NORTHWESTERN UNIVERSITY

The Design and Application of Novel Biochemical Assays Based on the SAMDI-MS Platform

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Chemistry

By

Patrick Thomas O'Kane

EVANSTON, ILLINOIS

June 2019

Abstract

SAMDI-MS (Self-Assembled Monolayers for MALDI-TOF Mass Spectrometry) couples the use of chemically-defined self-assembled monolayers of alkane thiolates on gold surfaces with MALDI-TOF mass spectrometry for rapid characterization of the surfaces. Reactions performed on the surfaces can be quantified directly by detection of the reaction substrates and products. This rapid detection by mass spectrometry allows SAMDI to offer a generalizable and quantitative platform for bioassay development that can easily be extended to a wide range of enzymatic reactions without the need for development of reaction-specific assays. In this work I present the extension of this SAMDI platform to the development and application of two novel bioassays.

The first, is an assay for the detection of protein-ligand interactions. In the work I present here, this assay is applied to the characterization of adaptor domains. Adaptor domains are small, protein domains that bind to sites of post-translational modification. These protein modules do not affect a measurable mass change on their own. Instead, this assay takes advantage of the rate enhancement of enzyme mediated reactions on surfaces that occurs when the enzyme and substrate are colocalized to the surface. Potential ligand peptides are patterned on a monolayer with substrate for a reporter enzyme. Adaptor domains fused to reporter enzyme will localize the reporter to the surface, in a manner that is dependent on the affinity for the ligand, where the reporter will more rapidly convert substrate to product. This covalent record of the transient bind interaction can be quantified, allowing the adaptor-ligand interactions to be rapidly profiled and rank-ordered by affinity. I use this assay to profile chromodomains, which bind to methylated lysine, on biologically relevant ligands derived from the histone 3 protein. Additionally, this assay

is applied to profiling SH2 domains, which bind to sites of phosphorylation on tyrosine, showing how it can be used to examine affinity of an adaptor domain to large libraries of potential ligands.

The second assay I will present in this work is a general platform for the rapid characterization of enzyme reaction networks which use the Coenzyme A (CoA) cofactor. CoA is essential to cellular metabolism and is involved in a huge number of reactions, the products of which have a wide range of desirable applications in industrial and pharmaceutical settings. There is a great deal of interest in using these biological pathways to manufacture small molecules industrially, but this goal is hindered by a lack of analytical techniques for rapidly characterizing and optimizing these complex reaction networks. Here, I present an assay which couples bio-specific capture chemistry with rapid analysis by SAMDI to create a general platform for the detection and quantification of CoA metabolites. The assay works by using a chemical probe to selectively capture the CoA metabolites from lysates and complex reaction mixtures; these captured metabolites are then selectively immobilized on arrays of self-assembled monolayer and characterized by SAMDI. This strategy offers a significant increase in the throughput of reaction characterization by eliminating the need for lengthy purification techniques and development of metabolite-specific detection methods. In this work, I apply this assay to the optimization of a cell-free reaction system for the bio-synthesis of HMG-CoA, the metabolic precursor to isoprenoid metabolites.

Acknowledgements

The process of completing a PhD is not always easy. Trying to accomplish something new and exciting that has never been done before (and sometimes for good reason) can be difficult. It was with the support of many people that I was able to get through this degree.

First, I would like to sincerely thank my advisor, Professor Milan Mrksich, for giving me the backing and freedom to explore a number of interesting and exciting problems. His support made all the work present here possible, enabling me to take a few interesting ideas and turn them into fully realized projects, and in the process creating several new tools that will hopeful serve to empower future work. I would also like to thank Professor Neil Kelleher and Professor Karl Scheidt for taking the time to serve on my dissertation committee, and to Professor Chad Mirkin for serving on my qualifying committee.

I would like to acknowledge and thank the collaborators with whom I've had the opportunity and pleasure of work with on the projects presented in this work. I would like to thank Aislinn McMillan, whom I mentored for several years as an undergraduate at Northwestern and who contributed to several of the experiments presented here. I would like to thank my collaborator, Quentin Dudley. His efforts were crucial to implementing the assay for CoA metabolism presented in this work. Finally, I would like to thank Jennifer Grant who is responsible for engineering the microfluidic systems and implementing the imaging mass spectrometry that made all the work in Chapter 8 of thesis possible.

I would like to thank all the past and present members of the Mrksich group who have helped me, taught me, and made the lab a place where I want to work every day. I'd like to start by thanking Yael Mayer, who singlehandedly keeps the Mrksich lab functioning. I'd like to thank

Andreea Stuparu, who was the first person I met when I started graduate school and the person who convinced me to look into joining Milan's lab. Heather Scott taught me most of what I know about protein expression after I joined the lab and was tasked with making half a dozen protein constructs. I'd like to thank Adam Eisenberg, who served as lab manager for the first several years of my grad school career, and who was always willing to help. Alexei Ten joined Milan's lab with me and who has been a friend, someone to vent frustrations to and celebrate successes with, throughout my tenure in the lab. Finally, I'd to thank Lindsey Szymczak, my friend and, for the last several years, my office desk neighbor who has served as my sounding board for all life problems, scientific and otherwise. She also convinced me to learn how to swim just because she didn't want to do a triathlon alone. Working alongside these and many other great people has made this PhD a worthwhile experience.

I would like to thank everyone else, friends and family, who helped support me through these years and who made all of this possible. The many friends I've made throughout this PhD have been one of the most rewarding parts of the experience. I'd like to thank Patrick Figliozzi, my roommate for much of my time in grad school. Thank you for always being there to commiserate with, discuss all my most nerdy hobbies with, and for constantly encouraging my craft beer obsession. Thanks to Zachary Hund for being a good friend and the instigator of many good times. Thank you to John Savage for going to all the concerts and music festivals with me, and for including me in his Irish hijinks that led to some unforgettable experiences. Thanks to Andrew Sand, without whom I can confidently say I would have easily gone the rest of my life without experiencing curling first hand. Thanks to Dmitriy Dolzhnikov for being a friend and all-around kind person.

Finally, I must give particular thanks to my family, my parents and my sisters. Thank you to my father, who has taken it upon himself to read up on epigenetics in order to understand some of my work, so much so that I think he may be ready to start a PhD of his own soon. Thank you to my mother for filling my inbox with photos, and for never forgetting to ask if there is anything I need. And thanks to my sisters, Katie and Maureen for all their love and support.

Table of Contents

List of Figures

Chapter 1

Introduction to Biosensors and SAMDI Mass Spectrometry

1.1 Biosensing, Arrays, and High-throughput Technologies

Many advances in modern biology and biochemistry have been enabled by the widespread development of sensing technologies and detection platforms. Even in our everyday lives, biosensing technologies are becoming increasingly ubiquitous; blood-glucose sensors are now a commonplace tool for managing diabetes and heartrate monitors are built into our watches. In the pharmaceutical industry, rapid screening methods are an essential tool for identifying potential drug molecules from compound libraries that routinely number in the hundreds-of-thousands to millions, and new assays for biochemical and phenotypic screening of these libraries are constantly in development. In the lab, the development of new, high-throughput biosensing strategies is enabling a greater understanding of complex cellular processes and allowing us to better characterize biochemical reactions in the context of cellular systems.

Strategies for analyzing biological systems can be generally classified by their target analyte, throughput, scalability, and the detection method they employ. Arguably, no single technology in this area has been as successful as DNA microarrays, as they offer both excellent sensitivity and scalability to hundreds of thousands of measurements. In just a few short decades a number of strategies for generating high density DNA arrays have been successfully developed and commercialized, which have proven to be invaluable with numerous applications in DNA and RNA analysis. Current techniques use photolithography[1], electrochemistry[2], random arrays

synthesis[3-4], and inkjet printing[5] to generate high density and affordable DNA arrays wherein tens of thousands of unique sequences can be contained on a single chip. In the laboratory, these microarrays have been used for large scale gene expression profiling, identifications of single nucleotide polymorphisms, genotyping, and analysis of transcription factors.[6] The ability to affordably manufacture such arrays has dramatically lowered the cost of genotyping, leading the way for new applications in research and medicine.

For characterizing proteins and enzymes, there is no single technology that we can point to as being so universally applicable as DNA arrays have been to genomic research. One reasons for this is the enormous diversity of functions that proteins possess, as well as their greater structural complexity relative to DNA. For DNA, information storage is the primary purpose and its function is dictated primarily by its sequence. Additionally, DNA possesses a relatively simple structure composed of just four primary nucleotide subunits and can be very easily amplified to allow samples as small as a single cell to be sequenced. For proteins, the need for functional characterization means that it is much more difficult to identify technologies applicable to all use cases and this limits the development of generalizable protein analytical tools.

Advances in the field of proteomics have made it possible to observe to global content of expressed protein in a cell sample. This involves separating a complex mixture of proteins from a biological sample, by means such as gel electrophoresis or chromatography, and then identifying individual proteins present by their unique characteristics. State-of-the-art methods rely on purification methods, such as chromatography, coupled to high-resolution mass spectrometry to separate and identify the unique profile of proteins present in samples of intact proteins (top-down approach) or proteolytically digest protein samples (bottom-up approach).[7-9] Functional

proteomics techniques, such as those developed by Cravatt and coworkers, have made it possible to identify proteins from a global population based on activity by building sensors to label proteins that possess a certain enzymatic activity.[10-12] These technologies have enabled mapping of global protein content in the same way that DNA microarrays enable global transcription mapping. However, the need for complex sample preparations and expensive instrumentation have been limiting factors hindering wider adoption.

For identification of a specific protein of interest in a sample, immunolabeling strategies are common. Western blotting is often used for smaller scale experiments, coupling gel electrophoresis for separation of proteins in a sample based on size and charge, followed by labeling of a protein of interest with an antibody that binds specifically to the target.[13] For high-throughput identification and quantification of a protein of interest, ELISA (enzyme-linked immunosorbent assay) offers a strategy for immunolabeling of a target protein that is compatible with array well plate formats and multiplexed detection usually by means of a colorimetric readout.[14] While many variations of the ELISA assay exist, in simplest form a protein sample of interest is bound to the surface of a well plate, a primary antibody is used to bind specifically to a targeted analyte, and then a secondary antibody conjugated to a reporter binds the primary antibody. Enzymatic reporters, such as horseradish peroxidase, enable signal amplification where a small amount of label is used to catalytically convert a supplied substrate to a readily detectible product. Assays based on this platform are scalable to hundreds of samples and require relatively inexpensive instrumentation but immunolabeling schemes increase the difficultly of designing and implementing assays, and this design process must be repeated for new target analytes.

These methods so far have focused on protein detection rather than functional characterization. The strategies and tools for characterizing the activity of proteins and enzymes are as diverse and varied as the functions of the proteins themselves. The sheer range of functions limits the general applicability of any one assay format. However, when we survey commercially available assays, several common strategies for activity quantification emerge. The most widespread strategy for high-throughput enzyme activity screening is the design of substrates whose reaction products can be monitored by absorbance or fluorescence (**Figure 1.1**). For some enzymes, non-native substrates can be designed around a fluorophore or colorimetric core, so that when an enzyme acts on the substrate it generates a product with a characteristic fluorescence or absorbance profile. Examples include small molecules like p-nitro-phenyl phosphate which turns bright yellow when enzymatically dephosphorylated, labeled peptide substrates for deacetylases which liberate a fluorophore when deacetylated, and fluorophore labeled sugars like fluorescein di-β-D-galactopyranoside that can be used to track the activity of glycosidases.[15-16] In some cases, labeling is achieved through antibodies against the enzyme product, a strategy used in the commercial LanthaScreen assay from ThermoFisher for kinase screening, which pairs antibodies against phosphorylated peptides and FRET (fluorescence resonance energy transfer) for detection. For proteases, disruption of a FRET pair by cleavage of a peptide is another available labeling scheme.[17]

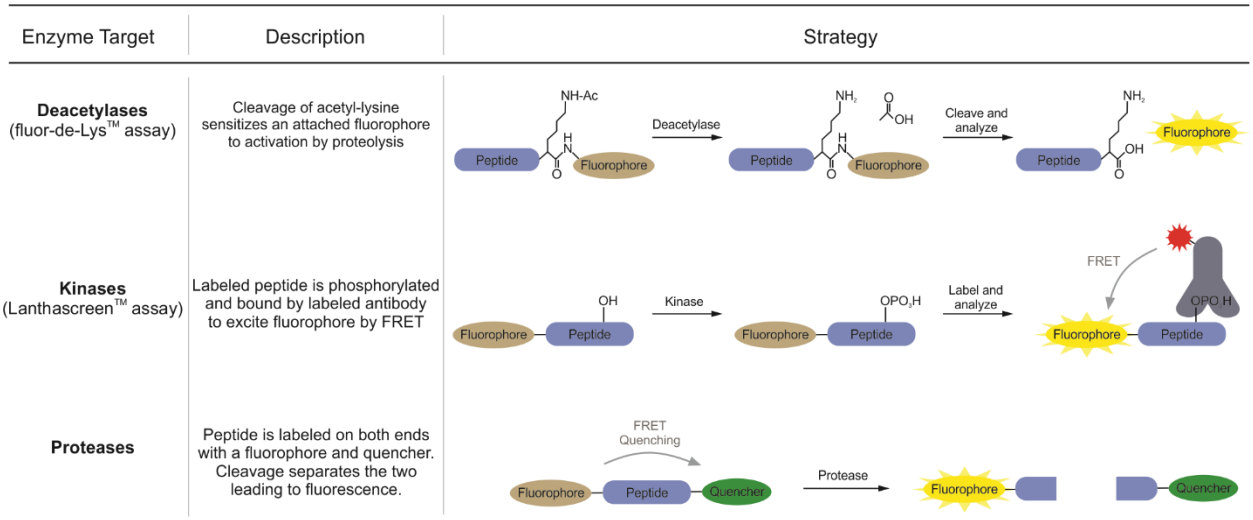


Figure 1.1. Examples of Commercial Assay Strategies. Labeled substrates, specific to an enzyme or enzyme family, are a common strategy for commercial assays for a wide range of target activities. Here, we can see several such labeling schemes.

The primary benefits of these methods are their commercial availability and straightforward analysis with affordable, and often readily available, fluorescence and absorbance plate readers. However, this diversity of labeling schemes, specific to a single enzyme or family of enzymes, highlights one of the major limitations. Namely, these labeling strategies must be painstakingly engineered on a case-by-case basis. They also require the use of modified, non-native, substrates which have in some cases changed the outcome of activity screens, identifying behavior that doesn't exist on the native substrate.[18] Some assay strategies seek to overcome the need for non-native substrates by tracking the consumption of cofactors instead, but not all enzymes use cofactors and these strategies do not work for complex systems such as lysates or cellular applications. Working towards the goal of more generalizable protein activity screening methods, our group has developed SAMDI mass spectrometry as a label-free and generalizable platform for high-throughput activity screening.

1.2 SAMDI Mass Spectrometry

In the Mrksich group, we have pioneered an assay platform called SAMDI (self-assembled monolayers for MALDI-TOF mass spectrometry). At its core, SAMDI combines the use of self-assembled monolayers for generating bio-arrays with MALDI-TOF mass spectrometry for rapid detection, creating a robust and generalizable biosensing platform.

Self-assembled monolayers (SAMs) offer a number of advantages that make them a powerful platform for the development of array-based biochemical assays. A key benefit is the ability to easily define and control the chemical structure of the reaction surface. The SAMs presented in this work utilize chemistry of alkyl disulfides on gold substrates that offer a number of benefits. The gold sulfur bond is relatively strong, resulting in a monolayer which is stable to a wide range of reaction conditions. Similar monolayers have been shown to be stable to temperatures as high as 80° C.[19] The monolayers can be rapidly generated by soaking the gold substrates in ethanolic solution of the alkyl disulfides and the chemistry at the terminal end of these disulfide molecules can be altered to control the composition of the surface, while mixed alkyl disulfides can be used to define the density of these functional groups. Surface composition is particularly important for application of monolayers to protein biochip assays. When using proteins, non-specific adsorption to surfaces can be an obstacle to effectively carrying out reactions on surfaces. The definable surface chemistry offered by SAMs provides a solution to this problem; monolayers terminated in ethylene glycol repeats have been shown to prevent non-specific adsorption, possibly as a result of interaction of the surface with the solvent and conformational entropy of the glycol groups providing a barrier to adsorption.[20]

The ability to control the chemistry at the surface also offers a convenient way to immobilize proteins, peptides, and small molecules for the purposes of creating biochips for assay development. A wide variety of ligation chemistries have been successfully used on SAM surfaces (**Figure 1.2**). These include maleimide-thiol ligation, click chemistry, Diels-Alder reactions, streptavidin-biotin binding, as well as others.[21-24] The work presented here uses primarily maleimide functionalized SAMs to immobilizes peptides and other small molecules of interest through incorporation of a thiol. The maleimide-thiol chemistry is a particularly attractive immobilization strategy for biochips as it is highly specific and does not require a catalyst. It is also very fast, with reported second order rate constants above 700 $M^{-1}s^{-1}$ in solution under mild, bio-compatible conditions, and we have observed rate constants on surfaces that reach 400 $M^{-1}s^{-1}$ in aqueous buffer at room temperature (see chapter 7).[25]

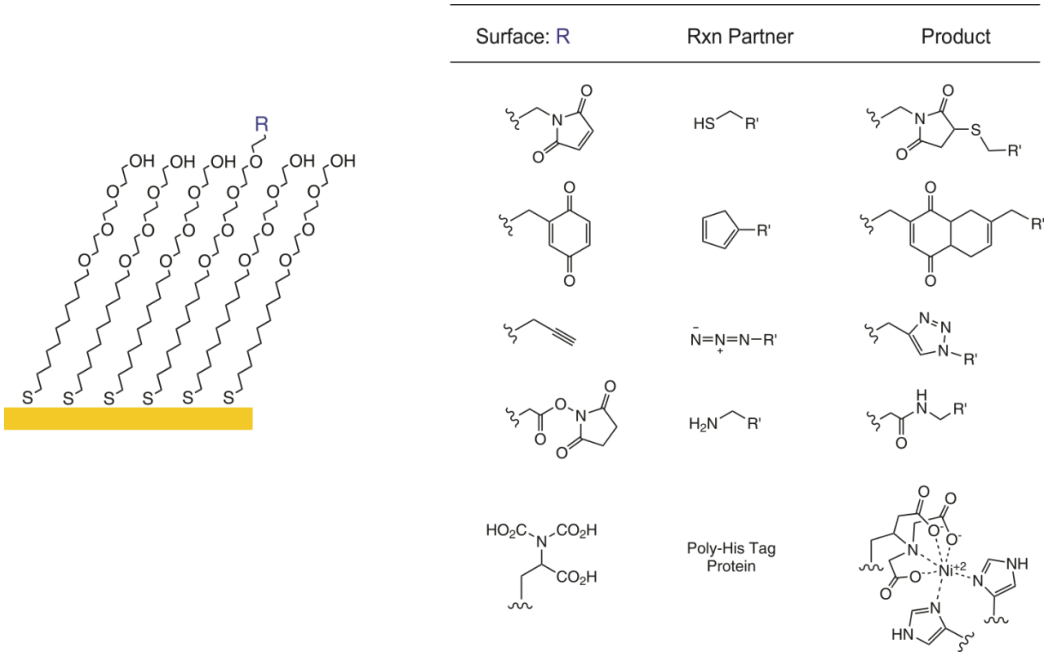


Figure 1.2. Chemistries for Surface Functionalization. A wide range of reaction schemes are compatible with SAMs for the purposes of generating functionalized biochips.

While SAMs provide a versatile platform for development of bioassays, MALDI-TOF mass spectrometry offers a rapid way to directly characterize the surfaces. MALDI-TOF (matrix assisted laser desorption ionization time-of-flight mass spectrometry) has become a popular form of mass spectrometry for detection of biomolecules, as it is effective over a wide mass range and uses a "soft ionization" method which results in little fragmentation, making it possible to visualize small molecules, peptides, and even intact proteins. In MALDI-TOF, purified samples are mixed with a solution of matrix compound and then dried onto a sample plate. The samples are ionized using a laser source, where the matrix serves to absorb the energy of the laser and transfer it to the analyte. While the details of this ionization process are not well understood, this strategy has been used for detection of small molecules weighing hundreds of Daltons to proteins with masses over 50 kDa.[26-27]

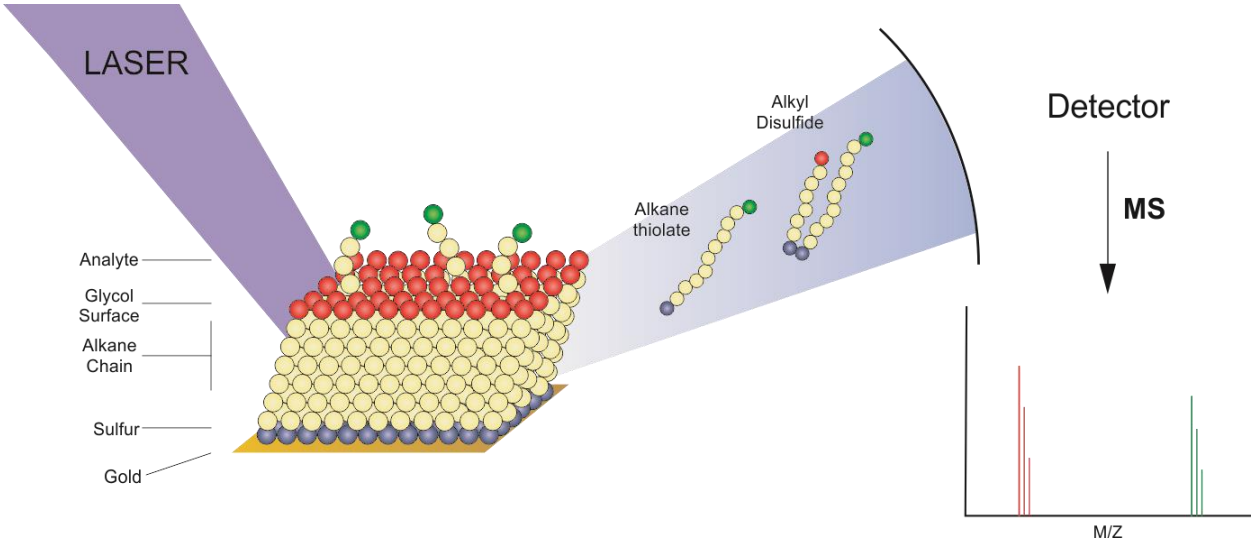


Figure 1.3. MALDI-TOF of Self-Assembled Monolayers. Self-Assembled monolayers of alkane thiolates on gold surfaces can be analyzed by MALDI-TF mass spectrometry. The laser ionization source cleaves the gold-sulfur bond and desorbs the monolayer. The resulting mass spectra shows peaks corresponding to the mass of any analytes immobilized to the surface

In 2002, our lab first reported that MALDI-TOF could be used to directly characterize self-assembled monolayers of alkanethiolates on gold surfaces.[28] This work showed that it was possible for the laser source of the MALDI-TOF to desorb the intact monolayer by cleaving the sulfur-gold bond at the surface, enabling any surface immobilized species to be easily characterized (**Figure 1.3**). Importantly, for the case of biochip assays, this makes it facile to directly detect a wide range of protein activities without the need for labels, tags, or the engineering of other activity-specific detection strategies. Since first publishing this concept, our group has shown that this strategy for assay development and reaction characterization is applicable to a broad range of biochemical activities. We have used this platform to characterize a diverse range of enzyme activities that include acetyl transferases, deacetylases, kinases, phosphatases, proteases, hydrolases, glycosyl transferases, and deaminases.[29-33] Additionally, we have shown that this strategy is applicable to immobilization and characterization of reactions on intact proteins.[22] Finally, in recent years, we have also demonstrated that this strategy is also applicable to quantifying protein activity in complex cell lysates, even showing that cells could be grown, lysed, and assayed directly on the SAMs.[34-35]

As we move beyond the early examples of SAMDI, which focused on biochemical systems wherein the activities of purified enzymes were characterized on surfaces presenting substrates, the work presented here describes my efforts to design novel assays that use the SAMDI platform while extending the types of biological and chemical processes that can be analyzed. Part of this story is the work done, by me and others, to scale this concept of reactions on chemically designed surface to increasingly-dense array formats, which enabled many of the large-scale experiments presented here (**Figure 1.4**). When I first began this work, most assays were done on small, 25 to

50-spot arrays, using glass microscope slides as a substrate. Moving to 384-spot arrays on steel substrates, that adopted a standard microtiter format, enabled compatibility with existing liquid handling robotics and made much larger experiments possible. More recently, 1536 arrays with a secondary hydrophobic monolayer on the steel background to facilitate liquid delivery, have become possible enabling even larger experiments to be done, at a scale of just 0.5 µL per reaction spot. The transition to even higher density 6144 arrays is currently limited primarily by the capabilities of liquid handling robotics. This progression of the underlying array technologies can be seen throughout this thesis, culminating with the application of MALDI Imaging Mass Spectrometry in chapter 8, wherein a continuous surface was "imaged" to generate pixels consisting of unique mass spectra at 200 µm resolution by the MALDI laser, rather than patterning discrete reaction "spots".

In this work, I will present the design and application of two high-throughput assays that take advantage of self-assembled monolayers as a controllable platform for performing biochemical reactions and MALDI-TOF as a rapid and generalizable detection strategy. In the first part of this thesis, I will describe an assay which uses SAMDI-MS to profile protein interaction domains and its application to the characterization of adaptor domains, protein domains that recognize and bind to post translational modifications. In the second part, I will demonstrate the design of an assay capable of rapidly characterizing all Coenzyme A metabolites from complex lysates and the application of this assay to optimizing a bioengineered, cell free reaction system.

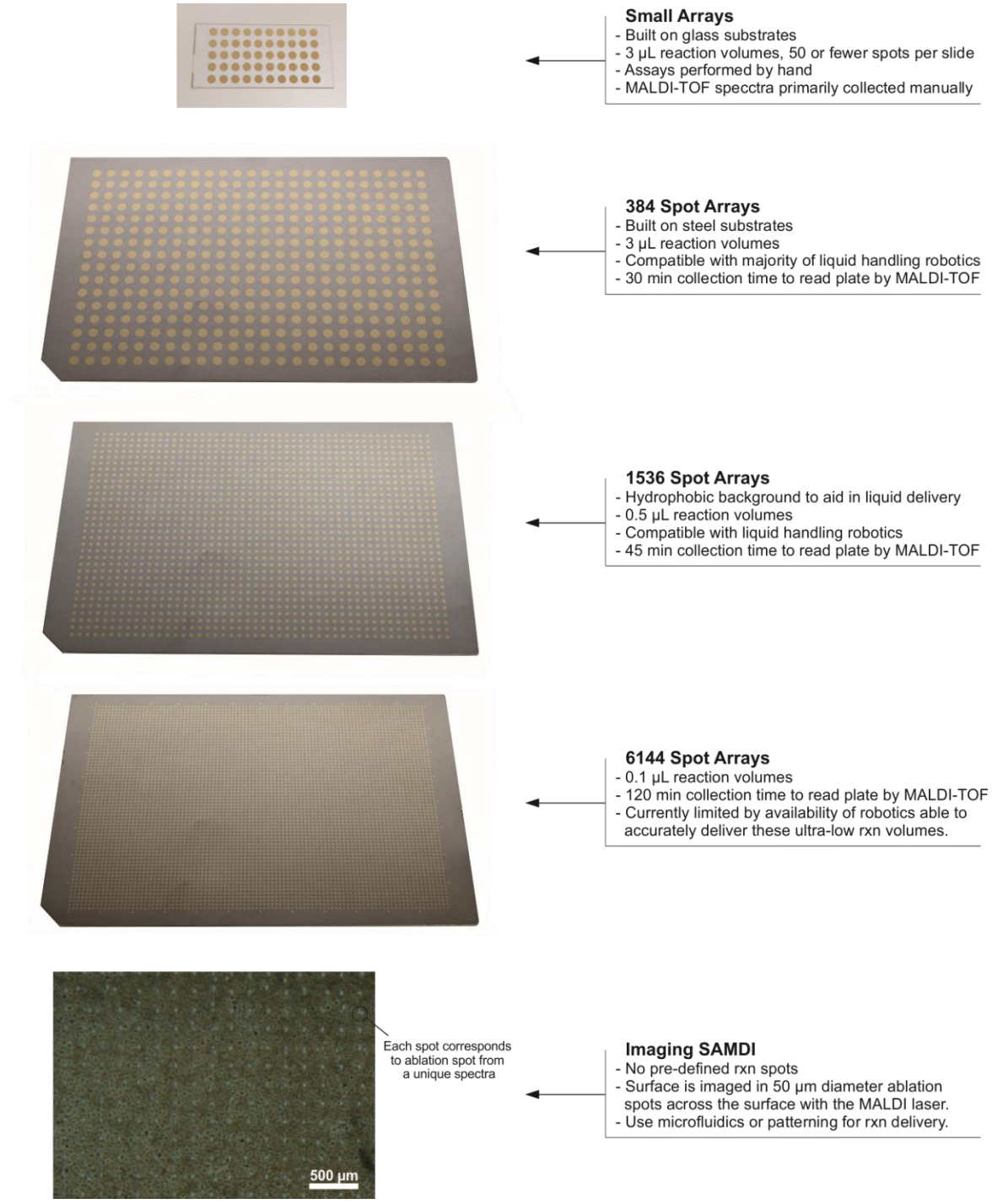


Figure 1.4. Advances in SAMDI Array Techniques. The design of surfaces and implementation of technologies has made higher density SAMDI arrays possible and enabled new kinds of experiments.

Part 1:

PI-SAMDI for High-throughput Characterization Of Protein Interaction Domains

Chapter 2

Adaptor Domains and their Role in the Histone Code

2.1 The Histone Code: Writers, Erasers, and Readers

Many of the enzymes that have been previously studied using SAMDI are involved in the addition or removal of post translational modifications (PTM's). PTM's are chemical modifications made to a protein after it has been expressed and often serve as an additional layer of regulation of the activity of these modified proteins. Common PTM's include phosphorylation, acetylation, methylation, ubiquitination, glycosylation, farnesylation, and others. These modifications are made in a site-specific manner to the side chains of amino acid residues in the protein and have the ability to affect the underlying activity of the protein.[36-38] Enzyme activity may be turned on or off by the addition of a PTM, or a PTM may act as a docking site in the formation of large, multiprotein complexes.

While PTM's can be found on proteins throughout the cell, they serve a particularly important role in DNA regulation and epigenetics. In eukaryotic organisms, DNA in the nucleus is packaged into a highly organized structure referred to as chromosomes or chromatin. Chromatin is made of DNA packaged around scaffolding proteins called histones.[39] The basic subunit of chromatin structure is the nucleosome, which refers to a length of DNA wrapped around an octameric complex of histone proteins. Nucleosomes, in turn, can form repeating structures that lead to higher order packing, and compacting, of the billions of base pairs worth of DNA present in every cell.[40]

However, these histones are not a passive packaging scaffold, but rather they provide a complex mechanism for regulating the underlying genes. While the DNA itself contains many

regulatory elements to control when and to what extent specific proteins are expressed, DNA packaging offers another layer of control by regulating whether a specific gene is physically accessible for expression. This is accomplished, in large part, through modification on the histone proteins. Histones contain long, structurally disordered, tails that can be heavily post-translationally modified.[41-42] This idea evolved into what is now referred to as the Histone Code; the concept that a specific set of modifications can be linked to a specific expression state of the underlying gene.[43-44] In the context of this code, enzymes that add or remove PTM's can be thought of as Histone Code "writers" and "erasers", respectively.

While the idea behind the histone code is rather straightforward, in practice it has proven to be a complex regulatory network and despite many groups contributing to the collective knowledge of the code, there remains much to explore. One aspect contributing to the complexity is crosstalk, the idea that PTM's on histone tails are not independent of each other. One PTM has the potential to affect the ability of enzymes to add and remove neighboring PTM's, both within the same histone tail and even across nearby nucleosomes. In our own group, we have identified the activity of the KDAC8 deacetylase on Histone 3 protein at the Lysine in position 12 (H3K12) is sensitive to secondary PTMs at both nearby and distal sites.[45] Another example comes in the form of the dependence of acetylation at H3K14 by the enzyme p300 on citrullination of H3R8.[33] Numerous other examples of such crosstalk have been identified, including cases where modifications on one histone affect those on a different histone protein entirely.[46]

Another area of interest with respect to the histone code is identifying the mechanisms by which post translational modifications lead to changes in gene expression. Originally, it was believed that PTM's acted by affecting the strength of the binding interaction between the

phosphate backbone of the DNA and the histone protein complex. In this model, PTM's that strengthen the interaction between the DNA and the histone proteins would lead to increased affinity and make it more difficult for the transcriptional machinery of the cell to access the underlying gene, leading to a decrease in expression. Conversely, PTM's that weaken the interact with DNA would allow the DNA to be more easily transcribed and the genes would be expressed at high levels. While this idea was intuitively attractive due to its simplicity, there is relatively little literature to support this as a major mechanism of action.

Instead, as more work was done, the more plausible mechanism through which PTM's affect gene expression became that they serve as docking sites. Specific PTM's on histone tails act as recognition sites for other proteins. Proteins, or multiprotein complexes, bind to the PTM and mediate changes in expression of the underlying genes. For most common PTM's, corresponding protein domains that bind them have been identified and many of them have their own sequence preferences, as do the enzyme writers and erasers.[47-48] These protein domains that recognize, bind specific PTM sites, and mediate signal transduction are referred to as "adaptor domains". For many of the known PTM's, families of adaptor domains have been discovered that recognize and bind to them (**Figure 2.1**), examples include the bromodomain which binds acetylated lysine and the SH2 domain which binds phosphorylated tyrosine.[49-50]

These adaptor domains constitute the final piece of the histone code. They serve as the "readers" that take the modifications made by the "writer" and "eraser" enzymes and translate those modifications into a change in gene expression. For instance, histone acetylation is generally associated with increased expression of the underlying genes and phosphorylation has been shown to help initiate and regulate cell division.[51]

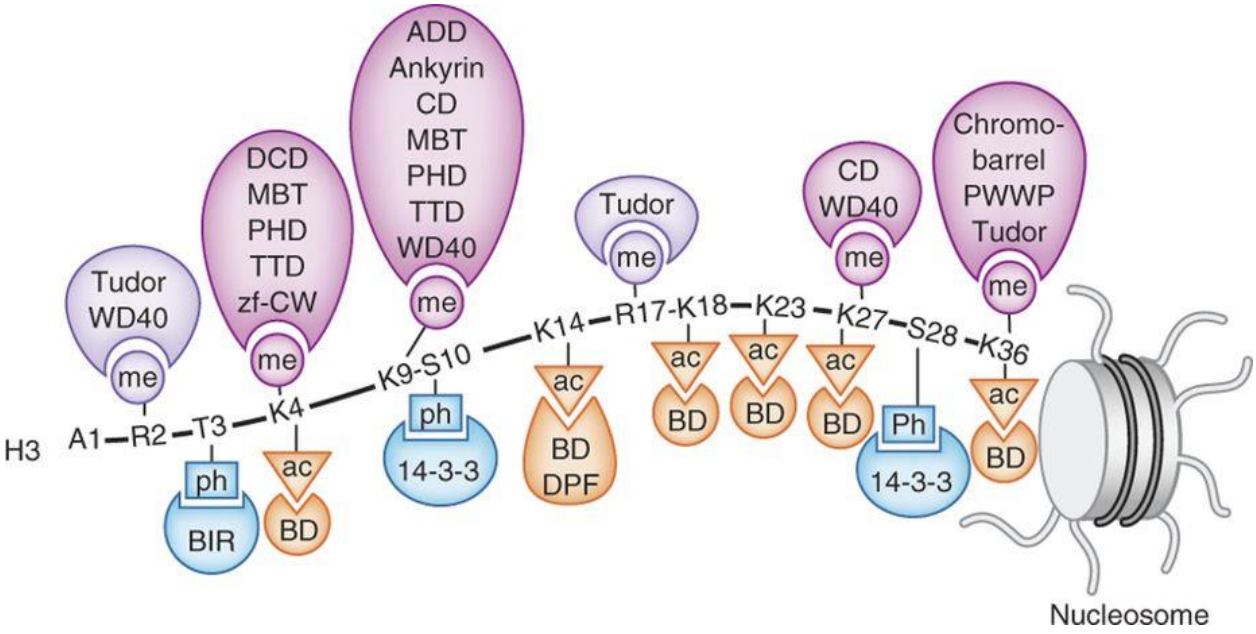


Figure 2.1. PTM Readers. For many common PTM's, shown here for histone H3, corresponding reader domains that bind them have been identified. (reproduced with permission[47])

2.2 Chromodomains, Methyl Lysine Readers

One such family of adaptor domains that I will explore in this work are the chromodomains. This family of evolutionarily related protein domains binds to sites of methylation on lysine residues. In humans, there are eight chromodomains, Cbx1 thru Cbx8 (chromobox protein homolog) but these can be further subdivided into two categories based on their homology to the canonical chromodomains, first identified in *Drosophila*.[52] Of the eight human chromodomains, three (Cbx1, Cbx3, and Cbx5) are homologous to the *Drosophila* HP1 (heterochromatin protein 1) and the remaining five are homologous to the *Drosophilia* Pc protein (polycomb group protein).[53]

While some adaptor domains are present in proteins with enzymatic domains, the proteins that contain both the HP1 and Pc chromodomains lack any enzymatic activity. Additionally, the HP1 chromodomains contain a nearby chromoshadow domain, which shares structural homology to the chromodomain but does not bind methylated lysine.[54] Instead, the chromoshadow domain is involved in homo and hetero-dimerization. The canonical binding site for the HP1 chromodomains is methylated H3K9 (that is, histone protein 3 at lysine residue in position 9), and the canonical binding site for the Pc chromodomains is methylation H3K27.[55-56] With the lysine methylation, three different states are possible; mono-, di-, or trimethylation. With respect to affinity, these chromodomains are generally known to favor higher methylation states.

The function or importance of these domains is not completely understood. Methylation is generally associated with lower levels of expression of the underlying genes. The Pc chromodomain is a member of the polycomb group proteins in *Drosophila*, where it was first discovered. Polycomb group proteins serve regulate homeotic genes, those required for early development, and may serve important roles in human embryonic development as well.[57] The HP1 proteins are found bound to tightly packaged heterochromatin regions and are enriched in centromeres and telomeres, regions of low gene expression.[58] HP1 may serve as a docking site for other proteins and there is evidence that it serves a role in DNA repair.[59]

2.3 Assays for Profiling Adaptor Domains

Adaptor domains serve diverse and important roles as their writer and eraser counterparts. However, while numerous assays exist for characterizing the enzyme writers and erasers, the adaptor domains that read these modifications are far less explored by comparison. The major

cause of this is the tools available to profile these interactions; while they do exist, each comes with their own limitations.

Among the most quantitative techniques available for binding interactions such as these is calorimetry. Modern isothermal calorimetry instrumentation can quantitatively describe the kinetics of binding, and measurements can be performed on a sample scale of a few hundred microliters.[60] However, throughput is limited to a rate of 1-2 hours per measurement and, when working with protein samples, 200 µL at concentrations in the hundreds of µM is still an enormous amount of material needed for each measurement. These relatively large sample sizes and low throughput means ITC is best suited for characterizing known ligands rather than screening potential ligands.

Fluorescence polarization is another technique that can be used to determine dissociation constants for protein interactions.[61] This technique provides an increase in throughput and is compatible with significantly smaller sample sizes. This technique uses excitation with a polarized light source and relies on measurable change in signal when a protein of interest binds to a fluorescently labeled ligand and limits its ability to freely rotate. The major disadvantage to this technique is the need for fluorescently labeled ligands. This adds extra complexity to the production of ligand libraries and fluorophores can significantly affect the affinity of ligands for the target. In some cases, the use of fluorophores can change the results of a screen and led to false positives.[62-63]

The final commonly used technique for characterizing adaptor domains are SPOT binding arrays. SPOT arrays utilize cellulose substrates, on which arrays of peptide binding ligands are synthesized.[64] These membrane arrays are blocked, incubated with a protein of interest, rinsed,

and labelled.[65] They have the potential to be relatively high throughput, with hundreds of peptide ligands on a single membrane, but they possess several limitations. The readout of this assay is primarily qualitative and comparative, with the ability to either see or not see binding. The nature of the detection means only large changes in affinity are observable and the dissociation constant for an interaction cannot be determined using this technique.

A handful of other techniques with limited throughput used in certain, more limited contexts, include NMR-based measurement of interactions and SPR (surface plasmon resonance). SPR, in particular, can be a powerful and analytical technique for binding interactions but it requires expensive, specialized instrumentation, throughput is limited to one measurement per 30-60 minutes, and material used can be relatively high.

Chapter 3

Characterization of Chromodomains by PI-SAMDI

This chapter is adapted from the following published work:
O'Kane, P.; Mrksich, M. *"An Assay Based on SAMDI Mass Spectrometry for Profiling Protein Interaction Domains"* JACS. 2017, 139 (30), 10320-10327.

3.1 Introduction

In the work presented here, we sought to develop an assay to profile protein interaction domains based on SAMDI-MS that could address some of the limitations in the tools currently available. We have demonstrated SAMDI to be a general and powerful assay platform for profiling a wide range of enzyme activities, one which offers several key advantages over other bioassay strategies. The surfaces used are made naturally resistant to protein adsorption using a tri(ethylene)glycol terminated background SAM; this improves reproducibility and obviates the need for blocking procedures used in other surface assays, such as SPOT arrays. Performing reactions on surfaces allows for simple purification by rinsing, and detection by mass spectrometry enables direct quantification without the need for labels.

However, unlike previously profiled enzyme activities, adaptor domains do not modify their ligands in a manner directly detectible by mass spectrometry. While high-affinity interactions can allow bound protein to be detected directly by SAMDI,[22] the interaction domains we are interested in here possess relatively weak affinities for their binding partners. Instead, our design strategy uses unique aspects of enzyme reactions on surfaces, by which an adaptor domain mediated binding can enhance the activity of an associated enzyme.

In previous work, it was observed that the activity of SRC kinase on surface immobilized substrates displayed unique kinetic characteristic due to the phospho-tyrosine binding SH2 domain

native to this protein.[31] A substrate for the kinase was patterned on a surface and the reaction rate was monitored by SAMDI-MS. It was found that the surface reaction demonstrated a unique kinetic profile, where the rate appeared to accelerate as the surface was phosphorylated. This effect was caused by the SH2 adaptor domain; as the surface was phosphorylated these became binding sites for the SH2 domain, which could localize the enzyme to the surface and increase the effective enzyme concentration experienced by the remaining substrate. This auto-catalytic cycle arises because the substrate is isolated to the surface and isn't observed when the reaction is done in solution.

A similar effect was observed when the SH2 domain of SRC kinase was fused to cutinase and surface activity of cutinase was profiled on an electroactive substrate in the presence of SH2 ligands.[66] In additional unpublished experiments the same SH2 domain was fused to the enzyme KDAC8 and it was shown that the magnitude of the effect was dependent on the affinity for the ligand. KDAC8 is a lysine deacetylase, so in this system the function of the enzyme and the function of the SH2 domain are orthogonal to one another. When a panel of KDAC8 substrates (peptides containing acetyl-lysine) were co-patterned on surfaces with a panel of SH2 ligand peptides. It was found that ligands for the SH2 domain could increase the observed rate of deacetylation in a manner that was dependent on the affinity for the ligand, with stronger affinity ligands giving greater increases in enzyme activity. The effect was most dramatic when the substrate for the enzyme was poor, so that the activity in the absence of this adaptor domain mediated interaction was low.

These observations shaped the design of this new assay that we used to profile the ligand preferences of human chromodomains (**Figure 3.1**). In this scheme, we prepare the adaptor

domain to be assayed as a fusion to an enzyme domain and we also prepare monolayers to which two peptides have been immobilized; one is a prospective ligand for the adaptor domain and the other is a substrate for the reporter enzyme. In this method, it is important to select a poor substrate, ensuring minimal modification of the substrate by the soluble enzyme but efficient modification of the substrate when the enzyme is brought to the surface by way of a protein-ligand interaction. The fraction of time that the protein-ligand interaction exists can then be determined by using SAMDI to measure the extent of substrate conversion, where the reporter enzyme activity on the surface serves as a permanent, covalent record to the interact. We refer to this strategy for monitoring binding interactions as PI-SAMDI, or protein interactions by SAMDI.

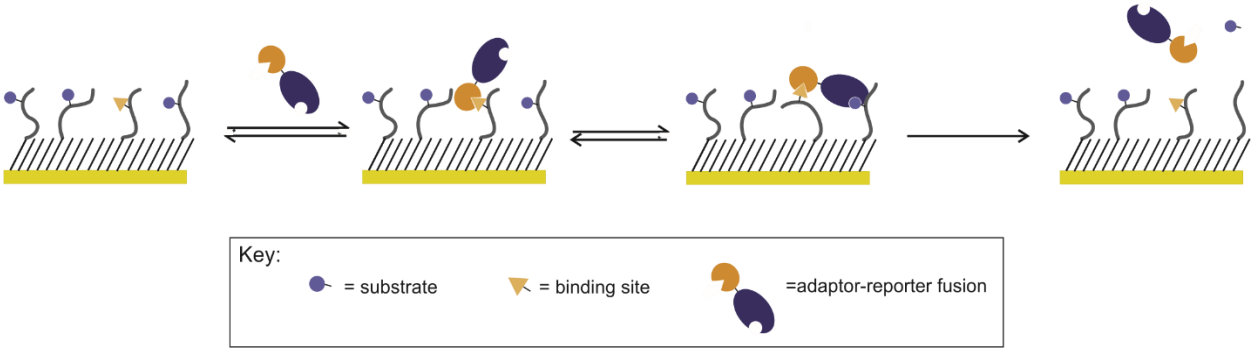


Figure 3.1. Surface Reaction Scheme. Adaptor domains can increase the rate of an enzymatic reaction on a surface in the presence of a ligand for it to bind. The adaptor domain localizes the enzyme to the surface when it binds, increasing the effective enzyme concentration at the surface and accelerating the observed rate of substrate conversion.

3.2 Results

Protein Expression and Assay Design

We generated artificial fusions of the human chromodomains to a reporter enzyme. Ligand peptides containing methyl-lysine could then be patterned onto SAM's presenting maleimide,

along with enzyme substrate. The surfaces were then incubated with the chromodomain-reporter enzyme construct and the relative affinity of the chromodomains for various ligands could be rapidly analyzed by looking at the change in perceived enzyme activity using SAMDI-MS. In essence, the reporter enzyme generates a covalent record of the transient interaction of the adaptor domain and its ligand. This covalent record allows even relatively weak interactions to be detected.

We prepared five of the eight human chromodomains as fusions to KDAC8. These included three of the HP1 homologs (Cbx1, Cbx3, and Cbx5) and both of the Pc homologs (Cbx4 and Cbx6). KDAC8 was chosen as a reporter enzyme for several reasons. It has robust activity, it has been previously expressed as a fusion protein without loss of activity, and the activity on a wide range of substrates has been well characterized in our group, a valuable advantage for selection of a poor substrate. In the work presented here, the peptide, with sequence GMK(Ac)FGC, was used as the reporter enzyme substrate across all assays.

Fusion proteins were constructed using the commercial pET-303 CT-His tag vector. We first used PCR to amplify the KDAC8 catalytic site while inserting a coding region for a linker sequence in front of the catalytic sites and restriction sites that were used to digest and ligate the sequence into the vector. Through similar amplification, digestion, and ligation the chromodomains were then introduced in front of the KDAC8 catalytic domain (**Figure 3.2**). We expressed the constructs in the BL21 (DH5α) *E.coli* strain using a standard IPTG induction method, and then purified the expressed protein by cobalt column affinity chromatography followed by FPLC using size exclusion chromatography.

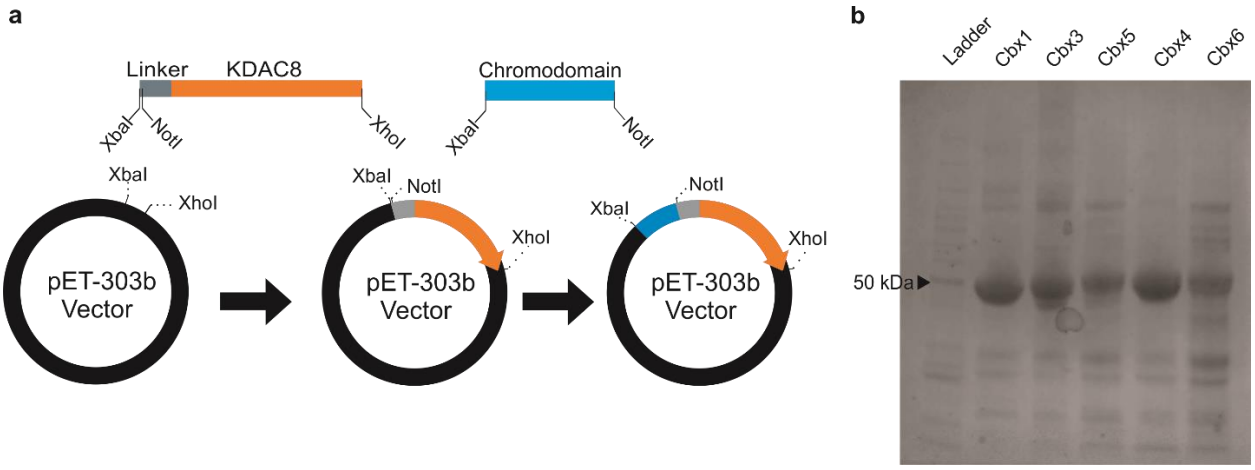
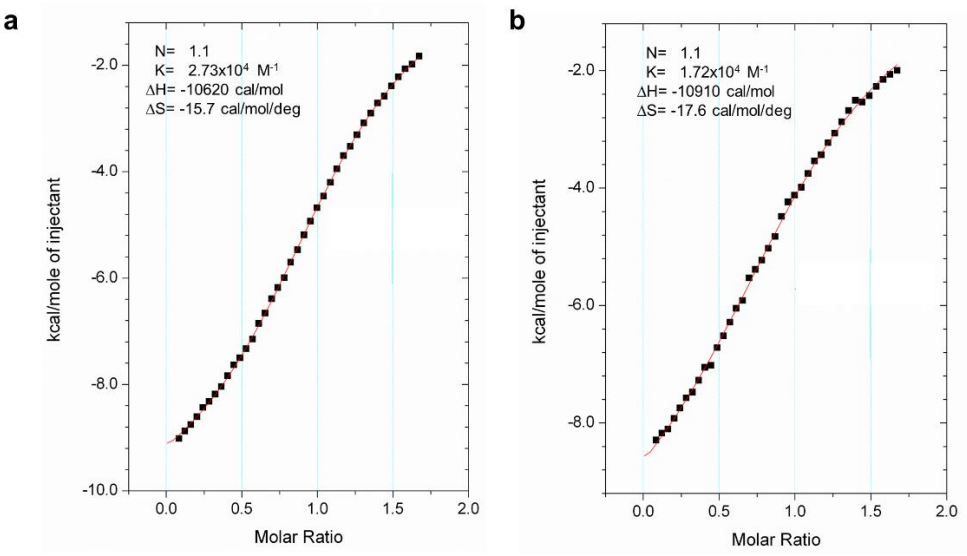


Figure 3.2. Design and expression of artificial fusions. a) Fusion proteins of chromodomains (CDs) to KDAC8 were designed by sequential digestion and ligation of the KDAC8 catalytic domain and the CD's into pET-303b vector. **b)** These artificial constructs are shown here on a SDS-PAGE gel after expression in *E.coli* and purification by cobalt column affinity chromatography. Proteins were further purified by FPLC on a size exclusion column before use.

For one of the fusions, Cbx3-KDAC8, we used isothermal calorimetry to verify affinity for a peptide possessing the sequence of the canonical binding sequence, H3K9, in both tri- and di-methylated forms (**Figure 3.3a-b**). Isothermal calorimetry uses the heat of association to quantitate the extent of binding, wherein the protein sample is loaded into a sealed chamber and small amounts of peptide ligand are injected sequentially. From this the dissociation constant, along with other parameters, can be determined. The Cbx3 was loaded into the 300 µL sample chamber at a concentration of 200 µM and the ligand peptides were injected at a concentration of 1.6 mM, with 40x 1 µL injections spaced 150 seconds apart. We found that the measured affinity agrees with available literature, as does the stronger observed affinity for the higher methylation

state suggesting that the chromodomain remains functional in the context of the fusion protein (**Figure 3.3c**).



c

Peptide	Dissociation Constant (μM)
H3K9Me3	36.6
H3H27Me2	58.1

Figure 3.3. Isothermal Calorimetry of Cbx-KDAC8 Fusion: Isothermal calorimetry was performed on the Cbx3 artificial fusion protein to determine the dissociation constants for the canonical H3K9 binding site; both tri- (**a**) and di- methylation (**b**) states were measured. The resulting binding curve for the trimethylated peptide and dimethylated peptide could then be used to calculate the affinity. **c)** Experimentally calculated dissociation constants show higher affinity for the trimethylation over dimethylation.

We prepared self-assembled monolayers presenting maleimide-terminated alkanethiolates at a density of 10% against a background of tri(ethylene glycol)-terminated alkanethiolates. The latter are very effective at preventing non-specific adsorption of protein. The maleimide group was used to immobilize cysteine-containing peptides by way of a Michael addition.[24] For the assay

presented here, it is important to choose appropriate densities for the ligand and reporter substrate. Previous work from our group has shown that a ligand density of approximately 2% provides the maximum increase in rate acceleration for the interfacial enzyme reaction, with diminishing benefit from increasing the density further. Therefore, we immobilized the two peptides (the GMK(Ac)FGC substrate for KDAC and a second peptide that is a potential ligand for the chromodomain) in a ratio of 4 to 1, for final surface densities of approximately 2% for the chromodomain ligand and 8% for the reporter enzyme substrate. This was done by preparing buffered solutions containing the two peptides at the appropriate molar ratio and incubating these solutions with the maleimide presenting surfaces (**Figure 3.4a**).

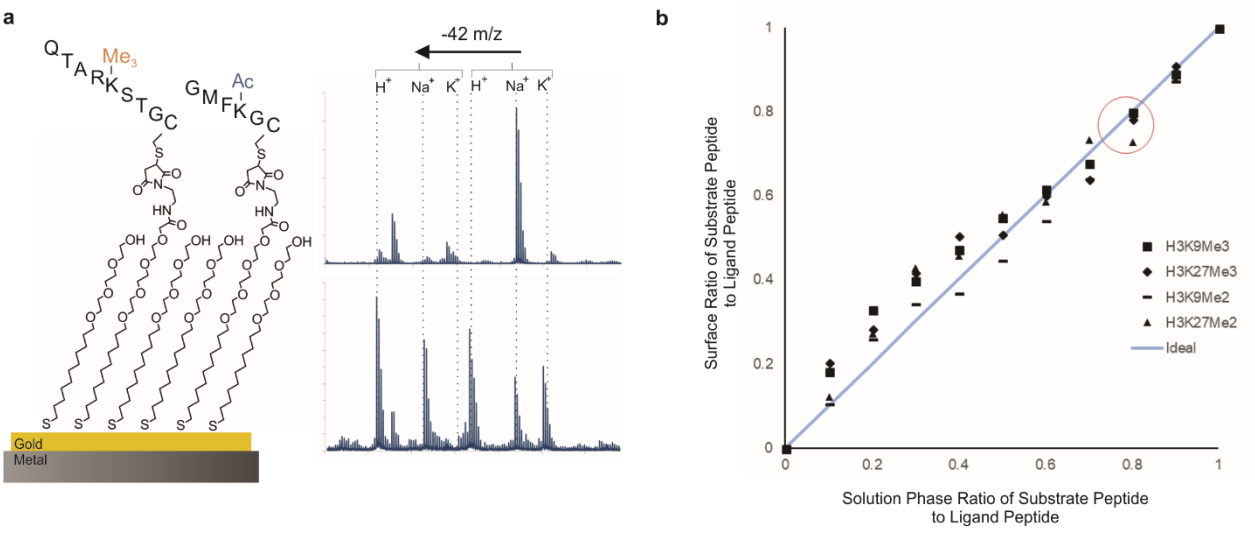


Figure 3.4. Description of Assay Surfaces. **a)** For assaying ligand affinity, surfaces were patterned with both substrate for KDAC8 and methylated ligands for chromodomains. Reporter activity was quantified by measuring deacetylation (sum of proton, sodium, and potassium adducts) by SAMDI-MS. **b)** Various ligands were co-immobilized with reporter peptide, GMK(Ac)FGC by mixing at stoichiometric ratios. Using the ethylene glycol background monolayer as an internal standard, we found relative solution phase concentration to track linearly with ratio of the two peptides on the surface.

We performed a control experiment to verify that the solution phase ratio did, indeed, translate well into the same ratio on the surface (**Figure 3.4b**). For four of the methylated ligand peptides used in this work, we prepared a gradient of molar ratios of the peptides with the KDAC8 substrate peptide (with sequence GMK(Ac)FGC), ranging from a ratio of 0 to 1 fractional content of the ligand peptide relative to the KDAC* substrate peptide. These solutions were immobilized to the same 10% maleimide presenting surfaces used in this work, and SAMDI mass spectra were collected for each. The surface ratio of the two peptides was determined for each condition using the ratio of the integrated area under the peak (AUP) for the two peptides. We normalized these values to the AUP of the tri-ethylene glycol background monolayer, present at a constant 90% density across all the surface, to account of differences in ionization efficiency between peptide sequences. We found the experimentally determined ratio of surface densities to track very closely with the corresponding solution phase molar ratios of the peptides.

Profiling Chromodomain Activity

Next, we used this platform to characterize affinity of these Cbx chromodomains on peptide ligands. All reactions were performed on plates patterned with 384 gold islands arranged in the geometry of a standard microtiter plate as described previously.[67-69] These array plates are compatible with liquid-handling automation, and only require 3 μL of reaction mixture for each spot. To initiate reactions, we used a Mutlidrop Combi benchtop robot (ThermoFisher) to deliver a solution of the chromodomain-KDAC fusion in a Tris buffer to each spot and allowed the reactions to proceed for up to 30 minutes at 37°C. The array plates were rinsed, treated with matrix (THAP in acetone) and analyzed by MALDI-TOF mass spectrometry to quantitate the amounts of

substrate and product. The extent of conversion of the substrate on the surface was determined using the AUP for the substrate and product.

$$Conversion = \frac{AUP_{product}}{AUP_{product} + AUP_{Substrate}} \quad (1)$$

We first performed the assay on monolayer arrays presenting peptides that represent the well-characterized ligands for the chromodomain proteins. These four peptides include either di- or tri-methylation at residues K9 or K27 (**Figure 3.5a**). In each case control spots are included that present the KDAC peptide substrate but that omit a second peptide ligand. In this way, we can measure the extent of enzyme activity that is due only to the action of a soluble (unbound) enzyme. We present the data as "fold enhancement" which corresponds to the average percent yield of the reaction in the presence of a given ligand peptide, divided by the average percent yield of the enzyme reaction in the absence of a ligand, i.e. the enhancement due to ligand binding (Equation 2). We performed at least four replicates of each experiment on the array plate.

$$Fold\ Enhancement = \frac{Yield\ with\ ligand}{Yield\ without\ ligand} \quad (2)$$

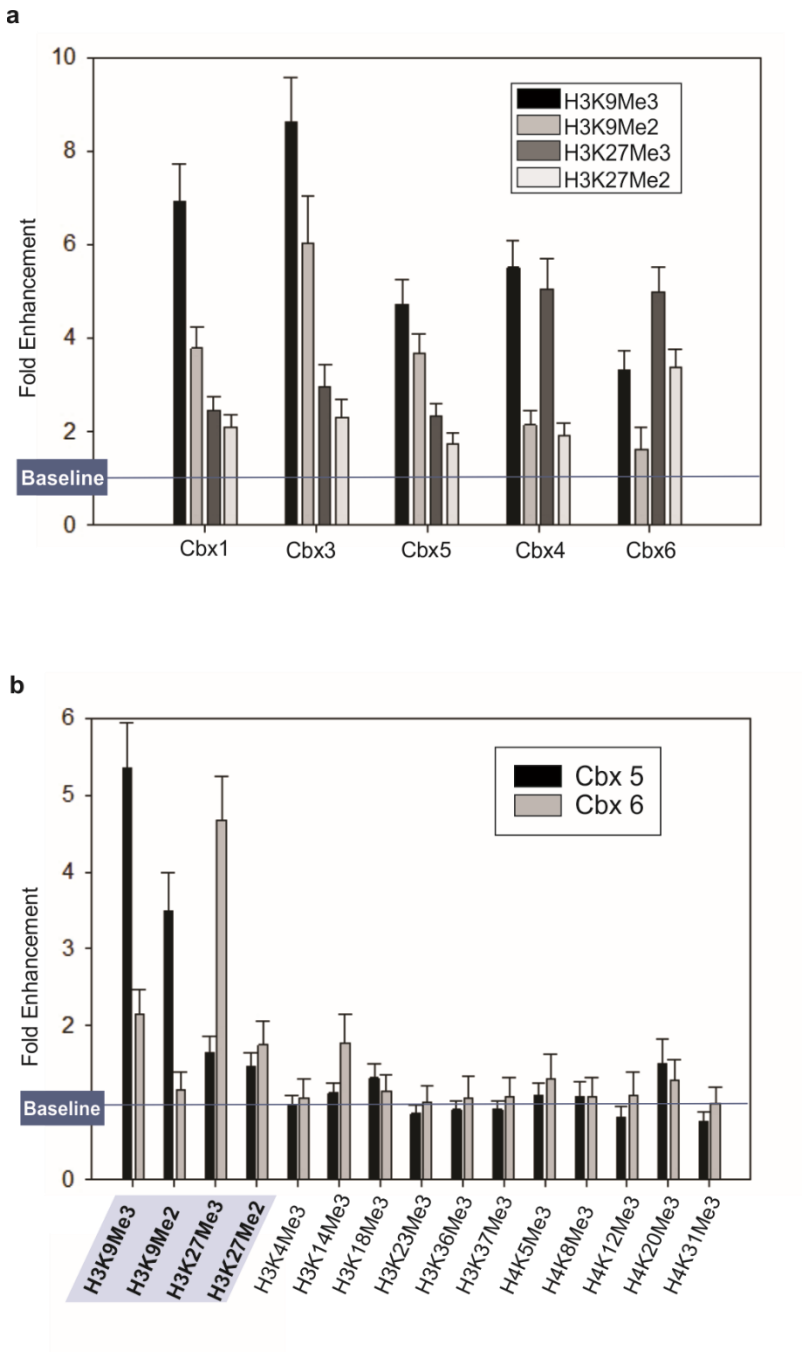


Figure 3.5. Cbx Fusion Activity on Canonical Substrates. a) The relative activity of each of the chromodomain fusions was measured on peptides representing the di- and tri-methylated canonical H3K9 and H3K27 binding sites. **b)** The activity was also observed on a small library of peptides representing all possible H3 and H4 methylation sites for the Cbx5 and Cbx6 chromodomains. Little activity was observed for the noncanonical ligands.

For the HP1 chromodomains (Cbx1, Cbx3, Cbx5) we observe greater rate enhancement for di- and tri-methylation at the H3K9 site, with the latter giving approximately eight-fold enhancements. These observations are consistent with previous reports, where it has been shown the HP1 chromodomains have greatest affinity for methylated H3K9 and a preference for higher methylation states. For the Pc chromodomains (Cbx4 and Cbx6), a preference for higher methylation is also observed. But while Cbx6 prefers the canonical H3K27 ligand, Cbx4 has a slight preference for the H3K9 methylation site. This lack of selectivity for H3K27 has also been previously reported, where it appears that the mammalian Pc homologs do not necessarily maintain the specificity of their *Drosophila* counterparts and certain isoforms lack specificity.[70]

These results validate the ability of PI-SAMDI assay to measure previously known protein-ligand interactions and to rank the interactions by affinity. Prior work has found no evidence that methylation at the other lysine residues in the amino-terminal tail of histone 3 gives sites that are ligands for the chromodomains. To explore the activities of these other sites, we next performed the PI-SAMDI assay on arrays that include peptides that are tri-methylated on each of the thirteen lysine residues derived from the Histone 3 and 4 tails (**Figure 3.5b**). These peptides represent methylation of residues 4, 14, 18, 23, 36 and 37 of Histone 3 and residues 5, 8, 12, 16 of Histone 4. We used the PI-SAMDI assay to measure binding of Cbx5 (from the HP1 family) and Cbx6 (from the Pc family) chromodomains to these peptides. As expected for these non-native positions, there is little to no observed increase in KDAC activity and, therefore, little binding of the chromodomains to these other sites.

We next used the PI-SAMDI assay to identify whether binding of chromodomains to the H3K9/K27 sites were influenced by a second post-translational modification in the peptide

ligands—that is, whether there existed a 'cross-talk' of distinct modifications to the peptide (**Figure 3.6**).[71] It is common for histone tails to possess multiple modifications *in vivo* and the residues surrounding the H3K9 and H3K27 sites are also frequently post-translationally modified. We synthesized nine peptides that were trimethylated at either H3K9 or H3K27 and that further had a secondary site of phosphorylation, acetylation, citrullination, or methylation; these modifications have all been reported to occur in chromosomal histones.[72-78] We find that some of these secondary modifications can strongly influence the affinity of the chromodomain. Specifically, nearby phosphorylation acts to dramatically decrease affinity, and in some cases, causes complete loss of binding all together. For the H3K9 methylation site, nearby phosphorylation at T6, S10, or T11 dramatically reduces the affinity of the chromodomain. Citrullination at R8 strongly reduces affinity as does methylation at R8, but to a lesser extent.

At the H3K27 methylation site, we again see strong reduction in affinity when nearby S28 or S31 residues are phosphorylated. Additionally, there is some sensitivity to methylation at K23, whereas acetylation at K23 has no observable effect on affinity of the PC chromodomains.

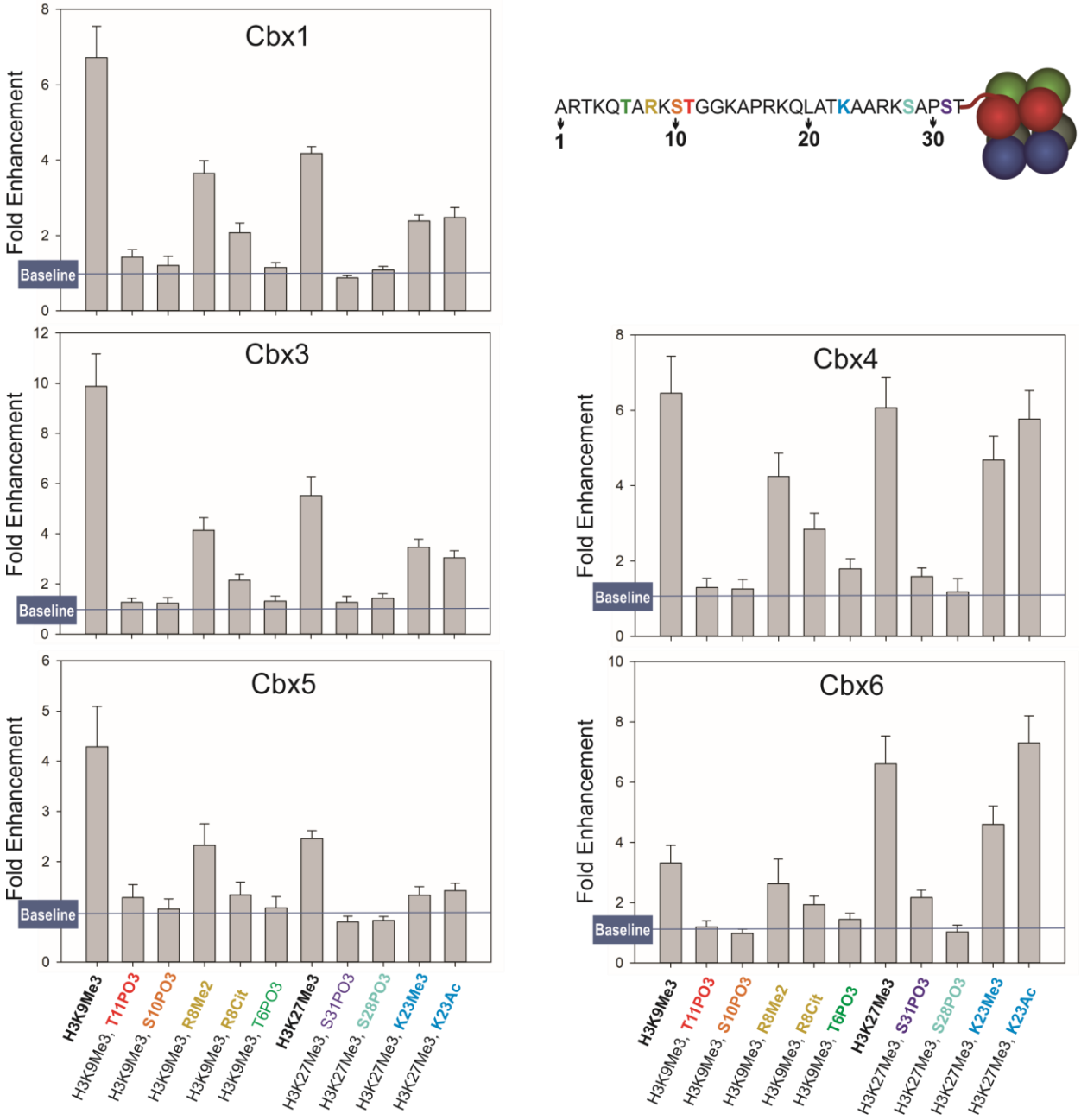


Figure 3.6. Effects of nearby PTMs on binding. The affinity of each chromodomain for the H3K9Me3 site was evaluated for peptides that additionally had a second post-translational modification. The HP1 isoforms Cbx1, Cbx3, and Cbx5 are shown in the left panel and the Pc isoforms Cbx4 and Cbx6 are shown in the right panel. The error bars represent the standard deviation of four replicates. The secondary modification on each peptide is color-coded to match the sequence shown in the upper right.

3.3 Conclusions

We describe a new label-free assay for measuring protein-ligand interactions. The assay relies on a coupling of the protein-ligand binding event to an enzyme-mediated modification of a peptide or small molecule on the monolayer. In this way, the enzyme can modify nearby substrates when the protein interaction domain binds to its ligand, leaving a 'covalent record' of the transient complex. The yield of the enzymatic reaction increases with the fraction of time the complex is present, allowing this method to rank affinities of a receptor for several prospective ligands. Because this assay does not use affinity methods to isolate the complex, it is useful for measuring low affinity interactions. Second, because it uses standard microtiter plate formats and automation, it can be performed in high throughput with small sample volumes.[69] In this way, the PI-SAMDI assay represents an important complement to the use of fluorescence-based assays, which also have substantially higher throughput than assays based on isothermal calorimetry and surface plasmon resonance. However, it can be challenging to optimize the attachment of fluorophores to proteins and the fluorophores can often lead to artefacts due to non-native interactions with the binding partner.[79-81] This assay strategy also possess the additional benefit of being directly compatible with small molecule screening, as will be demonstrated in the next chapter.

The use of self-assembled monolayers was critical in enabling the assay reported here. The oligo(ethylene glycol)-terminated monolayers are highly effective at preventing the non-specific adsorption of protein, which would block the ligand and enzyme substrate from interactions with the fusion protein. The monolayers also offer excellent control over density of the two peptides, which allows optimization of the sensitivity of the assay. The use of mass spectrometry to analyze the monolayers provides a sensitive detection method that benefits from the high throughput

implementation of SAMDI, which can measure tens of thousands of spots per day.[69, 82] The assay also allows the flexibility to use a broad range of reporter enzymes in the assay.

Prior work has established that the human homologs of the *Drosophila* HP1 protein (here Cbx1, Cbx3, and Cb5) share a specificity and strongly prefer binding the H3K9 position and higher methylation states; our data using the PI-SAMDI assay are in agreement with these trends. The specific affinities of these proteins have been examined with various methods, with the *Drosophilia* HP1 chromodomain generally reported to have a K_D in the range of 1-15 μM for the H3K9 trimethylated site. While mammalian HP1 homologs have reported K_d values as high as 30 μM.[53, 65, 83]

The human homologs of the *Drosophila* Pc proteins (here Cbx4 and Cbx6), in contrast, are promiscuous and bind both H3K9 and H3K27, unlike the *Drosophila* Pc protein which has selectivity for H3K9. For the *Drosophila* Pc homologues, we find that Cbx4 has a modest preference for H3K9, which has also been previously observed by fluorescence polarization. Where dissociation constants have been measured for mammalian Cbx6, the protein prefers the H3K27 position but with a weaker affinity as compared to other Pc homologs with a K_d well over 100 μM.[65, 70] The binding was measurable by our PI-SAMDI assay and we find that Cbx6 binds to both sites with a modest preference for H3K27 over H3K9. Generally, though, the trends that we observed in relative preferences agree with the picture developed from the various biochemical and biological experiments available in the literature on this family of proteins.

The compatibility of PI-SAMDI with peptide arrays makes it straightforward to evaluate binding interactions to libraries of prospective ligands. In this work, we prepared arrays wherein the peptides had a trimethylated lysine either at H3K9 or H3K27 and further had one additional

post-translational modification, with the goal of surveying potential 'cross-talk' interactions in chromodomain binding. We found that all five chromodomains had no measurable affinity for H3K9 in peptides that were phosphorylated at serine 10. Previous work by Peters and coworkers reported *in vivo* data that showed that inhibition of S10 phosphorylation prevented dissociation of HP1 from nucleosomes during mitosis.[84-85] They also used a radioactive labelling experiment to demonstrate a direct interaction of recombinant Cbx5 with H3-derived peptide that was inhibited with the S10 phosphorylated form. Our work provides direct evidence of the phospho-dependent binding of all three HP1 human homologues to H3-derived peptides. We also observed a reduction in binding affinity when H3S28 is phosphorylated. Evidence of this relationship *in vivo* comes from experiments showing genes normally repressed by the Pc complex are expressed when H3S28 is phosphorylated; suggesting a phosphorylation dependent removal of the Pc complex.[86] More recently, biochemical experiments have suggested that this regulation by nearby phosphorylation may be a widespread phenomenon for methyl-lysine reader domains, with *in situ* experiments showing that the affinity of several non-chromodomain methyl-lysine readers for their corresponding binding site on the H3 tail are also reduced in the presence of nearby phosphorylation.[87] Many of the biological studies are limited by the need to develop PTM-specific antibodies and often do not resolve the binding interactions of individual homologues. The PI-SAMDI assay provides a method to directly profile the binding of recombinant homologues to a large number of prospective ligands, both naturally occurring and non-native.

We describe the PI-SAMDI assay as appropriate for profiling or screening large numbers of protein-ligand combinations and from the results presented here, capable of rank-ordering affinities of ligands that span three orders of magnitude. We emphasize, however, that the assay

does not provide equilibrium binding constants for the interactions. Translating the enhancement factors to equilibrium constants is challenging, in part because analytical descriptions of enzyme kinetics at interfaces are challenging[88-89] and in part because of non-linearities that are intrinsic to the reaction scheme. For example, when the fusion proteins are bound to the monolayer at low density, each of the tethered KDAC8 domains can sweep a defined area of the monolayer and convert the substrates to products. But as the density of the bound fusion protein increases, there will be an overlap of the regions that neighboring enzymes can access, leading to a decrease in the enhancement factor. This complexity makes it difficult to develop an analytical expression for the binding affinity. In part for these reasons, the PI-SAMDI assay is well suited to profiling large numbers of prospective interactions, identifying interactions of interest, and providing a rank-order of affinity for these interactions. These potential applications will be explored further in the chapters to follow.

We believe this assay makes an important contribution to methods that profile protein interaction domains. Our strategy couples binding to an interfacial enzyme reaction that gives a 'covalent record' of transient binding interactions, allowing it to be applied to rapidly assess low affinity interactions that would otherwise be difficult to see by affinity-based methods.

.

3.4 Methods

*Preparation of Monolayer Arrays***.** Array plates with 384 gold spots on steel plates were soaked in a solution of disulfide molecules for 24 hours to allow formation of a self-assembled monolayer on the gold surface. The solution consisted of a mixture of EG3-alkane disulfide and a mixed disulfide of EG3-alkanethiol and a maleimide-terminated EG3-alkanethiol.[90] The solution of

disulfides contains an overall concentration of 1 mM of the two monolayer compounds in an appropriate stoichiometric ratio to yield a 10% maleimide surface density.

Substrate and Ligand Synthesis. The peptide substrate for KDAC8 and the peptide ligands for the chromodomains were synthesized using standard FMOC solid phase peptide synthesis on rink-amide resin. All peptides used were synthesized with a C-terminal cysteine for immobilization to the maleimide-terminated monolayer. For the experiments presented here, the substrate peptide has the sequence GMK(Ac)FGC. Additionally, the N-termini of all peptides were acetylated to improve stability and ionization efficiency. The peptides used as ligands are methylated versions of the sequences surrounding the H3K9 and H3K27 positions. These ligand peptides were synthesized by standard solid phase synthesis and are 10 residues long, containing the lysine of interest centered in the peptide. All peptides were purified by reverse phase HPLC after synthesis.

Surface Preparation. Peptide immobilization reactions were carried out by pre-mixing the two peptides in 1:4 stoichiometric ratio of methylated ligand to KDAC8 substrate and a total peptide concentration of 100 μM. For each spot, 3 μL of peptide was incubated on the surface presenting 10% maleimide in Tris buffer at pH 7.5 for one hour at 37ºC. With similar immobilization rates to the free maleimide, this yields a surface of 2% methylated ligand and 8% substrate peptide for the deacetylase.

Protein Design. Artificial fusions of KDAC8 to chromodomains Cbx1, Cbx3, Cbx4, Cb5, and Cbx6 were constructed in the commercial PET-303b bacterial vector. Each construct was made

by first using PCR to amplify the KDAC8 gene while introducing a poly-serine linker and restriction sites for insertion of the chromodomains on the N-terminal side. This was then inserted into the commercial PET-303b vector by restriction site digest and ligation using standard T7-ligase protocol. PCR was then used to amplify each chromodomain while introducing the appropriate restriction sites to allow for ligation into the KDAC8 containing plasmid. See appendix for additional information on sequence of fusion constructs.

*Protein Expression***.** For expression, each vector was transformed into the BL21 (DE3) strain of *E.coli* by electroporation. Cultures in 2XTY media containing carbenicillin (1 L) were grown to an optical density of 0.6. Expression of the fusion proteins was then induced with IPTG (1 mM) and incubated for 12-16 hours at 18ºC. The cultures were then pelleted by centrifugation, lysed by sonication in buffer A (50 mM Tris, 200 mM NaCl, 5% glycerol, 5 mM BME at pH 7.5), and the over-expressed fusion proteins, which contain a C-terminal HIS-tag were purified on a cobalt resin column. The proteins were then further purified by FPLC using a size exclusion column and then stored at -80ºC until use in a buffer containing 80 mM Tris, 150 mM NaCl, 2.7 mM KCl, 1 mM $MgCl_2$ and 10% glycerol at pH 7.5.

*Assays***.** When running assays, the fusion constructs were diluted to a concentration of 0.5 µM, as determined via Nanodrop, in running buffer (80mM Tris, 150 mM NaCl, 2.7 mM KCl, and 1 mM $MgCl_2$ at pH 7.5). The enzyme was plated onto array plates presenting substrate and ligand peptides at 3 µL per spot and the reactions were incubated at 37ºC for 10 to 15 min. At the end of a reaction, the surfaces were rinsed with water and ethanol, then dried. Matrix of THAP in acetone

was applied directly to the surface and after drying, each spot was analyzed by MALDI-TOF mass spectrometry using an AB Sciex 4800 or 5800 series instrument. The extent of enzymatic conversion of the substrate was then determined from the mass spectra for each reaction.

Chapter 4

Small Molecule Inhibitor Screening of Protein Interactions with PI-SAMDI

4.1 Introduction

In recent years, enzymes involved in regulation of the histone code, and epigenetic modifications more generally, have become desirable targets for pharmaceutical development. Their diverse and critical roles in signal transduction and gene regulation mean that they also serve important roles in many disease states, and a number of small molecule inhibitors have been developed targeting these families of enzymes. For example, the drug Vorinostat is a histone deacetylase inhibitor developed by Merck for the treatment of T cell lymphoma.[91] Kinase inhibitors have also become an increasingly important area of research for a number of diseases, including cancer, and there are over 25 FDA approved oncology drugs targeting kinases.[92-94] The families of proteins involved in regulation of PTM's are relatively straightforward targets for small molecule inhibitors and activators. The importance of these families of enzymes in epigenetics and cellular regulatory pathways make them attractive and useful druggable targets.

Adaptor domains and other PTM recognition domains have the potential to offer additional pharmaceutical targets, as they are involved in many of the same cellular processes as their writer and eraser counterparts. However, far less work has been done to develop small molecules targeting these proteins, when compared to the enzymes that add and remove the PTM's. The primary reason for this is a lack of tools for screening small molecule libraries for inhibition of protein interactions and the comparative abundance of assays for screening enzyme activity. It is only in the last decade that any such inhibitors have been reported. One of the more prominent

examples was the discovery of a small molecule inhibitors of the BRD4 bromodomain, an adaptor domain that binds acetyl-lysine motifs.[95-96] The BET family of bromodomains, to which BRD4 belongs, have been implicated in certain cancer processes and BET inhibitors are in development as treatments for multiple myeloma and other cancers. However, early bromodomain inhibitors were identified indirectly through phenotypic screens, rather than direct screening of small molecule libraries against the bromodomain specifically.

When it comes to methylation, a number of small molecule inhibitors have been identified for methyltransferases, the family of enzymes responsible for generating the methyl-lysine PTM.[97] Additionally, some work has been done to identify small molecule antagonists for methyl-lysine binding proteins. Using a medicinal chemistry and rational design approach, Arrowsmith and coworkers reported the design of antagonists targeting the L3MBTL3 protein, which binds to dimethyl-lysine. Instead of screening a large library of small molecules for activity, they designed a much smaller set of potential molecules which use a pyrollidine heterocycle as a mimic of the dimethyl-lysine binding partner. Using this strategy, they identified molecules that bind L3MBTL3 specifically, with dissociation constants of 100-200 nM.

While adaptor domains have become increasingly attractive drug targets, the lack of general and high throughput assay platform remains a major limiting factor. PI-SAMDI offers a straightforward way to directly screen for inhibition of adaptor domain binding, with throughput capable of scaling to tens of thousands of reactions. Here, we demonstrate this by screening a library of over 10,000 compounds for inhibitors of the Cbx1 chromodomain.

4.2 Results

As described in the previous chapter, PI-SAMDI functions by coupling the activity of a reporter enzyme on a surface bound substrate to binding affinity of a protein interaction domain for a ligand presented on the same surface. When the adaptor domain binding the surface-bound ligand, it recruits the enzyme to the surface and increases the effective concentration of the enzyme reporter experienced by the substrate. In this way, the observed rate of substrate conversion increases as the affinity of the adaptor domain for its ligand increases.

We demonstrated how this strategy could be used to rapidly profile the affinity of adaptor domains for libraries of biologically relevant ligands. We recognized that this same assay strategy could also be used as a tool for screening for small molecule inhibitors of a protein-ligand interaction and sought perform such a screen against the human Cbx1 chromodomain using the Cbx1-KDAC8 fusion, whose construction is detailed in the previous chapter. In the presence of a small molecule inhibitor of the interaction, the Cbx1 domain would be unable to bind to the ligand presented on the surface and we would expect to observe a decrease in the observed activity of the reporter enzyme (**Figure 4.1**).

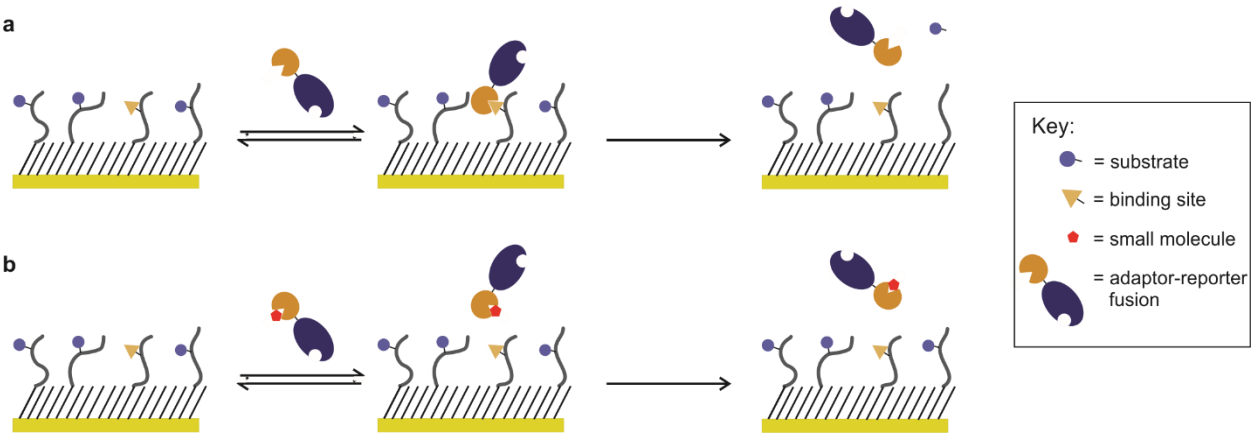


Figure 4.1. Small Molecule Screening with PI-SAMDI. a) In the absence of an inhibitor and in the presence of a ligand for binding, high levels of reporter activity are observed on the surface-bound substrate. **b)** An inhibitor of the adaptor domain-ligand interaction would prevent surface localization and be perceived by a decrease in the observed activity of the reporter on the surface.

We carried out the screening by preparing 384 spot SAMDI arrays with a self-assembled monolayers presenting maleimide-terminated alkanethiolates at a density of 10% against a background of tri(ethylene glycol)-terminated alkanethiolates. These surfaces were the same as those used for profiling chromodomain ligands. As before, two peptides were patterned on the surfaces, the KDAC8 reporter substrate of sequence GMK(Ac)FGC and the ligand for the Cbx1 chromodomain of sequence QTARK(Me_3)STGC, at densities of 8% and 2% respectively.

For the purposes of screening, these two peptides were held constant across the arrays. On each spot of the array, we incubated the Cbx1-KDAC8 fusion at a constant concentration, along with small molecules as potential inhibitors. The ligand peptide with the highest affinity was chosen and we previously observed a 7 to 8-fold increase in the activity of the reporter enzyme with this combination of ligand and adaptor domain (**Figure 3.5**). We sought to identify small molecule inhibitors by measuring the Cbx1-ligand peptide mediated activity of the reporter

enzyme (**Figure 4.2)**. A decrease in activity of the reporter enzyme signaled the potential of a small molecule inhibiting the Cbx1 interaction with the methylated ligand peptide.

For this pilot screen, we used four 384 spot arrays to screen 10, 240 small molecules from a purchased, commercially available, library (**Figure 4.2a)**. Here, each heatmap represents one of the array plates from the screen. Positive and negative controls were included on each array along the two outermost columns on each side of the arrays. The positive control contained no potential small molecule inhibitors, and so these spots should reach the maximal observed activity for the reporter enzyme. An ideal negative control would be an existing, high-affinity inhibitor. However, none have been reported for the Cbx1 chromodomain, so instead we used surfaces lacking the methylated ligand peptide. Without this ligand peptide, Cbx1 mediated rate enhancement is not possible, and so the observed activity should be equivalent to that of an ideal inhibitor.

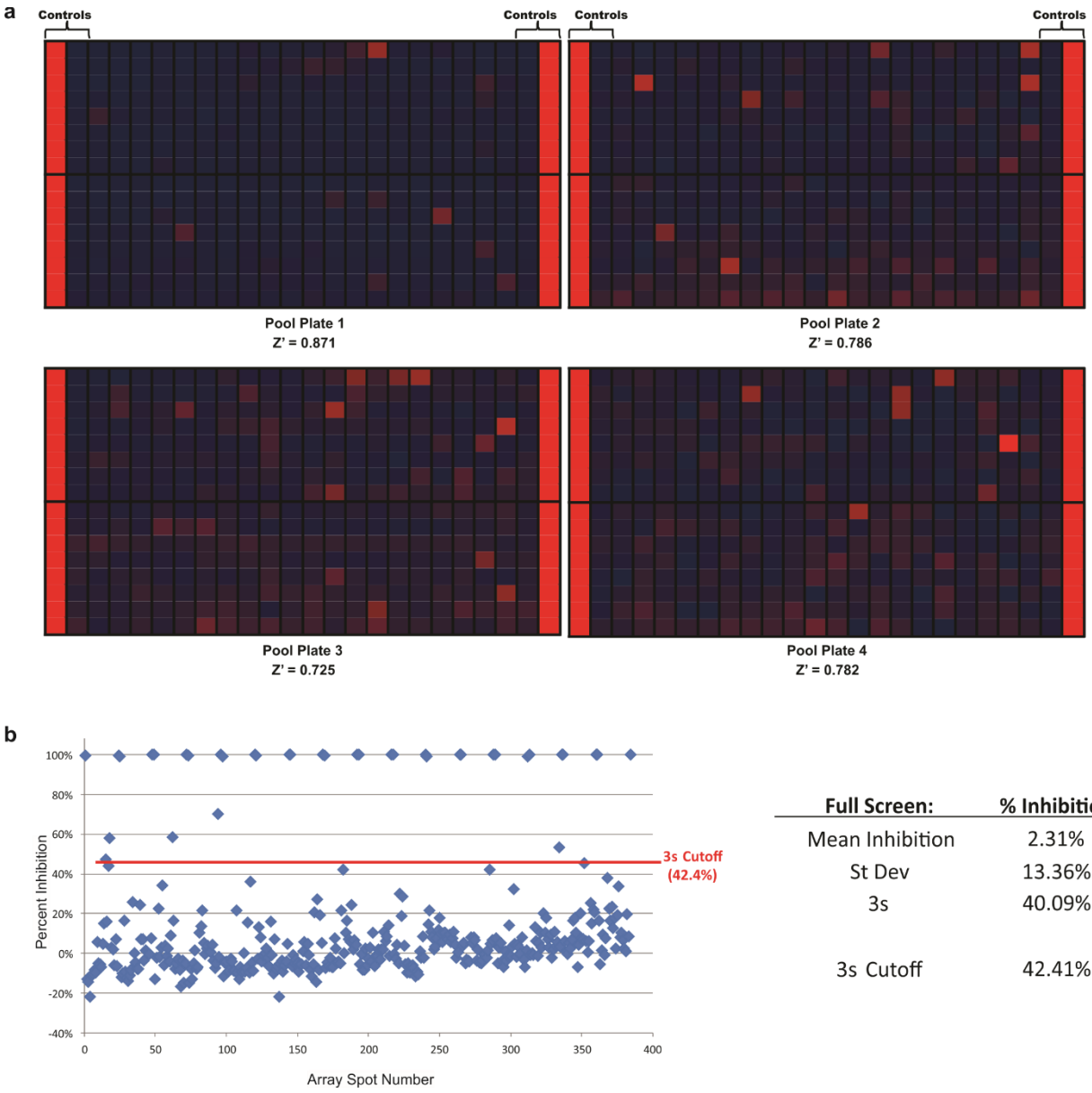


Figure 4.2. Results of Primary Cbx1 Screen. a) Over 10,000 small molecules were screened against Cbx1 across four 384 spot arrays in pools of 8 small molecules per reaction spot. The heatmaps here represent the observed activity for each plate, where red signifies increasing inhibition of the observed activity of reporter enzyme. On each plate, the outermost columns were reserved for positive and negative control reactions. **b)** The plot here shows the resulting inhibition data for plate 3 of the small molecule screen. Here, the resulting activity is plotted relative to the positive and negative controls. A statistical cutoff of 3 times the standard deviation was used and spots above this cutoff were identified as potential "hits".

We screened the potential inhibitor in pools of 8 compounds per spot, such that each reaction contains 8 unique small molecules incubated with the chromodomain fusion on the surface. This greatly reduces the number of reactions that need to be run, and then spots with potential hits can be deconvoluted with a much smaller secondary screen. For each array in the screen we determined the Z-factor as a measure of the efficacy of the assay using equation 3 below. Here, σ_p and σ_n represent the standard deviation in the positive and negative controls, respectively, while μ_p and μ_n the correspond to the mean of the of these controls.

$$Z\ factor = 1 - \frac{3(\sigma_p + \sigma_n)}{|\mu_p - \mu_n|} \qquad (3)$$

The Z-factor is used as a measure of the quality of a screening method, with 1 being ideal and the values above 0.7 seen here considered excellent and well suited for large-scale deployment. The resulting activity on each spot was plotted relative to the positive and negative controls as seen for array plate 3 in **Figure 4.2b**. The mean inhibition across the screen was determined. The standard deviation across the array was determined and 3 times the standard deviation was used as a statistical cutoff for determining a "hit" from a "non-hit".

However, this assay requires one additional step to properly identify hits from non-hits. Any small molecules which act by inhibition of the KDAC8 reporter enzyme will appear as hits in the primary screen as well. We checked each lead molecule by generating IC50 inhibition plots against wild type KDAC8 and the Cbx1-KDAC8 fusion. Activity on the fusion was measured on the surface presenting the reporter substrate and methylated ligand, where as the wild type KDAC8 activity was screened for activity on the peptide GRK(Ac)FGCA, which we have previously identified as a high-activity substrate from KDAC8 (**Figure 4.3)**.[67]

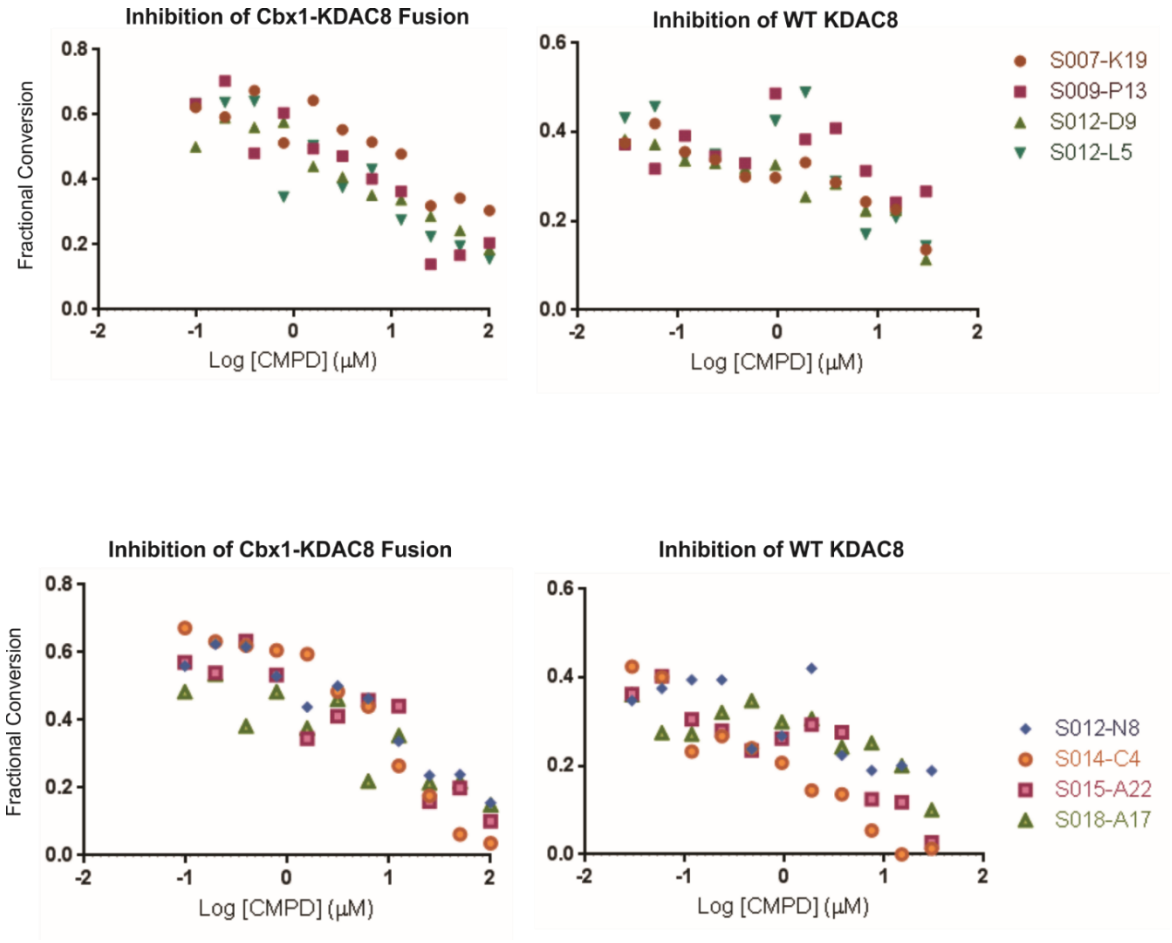


Figure 4.3. Characterization of Potential Hits. Potential hits identified by the primary screen were tested against both wild type KDAC8 and the Cbx1-KDAC8 fusion used here in order to identify which are acting on Cbx1 and which are modulating observed activity by acting on KDAC8, the reporter enzyme. All molecules identified in the screen were found to be inhibiting KDAC8 rather than Cbx1 binding.

True inhibitor of Cbx1 should show inhibition for activity of the Cba1-KDAC8 fusion, but no inhibition of KDAC8 alone. Examples of the tandem IC50 plots can be seen for 8 of the hit molecules in **Figure 4.3**. In this secondary characterization we found that the molecules from the primary screen are inhibiting KDAC8 rather than Cbx1, as demonstrated here by the matching

inhibition curves observed for both. Unfortunately, none of the 10,240 tested in this pilot screen were identified as chromodomain binding inhibitors.

4.3 Conclusions

The enzymes that regulate post translational modifications have become important targets for therapeutic applications, as their roles in regulating gene expression and other cellular functions make them important players in a number of disease states. Adaptor domains serve important roles, often in the same regulatory networks as the enzymes that produce the modifications that they recognize and bind. As such, adaptor domains have emerged as potential druggable targets, with examples such as inhibitors of the BRD bromodomain showing promise for treatment of several cancers.

However, available tools for screening inhibitors of adaptor domains are limited. While some high-throughput techniques exist for looking at adaptor domain-ligand interactions, such as the SPOT binding arrays described earlier, these assays are either incompatible or not well suited for high-throughput screening. The early inhibitors of bromodomains, which represent the first reported inhibitors of adaptor domains, were discovered primarily through phenotypic screening of cells and was later shown to function through inhibition of the bromodomain binding. Since then, other published work has taken a rational design strategy, using structural mimics of these protein's target binding sites as a starting point for designing a smaller and more target set of small molecules that can be screened with lower-throughput techniques.

In this work, we have shown the implementation of our SAMDI based assay to the screening of the Cbx1 chromodomain against a library of over 10,000 small molecules. While we did not identify any novel inhibitors from this study, this pilot screen demonstrates the straightforward application of our assay strategy to small molecule screens.

4.4 Methods

Preparation of Surfaces. The 384-spot surfaces were prepared as described in Chapter 3. Briefly, 5 nM of titanium and 30 nM of gold were deposited onto steel plates by electron beam evaporation, using masks to generate a 384-spot pattern in a standard microtiter format. These array plates were soaked in an ethanolic solution of EG3-alkane disulfide and a mixed disulfide of EG3-alkanethiol and a maleimide-terminated EG3-alkanethiol, with a 1 mM combined disulfide concentration for 48 hours. In order to immobilize the ligand peptide, of sequence QTARK(Me_3)STGC, and reporter enzyme substrate peptide, of sequence GMK(Ac)FGC, the two peptides were premixed at a 1:4 stoichiometric ratio and total concentration of 100 µM in Tris buffer at pH 7.2. 3 µL of peptide solution was plated onto each spot and the surface was incubated for 1 hr at 37°C. These peptides were synthesized by standard FMOC solid phase synthesis on rink amide resin.

Protein Design and Expression. The fusion protein of the Cbx1 chromodomain to the KDAC8 deacetylase was generated as previously described in Chapter 3. Briefly, the fusion was constructed in the pET-303b vector and transformed into BL21 (DE3) strain of *E.coli* for expression. For expression, 1 L cultures were grown in 2XTY media at 37°C until an optical

density of 0.5-0.7 was achieved, and then expression was induced by addition of 1 mM of IPTG and incubation for 12-16 hrs at 18°C. The overexpressed protein was isolated by lysis and purification by cobalt column affinity chromatography using the contained HIS-tag. The isolated protein was further purified by FPLC using a size exclusion column before use in the screen.

Selection of Reaction Conditions. Several parameters were tested to select the optimal reaction conditions for high-throughput screening. A HEPES based buffer (80mM HEPES, 150 mM NaCl, 2.7 mM KCl, and 1 mM $MgCl_2$ at pH 7.5) was selected for the reactions as the reporter enzyme was found to have longer-lived activity and while maintaining similar rate of activity in this buffer when compared to the Tris buffer system used in the previous chapter. We sought to minimize enzyme concentration to reduce background activity of the reporter, while maintaining enzymatic substrate conversion greater than 50% in the positive control. Optimizing these parameters improves the Z-factor and the statistical power of the screen. Activity of the Cbx1-KDAC8 fusion was monitored over time at varying concentrations. This was used to select 500 µM as an optimal concentration of enzyme for screening, which maximizes both signal and Z-factor.

Preforming High-throughput Screen. The library of small molecules used here was purchased commercially and stored at -80°C in 384 well plates in DMSO. The small molecules were pre-incubated with the enzyme before delivery to the reaction surfaces. The Cbx1-KDAC8 was prepared in 384 well plates in the HEPES running buffer at a concentration of 600 nM. Liquid handling robotics were used to transfer compounds into the well plate containing the enzyme, with

8 compounds delivered into each well at a final concentration of 10 µM per compound. 3 µL of the mixture containing compounds of interest and the Cbx1-KDAC8 protein was then transferred to the array surfaces using a TECAN liquid handling robot equipped with a 384-tipped head to deliver all reactions simultaneously. The reactions were incubated at 37°C for 20 minutes. The plates were rinsed with distilled water, dried under a stream of nitrogen, and matrix of 20 mg/mL of THAP in acetone was applied to the surface. Each spot was analyzed by MALDI-TOF mass spectrometry using an AB Sciex 5800 series instrument and the extent of enzymatic conversion of the substrate was determined from the resulting mass spectra for each reaction.

Chapter 5

Towards Mapping of Adaptor Domain Ligand Preferences with PI-SAMDI

5.1 Introduction

The preceding chapters have outlined the design of an assay for screening protein-ligand interactions. We applied this system to the screening of human chromodomains, methyl-lysine readers, against small libraries of biologically relevant ligand peptides. Then, we demonstrated how the scalability of the SAMDI platform could be used to screen small molecule inhibitors of these interaction by running over a thousand reactions to screen >10,000 small molecules, which involved screening the same interaction repeatedly in the presences of each of these small molecules and measuring changes in observed activity of a reporter. However, another possible application of this assay is using this throughput to screen large libraries of ligand sequences for adaptor domains where little is known about their binding preferences.

Here, we will look at SH2 domains (Src Homology 2 domains), which are a family of adaptor domains that bind to phosphorylation sites on tyrosine residues. Over 110 proteins in humans have been found to contain SH2 domains, including kinases, phosphatases, and transcription factors.[98] They play important roles in signaling cascades, where phosphorylation is a key feature of cellular signaling and readers of phosphorylation play critical roles in localization of proteins and transduction of these intracellular signals.[99] The SH2 domain examined in this work comes from the phosphatase SHP1, which has been shown to act as a negative regulator of cellular signaling pathways.[100-101] I will describe work towards developing PI-SAMDI as a method for screening large ligand peptide libraries using this SH2 domain as a model system.

5.2 Results

Optimizing PI-SAMDI for High-throughput Ligand Screening

In this chapter I discuss our work towards screening ligand preferences for an SH2 domain, an adaptor domain which binds phosphorylated tyrosine residues, from the human SHP1 protein. In our initial work with chromodomains, we chose the KDAC8 deacetylase as the reporter enzyme for the assay. This enzyme possesses a number of benefits as a reporter. It has robust activity, well characterized substrate preferences, and has been shown to maintain activity in artificial fusions. However, its use of peptide substrates becomes somewhat limiting when scaling to ligand screening. We have shown that co-patterning of both substrate and ligand peptides at specific densities could be done by controlling the relative molar ratio of the two peptides in the solution applied to the maleimide-presenting surface (**Figure 3.4**). While this strategy works well for smaller libraries of ligands, it requires each peptide be both purified and precisely weighed so that an accurate concentration can be obtained. This process becomes prohibitive when moving forward to libraries of hundreds, or thousands, of potential ligand peptides.

The solution to this problem was to look for a new reporter enzyme whose substrate could be easily pre-patterned across the entire arrays, effectively predefining a constant substrate density. Then, a pre-determined density of maleimide on the same surface could be used to immobilize ligand peptides of interest from large-scale library syntheses where the exact concentration of each peptide may not be known or easily determined.

We used cutinase (Cut) as the reporter enzyme. Cutinase is an ester hydrolase with robust and well characterized activity. We designed and synthesized a small molecule substrate for this reporter based on a phenolic ester, hindered by an α-methyl group to lower background activity of

the reporter in the absence of a ligand-mediated surface interaction (**Figure 5.1a**). This small molecule substrate was used to prefunctionalize maleimide-EG_3 disulfide compound in solution, before it was used to form a monolayer (**Figure 5.1b**). Then a monolayer presenting cutinase substrate at 2% density and maleimide at 8% density against a background of 90% tri-(ethylene) glycol was formed by soaking gold surfaces in ethanolic solution of these three disulfide species at the appropriate stoichiometric ratios. This results in a surface that presents both cutinase substrate and free maleimide at fixed densities (**Figure 5.1c**). The aromatic ester substrate possesses a mass of 1072 when bound to the monolayer and cleavage of the ester by the cutinase reporter leads to a shift of -84 mass units.

The SHP1 protein contains two tandem, homologous SH2 domains on the N-terminus of the proteins. For the purposes of this experiment, a fusion of a single SH2 domain (SH2-1, the one closest to the N-terminus) was prepared to cutinase. This construct was assembled in the commercially available pET-22b vector in using a similar strategy of PCR, digestion, and ligation as was used to generate the chromodomain-KDAC8 fusions described in chapter 3. However, this fusion was expressed periplasmically, instead of in the cytoplasm, as cutinase contains several disulfide bonds which require an oxidative environment for formation. The expressed protein was purified by cobalt column affinity chromatography and FPLC on a size exclusion column before use.

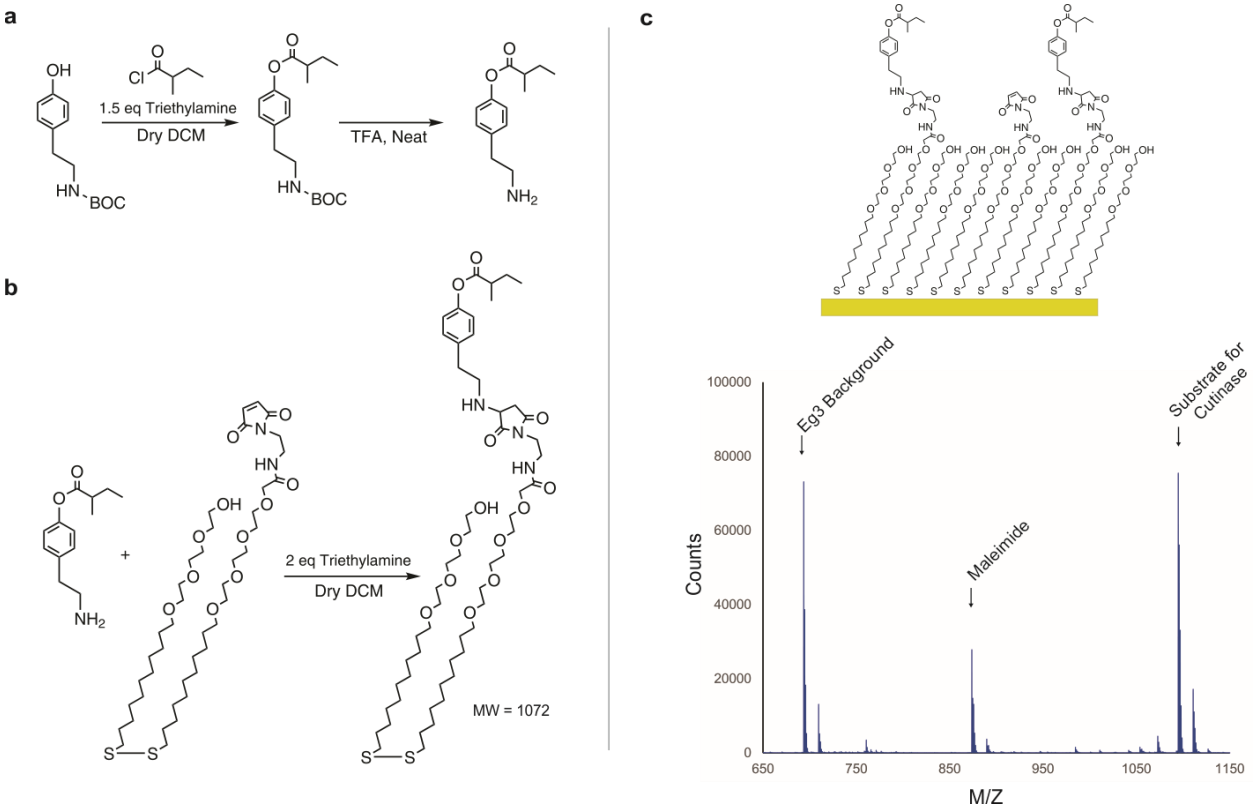


Figure 5.1. Surfaces for Cutinase-based Reporter. a) An aromatic hindered ester substrate for the cutinase reporter was synthesized. **b)** In solution, this substrate was coupled to monolayer compound. **c)** A mixed monolayer was formed the presented the reporter substrate and maleimide for immobilization of ligand peptides of interest, both at pre-defined densities. SAMDI spectra shows characterization of the mixed monolayer surface.

Profiling SH2 Ligand Sequence Preference

The fusion protein was screened against a library of possible ligand peptides containing a phosphorylated tyrosine with two variable position, one on either side of the phosphorylation site, which were varied with every natural amino acid except cysteine. This library possessed the general sequence structure of Ac-TRD**X**Y(PO_4)**Z**TC, where **X** and **Z** represent the two variable positions, for a total of 361 peptides. This library was meant to look at the effect of these immediate flanking residues on relative binding preference of the SH2 domain. The library was

printed in quadruplicate onto a 1536-spot SAMDI array presenting maleimide for immobilization and cutinase substrate. After immobilization of the library, Cut-SH2 fusion was spotted across the entire array at 50 nM in phosphate buffered saline (PBS) at pH 7.2, and incubated for 20 minutes at 37° C. The array was rinsed and read by MALDI-TOF mass spectrometry. The resulting spectra were used to calculate the fold enhancement (yield relative to spots lacking phosphorylated ligand peptides) as described in chapter 3. This data was combined into a heatmap showing relative affinity across the entire array of peptide ligands (**Figure 5.2a**). As a measure of quality of the screen we can look at the standard deviation across the replicate measurement for each ligand peptide (**Figure 5.2b**). We find that high statistic error is focused on regions of very low reporter yield, where very small changes in measured yield can have a dramatic effect on the standard deviation, and the regions of high reporter activity are highly repeatable.

While the heatmap shows no overwhelming selectivity for these two positions immediately surrounding the phosphorylation site appear. However, we can look at the sum of activity in each column or row (**Figure 5.2a**) to get a better sense of favored positions. There appears to be greater sensitivity to the Z (+1) position with hydrophobic (A, I, L) and aromatic (H, F, W) residues performing well. In the X (-1) position there is less variance overall but I, L, and F perform somewhat better than the others. This work shows that this surface functionalization strategy and reporter enzyme make it possible to generate arrays of peptide ligands for mapping of interaction domains by PI-SAMDI. However, the with this increase in throughput the primary limiting factor becomes the ability to synthesize large libraries of peptides to map a greater number of residues that may be playing a role in binding. One solution to this problem, which will be discussed below, is development of SAMDI-compatible split-pool methodologies for peptide libraries.

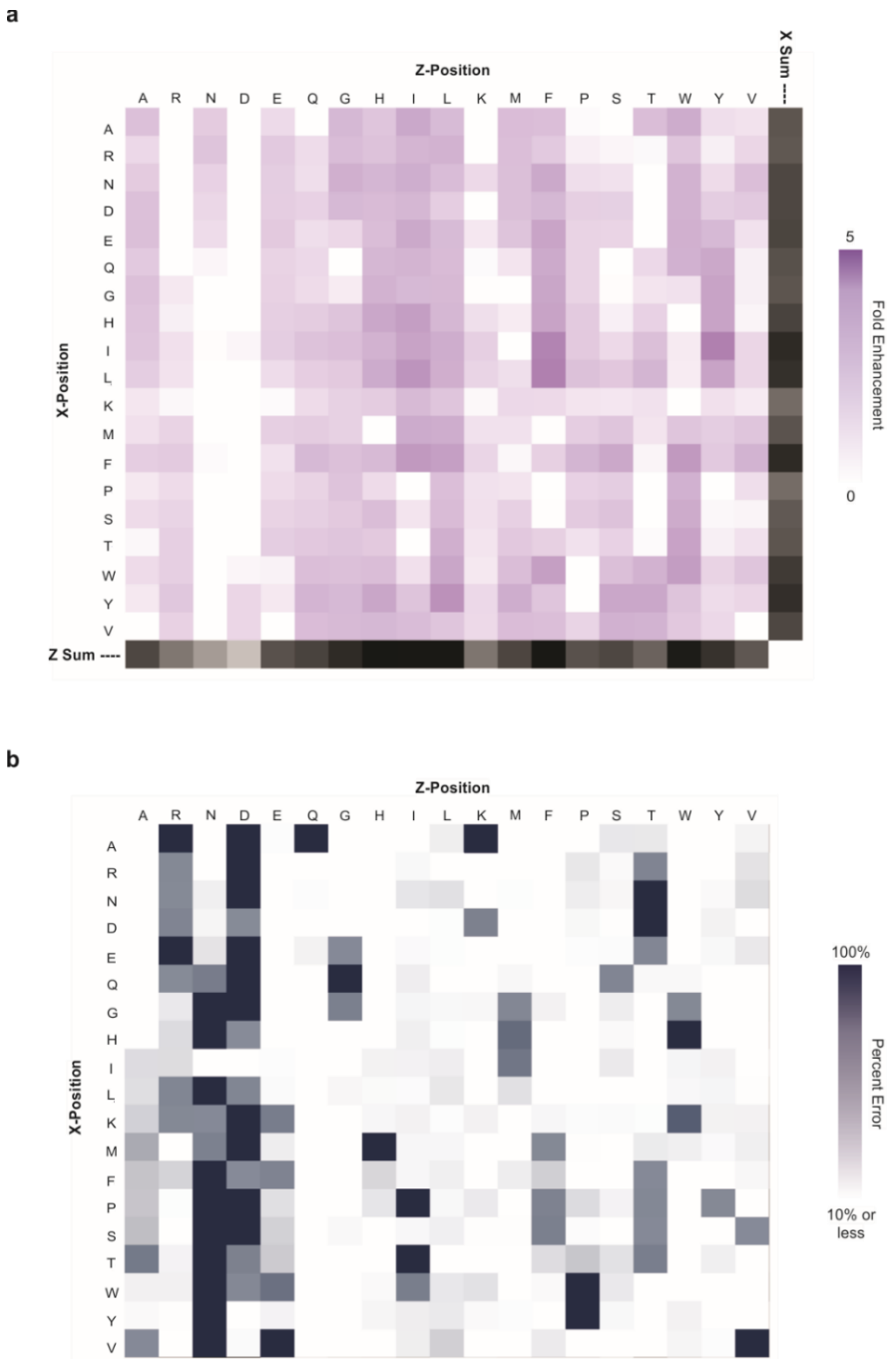


Figure 5.2. Mapping SH2 Binding Preference. a) Using cutinase as a reporter, the affinity of the SH2 domain from SHP1 for a library of phospho-tyrosine peptides of the general form Ac-TRD**X**Y(PO_4)**Z**TC was mapped. **b**) We can use the percent error across 4 replicate reactions as an estimate of quality. Here we see high error corresponds well to regions of low reporter yield, where small variations in measured conversion make a larger relative difference, and spots of high yield also show highly reproducible levels of activity.

5.3 Conclusion

Here, we have shown how the concept of PI-SAMDI can be adapted to the screening of ligands for adaptor domains where little is known about the preferred ligands or where it is desirable to examine large panels of potential ligands. The surface enhancement strategy we developed for PI-SAMDI is compatible with a wide range of potential reporters, and here we demonstrate how using certain selection criteria can expand the kinds of possible experiments. Particularly, the ability to generate a monolayer that is pre-functionalized with the reporter enzyme substrate greatly simplifies the process of functionalizing the surface with ligands of interest. Another possible strategy to achieve a similar outcome would be to use two orthogonal immobilization chemistries to separately immobilize reporter substrate and ligands.

Here, we have demonstrated the efficacy of this strategy by screening a phosphotyrosine binding SH2 domain from the SHP1 protein against a panel of ligand peptides which represent all possible permutations in the -1 and +1 positions surrounding the phosphorylation site. However, previous work in the literature has shown that SH2 domains are sensitive to the identity of residues further away,[102] and so a full mapping of the domain selectivity will require future work on larger libraries with more variable positions. While the PI-SAMDI strategy presented in this work is readily scalable to ligand libraries in the tens of thousands, the major limiting factor at the time this work was done was obtaining peptide libraries of those sizes in a cost-effect manner. Previous examples from the literature which look at these kinds of large ligand libraries with 3 or more variable positions use two primary methods to avoid the labor and costs of synthesizing such a library. The first is to utilize phage-display techniques to encode the expression of ligands and then a selection process so that only "hits" are identified and decoded.[103] The second is to use split

pool synthesis to synthesize much larger numbers of peptides in a systematic, but randomized way.[102, 104] In this method large tens of thousands of peptides are generated at lower cost and with far less labor, but their sequences are unknown until decoding. Split pool methods used in the literature also use a selection process to try and limit the number of sequences which must be laboriously decoded to only those identified as high-affinity ligands. And even then, many sequences are discarded, as much as 30% of the "hits", if they cannot be unambiguously decoded.[104]

Recent work in our group has led to the development of a split pool synthesis methodology which incorporates a mass ladder that enables facile synthesis and rapid decoding of the large peptide libraries (10,000 peptides and greater) by SAMDI, which then can be directly deployed for experiments. This could provide easy access to the larger synthetic throughput needed to profile many more variable positions. In the future, this technology paired with PI-SAMDI could enable not only the identification of high affinity ligands, as do other technologies used previously, but also the rank-ordering of all possible ligands by relative affinity.

5.4 Methods

Synthesis of 1-Methyl-Butyrl Tyramine Ester. All chemicals were purchased from Sigma Aldrich unless otherwise specified. 1 gram of N-BOC-Tyramine was dissolved in 100 mL of dry dichloromethane (DCM) solvent in an oven-dried 250 mL round bottom flask, to which 1.5 equivalents of triethylamine was added, and the solution was chilled on ice. 1.1 molar equivalent of 1-methyl-butyrl chloride, liquid, was added slowly by syringe over one minute while the

reaction was stirred. The reaction was allowed to warm to room temperature and proceeded until no starting material was detected by TLC, roughly 1 hour. The mixture was extracted over brine and water, the DCM layer was dried with $MgCl_2$, and the DCM was removed by rotovap. The BOC protecting group was removed by deprotection in neat trifluoracetic acid (TFA) for 5 minutes. The TFA was removed under vacuum and the crude product was purified on a silica gel column using a buffer system of ethyl-acetate and hexanes at a ratio of 2:3.

Solution Phase Coupling to Maleimide-EG3 Disulfide. The purified tyramine ester-substrate was coupled to a maleimide terminated SAM molecule by dissolving both dry DCM in the presence of 2 equivalents of triethylamine in order to deprotonate the amine and make it reactive to the maleimide. The reaction solution was extracted into brine and water, and then isolated from the organic layer by rotovap and purified on a silica gel column in a solvent system of methanol (MeOH) and DCM at a ratio of 1:9. The purified product was isolated by rotovap and dried under high vacuum.

Construction of Cutinase-SH2 Fusion. Plasmids containing SHP1 and Cutinase were already present in the lab. Cutinase and the N-terminal SH2 domain of SHP1 were amplified by PCR from the plasmids, which was also used to introduce restriction sites for ligation. The two sequences were ligated into the pET-22b(+) vector, Novagen, using stepwise digestion of restriction sites and a standard ligation protocol with T7 ligase. The full sequence of this insert can be found in the appendix. Importantly, this vector contains a pelB leader sequence to direct periplasmic

expression of the protein. This is necessary for proper folding of the expressed cutinase, which contains disulfide bridges and requires an oxidative environment for these to form.

Expression of Cut-SH2 Fusion. For expression, the vector was transformed into the BL21 (DE3) strain of *E.coli* by electroporation. Cultures in 2XTY media containing carbenicillin (1 L) were grown to an optical density of 0.6 at 32° C. Expression of the fusion proteins was then induced with IPTG (1 mM) and incubated for 12-16 hours at 18ºC. The cultures were then pelleted by centrifugation, lysed by incubation with BugBuster lysis detergent for protein expression from EMDMillipore, and the over-expressed fusion protein, which contain a C-terminal HIS-tag, was purified by cobalt column affinity chromatography. The protein was further purified by FPLC using a size exclusion column and then stored at -80ºC until use in PBS and 10% glycerol by volume.

Synthesis of Phosphotyrosine Peptide Library. The library used here was synthesized by Jeffery Huang. The peptides were synthesized using standard FMOC solid phase peptide synthesis on lanterns presenting rink amide functional groups for coupling. Commercially available Fmoc-Tyr(PO(OBzl)OH)-OH from NovaBioChem was used along with standard coupling conditions to introduce the phosphorylated tyrosine. The libraries were stored in 384 well plates in water and 0.1% TFA at -80° C. For immobilization, a TECAN liquid handling robot equipped with a 384-tipped head was used to dilute this stock into Tris buffer (pH 7.2) at a peptide concentration of roughly 20 µM.

Performing Assays on Phosphotyrosine Library. Arrays of 1534 gold spots were soaked in an ethanolic solution of maleimide-EG3 disulfide, tyramine substrate functionalized maleimide-EG3 disulfide, and the symmetric EG3 disulfide at a total disulfide concentration of 1 mM the proper stoichiometric ratios to generate a surface of 2% free maleimide, 8% reporter substrate, and 90% EG3 terminated monolayer. The peptide library was printed in quadruplicate onto this array from a 384 well plate using a Tecan liquid handling robot at 0.5 µL per spot. The peptides were incubated for 1 hr at 37° C and the surface was rinsed with water and dried. The Cut-SH2 fusion protein was diluted to 50 nM in PBS and distributed across the array at 0.5 µL per spot using a Multidrop COMBI liquid handling robot. The SH2-reporter enzyme construct was incubated on the ligand array for 20 min. at 37° C. The surface was rinsed with water, dried and matrix of 20 mg/mL of THAP in acetonitrile was applied to the surface. The plate was read by MADLI-TOF MS using an AB Sciex 5800 series instrument.

Part 2:

SAMDI-MS as a Tool for Rapid Characterization of Biosynthetic Pathways

Chapter 6

Bioengineered Pathways and their Applications

In recent years, bioengineered enzymatic pathways have become an increasingly attractive avenue for production of useful small molecules with a wide range of applications. This has been made possible by the convergence of significant advances in several fields. Decades worth of biochemical work to characterize enzyme activities and the rise of modern databases that organize the enormous existing knowledge of enzymes and enzymatic pathways has made it increasingly easy to design biosynthetic routes. Meanwhile, the development of powerful, standardized tools in molecular biology have enabled engineering of these pathways into a variety of easy to grow organisms.

Recent bioengineering efforts have spanned a wide range of applications. In medicine, biosynthetic pathways are being used to synthesizes complex molecules for use as drugs. Prominent example of this include the semi-synthetic production of the antimalarial drug artemisinin[105] and the bioengineering of *E.coli* for the production of Taxol[106], a potent anticancer drug. Drug molecules may require dozens of synthetic steps to access chemically; taxol for instance requires greater than 35 individual reactions. Replacing even parts of these complex syntheses with biosynthetic routes can improve yields and reduce production costs. In energy, bioengineering offers the potential for sustainable production of small molecules fuel sources as alternatives to petrochemicals. Early efforts in this field involved biosynthesis of compounds such as ethanol, butanol, and plant-oil based biodiesels.[107] While these compounds can be used as fuels, they are not necessarily compatible with existing infrastructure, which is designed to operate on

hydrocarbons with specific boiling point ranges and viscosities. More recent efforts toward bioengineered fuel production have targeted small molecules in fatty acid and terpenoid biosynthetic pathways which have the potential to be direct replacements for fossil fuels, without needing to replace or modify our current infrastructure.[108-109] Beyond these important targets for bioengineering, a host of unique and less obvious applications exist, such as the recently reported engineering of a brewer's yeast strain which is capable of producing the hop flavor compounds linalool and geraniol.[110]

However, while the field continues to advance, engineering cells is still a time-consuming and often non-intuitive process that can require dozens of person years to optimize a pathway for production of a target molecule. Many approaches use an iterative design-build-test (DBT) cycle to test numerous "weak hypotheses" and explore a large parameter space in which the underlying biology is not completely understood, or where the complexity of the system does not yet permit rational design. The effort required to successfully engineer a commercially viable biosynthetic pathway remains one of the primary obstacles to their wider adoption, and there is room for improvement at every stage of the DBT cycle.

While new examples of successfully engineered biosyntheses continue to be published, the process remains slow. A number of recent advances have accelerated our ability to generate biosynthetic pathways. Advances in large scaled DNA sequencing and growing libraries of functionally annotated enzymes have made identifying biosynthetic routes to desirable targets easier than ever before. DNA cloning, access to affordable commercial DNA synthesis, and identification of important biological "building blocks" have simplified the process of constructing

these pathways. And the development of tools such as cell-free synthesis has enabled the creation of biosynthetic pathways that bypass the need for laborious cellular engineering altogether.

However, the bulk of advances have been focused on simplifying the process of designing and building these biosynthetic systems. Meanwhile, the analytical tools available for characterization of these complex systems (the "test" part of the DBT cycle) have remained largely the same, usually relying on the tedious development of chromatography and mass spectrometry methodologies for each analyte of interest. This analytical throughput has become an increasingly limiting factor as the design process for bioengineered systems is simplified and shortened. In the work to follow, I present the development of a SAMDI-based strategy for the rapid characterization of biosynthetic pathways, where specific classes of metabolites can be quantified in parallel, and demonstrate, through application of this technique, how high throughput analysis can enable rapid optimization of bioengineered pathways and provide a more complete understanding of the underlying reaction network.

Chapter 7

A SAMDI-MS Assay for Optimizing CoA Dependent Pathways

This chapter is adapted from the following published work:
O'Kane, P.; Dudley, Q.; McMillan, A.; Jewett, M.; Mrksich, M. *High-throughput Mapping of CoA Metabolites by SAMDI-MS to Optimizing the Cell-free Biosynthesis of HMG-CoA.* In Press at Science Advances.

7.1 Introduction

In this work, we describe an assay based on SAMDI-MS (self-assembled monolayers for MALDI-TOF mass spectrometry) capable of rapidly detecting CoA-bound metabolites. In SAMDI, biochemical assays are performed on self-assembled monolayers (SAMs) of alkanethiolates on gold, and the resulting immobilized reaction products can be directly quantified using MALDI-TOF mass spectrometry.[21] In previous work, we have shown that SAMDI can be used as a general assay platform to profile a wide range of enzyme activities *in situ* and in complex lysates.[29, 34, 69, 111] Here, we demonstrate how SAMDI can be used to characterize biosynthetic pathways by immobilizing a specific class of metabolites to the surface. The example presented here uses chemo-selective capture of all acyl intermediates on CoA, followed by quantitative analysis by SAMDI-MS. The use of chemically defined monolayers enables the rapid isolation of all acyl-CoA species from complex lysates while simultaneously serving as the platform for detection.

Biosynthetic pathways offer unique opportunities for efficiently and renewably producing a wide range of molecules for medical, energy, and industrial applications. Many of these processes are based on Coenzyme A (CoA) dependent metabolic pathways, as CoA is involved in over 100 distinct reactions and it is an obligate cofactor for 4% of known enzymes.[112] Acetyl-CoA

serves as the universal precursor for numerous biosynthetic pathways including isoprenoids, fatty acids/alkanes, polyketides.[113] In yeast alone, acetyl-CoA is involved in 34 different metabolic reactions.[114] A diverse range of bioengineering efforts have used CoA dependent pathways, including many of the synthetic targets outlined in the previous chapter: taxol, terpenoid biofuels, and even the hop flavor compounds. The importance of CoA metabolism has motivated the engineering of organisms such as *S. cerevisiae* as "chassis" strains for overproducing acetyl-CoA [115] or other short chain acyl-CoA biosynthetic precursors.[116]

A key barrier to accelerating the engineering DBT cycle is the slow measurement methods available for detecting the products and intermediates of a pathway. For CoA-bound metabolites, the current state-of-the-art detection methods are column-based separation, followed by mass spectrometry.[117-121] These methods are highly sensitive (pmol of analyte per sample)[122] and can be adapted to measure multiple acyl-CoA compounds in a single analysis,[123] but have low throughput, generally requiring more than 15 minutes per analysis. Colorimetric screens and intracellular metabolite-sensing circuits offer increased throughput, but are typically specific for a single CoA-bound molecule and require laborious redesign to quantitate new target molecules.[124]

A significant benefit of SAMDI is the compatibility with microtiter formats and laboratory automation to allow evaluation of tens of thousands of reactions per day.[29, 69, 82] Here, we use this robust and high-throughput method to characterize a cell-free reaction system engineered to produce hydroxymethylglutaryl-CoA (HMG-CoA), the biosynthetic precursor to mevalonate and isoprenoid metabolites. Cell-free engineering is a powerful approach for accelerating the design and construction of biosynthetic pathways and has the further benefit that it is compatible with assay automation.[125] In cell-free reactions, biosynthetic pathways are

reconstituted in crude lysates, rather than living cells, bypassing the need for laborious metabolic engineering. These cell-free systems are liberated from the regulatory processes that, *in vivo*, support cell viability and growth, allowing for the design of synthetic pathways that may be difficult or impossible in living cells. [125-127] We demonstrate the combination of cell-free reactions and high-throughput characterization with SAMDI to perform thousands of biosynthetic reactions in a single experiment. This strategy, applied here for the optimization of HMG-CoA biosynthesis, greatly increase the speed at which a large reaction space can be mapped for complex networks of enzymes.

7.2 Results

Design of an Assay for Acyl-CoA Metabolites

We identified the thioester functional group, through which the CoA cofactor binds metabolites, as a unique chemical handle for bio-specific capture. Thioesters have been used extensively to ligate peptides together via the native chemical ligation chemistry (NCL) pioneered by Kent and coworkers.[128] We recognized that this ligation strategy could also be used to capture acyl moieties from CoA with the specificity required to operate in complex cell lysates. A peptide with an N-terminal cysteine was used for covalent capture, immobilization and detection of CoA metabolites. When the capture peptide is added to reactions containing CoA bound species, it readily scavenges the acyl groups from CoA by a trans-thioesterification reaction followed by a rapid internal rearrangement to irreversibly transfer the captured analyte to the N-terminal amine of the peptide. The reaction mixture is then applied to a monolayer presenting a maleimide group, where the free thiol of the cysteine undergoes immobilization to the monolayer and the lysate can

be removed by rinsing (**Figure 7.1a)**. The resulting monolayers have tethered to them all of the intermediates and products from the CoA biosynthesis and can be directly analyzed by SAMDI mass spectrometry.

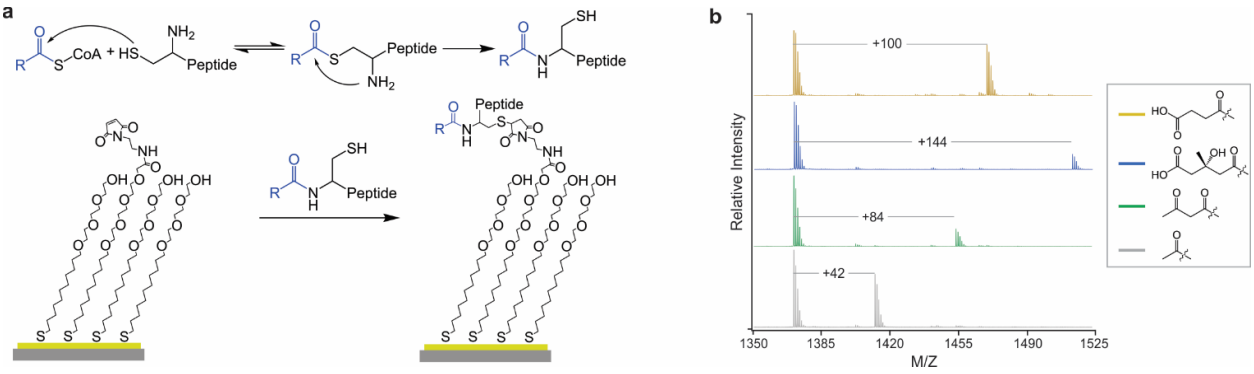


Figure 7.1. A generalized approach for capturing CoA-bound metabolites. a) A peptide with an N-terminal cysteine readily reacts with the thioester of RCO-CoA, creating a stable amide bond with the acyl group. After capture, the thiol of the peptide can then be used to immobilize the analyte and peptide onto a maleimide presenting monolayer. **b)** 500 µM of CoA conjugates of acetyl, acetoacetyl, HMG, and succinyl were reacted with the peptide CAK(Me_3)SA. The resulting SAMDI spectra show all analytes can be efficiently detected.

NCL has previously seen some use for bioconjugation *in vivo*, with some examples using this chemical strategy to label proteins containing N-terminal cysteines with fluorophores that possess thioester handles.[129] However, the possibility of cross reactivity of NCL chemistry with biological thioesters, such as acyl-CoA species and intermediates of fatty acid synthesis, has been cited as a limitation of the chemistry for use as a bioconjugation technique.[130] In this work, we take advantage of this potential shortcoming, using it to develop an assay targeting acyl-CoA metabolites.

We first validated this chemical strategy for capture and detection using a variety of commercially available, biologically relevant, CoA compounds. These CoA species were each reacted with the capture peptide, of sequence CAK$(Me)_3$SA, in phosphate buffered saline (PBS) at pH 7 and 40° C, at concentrations of 500 μM of CoA species and 1 mM of capture peptide for 2 hrs. The reactions were then applied to self-assembled monolayers presenting maleimide at a density of 20% against a background of tri(ethylene glycol) groups and analyzed by SAMDI mass spectrometry to reveal peaks that correspond to capture of the acyl groups (**Figure 7.1b**). For the purposes of this assay, the capture peptide requires an N-terminal cysteine, but the sequence does not strongly impact the reaction. The peptide used in this work was chosen because of its high ionization efficiency, a characteristic that is important for maximizing the limit of detection in mass spectrometry. We characterized the kinetics for the reaction of the peptide with both acetyl-CoA and HMG-CoA (See Figure S1 in the appendix). Both species were found to have similar second order rate constants, which fall between 0.10 $M^{-1}s^{-1}$ and 0.04 $M^{-1}s^{-1}$ over the pH range of 7.5 to 6.5, where the rate increases with increased pH.

Characterization of Pathway Kinetics and Flux

We applied this strategy to analyze the cell-free biosynthesis of HMG-CoA (**Figure 7.2a**)—the biosynthetic precursor to isoprenoids, which represent a diverse and useful class of molecules with applications in pharmaceuticals, flavorings, fragrances, commodity chemicals, and biofuels.[105, 131-134] This biosynthetic pathway starts with acetyl-CoA (Ac-CoA) and proceeds through the intermediate acetoacetyl-CoA (AA-CoA) to HMG-CoA. We perform the biosynthetic

reactions directly in crude lysate, allowing us to rapidly generate the pathways by mixing crude *E.coli* lysates that separately overexpressed the necessary acetyl-CoA acetyltransferase and HMG synthase enzymes. The cell-free reactions also allowed strict control of the concentrations of glucose, buffer, salts, and cofactors.[135]

All cell-free reactions were done in standard 384 microtiter well plates and were quenched by addition of formic acid. For capture of CoA metabolites, the pH was adjusted to 6.5 and the capture peptide was introduced. The crude reaction mixtures were diluted several fold and applied to arrays of monolayers presenting maleimide functional groups using liquid handling robotics, where captured analytes were selectively isolated from the reaction mixture (**Figure 7.2b**). Analysis was performed using SAMDI arrays of 1536 spots, where each spot is 1 mm in diameter. The reactions from the 384 well plates were each spotted in quadruplicate so that 4 replicate spectra could be averaged for each reaction. For each spot, 0.5 μL of the reaction mixture was used for immobilization.

We first characterized the kinetics of HMG-CoA production in these cell free reaction systems with concentrations of the supplied cofactors ATP, NAD^+, and CoA at 1 mM each (**Figure 7.2c-d**). We measured the rate of HMG-CoA production in two buffer systems utilizing either acetate or glutamate salts, both commonly used in cell-free reactions.[136] Both buffer systems reached similar yields of just over 600 μM of HMG-CoA, though HMG-CoA accumulated more rapidly in the acetate-based buffer, where it was detectible at the earliest reaction times. The same reaction in the glutamate-based buffer required approximately 30 min of reaction time to observe HMG-CoA.

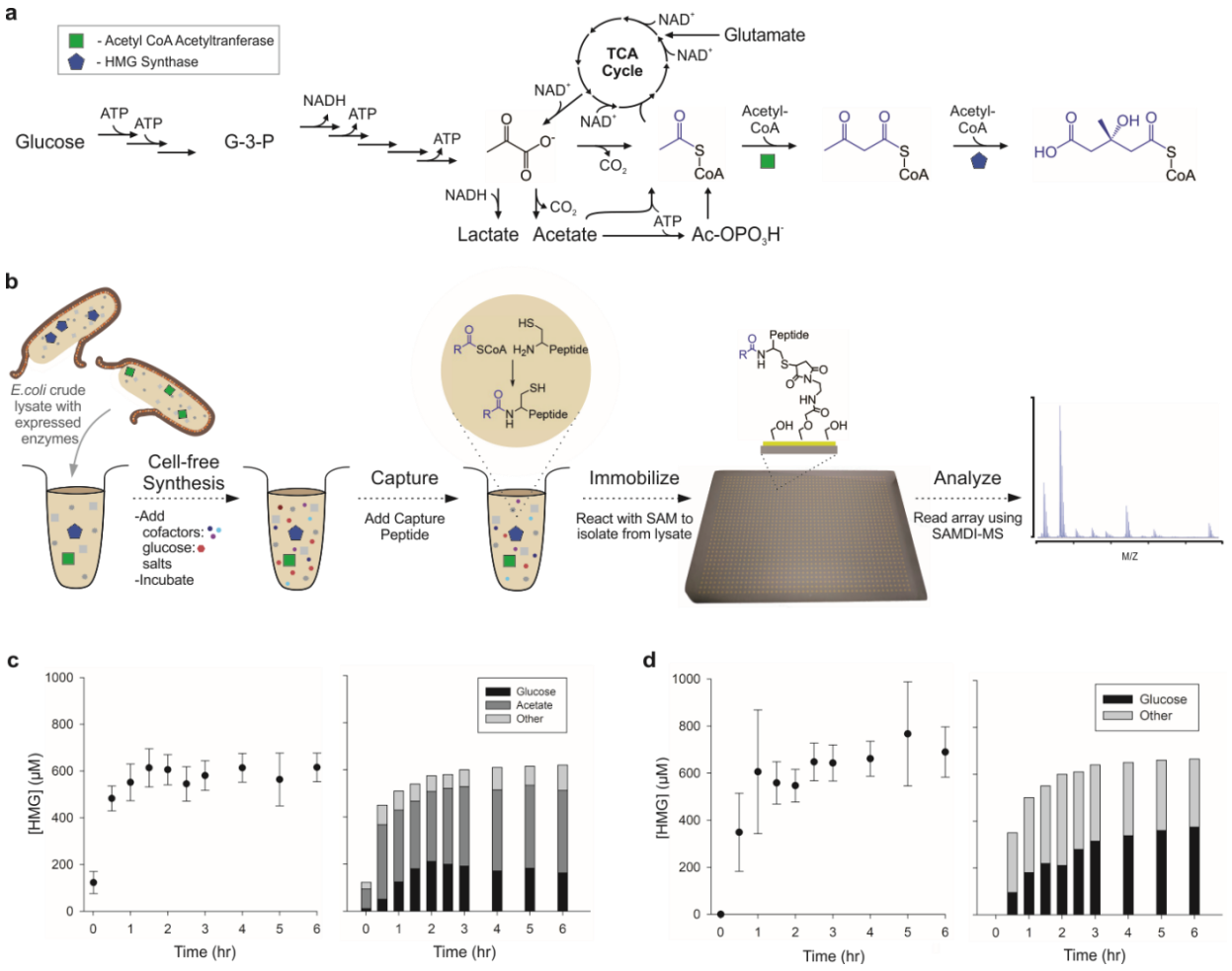


Figure 7.2. Pairing SAMDI-MS with Cell Free Biosynthesis. a) Cellular overexpression of ACAT and HMGS enables generation of enzyme-enriched lysates which can metabolize acetate and glutamate salts in the buffer, in addition to glucose. The pathway contains multiple CoA bound metabolites in addition to the HMG-CoA product**. b)** Cell-free reactions containing lysates, cofactors, buffers, and substrate were performed in standard 384 microtiter well plates. Reactions were then quenched, and any CoA products were captured by incubation with sensor peptide. Complex reaction mixtures were printed onto monolayer arrays using liquid handling robotics for isolation and detection. **HMG-CoA accumulates in the reaction over time** for both **c)** acetate salt buffer and **d)** glutamate salt buffer. The dominant carbon source used for HMG production was determined by feeding cell-free reactions ^{13}C-labeled glucose and ^{13}C-labeled acetate and monitoring isotopic incorporation into the HMG product**.** The concentration of cofactors ATP, NAD^+, and CoA were set to 1 mM each.

While we supplemented the lysate with glucose as a carbon source for metabolism, acetate and glutamate act as secondary sources of carbon and can also be metabolized by enzyme pathways native to *E.coli* and converted to HMG-CoA (**Figure 7.2a**). We determined the fraction of product coming from these different carbon sources by repeating the experiments with $^{13}C_6$-glucose and quantifying the isotopic incorporation into the final product. Similarly, $^{13}C_2$-acetate was used to monitor metabolic flux from acetate in the cell-free reactions where it was present. HMG-CoA is composed of three, 2-carbon acetyl subunits such that four possible label states (0, +2, +4, or +6 mass units) are observed (**Figure 7.3b**).

We quantified the relative abundance of each from the resulting mass spectra and then used these values to determine the fraction of labeled carbon, f^{C13}, and the corresponding fraction of unlabeled carbon, f^{C2}, in the HMG-CoA produced**.** These labeled states, with mass shifts of just +2 mass units each, overlap with the naturally occurring isotopic splitting pattern for the HMG-CoA product peak. However, the isotopic splitting for the native substrate can be easily calculated and determined experimentally (**Figure 7.3a**), and this known isotopic splitting pattern of the unlabeled species can be used to deconvolute spectra from labeling experiments and determine the contributions of the possible labeled states. This allows the AUP contributed from each label species to be determined and, from that, the fraction of labeled and unlabeled carbon in the HMG product to be calculated (**Figure 7.3c**).

We also used quantitation of the incorporation of label to determine the absolute concentration of HMG-CoA produced. After quantification of label incorporation, a known amount of commercially available, unlabeled HMG-CoA was spiked into each capture reaction.

We used the change in detected label content to determine the original concentration of HMG-CoA, as described by equation 1.

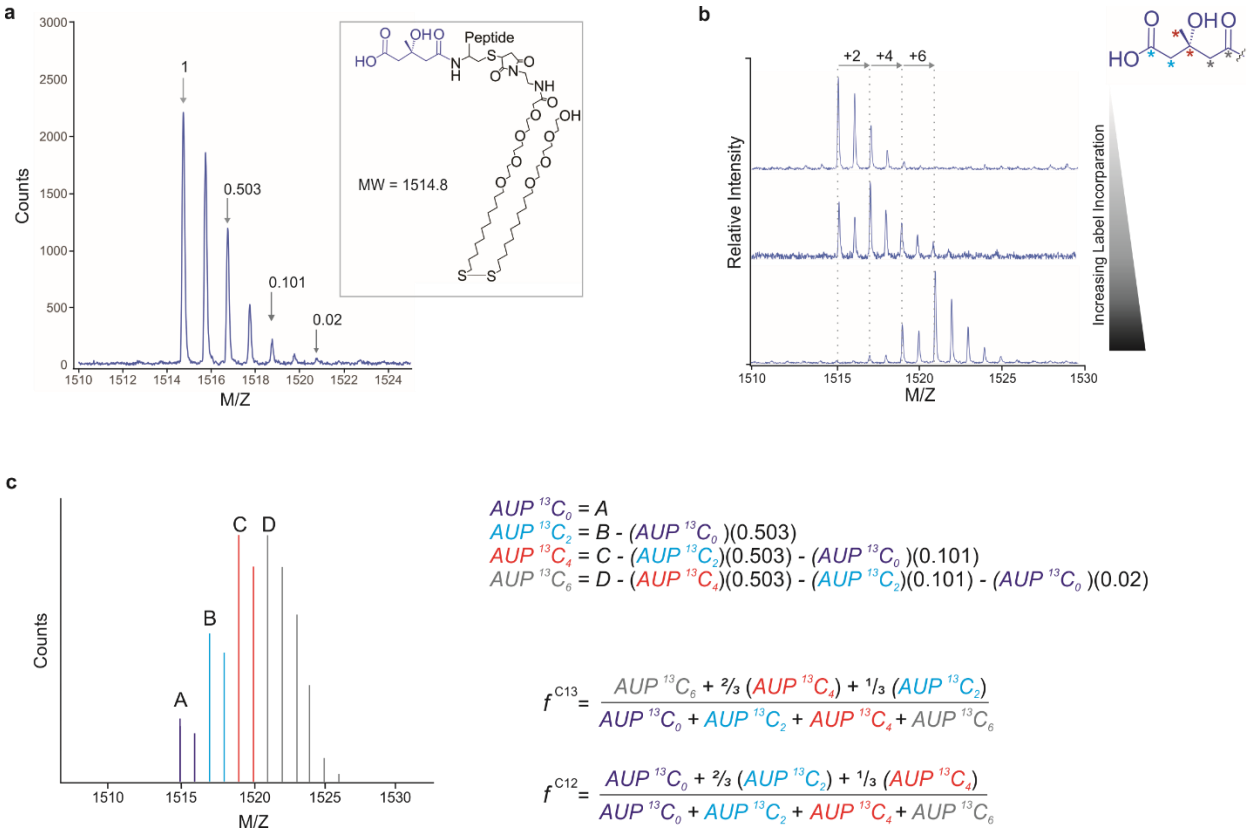


Figure 7.3. Interpreting the Assay Results. a) In SAMDI mass spectra, the entire monolayer is desorbed and detected as a disulfide to a molecule of the tri(ethylene)glycol background. For the HMG adduct, this observed product has a M/Z of just under 1515 mass units. The natural isotopic splitting pattern can be calculated or determined experimentally. **b)** When the cell-free reaction is provided isotopically labeled $^{13}C_6$ glucose, the incorporation of this label is observed as a peak splitting pattern of +2, +4, and +6 mass units, relative to the unlabeled species. The distribution of these possible HMG labeling states depends on the relative concentrations of unlabeled versus $^{13}C_2$ -labeled Ac-CoA in the cell-free reaction. **c)** The four label states (0, +2, +4, or +6) can be separately quantified from the convoluted spectra by accounting for the overlap from the natural isotopic splitting of the peaks in front of it. From the AUP of the four peaks A, B, C, and D, the AUP for each of the labeling states was calculated as shown. The deconvoluted AUP for each of the labeling states was then used to determine the fraction of ^{13}C label, f^{C13}, in the observed HMG product by weighting the AUP for each labeled state by the relative number of labeled carbons in it, then dividing by the sum of all four states.

Here, f_i^{C12} represents the fraction of unlabeled carbon before the spike and f_{spike}^{C12} represents the fraction of unlabeled carbon after the spike. The initial concentration of HMG, $[HMG]_i$, can be determined from these experimental measured values.

$$f_{spike}^{C12} = \frac{f_{initial}^{C12} * [HMG]_i + [HMG]_{spike}}{[HMG]_i + [HMG]_{spike}} \quad (1)$$

This method allows for determination of concentration by measuring the change in ^{12}C versus ^{13}C content and is ambivalent to variations in signal intensity across spectra, as it relies on the peak splitting pattern for the HMG product. This method enabled the tracking of both HMG-CoA yield and carbon usage simultaneously**.**

Identification of an Additional Metabolite

The cell-free reactions performed here contained 200 mM of glucose and 150 mM of either acetate or glutamate. We observed that acetate was the primary source of carbon in the acetate buffer system, and for early timepoints it was the source for the bulk of the HMG-CoA produced (**Figure 7.2c**). With longer reaction times, glucose utilization increased and leveled off at 30% of the incorporated carbon. In the glutamate buffer system, glutamate was also a major source of carbon, but the carbon fraction from glucose reached the higher level of 55% after 6 hours (**Figure 7. 2c-d)**. These results show both acetate and glutamate are important sources of carbon in these cell free systems and not just passive components of the buffer system.

It remains challenging to independently and quantitatively control the concentrations of cofactors in cell-based biosyntheses. The cell-free approach we use here makes it straightforward to independently optimize the buffer composition, cofactor concentrations, and carbon sources to

achieve the greatest yield of biosynthetic products. In particular, it has been observed that supplementation of ATP, NAD^+, and CoA can dramatically influence cell-free metabolism.[135] As described below, we fully mapped the dependency of HMG-CoA synthesis relative to these cofactors. However, during initial experiments, we observed an unidentified product bound to the peptide sensor that did not correspond to any of the expected CoA metabolites. This unidentified product gave a mass shift of +72 Da, and this peak increased by 3 mass units when the lysates were supplemented with ^{13}C glucose. This mass and carbon content was consistent with a lactate-functionalized peptide, but no enzyme in *E.coli* is known to produce lactyl-CoA.

We suspected this unidentified product derived from glycolysis and we identified the source as glyceraldehyde-3-phosphate (G3P) by biochemically reconstituting the initial steps of glycolysis using five purified enzymes (**Figure 7.4**). When the pathway was reconstructed stepwise and fed glucose, this species appeared only when the enzymes necessary to produce G3P were present. We suggest that the peak stems from a pyruvaldehyde (also known as methylglyoxal) intermediate, which is formed from G3P and dihydroxyacetone phosphate (DHAP) through elimination of the phosphate and is a known toxic byproduct of glycolysis. The enzyme that interconverts G3P and DHAP, triosephosphate isomerase, has also been shown to catalyze the formation of pyruvaldehyde from these species.[137-139] Previous work has shown that pyruvaldehyde reacts with thiols, through a thiohemiacetal rearrangement, to produce a lactyl-thioester (**Figure 7.5a**).[140-141] In this case, the lactyl-thioester can further react to yield a N-lactyl-peptide.

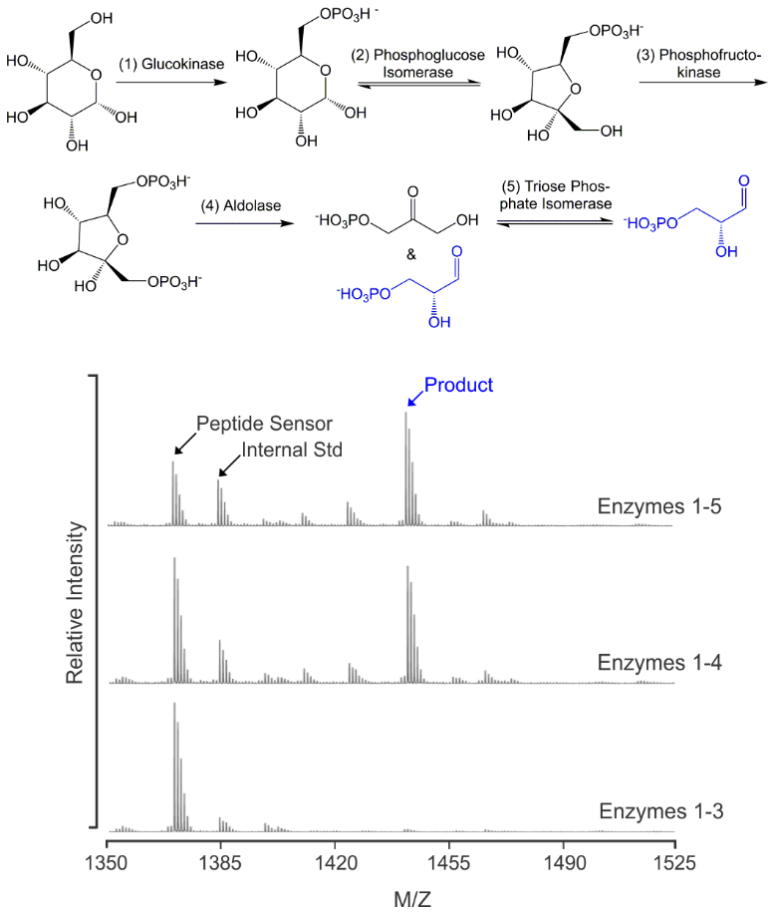


Figure 7.4. Detection of Glycolysis Intermediates. The initial steps of glycolysis were reconstituted *in vitro* using purified enzymes and fed glucose. When sufficient enzymes are present, G3P is captured and detected.

We also observed a peak corresponding to the aldol-condensation product of two molecules of pyruvaldehyde (**Figure 7.5b**). This dimeric adduct overlaps with that for HMG-CoA in the mass spectrum. However, it can be differentiated from HMG because it derives from two 3-carbon subunits and possesses a characteristic +3, +6 labeling pattern when isolated from lysates that are supplemented with ^{13}C glucose. We found that this aldol condensation product was only observed at a pH above 7. Therefore, to exclude capture of this adduct, we performed all capture reactions

at pH 6.5; we confirmed that the thioester rearrangement still proceeds at this pH, but the aldol condensation does not occur.

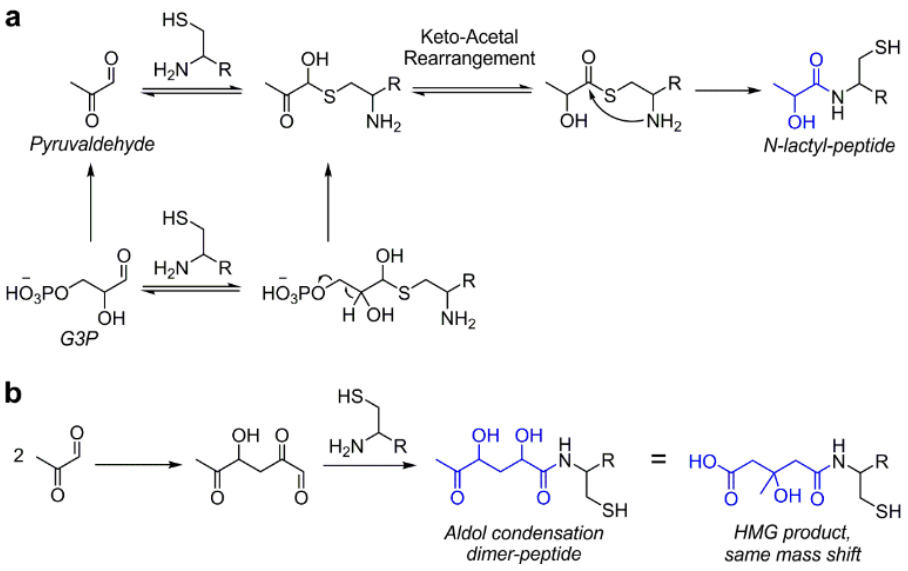


Figure 7.5. Derivatives of non-CoA bound, G3P were captured by the sensor peptide. **a)** A possible mechanism for detection of G3P as N-lactyl-peptide is via pyruvaldehyde, which is known to be generated from G3P and can undergo rearrangement with thiols to form lactyl-thioesters. **b)** Pyruvaldehyde can also undergo pH-dependent, aldol condensation to yield a 6-carbon species that overlaps in mass with the desired HMG product.

Mapping Metabolite Levels and Pathway Flux

In order to optimize HMG production and better understand the effect of cofactor concentrations on the reaction network, we performed a high throughput experiment that tested 768 unique cofactor conditions—where the concentrations of NAD, CoA and ATP were systematically varied. For each condition three replicate cell-free reactions were performed, and for each reaction four spectra were collected and averaged, totaling over 9,000 individual spectra (**Figure 7.6**). All cell-free reactions were allowed to proceed for 2 hours, quenched, reacted with the capture peptide, and immobilized as previously described. We did not observe AA-CoA under any conditions tested. In a separate control, we supplemented reactions with commercial AA-CoA and found that it was readily converted back to Ac-CoA (**Figure 7.7**). The ACAT enzyme is reversible and these results suggest that it strongly favors the reverse reaction. In the reactions with the full pathway, the small quantities of AA-CoA produced are likely rapidly converted to HMG-CoA, and so we do not observe any significant accumulation of AA-CoA.

For each spectrum, we determined the percent conversion for the three primary observed species (HMG-CoA, Ac-CoA, and G3P) using the integrated areas under the peaks (AUP) for the unreacted sensor and the products. Below, equation 2 shows how this calculation was done for HMG and analogous calculation were performed for acetyl and G3P products as well.

$$Percent\ conversion\ for\ HMG = \frac{AUP_{HMG}}{AUP_{HMG}+AUP_{Ac}+AUP_{G3P}+AUP_{sensor}} * 100\% \qquad (2)$$

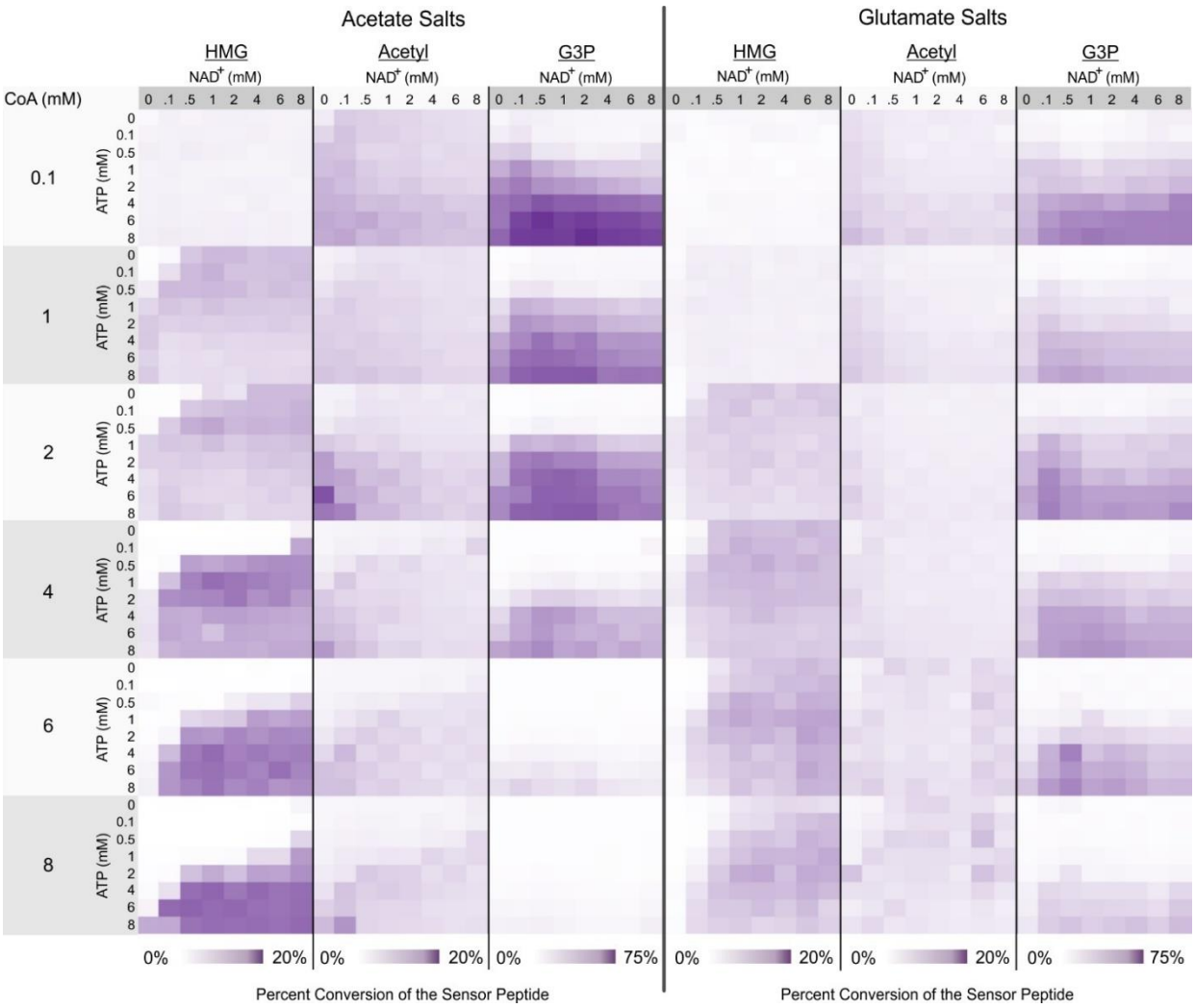


Figure 7.6. Parallel measurement of Ac-CoA, HMG-CoA, and G3P in 768 unique cofactor conditions. Each product was quantified by calculating percent conversion relative to the unreacted peptide sensor using the area under the curve for each species in the mass spectra. All reactions proceeded for 2 hrs at 37° C.

These high throughput experiments revealed several notable trends. First, when low concentrations of CoA and high concentrations of ATP were used, G3P dominated as the major product. This is perhaps not unexpected as low CoA limits the total amount of HMG-CoA the system is capable of synthesizing and the high ATP accelerates the early, pay-in phase of

glycolysis. Second, as the concentration of CoA is increased, the amount of G3P decreased, and this trade-off was much more significant when reactions were performed in the acetate buffer. Thus, G3P was prominently present in conditions favoring early glycolysis, but where that metabolic flux was unable to proceed to Ac-CoA. The relationship of HMG-CoA yield with cofactor concentrations differed significantly in the buffer systems based on acetate and glutamate salts, as described next.

We further analyzed the same reaction dataset to understand how the yield of HMG-CoA varied for all reaction conditions (**Figure 7.8a**). As described earlier, we used ^{13}C labeled glucose to monitor metabolic flux from glucose across the array. For each buffer condition, we identified the three reactions that produced the highest concentration of HMG-CoA among the set of 768 reactions. The averages of these highest yielding reactions were 6.4 ± 0.5 mM and 2.6 ± 0.1 mM for the acetate and glutamate systems, respectively. Across the array, the AUP for the HMG-CoA product was normalized to an internal standard present at a constant concentration across all reactions to control for variations in ionization across spectra. For this standard we used the peptide Ac-SK(Me)$_3$GGC which possesses similar ionization efficiency to the capture peptide but lacks an N-terminal cysteine and does not overlap in mass with any species of interest. We explored the limit of detect by performing reactions with known quantities of HMG-CoA under reaction conditions identical to those used for the array (see Figure S2 in the appendix). Under these conditions, HMG-CoA was detectible at concentrations as low as 5 μM, suggesting the observed concentration range for the array spans roughly 3 orders of magnitude.

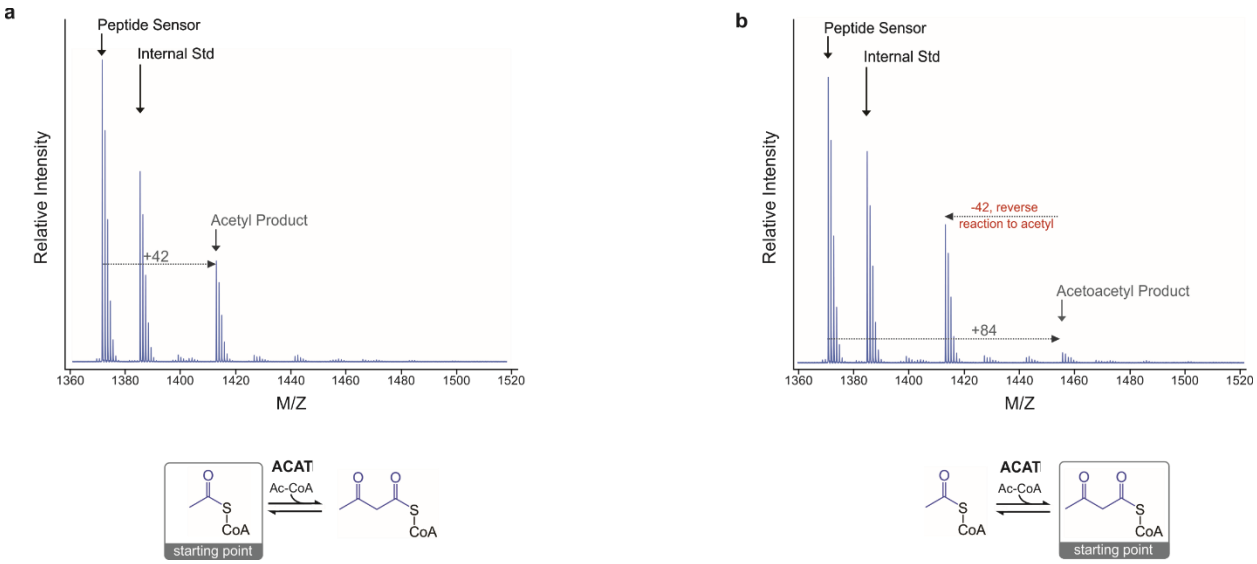


Figure 7.7. The pathway intermediate, acetoacetyl-CoA, is not observed. To explore why we do not observe the AA-CoA pathway intermediate, we incubated 1 mM of Ac-CoA or AA-CoA in cell-free lysate overexpressing the ACAT enzyme for 2 hours. The cell free reactions were given 1 mM each of CoA, NAD^+, and ATP, but no glucose was provided. **a)** When this cell-free reaction was incubated with Ac-CoA, this acetyl species is the primary species we capture and detect. **b)** Conversely, when the cell-free lysate is incubated with AA-CoA, a relatively small amount of acetoacetyl species is detected and the acetyl species is by far the primary product.

Metabolic flux and HMG-CoA production complexly interplayed with cofactor concentrations and we observed significantly different reaction profiles for the systems using acetate and glutamate salts. For the acetate system, higher CoA concentrations improved HMG-CoA titers but also required an associated increase in ATP concentration to achieve this maximal concentration. At low ATP concentrations, increasing available CoA inhibited HMG-CoA production (**Figure 7.8b**). Conversely, over the 2-hour reaction time, the glutamate system reached peak HMG-CoA production at 4 mM CoA. Increasing CoA concentration further to 6 and 8 mM did not increase production of HMG-CoA. However, this shifted the region where highest product

concentrations were observed, requiring a shift to higher NAD^+ concentration and somewhat higher ATP concentration to maintain the same titers (**Figure 7.8c**).

When product concentration is highest, the fraction of carbon derived from glucose accounted for 5% of product in the acetate buffer and 22% in the glutamate buffer (**Figure 7.8a**). As in Figure 3, we see significant incorporation of non-glucose carbon (i.e. acetate and glutamate) in the final product for all cofactor conditions tested. At high CoA concentrations, the glucose incorporation decreased. Counterintuitively, in the acetate system, the best cofactor conditions limit glycolytic flux in favor of optimizing the four-step reaction from acetate to HMG-CoA exclusively. These results suggest that multiple substrates are an important consideration when optimizing cofactors for crude-lysate cell-free metabolic engineering.

These results provide a stark reminder of the non-intuitive relationships that underlie the optimization of complex biological reaction networks such as the one presented here. We found a high level of interdependency between the system variables—concentration of CoA, ATP and NAD^+ cofactors and the carbon sources present in the reaction systems—and this interdependency means that an understanding of the reaction space can only be gained by methodical testing of a large number of reaction conditions. The assay we have presented here couples bio-specific chemistry for capture of CoA metabolites with detection by SAMDI-MS to greatly increase the throughput at which these biochemical pathways can be analyzed.

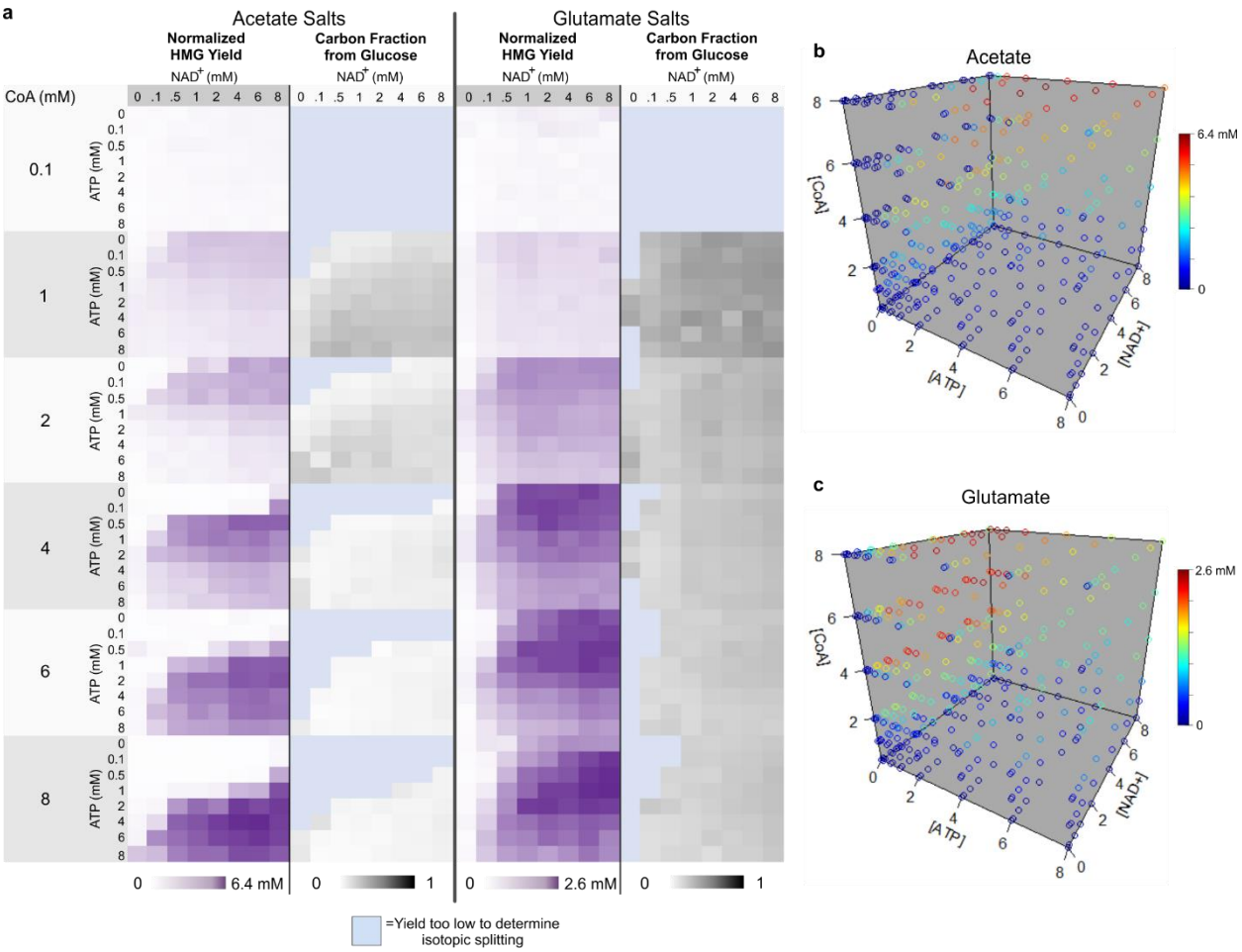


Figure 7.8. HMG-CoA concentration and carbon source shifts in response to cofactor conditions. a) To analyze HMG yield across all conditions, the dataset was normalized to an internal standard, a peptide of similar sequence to the sensor without an N-terminal Cys, present at a constant concentration across all reactions. C^{13} labeled glucose was used to concurrently monitor the fraction of HMG product coming from glucose. HMG production was also visualized as 4D plots for both **b)** acetate and **c)** glutamate buffer conditions. For the three highest yielding conditions in each buffer, average [HMG] was determined. In these plots each point represents a specific concentration condition of cofactors with [ATP], [NAD^+], [CoA] on the x, y, z axis respectively. Color of each point represents yield with highest yield represented by red. From acetate to glutamate; the red region shifts from high CoA and high ATP, to moderate CoA and low ATP.

7.3 Conclusions

In this work we present design of an assay that enables the high-throughput analysis of CoA utilizing metabolic pathways. This assay combines a general strategy for biospecific capture of CoA bound metabolites with simple purification by covalent immobilization onto self-assembled monolayers, that serve as the platform for detection by SAMDI mass spectrometry. We applied this assay to the characterization of a cell-free reaction system for the biosynthesis of HMG-CoA, demonstrating the ability to perform and analyze thousands of individual biosynthetic reactions and to map the reaction space for this complex system with respect to multiple variables.

We believe this assay represents an important contribution to methods available for the design of biosynthetic pathways. Non-intuitive relationships underlie the optimization of complex biological reaction networks such as the one presented here. Understanding the high level of interdependency between system variables that govern the function of these reaction networks can only be ascertained by methodical testing a large number of reaction conditions. Technologies for rapid characterization, such as this offer the potential to synergize with existing methods for accelerating the design and construction of biosynthetic system, as well as providing large experimental datasets to inform the future generations of design tools. We have presented here the design of an assay capable of tracking multiple metabolites simultaneously unambiguously by mass spectrometry, and compatible with straightforward tracking of metabolic flux using isotopic labeling. When compared to available methods, our strategy offers a leap in throughput representing an orders of magnitude increase, enabling dramatically faster characterization of CoA dependent metabolic pathways and to serve as a tool to speed up the bioengineering design cycle.

7.4 Methods

Preparation of Enzyme-enriched Lysates. Three strains of *E. coli* BL21(DE3) containing plasmids pETBCS-ACAT(Eco) and pETBCS-HMGS(Sce) were grown in a 1 liter fermenter at 37 °C (250 rpm) at constant pH (6.95) in rich media (18 $g \cdot L^{-1}$ glucose, 16 $g \cdot L^{-1}$ yeast extract, 10 $g \cdot L^{-1}$ tryptone, 5 $g \cdot L^{-1}$ NaCl, 7 $g \cdot L^{-1}$ potassium phosphate dibasic (K_2HPO_4), 3 $g \cdot L^{-1}$ potassium phosphate monobasic (KH_2PO_4)) containing 100 $\mu g \cdot mL^{-1}$ carbenicillin (IBI Scientific, Peosta, IA). After induction with 0.1 mM IPTG at OD_{600} 0.6, growth continued for four hours at 30 °C until harvest by centrifugation. See [135] for plasmid sequences and expression characterization. *E. coli* cells were lysed and crude lysates were generated using methods previously described. [135]

Cell Free Reactions. Cell free reactions were performed at a volume of 15 µL in 384-well plates and incubated at 37 °C. The standard reaction contained the following components: 200 mM glucose, 100 mM Bis Tris buffer, acetate or glutamate salts (8mM magnesium, 10mM ammonium, 134 mM potassium), and 10mM potassium phosphate (K_2HPO_4, pH 7.2). Unless specified, reactions also included 1 mM NAD^+, 1 mM ATP, and 1 mM CoA.[135] All reagents and chemicals were purchased from Sigma Aldrich (St. Louis, MO). Two lysates, enriched in ACAT(Eco) and HMGS(Sce) respectively, were mixed together at a total protein concentration of 5 $mg \cdot mL^{-1}$ each. Reactions were quenched by precipitating proteins using 2.25 µL of 10% formic acid and immediately stored at -80 °C until peptide incubation. Acetic acid-2-^{13}C was neutralized to acetate (pH 7.00) by titration with 45% w/w potassium hydroxide and diluted to 5 M stock concentration.

Preparation of Monolayer Arrays. Array plates were prepared by patterning 1536 gold spots, in a standard microtiter format, on steel plates using electron-beam metal evaporation to deposit 5 nM of titanium, followed by 30 nM of gold. These plates were soaked in a solution of disulfide molecules in ethanol for 24-48 hours to form a self-assembled monolayer on the gold surfaces. The solution consisted of a mixture of EG3-alkanethiol disulfide and a mixed disulfide of eg3-alkanethiol and maleimide-terminated EG3-alkanethiol.[90] The two disulfide molecules were present in a stoichiometric ratio to yield a 20% maleimide surface density, with an overall concentration of 1mM. After formation of this primary monolayer on the gold spots, the plates were rinsed with ethanol and dried. The array plates were then soaked in a 10 mM hexadecylphosphonic acid solution in ethanol for 15 minutes. The phosphonic acid terminated hydrocarbon molecules react selectively with the steel background giving it hydrophobic properties that aid with delivery of sub-microliter reaction volumes to the spots.

Synthesis of Peptide Reagents. The peptide of the sequence CAK$(Me)_3$SA was used in this work for capture of acyl-CoA species. This sequence was chosen after testing a small set of potential 5-mer sequences, all containing N-terminal cysteine residues. While the sequence identity did not strongly influence the reactivity towards CoA metabolites, it did impact the analysis by SAMDI-MS. We found the trimethyl-lysine residue to provide excellent ionization efficiency. Ionization efficiency is critical with respect to the assay's limit of detection, determining how abundant a species must be for a peak to be observed. Additionally, this sequence and all of its observed reaction products ionized as a single molecular-ion peak, rather than a set of salt adducts commonly observed with MALDI-TOF mass spectrometry. This greatly simplified the resulting spectra and

their interpretation. These advantageous properties likely stem from the enforced positive charge provided by the trimethyl-lysine residue. The peptide Ac-SK(Me)$_3$GGC was used as an internal standard for normalization of signal across reactions, as it possesses similar sequence and ionization efficiency to the capture peptide but is mass-resolved from all reaction species and lacks an N-terminal cysteine. Standard FMOC solid phase peptide synthesis on rink-amide resin was used to synthesize both peptides. To introduce the non-natural trimethyl-lysine residue, FMOC-Lys(Me)$_3$Cl was purchased from Novabiochem and used along with standard amino acid coupling conditions.

Capture of CoA Bound Moieties in Lysates. After completion, cell-free reactions were quenched with formic acid to denature the proteins and stop the reactions. The reactions, in 384 well plates, were centrifuged at 3,500 *x g* for 15 minutes to pellet any precipitated protein. For the acyl-CoA capture reactions, 3 µL of this cell-free reaction was transferred to a new 384-well plate and the following species were added, bringing the final reaction volume to 8 µL with the final concentrations as follows: 100 mM phosphate buffer at pH 6.5, 40 mM EDTA, 0.9 mM capture peptide, and 0.1 mM normalization peptide. The well plates were sealed, and the reaction mixtures were incubated at 42º C for 3 hours. It is important to choose and appropriate concentration of capture peptide. If the concentration is too high, the signal from unreacted sensor peptide may overwhelm the signal from any captured species. In this work the total added peptide concentration was chosen to be 1 mM across all reactions, a reasonable concentration relative to the expected yield from the best cell-free reactions, while also giving good dynamic range for detection of acyl-species in low-yield conditions.

Immobilization of Captured Products. The 8 µL capture reactions in 384 well plates were diluted 3-fold to 24 µL with 100 mM phosphate buffer at pH 7.2. This serves to dilute the concentration of peptide, which will immobilize efficiently at concentrations as low as 10 micromolar and adjust the pH to an optimal range for the reaction with maleimide. A TECAN liquid handling robot equipped with a 384-tipped head was used to print the reaction mixtures onto 1536-spot array surfaces, generating 4 replicates per reaction, at a volume of 0.5 µL per spot. The surfaces were placed in a humidified chamber and incubated at 37° C for 60 minutes to allow the cysteine-containing peptides to react with the maleimide functionalized SAM. After reaction, the surfaces were rinsed with 1% SDS detergent to remove excess protein that could hinder ionization and detection, then re-rinsed with distilled water and dried with a stream of nitrogen gas.

Analysis of Reactions. Matrix of THAP in acetonitrile (20 mg/mL) was applied directly to the 1536 spot-surface. The matrix was allowed to dry, and each spot was analyzed by MALDI-TOF mass spectrometry using an AB Sciex 5800 series instrument. Captured metabolites of the cell free reactions were identified by their mass shifts relative to the unreacted peptide. Analysis of the more than 10,000 spectra collected was performed using in-house, Python-based software package to automate the integration of peaks of interest from the mass spectra dataset. The software calculates the area under the curve at any m/z value given to it over a user-defined half-width mass window. For this work a half width mass window of 0.2 mass units was used across all spectra analyzed. For every spectrum in this dataset, the following peaks were integrated: 1371 Da

(capture peptide), 1385 Da (normalization peptide), 1413 Da (Ac-CoA reaction product), 1515 Da (HMG-CoA reaction product), 1443 Da (pyruvaldehyde adduct). Additionally, for each peak analyzed, regions immediately flanking the peak were integrated to determine a local spectral baseline, which was subtracted from the peak area. For isotopic labeling experiments, the area under the curve for various isotope peaks were also collected.

Chapter 8

Characterizing the Ligation of an α-Keto Aldehyde and Cysteine Using Microfluidics and Imaging SAMDI

This chapter is adapted from the following published work:
Grant, J.; O'Kane, P.; Kimmel, B.; Mrksich, M. *Using Microfluidics and Imaging SAMDI-MS To Characterize Reaction Kinetics*. ACS Central Science. 2019, *5* (3), 486-493

8.1 Introduction

In the previous chapter, we identified a novel mode of reactivity between pyruvaldehyde and the N-terminal cysteine of a peptide. Here we will explore this ligation reaction, characterizing the pH dependent kinetics, by combining microfluidic reactors with rapid analysis by SAMDI-MS. Microfluidic systems have shown unprecedented levels of precision, miniaturization, and operational control over traditional benchtop and batch-based procedures. For example, a chemical reaction with a rate constant of $k \sim 3 \times 10^4\ M^{-3}\ s^{-1}$ was analyzed by fusing two droplets containing the reactant and recording the reaction progress with a 2000 Hz camera.[142-143] Microfluidic devices have also been designed to screen reaction conditions, where one device can screen 36 temperatures for the synthesis of CdSe nanocrystals, 14 temperatures for studying lipid bilayer phase transition, and 14 temperatures for the activation energy of alkaline phosphatase.[144] However, analysis of reaction progress in microfluidic devices has been mainly limited to optical methods, which require the use of fluorescent or UV-active products that limit applicability towards a small subset of chemical reactions.

Mass spectrometry is a powerful analytical technique that addresses these limitations, where mixtures of compounds can be straightforwardly analyzed without the need for labels. Recently described approaches have coupled microfluidic outlet streams directly to ESI-MS[142, 145-

[146] or combined matrix with outlet droplets for subsequent analysis by MALDI-MS.[147-148] These methods highlight the benefits of integrating structurally unbiased detection strategies with microfluidics to provide high-throughput experimental data. However, ESI-MS and MALDI-MS require sampling an appreciable volume from the reactor outlet and are incompatible with salts and some solvents, thus sacrificing sensitivity and position-to-time information on product conversion. Mass spectrometric approaches that are sensitive, high-throughput, and do not require product workup will present new opportunities for on-chip chemical reaction monitoring.

Here, we use imaging self-assembled monolayers for matrix assisted laser-desorption ionization mass spectrometry (iSAMDI-MS) to describe the kinetic profile of a chemical reaction. In contrast to typical fluidic studies that sample product from the outlet stream or use optical tools for measuring reaction progress, we a designed a fluidic device that possesses chemically defined monolayers on the floor of the channel that can be directly analyzed with a MALDI-TOF mass spectrometer. The monolayer floor is functionalized with a self-assembled monolayer that covalently binds the reactant and product from the flow above, sampling the reaction mixture as it flows through the microfluidic channel. This chemically defined floor of the device collects a spatio-temporal record of the reaction that can later be read and quantified using iSAMDI-MS. Immobilizing the reactant and product to the floor under flow also allowed us to calculate the rate constants within a dispersive regime. In total, we quantified 15,720 data points from 107 μL of reagent for exploring the pH-dependent rate profile of the ligation of cysteine containing peptides with α-keto aldehyde species.

8.2 Results

Reaction Chemistry

Previously, we observed that pyruvaldehyde (also called methylglyoxal) derived from the metabolism of glucose, which contains α-keto aldehyde functionality, was capable of reacting with an n-terminal cysteine to produce a permanent ligation product (**Figure 8.1a**). Methylglyoxal is a known byproduct of glucose metabolism derived from glyceraldehyde-3-phosphate enzymatically and chemically through elimination of the phosphate.[137-139] It is also toxic, as it can react with nucleophilic amino acid side chains to crosslink proteins.[149-150] Methylglyoxal had been shown to react with free thiols in aqueous solution to generate a lactyl-thioester, that could readily hydrolyze to lactate.[140-141] In the previous chapter, we found that when reacting with the thiol sidechain of a cysteine at the N-terminus of a peptide, methylglyoxal could further react to generate lactyl-N-peptide.

Here, we explore this reaction pathway using microfluidics paired with MALDI imaging mass spectrometry to monitor the kinetics of product formation under biologically compatible conditions in buffer of 10 mM phosphate and 100 mM BisTris. In this work, we use phenyl-glyoxal instead of methylglyoxal. As discussed in the previous chapter, methylglyoxal is capable of undergoing dimerization by means of aldol condensation. This process is highly sensitive to pH and thus a significantly more complicated system for the purposes of profiling reaction kinetics. Phenylglyoxal can undergo the same ligation reaction with cysteine but cannot generate aldol condensation products, and so we used this simple α-keto-aldehyde as a model system.

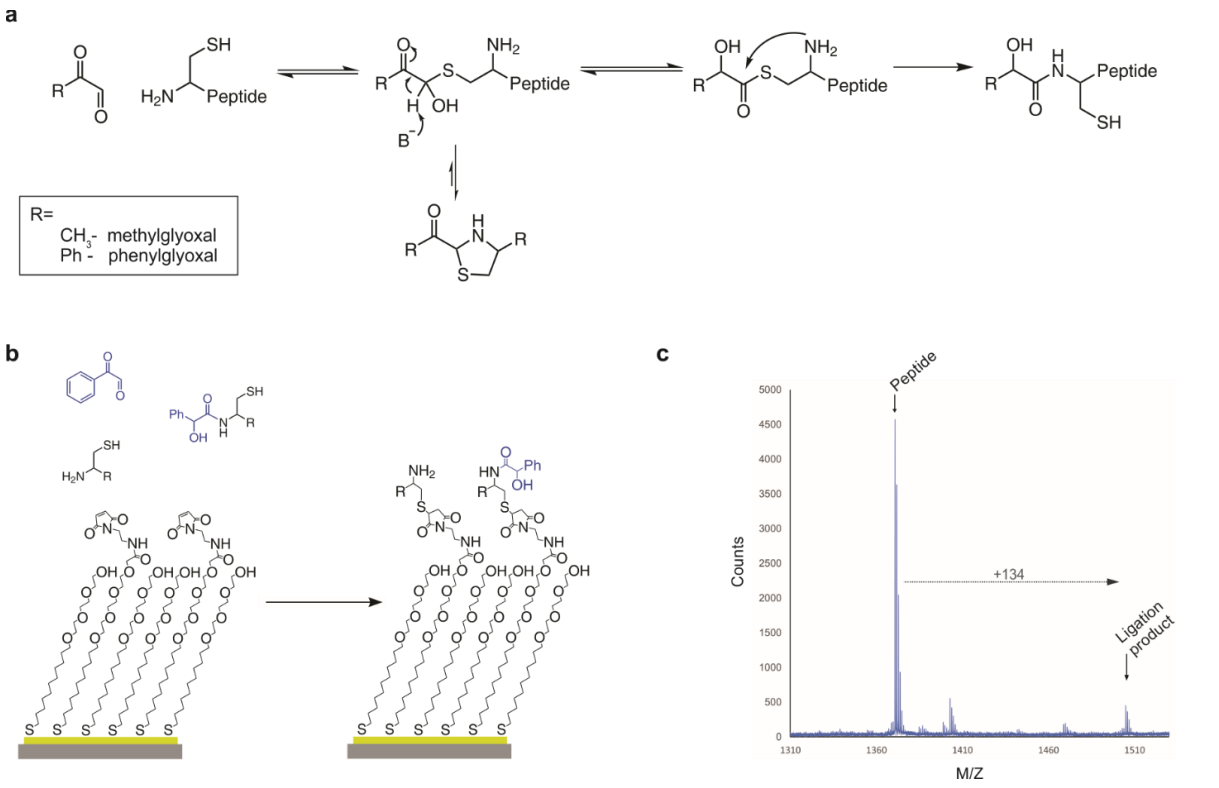


Figure 8.1. Reaction Mechanism. a) Mechanism for the ligation reaction studied in this work. **b)** Self-assembled monolayer presenting maleimide groups reacts with the cysteine-terminated peptide and its adduct. c) SAMDI-MS spectrum of the monolayer after immobilization shows peaks corresponding to the reactant and product.

It should be noted that the phenylglyoxal undergoes reversible thiazolidine ring formation, analogous to similarly observed species between cysteine and simple aldehydes.[140-141] For this reaction, the substrate peptide and functionalized product both possess free thiols and can be selectively isolated from the reaction mixture when exposed to self-assembled monolayers (SAMs) presenting maleimide (**Figure 8.1b**). The resulting monolayer with covalently captured species can be analyzed directly by SAMDI, generating a mass spectrum from which the conversion of substrate to product can be determined from the area under the corresponding peaks **(Figure 8.1c**).

Microfluidic Device Design and Operation

We performed the chemical reaction in a PDMS block containing a Y-shaped microfluidic channel with two inlets and one outlet (**Figure 8.2**). The PDMS block was molded from 3D printed masters to avoid the use of a cleanroom and enable rapid design prototyping.[151] Separate solutions of phenylglyoxal and the peptide, each at a concentration of 2 mM, were simultaneously injected into separate inlets and allowed to diffusively mix at the base of a Y-junction, where they continued to react as the solution flowed along a single channel that was 340 mm long, 550 μm wide, and 250 μm tall. In the flow, we included a peptide (0.25 mM) lacking an N-terminal cysteine that is unable to undergo internal rearrangement to form the permanent hydroxy-amide bond. This served as an internal standard for monitoring imaging performance and consistency. The thiol-containing peptide reactant, hydroxy-amide ligated product, and calibrant underwent immobilization to the monolayer as the fluid front flowed down the channel. The monolayer presents maleimide at a density of 20% relative to a tri(ethylene glycol) background and once saturated, it serves as a record of the reaction profile from the channel. After flowing the reactants through the device, we removed the PDMS block from the chip and applied matrix to the chip as previously described.

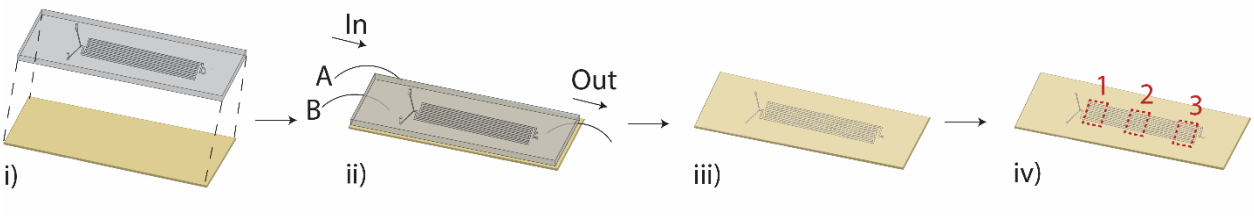


Figure 8.2. Reactor Design and Operation. i) A PDMS block with microfluidic channel features is aligned with a gold-coated slide functionalized with a self-assembled monolayer with terminal maleimide groups. ii) Reagent is simultaneously flowed into the two inlets. iii) The PDMS block is peeled off and the slide is coated with a matrix solution. iv) Three ROIs are identified for iSAMDI-MS acquisition.

We used imaging mass spectrometry (iSAMDI-MS) to generate mass intensity maps of the molecules immobilized to the monolayer, as described recently by our lab.[152] iSAMDI-MS acquires individual SAMDI-MS spectra corresponding to each pixel, where each pixel (i.e. spectrum) reports the chemical composition of the fluid above the self-assembled monolayer. The resulting dataset therefore serves as a spatial map of mass spectra containing maleimide-terminated alkanedisulfides functionalized with the reactant and product, and which represent the extent of reaction at each position in the channel. These pixels correspond to a distinct reaction time (depending on the position down the channel and the flow rate in the channel) and therefore provide a kinetic profile for the reaction (Figure 3A). Here, we used iSAMDI-MS to image three regions by acquiring spectra in reflector positive mode with a 200 μm lateral resolution. We performed experiments where the pH of the buffer was either 6.8, 7.2, 7.6, or 8.0; the spatial maps of product conversion at each pH are plotted in Figure 3B. The fractional yield of product was calculated using Equation 1:

$$\text{fractional yield} = \frac{I_P}{I_P + I_S} \quad (1)$$

In this equation, I_P is the monoisotopic peak intensity of the product-maleimide alkanedisulfide conjugate and I_S is the monoisotopic peak intensity of the reactant-maleimide alkanedisulfide conjugate. Each iSAMDI-MS dataset includes approximately 1,200 individually addressable reactions that occurred in the microfluidic device.

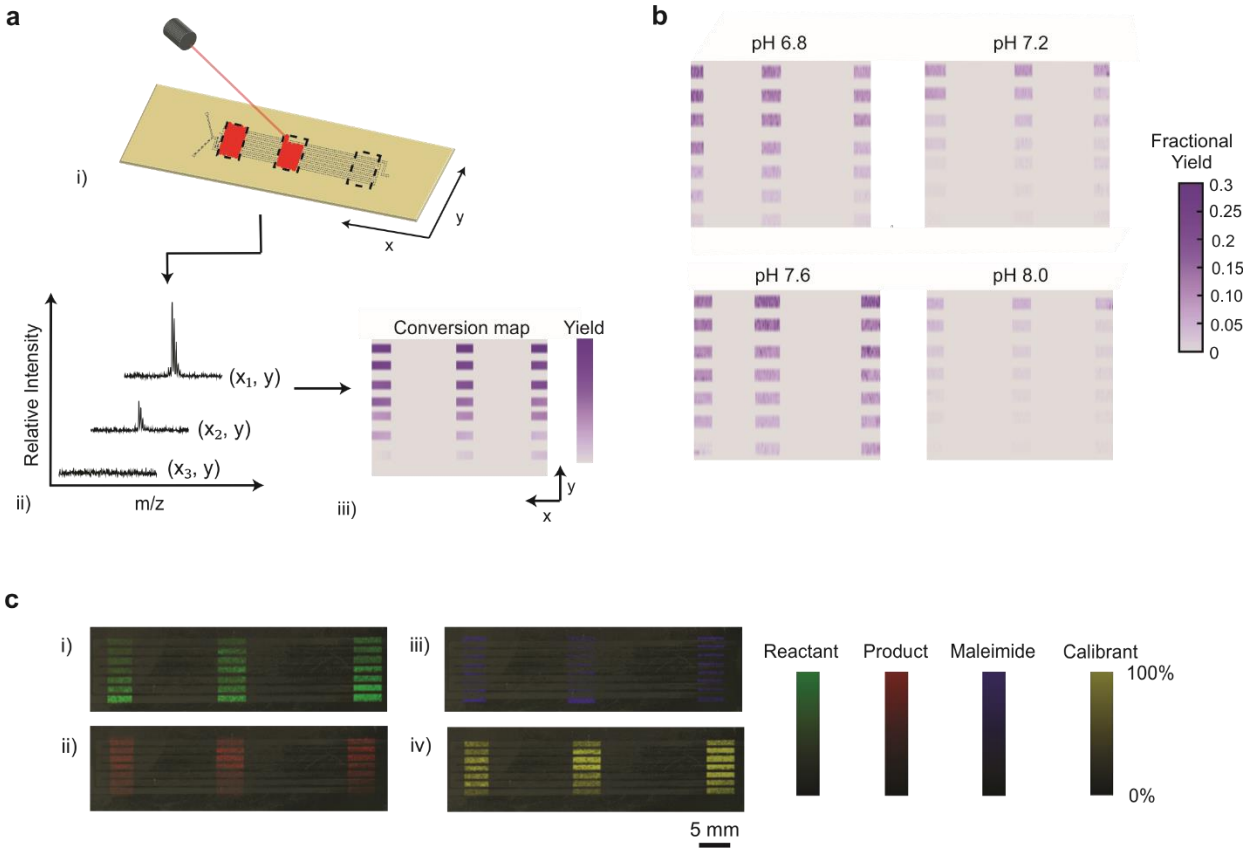


Figure 8.3. Reaction Monitoring in the Channel. a) iSAMDI-MS procedure. i) Three measurement regions on the chip are scanned with a MALDI-MS laser. ii) Each spectrum represents a pixel in the measurement region, and the product yield increases along the channel length. iii) The product yield is extracted from each pixel and plotted into a heatmap. **b)** Heatmaps of fractional product yield at pH 6.8, 7.2, 7.6, and 7.8. **c)** Heatmaps of spectra intensity of various species on the surface, generated from the raw imaging data.

Characterization of Dispersion and Immobilization Kinetics

We determined the fluid velocity inside the chip in order to determine the exact relationship between position on the surface and reaction time in the channel. A solution of 1 mM fluorescein was injected into both inlets at 0.1 µL min^{-1} and images of the fluid front passing through a 12 mm x 12 mm region of the reactor were obtained at 30-s intervals with a confocal microscope (**Figure 8.4a**). The average fluorescence intensities in 14 regions of interest (ROIs) (200 µm x 600 µm) were plotted this against time (**Figure 8.4b**). The resulting curves were fit with a sigmoidal function and the time at which the curves reached half the maximum intensity was plotted against the distance from the Y-junction (**Figure 8.4c**). Plotting the calculated time against the distance from the Y-junction and fitting the plot with linear regression yielded the velocity of the fluorophore through the channel. The experimentally determined fluid velocity of 1.96 mm min^{-1} is 34% greater than the fluid velocity calculated from Q = Av, where Q is the volumetric flow rate and A is the cross-sectional channel area. The 3D printed PDMS masters had rounded corners instead of sharp 90-degree edges and light pressure was applied from an external clamp to keep the PDMS and chip together. Both factors decreased the cross-sectional channel area of the channel and contributed to an increase in observed fluid velocity. We determined the average dispersion coefficient across the three fluorescein experiments to be 0.714 mm^2 s^{-1} using Equation 6 (shown in the methods section). The calculated Peclet number was 836, suggesting that the convective flow of the fluid had a greater effect on the reaction rate than dispersion of peptide species, and that the axial dispersion model (ADM) better represents this system.

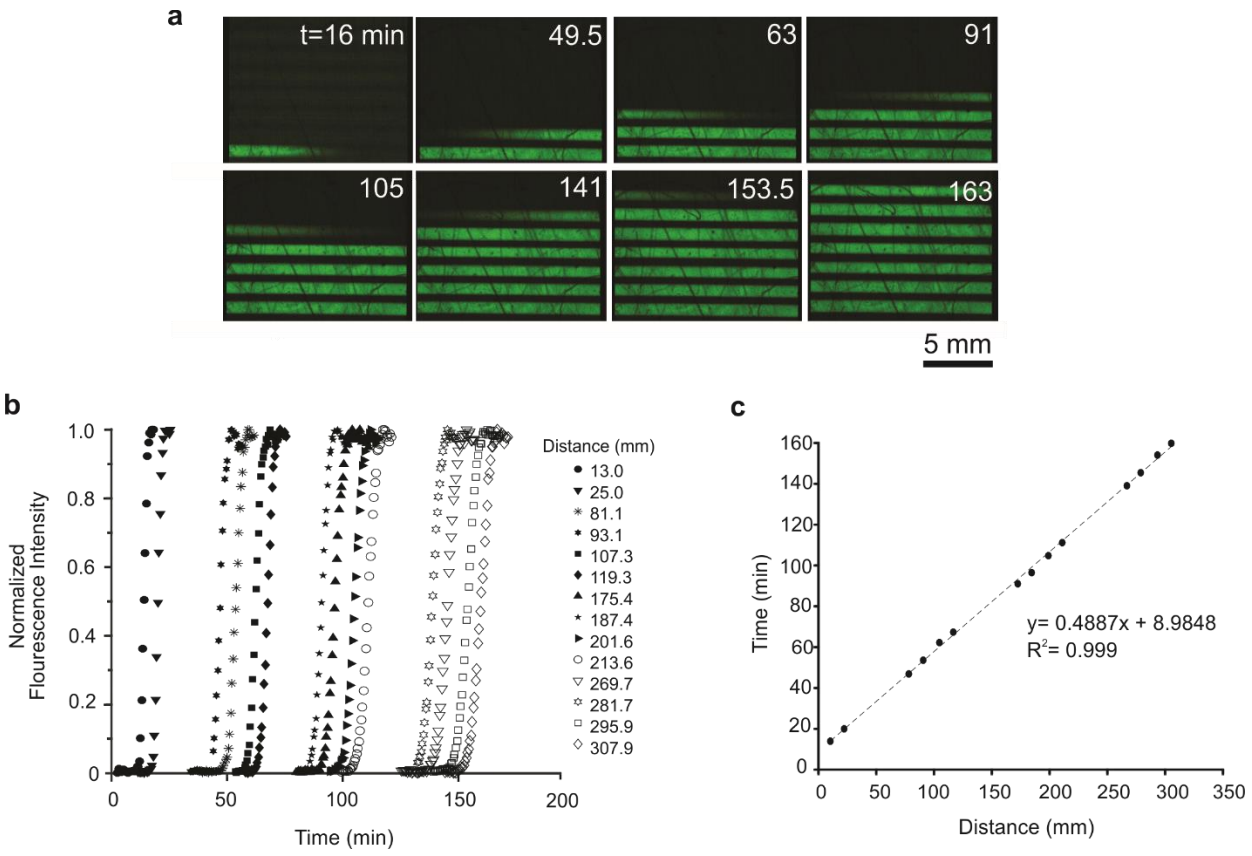
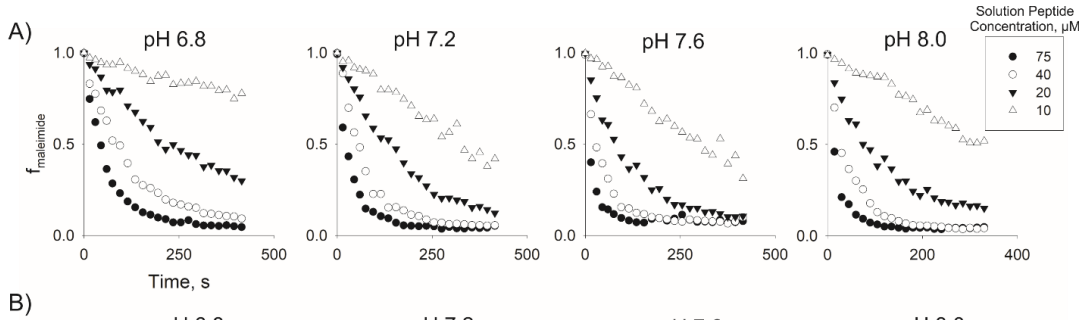
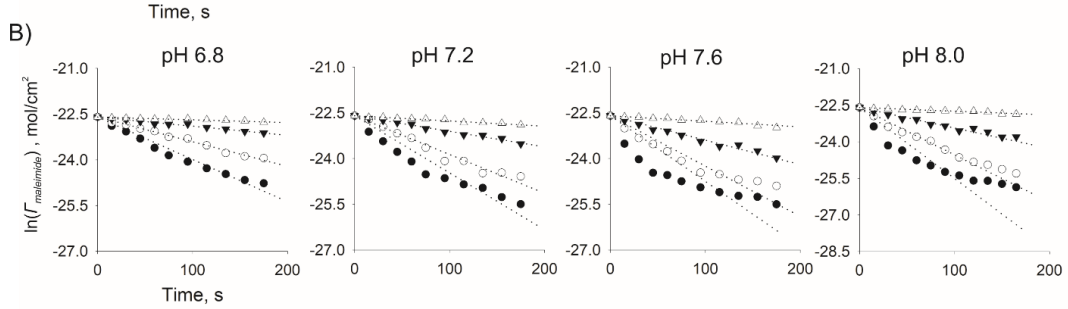


Figure 8.4. Determining Microfluidic Flow Characteristics. a) Tracking fluorescein in the microfluidic device over time. **b)** Plot of normalized fluorescein intensity over time at 14 imaging locations at different points throughout the channel. **c)** Plot of the time to half maximum fluorescence intensity at 14 imaging locations (i.e. distances along the channel) provides the fluid velocity.

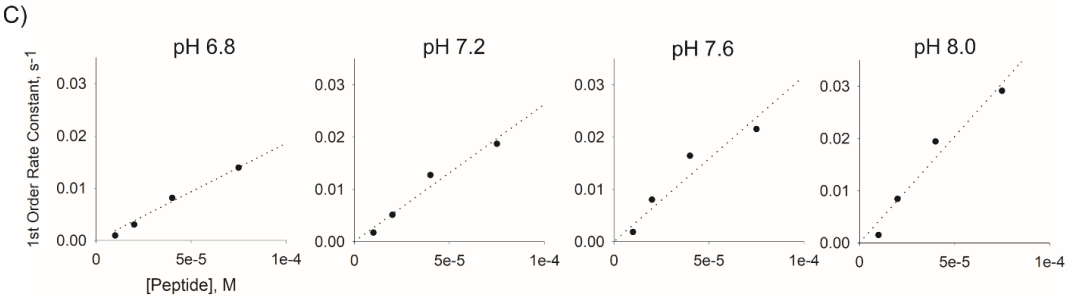
In addition to characterizing the dispersion front, we needed to determine the kinetics of surface immobilization to develop a complete understanding of the phenylglyoxal ligation reaction proceeding in the channel. We determined the kinetics of immobilization of the peptide to the maleimide functionalized self-assembled monolayer for each of the pH conditions used. Liquid handling robotics were used to dispense peptide solutions of varying concentrations across 384-spot arrays presenting monolayers identical to those used in the microfluidic device at 15 s intervals. For each pH condition, we generated kinetic profiles for the immobilization reaction over several peptide concentrations (**Figure 8.5a**). Each was fitted to a pseudo-first order rate law, as the solution-phase peptide is present in excess to the surface-bound maleimide. The pseudo-first order rate constants across these peptide concentrations were used to determine the second order rate constant of the maleimide immobilization reaction (**Figure 8.5b-c**).

We determined the rate constant for immobilization to be 186 M^{-1} s^{-1}, 263 M^{-1} s^{-1}, 317 M^{-1} s^{-1}, and 407 M^{-1} s^{-1} at the respective pH values of 6.8, 7.2, 7.6, and 8.0. This shows that the maleimide thiol capture chemistry is fast and that the rate increases with pH. The rapid rate of immobilization means that the self-assembled monolayer should sample from the dispersion front, where the concentration of the solution phase reactant and product are well below that of the bulk solution.





	Pseudo 1st Order Rate Constants, s^{-1}			
[Peptide], μM	pH 6.8	pH 7.2	pH 7.6	pH 8.0
10	0.0009 ± 0.0001	0.0017 ± 0.0001	0.0018 ± 0.0001	0.0015 ± 0.0001
20	0.0030 ± 0.0001	0.0051 ±0.0001	0.0080 ± 0.0003	0.0084 ± 0.0004
40	0.0081 ± 0.002	0.013 ± 0.001	0.016 ± 0.001	0.019 ± 0.001
75	0.014 ± 0.001	0.019 ± 0.001	0.022 ± 0.003	0.029 ± 0.003



2nd Order Rate Constant, $M^{-1}s^{-1}$			
pH 6.8	pH 7.2	pH 7.6	pH 8.0
186 ± 9	263 ± 17	317 ± 32	407 ± 29

Figure 8.5. Determination of kinetics of immobilization of analytes to the maleimide SAM. **a**) For each buffer condition, the immobilization of the peptide CAK(Me_3)SA to the surface was monitored by SAMDI-MS over time and at a range of concentrations, using the integrated area under the curve for the peaks corresponding to the maleimide SAM and ligated peptide product. **b**) This was used to determine the pseudo 1st order rate at each concentration of peptide, as the peptide is always present in large excess relative to the surface density of maleimide. Each reaction was plotted according to a standard 1st order kinetic plot, where the rate constant is given by the slope. **c**) For each pH, the pseudo 1st order rate constants were plotted relative to the concentration of peptide, giving a linear relationship where the slope is equal to the 2nd order rate constant for the reaction between the solution phase peptide and the maleimide surface. For all rate constants, the associated errors listed are the standard error in the slope of the regressed lines.

We also calculated the Damköhler number for this reaction. The Damköhler number is a dimensionless parameter defined by the ratio of the rate of immobilization to the rate of convective mass transfer (Equation 2).

$$Da = \frac{kC_{a,0}^{n-1}}{\frac{u}{L}} \quad (2)$$

Where k is the rate constant of immobilization, $C_{a,0}$ is the initial concentration of the reacting peptide species, n is the rate order of the immobilization reaction, L is the length of the channel, and u is the superficial velocity. We found the Damköhler number to be greater than 1 for all pH conditions described above, showing that the immobilization kinetics are not only fast, but that immobilization to the surface is rapid relative to the flow rate in the channel. This further suggests that it is necessary for us to develop a model of immobilization to account for the effects of dispersion.

*Incorporation of Dispersion with Immobilization Kinetics***.**

We used MATLAB to produce a model that simulates the concentration profile of the phenylglyoxal ligation reaction species in the dispersion front. The model also uses the rate constant for immobilization to build a description of the immobilization kinetics which incorporates the concentration gradient found in the dispersive regime. We first used the rate law describing the immobilization of the reaction species to the floor of the channel (Equation 3), where k is the second order rate constant for immobilization to the surface, Γ is the surface density of maleimide, and g(t) represents the total concentration of the reaction species. We modeled g(t) using the fluorescein imaging experiment presented in the previous section. The curves plotted in **Figure 8.4b** show normalized fluorescence intensity tracked at various positions along the channel

as fluorescein was injected at a concentration matching that of the peptide in the ligation reactions. The sigmoidal shape of the curves is due to the dispersion front and this front is elongated at positions further down the channel. Importantly, the fluorescence profiles demonstrate that the first region of interest used to quantitate the reaction progress occurs well past the chaotic mixing region of the channel inlet. To obtain g(t), a standard 4-parameter sigmoidal function (Equation 4) was fit to these fluorescence profiles. The monolayers used in this work present maleimide-terminated alkanethiolates at an initial density of 1.53×10^{-10} mol cm^{-2} (Γ).[31]

$$\frac{d\Gamma}{dt} = -kg(t)\Gamma \tag{3}$$

$$g(t) = \frac{a+b}{1+\left(\frac{t}{c}\right)^{d}} \tag{4}$$

Integration of Equation 3 gives the kinetic model that describes the immobilization of the reaction species to the channel floor in the dispersive region of flow. In **Figure 8.5**, we plot the time-dependent density of immobilized reaction species for each pH and at two fixed points in the channel (region 1 and region 14, representing positions near the beginning and end of the channel). The plots also show the solution phase concentration profile, g(t), for those same fixed positions in the channel.

This model shows that in all cases, the immobilization reaction is completed before the bulk concentration is achieved in the flow (that is, before the dispersion front has passed over the surface). When looking at a position near the end of the channel, broadening of the dispersion front leads to an even greater deviation between the concentration sampled by the surface and the concentration in the bulk flow. Additionally, the time required for complete immobilization increases as the dispersion front elongates. As the pH of the reaction increases, the rate constant

for immobilization also increases, leading to saturation of the surface at earlier points in the dispersion front (see Figure S3 in the appendix).

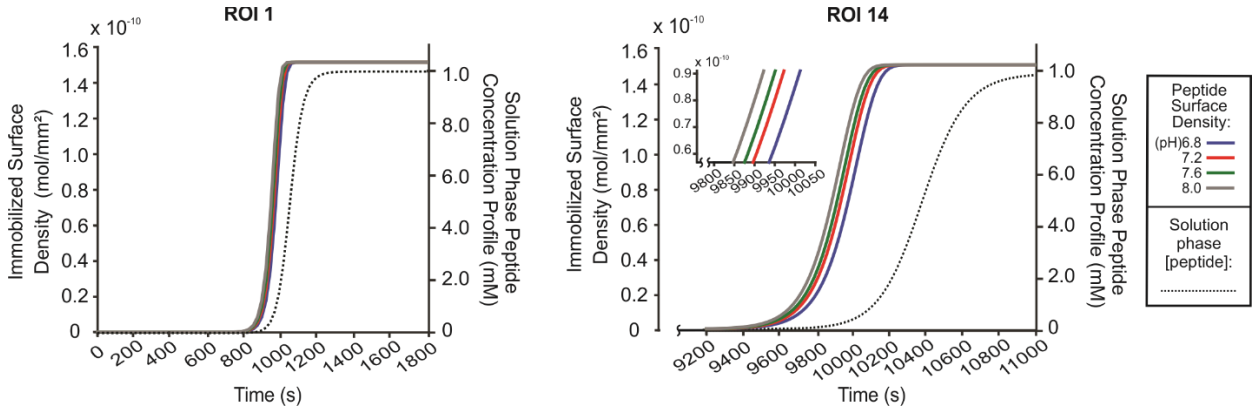


Figure 8.6. Modeling in-flow immobilization to the monolayer. The immobilization in the channel was modeled, incorporating both dispersion and the experimentally determined rate constants for the reaction between the maleimide surface and solution phase peptide. The plots here show pH dependent immobilization (colored lines, right y-axis) alongside the solution phase concentration and dispersion front (dashed line, left y-axis) for two fixed positions along the channel. ROI 1 is a position early in the channel and ROI 14 is a position at the end of the channel (see Figure S6 in the appendix). The inset highlights the effect of pH on the maleimide immobilization chemistry. Due to the rapid rate of immobilization, complete surface coverage is achieved before the dispersion front passes across the surface for all pH condition.

Determination of Phenylglyoxal Ligation Kinetics.

Finally, we used this model to analyze 14 positions in the channel for each pH condition, and to determine the time-weighted average concentration of reaction species during the immobilization at each position. To do so, we determined the time at which the reaction with the surface reached completion and then used this time as the upper limit for integration of the rate

law (Equation 3), giving the average concentration for the soluble reaction species at the surface (**Figure 8.7a**). In this analysis, time and position in the channel are equivalent representations of reaction progress.

Returning to the iSAMDI-MS heatmaps of the channel floors after the reaction, each pixel gives the fractional yield of the reaction at the corresponding position in the channel. To determine the second order rate constant for the reaction of phenylglyoxal and the peptide in the channel, we first determined the concentration of the product in the channel (at each pixel) by multiplying the fractional yield observed by mass spectrometry by the concentration determined by our model. This concentration represents the yield of the ligation reaction at any given point in the channel. To determine the second order reaction rate constant for the ligation reaction, we plotted the reciprocal of the concentration of the reactant peptide over time (**Figure 8.7b**). The slope of these plots gives the second order rate constant of the reactions, which we determined to be 0.38 ± 0.06 M^{-1} s^{-1}, 0.48 ± 0.04 M^{-1} s^{-1}, 0.62 ± 0.05 M^{-1} s^{-1}, and 0.62 ± 0.01 M^{-1} s^{-1}, at the respective pH values of 6.8, 7.2, 7.6, and 8.0. To assess the quality of our model, we performed similar reactions in bulk solution and analyzed the kinetics of those reactions using an identical SAMDI-MS strategy by pipetting a small amount of the reaction mixtures onto monolayer array to generate reaction timepoints (see Figure S4 in the appendix). We obtained rate constants from analysis of these bulk solution reactions which agree very well with those determined in the fluidic device.

These rates that we obtain are comparable to those for native chemical ligation reactions of thioesters and N-terminal cysteine residues.[153] Native chemical ligation also begins with nucleophilic attack by the thiol of a cysteine residue on a thioester and competes with an irreversible thiol-to-amine acyl transfer. We observe that our reaction is pH-dependent, where an

increase in pH from 6.8 to 7.6 yields an increase in rate but an increase in pH from 7.6 to 8.0 leads to no significant change. The pKa of an N-terminal amine and cysteine are 8.0 and 8.5, respectively, and explains our observed pH rate dependency.

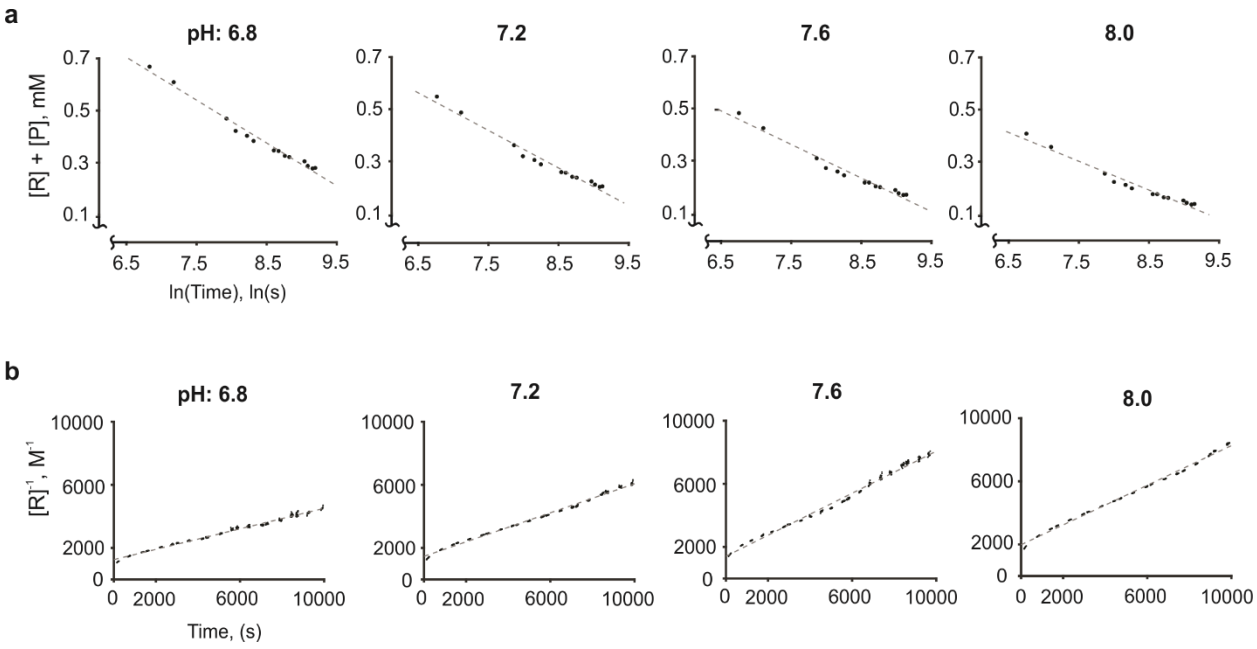


Figure 8.7. Determination of reaction kinetics from iSAMDI data a) Using our model of immobilization, the average concentration of peptide experienced by the saturated surface was plotted for all time positions along the channel. The model shows a logarithmic relationship between average concentration experienced by the surface and position in the channel. [R] and [P] represent the solution phase concentration of the reactant and product in mM, respectively. **b)** The concentration of peptide species at each position in the channel, given by the model, was incorporated with the ligation reaction yield at each position in the channel, determined experimentally by iSAMDI, to produce the second order rate plots for the ligation reaction between phenylglyoxal and the peptide as it occurred in the microfluidic device. [R] represents the solution phase concentration of reactant (mM).

8.3 Conclusion

iSAMDI-MS was used to study the progress of a chemical reaction at a level of detail that would otherwise be impractical to perform at the bench. This work combines the throughput of MALDI-IMS with the flexibility of microfluidics for obtaining high-density datasets. A MALDI mass spectrometer capable of imaging read the channel floor with 200 μm pixel resolution and generated a mass map of the reaction progress. We present an analytical solution for determining the reaction rate from the iSAMDI-MS dataset, which combines the kinetics of the immobilization reaction and the characteristics of the fluid front. A total of 15,720 spectra were analyzed with iSAMDI-MS to obtain the pH-dependent rate of the reaction between an N-terminal cysteine and phenylglyoxal.

Integrating the reaction and separation steps into one device eliminates both sample loss and the time associated with manual reagent handling. Here, the self-assembled monolayer sampled the reaction mixture from the bulk solution as the reaction mixture flowed down the fluidic channel. Microfluidic systems that use mass spectrometry have previously required pretreating the sample with matrix (for MALDI-MS) or directly interfacing the device with a mass spectrometer. This is time-consuming and can generate uncertainty in the kinetic profile if the reactions are not adequately quenched or are not sampled properly. In addition, it is difficult to integrate bulky equipment with microfluidic devices. Here, we simply rinse out the remaining reagent from the fluidic device, remove the chip entirely from the fluidic element, and acquire iSAMDI-MS directly from the chip.

Obtaining position-to-time information in microfluidic devices has traditionally been challenging due to dispersion. Segmented flow platforms have traditionally been used to

overcome dispersion by localizing reagents into droplets or plugs.[154-155] We used a self-assembled monolayer that allowed us to overcome this limitation by covalently immobilizing the reactant and product at known rates. The self-assembled monolayer maintained a covalent record of the reaction occurring in the microfluidic channel, allowing us to individually index the reactant concentration at each iSAMDI-MS measurement. The immobilization reaction happens very fast– on the order of M^{-1} s^{-1}– and we determined that the monolayer becomes saturated with reactant and product before experiencing the bulk solution concentration. Knowledge of the reactant concentration at each position along the channel enabled us to extract the second-order rate constant of the ligation reaction. This work is significant because it is the first example that uses a capture chemistry within a microfluidic channel to overcome dispersion. We recognize that the method presented in this work requires that the reactant and product contain a free thiol. Our laboratory has developed a variety of methods for capturing substrates to a self-assembled monolayer.[156-158] The mathematical modeling presented in this work can be straightforwardly adapted to different immobilization chemistries through changing the values of the immobilization rate.

An important advantage of iSAMDI-MS is the capability to conduct high-throughput experiments with a simple and scalable microfluidic device. We acquired approximately 1200 data points from each chip using 8 µL of reagent, where each data point only required 7 nL reagent. iSAMDI-MS can be easily extended to over 1 million data points using commercially available MALDI-MS instrumentation with 10 µm pixel resolution (rapifleX Tissuetyper, Bruker Daltonics). Reactions that consume picoliter volumes of reagent will be possible with devices that

contain nanometer-scale channels and reaction wells.[159-160] In addition, devices can be designed to contain multiple reaction channels for parallel reaction screening.

In summary, we developed a mass spectrometric approach that calculates the rate of a chemical reaction directly from the surface of a microfluidic channel. This approach allows for massively parallel experimentation using minimal reagent and does not require manual product separations or labels for product identification. The methodologies presented make significant contributions to microfluidics, chemical reaction monitoring, and MALDI Imaging Mass Spectrometry (MALDI-IMS). We envision that this work will have direct use in combinatorial chemistry and drug screening applications.

8.4 Methods

*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, Thermionics) 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 1-8 x 10^{-6} Torr. The slides were soaked overnight at 25 °C in an ethanolic solution (0.5 mm total disulfide concentration); the solution had a 1:4 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. The poly(dimethylsiloxane) (PDMS) master was rendered on SolidWorks software (Dassault Systémes SolidWorks Corporation) and had two 800 µm diameter inlets and one 800 µm diameter outlet (see Figure S5 in the appendix). The long, winding reaction channel was 250 µm tall and 550 µm wide. The 3D printed masters were prepared for PDMS polymerization as previously described.[151] A 0.8 mm biopsy punch was used to form the inlets and outlet of the PDMS block. The PDMS block was treated with 50 W air plasma for 35 s at 200 mTorr (Solarus Plasma Cleaner, Gatan, Inc.) and aligned with the Au slide functionalized with self-assembled monolayer. The PDMS and Au slide assembly was held together using light pressure from an external clamp made from extruded polycarbonate secured with 6 screws. The clamp had three 1.5 mm diameter holes that matched the location of the two inlets and outlet of the device.

Microfluidic device operation. Polytetrafluoroethylene (PTFE) tubing (0.042" outer diameter, Cole-Parmer) was inserted into the outlet via stainless steel catheter couplers (22 ga x 15 mm, Instech) and the device was primed with buffer using a syringe pump (PhD 2000, Harvard Apparatus). The outlet was disconnected from the syringe pump while keeping the tubing and catheter in the PDMS block. The priming process was repeated by flowing buffer through the two inlets. The device was placed inside a heat block at 37 ° C and the top of the heat block was covered with styrofoam and aluminum foil to fully enclose the device. Buffer was flowed through both device inlets at 0.1 uL min^{-1} for 1 h to allow the device to equilibrate. Simultaneously, a solution of 2 mM phenylglyoxal in buffer was injected into one inlet at 0.1 µL min^{-1} and a solution of 2 mM reactant peptide in buffer and calibrant peptide (0.5 mM) was injected into the other inlet

at 0.1 µL min^{-1}. After mixing, 1 mM of each reactant was present, as well as 0.25 mM of the calibrant, 100 mM BisTris, 10 mM Phosphate, 4 mM $MgCl_2$. Importantly, this buffer system contains no nucleophilic amines that may react with the phenylgloxal. After 4.5 h, buffer was flowed through both inlets at 100 µL min^{-1} to rinse the device. The PDMS block was removed from the chip and the chip was rinsed with ethanol, water, and acetone. This experiment was repeated across several pH conditions.

Mass spectrometry imaging. Matrix (25 mg mL^{-1} THAP in acetone) was applied by tilting the slide at a 45 ° angle, pipetting matrix (80 µL) on top of the slide and letting the matrix solution fall to the base of the slide where excess matrix was absorbed by a Kimwipe. Three ~ 5 mm x 9 mm ROIs were selected at the beginning, middle, and end of the device, covering all 7 serpentine channels, using FlexImaging software (Bruker Daltonics). MALDI Imaging Mass Spectrometry (MALDI-IMS) was acquired by creating an AutoXecute method as previously described.[161] Spectra were acquired with a 200 µm lateral resolution and with 125 laser shots accumulated per pixel. The laser pulse rate was 200 Hz and spectra were obtained with a mass window of 600-3000 *m/z*. The .mis file was loaded onto FlexImaging software and converted to imzML format. The FlexImaging output is shown in Supplementary Figure 2, which confirms that the device did not leak during operation, and that the quantity of reactant decreases and quantity of product increases along the channel length.

Image analysis and quantitation. The.imzML file was imported into MSiReader v1.00. The entire imaging region (including the three ROIs and all space in-between) was selected. Peak intensities of 1370.9 ± 0.2 Da, 1504.9 ± 0.2 Da, and 1384.8 ± 0.2 Da corresponding to the reactant-maleimide alkanedisulfide conjugate, product-maleimide alkanedisulfide conjugate, and peptide calibrant-maleimide alkanedisulfide conjugate, respectively, were exported into Excel. The fractional conversion of product was calculated using Equation 1, where I_P is the monoisotopic peak intensity of the product-maleimide alkanedisulfide conjugate and I_S is the monoisotopic peak intensity of the reactant-maleimide alkanedisulfide conjugate. 3-4 rows of pixels corresponded to the self-assembled monolayer in contact with the fluidic channel and contained immobilized analyte. The pixels in between the channels did not have bound analyte, and the fractional product conversion in these channels was set to 0. A ~218x56 pixel array of fractional hydroxy-amide conversion was plotted in a heatmap at each buffer pH (MATLAB R2016a, Mathworks).

iSAMDI-MS ROI distance determination. The slide was viewed under a stereozoom microscope (Nikon SMZ1500) after iSAMDI-MS acquisition. The distance between the first row of ablation spots of the top ROI and the Y-junction, and the distance between the last row of ablation spots of the bottom ROI and the outlet were measured using a NIS-Elements D 3.0 software (Nikon).

Ionization efficiency evaluation. The relative ionization efficiencies of the reactant and product were determined to validate that the fractional yields obtained from iSAMDI-MS accurately represented the ratios bound to the SAM. Solutions of pure reactant and product were individually

pipetted onto a chip functionalized with a self-assembled monolayer presenting maleimide at a density of 20% against a background of tri(ethylene glycol) groups. The solutions were incubated on the chip at room temperature for 1 h in a humidity chamber. The chip was rinsed with ethanol and water, and THAP matrix was applied as previously described. SAMDI-MS spectra were obtained from 125 accumulated shots with a fixed stage to simulate iSAMDI-MS acquisition conditions (Autoflex III, Bruker Daltonics). The peak intensities for the reactant-maleimide alkanedisulfide, product-maleimide alkanedisulfide, and tri(ethylene glycol) alkanedisulfide peaks were extracted with FlexAnalysis software (Bruker Daltonics). The ionization of the reactant peak and product peak relative to the tri(ethylene glycol) peak was calculated using Equation 5:

$$\% \; ionization = \frac{I_Z}{I_Z + I_{EG3}} \times 100 \tag{5}$$

Where I_Z represents the monoisotopic peak intensity of the product-maleimide alkanedisulfide conjugate or reactant-maleimide alkanedisulfide, and I_{EG3} represents the monoisotopic peak intensity of the tri(ethylene glycol)-alkanedisulfide conjugate. The % ionization for the reactant and product were 21.5 ± 0.59 and 26.1 ± 1.35, respectively, demonstrating that their ionization efficiencies are approximately equivalent and that the calculated fractional conversion from iSAMDI-MS accurately represented the ratio of reactant and product bound to the SAM.

Flow velocity determination. The microfluidic device was assembled as previously described and primed with 100 mM Bis-Tris, 10 mM phosphate buffer (pH 7.6). 1 mM solutions of fluorescein sodium salt (Sigma-Aldrich) were simultaneously flowed into both inlets at 0.1 μL min^{-1} for 4 h. Images of a 13 mm x 13 mm region were acquired every 30 s for 4 h using an inverted fluorescent

confocal Nikon Ti Eclipse microscope (Nikon) at 1 x magnification. The distance between the leftmost pixel in the image to the channel Y-junction was determined by moving the sample stage stepwise until the Y- junction appeared in the viewing region. The coordinates of both points were noted and used to calculate the distance of each ROI from the Y-junction. Three 200 µm x 200 µm ROIs were selected across the width of each channel on the left side and the right side of the image to yield 6 ROIs per channel (42 ROIs total) using NIS-Elements software (Nikon). The fluorescence intensity over time was extracted from each ROI and imported into Excel. The fluorescence intensity was normalized between 0 and 1. The three ROIs across each channel were averaged to yield a total of 14 normalized fluorescence intensity curves (two curves for each of the 7 channels taken from the left side and right side of the image). Sigmoidal fits were made using SigmaPlot 12.0 software (Systat Software) and the time to reach half maximal fluorescence intensity was recorded and plotted against distance from the Y-junction. Linear regression was used to calculate the velocity of the fluid. The time-dependent solution-phase concentration profile was then described by the function *g(t)*, a standard 4-parameter sigmoidal function (Equation 2). Where *a*, *b*, *c*, and *d* were fit by regression of the experimental fluorescein data and *t* represents time (MATLAB 2018 Curve Fitting Toolbox).

Dispersion coefficient determination. The stepwise function for the injected fluorescein concentration profile was modeled as a pulse by mirroring the fluid front at the maximum fluorescein concentration for each ROI in MATLAB 2018a (MathWorks). The average residence time for the fluid pulse at each spot was determined by the averaged integration of fluorescein concentration along the length of the channel. Dividing the length of the channel over the average

residence time yields the superficial velocity of the channel at that point, which we confirmed by comparative analysis to our experimentally determined flow velocity. The pulse shape was transformed into a dimensionless residence time distribution function and the curve maximum yielded the dispersion coefficient of the system.[162] The local maximum of this curve (E_{max}) has a defined relationship between the superficial velocity (u), length (L), and dispersion coefficient of the system (D) as shown in Equation 6:

$$E_{max} = \left(\frac{4\pi D}{uL}\right)^{-\frac{1}{2}} \quad (6)$$

Determination of Immobilization Kinetics. The rate of immobilization of the thiol-containing peptide to the maleimide presenting surface was determined experimentally using 384-spot monolayer arrays, presenting the same 20% maleimide surface used in the device. A Multi-Drop Combi liquid handling robot (Thermo Fisher) was used to dispense peptide solutions across the arrays, at varying concentrations and at each of the pH conditions used in the device. Reaction timepoints were generated every 15 to 20 seconds by automated deposition and the array was rinsed with distilled water at the end to stop all reactions. SAMDI-MS was performed using an AB ScieX 5800 series MALDI-TOF mass spectrometer and the yield of each reaction was quantified by integrating the area under the peak for the unreacted maleimide and the peptide coupled maleimide peaks. For each pH, the yield over time was fit to the pseudo 1st order rate law. The pseudo 1st order rate constants were then used across multiple peptide concentrations to determine the, pH dependent, second order rate constants for the immobilization of the peptide to the maleimide surface.[163]

Dimensionless parameters for modeling convective flow. We considered two distinct models to describe the fluid movement through the microfluidic systems; the axial dispersion model (ADM), where the convection of the fluid dominates, and the tank-in-series model (TISM), in which the dispersion of the fluid drives the mass transfer for the system.[164] Utilizing the dispersion coefficient for the microfluidic system, the axial Peclet number was determined as Pe = uL/D, where Pe is the Peclet number, u is the superficial velocity of the fluid, L is the channel length, and D is the dispersion coefficient.[165] Peclet numbers greater than 100 indicate that convective mass transfer acts as the primary driving force over axial diffusion.[166] To measure time scales between the reaction kinetics and the convective mass transfer, we leveraged the Damköhler number (Da) where k is the rate constant of immobilization, $C_{a,0}$ is the initial concentration of the reacting peptide species, n is the rate order of the immobilization reaction, L is the length of the channel, and u is the superficial velocity as utilized in Equation 7.

$$Da = \frac{kC_{a,0}^{n-1}}{\frac{u}{L}} \tag{7}$$

Chapter 9

Summary, Final Thoughts, and Future Directions

Since the inception of SAMDI as a general method for protein biochip assays, our group has worked to demonstrate the breadth of biological activities that can be characterized by this surface chemistry approach. Initial work focused on understanding the fundamentals of reactions of surfaces, an area that we continue to explore, and to show that a wide range of very different enzyme activities could be studied by SAMDI. We have shown the ability to look at the activities that include kinase and phosphatase, acetyl transferases and deacetylases, deaminases, hydrolases, proteases, glycosyl transferase, and others; all using the same fundamental strategy of monitoring biochemical reactions on surfaces with MALDI-TOF mass spectrometry. The work I have presented in this thesis showcases some of the ways in which we are now moving beyond the paradigm of simple biochemical assays, by focusing on the design and application of two very different assays that both take advantage of some of the unique benefits that controllable surface chemistry and characterization by mass spectrometry provide.

In the first part, I showed a strategy for detecting and profiling protein interaction domains using SAMDI. This assay takes advantage of the localization of substrate, a unique feature of surface reactions, to couple difficult to observe binding interactions to a much more easily quantified reporter enzyme. Here, I demonstrated how this assay could be readily applied to adaptor domains by profiling the activity of chromodomains, epigenetic readers of protein methylation sites. I characterized these domains on biologically relevant ligands and showed that these chromodomains appear to be sensitive to nearby PTMs other than lysine methylation, an idea

that is corroborated by literature suggesting *in vivo* sensitivity to nearby phosphorylation. This idea of "crosstalk" has become common for the enzymatic writers and erasers of the Histone Code, in some cases mediated by adaptor domains, but the adaptor domains themselves may also be experiencing a level of crosstalk via sensitivity to nearby PTMs. Additionally, I demonstrated the suitability of this assay for small molecule inhibitor screening of these protein interactions, a challenging task with available technologies.

In the second half of this thesis I outline the design of high-throughput assay strategy for profiling CoA-metabolites directly in lysates and complex reaction mixtures. This assay uses a chemical approach, taking advantage of biospecific chemistries to target a specific class of metabolites for capture, isolation, and detection. In this work I demonstrated how this could be a powerful tool for bioengineering pathways, using the throughput afforded by the assay to collect over 10,000 spectra for hundreds of unique reaction conditions and fully map the reaction space for the biosynthetic production of HMG-CoA. An exciting future direction of this technology is application to screening libraries of small molecule affecters of complex biosynthetic pathways. Additionally, this strategy of using biospecific chemistry to capture and immobilize a target class of metabolites from complex biological reaction mixtures may not be limited to the thioesters of acyl-CoA molecules. Other functional groups, such as aldehydes and ketones, are relatively uncommon in biology, but present in a host of interesting and desirable metabolites. Recent work in our group has shown it is possible to use chemistry with specificity for these groups, such as hydrazide presenting surfaces, to assay ketone-containing metabolites in complex lysates. This strategy shows promise for further generalizability beyond CoA metabolism.

Finally, I would like to finish by highlighting the importance that technology development has played in this work. While implementing new biochemical assays to tackle interesting problems, we as a group have also worked to develop the technologies required to scale SAMDI to ever more dense arrays with smaller reaction volumes, faster acquisition times, and greater compatibility with automation, empowering the design of ever more complex experiments. The first large-scale SAMDI arrays, using 384-spot surfaces and looking at the substrate specificity of deacetylases on libraries of substrates, was published less than 10 years ago. Since then, high-density arrays have become the norm, enabling rapid inhibitor screening and large-scale reaction profiling. In this thesis, I presented the first published use of 1536-spot arrays, which enabled the mapping of the reaction-space for the biosynthesis of HMG-CoA, as well as one of the first applications of imaging SAMDI. Imaging SAMDI in particular offers a way to move forward to ultra-high-density protein arrays as we approach the limitations of pipette-based liquid handling robotics. In the future, printing of surfaces coupled with imaging may offer a route to protein biochips with reaction densities that approach that of DNA biochips.

References

1. Fodor, S.; Read, J.; Pirrung, M.; Stryer, L.; Lu, A.; Solas, D., Light-directed, spatially addressable parallel chemical synthesis. *Science* **1991,** *251* (4995), 767-773.
2. Egeland, R. D.; Southern, E. M., Electrochemically directed synthesis of oligonucleotides for DNA microarray fabrication. *Nucleic Acids Research* **2005,** *33* (14), e125-e125.
3. Walt, D. R., Bead-based Fiber-Optic Arrays. *Science* **2000,** *287* (5452), 451-452.
4. Gunderson, K. L.; Kruglyak, S.; Graige, M. S.; Garcia, F.; Kermani, B. G.; Zhao, C.; Che, D.; Dickinson, T.; Wickham, E.; Bierle, J.; Doucet, D.; Milewski, M.; Yang, R.; Siegmund, C.; Haas, J.; Zhou, L.; Oliphant, A.; Fan, J.-B.; Barnard, S.; Chee, M. S., Decoding Randomly Ordered DNA Arrays. *Genome Research* **2004,** *14* (5), 870-877.
5. Lausted, C. G.; Warren, C. B.; Hood, L. E.; Lasky, S. R., [8] Printing Your Own Inkjet Microarrays. In *Methods in Enzymology*, Academic Press: 2006; Vol. 410, pp 168-189.
6. Bumgarner, R., DNA microarrays: Types, Applications and their future. *Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.]* **2013,** *0 22*, Unit-22.1.
7. Yates, J. R., Pivotal Role of Computers and Software in Mass Spectrometry – SEQUEST and 20 Years of Tandem MS Database Searching. *Journal of The American Society for Mass Spectrometry* **2015,** *26* (11), 1804-1813.
8. Zhang, Y.; Fonslow, B. R.; Shan, B.; Baek, M.-C.; Yates, J. R., 3rd, Protein analysis by shotgun/bottom-up proteomics. *Chemical reviews* **2013,** *113* (4), 2343-2394.
9. Skinner, O. S.; Haverland, N. A.; Fornelli, L.; Melani, R. D.; Do Vale, L. H. F.; Seckler, H. S.; Doubleday, P. F.; Schachner, L. F.; Srzentić, K.; Kelleher, N. L.; Compton, P. D., Top-down characterization of endogenous protein complexes with native proteomics. *Nature Chemical Biology* **2017,** *14*, 36.
10. Cravatt, B. F.; Sorensen, E. J., Chemical strategies for the global analysis of protein function. *Current Opinion in Chemical Biology* **2000,** *4* (6), 663-668.
11. Speers, A. E.; Adam, G. C.; Cravatt, B. F., Activity-Based Protein Profiling in Vivo Using a Copper(I)-Catalyzed Azide-Alkyne [3 + 2] Cycloaddition. *Journal of the American Chemical Society* **2003,** *125* (16), 4686-4687.
12. Barglow, K. T.; Cravatt, B. F., Activity-based protein profiling for the functional annotation of enzymes. *Nature Methods* **2007,** *4*, 822.
13. Mahmood, T.; Yang, P.-C., Western Blot: Technique, Theory, and Trouble Shooting. *North American Journal of Medical Sciences* **2012,** *4* (9), 429-434.
14. Engvall, E., [28] Enzyme immunoassay ELISA and EMIT. In *Methods in Enzymology*, Academic Press: 1980; Vol. 70, pp 419-439.
15. Bessey, O. A.; Lowry, O. H.; Brock, M. J., A method for the rapid determination of alkaline phosphates with five cubic millimeters of serum. *J Biol Chem* **1946,** *164*, 321-9.

16. Wegener, D.; Wirsching, F.; Riester, D.; Schwienhorst, A., A Fluorogenic Histone Deacetylase Assay Well Suited for High-Throughput Activity Screening. *Chemistry & Biology* **2003,** *10* (1), 61-68.

17. Zauner, T.; Berger-Hoffmann, R.; Müller, K.; Hoffmann, R.; Zuchner, T., Highly Adaptable and Sensitive Protease Assay Based on Fluorescence Resonance Energy Transfer. *Analytical chemistry* **2011,** *83* (19), 7356-7363.

18. Borra, M. T.; Smith, B. C.; Denu, J. M., Mechanism of Human SIRT1 Activation by Resveratrol. *Journal of Biological Chemistry* **2005,** *280* (17), 17187-17195.

19. Bensebaa, F.; Ellis, T. H.; Badia, A.; Lennox, R. B., Thermal Treatment of n-Alkanethiolate Monolayers on Gold, As Observed by Infrared Spectroscopy. *Langmuir* **1998,** *14* (9), 2361-2367.

20. Prime, K.; Whitesides, G., Self-assembled organic monolayers: model systems for studying adsorption of proteins at surfaces. *Science* **1991,** *252* (5009), 1164-1167.

21. Gurard-Levin, Z. A.; Mrksich, M., Combining Self-Assembled Monolayers and Mass Spectrometry for Applications in Biochips. *Annual Review of Analytical Chemistry* **2008,** *1* (1), 767-800.

22. Marin, V. L.; Bayburt, T. H.; Sligar, S. G.; Mrksich, M., Functional assays of membrane-bound proteins with SAMDI-TOF mass spectrometry. *Angewandte Chemie* **2007,** *46* (46), 8796-8.

23. Houseman, B. T.; Huh, J. H.; Kron, S. J.; Mrksich, M., Peptide chips for the quantitative evaluation of protein kinase activity. *Nature Biotechnology* **2002,** *20*, 270.

24. 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.

25. Saito, F.; Noda, H.; Bode, J. W., Critical Evaluation and Rate Constants of Chemoselective Ligation Reactions for Stoichiometric Conjugations in Water. *ACS chemical biology* **2015,** *10* (4), 1026-1033.

26. 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. *Nature Chemistry* **2013,** *5*, 423.

27. Webster, J.; Oxley, D., Protein Identification by MALDI-TOF Mass Spectrometry. In *Chemical Genomics and Proteomics: Reviews and Protocols*, Zanders, E. D., Ed. Humana Press: Totowa, NJ, 2012; pp 227-240.

28. Su, J.; Mrksich, M., Using mass spectrometry to characterize self-assembled monolayers presenting peptides, proteins, and carbohydrates. *Angewandte Chemie* **2002,** *41* (24), 4715-8.

29. 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. *Nature Chemical Biology* **2012,** *8*, 769.

30. Gurard-Levin, Z. A.; Mrksich, M., The activity of HDAC8 depends on local and distal sequences of its peptide substrates. *Biochemistry* **2008,** *47* (23), 6242-50.

31. Liao, X.; Su, J.; Mrksich, M., An adaptor domain-mediated autocatalytic interfacial kinase reaction. *Chemistry* **2009,** *15* (45), 12303-9.

32. E., W. S.; Gaurav, S.; Sushanth, G.; Madhavan, N.; Che-Fan, H.; Bo, L.; Milan, M., A Bottom-Up Proteomic Approach to Identify Substrate Specificity of Outer-Membrane Protease OmpT. *Angewandte Chemie International Edition* **2017,** *56* (52), 16531-16535.

33. 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 chemical biology* **2015,** *10* (1), 157-164.

34. Kuo, H.-Y.; DeLuca, T. A.; Miller, W. M.; Mrksich, M., Profiling Deacetylase Activities in Cell Lysates with Peptide Arrays and SAMDI Mass Spectrometry. *Analytical Chemistry* **2013,** *85* (22), 10635-10642.

35. J., B. E.; D., C. M.; Milan, M., Cellular Assays with a Molecular Endpoint Measured by SAMDI Mass Spectrometry. *Small* **2016,** *12* (28), 3811-3818.

36. Westermann, S.; Weber, K., Post-translational modifications regulate microtubule function. *Nature Reviews Molecular Cell Biology* **2003,** *4*, 938.

37. Bode, A. M.; Dong, Z., Post-translational modification of p53 in tumorigenesis. *Nature Reviews Cancer* **2004,** *4*, 793.

38. Ahearn, I. M.; Haigis, K.; Bar-Sagi, D.; Philips, M. R., Regulating the regulator: post-translational modification of RAS. *Nature Reviews Molecular Cell Biology* **2011,** *13*, 39.

39. Kornberg, R. D., Chromatin Structure: A Repeating Unit of Histones and DNA. *Science* **1974,** *184* (4139), 868-871.

40. McGinty, R. K.; Tan, S., Nucleosome structure and function. *Chemical reviews* **2015,** *115* (6), 2255-2273.

41. Grunstein, M., Histone acetylation in chromatin structure and transcription. *Nature* **1997,** *389*, 349.

42. Rea, S.; Eisenhaber, F.; O'Carroll, D.; Strahl, B. D.; Sun, Z.-W.; Schmid, M.; Opravil, S.; Mechtler, K.; Ponting, C. P.; Allis, C. D.; Jenuwein, T., Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **2000,** *406*, 593.

43. Jenuwein, T.; Allis, C. D., Translating the Histone Code. *Science* **2001,** *293* (5532), 1074-1080.

44. Strahl, B. D.; Allis, C. D., The language of covalent histone modifications. *Nature* **2000,** *403*, 41.

45. 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.

46. Weake, V. M.; Workman, J. L., Histone Ubiquitination: Triggering Gene Activity. *Molecular Cell* **2008,** *29* (6), 653-663.

47. Musselman, C. A.; Lalonde, M.-E.; Côté, J.; Kutateladze, T. G., Perceiving the epigenetic landscape through histone readers. *Nature Structural &Amp; Molecular Biology* **2012,** *19*, 1218.

48. Yun, M.; Wu, J.; Workman, J. L.; Li, B., Readers of histone modifications. *Cell Research* **2011,** *21* (4), 564-578.

49. Fujisawa, T.; Filippakopoulos, P., Functions of bromodomain-containing proteins and their roles in homeostasis and cancer. *Nature Reviews Molecular Cell Biology* **2017,** *18*, 246.

50. Filippakopoulos, P.; Müller, S.; Knapp, S., SH2 domains: modulators of nonreceptor tyrosine kinase activity. *Current opinion in structural biology* **2009,** *19* (6), 643-649.

51. Sterner, D. E.; Berger, S. L., Acetylation of histones and transcription-related factors. *Microbiology and molecular biology reviews : MMBR* **2000,** *64* (2), 435-459.

52. Paro, R.; Hogness, D. S., The Polycomb protein shares a homologous domain with a heterochromatin-associated protein of Drosophila. *Proceedings of the National Academy of Sciences* **1991,** *88* (1), 263-267.

53. Eissenberg, J. C., Structural biology of the chromodomain: Form and function. *Gene* **2012,** *496* (2), 69-78.

54. Cowieson, N. P.; Partridge, J. F.; Allshire, R. C.; McLaughlin, P. J., Dimerisation of a chromo shadow domain and distinctions from the chromodomain as revealed by structural analysis. *Current Biology* **2000,** *10* (9), 517-525.

55. Hughes, R. M.; Wiggins, K. R.; Khorasanizadeh, S.; Waters, M. L., Recognition of trimethyllysine by a chromodomain is not driven by the hydrophobic effect. *Proceedings of the National Academy of Sciences* **2007,** *104* (27), 11184-11188.

56. Bannister, A. J.; Zegerman, P.; Partridge, J. F.; Miska, E. A.; Thomas, J. O.; Allshire, R. C.; Kouzarides, T., Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* **2001,** *410*, 120.

57. Lee, T. I.; Jenner, R. G.; Boyer, L. A.; Guenther, M. G.; Levine, S. S.; Kumar, R. M.; Chevalier, B.; Johnstone, S. E.; Cole, M. F.; Isono, K.-i.; Koseki, H.; Fuchikami, T.; Abe, K.; Murray, H. L.; Zucker, J. P.; Yuan, B.; Bell, G. W.; Herbolsheimer, E.; Hannett, N. M.; Sun, K.; Odom, D. T.; Otte, A. P.; Volkert, T. L.; Bartel, D. P.; Melton, D. A.; Gifford, D. K.; Jaenisch, R.; Young, R. A., Control of Developmental Regulators by Polycomb in Human Embryonic Stem Cells. *Cell* **2006,** *125* (2), 301-313.

58. Eissenberg, J. C.; Elgin, S. C. R., The HP1 protein family: getting a grip on chromatin. *Current Opinion in Genetics & Development* **2000,** *10* (2), 204-210.

59. Dinant, C.; Luijsterburg, M. S., The Emerging Role of HP1 in the DNA Damage Response. *Molecular and Cellular Biology* **2009,** *29* (24), 6335-6340.

60. Freyer, M. W.; Lewis, E. A., Isothermal Titration Calorimetry: Experimental Design, Data Analysis, and Probing Macromolecule/Ligand Binding and Kinetic Interactions. In *Methods in Cell Biology*, Academic Press: 2008; Vol. 84, pp 79-113.

61. Rossi, A. M.; Taylor, C. W., Analysis of protein-ligand interactions by fluorescence polarization. *Nature protocols* **2011,** *6* (3), 365-387.

62. Thorne, N.; Auld, D. S.; Inglese, J., Apparent Activity in High-Throughput Screening: Origins of Compound-Dependent Assay Interference. *Current opinion in chemical biology* **2010,** *14* (3), 315-324.

63. Itoh, Y.; Suzuki, M.; Matsui, T.; Ota, Y.; Hui, Z.; Tsubaki, K.; Suzuki, T., False HDAC Inhibition by Aurone Compound. *Chemical and Pharmaceutical Bulletin* **2016,** *64* (8), 1124-1128.

64. Frank, R., Spot-synthesis: an easy technique for the positionally addressable, parallel chemical synthesis on a membrane support. *Tetrahedron* **1992,** *48* (42), 9217-9232.

65. Kaustov, L.; Ouyang, H.; Amaya, M.; Lemak, A.; Nady, N.; Duan, S.; Wasney, G. A.; Li, Z.; Vedadi, M.; Schapira, M.; Min, J.; Arrowsmith, C. H., Recognition and Specificity Determinants of the Human Cbx Chromodomains. *J Biol Chem* **2011,** *286* (1), 521-529.

66. Li, J.; Nayak, S.; Mrksich, M., Rate enhancement of an interfacial biochemical reaction through localization of substrate and enzyme by an adaptor domain. *The journal of physical chemistry. B* **2010,** *114* (46), 15113-8.

67. Gurard-Levin, Z. A.; Kim, J.; Mrksich, M., Combining mass spectrometry and peptide arrays to profile the specificities of histone deacetylases. *Chembiochem* **2009,** *10* (13), 2159-61.

68. Gurard-Levin, Z. A.; Kilian, K. A.; Kim, J.; Bahr, K.; Mrksich, M., Peptide arrays identify isoform-selective substrates for profiling endogenous lysine deacetylase activity. *ACS chemical biology* **2010,** *5* (9), 863-73.

69. Gurard-Levin, Z. A.; Scholle, M. D.; Eisenberg, A. H.; Mrksich, M., High-throughput screening of small molecule libraries using SAMDI mass spectrometry. *ACS combinatorial science* **2011,** *13* (4), 347-50.

70. Bernstein, E.; Duncan, E. M.; Masui, O.; Gil, J.; Heard, E.; Allis, C. D., Mouse Polycomb Proteins Bind Differentially to Methylated Histone H3 and RNA and Are Enriched in Facultative Heterochromatin. *Molecular and Cellular Biology* **2006,** *26* (7), 2560-2569.

71. Suganuma, T.; Workman, J. L., Crosstalk among Histone Modifications. *Cell* **2008,** *135* (4), 604-607.

72. Wang, Y.; Wysocka, J.; Sayegh, J.; Lee, Y.-H.; Perlin, J. R.; Leonelli, L.; Sonbuchner, L. S.; McDonald, C. H.; Cook, R. G.; Dou, Y.; Roeder, R. G.; Clarke, S.; Stallcup, M. R.; Allis, C. D.; Coonrod, S. A., Human PAD4 Regulates Histone Arginine Methylation Levels via Demethylimination. *Science* **2004,** *306* (5694), 279-283.

73. Metzger, E.; Imhof, A.; Patel, D.; Kahl, P.; Hoffmeyer, K.; Friedrichs, N.; Muller, J. M.; Greschik, H.; Kirfel, J.; Ji, S.; Kunowska, N.; Beisenherz-Huss, C.; Gunther, T.; Buettner,

R.; Schule, R., Phosphorylation of histone H3T6 by PKC[bgr]I controls demethylation at histone H3K4. *Nature* **2010,** *464* (7289), 792-796.

74. Goto, H.; Tomono, Y.; Ajiro, K.; Kosako, H.; Fujita, M.; Sakurai, M.; Okawa, K.; Iwamatsu, A.; Okigaki, T.; Takahashi, T.; Inagaki, M., Identification of a Novel Phosphorylation Site on Histone H3 Coupled with Mitotic Chromosome Condensation. *Journal of Biological Chemistry* **1999,** *274* (36), 25543-25549.

75. Hake, S. B.; Garcia, B. A.; Duncan, E. M.; Kauer, M.; Dellaire, G.; Shabanowitz, J.; Bazett-Jones, D. P.; Allis, C. D.; Hunt, D. F., Expression Patterns and Post-translational Modifications Associated with Mammalian Histone H3 Variants. *Journal of Biological Chemistry* **2006,** *281* (1), 559-568.

76. Garcia, B. A.; Hake, S. B.; Diaz, R. L.; Kauer, M.; Morris, S. A.; Recht, J.; Shabanowitz, J.; Mishra, N.; Strahl, B. D.; Allis, C. D.; Hunt, D. F., Organismal Differences in Post-translational Modifications in Histones H3 and H4. *Journal of Biological Chemistry* **2007,** *282* (10), 7641-7655.

77. Thomas, C. E.; Kelleher, N. L.; Mizzen, C. A., Mass Spectrometric Characterization of Human Histone H3: A Bird's Eye View. *Journal of Proteome Research* **2006,** *5* (2), 240-247.

78. Hake, S. B.; Garcia, B. A.; Kauer, M.; Baker, S. P.; Shabanowitz, J.; Hunt, D. F.; Allis, C. D., Serine 31 phosphorylation of histone variant H3.3 is specific to regions bordering centromeres in metaphase chromosomes. *Proceedings of the National Academy of Sciences of the United States of America* **2005,** *102* (18), 6344-6349.

79. Baur, J. A.; Sinclair, D. A., Therapeutic potential of resveratrol: the in vivo evidence. *Nature Reviews Drug Discovery* **2006,** *5* (6), 493-506.

80. Howitz, K. T.; Bitterman, K. J.; Cohen, H. Y.; Lamming, D. W.; Lavu, S.; Wood, J. G.; Zipkin, R. E.; Chung, P.; Kisielewski, A.; Zhang, L.-L.; Scherer, B.; Sinclair, D. A., Small molecule activators of sirtuins extend Saccharomyces cerevisiae lifespan. *Nature* **2003,** *425* (6954), 191-196.

81. Pacholec, M.; Bleasdale, J. E.; Chrunyk, B.; Cunningham, D.; Flynn, D.; Garofalo, R. S.; Griffith, D.; Griffor, M.; Loulakis, P.; Pabst, B.; Qiu, X.; Stockman, B.; Thanabal, V.; Varghese, A.; Ward, J.; Withka, J.; Ahn, K., SRT1720, SRT2183, SRT1460, and Resveratrol Are Not Direct Activators of SIRT1. *Journal of Biological Chemistry* **2010,** *285* (11), 8340-8351.

82. Patel, K.; Sherrill, J.; Mrksich, M.; Scholle, M. D., Discovery of SIRT3 Inhibitors Using SAMDI Mass Spectrometry. *Journal of Biomolecular Screening* **2015,** *20* (7), 842-848.

83. Jacobs, S. A.; Khorasanizadeh, S., Structure of HP1 Chromodomain Bound to a Lysine 9-Methylated Histone H3 Tail. *Science* **2002,** *295* (5562), 2080-2083.

84. Hirota, T.; Lipp, J. J.; Toh, B. H.; Peters, J. M., Histone H3 serine 10 phosphorylation by Aurora B causes HP1 dissociation from heterochromatin. *Nature* **2005,** *438* (7071), 1176-80.

85. Dormann, H. L.; Tseng, B. S.; Allis, C. D.; Funabiki, H.; Fischle, W., Dynamic regulation of effector protein binding to histone modifications: the biology of HP1 switching. *Cell cycle* **2006,** *5* (24), 2842-51.

86. Lau, P. N.; Cheung, P., Histone code pathway involving H3 S28 phosphorylation and K27 acetylation activates transcription and antagonizes polycomb silencing. *P. Natl. A. Sci. U.S.A.* **2011,** *108* (7), 2801-6.

87. Andrews, F. H.; Gatchalian, J.; Krajewski, K.; Strahl, B. D.; Kutateladze, T. G., Regulation of Methyllysine Readers through Phosphorylation. *ACS chemical biology* **2016,** *11* (3), 547-553.

88. Lagerholm, B. C.; Thompson, N. L., Temporal Dependence of Ligand Dissociation and Rebinding at Planar Surfaces. *J. Phys. Chem. B* **2000,** *104* (4), 863-868.

89. Lieto, A. M.; Cush, R. C.; Thompson, N. L., Ligand-Receptor Kinetics Measured by Total Internal Reflection with Fluorescence Correlation Spectroscopy. *Biophys. J.* **2003,** *85* (5), 3294-3302.

90. Houseman, B. T.; Mrksich, M., Towards quantitative assays with peptide chips: a surface engineering approach. *Trends in biotechnology* **2002,** *20* (7), 279-81.

91. Marks, P. A.; Breslow, R., Dimethyl sulfoxide to vorinostat: development of this histone deacetylase inhibitor as an anticancer drug. *Nature Biotechnology* **2007,** *25*, 84.

92. Zhang, J.; Yang, P. L.; Gray, N. S., Targeting cancer with small molecule kinase inhibitors. *Nature Reviews Cancer* **2009,** *9*, 28.

93. Wu, P.; Nielsen, T. E.; Clausen, M. H., FDA-approved small-molecule kinase inhibitors. *Trends in Pharmacological Sciences* **2015,** *36* (7), 422-439.

94. Gross, S.; Rahal, R.; Stransky, N.; Lengauer, C.; Hoeflich, K. P., Targeting cancer with kinase inhibitors. *The Journal of Clinical Investigation* **2015,** *125* (5), 1780-1789.

95. Chung, C.-w.; Coste, H.; White, J. H.; Mirguet, O.; Wilde, J.; Gosmini, R. L.; Delves, C.; Magny, S. M.; Woodward, R.; Hughes, S. A.; Boursier, E. V.; Flynn, H.; Bouillot, A. M.; Bamborough, P.; Brusq, J.-M. G.; Gellibert, F. J.; Jones, E. J.; Riou, A. M.; Homes, P.; Martin, S. L.; Uings, I. J.; Toum, J.; Clément, C. A.; Boullay, A.-B.; Grimley, R. L.; Blandel, F. M.; Prinjha, R. K.; Lee, K.; Kirilovsky, J.; Nicodeme, E., Discovery and Characterization of Small Molecule Inhibitors of the BET Family Bromodomains. *J Med Chem* **2011,** *54* (11), 3827-3838.

96. Filippakopoulos, P.; Qi, J.; Picaud, S.; Shen, Y.; Smith, W. B.; Fedorov, O.; Morse, E. M.; Keates, T.; Hickman, T. T.; Felletar, I.; Philpott, M.; Munro, S.; McKeown, M. R.; Wang, Y.; Christie, A. L.; West, N.; Cameron, M. J.; Schwartz, B.; Heightman, T. D.; La Thangue, N.; French, C. A.; Wiest, O.; Kung, A. L.; Knapp, S.; Bradner, J. E., Selective inhibition of BET bromodomains. *Nature* **2010,** *468* (7327), 1067-73.

97. Morera, L.; Lübbert, M.; Jung, M., Targeting histone methyltransferases and demethylases in clinical trials for cancer therapy. *Clinical epigenetics* **2016,** *8*, 57-57.

98. Kraskouskaya, D.; Duodu, E.; Arpin, C. C.; Gunning, P. T., Progress towards the development of SH2 domain inhibitors. *Chem. Soc. Rev.* **2013,** *42* (8), 3337-3370.

99. Schaffhausen, B., SH2 domain structure and function. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer* **1995,** *1242* (1), 61-75.

100. Lorenz, U., SHP-1 and SHP-2 in T cells: two phosphatases functioning at many levels. *Immunological Reviews* **2009,** *228* (1), 342-359.

101. Campbell, M. A.; Klinman Norman, R., Phosphotyrosine-dependent association between CD22 and protein tyrosine phosphatase 1C. *European Journal of Immunology* **1995,** *25* (6), 1573-1579.

102. Sweeney, M. C.; Wavreille, A.-S.; Park, J.; Butchar, J. P.; Tridandapani, S.; Pei, D., Decoding Protein−Protein Interactions through Combinatorial Chemistry: Sequence Specificity of SHP-1, SHP-2, and SHIP SH2 Domains. *Biochemistry* **2005,** *44* (45), 14932-14947.

103. Cochrane, D.; Webster, C.; Masih, G.; McCafferty, J., Identification of natural ligands for SH2 domains from a phage display cDNA library11Edited by J. A. Wells. *Journal of Molecular Biology* **2000,** *297* (1), 89-97.

104. Beebe, K. D.; Wang, P.; Arabaci, G.; Pei, D., Determination of the Binding Specificity of the SH2 Domains of Protein Tyrosine Phosphatase SHP-1 through the Screening of a Combinatorial Phosphotyrosyl Peptide Library. *Biochemistry* **2000,** *39* (43), 13251-13260.

105. Paddon, C.; Westfall, P.; Pitera, D.; Benjamin, K.; Fisher, K.; McPhee, D.; Leavell, M.; Tai, A.; Main, A.; Eng, D., High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature* **2013,** *496* (7446), 528-532.

106. Ajikumar, P. K.; Xiao, W.-H.; Tyo, K. E. J.; Wang, Y.; Simeon, F.; Leonard, E.; Mucha, O.; Phon, T. H.; Pfeifer, B.; Stephanopoulos, G., Isoprenoid Pathway Optimization for Taxol Precursor Overproduction in Escherichia coli. *Science* **2010,** *330* (6000), 70-74.

107. George, K. W.; Thompson, M. G.; Kang, A.; Baidoo, E.; Wang, G.; Chan, L. J. G.; Adams, P. D.; Petzold, C. J.; Keasling, J. D.; Soon Lee, T., Metabolic engineering for the high-yield production of isoprenoid-based C5 alcohols in E. coli. *Scientific Reports* **2015,** *5*, 11128.

108. Mewalal, R.; Rai, D. K.; Kainer, D.; Chen, F.; Külheim, C.; Peter, G. F.; Tuskan, G. A., Plant-Derived Terpenes: A Feedstock for Specialty Biofuels. *Trends in biotechnology* **2017,** *35* (3), 227-240.

109. Runguphan, W.; Keasling, J. D., Metabolic engineering of Saccharomyces cerevisiae for production of fatty acid-derived biofuels and chemicals. *Metabolic Engineering* **2014,** *21*, 103-113.

110. Denby, C. M.; Li, R. A.; Vu, V. T.; Costello, Z.; Lin, W.; Chan, L. J. G.; Williams, J.; Donaldson, B.; Bamforth, C. W.; Petzold, C. J.; Scheller, H. V.; Martin, H. G.; Keasling, J. D., Industrial brewing yeast engineered for the production of primary flavor determinants in hopped beer. *Nature Communications* **2018,** *9* (1), 965.

111. 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. *Analytical Chemistry* **2016,** *88* (17), 8604-8609.

112. Leonardi, R.; Zhang, Y.-M.; Rock, C. O.; Jackowski, S., Coenzyme A: Back in action. *Progress in Lipid Research* **2005,** *44* (2), 125-153.

113. Krivoruchko, A.; Zhang, Y.; Siewers, V.; Chen, Y.; Nielsen, J., Microbial acetyl-CoA metabolism and metabolic engineering. *Metabolic engineering* **2015,** *28*, 28-42.

114. Nielsen, J.; Keasling, J. D., Engineering Cellular Metabolism. *Cell* **2016,** *164* (6), 1185-1197.

115. Lian, J.; Si, T.; Nair, N. U.; Zhao, H., Design and construction of acetyl-CoA overproducing Saccharomyces cerevisiae strains. *Metabolic Engineering* **2014,** *24*, 139-149.

116. Krink-Koutsoubelis, N.; Loechner, A. C.; Lechner, A.; Link, H.; Denby, C. M.; Vögeli, B.; Erb, T. J.; Yuzawa, S.; Jakociunas, T.; Katz, L.; Jensen, M. K.; Sourjik, V.; Keasling, J. D., Engineered Production of Short-Chain Acyl-Coenzyme A Esters in Saccharomyces cerevisiae. *ACS Synthetic Biology* **2018**.

117. Magnes, C.; Suppan, M.; Pieber, T. R.; Moustafa, T.; Trauner, M.; Sinner, F. M., Validated comprehensive analytical method for quantification of coenzyme A activated compounds in biological tissues by online solid-phase extraction LC/MS/MS. *Analytical chemistry* **2008,** *80* (15), 5736-5742.

118. Lu, W.; Clasquin, M. F.; Melamud, E.; Amador-Noguez, D.; Caudy, A. A.; Rabinowitz, J. D., Metabolomic analysis via reversed-phase ion-pairing liquid chromatography coupled to a stand alone orbitrap mass spectrometer. *Analytical chemistry* **2010,** *82* (8), 3212.

119. Basu, S. S.; Mesaros, C.; Gelhaus, S. L.; Blair, I. A., Stable isotope labeling by essential nutrients in cell culture for preparation of labeled coenzyme A and its thioesters. *Analytical chemistry* **2011,** *83* (4), 1363.

120. Zimmermann, M.; Thormann, V.; Sauer, U.; Zamboni, N., Nontargeted profiling of coenzyme A thioesters in biological samples by tandem mass spectrometry. *Analytical chemistry* **2013,** *85* (17), 8284-8290.

121. Li, Q.; Zhang, S.; Berthiaume, J. M.; Simons, B.; Zhang, G.-F., Novel approach in LC-MS/MS using MRM to generate a full profile of acyl-CoAs: discovery of acyl-dephospho-CoAs. *Journal of lipid research* **2014,** *55* (3), 592-602.

122. Yang, S.; Sadilek, M.; Synovec, R. E.; Lidstrom, M. E., Liquid chromatography–tandem quadrupole mass spectrometry and comprehensive two-dimensional gas chromatography–time-of-flight mass spectrometry measurement of targeted metabolites of Methylobacterium extorquens AM1 grown on two different carbon sources. *Journal of Chromatography A* **2009,** *1216* (15), 3280-3289.

123. Liu, X.; Sadhukhan, S.; Sun, S.; Wagner, G. R.; Hirschey, M. D.; Qi, L.; Lin, H.; Locasale, J. W., High-resolution metabolomics with Acyl-CoA profiling reveals widespread remodeling in response to diet. *Molecular & Cellular Proteomics* **2015,** *14* (6), 1489-1500.

124. Zhang, F.; Carothers, J. M.; Keasling, J. D., Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids. *Nature Biotechnology* **2012,** *30* (4), 354-359.

125. Dudley, Q. M.; Karim, A. S.; Jewett, M. C., Cell-free metabolic engineering: Biomanufacturing beyond the cell. *Biotechnology Journal* **2015,** *10* (1), 69-82.

126. Hodgman, C. E.; Jewett, M. C., Cell-free synthetic biology: Thinking outside the cell. *Metabolic Engineering* **2012,** *14* (3), 261-269.

127. Karim, A. S.; Jewett, M. C., A cell-free framework for rapid biosynthetic pathway prototyping and enzyme discovery. *Metabolic Engineering* **2016,** *36*, 116-126.

128. Dawson, P.; Muir, T.; Clark-Lewis, I.; Kent, S., Synthesis of proteins by native chemical ligation. *Science* **1994,** *266* (5186), 776-779.

129. Yeo, D. S. Y.; Srinivasan, R.; Uttamchandani, M.; Chen, G. Y. J.; Zhu, Q.; Yao, S. Q., Cell-permeable small molecule probes for site-specific labeling of proteins. *Chemical Communications* **2003,** (23), 2870-2871.

130. Sletten, E. M.; Bertozzi, C. R., Bioorthogonal Chemistry: Fishing for Selectivity in a Sea of Functionality. *Angewandte Chemie International Edition* **2009,** *48* (38), 6974-6998.

131. Bohlmann, J.; Keeling, C. I., Terpenoid biomaterials. *The Plant Journal* **2008,** *54* (4), 656-669.

132. Leavell, M. D.; McPhee, D. J.; Paddon, C. J., Developing fermentative terpenoid production for commercial usage. *Current Opinion in Biotechnology* **2016,** *37*, 114-119.

133. George, K. W.; Alonso-Gutierrez, J.; Keasling, J. D.; Lee, T. S., Isoprenoid Drugs, Biofuels, and Chemicals—Artemisinin, Farnesene, and Beyond. *Advances in biochemical engineering/biotechnology* **2015,** *148*, 355-389.

134. Li, Y.; Pfeifer, B. A., Heterologous production of plant-derived isoprenoid products in microbes and the application of metabolic engineering and synthetic biology. *Current Opinion in Plant Biology* **2014,** *19*, 8-13.

135. Dudley, Q. M.; Anderson, K. C.; Jewett, M. C., Cell-free mixing of Escherichia coli crude extracts to prototype and rationally engineer high-titer mevalonate synthesis. *ACS Synthetic Biology* **2016,** *5* (12), 1578-1588.

136. Jewett, M. C.; Swartz, J. R., Mimicking the Escherichia coli cytoplasmic environment activates long-lived and efficient cell-free protein synthesis. *Biotechnology and Bioengineering* **2004,** *86* (1), 19-26.

137. Richard, J. P., Kinetic parameters for the elimination reaction catalyzed by triosephosphate isomerase and an estimation of the reaction's physiological significance. *Biochemistry* **1991,** *30* (18), 4581-4585.

138. Iyengar, R.; Rose, I. A., Concentration of activated intermediates of the fructose-1,6-bisphosphate aldolase and triosephosphate isomerase reactions. *Biochemistry* **1981,** *20* (5), 1223-1229.

139. Richard, J. P., Mechanism for the formation of methylglyoxal from triosephosphates. *Biochemical Society Transactions* **1993,** *21* (2), 549-553.

140. Weber, A. L., Formation of the thioester, N-acetyl, S-lactoylcysteine, by reaction of N-acetylcysteine with pyruvaldehyde in aqueous solution. *Journal of Molecular Evolution* **1982,** *18* (5), 354-359.

141. Okuyama, T.; Komoguchi, S.; Fueno, T., *Reaction of thiols with phenylglyoxal to give thiol esters of mandelic acid. II. Intramolecular general-base catalysis and change in rate-determining step*. 1982; Vol. 104.

142. Huebner, A. M.; Abell, C.; Huck, W. T. S.; Baroud, C. N.; Hollfelder, F., Monitoring a Reaction at Submillisecond Resolution in Picoliter Volumes. *Analytical chemistry* **2011,** *83* (4), 1462-1468.

143. Mao, H.; Yang, T.; Cremer, P. S., A Microfluidic Device with a Linear Temperature Gradient for Parallel and Combinatorial Measurements. *Journal of the American Chemical Society* **2002,** *124* (16), 4432-4435.

144. Fidalgo, L. M.; Whyte, G.; Ruotolo, B. T.; Benesch, J. L. P.; Stengel, F.; Abell, C.; Robinson, C. V.; Huck, W. T. S., Coupling Microdroplet Microreactors with Mass Spectrometry: Reading the Contents of Single Droplets Online. *Angewandte Chemie International Edition* **2009,** *48* (20), 3665-3668.

145. Wang, J.; Sui, G.; Mocharla Vani, P.; Lin Rachel, J.; Phelps Michael, E.; Kolb Hartmuth, C.; Tseng, H. R., Integrated Microfluidics for Parallel Screening of an In Situ Click Chemistry Library. *Angewandte Chemie* **2006,** *118* (32), 5402-5407.

146. Browne Duncan, L.; Wright, S.; Deadman Benjamin, J.; Dunnage, S.; Baxendale Ian, R.; Turner Richard, M.; Ley Steven, V., Continuous flow reaction monitoring using an on-line miniature mass spectrometer. *Rapid Communications in Mass Spectrometry* **2012,** *26* (17), 1999-2010.

147. Liu, Y.; Ismagilov, R. F., Dynamics of Coalescence of Plugs with a Hydrophilic Wetting Layer Induced by Flow in a Microfluidic Chemistrode. *Langmuir* **2009,** *25* (5), 2854-2859.

148. Hatakeyama, T.; Chen, D. L.; Ismagilov, R. F., Microgram-Scale Testing of Reaction Conditions in Solution Using Nanoliter Plugs in Microfluidics with Detection by MALDI-MS. *Journal of the American Chemical Society* **2006,** *128* (8), 2518-2519.

149. Lederer, M. O.; Klaiber, R. G., Cross-linking of proteins by maillard processes: characterization and detection of lysine–arginine cross-links derived from glyoxal and methylglyoxal. *Bioorganic & Medicinal Chemistry* **1999,** *7* (11), 2499-2507.

150. Allaman, I.; Bélanger, M.; Magistretti, P. J., Methylglyoxal, the dark side of glycolysis. *Frontiers in Neuroscience* **2015,** *9* (23).

151. 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.

152. Grant, J.; Goudarzi, S. H.; Mrksich, M., High-Throughput Enzyme Kinetics with 3D Microfluidics and Imaging SAMDI Mass Spectrometry. *Analytical Chemistry* **2018,** *90* (21), 13096-13103.

153. Pollock, S. B.; Kent, S. B. H., An investigation into the origin of the dramatically reduced reactivity of peptide-prolyl-thioesters in native chemical ligation. *Chemical Communications* **2011,** *47* (8), 2342-2344.

154. Song, H.; Chen, D. L.; Ismagilov, R. F., Reactions in Droplets in Microfluidic Channels. *Angewandte Chemie International Edition* **2006,** *45* (44), 7336-7356.

155. Mark, D.; Haeberle, S.; Roth, G.; von Stetten, F.; Zengerle, R., Microfluidic lab-on-a-chip platforms: requirements, characteristics and applications. *Chem. Soc. Rev.* **2010,** *39* (3), 1153-1182.

156. Helal, K. Y.; Alamgir, A.; Berns, E. J.; Mrksich, M., Traceless Immobilization of Analytes for High-Throughput Experiments with SAMDI Mass Spectrometry. *Journal of the American Chemical Society* **2018,** *140* (26), 8060-8063.

157. 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 combinatorial science* **2015,** *17* (11), 658-662.

158. Yousaf, M. N.; Mrksich, M., Diels−Alder Reaction for the Selective Immobilization of Protein to Electroactive Self-Assembled Monolayers. *Journal of the American Chemical Society* **1999,** *121* (17), 4286-4287.

159. Tyagi, S.; VanDelinder, V.; Banterle, N.; Fuertes, G.; Milles, S.; Agez, M.; Lemke, E. A., Continuous throughput and long-term observation of single-molecule FRET without immobilization. *Nature Methods* **2014,** *11*, 297.

160. Bouilly, D.; Hon, J.; Daly, N. S.; Trocchia, S.; Vernick, S.; Yu, J.; Warren, S.; Wu, Y.; Gonzalez, R. L.; Shepard, K. L.; Nuckolls, C., Single-Molecule Reaction Chemistry in Patterned Nanowells. *Nano Letters* **2016,** *16* (7), 4679-4685.

161. Casadonte, R.; Caprioli, R. M., Proteomic analysis of formalin-fixed paraffin-embedded tissue by MALDI imaging mass spectrometry. *Nature Protocols* **2011,** *6*, 1695.

162. Fogler, H. S., *Essentials of Chemical Reaction Engineering*. 2011.

163. Kwon, Y.; Mrksich, M., Dependence of the Rate of an Interfacial Diels−Alder Reaction on the Steric Environment of the Immobilized Dienophile: An Example of Enthalpy−Entropy Compensation. *Journal of the American Chemical Society* **2002,** *124* (5), 806-812.

164. Abu-Reesh, I. M.; Abu-Sharkh, B. F., Comparison of Axial Dispersion and Tanks-in-Series Models for Simulating the Performance of Enzyme Reactors. *Industrial & Engineering Chemistry Research* **2003,** *42* (22), 5495-5505.

165. Bošković, D.; Loebbecke, S.; Gross, G. A.; Koehler, J. M., Residence Time Distribution Studies in Microfluidic Mixing Structures. *Chemical Engineering & Technology* **2011,** *34* (3), 361-370.

166. Gervais, T.; Jensen, K. F., Mass transport and surface reactions in microfluidic systems. *Chemical Engineering Science* **2006,** *61* (4), 1102-1121.

Appendix

Design of Adaptor Domain Fusion Proteins:

The artificial fusion proteins of various chromodomains to KDAC8, used in the experiments presented in chapters 3 and 4, were generated in the commercially available vector pET-303/CT-His (Invitrogen/ThermoFisher Scientific). An artificial fusion of the SH2 domain from the phosphatase, SHP1, to cutinase was used in chapter 5 and was expressed in the commercial pET-22b(+) (Invitrogen/ThermoFisher Scientific). The for this latter construct NcoI restriction site was used to maintain the pelB leader sequence in the pET-22b(+) vector, which directs periplasmic expression. For the purposes of cloning and expression, the sequences below represent DNA sequences of the insert coding for each fusion protein. All sequences below are listed from 5' to 3'.

Key:
Restriction Sites: XbaI NotI XhoI NcoI
Blue text= adaptor domain
Black text= reporter enzyme
Grey text= linker region

CBX1-KDAC8 Sequence:

TCTAGAATGGAATATGTGGTGGAAAAAGTTCTCGACCGTCGAGTGGTAAAGGGCAA
AGTGGAGTACCTCCTAAAGTGGAAGGGATTCTCAGATGAGGACAACACATGGGAGC
CAGAAGAGAACCTGGATTGCCCCGACCTCATTGCTGAGTTTCTGCAGTCACAGAAAA
CAGCGGCCGCGGGCGGTTCTTCTATGGAGGAGCCGGAGGAACCGGCGGACAGTGGG
CAGTCGCTGGTCCCGGTTTATATCTATAGTCCCGAGTATGTCAGTATGTGTGACTCCC
TGGCCAAGATCCCCAAACGGGCCAGTATGGTGCATTCTTTGATTGAAGCATATGCAC
TGCATAAGCAGATGAGGATAGTTAAGCCTAAAGTGGCCTCCATGGAGGAGATGGCC
ACCTTCCACACTGATGCTTATCTGCAGCATCTCCAGAAGGTCAGCCAAGAGGGCGAT
GATGATCATCCGGACTCCATAGAATATGGGCTAGGTTATGACTGCCCAGCCACTGAA
GGGATATTTGACTATGCAGCAGCTATAGGAGGGGCTACGATCACAGCTGCCCAATG
CCTGATTGACGGAATGTGCAAAGTAGCAATTAACTGGTCTGGAGGGTGGCATCATG
CAAAGAAAGATGAAGCATCTGGTTTTTGTTATCTCAATGATGCTGTCCTGGGAATAT

TACGATTGCGACGGAAATTTGAGCGTATTCTCTACGTGGATTTGGATCTGCACCATG
GAGATGGTGTAGAAGACGCATTCAGTTTCACCTCCAAAGTCATGACCGTGTCCCTGC
ACAAATTCTCCCCAGGATTTTTCCCAGGAACAGGTGACGTGTCTGATGTTGGCCTAG
GGAAGGGACGGTACTACAGTGTAAATGTGCCCATTCAGGATGGCATACAAGATGAA
AAATATTACCAGATCTGTGAAAGTGTACTAAAGGAAGTATACCAAGCCTTTAATCCC
AAAGCAGTGGTCTTACAGCTGGGAGCTGACACAATAGCTGGGGATCCCATGTGCTC
CTTTAACATGACTCCAGTGGGAATTGGCAAGTGTCTTAAGTACATCCTTCAATGGCA
GTTGGCAACACTCATTTTGGGAGGAGGAGGCTATAACCTTGCCAACACGGCTCGATG
CTGGACATACTTGACCGGGGTCATCCTAGGGAAAACACTATCCTCTGAGATCCCAGA
TCATGAGTTTTTCACAGCATATGGTCCTGATTATGTGCTGGAAATCACGCCAAGCTG
CCGGCCAGACCGCAATGAGCCCCACCGAATCCAACAAATCCTCAACTACATCAAAG
GGAATCTGAAGCATGTGGTCCTCGAG

CBX3-KDAC8 Sequence:

TCTAGAATGGAATTTGTCGTGGAAAAAGTACTAGATCGACGTGTAGTGAATGGGAA
AGTGGAATATTTCCTGAAGTGGAAGGGATTTACAGATGCTGACAATACTTGGGAAC
CTGAAGAAAATTTAGATTGTCCAGAATTGATTGAAGCGTTTCTTAACTCTCAGAAAG
CTGGCAAAGAAAAAGCGGCCGCGGGCGGTTCTTCTATGGAGGAGCCGGAGGAACCG
GCGGACAGTGGGCAGTCGCTGGTCCCGGTTTATATCTATAGTCCCGAGTATGTCAGT
ATGTGTGACTCCCTGGCCAAGATCCCCAAACGGGCCAGTATGGTGCATTCTTTGATT
GAAGCATATGCACTGCATAAGCAGATGAGGATAGTTAAGCCTAAAGTGGCCTCCAT
GGAGGAGATGGCCACCTTCCACACTGATGCTTATCTGCAGCATCTCCAGAAGGTCAG
CCAAGAGGGCGATGATGATCATCCGGACTCCATAGAATATGGGCTAGGTTATGACT
GCCCAGCCACTGAAGGGATATTTGACTATGCAGCAGCTATAGGAGGGGCTACGATC
ACAGCTGCCCAATGCCTGATTGACGGAATGTGCAAAGTAGCAATTAACTGGTCTGG
AGGGTGGCATCATGCAAAGAAAGATGAAGCATCTGGTTTTTGTTATCTCAATGATGC
TGTCCTGGGAATATTACGATTGCGACGGAAATTTGAGCGTATTCTCTACGTGGATTT
GGATCTGCACCATGGAGATGGTGTAGAAGACGCATTCAGTTTCACCTCCAAAGTCAT
GACCGTGTCCCTGCACAAATTCTCCCCAGGATTTTTCCCAGGAACAGGTGACGTGTC
TGATGTTGGCCTAGGGAAGGGACGGTACTACAGTGTAAATGTGCCCATTCAGGATG
GCATACAAGATGAAAAATATTACCAGATCTGTGAAAGTGTACTAAAGGAAGTATAC
CAAGCCTTTAATCCCAAAGCAGTGGTCTTACAGCTGGGAGCTGACACAATAGCTGG
GGATCCCATGTGCTCCTTTAACATGACTCCAGTGGGAATTGGCAAGTGTCTTAAGTA
CATCCTTCAATGGCAGTTGGCAACACTCATTTTGGGAGGAGGAGGCTATAACCTTGC
CAACACGGCTCGATGCTGGACATACTTGACCGGGGTCATCCTAGGGAAAACACTAT
CCTCTGAGATCCCAGATCATGAGTTTTTCACAGCATATGGTCCTGATTATGTGCTGG
AAATCACGCCAAGCTGCCGGCCAGACCGCAATGAGCCCCACCGAATCCAACAAATC
CTCAACTACATCAAAGGGAATCTGAAGCATGTGGTCCTCGAG

CBX4-KDAC8 Sequence:

TCTAGAATGGAGCACGTCTTCGCGGTGGAGAGCATCGAGAAGAAGCGGATCCGCAA
GGGCAGAGTGGAGTATCTGGTGAAATGGAGAGGCTGGTCGCCCAAATATAACACGT
GGGAACCGGAGGAGAACATCCTGGACCCCAGGCTGCTGATCGCCTTCCAGAACAGG
GAACGGCAGGAGCAGGCGGGCCGCGGGCGGTTCTTCTATGGAGGAGCCGGAGGAAC
CGGCGGACAGTGGGCAGTCGCTGGTCCCGGTTTATATCTATAGTCCCGAGTATGTCA
GTATGTGTGACTCCCTGGCCAAGATCCCCAAACGGGCCAGTATGGTGCATTCTTTGA
TTGAAGCATATGCACTGCATAAGCAGATGAGGATAGTTAAGCCTAAAGTGGCCTCC
ATGGAGGAGATGGCCACCTTCCACACTGATGCTTATCTGCAGCATCTCCAGAAGGTC
AGCCAAGAGGGCGATGATGATCATCCGGACTCCATAGAATATGGGCTAGGTTATGA
CTGCCCAGCCACTGAAGGGATATTTGACTATGCAGCAGCTATAGGAGGGGCTACGA
TCACAGCTGCCCAATGCCTGATTGACGGAATGTGCAAAGTAGCAATTAACTGGTCTG
GAGGGTGGCATCATGCAAAGAAAGATGAAGCATCTGGTTTTTGTTATCTCAATGATG
CTGTCCTGGGAATATTACGATTGCGACGGAAATTTGAGCGTATTCTCTACGTGGATT
TGGATCTGCACCATGGAGATGGTGTAGAAGACGCATTCAGTTTCACCTCCAAAGTCA
TGACCGTGTCCCTGCACAAATTCTCCCCAGGATTTTTCCCAGGAACAGGTGACGTGT
CTGATGTTGGCCTAGGGAAGGGACGGTACTACAGTGTAAATGTGCCCATTCAGGAT
GGCATACAAGATGAAAAATATTACCAGATCTGTGAAAGTGTACTAAAGGAAGTATA
CCAAGCCTTTAATCCCAAAGCAGTGGTCTTACAGCTGGGAGCTGACACAATAGCTGG
GGATCCCATGTGCTCCTTTAACATGACTCCAGTGGGAATTGGCAAGTGTCTTAAGTA
CATCCTTCAATGGCAGTTGGCAACACTCATTTTGGGAGGAGGAGGCTATAACCTTGC
CAACACGGCTCGATGCTGGACATACTTGACCGGGGTCATCCTAGGGAAAACACTAT
CCTCTGAGATCCCAGATCATGAGTTTTTCACAGCATATGGTCCTGATTATGTGCTGG
AAATCACGCCAAGCTGCCGGCCAGACCGCAATGAGCCCCACCGAATCCAACAAATC
CTCAACTACATCAAAGGGAATCTGAAGCATGTGGTCCTCGAG

CBX5-KDAC8 Sequence:

TCTAGAATGGAGGAGTATGTTGTGGAGAAGGTGCTAGACAGGCGCGTGGTTAAGGG
ACAAGTGGAATATCTACTGAAGTGGAAAGGCTTTTCTGAGGAGCACAATACTTGGG
AACCTGAGAAAAACTTGGATTGCCCTGAGCTAATTTCTGAATTTATGAAAAAGTATA
AGAAGATGAAGGAGGCGGGCCGCGGGCGGTTCTTCTATGGAGGAGCCGGAGGAACC
GGCGGACAGTGGGCAGTCGCTGGTCCCGGTTTATATCTATAGTCCCGAGTATGTCAG
TATGTGTGACTCCCTGGCCAAGATCCCCAAACGGGCCAGTATGGTGCATTCTTTGAT
TGAAGCATATGCACTGCATAAGCAGATGAGGATAGTTAAGCCTAAAGTGGCCTCCA
TGGAGGAGATGGCCACCTTCCACACTGATGCTTATCTGCAGCATCTCCAGAAGGTCA
GCCAAGAGGGCGATGATGATCATCCGGACTCCATAGAATATGGGCTAGGTTATGAC
TGCCCAGCCACTGAAGGGATATTTGACTATGCAGCAGCTATAGGAGGGGCTACGAT
CACAGCTGCCCAATGCCTGATTGACGGAATGTGCAAAGTAGCAATTAACTGGTCTGG
AGGGTGGCATCATGCAAAGAAAGATGAAGCATCTGGTTTTTGTTATCTCAATGATGC
TGTCCTGGGAATATTACGATTGCGACGGAAATTTGAGCGTATTCTCTACGTGGATTT
GGATCTGCACCATGGAGATGGTGTAGAAGACGCATTCAGTTTCACCTCCAAAGTCAT
GACCGTGTCCCTGCACAAATTCTCCCCAGGATTTTTCCCAGGAACAGGTGACGTGTC
TGATGTTGGCCTAGGGAAGGGACGGTACTACAGTGTAAATGTGCCCATTCAGGATG

GCATACAAGATGAAAAATATTACCAGATCTGTGAAAGTGTACTAAAGGAAGTATAC
CAAGCCTTTAATCCCAAAGCAGTGGTCTTACAGCTGGGAGCTGACACAATAGCTGG
GGATCCCATGTGCTCCTTTAACATGACTCCAGTGGGAATTGGCAAGTGTCTTAAGTA
CATCCTTCAATGGCAGTTGGCAACACTCATTTTGGGAGGAGGAGGCTATAACCTTGC
CAACACGGCTCGATGCTGGACATACTTGACCGGGGTCATCCTAGGGAAAACACTAT
CCTCTGAGATCCCAGATCATGAGTTTTTCACAGCATATGGTCCTGATTATGTGCTGG
AAATCACGCCAAGCTGCCGGCCAGACCGCAATGAGCCCCACCGAATCCAACAAATC
CTCAACTACATCAAAGGGAATCTGAAGCATGTGGTCCTCGAG

CBX6-KDAC8 Sequence:

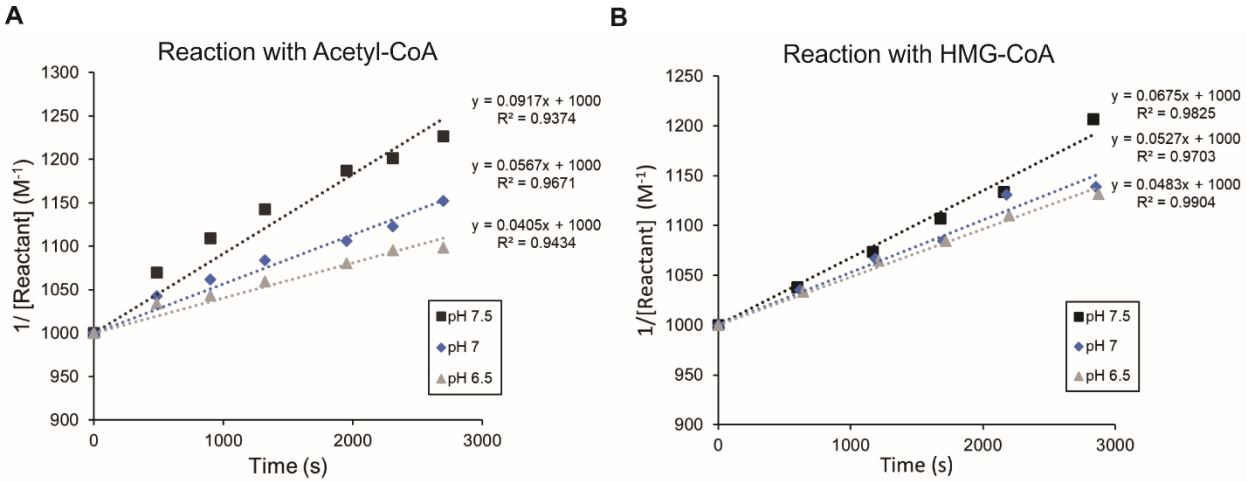
TCTAGAATGGAGCGGGTCTTCGCGGCCGAATCCATCATCAAACGGCGGATCCGAAA
GGGACGCATCGAGTACCTGGTGAAATGGAAGGGGTGGGCGATCAAGTACAGCACTT
GGGAGCCCGAGGAGAACATCCTGGACTCGCGGCTCATTGCAGCCTTCGAACAAAAG
GAGAGGGAGCGTGAGGCGGCCGCGGGCGGTTCTTCTATGGAGGAGCCGGAGGAACC
GGCGGACAGTGGGCAGTCGCTGGTCCCGGTTTATATCTATAGTCCCGAGTATGTCAG
TATGTGTGACTCCCTGGCCAAGATCCCCAAACGGGCCAGTATGGTGCATTCTTTGAT
TGAAGCATATGCACTGCATAAGCAGATGAGGATAGTTAAGCCTAAAGTGGCCTCCA
TGGAGGAGATGGCCACCTTCCACACTGATGCTTATCTGCAGCATCTCCAGAAGGTCA
GCCAAGAGGGCGATGATGATCATCCGGACTCCATAGAATATGGGCTAGGTTATGAC
TGCCCAGCCACTGAAGGGATATTTGACTATGCAGCAGCTATAGGAGGGGCTACGAT
CACAGCTGCCCAATGCCTGATTGACGGAATGTGCAAAGTAGCAATTAACTGGTCTGG
AGGGTGGCATCATGCAAAGAAAGATGAAGCATCTGGTTTTTGTTATCTCAATGATGC
TGTCCTGGGAATATTACGATTGCGACGGAAATTTGAGCGTATTCTCTACGTGGATTT
GGATCTGCACCATGGAGATGGTGTAGAAGACGCATTCAGTTTCACCTCCAAAGTCAT
GACCGTGTCCCTGCACAAATTCTCCCCAGGATTTTTCCCAGGAACAGGTGACGTGTC
TGATGTTGGCCTAGGGAAGGGACGGTACTACAGTGTAAATGTGCCCATTCAGGATG
GCATACAAGATGAAAAATATTACCAGATCTGTGAAAGTGTACTAAAGGAAGTATAC
CAAGCCTTTAATCCCAAAGCAGTGGTCTTACAGCTGGGAGCTGACACAATAGCTGG
GGATCCCATGTGCTCCTTTAACATGACTCCAGTGGGAATTGGCAAGTGTCTTAAGTA
CATCCTTCAATGGCAGTTGGCAACACTCATTTTGGGAGGAGGAGGCTATAACCTTGC
CAACACGGCTCGATGCTGGACATACTTGACCGGGGTCATCCTAGGGAAAACACTAT
CCTCTGAGATCCCAGATCATGAGTTTTTCACAGCATATGGTCCTGATTATGTGCTGG
AAATCACGCCAAGCTGCCGGCCAGACCGCAATGAGCCCCACCGAATCCAACAAATC
CTCAACTACATCAAAGGGAATCTGAAGCATGTGGTCCTCGAG

CUT-SH2 Sequence:

CCATGGATGGGCCTGCCTACTTCTAACCCTGCCCAGGAGCTTGAGGCGCGCCAGCTT
GGTAGAACAACTCGCGACGATCTGATCAACGGCAATAGCGCTTCCTGCGCCGATGT
CATCTTCATTTATGCCCGAGGTTCAACAGAGACGGGCAACTTGGGTACCCTCGGTCC
TAGCATTGCCTCCAACCTTGAGTCCGCGTTCGGCAAGGACGGTGTCTGGATTCAGGG
CGTTGGCGGTGCCTACCGTGCCACTCTTGGAGACAATGCTCTCCCTCGCGGAACCTC

TAGCGCCGCAATCAGGGAGATGCTCGGTCTCTTCCAGCAGGCCAACACCAAGTGCC
CTGACGCGACTTTGATCGCCGGTGGCTACAGCCAGGGTGCTGCACTTGCAGCCGCCT
CCATCGAGGACCTCGACTCGGCCATTCGTGACAAGATCGCCGGAACTGTTCTGTTCG
GCTATACCAAGAACCTACAGAACCGTGGCCGAATCCCCAACTACCCTGCCGATAGG
ACCAAGGTCTTCTGCAATACAGGGGATCTCGTTTGTACTGGTAGCTTGATCGTTGCT
GCACCTCACTTGGCGTATGGTCCTGATGCTCGTGGCCCTGCCCCTGAGTTCCTCATCG
AGAAGGCTCGGGCTGTCCGTGGTTCTGCTGCGGCCGCGGGCGGTTCTTCTTGGTTTC
ACCGAGACCTCAGTGGGCTGGATGCAGAGACCCTGCTCAAGGGCCGAGGTGTCCAC
GGTAGCTTCCTGGCTCGGCCCAGTCGCAAGAACCAGGGTGACTTCTCGCTCTCCGTC
AGGGTGGGGGATCAGGTGACCCATATTCGGATCCAGAACTCAGGGGATTTCTATGA
CCTGTATGGAGGGGAGAAGTTTGCGACTCTGACAGAGCTGGTGGAGTACTACACTC
AGCAGCAGGGTGTCCTGCAGGACCGCGACGGCACCATCATCCACCTCAAGTACCCG
CTGCTCGAG

Supplementary Figures:



C

	2nd Order Rate Constant (M^{-1} s^{-1})	
pH	Acetyl-CoA	HMG-CoA
7.5	0.092	0.068
7.0	0.057	0.053
6.5	0.041	0.048

Figure S1. Second Order Kinetics for the Reaction of CoA Metabolites with the Peptide. The peptide, of sequence CAK(Me_3)SA, was reacted with pure Acetyl-CoA and HMG-CoA (purchased from Sigma Aldrich). For each reaction, 1 mM of peptide was mixed with 1mM of CoA metabolite in buffer (100 mM BisTris, 10 mM Phosphate at the pH values listed). The reactions were performed at a 200 uL scale and incubated at 42°C. Reaction timepoints were obtained by sampling 10 µL of reaction mixture at desired times and spotting it, in triplicate, onto 384 spot array plates presenting a 20% maleimide SAM. The reaction progress at each timepoint was determined by performing SAMDI and analyzing the areas under the curve for the peaks corresponding to the unreacted peptide and ligated product. The yield was then fitted according to a standard second order rate plot for both (**A**) Acetyl-CoA and (**B**) HMG-CoA. The plots adopt linear relationships, where slope gives the second order rate constants of the reactions. **C**) The table summarizes the experimentally determined rate constants across pH conditions.

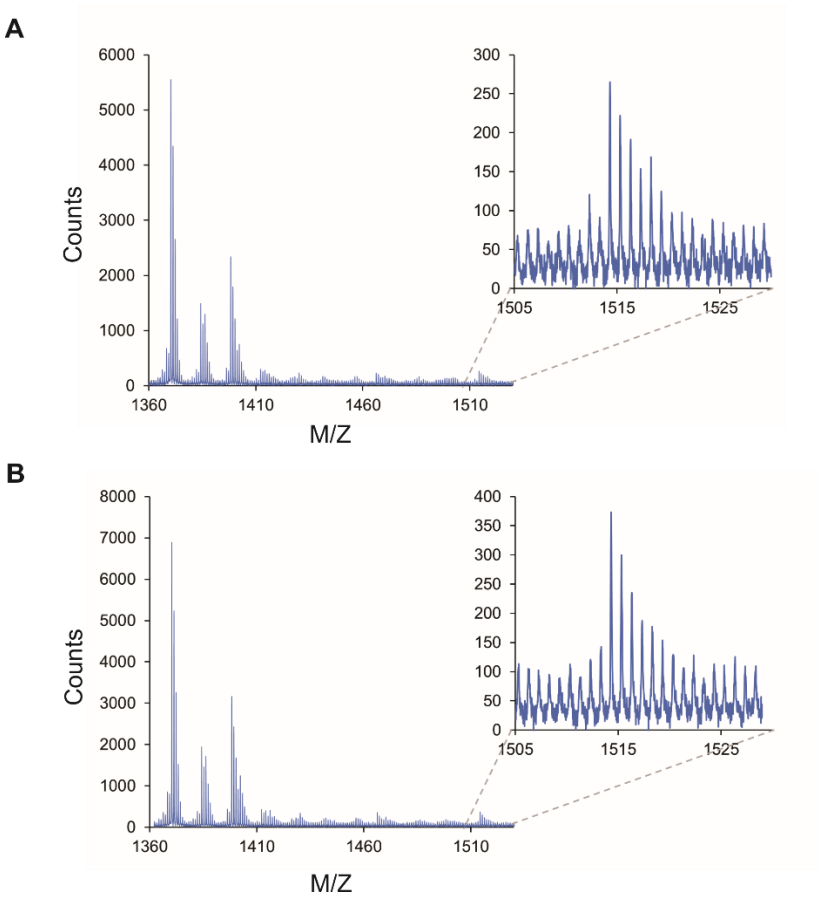


Figure S2. Limit of Detection for HMG-CoA. In order to explore the limit of detection, reactions were performed using known concentrations of HMG-CoA under conditions identical to those used in the large-scale, cell free reaction arrays (0.9 mM peptide and 0.1 mM internal standard, in a BisTris and phosphate buffer system at pH 6.5, incubated at 42° C for 3 hours). Shown here are two spectra corresponding to reactions with HMG-CoA concentrations of **A)** 10 µM and **B)** 5 µM. The peak corresponding to the peptide reactant appears at 1371 mass units and the peak corresponding to the captured HMG metabolite appears at 1515 mass units. As shown, the HMG ligated product is detectable at 5 µM. We also attempted a reaction with 1 µM of HMG-CoA, but at this concentration, the ligation product was no longer discernable.

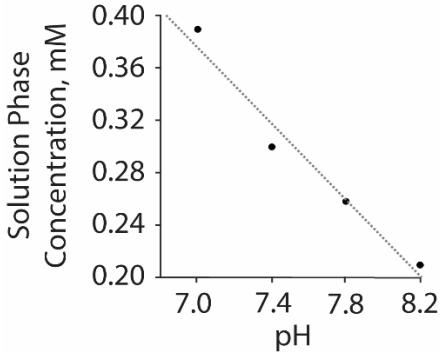
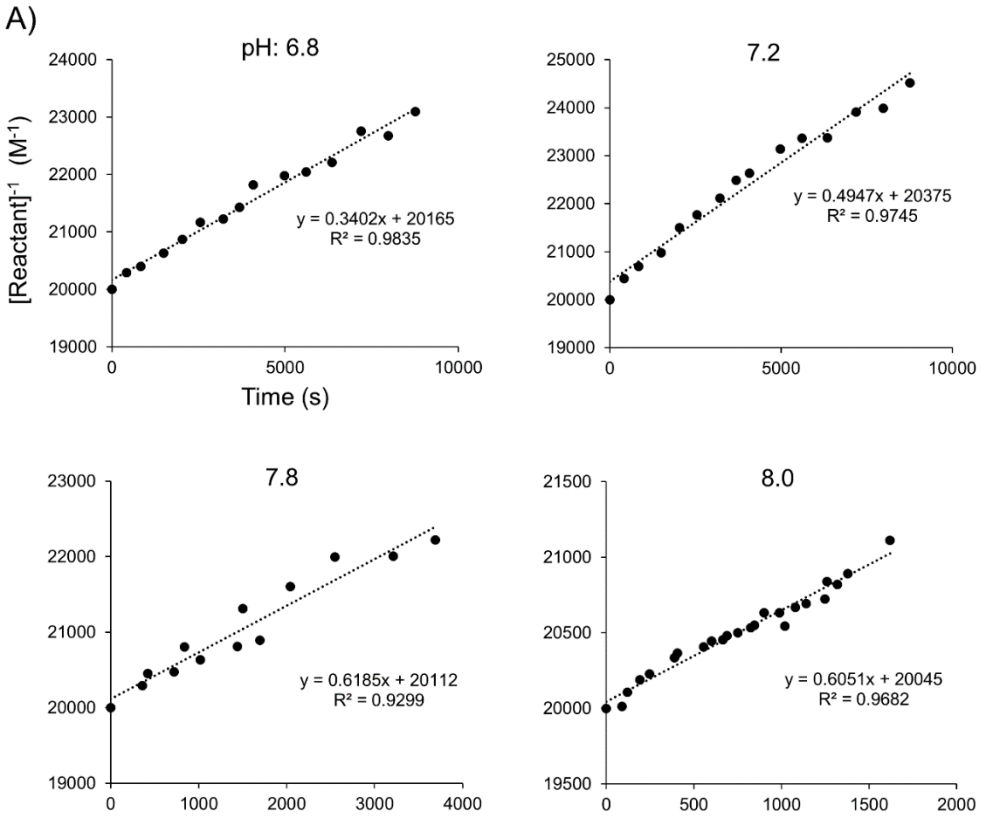


Figure S3. Average solution phase concentration at surface saturation v. solution pH. R^2= 0.9668.

A)



B)

pH	Rate Constant Determined from Fluidic Device ($M^{-1}s^{-1}$)	Rate Constant Determined from Bulk Solution ($M^{-1}s^{-1}$)
6.8	0.38 ± 0.06	0.34 ± 0.02
7.2	0.48 ± 0.04	0.49 ± 0.02
7.8	0.62 ± 0.05	0.62 ± 0.05
8.0	0.62 ± 0.01	0.61 ± 0.03

Figure S4. The rate constants determined from the reactions in the microfluidic device were compared to corresponding bulk solution reactions. Separate reactions containing 50 µM of phenyl glyoxal and 50 µM of N-terminal peptide (of sequence CAK(Me_3)SA) were generated for each of the four buffer conditions. These low concentrations are representative of the concentration regime sampled by the channels of the device. All reactions were performed at a 400 µL scale in standard, 500 µL Eppendorf tubes and were heated to 35° C using a heat block. At each desired timepoints, 10 µL of reaction mixture was pipetted by hand from the bulk reaction and spotted onto array surfaces possessing monolayers identical to those used in the microfluidic devices. These bulk reactions were then analyzed in a fashion identical to that of the fluidic device. The array surfaces were analyzed by SAMDI-MS and the relative yield was determined by integration of the reactant peptide and product peptide peaks. **A**) The the inverse of the reactant peptide concentration was

plotted versus time, in accordance with a standard 2nd order rate law. **B**) The 2nd order rate constants, given by the slopes, match closely with those determined in the fluidic device.

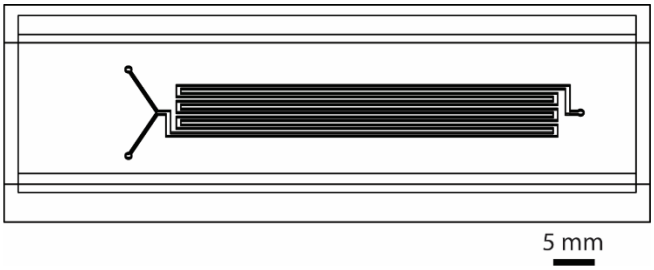


Figure S5. SolidWorks rendering of the PDMS master.

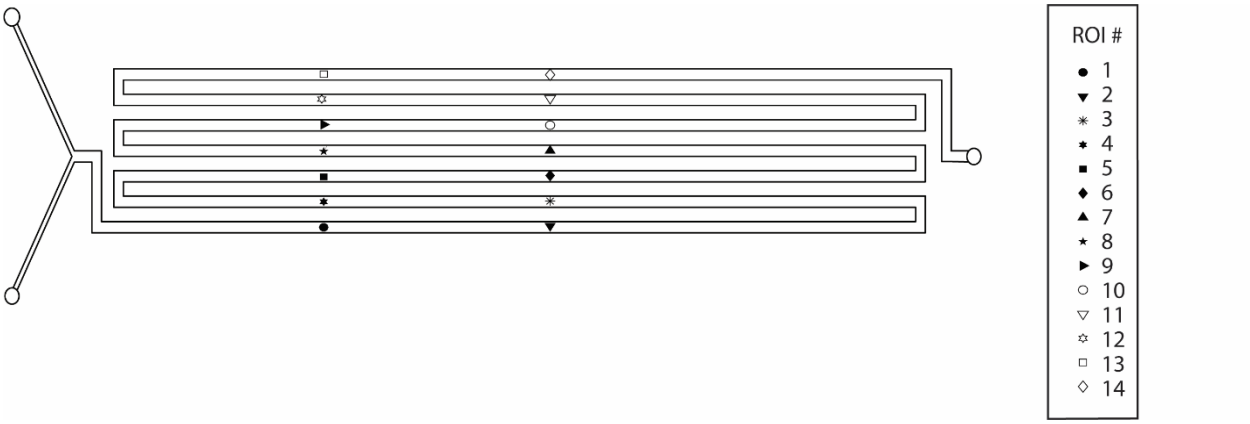


Figure S6. The location of each ROI in the device channel.