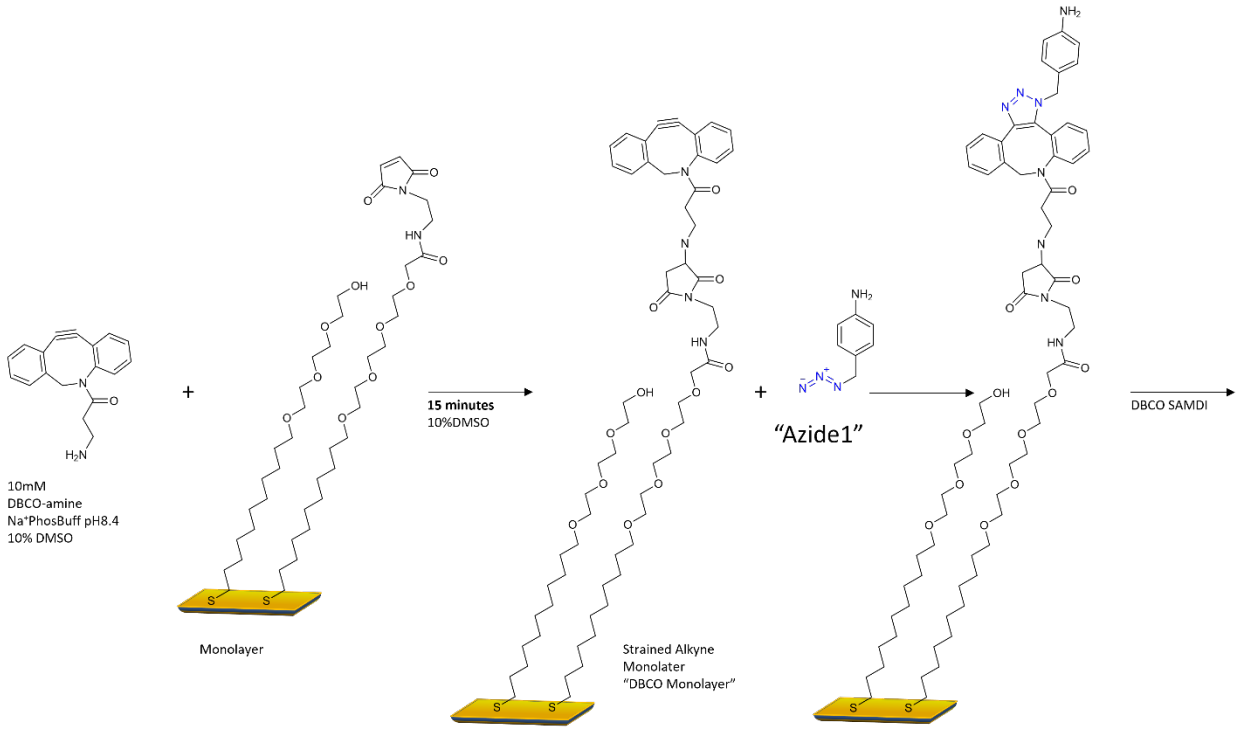


Figure 6.1. Formation of a DBCO monolayer with DBCO-amine. Commercially available DBCO-amine is dissolved in DMSO and diluted 1:10 in phosphate buffer pH 8.5 to 10mM. The solution is immediately put on the surface and rinsed off after 15 minutes.

In these experiments, immobilizing the DBCO-amine was attempted at a pH of both 7 and 11. The former require significantly longer immobilization times (~3 hours for full saturation of the maleimide) over which time the DBCO monolayer degraded into several peaks between *m/z* of 829 and 962. At a higher pH, the reaction completed more quickly even at lower concentrations of DBCO-amine, however, this monolayer was also seen to degrade much more quickly at higher

a higher pH. For this reason, a high concentration of DBCO-amine at 10mM and 15 minutes on the surface proved to be the most reliable and stable; however, it is noted that this methodology as the drawback that it is cost prohibitive, especially if needed for large screening applications. It should also be noted that this surface, either containing a free DBCO or one that has already reacted with a azide molecule of interest, is not stable in air. Simply keeping the monolayers under ethanol allows them to be stable for several hours and also helps to produce a better signal in the mass spectrometer.

When analyzing the surface by MALDI, the DBCO monolayer is present at a *m/z* value of 1149. In Figure 6.2, I show spectra that contains both the DBCO monolayer and a simple aminobenzyl with an azido group para to the amine.

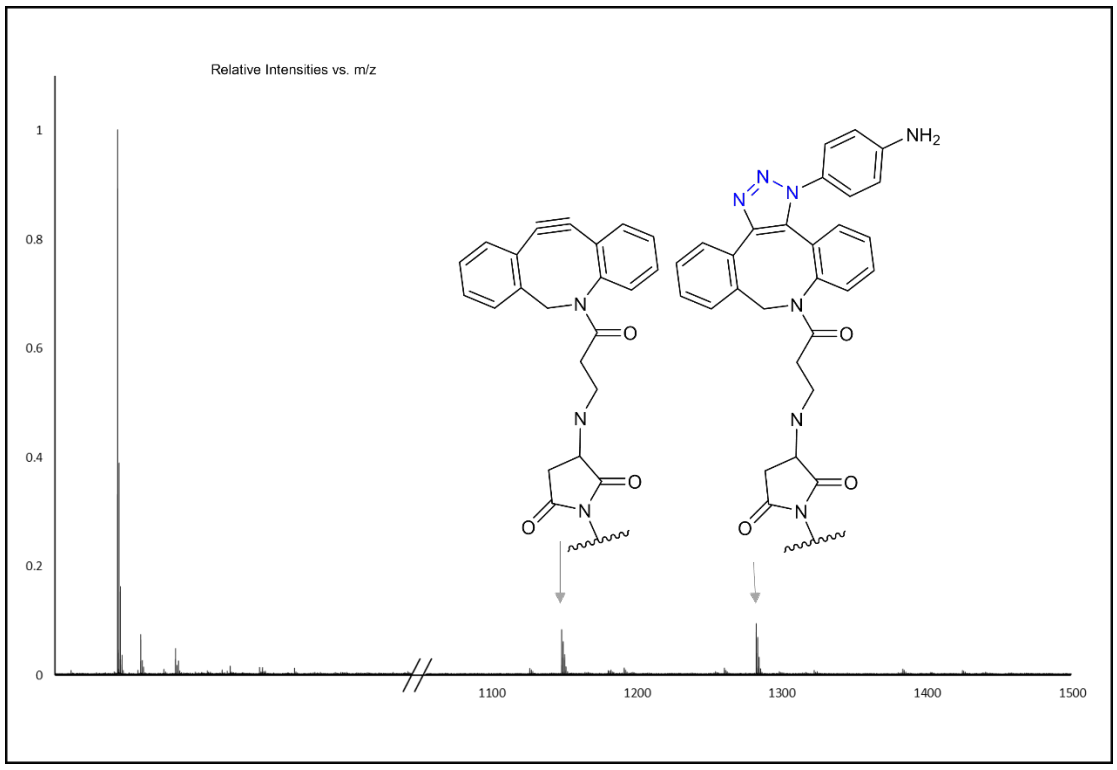


Figure 6.2. SAMDI-MS readout for both the unreacted DBCO-monolayer and a captured azide-tagged molecule.

Copper-Free Click Immobilization Strategy. Continuing with click strategies, we found that a strained alkyne can be used to drive the click reaction without the need for a catalyst. Taking advantage of the maleimide's reactivity with amines at a higher pH, I successfully functionalized the maleimide surface with the strained alkyne, dibenzyl-cyclooctyne amine (DBCO-amine) and tested its ability to immobilize a simple azide-containing molecule, p-azido aniline (Figure 3). Initial results revealed that this immobilization strategy is highly sensitive and capable of detecting the specified analyte at a concentration of 1µM allotting only 30 minutes for surface functionalization with the strained alkyne and 1 hour for immobilization.

In a second strategy for modification, I first modify the maleimide surface to present an amine, and then subsequently react the amine with a N-hydroxysuccinimide (NHS) ester group to add the strained alkyne. To modify the surface first to present an amine, a 100mM stock of cysteamine was prepared in phosphate buffer at a pH of 7. Spotting cysteamine at this concentration is only require for 20 minutes to saturate the maleimide. To further functionalize the surface, a 5mM stock of DBCO-sulfo-NHS-ester is prepared in phosphate buffer with a pH of about 8.5. At this concentration, the DBCO molecule requires an immobilization time of 1 hour. This methodology is slightly less preferable than the first as longer immobilization times provide more opportunity for the DBCO to degrade, and thus the signal is not as strong in this case. The strategy using the DBCO-sulfo-NHS-ester is shown in Figure 6.3. Using this approach produces a cysteamine *m/z* peak around 950 and an *m/z* value of 1265 for the DBCO surface.

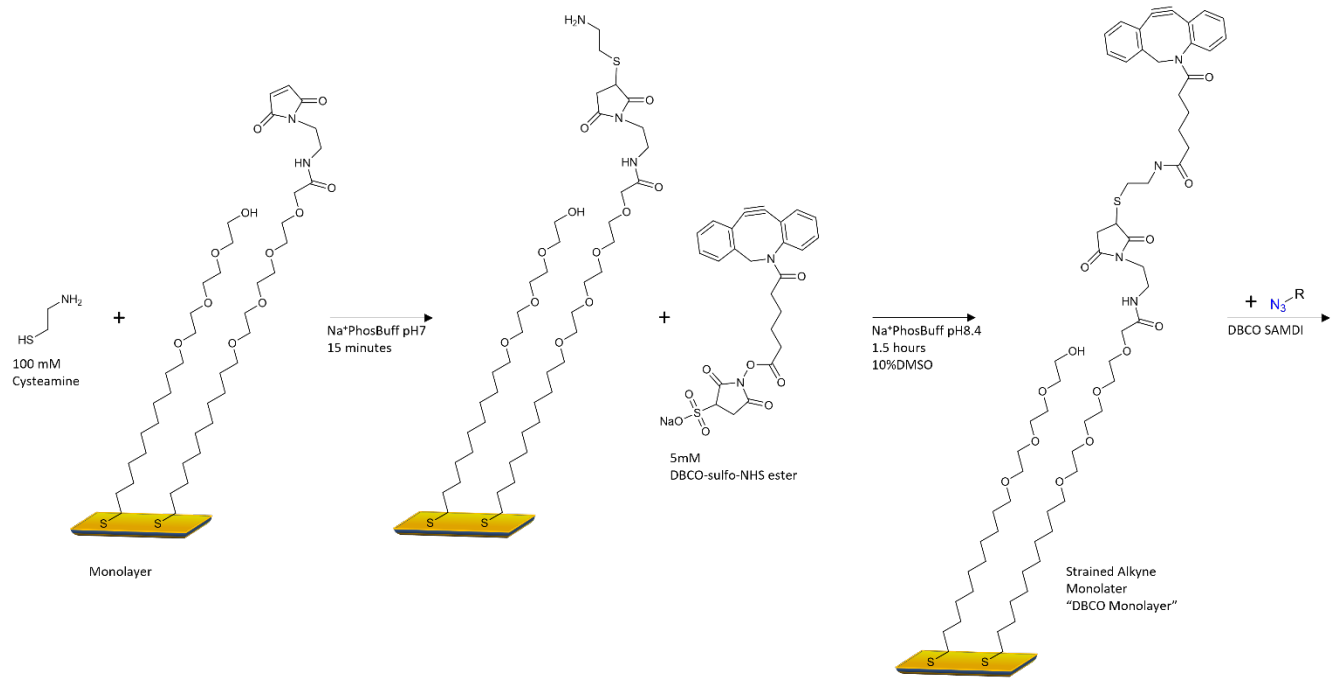


Figure 6.3. Formation of a DBCO monolayer with cysteamine and DBCO-NHS-sulfo-ester. Commercially available cysteamine is dissolved and spotted to a maleimide-presenting surface for 15-20 minutes. A DBCO-sulfo-NHS-ester is dissolved in DMSO and diluted 1:10 in phosphate buffer pH 8.5 to 5mM. The solution is immediately put on the surface and rinsed off after 1 to 1.5 hour.

In these experiments, the DBCO-sulfo-NHS-ester was also incubated with cysteamine first, prior to immobilization; however, while this method produces the same results, it is far less preferable as the DBCO-thiol stock is not stable, and the pH must be corrected from 7 to 8.5, prior to spotting to the surface. In both cases, longer immobilization times provides more time for the DBCO to degrade. Shorter immobilization times with higher concentrations of DBCO molecules also showed increased degradation of the molecules. In Figure 6.4, I show results from both of these approaches.

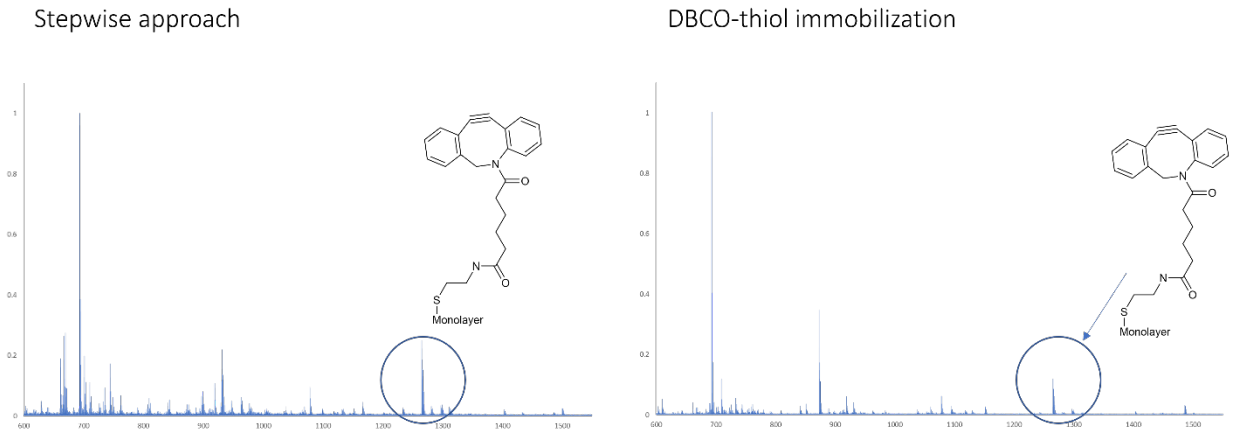


Figure 6.4. SAMDI-MS readout of an unreacted cysteamine-to-DBCO monolayer. The left spectrum shows results for the case that the maleimide monolayer is functionalized with cysteamine first and then DBCO. The right spectrum shows the case that the DBCO-sulfo-NHS-ester are joined in solution first, and then allowed to react with the maleimide monolayer. In both cases, when analyzing the surface by MALDI, the DBCO monolayer is present at a *m/z* value of 1265.

Overall, the copper-free click immobilization strategies are able to successfully detect molecules using SAMDI; however, their use in high-throughput experiments may be cost prohibitive, as high concentrations are needed for fast immobilization times. Here I showed three strategies for functionalizing the surface with DBCO strained alkynes, that are useful for immobilizing azide-containing molecules in aqueous conditions. I also showed that it is quite easy to re-functionalize the surface with an amine to ligate molecules via the NHS-sulfo-ester amine ligation reaction, which may be helpful for functionalizing the surface with other chemistries. In any case, the side reactions of these molecules must be considered. Due to the instability of the DBCO monolayer, it is still preferable to employ a copper-click surface for immobilizing analytes, when appropriate. If possible, performing the click reaction in-solution first, sequestering the click

reagents, and then pulling down the molecules may be a better option for immobilizing molecules while preserving the surface.

Chapter 7

Summary, Final Thoughts and Future Directions

For decades, the SAMDI-MS has been substantiated by its ability to characterize self-assembled monolayers, which may be engineered to capture analytes and enable analysis of chemical and biological reactions, by MALDI mass spectrometry. By positioning the self-assembled monolayers of alkane thiolates on a gold surface, resulting spectra are straightforward to interpret when only the semi-covalent bond is released in the mass spectrometer. Importantly, the surface of the monolayer can be modified to selectively capture samples from solution or immobilize samples for *in situ* reactions on the surface. These features, among many of the others discussed in chapter 1, has allowed the SAMDI-MS strategy to be used to develop high-throughput, generalizable and creative solutions to solve challenges in a myriad of research fields. Building off of SAMDI's ability to rapidly screen enzymatic reactions from cell lysates and iSAMDI's ability to spatiotemporally separate and visualize reactions, this dissertation introduces new assays that advance both the fields of directed evolution and sequencing.

In chapter 2, I create a powerful strategy for screening combinatorial libraries of enzyme variants in high throughput for the directed evolution of enzyme activities. The SAMDI-based assay is used to evolve a catalyst for an important and highly-challenging biotransformation, both in the enzyme's ability to catalyze the reaction and in the detection of the reaction products. Screening in high throughput remains a significant challenge in the field of directed evolution, especially for libraries of enzymes catalyzing non-natural reactions. Large-scale directed evolution campaigns can be infeasible due to the absence of a high-throughput screen or the extended time required to develop one. As a result, researchers can be forced to significantly reduce the sequence space they explore or discard the campaign altogether. By using self-assembled monolayers to

immobilize analytes of interest prior to mass spectrometric analysis, our immobilization strategy directly addresses this issue and has several advantages over current state of the art technologies. Selective immobilization of analytes allows for their direct analysis from very low volumes of cellular milieu without any purification. Additionally, our method is amenable to a variety of covalent capture strategies that make it generalizable outside of the specific reaction reported. Most importantly, because it is based on MALDI, it is inherently high-throughput. Thus, the method effectively stands to ameliorate the bottleneck in throughput for reactions that can be analyzed by MALDI and require a high-throughput screen. In this work, we reduced the time required for analysis by two orders of magnitude over prior techniques. As we can screen the activity of thousands of variants in a matter of hours relatively easily, we expect this technique to be applicable not only where high-throughput screens are otherwise impossible, but also to be broadly useful to members of the chemistry and biology communities that utilize directed evolution for small molecule synthesis where chromatographic screening methods have remained relatively ubiquitous. We anticipate that the throughput of this technique could be used to engineer the specificity of an enzyme for different substrates, evolve multiple enzyme reactions and reaction sites in unison, and may help in providing an understanding of why proteins evolved the way they did, but linking directed evolution to adaptive evolution. The technique may also find use in solving the next bottleneck in directed evolution, finding initial activity, as the SAMDI method is capable of discovering protein functions without needing to know what proteins do in advance.

In chapter 4 I also present a new technique for spatially separating multiple sequential reactions between an enzyme and its substrate. The technique combines self-assembled monolayers to provide a solid support for peptides, exopeptidase digestion, microfluidic liquid control, MALDI analysis and imaging mass spectrometry via iSAMDI to sequence peptides. Use

of exopeptidase digestion over chemical methods allow for sequencing to take place under aqueous conditions that are compatible with the most common materials in microfluidic fabrication. The microfluidic device automates the device such that all reactions can take place in one continuous flow of experiments and its combination with self-assembled monolayers negates the need to manually quench and take sample fractions of the reaction over time. Rather, a record of all of the reactions over time are permanently recorded on a surface for analysis by mass spectrometry. When analyzing the surface reactions with iSAMDI, the resolution is only determined by the resolution of the mass spectrometer, and thus thousands of reaction time points are able to be generated across the course of the entire experiment. This provides a resolution necessary to distinguish residues based on the differences in catalytic activity enzymes have for them in addition to just their masses alone. This feature allows us to distinguish between isobaric residues such as the most common, leucine and isoleucine. In this study, we also expanded the use of SAMDI for profiling post translational modifying activity on peptides, by revealing the specificity profile of an exopeptidase. These experiments were useful for selecting a suitable peptide candidate to develop our sequencing system, but we anticipate the technique will be used to uncover profiles for a variety of exopeptidases to increase our understanding of their physiological rolls.

With these contributions to different fields, I anticipate that research will be conducted at their intersection. Using the directed evolution screening methodology described in chapter 2, the ability of SAMDI to measure and profile the activities of enzymes on libraries of peptides containing hundreds of members as shown for a protease in chapter 4, it is now possible to engineer a protease to tailor its specificity for specific pairs of residues. This concept would be beneficial

for the evolution of non-specific exopeptidase for use in chapter 4 by expanding the platforms sequencing capability with a single enzyme.

Appendix

Nucleotide and amino acid sequences

All heme proteins disclosed below were cloned into a pET22b(+) vector.

DNA and amino acid sequence of P411-CHF(P74T), a previously reported cytochrome P411 variant which was used as the starting point for evolution:

DNA Sequence:

ATGACAATTAAAGAAATGCCTCAGCCAAAAACGTTTGGAGAGCTTAAAAATTTACC
GTTATTAAACACAGATAAACCGGTTCAAGCTTTGATGAAAATTGCGGATGAATTAGG
AGAAATCTTTAAATTCGAGGCGCCTGGTCGTGTAACGCGCTACTTATCAAGTCAGCG
TCTAATTAAAGAAGCATGCGATGAATCACGCTTTGATAAAGAGTTAAGTCAAACGCT
GAAATTTCTGCGTGATTTTCTTGGAGACGGGTTAGCCACAAGCTGGACGCATGAAAA
AAATTGGAAAAAAGCGCATAATATCTTACTTCCAAGCTTTAGTCAGCAGGCAATGA
AAGGCTATCATGCGATGATGGTCGATATCGCCGTGCAGCTTGTTCAAAAGTGGGAGC
GTCTAAATGCAGATGAGCATATTGAAGTATCGGAAGACATGACACGTTTAACGCTTG
ATACAATTGGTCTTTGCGGCTTTAACTATCGCTTTAACAGCTTTTACCGAGATCAGCC
TCATCCATTTATTATAAGTCTGGTCCGTGCACTGGATGAAGTAATGAACAAGCTGCA
GCGAGCAAATCCAGACGACCCAGCTTATGATGAAAACAAGCGCCAGTTTCAAGAAG
ATATCAAGGTGATGAACGACCTAGTAGATAAAATTATTGCAGATCGCAAAGCAAGG
GGTGAACAAAGCGATGATTTATTAACGCAGATGCTAAACGGAAAAGATCCAGAAAC
GGGTGAGCCGCTTGATGACGGGAACATTCGCTATCAAATTATTACATTCTTATATGC
GGGAGTTGAAGGTACAAGTGGTCTTTTATCATTTGCGCTGTATTTCTTAGTGAAAAA
TCCACATGTATTACAAAAAGTAGCAGAAGAAGCAGCACGAGTTCTAGTAGATCCTG
TTCCAAGCTACAAACAAGTCAAACAGCTTAAATATGTCGGCATGGTCTTAAACGAA

GCGCTGCGCTTATGGCCAACGGTTCCTTATTTTTCCCTATATGCAAAAGAAGATACG

GTGCTTGGAGGAGAATATCCTTTAGAAAAAGGCGACGAAGTAATGGTTCTGATTCCT

CAGCTTCACCGTGATAAAACAGTTTGGGGAGACGATGTGGAGGAGTTCCGTCCAGA

GCGTTTTGAAAATCCAAGTGCGATTCCGCAGCATGCGTTTAAACCGTTTGGAAACGG

TCAGCGTGCGTCTATCGGTCAGCAGTTCGCTCTTCATGAAGCAACGCTGGTACTTGG

TATGATGCTAAAACACTTTGACTTTGAAGATCATACAAACTACGAGCTCGATATTAA

AGAACTGCTTACGTTAAAACCTAAAGGCTTTGTGGTAAAAGCAAAATCGAAAAAAA

TTCCGCTTGGCGGTATTCCTTCACCTAGCACTGAACAGTCTGCTAAAAAAGTACGCA

AAAAGGCAGAAAACGCTCATAATACGCCGCTGCTTGTGCTATACGGTTCAAATATG

GGTACCGCTGAAGGAACGGCGCGTGATTTAGCAGATATTGCAATGAGCAAAGGATT

TGCACCGCAGGTCGCAACGCTTGATTCACACGCCGGAAATCTTCCGCGCGAAGGAG

CTGTATTAATTGTAACGGCGTCTTATAACGGTCATCCGCCTGATAACGCAAAGCAAT

TTGTCGACTGGTTAGACCAAGCGTCTGCTGATGAAGTAAAAGGCGTTCGCTACTCCG

TATTTGGATGCGGCGATAAAAACTGGGCTACTACGTATCAAAAAGTGCCTGCTTTTA

TCGATGAAACGCTTGCCGCTAAAGGGGCAGAAAACATCGCTGACCGCGGTGAAGCA

GATGCAAGCGACGACTTTGAAGGCACATATGAAGAATGGCGTGAACATATGTGGAG

TGACGTAGCAGCCTACTTTAACCTCGACATTGAAAACAGTGAAGATAATAAATCTAC

TCTTTCACTTCAATTTGTCGACAGCGCCGCGGATATGCCGCTTGCGAAAATGCACGG

TGCGTTTTCAACGCTCGAGCACCACCACCACCACTGA

Amino Acid Sequence:

MTIKEMPQPKTFGELKNLPLLNTDKPVQALMKIADELGEIFKFEAPGRVTRYLSSQRLIK
EACDESRFDKELSQTLKFLRDFLGDGLATSWTHEKNWKKAHNILLPSFSQQAMKGYHA
MMVDIAVQLVQKWERLNADEHIEVSEDMTRLTLDTIGLCGFNYRFNSFYRDQPHPFIISL
VRALDEVMNKLQRANPDDPAYDENKRQFQEDIKVMNDLVDKIIADRKARGEQSDDLLT
QMLNGKDPETGEPLDDGNIRYQIITFLYAGVEGTSGLLSFALYFLVKNPHVLQKVAEEA
ARVLVDPVPSYKQVKQLKYVGMVLNEALRLWPTVPYFSLYAKEDTVLGGEYPLEKGD
EVMVLIPQLHRDKTVWGDDVEEFRPERFENPSAIPQHAFKPFGNGQRASIGQQFALHEA
TLVLGMMLKHFDFEDHTNYELDIKELLTLKPKGFVVKAKSKKIPLGGIPSPSTEQSAKKV
RKKAENAHNTPLLVLYGSNMGTAEGTARDLADIAMSKGFAPQVATLDSHAGNLPREG
AVLIVTASYNGHPPDNAKQFVDWLDQASADEVKGVRYSVFGCGDKNWATTYQKVPAF
IDETLAAKGAENIADRGEADASDDFEGTYEEWREHMWSDVAAYFNLDIENSEDNKSTL
SLQFVDSAADMPLAKMHGAFSTLEHHHHHH*

DNA and amino acid sequence of our final variant, an evolved C–H alkylation enzyme:

DNA Sequence:

ATGACAATTAAAGAAATGCCTCAGCCAAAAACGTTTGGAGAGCTTAAAAATTTACC
GTTATTAAACACAGATAAACCGGTTCAAGCTTTGATGAAAATTGCGGATGAATTAGG
AGAAATCTTTAAATTCGAGGCGCCTGGTCGTGTAACGCGCTACTTATCAAGTCAGCG
TCTAATTAAAGAAGCATGCGATGAATCACGCTTTGATAAAGAGTTAAGTCAAACGCT
GAAATTTCTGCGTGATTTTCTTGGAGACGGGTTAGCCACAAGCTGGACGCATGAAAA
AAATTGGAAAAAAGCGCATAATATCTTACTTCCAAGCTTTAGTCAGCAGGCAATGA
AAGGCTATCATGCGATGATGGTCGATATCGCCGTGCAGCTTGTTCAAAAGTGGGAGC

GTCTAAATGCAGATGAGCACATTGAAGTATCGGAAGACATGACACGCTTAACGCTT

GATACAATTGGTCTTTGCGGCTTTAACTATCGCTTTAACAGCTTTTACCGAGATCAGC

CTCATCCATTTACTATAAGTCTGGTCCGTGCACTGGATGAAGTAATGAACAAGCTGC

AGCGAGCAAATCCAGACGACCCAGCTTATGATGAAAACAAGCGCCAGTTTCAAGAA

GATATCAAGGTGATGAACGACCTAGTAGATAAAATTATTGCAGATCGCAAAGCAAG

GGGTGAACAAAGCGATGATTTATTAACGCAGATGCTAAACGGAAAAGATCCAGAAA

CGGGTGAGCCGCTTGATGACGGGAACATTCGCTATCAAATTATTACATTCTTATATG

CGGGAGTTGAAGGTACAAGTGGTCTTTTATCATTTGCGCTGTATTTCTTAGTGAAAA

ATCCACATGTATTACAAAAAGTAGCAGAAGAAGCAGCACGAGTTCTAGTAGATCCT

GTTCCAAGCTACAAACAAGTCAAACAGCTTAAATATGTCGGCATGGTCTTAAACGA

AGCGCTGCGCTTATGGCCAACGGTTCCTTATTTTTCCCTATATGCGAAAGAAGATAC

GGTGCTTGGAGGAGAATATCCTTTAGAAAAAGGCGACGAAGTAATGGTTCTGATTC

CTCAGCTTCACCGTGATAAAACAGTTTGGGGAGACGATGTGGAGGAGTTCCGTCCA

GAGCGTTTTGAAAATCCAAGTGCGATTCCGCAGCATGCGTTTAAACCGTTTGGAAAC

GGTCAGCGTGCGTCTATCGGTCAGCAGTTCGCTCTTCATGAAGCAACGCTGGTACTT

GGTATGATGCTAAAACACTTTGACTTTGAAGATCATACAAACTACGAGCTCGATATT

AAAGAACTGCTTACGTTAAAACCTAAAGGCTTTGTGGTAAAAGCAAAATCGAAAAA

AATTCCGCTTGGCGGTATTCCTTCACCTAGCACTGAACAGTCTGCTAAAAAAGTACG

CAAAAAGGCAGAAAACGCTCATAATACGCCGCTGCTTGTGCTATACGGTTCAAATAT

GGGTACCGCTGAAGGAACGGCGCGTGATTTAGCAGATATTGCAATGAGCAAAGGAT

TTGCACCGCAGGTCGCAACGCTTGATTCACACGCCGGAAATCTTCCGCGCGAAGGA

GCTGTATTAATTGTAACGGCGTCTTATAACGGTCATCCGCCTGATAACGCAAAGCAA

TTTGTCGACTGGTTAGACCATGCGTCTGCTGATGAAGTAAAAGGCGTTCGCTACTCC

GTATTTGGATGCGGCGATAAAAACTGGGCTACTACGTATCAAAAAGTGCCTGCTTTT
ATCGATGAAACGCTTGCCGCTAAAGGGGCAGAAAACATCGCTGACCGCGGTGAAGC
AGATGCAAGCGACGACTTTGAAGGCACATATGAAGAATGGCGTGAACATATGTGGA
GTGACGTAGCAGCCTACTTTAACCTCGACATTGAAAACAGTGAAGATAATTAA

Amino Acid Sequence:

MTIKEMPQPKTFGELKNLPLLNTDKPVQALMKIADELGEIFKFEAPGRVTRYLSSQRLIK
EACDESRFDKELSQTLKFLRDFLGDGLATSWTHEKNWKKAHNILLPSFSQQAMKGYHA
MMVDIAVQLVQKWERLNADEHIEVSEDMTRLTLDTIGLCGFNYRFNSFYRDQPHPFTIS
LVRALDEVMNKLQRANPDDPAYDENKRQFQEDIKVMNDLVDKIIADRKARGEQSDDLL
TQMLNGKDPETGEPLDDGNIRYQIITFLYAGVEGTSGLLSFALYFLVKNPHVLQKVAEE
AARVLVDPVPSYKQVKQLKYVGMVLNEALRLWPTVPYFSLYAKEDTVLGGEYPLEKG
DEVMVLIPQLHRDKTVWGDDVEEFRPERFENPSAIPQHAFKPFGNGQRASIGQQFALHE
ATLVLGMMLKHFDFEDHTNYELDIKELLTLKPKGFVVKAKSKKIPLGGIPSPSTEQSAKK
VRKKAENAHNTPLLVLYGSNMGTAEGTARDLADIAMSKGFAPQVATLDSHAGNLPRE
GAVLIVTASYNGHPPDNAKQFVDWLDHASADEVKGVRYSVFGCGDKNWATTYQKVP
AFIDETLAAKGAENIADRGEADASDDFEGTYEEWREHMWSDVAAYFNLDIENSEDN

Chemical synthesis and characterization

General Procedure A: Methylation of alcohols

To a 250 mL round bottom flask was added NaH (60% dispersion in mineral oil, 15–30 mmol, 1.2–1.5 equiv.). The flask was evacuated and filled with argon (3 times). Anhydrous THF (45–80 mL) was added by syringe and the reaction mixture was cooled to 0 °C in an ice bath. Alcohol (10–20 mmol, 1.0 equiv.) in THF (5–10 mL) was added dropwise and the reaction mixture was allowed to warm to room temperature and stirred for 30 minutes. Following, iodomethane (20–40 mmol, 2.0 equiv.) in THF (10 mL) was added and the reaction was stirred at room temperature (8–15 hours). The reaction was quenched by the addition of brine (60 mL) or NH4Cl (sat. aq., 60 mL) and the phases were separated. The aqueous layer was extracted with diethyl ether (3 × 60 mL); the combined organics were washed with aq. sodium thiosulfate (10% w/v, 50 mL, when necessary), dried over Na2SO4 and concentrated under reduced pressure. Purification by silica column chromatography with hexanes / ethyl acetate or pentane / diethyl ether afforded compounds the desired products in 45–60% yield.

General Procedure B: Thioacetate Tagging of alkyl bromines

To a 100 mL round bottom flask was added alkyl bromide (1-10 mmol, 1 equiv.) in acetone (10-50 mL). Potassium thioacetate (6-60 mmol, 6.0 equiv.) in acetone was added and the reaction mixture was stirred for up two days until the bromide was fully consumed as indicated by TLC. Solvent was removed under reduced pressure and crude product was purified by silica column chromatography (hexanes / ethyl acetate) afforded the desired products in 95–99% yield.

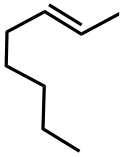
(E)-S-(7-methoxyhept-5-en-1-yl) ethanethioate

To a 100 mL flamed dried flask was added Grubbs' catalyst 2nd generation (85 mg, 1 mol%). The flask was then evacuated and backfilled with argon three times. Under argon, a dry CH_2Cl_2 solution containing 6-bromo-1-hexene (1.63 g, 10 mmol, 1.0 equiv.) and crotonaldehyde (3.50 g, 50 mmol, 5.0 equiv.) was added to the flask. The mixture was stirred under reflux for 20 hours and then cooled to room temperature and filtered through a silica plug. The solvent was removed under reduced pressure and the crude product was purified by flash chromatography (hexanes / ethyl acetate) to give (E)-7-bromohept-2-enal (1.6 g, 84% yield). This product was then dissolved in 10 mL dry THF and then added slowly to a suspension of $NaBH_4$ (375 mg, 10 mmol, 1.0 equiv.) in dry THF (10 mL) at 0 °C. To this reaction mixture, iodine (1.27 g, 5 mmol, 0.5 equiv.) in 10 mL of THF was slowly added at 0 °C. Reaction was stirred until the aldehyde was fully reduced as indicated by TLC. The reaction was quenched with NH_4Cl (sat. aq.), the phases were separated, and the aqueous phase was extracted with ethyl acetate (3 × 20 mL). The combined organic layers were washed with brine and dried over Na_2SO_4. The solvent was removed under reduced pressure and the crude alcohol product was used directly without purification. *General Procedure A* was used for the methylation step and *General Procedure B* was used for the thioacetate tagging. The final product was obtained with 50% overall yield (1.01g, 5 mmol).

^{1}H NMR (400 MHz, Chloroform-*d*) δ 5.67 (dtt, *J* = 15.4, 6.5, 1.1 Hz, 1H), 5.54 (dtt, *J* = 15.4, 6.1, 1.3 Hz, 1H), 3.85 (dq, *J* = 6.0, 1.0 Hz, 2H), 3.31 (s, 3H), 2.86 (t, *J* = 7.2 Hz, 2H), 2.32 (s, 3H), 2.06 (tdq, *J* = 7.5, 6.6, 1.1 Hz, 2H), 1.64 – 1.53 (m, 2H), 1.53 – 1.39 (m, 2H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ 195.98, 134.03, 126.66, 77.36, 77.04, 76.72, 73.18, 57.75, 31.75, 30.65, 29.02, 28.95, 28.17. **HRMS** (EI) m/z: 202.1031 ($M^{+\cdot}$); calc. for $C_{10}H_{18}SO_2$: 202.1028.

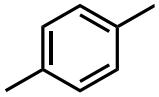
(E)-1-Methoxyoct-2-ene

Prepared from (E)-oct-2-en-1-ol using *General Procedure A.* This compound is known in the literature[6]. **^{1}H NMR** (400 MHz, CDCl3) δ 5.70 (dtt, J = 15.6, 6.6, 1.2 Hz, 1H), 5.54 (dtt, J = 15.3, 6.2, 1.4 Hz, 1H), 3.86 (dq, J = 6.2, 1.0 Hz, 2H), 3.31 (s, 3H), 2.08 – 1.99 (m, 2H), 1.43 – 1.34 (m, 2H), 1.34 – 1.21 (m, 4H), 0.88 (t, J = 7.0 Hz, 3H). **^{13}C NMR** (101 MHz, CDCl3) δ 135.2, 126.1, 73.5, 57.8, 32.4, 31.5, 28.9, 22.7, 14.2.



1-(Methoxymethyl)-4-methylbenzene

Prepared from *p*-tolylmethanol using *General Procedure A*. This compound is known in the literature[6]. **^{1}H NMR** (400 MHz, CDCl3) δ 7.23 (d, J = 8.0 Hz, 2H), 7.17 (d, J = 7.9 Hz, 2H), 4.43 (s, 2H), 3.37 (s, 3H), 2.35 (s, 3H).



Ethyl (E)-9-(acetylthio)-3-methoxynon-4-enoate

To a 100 mL flamed dried flask was added Grubbs' catalyst 2nd generation (85 mg, 1 mol%). The flask was then evacuated and backfilled with argon three times. Under argon, a dry CH_2Cl_2 solution containing 6-bromo-1-hexene (1.63 g, 10 mmol, 1.0 equiv.) and crotonaldehyde (3.50 g, 50 mmol, 5.0 equiv.) was added to the flask. The mixture was stirred under reflux for 20 hours and then cooled to room temperature and filtered through a silica plug. The solvent was removed under reduced pressure and the crude product was purified by flash chromatography (hexanes / ethyl acetate) to give (E)-7-bromohept-2-enal (1.6 g, 84% yield).

In a dry 100 mL round bottom flask, under argon, a solution of diisopropylamine (3.3 mmol, 1.1 equiv.) in THF (15 mL) was cooled to -78 °C and kept at this temperature for the remainder of the reaction. n-Butyllithium (3.3 mmol, 1.1 equiv., 1.6 or 2.5 M in hexanes) was added dropwise

and the resulting mixture was stirred for 15-30 min. Ethyl acetate (3 mmol, 1.0 equiv.) was added dropwise and the mixture was stirred for an additional 30-45 min. Then, neat aldehyde (3.3 mmol, 1.1 equiv.) was added slowly and the solution was stirred for a further 3 hours. The reaction mixture was quenched at -78 °C by the addition of NH_4Cl (sat. aq., 30 mL) and allowed to thaw to room temperature. Phases were separated and the aqueous phase was extracted with ethyl acetate or diethyl ether (3 × 20–30 mL). The combined organics were washed with NH4Cl (sat. aq., 2 × 10–15 mL), brine (10 mL), dried over Na2SO4 and concentrated under reduced pressure. Purification by silica column chromatography with hexanes / ethyl acetate afforded the desired aldol adducts in 75% yield. *General Procedure A* was used for the methylation step and *General Procedure B* was used for the thioacetate tagging. The final product was obtained with 40% overall yield (1.15g, 4 mmol).

^{1}H NMR (400 MHz, Chloroform-d) δ 5.68 (dtd, J = 15.4, 6.8, 0.8 Hz, 1H), 5.31 (ddt, J = 15.4, 8.1, 1.4 Hz, 1H), 4.14 (qd, J = 7.1, 0.7 Hz, 2H), 3.97 (tdd, J = 8.1, 5.5, 0.8 Hz, 1H), 3.25 (s, 3H), 2.86 (t, J = 7.2 Hz, 2H), 2.59 (dd, J = 14.9, 8.1 Hz, 1H), 2.42 (dd, J = 15.0, 5.5 Hz, 1H), 2.32 (s, 3H), 2.16 – 2.00 (m, 2H), 1.64 – 1.53 (m, 3H), 1.53 – 1.39 (m, 2H), 1.25 (t, J = 7.1 Hz, 3H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ 195.93, 171.04, 134.44, 129.21, 78.73, 77.34, 77.03, 76.71, 60.46, 56.16, 41.34, 31.61, 30.66, 28.99, 28.90, 28.16, 14.25. **HRMS** (EI) m/z: 288.1388 ($M^{+\cdot}$); calc. for $C_{14}H_{24}SO_4$: 288.1395.

Small scale enzymatic reactions and product calibration curve. Enzymatic reactions performed on analytical scale were conducted following the general procedure described below, also described in Section I (H). Product formation was quantified by GC based on the calibration curve of the chemically synthesized compound. TTN is defined as the amount of product divided by total heme protein as measured by the hemochrome assay (Section I (G)). Analysis data presented in this section are for results shown in Fig. 3 of the main text.

General procedure for biotransformations using whole E. coli cells. Suspensions of *E. coli* expressing the appropriate heme protein variant in M9-N buffer (OD600 = 30) were degassed by bubbling with argon in sealed vials for at least 40 minutes; the cells were kept on ice during this time. Separately, a solution of D-glucose (250 mM in M9-N) was degassed by sparging with argon for at least 30 minutes. To a 2 mL vial was added a suspension of *E. coli* expressing heme protein (OD600 = 1, 390 µL) in M9-N with glucose. In the anaerobic chamber, alkane substrate (5 µL of 400 mM stock solution in EtOH), and ethyl diazoacetate (5 µL of 400 mM stock solution in EtOH) were added in the listed order. Final reaction volume was 400 µL; final concentrations were 5 mM alkane substrate, 5 mM ethyl diazoacetate, and 30 mM D-glucose. The vials were sealed and shaken at room temperature and 500 rpm for 18 hours in the anaerobic chamber. The expression of heme protein was measured using the hemochrome assay (Section I (G)), and the concentration of heme protein in the biotransformation was calculated accordingly. Reactions in figure S7 were run at OD30 with 10 mM of each reactant, and the volumes of each part were adjusted to maintain the concentrations of each other reaction constituent.

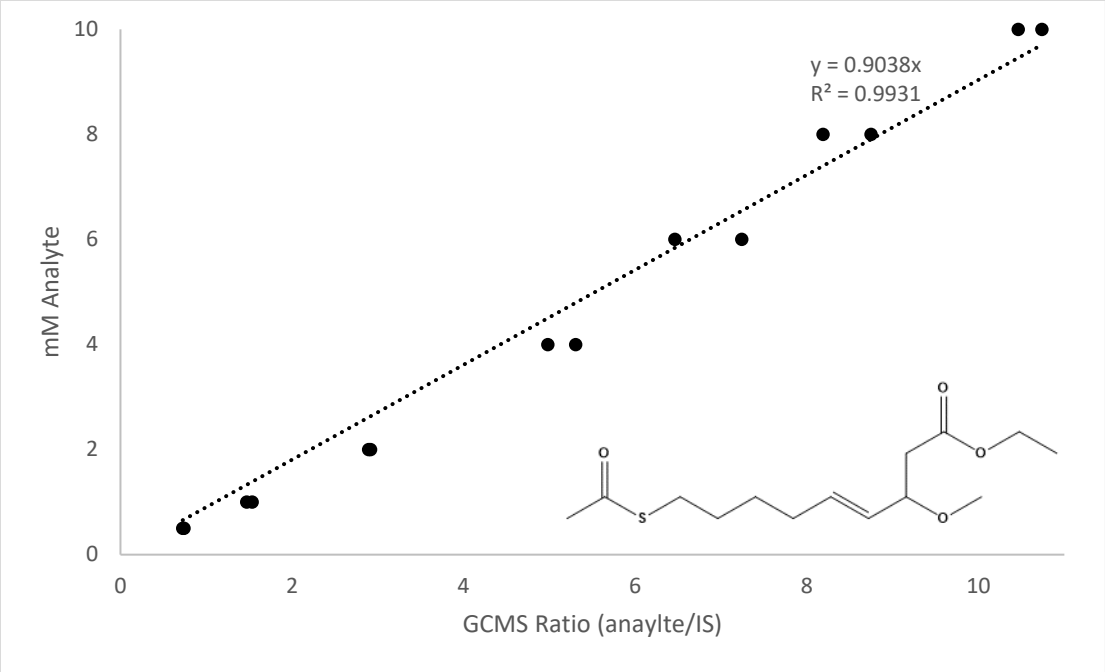
Reaction workup for quantitative GC analysis. Internal standard 1,3,5-trimethoxybenzene (20 µL of 20 mM stock solution in cyclohexane) was added to the reaction vial followed by mixed

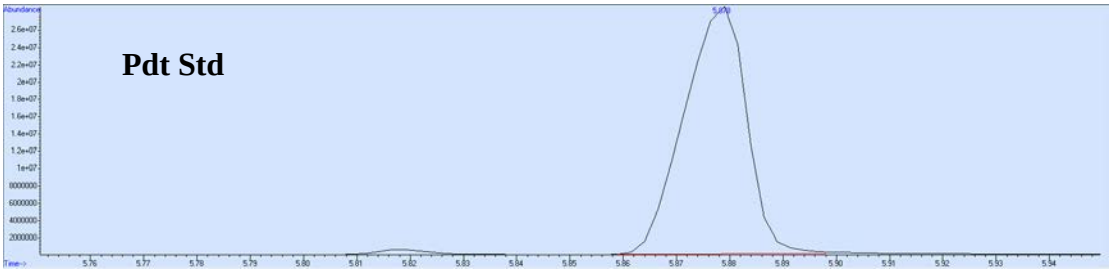
solvent (cyclohexane / ethyl acetate = 1 : 1, 600 µL). The mixture was transferred to a 1.5 mL microcentrifuge tube, vortexed (10 seconds, 3 times), and centrifuged (20,000 × g, 5 minutes) to completely separate the organic and aqueous layers. The organic layer was taken for GC analysis.

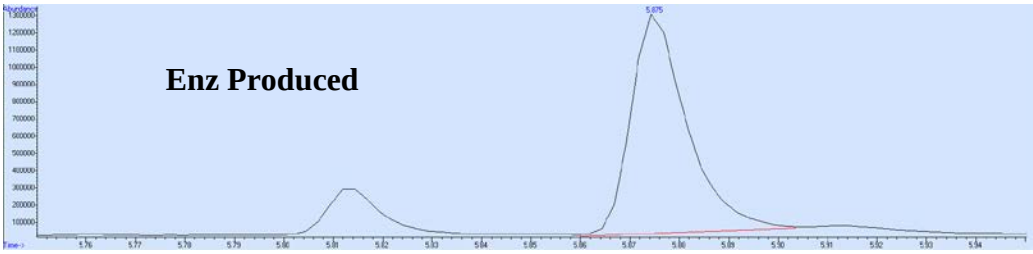
GC calibration curve preparation. Stock solutions of chemically synthesized products at various concentrations (0.2 to 200 mM in EtOH) were prepared. To a microcentrifuge tube were added 360 µL M9-N buffer, 20 µL product stock solution, 20 µL internal standard (20 mM 1,3,5-trimethoxybenzene in cyclohexane), and 600 µL mixed solvent system (cyclohexane : ethyl acetate = 1:1). The mixture was vortexed (10 seconds, 3 times) then centrifuged (20,000 × g, 5 min) to completely separate the organic and aqueous layers. The organic layer was removed for GC analysis. The standard curves plot product concentration in mM (y-axis) against the ratio of product area to internal standard area on the GC (x-axis).

Ethyl (E)-9-(acetylthio)-3-methoxynon-4-enoate

GC calibration curve with 1,3,5-trimethoxybenzene as internal standard (IS)


y = 0.9038x
R² = 0.9931
mM Analyte
GCMS Ratio (anaylte/IS)
0
2
4
6
8
10
O
O
O
O
S


Pdt Std

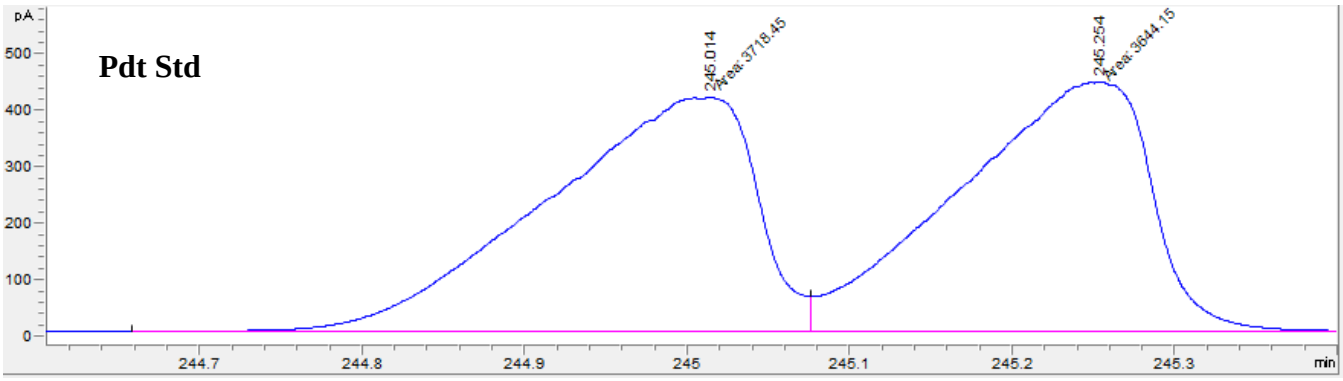

Enz Produced

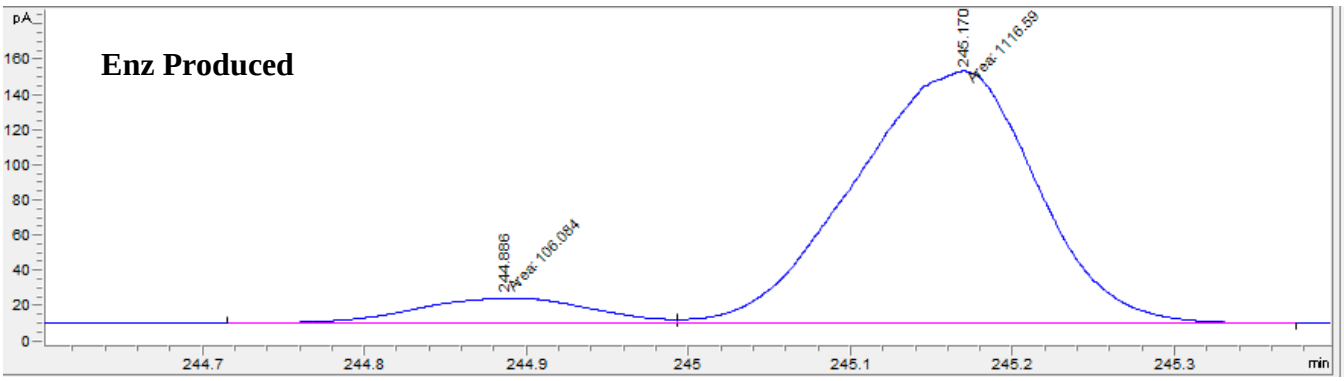
Determination of enantioselectivity

Enantioselectivity of the enzymatic C–H alkylation product was determined by chiral GC analysis. A representative trace, and conditions, are shown below. The absolute configuration of the synthesized product was assigned to be (S) by analogy to the chiral GC separation of Ethyl (E)-3-methoxydec-4-enoate reported in the literature.[4]

Ethyl (E)-9-(acetylthio)-3-methoxynon-4-enoate

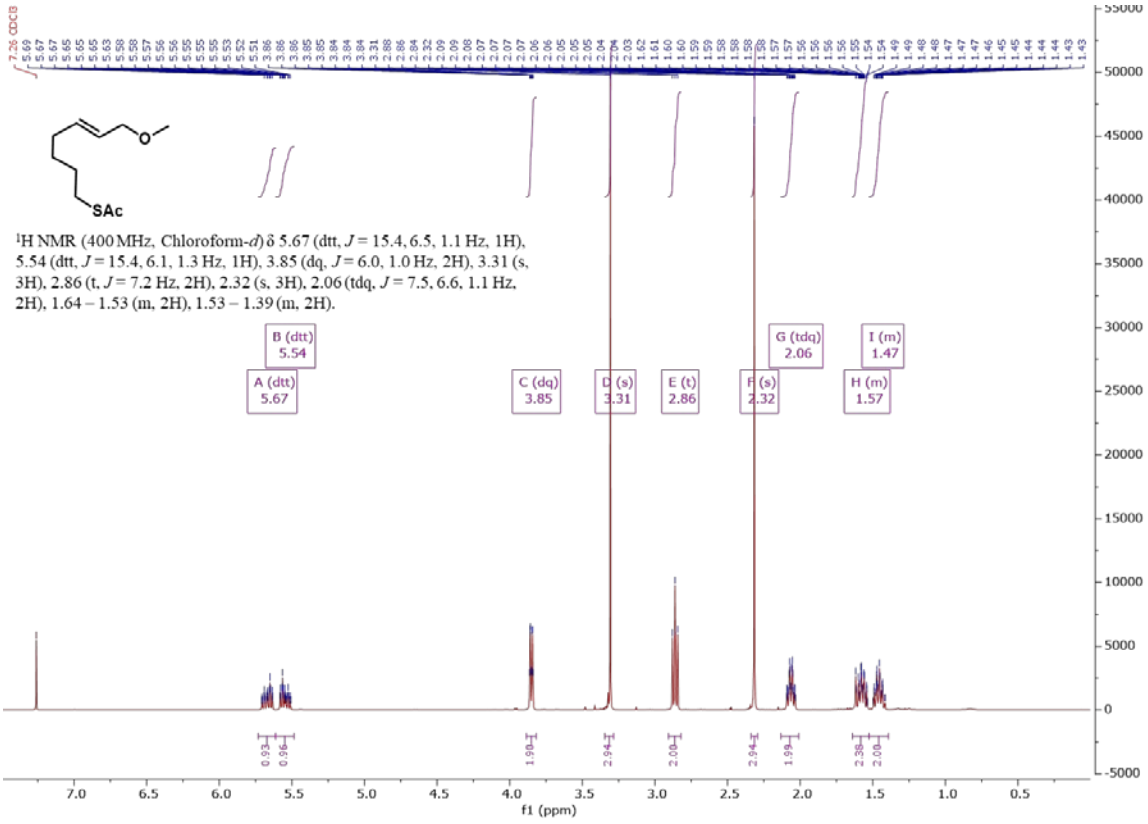
Cyclosil-B column: 140 °C

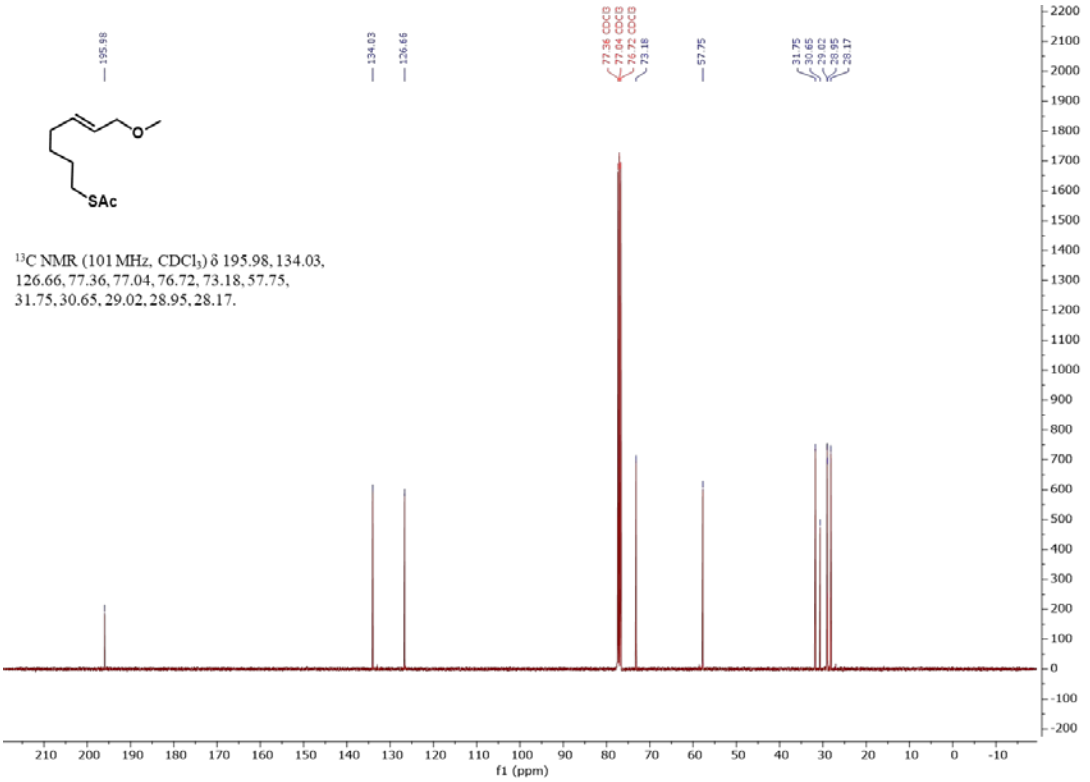




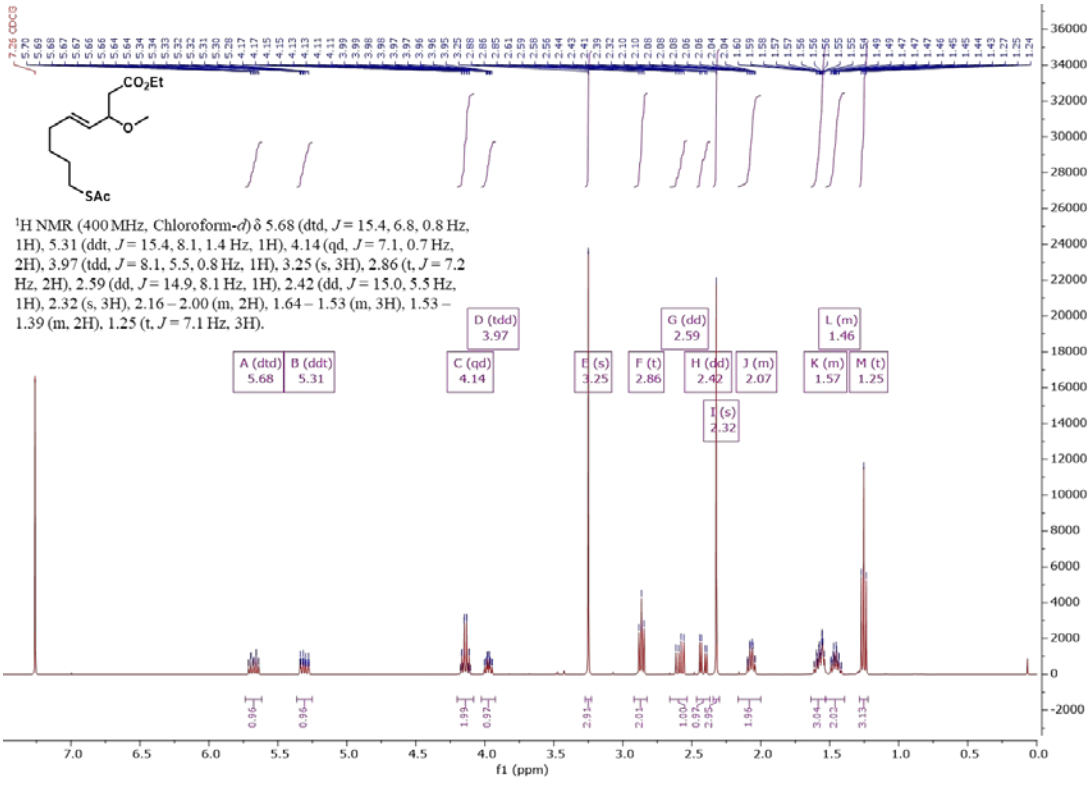
rac			**P411-CHF +P74T**		
Retention Time (min)	Area (mAu*s)	Area %	Retention Time (min)	Area (mAu*s)	Area %
245.014	3718.4	50.505	244.894	83.8	9.253
245.254	3644.2	49.495	245.168	821.4	90.747
Total	7362.6	100	Total	905.2	100
P411-CHF +P74T + I174T +Q553H			**P411-CHF +P74T + I174T +Q553H +K639***		
Retention Time (min)	Area (mAu*s)	Area %	Retention Time (min)	Area (mAu*s)	Area %
244.886	106.1	8.676	244.889	152.4	8.739
245.17	1116.6	91.324	245.171	1591.1	91.261
Total	1222.7	100	Total	1743.5	100

NMR Spectra

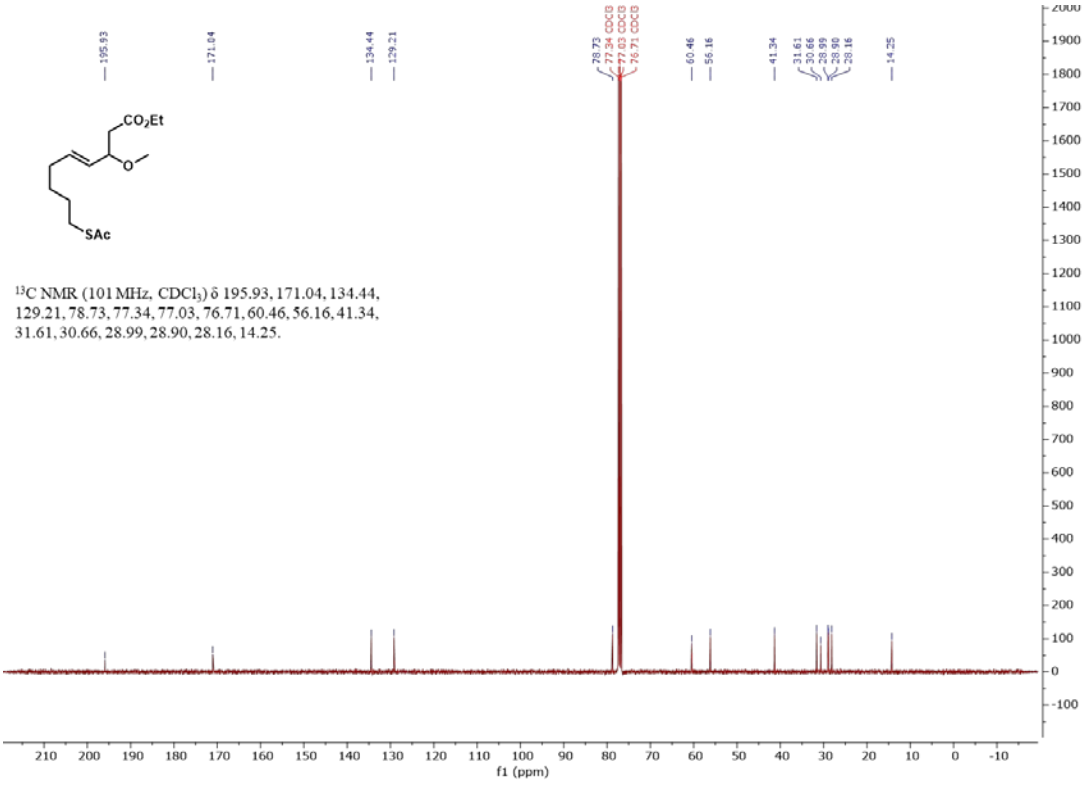




^{13}C NMR (101 MHz, $CDCl_3$) δ 195.98, 134.03, 126.66, 77.36, 77.04, 76.72, 73.18, 57.75, 31.75, 30.65, 29.02, 28.95, 28.17.



^{1}H NMR (400 MHz, Chloroform-*d*) δ 5.68 (dtd, *J* = 15.4, 6.8, 0.8 Hz, 1H), 5.31 (ddt, *J* = 15.4, 8.1, 1.4 Hz, 1H), 4.14 (qd, *J* = 7.1, 0.7 Hz, 2H), 3.97 (tdd, *J* = 8.1, 5.5, 0.8 Hz, 1H), 3.25 (s, 3H), 2.86 (t, *J* = 7.2 Hz, 2H), 2.59 (dd, *J* = 14.9, 8.1 Hz, 1H), 2.42 (dd, *J* = 15.0, 5.5 Hz, 1H), 2.32 (s, 3H), 2.16 – 2.00 (m, 2H), 1.64 – 1.53 (m, 3H), 1.53 – 1.39 (m, 2H), 1.25 (t, *J* = 7.1 Hz, 3H).



^{13}C NMR (101 MHz, $CDCl_3$) δ 195.93, 171.04, 134.44, 129.21, 78.73, 77.34, 77.03, 76.71, 60.46, 56.16, 41.34, 31.61, 30.66, 28.99, 28.90, 28.16, 14.25.

Supplementary Figures:

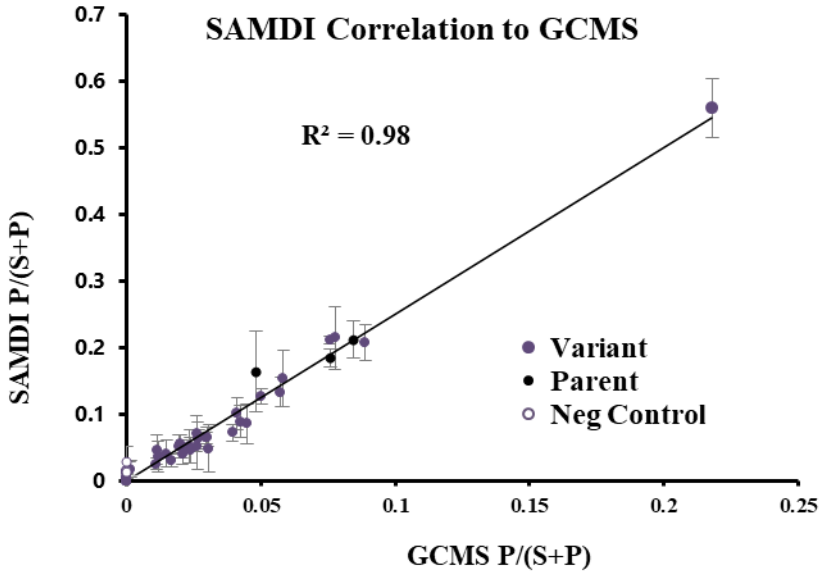


Figure S1a: Scatterplot of screening results from Figure 3.3 inclusive of an additional variant that was subsequently identified as a false positive by GCMS (Figure S1b). Data was collected via GCMS and SAMDI from a library of 70 mutants using P411-CHF(P74T) as the parent and 400 µM $MnCl_2$ during epPCR comparing product to remaining substrate. Values were calculated as a fraction of product over the total of the remaining starting material and product formed. Correlation was determined using least squares linear regression. Here a slope of 2.5 shows that the SAMDI and GC methods do not give the same uncalibrated measurement of yield, and is likely due to different ionization efficiencies of the reaction products. We have, however, verified that the SAMDI method does provide a quantitative measure of yield and allows identification of top variants and perform evolution.

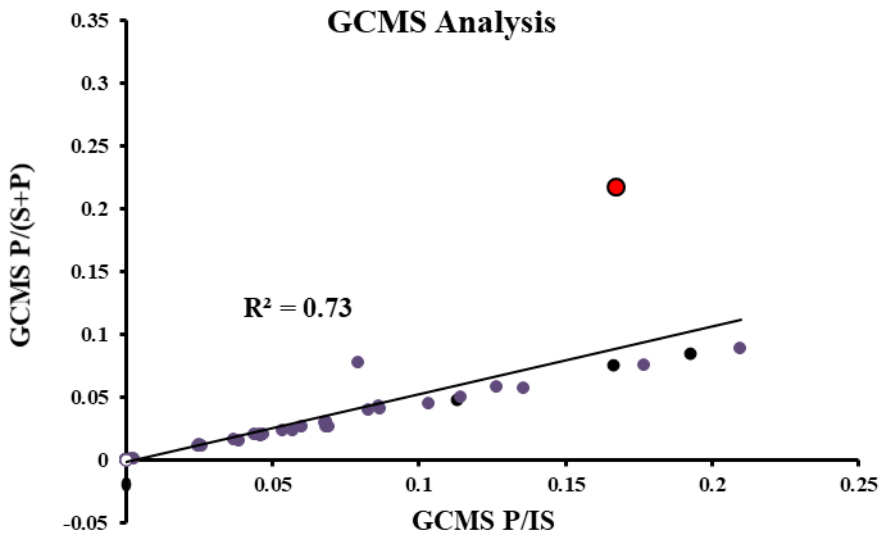


Figure S1b: Correlations of variants collected on GCMS from the random mutagenesis in Figure S1a library using P411-CHF(P74T) as the parent and 400 µM $MnCl_2$ during epPCR comparing product to remaining substrate with product concentration corrected by an internal standard. Data showed good correlation for most points; however, this direct comparison reveals a point (red) to be a false positive hit when only considering product and remaining substrate. Parent controls are shown in black and negative controls in white.

Gen.	Percent Yield	Fold Improvement	Mutation	TTN
0	2.44%	--	--	637
1	4.46%	1.83	**I174T Q553H**	1107
2	6.28%	1.4	**K639***	1310

Figure S2. Beneficial mutations discovered by SAMDI. Due to the large number of variants screened, other improved variants were identified in the course of this study with lower fold-improvement, however only the top variant from each round was chosen and characterized more rigorously.

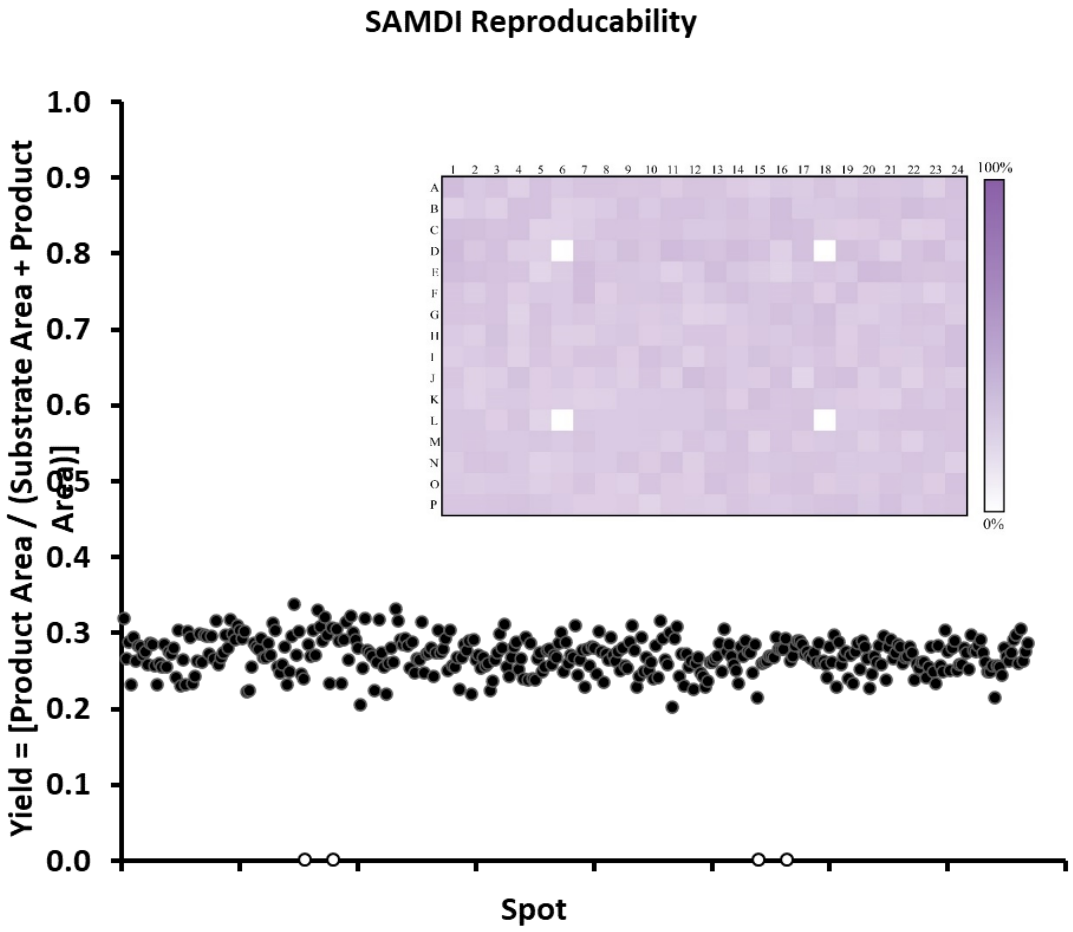


Figure S3. Reproducibility in the SAMDI method. Using SAMDI, a whole-cell reaction with thioacetate-tagged 1-methoxyhept-2-ene was deprotected and spotted 380 times on a SAMDI plate, without prior purification or centrifugation. 10 mL of reaction was conducted from 1 colony and 10 mM substrate at 25 ℃ for 24hours. Negative controls using P411-TrpB are located in rows D and L, columns 6 and 18. A standard deviation of 2.3% was found on an average yield of 27%, indicating the assay is able to discern active wells from non-active wells sufficiently. From these data, we subsequently calculated the resolving power of the technique using the m/z values for each peak and the average resolution output by the instrument. Here we found a value of 0.1 m/z, indicating the method's applicability to transformations of minute changes in mass.

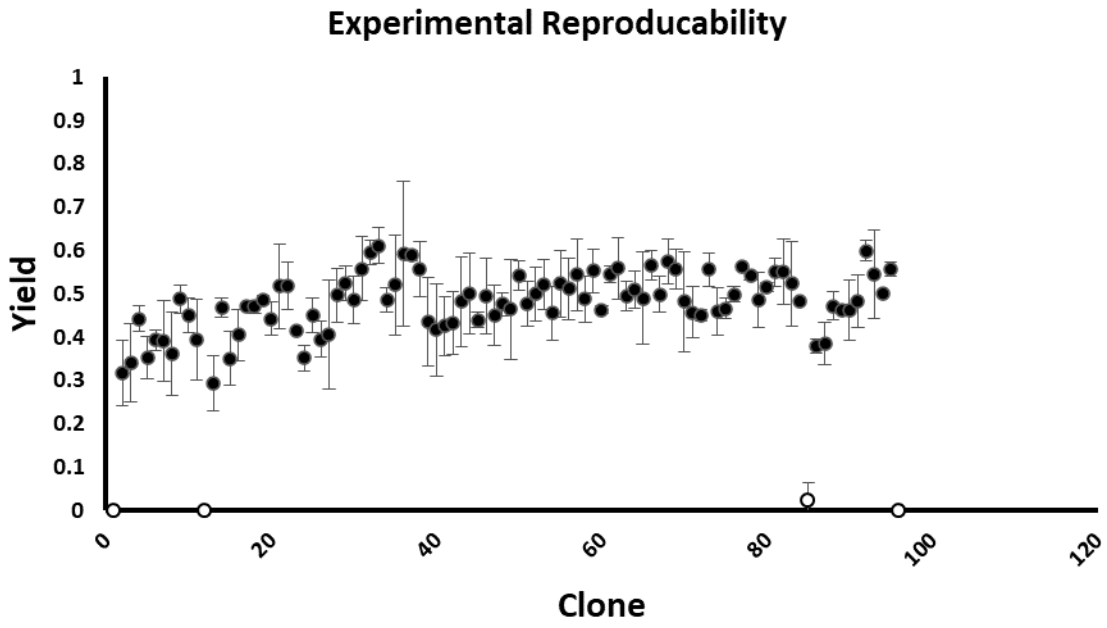


Figure S4. Results of screening a plate of the same clone to show the assay's variability. Measurement of P411-CHF(P74T) activity with thioacetate-tagged 1-methoxyhept-2-ene in a 96-deepwell format. 1 mL of expressed cells was centrifuged and resuspended in 0.4 mL M9-N with 10 mM substrate at 22 ℃ for 24 hours. Negative controls with no cells are located in rows A and H, columns 1 and 12. Each data point represents a separate reaction using a different clone from the same enzyme. Each reaction was deprotected and spotted in quadruplicates to SAMDI plates, without prior purification or centrifugation and averaged to give each data point. All values are then averaged. A coefficient of variation of 14% was found on an average yield of 48%, indicating the assay's reproducibility in screening. An increase in variability outside of the SAMDI technique is expected here as a result of differences in expression and all experimental errors.

MATLAB Relative Abundance Code

```
%USER INPUTS

%USER: enter filenames (format: 'NAME.txt')
substrate = ['MLASGC.txt'];
product1 =  [ 'LASGC.txt'];
product2 =  [  'ASGC.txt'];
product3 =  [   'SGC.txt'];
product4 =  [    'GC.txt'];
product5 =  [     'C.txt'];
internalstandard = ['EG3.txt'];
noise = ['noise.txt']; %I choose to export values at m/z=1700.


time_vector_factor =        ; % = time/data points

%USER: enter maximum values (format: VALUE)
substrate_max = 1;
product1_max = 1;
product2_max =1;
product3_max =1;
product4_max =1;
product5_max =1;
internalstandard_max = 1;

%% Internal Standard

%calldata
data = importdata(internalstandard);
lengthvar = length(data.textdata(:,1));

xpos = data.textdata(4:lengthvar,4);
ypos = data.textdata(4:lengthvar,5);
values = data.textdata(4:lengthvar,2);
xpos2 = zeros(lengthvar-3,1);
for i = 1:lengthvar-4
    str = cell2mat(xpos(i));
    if strlength(str) ~= 4
        str = strcat(num2str(0),str);
        xpos(i) = mat2cell(str,1);
    end
    num = str2num(str);
    xpos2(i) = num;
end
xpos = xpos2;
ypos = cell2mat(ypos);
ypos = str2num(ypos);
values = cell2mat(values);
values = str2num(values);

tbl = table(xpos,ypos,values);

%sort by x-val
tbl = sortrows(tbl,'xpos');
xpos = tbl{:,1};

%remove bad data
for l = 1:size(tbl,1)
    if xpos(l) == 0
        tbl(l,:) = [];
    end
end

xpos = tbl{:,1};
ypos = tbl{:,2};
values = tbl{:,3};

%%
```

```
block = 1;
blocks = [];
index = 0;
indices = [];
for j = 1:lengthvar-5
    if xpos(j) == xpos(j+1)
        block = block + 1;
    else
        index = j;
        indices = [indices;index];
        blocks = [blocks; block];
        block = 1;
    end
    blockmat = [blocks, indices];
end
%%

threshold_main = 4;
threshold_small = 14;
ROIs = [];
ind_ROIs = [];
for k = 1:size(blocks,1)-1
    if (abs(blocks(k) - blocks(k+1)) > threshold_main)
            if blocks(k) < threshold_small | blocks(k+1) < threshold_small
            ROI = indices(k);
            ROIs = [ROIs, ROI];
            ind_ROIs = [ind_ROIs, k];
        end
    end
end

if ROIs(1) ~= 0
    ROIs = [0,ROIs];
end

if ind_ROIs(1) ~= 0
    ind_ROIs = [0,ind_ROIs];
end

%sort into types of code, 1-4

data_store = [];
sd_store = [];
for m = 1:size(ROIs,2)-1
    %for 1, 3, 5 etc. - horizonal, avg over x-val
    section_indices = [];
    if rem(m,2) == 1
        avgs = zeros(1,size(section_indices,2));

        %take out just the data for this section
        section_data = tbl((ROIs(m)+1):(ROIs(m+1)),:);
        section_data = sortrows(section_data, 'xpos');
        sec_xpos = section_data{:,1};
        sec_ypos = section_data{:,2};
        sec_values = section_data{:,3};

        %define # of data points for each x-val
        sz_ind = ind_ROIs(m+1) - ind_ROIs(m);
        section_indices = zeros(1,sz_ind);
        for p = 1:ind_ROIs(m+1) - ind_ROIs(m)
            section_indices(p) = blockmat(p+ind_ROIs(m),1);
        end
        section_indices = [0,section_indices];
        %%

        %avg values of all data points at each x-val
        for n = 1:size(section_indices,2)-1
            starting_index = sum(section_indices(2:n))+1;
            ending_index = sum(section_indices(2:n+1));
            values_vec = sec_values(starting_index:ending_index);
            sum_section = sum(values_vec);
```

```
        avgs(n) = sum_section/section_indices(n+1);
        std_dev(n) = std(values_vec);
    end
    %added
    avgs = flip(avgs);
    %
    std_dev = rmmissing(std_dev);
    avg_sd = sum(std_dev)/length(std_dev);
    data_store = [data_store; avgs'];
    sd_store = [sd_store; avg_sd];
end

 %for 2, 6, 10 (organize by y-val, descending)
if rem(m,4) == 2
    avgs = [];

    %take out just the data for this section
    section_data = tbl((ROIs(m)+1):(ROIs(m+1)),:);
    section_data = sortrows(section_data, 'ypos');
    sec_xpos = section_data{:,1};
    sec_ypos = section_data{:,2};
    sec_values = section_data{:,3};
    %%
    %create indices (# of data points for each y-val)
    inds_y = [];
    i_y = 1;
    for n = 1:size(sec_ypos,1)-1
        if sec_ypos(n) == sec_ypos(n+1)
            i_y = i_y + 1;
        else
            inds_y = [inds_y, i_y];
            i_y = 1;
        end
    end
    section_indices = [0,inds_y];

    for n = 1:size(section_indices,2)-1
         starting_index = sum(section_indices(2:n))+1;
         ending_index = sum(section_indices(2:n+1));
         values_vec = sec_values(starting_index:ending_index);
         sum_section = sum(values_vec);
         avgs(n) = sum_section/section_indices(n+1);
         std_dev(n) = std(values_vec);
    end
    std_dev = rmmissing(std_dev);
    avg_sd = sum(std_dev)/length(std_dev);
    data_store = [data_store; avgs'];
    sd_store = [sd_store; avg_sd];
  end

  %for 4, 8, 12 (organize by y-val, ascending
   if rem(m,4) == 0
    avgs = [];

    %take out just the data for this section
    section_data = tbl((ROIs(m)+1):(ROIs(m+1)),:);
    section_data = sortrows(section_data, 'ypos', 'descend');
    sec_xpos = section_data{:,1};
    sec_ypos = section_data{:,2};
    sec_values = section_data{:,3};
    %%
    %create indices (# of data points for each y-val)
    inds_y = [];
    i_y = 1;
    for n = 1:size(sec_ypos,1)-1
        if sec_ypos(n) == sec_ypos(n+1)
            i_y = i_y + 1;
        else
            inds_y = [inds_y, i_y];
            i_y = 1;
        end
```

```
        end
        section_indices = [0,inds_y];

        for n = 1:size(section_indices,2)-1
             starting_index = sum(section_indices(2:n))+1;
             ending_index = sum(section_indices(2:n+1));
             values_vec = sec_values(starting_index:ending_index);
             sum_section = sum(values_vec);
             avgs(n) = sum_section/section_indices(n+1);
             std_dev(n) = std(values_vec);
        end
        std_dev = rmmissing(std_dev);
        avg_sd = sum(std_dev)/length(std_dev);
        data_store = [data_store; avgs'];
        sd_store = [sd_store; avg_sd];
      end
end

x = [1:size(data_store,1)];
x = x.*1.5;
STDDEV = sum(sd_store)/length(sd_store);

internalstandard_vec = data_store;
internalstandard_stdev = STDDEV;
%%

%% Noise

%calldata
data = importdata(noise);
lengthvar = length(data.textdata(:,1));

xpos = data.textdata(4:lengthvar,4);
ypos = data.textdata(4:lengthvar,5);
values = data.textdata(4:lengthvar,2);
xpos2 = zeros(lengthvar-3,1);
for i = 1:lengthvar-4
    str = cell2mat(xpos(i));
    if strlength(str) ~= 4
        str = strcat(num2str(0),str);
        xpos(i) = mat2cell(str,1);
    end
    num = str2num(str);
    xpos2(i) = num;
end
xpos = xpos2;
ypos = cell2mat(ypos);
ypos = str2num(ypos);
values = cell2mat(values);
values = str2num(values);

tbl = table(xpos,ypos,values);

%sort by x-val
tbl = sortrows(tbl,'xpos');
xpos = tbl{:,1};

%remove bad data
for l = 1:size(tbl,1)
    if xpos(l) == 0
        tbl(l,:) = [];
    end
end

xpos = tbl{:,1};
ypos = tbl{:,2};
values = tbl{:,3};

%%

block = 1;
```

```
blocks = [];
index = 0;
indices = [];
for j = 1:lengthvar-5
    if xpos(j) == xpos(j+1)
        block = block + 1;
    else
        index = j;
        indices = [indices;index];
        blocks = [blocks; block];
        block = 1;
    end
    blockmat = [blocks, indices];
end
%%

threshold_main = 4;
threshold_small = 14;
ROIs = [];
ind_ROIs = [];
for k = 1:size(blocks,1)-1
    if (abs(blocks(k) - blocks(k+1)) > threshold_main)
            if blocks(k) < threshold_small | blocks(k+1) < threshold_small
            ROI = indices(k);
            ROIs = [ROIs, ROI];
            ind_ROIs = [ind_ROIs, k];
        end
    end
end

if ROIs(1) ~= 0
    ROIs = [0,ROIs];
end

if ind_ROIs(1) ~= 0
    ind_ROIs = [0,ind_ROIs];
end

%sort into types of code, 1-4

data_store = [];
sd_store = [];
for m = 1:size(ROIs,2)-1
    %for 1, 3, 5 etc. - horizonal, avg over x-val
    section_indices = [];
    if rem(m,2) == 1
        avgs = zeros(1,size(section_indices,2));

        %take out just the data for this section
        section_data = tbl((ROIs(m)+1):(ROIs(m+1)),:);
        section_data = sortrows(section_data, 'xpos');
        sec_xpos = section_data{:,1};
        sec_ypos = section_data{:,2};
        sec_values = section_data{:,3};

        %define # of data points for each x-val
        sz_ind = ind_ROIs(m+1) - ind_ROIs(m);
        section_indices = zeros(1,sz_ind);
        for p = 1:ind_ROIs(m+1) - ind_ROIs(m)
            section_indices(p) = blockmat(p+ind_ROIs(m),1);
        end
        section_indices = [0,section_indices];
        %%

        %avg values of all data points at each x-val
        for n = 1:size(section_indices,2)-1
            starting_index = sum(section_indices(2:n))+1;
            ending_index = sum(section_indices(2:n+1));
            values_vec = sec_values(starting_index:ending_index);
            sum_section = sum(values_vec);
            avgs(n) = sum_section/section_indices(n+1);
```

```
        std_dev(n) = std(values_vec);
    end
    %added
    avgs = flip(avgs);
    %
    std_dev = rmmissing(std_dev);
    avg_sd = sum(std_dev)/length(std_dev);
    data_store = [data_store; avgs'];
    sd_store = [sd_store; avg_sd];
end

 %for 2, 6, 10 (organize by y-val, descending)
if rem(m,4) == 2
    avgs = [];

    %take out just the data for this section
    section_data = tbl((ROIs(m)+1):(ROIs(m+1)),:);
    section_data = sortrows(section_data, 'ypos');
    sec_xpos = section_data{:,1};
    sec_ypos = section_data{:,2};
    sec_values = section_data{:,3};
    %%
    %create indices (# of data points for each y-val)
    inds_y = [];
    i_y = 1;
    for n = 1:size(sec_ypos,1)-1
        if sec_ypos(n) == sec_ypos(n+1)
            i_y = i_y + 1;
        else
            inds_y = [inds_y, i_y];
            i_y = 1;
        end
    end
    section_indices = [0,inds_y];

    for n = 1:size(section_indices,2)-1
         starting_index = sum(section_indices(2:n))+1;
         ending_index = sum(section_indices(2:n+1));
         values_vec = sec_values(starting_index:ending_index);
         sum_section = sum(values_vec);
         avgs(n) = sum_section/section_indices(n+1);
         std_dev(n) = std(values_vec);
    end
    std_dev = rmmissing(std_dev);
    avg_sd = sum(std_dev)/length(std_dev);
    data_store = [data_store; avgs'];
    sd_store = [sd_store; avg_sd];
  end

  %for 4, 8, 12 (organize by y-val, ascending
   if rem(m,4) == 0
    avgs = [];

    %take out just the data for this section
    section_data = tbl((ROIs(m)+1):(ROIs(m+1)),:);
    section_data = sortrows(section_data, 'ypos', 'descend');
    sec_xpos = section_data{:,1};
    sec_ypos = section_data{:,2};
    sec_values = section_data{:,3};
    %%
    %create indices (# of data points for each y-val)
    inds_y = [];
    i_y = 1;
    for n = 1:size(sec_ypos,1)-1
        if sec_ypos(n) == sec_ypos(n+1)
            i_y = i_y + 1;
        else
            inds_y = [inds_y, i_y];
            i_y = 1;
        end
    end
```

```
        section_indices = [0,inds_y];

        for n = 1:size(section_indices,2)-1
             starting_index = sum(section_indices(2:n))+1;
             ending_index = sum(section_indices(2:n+1));
             values_vec = sec_values(starting_index:ending_index);
             sum_section = sum(values_vec);
             avgs(n) = sum_section/section_indices(n+1);
             std_dev(n) = std(values_vec);
        end
        std_dev = rmmissing(std_dev);
        avg_sd = sum(std_dev)/length(std_dev);
        data_store = [data_store; avgs'];
        sd_store = [sd_store; avg_sd];
      end
end

x = [1:size(data_store,1)];
x = x.*1.5;
STDDEV = sum(sd_store)/length(sd_store);

noise_vec = data_store;
noise_stdev = STDDEV;

%% Product 1

%calldata
data = importdata(product1);
lengthvar = length(data.textdata(:,1));

xpos = data.textdata(4:lengthvar,4);
ypos = data.textdata(4:lengthvar,5);
values = data.textdata(4:lengthvar,2);
xpos2 = zeros(lengthvar-3,1);
for i = 1:lengthvar-4
    str = cell2mat(xpos(i));
    if strlength(str) ~= 4
        str = strcat(num2str(0),str);
        xpos(i) = mat2cell(str,1);
    end
    num = str2num(str);
    xpos2(i) = num;
end
xpos = xpos2;
ypos = cell2mat(ypos);
ypos = str2num(ypos);
values = cell2mat(values);
values = str2num(values);

tbl = table(xpos,ypos,values);

%sort by x-val
tbl = sortrows(tbl,'xpos');
xpos = tbl{:,1};

%remove bad data
for l = 1:size(tbl,1)
    if xpos(l) == 0
        tbl(l,:) = [];
    end
end

xpos = tbl{:,1};
ypos = tbl{:,2};
values = tbl{:,3};

%%

block = 1;
blocks = [];
index = 0;
```

```
indices = [];
for j = 1:lengthvar-5
    if xpos(j) == xpos(j+1)
        block = block + 1;
    else
        index = j;
        indices = [indices;index];
        blocks = [blocks; block];
        block = 1;
    end
    blockmat = [blocks, indices];
end
%%

threshold_main = 4;
threshold_small = 14;
ROIs = [];
ind_ROIs = [];
for k = 1:size(blocks,1)-1
    if (abs(blocks(k) - blocks(k+1)) > threshold_main)
            if blocks(k) < threshold_small | blocks(k+1) < threshold_small
            ROI = indices(k);
            ROIs = [ROIs, ROI];
            ind_ROIs = [ind_ROIs, k];
        end
    end
end

if ROIs(1) ~= 0
    ROIs = [0,ROIs];
end

if ind_ROIs(1) ~= 0
    ind_ROIs = [0,ind_ROIs];
end

%sort into types of code, 1-4

data_store = [];
sd_store = [];
for m = 1:size(ROIs,2)-1
    %for 1, 3, 5 etc. - horizonal, avg over x-val
    section_indices = [];
    if rem(m,2) == 1
        avgs = zeros(1,size(section_indices,2));

        %take out just the data for this section
        section_data = tbl((ROIs(m)+1):(ROIs(m+1)),:);
        section_data = sortrows(section_data, 'xpos');
        sec_xpos = section_data{:,1};
        sec_ypos = section_data{:,2};
        sec_values = section_data{:,3};

        %define # of data points for each x-val
        sz_ind = ind_ROIs(m+1) - ind_ROIs(m);
        section_indices = zeros(1,sz_ind);
        for p = 1:ind_ROIs(m+1) - ind_ROIs(m)
            section_indices(p) = blockmat(p+ind_ROIs(m),1);
        end
        section_indices = [0,section_indices];
        %%

        %avg values of all data points at each x-val
        for n = 1:size(section_indices,2)-1
            starting_index = sum(section_indices(2:n))+1;
            ending_index = sum(section_indices(2:n+1));
            values_vec = sec_values(starting_index:ending_index);
            sum_section = sum(values_vec);
            avgs(n) = sum_section/section_indices(n+1);
            std_dev(n) = std(values_vec);
        end
```

```
    %added
    avgs = flip(avgs);
    %
    std_dev = rmmissing(std_dev);
    avg_sd = sum(std_dev)/length(std_dev);
    data_store = [data_store; avgs'];
    sd_store = [sd_store; avg_sd];
 end

 %for 2, 6, 10 (organize by y-val, descending)
if rem(m,4) == 2
    avgs = [];

    %take out just the data for this section
    section_data = tbl((ROIs(m)+1):(ROIs(m+1)),:);
    section_data = sortrows(section_data, 'ypos');
    sec_xpos = section_data{:,1};
    sec_ypos = section_data{:,2};
    sec_values = section_data{:,3};
    %%
    %create indices (# of data points for each y-val)
    inds_y = [];
    i_y = 1;
    for n = 1:size(sec_ypos,1)-1
        if sec_ypos(n) == sec_ypos(n+1)
            i_y = i_y + 1;
        else
            inds_y = [inds_y, i_y];
            i_y = 1;
        end
    end
    section_indices = [0,inds_y];

    for n = 1:size(section_indices,2)-1
         starting_index = sum(section_indices(2:n))+1;
         ending_index = sum(section_indices(2:n+1));
         values_vec = sec_values(starting_index:ending_index);
         sum_section = sum(values_vec);
         avgs(n) = sum_section/section_indices(n+1);
         std_dev(n) = std(values_vec);
    end
    std_dev = rmmissing(std_dev);
    avg_sd = sum(std_dev)/length(std_dev);
    data_store = [data_store; avgs'];
    sd_store = [sd_store; avg_sd];
  end

  %for 4, 8, 12 (organize by y-val, ascending
   if rem(m,4) == 0
    avgs = [];

    %take out just the data for this section
    section_data = tbl((ROIs(m)+1):(ROIs(m+1)),:);
    section_data = sortrows(section_data, 'ypos', 'descend');
    sec_xpos = section_data{:,1};
    sec_ypos = section_data{:,2};
    sec_values = section_data{:,3};
    %%
    %create indices (# of data points for each y-val)
    inds_y = [];
    i_y = 1;
    for n = 1:size(sec_ypos,1)-1
        if sec_ypos(n) == sec_ypos(n+1)
            i_y = i_y + 1;
        else
            inds_y = [inds_y, i_y];
            i_y = 1;
        end
    end
    section_indices = [0,inds_y];
```

```
        for n = 1:size(section_indices,2)-1
             starting_index = sum(section_indices(2:n))+1;
             ending_index = sum(section_indices(2:n+1));
             values_vec = sec_values(starting_index:ending_index);
             sum_section = sum(values_vec);
             avgs(n) = sum_section/section_indices(n+1);
             std_dev(n) = std(values_vec);
        end
        std_dev = rmmissing(std_dev);
        avg_sd = sum(std_dev)/length(std_dev);
        data_store = [data_store; avgs'];
        sd_store = [sd_store; avg_sd];
      end
end

x = [1:size(data_store,1)];
x = x.*1.5;
STDDEV = sum(sd_store)/length(sd_store);

product1_vec = data_store;
product1_stdev = STDDEV;
%% Product 2

%calldata
data = importdata(product2);
lengthvar = length(data.textdata(:,1));

xpos = data.textdata(4:lengthvar,4);
ypos = data.textdata(4:lengthvar,5);
values = data.textdata(4:lengthvar,2);
xpos2 = zeros(lengthvar-3,1);
for i = 1:lengthvar-4
    str = cell2mat(xpos(i));
    if strlength(str) ~= 4
        str = strcat(num2str(0),str);
        xpos(i) = mat2cell(str,1);
    end
    num = str2num(str);
    xpos2(i) = num;
end
xpos = xpos2;
ypos = cell2mat(ypos);
ypos = str2num(ypos);
values = cell2mat(values);
values = str2num(values);

tbl = table(xpos,ypos,values);

%sort by x-val
tbl = sortrows(tbl,'xpos');
xpos = tbl{:,1};

%remove bad data
for l = 1:size(tbl,1)
    if xpos(l) == 0
        tbl(l,:) = [];
    end
end

xpos = tbl{:,1};
ypos = tbl{:,2};
values = tbl{:,3};

%%

block = 1;
blocks = [];
index = 0;
indices = [];
for j = 1:lengthvar-5
    if xpos(j) == xpos(j+1)
```

```
        block = block + 1;
    else
        index = j;
        indices = [indices;index];
        blocks = [blocks; block];
        block = 1;
    end
    blockmat = [blocks, indices];
end
%%

threshold_main = 4;
threshold_small = 14;
ROIs = [];
ind_ROIs = [];
for k = 1:size(blocks,1)-1
    if (abs(blocks(k) - blocks(k+1)) > threshold_main)
            if blocks(k) < threshold_small | blocks(k+1) < threshold_small
            ROI = indices(k);
            ROIs = [ROIs, ROI];
            ind_ROIs = [ind_ROIs, k];
        end
    end
end

if ROIs(1) ~= 0
    ROIs = [0,ROIs];
end

if ind_ROIs(1) ~= 0
    ind_ROIs = [0,ind_ROIs];
end

%sort into types of code, 1-4

data_store = [];
sd_store = [];
for m = 1:size(ROIs,2)-1
    %for 1, 3, 5 etc. - horizonal, avg over x-val
    section_indices = [];
    if rem(m,2) == 1
        avgs = zeros(1,size(section_indices,2));

        %take out just the data for this section
        section_data = tbl((ROIs(m)+1):(ROIs(m+1)),:);
        section_data = sortrows(section_data, 'xpos');
        sec_xpos = section_data{:,1};
        sec_ypos = section_data{:,2};
        sec_values = section_data{:,3};

        %define # of data points for each x-val
        sz_ind = ind_ROIs(m+1) - ind_ROIs(m);
        section_indices = zeros(1,sz_ind);
        for p = 1:ind_ROIs(m+1) - ind_ROIs(m)
            section_indices(p) = blockmat(p+ind_ROIs(m),1);
        end
        section_indices = [0,section_indices];
        %%

        %avg values of all data points at each x-val
        for n = 1:size(section_indices,2)-1
            starting_index = sum(section_indices(2:n))+1;
            ending_index = sum(section_indices(2:n+1));
            values_vec = sec_values(starting_index:ending_index);
            sum_section = sum(values_vec);
            avgs(n) = sum_section/section_indices(n+1);
            std_dev(n) = std(values_vec);
        end
        %added
        avgs = flip(avgs);
        %
```

```
    std_dev = rmmissing(std_dev);
    avg_sd = sum(std_dev)/length(std_dev);
    data_store = [data_store; avgs'];
    sd_store = [sd_store; avg_sd];
 end

 %for 2, 6, 10 (organize by y-val, descending)
if rem(m,4) == 2
    avgs = [];

    %take out just the data for this section
    section_data = tbl((ROIs(m)+1):(ROIs(m+1)),:);
    section_data = sortrows(section_data, 'ypos');
    sec_xpos = section_data{:,1};
    sec_ypos = section_data{:,2};
    sec_values = section_data{:,3};
    %%
    %create indices (# of data points for each y-val)
    inds_y = [];
    i_y = 1;
    for n = 1:size(sec_ypos,1)-1
        if sec_ypos(n) == sec_ypos(n+1)
            i_y = i_y + 1;
        else
            inds_y = [inds_y, i_y];
            i_y = 1;
        end
    end
    section_indices = [0,inds_y];

    for n = 1:size(section_indices,2)-1
         starting_index = sum(section_indices(2:n))+1;
         ending_index = sum(section_indices(2:n+1));
         values_vec = sec_values(starting_index:ending_index);
         sum_section = sum(values_vec);
         avgs(n) = sum_section/section_indices(n+1);
         std_dev(n) = std(values_vec);
    end
    std_dev = rmmissing(std_dev);
    avg_sd = sum(std_dev)/length(std_dev);
    data_store = [data_store; avgs'];
    sd_store = [sd_store; avg_sd];
  end

  %for 4, 8, 12 (organize by y-val, ascending
   if rem(m,4) == 0
    avgs = [];

    %take out just the data for this section
    section_data = tbl((ROIs(m)+1):(ROIs(m+1)),:);
    section_data = sortrows(section_data, 'ypos', 'descend');
    sec_xpos = section_data{:,1};
    sec_ypos = section_data{:,2};
    sec_values = section_data{:,3};
    %%
    %create indices (# of data points for each y-val)
    inds_y = [];
    i_y = 1;
    for n = 1:size(sec_ypos,1)-1
        if sec_ypos(n) == sec_ypos(n+1)
            i_y = i_y + 1;
        else
            inds_y = [inds_y, i_y];
            i_y = 1;
        end
    end
    section_indices = [0,inds_y];

    for n = 1:size(section_indices,2)-1
         starting_index = sum(section_indices(2:n))+1;
         ending_index = sum(section_indices(2:n+1));
```

```
            values_vec = sec_values(starting_index:ending_index);
            sum_section = sum(values_vec);
            avgs(n) = sum_section/section_indices(n+1);
            std_dev(n) = std(values_vec);
        end
        std_dev = rmmissing(std_dev);
        avg_sd = sum(std_dev)/length(std_dev);
        data_store = [data_store; avgs'];
        sd_store = [sd_store; avg_sd];
      end
end

x = [1:size(data_store,1)];
x = x.*1.5;
STDDEV = sum(sd_store)/length(sd_store);

product2_vec = data_store;
product2_stdev = STDDEV;
%%
%% Product 3

%calldata
data = importdata(product3);
lengthvar = length(data.textdata(:,1));

xpos = data.textdata(4:lengthvar,4);
ypos = data.textdata(4:lengthvar,5);
values = data.textdata(4:lengthvar,2);
xpos2 = zeros(lengthvar-3,1);
for i = 1:lengthvar-4
    str = cell2mat(xpos(i));
    if strlength(str) ~= 4
        str = strcat(num2str(0),str);
        xpos(i) = mat2cell(str,1);
    end
    num = str2num(str);
    xpos2(i) = num;
end
xpos = xpos2;
ypos = cell2mat(ypos);
ypos = str2num(ypos);
values = cell2mat(values);
values = str2num(values);

tbl = table(xpos,ypos,values);

%sort by x-val
tbl = sortrows(tbl,'xpos');
xpos = tbl{:,1};

%remove bad data
for l = 1:size(tbl,1)
    if xpos(l) == 0
        tbl(l,:) = [];
    end
end

xpos = tbl{:,1};
ypos = tbl{:,2};
values = tbl{:,3};

%%

block = 1;
blocks = [];
index = 0;
indices = [];
for j = 1:lengthvar-5
    if xpos(j) == xpos(j+1)
        block = block + 1;
    else
```

```
        index = j;
        indices = [indices;index];
        blocks = [blocks; block];
        block = 1;
    end
    blockmat = [blocks, indices];
end
%%

threshold_main = 4;
threshold_small = 14;
ROIs = [];
ind_ROIs = [];
for k = 1:size(blocks,1)-1
    if (abs(blocks(k) - blocks(k+1)) > threshold_main)
            if blocks(k) < threshold_small | blocks(k+1) < threshold_small
            ROI = indices(k);
            ROIs = [ROIs, ROI];
            ind_ROIs = [ind_ROIs, k];
        end
    end
end

if ROIs(1) ~= 0
    ROIs = [0,ROIs];
end

if ind_ROIs(1) ~= 0
    ind_ROIs = [0,ind_ROIs];
end

%sort into types of code, 1-4

data_store = [];
sd_store = [];
for m = 1:size(ROIs,2)-1
    %for 1, 3, 5 etc. - horizonal, avg over x-val
    section_indices = [];
    if rem(m,2) == 1
        avgs = zeros(1,size(section_indices,2));

        %take out just the data for this section
        section_data = tbl((ROIs(m)+1):(ROIs(m+1)),:);
        section_data = sortrows(section_data, 'xpos');
        sec_xpos = section_data{:,1};
        sec_ypos = section_data{:,2};
        sec_values = section_data{:,3};

        %define # of data points for each x-val
        sz_ind = ind_ROIs(m+1) - ind_ROIs(m);
        section_indices = zeros(1,sz_ind);
        for p = 1:ind_ROIs(m+1) - ind_ROIs(m)
            section_indices(p) = blockmat(p+ind_ROIs(m),1);
        end
        section_indices = [0,section_indices];
        %%

        %avg values of all data points at each x-val
        for n = 1:size(section_indices,2)-1
            starting_index = sum(section_indices(2:n))+1;
            ending_index = sum(section_indices(2:n+1));
            values_vec = sec_values(starting_index:ending_index);
            sum_section = sum(values_vec);
            avgs(n) = sum_section/section_indices(n+1);
            std_dev(n) = std(values_vec);
        end
        %added
        avgs = flip(avgs);
        %
        std_dev = rmmissing(std_dev);
        avg_sd = sum(std_dev)/length(std_dev);
```

```
    data_store = [data_store; avgs'];
    sd_store = [sd_store; avg_sd];
 end

 %for 2, 6, 10 (organize by y-val, descending)
if rem(m,4) == 2
    avgs = [];

    %take out just the data for this section
    section_data = tbl((ROIs(m)+1):(ROIs(m+1)),:);
    section_data = sortrows(section_data, 'ypos');
    sec_xpos = section_data{:,1};
    sec_ypos = section_data{:,2};
    sec_values = section_data{:,3};
    %%
    %create indices (# of data points for each y-val)
    inds_y = [];
    i_y = 1;
    for n = 1:size(sec_ypos,1)-1
        if sec_ypos(n) == sec_ypos(n+1)
            i_y = i_y + 1;
        else
            inds_y = [inds_y, i_y];
            i_y = 1;
        end
    end
    section_indices = [0,inds_y];

    for n = 1:size(section_indices,2)-1
         starting_index = sum(section_indices(2:n))+1;
         ending_index = sum(section_indices(2:n+1));
         values_vec = sec_values(starting_index:ending_index);
         sum_section = sum(values_vec);
         avgs(n) = sum_section/section_indices(n+1);
         std_dev(n) = std(values_vec);
    end
    std_dev = rmmissing(std_dev);
    avg_sd = sum(std_dev)/length(std_dev);
    data_store = [data_store; avgs'];
    sd_store = [sd_store; avg_sd];
  end

  %for 4, 8, 12 (organize by y-val, ascending
   if rem(m,4) == 0
    avgs = [];

    %take out just the data for this section
    section_data = tbl((ROIs(m)+1):(ROIs(m+1)),:);
    section_data = sortrows(section_data, 'ypos', 'descend');
    sec_xpos = section_data{:,1};
    sec_ypos = section_data{:,2};
    sec_values = section_data{:,3};
    %%
    %create indices (# of data points for each y-val)
    inds_y = [];
    i_y = 1;
    for n = 1:size(sec_ypos,1)-1
        if sec_ypos(n) == sec_ypos(n+1)
            i_y = i_y + 1;
        else
            inds_y = [inds_y, i_y];
            i_y = 1;
        end
    end
    section_indices = [0,inds_y];

    for n = 1:size(section_indices,2)-1
         starting_index = sum(section_indices(2:n))+1;
         ending_index = sum(section_indices(2:n+1));
         values_vec = sec_values(starting_index:ending_index);
         sum_section = sum(values_vec);
```

```
            avgs(n) = sum_section/section_indices(n+1);
            std_dev(n) = std(values_vec);
        end
        std_dev = rmmissing(std_dev);
        avg_sd = sum(std_dev)/length(std_dev);
        data_store = [data_store; avgs'];
        sd_store = [sd_store; avg_sd];
    end
end

x = [1:size(data_store,1)];
x = x.*1.5;
STDDEV = sum(sd_store)/length(sd_store);

product3_vec = data_store;
product3_stdev = STDDEV;
%%
%% Product 4

%calldata
data = importdata(product4);
lengthvar = length(data.textdata(:,1));

xpos = data.textdata(4:lengthvar,4);
ypos = data.textdata(4:lengthvar,5);
values = data.textdata(4:lengthvar,2);
xpos2 = zeros(lengthvar-3,1);
for i = 1:lengthvar-4
    str = cell2mat(xpos(i));
    if strlength(str) ~= 4
        str = strcat(num2str(0),str);
        xpos(i) = mat2cell(str,1);
    end
    num = str2num(str);
    xpos2(i) = num;
end
xpos = xpos2;
ypos = cell2mat(ypos);
ypos = str2num(ypos);
values = cell2mat(values);
values = str2num(values);

tbl = table(xpos,ypos,values);

%sort by x-val
tbl = sortrows(tbl,'xpos');
xpos = tbl{:,1};

%remove bad data
for l = 1:size(tbl,1)
    if xpos(l) == 0
        tbl(l,:) = [];
    end
end

xpos = tbl{:,1};
ypos = tbl{:,2};
values = tbl{:,3};

%%

block = 1;
blocks = [];
index = 0;
indices = [];
for j = 1:lengthvar-5
    if xpos(j) == xpos(j+1)
        block = block + 1;
    else
        index = j;
        indices = [indices;index];
```

```
        blocks = [blocks; block];
        block = 1;
    end
    blockmat = [blocks, indices];
end
%%

threshold_main = 4;
threshold_small = 14;
ROIs = [];
ind_ROIs = [];
for k = 1:size(blocks,1)-1
    if (abs(blocks(k) - blocks(k+1)) > threshold_main)
            if blocks(k) < threshold_small | blocks(k+1) < threshold_small
            ROI = indices(k);
            ROIs = [ROIs, ROI];
            ind_ROIs = [ind_ROIs, k];
        end
    end
end

if ROIs(1) ~= 0
    ROIs = [0,ROIs];
end

if ind_ROIs(1) ~= 0
    ind_ROIs = [0,ind_ROIs];
end

%sort into types of code, 1-4

data_store = [];
sd_store = [];
for m = 1:size(ROIs,2)-1
    %for 1, 3, 5 etc. - horizonal, avg over x-val
    section_indices = [];
    if rem(m,2) == 1
        avgs = zeros(1,size(section_indices,2));

        %take out just the data for this section
        section_data = tbl((ROIs(m)+1):(ROIs(m+1)),:);
        section_data = sortrows(section_data, 'xpos');
        sec_xpos = section_data{:,1};
        sec_ypos = section_data{:,2};
        sec_values = section_data{:,3};

        %define # of data points for each x-val
        sz_ind = ind_ROIs(m+1) - ind_ROIs(m);
        section_indices = zeros(1,sz_ind);
        for p = 1:ind_ROIs(m+1) - ind_ROIs(m)
            section_indices(p) = blockmat(p+ind_ROIs(m),1);
        end
        section_indices = [0,section_indices];
        %%

        %avg values of all data points at each x-val
        for n = 1:size(section_indices,2)-1
            starting_index = sum(section_indices(2:n))+1;
            ending_index = sum(section_indices(2:n+1));
            values_vec = sec_values(starting_index:ending_index);
            sum_section = sum(values_vec);
            avgs(n) = sum_section/section_indices(n+1);
            std_dev(n) = std(values_vec);
        end
        %added
        avgs = flip(avgs);
        %
        std_dev = rmmissing(std_dev);
        avg_sd = sum(std_dev)/length(std_dev);
        data_store = [data_store; avgs'];
        sd_store = [sd_store; avg_sd];
```

```
 end

 %for 2, 6, 10 (organize by y-val, descending)
if rem(m,4) == 2
     avgs = [];

     %take out just the data for this section
     section_data = tbl((ROIs(m)+1):(ROIs(m+1)),:);
     section_data = sortrows(section_data, 'ypos');
     sec_xpos = section_data{:,1};
     sec_ypos = section_data{:,2};
     sec_values = section_data{:,3};
     %%
     %create indices (# of data points for each y-val)
     inds_y = [];
     i_y = 1;
     for n = 1:size(sec_ypos,1)-1
         if sec_ypos(n) == sec_ypos(n+1)
             i_y = i_y + 1;
         else
             inds_y = [inds_y, i_y];
             i_y = 1;
         end
     end
     section_indices = [0,inds_y];

     for n = 1:size(section_indices,2)-1
          starting_index = sum(section_indices(2:n))+1;
          ending_index = sum(section_indices(2:n+1));
          values_vec = sec_values(starting_index:ending_index);
          sum_section = sum(values_vec);
          avgs(n) = sum_section/section_indices(n+1);
          std_dev(n) = std(values_vec);
     end
     std_dev = rmmissing(std_dev);
     avg_sd = sum(std_dev)/length(std_dev);
     data_store = [data_store; avgs'];
     sd_store = [sd_store; avg_sd];
   end

  %for 4, 8, 12 (organize by y-val, ascending
   if rem(m,4) == 0
     avgs = [];

     %take out just the data for this section
     section_data = tbl((ROIs(m)+1):(ROIs(m+1)),:);
     section_data = sortrows(section_data, 'ypos', 'descend');
     sec_xpos = section_data{:,1};
     sec_ypos = section_data{:,2};
     sec_values = section_data{:,3};
     %%
     %create indices (# of data points for each y-val)
     inds_y = [];
     i_y = 1;
     for n = 1:size(sec_ypos,1)-1
         if sec_ypos(n) == sec_ypos(n+1)
             i_y = i_y + 1;
         else
             inds_y = [inds_y, i_y];
             i_y = 1;
         end
     end
     section_indices = [0,inds_y];

     for n = 1:size(section_indices,2)-1
          starting_index = sum(section_indices(2:n))+1;
          ending_index = sum(section_indices(2:n+1));
          values_vec = sec_values(starting_index:ending_index);
          sum_section = sum(values_vec);
          avgs(n) = sum_section/section_indices(n+1);
          std_dev(n) = std(values_vec);
```

```
        end
        std_dev = rmmissing(std_dev);
        avg_sd = sum(std_dev)/length(std_dev);
        data_store = [data_store; avgs'];
        sd_store = [sd_store; avg_sd];
      end
end

x = [1:size(data_store,1)];
x = x.*1.5;
STDDEV = sum(sd_store)/length(sd_store);

product4_vec = data_store;
product4_stdev = STDDEV;

%%
%% Product 5

%calldata
data = importdata(product5);
lengthvar = length(data.textdata(:,1));

xpos = data.textdata(4:lengthvar,4);
ypos = data.textdata(4:lengthvar,5);
values = data.textdata(4:lengthvar,2);
xpos2 = zeros(lengthvar-3,1);
for i = 1:lengthvar-4
    str = cell2mat(xpos(i));
    if strlength(str) ~= 4
        str = strcat(num2str(0),str);
        xpos(i) = mat2cell(str,1);
    end
    num = str2num(str);
    xpos2(i) = num;
end
xpos = xpos2;
ypos = cell2mat(ypos);
ypos = str2num(ypos);
values = cell2mat(values);
values = str2num(values);

tbl = table(xpos,ypos,values);

%sort by x-val
tbl = sortrows(tbl,'xpos');
xpos = tbl{:,1};

%remove bad data
for l = 1:size(tbl,1)
    if xpos(l) == 0
        tbl(l,:) = [];
    end
end

xpos = tbl{:,1};
ypos = tbl{:,2};
values = tbl{:,3};

%%

block = 1;
blocks = [];
index = 0;
indices = [];
for j = 1:lengthvar-5
    if xpos(j) == xpos(j+1)
        block = block + 1;
    else
        index = j;
        indices = [indices;index];
        blocks = [blocks; block];
```

```
        block = 1;
    end
    blockmat = [blocks, indices];
end
%%

threshold_main = 4;
threshold_small = 14;
ROIs = [];
ind_ROIs = [];
for k = 1:size(blocks,1)-1
    if (abs(blocks(k) - blocks(k+1)) > threshold_main)
            if blocks(k) < threshold_small | blocks(k+1) < threshold_small
            ROI = indices(k);
            ROIs = [ROIs, ROI];
            ind_ROIs = [ind_ROIs, k];
        end
    end
end

if ROIs(1) ~= 0
    ROIs = [0,ROIs];
end

if ind_ROIs(1) ~= 0
    ind_ROIs = [0,ind_ROIs];
end

%sort into types of code, 1-4

data_store = [];
sd_store = [];
for m = 1:size(ROIs,2)-1
    %for 1, 3, 5 etc. - horizonal, avg over x-val
    section_indices = [];
    if rem(m,2) == 1
        avgs = zeros(1,size(section_indices,2));

        %take out just the data for this section
        section_data = tbl((ROIs(m)+1):(ROIs(m+1)),:);
        section_data = sortrows(section_data, 'xpos');
        sec_xpos = section_data{:,1};
        sec_ypos = section_data{:,2};
        sec_values = section_data{:,3};

        %define # of data points for each x-val
        sz_ind = ind_ROIs(m+1) - ind_ROIs(m);
        section_indices = zeros(1,sz_ind);
        for p = 1:ind_ROIs(m+1) - ind_ROIs(m)
            section_indices(p) = blockmat(p+ind_ROIs(m),1);
        end
        section_indices = [0,section_indices];
        %%

        %avg values of all data points at each x-val
        for n = 1:size(section_indices,2)-1
            starting_index = sum(section_indices(2:n))+1;
            ending_index = sum(section_indices(2:n+1));
            values_vec = sec_values(starting_index:ending_index);
            sum_section = sum(values_vec);
            avgs(n) = sum_section/section_indices(n+1);
            std_dev(n) = std(values_vec);
        end
        %added
        avgs = flip(avgs);
        %
        std_dev = rmmissing(std_dev);
        avg_sd = sum(std_dev)/length(std_dev);
        data_store = [data_store; avgs'];
        sd_store = [sd_store; avg_sd];
    end
```

```
 %for 2, 6, 10 (organize by y-val, descending)
if rem(m,4) == 2
     avgs = [];

     %take out just the data for this section
     section_data = tbl((ROIs(m)+1):(ROIs(m+1)),:);
     section_data = sortrows(section_data, 'ypos');
     sec_xpos = section_data{:,1};
     sec_ypos = section_data{:,2};
     sec_values = section_data{:,3};
     %%
     %create indices (# of data points for each y-val)
     inds_y = [];
     i_y = 1;
     for n = 1:size(sec_ypos,1)-1
         if sec_ypos(n) == sec_ypos(n+1)
             i_y = i_y + 1;
         else
             inds_y = [inds_y, i_y];
             i_y = 1;
         end
     end
     section_indices = [0,inds_y];

     for n = 1:size(section_indices,2)-1
          starting_index = sum(section_indices(2:n))+1;
          ending_index = sum(section_indices(2:n+1));
          values_vec = sec_values(starting_index:ending_index);
          sum_section = sum(values_vec);
          avgs(n) = sum_section/section_indices(n+1);
          std_dev(n) = std(values_vec);
     end
     std_dev = rmmissing(std_dev);
     avg_sd = sum(std_dev)/length(std_dev);
     data_store = [data_store; avgs'];
     sd_store = [sd_store; avg_sd];
   end

  %for 4, 8, 12 (organize by y-val, ascending
   if rem(m,4) == 0
     avgs = [];

     %take out just the data for this section
     section_data = tbl((ROIs(m)+1):(ROIs(m+1)),:);
     section_data = sortrows(section_data, 'ypos', 'descend');
     sec_xpos = section_data{:,1};
     sec_ypos = section_data{:,2};
     sec_values = section_data{:,3};
     %%
     %create indices (# of data points for each y-val)
     inds_y = [];
     i_y = 1;
     for n = 1:size(sec_ypos,1)-1
         if sec_ypos(n) == sec_ypos(n+1)
             i_y = i_y + 1;
         else
             inds_y = [inds_y, i_y];
             i_y = 1;
         end
     end
     section_indices = [0,inds_y];

     for n = 1:size(section_indices,2)-1
          starting_index = sum(section_indices(2:n))+1;
          ending_index = sum(section_indices(2:n+1));
          values_vec = sec_values(starting_index:ending_index);
          sum_section = sum(values_vec);
          avgs(n) = sum_section/section_indices(n+1);
          std_dev(n) = std(values_vec);
     end
```

```
        std_dev = rmmissing(std_dev);
        avg_sd = sum(std_dev)/length(std_dev);
        data_store = [data_store; avgs'];
        sd_store = [sd_store; avg_sd];
      end
end

x = [1:size(data_store,1)];
x = x.*1.5;
STDDEV = sum(sd_store)/length(sd_store);

product5_vec = data_store;
product5_stdev = STDDEV;
%%
%% substrate

%calldata
data = importdata(substrate);
lengthvar = length(data.textdata(:,1));

xpos = data.textdata(4:lengthvar,4);
ypos = data.textdata(4:lengthvar,5);
values = data.textdata(4:lengthvar,2);
xpos2 = zeros(lengthvar-3,1);
for i = 1:lengthvar-4
    str = cell2mat(xpos(i));
    if strlength(str) ~= 4
        str = strcat(num2str(0),str);
        xpos(i) = mat2cell(str,1);
    end
    num = str2num(str);
    xpos2(i) = num;
end
xpos = xpos2;
ypos = cell2mat(ypos);
ypos = str2num(ypos);
values = cell2mat(values);
values = str2num(values);

tbl = table(xpos,ypos,values);

%sort by x-val
tbl = sortrows(tbl,'xpos');
xpos = tbl{:,1};

%remove bad data
for l = 1:size(tbl,1)
    if xpos(l) == 0
        tbl(l,:) = [];
    end
end

xpos = tbl{:,1};
ypos = tbl{:,2};
values = tbl{:,3};

%%

block = 1;
blocks = [];
index = 0;
indices = [];
for j = 1:lengthvar-5
    if xpos(j) == xpos(j+1)
        block = block + 1;
    else
        index = j;
        indices = [indices;index];
        blocks = [blocks; block];
        block = 1;
    end
```

```
    blockmat = [blocks, indices];
end
%%

threshold_main = 4;
threshold_small = 14;
ROIs = [];
ind_ROIs = [];
for k = 1:size(blocks,1)-1
    if (abs(blocks(k) - blocks(k+1)) > threshold_main)
            if blocks(k) < threshold_small | blocks(k+1) < threshold_small
            ROI = indices(k);
            ROIs = [ROIs, ROI];
            ind_ROIs = [ind_ROIs, k];
        end
    end
end

if ROIs(1) ~= 0
    ROIs = [0,ROIs];
end

if ind_ROIs(1) ~= 0
    ind_ROIs = [0,ind_ROIs];
end

%sort into types of code, 1-4

data_store = [];
sd_store = [];
for m = 1:size(ROIs,2)-1
    %for 1, 3, 5 etc. - horizonal, avg over x-val
    section_indices = [];
    if rem(m,2) == 1
        avgs = zeros(1,size(section_indices,2));

        %take out just the data for this section
        section_data = tbl((ROIs(m)+1):(ROIs(m+1)),:);
        section_data = sortrows(section_data, 'xpos');
        sec_xpos = section_data{:,1};
        sec_ypos = section_data{:,2};
        sec_values = section_data{:,3};

        %define # of data points for each x-val
        sz_ind = ind_ROIs(m+1) - ind_ROIs(m);
        section_indices = zeros(1,sz_ind);
        for p = 1:ind_ROIs(m+1) - ind_ROIs(m)
            section_indices(p) = blockmat(p+ind_ROIs(m),1);
        end
        section_indices = [0,section_indices];
        %%

        %avg values of all data points at each x-val
        for n = 1:size(section_indices,2)-1
            starting_index = sum(section_indices(2:n))+1;
            ending_index = sum(section_indices(2:n+1));
            values_vec = sec_values(starting_index:ending_index);
            sum_section = sum(values_vec);
            avgs(n) = sum_section/section_indices(n+1);
            std_dev(n) = std(values_vec);
        end
        %added
        avgs = flip(avgs);
        %
        std_dev = rmmissing(std_dev);
        avg_sd = sum(std_dev)/length(std_dev);
        data_store = [data_store; avgs'];
        sd_store = [sd_store; avg_sd];
    end

    %for 2, 6, 10 (organize by y-val, descending)
```

```
if rem(m,4) == 2
     avgs = [];

     %take out just the data for this section
     section_data = tbl((ROIs(m)+1):(ROIs(m+1)),:);
     section_data = sortrows(section_data, 'ypos');
     sec_xpos = section_data{:,1};
     sec_ypos = section_data{:,2};
     sec_values = section_data{:,3};
     %%
     %create indices (# of data points for each y-val)
     inds_y = [];
     i_y = 1;
     for n = 1:size(sec_ypos,1)-1
         if sec_ypos(n) == sec_ypos(n+1)
             i_y = i_y + 1;
         else
             inds_y = [inds_y, i_y];
             i_y = 1;
         end
     end
     section_indices = [0,inds_y];

     for n = 1:size(section_indices,2)-1
          starting_index = sum(section_indices(2:n))+1;
          ending_index = sum(section_indices(2:n+1));
          values_vec = sec_values(starting_index:ending_index);
          sum_section = sum(values_vec);
          avgs(n) = sum_section/section_indices(n+1);
          std_dev(n) = std(values_vec);
     end
     std_dev = rmmissing(std_dev);
     avg_sd = sum(std_dev)/length(std_dev);
     data_store = [data_store; avgs'];
     sd_store = [sd_store; avg_sd];
   end

  %for 4, 8, 12 (organize by y-val, ascending
   if rem(m,4) == 0
     avgs = [];

     %take out just the data for this section
     section_data = tbl((ROIs(m)+1):(ROIs(m+1)),:);
     section_data = sortrows(section_data, 'ypos', 'descend');
     sec_xpos = section_data{:,1};
     sec_ypos = section_data{:,2};
     sec_values = section_data{:,3};
     %%
     %create indices (# of data points for each y-val)
     inds_y = [];
     i_y = 1;
     for n = 1:size(sec_ypos,1)-1
         if sec_ypos(n) == sec_ypos(n+1)
             i_y = i_y + 1;
         else
             inds_y = [inds_y, i_y];
             i_y = 1;
         end
     end
     section_indices = [0,inds_y];

     for n = 1:size(section_indices,2)-1
          starting_index = sum(section_indices(2:n))+1;
          ending_index = sum(section_indices(2:n+1));
          values_vec = sec_values(starting_index:ending_index);
          sum_section = sum(values_vec);
          avgs(n) = sum_section/section_indices(n+1);
          std_dev(n) = std(values_vec);
     end
     std_dev = rmmissing(std_dev);
     avg_sd = sum(std_dev)/length(std_dev);
```

```
        data_store = [data_store; avgs'];
        sd_store = [sd_store; avg_sd];
      end
end

x = [1:size(data_store,1)];
x = x.*1.5;
STDDEV = sum(sd_store)/length(sd_store);

substrate_vec = data_store;
substrate_stdev = STDDEV;
%%
%% Plot data

%create time vector, normalize data vectors
x = [1:size(data_store,1)-4];
x = x*time_vector_factor;

substrate_vec = substrate_max * substrate_vec;
product1_vec = product1_max * product1_vec;
product2_vec = product2_max * product2_vec;
product3_vec = product3_max * product3_vec;
product4_vec = product4_max * product4_vec;
product5_vec = product5_max * product5_vec;
internalstandard_vec = internalstandard_max * internalstandard_vec;

% subtracting noise
avg_noise = mean(noise_vec);

substrate_vec = substrate_vec - avg_noise;
product1_vec = product1_vec - avg_noise;
product2_vec = product2_vec - avg_noise;
product3_vec = product3_vec - avg_noise;
product4_vec = product4_vec - avg_noise;
product5_vec = product5_vec - avg_noise;
internalstandard_vec = internalstandard_vec - avg_noise;

%This removes the first 4 rows of data points. I don't know why this is
%here but the code doesn't work without it.
substrate_vec = substrate_vec(5:length(substrate_vec));
product1_vec = product1_vec(5:length(product1_vec));
product2_vec = product2_vec(5:length(product2_vec));
product3_vec = product3_vec(5:length(product3_vec));
product4_vec = product4_vec(5:length(product4_vec));
product5_vec = product5_vec(5:length(product5_vec));
internalstandard_vec = internalstandard_vec(5:length(internalstandard_vec));

% normalizing the points to the EG3 peak
substrate_norm = substrate_vec./(internalstandard_vec+substrate_vec);
prod1_norm = product1_vec./(internalstandard_vec+product1_vec);
prod2_norm = product2_vec./(internalstandard_vec+product2_vec);
prod3_norm = product3_vec./(internalstandard_vec+product3_vec);
prod4_norm = product4_vec./(internalstandard_vec+product4_vec);
prod5_norm = product5_vec./(internalstandard_vec+product5_vec);



%% Plot raw + normalized data

% raw data
figure(1)
plot(x,substrate_vec,'.','Color',[0.67,0,0]);
title({'Raw Data: Substrate'},'fontsize', 30);
xlabel('Time (min)','fontsize',20);
ylabel('AUC','fontsize',20);

figure(2)
plot(x,product1_vec,'.','Color',[0,0.67,0]);
title({'Raw Data: Product 1'},'fontsize', 30);
xlabel('Time (min)','fontsize',20);
ylabel('AUC','fontsize',20);
```

```
figure(3)
plot(x,product2_vec,'.','Color',[0,0,0.67]);
title({'Raw Data: Product 2'},'fontsize', 30);
xlabel('Time (min)','fontsize',20);
ylabel('AUC','fontsize',20);

figure(4)
plot(x,product3_vec,'.','Color',[0.67,0.33,0]);
title({'Raw Data: Product 3'},'fontsize', 30);
xlabel('Time (min)','fontsize',20);
ylabel('AUC','fontsize',20);

figure(5)
plot(x,product4_vec,'.','Color',[0,0.67,0.33]);
title({'Raw Data: Product 4'},'fontsize', 30);
xlabel('Time (min)','fontsize',20);
ylabel('AUC','fontsize',20);

figure(6)
plot(x,product5_vec,'.','Color',[0.33,0,0.67]);
title({'Raw Data: Product 5'},'fontsize', 30);
xlabel('Time (min)','fontsize',20);
ylabel('AUC','fontsize',20);

figure(7)
plot(x,internalstandard_vec,'.','Color','r');
title({'Raw Data: Internal Standard'},'fontsize', 30);
xlabel('Time (min)','fontsize',20);
ylabel('AUC','fontsize',20);

figure(8)
hold on
plot(x,substrate_vec,'.','Color',[0.67,0,0])
plot(x,product1_vec,'.','Color',[0,0.67,0])
plot(x,product2_vec,'.','Color',[0,0,0.67])
plot(x,product3_vec,'.','Color',[0.67,0.33,0])
plot(x,product4_vec,'.','Color',[0,0.67,0.33])
plot(x,product5_vec,'.','Color',[0.33,0,0.67])
title({'Raw Data: All Species'},'fontsize', 30);
xlabel('Time (min)','fontsize',20);
ylabel('% Present','fontsize',20);
hold off
legend('MIASGC','IASGC','ASGC','SGC','GC','C');

%% Smoothed plot (not normalized)

smoothed_substrate = smooth(substrate_vec,0.25);
smoothed_prod1 = smooth(product1_vec,0.25);
smoothed_prod2 = smooth(product2_vec,0.25);
smoothed_prod3 = smooth(product3_vec,0.25);
smoothed_prod4 = smooth(product4_vec,0.25);
smoothed_prod5 = smooth(product5_vec,0.25);

figure(9)
hold on
plot(x,smoothed_substrate,'.','Color',[0.67,0,0])
plot(x,smoothed_prod1,'.','Color',[0,0.67,0])
plot(x,smoothed_prod2,'.','Color',[0,0,0.67])
plot(x,smoothed_prod3,'.','Color',[0.67,0.33,0])
plot(x,smoothed_prod4,'.','Color',[0,0.67,0.33])
plot(x,smoothed_prod5,'.','Color',[0.33,0,0.67])
title({'Raw Smooth Data: All Species'},'fontsize', 30);
xlabel('Time (min)','fontsize',20);
ylabel('% Present','fontsize',20);
hold off
legend('MIASGC','IASGC','ASGC','SGC','GC','C');

%% normalized data
figure(10)
plot(x,substrate_norm,'.','Color',[0.67,0,0]);
```

```
title({'Normalized Data: Substrate'},'fontsize', 30);
xlabel('Time (min)','fontsize',20);
ylabel('% Present','fontsize',20);

figure(11)
plot(x,prod1_norm,'.','Color',[0,0.67,0]);
title({'Normalized Data: Product 1'},'fontsize', 30);
xlabel('Time (min)','fontsize',20);
ylabel('% Present','fontsize',20);

figure(12)
plot(x,prod2_norm,'.','Color',[0,0,0.67]);
title({'Normalized Data: Product 2'},'fontsize', 30);
xlabel('Time (min)','fontsize',20);
ylabel('% Present','fontsize',20);

figure(13)
plot(x,prod3_norm,'.','Color',[0.67,0.33,0]);
title({'Normalized Data: Product 3'},'fontsize', 30);
xlabel('Time (min)','fontsize',20);
ylabel('% Present','fontsize',20);

figure(14)
plot(x,prod4_norm,'.','Color',[0,0.67,0.33]);
title({'Normalized Data: Product 4'},'fontsize', 30);
xlabel('Time (min)','fontsize',20);
ylabel('% Present','fontsize',20);

figure(15)
plot(x,prod5_norm,'.','Color',[0.33,0,0.67]);
title({'Normalized Data: Product 5'},'fontsize', 30);
xlabel('Time (min)','fontsize',20);
ylabel('% Present','fontsize',20);

figure(16)
hold on
plot(x,substrate_norm,'.','Color',[0.67,0,0])
plot(x,prod1_norm,'.','Color',[0,0.67,0])
plot(x,prod2_norm,'.','Color',[0,0,0.67])
plot(x,prod3_norm,'.','Color',[0.67,0.33,0])
plot(x,prod4_norm,'.','Color',[0,0.67,0.33])
plot(x,prod5_norm,'.','Color',[0.33,0,0.67])
title({'Normalized Data: All Species'},'fontsize', 30);
xlabel('Time (min)','fontsize',20);
ylabel('% Present','fontsize',20);
hold off
legend('MIASGC','IASGC','ASGC','SGC','GC','C');

%% Smoothed plot

smoothed_substrate = smooth(substrate_norm,0.25);
smoothed_prod1 = smooth(prod1_norm,0.25);
smoothed_prod2 = smooth(prod2_norm,0.25);
smoothed_prod3 = smooth(prod3_norm,0.25);
smoothed_prod4 = smooth(prod4_norm,0.25);
smoothed_prod5 = smooth(prod5_norm,0.25);

figure(17)
hold on
plot(x,smoothed_substrate,'.','Color',[0.67,0,0])
plot(x,smoothed_prod1,'.','Color',[0,0.67,0])
plot(x,smoothed_prod2,'.','Color',[0,0,0.67])
plot(x,smoothed_prod3,'.','Color',[0.67,0.33,0])
plot(x,smoothed_prod4,'.','Color',[0,0.67,0.33])
plot(x,smoothed_prod5,'.','Color',[0.33,0,0.67])
title({'Normalized Data: All Species'},'fontsize', 30);
xlabel('Time (min)','fontsize',20);
ylabel('% Present','fontsize',20);
hold off
legend('MIASGC','IASGC','ASGC','SGC','GC','C');
```

```
%%
%%
%% SimBiology Data Processing

%removing front and back end periodicity
vector_length = length(smoothed_substrate);
three_percent_vector_length = 0.03*vector_length;
three_percent_vector_length = round(three_percent_vector_length)

cut_substrate = substrate_norm(three_percent_vector_length+1:end-three_percent_vector_length);
cut_prod1 = prod1_norm(three_percent_vector_length+1:end-three_percent_vector_length);
cut_prod2 = prod2_norm(three_percent_vector_length+1:end-three_percent_vector_length);
cut_prod3 = prod3_norm(three_percent_vector_length+1:end-three_percent_vector_length);
cut_prod4 = prod4_norm(three_percent_vector_length+1:end-three_percent_vector_length);
cut_prod5 = prod5_norm(three_percent_vector_length+1:end-three_percent_vector_length);

cut_smoothed_substrate = smoothed_substrate(three_percent_vector_length+1:end-
three_percent_vector_length);
cut_smoothed_prod1 = smoothed_prod1(three_percent_vector_length+1:end-
three_percent_vector_length);
cut_smoothed_prod2 = smoothed_prod2(three_percent_vector_length+1:end-
three_percent_vector_length);
cut_smoothed_prod3 = smoothed_prod3(three_percent_vector_length+1:end-
three_percent_vector_length);
cut_smoothed_prod4 = smoothed_prod4(three_percent_vector_length+1:end-
three_percent_vector_length);
cut_smoothed_prod5 = smoothed_prod5(three_percent_vector_length+1:end-
three_percent_vector_length);

cutx = x(three_percent_vector_length+1:end-three_percent_vector_length);

figure(18)
hold on
plot(cutx,cut_smoothed_substrate,'.','Color',[0.67,0,0])
plot(cutx,cut_smoothed_prod1,'.','Color',[0,0.67,0])
plot(cutx,cut_smoothed_prod2,'.','Color',[0,0,0.67])
plot(cutx,cut_smoothed_prod3,'.','Color',[0.67,0.33,0])
plot(cutx,cut_smoothed_prod4,'.','Color',[0,0.67,0.33])
plot(cutx,cut_smoothed_prod5,'.','Color',[0.33,0,0.67])
title({'Cut Data Smoothed: All Species'},'fontsize', 30);
xlabel('Time (min)','fontsize',20);
ylabel('% Present','fontsize',20);
hold off
legend('MIASGC','IASGC','ASGC','SGC','GC','C');

figure(19)
hold on
plot(cutx,cut_substrate,'.','Color',[0.67,0,0])
plot(cutx,cut_prod1,'.','Color',[0,0.67,0])
plot(cutx,cut_prod2,'.','Color',[0,0,0.67])
plot(cutx,cut_prod3,'.','Color',[0.67,0.33,0])
plot(cutx,cut_prod4,'.','Color',[0,0.67,0.33])
plot(cutx,cut_prod5,'.','Color',[0.33,0,0.67])
title({'Cut Data: All Species'},'fontsize', 30);
xlabel('Time (min)','fontsize',20);
ylabel('% Present','fontsize',20);
hold off
legend('MIASGC','IASGC','ASGC','SGC','GC','C');

table1 =[x', smoothed_substrate, smoothed_prod1, smoothed_prod2, smoothed_prod3, smoothed_prod4,
smoothed_prod5];
table2 =[cutx', cut_smoothed_substrate, cut_smoothed_prod1, cut_smoothed_prod2,
cut_smoothed_prod3, cut_smoothed_prod4, cut_smoothed_prod5];
tablea =[cutx', cut_substrate, cut_prod1, cut_prod2, cut_prod3, cut_prod4, cut_prod5];
% %These tables are required for SimBiology. Note, data can only be added
% %to the program once. Everything needs to be contained in one table.
% table1 =[x', smoothed_substrate, smoothed_prod1, smoothed_prod2, smoothed_prod3,
smoothed_prod4, smoothed_prod5];
```

References

(1) Asyuda, A.; Das, S.; Zharnikov, M. Thermal Stability of Alkanethiolate and Aromatic Thiolate Self-Assembled Monolayers on Au(111): An X-Ray Photoelectron Spectroscopy Study. *J. Phys. Chem. C* **2021**, *125* (39), 21754–21763. https://doi.org/10.1021/acs.jpcc.1c06984.

(2) Self-Assembled Organic Monolayers: Model Systems for Studying Adsorption of Proteins at Aurfaces https://www.science.org/doi/10.1126/science.252.5009.1164 (accessed 2022 -01 -03).

(3) Gurard-Levin, Z. A.; Mrksich, M. Combining Self-Assembled Monolayers and Mass Spectrometry for Applications in Biochips. *Annu. Rev. Anal. Chem.* **2008**, *1* (1), 767–800. https://doi.org/10.1146/annurev.anchem.1.031207.112903.

(4) Houseman, B. T.; Huh, J. H.; Kron, S. J.; Mrksich, M. Peptide Chips for the Quantitative Evaluation of Protein Kinase Activity. *Nat. Biotechnol.* **2002**, *20* (3), 270–274. https://doi.org/10.1038/nbt0302-270.

(5) Marin, V. L.; Bayburt, T. H.; Sligar, S. G.; Mrksich, M. Functional Assays of Membrane-Bound Proteins with SAMDI-TOF Mass Spectrometry. *Angew. Chem. Int. Ed Engl.* **2007**, *46* (46), 8796–8798. https://doi.org/10.1002/anie.200702694.

(6) Houseman, B. T.; Mrksich, M. Towards Quantitative Assays with Peptide Chips: A Surface Engineering Approach. *Trends Biotechnol.* **2002**, *20* (7), 279–281. https://doi.org/10.1016/S0167-7799(02)01984-4.

(7) Webster, J.; Oxley, D. Protein Identification by MALDI-TOF Mass Spectrometry. *Methods Mol. Biol. Clifton NJ* **2012**, *800*, 227–240. https://doi.org/10.1007/978-1-61779-349-3_15.

(8) Cabrera-Pardo, J. R.; Chai, D. I.; Liu, S.; Mrksich, M.; Kozmin, S. A. Label-Assisted Mass Spectrometry for the Acceleration of Reaction Discovery and Optimization. *Nat. Chem.* **2013**, *5* (5), 423–427. https://doi.org/10.1038/nchem.1612.

(9) Yergey, A. L.; Coorssen, J. R.; Backlund, P. S.; Blank, P. S.; Humphrey, G. A.; Zimmerberg, J.; Campbell, J. M.; Vestal, M. L. De Novo Sequencing of Peptides Using MALDI/TOF-TOF. *J. Am. Soc. Mass Spectrom.* **2002**, *13* (7), 784–791. https://doi.org/10.1016/S1044-0305(02)00393-8.

(10) Milan Mrksich. Mass Spectrometry of Self-Assembled Monolayers: A New Tool for Molecular Surface Science. *ACS Nano* **2008**, *2* (1), 7–18. https://doi.org/10.1021/nn7004156.

(11) Su, J.; Mrksich, M. Using Mass Spectrometry to Characterize Self-Assembled Monolayers Presenting Peptides, Proteins, and Carbohydrates. *Angew. Chem. Int. Ed. 41* (24), 4715–4718. https://doi.org/10.1002/anie.200290026.

(12) Gurard-Levin, Z. A.; Scholle, M. D.; Eisenberg, A. H.; Mrksich, M. High-Throughput Screening of Small Molecule Libraries Using SAMDI Mass Spectrometry. *ACS Comb. Sci.* **2011**, *13* (4), 347–350. https://doi.org/10.1021/co2000373.

(13) Diagne, A. B.; Li, S.; Perkowski, G. A.; Mrksich, M.; Thomson, R. J. SAMDI Mass Spectrometry-Enabled High-Throughput Optimization of a Traceless Petasis Reaction. *ACS Comb. Sci.* **2015**, *17* (11), 658–662. https://doi.org/10.1021/acscombsci.5b00131.

(14) Ban, L.; Pettit, N.; Li, L.; Stuparu, A. D.; Cai, L.; Chen, W.; Guan, W.; Han, W.; Wang, P. G.; Mrksich, M. Discovery of Glycosyltransferases Using Carbohydrate Arrays and Mass Spectrometry. *Nat. Chem. Biol.* **2012**, *8* (9), 769–773. https://doi.org/10.1038/nchembio.1022.

(15) Kornacki, J. R.; Stuparu, A. D.; Mrksich, M. Acetyltransferase P300/CBP Associated Factor (PCAF) Regulates Crosstalk-Dependent Acetylation of Histone H3 by Distal Site Recognition. *ACS Chem. Biol.* **2015**, *10* (1), 157–164. https://doi.org/10.1021/cb5004527.

(16) Wood, S. E.; Sinsinbar, G.; Gudlur, S.; Nallani, M.; Huang, C.-F.; Liedberg, B.; Mrksich, M. A Bottom-Up Proteomic Approach to Identify Substrate Specificity of Outer-Membrane Protease OmpT. *Angew. Chem.* **2017**, *129* (52), 16758–16762. https://doi.org/10.1002/ange.201707535.

(17) Liao, X.; Su, J.; Mrksich, M. An Adaptor Domain-Mediated Auto-Catalytic Interfacial Kinase Reaction. *Chem. Weinh. Bergstr. Ger.* **2009**, *15* (45), 12303–12309. https://doi.org/10.1002/chem.200901345.

(18) Gurard-Levin, Z. A.; Mrksich, M. The Activity of HDAC8 Depends on Local and Distal Sequences of Its Peptide Substrates. *Biochemistry* **2008**, *47* (23), 6242–6250. https://doi.org/10.1021/bi800053v.

(19) Szymczak, L. C.; Sykora, D. J.; Mrksich, M. Using Peptide Arrays to Profile Phosphatase Activity in Cell Lysates. *Chem. – Eur. J.* **2020**, *26* (1), 165–170. https://doi.org/10.1002/chem.201904364.

(20) Min, D.-H.; Su, J.; Mrksich, M. Profiling Kinase Activities by Using a Peptide Chip and Mass Spectrometry. *Angew. Chem. Int. Ed.* **2004**, *43* (44), 5973–5977. https://doi.org/10.1002/anie.200461061.

(21) Kuo, H.-Y.; DeLuca, T. A.; Miller, W. M.; Mrksich, M. Profiling Deacetylase Activities in Cell Lysates with Peptide Arrays and SAMDI Mass Spectrometry. *Anal. Chem.* **2013**, *85* (22), 10635–10642. https://doi.org/10.1021/ac402614x.

(22) Min, D.-H.; Tang, W.-J.; Mrksich, M. Chemical Screening by Mass Spectrometry to Identify Inhibitors of Anthrax Lethal Factor. *Nat. Biotechnol.* **2004**, *22* (6), 717–723. https://doi.org/10.1038/nbt973.

(23) Patel, K.; Sherrill, J.; Mrksich, M.; Scholle, M. D. Discovery of SIRT3 Inhibitors Using SAMDI Mass Spectrometry. *J. Biomol. Screen.* **2015**, *20* (7), 842–848. https://doi.org/10.1177/1087057115588512.

(24) E. Anderson, S.; S. Fahey, N.; Park, J.; T. O'Kane, P.; A. Mirkin, C.; Mrksich, M. A High-Throughput SAMDI-Mass Spectrometry Assay for Isocitrate Dehydrogenase 1. *Analyst* **2020**, *145* (11), 3899–3908. https://doi.org/10.1039/D0AN00174K.

(25) Anderson, L. L.; Berns, E. J.; Bugga, P.; George, A. L.; Mrksich, M. Measuring Drug Metabolism Kinetics and Drug–Drug Interactions Using Self-Assembled Monolayers for Matrix-Assisted Laser Desorption-Ionization Mass Spectrometry. *Anal. Chem.* **2016**, *88* (17), 8604–8609. https://doi.org/10.1021/acs.analchem.6b01750.

(26) Bayly, A. A.; McDonald, B. R.; Mrksich, M.; Scheidt, K. A. High-Throughput Photocapture Approach for Reaction Discovery. *Proc. Natl. Acad. Sci.* **2020**, *117* (24), 13261–13266. https://doi.org/10.1073/pnas.2003347117.

(27) Montavon, T. J.; Li, J.; Cabrera-Pardo, J. R.; Mrksich, M.; Kozmin, S. A. Three-Component Reaction Discovery Enabled by Mass Spectrometry of Self-Assembled Monolayers. *Nat. Chem.* **2012**, *4* (1), 45–51. https://doi.org/10.1038/nchem.1212.

(28) Berns, E. J.; Cabezas, M. D.; Mrksich, M. Cellular Assays with a Molecular Endpoint Measured by SAMDI Mass Spectrometry. *Small Weinh. Bergstr. Ger.* **2016**, *12* (28), 3811–3818. https://doi.org/10.1002/smll.201502940.

(29) Moully, E. H.; Berns, E. J.; Mrksich, M. Label-Free Assay of Protein Tyrosine Phosphatase Activity in Single Cells. *Anal. Chem.* **2019**. https://doi.org/10.1021/acs.analchem.9b03640.

(30) Mukherjee, P.; Berns, E. J.; Patino, C. A.; Hakim Moully, E.; Chang, L.; Nathamgari, S. S. P.; Kessler, J. A.; Mrksich, M.; Espinosa, H. D. Temporal Sampling of Enzymes from Live Cells by Localized Electroporation and Quantification of Activity by SAMDI Mass Spectrometry. *Small* **2020**, *16* (26), 2000584. https://doi.org/10.1002/smll.202000584.

(31) Yamankurt, G.; Berns, E. J.; Xue, A.; Lee, A.; Bagheri, N.; Mrksich, M.; Mirkin, C. A. Exploration of the Nanomedicine-Design Space with High-Throughput Screening and Machine Learning. *Nat. Biomed. Eng.* **2019**, *3* (4), 318–327. https://doi.org/10.1038/s41551-019-0351-1.

(32) Cafferty, B. J.; Ten, A. S.; Fink, M. J.; Morey, S.; Preston, D. J.; Mrksich, M.; Whitesides, G. M. Storage of Information Using Small Organic Molecules. *ACS Cent. Sci.* **2019**, *5* (5), 911–916. https://doi.org/10.1021/acscentsci.9b00210.

(33) O'Kane, P. T.; Mrksich, M. An Assay Based on SAMDI Mass Spectrometry for Profiling Protein Interaction Domains. *J. Am. Chem. Soc.* **2017**, *139* (30), 10320–10327. https://doi.org/10.1021/jacs.7b03805.

(34) Stone, M. J.; Williams, D. H. On the Evolution of Functional Secondary Metabolites (Natural Products). *Mol. Microbiol.* **1992**, *6* (1), 29–34. https://doi.org/10.1111/j.1365-2958.1992.tb00834.x.

(35) Toprak, E.; Veres, A.; Michel, J.-B.; Chait, R.; Hartl, D. L.; Kishony, R. Evolutionary Paths to Antibiotic Resistance under Dynamically Sustained Drug Selection. *Nat. Genet.* **2012**, *44* (1), 101–105. https://doi.org/10.1038/ng.1034.

(36) Luzzatto, L. Sickle Cell Anaemia and Malaria. *Mediterr. J. Hematol. Infect. Dis.* **2012**, *4* (1), e2012065. https://doi.org/10.4084/MJHID.2012.065.

(37) Shriner, D.; Rotimi, C. N. Whole-Genome-Sequence-Based Haplotypes Reveal Single Origin of the Sickle Allele during the Holocene Wet Phase. *Am. J. Hum. Genet.* **2018**, *102* (4), 547–556. https://doi.org/10.1016/j.ajhg.2018.02.003.

(38) Bartlett, M. D.; Crosby, A. J. High Capacity, Easy Release Adhesives From Renewable Materials. *Adv. Mater.* **2014**, *26* (21), 3405–3409. https://doi.org/10.1002/adma.201305593.

(39) Hassabis, D.; Kumaran, D.; Summerfield, C.; Botvinick, M. Neuroscience-Inspired Artificial Intelligence. *Neuron* **2017**, *95* (2), 245–258. https://doi.org/10.1016/j.neuron.2017.06.011.

(40) Goldsmith, M.; Tawfik, D. S. Directed Enzyme Evolution: Beyond the Low-Hanging Fruit. *Curr. Opin. Struct. Biol.* **2012**, *22* (4), 406–412. https://doi.org/10.1016/j.sbi.2012.03.010.

(41) Arnold, F. H.; Moore, J. C. Optimizing Industrial Enzymes by Directed Evolution. *Adv. Biochem. Eng. Biotechnol.* **1997**, *58*, 1–14.

(42) Schmid, A.; Dordick, J. S.; Hauer, B.; Kiener, A.; Wubbolts, M.; Witholt, B. Industrial Biocatalysis Today and Tomorrow. *Nature* **2001**, *409* (6817), 258–268. https://doi.org/10.1038/35051736.

(43) Cobb, R. E.; Chao, R.; Zhao, H. Directed Evolution: Past, Present, and Future. *AIChE J.* **2013**, *59* (5), 1432–1440. https://doi.org/10.1002/aic.13995.

(44) Cobb, R. E.; Si, T.; Zhao, H. Directed Evolution: An Evolving and Enabling Synthetic Biology Tool. *Curr. Opin. Chem. Biol.* **2012**, *16* (3), 285–291. https://doi.org/10.1016/j.cbpa.2012.05.186.

(45) Kumar, A.; Singh, S. Directed Evolution: Tailoring Biocatalysts for Industrial Applications. *Crit. Rev. Biotechnol.* **2013**, *33* (4), 365–378. https://doi.org/10.3109/07388551.2012.716810.

(46) Davis, A. M.; Plowright, A. T.; Valeur, E. Directing Evolution: The next Revolution in Drug Discovery? *Nat. Rev. Drug Discov.* **2017**, *16* (10), 681–698. https://doi.org/10.1038/nrd.2017.146.

(47) Prier, C. K.; Arnold, F. H. Chemomimetic Biocatalysis: Exploiting the Synthetic Potential of Cofactor-Dependent Enzymes To Create New Catalysts. *J. Am. Chem. Soc.* **2015**, *137* (44), 13992–14006. https://doi.org/10.1021/jacs.5b09348.

(48) Sen, S.; Dasu, V. V.; Mandal, B. Developments in Directed Evolution for Improving Enzyme Functions. *Appl. Biochem. Biotechnol.* **2007**, *143* (3), 212–223. https://doi.org/10.1007/s12010-007-8003-4.

(49) Coelho, P. S.; Wang, Z. J.; Ener, M. E.; Baril, S. A.; Kannan, A. A.; Arnold, F. H.; Brustad, E. M. A Serine-Substituted P450 Catalyzes Highly Efficient Carbene Transfer to Olefins In Vivo. *Nat. Chem. Biol.* **2013**, *9* (8), 485–487. https://doi.org/10.1038/nchembio.1278.

(50) Wang, Z. J.; Renata, H.; Peck, N. E.; Farwell, C. C.; Coelho, P. S.; Arnold, F. H. Improved Cyclopropanation Activity of Histidine-Ligated Cytochrome P450 Enables Enantioselective Formal Synthesis of Levomilnacipran. *Angew. Chem. Int. Ed Engl.* **2014**, *53* (26), 6810–6813. https://doi.org/10.1002/anie.201402809.

(51) Hernandez, K. E.; Renata, H.; Lewis, R. D.; Kan, S. B. J.; Zhang, C.; Forte, J.; Rozzell, D.; McIntosh, J. A.; Arnold, F. H. Highly Stereoselective Biocatalytic Synthesis of Key Cyclopropane Intermediate to Ticagrelor. *ACS Catal.* **2016**, *6* (11), 7810–7813. https://doi.org/10.1021/acscatal.6b02550.

(52) Bajaj, P.; Sreenilayam, G.; Tyagi, V.; Fasan, R. Gram-Scale Synthesis of Chiral Cyclopropane-Containing Drugs and Drug Precursors with Engineered Myoglobin Catalysts Featuring Complementary Stereoselectivity. *Angew. Chem. Int. Ed Engl.* **2016**, *55* (52), 16110–16114. https://doi.org/10.1002/anie.201608680.

(53) McIntosh, J. A.; Coelho, P. S.; Farwell, C. C.; Wang, Z. J.; Lewis, J. C.; Brown, T. R.; Arnold, F. H. Enantioselective Intramolecular C□H Amination Catalyzed by Engineered Cytochrome P450 Enzymes In Vitro and In Vivo. *Angew. Chem. 125* (35), 9479–9482. https://doi.org/10.1002/ange.201304401.

(54) Xiao, H.; Bao, Z.; Zhao, H. High Throughput Screening and Selection Methods for Directed Enzyme Evolution. *Ind. Eng. Chem. Res.* **2015**, *54* (16), 4011–4020. https://doi.org/10.1021/ie503060a.

(55) McLachlan, M. J.; Sullivan, R. P.; Zhao, H. Directed Enzyme Evolution and High-Throughput Screening. In *Biocatalysis for the Pharmaceutical Industry*; Tao, J. (Alex), Lin, G.-Q., Liese, A., Eds.; John Wiley & Sons, Ltd: Chichester, UK, 2009; pp 45–64. https://doi.org/10.1002/9780470823163.ch3.

(56) Schaerli, Y.; Hollfelder, F. The Potential of Microfluidic Water-in-Oil Droplets in Experimental Biology. *Mol. Biosyst.* **2009**, *5* (12), 1392–1404. https://doi.org/10.1039/B907578J.

(57) Rockah-Shmuel, L.; Tawfik, D. S.; Goldsmith, M. Generating Targeted Libraries by the Combinatorial Incorporation of Synthetic Oligonucleotides During Gene Shuffling (ISOR). In *Directed Evolution Library Creation*; Methods in Molecular Biology; Springer, New York, NY, 2014; pp 129–137. https://doi.org/10.1007/978-1-4939-1053-3_8.

(58) Yang, G.; Withers, S. G. Ultrahigh-Throughput FACS-Based Screening for Directed Enzyme Evolution. *ChemBioChem* **2009**, *10* (17), 2704–2715. https://doi.org/10.1002/cbic.200900384.

(59) Zhang, R. K.; Chen, K.; Huang, X.; Wohlschlager, L.; Renata, H.; Arnold, F. H. Enzymatic Assembly of Carbon–Carbon Bonds via Iron-Catalysed Sp 3 C–H Functionalization. *Nature* **2019**, *565* (7737), 67–72. https://doi.org/10.1038/s41586-018-0808-5.

(60) Hoebenreich, S.; Zilly, F. E.; Acevedo-Rocha, C. G.; Zilly, M.; Reetz, M. T. Speeding up Directed Evolution: Combining the Advantages of Solid-Phase Combinatorial Gene Synthesis with Statistically Guided Reduction of Screening Effort. *ACS Synth. Biol.* **2015**, *4* (3), 317–331. https://doi.org/10.1021/sb5002399.

(61) Diefenbach, X. W.; Farasat, I.; Guetschow, E. D.; Welch, C. J.; Kennedy, R. T.; Sun, S.; Moore, J. C. Enabling Biocatalysis by High-Throughput Protein Engineering Using Droplet

Microfluidics Coupled to Mass Spectrometry. *ACS Omega* **2018**, *3* (2), 1498–1508. https://doi.org/10.1021/acsomega.7b01973.

(62) Dalby, P. A. Strategy and Success for the Directed Evolution of Enzymes. *Curr. Opin. Struct. Biol.* **2011**, *21* (4), 473–480. https://doi.org/10.1016/j.sbi.2011.05.003.

(63) Duetz, W. A. Microtiter Plates as Mini-Bioreactors: Miniaturization of Fermentation Methods. *Trends Microbiol.* **2007**, *15* (10), 469–475. https://doi.org/10.1016/j.tim.2007.09.004.

(64) Weis, R.; Luiten, R.; Skranc, W.; Schwab, H.; Wubbolts, M.; Glieder, A. Reliable High-Throughput Screening with Pichia Pastoris by Limiting Yeast Cell Death Phenomena. *FEMS Yeast Res.* **2004**, *5* (2), 179–189. https://doi.org/10.1016/j.femsyr.2004.06.016.

(65) Behrendorff, J. B.; Vickers, C. E.; Chrysanthopoulos, P.; Nielsen, L. K. 2,2-Diphenyl-1-Picrylhydrazyl as a Screening Tool for Recombinant Monoterpene Biosynthesis. *Microb. Cell Factories* **2013**, *12* (1), 76. https://doi.org/10.1186/1475-2859-12-76.

(66) Kim, T.-W.; Chokhawala, H. A.; Hess, M.; Dana, C. M.; Baer, Z.; Sczyrba, A.; Rubin, E. M.; Blanch, H. W.; Clark, D. S. High-Throughput In Vitro Glycoside Hydrolase (HIGH) Screening for Enzyme Discovery. *Angew. Chem. Int. Ed. 50* (47), 11215–11218. https://doi.org/10.1002/anie.201104685.

(67) Gielen, F.; Hours, R.; Emond, S.; Fischlechner, M.; Schell, U.; Hollfelder, F. Ultrahigh-Throughput–Directed Enzyme Evolution by Absorbance-Activated Droplet Sorting (AADS). *Proc. Natl. Acad. Sci.* **2016**, *113* (47), E7383–E7389. https://doi.org/10.1073/pnas.1606927113.

(68) Becker, S.; Höbenreich, H.; Vogel, A.; Knorr, J.; Wilhelm, S.; Rosenau, F.; Jaeger, K.-E.; Reetz, M. T.; Kolmar, H. Single-Cell High-Throughput Screening To Identify Enantioselective Hydrolytic Enzymes. *Angew. Chem. Int. Ed. 47* (27), 5085–5088. https://doi.org/10.1002/anie.200705236.

(69) Aharoni, A.; Thieme, K.; Chiu, C. P. C.; Buchini, S.; Lairson, L. L.; Chen, H.; Strynadka, N. C. J.; Wakarchuk, W. W.; Withers, S. G. High-Throughput Screening Methodology for the Directed Evolution of Glycosyltransferases. *Nat. Methods* **2006**, *3* (8), 609–614. https://doi.org/10.1038/nmeth899.

(70) Leemhuis Hans; Kelly Ronan M.; Dijkhuizen Lubbert. Directed Evolution of Enzymes: Library Screening Strategies. *IUBMB Life* **2009**, *61* (3), 222–228. https://doi.org/10.1002/iub.165.

(71) Sciambi, A.; R. Abate, A. Accurate Microfluidic Sorting of Droplets at 30 KHz. *Lab. Chip* **2015**, *15* (1), 47–51. https://doi.org/10.1039/C4LC01194E.

(72) Borra, M. T.; Smith, B. C.; Denu, J. M. Mechanism of Human SIRT1 Activation by Resveratrol. *J. Biol. Chem.* **2005**, *280* (17), 17187–17195. https://doi.org/10.1074/jbc.M501250200.

(73) Si, T.; Li, B.; Comi, T. J.; Wu, Y.; Hu, P.; Wu, Y.; Min, Y.; Mitchell, D. A.; Zhao, H.; Sweedler, J. V. Profiling of Microbial Colonies for High-Throughput Engineering of

Multistep Enzymatic Reactions via Optically Guided Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry. *J. Am. Chem. Soc.* **2017**, *139* (36), 12466–12473. https://doi.org/10.1021/jacs.7b04641.

(74) Bothner, B.; Chavez, R.; Wei, J.; Strupp, C.; Phung, Q.; Schneemann, A.; Siuzdak, G. Monitoring Enzyme Catalysis with Mass Spectrometry *. *J. Biol. Chem.* **2000**, *275* (18), 13455–13459. https://doi.org/10.1074/jbc.275.18.13455.

(75) Greis, K. D. Mass Spectrometry for Enzyme Assays and Inhibitor Screening: An Emerging Application in Pharmaceutical Research. *Mass Spectrom. Rev.* **2007**, *26* (3), 324–339. https://doi.org/10.1002/mas.20127.

(76) Lim, J. W.; Shin, K. S.; Moon, J.; Lee, S. K.; Kim, T. A Microfluidic Platform for High-Throughput Screening of Small Mutant Libraries. *Anal. Chem.* **2016**, *88* (10), 5234–5242. https://doi.org/10.1021/acs.analchem.6b00317.

(77) Si, T.; Li, B.; Comi, T. J.; Wu, Y.; Hu, P.; Wu, Y.; Min, Y.; Mitchell, D. A.; Zhao, H.; Sweedler, J. V. Profiling of Microbial Colonies for High-Throughput Engineering of Multistep Enzymatic Reactions via Optically Guided Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry. *J. Am. Chem. Soc.* **2017**, *139* (36), 12466–12473. https://doi.org/10.1021/jacs.7b04641.

(78) Ma, F.; Chung, M. T.; Yao, Y.; Nidetz, R.; Lee, L. M.; Liu, A. P.; Feng, Y.; Kurabayashi, K.; Yang, G.-Y. Efficient Molecular Evolution to Generate Enantioselective Enzymes Using a Dual-Channel Microfluidic Droplet Screening Platform. *Nat. Commun.* **2018**, *9* (1), 1–8. https://doi.org/10.1038/s41467-018-03492-6.

(79) Debon, A.; Pott, M.; Obexer, R.; Green, A. P.; Friedrich, L.; Griffiths, A. D.; Hilvert, D. Ultrahigh-Throughput Screening Enables Efficient Single-Round Oxidase Remodelling. *Nat. Catal.* **2019**, *2* (9), 740–747. https://doi.org/10.1038/s41929-019-0340-5.

(80) de Rond, T.; Gao, J.; Zargar, A.; de Raad, M.; Cunha, J.; Northen, T. R.; Keasling, J. D. A High-Throughput Mass Spectrometric Enzyme Activity Assay Enabling the Discovery of Cytochrome P450 Biocatalysts. *Angew. Chem.* **2019**, *131* (30), 10220–10225. https://doi.org/10.1002/ange.201901782.

(81) Romney, D. K.; Murciano-Calles, J.; Wehrmüller, J. E.; Arnold, F. H. Unlocking Reactivity of TrpB: A General Biocatalytic Platform for Synthesis of Tryptophan Analogues. *J. Am. Chem. Soc.* **2017**, *139* (31), 10769–10776. https://doi.org/10.1021/jacs.7b05007.

(82) Kiss, G.; Çelebi-Ölçüm, N.; Moretti, R.; Baker, D.; Houk, K. N. Computational Enzyme Design. *Angew. Chem. Int. Ed.* **2013**, *52* (22), 5700–5725. https://doi.org/10.1002/anie.201204077.

(83) Zhuang, Y.; Yang, G.-Y.; Chen, X.; Liu, Q.; Zhang, X.; Deng, Z.; Feng, Y. Biosynthesis of Plant-Derived Ginsenoside Rh2 in Yeast via Repurposing a Key Promiscuous Microbial Enzyme. *Metab. Eng.* **2017**, *42*, 25–32. https://doi.org/10.1016/j.ymben.2017.04.009.

(84) Acevedo-Rocha, C. G.; Agudo, R.; Reetz, M. T. Directed Evolution of Stereoselective Enzymes Based on Genetic Selection as Opposed to Screening Systems. *J. Biotechnol.* **2014**, *191*, 3–10. https://doi.org/10.1016/j.jbiotec.2014.04.009.

(85) Zhou, S.; Alper, H. S. Strategies for Directed and Adapted Evolution as Part of Microbial Strain Engineering. *J. Chem. Technol. Biotechnol.* **2019**, *94* (2), 366–376. https://doi.org/10.1002/jctb.5746.

(86) Feng, X.; Sanchis, J.; Reetz, M. T.; Rabitz, H. Enhancing the Efficiency of Directed Evolution in Focused Enzyme Libraries by the Adaptive Substituent Reordering Algorithm. *Chem. – Eur. J.* **2012**, *18* (18), 5646–5654. https://doi.org/10.1002/chem.201103811.

(87) Bunzel, H. A.; Garrabou, X.; Pott, M.; Hilvert, D. Speeding up Enzyme Discovery and Engineering with Ultrahigh-Throughput Methods. *Curr. Opin. Struct. Biol.* **2018**, *48*, 149–156. https://doi.org/10.1016/j.sbi.2017.12.010.

(88) Markel, U.; Essani, K. D.; Besirlioglu, V.; Schiffels, J.; Streit, W. R.; Schwaneberg, U. Advances in Ultrahigh-Throughput Screening for Directed Enzyme Evolution. *Chem. Soc. Rev.* **2020**, *49* (1), 233–262. https://doi.org/10.1039/C8CS00981C.

(89) Fulton, A.; Hayes, M. R.; Schwaneberg, U.; Pietruszka, J.; Jaeger, K.-E. High-Throughput Screening Assays for Lipolytic Enzymes. In *Protein Engineering*; Methods in Molecular Biology; Humana Press, New York, NY, 2018; pp 209–231. https://doi.org/10.1007/978-1-4939-7366-8_12.

(90) Lauchli, R.; Rabe, K. S.; Kalbarczyk, K. Z.; Tata, A.; Heel, T.; Kitto, R. Z.; Arnold, F. H. High-Throughput Screening for Terpene-Synthase-Cyclization Activity and Directed Evolution of a Terpene Synthase. *Angew. Chem. Int. Ed. 52* (21), 5571–5574. https://doi.org/10.1002/anie.201301362.

(91) Pluchinsky, A. J.; Wackelin, D. J.; Huang, X.; Arnold, F. H.; Mrksich, M. High Throughput Screening with SAMDI Mass Spectrometry for Directed Evolution. *J. Am. Chem. Soc.* **2020**, *142* (47), 19804–19808. https://doi.org/10.1021/jacs.0c07828.

(92) Burke, J. R.; La Clair, J. J.; Philippe, R. N.; Pabis, A.; Corbella, M.; Jez, J. M.; Cortina, G. A.; Kaltenbach, M.; Bowman, M. E.; Louie, G. V.; Woods, K. B.; Nelson, A. T.; Tawfik, D. S.; Kamerlin, S. C. L.; Noel, J. P. Bifunctional Substrate Activation via an Arginine Residue Drives Catalysis in Chalcone Isomerases. *ACS Catal.* **2019**, *9* (9), 8388–8396. https://doi.org/10.1021/acscatal.9b01926.

(93) Tong, Y.; Wei, Y.; Hu, Y.; Ang, E. L.; Zhao, H.; Zhang, Y. A Pathway for Isethionate Dissimilation in Bacillus Krulwichiae. *Appl. Environ. Microbiol.* **2019**, *85* (15). https://doi.org/10.1128/AEM.00793-19.

(94) Cen, Y.; Singh, W.; Arkin, M.; Moody, T. S.; Huang, M.; Zhou, J.; Wu, Q.; Reetz, M. T. Artificial Cysteine-Lipases with High Activity and Altered Catalytic Mechanism Created by Laboratory Evolution. *Nat. Commun.* **2019**, *10* (1), 1–10. https://doi.org/10.1038/s41467-019-11155-3.

(95) Yu, H.; López, R. I. H.; Steadman, D.; Méndez-Sánchez, D.; Higson, S.; Cázares-Körner, A.; Sheppard, T. D.; Ward, J. M.; Hailes, H. C.; Dalby, P. A. Engineering Transketolase to Accept Both Unnatural Donor and Acceptor Substrates and Produce α-Hydroxyketones. *FEBS J.* **2020**, *287* (9), 1758–1776. https://doi.org/10.1111/febs.15108.

(96) Chen, K.; Arnold, F. H. Engineering New Catalytic Activities in Enzymes. *Nat. Catal.* **2020**, 1–11. https://doi.org/10.1038/s41929-019-0385-5.

(97) Leveson-Gower, R. B.; Mayer, C.; Roelfes, G. The Importance of Catalytic Promiscuity for Enzyme Design and Evolution. *Nat. Rev. Chem.* **2019**, 1–19. https://doi.org/10.1038/s41570-019-0143-x.

(98) Chandgude, A. L.; Ren, X.; Fasan, R. Stereodivergent Intramolecular Cyclopropanation Enabled by Engineered Carbene Transferases. *J. Am. Chem. Soc.* **2019**, *141* (23), 9145–9150. https://doi.org/10.1021/jacs.9b02700.

(99) Grimm, A. R.; Sauer, D. F.; Polen, T.; Zhu, L.; Hayashi, T.; Okuda, J.; Schwaneberg, U. A Whole Cell E. Coli Display Platform for Artificial Metalloenzymes: Poly(Phenylacetylene) Production with a Rhodium–Nitrobindin Metalloprotein. *ACS Catal.* **2018**, *8* (3), 2611–2614. https://doi.org/10.1021/acscatal.7b04369.

(100) Emmanuel, M. A.; Greenberg, N. R.; Oblinsky, D. G.; Hyster, T. K. Accessing Non-Natural Reactivity by Irradiating Nicotinamide-Dependent Enzymes with Light. *Nature* **2016**, *540* (7633), 414–417. https://doi.org/10.1038/nature20569.

(101) Gu, Y.; Natoli, S. N.; Liu, Z.; Clark, D. S.; Hartwig, J. F. Site-Selective Functionalization of (Sp3)C−H Bonds Catalyzed by Artificial Metalloenzymes Containing an Iridium-Porphyrin Cofactor. *Angew. Chem. Int. Ed.* **2019**, *58* (39), 13954–13960. https://doi.org/10.1002/anie.201907460.

(102) Simon, A. J.; d'Oelsnitz, S.; Ellington, A. D. Synthetic Evolution. *Nat. Biotechnol.* **2019**, *37* (7), 730–743. https://doi.org/10.1038/s41587-019-0157-4.

(103) de Rond, T.; Danielewicz, M.; Northen, T. High Throughput Screening of Enzyme Activity with Mass Spectrometry Imaging. *Curr. Opin. Biotechnol.* **2015**, *31*, 1–9. https://doi.org/10.1016/j.copbio.2014.07.008.

(104) Novoa, C.; Dhoke, G. V.; Mate, D. M.; Martínez, R.; Haarmann, T.; Schreiter, M.; Eidner, J.; Schwerdtfeger, R.; Lorenz, P.; Davari, M. D.; Jakob, F.; Schwaneberg, U. KnowVolution of a Fungal Laccase toward Alkaline PH. *ChemBioChem* **2019**, *20* (11), 1458–1466. https://doi.org/10.1002/cbic.201800807.

(105) Watkins, E. J.; Almhjell, P. J.; Arnold, F. H. Direct Enzymatic Synthesis of a Deep-Blue Fluorescent Noncanonical Amino Acid from Azulene and Serine. *ChemBioChem* **2020**, *21* (1–2), 80–83. https://doi.org/10.1002/cbic.201900497.

(106) Yan, C.; Parmeggiani, F.; Jones, E. A.; Claude, E.; Hussain, S. A.; Turner, N. J.; Flitsch, S. L.; Barran, P. E. Real-Time Screening of Biocatalysts in Live Bacterial Colonies. *J. Am. Chem. Soc.* **2017**, *139* (4), 1408–1411. https://doi.org/10.1021/jacs.6b12165.

(107) Tarallo, V.; Sudarshan, K.; Nosek, V.; Míšek, J. Development of a Simple High-Throughput Assay for Directed Evolution of Enantioselective Sulfoxide Reductases. *Chem. Commun.* **2020**, *56* (40), 5386–5388. https://doi.org/10.1039/D0CC01660H.

(108) Kan, S. B. J.; Huang, X.; Gumulya, Y.; Chen, K.; Arnold, F. H. Genetically Programmed Chiral Organoborane Synthesis. *Nature* **2017**, *552* (7683), 132–136. https://doi.org/10.1038/nature24996.

(109) Welch, C. J.; Gong, X.; Schafer, W.; Pratt, E. C.; Brkovic, T.; Pirzada, Z.; Cuff, J. F.; Kosjek, B. MISER Chromatography (Multiple Injections in a Single Experimental Run): The Chromatogram Is the Graph. *Tetrahedron Asymmetry* **2010**, *21* (13), 1674–1681. https://doi.org/10.1016/j.tetasy.2010.05.029.

(110) Knorrscheidt, A.; Püllmann, P.; Schell, E.; Homann, D.; Freier, E.; Weissenborn, M. J. Identification of Novel Unspecific Peroxygenase Chimeras and Unusual YfeX Axial Heme Ligand by a Versatile High-Throughput GC-MS Approach. *ChemCatChem* **2020**, *n/a* (n/a). https://doi.org/10.1002/cctc.202000618.

(111) Sheludko, Y. V.; Fessner, W.-D. Winning the Numbers Game in Enzyme Evolution – Fast Screening Methods for Improved Biotechnology Proteins. *Curr. Opin. Struct. Biol.* **2020**, *63*, 123–133. https://doi.org/10.1016/j.sbi.2020.05.003.

(112) E. Kempa, E.; A. Hollywood, K.; A. Smith, C.; E. Barran, P. High Throughput Screening of Complex Biological Samples with Mass Spectrometry – from Bulk Measurements to Single Cell Analysis. *Analyst* **2019**, *144* (3), 872–891. https://doi.org/10.1039/C8AN01448E.

(113) Gul, I.; Bogale, T. F.; Chen, Y.; Yang, X.; Fang, R.; Feng, J.; Gao, H.; Tang, L. A Paper-Based Whole-Cell Screening Assay for Directed Evolution-Driven Enzyme Engineering. *Appl. Microbiol. Biotechnol.* **2020**, *104* (13), 6013–6022. https://doi.org/10.1007/s00253-020-10615-x.

(114) Dai, R.; Ten, A. S.; Mrksich, M. Profiling Protease Activity in Laundry Detergents with Peptide Arrays and SAMDI Mass Spectrometry. *Ind. Eng. Chem. Res.* **2019**, *58* (25), 10692–10697. https://doi.org/10.1021/acs.iecr.9b00057.

(115) O'Kane, P. T.; Dudley, Q. M.; McMillan, A. K.; Jewett, M. C.; Mrksich, M. High-Throughput Mapping of CoA Metabolites by SAMDI-MS to Optimize the Cell-Free Biosynthesis of HMG-CoA. *Sci. Adv.* **2019**, *5* (6), eaaw9180. https://doi.org/10.1126/sciadv.aaw9180.

(116) Kightlinger, W.; Lin, L.; Rosztoczy, M.; Li, W.; DeLisa, M. P.; Mrksich, M.; Jewett, M. C. Design of Glycosylation Sites by Rapid Synthesis and Analysis of Glycosyltransferases. *Nat. Chem. Biol.* **2018**, *14* (6), 627–635. https://doi.org/10.1038/s41589-018-0051-2.

(117) Helal, K. Y.; Alamgir, A.; Berns, E. J.; Mrksich, M. Traceless Immobilization of Analytes for High-Throughput Experiments with SAMDI Mass Spectrometry. *J. Am. Chem. Soc.* **2018**, *140* (26), 8060–8063. https://doi.org/10.1021/jacs.8b02918.

(118) Knorrscheidt, A.; Püllmann, P.; Schell, E.; Homann, D.; Freier, E.; Weissenborn, M. Development of 96 Multiple Injection-GC-MS Technique and Its Application in Protein Engineering of Natural and Non-Natural Enzymatic Reactions. **2019**. https://doi.org/10.26434/chemrxiv.10314239.v1.

(119) Techner, J.-M.; Hershewe, J.; Kightlinger, W.; Lin, L.; Ramesh, A.; DeLisa, M. P.; Jewett, M. C.; Mrksich, M. High-Throughput Synthesis and Analysis of Intact Glycoproteins Using SAMDI-MS. *Anal. Chem.* **2020**, *92* (2), 1963–1971. https://doi.org/10.1021/acs.analchem.9b04334.

(120) Hammer, S. C.; Knight, A. M.; Arnold, F. H. Design and Evolution of Enzymes for Non-Natural Chemistry. *Curr. Opin. Green Sustain. Chem.* **2017**, *7*, 23–30. https://doi.org/10.1016/j.cogsc.2017.06.002.

(121) Gibson, D. G.; Young, L.; Chuang, R.-Y.; Venter, J. C.; Hutchison, C. A.; Smith, H. O. Enzymatic Assembly of DNA Molecules up to Several Hundred Kilobases. *Nat. Methods* **2009**, *6* (5), 343–345. https://doi.org/10.1038/nmeth.1318.

(122) Berry, E. A.; Trumpower, B. L. Simultaneous Determination of Hemes a, b, and c from Pyridine Hemochrome Spectra. *Anal. Biochem.* **1987**, *161* (1), 1–15. https://doi.org/10.1016/0003-2697(87)90643-9.

(123) Kan, S. B. J.; Lewis, R. D.; Chen, K.; Arnold, F. H. Directed Evolution of Cytochrome c for Carbon–Silicon Bond Formation: Bringing Silicon to Life. *Science* **2016**. https://doi.org/10.1126/science.aah6219.

(124) Knight, A. M.; Kan, S. B. J.; Lewis, R. D.; Brandenberg, O. F.; Chen, K.; Arnold, F. H. Diverse Engineered Heme Proteins Enable Stereodivergent Cyclopropanation of Unactivated Alkenes. *ACS Cent. Sci.* **2018**, *4* (3), 372–377. https://doi.org/10.1021/acscentsci.7b00548.

(125) Wojtkiewicz, M.; Barnett, K.; Ciborowski, P. Protein Identification by Mass Spectrometry: Proteomics. In *Current Laboratory Methods in Neuroscience Research*; Xiong, H., Gendelman, H. E., Eds.; Springer Protocols Handbooks; Springer New York: New York, NY, 2014; pp 399–409. https://doi.org/10.1007/978-1-4614-8794-4_28.

(126) Chen, W.; Yin, X.; Mu, J.; Yin, Y. Subfemtomole Level Protein Sequencing by Edman Degradation Carried out in a Microfluidic Chip. *Chem. Commun.* **2007**, *0* (24), 2488–2490. https://doi.org/10.1039/B700200A.

(127) Gupta, K.; Kumar, M.; Chandrashekara, K.; Krishnan, K. S.; Balaram, P. Combined Electron Transfer Dissociation–Collision-Induced Dissociation Fragmentation in the Mass Spectrometric Distinction of Leucine, Isoleucine, and Hydroxyproline Residues in Peptide Natural Products. *J. Proteome Res.* **2012**, *11* (2), 515–522. https://doi.org/10.1021/pr200091v.

(128) Edman, P. A Method for the Determination of Amino Acid Sequence in Peptides. *Arch. Biochem.* **1949**, *22* (3), 475.

(129) Edman, P.; Begg, G. A Protein Sequenator. *Eur. J. Biochem.* **1967**, *1* (1), 80–91. https://doi.org/10.1111/j.1432-1033.1967.tb00047.x.

(130) Aguiar, M.; Haas, W.; Beausoleil, S. A.; Rush, J.; Gygi, S. P. Gas-Phase Rearrangements Do Not Affect Site Localization Reliability in Phosphoproteomics Data Sets. *J. Proteome Res.* **2010**, *9* (6), 3103–3107. https://doi.org/10.1021/pr1000225.

(131) Doherty, M. K.; Hammond, D. E.; Clague, M. J.; Gaskell, S. J.; Beynon, R. J. Turnover of the Human Proteome: Determination of Protein Intracellular Stability by Dynamic SILAC. *J. Proteome Res.* **2009**, *8* (1), 104–112. https://doi.org/10.1021/pr800641v.

(132) Choudhary, C.; Kumar, C.; Gnad, F.; Nielsen, M. L.; Rehman, M.; Walther, T. C.; Olsen, J. V.; Mann, M. Lysine Acetylation Targets Protein Complexes and Co-Regulates Major Cellular Functions. *Science* **2009**. https://doi.org/10.1126/science.1175371.

(133) Olsen, J. V.; Blagoev, B.; Gnad, F.; Macek, B.; Kumar, C.; Mortensen, P.; Mann, M. Global, In Vivo, and Site-Specific Phosphorylation Dynamics in Signaling Networks. *Cell* **2006**, *127* (3), 635–648. https://doi.org/10.1016/j.cell.2006.09.026.

(134) Zhang, N.; Wu, G.; Wu, H.; Chalmers, M. J.; Gaskell, S. J. Purification, Characterization and Sequence Determination of BmKK4, a Novel Potassium Channel Blocker from Chinese Scorpion Buthus Martensi Karsch. *Peptides* **2004**, *25* (6), 951–957. https://doi.org/10.1016/j.peptides.2004.03.014.

(135) Nair, S. S.; Nilsson, C. L.; Emmett, M. R.; Schaub, T. M.; Gowd, K. H.; Thakur, S. S.; Krishnan, K. S.; Balaram, P.; Marshall, A. G. De Novo Sequencing and Disulfide Mapping of a Bromotryptophan-Containing Conotoxin by Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Anal. Chem.* **2006**, *78* (23), 8082–8088. https://doi.org/10.1021/ac0607764.

(136) Maselli, V. M.; Bilusich, D.; Bowie, J. H.; Tyler, M. J. Host-Defence Skin Peptides of the Australian Streambank Froglet Crinia Riparia: Isolation and Sequence Determination by Positive and Negative Ion Electrospray Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **2006**, *20* (5), 797–803. https://doi.org/10.1002/rcm.2360.

(137) Ma, M.; Chen, R.; Ge, Y.; He, H.; Marshall, A. G.; Li, L. Combining Bottom-Up and Top-Down Mass Spectrometric Strategies for De Novo Sequencing of the Crustacean Hyperglycemic Hormone from Cancer Borealis. *Anal. Chem.* **2009**, *81* (1), 240–247. https://doi.org/10.1021/ac801910g.

(138) Medzihradszky, K. F. Peptide Sequence Analysis. In *Methods in Enzymology*; Biological Mass Spectrometry; Academic Press, 2005; Vol. 402, pp 209–244. https://doi.org/10.1016/S0076-6879(05)02007-0.

(139) Bean, M. F.; Carr, S. A.; Thorne, G. C.; Reilly, M. H.; Gaskell, S. J. Tandem Mass Spectrometry of Peptides Using Hybrid and Four-Sector Instruments: A Comparative Study. *Anal. Chem.* **1991**, *63* (14), 1473–1481. https://doi.org/10.1021/ac00014a024.

(140) Johnson, R. S.; Martin, S. A.; Biemann, Klaus.; Stults, J. T.; Watson, J. Throck. Novel Fragmentation Process of Peptides by Collision-Induced Decomposition in a Tandem Mass

Spectrometer: Differentiation of Leucine and Isoleucine. *Anal. Chem.* **1987**, *59* (21), 2621–2625. https://doi.org/10.1021/ac00148a019.

(141) Han, H.; Xia, Y.; McLuckey, S. A. Ion Trap Collisional Activation of c and Z• Ions Formed via Gas-Phase Ion/Ion Electron-Transfer Dissociation. *J. Proteome Res.* **2007**, *6* (8), 3062–3069. https://doi.org/10.1021/pr070177t.

(142) Savitski, M. M.; Nielsen, M. L.; Zubarev, R. A. Side-Chain Losses in Electron Capture Dissociation To Improve Peptide Identification. *Anal. Chem.* **2007**, *79* (6), 2296–2302. https://doi.org/10.1021/ac0619332.

(143) Kjeldsen, F.; Haselmann, K. F.; Sørensen, E. S.; Zubarev, R. A. Distinguishing of Ile/Leu Amino Acid Residues in the PP3 Protein by (Hot) Electron Capture Dissociation in Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Anal. Chem.* **2003**, *75* (6), 1267–1274. https://doi.org/10.1021/ac020422m.

(144) Tran, N. H.; Zhang, X.; Xin, L.; Shan, B.; Li, M. De Novo Peptide Sequencing by Deep Learning. *Proc. Natl. Acad. Sci.* **2017**, *114* (31), 8247–8252. https://doi.org/10.1073/pnas.1705691114.

(145) Chait, B. T.; Wang, R.; Beavis, R. C.; Kent, S. B. H. Protein Ladder Sequencing. *Science* **1993**, *262* (5130), 89–92. https://doi.org/10.1126/science.8211132.

(146) Tanco, S.; Gevaert, K.; Van Damme, P. C-Terminomics: Targeted Analysis of Natural and Posttranslationally Modified Protein and Peptide C-Termini. *PROTEOMICS* **2015**, *15* (5–6), 903–914. https://doi.org/10.1002/pmic.201400301.

(147) Nakazawa, T.; Yamaguchi, M.; Okamura, T.; Ando, E.; Nishimura, O.; Tsunasawa, S. Terminal Proteomics: N- and C-Terminal Analyses for High-Fidelity Identification of Proteins Using MS. *PROTEOMICS* **2008**, *8* (4), 673–685. https://doi.org/10.1002/pmic.200700084.

(148) Thakkar, A.; Wavreille, A.-S.; Pei, D. Traceless Capping Agent for Peptide Sequencing by Partial Edman Degradation and Mass Spectrometry. *Anal. Chem.* **2006**, *78* (16), 5935–5939. https://doi.org/10.1021/ac0607414.

(149) Hamberg, A.; Kempka, M.; Sjödahl, J.; Roeraade, J.; Hult, K. C-Terminal Ladder Sequencing of Peptides Using an Alternative Nucleophile in Carboxypeptidase Y Digests. *Anal. Biochem.* **2006**, *357* (2), 167–172. https://doi.org/10.1016/j.ab.2006.07.025.

(150) Sweeney, M. C.; Pei, D. An Improved Method for Rapid Sequencing of Support-Bound Peptides by Partial Edman Degradation and Mass Spectrometry. *J. Comb. Chem.* **2003**, *5* (3), 218–222. https://doi.org/10.1021/cc020113+.

(151) Kratzer, R.; Eckerskorn, C.; Karas, M.; Lottspeich, F. Suppression Effects in Enzymatic Peptide Ladder Sequencing Using Ultraviolet - Matrix Assisted Laser Desorption/Ionization - Mass Spectrometry. *ELECTROPHORESIS* **1998**, *19* (11), 1910–1919. https://doi.org/10.1002/elps.1150191109.

(152) Bonetto, V.; Bergman, A.-C.; Jörnvall, H.; Sillard, R. C-Terminal Sequence Analysis of Peptides and Proteins Using Carboxypeptidases and Mass Spectrometry after Derivatization

of Lys and Cys Residues. *Anal. Chem.* **1997**, *69* (7), 1315–1319. https://doi.org/10.1021/ac960896j.

(153) Gu, Q.-M.; Prestwich, G. D. Efficient Peptide Ladder Sequencing by MALDI-TOF Mass Spectrometry Using Allyl Isothiocyanate. *J. Pept. Res.* **1997**, *49* (6), 484–491. https://doi.org/10.1111/j.1399-3011.1997.tb01155.x.

(154) Bartlet-Jones, M.; Jeffery, W. A.; Hansen, H. F.; Pappin, D. J. C.; Cottrell, J. Peptide Ladder Sequencing by Mass Spectrometry Using a Novel, Volatile Degradation Reagent. *Rapid Commun. Mass Spectrom.* **1994**, *8* (9), 737–742. https://doi.org/10.1002/rcm.1290080916.

(155) Thiede, B.; Wittmann-Liebold, B.; Bienert, M.; Krause, E. MALDI-MS for C-Terminal Sequence Determination of Peptides and Proteins Degraded by Carboxypeptidase Y and P. *FEBS Lett.* **1995**, *357* (1), 65–69. https://doi.org/10.1016/0014-5793(94)01323-S.

(156) Xu, Y.; Little, M. W.; Rousell, D. J.; Laboy, J. L.; Murray, K. K. Direct from Polyacrylamide Gel Infrared Laser Desorption/Ionization. *Anal. Chem.* **2004**, *76* (4), 1078–1082. https://doi.org/10.1021/ac034879n.

(157) Xu, Y.; Little, M. W.; Murray, K. K. Interfacing Capillary Gel Microfluidic Chips with Infrared Laser Desorption Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2006**, *17* (3), 469–474. https://doi.org/10.1016/j.jasms.2005.12.003.

(158) Fujita, M.; Hattori, W.; Sano, T.; Baba, M.; Someya, H.; Miyazaki, K.; Kamijo, K.; Takahashi, K.; Kawaura, H. High-Throughput and High-Resolution Two Dimensional Mapping of PI and m/z Using a Microchip in a Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometer. *J. Chromatogr. A* **2006**, *1111* (2), 200–205. https://doi.org/10.1016/j.chroma.2005.11.117.

(159) Bakry, R.; Bonn, G. K.; Mair, D.; Svec, F. Monolithic Porous Polymer Layer for the Separation of Peptides and Proteins Using Thin-Layer Chromatography Coupled with MALDI-TOF-MS. *Anal. Chem.* **2007**, *79* (2), 486–493. https://doi.org/10.1021/ac061527i.

(160) Lee, J.; Soper, S. A.; Murray, K. K. Microfluidics with MALDI Analysis for Proteomics—A Review. *Anal. Chim. Acta* **2009**, *649* (2), 180–190. https://doi.org/10.1016/j.aca.2009.07.037.

(161) Lee, J.; Musyimi, H. K.; Soper, S. A.; Murray, K. K. Development of an Automated Digestion and Droplet Deposition Microfluidic Chip for MALDI-TOF MS. *J. Am. Soc. Mass Spectrom.* **2008**, *19* (7), 964–972. https://doi.org/10.1016/j.jasms.2008.03.015.

(162) Sim, T. S.; Kim, E.-M.; Joo, H. S.; Kim, B. G.; Kim, Y.-K. Application of a Temperature-Controllable Microreactor to Simple and Rapid Protein Identification Using MALDI-TOF MS. *Lab. Chip* **2006**, *6* (8), 1056–1061. https://doi.org/10.1039/B607769M.

(163) Guo, X.; Chan-Park, M. B.; Yoon, S. F.; Chun, J.; Hua, L.; Sze, N. S.-K. UV Embossed Polymeric Chip for Protein Separation and Identification Based on Capillary Isoelectric Focusing and MALDI-TOF-MS. *Anal. Chem.* **2006**, *78* (10), 3249–3256. https://doi.org/10.1021/ac051773e.

(164) Colangelo, J.; Orlando, R. On-Target Exoglycosidase Digestions/MALDI-MS for Determining the Primary Structures of Carbohydrate Chains. *Anal. Chem.* **1999**, *71* (7), 1479–1482. https://doi.org/10.1021/ac980980u.

(165) Zhao, Z.-G.; Cordovez, L. A.; Johnston, S. A.; Woodbury, N. Peptide Sequencing Directly on Solid Surfaces Using MALDI Mass Spectrometry. *Sci. Rep.* **2017**, *7* (1), 17811. https://doi.org/10.1038/s41598-017-18105-3.

(166) Semmler, A.; Weber, R.; Przybylski, M.; Wittmann, V. De Novo Sequencing of Peptides on Single Resin Beads by MALDI-FTICR Tandem Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2010**, *21* (2), 215–219. https://doi.org/10.1016/j.jasms.2009.10.004.

(167) Lin, S.; Liu, Y.; Zhang, M.; Xu, X.; Chen, Y.; Zhang, H.; Yang, C. Microfluidic Single-Cell Transcriptomics: Moving towards Multimodal and Spatiotemporal Omics. *Lab. Chip* **2021**, *21* (20), 3829–3849. https://doi.org/10.1039/D1LC00607J.

(168) Liu, Y.; Yang, M.; Deng, Y.; Su, G.; Enninful, A.; Guo, C. C.; Tebaldi, T.; Zhang, D.; Kim, D.; Bai, Z.; Norris, E.; Pan, A.; Li, J.; Xiao, Y.; Halene, S.; Fan, R. High-Spatial-Resolution Multi-Omics Sequencing via Deterministic Barcoding in Tissue. *Cell* **2020**, *183* (6), 1665-1681.e18. https://doi.org/10.1016/j.cell.2020.10.026.

(169) Qiu, Q.; Hu, P.; Qiu, X.; Govek, K. W.; Cámara, P. G.; Wu, H. Massively Parallel and Time-Resolved RNA Sequencing in Single Cells with ScNT-Seq. *Nat. Methods* **2020**, *17* (10), 991–1001. https://doi.org/10.1038/s41592-020-0935-4.

(170) Bhattacharjee, N.; Folch, A. Large-Scale Microfluidic Gradient Arrays Reveal Axon Guidance Behaviors in Hippocampal Neurons. *Microsyst. Nanoeng.* **2017**, *3* (1), 1–14. https://doi.org/10.1038/micronano.2017.3.

(171) Ono, Y.; Kitajima, M.; Daikoku, S.; Shiroya, T.; Nishihara, S.; Kanie, Y.; Suzuki, K.; Goto, S.; Kanie, O. Sequential Enzymatic Glycosyltransfer Reactions on a Microfluidic Device: Synthesis of a Glycosaminoglycan Linkage Region Tetrasaccharide. *Lab. Chip* **2008**, *8* (12), 2168–2173. https://doi.org/10.1039/B809316D.

(172) Maier, M.; Radtke, C. P.; Hubbuch, J.; Niemeyer, C. M.; Rabe, K. S. On-Demand Production of Flow-Reactor Cartridges by 3D Printing of Thermostable Enzymes. *Angew. Chem. Int. Ed.* **2018**, *57* (19), 5539–5543. https://doi.org/10.1002/anie.201711072.

(173) Obst, F.; Mertz, M.; Mehner, P. J.; Beck, A.; Castiglione, K.; Richter, A.; Voit, B.; Appelhans, D. Enzymatic Synthesis of Sialic Acids in Microfluidics to Overcome Cross-Inhibitions and Substrate Supply Limitations. *ACS Appl. Mater. Interfaces* **2021**, *13* (41), 49433–49444. https://doi.org/10.1021/acsami.1c12307.

(174) Pröschel, M.; Detsch, R.; Boccaccini, A. R.; Sonnewald, U. Engineering of Metabolic Pathways by Artificial Enzyme Channels. *Front. Bioeng. Biotechnol.* **2015**, *3*, 168. https://doi.org/10.3389/fbioe.2015.00168.

(175) Jia, F.; Narasimhan, B.; Mallapragada, S. Materials-Based Strategies for Multi-Enzyme Immobilization and Co-Localization: A Review. *Biotechnol. Bioeng.* **2014**, *111* (2), 209–222. https://doi.org/10.1002/bit.25136.

(176) Betancor, L.; Luckarift, H. Co-Immobilized Coupled Enzyme Systems in Biotechnology. *Biotechnol. Genet. Eng. Rev.* **2010**, *27*, 95–114. https://doi.org/10.1080/02648725.2010.10648146.

(177) Rabe, K. S.; Müller, J.; Skoupi, M.; Niemeyer, C. M. Cascades in Compartments: En Route to Machine-Assisted Biotechnology. *Angew. Chem. Int. Ed.* **2017**, *56* (44), 13574–13589. https://doi.org/10.1002/anie.201703806.

(178) Weake, V. M.; Workman, J. L. Histone Ubiquitination: Triggering Gene Activity. *Mol. Cell* **2008**, *29* (6), 653–663. https://doi.org/10.1016/j.molcel.2008.02.014.

(179) Grant, J.; Modica, J. A.; Roll, J.; Perkovich, P.; Mrksich, M. An Immobilized Enzyme Reactor for Spatiotemporal Control over Reaction Products. *Small* **2018**, *14* (31), 1800923. https://doi.org/10.1002/smll.201800923.

(180) Gong, A.; Zhu, C.-T.; Xu, Y.; Wang, F.-Q.; Tsabing, D. K.; Wu, F.-A.; Wang, J. Moving and Unsinkable Graphene Sheets Immobilized Enzyme for Microfluidic Biocatalysis. *Sci. Rep.* **2017**, *7* (1), 4309. https://doi.org/10.1038/s41598-017-04216-4.

(181) Seong, G. H.; Heo, J.; Crooks, R. M. Measurement of Enzyme Kinetics Using a Continuous-Flow Microfluidic System. *Anal. Chem.* **2003**, *75* (13), 3161–3167. https://doi.org/10.1021/ac034155b.

(182) Casadonte, R.; Caprioli, R. M. Proteomic Analysis of Formalin-Fixed Paraffin-Embedded Tissue by MALDI Imaging Mass Spectrometry. *Nat. Protoc.* **2011**, *6* (11), 1695–1709. https://doi.org/10.1038/nprot.2011.388.

(183) Garrett, T. J.; Prieto-Conaway, M. C.; Kovtoun, V.; Bui, H.; Izgarian, N.; Stafford, G.; Yost, R. A. Imaging of Small Molecules in Tissue Sections with a New Intermediate-Pressure MALDI Linear Ion Trap Mass Spectrometer. *Int. J. Mass Spectrom.* **2007**, *260* (2), 166–176. https://doi.org/10.1016/j.ijms.2006.09.019.

(184) Kompauer, M.; Heiles, S.; Spengler, B. Atmospheric Pressure MALDI Mass Spectrometry Imaging of Tissues and Cells at 1.4-Mm Lateral Resolution. *Nat. Methods* **2017**, *14* (1), 90–96. https://doi.org/10.1038/nmeth.4071.

(185) Andersson, M.; Groseclose, M. R.; Deutch, A. Y.; Caprioli, R. M. Imaging Mass Spectrometry of Proteins and Peptides: 3D Volume Reconstruction. *Nat. Methods* **2008**, *5* (1), 101–108. https://doi.org/10.1038/nmeth1145.

(186) Altelaar, A. F. M.; Taban, I. M.; McDonnell, L. A.; Verhaert, P. D. E. M.; de Lange, R. P. J.; Adan, R. A. H.; Mooi, W. J.; Heeren, R. M. A.; Piersma, S. R. High-Resolution MALDI Imaging Mass Spectrometry Allows Localization of Peptide Distributions at Cellular Length Scales in Pituitary Tissue Sections. *Int. J. Mass Spectrom.* **2007**, *260* (2), 203–211. https://doi.org/10.1016/j.ijms.2006.09.028.

(187) Sinha, T. K.; Khatib-Shahidi, S.; Yankeelov, T. E.; Mapara, K.; Ehtesham, M.; Cornett, D. S.; Dawant, B. M.; Caprioli, R. M.; Gore, J. C. Integrating Spatially Resolved Three-Dimensional MALDI IMS with in Vivo Magnetic Resonance Imaging. *Nat. Methods* **2008**, *5* (1), 57–59. https://doi.org/10.1038/nmeth1147.

(188) Zavalin, A.; Yang, J.; Hayden, K.; Vestal, M.; Caprioli, R. M. Tissue Protein Imaging at 1 Mm Laser Spot Diameter for High Spatial Resolution and High Imaging Speed Using Transmission Geometry MALDI TOF MS. *Anal. Bioanal. Chem.* **2015**, *407* (8), 2337–2342. https://doi.org/10.1007/s00216-015-8532-6.

(189) Grant, J.; O'Kane, P. T.; Kimmel, B. R.; Mrksich, M. Using Microfluidics and Imaging SAMDI-MS To Characterize Reaction Kinetics. *ACS Cent. Sci.* **2019**, *5* (3), 486–493. https://doi.org/10.1021/acscentsci.8b00867.

(190) Tran, N. H.; Rahman, M. Z.; He, L.; Xin, L.; Shan, B.; Li, M. Complete De Novo Assembly of Monoclonal Antibody Sequences. *Sci. Rep.* **2016**, *6* (1), 31730. https://doi.org/10.1038/srep31730.

(191) Peng, W.; Pronker, M. F.; Snijder, J. Mass Spectrometry-Based De Novo Sequencing of Monoclonal Antibodies Using Multiple Proteases and a Dual Fragmentation Scheme. *J. Proteome Res.* **2021**, *20* (7), 3559–3566. https://doi.org/10.1021/acs.jproteome.1c00169.

(192) Vitorino, R.; Guedes, S.; Trindade, F.; Correia, I.; Moura, G.; Carvalho, P.; Santos, M. A. S.; Amado, F. De Novo Sequencing of Proteins by Mass Spectrometry. *Expert Rev. Proteomics* **2020**, *17* (7–8), 595–607. https://doi.org/10.1080/14789450.2020.1831387.

(193) Matsueda, G. R.; Haber, E.; Margolies, M. N. Quantitative Solid-Phase Edman Degradation for Evaluation of Extended Solid-Phase Peptide Synthesis. *Biochemistry* **1981**, *20* (9), 2571–2580. https://doi.org/10.1021/bi00512a032.

(194) Lobas, A. A.; Verenchikov, A. N.; Goloborodko, A. A.; Levitsky, L. I.; Gorshkov, M. V. Combination of Edman Degradation of Peptides with Liquid Chromatography/Mass Spectrometry Workflow for Peptide Identification in Bottom-up Proteomics. *Rapid Commun. Mass Spectrom.* **2013**, *27* (3), 391–400. https://doi.org/10.1002/rcm.6462.

(195) Nemeth-Cawley, J. F.; Rouse, J. C. Identification and Sequencing Analysis of Intact Proteins via Collision-Induced Dissociation and Quadrupole Time-of-Flight Mass Spectrometry. *J. Mass Spectrom.* **2002**, *37* (3), 270–282. https://doi.org/10.1002/jms.281.

(196) Lin, T.; Glish, G. L. C-Terminal Peptide Sequencing via Multistage Mass Spectrometry. *Anal. Chem.* **1998**, *70* (24), 5162–5165. https://doi.org/10.1021/ac980823v.

(197) Mao, Y.; Zhang, L.; Kleinberg, A.; Xia, Q.; Daly, T. J.; Li, N. Fast Protein Sequencing of Monoclonal Antibody by Real-Time Digestion on Emitter during Nanoelectrospray. *mAbs* **2019**, *11* (4), 767–778. https://doi.org/10.1080/19420862.2019.1599633.

(198) Armirotti, A.; Millo, E.; Damonte, G. How to Discriminate Between Leucine and Isoleucine by Low Energy ESI-TRAP MSn. *J. Am. Soc. Mass Spectrom.* **2007**, *18* (1), 57–63. https://doi.org/10.1016/j.jasms.2006.08.011.

(199) Billeci, T. M.; Stults, J. T. Tryptic Mapping of Recombinant Proteins by Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry. *Anal. Chem.* **1993**, *65* (13), 1709–1716. https://doi.org/10.1021/ac00061a013.

(200) Keough, T.; Youngquist, R. S.; Lacey, M. P. A Method for High-Sensitivity Peptide Sequencing Using Postsource Decay Matrix-Assisted Laser Desorption Ionization Mass

Spectrometry. *Proc. Natl. Acad. Sci.* **1999**, *96* (13), 7131–7136. https://doi.org/10.1073/pnas.96.13.7131.

(201) Zhong, H.; Zhang, Y.; Wen, Z.; Li, L. Protein Sequencing by Mass Analysis of Polypeptide Ladders after Controlled Protein Hydrolysis. *Nat. Biotechnol.* **2004**, *22* (10), 1291–1296. https://doi.org/10.1038/nbt1011.

(202) Doucette, A.; Li, L. Investigation of the Applicability of a Sequential Digestion Protocol Using Trypsin and Leucine Aminopeptidase M for Protein Identification by Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry. *Eur. J. Mass Spectrom.* **2001**, *7* (2), 157–170. https://doi.org/10.1255/ejms.400.

(203) Samyn, B.; Sergeant, K.; Castanheira, P.; Faro, C.; Van Beeumen, J. A New Method for C-Terminal Sequence Analysis in the Proteomic Era. *Nat. Methods* **2005**, *2* (3), 193–200. https://doi.org/10.1038/nmeth738.

(204) Mano, N.; Iijima, S.; Kasuga, K.; Goto, J. Exopeptidase Degradation for the Analysis of Phosphorylation Site in a Mono-Phosphorylated Peptide with Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry. *Anal. Sci.* **2003**, *19* (11), 1469–1472. https://doi.org/10.2116/analsci.19.1469.

(205) Jiang, H.; Zou, H.; Wang, H.; Zhang, Q.; Ni, J.; Zhang, Q.; Guo, Z.; Chen, X. Combination of MALDI-TOF Mass Spectrometry with Immobilized Enzyme Microreactor for Peptide Mapping. *Sci. China Ser. B Chem.* **2000**, *43* (6), 625–633. https://doi.org/10.1007/BF02969510.

(206) Chen, W.; Yin, X.; Yin, Y. Rapid and Reliable Peptide de Novo Sequencing Facilitated by Microfluidic Chip-Based Edman Degradation. *J. Proteome Res.* **2008**, *7* (2), 766–770. https://doi.org/10.1021/pr070465p.

(207) Fallahi, H.; Zhang, J.; Phan, H.-P.; Nguyen, N.-T. Flexible Microfluidics: Fundamentals, Recent Developments, and Applications. *Micromachines* **2019**, *10* (12), 830. https://doi.org/10.3390/mi10120830.

(208) Grant, J.; Goudarzi, S. H.; Mrksich, M. High-Throughput Enzyme Kinetics with 3D Microfluidics and Imaging SAMDI Mass Spectrometry. *Anal. Chem.* **2018**, *90* (21), 13096–13103. https://doi.org/10.1021/acs.analchem.8b04391.

(209) Gurard-Levin, Z. A.; Kim, J.; Mrksich, M. Combining Mass Spectrometry and Peptide Arrays to Profile the Specificities of the Histone Deacetylases. *Chembiochem Eur. J. Chem. Biol.* **2009**, *10* (13), 2159–2161. https://doi.org/10.1002/cbic.200900417.

(210) Ben-Meir, D.; Spungin, A.; Ashkenazi, R.; Blumberg, S. Specificity of Streptomyces Griseus Aminopeptidase and Modulation of Activity by Divalent Metal Ion Binding and Substitution. *Eur. J. Biochem.* **1993**, *212* (1), 107–112. https://doi.org/10.1111/j.1432-1033.1993.tb17639.x.

(211) Spungin, A.; Blumberg, S. Streptomyces Griseus Aminopeptidase Is a Calcium-Activated Zinc Metalloprotein. *Eur. J. Biochem.* **1989**, *183* (2), 471–477. https://doi.org/10.1111/j.1432-1033.1989.tb14952.x.

(212) Maras, B.; Greenblatt, H. M.; Shoham, G.; Spungin-Bialik, A.; Blumberg, S.; Barra, D. Aminopeptidase from Streptomyces Griseus. *Eur. J. Biochem.* **1996**, *236* (3), 843–846. https://doi.org/10.1111/j.1432-1033.1996.00843.x.

(213) Jankiewicz, U.; Bielawski, W. The Properties and Functions of Bacterial Aminopeptidases. *Acta Microbiol. Pol.* **2003**, *52* (3), 217–231.

(214) Moerschell, R. P.; Hosokawa, Y.; Tsunasawa, S.; Sherman, F. The Specificities of Yeast Methionine Aminopeptidase and Acetylation of Amino-Terminal Methionine in Vivo. Processing of Altered Iso-1-Cytochromes c Created by Oligonucleotide Transformation. *J. Biol. Chem.* **1990**, *265* (32), 19638–19643. https://doi.org/10.1016/S0021-9258(17)45419-6.

(215) Niven, G. W.; Holder, S. A.; Strøman, P. A Study of the Substrate Specificity of Aminopeptidase N from Lactococcus Lactis Subsp. Cremoris Wg2. *Appl. Microbiol. Biotechnol.* **1995**, *44* (1), 100–105. https://doi.org/10.1007/BF00164487.

(216) Nandan, A. S.; Nampoothiri, K. M. Unveiling Aminopeptidase P from Streptomyces Lavendulae: Molecular Cloning, Expression and Biochemical Characterization. *Enzyme Microb. Technol.* **2014**, *55*, 7–13. https://doi.org/10.1016/j.enzmictec.2013.11.003.

(217) Cheung, H. S.; Cushman, D. W. A Soluble Aspartate Aminopeptidase from Dog Kidney. *Biochim. Biophys. Acta BBA - Enzymol.* **1971**, *242* (1), 190–193. https://doi.org/10.1016/0005-2744(71)90098-2.

(218) Nandan, A.; Pandey, A.; Nampoothiri, K. M. Proline-Specific Extracellular Aminopeptidase Purified from Streptomyces Lavendulae. *Appl. Biochem. Biotechnol.* **2011**, *163* (8), 994–1001. https://doi.org/10.1007/s12010-010-9103-0.

(219) Blinkovsky, A. M.; Byun, T.; Brown, K. M.; Golightly, E. J.; Klotz, A. V. A Non-Specific Aminopeptidase from Aspergillus. *Biochim. Biophys. Acta BBA - Protein Struct. Mol. Enzymol.* **2000**, *1480* (1), 171–181. https://doi.org/10.1016/S0167-4838(00)00064-9.

(220) Breddam, K. Serine Carboxypeptidases. A Review. *Carlsberg Res. Commun.* **1986**, *51* (2), 83–128. https://doi.org/10.1007/BF02907561.

(221) Borgo, B.; Havranek, J. J. Computer-Aided Design of a Catalyst for Edman Degradation Utilizing Substrate-Assisted Catalysis. *Protein Sci. Publ. Protein Soc.* **2015**, *24* (4), 571–579. https://doi.org/10.1002/pro.2633.

(222) Houseman, B. T.; Gawalt, E. S.; Mrksich, M. Maleimide-Functionalized Self-Assembled Monolayers for the Preparation of Peptide and Carbohydrate Biochips. *Langmuir* **2003**, *19* (5), 1522–1531. https://doi.org/10.1021/la0262304.

(223) Yousaf, M. N.; Mrksich, M. Diels−Alder Reaction for the Selective Immobilization of Protein to Electroactive Self-Assembled Monolayers. *J. Am. Chem. Soc.* **1999**, *121* (17), 4286–4287. https://doi.org/10.1021/ja983529t.

(224) Ban, L.; Mrksich, M. On-Chip Synthesis and Label-Free Assays of Oligosaccharide Arrays. *Angew. Chem.* **2008**, *120* (18), 3444–3447. https://doi.org/10.1002/ange.200704998.

(225) Wollman, E. W.; Kang, D.; Frisbie, C. D.; Lorkovic, I. M.; Wrighton, M. S. Photosensitive Self-Assembled Monolayers on Gold: Photochemistry of Surface-Confined Aryl Azide and Cyclopentadienylmanganese Tricarbonyl. *J. Am. Chem. Soc.* **1994**, *116* (10), 4395–4404. https://doi.org/10.1021/ja00089a030.

(226) Chechik, V.; Crooks, R. M.; Stirling, C. J. M. Reactions and Reactivity in Self-Assembled Monolayers. *Adv. Mater.* **2000**, *12* (16), 1161–1171. https://doi.org/10.1002/1521-4095(200008)12:16<1161::AID-ADMA1161>3.0.CO;2-C.

(227) Howard, C. J.; Floyd, B. M.; Bardo, A. M.; Swaminathan, J.; Marcotte, E. M.; Anslyn, E. V. Solid-Phase Peptide Capture and Release for Bulk and Single-Molecule Proteomics. *ACS Chem. Biol.* **2020**, *15* (6), 1401–1407. https://doi.org/10.1021/acschembio.0c00040.

(228) Howard, K. L.; Boyer, G. L. Adduct Simplification in the Analysis of Cyanobacterial Toxins by Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **2007**, *21* (5), 699–706. https://doi.org/10.1002/rcm.2887.

(229) Lutz, J.-F. Copper-Free Azide–Alkyne Cycloadditions: New Insights and Perspectives. *Angew. Chem. Int. Ed.* **2008**, *47* (12), 2182–2184. https://doi.org/10.1002/anie.200705365.

(230) Agard, N. J.; Prescher, J. A.; Bertozzi, C. R. A Strain-Promoted [3 + 2] Azide−Alkyne Cycloaddition for Covalent Modification of Biomolecules in Living Systems. *J. Am. Chem. Soc.* **2004**, *126* (46), 15046–15047. https://doi.org/10.1021/ja044996f.

(231) Dommerholt, J.; Rutjes, F. P. J. T.; van Delft, F. L. Strain-Promoted 1,3-Dipolar Cycloaddition of Cycloalkynes and Organic Azides. *Top. Curr. Chem. J.* **2016**, *374* (2). https://doi.org/10.1007/s41061-016-0016-4.

(232) Debets, M. F.; van Berkel, S. S.; Dommerholt, J.; Dirks, A. (Ton) J.; Rutjes, F. P. J. T.; van Delft, F. L. Bioconjugation with Strained Alkenes and Alkynes. *Acc. Chem. Res.* **2011**, *44* (9), 805–815. https://doi.org/10.1021/ar200059z.

(233) Mbua, N. E.; Guo, J.; Wolfert, M. A.; Steet, R.; Boons, G.-J. Strain-Promoted Alkyne-Azide Cycloadditions (SPAAC) Reveal New Features of Glycoconjugate Biosynthesis. *Chembiochem Eur. J. Chem. Biol.* **2011**, *12* (12), 1912–1921. https://doi.org/10.1002/cbic.201100117.

(234) Carpenter, R. D.; Hausner, S. H.; Sutcliffe, J. L. Copper-Free Click for PET: Rapid 1,3-Dipolar Cycloadditions with a Fluorine-18 Cyclooctyne. *ACS Med. Chem. Lett.* **2011**, *2* (12), 885–889. https://doi.org/10.1021/ml200187j.

(235) Sletten, E. M.; Bertozzi, C. R. A Hydrophilic Azacyclooctyne for Cu-Free Click Chemistry. *Org. Lett.* **2008**, *10* (14), 3097–3099. https://doi.org/10.1021/ol801141k.

(236) Jewett, J. C.; Sletten, E. M.; Bertozzi, C. R. Rapid Cu-Free Click Chemistry with Readily Synthesized Biarylazacyclooctynones. *J. Am. Chem. Soc.* **2010**, *132* (11), 3688–3690. https://doi.org/10.1021/ja100014q.

Vita

Adam J. Pluchinsky grew up in Catasauqua, Pennsylvania. He attended elementary school at Our Lady of Hungary Regional School and graduated from Catasauqua Area High School as the Valedictorian of his class in June 2011. For his undergraduate education, he attended both Fordham and Columbia University from August 2011 to June 2016, where he graduated with honors and earned a B.S. in Chemistry and a second in Biomedical Engineering, respectively, as part of a specialized dual enrollment 3-2 cooperative engineering combined plan program. The following autumn, he entered Northwestern University to complete his doctorate work over the course of 5 years.