NORTHWESTERN UNIVERSITY

Application of SAMDI Mass Spectrometry for Molecular Data Storage and Random Peptide Arrays

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Chemistry

By

Alexei Sergeevich Ten

EVANSTON, ILLINOIS

March 2020

Abstract

Application of SAMDI Mass Spectrometry on Molecular Data Storage

and Random Peptide Arrays

Alexei Sergeevich Ten

Self-Assembled Monolayers for MALDI-TOF Mass Spectrometry (SAMDI-MS) is a technique that combines self-assembling molecules of alkane disulfides on gold and MALDI-TOF mass spectrometry. By using well-defined monolayers with functionalizations that both prevent non-specific adsorption onto the surface and presents immobilization handles, it is possible to pull out analytes of interest, selectively, onto the surface from reaction mixtures. The use of the MALDI-TOF enables rapid and quantitative characterization of immobilized molecules, by their mass, enabling quantification of both reactants and products, of reactions that incur a mass change. In this work I will demonstrate the utility of SAMDI-MS with peptides and peptide arrays to demonstrate two novel uses of this method.

First, I will demonstrate an approach that uses peptide mixtures with SAMDI to store digital information in molecules. Currently, digital information is stored on many different media, including magnetic tape, optical discs, hard disks, and flash. Although each of these methods has its benefits, they also have drawbacks. To combat some of these possible drawbacks, such as longevity, the use of molecules to storage information has been proposed. The molecules predominantly considered when information storage is uttered is DNA. DNA has been used by nature to store genetic information for billions of years, and now with the advances in DNA synthesis and sequencing technology it is possible to synthesize artificial strands of DNA to encode digital data, albeit at a cost many times more expensive than currently available technology. To

provide another possible method for storing information using molecules, I will demonstrate how it is possible to use the feature of SAMDI surfaces to pull out all analytes with the proper handles out of a solution and distinguish them by mass. I use this feature and show that it is possible to immobilize 32 peptides onto a single spot, giving us an information density of 4 bytes/spot in binary. I will further expand on this ability and incorporate SAMDI's ability to quantitate molecule abundance in solution and demonstrate that through the use of concentrations it is also possible to store information in higher-than-binary manner.

In the second portion of this work I will show two novel applications of SAMDI and peptide arrays, as well as introduce a method for the synthesis, analysis, and utilization of random peptide arrays with a novel capping scheme. The two novel applications of SAMDI are to first study the proteases contained within detergents, and second, to look at initial rates of related isoforms of KDACs, on a peptide library, to provide a method to distinguish activity profiles of these isoforms that have similar activity. The final approach will provide a method to utilize a split-pool synthesis method to synthesize random peptide libraries. I will show that it is possible to combine this method with a novel capping scheme, which uses only 4 capping groups, and SAMDI to fully decode a library of peptides without any ambiguity, which has been the drawback of previous applications. I will also demonstrate the use of these random arrays with SAMDI surfaces in an attempt to discover a novel short peptide adhesion ligand using a 5,832-member library.

Acknowledgements

First and foremost, I would like to thank my advisor, Professor Milan Mrksich, for giving me the support, encouragement, and motivation needed to complete my thesis work. Without his support and input I would not have grown as a scientist the way I have. I would also like to thank Professor Richard Silverman and Professor Even Scott for taking the time out of their busy schedules to serve on my dissertation committee. I would also like to thank Professor Brian Hoffman and Professor Alexander Statsyuk for being on my qualifying exam committee.

Next, I would like to thank all the collaborators with whom I've had the pleasure of working with, and without whom many of these projects would have taken a much longer time to complete. I would like to thank Raymond Dai, who I've had the pleasure of mentoring throughout his 3 years in the lab, and who has contributed greatly to the random array and molecular data storage projects, as well as leading the detergent protease project. I would also to thank my three summer research students, Matthew Chow, Carolyn Ma, and Scott Morey. Matthew and Carolyn were instrumental in the initial stages of the random array project, and Scott simplified the completion of the molecular data storage project through his knowledge of python coding. I would also like to thank Dr. Andreea Stuparu, Dr. Adam Eisenberg, Heather Scott, and Matthew Culpepper, with whom I've had the pleasure of working on the KDAC initial slopes project with. Further, I would like to thank the Whitesides group, especially the members that are part of the molecular informatics programs, Dr. Michael Fink, Dr. Amit Nagarkar, Dr. Brian Cafferty, Dr. Albert Wong, and Sarah Battat. I must also thank Dr. George Whitesides, who throughout the last 2 years essentially served as a second advisor, through the molecular informatics collaboration, and who has graciously agreed to serve as my postdoc advisor. I would also like to thank Dr. Pradeep Bugga, without whom I would not have learned much of the cell culturing that I have, and without who's help,

none of the cell adhesion studies could have been done. Lastly, I would like to thank all the collaborators who have worked with me on many of the projects that have been explored in the past, including Garima Goyal, Gaurav Sinsinbar, and Juliet Roll.

I would also like to extend my sincerest thanks to all current and past Mrksich group members, who throughout my time in the lab have served as support figures, mentors, and great friends, making the lab a place I looked forward to working everyday. I would especially like to thank Dr. Justin Modica, who is a wealth of knowledge, and to whom I could also reach out to seek advice on many topics. I would also like to thank Dr. Patrick O'Kane, whom I've had the great pleasure starting my graduate school career at University of Chicago with, and who has been my most trusted confidant and friend within the Mrksich group. I would also like to thank Dr. Lindsey Szymczak, who was like a younger sister I never knew I wanted, but had the pleasure of having, and without whom I probably would not have gotten myself into as much trouble on our group expeditions through Chicago. I would also like to thank Dr. Jennifer Grant, Dr. Sarah Wood, Dan Sykora, and Dr. Courtney Sobers, who have made my time fun not only in the lab, but also out of lab. Finally, I would like to thank Yael Mayer, without whom my entire experience in graduate school would have been different. Not only was she a friend, confidant, and our administrator, but she was an all-around awesome person, without whom happy hours and group events would not have been the same.

I would also like to thank the entire University of Chicago chemistry entering class of 2010. If it wasn't for the support of many of these wonderful people, I would not be where I am today. From late night study sessions, to nights out on the town, everyone in this class made my first and subsequent years in graduate school fun and tolerable. I would especially like Dr. Zack Hund, Dr.

Patrick Figliozzi, Dr. Dmitry Dolzhikov, Erica Stevenson, Dr. John Savage, and Dr. Andrew Sand, all of whom have played pivotal roles in my time in graduate school, and with whom I hope I can stay connected as I close this chapter in my life.

I would also like to thank all of the wonderful people that I have met while residing in Chicago. Although there are too many to name, I would like to thank Sara Gehrdes, Liam Page, Mike Healy, Lendi Hubbard, and Mellisa Hund. I would also like to thank all of my friends from before my time in graduate school, with whom I have remained close with. I would like to thank Michael Gostintsev, who I've known for almost as long as I've been in the United States, and with whom I've remained close to this date. I would also like to thank Brian Ito, who is another one of my close childhood friends, and who I've enjoyed having visit me in Chicago, and visiting him in Los Angeles.

I also would like to thank my wonderful girlfriend Patricia Lawley. She has been one of the best things to happen in my life and has been the pillar that has supported me through my later half of graduate school. I would also like to extend my thanks to her entire family, who have acted like a second family to me and have allowed me to enter their lives. Finally, I must thank my entire family. Without my parent's encouragements, and sometime nagging, I would not have accomplished the goal that I've set out to accomplish almost 10 years ago. It is through their support, unconditional love, and wisdom, that I've managed to grow as a person. I must also thank my brother, with whom I've grown closer, even though we have been separated greater than ever physically.

Table of Contents

List of Figures

Chapter 1: Introduction

This dissertation combines the use of self-assembled monolayers for MALDI-TOF mass spectrometry (SAMDI-MS) with peptides and peptide arrays to demonstrate examples of molecular data storage, profiling enzymatic activity on peptide substrates, and a method for synthesizing large peptide libraries to be used with SAMDI-MS. In this chapter I will introduce the concept and development of SAMDI-MS, molecular data storage, and peptide arrays.

SAMDI-MS

Over the past two decades, the Mrksich group has developed a robust and widely applicable assay technique, SAMDI-MS. At the core of this technique are self-assembled monoloayer (SAMs) and MALDI-TOF mass spectrometry. As the name suggest, SAMs, are a set of molecules that, given correct conditions, will self-assemble to form a well-defined single molecule thick layer on surfaces. [1, 2] The Mrksich group uses monolayers consisting of a mixture of symmetrical alkane disulfides terminated with tri(ethylene glycol) groups and an asymmetrical alkane disulfide, one chain terminated in tri(ethylene glycol) and the other in a reactive immobilization handle. This mixture of disulfides readily forms a monolayer when reacted with a gold surface. Using a mixture of SAM molecules also allows for easy manipulation of the concentration of immobilization handles on the surface, and the use of tri(ethylene glycol) groups effectively passivates the monolayer surfaces, rendering them resistant to non-specific adsorption of proteins.[3] Passivating the surface against non-specific adsorption enables the use of many different immobilization chemistries on the monolayers. The analytes of interest be can easily pulled out of reaction mixtures by using specific immobilization handles that are selective to the analytes of interest. Previously we have shown the ability to immobilize molecules through click chemistry, Diels-

Alder reaction, biotin-streptavidin binding, and various nucleophilic additions.[4-8] However, the most used immobilization chemistry, and one that will be used through this thesis, is the Michael addition of a thiol to maleimide (Figure 1.1).[9] The maleimide functionalization on SAMs enables easy immobilization of cysteine containing peptides, through the thiol on the residue's sidechain. This reaction occurs rapidly at neutral pH and in high-yield. It is also highly selective in the presence of other functionalities.[10, 11]

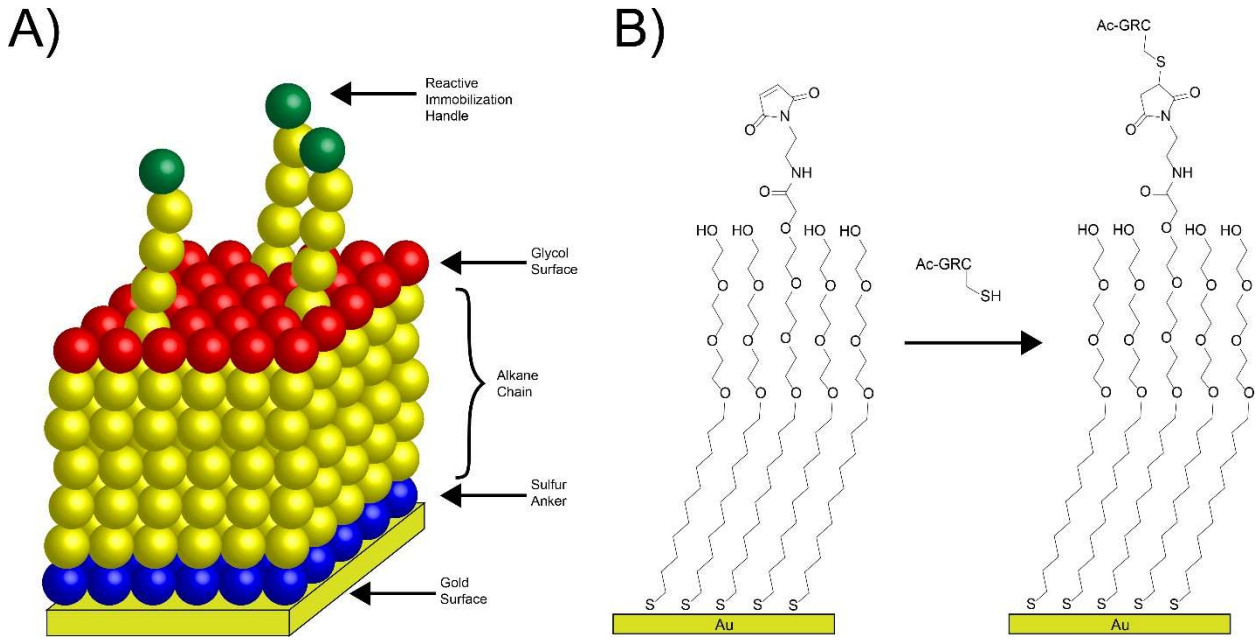


Figure 1.1: 3D and chemical structures of self-assembled monolayers used for SAMDI. A) 3D rendering of the self-assembled monolayers used with SAMDI. B) Chemical structure of the monolayers. Here the immobilization handle is a maleimide group, which undergoes a Michael addition with a thiol containing molecule, such as a peptide.

The use of the above-mentioned monolayers gave us the platform for immobilizing analytes of interest, however it was MALDI-TOF that enables the characterization of these monolayers. MALDI-TOF is a method whereby a molecule of interest is mixed into a solution of matrix compounds and then spotted onto a surface, where it is dried. The substrate with the dry spots is then inserted into the instrument and a laser is shone onto the spots of matrix and molecule.

The matrix transfers the energy from the laser into the molecules, ionizing them which then fly toward a detector, and their mass-to-charge ratio is then determined by their time of flight. The use of matrix to transfer the energy into molecules of interest makes MALDI a "soft ionization" technique, whereby the molecules that are excited by the laser do not undergo fragmentation. Thus, it is possible to ionize entire molecules, such as organic compounds, peptides, or even proteins, and obtain information on their mass to charge ratios. Although MALDI-TOF is a soft ionization technique, which does not promote fragmentation, it was our belief that the laser of a MALDI instrument is powerful enough to cleave the gold-sulfur bond, per previously experiments.[12] In fact, when the gold surfaces with SAMs were covered in matrix and inserted into a MALDI instrument, we observed clean cleavage of the gold-thiol bond, without any apparent fragmentation of the monolayers, which flew as either alkane thiols or disulfides (Figure 1.2).[13] This meant that we can readily obtain a mass spectrum for the monolayers, along with any molecule that is immobilized onto it. And because MALDI is a mass spectrometry approach, we can track any changes that have been made to the monolayer or molecules bound to it, as long as there is a mass change. In fact, the first example of SAMDI-MS was demonstrated with the addition of a galactose to a surface bound β-GlcNAc by the enzyme GalTase.[13]

After showing the possibility of seeing peptides immobilized to SAMs as well as the ability to monitor enzymatic reactions on the surfaces using SAMDI, the method became one of the major techniques utilized within the group. To this date SAMDI has been used with a wide range of reactions, both enzymatic and chemical, as well as being used for cell adhesion work. [7, 9, 13-24] With SAMDI's utility becoming clear over the years, it was only a matter of time that a realization was made that SAMDI could become the next high-throughput analysis method. Between the clear advantage that it provided with defined SAMs, which enabled clean and selective immobilizations

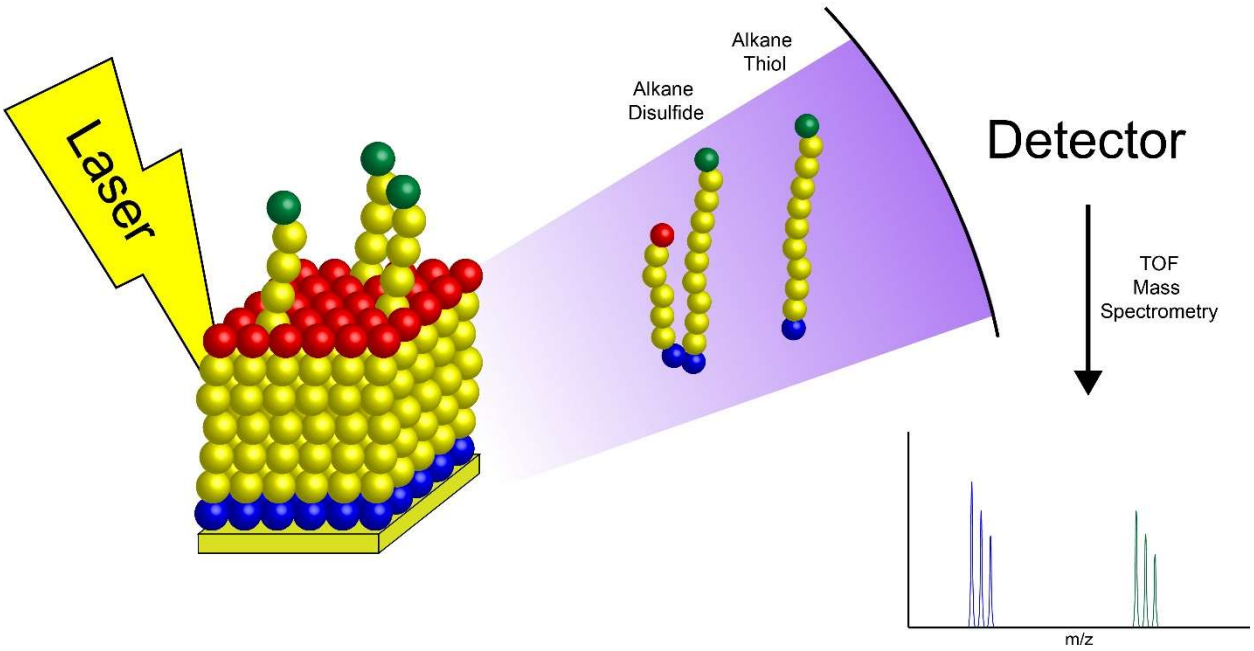


Figure 1.2: SAMDI-MS of alkane disulfides immobilized onto gold surfaces. Alkane disulfides SAMs that are attached to gold are excited by a laser within the MALDI-TOF instrument, this breaks the gold-thiol bond releasing either intact disulfides or individual thiols. These SAMs then fly towards a detector, the time of flight determines the m/z ratio of the molecule.

of molecules of interest, as well as the inert background, and the ability to monitor reactions through mass-change without the need for labels, approaches needed to be developed to increase the number of experiments that could be done with SAMDI. For this purpose, a decision was made to make SAMDI compatible with standard liquid handling robots. To accomplish this, the glass slides with 50 gold- spot, which were first in use when I joined the lab, were slowly phased out for steel plates with 384 gold islands. The 384 SAMDI plates were designed to have the same geometry as standard microtiter plates, enabling the use of these plates with standard liquid handlers. This brought on a paradigm shift in the way the members of the lab thought. The over 7 fold increase in the number of spots, as well as automation provided by liquid handling robots opened up possibilities to run more reactions in a single day, enabled screening of many more reaction conditions.[7] Ultimately though, the increase in throughput meant that more materials were

used. Because each spot of a 384 required a minimum of 2 μL of reaction solution, a decision was made to move to the next higher microtiter plate densities for SAMDI, 1536, which would require 4-fold less volume for each spot (0.5 μL). The increasing in density of spots on the surfaces also had the benefit of decreasing the time it took to analyze 1536 spots. By running 4-equivalents of 384 plates in a single run, the time that was required for loading and unloading of each plate was negated, which could take up to 5 minutes per cycle. With the realization of what increase in density brings, a new directive was setup in the group, to further increase the density of spots on a single SAMDI plate.

To achieve this goal two separate paths were paved. One was to continue increasing the number of discrete spots per plate, and the other, to forgo discrete spots, and develop a way to sample the surfaces. To accommodate the first path, plates presenting 6,144 spots have been made and used within the group (Figure 1.3). However, the use of these plates is not as prevalent in the group as the 1536 plate, due to the lack of reliable liquid handlers which can reliably hit each discrete gold spot. On the other hand, the second path, which removed the need to use distinct gold spots, has seen a lot of development and usage within the lab. This method was to use solid gold surfaces with monolayer, whereupon molecules would be deposited and allowed to immobilize. Following immobilization an imaging MALDI instrument would used to run imaging mass spectrometry (IMS) on our SAMs, a method we termed imaging SAMDI (i-SAMDI). With i-SAMDI it is possible to image spots down to the lowest laser spots provided by the MALDI instruments, 10 μm. It is our belief that by combing i-SAMDI and appropriate liquid dispensing methods, it would be possible to greatly exceed the 6,144 densities possible with discrete spots.

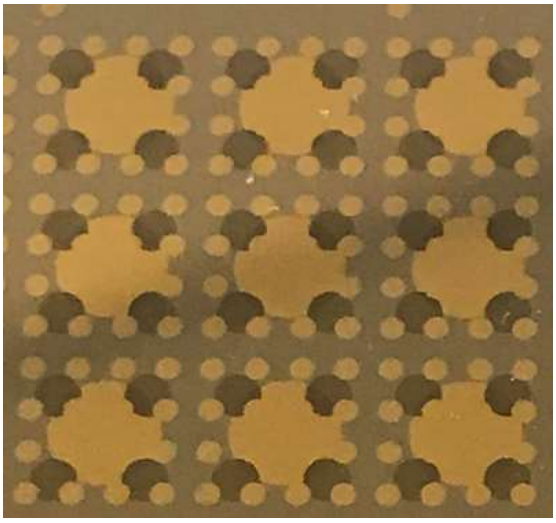


Figure 1.3: Size comparison of 384, 1536, and 6144 density spots on a SAMDI plate. Here we have evaporated all three densities of spots onto a single target plate. The largest circles in gold are the spots of a 384 array. This is followed by the smaller dark spots in 1536 density. Finally, the smallest gold circles are those of the 6144 spots.

Molecular Information Storage

Ever since the development of human society, we humans have been storing information in one way or another. Whether it be on cave walls, papyrus, leather, clay tablets, large monuments, or paper, information storage has been one of the constants throughout our species' history. Nowadays the methods of information storage vary widely in media and sophistication. From book, drawings, magnetic tape, optical discs, hard disks, flash, and cloud storage, each of these methods have their strengths and weaknesses.[25-30] One constant amongst these methods however is their transience. Paper can burn or decay, magnetic tape and hard disks demagnetize, optical devices can scratch, and flash and cloud storage is susceptible to electronic damage. However, nature has had an alternative for information storage for billions of years (that we know of), a method that has been shown to have a long lifetime. This of course is the use of nucleotides to store genetic information in RNA or DNA. Although the exact timeline and method by which RNA and DNA evolved is unknown, the knowledge that DNA holds all the genetic information

of a living organism is one that is known even by the youngest children.[31-36] In fact, humans have been storing genetic data in DNA for thousands of years, though sometimes unknowingly. First, by manipulating the crops and livestock they possess. By crossbreeding and selecting for desirable traits, humans essentially wrote and changed the DNA inside living organisms, modifying species to become something that better suited our needs. It is possible to look back in the archeological and DNA records to see exactly how our storage of genetic information within species has influenced them. Secondly, through genetic manipulation that occurs in many laboratories across the world, we are writing and storing genetic information into organisms, such as bacteria and yeast. Each time a gene is spliced into the genome of a lab organism, we are using methods provided by nature to store information.

However, the same technology that enabled the sequencing and synthesis of genes for splicing into lab based organisms also enabled the use of DNA as digital data storage.[37, 38] By synthesizing DNA of non-sense sequences, genetically speaking, it was possible to store binary and ternary digital information in DNA, and read it back when necessary. By removing the need for a host to carry the DNA, the information that was stored within its sequences was immune from degradation and mutations, given that it is stored correctly.[39] Since the publication of the papers by Church and Goldman in 2012, interest and research into DNA data storage has greatly increased.[40-43] There is even enough interest in the benefits that DNA based storage can provide, that several companies have either started or opened new branches to capitalize on DNA as a storage medium. Two of the largest players in this field are Twist, which is predominantly a DNA synthesis company, which stands to profit from the need of a lot of synthetic DNA to store data with, and Catalog, a company purely entrenched in DNA as a data storage medium.

Although much research has been done in achieving faster reading and writing speeds of DNA, and the longevity of DNA stored under ideal conditions is astounding, cost to synthesize new strands of DNA for each minutia of information is exorbitant, compared to available technology.[44, 45] Thus several other molecular approaches that do not involve DNA have been explored. One such example is the use of non-DNA based molecule polymers, which much like DNA would be read in a typical "beads on a string" manner, and require sequencing and synthesis of each strand to determine the information stored.[46] Others have deviated away from using a "bead on a string" storage method and instead concentrated on storing information using properties of a molecule, such as conductive or magnetic properties, though these approaches require use of complicated systems and conditions not readily achievable in everyday situations.[47-51] Yet another approach that has been proposed is the use of disc-like optical devices, that use fluorescent molecules to store the information.[52-54] In Chapters 2 through 4 I will introduce yet another method that enables molecular information storage by combining peptides and SAMDI-MS.

Peptide arrays

Ever since the introduction of solid-phase peptide synthesis (SPPS) by Merrifield, which drastically reduced the cost and effort needed to synthesizing peptides, the technology has grown in utility and usefulness.[55, 56] The introduction of new N-terminal protection and deprotection schemes, such as tert-butoxycarbonyl (Boc) and fluorenylmethoxycarbony (Fmoc), which enabled faster and milder reaction conditions, and the introduction of wide range of solid supports, attachment chemistries, and an increase in available natural and un-natural amino acids and their derivatives, have broadened the possibilities of peptides that can be synthesized in a laboratory setting.[57, 58] These advances enabled the synthesis and use of large number of peptides, or peptide arrays, with numbers in the hundreds or even thousands. These peptide arrays in turn simplified

the study of various questions previously answered through the use of protein arrays.[59, 60] These include profiling enzymes, antibodies, and ligand-receptor interactions, as well as discovering new cell adhesion ligands, to name a few. [8, 18, 20, 24, 61-70]

However, even with all the advances since Merrifield's original SPPS paper, currently lab made peptide libraries are often limited to numbers in the hundreds, primarily due to cost. Although the cost of a single peptide is relatively low, making an arrays of thousands or even tens of thousands quickly become prohibitively expensive and extremely time consuming.[14] Of course, commercial sources exist, which can minimize cost and eliminate personnel time spent on synthesis. These sources provide peptides in several forms, including surface bound SPOT microarrays by PEPperPRINT, a format similar to DNA microchip arrays, or more familiar lyophilized powders in microwells.[71] Both commercial sources of peptides have their advantages and drawbacks. For example, the microarray formats provide fast synthesis with specially built inkjet printers, however the arrays are single use and must be discarded after use, and any reaction must be done on the solid support surfaces, which can limit possible reactions and analysis methods. On the other hand, the lyophilized peptides are slower to synthesize, but can be used multiple times and not limited in their usage method.

Yet another method for generating large peptide libraries is through the use of phage displays.[72] Although predominantly used for expressing full proteins of interest on the surfaces of bacteria phages, methods for synthesizing peptide libraries have also been developed. By inserting genetic information into the phage coat protein gene, many different peptide sequences of interest can be generated on the surfaces of the phages. It has also been shown that through random mutation of the genes that are being inserted, it is possible to synthesize peptide libraries with

random sequences, which enables the expression of many related but different sequences of peptides.[73-76] However, as the phage display method utilizes a living organism for the synthesis and display of these peptides, the randomness of the presented peptides have come into question.[77]

We in the Mrksich group predominantly synthesize our peptide in a one-at-a-time fashion using standard SPPS methods, though parallelization is employed to increase the throughput. And through the synthesis of these peptide arrays and combining them with SAMDI-MS we have shown the usefulness of peptide arrays to monitor many different reactions. [8, 18, 20, 21, 23, 61-65, 78] In this thesis we will further expand on the work of using peptide arrays with SAMDI-MS and demonstrate two novel uses for peptide arrays in Chapter 5 and 6. Furthermore, in Chapter 7 we will demonstrate a method for synthesizing very large random peptide arrays, which through the use of a novel capping scheme and SAMDI-MS can be fully decoded and used for any SAMDI experiment.

Part 1:

Information Storage Utilizing Peptides and SAMDI-MS

Chapter 2: Storage of Information Using Small Organic Molecules

This chapter is adapted from the following published work:
Cafferty, B.J.; Ten, A.S.; Fink, M.J.; Morey, S.; Preston, D.J.; Mrksich, M.; Whitesides, G.M., *"Storage of Information Using Small Organic Molecules"* ACS Central Science, 2019. **5**(5): p. 911-916

Introduction

Technologies from printing with ink on paper, to very sophisticated electronic, optical, and magnetic methods, are used to store information. The importance (across a range of parameters: cost, space, energy use, rate of reading and writing, rate of degradation on storage, potential for corruption through tampering, independence of protocols and hardware for reading) is such that each of these methods has weaknesses in addition to its strengths,[25-30] and there remains a need to evaluate possible alternatives.[79-83] New methods would not necessarily replace the amazingly highly engineered methods currently used, but might circumvent some of their weaknesses, and perhaps open new applications.

The use of molecules for storing information[46-48, 84, 85] has in large part been stimulated by the ability of cells to store very large amounts of information in molecules (especially macromolecules: DNA, RNA, proteins, and carbohydrates) and metabolic networks. Most macromolecules use a common strategy of ordering the information along a one-dimensional array of covalently linked monomers ("beads on a string"). They have also raised enormously interesting question about the meaning of "information" in living cells: for example, is the cell a Turing machine?

Research is developing strategies that use high molecular weight, biologically-derived systems, largely based on the sequence of synthetic DNA strands.[37, 38, 43] Our objective is to explore a different strategy—one not modeled on biology—that uses *low* molecular weight molecules. We

especially wished to avoid macromolecules that use (often repetitive) organic synthetic steps, and that require the synthesis of a unique macromolecule for each separate message. We have instead used sets of peptides having distinguishable molecular weights to store information. Overall, this system requires a set of a maximum of eight peptides, as a mixture, in a microwell, to store one byte, and a mixture of 32 peptides to store four bytes. Using larger mixtures of peptides enables storage of larger sets of data. These systems are capable of writing any arbitrary binary information using the same set of small molecules. In this work, reading is accomplished by identifying the masses of the molecules that are immobilized to a self-assembled monolayer (primarily as disulfides from the laser desorption process) using mass spectrometry. MS provides both high precision (enabling accurate determination of the composition of mixtures of peptides in a single sub-millimeter spot of an immobilized array, without separation, and with few errors) and high rates of reading.

Results

Our initial demonstration has been to write messages in eight-bit ASCII code, convert them to an equivalent molecular code, store them on an array plate (four bytes per spot), and read them using SAMDI (self-assembled monolayers for matrix-assisted laser desorption/ionization) mass spectrometry.[86] ASCII (American Standard Code for Information Interchange) is a look-up table that includes the alphabet, numbers, punctuation, and special characters—a maximum of 256 characters—and is used primarily for alphanumeric text. Table 2.1 summarizes this strategy for the letter “K,” and Appendix Table 2.1 summarizes a complete assignment of peptides sufficient to encode four bytes in a single mixture, with their assignments to a binary molecular representation.

To differentiate the two systems with which we are working (*electronic* storage and its theoretic foundation in Boolean algebra, and *molecular* storage), we call the equivalent of a bit, and of an eight-bit byte, of information—in the form of mixtures of molecules—a "molbit" and a "molbyte." To store information in molecules, we first designed a method that would allow us to encode ASCII in molecules distinguishable by mass spectrometry. For example, the letter "K" in ASCII is represented by one byte (01001011) in binary. We convert that binary representation to a molecular one by assigning a peptide to each of the eight bits in a byte, and include that peptide on the spot if the bit value is "1" and omit it if the bit value is "0" (Table 2.1).

These peptides were selected to have four characteristics: i) All were resolvable by mass using SAMDI as components of a common mixture (Figure 2.1). The different amino acids in each peptide are covalently bonded, but their order is not relevant—only the total mass. The peptides are *not* covalently bonded to one another, and do not form macromolecules. Information is thus stored as mixtures of low molecular weight (MW < 1,000 g mol^{-1}) molecules, in arrays, specifying "1" and "0" in a binary representation, rather than as a sequence of groups in a linear polymer. ii) All peptides terminate in a cysteine to allow efficient immobilization by Michael addition to the reactive maleimide group present in the 1.25-mm diameter spot of the SAMDI plate. iii) Each peptide includes a trimethyllysine (K^{Me3}) with a fixed positive charge to aid in mass spectrometry (positive mode). By using the set of 32 peptides listed in Appendix Table 2.1, each of which is distinguishable in a mixture containing the others, we can store the information for four molbytes (e.g., four letters in ASCII) in one spot.

Using this method, the *presence* of a particular peptide in a mixture indicates three parameters: i) The byte (out of the four bytes, when 32 peptides are used) to which it is contributing

Table 2.1: Correspondence of an alphanumeric character (the letter "K") encoded in ASCII in binary, and in four molbits as peptides.

Byte									
I	Letter	K							
	Binary String in ASCII[†]	**0**	**1**	**0**	**0**	**1**	**0**	**1**	**1**
	Molbit Assignment to Binary String		Ac-ZK$^{(me3)}$C			Ac-GVK$^{(me3)}$C		Ac-ALK$^{(me3)}$C	Ac-GFK$^{(me3)}$C
	Peptide Number[‡]	_	**2**	_	_	**5**	_	**7**	**8**
II[*]		_	**10**	_	_	**13**	_	**15**	**16**
III		_	**18**	_	_	**21**	_	**23**	**24**
IV		_	**26**	_	_	**29**	_	**31**	**32**

Ac: acetyl, A: alanine, G: glycine, C: cysteine, F: phenylalanine, K$^{(me3)}$: N$^{\varepsilon}$,N$^{\varepsilon}$,N$^{\varepsilon}$-trimethyl lysine, L: leucine, V: valine, Z: 2-aminobutyric acid.
[†]See Table S2 for binary assignment of printable character in ASCII.
[‡]Numbering scheme for the 32 peptides used in this work (Table S1 gives complete sequences).
[*]The peptides that correspond to the molbits that represent the letter "K" in bytes II, III, and IV are also shown.

information; ii) its location (which is assigned based on the molecular weight of the corresponding peptide) in the bitstring of that byte; and iii) its value ("1"). The *absence* of that peptide indicates that that position in the molbyte is "0". The presence of the four peptides listed in Table 2.1 are thus assigned to bits with the value 1 and the four peptides absent from the mixture are assigned to bits with the value 0. The one remaining parameter to be defined is the position of this byte (or bytes when more than eight molbits are used per spot) in the sequence of the entire message: this information is provided by the position of the spot in the sequence of spots on the SAMDI array plate. The attractive feature of this method is that only eight peptides allows the specification of all of the characters of one byte, and thus allows an arbitrary message to be written in ASCII (or any character set of 256 members); by using 32 distinguishable peptides we can specify four bytes in one spot.

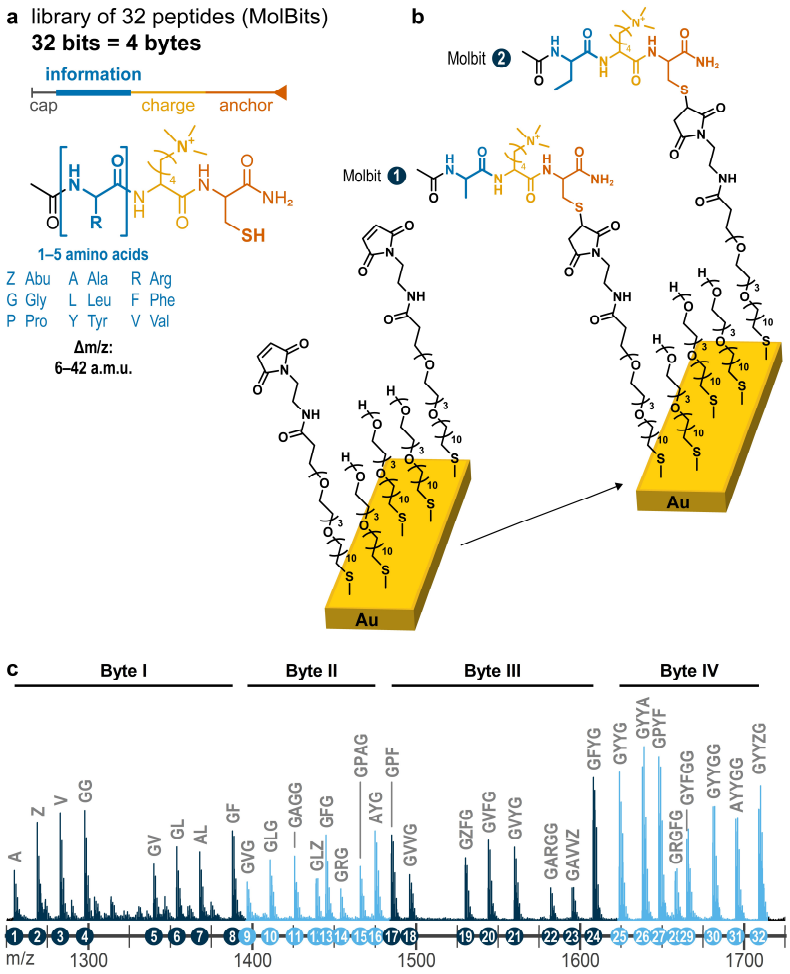


Figure 2.1: Design of peptide molbits and spectrum of all 32 molbits in a single mixture. (a) Peptide molbits contain an information region that consists of one to five amino acids (chosen from 2-aminobutyric acid, alanine, arginine, glycine, leucine, phenylalanine, proline, tyrosine, valine), which provides a distinguishable mass-to-charge ratio for each peptide (a difference of 6-42 a.m.u.), a charge residue (trimethyl lysine), and an anchor residue (terminal cysteine). The N-terminus is capped by an acetyl group for chemical stability. (b) Schematic showing an example of the immobilization of two peptides (corresponding to molbit 1 and molbit 2 in panel c) to a maleimide-terminated monolayer for storage. Prior to conjugation of peptide(s), the monolayer consists of a mixture of triethyleneglycol undecanethiol (EG_3-capped alkanethiol) terminating in either an alcohol or maleimide. (c) A spectrum of a SAMDI spot containing all 32 molbits; the intensity was normalized to the highest signal. Peptides were grouped by molecular weight into sets of eight, representing a byte of information (4 bytes total). The single-letter codes of residues in the information region are listed above each peak in the mass spectrum (see Appendix Table 2.1 for full list of peptide sequence and corresponding masses). The observed masses are for mixed disulfides derived from a EG_3-capped alkanethiol and the peptide conjugated to a maleimide-terminated EG_3-capped alkanethiol.

Figure 2.2 outlines the process we used to 'write', 'store' and 'read' text using this set of 32 peptides. For a particular byte, the appropriate set of peptides representing '1's in the bitstring is deposited and mixed in wells of a 384 well plate using an Echo® 555 liquid handler. A Tecan® liquid handler than transfers these mixtures to an array plate having 1,536 gold islands ("spots"), each presenting a self-assembled monolayer. The peptides react covalently with the terminal maleimide groups present on the monolayers of the array plate. Covalent coupling prevents the components of the mixture from spreading on the surface and allows their analysis with SAMDI mass spectrometry. The plate, with the completed text encoded as mixtures of peptides in spots ordered on the plate, is stored. Reading by SAMDI is as described previously.[86]

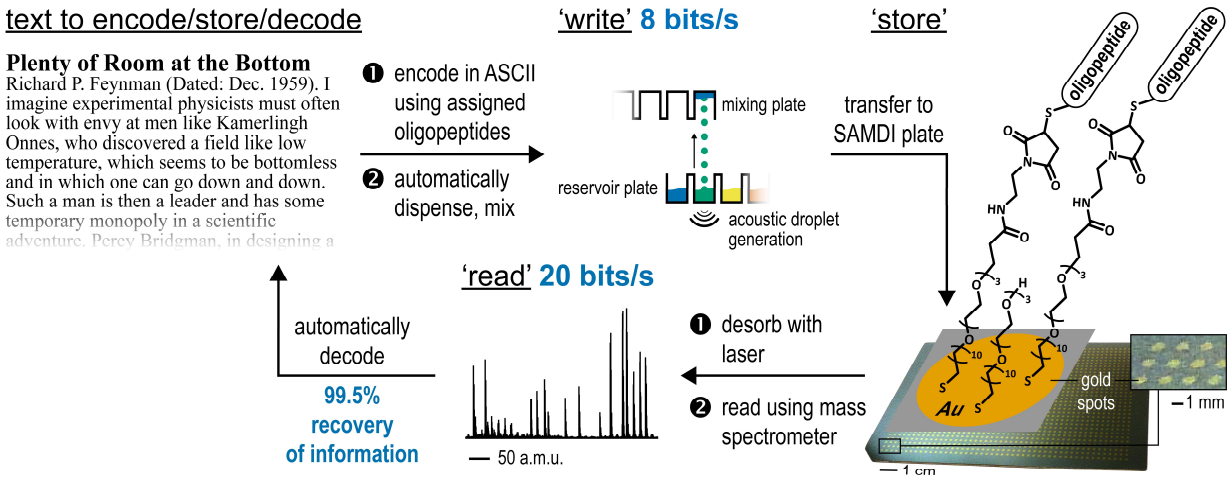


Figure 2.2: Overview of 'writing' using mixtures of molbits and 'reading' using SAMDI MS. 'Writing' is performed by first translating information (here, the alphanumeric characters of Feynman's lecture "There is plenty of room at the bottom") into binary. Binary information is converted to peptides immobilized on a self-assembled monolayer, for storage. A MALDI-TOF mass spectrometer analyzes ('reads') these plates. A program decodes the information in the spectra and generates a bitstring that is used to regenerate the original text. Recovery of information was determined by (number of correctly identified molbits)/(total number of molbits) × 100.

This strategy for writing and reading bytes allows a small number of low molecular weight molecules to encode many forms of information and, once synthesized, avoids the need for further synthesis to store a new message. (In this demonstration, to order these molbytes, we use an array

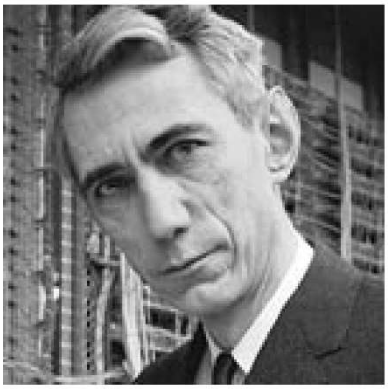
plate in the format of a conventional microwell plate, but a number of other formats are also possible). The density of information (D) we can put on a plate depends on the representation, but here is given by D = (molbytes/cm^2) = (wells/cm^2)/plate)(molebyte/well). For the current system, this number is D = 64 bytes/cm^2.

Our examples here include text (Appendix) and JPEG images (Figure 2.3). The procedure we use is operationally simple. The small number of molecules required (within a given set such as peptides) need only be synthesized once (or, more probably, purchased, since there are many custom commercial suppliers), and serve to encode a very wide range of information. The text of Feynman's famous lecture "There is plenty of room at the bottom" is a demonstration of current capability (the text and errors are shown in the Appendix). It was written, stored, and read with 99.9% recovery of information. This text (38,313 bytes or alphanumeric characters) was written and read using one set of devices (Figure 2.2) in 20 h. The images (Figure 2.3) are another. The speed of 'writing' is 8 bits/s, and 'reading' is 20 bits/s, without parallelization. This process is obviously amenable to simple linear parallelization, particularly since each line of instruments could be writing different information at the same time, using a shared set of molecules for storage: the speed could thus easily be increased by a factor of ten or more, albeit at ten times the capital cost.

Discussion

This paper demonstrates one method of encoding information for storage in molecules. It represents one of two limiting strategies for this purpose. The first (this method) encodes information in simple, separate molecules that are designed to minimize synthesis, and to fit naturally as the molecular equivalent of existing methods of storing digital electronic information.

a



b

Figure 2.3: JPEG images encoded using mixtures of molbits. (a) Image of Claude Shannon, "the father of information theory," encoded in 6,094 molbytes, and (b) image of the woodblock print entitled *The Great Wave off Kanagawa*, by Hokusai, encoded in 5,953 molbytes. Each image was encoded on a single plate and decoded with zero errors (100% recovery). Images were encoded as JPEG files as described in the methods section. See Appendix Figure 2.1 for an example of the resolution of an encoded JPEG file. Photograph of Claude Shannon by Alfred Eisenstaedt / The LIFE Picture Collection / Getty.

It is not intended (at least at this stage in development) to compete with existing electronic, optical, or magnetic methods of storage. Instead, its immediate objective is to provide an alternative method for archival storage that is stable for long times, does not require energy for storage, and

is secure. The molecules are used as mixtures, and ordered both by physical location (on an array plate) and by mass of the molecules within each spot of the array. This method is designed for flexibility in writing and use, for simplicity, and for long-term stability in storage. It enables the encoding of *any* binary data, using the same procedures and a constant set of small molecules (which could easily be available on a multi-kg scale). No additional synthesis is required for each new message. It depends entirely on simple physical manipulations for sampling, liquid transfer, mixing, separation, and reading at all of its steps. For reading, it uses (in the demonstration shown here) a mass spectrometer—a technique that provides dramatically more information than reading charge on capacitors. The sensitivity in MALDI-TOF MS is highly dependent on the specific analyte and on sample preparation and have not been rigorously determined for monolayers of alkanethiolates on gold. In favorable cases, however, as few as thousands of analyte molecules are sufficient to produce a spectrum.[87] The capability of mass spectrometry continues to increase rapidly, and thus will enable the further extension of these techniques. The higher the resolution of the spectrometer, the more complex the mixtures that can be analyzed, and thus the greater the amount of information that can be stored per array.

One method of distinguishing different procedures for storing information is the number and density of "locations" (that is, places where information can be stored: i.e., capacitors in solid-state devices, grooves in optical disks, magnetic domains, visible letters), the amount of information that can be stored in each location, and the cost and time of writing, storage, and reading. For electronic storage, transistors are very small (approx. $10^{11}/cm^2$ at the current 11-nm minimum feature size) and inexpensive, but they store only one bit. Methods for storing information in molbytes—as outlined here—offer a high density of information per location, but—using currently available plates—a density of locations that is modest. Molecular storage by this

method will improve rapidly with advances in technology for spotting. Higher density of spots in arrays and faster liquid transfer could be achieved by inkjet printing.[38, 88] For example, inkjet printers can generate drops at rates of ~1,000 per second and position drops on a surface with center-to-center distances of 10 μm.[89] Using a set of 32 molbits (as described in this work), this spotting diameter would give a density of information of 4 MB/cm^2. Optimization of inkjet printers for the type of molecular ink used and speed would further increase the density and rate of writing information. We have not yet analyzed and redesigned this system for efficiency.

The current demonstrations have used peptides, but many other classes of organic molecules (additional unnatural amino acids, fatty acids, aromatics including heterocycles, saturated terpenes, and others) are also possible: the method thus has broad scope. Although we have designed the current system for simplicity, the combination of molecular design, organic synthesis, and advanced methods of separations and analysis also has the potential greatly to increase the amount of information that can be stored per molecule and per location (e.g., spots, wells).

Choosing classes of molecules for information storage that offer long-term stability, with no energy required for storage, is one long-term objective of this area of research. Long-term stability of appropriate organic molecules with appropriate structures over hundreds of years has not been systematically explored, but is commonly assumed. Peptides have stabilities of hundreds or thousands of years under suitable conditions;[90] i.e., in the absence of light (or ionizing radiation), oxygen or other oxidants, and high temperatures, and possibly in the absence of water, in inert containers. Importantly, occasional breaks in individual molecules would (unlike breaks in DNA) not significantly damage the fidelity of reading, since they would appear at masses that are not

coded by the molbits. Molecular storage of information should be especially resistant to tampering electrically, magnetically, or optically, since the only way to read or rewrite the composition of information stored molecularly would be to access the molecules physically, and then to do chemistry.

A second strategy for storing information in molecules is using organic polymers (DNA, synthetic polymers, proteins, oligosaccharides).[43, 46] These methods —in principle—have other attractive characteristics. DNA has the specific ability (and the requirement) to order the information in a message in terms of the position of its nucleotide components along a covalently linked chain ("beads on a string"), and thus does not require spatial ordering, but it has the accompanying substantial disadvantage of having to synthesize a new strand and sequence of DNA for each different message. DNA also offers the potential of low cost for reading (using Gen 4 and future methods), and amplification by replication (albeit with unknown problems, for long unnatural and non-biological sequences).

Although methods of synthesis are improving,[91] synthesis of DNA (particularly of long strands) by chemical synthetic methods remains slow and expensive. Cycle times for the coupling of nucleotides using phosphoramidite chemistry[92] (the most common approach used for synthesis in DNA-based storage systems) are on the order of ten minutes. This amount of time, which does not account for additional processing of the DNA strands (cleavage from support and deprotection of nucleobases), is equivalent to a rate of writing of 0.001 bits/sec—a rate that is significantly slower than the (unoptimized) approach described here (8 bits/sec). The cost of writing a bit of information using DNA as a storage medium has been reported to be as low as $\$5\times10^{-4}$,[45] which is $\sim1\times10^{8}$ times more expensive than the price to store a bit in a hard disk drive.[93] The peptides

used in this work are relatively expensive (~$\$1\times10^{-3}$ per bit) because they were custom-synthesized, but they are not intrinsically expensive, even at the multi-kg scale. A significant advantage in cost for the approach described here over approaches that use DNA (or other sequence-controlled polymers) is that inexpensive commodity chemicals can be used as information carriers (for example, we estimate that using alkanethiols as molbits would reduce the cost to below $\$1\times10^{-10}$ per bit). The price of liquid handlers and analytical instruments (here, MALDI-TOF MS) required for this approach can be expensive but they need only be purchased once.

It is too early in the development of either strategy, or of others, to compare them in specific applications. Their ultimate applications may also be quite different. It is also impossible to compare them with the well-established electrical, optical, and magnetic methods, which are the products of exceptionally successful, multi-decade programs in technology development. It is, however, clear that the chemistry of small molecules, and the analytical and synthetic methods that have been developed for synthesizing, separating, and identifying them, offer an exceptionally rich array of scientific and technological methods to apply in new approaches to information storage.

Conclusion

We have in this chapter shown that it is possible to use 32 peptides with distinct masses to encode 4 bytes of information for every SAMDI spot. We have demonstrated this method by mixing, immobilizing, reading, and decoding peptides on SAMDI surfaces and were able to read 400,000 bits with greater than 99.9% accuracy. We believe that this could be a viable method to store information on molecules.

Methods

Preparation of Solutions of Peptides (molbits). Peptides were synthesized using standard Fmoc chemistry on rink-amide resin and purified by HPLC. Stock solutions of each peptide were made in 0.1% TFA with DI water and stored at 4 °C. To prepare the peptides and peptide mixtures for immobilization, each peptide stock solution was distributed into a source plate. Mixing of peptides to form binary data sets was performed using these peptide stock solutions and an Echo® 555 (Labcyte Inc.) liquid handler, with the final concentration of each peptide, when present, at 20 μM (some sequences had to be diluted further to maintain comparable ionization to the other analytes). A Python program written in-house was used to assign peptides from alphanumeric character inputs (translated to ASCII) and bitstrings.

Generating Input Tables for Automated Encoding of Text. To generate an input table for alphanumeric text for the Echo® 555 liquid handler, a given text was first divided into sections of 6,144 characters (the maximum number of characters that fit on SAMDI 1,536-spot target plate). These blocks of text are then run through a program ("Text Split"), which further divides the 6,144 characters of each block into four sections of 1,536 characters. Each section of 1,536 characters was then assigned to a 384 well plate, with 4 characters (bytes) per well, and a text file (extension .txt) was generated containing the string of characters for each well plate. This file was then used in the program titled "Molbit Encoding". The program also requires inputs for the volume for each stock solution of peptide to be transferred (in nL), the total capacity per source well (the location of a given peptide to be transferred), the name of the destination plate, and a list of the ASCII binary combinations for each of the characters used (see Appendix Table 2.2 for list of printable ASCII characters and their representation in binary). Once it receives the required inputs, the program matches each character in the .txt file to the appropriate binary ASCII combination and

generates an input table for the Echo instrument, including information on source well, transfer volume, destination well, and destination plate name.

Generating Input Tables for Automated Encoding of an Arbitrary Bitstream. To generate an input table for non-ASCII data for the Echo® 555 liquid handler, a bitstream was first generated. The bits were then sequentially numbered 1 through 32. After this process the "Vlookup" function in excel was used to assign a predefined source well for each number. Each group of 32 bits was next assigned with a well of a 1,536-well destination plate. The bitstream, with each entry's associated bit number, source well, and destination well, was then reduced to include only those entries with a bitstream value of 1. Next we used the "Vlookup" function to assign the transfer volume for each entry, based on the source well. Finally, these entries were transferred into an Echo input table, with information on source well, transfer volume, destination well and destination plate name.

Automated Encoding via Liquid Transfer. Prior to initializing a run on the Echo® 555 liquid handler robot, a source plate (Labcyte Echo Qualified 384-well plates, Cat #: PP0200) was prepared with the desired peptides to be transferred. Each well of the source plate contained 65 μL of each of the 32 stock solutions (2 mM in peptide). The number of wells needed for each peptide can be determined from the input table generated via the encoding program. The source plate and destination plate (Greiner BioOne 384-well plates Cat#: 784201) were placed in storage towers in the Access Laboratory Workstation attached to the liquid handler. To initiate the run, the input table was imported, which defines the locations of the source and destination plates, and the protocol was executed. Once the peptides were transferred, the destination and source plates were

covered with lids (Labcyte MicroClime Environmental Microplate Lid Cat#: LL-0310) to ensure that the contents of the plates did not dry.

Preparation of Monolayer Arrays[86]. Array plates with 384 and 1536 gold spots on steel plates were soaked in a solution of a mixture of EG3-capped alkane disulfide and a mixed disulfide of EG3-capped alkanethiol and a maleimide-terminated EG3-capped alkanethiol for 24 h, at room temperature, to allow formation of a self-assembled monolayer on the gold surface. The solution of disulfides contained an overall concentration of 1 mM of the two monolayer compounds in a stoichiometric ratio (2 to 3) to yield a monolayer wherein the maleimide groups are present at a density of 20%. Following monolayer formation, the plates were soaked in a solution of hexadecyl phosphonic acid (10 mM) for 5 minutes, and rinsed with ethanol, water, ethanol, dried with nitrogen and stored dry under vacuum. SAMDI plates were used within one week of forming monolayers.

Immobilization of Peptides onto Plates. Prior to immobilization, the peptide mixture plates generated by the Echo® 555 liquid handler were filled with 4 μL of 100 mM Tris buffer at pH 8.0, with a ThermoFisher Multidrop Combi, to ensure the solutions of mixed peptides are at the correct pH and appropriate concentration for conjugation to the monolayer. Each set of four 384-multiwell plates were then transferred to a 1,536-spot SAMDI plate functionalized with 20% maleimide and displaying a hexadecyl phosphonic acid background between spots. Samples (0.75 μL) from each well of the 384-multiwell plate that contained solution were transferred onto the 1536-spot SAMDI plate utilizing the TECAN Fluent/Freedom Evo instruments, with a MCA 384 head utilizing 15 μL tips, such that each 384-multiwell plate is transferred to one quadrant of a 1536-spot SAMDI plate. In this way the spots are read left to right and top to bottom, and allow the original encoded

text to be read. Once transferred, the peptide solutions react with the maleimide groups on the surface of the plate for 10-30 minutes, in a humidified chamber, to covalently immobilize the mixture of peptides. After immobilization, the plate was washed with ethanol, water, ethanol and dried under a stream of nitrogen.

MALDI-TOF MS Analysis. SAMDI plates with immobilized peptides were first treated with 2',4',6'-trihydroxyacetophenone matrix solution (THAP, 12 mg/ml in acetone) and then were loaded into an ABSciex TOF-TOF 5800 instrument. Matrix-assisted laser desorption/ionization time-of-flight mass spectra were collected for each spot in positive mode with the instrument setting of 700 shots/spectrum, 5300 laser intensity, stage velocity of 1500 μm/s, 0.61 digitizer setting, and a laser pulse rate of 400 Hz.

Analysis of Apectra with Program. Prior to analysis of the SAMDI spectra, an input table was generated containing the peptide mass combinations for each of the 95 printable ASCII characters used for each of the 4 bytes (Appendix Table 2.2). This input table was then divided so that each contained only the peptide combinations for the corresponding byte. This division was done using the "Molbit Decoding" program along with an input of the 95 ASCII characters in quadruplicate, once per byte, and a list of the peptides for each character and byte. The SAMDI spectra were exported from the instrument computer and analyzed using the "new profiler" program. This program required the following inputs to run; location of the mass spectrum files, location for the output of generated files, an input table for the byte (1-4) being analyzed, as well as the background threshold. The background threshold is a user-determined value; it is based on the absolute peak intensity relative to the highest peak in the spectrum and is usually set between 20-30%. The background threshold helps avoid false positives in detecting presence of molbits due to the noise

in the spectra. The program functions in the following way. It first scans the spectrum and identifies the maximum intensity value (arbitrary units) and sets this value to 1. It then converts each of the other intensities to relative intensity units based on this parent value. The software then removes any value below the threshold set by the user and generates a new list containing only those peaks remaining above the threshold. Following the generation of the new list, it sums the values of the intensities by rounding to the nearest integer mass value. It then attempts to generate groups of masses based on the two highest consecutive intensity units, followed by single mass intensity groups that cannot be combined. At this point the program scans the input table to find an entry that provides the highest sum of intensities based on mass groups present. Once it finds the entry, it returns the value for the character for which it has decoded. If it fails to match an entry in the input table it will return a "FAILED" response and move on to the next spectrum. Once the software has finished running through the entire dataset, it produces a file that lists the label of the data spot, the decoded character (if applicable), as well as the masses that were identified for that character. Recovery of information was determined by the number of correctly identified molbits by spectral analysis, divided by the total number of molbits originally encoded, multiplied by 100.

Image Compression, Encoding, Storage, Retrieval, and Reconstitution. First, if the original copy of an image was larger than the storage space available on one SAMDI 1,536-spot plate (6,144 bytes), that image was compressed, via the JPEG algorithm, to fit on one well plate. The JPEG algorithm was implemented with Adobe Photoshop CS4, version 11.0, with the JPEG quality and blur settings indicated in Appendix Table 2.3 using the "Save for Web and Devices" function. After compression, the JPEG files were encoded as bitstreams using the program titled "Image Encoding", run in Matlab R2015b. The code reads the bytes stored on the local computer hard drive that comprise the JPEG file, and converts these bits to a bitstream. The length of the

data contained in the bitstream, in bits, was also read by the code and prepended (as a 16-bit segment) to the front of the bitstream, which was then encoded onto the well plate using the automatic molecular encoding process described above. Retrieval of data from the well plate was performed as described above, where the output from reading the SAMDI plate was a bitstream. This bitstream, in the form of a text (.txt) file of "1" and "0" with no other characters, was read by a program titled "Image Extraction", which extracts the length of the image file from the first 16 bits of the bitstream and then retrieves that quantity of bits from the bitstream, starting at the 17th bit (after the string of bits that records the length of the file). This image data was reconstituted into an image file in JPEG format which can be interpreted and displayed by a computer. The error rate during retrieval and reconstitution of each image is included in Appendix Table 2.3.

Chapter 3: Non-Binary Storage of Information in Mixtures of Small Molecules

The work in this chapter was accomplished with the help and input from Michael Fink and Brian Cafferty

Introduction

The use of molecules to store information now offers a new opportunity for archival data storage.[37, 38, 46-48, 84, 85, 94] The first demonstrations used long DNA strands where the presence of certain nucleobases correspond to a '0' or '1' in binary encoding.[37] These methods have advanced substantially, but still have the limitation that they require significant cost and time to synthesize DNA each time information is stored.[45] Chapter 2 demonstrated the use of mixtures of peptides to store information. In this approach, each peptide represents a bit—which we refer to as a molecular bit, or a 'molbit'—and the presence and absence of each of eight peptides within a mixture is directly associated with a byte—a sequence of eight molbits and again, termed a 'molbyte'. In this way, the eight peptides can be synthesized once and then arranged into mixtures that are arrayed onto a self-assembled monolayer. The molbytes can be read using mass spectrometry to identify which of the peptides are present at each location on the substrate.[86] Because mass spectrometry can measure the relative amounts of the molecules within each spot, we reasoned that it should be possible to use higher base storage—that is, instead of the absence or presence of a molecule corresponding to a 0 or 1, different intensities of the peak can correspond to each of several levels of molecule and give a more efficient storage density (Figure 3.1).[95-97] Here, we demonstrate this concept and show that storage of molecules in base 8 is feasible.

In Chapter 2 we demonstrated using peptides to encode each molbit. We designed a set of peptide sequences such that the masses were appropriately spaced and to include a terminal

cysteine residue to allow immobilization to a self-assembled monolayer of alkanethiolates on gold. In this way, we could assemble mixtures of peptides that corresponded to a molbyte in a multi-well plate and simultaneously transfer the solutions to a steel plate having an array of 1,536 gold islands that each was modified with a monolayer presenting maleimide groups against an inert background of tri(ethylene glycol) groups.[3, 9] Once immobilized, the monolayers are rinsed and stored. The information can later be accessed by analyzing the plate with matrix-assisted laser desorption-ionization mass spectrometry (in a technique we term SAMDI MS) to generate spectra for each of the gold islands.[13] The spectra revealed which of the eight peaks—which correspond to each of the peptides—were present and therefore the identity of the molbyte. The identification of a peak required that it was above a threshold intensity (that corresponds to baseline noise), and where the molbit is recorded as a '1' and if it is below this threshold, it is recorded as a '0'. In this work we demonstrate that using the aforementioned molecular storage technique, we can expand the data storage into higher base systems.

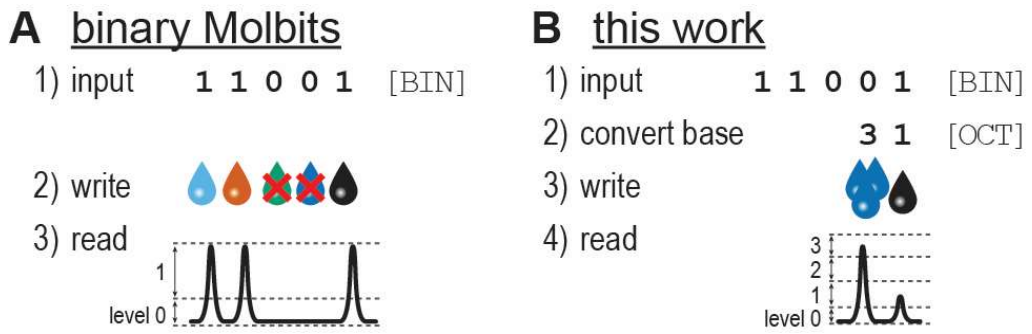


Figure 3.1: Basic concept of binary vs. non-binary molecular data storage. A) In the binary method, the presence of absence of a specific peptide determined if the bit corresponding to that peptide was either a '1' or a '0'. B) In this method, the concentration of the peptide present determines the level of that peptide in the specific base system.

Results and Discussion

Storage of Information in Quaternary (Base-4): We first demonstrate storage in a base 4, quaternary, system using concentrations of peptides. For this work eight peptides were chosen, to directly compare a single byte in binary to other base systems. We also included a 9th peptide at a constant concentration to serve as a reference peptide against which the areas under the curves, and thus concentrations, of the others are measured. Mixtures of peptides were assembled in a 1536 well plate, where each peptide could be present at one of 4 concentrations; 0, 7, 15 and 30 μM, and where each concentration corresponds to a molbit value of 0, 1, 2 or 3, respectively. These concentrations were chosen by simply halving the concentrations between each level. We also prepared 1536 SAMDI plates by evaporating gold islands onto steel in a standard 1536 microtiter format. Self-assembled monolayers (SAMs) of alkanethiolates terminated in maleimide groups, at a density of 20% against a background of tri(ethylene glycol)-groups, were assembled by soaking the SAMDI plates in a SAMs solution.[9] The maleimide functionality allows for the selective immobilization of cysteine-terminated peptides, while the glycol background prevents non-specific adsorption of proteins to the surface.[23] We then immobilized the mixtures of peptides onto the SAMDI array plate and analyzed the array with SAMDI-MS. The average area under the curves (AUC), relative to the internal standard AUC, for each peptide, at each level were determined. The ranges of each level were then set by simply taking the midpoints between adjacent levels. These values were then set as the reference points for all future observations of the quaternary system (Figure 3.2).

Higher than Base-4 Information Storage with Peptides. Next, we evaluated still higher base systems, with the goal of understanding the trade-off between data density and error rate. The two additional base systems chosen were octal (8) and hexadecimal (16). As with the quaternary system

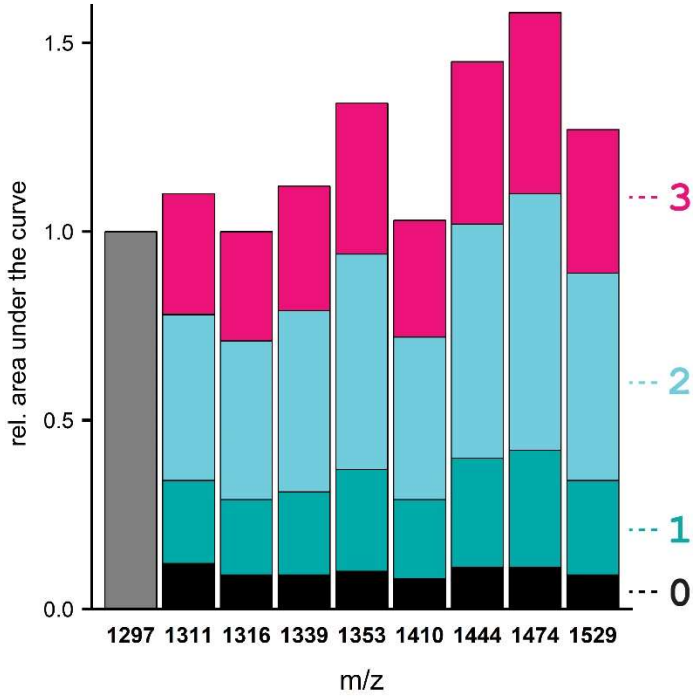


Figure 3.2: Calibration values for each peptide for the quaternary base system. The average relative area under the curve value for each level, as determined from experiments, was set as the midpoint of each level, and the midpoint between adjacent levels was set as the boundary between each level.

the equivalent of a binary '1' was assigned to a '7' and 'f' in the octal and hexadecimal systems, respectively. Next, we determined the concentrations associated with the additional levels, '1' – '6' for octal and '1' – '9', 'a' – 'e' for hexadecimal. Care had to be taken when considering the concentrations of each peptide to use per level, due to the way the liquid handling instrument, Labcyte Echo 555, dispenses liquids. The instrument transfers liquids in multiples of 2.5 nL, and each drop has an associated error, albeit small, thus it was determined that a top-heavy distribution would be best. In other words, having larger separation of concentrations for higher levels. Adopting a top-heavy distribution allows for greater difference in concentration between adjacent levels, where the compounding error from dispensing increasing number of drops would be most evident. Table 3.1 show the levels for each base system, along with their transfer volumes and concentrations of peptide. Following the determination of the peptide concentrations per level,

Binary			Hexadecimal		
Level	Volume (nL)	Concentratoin (μM)	Level	Volume (nL)	Concentratoin (μM)
0	0	0	0	0	0
1	75	30	1	5	2
Quaternary			2	7.5	3
0	0	0	3	10	4
1	17.5	7	4	12.5	5
2	37.5	15	5	17.5	7
3	75	30	6	22.5	9
Octal			7	27.5	11
0	0	0	8	32.5	13
1	5	2	9	37.5	15
2	12.5	5	a	42.5	17
3	22.5	9	b	47.5	19
4	32.5	13	c	52.5	21
5	45	18	d	60	24
6	60	24	e	67.5	27
7	75	30	f	75	30

Table 3.1: Transfer volumes and concentrations of peptides for each base system and level.

appropriate amounts of each peptide were assembled in a 1536 well plate and the mixtures were immobilized onto the monolayers. The immobilized mixtures were then analyzed by SAMDI-MS.

The average AUC values per level, along with level boundaries were determined as stated previously. The generated calibration curves, along with examples of encoded ASCII characters for all four base systems are shown in Figure 3.3.

Impact of Using Higher than Binary Base Storage. By expanding the previously reported molecular data storage technique into a higher-base system, it is possible to store an increased amount of information using less material per byte and with a minimal impact on accuracy. By moving away from a simple presence or absence of a peptide bound to a surface to denote a '0' or a '1' in binary, and instead moving to a scheme where the relative abundance of a peptide in relation to an internal standard denotes the base "value" of said peptide, it is possible to store an increasing amount of information, per peptide, based on the base system used. For example, in a standard

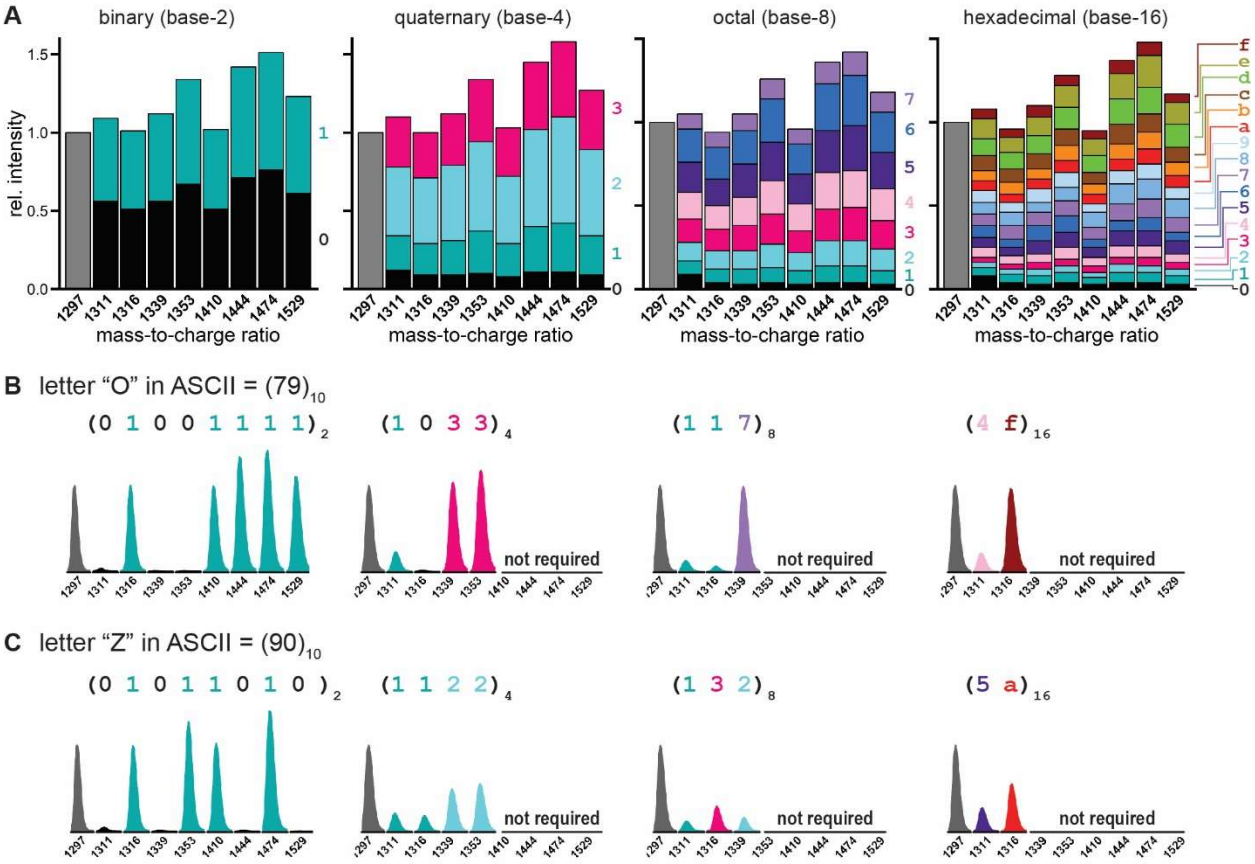


Figure 3.3: Calibration curves for all 4 base systems. A) The calibration curves for 4 base systems, binary, quaternary, octal, and hexadecimal. B) Example encoding of the capital letter "O" using all 4 base systems. The color of the peak represents the level of each peptide in that base system. For quaternary, octal, and hexadecimal not all peptides are required to encode this letter. C) Example encoding of the capital letter "Z" in all 4 base systems.

binary system, each peptide can only hold two values, a '0' and a '1'. This means that 8 peptides are needed to store a single byte (2^8 = 256). However, in a base 4 system (quaternary), each peptide can represent one of 4 values, '0', '1', '2', and '3', and thus only 4 peptides are needed to represent a byte (4^4 =256), effectively doubling the storage capacity of the 8 data peptides originally needed for binary. Similarly, for a base 8 (octal) and base 16 (hexadecimal) systems, 3 and 2 peptides are needed to represent a single byte, increasing the storage capacity by 8/3 and 4-fold over binary. Incidentally, the higher base storage systems also provide a benefit of needing

less material to store the same information as in binary. This is a result of the higher base systems requiring less peptides to encode a single byte.

	1,536 Binary	1536 Quaternary	1536 Octal	1536 Hexadecimal
Information Density (byte/cm^2)	16	32	42.67	64
Read rate (B/s)	0.63	1.25	1.67	2.5
Write rate (B/s)	1	2	2.67	4
Amount Stored per Plate	12,288 bits	24,576 bits	32,768 bits	49,152 bits
Observed Accuracy	100%	100%	98.56%	81.99%

Table 3.2: Important metrics of the four base-systems explored. Information density, read and write rate, amount of information stored, and observed accuracy for a single 1536 plate example.

Although using higher bases increase the amount of information stored in a peptide mixture, it is also important to take notice of possible increases in error rates when discussing information storage. In our original binary demonstration, utilizing 32 peptides, we were able to decode information with 99.9% accuracy. However, in this example, with the inclusion of an internal standard, the accuracy of the binary example was increased to 100% (Table 3.2). Furthermore, it was also determined that it is possible to obtain 100% accuracy with quaternary example as well. This indicates that with the method outlined here, it is possible to increase the data storage density by two-fold without any increase in readout error. However, going past the quaternary system, we observe rising error rates. For the octal example our error rate rises to 1.44% and for the hexadecimal case it was 18.01%. The possible sources of this error include deviation in the liquid handler (Labcytre Echo 555), or slight variations on the ion counts hitting the detector within the MALDI.

Although there is an increase in the error rate for higher base systems, it is undisputed that it is possible to store an information using concentrations of molecules. For this example, a

quaternary system provides a two-fold increase in density of information, while not incurring any additional readout errors. And in the case of the octal system, it is still feasible to use the storage system, providing an 8/3 enhancement in storage density over binary, while only suffering from a small amount of error. This error could of course be minimized by utilizing error correction coding, or oversampling, or by simply limiting the data encoded to text, which inherently has its own error correction, as a nonsense letter in the middle of a word is easy to spot, and our brains naturally correct this when reading. However, with the case of an 18% error rate in the hexadecimal example, the current method would be difficult to use, without a very rigorous error correction code, though with advances in both liquid handling and mass-spectroscopy, it is possible that this error rate can be reduced in the near future.

Conclusion

Regardless of the current error rates associated with higher-base storage systems using small molecules, chemical storage enables storage of information in higher than binary base format. With chemistry it is possible to utilize absolute concentrations of molecules, relative to other molecules, to encode information, which with appropriate techniques can be measured to very high precision. However, the classical storage techniques, such as magnetic tape, HDD's, solid state drives, and optical devices are usually stuck utilizing a binary format. With magnetic tape, HDD's and optical devices, the state of the media can only be in one of two states, negative/positive or not-charged/ charged, giving the domains a '0' or a '1' reading, though some research has been conducted into methods of expending past binary.[98] And with electrical potential based media, such as solid-state drives, the state of a semiconductor is not measured in an absolute sense, but instead depends on a range of charge.[99] This inherently does not pose a large problem for binary. However, when more than two states are desired, with each additional level the range

of voltages that represents a single state shrinks and any fluctuation in voltage will cause inaccurate readout.[100] This is further accentuated with the trends of making these semiconductors smaller and use less energy, by default this shrinks the voltage potential difference between the '0' and '1' state in a binary setting, which then also makes the possibility of using higher-base storage more difficult.[101] Of course attempts have been made to use a higher than base 2 systems in encoding and storing information, such as ternary computing or multilevel recordings for optical devices, however they have never gained a significant traction.[98, 102]

Another great reason for using molecules for storing information instead of classical techniques revolves around the idea of a spontaneous demagnetization, or flipping, of the digital domains.[103] Although optical devices do not suffer from this problem, anything that is either magnetism or potential based will. This is an idea that under non-ideal conditions, over time, or through wear, a magnetic domain can spontaneous flip from one state to another, or a capacitor that is charged could lose its charge. This would obviously cause a loss of the binary information from a '0' to a '1' or vice versa. This thus causes a problem that the devices that utilize these methods of storage must always be kept at ideal conditions, and in some cases also require an electronic power input to maintain the states that they currently hold. However, with a chemical approach this is not something that would pose a problem. With appropriate selection of molecules, such as those that are non-reactive or no-volatile, there is no concern of it spontaneously leaving a mixture or changing to a different molecule. Thus, once a mixture is made and the information is stored it can be assumed that those molecules will continue to be present in that mixture without any outside intervention or input.

Here we have shown that using mixtures and the ability to measure relative concentrations of molecules gives rise to a method for storing information utilizing higher-base encoding systems than the previously reported binary system in Chapter 2, which are also difficult to achieve with modern magnetic and potential based techniques.[94] We have shown that it is possible to increase the data storage density by 2 and 8/3 fold by utilizing quaternary and octal systems, respectively, with minimal error rates, 0% and 1.44%. We have also shown, that in principle, this technique could be extended to hexadecimal storage. Though the current error rate demonstrated within this work is 18%, with improvements in both liquid handling and analysis techniques we believe the observed error rates can be minimized.

Methods

Selection of Peptides for Data Encoding. All peptides that were previously synthesized, but not necessarily used, using standard Fmoc chemistry for the work reported in Chapter 2 were considered. 9 peptides, 8 data and 1 internal standard, that gave good signal-to-noise and did not cause spectral overlap with any other peptide were chosen. These were Ac-AGK$^{(me3)}$C (1311 m/z), Ac-A^{d3}G^{d2}K$^{(me3)}$C (1316 m/z), Ac-GVK$^{(me3)}$C (1339 m/z), Ac-GLK$^{(me3)}$C (1353 m/z), Ac-GLGK$^{(me3)}$C (1410 m/z), Ac-GFGK$^{(me3)}$C (1444 m/z), Ac-AYGK$^{(me3)}$C (1474 m/z), and Ac-G(Abu)FK$^{(me3)}$C (1529 m/z) as the data peptides and Ac-GGK$^{(me3)}$C (1298 m/z) as the internal standard peptide.

Preparation of Solutions of Peptides. To prepare peptide mixtures for immobilization, each peptide stock solution was distributed into a source plate. Mixing of peptides to form binary, quaternary, octal and hexadecimal data sets was performed using these peptide stock solutions and an Echo® 555 (Labcyte Inc.) liquid handler, in predetermined concentrations.

Generating Input Tables for Automated Encoding of ASCII Characters. To generate an input table of the standard 95 ASCII characters for the Echo® 555 liquid handler, a list of the 95 characters and their binary, quaternary, octal, and hexadecimal codes were generated (Appendix Table 3.1). Next, the bitstreams (here we define bitstream as any combination of numbers + letters that encode an ASCII character in any of the base systems) for all 95 characters was combined. This combined bitstream was then duplicated an appropriate number of times to reach a number greater than or equal to 12,288, or the number of spots (1536) per SAMDI plate times the number of data peptides per spot (8). Once this was done, each of the base system's bitstreams were divided into 8 number long segments. Each digit in the segment was then assigned to one of the 8 data peptides, assigning a source well for each. The bitstream, with each entry's associated bit number, source well, and destination well, was then reduced to include only those entries with a bitstream value that was not '0'. Next, we used the "Vlookup" function in Microsoft Excel to assign the transfer volume for each entry, based on the base system used. Finally, these entries were transferred into an Echo input table, with information on source well, transfer volume, destination well and destination plate name.

Automated Encoding via Liquid Transfer. Prior to initializing a run on the Echo® 555 liquid handler robot, a source plate (Labcyte Echo Qualified 384-well plates, Cat #: PP0200) was prepared with the desired peptides to be transferred. Each well of the source plate contained 65 µL of each of the 8 data peptides and 1 internal standard peptide stock solutions (2 mM in peptide). The source plate and destination plate (Greiner BioOne 1536-well plates Cat#: 782270) were placed in storage towers in the Access Laboratory Workstation attached to the liquid handler. To initiate the run, the input table was imported, which defines the locations of the source and destination plates, and the protocol was executed. Once the peptides were transferred, the

destination and source plates were covered with lids (Labcyte MicroClime Environmental Microplate Lid Cat#: LL-0310) to ensure that the contents of the plates did not dry.

Preparation of SAMDI-MS Plates. Array plates with 1536 gold spots on steel plates were soaked in a solution of a mixture of EG3-capped alkane disulfide and a mixed disulfide of EG3-capped alkanethiol and a maleimide-terminated EG3-capped alkanethiol for 24 h, at room temperature, to allow formation of a self-assembled monolayer on the gold surface. The solution of disulfides contained an overall concentration of 1 mM of the two monolayer compounds in a stoichiometric ratio (2 to 3) to yield a monolayer wherein the maleimide groups are present at a density of 20%. Following monolayer formation, the plates were soaked in a solution of hexadecyl phosphonic acid (10 mM) for 5 minutes, and rinsed with ethanol, water, ethanol, dried with nitrogen and stored dry under vacuum.

Immobilization of Peptides onto Plates. Prior to immobilization, the peptide mixture plates generated by the Echo® 555 liquid handler were filled with 5 μL of 100 mM Tris buffer at pH 8.0, with a ThermoFisher Multidrop Combi, to ensure the solutions of mixed peptides are at the correct pH and appropriate concentration for immobilization. 0.5 μL from each well of the 1536-multiwell plate was then transferred to a 1,536-spot SAMDI plate functionalized with 20% maleimide and displaying a hexadecyl phosphonic acid background between spots. The transfers were completed utilizing the TECAN Freedom Evo instruments, with a MCA 384 head utilizing 15 μL tips. Once transferred, the peptide solutions react with the maleimide groups on the surface of the plate for 15 minutes, in a humidified chamber, to covalently immobilize the mixture of peptides. After immobilization, the plate was washed with ethanol, water, ethanol and dried under nitrogen.

SAMDI-MS Analysis. SAMDI plates with immobilized peptides were first treated with 2',4',6'-trihydroxyacetophenone matrix solution (THAP, 12 mg/ml in acetone) and then were loaded into an ABSciex TOF-TOF 5800 instrument. Matrix assisted laser desorption/ionization time-of-flight mass spectra were collected for each spot in reflector positive mode.

Levels Calibration Curve Generation. To generate calibration curves for each peptide for all four base systems, binary, quaternary, octal, and hexadecimal, peptides were immobilized at all possible concentration levels for each system. Following this, the areas of each peptide's peaks was exported using the built-in analysis function within the TOF/TOF Series Explorer Software. The area under the curve (AUC) for each peptide was then compared to the AUC for the internal standard peptide. Once the relative AUC for the peptides were determined at each level, the halfway points between each level were set as the boundary between each level.

Analysis of Spectra. Once all the spectra on a single 1536 SAMDI plate were acquired, the built-in analysis function within the TOF/TOF Series Explorer Software was used to ascertain the AUC for each peptide. These values were then imported into Microsoft Excel and the relative AUC were calculated for each data peptide. To automatically determine the level associated with each of the peptide in each spectra the following equation was used, =INDEX([level labels],MATCH(MIN(ABS([relative ratio for unknown]-[relative ratio for calibration])),ABS([relative ratio for unknown]-[relative ratio for calibration]),0)), where values within the square brackets are user inputs. The levels determined by this equation were then compared to the actual levels written and any mistakes were counted and % error determined.

Chapter 4: Increased Information Storage Using Inkjet Printing.

The work in this chapter was accomplished with the help and input from Amit Nagarkar

Introduction

As proven in Chapter 3, it is possible to increase the amount of information stored using molecules on SAMDI surfaces utilizing relative concentrations of these molecules in relation to an internal standard. However, there is another method by which an increased amount of information can be stored in the same physical space. This method moves away from immobilizing multiple molecules in a location, which only linearly increases the amount of information that can be stored, and instead moves to storing one molecule in many locations, which scales as a square for every halving of the distance between locations.

However, the use of conventional liquid handling instruments limits the density with which these molecules can be delivered to a SAMDI surface. These limits have two sources, the spatial inaccuracy of the instruments and the minimum volume of transferable liquid. Utilizing common 384 pipette instruments, such as a TECAN FreedomEvo or Fluent, allows for non-trivial dispensing of liquids to 384 target plates, both on surface and well formats. However, it requires increased effort to accomplish dispensing onto 1,536 plates. This is mostly due to the variance in the mounting angle of the tips, along with variance in the placement of the destination plates. Another source of error arises from the requirement for the instrument to make 4 transfers to deposit liquid into each of the 1,536 wells, as with each transfer, and without precise tuning and calibration of the instrument's movements, a small step-loss could cause misalignment which will be amplified through the course of the instrument's use, though with careful setup and execution, 1,536 plates can and have been used, as we have previously reported.[19, 94] However, a plate

presenting the next higher density – 6,144 – would necessitate 16 individual transfers, each of which could suffer from misalignment, decreasing the probability of correct dispensions of all 6,144 solutions. Of course, the misalignments due to tip mounting error can be overcome utilizing permanently mounted tips, such as those utilized by the AnalytikJena CyBio Well Vario.[104] Yet, these instruments could nonetheless still suffer from misalignment issues stemming from the movement of the head relative to the substrate. Even with an increased number of dispensing tips, such as the 1,536-tip head available on the vario, anything above 6,144 is not commonly done. A possible workaround of having to use an instrument that utilizes multiple tips, with fixed spacing, is to move to an instrument that does not require mounted tips and can transfer from a single source to many destination locations on demand. One of the best examples of these instruments is the LabCyte Echo series, an instrument we have used extensively in Chapter 2 and 3. These instruments utilize acoustic wave generators to transfer liquid from a source well to a destination location. These instruments have precise movement servo's that enable transference of a liquid from the source to a destination with high precision, albeit one source and one drop at a time. The size of these drops is of course dictated by the model of the instrument used, but currently it is available in 2.5 and 25 nL drop sizes. And although each transfer takes only a fraction of a second, transferring multiple drops into all the destination location still takes a great deal of time.

However, all of the above-mentioned instruments have one big issue when it comes to increasing the density of spots on a surface, volume. It may seem that a drop of 2.5 nL is miniscule in the scheme of everyday life, yet when one considers the volume of an ideal-halfsphere as $V = {}^{2}/_{3}\,\pi r^{3}$, we determine that the radius of a 2.5 nL drop would be roughly 100 µm, or a diameter of 200 µm. If we now look diameters of gold islands on standard SAMDI plates we have 2.5 mm, 1.25 mm, and 0.625 mm for 384, 1,536, and 6,144 densities, respectively, with center to center

distances of twice the diameter. Putting it in other terms, the diameter of the spots halves as we quadruple in density. Continuing this trend, if we halve the 625 μm spot diameter of the 6,144 plates, we obtain a diameter of 312.5 μm for the 24,576 density, and roughly 156 μm for 98,304. This means that if we use the smallest drop size possible with the Echo, we should not expect to dispense drops at a density somewhere between 24,576 and 98,304 spots per standard SAMDI plate. In this chapter we will discuss techniques through which the goal of delivering molecules in ever increasing densities, with the intent of increasing information storage, can be accomplished.

Results and Discussion

Importance of I-SAMDI, Dispensation Technique, and Volume on Density. Although evaporating, dispensing to, and sampling 24,576 individual gold islands on a SAMDI plate may seem like a daunting task, we have previously developed a method that enables sampling of far greater number of locations, imaging-SAMDI (i-SAMDI). I-SAMDI is a method which combines imaging mass spectrometry (IMS) functionality of a Bruker autoflex or RapifleX with SAMDI-MS.[13, 97, 105] Utilizing this mode on the MALDI instruments allows one to visualize the location of molecules of interest.[105] Initially developed to image the localization of molecules of interest within tissue samples, we have adopted this technique to image SAMDI surfaces. For this purpose, the entirety of the surface is covered in gold and our monolayer molecules are allowed to self-assemble. This primes the entire surface of the gold to be receptive to the immobilization of molecules of interest. In previous examples we have utilized these surfaces to monitor the reaction of enzymes in microfluidic channels.[96, 97] However, there is no restriction on how the molecules can be delivered onto the surface. Utilizing any applicable method that can deposit solutions containing molecules of interest onto the surface, arrays of molecules could be immobilize to form an array of spots, similar to the SAMDI arrays displaying individual gold-islands on a plate. In

fact when we dispense 2.5 nL drops, using the Echo, in a 24,576 grid pattern and visualize them using i-SAMDI we obtain Figure 4.1. Here we see that that these 2.5 nL drops have a diameter of roughly 300 μm. The increased diameter of these drops is most likely caused by a flattening of the half-sphere due to the velocity with which the droplet hits the target plate.

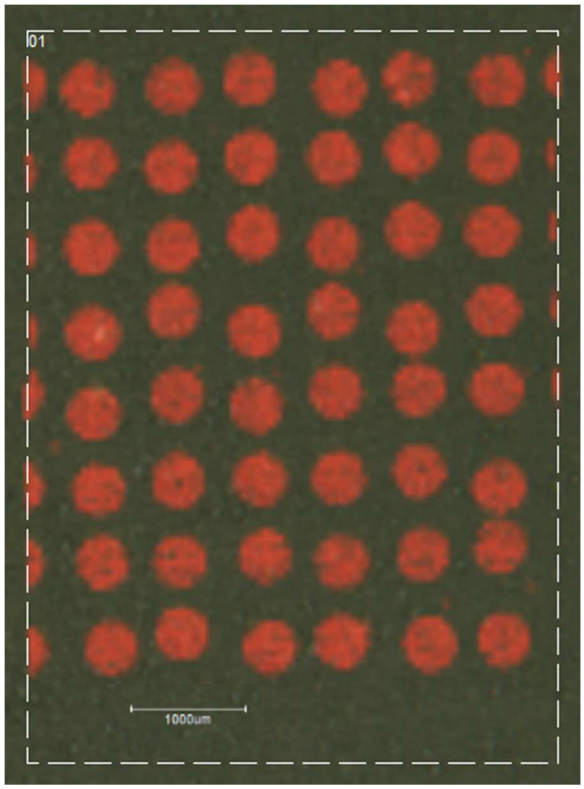


Figure 4.1: Image of a RapifleX i-SAMDI run of 2.5 nL droplets dispensed in a 24,576 density onto a solid gold surface. In this figure red represents the locations where a peptide immobilized after being delivered in the 2.5 nL droplet dispensed by the Echo.

Thus, using the Echo, we can only access densities of 24,576 – or about 98,304 bytes/plate using the binary approach outlined in Chapter 2. However, it should be possible to obtain higher spot densities if we move away from mainstream liquid-handling instrumentation utilized in a laboratory setting, and towards a technology that has been developed for printing. This technology is of course inkjet printing, a technology that was first introduced in the 1970's and has since seen many developments. Many consumer inkjet printers, such as those found in many households can

already dispense droplets that are smaller and with better resolution than some of the previously mentioned lab-grade instrumentation. A standard $100 inkjet printer can dispense droplets as small as sub-10 picoliter and with a resolution down to sub-10 micron. Now, we must take the resolution given by the manufacturer with a grain of salt, as inkjet printers are made to deceive the eye. By mixing different quantities of 4 inks, black, cyan, magenta, and yellow, on top of each other it makes all the colors in the visible spectrum. However, for data storage, each spot of peptide must still be spatially separated. Thus, for our application, the volume of the droplets determine what the minimum resolvable distance is. Following the equation for the volume of the sphere mentioned above, a 10 pL and 1 pL droplets would produce spots with diameters of roughly 34 and 16 microns respectively. Given a center-to-center distance of twice the diameter, this would give us about 2.1 and 9.4 million spots on a single SAMDI plate of 12 cm x 8 cm, respectively.

Using Consumer Grade Inkjet Printers. To explore the ability of a consumer grade inkjet printer to print at or near these densities, we purchased an Epson Stylus C88+ from Amazon and loaded refillable ink cartridges with fluorescent inks manufactured to be printed with standard inkjet printers. We utilized the software developed by Wassdorp et.al. and printed a representative matrix of dots.[106] A decision was made to print the matrix of spots on a paper-like substrate, to negate the need for any engineering needed to accommodate SAMDI plates. The media chosen was cotton photopaper (absent of any optical brighteners to not interfere with fluorescence). After printing the matrix, a portion of the dots were imaged on a Nikon Ti confocal microscope under 10x magnification. We utilized the built-in measuring capabilities of the NIS-Elements AR software to measure the diameters of the spots, as well as center-to-center distances of the spots (Figure 4.2). What we observed is that the drops, which measure 3 pL in volume, had a diameter of roughly 30 micron, more than the theoretical 22.5 micron predicted by the equation, most likely due to the

wicking properties of the paper, however no mixing of the adjacent spots was observed under these settings. The program allowed for adjustment of the horizontal spacing, which is controlled by the timing of the spot being dispensed while the printhead moves across the media. However the larger vertical distance between the rows of spots is fixed due to the physical distance between the nozzles on the printer head. The control of the vertical distance between spots is typically accomplished by physically moving the media during the printing process.

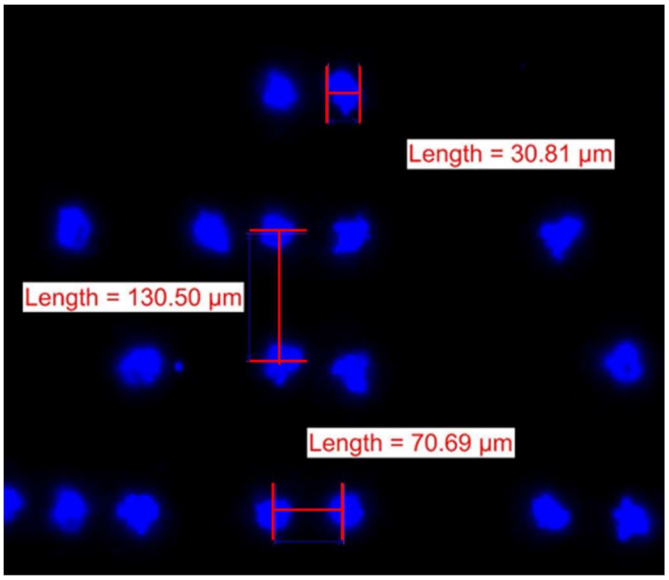


Figure 4.2: Dispension of fluorescent dyes onto a cotton photopaper. Here UV-fluorescent dyes were dispensed using a Epson Stylus C88 + inkjet printer to determine the possible resolutions of printing using ideal printing materials. Observed under 10X optical magnification.

The use of commercial inkjet printers to efficiently print SAMDI "inks", which consist of a buffered aqueous solution with 20 µM peptide, as reliably as those developed for inkjet printers require two conditions to be met. First, the formulation of the inks much be such that they have a high viscosity while maintaining low surface tension. And second, the waveform used to generate a pulse to eject the inks out of the nozzles needs to be suitable. Much research has been done on ink additives which has allowed for the printing of many different compound-containing inks. However, the consumer grade inkjet printers' ability to change the waveform used is limited to

those that have been preset by the manufacturer, typically three setting for three different size droplets. Thus, if the utilization of consumer-grade inkjet printers is desired, the only option is to formulate the "ink" to be usable with one of the standard waveforms.

Benefits of Commercial Grade Materials Printers. However, non-commercial grade inkjet or "materials printers" exist. These include the Fujifilm Dimatix Materials Printer. Not only does this printer boast a lower minimum volume per drop, 1 pL, but it also can tailor the waveform at will. Using the Dimatix, our collaborators have previously shown that it is possible to print a nearly perfect grid pattern utilizing Alexa Fluor dyes in DMSO (Figure 4.3). We can also see that these 1 pL drops have a diameter close to that of the expected 16 μm, and can be printed in a grid with a center-to-center distance of roughly twice the diameter. However, we were also curious to determine if it would be possible to print aqueous peptide containing "inks" without the use of additives, simply by tailoring the waveform. To accomplish this, we utilized the waveform developed for printing of the Alexa Fluor's in DMSO and began tuning until droplets were seen being ejected from the printer nozzles. Once a waveform that could produce drops was determined a sample grid was printed onto a blank metal surface and the plate was imaged using the onboard camera (Figure 4.4). Through post-processing of the image, we superimposed vertical and horizontal lines in a fixed 30 μm grid pattern. At the center of the intersecting lines we drew a green circle with a diameter of 20 μm, indicating the coverage of a 20 μm wide laser spot that can be easily generated within the RapifleX. When we examine all of the green circles, we see that there are small white dots within each one. These are the aqueous "inks" that the Dimatix generated and printed. Clearly their location is not as precise as those printed in Figure 4.3; however, each of them lies within the theoretical 20 μm laser spot. We must also remember that these picliter sized drops dry and contract as they do, so the size of the drops and their location may not be the

original size nor location. However, because each of the spots lies within the 20 µm circles, we believe, even without any further optimization, it should be possible to print our SAMDI "inks" in a grid patterns as small as 30 µm apart. This density of spots would give us roughly 9.4 million spots within a regular SAMDI plate, increasing the density of information from 64 bytes/cm^2 previously reported to 12,215 bytes/cm^2.[94]

Reaching Maximum Possible Densities. However, even at these spot densities, we do not reach the minimal laser and raster distances available on the RapifleX. The smallest laser diameter achievable on the RapifleX is 10 µm. At this laser spot size, and assuming a raster distance of 10 µm, we would achieve 96 million spots on a single plate. In other words, each SAMDI plate could hold 96 million bits, or 12 million bytes (12MB), giving it a density of 125,000 bytes/cm^2. To achieve these densities, we would of course need to use printers capable of generating drops that are even smaller than 1 pL, a feat achievable using SIJ Technology's super-inkjet printers.[107] Using this technology it is possible to print droplets as small as 1-10 femtoliters (fL) and create features

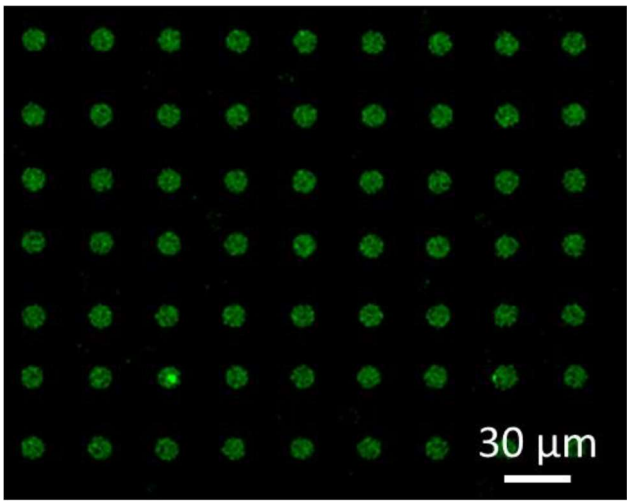


Figure 4.3: Alexa Fluor dyes printed with Fujifilm Dimatix. 1 pl droplets printed in a grid pattern to demonstrate the consistent drop size and spacing when using the appropriate waveform with the printer.

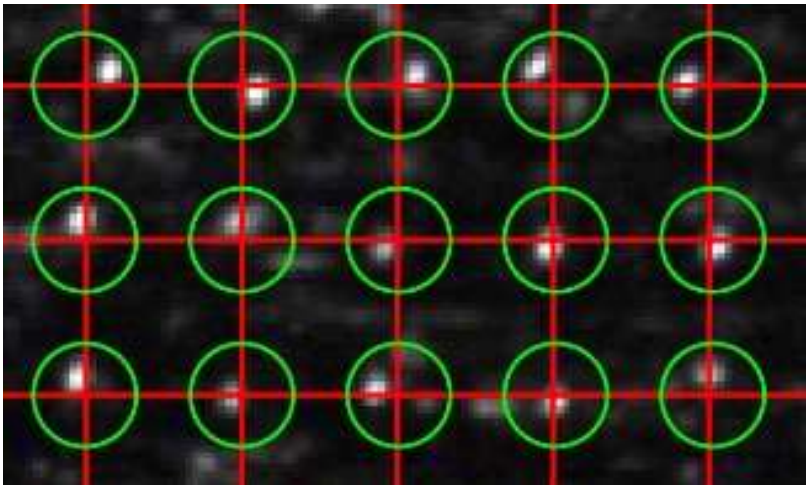


Figure 4.4: The grid printed by the Dimatix printer utilizing a modified waveform for aqueous peptide solutions. The red grid pattern is composed of lines drawn at 30 µm spacing, where perfectly placed droplets would be visible at the crosshairs of the vertical and horizontal lines. The green circles are representative of a 20 µm laser spot with the raster distance of 30 µm. And the bright white dots within each of the green circles is the small 1 pL drop of aqueous solution containing peptide that was printer by the printer.

into the nm range, far smaller than the resolution of the top of the line MALDI instruments. It is not unimaginable that given time it would be possible to develop MALDI instruments with lasers capable of reaching single digit µm laser spots or even smaller, and with this development, the amount of information that can be stored on a single SAMDI plate will continue to drastically increase.

Conclusion

With the currently available liquid handling instrumentation, based on pipetting and acoustics, we are limited in the densities of information that can be deposited onto a single SAMDI plate. This is due to two factors, the spatial inaccuracies of these instruments, and volumetric constraints. However, we believe that if we can transfer substantially smaller volumes, with greater accuracy, and with smaller pitch, it would be possible to reach higher areal density of information stored, even if we transition to 1 molecule per spot. Inkjet printing provides us with a tool that can checkoff each of those requirements, and we have demonstrated that in principle it would be

possible to increase densities by a factor of 40 with simple consumer-grade desktop printers. However, this would require engineering of the peptide inks to be jettable by the standard waveforms of a consumer inkjet printer. On the other hand, it is possible to also go beyond consumer grade, and use commercial printers specifically manufactured for customizability and utility. A commercial-grade printer could have its waveform edited to print even the standard peptide inks used previously, and with its lower drop volume and increased accuracy can increase the density of storage by a factor of almost 200, from those reported in Chapter 2. However, even this does not reach the limit of the density that can be achieved by the most advanced MALDI instruments. With laser resolutions down to 10μm, the most advanced MALDI's can read spots at densities of 1,000,000 spots/cm^2, giving a factor of 2000 increase in information storage from that reported in Chapter 2. And with further advancements in MALDI instrumentation and developments in inkjet printers, such as the super-inkjet printers, it is not unimaginable that the information that can be stored on a single plate would increase further.

Methods

Materials. The Epson Stylus C88 + printer, 365 nm flashlight, and all dyes and cartridges were bought on Amazon. Cotton photopaper (Canson Infinity RAG Photographique) was purchased at BLICK Art Supplies. Fmoc Rink Amide MBHA resin (0.42 mmol/g) was purchased from AnaSpec Inc. Fmoc amino acids were purchased from AAPPTec and Chem-Impex International Inc. All other solvents and reagents for Fmoc solid-phase peptide synthesis were purchased from Thermo Fisher and Sigma Aldrich. AlexaFluor dyes were bought from Thermo Fischer and were used without further purification. A virtual Linux machine running Ubuntu was setup using the virtualbox software for windows.

Preparation of SAMDI-MS Plates for RapifleX. Steel plates, with modified dimentions that fit the Bruker RapifleX, were washed with hexanes, ethanol, water, ethanol, and dried with nitrogen gas. An electron beam evaporator was used to deposit the gold onto the entire plate. The gold covered plate was then soaked in a solution containing a mixture of tri(ethylene glycol)-terminated (EG3) alkanedisulfides and a mixed disulfide of EG3-alkanethiol and a maleimide-terminated EG3-alkanethiol for 24 hours at room temperature to form self-assembled monolayers on the gold surface. The solution of disulfides (1 mM total concentration) had the two disulfides present in an appropriate ratio to yield a final density of maleimide groups of 20%

24,576 Array Density Peptide Deposition by Echo. A new array of 192 x 128 spots, for the Labcyte Echo was designed. The fully gold covered Bruker SAMDI plate with 20% maleimide functionality was inserted into the Echo. Using the automated transfer program, 2.5 nL of a peptide solution was delivered to each of the 24,576 locations on the array. These peptides then underwent immobilization, via Michael addition, to the maleimide functionality, with the thiol on the cysteine sidechain. After the 5 minutes all of the delivered peptide solutions were dry. The remaining buffer salts were washed away with water and ethanol, and the plate was covered with THAP matrix (12 mg/mL in acetone).

i-SAMDI-MS Analysis. SAMDI plates with immobilized peptides were first scanned in a desktop scanner using 2400 DPI setting. These images along with the plates were loaded into a Bruker RapifleX. The Fleximaging software was opened and 3 corners of the plates were taught to align the image to the physical location of the plate. An area to scan was then selected in the Fleximaging software. To initiate the run an acquisition method, along with laser spot, raster distance, and laser

intensity were selected. Upon completion, the location of the immobilized peptides were visualized via color coding of the image to the presence of immobilized peptide mass.

Use of ESCP2-Client. The use of this software has been previously reported.[106] Briefly, the header, and body of the code was written into the python script. The header section contained the printer specifications, including printer resolution, nozzle count, print bar count and identification, and location of the print. The body contained the information on the matrix of the spots to be printed. To generate this matrix, ASCII text was converted to binary and the string of bits were divided into appropriate number of columns and rows. The number of columns depended on the spacing for the spots, but for an entire page was determined to be 2970 spots. The number of rows was fixed by the number of nozzles per color, 59 for each color, and 180 for black. Upon running the python script in the terminal of the Linux virtual machine a .prn file was generated.

Dispensing Fluorescent Ink with Consumer-Grade Inkjet Printer. Fluorescent dyes purchased from Amazon were loaded into empty ink cartridges. These cartridges were then loaded into the printer. The old inks were then cleared out of the lines and print head by performing the head cleaning procedure 5 times. Each of the nozzles was confirmed to be primed by printing a simple test page. If any nozzle was not primed or appeared to be clogged the head cleaning procedure was done again. Once each of the nozzles has been cleared and primed, the .prn file produced by the python script was run using the "lp <file-name>.prn" function in the Linux terminal.

Dispensing Material with Dimatix Materias Printer. Dispensing of fluorescent inks was accomplished with the use of a Fujifilm Dimatix 2831 printer, using the 1 pL printing cartridge. A TIFF file for the array to be printed was loaded into the program and the array of spots was printed.

The firing conditions used for the active nozzles were: 16V firing voltage, 0.5 mm height off the substrate.

Part 2:

Utilization of Peptide Arrays with SAMDI-MS

Chapter 5: Profiling Protease Activity in Laundry Detergents with Peptide Arrays and SAMDI Mass Spectrometry

This chapter is adapted from the following published work:
Dai, R.; Ten, A.S.; Mrksich, M., *"Profiling Protease Activity in Laundry Detergents with Peptide Arrays and SAMDI Mass Spectrometry"* Industrial & Engineering Chemistry Research, 2019. **58**(25): p. 10692-10697.

Introduction

Proteases are important in a wide range of industrial applications—including food manufacturing, leather making, and personal care—and they account for 60% of the market for industrial enzymes.[108, 109] The primary application is found in laundry detergents, where the enzymes remove proteinaceous stains, and have been engineered to operate efficiently at low temperatures, thereby eliminating the need for energy-consuming high temperature washes. The use of enzymes has also relieved the inclusion of phosphates, resulting in more environmentally benign detergents.[110] Continued development of proteases for laundry detergents will benefit from technologies that have been developed for protein engineering and characterization in the life sciences. In this paper, we demonstrate the use of peptide arrays to profile the activities and specificities of several commercial detergents, and we suggest this method will be important in comparing and optimizing proteases for detergents.

Proteases are used in laundry detergents to digest protein-based stains including those from blood, grass, sweat, milk, and food soils. Among the important characteristics for selecting proteases are thermal stability, broad alkaline pH stability, and compatibility with detergent ingredients and mild bleaching agents.[111] Serine endoproteases have proven the most effective, as other proteases are incompatible; thiol proteases are oxidized by bleaching agents; metalloproteases lose their metal cofactors due to the binding of water softening agents and

hydroxide ions.[112] The subtilisin class of serine proteases from Bacillus species is favored for its high stability and broad substrate specificity, and has been used extensively in laundry detergents for decades.[113]

Due to their significance in industry, subtilisin proteases have been the focus of much study, with the goals of improving the production of subtilisin as well as characterizing the activities of different variants on adsorbed substrates.[114, 115] Protein engineering efforts have focused on optimizing subtilisin proteases via directed evolution to enhance thermal stability as well as efficiency at low temperatures, alkaline stability, and compatibility with detergent components.[111, 116, 117] These efforts have revealed a plasticity of protease active sites that allow them to be engineered for a wide range of activities under varying conditions.[108] While assays are used to ensure that proteases remain active at these conditions, experiments are generally not performed to characterize the substrate specificities of the proteases in detergents.[118, 119] In principle, it would be useful to engineer proteases that have high activity and little specificity, to make them broadly effective at digesting and removing protein stains.

Combinatorial methods have proven important for assessing the specificities of engineered proteases and identifying optimal substrates.[120, 121] In this paper, we describe the use of peptide arrays to profile protease activity and specificity in commercial laundry detergents. We separately apply detergent solutions to an array having 324 peptide sequences and allow proteases in the detergent to cleave the peptides. The peptide array is then analyzed with SAMDI (self-assembled monolayers for matrix-assisted laser desorption/ionization) mass spectrometry to identify the cleavage sites in the peptides and to quantitatively measure the extent of cleavage.[13, 78] This work

provides activity profiles that allow a comparison of the detergents and suggest that this assay may be important in guiding the development of new formulations.

Results and Discussion

Preparation of Peptide Arrays. We individually synthesized 324 peptides based on the sequence Ac-GXZAFC, where the variable positions X and Z represent all 18 natural amino acids excluding cysteine and methionine. We also prepared an array of self-assembled monolayers of alkane thiolates on gold terminated in maleimide groups at a density of 10% against a background of tri(ethylene glycol)-groups (Figure 5.1).[9] The maleimide functionality allows for the selective immobilization of cysteine-terminated peptides, while the glycol background prevents non-specific adsorption of proteins to the surface.[23] The monolayers were prepared on a metal plate having an array of 384 gold islands in the standard format used in the life sciences, as previously described.[122] The use of self-assembled monolayers has the further benefits that the peptides are immobilized in a defined orientation and a uniform density across the array, giving a consistent activity and good reproducibility.[56, 63, 123] The 324 peptides were transferred onto the monolayers within the 384-array plates using an automated multichannel liquid transferring robot with a 384-tip head. SAMDI mass spectrometry was then used to confirm the immobilization and purity of each peptide (each of the 324 spectra are presented in Appendix Figure 5.1).[13]

Characterization of Protease Activity and Selectivity. To measure the proteolytic activity of a laundry detergent, we prepared a working concentration of the detergent by diluting it in water according to the manufacturer's instructions for a medium-load cold wash; note that we are labeling the detergents as A-E and not disclosing the specific detergent brand for each proteolytic

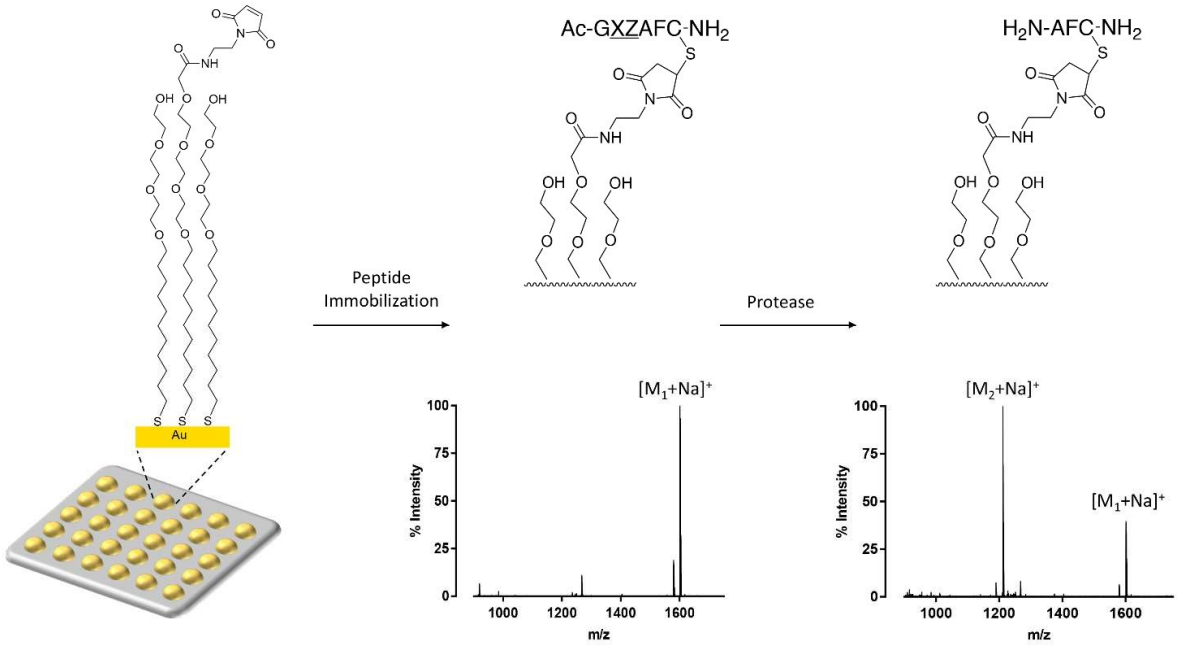


Figure 5.1: Overview of immobilization and protease reaction with SAMDI-MS. Maleimide-terminated self-assembled monolayers were used to covalently immobilize cysteine-terminated peptide substrates on the surface in an array format. The array was placed in a solution of laundry detergent (using the manufacturer's recommended dilution) and then analyzed by SAMDI MS to quantitate the proteolytic product. Mass spectra corresponding to the substrate and product peptides are shown in the lower left and right, respectively.

profile. We separately immersed immobilized peptide array plates into solutions of detergent in sealed reaction chambers and placed the containers on a shaker, to mimic the conditions of a cold laundry wash. After 20 minutes, we removed the plates from the solutions and rinsed them with water, ethanol, dried under nitrogen, and treated with THAP matrix in acetone. The plates were analyzed by SAMDI mass spectrometry and protease activity was quantified by quantitating the areas under the peaks of the parent peptide and cleavage products. Each detergent condition was repeated in triplicate and individually analyzed. The standard deviation in these measurements was smaller than 6% for 88% of the peptides.

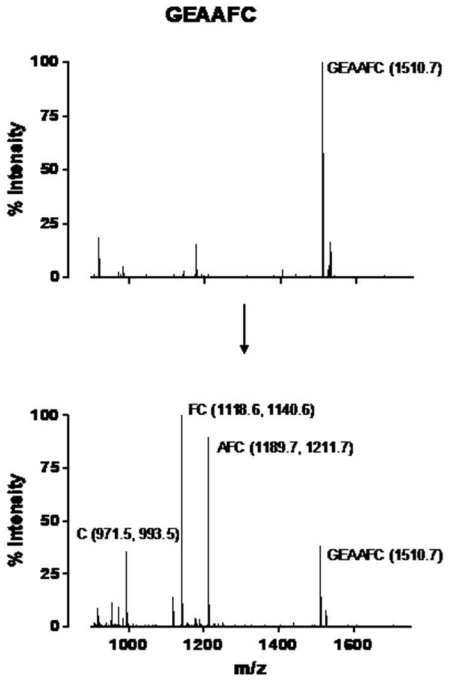


Figure 5.2: Example protease activity SAMDI-MS spectrum. Mass spectra of a monolayer presenting the peptide GEAAFC before (top) and after (bottom) treatment with detergent A. Multiple cleavage sites reveal the enzyme activity on this sequence.

We first profiled detergent A and observed a wide range of cleavage activity that showed a strong dependence on the peptide sequence. For the peptide substrate GEAAFC, we observed three distinct product peaks resulting from proteolytic activity (Figure 5.2). First, the peaks at m/z 971.5 (H+) and m/z 993.5 (Na+) correspond to the fragment NH_2-Cys-maleimide, and therefore represent cleavage of the Phe-Cys bond near the C-terminus of the peptide. Second, the peaks at m/z 1118.6 (H+) and m/z 1140.6 (Na+) correspond to the fragment NH_2-Phe-Cys-maleimide, and represent cleavage of the Ala-Phe bond. Finally, the peaks at m/z 1189.7 (H+) and m/z 1211.7 (Na+) correspond to NH2-Ala-Phe-Cys-maleimide, and represent cleavage between the variable position alanine and fixed position alanine in the sequence. Certain peptides in the array showed variations of the three cleavage sites, others were cleaved at only one or two of the positions, and still others were inactive towards any proteolytic activity. Since there existed a possibility for

multiple cleavage sites, total protease activity on each peptide was quantitated using the sum of the area under the peaks (AUP) of the products with the following equation:

$$\text{Activity} = \frac{\sum_{i=1}^{n} \text{AUP}_{\text{product,i}}}{\sum_{i=1}^{n} \text{AUP}_{\text{product,i}} + \text{AUP}_{\text{substrate}}}$$

To compare the total proteolytic activity of detergent A on the peptide array, we generated a heat map that showed the extent of cleavage for each peptide, where each peptide is represented by one square and the degree of purple intensity corresponds to the yield for cleavage products (Figure 5.3A). The proteases in detergent A were observed to have higher activity towards substrates with hydrophobic amino acids in the variable positions. This trend correlated with previously known selectivities of subtilisin proteases, including a slight preference for hydrophobic residues in the -1 position.[113] Furthermore, the peptide arrays reveal a lower proteolytic activity for substrates having a glycine or histidine in either variable position and almost no activity with a proline in the Z positions, as well as significantly lower activity with asparagine, and arginine in the variable X position. Finally, a control experiment where the array was not treated with detergent showed essentially no proteolysis of the immobilized peptides (Appendix Figure 5.2).

We also analyzed the data to present a cleavage site-specific profile for detergent A (Figure 5.4A). This detergent's proteases displayed a high relative specificity for the bond between the variable position Z and the fixed position alanine in the GXZAFC array, shown through the orange bars that denote NH2-Ala-Phe-Cys as the final product. The second most abundant product was NH2-Cys shown in green, followed by NH2-Phe-Cys shown in yellow (the fraction of each cleavage product for each of the detergents is tabulated in Appendix Table 5.1A). These results,

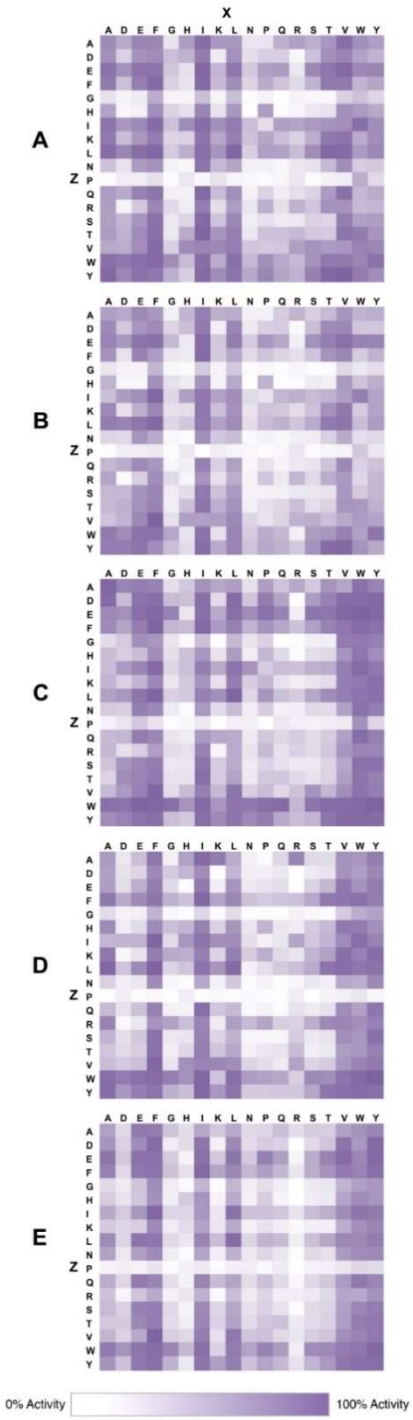


Figure 5.3: Heat maps represent total proteolytic activity of the peptide library GXZAFC in each of five anonymized detergents, A-E

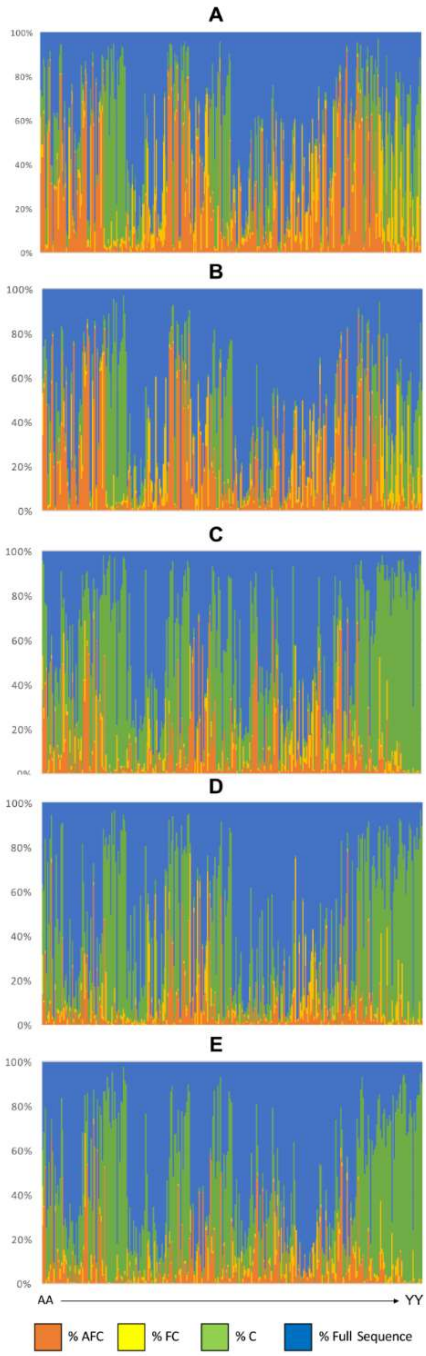


Figure 5.4: Site-specific cleavage activity of each protease. 100% stacked bar graphs display sequence-specific cleavage activity of each detergent protease across the 324-peptide library. The variable dipeptides in GXZAFC on the X-axis are arranged in alphabetical order, from AA to YY (left to right).

together with the heat map, show that both total cleavage activity and specific cleavage activity can be deduced using SAMDI to analyze peptide arrays.

We also used the peptide arrays to profile the other four commercial laundry detergents (Figures 5.3B-E and 5.4B-E, Appendix Table 5.1B-E). We again prepared detergents at the suggested concentrations, treated separate peptide arrays with each solution, and analyzed the arrays with SAMDI to generate heat maps and site-specific profiles as described above to show cleavage activity and specificity for each. A sixth detergent that lacks enzymes was also tested and no proteolytic activity was observed (data not shown). A brief comparison of the heat maps reveals that detergent C has the highest overall activity of the 5 detergents. Also, the heat maps showed similar overall activity and specificity for the 5 detergents, such as the preference for hydrophobic residues in the variable positons, and the strong loss of activity for peptides having glycine and proline in positions X and Z, respectively. However, the 5 detergents can be divided into 2 groups, based on a slight variation in activity, with detergents A and B in the first group, and C, D, and E in the second. The two groups have a substantially similar specificity, though proteases in the second group show a strong preference for peptides having tryptophan and tyrosine in both variable positions, which is not seen with detergents A and B.

The site-specific cleavage activity depicted in Figure 5.4 (which quantitates the possible products from proteolysis of the peptide at different amide bonds) gives further information on the relatedness of the detergents. The first group (detergents A and B) have a strong preference for cleaving the amide that is C-terminal to the variable positions (Figure 5.4, orange bars). The three detergents in the second group are also quite similar (though, as explained further below, there are differences). The differences between the two groups, however, become significantly more

apparent when site-specific activities are analyzed. Unlike the proteases in the first group, those in the second show a strong preference for cleavage at the Phe-Cys amide bond near the C-terminus of the peptide (green bars).

Interestingly, although the proteases in the second group (C, D, and E) have similar specificities, careful examination of the activity profiles reveals slight differences. For example, detergent C shows high activity on peptides with Asp and Glu in both variable positions and Pro in position X, while detergent D lacks a preference for those residues, and even shows lower activity for Glu and Pro in both variable positions. Likewise, detergent E only shows an increased preference for Glu, but not for Asp and, unlike C and D, disfavors Arg in the X position.

Finally, we quantitated the total proteolytic activity for each detergent ─ by taking the average total cleavage of each peptide within the array ─ we observe similar average levels of activity; A: 52.9, B: 43.2, C: 56.6, D: 45.9, E: 46.3%. This result is not unexpected, as the formulations are developed to have a similar overall activity for removing stains.

Thus, in this study, we showed that the combination of peptide arrays and SAMDI mass spectrometry allowed us to profile not only overall proteolytic activity in commercial detergents, but also allowed us to identify differences in the site-specific proteolytic activities of the different compositions. Furthermore, the use of self-assembled monolayers allowed for the selective capture of peptide substrates to the surface in array format while preventing nonspecific adsorption, and the use of MALDI mass spectrometry provided quantitative, label-free detection of proteolytic activity directly on the surface. At ~1.5 seconds per spot, MALDI mass spectrometry also allowed for fast readouts, with one 324-peptide array plate analysis taking 15 minutes. This, in addition to

the compatibility of this method with standard microtiter plate formats and automated liquid handling robots, allowed for thousands of individual reactions to be performed in a day.[122]

Conclusions

This study is the first to use peptide arrays to profile laundry detergents, and to characterize the protease specificities across a large number of peptide substrates. This work allowed a direct comparison of proteases across five commercial detergent brands. We were able to identify detergents having higher protease activity, which may correspond to better protein stain-removing ability. We were also able to compare the specificities of the protease activities for their peptidic substrates, and this demonstration will be relevant to future work that optimizes protease variants for high activity and low specificity. Furthermore, this approach may be useful for developing detergents that are efficient at specific protein stains based on the sequences of those proteins. This demonstration of screening proteases in commercial formulations may, in addition to laundry detergents, be important for screening products in other areas, as well as other enzyme activities. As efforts in protease engineering advance, SAMDI proves a promising method for screening and optimizing new variants for detergents and more.

Methods

Materials. Metal plates having an array of gold islands were prepared as previously described.[122] Fmoc Rink Amide MBHA resin (0.42 mmol/g) was purchased from AnaSpec Inc. Fmoc amino acids were purchased from AAPPTec and Chem-Impex International Inc. All other solvents and reagents for Fmoc solid-phase peptide synthesis were purchased from Thermo Fisher and Sigma Aldrich. Arm & Hammer, Persil, Kirkland, Gain, and Tide laundry detergents were purchased from local grocery stores.

Preparation of Self-Assembled Monolayers on Gold. Steel plates were washed with hexanes, ethanol, water, ethanol, and dried with nitrogen gas. An electron beam evaporator was used to deposit the array of gold spots in a 384 format.[122] The array plates were then soaked in a solution containing a mixture of tri(ethylene glycol)-terminated (EG3) alkanedisulfides and a mixed disulfide of EG3-alkanethiol and a maleimide-terminated EG3-alkanethiol for 24 hours at room temperature to form self-assembled monolayers on the gold surface. The solution of disulfides (1 mM total concentration) had the two disulfides present in an appropriate ratio to yield a final density of maleimide groups of 10%.

Synthesis and Immobilization of Peptide Library. 324 peptides having the general sequence Ac-GXZAFC where X and Z represent all 18 natural amino acids except cysteine and methionine were synthesized using standard Fmoc solid-phase peptide synthesis on Rink Amide resin (0.42 mmol/g) as previously described.[9] The peptides were dissolved in 1% dimethyl sulfoxide (>99.5% pure) in deionized water. The library was then transferred into 384 microtiter well plates and diluted to 100 μM in 200 mM Tris HCl buffer (pH 8.0) using a Multidrop Combi benchtop robot. A 0.25 mM Tris[2-carboxyethyl] phosphine hydrochloride reducing resin in Tris HCl buffer was added to reduce any disulfide bonds from the C-terminal cysteine thiol groups. The peptide library was transferred onto SAMDI array plates (2.5 μL per spot) using the TECAN Freedom EVO automated liquid handling robot with a 384-tip head and the peptides were allowed to react with the maleimide-terminated monolayers in a humidified chamber for 20 minutes at 37°C. The resulting array plate was then rinsed with ethanol, water, ethanol, and dried with nitrogen.

Application of Proteases and Profiling of Enzyme Activity. A Speed Queen top load washer (Model number: AWNE82SP113TW01) was chosen as a representative washing machine,

utilizing 19.5 gallons (73.8 liters) of water for a cold medium wash. Detergent solutions, in water, were prepared as directed by each detergent brand's recommendation for a medium-load cold wash cycle (equivalent of 45-60 mL of each detergent in 73.8 liters). The immobilized peptide arrays were then placed in 100 mL of detergent solution in a reaction chamber, the reaction chamber was sealed and covered with aluminum foil and placed on a shaker for agitation for 20 minutes at room temperature (22-26°C) to mimic the conditions of the cold wash settings. The array plates were then rinsed with ethanol, water, ethanol, dried with nitrogen, and treated with 2',4',6'-trihydroxyacetophenone matrix solution (THAP, 12 mg/ml in acetone). Matrix-assisted laser desorption/ionization time-of-flight mass spectra were acquired on an AB Sciex TOF-TOF 5800 instrument (~1.5 seconds per spot) in reflector positive mode to quantitate the cleavage activity on individual peptide substrates directly on the surface. The data were analyzed using Data Explorer Software to retrieve the integrated areas under the peaks (AUP) of the substrate and product, as follows:

$$\text{Activity} = \frac{\text{AUP}_{\text{product}}}{\text{AUP}_{\text{product}} + \text{AUP}_{\text{substrate}}}$$

Chapter 6: Peptide Arrays and SAMDI-MS to Characterize Enzyme Kinetics of KDACs

This work was completed with the help of Dr. Andreea Stuparu, and Dr. Adam Eisenberg.

Introduction

In Chapter 5 we demonstrated using peptide arrays with SAMDI-MS to monitor the proteolytic activity of enzymes contained within detergents.[17] In this chapter we will return to the roots of the Mrksich group's work with peptide arrays and explore the activity of enzymes responsible for post-translational modifications (PTMs). As mentioned in the introduction, we have looked at many different PTMs, including acetylation, phosphorylation, methylation, glycosylation, and citrullination.[8, 18, 20, 61-65] For this work we will specifically look at lysine deacetylases (KDACs) on an acetylated lysine array. As the name suggests, KDACs are a family of enzymes responsible for the removal of the acetylation PTM on various proteins.[124] KDACs were previously known as HDACs, or Histone deacetylases, as it was originally believed that they were responsible for only the removal of the acetylation PTM on histone tails.[125-127] The KDAC family of enzymes are divided into four classification classes, class I, II, III, and IV, based on sequence homology and localization.[128, 129] Classes I, II, and IV are classified as the zinc-dependent, or classical KDACs. Class I are predominantly nuclear and include KDAC 1-3 and 8. Class II is further divided into two groups, class IIA and IIB, with IIA containing KDAC 4, 5, 7, and 9, being prevalent in both the nucleus and the cytoplasm, and IIB being mostly cytoplasmic, and containing KDAC 6 and 10.[130] Class IV contains a single member, KDAC 11, which shows some sequence overlap, but is not highly homologous to either class I or II KDACs.[131] Class III KDACs differ from the other three classes in that they are a non-zinc dependent KDACs, and instead depend on a molecule of NAD+ as a cofactor, and are known as the Sirtuins.[132] There are

7 Sirtuins, SIRT 1 – 7, unlike members of the other classes however, the sirtuins are not colocalized in a single location, being found in the nucleus, cytoplasm and the mitochondria.[133]

In the past, to distinguish between isoforms of KDACs, the Mrksich group has utilized peptide arrays with SAMDI-MS at fixed concentrations of enzyme and reaction time to obtain a "fingerprint" of the activity of the enzymes.[61] This allowed us to look at the endpoint reactivities of these enzymes on the substrates and elucidate what we believed to be the specificities of the enzymes. However, as with many high-throughput endpoint experiments, a lot of the information, such as the kinetics of the enzymatic reaction are lost in favor of getting a broader picture, and for many enzymes with closely related isoforms, we observe similar reactivity fingerprints. Even for enzymes with unique activity, we observe high activity on multiple peptide substrates, which made determining the best peptide substrates difficult. Due to observing similar endpoint activities with endpoint studies, it made us contemplate a popular quote usually attributed to Ralph Waldo Emerson "it's the not the destination, it's the journey". As the quote suggests, if we allow several boulders, of varying sizes to roll down a hill, eventually most will reach the bottom, however we would not know the path or with what speed they reached their resting positions. Thus, we believe it is important to not only look at the endpoint measurements of enzymatic reactions, but also look at the kinetics of how those measurements were reached. And this is where the benefit of SAMDI-MS becomes apparent.

With SAMDI-MS we can rapidly prepare multiple plates, run enzymatic reactions on each, and analyze in less time than it would take some commercial methods to setup a single run. Also because with SAMDI-MS we monitor the change in mass between substrate and product, we do not need to worry about incorporating labels, which have caused inaccuracies in assays done in

the past, and can use peptide substrates more closely resembling native proteins.[134-136] For this study we chose three KDACs, these are two closely related sirtuin isoforms with similar activities, SIRT2 and SIRT3, and KDAC8 which we have studied extensively in the past.

Results and Discussion

Preparation of Peptide Arrays for use with KDAC Activity Profiling using SAMDI-MS. To demonstrate the usefulness of looking at time dependent activity studies of enzyme isoforms with similar activities or broadly high activity, we first synthesized a 361-member peptide library with the general sequence Ac-GXK^{Ac}ZGC-NH_2, where X and Z are all natural amino acids excluding cysteine. These peptides were synthesized using standard FMOC solid state peptide synthesis techniques on rink-amide resin, cleaved, lyophilized, reconstituted in 0.1% TFA aqueous solution, and arrayed into a standard 384-well plate. Prior to use, the peptides were diluted and buffered to pH 8.0 with a tris buffer. Next, SAMDI plates with pre-evaporated 384 gold islands in standard microtiter plates format were soaked in an ethanolic solution of our SAMs with a 10% maleimide density against a background of tri(ethylene glycol) groups.[9] The maleimide functionality allows for the selective immobilization of the cysteine terminated peptide, while tri(ethylene glycol) groups provide resistance to non-specific adsorption during peptide immobilization along with further enzyme reaction steps.[23] Once the monolayers and the peptides were ready, the buffered solutions of peptides were transferred onto the SAMDI plates and the peptides were allowed to immobilize for half an hour. Upon immobilization the remaining peptide solutions were washed away with water and ethanol, and the residual ethanol was dried under nitrogen. The SAMDI plates with immobilized peptides were then coated with matrix (30 mg/mL 2,4,6-trihydroxyacetophenone (THAP) in acetone) and analyzed inside the MALDI instrument to determine the quality of the plates, and to also obtain baseline activity for each plate. Upon

completion of the run, each plate was removed from the MALDI instrument, the matrix washed off, and each plate was stored under nitrogen ready for use with enzymes. By running each plate individually prior to enzymatic reactions, we were able to obtain a zero-time time point, as well as confirm the quality of each SAMDI plate prior to reacting with enzyme of interest.

KDAC Kinetic Assay. KDACs were expressed in-house using E. Coli and their activities were confirmed using conditions and peptides that have shown good activity in the past.[61] Once the activity of the enzymes were confirmed, we next moved to optimize the enzyme and, when necessary, NAD+ cofactor, concentrations. Our goal was to determine conditions which would give each enzyme roughly 90% deacetylation activity on the best peptide substrates after a 30 minute reaction at room temperature (Figure 6.1). We then ran each enzyme once more, for 10 minutes, using the same conditions to confirm we observe deacetylation of the best substrate roughly in the 40-50% range, to ensure we can obtain useful kinetic data, adjusting concentration of enzyme appropriately if needed. The final concentrations of enzymes and cofactors can be found in Table 6.1.

Once appropriate SIRT3 and cofactor concentrations were determined, deacetylation reactions were run with enzyme at four different timepoints, 2, 4, 6, and 10 minutes, in triplicate. To ensure quenching of the reactions at the appropriate timepoints, SAMDI plates with spotted enzyme solutions were dropped into a container of ethanol, diluting and denaturing the enzyme. The plates were then further washed with water and ethanol, covered in matrix and SAMDI spectra were obtained. The obtained spectra were then exported and analyzed with our in-house "Profiler" software to give the percent deacetylation. The deacetylation data was then imported into

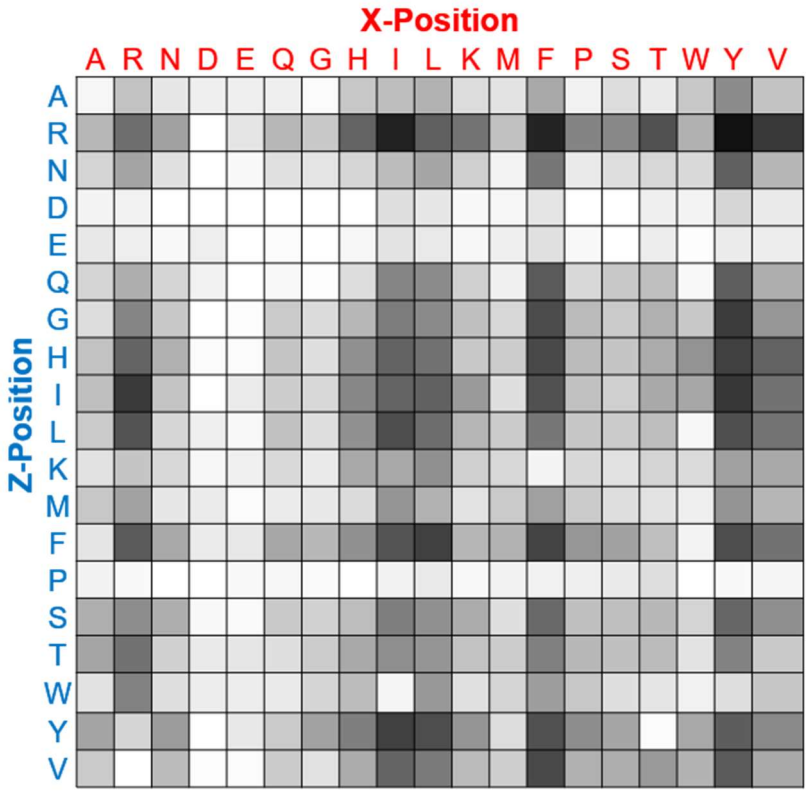


Figure 6.1: SIRT3 30 minute endpoint enzymatic activity heatmap of the Ac-GXK^{Ac}ZGC-NH_2 peptide library. 250 nM SIRT3 and 20 µM NAD+ were incubated for 30 minutes at room temperature and analyzed via SAMDI-MS. The deacetylation activity of the enzyme were calculated and mapped onto the heatmap grid with the grayscale ranging from white (0% activity) to black (100% activity). The maximum activity of 94% deacetylation can be found on this heatmap at the sequence Ac-GYK^{Ac}RGC-NH_2.

	Enzyme Concentration	Cofactor Concentration
KDAC8	250 nM	N/A
SIRT2	250 nM	200 µM
SIRT3	250 nM	20 µM

Table 6.1: The concentration of KDAC enzyme and cofactor, when necessary, determined to give good range of activity for kinetic studies.

Excel and further processed by subtracting the background and averaging the data for each timepoint. Following this, the initial velocity slopes were calculated, and their values were inserted

into our heatmap format, clearly showing a preference for arginine in the Z position by SIRT3 (Figure 6.2).

Table 6.2 summarizes the 8 peptides with the highest slope for SIRT3. Although the four most active peptides remain the same for the endpoint study and this kinetic study, we do notice they have quite a difference in the slope. For example, the two peptides Ac-GYK^{Ac}RGC and Ac-GLK^{Ac}RGC are deacetylated to a high degree by SIRT3, 94% and 92% respectively, but their slopes differ by 30%, 0.061 and 0.043. Furthermore, we see some peptides show a high initial slope, such as Ac-GYK^{Ac}TGC, but show lower endpoint conversion. We hypothesize this could be due to auto-inhibition of the enzymes by already deacetylated peptides found on the surface, similar to activation of KDAC8 by allosteric site interactions with longer substrate peptides.[137]

We also repeated the kinetic studies on KDAC8 and SIRT2, and as with SIRT3 can clearly see that the slopes of only few peptides dominate the heatmaps. For KDAC8 we can clearly see that the three most active peptides all have arginine in the X position and an aromatic residue in the Z position (Figure 6.3), matching the endpoint experiments that have been reported before.[137] For SIRT2 we see a different preference of peptide, compared to SIRT3, preferring aromatic residues in both the X and Z positions, and indifferent for the presence of arginine in the Z position (Figure 6.4).

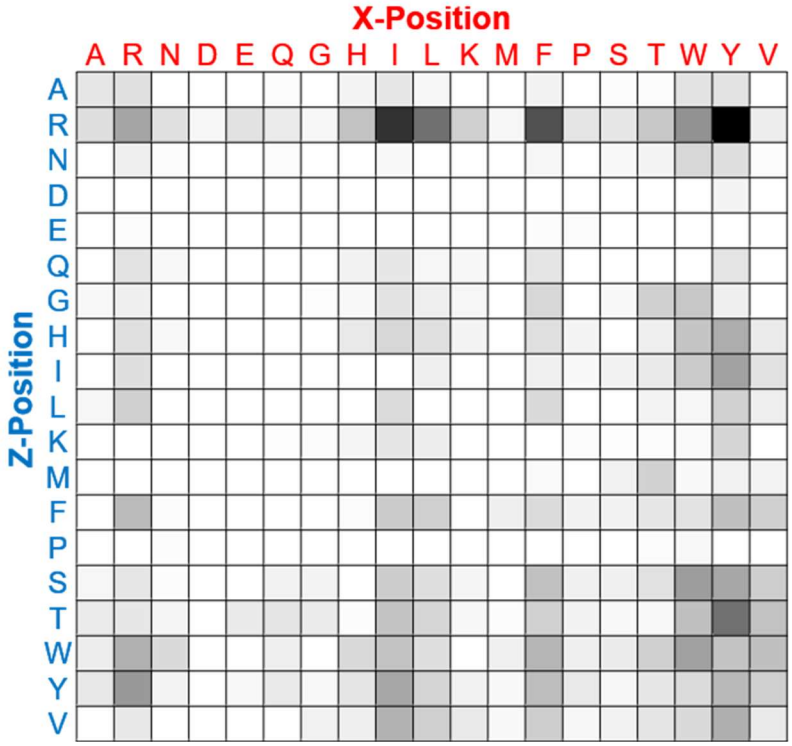


Figure 6.2: SIRT3 initial velocity slope heatmap. The slopes of the initial velocity curves for 250 nM SIRT3 with 20 µM NAD+ were calculated from 0, 2, 4, 8, and 10-minute reactions. These were then inserted into the heatmap format for visualization. Here white squares represent no deacetylation activity and the black square at Ac-GYK^{Ac}RGC-NH_2 represents the highest slope value (0.061, 55% deacetylation activity at 10 minutes).

Peptide Sequence	Slope	% Yield at 30 minutes
Ac-GYK^{Ac}RGC	0.061	94%
Ac-GIK^{Ac}RGC	0.053	93%
Ac-GFK^{Ac}RGC	0.049	93%
Ac-GLK^{Ac}RGC	0.043	92%
Ac-GYK^{Ac}TGC	0.043	65%
Ac-GWK^{Ac}RGC	0.038	88%
Ac-GRK^{Ac}YGC	0.037	92%
Ac-GWK^{Ac}SGC	0.036	64%

Table 6.2: Eight peptides with the highest slopes for SIRT3 with their sequence and % deacetylation at 30 minute timepoint.

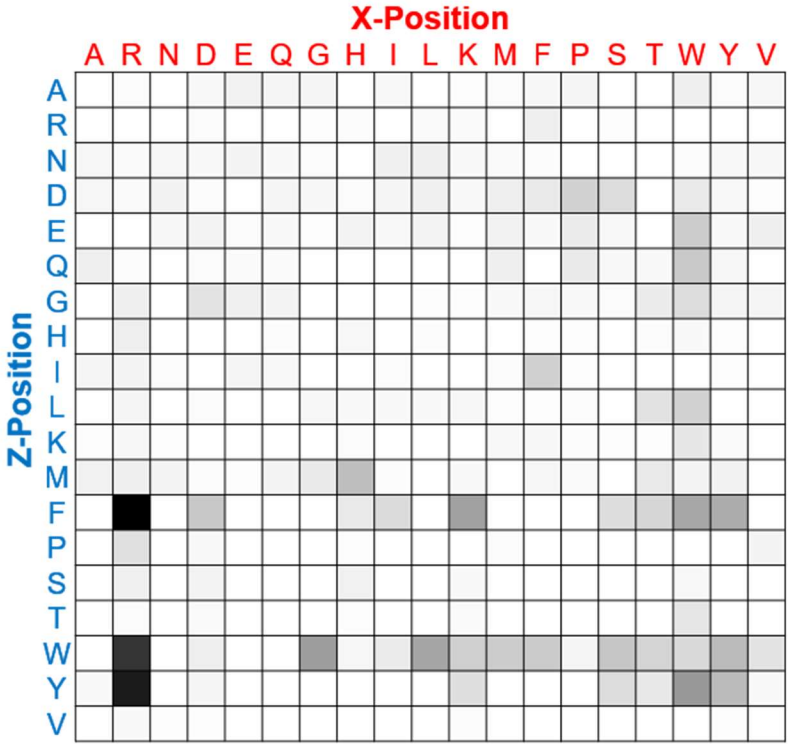


Figure 6.3: KDAC8 initial velocity slope heatmap. The slopes of the initial velocity curves for 250 nM KDAC8 were calculated from 0, 2, 4, 8, and 10-minute reactions. These were then inserted into the heatmap format for visualization. Here white squares represent no deacetylation activity and the black square at Ac-GRK^{Ac}FGC-NH_2 represents the highest slope value (0.063, 32% deacetylation activity at 10 minutes).

Conclusion

While we have shown that SAMDI-MS and peptide arrays are an invaluable tool at looking at endpoint enzymatic assays, looking at endpoint studies only give us a "fingerprint" of the enzyme's activity on the peptide array. However, SAMDI-MS does not limit us to only looking at endpoint studies. Because SAMDI-MS assays do not require purification steps prior to analysis, running a kinetic study is as simple as running extra SAMDI plates at each timepoints of interest and analyzing it with the MALDI instrument. We believe that by looking at the kinetics of an enzyme, the slopes of the initial velocity curves in this example, we can better determine the best substrates for the enzymes of interest. Also, by looking at initial rate slopes of related enzyme

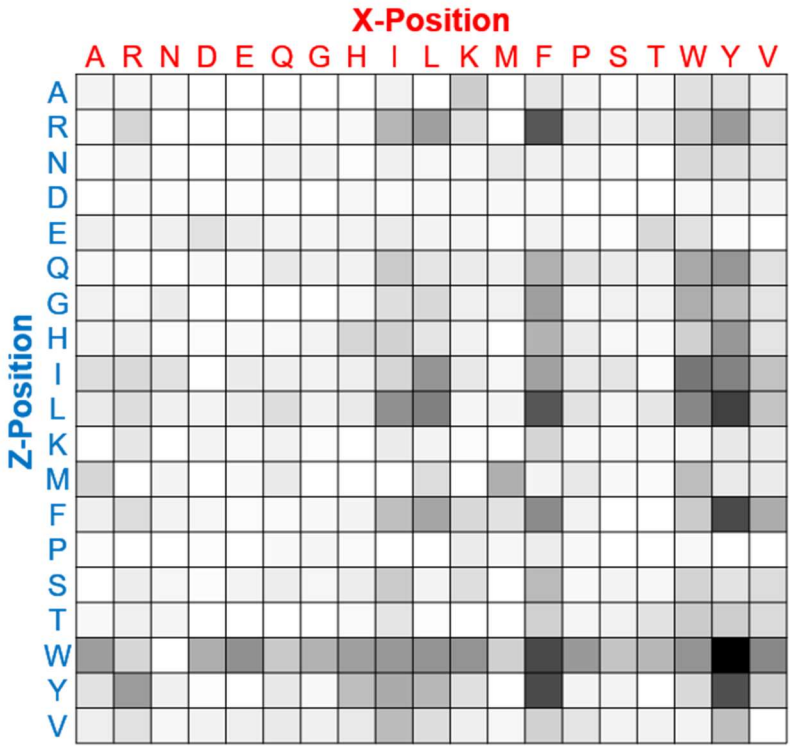


Figure 6.4: SIRT2 initial velocity slope heatmap. The slopes of the initial velocity curves for 250 nM SIRT2 with 200 µM NAD+ were calculated from 0, 2, 4, 8, and 10-minute reactions. These were then inserted into the heatmap format for visualization. Here white squares represent no deacetylation activity and the black square at Ac-GYK^{Ac}WGC-NH_2 represents the highest slope value (0.034, 35% deacetylation activity at 10 minutes).

isoforms, such as SIRT2 and SIRT3, it might be possible to determine the ratio of these two enzymes in a mixture by looking at the deacetylation of certain peptides and deducing the contribution of each enzyme using initial slopes.

Methods

Materials: Fmoc-amino acids, Fmoc-Rink amide resin, and N-acetylglycine were purchased from Anaspec. Fmoc-Lys(Ac)-OH was purchased from Novabiochem. Triethylsilane (TES), ethanedithiol (EDT), piperidine, and N-methylmorpholine (NMM) were purchased from Sigma Aldrich. All other peptide synthesis reagents were purchased from VWR.

Peptide Synthesis and Peptide Solution Preparation: Peptides were synthesized by hand using standard Fmoc solid-phase peptide synthesis techniques on Fmoc Rink-Amide resin. The initial and subsequent Fmoc deprotection steps were performed by adding enough solution of 20% piperidine in N,N-dimethylformamide (DMF) to cover the resin, and allowing it to react for 15 minutes. The amino acids were coupled by adding 4-molar equivalents of Fmoc-amino acids, PyBop, and 8-molar equivalents of NMM, in DMF and allowing to couple for 20 minutes. The resin was rinsed between each coupling and deprotection steps by rinsing with DMF 5 times. To streamline synthesis, the first two residues, cysteine and glycine, were coupled to all the resin in one large batch. Following the coupling of the glycine, the resin was split into 19 separate containers, one for each amino acid in the Z position. Then the Z amino acid and acetylated lysine were coupled, followed by splitting of each container into further 19 containers. The synthesis was then finished off by coupling the X position amino acids and finally with N-acetylglycine. Following the synthesis of the peptides, the resin in each tube was washed with DCM and dried under vacuum. To cleave the peptides from the resin and deprotect the sidechain protecting groups, a solution of 95% TFA, 2.5% water, 1.25% TES, and 1.25% EDT were added and allowed to sit for 2 hours at room temperature. The TFA solution was then filtered, to remove the resin, and the peptides were precipitated out of solution with cold diethylether. The precipitant was centrifuge down, ether decanted and the slurry washed again with cold ether. After the 2^{nd} wash, the slurry was dried, then dissolved in water and lyophilized to produce a peptide powder. This peptide powder was then dissolved in an aqueous 0.1% TFA solution, with the addition of minimal amount of DMSO to increase solubility. The peptides were then arrayed into 384 well plates and frozen until needed. When required, peptide solutions were diluted 100-fold in KDAC Classic buffer (50 mM Tris pH8, 137 mM NaCl, 2.7 mM KCl, 1mM $MgCl_2$) prior to immobilization.

Preparation of Self-Assembled Monolayers on Gold. Steel plates were washed with hexanes, ethanol, water, ethanol, and dried under nitrogen gas. An electron beam evaporator was used to deposit the array of gold spots in a 384 format.[122] The array plates were then soaked in a solution containing a mixture of tri(ethylene glycol)-terminated (EG3) alkanedisulfides and a mixed disulfide of EG3-alkanethiol and a maleimide-terminated EG3-alkanethiol for 48 hours at -20 °C to form self-assembled monolayers on the gold surface. The solution of disulfides (1 mM total concentration) had the two disulfides present in an appropriate ratio to yield a final density of maleimide groups of 10%. Once the monolayers were formed, the plates were removed from the solution, washed with ethanol, water, ethanol, and dried under nitrogen. They were stored under vacuum awaiting peptide immobilization.

Peptide Immobilization. The arrayed and diluted 361 member library of peptides were transferred to the prepared gold surfaces with monolayers using a 96-head liquid handling robot (TECAN EVO). 2 μL of peptide solution from each well was transferred onto the surfaces and allowed to immobilize for 30 minutes at room temperature in a humidified chamber. Following immobilization, the peptide solutions were washed off the plate with water and ethanol, dried and set aside, awaiting enzymatic reactions or analysis by MALD-TOF.

SIRT2 and SIRT3 Expression and Purification. The genes encoding SIRT2 and 3 genes were purchased from Addgene and introduced, via ligation independent cloning, into pTriEx-3 neo (Novagen EMD Millipore) with the following primers: SIRT2 forward 5'-GGCACAACACCCATG-3', and reverse 5'-GACTTCCTGCGGAACTTATTCTCCCAG-3'; SIRT3 forward 5'-GGCACAACACCCATG-3', and reverse 5'-GCGTTCTGGGGTTGGCGCGCC-3'. The plasmid was then transformed into E. coli (DH5α,

Novagen) and colonies bearing the correct inserts were confirmed via DNA sequencing. The recombinant plasmid was then transformed into E. coli [BL21(DE3), Invitrogen]. The cells were grown in 2XTY media and induced using a 0.5 mM isopropyl-B-D-galactopyranoside (IPTG) overnight at 18 ºC after reaching an OD600 of 0.4-0.6. The cells were then pelleted via centrifugation and lysed by sonication in the presence of protease inhibitors (Roche). The His-tagged protein was purified via Co^{2+} HisPur affinity resin (Thermo) and eluted using 300 mM imidazole in 50 mM Tris-HCl pH 7.5, 200 mM NaCl, 5% glycerol, and 5 mM beta-mercaptoethanol. Pure fractions were determined by SDS-PAGE on a 4-12% polyacrylimide gel (BioRad) and then concentrated and dialyzed against storage buffer (50 mM Tris-HCl pH 8.0, 10% glycerol, 0.1 mM EDTA). The protein was flash frozen and stored at -80 °C until use. Protein concentration was determined on a NanoDrop spectrophotometer (Thermo Scientific).

KDAC8 expression and purification. KDAC8 was cloned and expressed as previously reported.[138] Briefly, KDAC8 was amplified by PCR using primers allowing for Nhe1 and Xho1 restriction sites: forward 5'-3' TATTCTCGAGGACCACATGCTTCA Xho1 and reverse 5'-3' ATAAGCTAGCATGGAGGAGCCGGA Nhe1. After digestion with Nhe1 and Xho1 the products were ligated into vector pet21a and transformed into E. coli [BL21(DE3), Invitrogen] for protein expression. The cells were grown in 2XTY media until an OD600 of 0.4 – 0.6 and then induced with 0.5 mM IPTG and 0.2 mM ZnCl2 for 4 hours at 37 ºC. The cells were then centrifuged down and lysed via sonication in the presence of protease inhibitors (Roche). KDAC8 was purified via His-tag affinity chromatography using a Co^{2+} HisPur affinity resin. Pure fractions were determined by SDS-PAGE on a 4-12% polyacrylimide gel and then concentrated and dialyzed against storage buffer (20 mM Tris-HCl pH 8.0, 50 mM NaCl, 3mM $MgCl_2$, 7mM KCl, 10% glycerol). The

protein was then flash frozen and stored at -80 ºC. Protein concentration was determined on a NanoDrop spectrophotometer.

KDAC Enzymatic Reactions. In-house expressed KDACs were removed from -80 ºC storage and allowed to thaw on ice. Once thawed, desired amounts of KDACs were transferred to a new tube, along with a stock of NAD+ and diluted using KDAC classic buffer (50 mM Tris pH8, 137 mM NaCl, 2.7 mM KCl, 1mM $MgCl_2$) to a final working concentration. Following this a Multidrop Combi (Thermo Scientific) was used to dispense 2.5 μL of enzyme solution onto each spot of a 384 spot SAMDI array plate with immobilized peptides. The enzyme was allowed to sit on the surface, in a humidified chamber at room temperature, for a predetermined length of time and then the plate submerged in a bath of ethanol, to quench the reaction. The plate was then removed from the ethanol bath, rinsed with water and ethanol, dried under nitrogen and covered in 2',4',6'-trihydroxyacetophenone (THAP) matrix (30 mg/mL in acetone) in preparation for analysis by the MALDI instrument.

SAMDI-MS. The gold array plates with immobilized peptides, before or after enzyme reaction, covered with THAP matrix, were inserted into the ABSciex 5800 MALDI TOF/TOF instrument. Using the automatic reflector positive mode, 400 shots/spot were acquired for each spot with immobilized peptide. The spectra for each plate was then exported and deacetylation activity was analyzed using the Profiler software.

Spectral Analysis via Profiler Software. The automatically acquired SAMDI spectra were exported using the TOF/TOF Series Explorer Software (ABSciex). The exported spectra files were then run through the Profiler software, generating a list of areas under the curve (AUC) for each

product and reactant peptides of interest. These values were then used to determine the activity of the enzyme using the equation: $\text{Activity} = \frac{\text{AUP}_{\text{product}}}{\text{AUP}_{\text{product}} + \text{AUP}_{\text{substrate}}}$.

Chapter 7: Synthesis, Analysis, and Utilization of Random Peptide Arrays for SAMDI-MS

The work in this chapter was accomplished with the help of Raymond Dai, Matthew Chow, Carolyn Ma, Dr. Caleb Joseph, and Dr. Pradeep Bugga

Introduction

With advances to peptide synthesis technology, the utilization of peptide arrays to answer important scientific questions has been on the rise.[56] These uses include profiling enzymes, antibodies, and ligand-receptor interactions, as well as discovering new cell adhesion ligands, to name a few. [8, 18, 20, 24, 61-70] However, the commonly used techniques to isolate and identify peptides of interest are both labor and time intensive.[139] Thus, we have introduced the use of peptide arrays with SAMDI-MS (self-assembled monolayers for matrix-assisted laser desorption/ionization MS).[9, 13-16] SAMDI utilizes self-assembled monolayers (SAMs) of alkane disulfide on gold that present an immobilization functionalization, at a set density, against an inert background of polyethylene glycol (PEG).[3, 86] The SAM functionalization, most commonly maleimide for peptide arrays, allows for specific pulldown of cysteine containing peptides from complex reaction mixtures and the PEG background prevents non-specific adhesion to the surfaces, allowing impurities and non-attached molecules to be simply washed away.[3, 56] The immobilized peptides can then be analyzed using a MALDI-TOF. Because the monitoring technique is a MS based one, any native enzymatic activity that incur a mass change can be monitored quantitatively and without the use of labels, as has previously been shown to be important when looking at enzymatic activity.[140, 141] We have also previously shown that peptide-protein interactions that do not directly incur a mass change can be analyzed by conjugating such proteins to enzymes that do, or by further modifying the resulting peptides.[18, 62, 96] Of course SAMDI-MS's utility is not only limited to

peptides with enzymatic reactions, the technique and peptides can also be used to monitor chemical reactions, as well as used in determining minimal cell binding motifs.[5, 21, 142-145]

Although the usefulness of peptide arrays has become widely apparent, the synthesis of large arrays remains a problem, even with recent advances.[56, 146] Some of the major techniques introduced in the last half century include solid-phase peptide synthesis (SPPS), which allowed for the synthesis of peptides on solid scaffolds and negated the need for tedious purification between each synthesis steps.[55] Also, the introduction of new N-terminal protection and deprotection schemes, such as tert-butoxycarbonyl (Boc) and fluorenylmethoxycarbony (Fmoc), which enabled faster and milder reaction conditions.[57] These advances along with the introduction of a wide range of solid supports, attachment chemistries, and an increase in available natural and unnatural amino acids and their derivatives, have broadened the possibilities of the libraries that can be synthesized in a laboratory setting.[58] However, even with all of the advances since Merrifield's original SPPS paper, currently lab made peptide libraries are often limited to numbers in the hundreds. Although the cost of a single peptide is relatively low, making an arrays of thousands or even tens of thousands quickly become prohibitively expensive and extremely time consuming.[14] Of course, commercial sources exist, which can minimize cost and eliminate personnel time spent on synthesis. These sources provide peptides in many forms, including surface bound SPOT microarrays by PEPperPRINT, a format similar to DNA microchip arrays, or more familiar lyophilized powders in microwells.[71] Both commercial sources of peptides have their advantages and drawbacks. For example, the microarray formats provide fast synthesis with specially built inkjet printers, however the arrays are single use and must be discarded after use, and any reaction must be done on the solid support surfaces, which can limit possible reactions and analysis methods. On the other hand, the lyophilized peptides are slower to synthesize, but can be used

multiple times and not limited in usage. Although both commercial sources formats have their pluses and minuses, ultimately the costs associated with purchasing very large peptide arrays can be prohibitively expensive.

One proposed way to combat the bubbling price of synthesizing large arrays is to employ a single-molecule-per-bead synthesis strategy.[147, 148] This technique is commonly associated with utilizing the split-pool technique, whereby steps of combining and randomly distributing beads into reaction vessels are repeated to achieve the desired diversity in the library.[149] Using this method's molecules, such as oligonucleotides or peptides, are synthesized in batch, with beads expressing a single random sequence from those synthesized. Because each bead contains a random sequence, a decoding step must be taken to determine the identity of the sequence. Two commonly used methods are fluorescence and mass spectrometry.[150-152] Typically, the fluorescence method is utilized with oligonucleotides, whereby a complementary DNA sequence with a fluorescent marker is attached to a probe bead, which in itself contains varying concentrations of fluorescent dyes. These complementary DNA sequences are then allowed to hybridize to immobilized DNA sequences on target beads. The target beads rest in etched microwells on a fiber-optic cable, allowing for direct readout and identification of each bead via fluorescence markers of the probing beads.[151] On the other hand the mass spectrometry techniques deal with sequence determination in one of two ways, tandem MS, or by building in the sequence history into the synthesis itself.[147, 148, 152-155]

In the second method, a small portion of the growing peptide chain is capped to give a mass-ladder, which as the name suggests, could be followed to determine the synthesis history of the peptide. Upon completion of the synthesis the full-length peptide, along with the truncated

sequences, are cleaved off the bead and the sequence of the peptides are determined, typically with a MALDI-TOF based technique. Historically however, the capping schemes used for this technique either provided ambiguity in decoding sequences with isobaric amino acids or employed a capping scheme that required a large number of capping groups.[148, 152, 153]

In this chapter we report a method for synthesizing random peptide arrays which combines a split-pool combinatorial method with a novel mass encoding scheme in association with SAMDI-MS for analysis and utilization of said arrays. This enables the synthesis of large peptide libraries cheaply and quickly, which enables data driven decisions for subsequent peptide arrays. We demonstrate this approach by synthesizing a 5,832 member random peptide in an attempt to identify a novel cell adhesion substrate for the HUVEC cell line.

Results and Discussion

General Random Array Peptide Synthesis Method. For synthesizing random peptide arrays, the first amino acid is first immobilized onto Tentagel-MB-OH beads via an ester bond using the Steglich esterification reaction.[156] Using Tentagel-MB-OH beads provides two benefits; first, the large size of the beads, approximately 300 micron, simplifies bead sorting, and secondly, the ester linkage of the first amino acid to the solid scaffold is orthogonal to the standard TFA deprotection of the sidechain protecting groups. This enables the deprotection of the sidechain groups without cleavage of the peptides from the resin, eliminating complex purification. A minimum of ten-fold excess of beads, compared to the number of peptides to be synthesized, calculated from a given value of 65,550 beads/gram, is used for the synthesis of each array – rationale to follow. The growing peptide chains are then elongated utilizing standard Fmoc chemistry. Once the random position of the peptide sequence is reached the split-pool technique is employed, whereby the pool

of beads is split into appropriate number of vessels, and the random position residues coupled. With each addition of a residue in the random position, a small portion of the growing peptide chain is terminated with a chemical capping group, as with previously reported methods (Figure 7.1).[148, 152, 153] Following this, the beads are grouped again and any non-random residues are added, the sidechains deprotected via standard TFA cleavage, and beads are dried. The beads are then sorted into wells of a 384-well plate, using a COPAS FP-500 instrument, and a cleavage solution of 7% triethylamine (TEA) in water is added, which cleaves the ester bond, releasing the peptides from the beads. Using the procedure outlined here, it is possible to synthesize 5,500 single peptide beads per a standard SPPS synthesis using 50 mg of rink-amide resin, effectively reducing the cost of a single peptide by more than 5,000-fold.

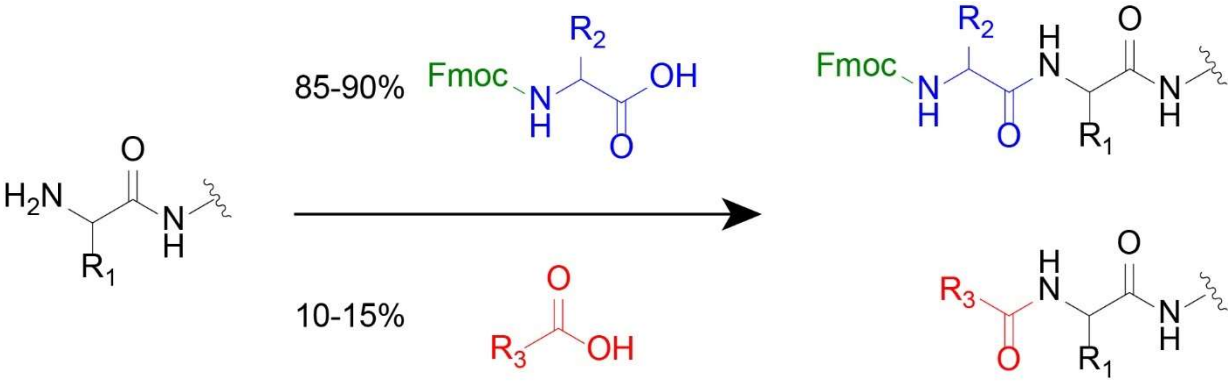


Figure 7.1: General synthesis scheme during random position amino acid addition. The coupling is accomplished using regular Fmoc SPPS methods.

Method and Capping Group Validation. To first confirm the general method and the viability of the selected capping groups, we synthesized a 1000-member peptide library with the sequence Ac-G($AA_3AA_2AA_1$)-GRC, where $AA_{1\text{-}3}$ are random amino acid positions, and arginine was included to help with ionization. 10 amino acids were chosen for the random positions, which were: glycine, alanine, serine, proline, valine, leucine, isoleucine, glutamine, glutamic acid, and

phenylalanine. Our goal with the proposed capping scheme was to select the minimal number of capping groups that would allow us to fully decode the library, that is to make sure that each peptide within the array has a unique combination of full-length and mass-ladder masses. For this purpose, four commercially available carboxylic acids were chosen as the capping groups. Using these four commercially available carboxylic acids provided adequate separation of masses, even for isobaric residues, which led to a fully decodable random peptide library, while also decreasing synthesis cost, as the carboxylic acids chosen here are less expensive than the Fmoc amino acids. The selected capping groups and their associated amino acids can be found in Table 7.1.

Amino Acid	Capping Group	Δm
Glycine	Acetic acid	-15
Alanine	Acetic acid	-29
Leucine	Acetic acid	-71
Phenylalanine	Acetic acid	-105
Arginine	Acetic acid	-114
Proline	Propanoic acid	-41
Thrionine	Propanoic acid	-45
Lysine	Propanoic acid	-72
Histidine	Propanoic acid	-81
Tyrosine	Propanoic acid	-107
Tryptophan	Propanoic acid	-130
Asparagine	Benzoic acid	-10
Glutamine	Benzoic acid	-24
Serine	p-Toluic acid	31
Isoleucine	p-Toluic acid	5
Aspartic acid	p-Toluic acid	3
Glutamic acid	p-Toluic acid	-11
Valine	p-Toluic acid	19

Table 7.1: Amino acids and their corresponding protecting groups for random array synthesis. Each of the amino acids is associated with a capping group for the capping of the growing peptide chain; here we also include a Δm value as the mass-difference from the expected free amine peak associated with the addition of the amino acid.

After completing the synthesis, the sidechains of the amino acids were cleaved using standard TFA cleavage protocol. The beads were then washed to remove any TFA, dried and

prepared for dispensation into well plates. A single bead was sorted into each well of a 384-well plate using a COPAS FP-500 instrument, to ensure each well only had one full-length peptide and the truncated mass-ladder peptides. 5 μL of an aqueous 7% triethylamine (TEA) solution was then distributed into each well using a multidrop combi robot. This caused the ester bonds between the peptides and the resin to cleave, releasing the peptides into solution. The peptides were then buffered with 25 μL 100 mM tris pH 8.0 (~100 μM final peptide concentration) and treated with TCEP beads. The supernatant from each well was transferred onto a SAMDI plate, onto which a self-assembled monolayers of alkanethiolates terminated in maleimide groups at a density of 20% against a background of tri(ethylene glycol)-groups was formed.[9] After a 30 minute reaction with the surface, the liquid was washed off with ethanol, water, ethanol, and the plate dried under nitrogen. The dry plate was then covered in THAP matrix (12 mg/mL in Acetone) and loaded into an ABSciex 5800 MALDI instrument. Spectra of each spot were obtained in automated reflector positive mode. The obtained spectra were then analyzed by observing the peaks present in the spectra. The general workflow of the process, as well as the ladder formation can be found in Figure 7.2.

For this library, we expect to see four peaks, one corresponding to the full-length peptide, and three for the truncated mass-ladder peptides. Figure 7.3 provides an example of an acquired and decoded spectra corresponding to the peptide Ac-GLVIGRC, for ease of visualizing the capping scheme, we have color-coded each of the peaks, residue and capping group in the spectra. For this example, the full-length peptide mass is 1610 (black peak), this mass corresponds to one of 18 possible peptides that were synthesized within this array (Table 7.2). To decode which peptide is represented by the spectrum, we simply have to match the peaks that are present to those

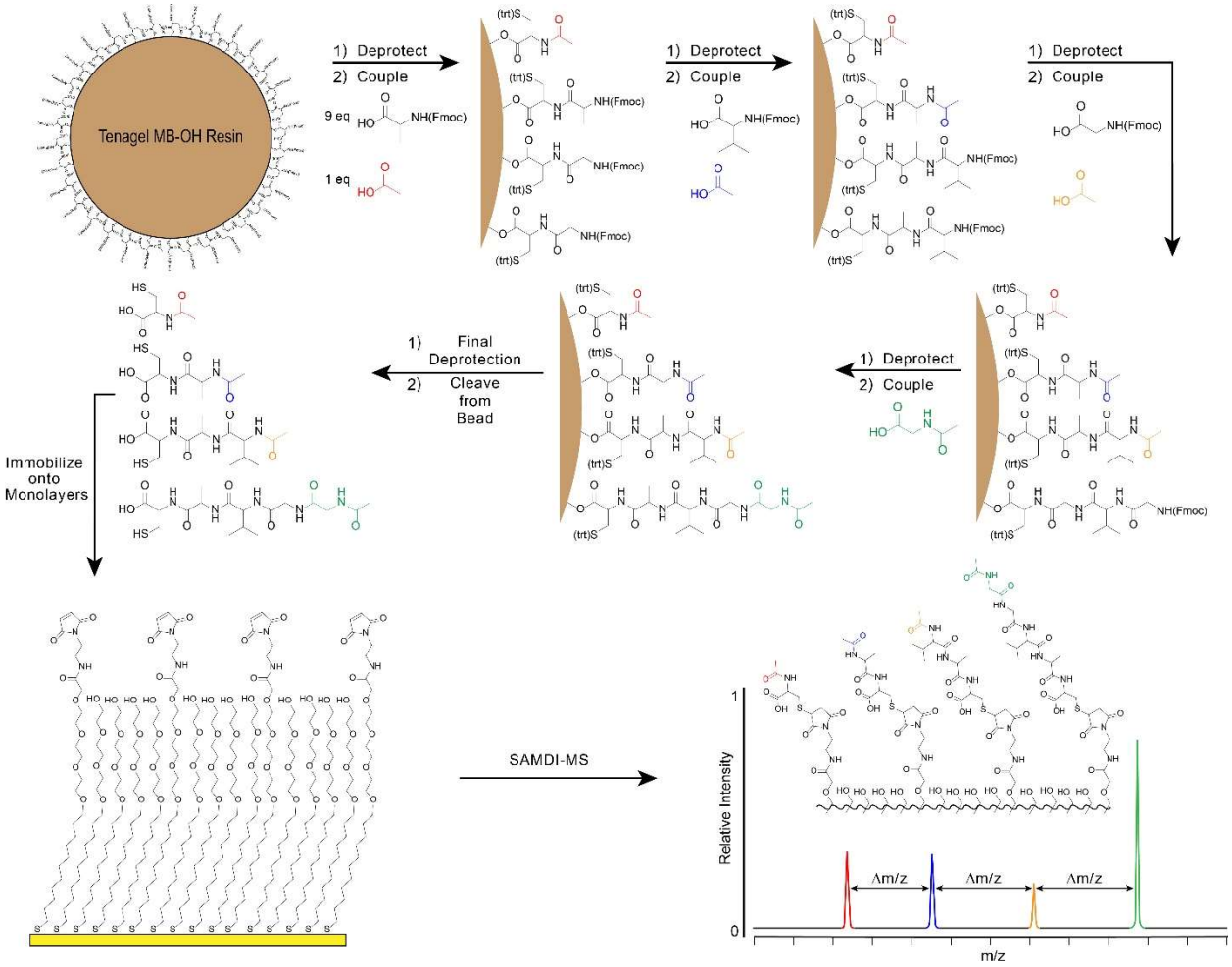


Figure 7.2: Overview of the random peptide array synthesis method. The first amino acid is attached to the Tentagel MB-OH resin through an ester bond. The peptide chain is then elongated via standard Fmoc peptide synthesis methods. Once the random positions of the peptide library are reached, a carboxylic acid, which acts as a chain terminating group, is co-reacted with an Fmoc amino acid to terminate a portion of the growing peptide chain, forming a mass-ladder peptide. This is repeated several more times, depending on the number of random positions. The sidechains of the peptides are then cleaved using standard TFA cleavage, which is orthogonal to the peptide-bead attachment chemistry. Beads are then sorted into wells, peptides cleaved off, and the free peptides are immobilized onto SAMDI plates. The immobilized peptides are then read via SAMDI-MS and the identity of the peptides determined by decoding the mass-ladder generated during the random position coupling steps.

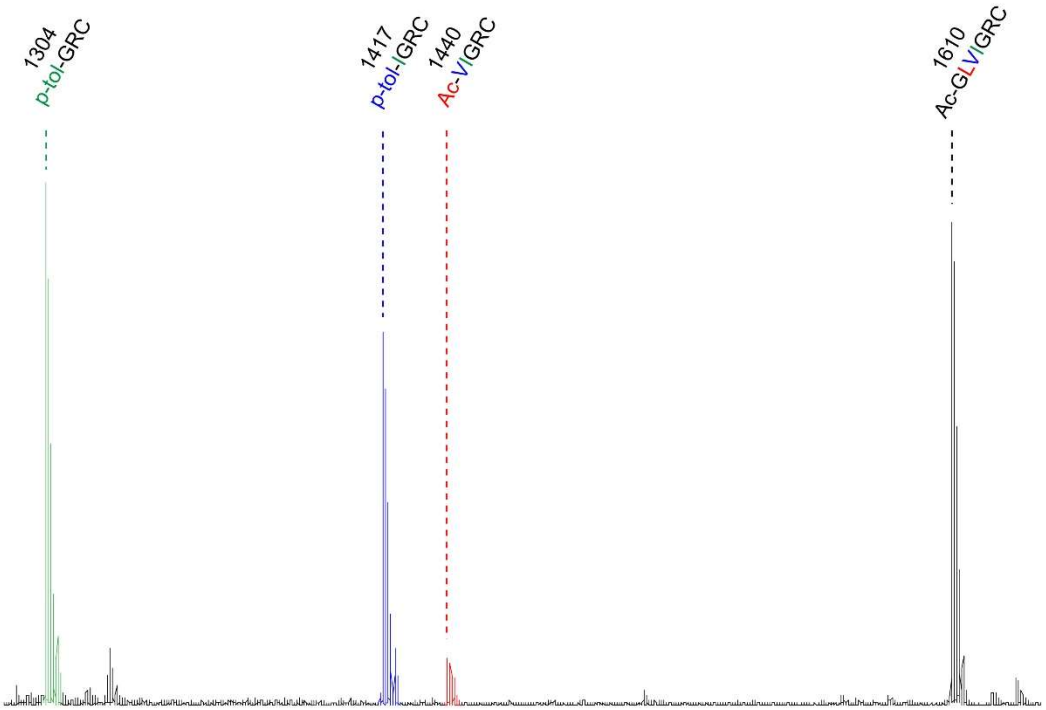


Figure 7.3: Example decoded spectra of Ac-GLVIGRC peptide synthesized via random peptide synthesis. The four important peaks are labeled, the colors of the peaks correspond to the random position residues and protecting group that were added during the random position synthesis. These protecting group allows for the visualization of the mass ladder by terminating a small portion of the growing peptide chain.

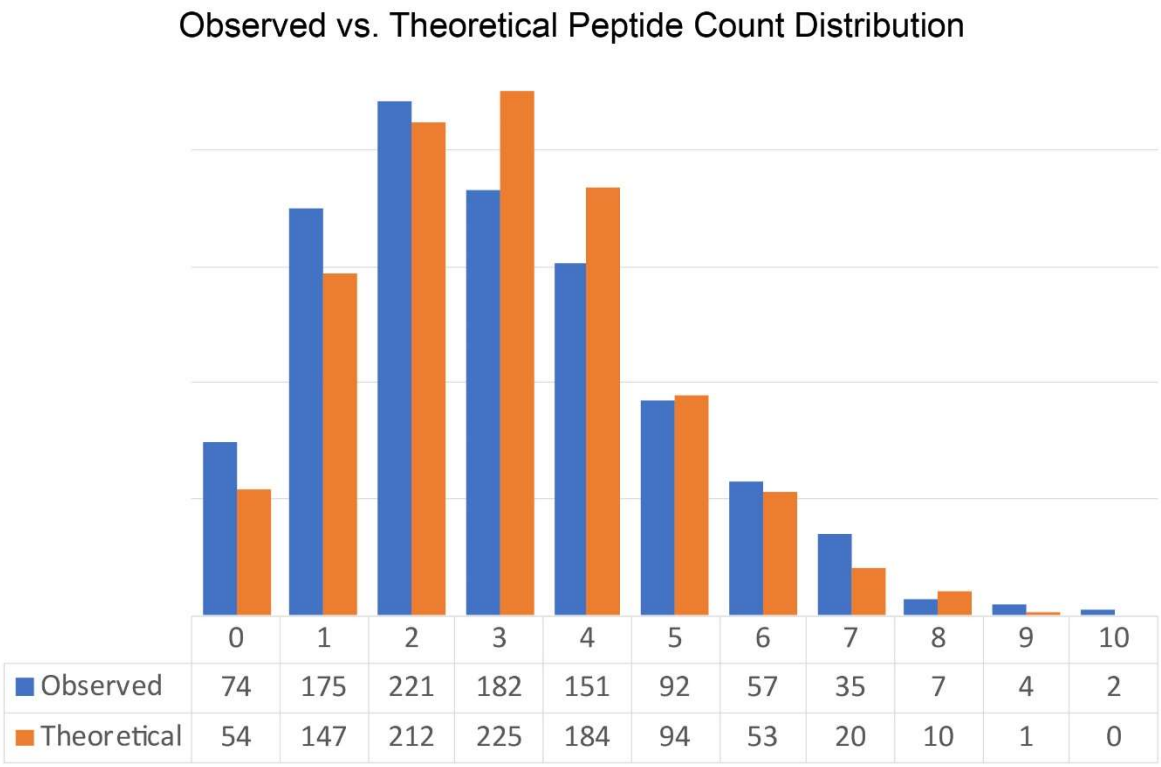
Random Sequence	Full Length Mass	1st Ladder Mass	2nd Ladder Mass	3rd Ladder Mass
LLV	1610	1440	1327	1304
LIV	1610	1440	1403	1304
LVL	1610	1440	1417	1228
LVI	1610	1440	1417	1304
PEV	1610	1470	1403	1304
PVE	1610	1470	1433	1304
EPV	1610	1500	1341	1304
EVP	1610	1500	1401	1242
ILV	1610	1516	1327	1304
IIV	1610	1516	1403	1304
IVL	1610	1516	1417	1228
IVI	1610	1516	1417	1304
VLL	1610	1530	1341	1228
VLI	1610	1530	1341	1304
VPE	1610	1530	1371	1304
VEP	1610	1530	1401	1242
VIL	1610	1530	1417	1228
VII	1610	1530	1417	1304

Table 7.2: All possible peptides matching the full-length mass of Figure 7.3.

in Table 7.2. The decoding table for all 1000 peptides (Appendix Table 7.1) was constructed by calculating the mass of the full-length peptide along with the mass of the free-amine peptide, corresponding to the addition of the random position residue, and adding or subtracting the Δm value found in Table 7.1. With this table in hand, the peptide with the three mass-ladder peaks, 1440, 1417, and 1304, is the peptide with LVI in the random positions. It must be noted that the apparent abundance of each peak within the spectra does not necessarily correspond to the actual percentage of capping group used in each position. We believe that the most likely reasoning behind this is the ionization efficiency differences of each peptide and capping group mixture.[157]

Next, to confirm there is no synthesis bias, as well as to ensure there is no difference in the difficulty to decode the mass-ladder peptides with different capping groups, we employed the same method used to decode the LVI peptide, successfully decoded another 2,999 spectra and determined their identity. We then tallied the number of times each peptide was found within those 3,000 spectra and plotted those numbers versus a theoretical gaussian distribution of the same number of spectra with an equal chance at finding each peptide (Figure 7.4). What we observed is that in our census of 3,000 spectra, the distribution's median is slightly skewed to the left. However, we do not believe that this skew is caused by the synthesis procedure, nor an increase in the difficulty of decoding a peptide terminated in any one capping group, we believe this is simply part of the randomness at the core of this method. Next, we further analyzed another 3,000 spectra, specifically looking for the 74 peptides that were not found in the first 3,000 analyzed spectra. In the end, the number and identity of peptides not found was 5; FSV, IVE, AFE, FPI, QFI, while the theoretical number should have been 4, further reinforcing the fact that neither this method nor the capping scheme instills any bias towards the synthesis or decoding of some peptides over others. After analyzing 6,000 spectra, it was also determined that the capping groups chosen to form the

mass-ladder gave good separation of the mass peaks within all spectra and left no ambiguity as to the identity of the peptides.



	0	1	2	3	4	5	6	7	8	9	10
Observed	74	175	221	182	151	92	57	35	7	4	2
Theoretical	54	147	212	225	184	94	53	20	10	1	0

Figure 7.4: Observed vs. theoretical number of times a peptide was found. We successfully decoded a total of 3,000 spectra and found that although the observed distribution's median is skewed to the left, we believe this is within the randomness of the synthesis and bead picking process and is not indicative of bias in the synthesis or decoding process.

Another important point that was determined from the synthesis and analysis of the 1,000-member random peptide library was that it should be possible, theoretically, to find close to 95% of the peptides synthesized in any random array of size N if 3N spectra are analyzed. This means that if higher coverage is desired, blindly screening 3N beads will yield 95% coverage of the library. Furthermore, 6N will yield 99.5% representation of the entire library, and 10N will assure that 100% of the library will be found. Thus 10-fold excess of beads was chosen as a minimum number of beads to be used for future array synthesis.

Automation of Spectral Analysis. After confirming the random array synthesis protocol and the capping scheme were working as intended, one major bottleneck was remaining in our goal to develop a quick and cheap method for the synthesis of large peptide libraries, and their use with SAMDI-MS, the spectral analysis. Although the spectral analysis process is straight forward, the manual decoding of a spectra takes between 5 to 10 seconds each. The cooperation of two people is also required, one looking at the mass-table and another at the spectrum. At these speeds, to simply decode 1,000 spectra would take approximately 1-2 hours. Thus, to accelerate the decoding process a computer program is needed that can automatically decode the spectra and match the peaks present within them to the mass-table used for manual decoding.

To accomplish this task, we began with the in-house "profiler" software that is typically used for enzymatic activity analysis. We used the software's ability to read .t2d, the standard spectrum file format for our ABSciex MALDI, and built the software around this function. The process by which the software analyzes data is outlined in detail in the methods section, but briefly; the new software opens each .t2d file and finds the mass and intensity data within the files. It then scans through the mass table provided by the user and generates a list of all masses of interest. The masses of interest are then matched to each spectrum, and if a peak is present at one of the masses of interest, as defined by a threshold in the intensity, it includes it in a list of all present peaks within that spectrum. Once the list of present peaks is generated, the program attempts to match the full-length peak plus any mass-ladder peaks to a possible list of peptides within the mass-table provided by the user. If more than one combination is possible, it determines which of the multiple combinations gives the greatest sum of the intensities. If no combination is present, or if peaks are missing, the software returns "failed" and moves on to the next spectrum. The implementation of this software shortened the time to decode a single 1536 plate's worth of spectra from

approximately 2 hours, to just under 5 minutes. We then ran the spectra that were previously decoded by hand using the software and validated the software's ability to decode the spectra automatically, with each spectrum that was successfully decoded by the software matching the identity of the hand decoding process.

Preparation and Analysis of the First Peptide Array for Cell-Adhesion Ligand Study. Having confirmed the viability of the synthesis, capping, and automation of decoding, we decided to tackle the synthesis of a large peptide library. Specifically, this was the 5,832 member peptide library with the sequence Ac-G($AA_3AA_2AA_1$)GRC, where AA_n are the random positions which represent all natural amino acids, excluding methionine and cysteine. Again, arginine was incorporated into the sequence to help with ionization. The purpose of synthesizing such a large library was two-fold; one, to demonstrate that a large library, which would normally take several months to synthesize, using standard SPPS techniques, can be done quickly and cheaply, and two, to investigate if a novel 3-mer peptide can be found that could modulate cell adhesion in HUVEC cells. Since the discovery of the RGD binding domain and its interaction with adhesion proteins, known as integrins, much research has been done on the topic to find other binding proteins.[158-163] We have previously reported on the use of peptides and SAMDI surfaces to study cell adhesion, as well as reporting the discovery of the truncated FEI ligand found within nephronectin.[24, 164-167] However, to date we have not looked at a large library of peptides to discover novel adhesion ligands.

Before performing any analysis, a decision was made that no decoding of the peptide library would be performed prior to looking for adhesion peptides. Instead the adhesion study would be run with 4-times the number of beads as the number of peptides synthesized, giving us

a theoretical coverage of over 98% of the library, and decoding any possible adhesion ligand hits. To search for novel adhesion ligands in HUVEC cells, the three-random-position library was synthesized, a process that required only 3 days, and beads were sorted into 384 well plates. The cleaved peptides solutions in these plates were then immobilized onto 1536 SAMDI plates. These SAMDI plates presented a 1% maleimide terminated monoloayers on the gold islands, as per previous cell adhesion experiments, and the plate was backfilled with hexadecylphosphonic acid between the islands. The phosphonic acid background served two purposes, to make the areas between gold islands hydrophobic, to assist in dispensing of peptide solutions onto the gold, and as a positive adhesion surface for the cells. Making a positive adhesion surface between the islands allowed for easy determination of areas on the plate that were and were not covered in gold. It also made the detection of non-adhesion and adhesion peptide containing locations simple, by looking for the lack of and presence of adhered cells, at densities different than the surrounding areas (Figure 7.5).

The HUVEC cells were treated with CellTracker™ Green CMFDA Dye, which allowed for the visualization of the cells via fluorescence when excited with a 488 nm laser. These cells were then transferred to the 1536 plate with immobilized peptides, and allowed to adhere and spread overnight. The cells were then fixed with a 4% formic acid solution and visualized on a Nikon® Eclipse Ti-E microscope. After immobilizing the first 1536 plate, and performing the adhesion study, we saw 24 gold spots with possible adhesion ligands. The solutions corresponding to these 24 spots were pulled out of their wells and their identity was determined. We found a single well with RGD in the random positions and another 19 unique sequences (Appendix Figure 7.1).

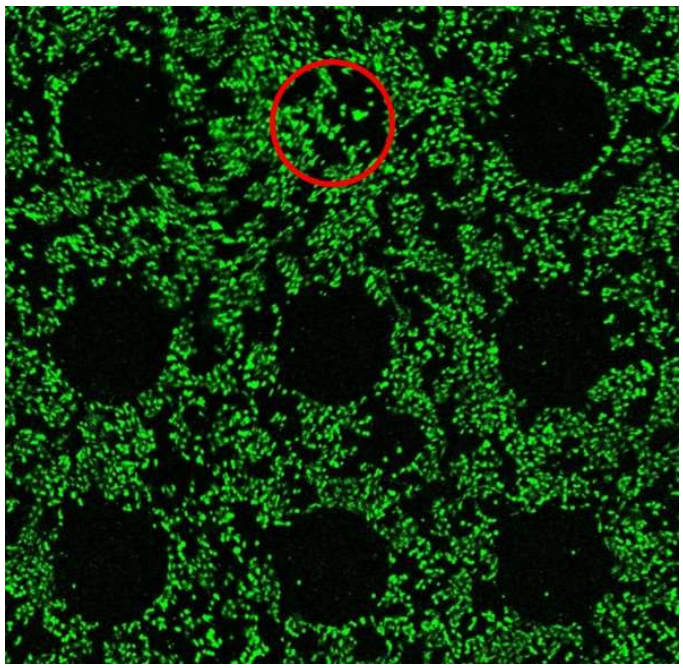


Figure 7.5: Image of HUVEC cells adhered to a 1536 SAMDI plate with immobilized peptides. In this image, the cells contain the CellTracker green dye, this allows us to visualize the cells on the confocal microscope by exciting the fixed cells with a 488 nm laser. The black circular regions are the gold islands that do not have a cell adhesion peptide immobilized, and the red circular region is a positive control, with cyclic-RGD immobilized. The areas in between the circles are the bare steel areas of the SAMDI plate that have had hexadecylphosphonic acid immobilized and serve as a positive cell adhesion surface.

However, when these 19 unique sequences were examined, a single commonality was discovered among all the sequences, each peptide had an asparagine in the first random position, giving each peptide a common sequence of Ac-GAA_3AA_2NGRC. This made us believe that the 3 residues "NGR" were an important factor to the adhesion-ligand like activity of these 19 sequences. When this idea was further explored, it was discovered that Corti and Curnis have previously shown a mechanism by which an NGR containing peptides would readily convert to iso-DGR, by deamidation of the asparagine into an iso-aspartate (Figure 7.5).[168] This iso-DGR peptide then in turn acts as an integrin binding peptide and is capable of binding to αvβ3 integrin in a similar manner to an RGD peptide. It was also reported that this deamidation process is fastest under basic conditions and a glycine next to the asparagine further accelerated the process. Thus,

it was unfortunate that the peptide cleavage conditions had a pH greater than 10, and that our peptide design had both a glycine next to the first random position, and an arginine 2 amino acids away, essentially guaranteeing any peptide with an asparagine in the first random position would generate an iso-DGR peptide and cause cells to adhere. The generation of the iso-DGR peptide was confirmed when the mass spectrum of the peptides was analyzed more carefully. Each peptide containing an asparagine in the first random position showed an isotope splitting pattern that gave a much smaller than expected +1 peak, indicative of the conversion of an asparagine to an iso-aspartate, increasing the mass by 1 (Appendix Figure 7.2). At this point the decision was made to

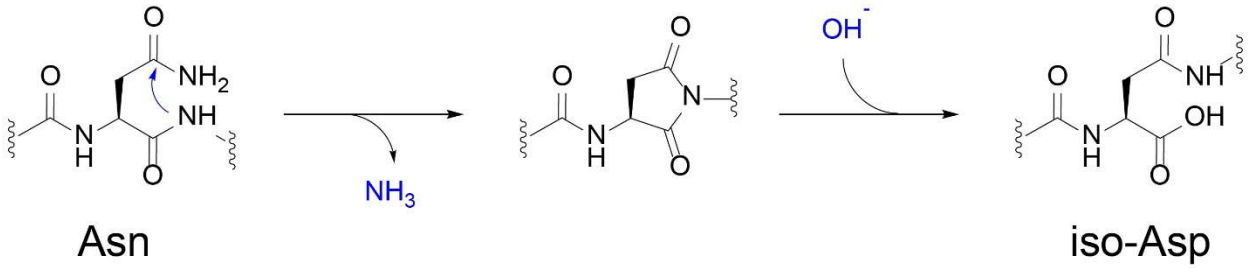


Figure 7.5: Deamidation converts asparagine to iso-aspartate under basic conditions. It had been previously shown that under basic conditions, an asparagine in the peptide can readily convert to an iso-aspartate, especially with a glycine on the c-terminus of the asparagine. It has also been shown that a sequence containing iso-DGR is an integrin binding peptide as good as RGD.[168]

forgo screening the rest of this first library and instead design a second library that would not cause 1/18th of the library to give positive adhesion results due to formation of an iso-DGR peptide. This decision to abandon this array and synthesize a new peptide array was only possible because the synthesis of a new array can be done quickly and at a low cost, compared to standard SPPS methods.

Preparation and Analysis of the Second Peptide Array for Cell-Adhesion Ligand Study. After the decision was made to abandon the first array, a new sequence had to be determined, one that would not produce an iso-DGR sequence to be present in a large portion of the array. The chosen

sequence was Ac-G($AA_3AA_2AA_1$)AGK$^{(me3)}$C, where AA_n are again the random positions which represent all natural amino acids, excluding methionine and cysteine. However this time the arginine was replaced with a trimethyl lysine, which provides better ionization as well as pushing all of the peptides to favor the proton adducts.[19, 94] With this library, only the peptides with NGR in the random positions would be converted to iso-DGR, and thus only a single peptide, other than RGD and FEI, out of the 5,832 would be a known adhesion ligand. The peptide library was synthesized as previously stated, beads sorted, and peptides cleaved. These peptides were then immobilized onto 1,536 SAMDI plates and HUVEC cells were seeded onto those plates. As with the previous array, the decision was made to perform adhesion studies prior to decoding. 16 – 1,536 plates worth of beads would be tested, giving a grand total of 24,576 beads, a number greater than four-fold the number of peptides synthesized, or a coverage of greater than 98%.

The adhesion studies were done in the same manner as with the previous array, and each 1,536 plate was imaged with the Nikon® Eclipse Ti-E microscope using a 1X objective and NIS-Elements software, generating an image of the entire plate (Appendix Figure 7.3). If a spot had adherent cells, an image of the specific spot was acquired using 10X magnification, to confirm the presence of attached and well-spread cells. After confirming the presence of adherent and spread cells, the address of the corresponding well was determined and solution from those wells were extracted and immobilized onto SAMDI plates with 20% maleimide functionalization for decoding purposes. In total, of the 24,576 peptides tested, 6 were confirmed to promote adhesion. After decoding, it was determined that 2 of these peptides were the RGD peptide, and four were iso-DGR (Figure 7.6). We did not seem to observe a spot that represented the sequence FEI, which we have shown to be an adhesion ligand previously. However, with the possibility of having close to 100 sequences not present (2% of 5,832), it is not unimaginable that the FEI peptide is one of those

100. Also, although we were not able to find a novel adhesion ligand from our 3-random position, 5,832 member peptide library, the results are in line with other methods that have screened large peptide libraries for adhesion peptides, such as yeast and phage displays.[76]

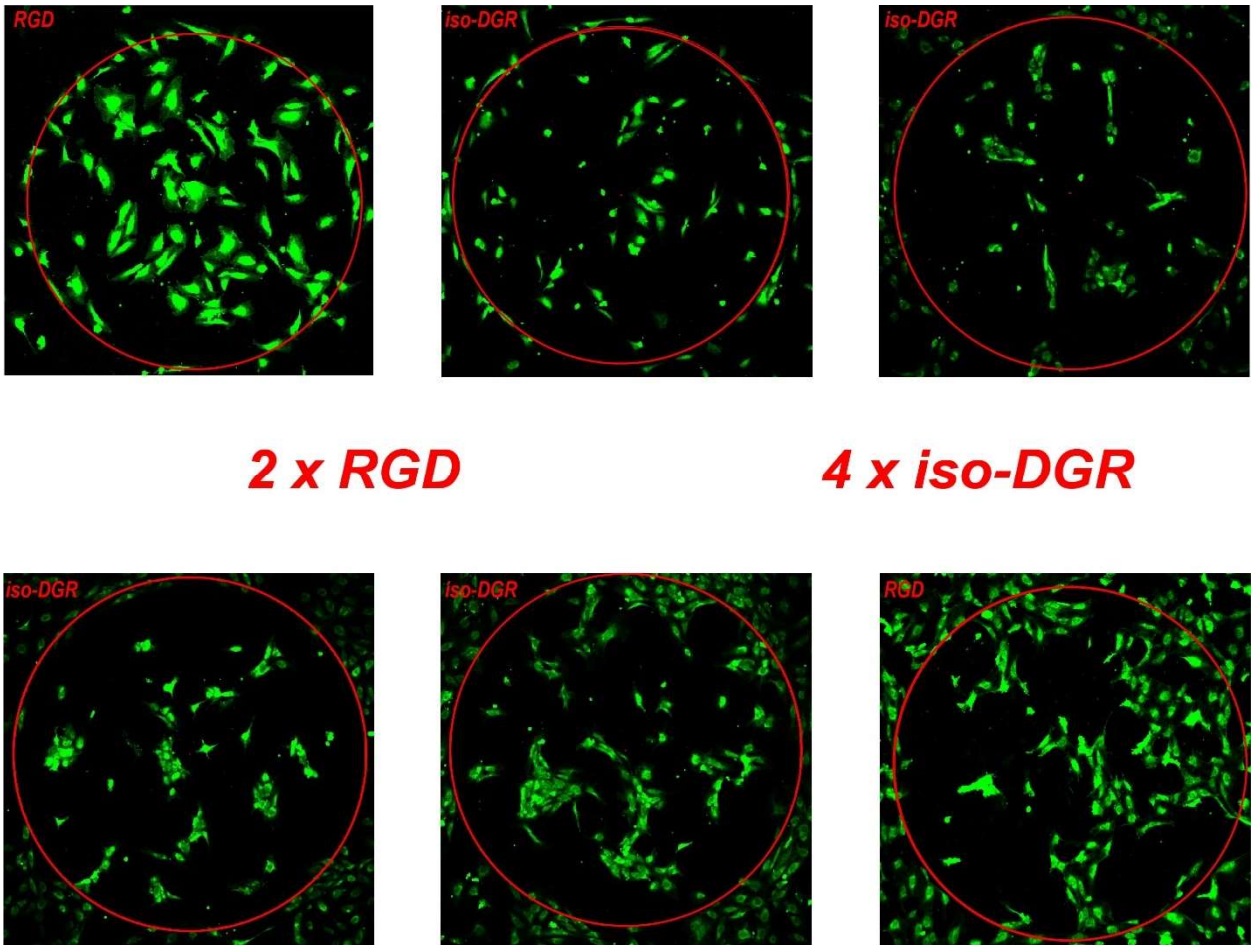


Figure 7.6: Adhesion peptides found and decoded after scanning 24,576 peptides of the 2nd 3-position random array. After scanning a total of 16 – 1,536 plates, only 6 spots were found to have peptides that induced adhesion of HUVEC cells. These were 2 spots with the RGD sequence, and 4 with iso-DGR.

Conclusion

While the use of peptide arrays with SAMDI-MS enables monitoring many enzymatic and non-enzymatic reactions, the cost associated with synthesizing these libraries, both in terms of money and time, makes it impossible to synthesize and examine large or one-off arrays of peptides.

[9, 13-17] However, here we have introduced a method that combines the synthesis of random arrays with SAMDI-MS to synthesize, decode, and utilize large peptide libraries quickly and cost efficiently. With this method we were able to circumvent the previous approaches' requirements for complicated or orthogonal decoding, and synthesis processes.[76, 148, 149, 152, 153] We also demonstrated the ability to synthesize a large peptide library for cell adhesion ligand studies, and through the time and cost savings, provided by this method, showed the ability to make result driven decisions. Taking the knowledge gained we modified the sequence of the first peptide library, and synthesized a second library of a previously unobtainable size quicker and cheaper than standard SPPS methods would have taken to synthesize one of these libraries. And although a novel adhesion ligand peptide was not discovered, it is our belief that had one been present within the array, our approach would be able to find and decode it. Through the use of SAMDI-MS and the random peptide synthesis method outline in this chapter, it is our belief that libraries of thousands to tens of thousands of peptides can be easily synthesized and utilized in days to weeks, instead of months to years.

Methods

Materials. Fmoc amino acids and peptide synthesis reagents were purchased from AAPPTec, AnaSpec, and Chem-Impex International Inc. The Tentagel MB-OH resin was ordered from Rapp Polymere. All other solvents and reagents for Fmoc solid-phase peptide synthesis were purchased from Thermo Fisher and Sigma Aldrich. HUVEC cells and cell culture media were purchased from ATCC.

Synthesis of Random Peptide Array on Beads. The random peptide library were synthesized in the following way; for the sequence Ac-G($AA_1AA_2AA_3$)AGK$^{(me3)}$C, where AA_n are the random

positions, was synthesized by first chemically functionalizing the Tentagel MB-OH resin with Fmoc-cysteine via an ester bond utilizing the Steglich esterification, though DIC was used instead of DCC.[156] The rest of the amino acids, up to the random positions were added using standard Fmoc chemistry, as previously reported.[23] The random position amino acids were added by first splitting the beads into an appropriate number of reaction vessels, corresponding to number of different amino acids per position. The growing peptide chains were then deprotected using standard chemistry and the new amino acid were added in an 85:15 ratio of Fmoc-amino acid to capping group with standard fmoc synthesis reagents. Following the coupling of the random position, the beads were washed, pooled into a single vessel, mixed, and split again into individual reaction vessels, repeating the deprotection, coupling, pooling steps. Upon completion of synthesis, the beads were washed with dimethylformamide (DMF), dichloromethane (DCM), ethanol (EtOH) and allowed to dry under vacuum. After drying, the beads were transferred to a 20 mL scintillation vial and the sidechains protecting groups of the amino acids were cleaved using a standard trifluoroacetic acid (TFA) protocol. Upon completion of the cleavage protocol, the TFA was drained. Beads were then washed with water, methanol, DCM, ethanol, and dried in vacuo.

Bead Sorting and Cleavage of Peptide Chains. The dry beads, with peptides that have had their sidechains removed, were first soaked in acetonitrile (ACN) and sonicated for 30 minutes, this helped breakup bead clusters. Once the beads have been adequately separated, water is added to bring the final concentration of ACN to below 10%. At this time, 5 µl of a cleavage solution of 7% triethylamine (TEA) in water is dispensed into each well of a 384 well plate (Greiner #781201) using the Multidrop Combi reagent dispenser (Thermo Scientific). The slurry of beads and the well plate are then loaded into the COPAS FP-500 instrument (Union Biometrica) and 1 bead is sorted into each well of a 384 well plate.[148, 169] The plates are then sealed and centrifuged down at 4,000

rpm for 1 minute, to ensure each bead is submerged in the cleavage solution. The beads are allowed to react with the cleavage solution over night, to ensure maximum cleavage of the ester bond.

Preparation of SAMDI Plates. Specially made blank MALDI target plates (Ace Metal Crafts) were cleaned in hexanes, ethanol, acetone, ethanol. An electron beam evaporator (Thermionics VE-100) was then used to deposit 5 nm of Ti followed by 30 nm of Au through a mask with 1536 holes in a standard microtiter format. The plates are then submerged in a solution of 0.5 mM alkane disulfides in ethanol in appropriate concentrations of an symmetric disulfide terminated in tri(ethylene glycol) and a asymmetric disulfide terminated with maleimide and tri(ethylene glycol). 20% maleimide was used for peptide sequence determination and 1% for cell adhesion studies. Following this the plates were washed with ethanol, water, ethanol, dried under nitrogen and placed in a second solution of 10 mM hexadecylphosphonic acid (Sigma 736244) in ethanol for 5 minutes to back fill the bare steel areas with hydrophobic molecules, to promote liquid retention on gold spots, and produce a positive adhesion surface.

Peptide Immobilization onto SAMDI Plates. Well plates with cleaved peptides were first buffered with 25 μL of 100 mM Tris pH 8.0 buffer. The buffered solutions were then treated with 3 μL of TCEP bead slurry (Thermo Scientific #77712). After 30 minutes at room temperature the plates were centrifuged to pellet down the TCEP beads. 0.5 μL of supernatant from each well was then transferred onto the target 1536 SAMDI plates using a 384-tip liquid handling robot (TECAN Freedom EVO). For the cell adhesion study, any spots that were missed by the TECAN or did not have liquid transferred to the surface were backfilled with cyclic RGD peptide, as a positive adhesion control. The solutions of peptides were allowed to react for 30 minutes, then washed with

ethanol, water, ethanol, and dried under nitrogen. Immobilized peptide plates were stored under vacuum until needed.

SAMDI-MS. SAMDI plates with immobilized peptides were covered in 2',4',6'-trihydroxyacetophenone matrix solution (THAP, 12 mg/mL in acetone) and loaded into an ABSciex ToF/ToF 5800 instrument. Spectra of each spot were obtained via reflector positive mode in automated mode.

Generating the Decoding Mass Table. To generate the mass table used for decoding the array of random peptides synthesize, we first compiled a list of all the possible peptides. To do this we made every combination of amino acids possible, which gave us x^n unique sequences, where x is the number of amino acids used per position and n is the number of positions. Following this the masses of truncated sequences were determined by calculating the total mass of the amino acids added through the step and adding the Δm values corresponding to each amino acid in table 7.1, as well as +851 for the mass of the monolayer disulfide. Finally the mass of the full-length peptide was calculated and added to the mass of the monolayer disulfide. The sequence, full-length mass, and truncated masses were then arrayed into an Microsoft Excel spread sheet.

Sequence Determination (Manual). Spectra automatically acquired in reflector positive mode on the ABSciex 5800 MALDI-TOF/TOF were opened using the standard TOF/TOF Series Explorer Software (ABSciex). The appropriate decoding Excel sheet was opened and sorted by the mass of the full-length peptide. A peak corresponding to the full-length peptide was then found (usually highest mass peak) and matched to the mass table. This was followed by matching any truncation peaks. Once all the peaks were matched, the sequence of the peptide was noted and the next spectra was analyzed.

Sequence Determination (Automated). To decode the peptide spectra automatically, the .t2d files associated with the mass spectrum of each peptide was exported to a folder using the standard TOF/TOF Series Explorer Software. Following this the "Peptide Processor" software that was specifically written for this application was loaded. For this program to function three inputs had to be specified; input path for the spectra files, an output path for processed spectra and logs, and a path for the mass table. The program then takes these inputs and determines which peaks are present within the spectra and matches those to the possible combinations within the mass table. There were also four optional inputs within the program to further assist in decoding. "Adjustment mass", this is a real number value to adjust masses that are shifted (due to calibration of the instrument). "Group intensity threshold", where if the sum of intensities of peaks do not reach a certain threshold they would be disregarded. "Background intensity threshold", a value which removes any intensity value below a certain threshold, to assist in speeding up the analysis. And "extra logging", which produces a .txt output, with detailed analysis, for each spectrum. This was predominantly used for debugging of the program, along with determining errors in input. Once all the necessary and optional inputs have been filled in, the program runs through every spectra, remove the background signal and generate a group of peaks. The software will then attempt to decode the sequence of the peptides by finding the combination of peaks that match an entry in the mass table and also give the highest sum of the peak intensities. If it finds such a grouping of peaks it will return "success" and the sequence of the random positions associated with the entry in the mass table.

Mamalian Cell Culture. Human umbilical vein endothelial cells (HUVEC) were obtained from ATCC® (CRL-1730™) and cultured using vascular cell basal medium (ATCC® PCS-100-030™) supplemented with microvascular endothelial cell growth kit-BBE (ATCC® PCS-110-040™).

Prior to cell adhesion studies, the HUVEC cells were cultured with CellTracker™ Green CMFDA Dye (Thermofisher C2925) in fetal bovine serum (FBS) free media for 15 minutes. Following this, the media was aspirated and TrypLE Express (Gibco #12604021) was added to trypsinize the cells. After 5 minutes the cells were transferred to a centrifuge tube, washed with FBS containing media and spun down. The washing step was performed once more, with FBS free media. Following this, the cells were resuspended in 10 mL of FBS free media and counted.

Creating a Plate Holder for Cell Adhesion Study. A mold to hold a SAMDI plate for cell adhesion studies was made by first taping 5 SAMDI plates together. These plates were then placed in a container that is 4 x 6 inches and enough PDMS (SYLGARD® 184 silicone elastomer, DOW Chemical) was added to submerge the plates under 3 mm of mixture. The mold was then placed in a vacuum chamber to remove bubbles and placed in a 40 °C oven to cure overnight. After curing, the mold with the plates was removed from the container and placed in an 80 °C oven overnight to further cure. After this second cure, the plates were removed, and the PDMS mold was ready to use with cell adhesion studies.

Cell Adhesion Study. 1536 spot SAMDI plates with a 1% maleimide density which have had peptide already immobilized were placed in the PDMS mold. Enough cell suspension in FBS free media was pulled out to give 1-2 million cells for each SAMDI plate. The suspension was then further diluted in FBS free media to give a volume of 15 mL. The cells were then transferred to the SAMDI plate within the mold, ensuring coverage of the entire plate under media, and agitated to ensure good cell coverage. The mold was then placed in a separate chamber within the cell culture oven and allowed to sit for 30 minutes. Following the 30-minute rest, the molds were further agitated to further promote good cell coverage and allowed to rest for another 30 minutes.

After the second rest, the FBS free media was gently aspirated and 15 mL of media with FBS was added to cover the plates. The plates and molds were then placed back in the oven and allowed to sit overnight to allow the cells to spread.

Cell Imaging and Adhesion Determination. SAMDI plates that have had cells cultured overnight were removed from the oven and the media aspirated off. The plates were then gently rinsed with Dulbecco's Phosphate Buffered Saline (DPBS Gibco #14190-144) to remove any remaining media and unbound cells. After aspirating the DPBS wash, a 4% formaldehyde solution in DPBS was applied and allowed to sit on the surface for several minutes, to fix the cells, and then aspirated. The plate was then fully dried prior to imaging. The imaging was done on a Nikon® Eclipse Ti-E microscope with a 1X objective using NIS-Elements software. Using the software a large image scan of the entire plate was obtained using a preset 10% overlap for stitching (Appendix Figure 7.3). After imaging, dark circles were seen amongst a sea of green cells, bound to the positive adhesion background. These are the gold spots with immobilized peptides that do not promote adhesion. Regions that break the array of black circles were spots that possibly contain cell adhesion peptides. Locations of these spots were recorded, a 10X image taken, and the solution of these wells were decoded to determine the sequence.

Chapter 8: Summary and Future Directions

At the beginning of my tenure in the Mrksich Lab, it was not uncommon for most lab members to run only tens to several hundred SAMDI experiments in a single day. In those days most people still utilized glass slides with 50 gold islands and using a 384 gold island steel plate was rare. At the time, work in the lab mainly concentrated on a wide range of enzymatic reactions which induce a mass change, including acetyl transferases, deacetylases, kinases, phosphatases, deaminases, hydrolases, proteases, and glycosyl transferases on both surface immobilized and in-solution peptides. [8, 17, 18, 20, 61-65, 78] We have also explored the activity of proteins that do not induce a mass change, such as adhesion domains, by fusing these proteins to enzymes that cause a mass change.[18] However, as my time in the group passed, my use of glass slides with 50 gold islands slowly faded away and 384-spot plates became the norm, and eventually those started being superseded by 1536 spot plates, and the reactions being monitored by SAMDI, although still mostly biochemical, because more complex.

However, in the first section of this thesis I demonstrated examples for the use of peptides with SAMDI-MS that does not involve exploring biochemical assays, and unlike most previous experiments, uses more than a single peptide immobilized to the surface. Here I presented a novel method to store information with SAMDI-MS and peptides. In Chapter 2 I demonstrated a method to store binary information using 32 peptides for every SAMDI spot, giving us a density of 4 bytes/spot. Along with being one of the first demonstrations of using 1536 spot plates for SAMDI, this method provided a first use of mass spectrometry to store digital information in molecular form, possibly providing a new method to store digital information offline. I further expended on

this idea in Chapter 3 by exploring the use of higher-than-binary base system using SAMDI and peptides, a hard to achieve feat using digital means of storage. Here I showed the benefit of using concentrations to store information. By leveraging the ability to store peptides in varying concentrations, relative to an internal standard, I demonstrated a method that could store information in base 4, 8, and 16, increasing the digital information stored on a single spot by 2, 8/3, and 4-fold over binary, respectively. In Chapter 4 I further explored the ability of SAMDI to store an increased amount of information. However, in this example I moved away from storing multiple peptides per spot, or using concentrations, and instead concentrated on the ability of modern MALDI-TOF instruments to read spectra from very small areas and at much higher rates. I showed how it could be possible to utilize the new top of the line MALDI, the Bruker RapifleX, along with imaging SAMDI and inkjet printers to increase the storage capacity of a single SAMDI plate by 2,000-fold.

In the second half of this thesis I first outlined two new applications of SAMDI with peptide arrays. In Chapter 5 I demonstrated the ability of SAMDI to monitor proteolytic activity of detergents containing proteases. Previously, most of the studies of proteolytic activity within detergents were done in real life conditions, whereby a protein stain was actually applied to a piece of cloth and the cloth was washed with the detergent in a washing machine, a process that took a long time. However, by combining immobilized peptides and SAMDI I demonstrated that we can run multiple detergents, at normal detergent concentrations, with a 324-member peptide library, and not only determine the overall cleavage activity on each peptide, but also get cleavage site preference information. In Chapter 6 I returned to one of the staples of SAMDI and peptide arrays, the deacetylation reactions on immobilized peptides. However, here I explored the importance of looking at the activity of enzymes not through a final time point reaction, but through a kinetic

study, whereby the activity of enzymes on peptides is ranked through the slopes of the initial activity curves. By looking at these slopes, it was possible to further deduce the best substrate for 3 KDACs and show how even a related isoform of enzymes can be easily distinguished, by looking at kinetics. Finally, in Chapter 7 I explored and discussed a method to synthesize large peptide arrays quickly and cost efficiently. By utilizing split-pool methods, along with a novel mass-ladder generating capping scheme, we can use standard Fmoc chemistry along with readily available carboxylic acids to synthesize muti-thousand peptide arrays in the same time that it would take to synthesize a handful of peptides, and at a cost per peptide that is drastically cheaper. I utilized this random peptide array synthesis method to first make a 3-random position library of 5,832 peptides with the sequence Ac-G($AA_3AA_2AA_1$)GRC, where AA_n are the random positions which represent all natural amino acids, excluding methionine and cysteine. This array only took 3 days to synthesize, a process that would have taken months to accomplish in any other method other than SPOT arrays. I then used this array in an attempt to discover novel integrin binding peptide ligands using HUVEC cells. However, after the first 1536 plate was analyzed, it was determined that the synthesized sequence was susceptible to deamidation of asparagines in the AA_3 position. However, because the synthesis of the initial 5,832-member array was fast and cheap, I was able to use the information acquired from the first 1536 plate and redesign the peptide sequence. I proposed and synthesized a second array with the sequence Ac-G($AA_1AA_2AA_3$)AGK$^{(me3)}$C, again where AA_n are the random positions which represent all natural amino acids, excluding methionine and cysteine, and this time including a trimethyl lysine to improve ionization and ease of decoding. This array was again synthesized in approximately 3 days, and the cell adhesion studies were started again. After running cell adhesion studies on 24,576 spots however, no new adhesion ligand was discovered, though previously know adhesion ligands RGD and iso-DGR were found,

reinforcing the belief that with the correct peptide library sequence, the use of random peptide arrays and SAMDI would be able to discover new integrin adhesion ligands.

Finally, I would like to share my thoughts on the future direction of SAMDI with peptides and peptide arrays. As I've mentioned previously, at the start of my tenure in the group, many people, myself included, performed SAMDI experiments in small batches, utilizing glass slides with 50 spots of gold, many times cutting those glass slides into smaller slivers, using only 5 or 10 of those spots for a single day's worth of experiments. However, as time passed, many people transitioned into experiments more heavily reliant on peptide arrays, instead of a small group of peptides. As these experiments became the norm, the use of robotics, such as TECAN's and Combi's became more and more pervasive. And with the increased throughput provided by the use of liquid handling robot, the number of experiments that a single graduate student can run in a day increased drastically. The dozens of experiments a student would run in a day gave way to hundreds, to the point that a single 384 plate could not contain all the experiments one wanted to run, and thus giving us the 1536 spot plates. With the introduction of those plates we had to develop our liquid handling techniques further, pushing the limits of available liquid handling robot and high-throughput techniques to their limits. Now days a 1536 plate is as standard as walking to the breakroom and grabbing a coffee, and a 384 plate seems like a slow day.

However, it was not only our ability to run experiments that increased over time. As we ran an ever-increasing number of experiments, we had to conceive ways of analyzing the data produced. This is where a lot of the automation in data analysis arose. I believe it would not be farfetched to say that some of us became wizards with Excel, setting up equations where all we had to do was copy our raw data, and our equations would do the rest. These Excel sheets, along

with our in-house programs written specifically to assist in the analysis of massive amounts of data further pushed the need to acquire more and more data. And this is where we find ourselves now; we are always looking for ways to run greater number or experiments. I believe that with incorporation of inkjet printers, the density of reading provided by methods such as imaging SAMDI, and with a wider implementation of random peptide arrays, we will soon see experiments numbering 100,000's in a single day, run by a single graduate student, all analyzed with a single press of a button, or a click of a mouse.

Using SAMDI-MS and random peptide arrays outlined in this thesis, it will become possible to formulate a hypothesis, synthesize peptide arrays, and run preliminary experiments all in a matter of weeks. And because the random peptide synthesis method does not incur a large cost, both in time and materials, many "what-if" experiments can be explored. This can then lead to new and exciting application of peptide arrays, which would not have been considered previously, due to the large upfront costs of synthesis of the peptide arrays for these applications. Also, with the decrease in cost and time needed to synthesize peptide arrays, it would enable rapid modifications of peptide arrays design to reflect experimental results. I have also explored, though not outlined in this thesis, the possibility of removing the mass-ladder peptides. Through the use of azide containing capping groups, I believe it will be possible, in the very near future, to incorporate a pull-out, or purification step, whereby dibenzylcyclooctyne (DBCO) immobilized beads can be introduced into the peptide solution post decoding and remove any mass-ladder peptides from solution by undergoing click-chemistry with the azide containing peptides. This would then enable the use of the method outlined here to synthesize large random peptide arrays, use the mass-ladder to identify the peptides, and then by pull out the mass-ladder peptides from

solution, leaving the full-length peptides. These peptides can then be used without the worry of unintended interactions of the truncated peptide sequences within the assay.

As for the information storage project, I believe future advancements in monolayer design, liquid handlers, and MALDI instruments can help make the molecular information storage with SAMDI a viable method for storing data long-term. Though I do not envision it being a replacement for current digital storage techniques such as HDD's, flash, and magnetic tape storage, I do believe that this technology could find a niche application where the storage of information is desired on a longer timescale than the currently available digital media. Due to their transient nature, the digital data storage techniques suffer from data loss in timescales of decades. However, I believe that with proper preparation and storage of SAMDI based information storage media, the desired information could be stored for a longer period of time. Also, through the use of thinner substrates, such as micron-thick metal films, and the combination of sub-micron scale inkjet printers, the densities of this medium could compete with some of the available technology on the market today, all the while offering benefits in lifetime as well as providing resistance to electromagnetic damage.

References

1. Poirier, G.E. and Pylant, E.D., *The Self-Assembly Mechanism of Alkanethiols on Au(111).* Science, 1996. **272**(5265): p. 1145.
2. Whitesides, G.M., Kriebel, J.K., et al., *Molecular engineering of Surfaces Using Self-Assembled Monolayers.* Science Progress, 2005. **88**(1): p. 17-48.
3. Mrksich, M. and Whitesides, G.M., *Using Self-Assembled Monolayers to Understand the Interactions of Man-made Surfaces with Proteins and Cells.* Annual Review of Biophysics and Biomolecular Structure, 1996. **25**(1): p. 55-78.
4. Anderson, L.L., Berns, E.J., et al., *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): p. 8604-8609.
5. Yousaf, M.N., Chan, E.W.L., et al., *The Kinetic Order of an Interfacial Diels–Alder Reaction Depends on the Environment of the Immobilized Dienophile.* Angewandte Chemie International Edition, 2000. **39**(11): p. 1943-1946.
6. Kim, J. and Mrksich, M., *Profiling the selectivity of DNA ligases in an array format with mass spectrometry.* Nucleic acids research, 2010. **38**(1): p. e2-e2.
7. Diagne, A.B., Li, S., et al., *SAMDI Mass Spectrometry-Enabled High-Throughput Optimization of a Traceless Petasis Reaction.* ACS Combinatorial Science, 2015. **17**(11): p. 658-662.
8. Ban, L., Pettit, N., et al., *Discovery of glycosyltransferases using carbohydrate arrays and mass spectrometry.* Nature chemical biology, 2012. **8**(9): p. 769-773.
9. Houseman, B.T., Gawalt, E.S., et al., *Maleimide-Functionalized Self-Assembled Monolayers for the Preparation of Peptide and Carbohydrate Biochips.* Langmuir, 2003. **19**(5): p. 1522.
10. Kim, Y., Ho, S.O., et al., *Efficient Site-Specific Labeling of Proteins via Cysteines.* Bioconjugate Chemistry, 2008. **19**(3): p. 786-791.
11. Henkel, M., Röckendorf, N., et al., *Selective and Efficient Cysteine Conjugation by Maleimides in the Presence of Phosphine Reductants.* Bioconjugate Chemistry, 2016. **27**(10): p. 2260-2265.
12. Trevor, J.L., Lykke, K.R., et al., *Two-Laser Mass Spectrometry of Thiolate, Disulfide, and Sulfide Self-Assembled Monolayers.* Langmuir, 1998. **14**(7): p. 1664-1673.
13. Su, J. and Mrksich, M., *Using Mass Spectrometry to Characterize Self-Assembled Monolayers Presenting Peptides, Proteins, and Carbohydrates.* Angew. Chem., Int. Ed., 2002. **41**(24): p. 4715.
14. Min, D.-H. and Mrksich, M., *Peptide arrays: towards routine implementation.* Current Opinion in Chemical Biology, 2004. **8**(5): p. 554-558.
15. Min, D.-H., Su, J., et al., *Profiling Kinase Activities by Using a Peptide Chip and Mass Spectrometry.* Angewandte Chemie International Edition, 2004. **43**(44): p. 5973-5977.
16. Gurard-Levin, Z.A., Kim, J., et al., *Combining Mass Spectrometry and Peptide Arrays to Profile the Specificities of Histone Deacetylases.* ChemBioChem, 2009. **10**(13): p. 2159-2161.
17. Dai, R., Ten, A.S., et al., *Profiling Protease Activity in Laundry Detergents with Peptide Arrays and SAMDI Mass Spectrometry.* Industrial & Engineering Chemistry Research, 2019. **58**(25): p. 10692-10697.

18. O'Kane, P.T. and Mrksich, M., *An Assay Based on SAMDI Mass Spectrometry for Profiling Protein Interaction Domains.* Journal of the American Chemical Society, 2017. **139**(30): p. 10320-10327.
19. O'Kane, P.T., Dudley, Q.M., et al., *High-throughput mapping of CoA metabolites by SAMDI-MS to optimize the cell-free biosynthesis of HMG-CoA.* Science Advances, 2019. **5**(6): p. eaaw9180.
20. Szymczak, L.C., Huang, C.-F., et al., *Chapter Fourteen - Combining SAMDI Mass Spectrometry and Peptide Arrays to Profile Phosphatase Activities*, in *Methods in Enzymology*, K.N. Allen, Editor. 2018, Academic Press. p. 389-403.
21. Szymczak, L.C. and Mrksich, M., *Using Peptide Arrays To Discover the Sequence-Specific Acetylation of the Histidine-Tyrosine Dyad.* Biochemistry, 2019. **58**(13): p. 1810-1817.
22. Marin, V.L., Bayburt, T.H., et al., *Functional Assays of Membrane-Bound Proteins with SAMDI-TOF Mass Spectrometry.* Angewandte Chemie International Edition, 2007. **46**(46): p. 8796-8798.
23. Kilian, K.A. and Mrksich, M., *Directing Stem Cell Fate by Controlling the Affinity and Density of Ligand-Receptor Interactions at the Biomaterials Interface.* Angew. Chem., Int. Ed., 2012. **51**(20): p. 4891.
24. Sánchez-Cortés, J. and Mrksich, M., *Using Self-Assembled Monolayers To Understand α8β1-Mediated Cell Adhesion to RGD and FEI Motifs in Nephronectin.* ACS Chemical Biology, 2011. **6**(10): p. 1078-1086.
25. Lloyd, S., *Ultimate physical limits to computation.* Nature, 2000. **406**: p. 1047.
26. Shulaker, M.M., Hills, G., et al., *Three-dimensional integration of nanotechnologies for computing and data storage on a single chip.* Nature, 2017. **547**: p. 74.
27. Salahuddin, S., Ni, K., et al., *The era of hyper-scaling in electronics.* Nature Electronics, 2018. **1**(8): p. 442-450.
28. Fettweis, G. and Zimmermann, E. *ICT Energy Consumption - Trends and Challenges*. in *The 11th International Symposium on Wireless Personal Multimedia Communications (WPMC 2008)*. 2008. Saariselkä, Finland.
29. Baliga, J., Ayre, R., et al., *Green Cloud Computing: Balancing Energy in Processing, Storage, and Transport.* Proceedings of the IEEE, 2011. **99**: p. 149-167.
30. Brandner, R., Pordesch, U., et al., *Long-Term Archive Service Requirements*, in *Request for Comments*, R. Editor, Editor. 2007, The IETF Trust: Internet Requests for Comments.
31. Trevino, S.G., Zhang, N., et al., *Evolution of functional nucleic acids in the presence of nonheritable backbone heterogeneity.* Proceedings of the National Academy of Sciences, 2011. **108**(33): p. 13492.
32. Joyce, G.F., *RNA evolution and the origins of life.* Nature, 1989. **338**(6212): p. 217-224.
33. Piccirilli, J.A., *RNA seeks its maker.* Nature, 1995. **376**(6541): p. 548-549.
34. Joyce, G.F., Schwartz, A.W., et al., *The case for an ancestral genetic system involving simple analogues of the nucleotides.* Proceedings of the National Academy of Sciences, 1987. **84**(13): p. 4398.
35. Engelhart, A.E. and Hud, N.V., *Primitive genetic polymers.* Cold Spring Harb Perspect Biol, 2010. **2**(12): p. a002196.
36. Cartwright, J.H.E., Giannerini, S., et al., *DNA as information: at the crossroads between biology, mathematics, physics and chemistry.* Philosophical transactions. Series A, Mathematical, physical, and engineering sciences, 2016. **374**(2063): p. 20150071.

37. Church, G.M., Gao, Y., et al., *Next-generation digital information storage in DNA.* Science, 2012. **337**(6102): p. 1628.
38. Goldman, N., Bertone, P., et al., *Towards practical, high-capacity, low-maintenance information storage in synthesized DNA.* Nature, 2013. **494**(7435): p. 77-80.
39. Cox, J.P.L., *Long-term data storage in DNA.* Trends in Biotechnology, 2001. **19**(7): p. 247-250.
40. Organick, L., Ang, S.D., et al., *Random access in large-scale DNA data storage.* Nature Biotechnology, 2018. **36**(3): p. 242-248.
41. Tabatabaei Yazdi, S.M.H., Yuan, Y., et al., *A Rewritable, Random-Access DNA-Based Storage System.* Scientific Reports, 2015. **5**(1): p. 14138.
42. Bornholt, J., Lopez, R., et al., *A DNA-Based Archival Storage System*. Proceedings of the Twenty-First International Conference on Architectural Support for Programming Languages and Operating Systems. 2016, Atlanta, Georgia, USA: Association for Computing Machinery. 637–649.
43. Organick, L., Ang, S.D., et al., *Random access in large-scale DNA data storage.* Nat Biotechnol, 2018. **36**(3): p. 242-248.
44. Willerslev, E., Hansen, A.J., et al., *Long-term persistence of bacterial DNA.* Current Biology, 2004. **14**(1): p. R9-R10.
45. Erlich, Y. and Zielinski, D., *DNA Fountain enables a robust and efficient storage architecture.* Science, 2017. **355**(6328): p. 950-954.
46. Colquhoun, H. and Lutz, J.F., *Information-containing macromolecules.* Nat Chem, 2014. **6**(6): p. 455-6.
47. Green, J.E., Choi, J.W., et al., *A 160-kilobit molecular electronic memory patterned at 10(11) bits per square centimetre.* Nature, 2007. **445**(7126): p. 414-7.
48. Goodwin, C.A.P., Ortu, F., et al., *Molecular magnetic hysteresis at 60 kelvin in dysprosocenium.* Nature, 2017. **548**(7668): p. 439-442.
49. Shoshani, S., Ratner, T., et al., *Biologically Relevant Molecular Finite Automata.* Israel Journal of Chemistry, 2011. **51**(1): p. 67-86.
50. Ratner, T., Reany, O., et al., *Encoding and processing of alphanumeric information by chemical mixtures.* Chemphyschem, 2009. **10**(18): p. 3303-9.
51. Sarkar, T., Selvakumar, K., et al., *Message in a molecule.* Nat Commun, 2016. **7**: p. 11374.
52. Saini, A., Christenson, C.W., et al., *Threshold response using modulated continuous wave illumination for multilayer 3D optical data storage.* Journal of Applied Physics, 2017. **121**(4): p. 043101.
53. Christenson, C.W., Saini, A., et al., *Nonlinear fluorescence modulation of an organic dye for optical data storage.* Journal of the Optical Society of America B, 2014. **31**(3): p. 637-641.
54. Ryan, C., Christenson, C.W., et al., *Roll-to-Roll Fabrication of Multilayer Films for High Capacity Optical Data Storage.* Advanced Materials, 2012. **24**(38): p. 5222-5226.
55. Merrifield, R.B., *Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide.* Journal of the American Chemical Society, 1963. **85**(14): p. 2149-2154.
56. Szymczak, L.C., Kuo, H.Y., et al., *Peptide Arrays: Development and Application.* Anal. Chem., 2018. **90**(1): p. 266.
57. Kates, S.A. and Albericio, F., *Solid-phase synthesis : a practical guide*. 2004, New York.
58. Amblard, M., Fehrentz, J.-A., et al., *Methods and protocols of modern solid phase peptide synthesis.* Molecular Biotechnology, 2006. **33**(3): p. 239-254.

59. Zhu, H., Klemic, J.F., et al., *Analysis of yeast protein kinases using protein chips.* Nature Genetics, 2000. **26**(3): p. 283-289.
60. MacBeath, G. and Schreiber, S.L., *Printing Proteins as Microarrays for High-Throughput Function Determination.* Science, 2000. **289**(5485): p. 1760.
61. Gurard-Levin, Z.A., Kilian, K.A., et al., *Peptide Arrays Identify Isoform-Selective Substrates for Profiling Endogenous Lysine Deacetylase Activity.* ACS Chemical Biology, 2010. **5**(9): p. 863-873.
62. Liao, X., Su, J., et al., *An Adaptor Domain-Mediated Autocatalytic Interfacial Kinase Reaction.* Chemistry – A European Journal, 2009. **15**(45): p. 12303-12309.
63. Kightlinger, W., Lin, L., et al., *Design of Glycosylation Sites by Rapid Synthesis and Analysis of Glycosyltransferases.* Nat. Chem. Biol., 2018. **14**: p. 627.
64. Grant, J., Modica, J.A., et al., *An Immobilized Enzyme Reactor for Spatiotemporal Control over Reaction Products.* Small, 2018. **14**(31): p. 1800923.
65. Grant, J., *Microfluidics on Self-Assembled Monolayers for Analyzing Biological and Chemical Reactions*, in *Chemistry*. 2018, Northwestern University: Evanston, Illinois. p. 168.
66. Rasmussen, M.S., Birgisdottir, Å.B., et al., *Use of Peptide Arrays for Identification and Characterization of LIR Motifs*, in *Autophagy: Methods and Protocols*, N. Ktistakis and O. Florey, Editors. 2019, Springer New York: New York, NY. p. 149-161.
67. Katz, C., Levy-Beladev, L., et al., *Studying protein–protein interactions using peptide arrays.* Chemical Society Reviews, 2011. **40**(5): p. 2131-2145.
68. Ayoglu, B., Schwenk, J.M., et al., *Antigen arrays for profiling autoantibody repertoires.* Bioanalysis, 2016. **8**(10): p. 1105-1126.
69. Cretich, M., Damin, F., et al., *Protein and peptide arrays: Recent trends and new directions.* Biomolecular Engineering, 2006. **23**(2): p. 77-88.
70. Thiele, A., Stangl, G.I., et al., *Deciphering Enzyme Function Using Peptide Arrays.* Molecular Biotechnology, 2011. **49**(3): p. 283.
71. Szymczak, L.C., Kuo, H.-Y., et al., *Peptide Arrays: Development and Application.* Analytical Chemistry, 2018. **90**(1): p. 266-282.
72. Smith, G.P., *Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface.* Science, 1985. **228**(4705): p. 1315.
73. Szardenings, M., *Phage Display of Random Peptide Libraries: Applications, Limits, and Potential.* Journal of Receptors and Signal Transduction, 2003. **23**(4): p. 307-349.
74. Zou, Q. and Yang, K.-L., *Identification of peptide inhibitors of penicillinase using a phage display library.* Analytical Biochemistry, 2016. **494**: p. 4-9.
75. Lipp, A.-M., Ji, B., et al., *Micro-structured peptide surfaces for the detection of high-affinity peptide–receptor interactions in living cells.* Biosensors and Bioelectronics, 2015. **74**: p. 757-763.
76. Hetrick, K.J., Walker, M.C., et al., *Development and Application of Yeast and Phage Display of Diverse Lanthipeptides.* ACS Central Science, 2018. **4**(4): p. 458-467.
77. Ryvkin, A., Ashkenazy, H., et al., *Phage display peptide libraries: deviations from randomness and correctives.* Nucleic acids research, 2018. **46**(9): p. e52-e52.
78. Wood, S.E., Sinsinbar, G., et al., *A Bottom-Up Proteomic Approach to Identify Substrate Specificity of Outer-Membrane Protease OmpT.* Angew. Chem., Int. Ed., 2017. **56**(52): p. 16531.

79. Kalff, F.E., Rebergen, M.P., et al., *A kilobyte rewritable atomic memory.* Nat Nanotechnol, 2016. **11**(11): p. 926-929.
80. Ahn, J., *Information Storage and Retrieval Through Quantum Phase.* Science, 2000. **287**(5452): p. 463-465.
81. Zhang, J., Gecevicius, M., et al., *Seemingly unlimited lifetime data storage in nanostructured glass.* Phys Rev Lett, 2014. **112**(3): p. 033901.
82. Gu, M., Zhang, Q., et al., *Nanomaterials for optical data storage.* Nature Reviews Materials, 2016. **1**(12).
83. Begtrup, G.E., Gannett, W., et al., *Nanoscale reversible mass transport for archival memory.* Nano Lett, 2009. **9**(5): p. 1835-8.
84. Clelland, C.T., Risca, V., et al., *Hiding messages in DNA microdots.* Nature, 1999. **399**(6736): p. 533-4.
85. Zhirnov, V., Zadegan, R.M., et al., *Nucleic acid memory.* Nat Mater, 2016. **15**(4): p. 366-70.
86. Mrksich, M., *Mass spectrometry of self-assembled monolayers: a new tool for molecular surface science.* ACS Nano, 2008. **2**(1): p. 7-18.
87. Keller, B.O. and Li, L., *Detection of 25,000 molecules of Substance P by MALDI-TOF mass spectrometry and investigations into the fundamental limits of detection in MALDI.* Journal of the American Society for Mass Spectrometry, 2001. **12**(9): p. 1055-1063.
88. Singh, M., Haverinen, H.M., et al., *Inkjet printing-process and its applications.* Adv Mater, 2010. **22**(6): p. 673-85.
89. Derby, B., *Inkjet Printing of Functional and Structural Materials: Fluid Property Requirements, Feature Stability, and Resolution.* Annual Review of Materials Research, 2010. **40**(1): p. 395-414.
90. Martin, R.B., *Free energies and equilibria of peptide bond hydrolysis and formation.* Biopolymers, 1998. **45**(5): p. 351-353.
91. Hughes, R.A. and Ellington, A.D., *Synthetic DNA Synthesis and Assembly: Putting the Synthetic in Synthetic Biology.* Cold Spring Harb Perspect Biol, 2017. **9**(1).
92. Beaucage, S.L. and Caruthers, M.H., *Deoxynucleoside phosphoramidites—A new class of key intermediates for deoxypolynucleotide synthesis.* Tetrahedron Letters, 1981. **22**(20): p. 1859-1862.
93. Panda, D., Molla, K.A., et al., *DNA as a digital information storage device: hope or hype?* 3 Biotech, 2018. **8**(5): p. 239.
94. Cafferty, B.J., Ten, A.S., et al., *Storage of Information Using Small Organic Molecules.* ACS Central Science, 2019. **5**(5): p. 911-916.
95. Hankin, J.A. and Murphy, R.C., *Relationship between MALDI IMS intensity and measured quantity of selected phospholipids in rat brain sections.* Analytical chemistry, 2010. **82**(20): p. 8476-8484.
96. Grant, J., O'Kane, P.T., et al., *Using Microfluidics and Imaging SAMDI-MS To Characterize Reaction Kinetics.* ACS Central Science, 2019. **5**(3): p. 486-493.
97. Grant, J., Goudarzi, S.H., et al., *High-Throughput Enzyme Kinetics with 3D Microfluidics and Imaging SAMDI Mass Spectrometry.* Analytical Chemistry, 2018. **90**(21): p. 13096-13103.
98. Tanaka, K., Murakami, L., et al., *Multilevel Recording with Multilayer Magneto Optical Media by Light Intensity Modulation.* Japanese Journal of Applied Physics, 2002. **41**(Part 1, No. 3B): p. 1647-1649.

99. Pavan, P., Bez, R., et al., *Flash memory cells-an overview.* Proceedings of the IEEE, 1997. **85**(8): p. 1248-1271.
100. Bez, R., Camerlenghi, E., et al., *Introduction to flash memory.* Proceedings of the IEEE, 2003. **91**(4): p. 489-502.
101. Cai, Y., Yalcin, G., et al., *ERROR ANALYSIS AND RETENTION-AWARE ERROR MANAGEMENT FOR NAND FLASH MEMORY.* Intel Technology Journal, 2013. **17**(1).
102. Cambou, B., Flikkema, G.P., et al., *Can Ternary Computing Improve Information Assurance?* Cryptography, 2018. **2**(1).
103. Van Bogart, J.W.C., Commission on, P., et al., *Magnetic tape storage and handling : a guide for libraries and archives*. 1995, Washington, DC; St. Paul, MN: Commission on Preservation and Access ; National Media Laboratory.
104. Winter, M., Bretschneider, T., et al., *Establishing MALDI-TOF as Versatile Drug Discovery Readout to Dissect the PTP1B Enzymatic Reaction.* SLAS DISCOVERY: Advancing Life Sciences R&D, 2018. **23**(6): p. 561-573.
105. Cornett, D.S., Reyzer, M.L., et al., *MALDI imaging mass spectrometry: molecular snapshots of biochemical systems.* Nature Methods, 2007. **4**(10): p. 828-833.
106. Waasdorp, R., van den Heuvel, O., et al., *Accessing individual 75-micron diameter nozzles of a desktop inkjet printer to dispense picoliter droplets on demand.* RSC Advances, 2018. **8**(27): p. 14765-14774.
107. Brossard, F.S.F., Pecunia, V., et al., *Inkjet-Printed Nanocavities on a Photonic Crystal Template.* Advanced Materials, 2017. **29**(47): p. 1704425.
108. Li, Q., Yi, L., et al., *Commercial proteases: Present and future.* FEBS Lett., 2013. **587**(8): p. 1155.
109. Sawant, R. and Nagendran, S., *Protease: An Enzyme with Multiple Industrial Applications.* World J. Pharm. Pharm. Sci., 2014. **3**(6): p. 568.
110. Cherry, J.R. and Fidantsef, A.L., *Directed evolution of industrial enzymes: an update.* Curr. Opin. Biotechnol., 2003. **14**(4): p. 438.
111. Vojcic, L., Pitzler, C., et al., *Advances in protease engineering for laundry detergents.* New Biotechnol., 2015. **32**(6): p. 629.
112. Shahid, S.M.T. and Ahmed, K., *Enzyme Proteases Used In Laundry Detergents Engineering.* Sci. Int., 2016. **28**(3): p. 2711.
113. Maurer, K.H., *Detergent proteases.* Curr. Opin. Biotechnol., 2004. **15**(4): p. 330.
114. Kim, J.H., Roy, S., et al., *Protease Adsorption and Reaction on an Immobilized Substrate Surface.* Langmuir, 2002. **18**(16): p. 6312.
115. Pierce, J.A., Robertson, C.R., et al., *Physiological and Genetic Strategies for Enhanced Subtilisin Production by Bacillus subtilis.* Biotechnol. Prog., 1992. **8**(3): p. 211.
116. Kirk, O., Borchert, T.V., et al., *Industrial enzyme applications.* Curr. Opin. Biotechnol., 2002. **13**(4): p. 345.
117. Wintrode, P.L., Miyazaki, K., et al., *Cold Adaptation of a Mesophilic Subtilisn-like Protease by Laboratory Evolution.* J. Biol. Chem., 2000. **275**(41): p. 31635.
118. Banik, R.M. and Prakash, M., *Laundry detergent compatibility of the alkaline protease from Bacillus cereus.* Microbiol. Res., 2004. **159**(2): p. 135.
119. Oberoi, R., Beg, Q.K., et al., *Characterization and wash performance analysis of an SDS-stable lkaline protease from a Bacillus sp.* World J. Microbiol. Biotechnol., 2001. **17**(5): p. 493.

120. Choe, Y., Leonetti, F., et al., *Substrate Profiling of Cysteine Proteases Using a Combinatorial Peptide Library Identifies Functionally Unique Specificities.* J. Biol. Chem., 2006. **281**(18): p. 12824.
121. Li, Q., Yi, L., et al., *Profiling Protease Specificity: Combining Yeast ER Sequestration Screening (YESS) with Next Generation Sequencing.* ACS Chem. Biol., 2017. **12**(2): p. 510.
122. Gurard-Levin, Z.A., Scholle, M.D., et al., *High-Throughput Screening of Small Molecule Libraries using SAMDI Mass Spectrometry.* ACS Comb. Sci., 2011. **13**(4): p. 347.
123. Castaneda, C.A., Lopez, J.E., et al., *Fierke, Active Site Metal Identity Alters HDAC8 Substrate Selectivity: A Potential Novel Regulatory Mechanism.* Biochemistry, 2017. **56**(42): p. 5663.
124. Füßl, M., Lassowskat, I., et al., *Beyond Histones: New Substrate Proteins of Lysine Deacetylases in Arabidopsis Nuclei.* Frontiers in plant science, 2018. **9**: p. 461-461.
125. Allfrey, V.G., Faulkner, R., et al., *ACETYLATION AND METHYLATION OF HISTONES AND THEIR POSSIBLE ROLE IN THE REGULATION OF RNA SYNTHESIS.* Proceedings of the National Academy of Sciences of the United States of America, 1964. **51**(5): p. 786-794.
126. Brownell, J.E., Zhou, J., et al., *Tetrahymena Histone Acetyltransferase A: A Homolog to Yeast Gcn5p Linking Histone Acetylation to Gene Activation.* Cell, 1996. **84**(6): p. 843-851.
127. Taunton, J., Hassig, C.A., et al., *A Mammalian Histone Deacetylase Related to the Yeast Transcriptional Regulator Rpd3p.* Science, 1996. **272**(5260): p. 408.
128. Roth, S.Y., Denu, J.M., et al., *Histone Acetyltransferases.* Annual Review of Biochemistry, 2001. **70**(1): p. 81-120.
129. Haberland, M., Montgomery, R.L., et al., *The many roles of histone deacetylases in development and physiology: implications for disease and therapy.* Nature Reviews Genetics, 2009. **10**(1): p. 32-42.
130. Dokmanovic, M., Clarke, C., et al., *Histone Deacetylase Inhibitors: Overview and Perspectives.* Molecular Cancer Research, 2007. **5**(10): p. 981.
131. Yang, X.-J. and Seto, E., *The Rpd3/Hda1 family of lysine deacetylases: from bacteria and yeast to mice and men.* Nature Reviews Molecular Cell Biology, 2008. **9**(3): p. 206-218.
132. Sauve, A.A. and Youn, D.Y., *Sirtuins: NAD+-dependent deacetylase mechanism and regulation.* Current Opinion in Chemical Biology, 2012. **16**(5): p. 535-543.
133. Kupis, W., Pałyga, J., et al., *The role of sirtuins in cellular homeostasis.* Journal of physiology and biochemistry, 2016. **72**(3): p. 371-380.
134. Kölle, D., Brosch, G., et al., *Biochemical Methods for Analysis of Histone Deacetylases.* Methods, 1998. **15**(4): p. 323-331.
135. Fatkins, D.G. and Zheng, W., *A spectrophotometric assay for histone deacetylase 8.* Analytical Biochemistry, 2008. **372**(1): p. 82-88.
136. Riester, D., Hildmann, C., et al., *Histone deacetylase inhibitor assay based on fluorescence resonance energy transfer.* Analytical Biochemistry, 2007. **362**(1): p. 136-141.
137. Gurard-Levin, Z.A. and Mrksich, M., *The Activity of HDAC8 Depends on Local and Distal Sequences of Its Peptide Substrates.* Biochemistry, 2008. **47**(23): p. 6242-6250.
138. Hu, E., Chen, Z., et al., *Cloning and characterization of a novel human class I histone deacetylase that functions as a transcription repressor.* J Biol Chem, 2000. **275**(20): p. 15254-64.

139. Okada, H., Uezu, A., et al., *Peptide Array X-Linking (PAX): A New Peptide-Protein Identification Approach.* PLOS ONE, 2012. **7**(5): p. e37035.
140. Howitz, K.T., Bitterman, K.J., et al., *Small molecule activators of sirtuins extend Saccharomyces cerevisiae lifespan.* Nature, 2003. **425**(6954): p. 191-196.
141. Beher, D., Wu, J., et al., *Resveratrol is Not a Direct Activator of SIRT1 Enzyme Activity.* Chemical Biology & Drug Design, 2009. **74**(6): p. 619-624.
142. Chan, E.W.L., Yousaf, M.N., et al., *Understanding the Role of Adsorption in the Reaction of Cyclopentadiene with an Immobilized Dienophile.* The Journal of Physical Chemistry A, 2000. **104**(41): p. 9315-9320.
143. Li, S. and Mrksich, M., *An Unusual Salt Effect in an Interfacial Nucleophilic Substitution Reaction.* Langmuir, 2018. **34**(23): p. 6713-6718.
144. Eisenberg, J.L., Piper, J.L., et al., *Using Self-Assembled Monolayers to Model Cell Adhesion to the 9th and 10th Type III Domains of Fibronectin.* Langmuir, 2009. **25**(24): p. 13942-13951.
145. Sánchez-Cortés, J., Bähr, K., et al., *Cell Adhesion to Unnatural Ligands Mediated by a Bifunctional Protein.* Journal of the American Chemical Society, 2010. **132**(28): p. 9733-9737.
146. Behrendt, R., White, P., et al., *Advances in Fmoc solid-phase peptide synthesis.* Journal of Peptide Science, 2016. **22**(1): p. 4-27.
147. Camperi, S.A., Marani, M.M., et al., *An efficient strategy for the preparation of one-bead-one-peptide libraries on a new biocompatible solid support.* Tetrahedron Letters, 2005. **46**(9): p. 1561-1564.
148. Hu, B.-H., Jones, M.R., et al., *Method for screening and MALDI-TOF MS sequencing of encoded combinatorial libraries.* Analytical chemistry, 2007. **79**(19): p. 7275-7285.
149. Sternson, S.M., Louca, J.B., et al., *Split−Pool Synthesis of 1,3-Dioxanes Leading to Arrayed Stock Solutions of Single Compounds Sufficient for Multiple Phenotypic and Protein-Binding Assays.* Journal of the American Chemical Society, 2001. **123**(8): p. 1740-1747.
150. Michael, K.L., Taylor, L.C., et al., *Randomly Ordered Addressable High-Density Optical Sensor Arrays.* Analytical Chemistry, 1998. **70**(7): p. 1242-1248.
151. Ferguson, J.A., Steemers, F.J., et al., *High-Density Fiber-Optic DNA Random Microsphere Array.* Analytical Chemistry, 2000. **72**(22): p. 5618-5624.
152. Youngquist, R.S., Fuentes, G.R., et al., *Generation and screening of combinatorial peptide libraries designed for rapid sequencing by mass spectrometry.* Journal of the American Chemical Society, 1995. **117**(14): p. 3900-3906.
153. St. Hilaire, P.M., Lowary, T.L., et al., *Oligosaccharide Mimetics Obtained by Novel, Rapid Screening of Carboxylic Acid Encoded Glycopeptide Libraries.* Journal of the American Chemical Society, 1998. **120**(51): p. 13312-13320.
154. Lee, S.S., Lim, J., et al., *Accurate MALDI-TOF/TOF Sequencing of One-Bead−One-Compound Peptide Libraries with Application to the Identification of Multiligand Protein Affinity Agents Using in Situ Click Chemistry Screening.* Analytical Chemistry, 2010. **82**(2): p. 672-679.
155. Cha, J., Lim, J., et al., *Process Automation toward Ultra-High-Throughput Screening of Combinatorial One-Bead-One-Compound (OBOC) Peptide Libraries.* Journal of Laboratory Automation, 2012. **17**(3): p. 186-200.

156. Neises, B. and Steglich, W., *Simple Method for the Esterification of Carboxylic Acids.* Angewandte Chemie International Edition in English, 1978. **17**(7): p. 522-524.
157. Xue, A.Y., Szymczak, L.C., et al., *Machine Learning on Signal-to-Noise Ratios Improves Peptide Array Design in SAMDI Mass Spectrometry.* Analytical Chemistry, 2017. **89**(17): p. 9039-9047.
158. Pierschbacher, M.D. and Ruoslahti, E., *Cell attachment activity of fibronectin can be duplicated by small synthetic fragments of the molecule.* Nature, 1984. **309**(5963): p. 30-33.
159. Pytela, R., Pierschbacher, M.D., et al., *Platelet membrane glycoprotein IIb/IIIa: member of a family of Arg-Gly-Asp--specific adhesion receptors.* Science, 1986. **231**(4745): p. 1559.
160. Pytela, R., Pierschbacher, M.D., et al., *Identification and isolation of a 140 kd cell surface glycoprotein with properties expected of a fibronectin receptor.* Cell, 1985. **40**(1): p. 191-198.
161. Hynes, R.O., *Integrins: A family of cell surface receptors.* Cell, 1987. **48**(4): p. 549-554.
162. Ruoslahti, E. and Pierschbacher, M.D., *New perspectives in cell adhesion: RGD and integrins.* Science, 1987. **238**(4826): p. 491.
163. Hynes, R.O., *The emergence of integrins: a personal and historical perspective.* Matrix biology : journal of the International Society for Matrix Biology, 2004. **23**(6): p. 333-340.
164. Mrksich, M., *A surface chemistry approach to studying cell adhesion.* Chemical Society Reviews, 2000. **29**(4): p. 267-273.
165. Kato, M. and Mrksich, M., *Using Model Substrates To Study the Dependence of Focal Adhesion Formation on the Affinity of Integrin−Ligand Complexes.* Biochemistry, 2004. **43**(10): p. 2699-2707.
166. Eisenberg, J.L., Beaumont, K.G., et al., *Plectin-containing, centrally localized focal adhesions exert traction forces in primary lung epithelial cells.* Journal of Cell Science, 2013. **126**(16): p. 3746.
167. Beaumont, K.G. and Mrksich, M., *The mechanostability of isolated focal adhesions is strongly dependent on pH.* Chemistry & biology, 2012. **19**(6): p. 711-720.
168. Corti, A. and Curnis, F., *Isoaspartate-dependent molecular switches for integrin–ligand recognition.* Journal of Cell Science, 2011. **124**(4): p. 515.
169. Burns, A.R., Kwok, T.C.Y., et al., *High-throughput screening of small molecules for bioactivity and target identification in Caenorhabditis elegans.* Nature Protocols, 2006. **1**(4): p. 1906-1914.

Appendix

Appendix Table 2.1: Assignment of 32 peptides to 32 molbits (4 molbytes) by ascending mass-to-charge ratio

Molbyte Number	Molbit Number	Peptide Number	sequence	m.w. (g mol^{-1})	m/z obs.* (a.m.u.)
I	1	1	Ac-AK$^{(me3)}$C	404	1255
	2	2	Ac-(abu)K$^{(me3)}$C	418	1269
	3	3	Ac-VK$^{(me3)}$C	432	1283
	4	4	Ac-GGK$^{(me3)}$C	447	1298
	5	5	Ac-GVK$^{(me3)}$C	489	1340
	6	6	Ac-GLK$^{(me3)}$C	503	1354
	7	7	Ac-ALK$^{(me3)}$C	517	1368
	8	8	Ac-GFK$^{(me3)}$C	537	1388
II	1	9	Ac-GVGK$^{(me3)}$C	546	1397
	2	10	Ac-GLGK$^{(me3)}$C	560	1411
	3	11	Ac-GAGGK$^{(me3)}$C	575	1426
	4	12	Ac-GL(abu)K$^{(me3)}$C	588	1439
	5	13	Ac-GFGK$^{(me3)}$C	594	1445
	6	14	Ac-GRGK$^{(me3)}$C	603	1454
	7	15	Ac-GPAGK$^{(me3)}$C	615	1466
	8	16	Ac-AYGK$^{(me3)}$C	624	1475
III	1	17	Ac-GPFK$^{(me3)}$C	634	1485
	2	18	Ac-GVVGK$^{(me3)}$C	645	1496
	3	19	Ac-G(abu)FGK$^{(me3)}$C	679	1530
	4	20	Ac-GVFGK$^{(me3)}$C	693	1544
	5	21	Ac-GVYGK$^{(me3)}$C	709	1560
	6	22	Ac-GARGGK$^{(me3)}$C	731	1582
	7	23	Ac-GAVV(abu)K$^{(me3)}$C	744	1595
	8	24	Ac-GFYGK$^{(me3)}$C	757	1608
IV	1	25	Ac-GYYGK$^{(me3)}$C	773	1624
	2	26	Ac-GYYAK$^{(me3)}$C	787	1638
	3	27	Ac-GPYFK$^{(me3)}$C	797	1648
	4	28	Ac-GRGFGK$^{(me3)}$C	807	1658
	5	29	Ac-GYFGGK$^{(me3)}$C	814	1665
	6	30	Ac-GYYGGK$^{(me3)}$C	830	1681
	7	31	Ac-AYYGGK$^{(me3)}$C	844	1695
	8	32	Ac-GYY(abu)GK$^{(me3)}$C	858	1709

Ac: acetyl, Abu: 2-aminobutyric acid, A: alanine, G: glycine, C: cysteine, F: phenylalanine, K$^{(me3)}$: N$^{\varepsilon}$,N$^{\varepsilon}$,N$^{\varepsilon}$-trimethyl lysine L: leucine, P: proline, R: arginine, V: valine, Y: tyrosine, a.m.u.: atomic mass unit.

*The observed mass-to-charge ratio (m/z) includes the peptide plus the mixed disulfide derived from the SAM, which is formed during laser-assisted desorption/ionization (+ 851 a.m.u.).

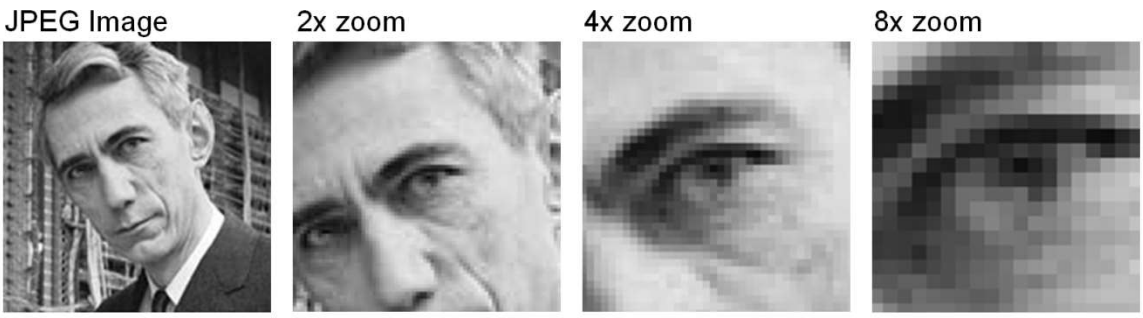
Appendix Table 2.2: List of printable characters encoded using mixtures of molbits and their 8-bit ASCII character code*

A	01000001	Q	01010001	g	01100111	w	01110111	!	00100001	;	00111011
B	01000010	R	01010010	h	01101000	x	01111000	“	00100010	<	00111100
C	01000011	S	01010011	i	01101001	y	01111001	#	00100011	=	00111101
D	01000100	T	01010100	j	01101010	z	01111010	$	00100100	>	00111110
E	01000101	U	01010101	k	01101011	0	00110000	%	00100101	?	00111111
F	01000110	V	01010110	l	01101100	1	00110001	&	00100110	[	01011011
G	01000111	W	01010111	m	01101101	2	00110010	‘	00100111	\	01011100
H	01001000	X	01011000	n	01101110	3	00110011	(	00101000	]	01011101
I	01001001	Y	01011001	o	01101111	4	00110100	)	00101001	^	01011110
J	01001010	Z	01011010	p	01110000	5	00110101	*	00101010	_	01011111
K	01001011	a	01100001	q	01110001	6	00110110	+	00101011	`	01100000
L	01001100	b	01100010	r	01110010	7	00110111	,	00101100	{	01111011
M	01001101	c	01100011	s	01110011	8	00111000	-	00101101	|	01111100
N	01001110	d	01100100	t	01110100	9	00111001	.	00101110	}	01111101
O	01001111	e	01100101	u	01110101	“space”	00100000	/	00101111	~	01111110
P	01010000	f	01100110	v	01110110	@	01000000	:	00111010		

* These are the characters that *have* been written, not all possible character that *can* be written.

Appendix Table 2.3: Properties of images and JPEG conversion parameters.

Image name	Compressed Dimensions (px)	JPEG Quality	JPEG Blur	Compressed Size (bits(b) /bytes(B))	Retrieval Error
Claud Shannon	150 x 150	49	0.2	48,752b/6,094B	0.00 %
The Great Wave off Kanagawa	240 x 200	11	0.2	47,624b/5,953B	0.00 %



Appendix Figure 2.1: Losses during compression of an image to a JPEG file. The original and reconstituted JPEG images exhibit pixilation that is more visible at high magnification, a consequence of the compression required to reduce the image size.

Text of Feynman's "There is plenty of room at the bottom" written using 32-molbit encoding strategy on 7 MALDI target plates. Characters (bytes) that were misidentified are in red, underlined and bolded. 301 bits of 306,504 total bits (0.01%) were misidentified.

Plate 1

Plenty of Room at the Bottom Richard P. Feynman (Dated: Dec. 1959). I imagine experimental physicists must often look with envy at men like Kamerlingh Onnes, who discovered a field like low temperature, which seems to be bottomless and in which one can go down and down. Such a man is then a leader and has some temporary monopoly in a scientific adventure. Percy Bridgman, in designing a way to obtain higher pressures, opened up another new field and was able to move into it and to lead us all along. The development of ever higher vacuum was a continuing development of the same kind. I would like to describe a field, in which little has been done, but in which an enormous amount can be done in principle. This field is not quite the same as the others in that it will not tell us much of fundamental physics (in the sense of, "What are the strange particles?") but it is more like solid-state physics in the sense that it might tell us much of great interest about the strange phenomena that occur in complex situations. Furthermore, a point that is most important is that it would have an enormous number of technical applications. What I want to talk about is the problem of manipulating and controlling things on a small scale. As soon as I mention this, people tell me about miniaturization, and how far it has progressed today. They tell me about electric motors that are the size of the nail on your small finger. And there is a device on the market, they tell me, by which you can write the Lord's Prayer on the head of a pin. But that's

nothing; that's the most primitive, halting step in the direction I intend to discuss. It is a staggeringly small world that is below. In the year 2000, when they look back at this age, they will wonder why it was not until the year 1960 that anybody began seriously to move in this direction. Why cannot we write the entire 24 volumes of the Encyclopedia Brittanica on the head of a pin? Let's see what would be involved. The head of a pin is a sixteenth of an inch across. If you magnify it by 25,000 diameters, the area of the head of the pin is then equal to the area of all the pages of the Encyclopedia Brittanica. Therefore, all it is necessary to do is to reduce in size all the writing in the Encyclopedia by 25,000 times. Is that possible? The resolving power of the eye is about 1/120 of an inch-that is roughly the diameter of one of the little dots on the fine half-tone reproductions in the Encyclopedia. This, when you demagnify it by 25,000 times, is still 80 angstroms in diameter-32 atoms across, in an ordinary metal. In other words, one of those dots still would contain in its area 1,000 atoms. So, each dot can easily be adjusted in size as required by the photoengraving, and there is no question that there is enough room on the head of a pin to put all of the Encyclopedia Brittanica. Furthermore, it can be read if it is so written. Let's imagine that it is written in raised letters of metal; that is, where the black is in the Encyclopedia, we have raised letters of metal that are actually 1/25,000 of their ordinary size. How would we read it? If we had something written in such a way, we could read it using techniques in common use today. (They will undoubtedly find a better way when we do actually have it written, but to make my point conservatively I shall just take techniques we know today.) We would press the metal into a plastic material and make a mold of it, then peel the plastic off very carefully, evaporate silica into the plastic to get a very thin film, then shadow it by evaporating gold at an angle against the silica so that all the little letters will appear clearly, dissolve the plastic away from the silica film, and then look through it with an electron microscope! There is no question that if the thing were reduced by 25,000 times in the form of raised letters on the pin, it would be easy for us to read it today. Furthermore; there is no question that we would find it easy to make copies of the master; we would just need to press the same metal plate again into plastic and we would have another copy. How do we write small? The next question is: How do we write it? We have no standard technique to do this now. But let me argue that it is not as difficult as it first appears to be. We can reverse the lenses of the electron microscope in order to demagnify as well as magnify. A source of ions, sent through the microscope lenses in reverse, could be focused to a very small spot. We could write with that spot like we write in a TV cathode ray oscilloscope, by going across in lines, and

having an adjustment which determines the amount of material which is going to be deposited as we scan in lines. This method might be very slow because of space charge limitations. There will be more rapid methods. We could first make, perhaps by some photo process, a screen which has holes in it in the form of the letters. Then we would strike an arc behind the holes and draw metallic ions through the holes; then we could again use our system of lenses and make a small image in the form of ions, which would deposit the metal on the pin. A simpler way might be this (though I am not sure it 2 would work): We take light and, through an optical microscope running backwards, we focus it onto a very small photoelectric screen. Then electrons come away from the screen where the light is shining. These electrons are focused down in size by the electron microscope lenses to impinge directly upon the surface of the metal. Will such a beam etch away the metal if it is run long enough? I don't know. If it doesn't work for a metal surface, it must be possible to find some surface with which to coat the original pin so that, where the electrons bombard, a change is made which we could recognize later. There is no intensity problem in these devices-not what you are used to in magnification, where you have to take a few electrons and spread them over a bigger and bigger screen; it is just the opposite. The light which we get from a page is concentrated onto a very small area so it is very intense. The few electrons which com

Plate 2

e from the photoelectric screen are demagnified down to a very tiny area so that, again, they are very intense. I don't know why this hasn't been done yet! That's the Encyclopedia Brittanica on the head of a pin, but let's consider all the books in the world. The Library of Congress has approximately 9 million volumes; the British Museum Library has 5 million volumes; there are also 5 million volumes in the National Library in France. Undoubtedly there are duplications, so let us say that there are some 24 million volumes of interest in the world. What would happen if I print all this down at the scale we have been discussing? How much space would it take? It would take, of course, the area of about a million pinheads because, instead of there being just the 24 volumes of the Encyclopaedia, there are 24 million volumes. The million pinheads can be put in a square of a thousand pins on a side, or an area of about 3 square yards. That is to say, the silica replica with the paper-thin backing of plastic, with which we have made the copies, with all this

information, is on an area of approximately the size of 35 pages of the Encyclopedia. That is about half as many pages as there are in this magazine. All of the information which all of mankind has every recorded in books can be carried around in a pamphlet in your hand-and not written in code, but a simple reproduction of the original pictures, engravings, and everything else on a small scale without loss of resolution. What would our librarian at Caltech say, as she runs all over from one building to another, if I tell her that, ten years from now, all of the information that she is struggling to keep track of- 120,000 volumes, stacked from the floor to the ceiling, drawers full of cards, storage rooms full of the older books-can be kept on just one library card! When the University of Brazil, for example, finds that their library is burned, we can send them a copy of every book in our library by striking off a copy from the master plate in a few hours and mailing it in an envelope no bigger or heavier than any other ordinary air mail letter. Now, the name of this talk is "There is Plenty of Room at the Bottom"-not just "There is Room at the Bottom." What I have demonstrated is that there is room- that you can decrease the size of things in a practical way. I now want to show that there is plenty of room. I will not now discuss how we are going to do it, but only what is possible in principle-in other words, what is possible according to the laws of physics. I am not inventing anti-gravity, which is possible someday only if the laws are not what we think. I am telling you what could be done if the laws are what we think; we are not doing it simply because we haven't yet gotten around to it. - Information on a small scale - Suppose that, instead of trying to reproduce the pictures and all the information directly in its present form, we write only the information content in a code of dots and dashes, or something like that, to represent the various letters. Each letter represents six or seven "bits" of information; that is, you need only about six or seven dots or dashes for each letter. Now, instead of writing everything, as I did before, on the surface of the head of a pin, I am going to use the interior of the material as well. Let us represent a dot by a small spot of one metal, the next dash, by an adjacent spot of another metal, and so on. Suppose, to be conservative, that a bit of information is going to require a little cube of atoms 5 times 5 times 5-that is 125 atoms. Perhaps we need a hundred and some odd atoms to make sure that the information is not lost through diffusion, or through some other process. I have estimated how many letters there are in the Encyclopaedia, and I have assumed that each of my 24 million books is as big as an Encyclopaedia volume, and have calculated, then, how many bits of information there are (1015). For each bit I allow 100 atoms. And it turns out that all of the information that man has carefully accumulated in all the books in the world can be written in this

form in a cube of material one two-hundredth of an inch wide- which is the barest piece of dust that can be made out by the human eye. So there is plenty of room at the bottom! Don't tell me about microfilm! This fact-that enormous amounts of information can be carried in an exceedingly small space-is, of course, well known to the biologists, and resolves the mystery which existed before we understood all this clearly, of how it could be that, in the tiniest cell, all of the information for the organization of a complex creature such as ourselves can be stored. All this information-whether we have brown eyes, or whether we think at all, or that in the embryo the jawbone should first develop with a little hole in the side so that later a nerve can grow through it-all this information is contained in a very tiny fraction of the cell in the form of long-chain DNA molecules in which approximately 50 atoms are used for one bit of 3 information about the cell. - Better electron microscopes - If I have written in a code, with 5 times 5 times 5 atoms to a bit, the question is: How could I read it today? The electron microscope is not quite good enough, with the greatest care and effort, it can only resolve about 10 angstroms. I would like to try and impress upon you while I am talking about all of these things on a small scale, the importance of improving the electron microscope by a hundred times. It is not impossible; it is not against the laws of diffraction of the electron. The wave length of the electron in such a microscope is only 1/20 of an angstrom. So it should be possible to see the individual atoms. What good would it be to see individual atoms distinctly? We have friends in other fields-in biology, for instance. We physicists often look at them and say, "You know the reason you fellows are making so little progress?" (Actually I don't know any field where they are making more rapid progress than they are in biology today.) "You should use more mathematics, like we do." They coul

Plate 3

d answer us-but they're polite, so I'll answer for them: "What you should do in order for us to make more rapid progress is to make the electron microscope 100 times better." What are the most central and fundamental problems of biology today? They are questions like: What is the sequence of bases in the DNA? What happens when you have a mutation? How is the base order in the DNA connected to the order of amino acids in the protein? What is the structure of the RNA; is it single-chain or double-chain, and how is it related in its order of bases to the DNA? What is the

organization of the microsomes? How are proteins synthesized? Where does the RNA go? How does it sit? Where do the proteins sit? Where do the amino acids go in? In photosynthesis, where is the chlorophyll; how is it arranged; where are the carotenoids involved in this thing? What is the system of the conversion of light into chemical energy? It is very easy to answer many of these fundamental biological questions; you just look at the thing! You will see the order of bases in the chain; you will see the structure of the microsome. Unfortunately, the present microscope sees at a scale which is just a bit too crude. Make the microscope one hundred times more powerful, and many problems of biology would be made very much easier. I exaggerate, of course, but the biologists would surely be very thankful to you-and they would prefer that to the criticism that they should use more mathematics. The theory of chemical processes today is based on theoretical physics. In this sense, physics supplies the foundation of chemistry. But chemistry also has analysis. If you have a strange substance and you want to know what it is, you go through a long and complicated process of chemical analysis. You can analyze almost anything today, so I am a little late with my idea. But if the physicists wanted to, they could also dig under the chemists in the problem of chemical analysis. It would be very easy to make an analysis of any complicated chemical substance; all one would have to do would be to look at it and see where the atoms are. The only trouble is that the electron microscope is one hundred times too poor. (Later, I would like to ask the question: Can the physicists do something about the third problem of chemistry-namely, synthesis? Is there a physical way to synthesize any chemical substance? The reason the electron microscope is so poor is that the f- value of the lenses is only 1 part to 1,000; you don't have a big enough numerical aperture. And I know that there are theorems which prove that it is impossible, with axially symmetrical stationary field lenses, to produce an f-value any bigger than so and so; and therefore the resolving power at the present time is at its theoretical maximum. But in every theorem there are assumptions. Why must the field be symmetrical? I put this out as a challenge: Is there no way to make the electron microscope more powerful? - The marvelous biological system - The biological example of writing information on a small scale has inspired me to think of something that should be possible. Biology is not simply writing information; it is doing something about it. A biological system can be exceedingly small. Many of the cells are very tiny, but they are very active; they manufacture various substances; they walk around; they wiggle; and they do all kinds of marvelous things-all on a very small scale. Also, they store information. Consider the possibility that we too can make a thing very small which does what we want-that

we can manufacture an object that maneuvers at that level! There may even be an economic point to this business of making things very small. Let me remind you of some of the problems of computing machines. In computers we have to store an enormous amount of information. The kind of writing that I was mentioning before, in which I had everything down as a distribution of metal, is permanent. Much more interesting to a computer is a way of writing, erasing, and writing something else. (This is usually because we don't want to waste the material on which we have just written. Yet if we could write it in a very small space, it wouldn't make any difference; it could just be thrown away after it was read. It doesn't cost very much for the material). - Miniaturizing the computer - I don't know how to do this on a small scale in a practical way, but I do know that computing machines are very large; they fill rooms. Why can't we make them very small, make them of little wires, little elements-and by little, I mean little. For instance, the wires should be 10 or 100 atoms in diameter, and the circuits should be a few thousand angstroms across. Everybody who has analyzed the logical theory of computers has come to the 4 conclusion that the possibilities of computers are very interesting-if they could be made to be more complicated by several orders of magnitude. If they had millions of times as many elements, they could make judgments. They would have time to calculate what is the best way to make the calculation that they are about to make. They could select the method of analysis which, from their experience, is better than the one that we would give to them. And in many other ways, they would have new qualitative features. If I look at your face I immediately recognize that I have seen it before. (Actually, my friends will say I have chosen an unfortunate example here for the subject of this illustration. At least I recognize that it is a man and not an apple.) Yet there is no machine which, with that speed, can take a picture of a face and say even that it is a man; and much less that it is the same man that you showed it before-unless it is exactly the same pict**u**re. If the face is changed; if I am closer to the face; if I am further from the face; if the light changes-I recognize it anyway. Now, this little computer I carry in my head is easily able to do that. The computers that we build are not able to do that. The number of elements in this bone box of mine are enormously greater than the number of elements in our "wonderful" com

Plate 4

puters. But our mechanical computers are too big; the elements in this box are microscopic. I want to make some that are submicroscopic. If we wanted to make a computer that had all these marvelous extra qualitative abilities, we would have to make it, perhaps, the size of the Pentagon. This has several disadvantages. First, it requires too much material; there may not be enough germanium in the world for all the transistors which would have to be put into this enormous thing. There is also the problem of heat generation and power consumption; TVA would be needed to run the computer. But an even more practical difficulty is that the computer would be limited to a certain speed. Because of its large size, there is finite time required to get the information from one place to another. The information cannot go any faster than the speed of light-so, ultimately, when our computers get faster and faster and more and more elaborate, we will have to make them smaller and smaller. But there is plenty of room to make them smaller. There is nothing that I can see in the physical laws that says the computer elements cannot be made enormously smaller than they are now. In fact, there may be certain advantages. - Miniaturization by evaporation - How can we make such a device? What kind of manufacturing processes would we use? One possibility we might consider, since we have talked about writing by putting atoms down in a certain arrangement, would be to evaporate the material, then evaporate the insulator next to it. Then, for the next layer, evaporate another position of a wire, another insulator, and so on. So, you simply evaporate until you have a block of stuff which has the elements- coils and condensers, transistors and so on-of exceedingly fine dimensions. But I would like to discuss, just for amusement, that there are other possibilities. Why can't we manufacture these small computers somewhat like we manufacture the big ones? Why can't we drill holes, cut things, solder things, stamp things out, mold different shapes all at an infinitesimal level? What are the limitations as to how small a thing has to be before you can no longer mold it? How many times when you are working on something frustratingly tiny like your wife's wristwatch, have you said to yourself, "If I could only train an ant to do this!" What I would like to suggest is the possibility of training an ant to train a mite to do this. What are the possibilities of small but movable machines? They may or may not be useful, but they surely would be fun to make. Consider any machine-for example, an automobile- and ask about the problems of making an infinitesimal machine like it. Suppose, in the particular design of the automobile, we need a certain precision of the parts; we need an accuracy, let's suppose, of 4/10,000 of an inch. If things are more inaccurate than that in the shape of the cylinder and so on, it isn't going to work very well. If I make the thing too small, I have to worry about the size of the

atoms; I can't make a circle of "balls" so to speak, if the circle is too small. So, if I make the error, corresponding to 4/10,000 of an inch, correspond to an error of 10 atoms, it turns out that I can reduce the dimensions of an automobile 4,000 times, approximately-so that it is 1 mm. across. Obviously, if you redesign the car so that it would work with a much larger tolerance, which is not at all impossible, then you could make a much smaller device. It is interesting to consider what the problems are in such small machines. Firstly, with parts stressed to the same degree, the forces go as the area you are reducing, so that things like weight and inertia are of relatively no importance. The strength of material, in other words, is very much greater in proportion. The stresses and expansion of the flywheel from centrifugal force, for example, would be the same proportion only if the rotational speed is increased in the same proportion as we decrease the size. On the other hand, the metals that we use have a grain structure, and this would be very annoying at small scale because the material is not homogeneous. Plastics and glass and things of this amorphous nature are very much more homogeneous, and so we would have to make our machines out of such materials. There are problems associated with the electrical part of the system-with the copper wires and the magnetic parts. The magnetic properties on a very small scale are not the same as on a large scale; there is the "domain" problem involved. A big magnet made of millions 5 of domains can only be made on a small scale with one domain. The electrical equipment won't simply be scaled down; it has to be redesigned. But I can see no reason why it can't be redesigned to work again. - Problems of lubrication - Lubrication involves some interesting points. The effective viscosity of oil would be higher and higher in proportion as we went down (and if we increase the speed as much as we can). If we don't increase the speed so much, and change from oil to kerosene or some other fluid, the problem is not so bad. But actually we may not have to lubricate at all! We have a lot of extra force. Let the bearings run dry; they won't run hot because the heat escapes away from such a small device very, very rapidly. This rapid heat loss would prevent the gasoline from exploding, so an internal combustion engine is impossible. Other chemical reactions, liberating energy when cold, can be used. Probably an external supply of electrical power would be most convenient for such small machines. What would be the utility of such machines? Who knows? Of course, a small automobile would only be useful for the mites to drive around in, and I suppose our Christian interests don't go that far. However, we did note the possibility of the manufacture of small elements for computers in completely automatic factories, containing lathes and other machine tools at the very small level. The small lathe would not have

to be exactly like our big lathe. I leave to your imagination the improvement of the design to take full advantage of the properties o

Plate 5

f things on a small scale, and in such a way that the fully automatic aspect would be easiest to manage. A friend of mine (Albert R. Hibbs) suggests a very interesting possibility for relatively small machines. He says that, although it is a very wild idea, it would be interesting in surgery if you could swallow the surgeon. You put the mechanical surgeon inside the blood vessel and it goes into the heart and "looks" around. (Of course the information has to be fed out.) It finds out which valve is the faulty one and takes a little knife and slices it out. Other small machines might be permanently incorporated in the body to assist some inadequately functioning organ. Now comes the interesting question: How do we make such a tiny mechanism? I leave that to you. However, let me suggest one weird possibility. You know, in the atomic energy plants they have materials and machines that they can't handle directly because they have become radioactive. To unscrew nuts and put on bolts and so on, they have a set of master and slave hands, so that by operating a set of levers here, you control the "hands" there, and can turn them this way and that so you can handle things quite nicely. Most of these devices are actually made rather simply, in that there is a particular cable, like a marionette string, that goes directly from the controls to the "hands." But, of course, things also have been made using servomotors, so that the connection between the one thing and the other is electrical rather than mechanical. When you turn the levers, they turn a servomotor, and it changes the electrical currents in the wires, which repositions a motor at the other end. Now, I want to build much the same device-a master slave system which operates electrically. But I want the slaves to be made especially carefully by modern large-scale machinists so that they are one-fourth the scale of the "hands" that you ordinarily maneuver. So you have a scheme by which you can do things at one- quarter scale anyway-the little servo motors with little hands play with little nuts and bolts; they drill little holes; they are four times smaller. Aha! So I manufacture a quarter-size lathe; I manufacture quarter-size tools; and I make, at the one-quarter scale, still another set of hands again relatively one-quarter size! This is one-sixteenth size, from my point of view. And after I finish doing this I wire directly from my large-scale

system, through transformers perhaps, to the one-sixteenth-size servomotors. Thus I can now manipulate the one-sixteenth size hands. Well, you get the principle from there on. It is rather a difficult program, but it is a possibility. You might say that one can go much farther in one step than from one to four. Of course, this has all to be designed very carefully and it is not necessary simply to make it like hands. If you thought of it very carefully, you could probably arrive at a much better system for doing such things. If you work through a pantograph, even today, you can get much more than a factor of four in even one step. But you can't work directly through a pantograph which makes a smaller pantograph which then makes a smaller pantograph-because of the looseness of the holes and the irregularities of construction. The end of the pantograph wiggles with a relatively greater irregularity than the irregularity with which you move your hands. In going down this scale, I would find the end of the pantograph on the end of the pantograph on the end of the pantograph shaking so badly that it wasn't doing anything sensible at all. At each stage, it is necessary to improve the precision of the apparatus. If, for instance, having made a small lathe with a pantograph, we find its lead screw irregular-more irregular than the large-scale one-we could lap the lead screw against breakable nuts that you can reverse in the usual way back and forth until this lead screw is, at its scale, as accurate as our original lead screws, at our scale. We can make flats by rubbing unflat surfaces in triplicates together-in three pairs-and the flats then become flatter than the thing you started with. Thus, it is not impossible to improve precision on a small scale by the correct operations. So, when we build this stuff, it is necessary at each step to improve the accuracy of the equipment by working for awhile down there, making accurate lead screws, Johansen blocks, and all the other materials which we use in accurate machine work at the higher level. We have to stop at each level and manufacture all the stuff to go to the next level-a very long and very difficult program. Perhaps you can figure a better way than that to get down to small scale more rapidly. Yet, after all this, you have just got one little baby lathe four thousand times smaller than usual. But we were thinking of making an enormous computer, which we were going to build by drilling holes on this lathe to make little washers for the computer. How many washers can you manufacture on this one lathe? - A hundred tiny hands - When I make my first set of slave "hands" at one fourth scale, I am going to make ten sets. I make ten sets of "hands," and I wire them to my original levers so they each do exactly the same thing at the same time in parallel. Now, when I am making my new devices one quarter again as small, I let each one manufacture ten copies, so that I would have a hundred "hands" at the 1/16th size. Where am I

going to put the million lathes that I am going to have? Why, there is nothing to it; the volume is much less than that of even one full-scale lathe. For instance, if I made a billion little lathes, each 1/4000 of the scale of a regular lathe, there are plenty of materials and space available because in the billion little ones there is less than 2 percent of the materials in one big lathe. It doesn't cost anything for materials, you see. So I want to build a billion tiny factories, models of each other, which are manufacturing simultaneously, drilling holes, stamping parts, and so on. As we go down in size, there are a number of interesting problems that arise. All things do not

Plate 6

simply scale down in proportion. There is the problem that materials stick together by the molecular (Van der Waals) attractions. It would be like this: After you have made a part and you unscrew the nut from a bolt, it isn't going to fall down because the gravity isn't appreciable; it would even be hard to get it off the bolt. It would be like those old movies of a man with his hands full of molasses, trying to get rid of a glass of water. There will be several problems of this nature that we will have to be ready to design for. - Rearranging the atoms - But I am not afraid to consider the final question as to whether, ultimately-in the great future-we can arrange the atoms the way we want; the very atoms, all the way down! What would happen if we could arrange the atoms one by one the way we want them (within reason, of course; you can't put them so that they are chemically unstable, for example). Up to now, we have been content to dig in the ground to find minerals. We heat them and we do things on a large scale with them, and we hope to get a pure substance with just so much impurity, and so on. But we must always accept some atomic arrangement that nature gives us. We haven't got anything, say, with a "checkerboard" arrangement, with the impurity atoms exactly arranged 1,000 angstroms apart, or in some other particular pattern. What could we do with layered structures with just the right layers? What would the properties of materials be if we could really arrange the atoms the way we want them? They would be very interesting to investigate theoretically. I can't see exactly what would happen, but I can hardly doubt that when we have some control of the arrangement of things on a small scale we will get an enormously greater range of possible properties that substances can have, and of different things that we can do. Consider, for example, a piece of material in which we make little

coils and condensers (or their solid state analogs) 1,000 or 10,000 angstroms in a circuit, one right next to the other, over a large area, with little antennas sticking out at the other end-a whole series of circuits. Is it possible, for example, to emit light from a whole set of antennas, like we emit radio waves from an organized set of antennas to beam the radio programs to Europe? The same thing would be to beam the light out in a definite direction with very high intensity. (Perhaps such a beam is not very useful technically or economically.) I have thought about some of the problems of building electric circuits on a small scale, and the problem of resistance is serious. If you build a corresponding circuit on a small scale, its natural frequency goes up, since the wave length goes down as the scale; but the skin depth only decreases with the square root of the scale ratio, and so resistive problems are of increasing difficulty. Possibly we can beat resistance through the use of superconductivity if the frequency is not too high, or by other tricks. - Atoms in a small world - When we get to the very, very small world-say circuits of seven atoms-we have a lot of new things that would happen that represent completely new opportunities for design. Atoms on a small scale behave like nothing on a large scale, for they satisfy the laws of quantum mechanics. So, as we go down and fiddle around with the atoms down there, we are working with different laws, and we can expect to do different things. We can manufacture in different ways. We can use, not just circuits, but some system involving the quantized energy levels, or the interactions of quantized spins, etc. Another thing we will notice is that, if we go down far enough, all of our devices can be mass produced so that they are absolutely perfect copies of one another. We cannot build two large machines so that the dimensions are exactly the same. But if your machine is only 100 atoms high, you only have to get it correct to one-half of one percent to make sure the other machine is exactly the same size-namely, 100 atoms high! At the atomic level, we have new kinds of forces and 7 new kinds of possibilities, new kinds of effects. The problems of manufacture and reproduction of materials will be quite different. I am, as I said, inspired by the biological phenomena in which chemical forces are used in repetitious fashion to produce all kinds of weird effects (one of which is the author). The principles of physics, as far as I can see, do not speak against the possibility of maneuvering things atom by atom. It is not an attempt to violate any laws; it is something, in principle, that can be done; but in practice, it has not been done because we are too big. Ultimately, we can do chemical synthesis. A chemist comes to us and says, "Look, I want a molecule that has the atoms arranged thus and so; make me that molecule." The chemist does a mysterious thing when he wants to make a molecule. He sees that it has got that ring, so he mixes

this and that, and he shakes it, and he fiddles around. And, at the end of a difficult process, he usually does succeed in synthesizing what he wants. By the time I get my devices working, so that we can do it by physics, he will have figured out how to synthesize absolutely anything, so that this will really be useless. But it is interesting that it would be, in principle, possible (I think) for a physicist to synthesize any chemical substance that the chemist writes down. Give the orders and the physicist synthesizes it. How? Put the atoms down where the chemist says, and so you make the substance. The problems of chemistry and biology can be greatly helped if our ability to see what we are doing, and to do things on an atomic level, is ultimately developed- a development which I think cannot be avoided. Now, you might say, "Who should do this and why should they do it?" Well, I pointed out a few of the economic applications, but I know that the reason that you would do it might be just for fun. But have some fun! Let's have a competition between laboratories. Let one laboratory make a tiny motor which it sends to another lab which sends it back with a thing that

Plate 7

fits inside the shaft of the first motor. - High school competition - Just for the fun of it, and in order to get kids interested in this field, I would propose that someone who has some contact with the high schools think of making some kind of high school competition. After all, we haven't even started in this field, and even the kids can write smaller than has ever been written before. They could have competition in high schools. The Los Angeles high school could send a pin to the Venice high school on which it says, "How's this?" They get the pin back, and in the dot of the "i" it says, "Not so hot." Perhaps this doesn't excite you to do it, and only economics will do so. Then I want to do something; but I can't do it at the present moment, because I haven't prepared the ground. It is my intention to offer a prize of 1,000 dollars to the first guy who can take the information on the page of a book and put it on an area 1/25,000 smaller in linear scale in such manner that it can be read by an electron microscope. And I want to offer another prize-if I can figure out how to phrase it so that I don't get into a mess of arguments about definitions-of another 1,000 dollars to the first guy who makes an operating electric motor-a rotating electric motor which

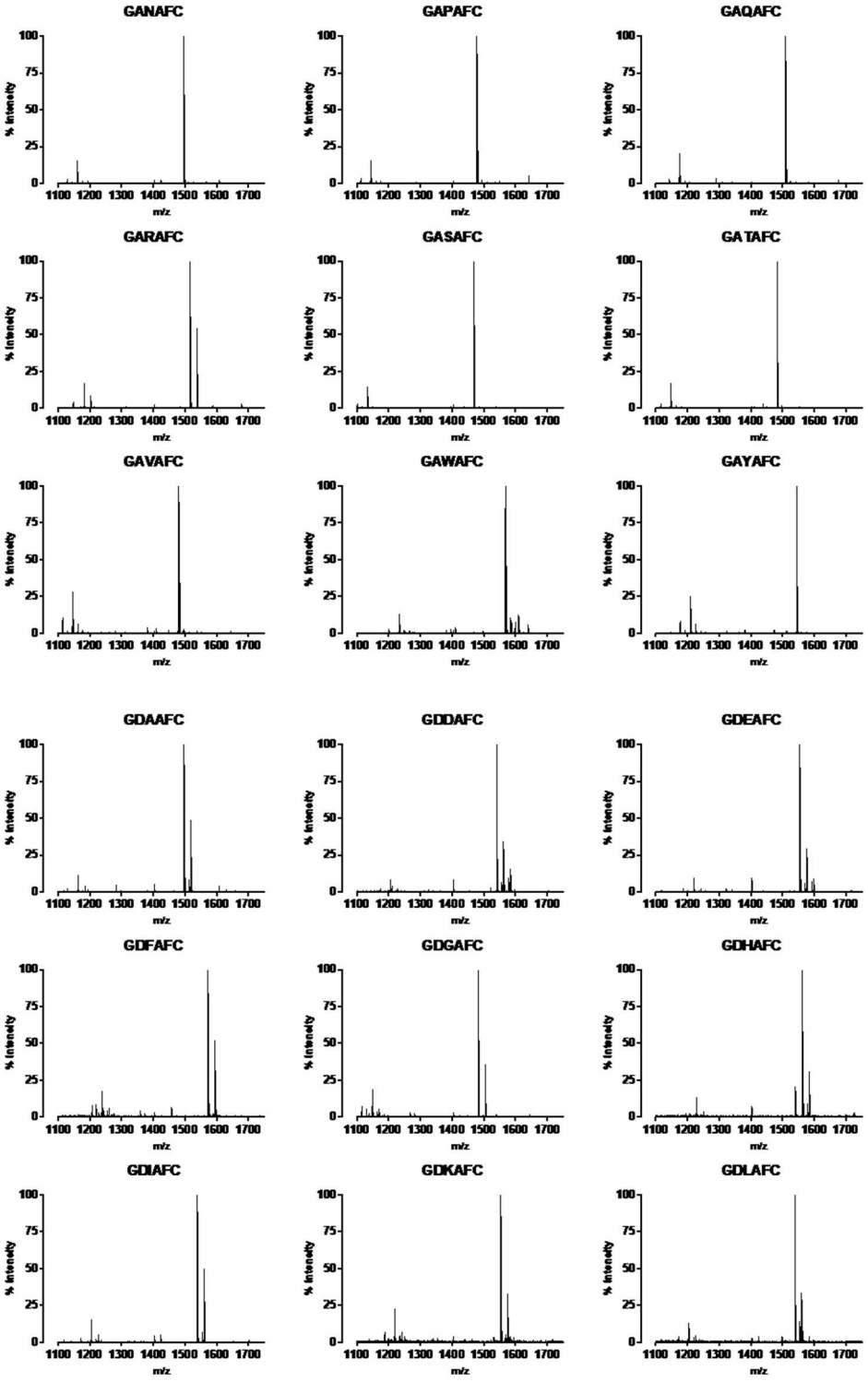
can be controlled from the outside and, not counting the lead-in wires, is only 1/64 inch cube. I do not expect that such prizes will have to wait very long for claimants.

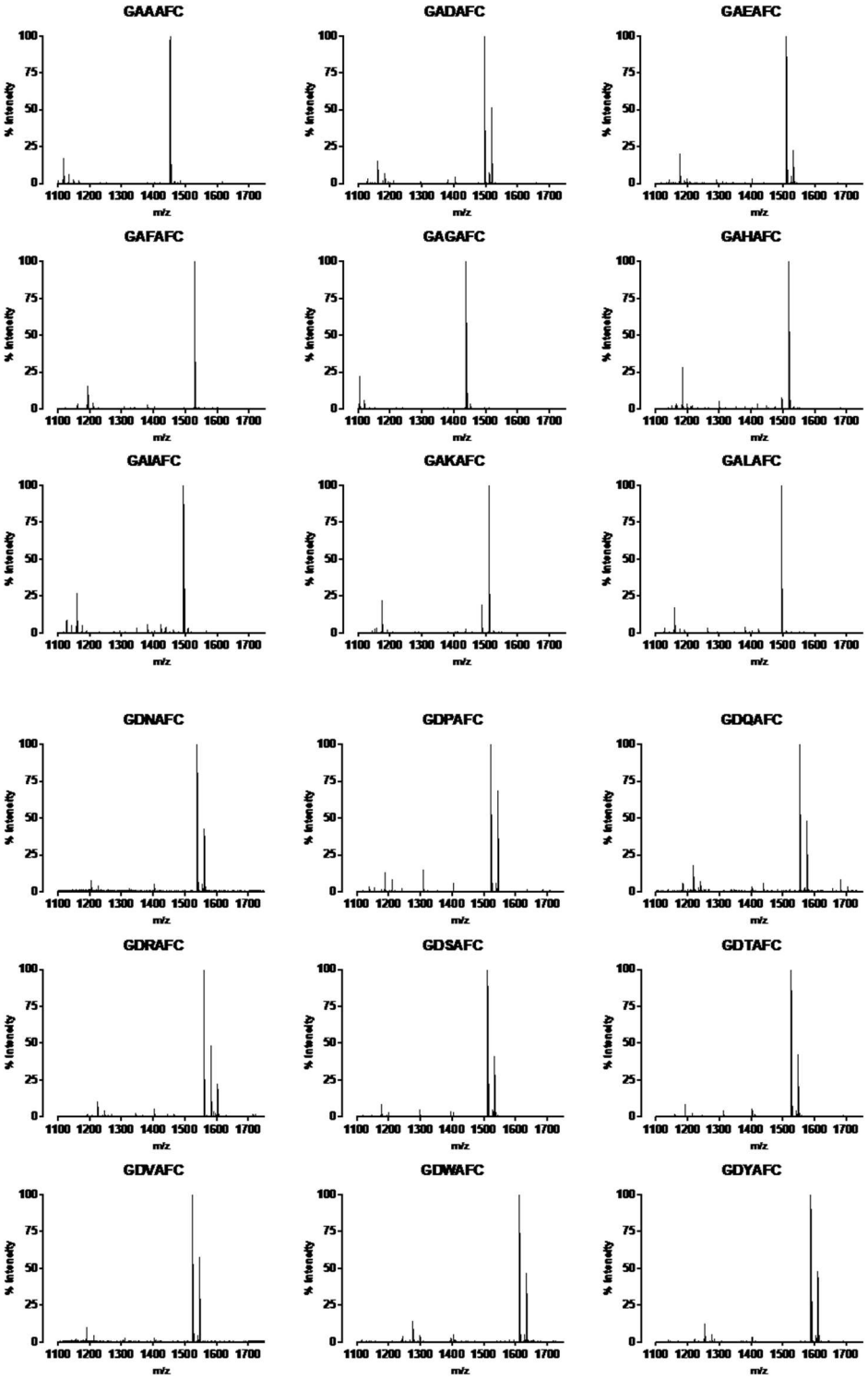
Appendix Table 3.1: List of 95 standard ASCII characters and their encoding in binary, quaternary, octal and hexadecimal.

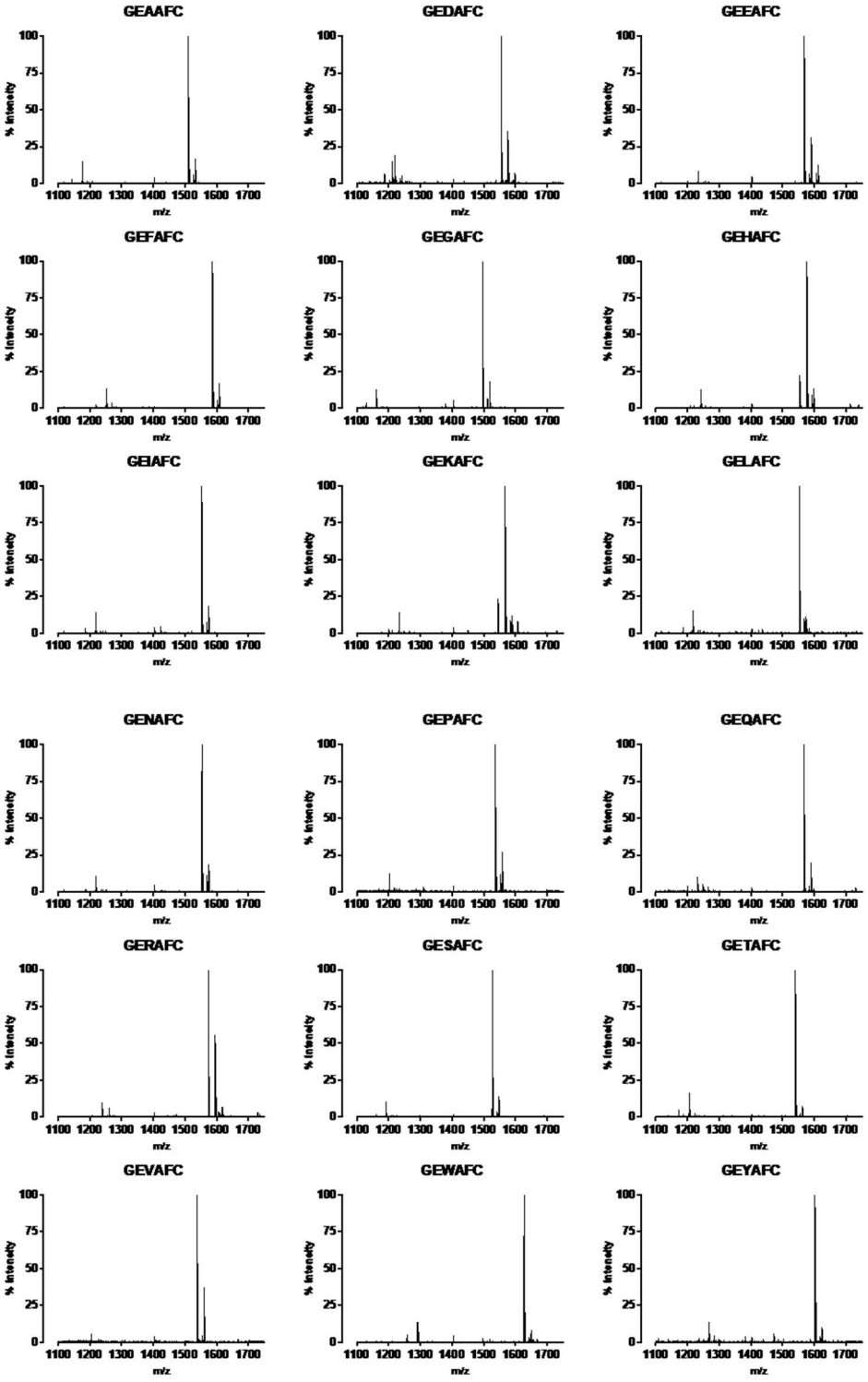
	Binary	Quaternary	Octal	Hexadecimal
(space)	00100000	0200	040	20
!	00100001	0201	041	21
"	00100010	0202	042	22
#	00100011	0203	043	23
$	00100100	0210	044	24
%	00100101	0211	045	25
&	00100110	0212	046	26
'	00100111	0213	047	27
(	00101000	0220	050	28
)	00101001	0221	051	29
*	00101010	0222	052	2a
+	00101011	0223	053	2b
+	00101011	0223	053	2b
,	00101100	0230	054	2c
-	00101101	0231	055	2d
.	00101110	0232	056	2e
/	00101111	0233	057	2f
0	00110000	0300	060	30
1	00110001	0301	061	31
2	00110010	0302	062	32
3	00110011	0303	063	33
4	00110100	0310	064	34
5	00110101	0311	065	35
6	00110110	0312	066	36
7	00110111	0313	067	37
8	00111000	0320	070	38
9	00111001	0321	071	39
:	00111010	0322	072	3a
;	00111011	0323	073	3b

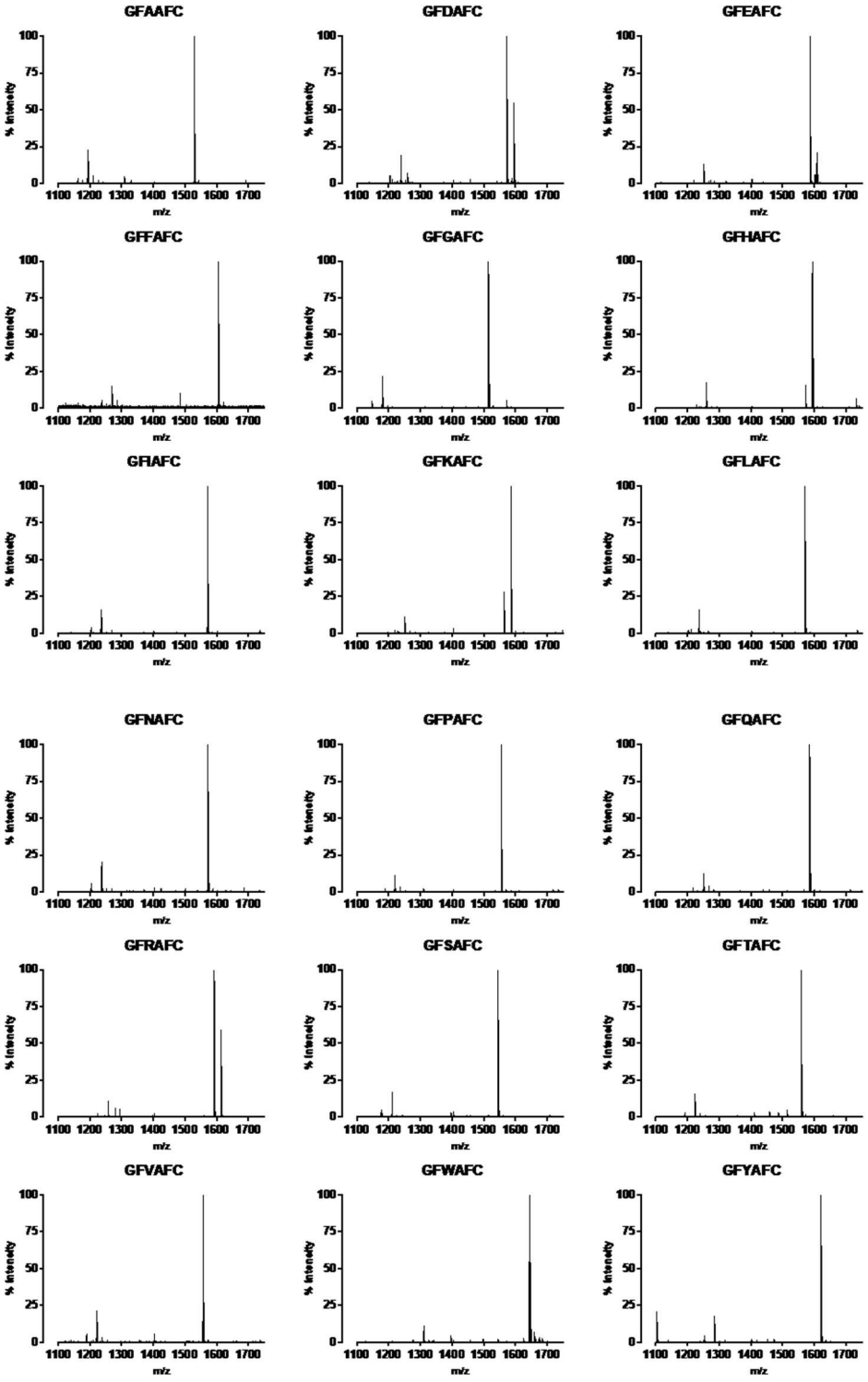
<	00111100	0330	074	3c
>	00111110	0332	076	3e
?	00111111	0333	077	3f
@	01000000	1000	100	40
A	01000001	1001	101	41
B	01000010	1002	102	42
C	01000011	1003	103	43
D	01000100	1010	104	44
E	01000101	1011	105	45
F	01000110	1012	106	46
G	01000111	1013	107	47
H	01001000	1020	110	48
I	01001001	1021	111	49
J	01001010	1022	112	4a
K	01001011	1023	113	4b
L	01001100	1030	114	4c
M	01001101	1031	115	4d
N	01001110	1032	116	4e
O	01001111	1033	117	4f
P	01010000	1100	120	50
Q	01010001	1101	121	51
R	01010010	1102	122	52
S	01010011	1103	123	53
T	01010100	1110	124	54
U	01010101	1111	125	55
V	01010110	1112	126	56
W	01010111	1113	127	57
X	01011000	1120	130	58
Y	01011001	1121	131	59
Z	01011010	1122	132	5a
[	01011011	1123	133	5b
\	01011100	1130	134	5c
]	01011101	1131	135	5d
^	01011110	1132	136	5e
(underscore) _	01011111	1133	137	5f
`	01100000	1200	140	60

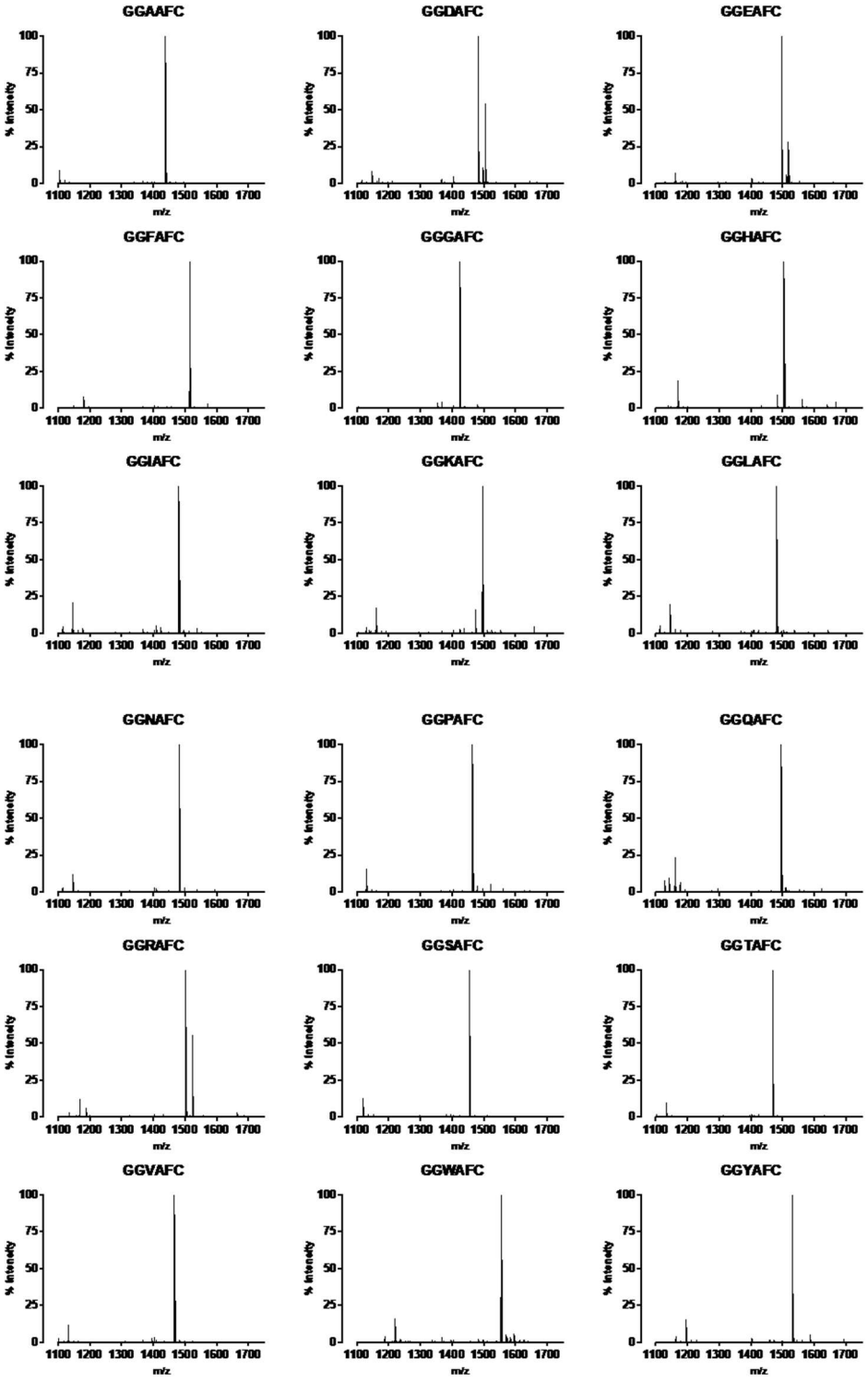
a	01100001	1201	141	61
b	01100010	1202	142	62
c	01100011	1203	143	63
d	01100100	1210	144	64
e	01100101	1211	145	65
f	01100110	1212	146	66
g	01100111	1213	147	67
h	01101000	1220	150	68
i	01101001	1221	151	69
j	01101010	1222	152	6a
k	01101011	1223	153	6b
l	01101100	1230	154	6c
m	01101101	1231	155	6d
n	01101110	1232	156	6e
o	01101111	1233	157	6f
p	01110000	1300	160	70
q	01110001	1301	161	71
r	01110010	1302	162	72
s	01110011	1303	163	73
t	01110100	1310	164	74
u	01110101	1311	165	75
v	01110110	1312	166	76
w	01110111	1313	167	77
x	01111000	1320	170	78
y	01111001	1321	171	79
z	01111010	1322	172	7a
{	01111011	1323	173	7b
|	01111100	1330	174	7c
}	01111101	1331	175	7d
~	01111110	1332	176	7e

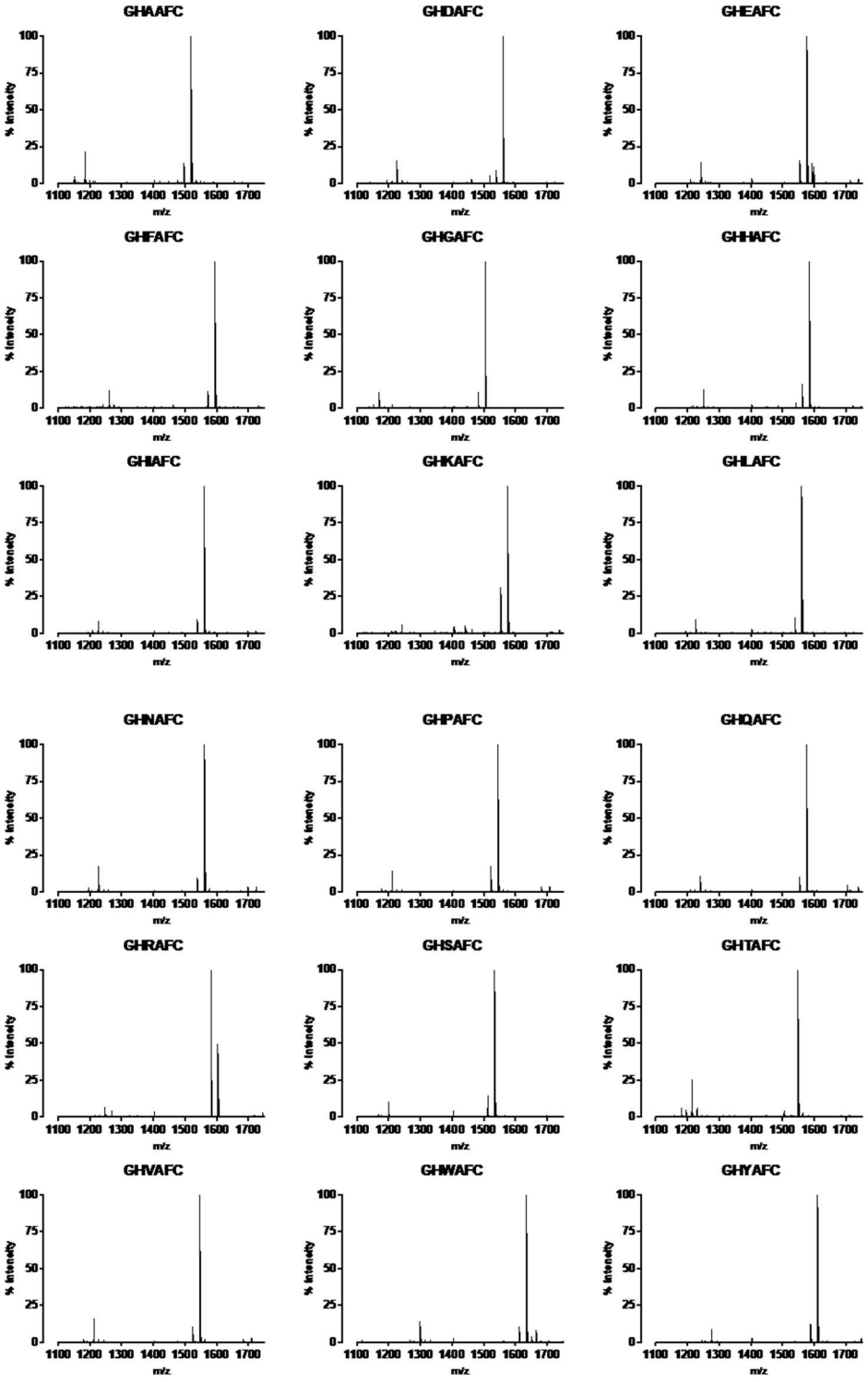

GANAFC
GAPAFC
GAQAFC
GARAFC
GASAFC
GATAFC
GAVAFC
GAWAFC
GAYAFC
GDAAFC
GDDAFC
GDEAFC
GDFAFC
GDGAFC
GDHAFC
GDIAFC
GDKAFC
GDLAFC
% Intensity
m/z
100
75
50
25
0
1100 1200 1300 1400 1500 1600 1700

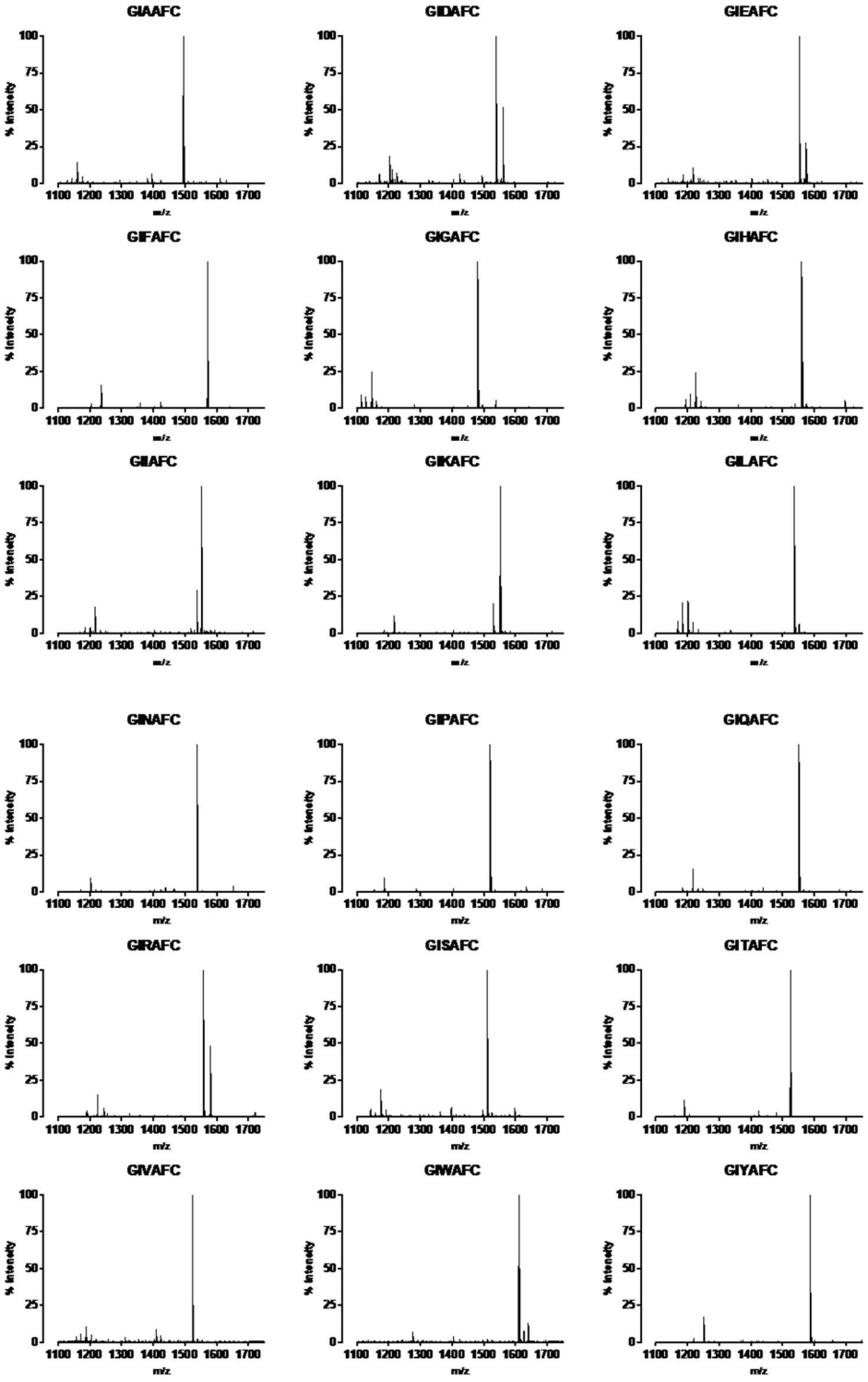

GAAAFC
GADAFC
GAEAFC
GAFAFC
GAGAFC
GAHAFC
GAIAFC
GAKAFC
GALAFC
GDNAFC
GDPAFC
GDQAFC
GDRAFC
GDSAFC
GDTAFC
GDVAFC
GDWAFC
GDYAFC
% Intensity
m/z
100
75
50
25
0
1100 1200 1300 1400 1500 1600 1700

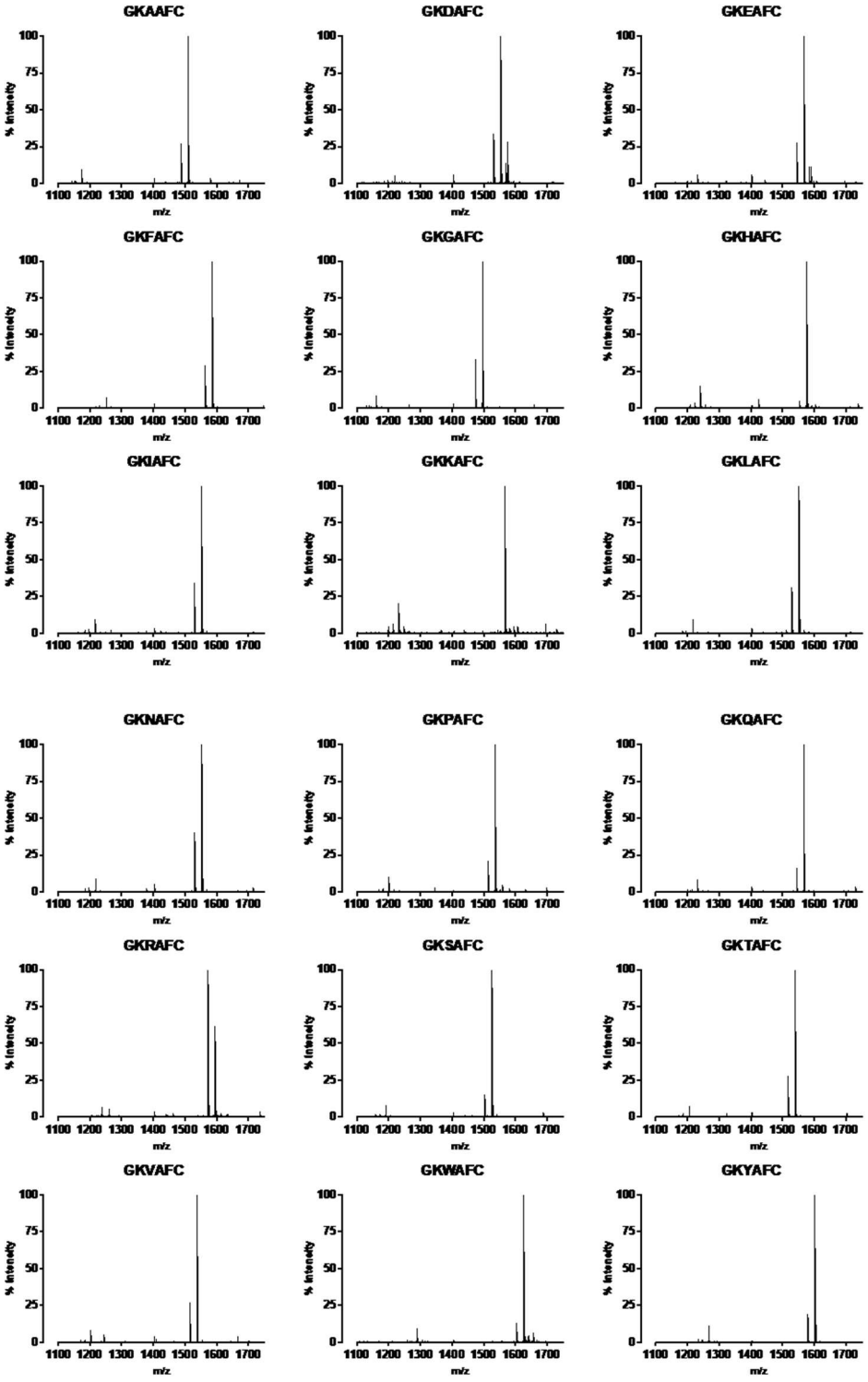

GEAAFC
GEDAFC
GEEAFC
GEFAFC
GEGAFC
GEHAFC
GEIAFC
GEKAFC
GELAFC
GENAFC
GEPAFC
GEQAFC
GERAFC
GESAFC
GETAFC
GEVAFC
GEWAFC
GEYAFC
% Intensity
m/z
100
75
50
25
0
1100 1200 1300 1400 1500 1600 1700

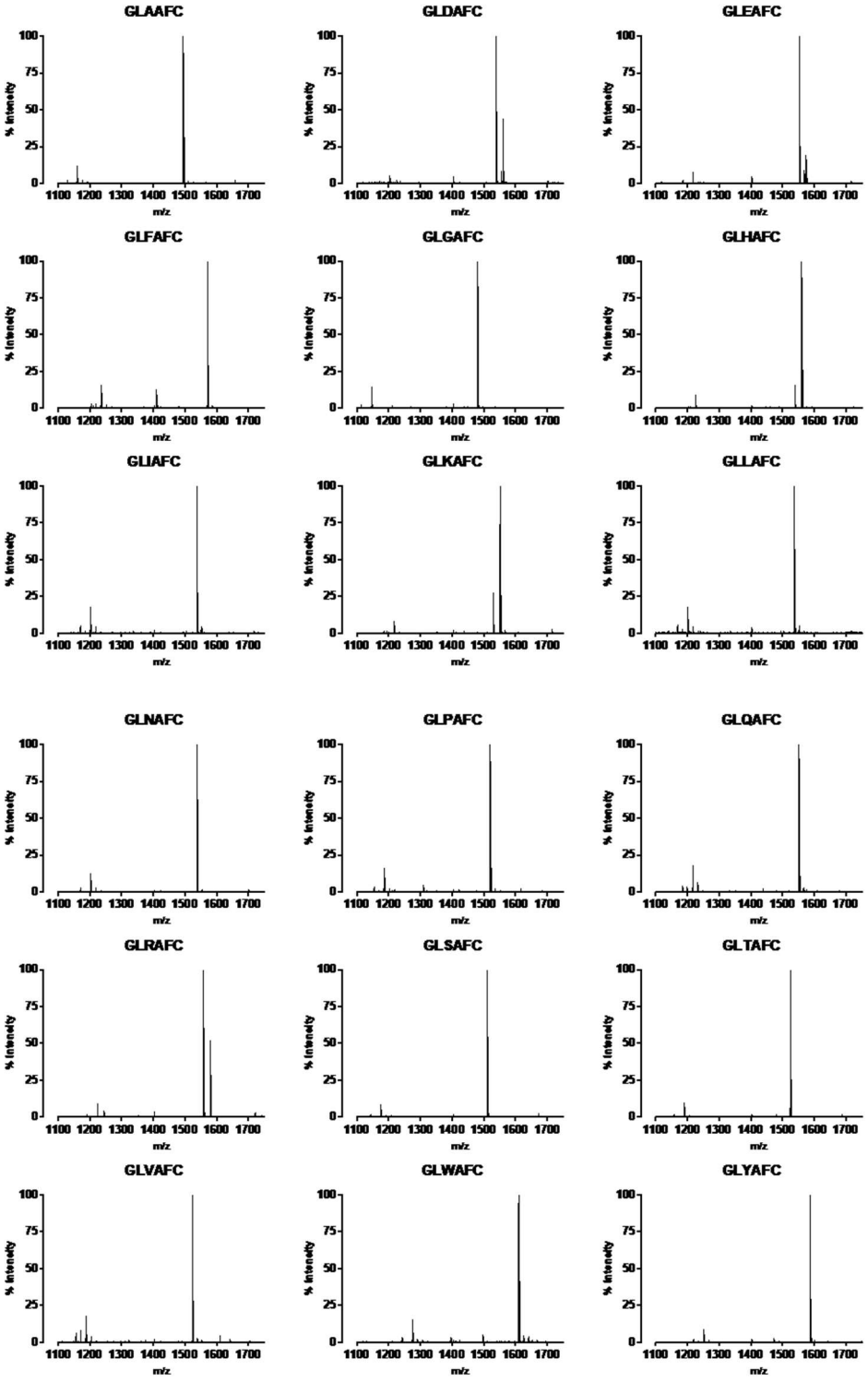

GFAAFC
GFDAFC
GFEAFC
GFFAFC
GFGAFC
GFHAFC
GFIAFC
GFKAFC
GFLAFC
GFNAFC
GFPAFC
GFQAFC
GFRAFC
GFSAFC
GFTAFC
GFVAFC
GFWAFC
GFYAFC
% Intensity
m/z
100
75
50
25
0
1100 1200 1300 1400 1500 1600 1700

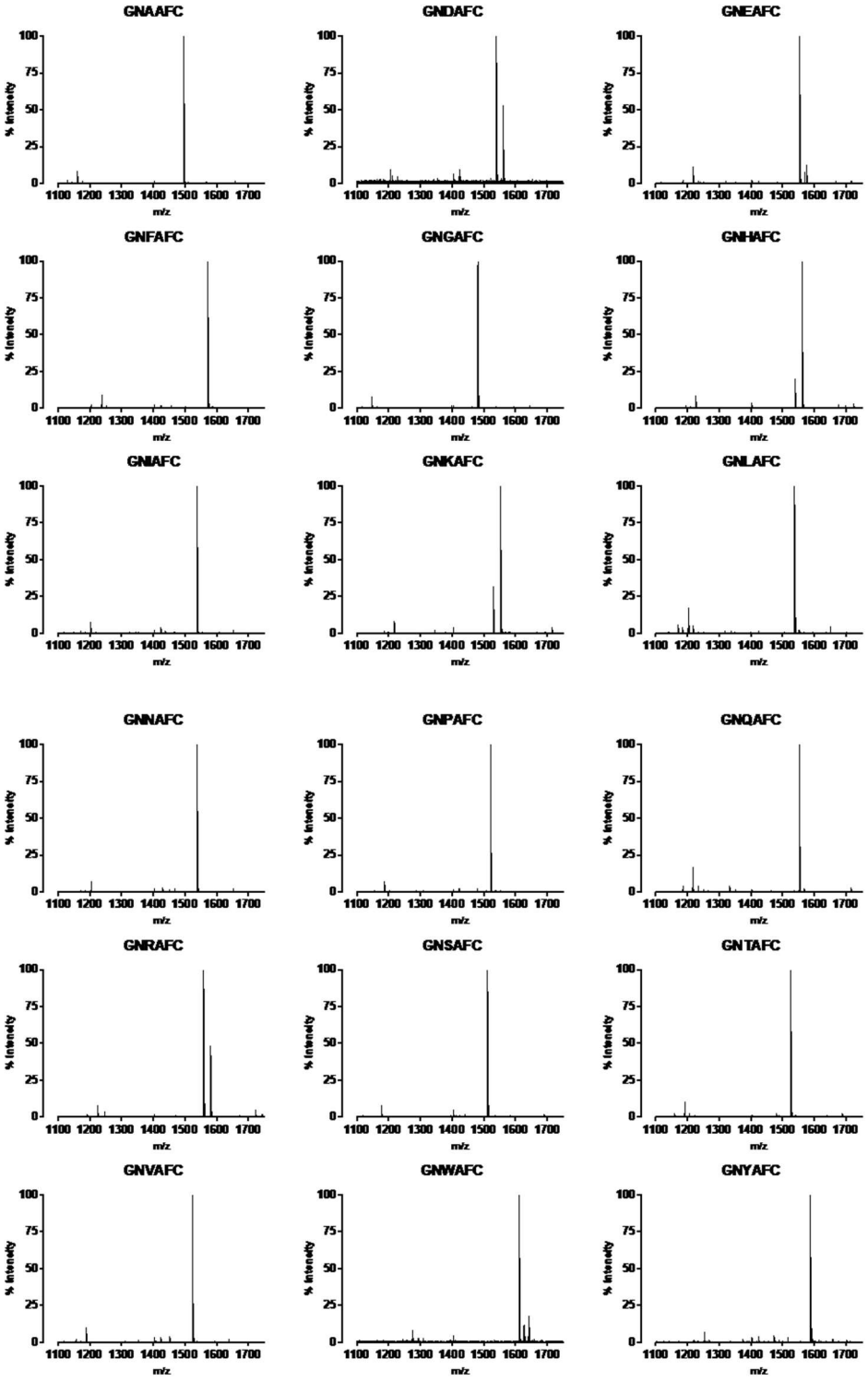

GGAAFC
GGDAFC
GGEAFC
GGFAFC
GGGAFC
GGHAFC
GGIAFC
GGKAFC
GGLAFC
GGNAFC
GGPAFC
GGQAFC
GGRAFC
GGSAFC
GGTAFC
GGVAFC
GGWAFC
GGYAFC
% Intensity
m/z
100
75
50
25
0
1100 1200 1300 1400 1500 1600 1700

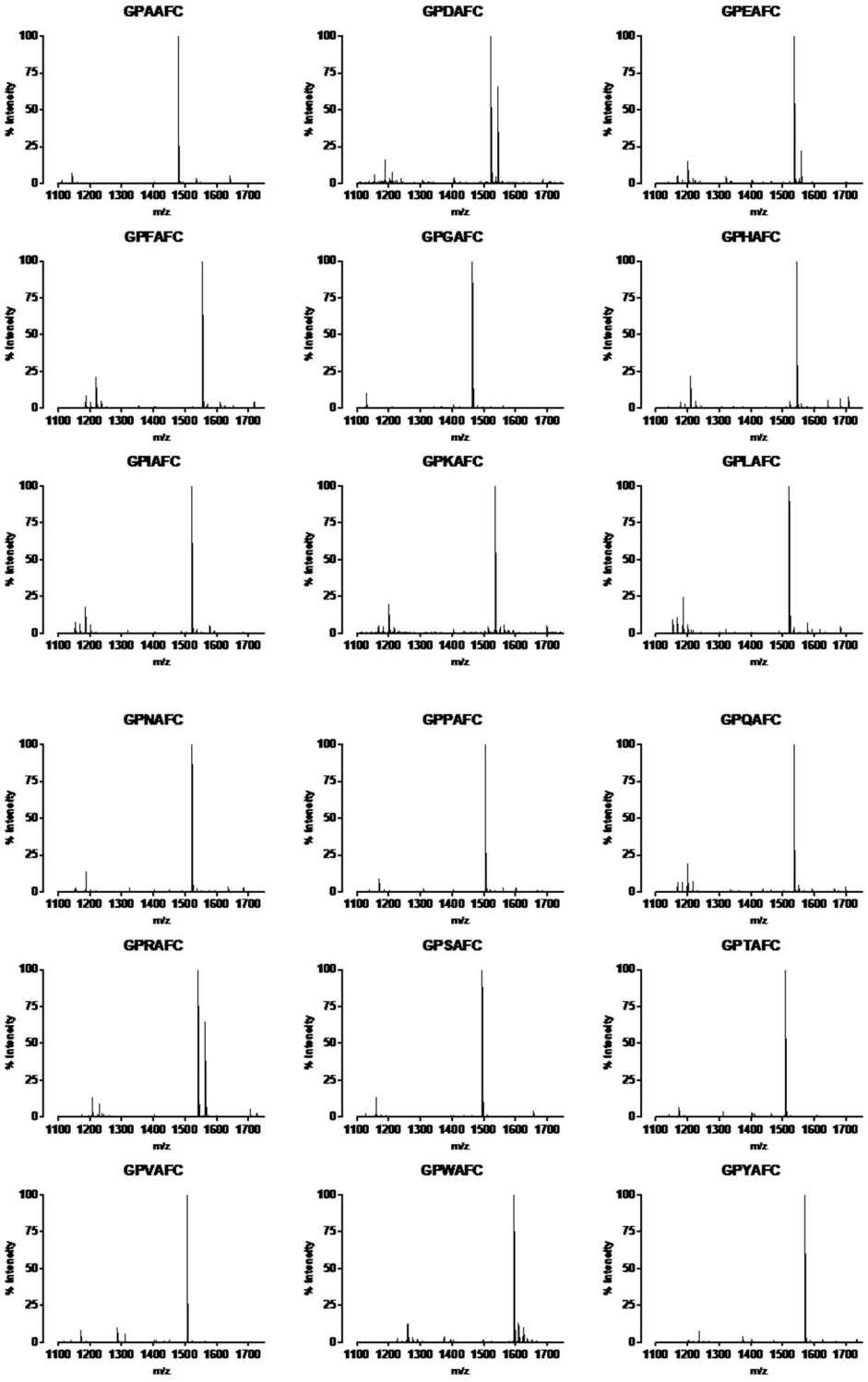

GHAAFC
GHDAFC
GHEAFC
GHFAFC
GHGAFC
GHHAFC
GHIAFC
GHKAFC
GHLAFC
GHNAFC
GHPAFC
GHQAFC
GHRAFC
GHSAFC
GHTAFC
GHVAFC
GHWAFC
GHYAFC
% Intensity
m/z
100
75
50
25
0
1100 1200 1300 1400 1500 1600 1700

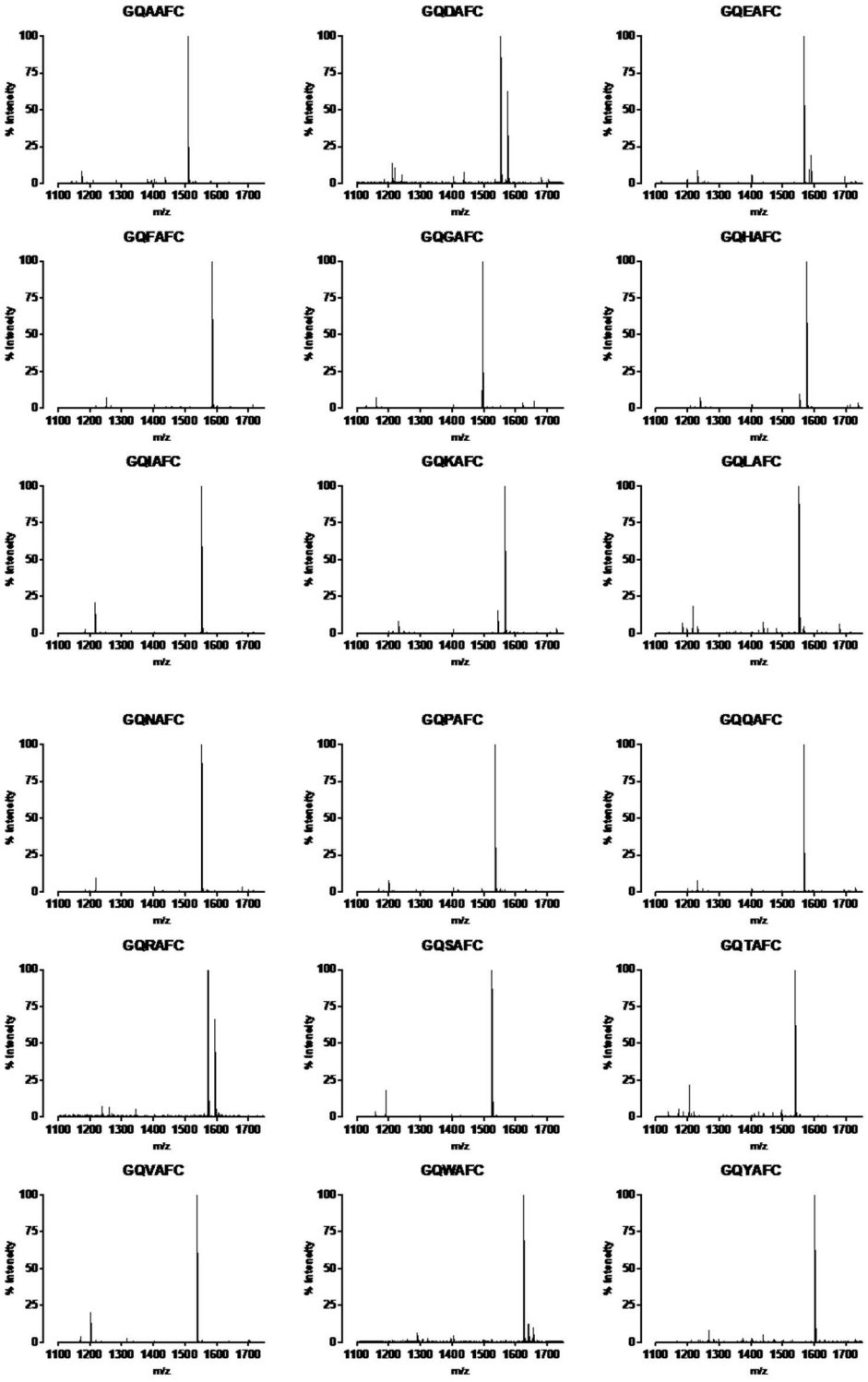

GIAAFC
GIDAFC
GIEAFC
GIFAFC
GIGAFC
GIHAFC
GIIAFC
GIKAFC
GILAFC
GINAFC
GIPAFC
GIQAFC
GIRAFC
GISAFC
GITAFC
GIVAFC
GIWAFC
GIYAFC
% Intensity
m/z
100
75
50
25
0
1100 1200 1300 1400 1500 1600 1700

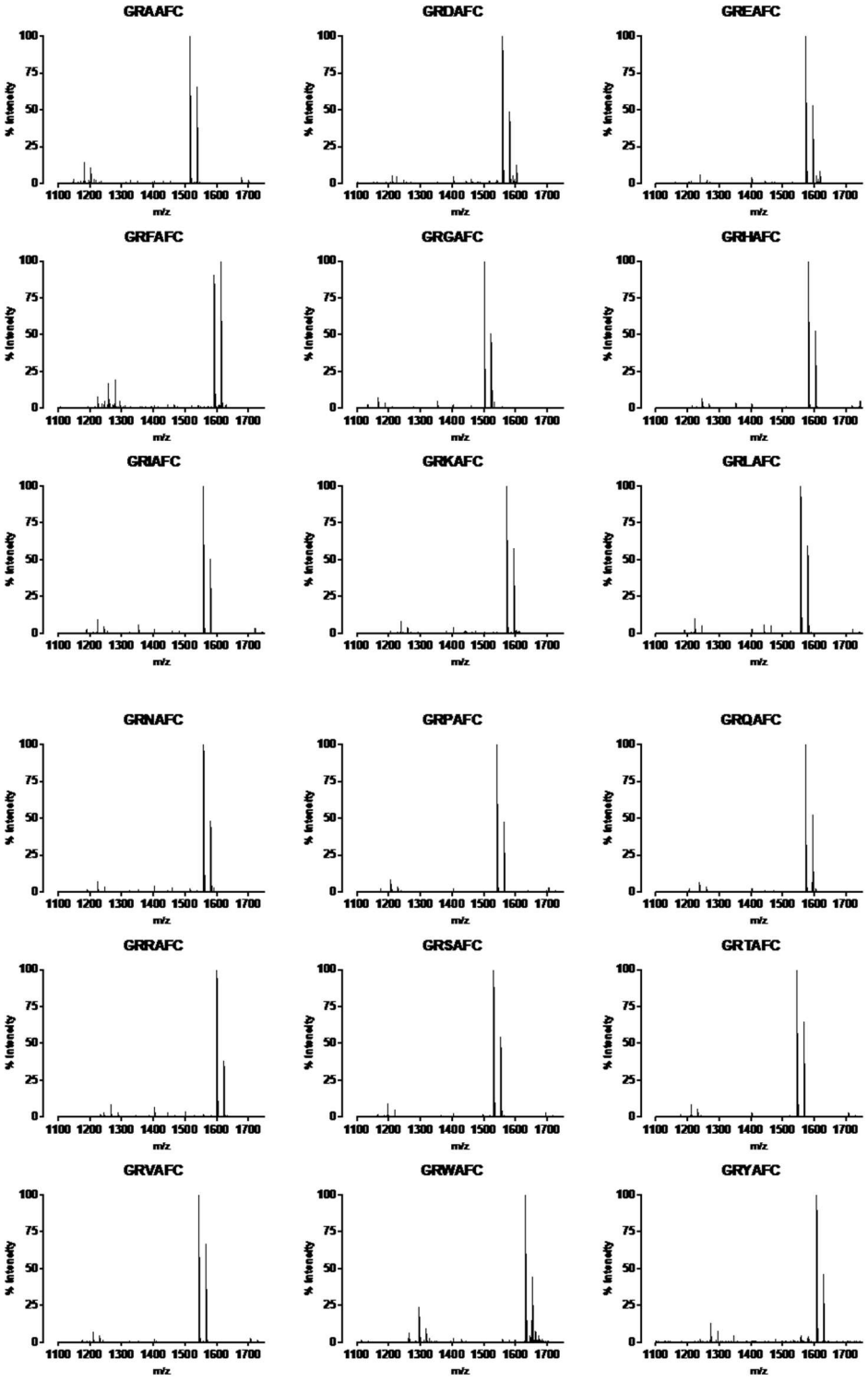

GKAAFC
GKDAFC
GKEAFC
GKFAFC
GKGAFC
GKHAFC
GKIAFC
GKKAFC
GKLAFC
GKNAFC
GKPAFC
GKQAFC
GKRAFC
GKSAFC
GKTAFC
GKVAFC
GKWAFC
GKYAFC
% Intensity
m/z
100
75
50
25
0
1100 1200 1300 1400 1500 1600 1700

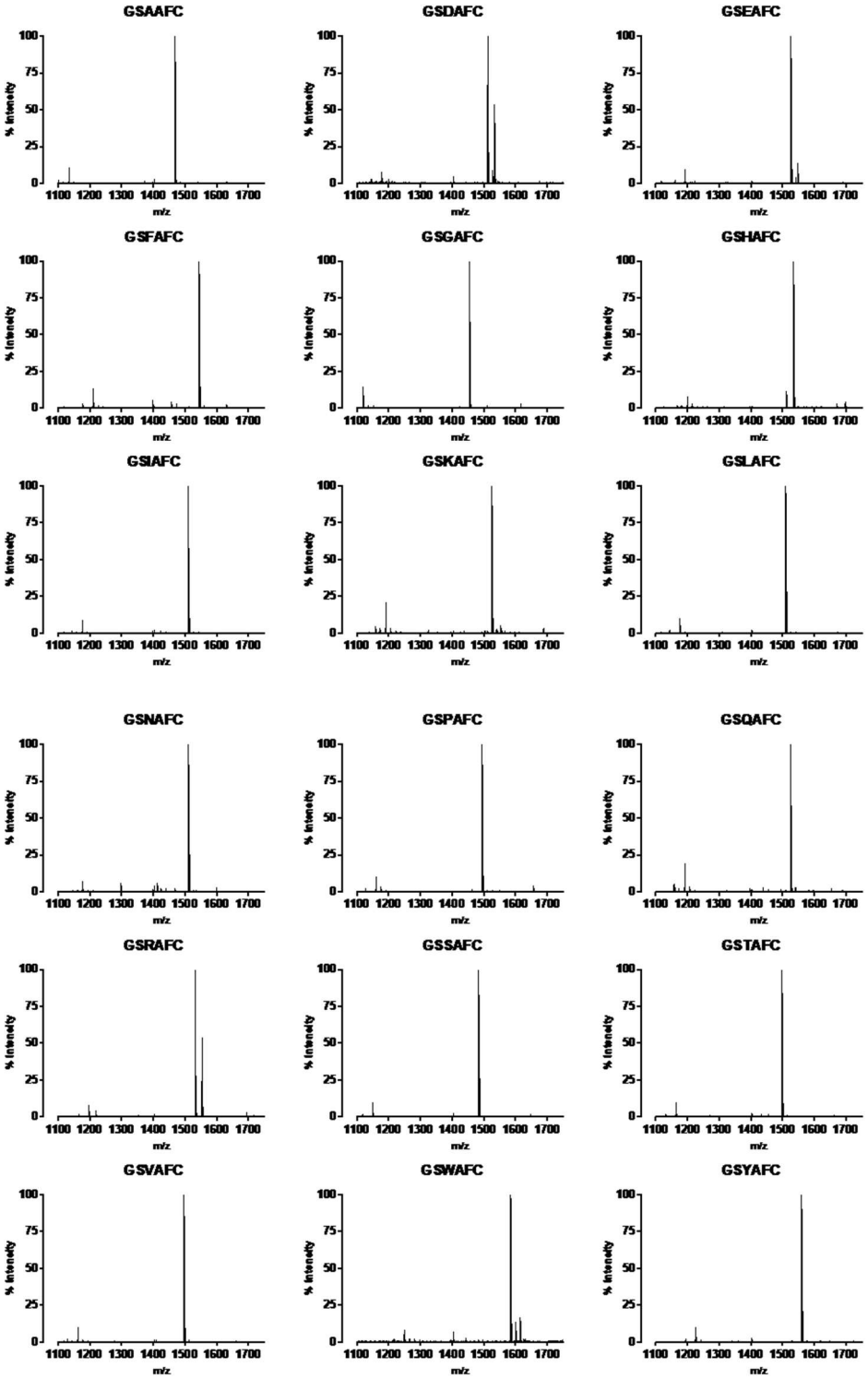

GLAAFC
GLDAFC
GLEAFC
GLFAFC
GLGAFC
GLHAFC
GLIAFC
GLKAFC
GLLAFC
GLNAFC
GLPAFC
GLQAFC
GLRAFC
GLSAFC
GLTAFC
GLVAFC
GLWAFC
GLYAFC
% Intensity
m/z
100
75
50
25
0
1100 1200 1300 1400 1500 1600 1700

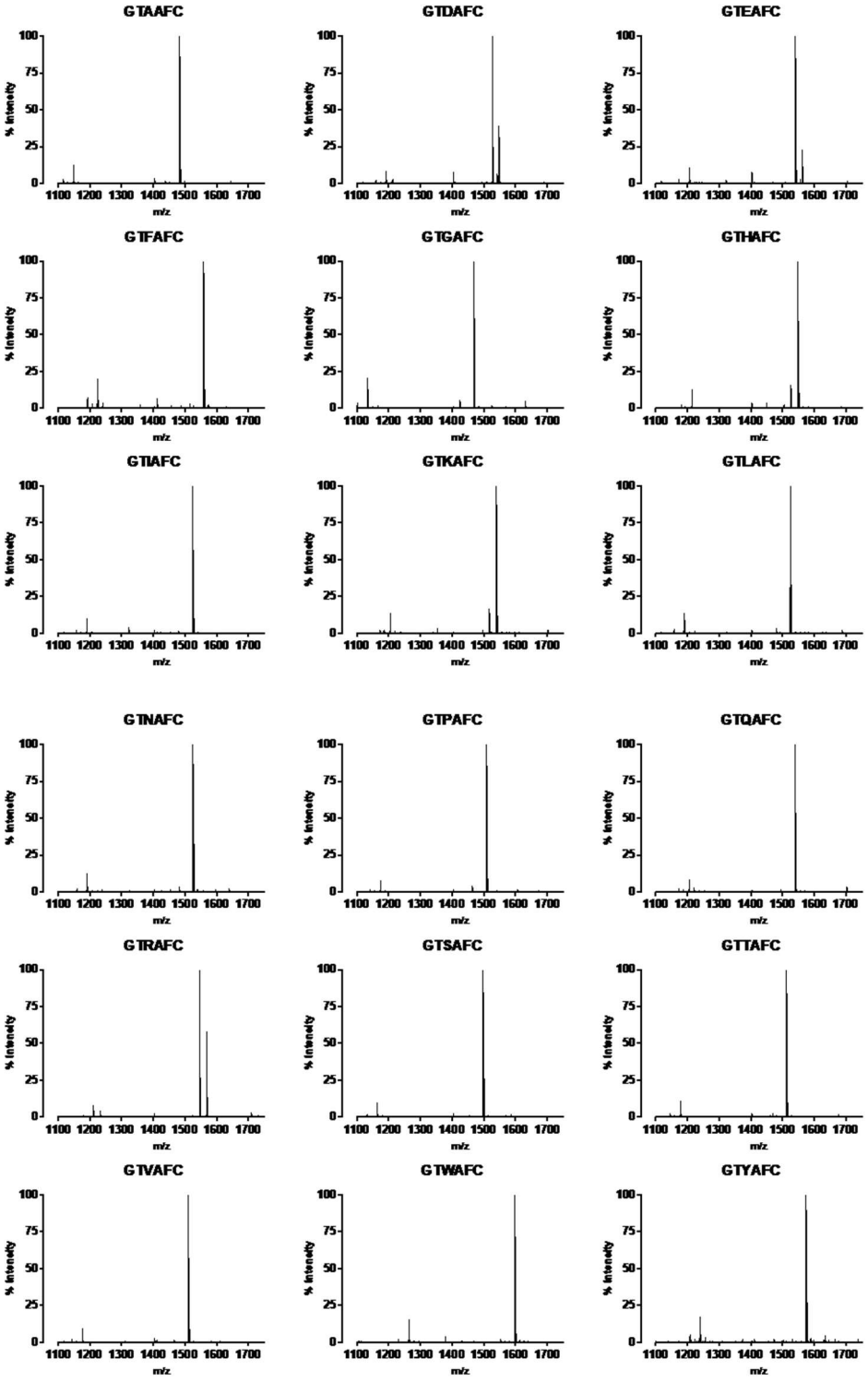

GNAAFC
GNDAFC
GNEAFC
GNFAFC
GNGAFC
GNHAFC
GNIAFC
GNKAFC
GNLAFC
GNNAFC
GNPAFC
GNQAFC
GNRAFC
GNSAFC
GNTAFC
GNVAFC
GNWAFC
GNYAFC
% Intensity
m/z
100
75
50
25
0
1100 1200 1300 1400 1500 1600 1700

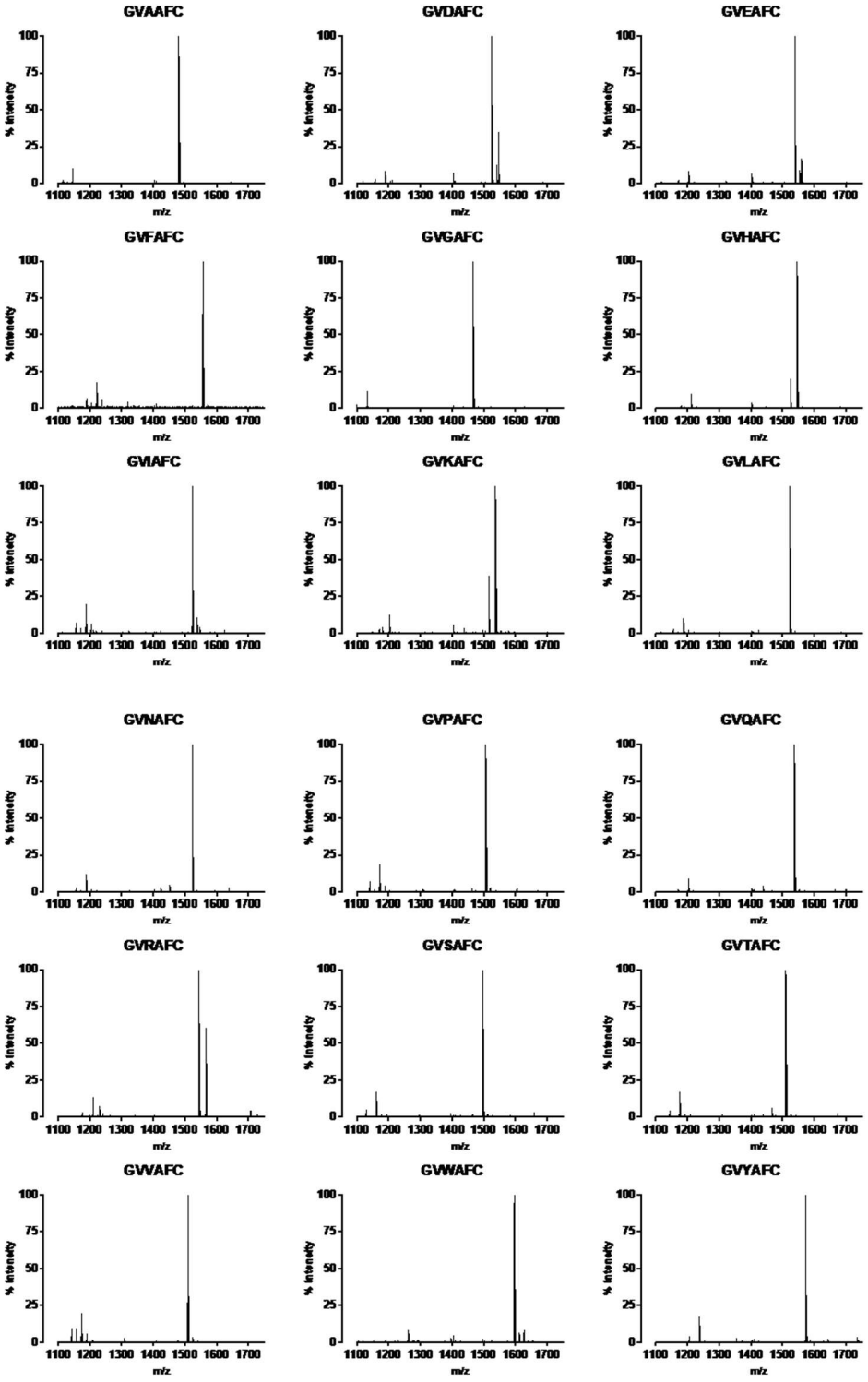

GPAAFC
GPDAFC
GPEAFC
GPFAFC
GPGAFC
GPHAFC
GPIAFC
GPKAFC
GPLAFC
GPNAFC
GPPAFC
GPQAFC
GPRAFC
GPSAFC
GPTAFC
GPVAFC
GPWAFC
GPYAFC
% Intensity
m/z
100
75
50
25
0
1100 1200 1300 1400 1500 1600 1700

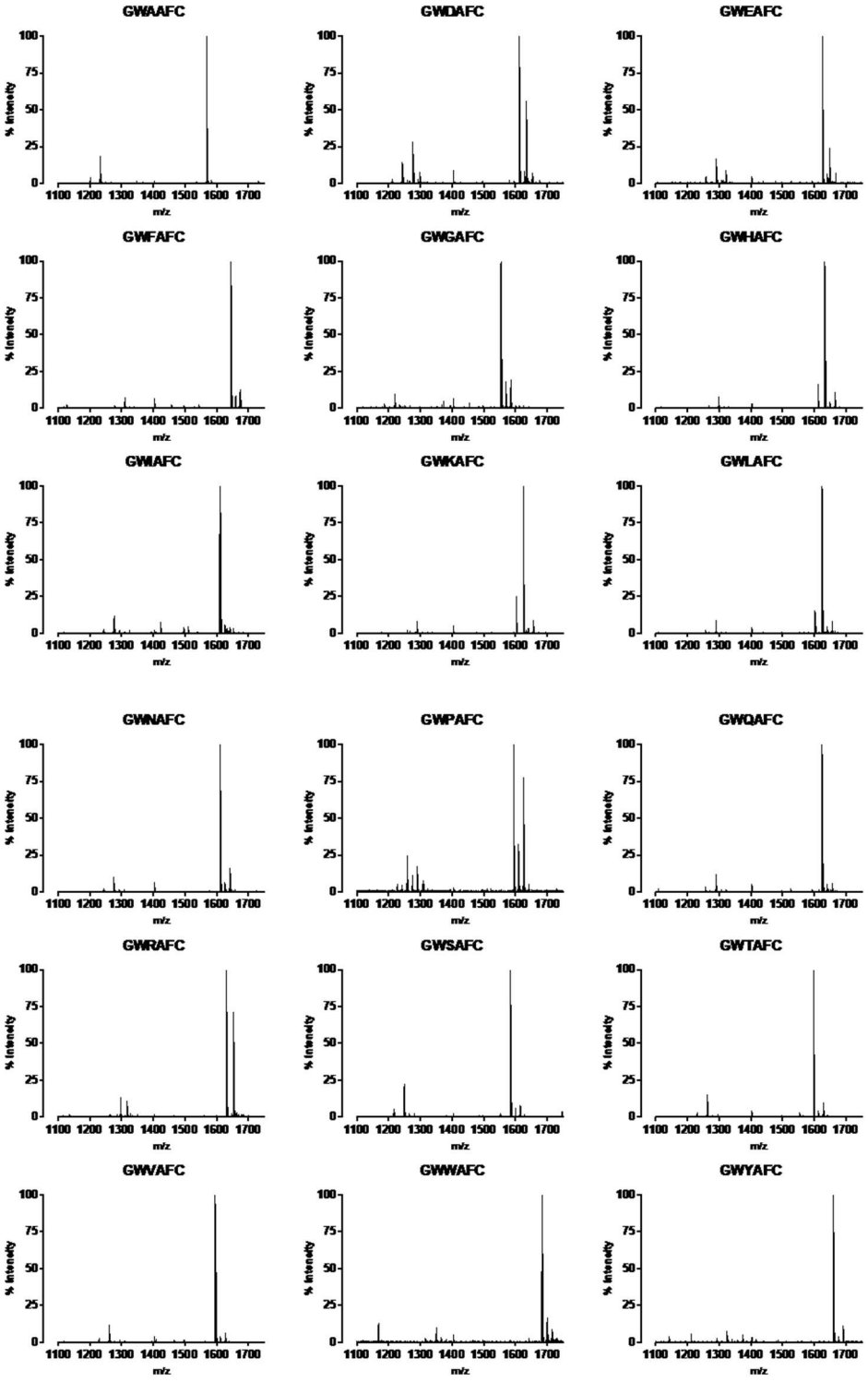

GQAAFC
GQDAFC
GQEAFC
GQFAFC
GQGAFC
GQHAFC
GQIAFC
GQKAFC
GQLAFC
GQNAFC
GQPAFC
GQQAFC
GQRAFC
GQSAFC
GQTAFC
GQVAFC
GQWAFC
GQYAFC
% Intensity
m/z
100
75
50
25
0
1100 1200 1300 1400 1500 1600 1700

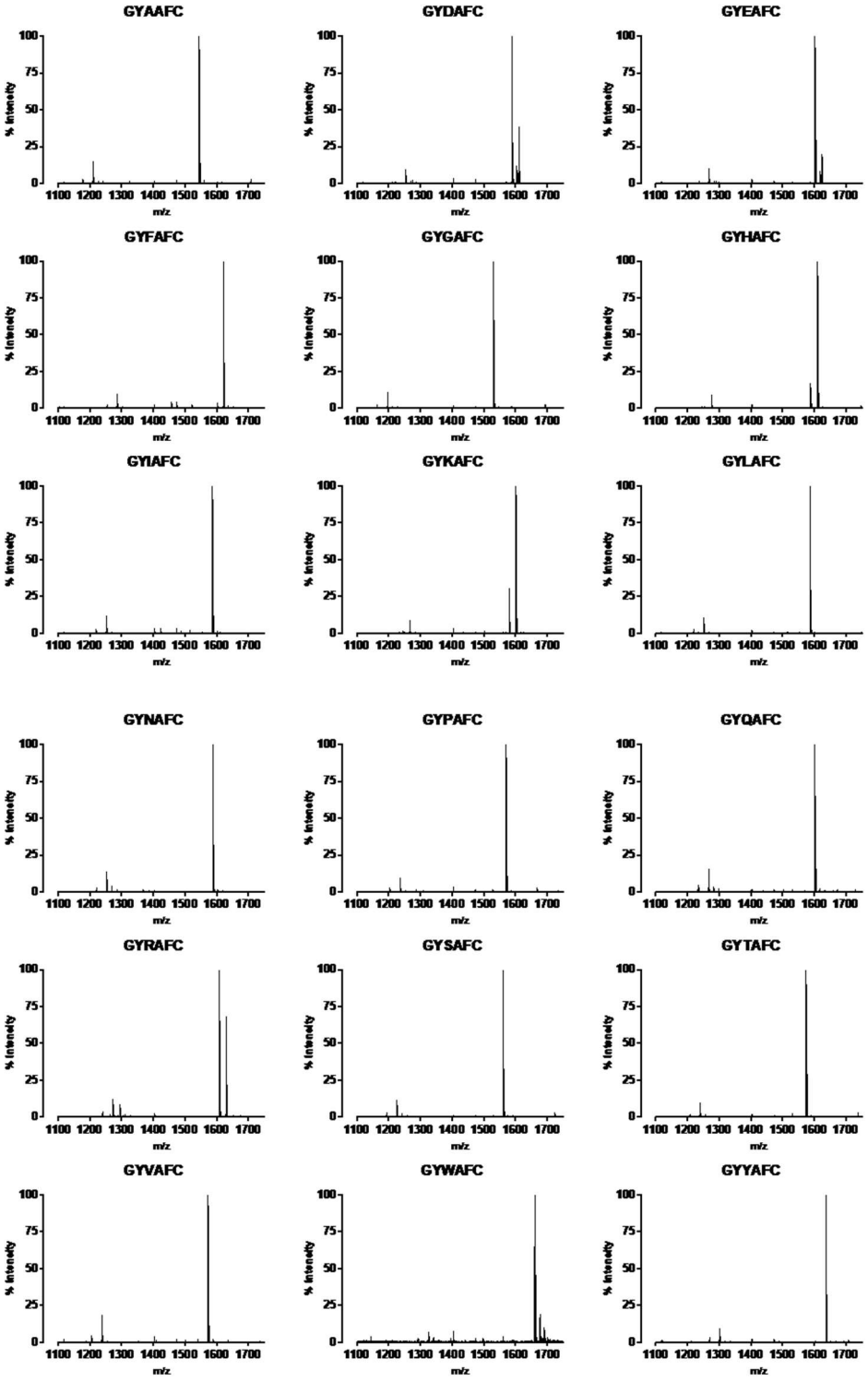

GRAAFC
GRDAFC
GREAFC
GRFAFC
GRGAFC
GRHAFC
GRIAFC
GRKAFC
GRLAFC
GRNAFC
GRPAFC
GRQAFC
GRRAFC
GRSAFC
GRTAFC
GRVAFC
GRWAFC
GRYAFC
% Intensity
m/z
100
75
50
25
0
1100 1200 1300 1400 1500 1600 1700


GSAAFC
GSDAFC
GSEAFC
GSFAFC
GSGAFC
GSHAFC
GSIAFC
GSKAFC
GSLAFC
GSNAFC
GSPAFC
GSQAFC
GSRAFC
GSSAFC
GSTAFC
GSVAFC
GSWAFC
GSYAFC
% Intensity
m/z
100
75
50
25
0
1100 1200 1300 1400 1500 1600 1700


GTAAFC
GTDAFC
GTEAFC
GTFAFC
GTGAFC
GTHAFC
GTIAFC
GTKAFC
GTLAFC
GTNAFC
GTPAFC
GTQAFC
GTRAFC
GTSAFC
GTTAFC
GTVAFC
GTWAFC
GTYAFC
% Intensity
m/z
100
75
50
25
0
1100 1200 1300 1400 1500 1600 1700


GVAAFC
GVDAFC
GVEAFC
GVFAFC
GVGAFC
GVHAFC
GVIAFC
GVKAFC
GVLAFC
GVNAFC
GVPAFC
GVQAFC
GVRAFC
GVSAFC
GVTAFC
GVVAFC
GVWAFC
GVYAFC
% Intensity
m/z
100
75
50
25
0
1100 1200 1300 1400 1500 1600 1700


GWAAFC
GWDAFC
GWEAFC
GWFAFC
GWGAFC
GWHAFC
GWIAFC
GWKAFC
GWLAFC
GWNAFC
GWPAFC
GWQAFC
GWRAFC
GWSAFC
GWTAFC
GWVAFC
GWWAFC
GWYAFC
% Intensity
m/z
100
75
50
25
0
1100 1200 1300 1400 1500 1600 1700



Appendix Figure 5.1: SAMDI mass spectra of 324-peptide library GXZAFC in alphabetical order of variable position amino acids.

Appendix Figure 5.2: Heat map of the negative controls (unreacted peptides) for the GXZAFC peptide library.

Appendix Table 5.1A: Site specific cleavage data of 324-peptide library GXZAFC in alphabetical order of variable position amino acids for detergent A.

Sequence	% AFC	% FC	% C	% Full	Sequence	% AFC	% FC	% C	% Full	Sequence	% AFC	% FC	% C	% Full
AA	36.0	25.7	12.1	26.2	FE	0.9	1.5	84.9	12.7	IG	2.7	0.7	38.1	58.5
AD	43.4	5.5	22.0	29.1	FF	3.9	2.4	87.3	6.4	IH	1.2	0.7	74.6	23.5
AE	62.8	1.8	23.8	11.6	FG	0.9	1.9	50.4	46.8	II	59.7	3.6	27.2	9.5
AF	53.4	0.4	30.4	15.8	FH	0.2	3.4	67.7	28.7	IK	63.9	0.7	21.6	13.8
AG	0.4	3.4	16.3	79.9	FI	1.9	12.1	78.6	7.4	IL	83.8	1.2	4.2	10.8
AH	2.1	1.3	55.0	41.6	FK	0.5	0.3	80.8	18.4	IN	69.3	3.0	13.8	13.9
AI	1.4	53.5	32.4	12.7	FL	3.2	2.5	89.0	5.3	IP	1.6	3.5	2.0	92.9
AK	55.0	5.2	23.7	16.1	FN	2.9	2.1	64.7	30.3	IQ	78.4	0.8	14.6	6.2
AL	86.8	1.1	3.7	8.4	FP	2.5	16.4	2.6	78.5	IR	49.0	0.0	33.5	17.5
AN	31.0	0.8	7.8	60.4	FQ	5.2	1.0	81.1	12.7	IS	57.8	1.6	34.3	6.3
AP	0.3	2.9	2.0	94.8	FR	0.0	0.1	76.2	23.7	IT	64.4	4.2	27.4	4.0
AQ	28.7	2.0	32.7	36.6	FS	2.2	1.8	84.9	11.1	IV	16.0	11.8	49.1	23.1
AR	29.4	0.8	31.8	38.0	FT	2.9	3.6	82.9	10.6	IW	31.4	2.0	57.6	9.0
AS	28.1	1.9	31.3	38.7	FV	0.0	6.3	87.7	6.0	IY	78.8	1.4	13.4	6.4
AT	42.6	2.4	18.4	36.6	FW	4.6	2.7	85.8	6.9	KA	17.8	46.0	1.8	34.4
AV	10.3	22.6	22.9	44.2	FY	3.9	1.3	79.6	15.2	KD	8.7	4.4	2.0	84.9
AW	31.4	3.1	55.2	10.3	GA	2.0	9.9	7.2	80.9	KE	33.7	10.7	2.8	52.8
AY	80.6	0.6	13.4	5.4	GD	18.3	0.0	23.4	58.3	KF	39.8	1.8	2.9	55.5
DA	36.6	12.8	6.0	44.6	GE	13.0	4.6	19.7	62.7	KG	0.2	8.6	1.1	90.1
DD	23.7	1.7	5.3	69.3	GF	5.3	2.0	24.3	68.4	KH	4.9	8.8	6.4	79.9
DE	59.4	1.8	7.3	31.5	GG	0.6	2.2	7.1	90.1	KI	0.2	69.6	1.4	28.8
DF	29.3	2.2	9.4	59.1	GH	0.2	1.2	28.0	70.6	KK	51.3	1.6	3.7	43.4
DG	0.0	2.5	7.4	90.1	GI	2.0	29.2	9.0	59.8	KL	63.1	4.2	1.1	31.6
DH	1.8	0.0	5.8	92.4	GK	8.8	2.9	16.4	71.9	KN	41.1	1.8	2.0	55.1
DI	6.4	50.5	8.7	34.4	GL	18.3	10.0	6.8	64.9	KP	0.9	16.8	3.2	79.1
DK	36.5	0.0	6.1	57.4	GN	1.2	2.1	7.2	89.5	KQ	39.5	1.3	3.2	56.0
DL	67.5	6.0	3.0	23.5	GP	0.1	2.4	3.1	94.4	KR	18.8	1.6	2.9	76.7
DN	12.0	4.5	2.9	80.6	GQ	2.5	1.4	8.2	87.9	KS	26.4	10.5	2.2	60.9
DP	8.9	1.4	3.8	85.9	GR	4.3	0.1	20.9	74.7	KT	34.5	29.5	3.1	32.9
DQ	45.2	1.0	8.3	45.5	GS	0.8	1.3	14.9	83.0	KV	2.1	69.8	3.0	25.1
DR	6.0	0.2	3.4	90.4	GT	2.0	3.5	10.5	84.0	KW	29.7	5.5	17.5	47.3
DS	26.4	1.9	18.8	52.9	GV	1.3	21.1	8.8	68.8	KY	68.4	1.3	2.4	27.9
DT	34.1	3.6	12.2	50.1	GW	5.8	2.6	63.9	27.7	LA	22.8	2.7	40.4	34.1
DV	15.8	20.5	16.6	47.1	GY	26.1	0.9	24.8	48.2	LD	14.1	1.4	54.3	30.2
DW	21.2	5.4	58.3	15.1	HA	2.2	58.8	2.0	37.0	LE	26.8	0.9	58.2	14.1
DY	68.3	1.0	6.6	24.1	HD	21.1	26.5	2.1	50.3	LF	18.4	1.1	52.7	27.8
EA	56.5	17.1	3.5	22.9	HE	4.2	16.3	1.0	78.5	LG	1.2	1.6	39.4	57.8
ED	61.7	2.8	4.1	31.4	HF	11.0	6.0	2.2	80.8	LH	0.7	0.8	64.1	34.4
EE	80.7	2.9	3.2	13.2	HG	0.9	7.2	0.9	91.0	LI	0.9	20.4	60.9	17.8
EF	76.7	0.8	3.6	18.9	HH	1.9	6.4	13.2	78.5	LK	21.4	0.0	58.8	19.8
EG	2.4	3.4	8.7	85.5	HI	0.4	70.7	1.4	27.5	LL	81.9	6.3	7.6	4.2
EH	2.9	0.2	9.0	87.9	HK	18.9	7.7	3.3	70.1	LN	8.6	0.8	48.2	42.4
EI	1.8	55.9	17.2	25.1	HL	23.5	27.0	1.2	48.3	LP	1.2	4.9	2.5	91.4
EK	62.2	0.5	2.2	35.1	HN	4.7	2.1	0.5	92.7	LQ	14.6	0.6	56.2	28.6
EL	84.6	2.0	4.4	9.0	HP	7.6	9.9	0.0	82.5	LR	6.8	0.2	55.7	37.3
EN	57.5	1.5	5.0	36.0	HQ	7.0	3.5	0.9	88.6	LS	4.7	0.3	70.4	24.6
EP	0.3	9.0	4.2	86.5	HR	5.2	5.2	3.3	86.3	LT	10.9	3.0	66.3	19.8
EQ	76.3	0.7	2.4	20.6	HS	2.6	24.0	1.6	71.8	LV	2.8	9.0	69.2	19.0
ER	30.5	0.0	5.0	64.5	HT	3.5	47.0	2.9	46.6	LW	0.7	0.5	88.9	9.9
ES	59.0	2.3	6.6	32.1	HV	5.7	64.9	2.3	27.1	LY	50.2	1.1	35.5	13.2
ET	63.3	5.3	4.9	26.5	HW	11.9	5.0	27.4	55.7	NA	5.2	9.9	3.9	81.0
EV	11.6	52.5	6.9	29.0	HY	27.2	3.1	2.2	67.5	ND	11.0	3.6	8.0	77.4

EW	64.0	2.9	23.1	10.0	IA	80.1	2.6	8.5	8.8	NE	22.8	6.0	6.6	64.6
EY	81.8	1.5	5.5	11.2	ID	69.5	1.3	11.1	18.1	NF	13.2	5.5	9.8	71.5
FA	2.1	9.3	64.8	23.8	IE	84.8	0.0	10.0	5.2	NG	0.2	1.7	3.6	94.5
FD	1.5	2.1	69.7	26.7	IF	73.3	1.8	18.8	6.1	NH	0.8	0.1	2.9	96.2
NI	1.3	37.5	3.5	57.7	RG	3.8	5.5	0.4	90.3	VE	72.4	1.3	20.5	5.8
NK	37.2	0.0	2.7	60.1	RH	0.4	6.4	1.5	91.7	VF	59.9	1.1	22.9	16.1
NL	27.8	2.2	2.2	67.8	RI	5.4	52.4	0.9	41.3	VG	1.1	1.4	28.2	69.3
NN	4.3	2.2	5.3	88.2	RK	27.7	2.5	0.9	68.9	VH	4.7	0.3	66.5	28.5
NP	3.6	3.1	1.5	91.8	RL	39.2	4.3	0.0	56.5	VI	9.9	15.0	46.1	29.0
NQ	4.6	1.6	4.0	89.8	RN	15.9	0.6	1.0	82.5	VK	61.1	0.4	26.1	12.4
NR	15.6	0.1	3.0	81.3	RP	0.9	17.3	0.0	81.8	VL	82.1	1.9	9.2	6.8
NS	4.9	3.2	5.6	86.3	RQ	17.8	2.9	1.3	78.0	VN	51.0	4.2	17.3	27.5
NT	8.2	4.6	3.9	83.3	RR	36.6	13.1	0.0	50.3	VP	2.4	8.9	2.1	86.6
NV	10.7	22.6	5.0	61.7	RS	6.9	9.8	0.3	83.0	VQ	62.7	0.5	26.2	10.6
NW	6.4	2.5	36.1	55.0	RT	18.8	26.2	0.3	54.7	VR	31.7	0.4	36.6	31.3
NY	50.7	0.6	4.5	44.2	RV	1.9	58.5	0.7	38.9	VS	36.2	0.3	44.5	19.0
PA	7.9	0.9	10.2	81.0	RW	23.7	3.5	18.1	54.7	VT	53.6	0.6	33.6	12.2
PD	24.5	0.4	7.8	67.3	RY	47.3	1.0	1.4	50.3	VV	6.1	11.9	48.7	33.3
PE	30.2	0.8	30.3	38.7	SA	13.6	34.8	3.4	48.2	VW	16.1	1.5	66.5	15.9
PF	31.7	0.8	21.3	46.2	SD	23.6	10.0	6.6	59.8	VY	79.7	1.2	16.1	3.0
PG	0.8	1.9	7.8	89.5	SE	60.0	10.2	5.4	24.4	WA	7.1	55.5	7.8	29.6
PH	2.1	0.6	59.2	38.1	SF	42.9	1.2	5.3	50.6	WD	18.8	27.2	13.5	40.5
PI	1.6	1.2	19.6	77.6	SG	0.3	4.2	3.3	92.2	WE	3.9	47.4	23.7	25.0
PK	16.9	0.1	40.2	42.8	SH	1.1	0.7	14.4	83.8	WF	5.3	5.9	37.0	51.8
PL	55.0	1.2	4.6	39.2	SI	2.0	52.2	10.7	35.1	WG	0.0	9.7	30.4	59.9
PN	6.4	0.6	5.2	87.8	SK	48.0	1.5	3.6	46.9	WH	0.0	0.7	63.1	36.2
PP	1.3	2.1	1.6	95.0	SL	65.2	4.0	3.7	27.1	WI	26.2	48.8	14.9	10.1
PQ	8.5	1.0	19.5	71.0	SN	7.8	4.1	8.5	79.6	WK	5.2	7.7	40.9	46.2
PR	5.4	0.3	46.2	48.1	SP	0.0	2.8	2.4	94.8	WL	5.0	10.6	47.9	36.5
PS	3.3	0.9	17.0	78.8	SQ	23.4	1.7	4.1	70.8	WN	2.3	3.6	27.0	67.1
PT	13.0	6.7	20.7	59.6	SR	31.7	1.0	5.8	61.5	WP	0.0	49.4	3.6	47.0
PV	2.1	7.3	25.3	65.3	SS	16.7	7.7	5.3	70.3	WQ	6.2	10.0	29.6	54.2
PW	6.2	1.0	62.6	30.2	ST	22.9	11.2	2.5	63.4	WR	0.0	3.2	41.0	55.8
PY	64.3	0.8	11.1	23.8	SV	2.1	64.0	3.4	30.5	WS	4.0	24.9	30.7	40.4
QA	14.1	31.7	5.5	48.7	SW	34.7	1.3	33.0	31.0	WT	3.6	67.9	10.7	17.8
QD	19.6	8.5	4.5	67.4	SY	76.0	0.3	3.0	20.7	WV	0.5	69.6	8.7	21.2
QE	48.4	11.8	3.6	36.2	TA	52.0	11.5	2.0	34.5	WW	0.0	3.7	73.3	23.0
QF	30.7	1.3	7.0	61.0	TD	66.3	2.0	2.1	29.6	WY	28.6	5.3	48.4	17.7
QG	0.0	1.0	1.7	97.3	TE	78.3	3.4	3.7	14.6	YA	6.9	38.5	16.8	37.8
QH	0.2	0.0	4.7	95.1	TF	79.6	0.9	4.2	15.3	YD	5.3	10.3	33.4	51.0
QI	0.8	58.4	1.7	39.1	TG	0.7	2.8	4.1	92.4	YE	11.8	13.2	52.2	22.8
QK	40.5	0.7	0.6	58.2	TH	3.2	0.5	16.6	79.7	YF	4.8	2.3	50.0	42.9
QL	51.1	4.2	4.2	40.5	TI	5.6	55.1	7.0	32.3	YG	2.7	2.6	22.0	72.7
QN	2.8	1.2	4.0	92.0	TK	83.1	0.2	0.8	15.9	YH	0.0	0.6	55.2	44.2
QP	0.6	3.7	4.2	91.5	TL	91.7	0.9	1.5	5.9	YI	0.3	58.1	13.5	28.1
QQ	9.6	1.4	3.2	85.8	TN	41.9	0.8	2.2	55.1	YK	10.2	0.8	60.6	28.4
QR	15.2	0.9	5.2	78.7	TP	3.3	4.6	2.7	89.4	YL	20.1	7.9	48.3	23.7
QS	7.1	3.3	4.3	85.3	TQ	59.0	0.9	3.4	36.7	YN	5.7	1.4	28.7	64.2
QT	17.8	14.3	7.8	60.1	TR	67.3	0.5	2.1	30.1	YP	0.0	23.9	1.4	74.7
QV	1.0	35.2	1.6	62.2	TS	24.1	2.9	21.8	51.2	YQ	7.5	4.4	41.5	46.6
QW	27.6	4.8	23.3	44.3	TT	30.9	3.5	17.4	48.2	YR	1.6	0.6	45.1	52.7
QY	49.3	2.7	6.6	41.4	TV	9.4	50.1	3.9	36.6	YS	2.5	3.6	56.5	37.4
RA	5.4	55.1	0.8	38.7	TW	62.8	1.9	19.1	16.2	YT	4.6	20.6	42.1	32.7
RD	4.6	3.0	0.5	91.9	TY	88.7	1.5	4.9	4.9	YV	0.2	55.6	19.9	24.3
RE	5.9	10.2	0.8	83.1	VA	72.9	4.1	13.4	9.6	YW	0.9	2.2	85.2	11.7
RF	22.9	2.1	1.4	73.6	VD	56.5	2.0	18.7	22.8	YY	19.8	1.4	51.0	27.8

Appendix Table 5.1B: Site specific cleavage data of 324-peptide library GXZAFC in alphabetical order of variable position amino acids for detergent B.

Sequence	% AFC	% FC	% C	% Full	Sequence	% AFC	% FC	% C	% Full	Sequence	% AFC	% FC	% C	% Full
AA	33.7	19.2	6.9	40.2	FE	0.3	0.6	88.2	10.9	IG	2.6	0.7	30.1	66.6
AD	55.5	1.8	16.9	25.8	FF	2.1	0.0	60.5	37.4	IH	1.3	0.5	67.0	31.2
AE	61.9	1.1	14.2	22.8	FG	0.6	1.0	44.9	53.5	II	59.7	3.4	22.3	14.6
AF	47.7	0.4	20.2	31.7	FH	0.0	2.1	68.7	29.2	IK	71.9	0.0	13.4	14.7
AG	0.5	1.5	14.4	83.6	FI	0.0	15.9	79.5	4.6	IL	81.8	0.3	2.0	15.9
AH	1.3	1.3	45.6	51.8	FK	0.3	0.1	75.6	24.0	IN	63.2	0.4	6.6	29.8
AI	1.5	24.2	14.5	59.8	FL	2.6	1.4	89.8	6.2	IP	0.0	0.7	0.0	99.3
AK	57.5	1.2	15.9	25.4	FN	1.3	0.7	51.8	46.2	IQ	69.1	0.2	10.0	20.7
AL	78.6	0.8	1.7	18.9	FP	1.8	12.2	1.0	85.0	IR	46.0	0.3	21.2	32.5
AN	19.3	0.8	6.7	73.2	FQ	3.3	0.9	61.8	34.0	IS	57.5	1.0	26.3	15.2
AP	0.3	0.9	1.2	97.6	FR	0.0	0.0	72.0	28.0	IT	63.6	2.5	19.8	14.1
AQ	31.6	2.7	10.9	54.8	FS	2.7	1.2	72.3	23.8	IV	13.1	8.8	36.9	41.2
AR	30.8	0.1	26.9	42.2	FT	2.2	2.5	73.2	22.1	IW	46.2	1.6	39.6	12.6
AS	21.2	1.1	22.4	55.3	FV	3.1	4.6	89.5	2.8	IY	79.8	0.4	10.3	9.5
AT	31.2	1.6	11.8	55.4	FW	4.0	0.6	84.4	11.0	KA	14.7	35.4	0.6	49.3
AV	7.1	15.4	15.5	62.0	FY	3.5	0.8	67.5	28.2	KD	7.3	3.0	0.8	88.9
AW	38.4	1.2	43.7	16.7	GA	1.2	5.4	4.0	89.4	KE	25.5	7.6	1.9	65.0
AY	70.9	0.4	10.0	18.7	GD	13.6	4.5	8.4	73.5	KF	32.1	0.6	1.8	65.5
DA	37.9	11.4	6.1	44.6	GE	17.0	3.7	23.2	56.1	KG	0.0	7.4	0.3	92.3
DD	26.3	2.2	6.6	64.9	GF	3.1	0.5	12.1	84.3	KH	2.6	3.1	3.0	91.3
DE	56.9	1.3	6.3	35.5	GG	0.3	0.8	4.6	94.3	KI	0.1	57.0	1.3	41.6
DF	11.3	1.7	5.9	81.1	GH	0.0	0.3	18.8	80.9	KK	38.2	1.6	1.0	59.2
DG	0.1	2.6	5.6	91.7	GI	0.6	19.7	5.1	74.6	KL	55.2	3.3	0.4	41.1
DH	1.2	0.0	4.6	94.2	GK	7.7	1.2	11.1	80.0	KN	34.4	1.1	0.4	64.1
DI	22.8	29.9	9.9	37.4	GL	14.9	7.1	3.5	74.5	KP	0.2	10.7	1.2	87.9
DK	18.3	0.0	1.1	80.6	GN	0.9	0.9	3.6	94.6	KQ	29.8	0.5	0.2	69.5
DL	74.8	0.5	1.6	23.1	GP	0.0	1.1	1.7	97.2	KR	17.8	1.1	1.0	80.1
DN	17.5	1.0	6.2	75.3	GQ	1.4	0.7	4.9	93.0	KS	29.3	9.6	0.3	60.8
DP	4.3	2.7	4.5	88.5	GR	5.1	0.0	18.6	76.3	KT	28.9	21.8	0.4	48.9
DQ	43.3	1.0	7.6	48.1	GS	0.7	1.5	10.3	87.5	KV	1.8	58.5	1.1	38.6
DR	2.2	0.0	0.8	97.0	GT	1.8	2.2	7.1	88.9	KW	14.5	0.0	4.3	81.2
DS	28.0	1.2	15.2	55.6	GV	0.7	13.6	5.3	80.4	KY	51.3	0.4	1.2	47.1
DT	35.4	3.4	11.5	49.7	GW	9.0	0.4	46.3	44.3	LA	19.5	2.9	29.5	48.1
DV	14.3	19.3	14.0	52.4	GY	22.4	0.5	18.5	58.6	LD	17.0	0.6	58.0	24.4
DW	17.9	2.6	51.6	27.9	HA	1.7	44.7	0.9	52.7	LE	27.8	0.7	53.2	18.3
DY	76.1	0.3	4.2	19.4	HD	17.4	22.6	1.5	58.5	LF	14.6	0.6	39.9	44.9
EA	55.7	15.4	3.2	25.7	HE	3.1	10.5	0.4	86.0	LG	0.5	0.7	33.8	65.0
ED	61.3	1.5	2.7	34.5	HF	8.2	4.4	0.7	86.7	LH	0.5	0.6	54.0	44.9
EE	82.4	2.2	1.9	13.5	HG	0.0	3.7	0.0	96.3	LI	1.3	16.8	53.3	28.6
EF	76.1	0.2	2.4	21.3	HH	1.0	3.5	7.6	87.9	LK	21.9	0.0	46.8	31.3
EG	1.9	1.8	5.3	91.0	HI	0.1	59.9	0.6	39.4	LL	64.1	2.9	6.0	27.0
EH	1.1	0.0	5.1	93.8	HK	14.3	5.3	1.9	78.5	LN	6.7	0.4	34.3	58.6
EI	0.8	57.5	12.4	29.3	HL	15.6	19.6	0.7	64.1	LP	1.4	2.9	2.6	93.1
EK	49.1	0.0	1.0	49.9	HN	1.5	0.5	0.4	97.6	LQ	12.2	0.5	47.6	39.7
EL	82.1	1.1	2.6	14.2	HP	7.8	6.4	0.3	85.5	LR	7.7	0.2	43.4	48.7
EN	65.0	0.5	2.3	32.2	HQ	6.7	3.5	0.5	89.3	LS	4.0	0.5	58.8	36.7
EP	0.2	9.7	2.0	88.1	HR	2.6	2.9	1.4	93.1	LT	12.1	2.5	53.8	31.6
EQ	70.8	0.4	2.5	26.3	HS	1.5	19.4	0.5	78.6	LV	3.0	7.0	65.2	24.8
ER	27.5	0.1	3.0	69.4	HT	2.4	35.2	1.6	60.8	LW	1.2	0.4	75.6	22.8
ES	63.0	1.6	3.9	31.5	HV	7.6	52.0	0.7	39.7	LY	47.3	0.5	29.1	23.1
ET	63.0	4.5	2.4	30.1	HW	4.0	3.7	10.3	82.0	NA	2.7	5.0	1.4	90.9
EV	9.8	61.4	3.9	24.9	HY	19.9	1.7	0.6	77.8	ND	12.3	0.0	1.7	86.0

EW	70.2	1.4	16.7	11.7	IA	73.2	1.8	5.6	19.4	NE	22.9	5.1	5.6	66.4
EY	81.8	0.7	3.4	14.1	ID	72.9	0.7	9.1	17.3	NF	11.8	4.6	7.2	76.4
FA	1.6	7.1	56.3	35.0	IE	82.7	0.4	9.2	7.7	NG	0.0	1.3	3.1	95.6
FD	1.3	1.6	70.9	26.2	IF	76.0	0.5	16.1	7.4	NH	0.2	0.0	0.9	98.9
NI	1.2	20.6	2.5	75.7	RG	3.2	3.8	0.2	92.8	VE	67.3	1.0	20.2	11.5
NK	16.6	0.0	1.1	82.3	RH	0.4	2.8	0.5	96.3	VF	55.4	0.7	19.8	24.1
NL	18.8	1.2	1.9	78.1	RI	2.3	47.2	0.3	50.2	VG	0.6	1.0	22.9	75.5
NN	1.4	0.6	2.9	95.1	RK	20.5	1.5	0.7	77.3	VH	4.4	0.5	59.2	35.9
NP	3.1	1.1	1.7	94.1	RL	31.3	5.0	0.2	63.5	VI	11.0	12.4	37.6	39.0
NQ	1.9	1.1	3.0	94.0	RN	10.3	0.6	0.0	89.1	VK	63.6	0.5	20.4	15.5
NR	7.1	0.0	1.4	91.5	RP	1.2	14.0	0.0	84.8	VL	80.0	1.2	4.9	13.9
NS	2.5	2.1	4.3	91.1	RQ	12.7	2.4	0.0	84.9	VN	40.3	2.4	10.8	46.5
NT	4.1	2.5	3.4	90.0	RR	33.5	11.7	0.8	54.0	VP	0.8	5.9	1.3	92.0
NV	9.3	11.5	5.9	73.3	RS	6.6	8.0	0.0	85.4	VQ	52.7	1.2	21.0	25.1
NW	11.3	2.4	19.2	67.1	RT	17.9	22.6	0.0	59.5	VR	35.2	0.1	29.4	35.3
NY	33.8	1.0	3.6	61.6	RV	1.8	46.5	0.4	51.3	VS	29.4	0.5	35.7	34.4
PA	3.6	0.9	6.0	89.5	RW	5.6	0.5	3.7	90.2	VT	42.1	0.8	23.7	33.4
PD	23.4	0.8	6.7	69.1	RY	33.1	0.6	0.9	65.4	VV	4.1	9.5	40.2	46.2
PE	31.5	0.5	33.5	34.5	SA	7.8	21.2	2.1	68.9	VW	22.2	0.6	56.7	20.5
PF	22.7	0.6	14.6	62.1	SD	26.2	10.1	5.5	58.2	VY	78.9	0.0	15.3	5.8
PG	1.0	0.7	4.0	94.3	SE	56.8	8.8	3.4	31.0	WA	4.7	38.4	5.0	51.9
PH	3.4	0.1	44.5	52.0	SF	32.2	0.9	4.2	62.7	WD	1.9	22.2	12.7	63.2
PI	1.4	0.9	16.0	81.7	SG	0.0	2.3	2.2	95.5	WE	7.9	42.6	27.6	21.9
PK	12.1	0.0	25.5	62.4	SH	0.2	0.1	12.7	87.0	WF	4.5	3.9	24.0	67.6
PL	38.4	0.4	2.9	58.3	SI	0.7	37.4	5.6	56.3	WG	0.0	5.0	9.6	85.4
PN	4.4	0.7	3.2	91.7	SK	37.8	0.5	0.5	61.2	WH	0.0	0.2	37.9	61.9
PP	0.5	0.8	1.4	97.3	SL	47.8	2.3	2.7	47.2	WI	17.7	51.8	10.6	19.9
PQ	4.8	0.5	13.1	81.6	SN	4.1	3.3	5.0	87.6	WK	2.3	2.7	28.6	66.4
PR	4.2	0.1	31.6	64.1	SP	0.0	1.4	2.0	96.6	WL	4.4	3.2	34.1	58.3
PS	1.5	0.6	7.8	90.1	SQ	13.9	0.9	3.0	82.2	WN	3.8	5.5	12.1	78.6
PT	6.9	3.0	9.6	80.5	SR	21.9	0.2	2.4	75.5	WP	0.4	31.1	3.6	64.9
PV	1.4	4.5	12.2	81.9	SS	7.3	5.4	4.1	83.2	WQ	3.0	4.8	21.9	70.3
PW	10.6	0.8	41.6	47.0	ST	11.3	6.8	3.3	78.6	WR	0.2	2.0	23.9	73.9
PY	46.4	0.7	8.2	44.7	SV	1.7	46.4	2.4	49.5	WS	2.3	13.6	17.8	66.3
QA	9.2	19.3	5.3	66.2	SW	34.9	1.9	25.1	38.1	WT	4.0	53.2	7.2	35.6
QD	19.7	7.1	5.9	67.3	SY	65.1	0.6	2.4	31.9	WV	0.4	56.1	6.2	37.3
QE	41.3	10.6	3.5	44.6	TA	33.8	7.6	1.2	57.4	WW	2.1	5.2	50.8	41.9
QF	20.7	1.4	5.6	72.3	TD	69.0	1.8	2.2	27.0	WY	21.0	4.4	43.0	31.6
QG	0.1	1.3	2.3	96.3	TE	76.4	2.9	1.5	19.2	YA	6.7	31.4	13.6	48.3
QH	0.4	0.0	2.1	97.5	TF	64.0	0.6	2.9	32.5	YD	3.7	6.5	25.5	64.3
QI	1.1	43.3	1.5	54.1	TG	0.2	1.5	2.5	95.8	YE	10.6	11.1	55.0	23.3
QK	34.9	0.0	0.0	65.1	TH	2.4	0.2	7.8	89.6	YF	4.5	1.0	43.4	51.1
QL	44.7	2.7	2.2	50.4	TI	7.4	36.7	3.8	52.1	YG	0.6	1.4	19.7	78.3
QN	3.2	1.1	2.1	93.6	TK	67.7	0.2	0.8	31.3	YH	0.0	0.3	51.5	48.2
QP	0.2	3.5	2.1	94.2	TL	81.6	0.4	0.6	17.4	YI	0.8	46.7	11.0	41.5
QQ	5.4	0.9	1.7	92.0	TN	26.0	0.6	2.3	71.1	YK	11.1	0.7	46.4	41.8
QR	12.8	0.3	1.9	85.0	TP	1.5	3.2	1.5	93.8	YL	16.4	4.2	41.8	37.6
QS	4.9	2.8	2.5	89.8	TQ	35.9	0.5	1.6	62.0	YN	3.6	1.1	20.5	74.8
QT	10.9	7.5	6.7	74.9	TR	50.9	0.1	1.2	47.8	YP	0.2	14.8	1.5	83.5
QV	1.0	24.2	2.1	72.7	TS	15.9	0.5	1.3	82.3	YQ	5.3	2.7	31.9	60.1
QW	36.3	0.7	8.4	54.6	TT	24.8	1.5	2.2	71.5	YR	1.2	0.5	43.8	54.5
QY	43.7	1.6	4.6	50.1	TV	5.6	32.8	2.8	58.8	YS	1.3	2.0	42.6	54.1
RA	3.2	44.1	0.1	52.6	TW	50.9	0.9	13.7	34.5	YT	3.4	15.7	30.5	50.4
RD	3.2	1.8	0.2	94.8	TY	88.1	0.5	2.8	8.6	YV	0.3	42.2	15.5	42.0
RE	3.9	6.4	0.1	89.6	VA	55.6	2.2	8.7	33.5	YW	3.2	4.6	77.3	14.9
RF	19.1	1.2	0.4	79.3	VD	58.5	1.2	15.8	24.5	YY	14.8	1.1	38.3	45.8

Appendix Table 5.1C: Site specific cleavage data of 324-peptide library GXZAFC in alphabetical order of variable position amino acids for detergent C.

Sequence	% AFC	% FC	% C	% Full	Sequence	% AFC	% FC	% C	% Full	Sequence	% AFC	% FC	% C	% Full
AA	37.9	14.7	43.4	4.0	FE	0.0	0.0	95.5	4.5	IG	0.0	0.5	42.6	56.9
AD	24.8	4.7	64.8	5.7	FF	2.5	1.6	94.0	1.9	IH	0.6	0.0	67.0	32.4
AE	39.9	1.3	35.5	23.3	FG	0.2	0.5	67.5	31.8	II	3.4	1.1	78.0	17.5
AF	46.7	0.7	28.0	24.6	FH	0.1	0.2	73.7	26.0	IK	4.0	0.6	65.5	29.9
AG	0.0	1.4	24.3	74.3	FI	0.0	7.1	87.9	5.0	IL	50.7	0.6	29.6	19.1
AH	2.4	0.9	41.7	55.0	FK	0.0	0.3	76.3	23.4	IN	4.5	0.9	61.6	33.0
AI	0.9	5.4	38.8	54.9	FL	1.0	0.9	95.3	2.8	IP	2.2	8.3	7.5	82.0
AK	8.8	4.1	37.1	50.0	FN	0.9	0.9	75.4	22.8	IQ	24.5	0.6	41.3	33.6
AL	51.8	1.4	8.8	38.0	FP	1.3	16.0	4.7	78.0	IR	1.0	0.2	78.3	20.5
AN	5.6	1.3	29.4	63.7	FQ	5.8	0.9	71.0	22.3	IS	12.0	1.6	75.0	11.4
AP	0.1	10.3	1.4	88.2	FR	0.0	0.3	62.7	37.0	IT	6.9	0.3	84.4	8.4
AQ	28.0	2.2	16.1	53.7	FS	2.2	1.2	79.5	17.1	IV	1.0	1.3	79.3	18.4
AR	0.6	0.0	47.4	52.0	FT	2.5	1.2	81.3	15.0	IW	3.3	0.6	91.4	4.7
AS	11.8	1.3	27.9	59.0	FV	0.0	1.5	95.8	2.7	IY	21.2	0.1	71.0	7.7
AT	7.0	1.5	23.0	68.5	FW	2.9	0.8	92.8	3.5	KA	24.7	32.3	1.3	41.7
AV	2.1	4.3	40.7	52.9	FY	0.4	0.3	90.8	8.5	KD	25.5	0.2	2.0	72.3
AW	6.5	2.0	82.4	9.1	GA	2.2	9.0	12.5	76.3	KE	49.3	2.8	3.6	44.3
AY	48.6	0.8	25.0	25.6	GD	2.2	3.4	40.4	54.0	KF	60.8	0.4	4.4	34.4
DA	9.8	53.3	15.6	21.3	GE	15.4	2.3	63.1	19.2	KG	0.0	5.5	1.4	93.1
DD	13.8	2.0	29.2	55.0	GF	12.9	0.6	34.7	51.8	KH	13.4	0.0	7.7	78.9
DE	26.2	4.9	40.4	28.5	GG	0.5	1.2	19.1	79.2	KI	0.0	51.9	4.7	43.4
DF	26.4	1.2	32.3	40.1	GH	0.0	0.3	23.6	76.1	KK	16.8	3.3	3.2	76.7
DG	0.0	2.4	34.4	63.2	GI	0.8	5.3	23.6	70.3	KL	68.0	3.0	1.3	27.7
DH	1.0	0.0	33.4	65.6	GK	0.3	0.6	19.6	79.5	KN	23.3	1.4	1.6	73.7
DI	9.8	9.5	47.1	33.6	GL	14.0	2.1	17.3	66.6	KP	0.0	5.3	0.0	94.7
DK	1.9	1.5	29.3	67.3	GN	0.2	0.6	16.0	83.2	KQ	56.9	0.0	1.0	42.1
DL	41.0	4.8	21.6	32.6	GP	0.1	4.2	1.9	93.8	KR	3.3	1.5	6.1	89.1
DN	0.6	0.3	33.1	66.0	GQ	3.7	0.7	15.6	80.0	KS	29.7	4.2	1.7	64.4
DP	4.9	11.9	3.9	79.3	GR	2.2	0.0	23.7	74.1	KT	28.0	5.9	2.3	63.8
DQ	21.4	0.9	26.4	51.3	GS	0.8	0.9	17.3	81.0	KV	1.4	41.6	4.5	52.5
DR	0.0	0.0	23.2	76.8	GT	0.8	1.4	19.9	77.9	KW	8.2	0.9	58.6	32.3
DS	10.9	4.5	53.8	30.8	GV	0.6	4.1	20.8	74.5	KY	71.5	0.7	3.6	24.2
DT	6.4	3.5	64.2	25.9	GW	0.4	0.2	90.7	8.7	LA	16.4	7.0	40.6	36.0
DV	5.5	11.0	49.5	34.0	GY	12.6	0.7	31.2	55.5	LD	12.6	0.1	80.2	7.1
DW	0.3	1.6	87.2	10.9	HA	1.2	32.8	4.3	61.7	LE	27.7	0.1	63.2	9.0
DY	41.6	0.4	37.4	20.6	HD	11.0	13.7	18.3	57.0	LF	33.2	0.3	47.0	19.5
EA	32.3	36.4	14.1	17.2	HE	4.0	6.0	9.8	80.2	LG	0.5	0.9	45.6	53.0
ED	41.8	0.2	47.6	10.4	HF	10.5	0.3	34.7	54.5	LH	0.5	0.1	63.8	35.6
EE	57.1	2.6	31.4	8.9	HG	0.4	1.8	6.0	91.8	LI	1.4	3.5	83.5	11.6
EF	53.1	0.6	29.6	16.7	HH	0.4	0.2	41.8	57.6	LK	1.9	0.2	80.6	17.3
EG	1.0	2.5	68.6	27.9	HI	0.0	31.7	11.0	57.3	LL	35.6	4.5	57.2	2.7
EH	1.4	0.1	42.6	55.9	HK	0.7	3.4	34.5	61.4	LN	0.7	0.3	44.3	54.7
EI	0.8	24.7	52.5	22.0	HL	10.9	13.2	13.7	62.2	LP	1.5	8.8	4.2	85.5
EK	8.3	1.8	39.3	50.6	HN	0.5	0.1	7.7	91.7	LQ	19.3	0.4	35.9	44.4
EL	69.4	3.5	14.6	12.5	HP	6.3	3.3	0.2	90.2	LR	0.7	0.0	54.5	44.8
EN	14.1	0.7	67.6	17.6	HQ	6.5	0.2	9.7	83.6	LS	4.3	0.3	49.9	45.5
EP	0.2	34.8	2.8	62.2	HR	0.0	0.3	22.9	76.8	LT	6.4	0.6	53.9	39.1
EQ	59.1	0.7	23.5	16.7	HS	1.3	5.2	13.7	79.8	LV	1.3	1.4	86.2	11.1
ER	0.0	0.0	33.3	66.7	HT	0.5	6.2	15.5	77.8	LW	0.7	0.1	87.2	12.0
ES	32.3	3.2	42.6	21.9	HV	5.8	28.1	12.0	54.1	LY	32.9	1.2	53.6	12.3
ET	23.8	3.8	52.1	20.3	HW	0.0	1.5	68.1	30.4	NA	2.2	16.2	4.1	77.5
EV	2.9	22.8	54.9	19.4	HY	8.3	0.8	29.1	61.8	ND	11.1	2.3	26.8	59.8

EW	2.1	1.1	95.0	1.8	IA	26.6	6.5	37.4	29.5	NE	20.3	8.5	45.0	26.2
EY	52.6	1.0	35.7	10.7	ID	9.0	0.3	81.6	9.1	NF	10.2	2.7	38.2	48.9
FA	1.0	7.3	71.7	20.0	IE	22.5	0.4	68.7	8.4	NG	0.0	1.5	20.9	77.6
FD	0.5	0.1	91.7	7.7	IF	19.9	0.1	75.3	4.7	NH	0.3	0.0	16.1	83.6
NI	0.7	22.3	46.5	30.5	RG	4.0	3.5	1.0	91.5	VE	12.0	0.8	83.2	4.0
NK	0.0	0.3	34.9	64.8	RH	1.6	0.0	5.9	92.5	VF	13.3	0.1	73.8	12.8
NL	14.7	3.6	11.1	70.6	RI	1.3	39.0	2.3	57.4	VG	0.0	0.4	84.6	15.0
NN	0.9	0.7	14.0	84.4	RK	5.2	6.7	4.0	84.1	VH	1.2	0.0	68.4	30.4
NP	2.3	10.0	2.2	85.5	RL	28.6	5.6	1.7	64.1	VI	2.1	1.7	78.4	17.8
NQ	4.3	0.6	12.6	82.5	RN	4.9	1.7	8.1	85.3	VK	0.8	0.2	81.2	17.8
NR	0.9	0.0	20.6	78.5	RP	4.0	8.2	1.7	86.1	VL	28.9	1.3	52.3	17.5
NS	2.5	3.0	12.1	82.4	RQ	27.8	0.9	7.1	64.2	VN	4.7	2.0	54.6	38.7
NT	2.5	2.0	13.8	81.7	RR	2.3	14.8	14.5	68.4	VP	0.6	12.2	4.6	82.6
NV	6.4	7.8	20.5	65.3	RS	7.6	9.5	0.4	82.5	VQ	9.8	0.9	51.1	38.2
NW	0.5	1.3	84.7	13.5	RT	17.7	9.7	0.5	72.1	VR	1.0	0.0	67.6	31.4
NY	16.8	0.5	59.0	23.7	RV	1.1	40.3	0.8	57.8	VS	2.5	0.8	55.0	41.7
PA	39.1	0.4	24.0	36.5	RW	1.4	0.5	46.0	52.1	VT	1.8	0.7	60.0	37.5
PD	42.6	0.2	40.8	16.4	RY	39.5	0.8	4.7	55.0	VV	0.7	2.6	70.0	26.7
PE	50.4	0.0	39.8	9.8	SA	8.0	24.9	4.4	62.7	VW	0.7	0.5	92.2	6.6
PF	57.7	0.1	12.8	29.4	SD	28.3	2.0	35.5	34.2	VY	14.8	0.4	79.8	5.0
PG	1.2	1.3	29.6	67.9	SE	50.3	6.4	31.4	11.9	WA	6.1	46.4	38.0	9.5
PH	5.0	0.0	43.2	51.8	SF	45.8	0.5	16.5	37.2	WD	11.8	0.9	80.9	6.4
PI	0.9	0.6	34.6	63.9	SG	0.0	1.8	13.9	84.3	WE	7.1	11.4	79.9	1.6
PK	5.5	0.0	42.3	52.2	SH	1.2	0.4	23.8	74.6	WF	6.1	0.6	86.8	6.5
PL	60.8	0.8	6.1	32.3	SI	0.9	21.2	30.5	47.4	WG	0.0	1.9	84.5	13.6
PN	10.6	1.2	23.4	64.8	SK	4.4	1.3	15.0	79.3	WH	0.0	5.1	82.2	12.7
PP	1.5	1.8	1.8	94.9	SL	46.7	3.0	8.4	41.9	WI	23.8	17.9	52.3	6.0
PQ	25.4	0.5	12.4	61.7	SN	6.0	5.0	17.5	71.5	WK	0.5	5.3	80.7	13.5
PR	1.7	0.0	44.5	53.8	SP	0.2	11.5	1.3	87.0	WL	1.1	3.7	83.5	11.7
PS	10.9	0.8	19.3	69.0	SQ	18.0	0.8	7.4	73.8	WN	3.3	2.1	84.6	10.0
PT	10.2	2.1	25.1	62.6	SR	0.8	0.5	26.1	72.6	WP	0.0	62.7	8.1	29.2
PV	1.3	2.6	41.4	54.7	SS	4.7	4.4	9.5	81.4	WQ	6.0	4.0	82.4	7.6
PW	4.0	0.0	87.7	8.3	ST	6.2	2.8	13.5	77.5	WR	0.0	1.1	77.2	21.7
PY	68.8	0.1	15.1	16.0	SV	1.3	32.7	37.0	29.0	WS	4.3	5.8	80.4	9.5
QA	8.5	25.7	7.0	58.8	SW	1.5	2.4	84.6	11.5	WT	3.4	6.8	78.0	11.8
QD	27.6	2.2	30.6	39.6	SY	40.2	0.6	8.9	50.3	WV	0.5	14.1	74.2	11.2
QE	47.1	6.4	25.0	21.5	TA	29.0	37.5	3.2	30.3	WW	0.9	2.3	92.5	4.3
QF	31.8	1.5	24.7	42.0	TD	47.2	0.5	32.7	19.6	WY	12.7	1.0	81.7	4.6
QG	0.0	1.1	5.1	93.8	TE	64.3	4.3	18.1	13.3	YA	1.9	7.4	70.0	20.7
QH	0.7	0.0	26.6	72.7	TF	64.5	0.6	13.9	21.0	YD	0.5	0.3	95.0	4.2
QI	0.8	29.8	14.2	55.2	TG	0.0	1.9	12.8	85.3	YE	0.9	0.6	94.7	3.8
QK	3.4	0.9	17.9	77.8	TH	3.0	0.4	25.1	71.5	YF	0.8	0.6	88.8	9.8
QL	33.3	2.8	6.4	57.5	TI	5.1	22.8	27.0	45.1	YG	0.0	0.5	81.2	18.3
QN	1.9	1.4	10.4	86.3	TK	7.2	1.1	17.9	73.8	YH	0.1	0.1	73.9	25.9
QP	0.5	15.8	2.9	80.8	TL	67.4	1.9	5.3	25.4	YI	0.3	1.8	84.5	13.4
QQ	20.0	0.9	7.9	71.2	TN	8.1	1.0	16.7	74.2	YK	0.0	0.1	91.1	8.8
QR	0.7	0.4	21.8	77.1	TP	1.3	12.7	1.5	84.5	YL	0.5	0.5	89.1	9.9
QS	8.3	4.3	11.4	76.0	TQ	27.5	1.1	6.7	64.7	YN	0.0	0.4	83.3	16.3
QT	8.5	6.4	24.0	61.1	TR	9.0	0.2	29.8	61.0	YP	0.2	21.4	8.3	70.1
QV	0.6	18.6	9.0	71.8	TS	16.9	2.0	11.8	69.3	YQ	1.1	1.1	79.3	18.5
QW	6.0	2.5	84.4	7.1	TT	12.3	2.7	16.3	68.7	YR	0.0	0.1	77.5	22.4
QY	32.3	1.6	24.0	42.1	TV	1.8	22.4	25.0	50.8	YS	0.0	0.8	89.2	10.0
RA	2.7	54.5	0.2	42.6	TW	3.5	2.4	87.9	6.2	YT	0.5	0.9	80.1	18.5
RD	5.2	0.0	0.8	94.0	TY	67.2	0.8	19.6	12.4	YV	0.0	1.3	81.4	17.3
RE	10.9	3.3	1.0	84.8	VA	9.2	7.7	42.5	40.6	YW	0.0	0.4	97.9	1.7
RF	36.7	0.8	4.2	58.3	VD	6.5	0.6	86.5	6.4	YY	0.8	0.5	92.4	6.3

Appendix Table 5.1D: Site specific cleavage data of 324-peptide library GXZAFC in alphabetical order of variable position amino acids for detergent D.

Sequence	% AFC	% FC	% C	% Full	Sequence	% AFC	% FC	% C	% Full	Sequence	% AFC	% FC	% C	% Full
AA	13.9	17.6	31.9	36.6	FE	0.7	0.7	79.8	18.8	IG	0.0	0.8	25.9	73.3
AD	9.9	0.9	49.6	39.6	FF	1.7	0.0	83.6	14.7	IH	4.7	0.0	73.3	22.0
AE	13.6	1.2	46.8	38.4	FG	0.7	1.2	47.3	50.8	II	13.1	1.9	71.5	13.5
AF	32.2	0.8	51.1	15.9	FH	0.9	0.3	71.5	27.3	IK	15.6	0.2	64.1	20.1
AG	0.2	1.4	21.7	76.7	FI	0.0	8.1	87.2	4.7	IL	73.1	0.3	13.7	12.9
AH	2.6	1.6	69.1	26.7	FK	0.0	0.2	81.1	18.7	IN	10.8	0.8	43.5	44.9
AI	1.0	4.9	39.8	54.3	FL	9.8	6.5	80.2	3.5	IP	1.0	0.9	1.4	96.7
AK	9.3	1.5	73.1	16.1	FN	1.0	0.8	71.3	26.9	IQ	54.7	0.2	35.0	10.1
AL	72.5	2.2	19.4	5.9	FP	2.7	1.8	1.5	94.0	IR	6.0	0.1	70.5	23.4
AN	2.6	1.3	25.7	70.4	FQ	5.7	0.9	71.9	21.5	IS	27.5	1.4	49.9	21.2
AP	0.0	1.5	1.5	97.0	FR	0.0	0.0	67.8	32.2	IT	22.4	2.2	56.1	19.3
AQ	15.5	2.6	25.7	56.2	FS	2.8	1.3	83.3	12.6	IV	3.9	5.0	68.2	22.9
AR	1.5	0.0	75.1	23.4	FT	2.1	1.6	83.2	13.1	IW	6.6	0.4	87.7	5.3
AS	5.6	1.5	40.9	52.0	FV	0.0	7.2	87.3	5.5	IY	46.5	0.6	47.9	5.0
AT	4.8	1.8	30.3	63.1	FW	3.4	1.6	87.2	7.8	KA	9.6	67.7	4.3	18.4
AV	0.9	3.3	37.8	58.0	FY	1.2	0.2	85.5	13.1	KD	2.6	0.0	0.7	96.7
AW	3.9	2.0	84.2	9.9	GA	0.1	2.9	8.4	88.6	KE	10.8	1.2	0.6	87.4
AY	38.7	0.9	51.2	9.2	GD	1.5	1.5	30.3	66.7	KF	50.8	2.2	4.7	42.3
DA	2.0	5.4	9.2	83.4	GE	0.7	1.2	21.6	76.5	KG	0.0	5.4	0.0	94.6
DD	9.1	2.0	8.7	80.2	GF	2.0	0.6	46.6	50.8	KH	14.4	3.4	12.3	69.9
DE	5.9	2.5	14.0	77.6	GG	0.0	1.2	5.1	93.7	KI	0.0	65.2	5.1	29.7
DF	21.5	1.2	31.0	46.3	GH	0.0	0.8	39.8	59.4	KK	20.8	1.6	5.3	72.3
DG	0.0	2.7	6.9	90.4	GI	1.1	4.7	18.3	75.9	KL	57.2	7.3	2.5	33.0
DH	5.5	0.0	7.5	87.0	GK	0.6	1.2	52.0	46.2	KN	6.8	2.3	0.6	90.3
DI	5.7	2.5	39.6	52.2	GL	9.0	3.0	16.0	72.0	KP	0.0	5.4	0.9	93.7
DK	0.5	0.0	7.1	92.4	GN	0.0	1.4	5.6	93.0	KQ	34.1	0.5	1.8	63.6
DL	25.3	2.0	10.6	62.1	GP	0.0	0.6	1.9	97.5	KR	4.4	1.5	7.1	87.0
DN	1.1	1.7	9.5	87.7	GQ	0.4	0.5	9.4	89.7	KS	14.2	10.8	2.2	72.8
DP	5.7	1.7	5.1	87.5	GR	1.0	0.0	57.4	41.6	KT	10.5	24.7	2.6	62.2
DQ	6.1	1.1	12.8	80.0	GS	0.0	2.4	14.0	83.6	KV	1.1	67.5	7.6	23.8
DR	0.5	0.0	6.0	93.5	GT	0.0	0.9	11.7	87.4	KW	2.0	0.0	35.9	62.1
DS	2.5	1.8	17.2	78.5	GV	0.0	2.1	15.7	82.2	KY	57.0	2.5	5.5	35.0
DT	3.8	1.7	18.5	76.0	GW	0.0	0.5	79.8	19.7	LA	8.2	8.1	36.2	47.5
DV	15.7	3.1	16.0	65.2	GY	4.1	0.6	45.7	49.6	LD	2.6	0.8	43.2	53.4
DW	0.0	2.9	78.7	18.4	HA	1.7	52.9	4.8	40.6	LE	8.6	0.6	58.1	32.7
DY	33.9	0.6	29.8	35.7	HD	19.3	17.2	10.1	53.4	LF	14.2	0.4	61.1	24.3
EA	11.8	20.4	9.1	58.7	HE	3.3	0.4	0.9	95.4	LG	0.7	1.0	21.1	77.2
ED	14.6	0.7	12.5	72.2	HF	9.9	2.0	27.3	60.8	LH	0.6	0.4	65.2	33.8
EE	30.2	1.3	19.8	48.7	HG	0.2	2.0	1.4	96.4	LI	0.5	4.2	66.9	28.4
EF	41.1	0.4	31.2	27.3	HH	1.6	0.8	44.8	52.8	LK	2.2	0.0	78.1	19.7
EG	0.0	0.7	11.1	88.2	HI	0.0	58.5	7.2	34.3	LL	57.8	0.0	33.9	8.3
EH	2.6	0.0	21.8	75.6	HK	2.6	3.4	44.2	49.8	LN	0.6	0.6	34.5	64.3
EI	0.8	14.8	32.9	51.5	HL	16.5	13.8	7.2	62.5	LP	1.6	0.7	1.5	96.2
EK	5.4	0.3	13.1	81.2	HN	0.9	0.0	2.3	96.8	LQ	6.6	0.5	37.2	55.7
EL	62.5	1.8	8.5	27.2	HP	9.6	1.8	0.2	88.4	LR	1.7	0.0	59.3	39.0
EN	2.9	1.4	13.0	82.7	HQ	5.6	0.2	3.4	90.8	LS	2.6	0.7	45.6	51.1
EP	0.0	1.9	3.0	95.1	HR	0.3	0.0	40.6	59.1	LT	7.4	1.1	47.3	44.2
EQ	24.6	0.6	14.4	60.4	HS	1.6	8.3	8.2	81.9	LV	3.3	3.2	64.6	28.9
ER	0.0	0.0	17.3	82.7	HT	2.0	14.2	9.9	73.9	LW	0.2	0.2	88.5	11.1
ES	11.2	1.2	24.1	63.5	HV	5.4	57.1	7.4	30.1	LY	15.7	0.6	70.2	13.5
ET	7.5	1.4	23.3	67.8	HW	0.0	2.1	67.0	30.9	NA	0.9	5.9	3.7	89.5
EV	2.6	16.0	22.5	58.9	HY	12.6	1.8	28.2	57.4	ND	5.2	1.2	6.6	87.0

EW	1.8	1.9	84.8	11.5	IA	52.2	7.9	32.3	7.6	NE	3.4	2.3	9.2	85.1
EY	46.9	1.0	32.4	19.7	ID	16.5	0.3	55.1	28.1	NF	4.5	4.7	30.7	60.1
FA	1.3	22.7	60.2	15.8	IE	39.6	0.2	42.7	17.5	NG	0.0	1.8	2.8	95.4
FD	1.1	0.6	72.6	25.7	IF	36.6	0.3	56.8	6.3	NH	0.3	0.0	17.9	81.8
NI	0.0	9.0	12.5	78.5	RG	2.6	7.1	0.6	89.7	VE	11.2	0.7	72.1	16.0
NK	0.5	0.0	32.3	67.2	RH	3.7	2.6	6.5	87.2	VF	14.6	0.0	71.4	14.0
NL	5.6	1.2	3.8	89.4	RI	1.9	53.3	1.8	43.0	VG	0.5	0.7	33.0	65.8
NN	0.3	0.8	3.8	95.1	RK	6.6	6.5	3.8	83.1	VH	3.0	0.0	80.9	16.1
NP	3.5	0.8	1.4	94.3	RL	26.7	13.0	1.5	58.8	VI	4.5	3.8	64.7	27.0
NQ	0.7	0.7	4.2	94.4	RN	5.0	2.5	0.3	92.2	VK	3.1	0.0	85.9	11.0
NR	2.8	0.0	40.6	56.6	RP	1.1	12.3	0.0	86.6	VL	46.5	1.2	43.7	8.6
NS	0.9	0.7	6.1	92.3	RQ	15.1	3.6	0.8	80.5	VN	11.3	3.1	41.7	43.9
NT	2.8	1.4	5.8	90.0	RR	4.5	15.1	9.3	71.1	VP	1.3	7.3	1.7	89.7
NV	10.9	6.3	10.2	72.6	RS	5.9	26.7	0.7	66.7	VQ	13.8	0.7	57.3	28.2
NW	0.0	1.2	60.6	38.2	RT	11.4	31.2	0.7	56.7	VR	1.7	0.1	74.4	23.8
NY	7.4	0.5	26.5	65.6	RV	2.3	54.2	1.8	41.7	VS	3.9	0.9	65.9	29.3
PA	0.4	0.8	4.2	94.6	RW	1.9	1.9	26.6	69.6	VT	3.4	1.2	56.9	38.5
PD	7.0	0.6	4.1	88.3	RY	28.2	4.7	3.6	63.5	VV	1.0	6.4	58.1	34.5
PE	2.0	0.0	17.4	80.6	SA	2.7	17.6	4.7	75.0	VW	1.4	0.3	85.0	13.3
PF	19.0	0.3	14.9	65.8	SD	3.9	1.6	8.8	85.7	VY	18.1	0.8	70.2	10.9
PG	0.6	0.8	3.2	95.4	SE	9.1	2.3	16.7	71.9	WA	1.2	39.8	25.0	34.0
PH	4.1	0.0	44.3	51.6	SF	27.2	1.3	26.7	44.8	WD	7.7	0.0	53.6	38.7
PI	1.4	1.0	8.2	89.4	SG	0.0	2.2	4.3	93.5	WE	0.9	3.0	64.4	31.7
PK	0.4	0.0	51.2	48.4	SH	1.7	0.0	32.6	65.7	WF	3.4	1.7	82.5	12.4
PL	16.5	1.0	4.2	78.3	SI	1.0	14.1	21.4	63.5	WG	0.0	3.5	48.4	48.1
PN	3.5	1.3	4.0	91.2	SK	7.7	0.8	39.3	52.2	WH	0.0	0.3	81.4	18.3
PP	1.4	1.7	1.4	95.5	SL	43.1	1.7	7.7	47.5	WI	17.6	26.0	46.5	9.9
PQ	1.4	0.5	7.6	90.5	SN	3.6	2.9	9.7	83.8	WK	0.0	1.1	78.6	20.3
PR	1.3	0.3	55.7	42.7	SP	0.0	0.7	1.6	97.7	WL	0.0	1.3	83.6	15.1
PS	0.3	0.7	7.6	91.4	SQ	6.5	0.8	6.2	86.5	WN	0.0	0.9	50.9	48.2
PT	2.9	3.8	9.9	83.4	SR	1.8	0.3	49.4	48.5	WP	0.0	11.2	6.5	82.3
PV	1.1	5.6	15.8	77.5	SS	2.0	6.2	7.7	84.1	WQ	1.2	1.3	57.5	40.0
PW	0.9	0.5	57.3	41.3	ST	2.2	2.1	7.6	88.1	WR	0.0	0.0	70.9	29.1
PY	19.0	0.7	14.9	65.4	SV	2.0	15.8	9.9	72.3	WS	1.0	7.9	56.4	34.7
QA	4.7	14.4	6.8	74.1	SW	1.0	0.0	72.8	26.2	WT	1.9	8.3	55.1	34.7
QD	7.5	1.2	7.6	83.7	SY	31.7	0.7	24.2	43.4	WV	0.1	29.0	47.5	23.4
QE	10.2	2.3	9.6	77.9	TA	5.5	14.0	4.6	75.9	WW	0.0	0.9	92.1	7.0
QF	13.5	0.8	23.4	62.3	TD	9.0	0.5	9.0	81.5	WY	11.5	2.3	77.9	8.3
QG	0.0	0.8	1.7	97.5	TE	26.9	3.5	14.9	54.7	YA	3.7	24.0	53.7	18.6
QH	0.9	0.0	16.5	82.6	TF	53.1	0.4	29.2	17.3	YD	0.9	0.2	73.5	25.4
QI	1.1	12.7	3.2	83.0	TG	0.0	1.7	3.4	94.9	YE	1.0	1.0	80.3	17.7
QK	4.8	0.0	19.7	75.5	TH	4.5	0.2	39.4	55.9	YF	1.2	1.2	78.9	18.7
QL	26.7	2.3	5.5	65.5	TI	8.5	18.6	21.4	51.5	YG	0.9	1.0	57.0	41.1
QN	0.0	0.9	3.1	96.0	TK	15.2	0.6	43.7	40.5	YH	0.7	0.2	68.6	30.5
QP	0.3	1.2	2.2	96.3	TL	77.9	1.5	7.4	13.2	YI	0.0	10.7	67.5	21.8
QQ	1.9	0.9	2.6	94.6	TN	3.5	1.2	7.0	88.3	YK	0.2	0.1	90.1	9.6
QR	1.0	0.4	29.9	68.7	TP	2.0	4.3	1.8	91.9	YL	2.3	0.5	85.6	11.6
QS	2.5	1.6	3.2	92.7	TQ	11.6	0.7	7.3	80.4	YN	0.0	0.4	73.7	25.9
QT	4.4	4.1	10.6	80.9	TR	9.2	0.2	64.6	26.0	YP	0.0	3.9	3.3	92.8
QV	0.8	9.0	4.0	86.2	TS	7.9	1.0	10.3	80.8	YQ	2.2	0.0	82.6	15.2
QW	12.3	3.1	40.7	43.9	TT	4.8	1.7	9.2	84.3	YR	0.0	0.1	78.9	21.0
QY	14.2	2.3	19.8	63.7	TV	1.3	16.1	15.7	66.9	YS	0.0	0.3	88.8	10.9
RA	1.2	73.9	1.0	23.9	TW	0.7	0.9	69.1	29.3	YT	1.2	1.2	76.7	20.9
RD	1.8	0.0	0.0	98.2	TY	55.3	0.5	30.0	14.2	YV	0.0	10.4	73.3	16.3
RE	1.4	1.1	0.0	97.5	VA	12.8	6.7	49.2	31.3	YW	0.0	0.3	96.3	3.4
RF	32.1	5.0	3.1	59.8	VD	10.6	0.7	57.6	31.1	YY	2.0	0.4	86.8	10.8

Appendix Table 5.1E: Site specific cleavage data of 324-peptide library GXZAFC in alphabetical order of variable position amino acids for detergent E.

Sequence	% AFC	% FC	% C	% Full	Sequence	% AFC	% FC	% C	% Full	Sequence	% AFC	% FC	% C	% Full
AA	25.9	18.1	16.3	39.7	FE	0.3	0.5	88.8	10.4	IG	0.0	0.7	47.5	51.8
AD	34.6	0.4	33.0	32.0	FF	0.7	0.0	87.5	11.8	IH	0.2	0.0	67.1	32.7
AE	41.3	0.8	44.4	13.5	FG	0.0	0.5	72.3	27.2	II	2.6	1.5	63.3	32.6
AF	52.3	0.3	26.4	21.0	FH	0.0	0.2	70.0	29.8	IK	2.8	0.0	56.2	41.0
AG	0.0	1.7	23.6	74.7	FI	0.0	0.0	95.5	4.5	IL	44.1	0.9	32.9	22.1
AH	1.6	1.2	30.7	66.5	FK	0.0	0.0	73.9	26.1	IN	4.2	2.0	52.9	40.9
AI	0.4	5.3	49.4	44.9	FL	0.4	0.0	86.2	13.4	IP	0.3	6.5	8.6	84.6
AK	6.1	2.4	26.5	65.0	FN	1.2	1.7	71.2	25.9	IQ	26.0	0.6	50.9	22.5
AL	65.7	0.0	8.5	25.8	FP	1.6	11.2	5.2	82.0	IR	3.3	0.0	47.5	49.2
AN	10.2	0.0	49.3	40.5	FQ	6.0	0.7	70.2	23.1	IS	9.2	1.1	65.2	24.5
AP	0.0	7.0	2.7	90.3	FR	0.0	0.1	55.1	44.8	IT	8.8	1.5	66.9	22.8
AQ	32.4	2.6	18.3	46.7	FS	2.2	1.0	78.9	17.9	IV	1.4	2.4	78.5	17.7
AR	2.2	0.3	37.9	59.6	FT	1.5	1.4	78.8	18.3	IW	4.3	0.7	84.6	10.4
AS	13.9	1.8	28.9	55.4	FV	0.0	0.0	97.5	2.5	IY	20.1	0.4	66.2	13.3
AT	9.3	1.5	24.8	64.4	FW	4.1	1.8	85.8	8.3	KA	22.9	10.5	2.2	64.4
AV	1.6	5.1	41.0	52.3	FY	0.3	0.1	89.5	10.1	KD	8.1	0.0	3.0	88.9
AW	7.7	1.7	74.2	16.4	GA	2.7	13.0	23.3	61.0	KE	31.0	0.0	4.9	64.1
AY	47.3	0.6	21.5	30.6	GD	0.5	1.0	12.2	86.3	KF	54.7	0.0	3.7	41.6
DA	4.3	10.8	10.8	74.1	GE	8.0	2.0	24.6	65.4	KG	0.0	4.0	1.7	94.3
DD	12.1	0.9	16.7	70.3	GF	11.6	1.3	27.7	59.4	KH	3.6	0.0	6.3	90.1
DE	10.3	3.4	23.2	63.1	GG	0.0	1.4	13.3	85.3	KI	0.0	14.7	2.8	82.5
DF	8.2	2.3	14.8	74.7	GH	0.0	0.0	12.9	87.1	KK	15.3	1.1	2.7	80.9
DG	0.0	2.4	13.1	84.5	GI	0.4	5.5	22.4	71.7	KL	40.6	0.5	2.1	56.8
DH	3.1	0.0	17.3	79.6	GK	0.2	2.6	11.5	85.7	KN	9.7	0.0	3.1	87.2
DI	5.7	1.9	28.5	63.9	GL	8.4	3.3	14.0	74.3	KP	0.2	2.2	2.7	94.9
DK	0.0	0.7	27.0	72.3	GN	0.0	0.9	12.9	86.2	KQ	41.3	0.0	2.4	56.3
DL	13.0	1.7	13.6	71.7	GP	0.0	3.0	2.7	94.3	KR	3.1	0.0	3.9	93.0
DN	0.0	0.6	17.0	82.4	GQ	4.9	0.9	16.7	77.5	KS	24.1	0.9	2.7	72.3
DP	4.4	3.4	4.9	87.3	GR	2.7	0.0	9.0	88.3	KT	23.8	2.2	4.0	70.0
DQ	10.2	0.0	17.5	72.3	GS	1.1	2.0	23.2	73.7	KV	1.0	15.0	5.7	78.3
DR	0.0	0.0	8.6	91.4	GT	1.5	1.7	21.1	75.7	KW	8.3	0.0	22.6	69.1
DS	4.0	2.2	24.7	69.1	GV	0.2	5.9	26.5	67.4	KY	58.0	0.0	3.2	38.8
DT	4.1	2.5	28.2	65.2	GW	0.0	1.1	75.4	23.5	LA	14.2	6.9	34.5	44.4
DV	7.6	1.7	21.2	69.5	GY	16.9	0.6	32.2	50.3	LD	13.6	0.8	60.9	24.7
DW	0.0	7.3	61.2	31.5	HA	1.1	18.4	3.9	76.6	LE	24.0	1.3	61.8	12.9
DY	15.3	1.2	23.6	59.9	HD	6.6	5.7	10.8	76.9	LF	28.9	1.1	45.7	24.3
EA	34.3	33.4	17.1	15.2	HE	2.6	2.5	4.7	90.2	LG	0.8	1.3	45.2	52.7
ED	38.3	0.0	34.8	26.9	HF	5.1	0.0	16.4	78.5	LH	0.6	0.1	54.0	45.3
EE	55.2	2.6	31.1	11.1	HG	0.5	0.7	4.9	93.9	LI	0.4	2.4	86.6	10.6
EF	53.5	0.4	32.0	14.1	HH	0.3	0.0	34.4	65.3	LK	1.3	0.0	53.5	45.2
EG	0.0	1.5	37.2	61.3	HI	0.0	18.8	8.5	72.7	LL	49.4	0.0	43.6	7.0
EH	1.1	0.0	36.1	62.8	HK	0.5	1.7	21.7	76.1	LN	1.8	0.9	47.9	49.4
EI	1.3	11.6	52.2	34.9	HL	4.9	4.6	8.0	82.5	LP	1.7	5.1	6.1	87.1
EK	7.5	0.9	24.1	67.5	HN	0.4	0.0	4.6	95.0	LQ	16.9	1.3	40.2	41.6
EL	71.7	2.5	13.1	12.7	HP	8.9	0.9	1.6	88.6	LR	2.9	0.0	39.7	57.4
EN	11.6	0.5	49.6	38.3	HQ	3.8	0.0	7.7	88.5	LS	5.2	0.9	51.5	42.4
EP	0.0	7.0	2.6	90.4	HR	0.0	0.0	18.0	82.0	LT	7.3	1.3	50.9	40.5
EQ	59.4	0.8	23.0	16.8	HS	0.3	2.3	7.1	90.3	LV	1.2	1.1	84.1	13.6
ER	0.8	0.0	37.5	61.7	HT	1.8	2.7	8.7	86.8	LW	0.5	1.2	81.5	16.8
ES	31.1	3.3	45.6	20.0	HV	11.6	10.1	5.5	72.8	LY	27.8	1.6	51.9	18.7
ET	27.8	3.1	45.3	23.8	HW	0.0	1.5	45.3	53.2	NA	2.6	16.5	6.5	74.4
EV	3.6	6.9	45.2	44.3	HY	3.5	0.2	12.6	83.7	ND	5.3	2.7	16.1	75.9

EW	3.6	0.5	82.4	13.5	IA	17.8	5.7	32.9	43.6	NE	7.5	4.6	23.3	64.6
EY	52.3	0.6	33.4	13.7	ID	9.1	0.6	76.7	13.6	NF	4.9	4.3	37.9	52.9
FA	0.9	7.2	74.5	17.4	IE	18.2	0.6	70.6	10.6	NG	0.0	3.0	10.9	86.1
FD	0.6	0.0	92.5	6.9	IF	16.9	0.3	77.6	5.2	NH	0.0	0.0	11.8	88.2
NI	0.5	11.3	28.5	59.7	RG	2.0	0.4	1.0	96.6	VE	10.5	0.9	81.4	7.2
NK	0.0	0.0	11.4	88.6	RH	0.0	0.0	2.8	97.2	VF	12.1	0.6	71.3	16.0
NL	9.6	4.8	13.8	71.8	RI	3.5	15.9	1.0	79.6	VG	0.2	0.8	48.4	50.6
NN	0.9	1.3	14.5	83.3	RK	2.2	3.2	1.2	93.4	VH	6.5	0.0	52.0	41.5
NP	3.0	4.8	5.6	86.6	RL	12.1	0.7	0.4	86.8	VI	4.1	2.9	65.7	27.3
NQ	4.1	1.9	17.2	76.8	RN	5.0	0.0	0.0	95.0	VK	1.4	0.3	51.9	46.4
NR	5.5	0.0	10.0	84.5	RP	1.4	1.3	0.5	96.8	VL	23.5	1.7	50.8	24.0
NS	2.8	4.0	14.4	78.8	RQ	7.6	0.0	0.0	92.4	VN	5.6	2.0	57.2	35.2
NT	2.8	3.9	18.1	75.2	RR	2.7	4.3	7.0	86.0	VP	1.0	10.3	6.4	82.3
NV	7.7	8.9	21.6	61.8	RS	2.5	2.1	0.4	95.0	VQ	10.4	1.0	52.7	35.9
NW	0.0	5.3	53.6	41.1	RT	13.8	2.1	0.0	84.1	VR	2.2	0.0	55.5	42.3
NY	7.6	2.5	30.0	59.9	RV	2.8	15.4	0.0	81.8	VS	3.1	1.4	56.0	39.5
PA	13.7	2.2	11.6	72.5	RW	0.0	0.0	13.4	86.6	VT	2.5	0.9	58.7	37.9
PD	20.4	2.0	11.7	65.9	RY	18.8	0.1	1.6	79.5	VV	0.7	3.2	67.9	28.2
PE	46.1	0.7	34.2	19.0	SA	6.0	21.5	6.0	66.5	VW	1.2	1.5	80.5	16.8
PF	48.8	1.3	12.0	37.9	SD	9.4	1.5	14.8	74.3	VY	13.3	0.4	71.8	14.5
PG	0.6	2.1	20.9	76.4	SE	28.0	1.8	24.0	46.2	WA	6.3	31.0	32.0	30.7
PH	5.3	0.0	27.9	66.8	SF	34.2	1.6	17.6	46.6	WD	13.1	2.6	49.9	34.4
PI	1.5	1.5	37.1	59.9	SG	0.0	3.7	12.3	84.0	WE	6.3	7.7	62.6	23.4
PK	2.2	0.0	18.1	79.7	SH	0.3	0.0	13.6	86.1	WF	6.9	2.7	69.0	21.4
PL	41.4	1.5	6.9	50.2	SI	0.6	16.3	26.2	56.9	WG	0.3	3.9	57.2	38.6
PN	9.3	1.6	16.9	72.2	SK	2.9	0.6	9.1	87.4	WH	0.0	0.9	73.4	25.7
PP	1.6	2.3	3.0	93.1	SL	28.5	4.1	9.7	57.7	WI	18.9	12.4	56.1	12.6
PQ	25.7	1.1	15.3	57.9	SN	4.0	5.4	16.2	74.4	WK	0.0	1.7	69.5	28.8
PR	1.6	0.0	27.4	71.0	SP	0.0	5.9	4.6	89.5	WL	0.0	2.7	73.1	24.2
PS	9.2	1.5	19.5	69.8	SQ	15.0	1.7	10.4	72.9	WN	1.1	1.9	70.2	26.8
PT	8.1	3.6	25.7	62.6	SR	0.7	0.0	15.2	84.1	WP	0.0	22.8	4.9	72.3
PV	1.4	4.1	34.9	59.6	SS	4.6	6.3	14.0	75.1	WQ	6.2	2.5	76.1	15.2
PW	3.2	0.0	75.8	21.0	ST	4.7	4.0	15.3	76.0	WR	0.0	0.0	56.3	43.7
PY	45.7	1.5	11.2	41.6	SV	1.6	17.3	14.1	67.0	WS	3.3	4.5	69.2	23.0
QA	8.6	23.8	9.3	58.3	SW	0.8	3.9	69.2	26.1	WT	4.1	6.8	71.9	17.2
QD	12.9	2.1	14.7	70.3	SY	29.7	2.1	16.6	51.6	WV	0.9	10.8	66.8	21.5
QE	34.9	5.2	21.1	38.8	TA	10.2	17.3	5.7	66.8	WW	0.0	3.2	82.7	14.1
QF	21.3	2.7	22.9	53.1	TD	22.2	2.4	14.0	61.4	WY	13.8	1.2	74.7	10.3
QG	0.0	2.9	9.7	87.4	TE	56.4	4.4	18.3	20.9	YA	2.8	7.3	66.1	23.8
QH	0.2	0.0	9.4	90.4	TF	52.6	1.6	14.9	30.9	YD	0.3	0.3	93.3	6.1
QI	1.7	23.2	12.5	62.6	TG	0.0	3.1	14.8	82.1	YE	0.4	0.8	93.1	5.7
QK	3.6	0.3	9.3	86.8	TH	1.6	0.0	13.8	84.6	YF	0.7	0.4	89.3	9.6
QL	29.5	4.8	9.9	55.8	TI	4.4	19.1	29.3	47.2	YG	0.2	0.6	65.4	33.8
QN	2.1	2.1	14.3	81.5	TK	3.8	0.8	7.7	87.7	YH	0.0	0.0	68.0	32.0
QP	0.3	8.4	6.0	85.3	TL	46.7	2.9	7.2	43.2	YI	0.0	1.4	82.0	16.6
QQ	16.2	1.3	8.2	74.3	TN	5.0	2.0	14.4	78.6	YK	0.3	0.5	75.9	23.3
QR	0.0	0.0	13.6	86.4	TP	1.0	7.5	3.6	87.9	YL	0.9	0.5	81.0	17.6
QS	5.1	4.5	11.5	78.9	TQ	28.4	1.1	7.4	63.1	YN	0.0	0.3	74.4	25.3
QT	9.7	6.6	25.0	58.7	TR	14.2	0.0	12.0	73.8	YP	0.0	15.3	8.6	76.1
QV	0.8	16.1	9.7	73.4	TS	13.5	3.0	12.5	71.0	YQ	1.0	1.4	74.4	23.2
QW	7.8	4.2	51.9	36.1	TT	11.0	3.1	15.9	70.0	YR	0.1	0.0	68.5	31.4
QY	23.2	2.9	20.3	53.6	TV	2.3	16.5	19.8	61.4	YS	0.0	0.3	82.9	16.8
RA	2.8	21.2	0.0	76.0	TW	3.1	2.5	69.0	25.4	YT	0.0	0.5	84.4	15.1
RD	1.1	0.0	2.2	96.7	TY	46.5	1.8	17.2	34.5	YV	0.0	2.1	82.7	15.2
RE	5.5	1.1	2.4	91.0	VA	8.2	7.8	44.7	39.3	YW	0.0	0.0	94.7	5.3
RF	16.9	0.0	1.8	81.3	VD	7.8	0.7	79.2	12.3	YY	0.8	1.0	88.8	9.4

Appendix Table 7.1: Decoding mass-table for all possible full-length and ladder masses of the 1000 member Ac-G($AA_3AA_2AA_1$)GRC library. The table is sorted from lowest to highest mass ordered by full-length, 1st ladder, 2nd ladder, and 3rd ladder masses.

Random Sequence	Full Length Mass	1st Ladder Mass	2nd Ladder Mass	3rd Ladder Mass
GGG	1456	1342	1285	1228
AGG	1470	1342	1285	1228
GAG	1470	1356	1285	1228
GGA	1470	1356	1299	1228
AAG	1484	1356	1285	1228
AGA	1484	1356	1299	1228
GAA	1484	1370	1299	1228
GGS	1486	1372	1315	1304
GSG	1486	1372	1361	1228
SGG	1486	1418	1285	1228
PGG	1496	1356	1285	1228
GPG	1496	1382	1299	1228
GGP	1496	1382	1325	1242
AAA	1498	1370	1299	1228
GGV	1498	1384	1327	1304
GVG	1498	1384	1361	1228
VGG	1498	1418	1285	1228
AGS	1500	1372	1315	1304
ASG	1500	1372	1361	1228
GAS	1500	1386	1315	1304
GSA	1500	1386	1375	1228
SAG	1500	1432	1285	1228
SGA	1500	1432	1299	1228
PAG	1510	1370	1285	1228
PGA	1510	1370	1299	1228
APG	1510	1382	1299	1228
AGP	1510	1382	1325	1242
GPA	1510	1396	1313	1228
GAP	1510	1396	1325	1242
LGG	1512	1342	1285	1228
AGV	1512	1384	1327	1304
AVG	1512	1384	1361	1228

GLG	1512	1398	1285	1228
GAV	1512	1398	1327	1304
GGL	1512	1398	1341	1228
GGI	1512	1398	1341	1304
GIG	1512	1398	1361	1228
GVA	1512	1398	1375	1228
IGG	1512	1418	1285	1228
VAG	1512	1432	1285	1228
VGA	1512	1432	1299	1228
AAS	1514	1386	1315	1304
ASA	1514	1386	1375	1228
SAA	1514	1446	1299	1228
GSS	1516	1402	1391	1304
SGS	1516	1448	1315	1304
SSG	1516	1448	1361	1228
PAA	1524	1384	1299	1228
APA	1524	1396	1313	1228
AAP	1524	1396	1325	1242
LAG	1526	1356	1285	1228
LGA	1526	1356	1299	1228
PGS	1526	1386	1315	1304
PSG	1526	1386	1361	1228
ALG	1526	1398	1285	1228
AAV	1526	1398	1327	1304
AGL	1526	1398	1341	1228
AGI	1526	1398	1341	1304
AIG	1526	1398	1361	1228
AVA	1526	1398	1375	1228
GLA	1526	1412	1299	1228
GPS	1526	1412	1329	1304
GAL	1526	1412	1341	1228
GAI	1526	1412	1341	1304
GIA	1526	1412	1375	1228
GSP	1526	1412	1401	1242
IAG	1526	1432	1285	1228

IGA	1526	1432	1299	1228
VAA	1526	1446	1299	1228
SPG	1526	1458	1299	1228
SGP	1526	1458	1325	1242
QGG	1527	1404	1285	1228
GQG	1527	1413	1347	1228
GGQ	1527	1413	1356	1290
GGE	1528	1414	1357	1304
GEG	1528	1414	1361	1228
GVS	1528	1414	1391	1304
GSV	1528	1414	1403	1304
EGG	1528	1418	1285	1228
VGS	1528	1448	1315	1304
VSG	1528	1448	1361	1228
SGV	1528	1460	1327	1304
SVG	1528	1460	1361	1228
ASS	1530	1402	1391	1304
SAS	1530	1462	1315	1304
SSA	1530	1462	1375	1228
PPG	1536	1396	1299	1228
PGP	1536	1396	1325	1242
GPP	1536	1422	1339	1242
PGV	1538	1398	1327	1304
PVG	1538	1398	1361	1228
GPV	1538	1424	1341	1304
GVP	1538	1424	1401	1242
VPG	1538	1458	1299	1228
VGP	1538	1458	1325	1242
LAA	1540	1370	1299	1228
PAS	1540	1400	1315	1304
PSA	1540	1400	1375	1228
ALA	1540	1412	1299	1228
APS	1540	1412	1329	1304
AAL	1540	1412	1341	1228
AAI	1540	1412	1341	1304

AIA	1540	1412	1375	1228
ASP	1540	1412	1401	1242
GVV	1540	1426	1403	1304
IAA	1540	1446	1299	1228
VGV	1540	1460	1327	1304
VVG	1540	1460	1361	1228
SPA	1540	1472	1313	1228
SAP	1540	1472	1325	1242
AQG	1541	1413	1347	1228
AGQ	1541	1413	1356	1290
QAG	1541	1418	1285	1228
QGA	1541	1418	1299	1228
GAQ	1541	1427	1356	1290
GQA	1541	1427	1361	1228
LGS	1542	1372	1315	1304
LSG	1542	1372	1361	1228
AGE	1542	1414	1357	1304
AEG	1542	1414	1361	1228
AVS	1542	1414	1391	1304
ASV	1542	1414	1403	1304
GLS	1542	1428	1315	1304
GAE	1542	1428	1357	1304
GEA	1542	1428	1375	1228
GIS	1542	1428	1391	1304
GSL	1542	1428	1417	1228
GSI	1542	1428	1417	1304
EAG	1542	1432	1285	1228
EGA	1542	1432	1299	1228
IGS	1542	1448	1315	1304
ISG	1542	1448	1361	1228
VAS	1542	1462	1315	1304
VSA	1542	1462	1375	1228
SLG	1542	1474	1285	1228
SAV	1542	1474	1327	1304
SGL	1542	1474	1341	1228

SGI	1542	1474	1341	1304
SIG	1542	1474	1361	1228
SVA	1542	1474	1375	1228
FGG	1546	1342	1285	1228
GFG	1546	1432	1285	1228
GGF	1546	1432	1375	1228
SSS	1546	1478	1391	1304
PPA	1550	1410	1313	1228
PAP	1550	1410	1325	1242
APP	1550	1422	1339	1242
LPG	1552	1382	1299	1228
LGP	1552	1382	1325	1242
PLG	1552	1412	1285	1228
PAV	1552	1412	1327	1304
PGL	1552	1412	1341	1228
PGI	1552	1412	1341	1304
PIG	1552	1412	1361	1228
PVA	1552	1412	1375	1228
APV	1552	1424	1341	1304
AVP	1552	1424	1401	1242
GLP	1552	1438	1325	1242
GPL	1552	1438	1355	1228
GPI	1552	1438	1355	1304
GIP	1552	1438	1401	1242
IPG	1552	1458	1299	1228
IGP	1552	1458	1325	1242
VPA	1552	1472	1313	1228
VAP	1552	1472	1325	1242
LGV	1554	1384	1327	1304
LVG	1554	1384	1361	1228
AVV	1554	1426	1403	1304
GLV	1554	1440	1327	1304
GIV	1554	1440	1403	1304
GVL	1554	1440	1417	1228
GVI	1554	1440	1417	1304

IGV	1554	1460	1327	1304
IVG	1554	1460	1361	1228
VLG	1554	1474	1285	1228
VAV	1554	1474	1327	1304
VGL	1554	1474	1341	1228
VGI	1554	1474	1341	1304
VIG	1554	1474	1361	1228
VVA	1554	1474	1375	1228
AAQ	1555	1427	1356	1290
AQA	1555	1427	1361	1228
QAA	1555	1432	1299	1228
LAS	1556	1386	1315	1304
LSA	1556	1386	1375	1228
PSS	1556	1416	1391	1304
ALS	1556	1428	1315	1304
AAE	1556	1428	1357	1304
AEA	1556	1428	1375	1228
AIS	1556	1428	1391	1304
ASL	1556	1428	1417	1228
ASI	1556	1428	1417	1304
EAA	1556	1446	1299	1228
IAS	1556	1462	1315	1304
ISA	1556	1462	1375	1228
SLA	1556	1488	1299	1228
SPS	1556	1488	1329	1304
SAL	1556	1488	1341	1228
SAI	1556	1488	1341	1304
SIA	1556	1488	1375	1228
SSP	1556	1488	1401	1242
QGS	1557	1434	1315	1304
QSG	1557	1434	1361	1228
GQS	1557	1443	1377	1304
GSQ	1557	1443	1432	1290
SQG	1557	1489	1347	1228
SGQ	1557	1489	1356	1290

GES	1558	1444	1391	1304
GSE	1558	1444	1433	1304
EGS	1558	1448	1315	1304
ESG	1558	1448	1361	1228
VSS	1558	1478	1391	1304
SGE	1558	1490	1357	1304
SEG	1558	1490	1361	1228
SVS	1558	1490	1391	1304
SSV	1558	1490	1403	1304
FAG	1560	1356	1285	1228
FGA	1560	1356	1299	1228
AFG	1560	1432	1285	1228
AGF	1560	1432	1375	1228
GFA	1560	1446	1299	1228
GAF	1560	1446	1375	1228
LPA	1566	1396	1313	1228
LAP	1566	1396	1325	1242
PLA	1566	1426	1299	1228
PPS	1566	1426	1329	1304
PAL	1566	1426	1341	1228
PAI	1566	1426	1341	1304
PIA	1566	1426	1375	1228
PSP	1566	1426	1401	1242
ALP	1566	1438	1325	1242
APL	1566	1438	1355	1228
API	1566	1438	1355	1304
AIP	1566	1438	1401	1242
IPA	1566	1472	1313	1228
IAP	1566	1472	1325	1242
SPP	1566	1498	1339	1242
PQG	1567	1427	1347	1228
PGQ	1567	1427	1356	1290
QPG	1567	1444	1299	1228
QGP	1567	1444	1325	1242
GPQ	1567	1453	1370	1290

GQP	1567	1453	1387	1242
LLG	1568	1398	1285	1228
LAV	1568	1398	1327	1304
LGL	1568	1398	1341	1228
LGI	1568	1398	1341	1304
LIG	1568	1398	1361	1228
LVA	1568	1398	1375	1228
PGE	1568	1428	1357	1304
PEG	1568	1428	1361	1228
PVS	1568	1428	1391	1304
PSV	1568	1428	1403	1304
ALV	1568	1440	1327	1304
AIV	1568	1440	1403	1304
AVL	1568	1440	1417	1228
AVI	1568	1440	1417	1304
GLL	1568	1454	1341	1228
GLI	1568	1454	1341	1304
GPE	1568	1454	1371	1304
GEP	1568	1454	1401	1242
GIL	1568	1454	1417	1228
GII	1568	1454	1417	1304
EPG	1568	1458	1299	1228
EGP	1568	1458	1325	1242
ILG	1568	1474	1285	1228
IAV	1568	1474	1327	1304
IGL	1568	1474	1341	1228
IGI	1568	1474	1341	1304
IIG	1568	1474	1361	1228
IVA	1568	1474	1375	1228
VLA	1568	1488	1299	1228
VPS	1568	1488	1329	1304
VAL	1568	1488	1341	1228
VAI	1568	1488	1341	1304
VIA	1568	1488	1375	1228
VSP	1568	1488	1401	1242

SPV	1568	1500	1341	1304
SVP	1568	1500	1401	1242
QGV	1569	1446	1327	1304
QVG	1569	1446	1361	1228
GQV	1569	1455	1389	1304
GVQ	1569	1455	1432	1290
VQG	1569	1489	1347	1228
VGQ	1569	1489	1356	1290
GEV	1570	1456	1403	1304
GVE	1570	1456	1433	1304
EGV	1570	1460	1327	1304
EVG	1570	1460	1361	1228
VGE	1570	1490	1357	1304
VEG	1570	1490	1361	1228
VVS	1570	1490	1391	1304
VSV	1570	1490	1403	1304
SVV	1570	1502	1403	1304
AQS	1571	1443	1377	1304
ASQ	1571	1443	1432	1290
QAS	1571	1448	1315	1304
QSA	1571	1448	1375	1228
SAQ	1571	1503	1356	1290
SQA	1571	1503	1361	1228
LSS	1572	1402	1391	1304
AES	1572	1444	1391	1304
ASE	1572	1444	1433	1304
EAS	1572	1462	1315	1304
ESA	1572	1462	1375	1228
ISS	1572	1478	1391	1304
SLS	1572	1504	1315	1304
SAE	1572	1504	1357	1304
SEA	1572	1504	1375	1228
SIS	1572	1504	1391	1304
SSL	1572	1504	1417	1228
SSI	1572	1504	1417	1304

FAA	1574	1370	1299	1228
AFA	1574	1446	1299	1228
AAF	1574	1446	1375	1228
FGS	1576	1372	1315	1304
FSG	1576	1372	1361	1228
PPP	1576	1436	1339	1242
GFS	1576	1462	1315	1304
GSF	1576	1462	1451	1228
SFG	1576	1508	1285	1228
SGF	1576	1508	1375	1228
PPV	1578	1438	1341	1304
PVP	1578	1438	1401	1242
VPP	1578	1498	1339	1242
PVV	1580	1440	1403	1304
VPV	1580	1500	1341	1304
VVP	1580	1500	1401	1242
PAQ	1581	1441	1356	1290
PQA	1581	1441	1361	1228
APQ	1581	1453	1370	1290
AQP	1581	1453	1387	1242
QPA	1581	1458	1313	1228
QAP	1581	1458	1325	1242
LLA	1582	1412	1299	1228
LPS	1582	1412	1329	1304
LAL	1582	1412	1341	1228
LAI	1582	1412	1341	1304
LIA	1582	1412	1375	1228
LSP	1582	1412	1401	1242
PLS	1582	1442	1315	1304
PAE	1582	1442	1357	1304
PEA	1582	1442	1375	1228
PIS	1582	1442	1391	1304
PSL	1582	1442	1417	1228
PSI	1582	1442	1417	1304
ALL	1582	1454	1341	1228

ALI	1582	1454	1341	1304
APE	1582	1454	1371	1304
AEP	1582	1454	1401	1242
AIL	1582	1454	1417	1228
AII	1582	1454	1417	1304
EPA	1582	1472	1313	1228
EAP	1582	1472	1325	1242
ILA	1582	1488	1299	1228
IPS	1582	1488	1329	1304
IAL	1582	1488	1341	1228
IAI	1582	1488	1341	1304
IIA	1582	1488	1375	1228
ISP	1582	1488	1401	1242
VVV	1582	1502	1403	1304
SLP	1582	1514	1325	1242
SPL	1582	1514	1355	1228
SPI	1582	1514	1355	1304
SIP	1582	1514	1401	1242
LQG	1583	1413	1347	1228
LGQ	1583	1413	1356	1290
AQV	1583	1455	1389	1304
AVQ	1583	1455	1432	1290
QLG	1583	1460	1285	1228
QAV	1583	1460	1327	1304
QGL	1583	1460	1341	1228
QGI	1583	1460	1341	1304
QIG	1583	1460	1361	1228
QVA	1583	1460	1375	1228
GLQ	1583	1469	1356	1290
GQL	1583	1469	1403	1228
GQI	1583	1469	1403	1304
GIQ	1583	1469	1432	1290
IQG	1583	1489	1347	1228
IGQ	1583	1489	1356	1290
VAQ	1583	1503	1356	1290

VQA	1583	1503	1361	1228
LGE	1584	1414	1357	1304
LEG	1584	1414	1361	1228
LVS	1584	1414	1391	1304
LSV	1584	1414	1403	1304
AEV	1584	1456	1403	1304
AVE	1584	1456	1433	1304
GLE	1584	1470	1357	1304
GEL	1584	1470	1417	1228
GEI	1584	1470	1417	1304
GIE	1584	1470	1433	1304
ELG	1584	1474	1285	1228
EAV	1584	1474	1327	1304
EGL	1584	1474	1341	1228
EGI	1584	1474	1341	1304
EIG	1584	1474	1361	1228
EVA	1584	1474	1375	1228
IGE	1584	1490	1357	1304
IEG	1584	1490	1361	1228
IVS	1584	1490	1391	1304
ISV	1584	1490	1403	1304
VLS	1584	1504	1315	1304
VAE	1584	1504	1357	1304
VEA	1584	1504	1375	1228
VIS	1584	1504	1391	1304
VSL	1584	1504	1417	1228
VSI	1584	1504	1417	1304
SLV	1584	1516	1327	1304
SIV	1584	1516	1403	1304
SVL	1584	1516	1417	1228
SVI	1584	1516	1417	1304
FPG	1586	1382	1299	1228
FGP	1586	1382	1325	1242
PFG	1586	1446	1285	1228
PGF	1586	1446	1375	1228

GFP	1586	1472	1325	1242
GPF	1586	1472	1389	1228
QSS	1587	1464	1391	1304
SQS	1587	1519	1377	1304
SSQ	1587	1519	1432	1290
FGV	1588	1384	1327	1304
FVG	1588	1384	1361	1228
GFV	1588	1474	1327	1304
GVF	1588	1474	1451	1228
ESS	1588	1478	1391	1304
VFG	1588	1508	1285	1228
VGF	1588	1508	1375	1228
SES	1588	1520	1391	1304
SSE	1588	1520	1433	1304
FAS	1590	1386	1315	1304
FSA	1590	1386	1375	1228
AFS	1590	1462	1315	1304
ASF	1590	1462	1451	1228
SFA	1590	1522	1299	1228
SAF	1590	1522	1375	1228
LPP	1592	1422	1339	1242
PLP	1592	1452	1325	1242
PPL	1592	1452	1355	1228
PPI	1592	1452	1355	1304
PIP	1592	1452	1401	1242
IPP	1592	1498	1339	1242
LPV	1594	1424	1341	1304
LVP	1594	1424	1401	1242
PLV	1594	1454	1327	1304
PIV	1594	1454	1403	1304
PVL	1594	1454	1417	1228
PVI	1594	1454	1417	1304
IPV	1594	1500	1341	1304
IVP	1594	1500	1401	1242
VLP	1594	1514	1325	1242

VPL	1594	1514	1355	1228
VPI	1594	1514	1355	1304
VIP	1594	1514	1401	1242
LVV	1596	1426	1403	1304
IVV	1596	1502	1403	1304
VLV	1596	1516	1327	1304
VIV	1596	1516	1403	1304
VVL	1596	1516	1417	1228
VVI	1596	1516	1417	1304
LAQ	1597	1427	1356	1290
LQA	1597	1427	1361	1228
PQS	1597	1457	1377	1304
PSQ	1597	1457	1432	1290
ALQ	1597	1469	1356	1290
AQL	1597	1469	1403	1228
AQI	1597	1469	1403	1304
AIQ	1597	1469	1432	1290
QLA	1597	1474	1299	1228
QPS	1597	1474	1329	1304
QAL	1597	1474	1341	1228
QAI	1597	1474	1341	1304
QIA	1597	1474	1375	1228
QSP	1597	1474	1401	1242
IAQ	1597	1503	1356	1290
IQA	1597	1503	1361	1228
SPQ	1597	1529	1370	1290
SQP	1597	1529	1387	1242
LLS	1598	1428	1315	1304
LAE	1598	1428	1357	1304
LEA	1598	1428	1375	1228
LIS	1598	1428	1391	1304
LSL	1598	1428	1417	1228
LSI	1598	1428	1417	1304
PES	1598	1458	1391	1304
PSE	1598	1458	1433	1304

ALE	1598	1470	1357	1304
AEL	1598	1470	1417	1228
AEI	1598	1470	1417	1304
AIE	1598	1470	1433	1304
QQG	1598	1475	1347	1228
QGQ	1598	1475	1356	1290
GQQ	1598	1484	1418	1290
ELA	1598	1488	1299	1228
EPS	1598	1488	1329	1304
EAL	1598	1488	1341	1228
EAI	1598	1488	1341	1304
EIA	1598	1488	1375	1228
ESP	1598	1488	1401	1242
ILS	1598	1504	1315	1304
IAE	1598	1504	1357	1304
IEA	1598	1504	1375	1228
IIS	1598	1504	1391	1304
ISL	1598	1504	1417	1228
ISI	1598	1504	1417	1304
SLL	1598	1530	1341	1228
SLI	1598	1530	1341	1304
SPE	1598	1530	1371	1304
SEP	1598	1530	1401	1242
SIL	1598	1530	1417	1228
SII	1598	1530	1417	1304
QGE	1599	1476	1357	1304
QEG	1599	1476	1361	1228
QVS	1599	1476	1391	1304
QSV	1599	1476	1403	1304
GQE	1599	1485	1419	1304
GEQ	1599	1485	1432	1290
EQG	1599	1489	1347	1228
EGQ	1599	1489	1356	1290
VQS	1599	1519	1377	1304
VSQ	1599	1519	1432	1290

SQV	1599	1531	1389	1304
SVQ	1599	1531	1432	1290
FPA	1600	1396	1313	1228
FAP	1600	1396	1325	1242
PFA	1600	1460	1299	1228
PAF	1600	1460	1375	1228
AFP	1600	1472	1325	1242
APF	1600	1472	1389	1228
GEE	1600	1486	1433	1304
EGE	1600	1490	1357	1304
EEG	1600	1490	1361	1228
EVS	1600	1490	1391	1304
ESV	1600	1490	1403	1304
VES	1600	1520	1391	1304
VSE	1600	1520	1433	1304
SEV	1600	1532	1403	1304
SVE	1600	1532	1433	1304
FLG	1602	1398	1285	1228
FAV	1602	1398	1327	1304
FGL	1602	1398	1341	1228
FGI	1602	1398	1341	1304
FIG	1602	1398	1361	1228
FVA	1602	1398	1375	1228
LFG	1602	1432	1285	1228
LGF	1602	1432	1375	1228
AFV	1602	1474	1327	1304
AVF	1602	1474	1451	1228
GFL	1602	1488	1341	1228
GFI	1602	1488	1341	1304
GLF	1602	1488	1375	1228
GIF	1602	1488	1451	1228
IFG	1602	1508	1285	1228
IGF	1602	1508	1375	1228
VFA	1602	1522	1299	1228
VAF	1602	1522	1375	1228

FSS	1606	1402	1391	1304
SFS	1606	1538	1315	1304
SSF	1606	1538	1451	1228
PPQ	1607	1467	1370	1290
PQP	1607	1467	1387	1242
QPP	1607	1484	1339	1242
LLP	1608	1438	1325	1242
LPL	1608	1438	1355	1228
LPI	1608	1438	1355	1304
LIP	1608	1438	1401	1242
PLL	1608	1468	1341	1228
PLI	1608	1468	1341	1304
PPE	1608	1468	1371	1304
PEP	1608	1468	1401	1242
PIL	1608	1468	1417	1228
PII	1608	1468	1417	1304
EPP	1608	1498	1339	1242
ILP	1608	1514	1325	1242
IPL	1608	1514	1355	1228
IPI	1608	1514	1355	1304
IIP	1608	1514	1401	1242
PQV	1609	1469	1389	1304
PVQ	1609	1469	1432	1290
QPV	1609	1486	1341	1304
QVP	1609	1486	1401	1242
VPQ	1609	1529	1370	1290
VQP	1609	1529	1387	1242
LLV	1610	1440	1327	1304
LIV	1610	1440	1403	1304
LVL	1610	1440	1417	1228
LVI	1610	1440	1417	1304
PEV	1610	1470	1403	1304
PVE	1610	1470	1433	1304
EPV	1610	1500	1341	1304
EVP	1610	1500	1401	1242

ILV	1610	1516	1327	1304
IIV	1610	1516	1403	1304
IVL	1610	1516	1417	1228
IVI	1610	1516	1417	1304
VLL	1610	1530	1341	1228
VLI	1610	1530	1341	1304
VPE	1610	1530	1371	1304
VEP	1610	1530	1401	1242
VIL	1610	1530	1417	1228
VII	1610	1530	1417	1304
QVV	1611	1488	1403	1304
VQV	1611	1531	1389	1304
VVQ	1611	1531	1432	1290
AQQ	1612	1484	1418	1290
QAQ	1612	1489	1356	1290
QQA	1612	1489	1361	1228
EVV	1612	1502	1403	1304
VEV	1612	1532	1403	1304
VVE	1612	1532	1433	1304
LQS	1613	1443	1377	1304
LSQ	1613	1443	1432	1290
AQE	1613	1485	1419	1304
AEQ	1613	1485	1432	1290
QLS	1613	1490	1315	1304
QAE	1613	1490	1357	1304
QEA	1613	1490	1375	1228
QIS	1613	1490	1391	1304
QSL	1613	1490	1417	1228
QSI	1613	1490	1417	1304
EAQ	1613	1503	1356	1290
EQA	1613	1503	1361	1228
IQS	1613	1519	1377	1304
ISQ	1613	1519	1432	1290
SLQ	1613	1545	1356	1290
SQL	1613	1545	1403	1228

SQI	1613	1545	1403	1304
SIQ	1613	1545	1432	1290
LES	1614	1444	1391	1304
LSE	1614	1444	1433	1304
AEE	1614	1486	1433	1304
ELS	1614	1504	1315	1304
EAE	1614	1504	1357	1304
EEA	1614	1504	1375	1228
EIS	1614	1504	1391	1304
ESL	1614	1504	1417	1228
ESI	1614	1504	1417	1304
IES	1614	1520	1391	1304
ISE	1614	1520	1433	1304
SLE	1614	1546	1357	1304
SEL	1614	1546	1417	1228
SEI	1614	1546	1417	1304
SIE	1614	1546	1433	1304
FLA	1616	1412	1299	1228
FPS	1616	1412	1329	1304
FAL	1616	1412	1341	1228
FAI	1616	1412	1341	1304
FIA	1616	1412	1375	1228
FSP	1616	1412	1401	1242
LFA	1616	1446	1299	1228
LAF	1616	1446	1375	1228
PFS	1616	1476	1315	1304
PSF	1616	1476	1451	1228
AFL	1616	1488	1341	1228
AFI	1616	1488	1341	1304
ALF	1616	1488	1375	1228
AIF	1616	1488	1451	1228
IFA	1616	1522	1299	1228
IAF	1616	1522	1375	1228
SFP	1616	1548	1325	1242
SPF	1616	1548	1389	1228

FQG	1617	1413	1347	1228
FGQ	1617	1413	1356	1290
QFG	1617	1494	1285	1228
QGF	1617	1494	1375	1228
GFQ	1617	1503	1356	1290
GQF	1617	1503	1437	1228
FGE	1618	1414	1357	1304
FEG	1618	1414	1361	1228
FVS	1618	1414	1391	1304
FSV	1618	1414	1403	1304
GFE	1618	1504	1357	1304
GEF	1618	1504	1451	1228
EFG	1618	1508	1285	1228
EGF	1618	1508	1375	1228
VFS	1618	1538	1315	1304
VSF	1618	1538	1451	1228
SFV	1618	1550	1327	1304
SVF	1618	1550	1451	1228
LPQ	1623	1453	1370	1290
LQP	1623	1453	1387	1242
PLQ	1623	1483	1356	1290
PQL	1623	1483	1403	1228
PQI	1623	1483	1403	1304
PIQ	1623	1483	1432	1290
QLP	1623	1500	1325	1242
QPL	1623	1500	1355	1228
QPI	1623	1500	1355	1304
QIP	1623	1500	1401	1242
IPQ	1623	1529	1370	1290
IQP	1623	1529	1387	1242
LLL	1624	1454	1341	1228
LLI	1624	1454	1341	1304
LPE	1624	1454	1371	1304
LEP	1624	1454	1401	1242
LIL	1624	1454	1417	1228

LII	1624	1454	1417	1304
PLE	1624	1484	1357	1304
PEL	1624	1484	1417	1228
PEI	1624	1484	1417	1304
PIE	1624	1484	1433	1304
ELP	1624	1514	1325	1242
EPL	1624	1514	1355	1228
EPI	1624	1514	1355	1304
EIP	1624	1514	1401	1242
ILL	1624	1530	1341	1228
ILI	1624	1530	1341	1304
IPE	1624	1530	1371	1304
IEP	1624	1530	1401	1242
IIL	1624	1530	1417	1228
III	1624	1530	1417	1304
LQV	1625	1455	1389	1304
LVQ	1625	1455	1432	1290
QLV	1625	1502	1327	1304
QIV	1625	1502	1403	1304
QVL	1625	1502	1417	1228
QVI	1625	1502	1417	1304
IQV	1625	1531	1389	1304
IVQ	1625	1531	1432	1290
VLQ	1625	1545	1356	1290
VQL	1625	1545	1403	1228
VQI	1625	1545	1403	1304
VIQ	1625	1545	1432	1290
FPP	1626	1422	1339	1242
LEV	1626	1456	1403	1304
LVE	1626	1456	1433	1304
PFP	1626	1486	1325	1242
PPF	1626	1486	1389	1228
ELV	1626	1516	1327	1304
EIV	1626	1516	1403	1304
EVL	1626	1516	1417	1228

EVI	1626	1516	1417	1304
IEV	1626	1532	1403	1304
IVE	1626	1532	1433	1304
VLE	1626	1546	1357	1304
VEL	1626	1546	1417	1228
VEI	1626	1546	1417	1304
VIE	1626	1546	1433	1304
FPV	1628	1424	1341	1304
FVP	1628	1424	1401	1242
PFV	1628	1488	1327	1304
PVF	1628	1488	1451	1228
QQS	1628	1505	1377	1304
QSQ	1628	1505	1432	1290
VFP	1628	1548	1325	1242
VPF	1628	1548	1389	1228
SQQ	1628	1560	1418	1290
QES	1629	1506	1391	1304
QSE	1629	1506	1433	1304
EQS	1629	1519	1377	1304
ESQ	1629	1519	1432	1290
SQE	1629	1561	1419	1304
SEQ	1629	1561	1432	1290
FVV	1630	1426	1403	1304
EES	1630	1520	1391	1304
ESE	1630	1520	1433	1304
VFV	1630	1550	1327	1304
VVF	1630	1550	1451	1228
SEE	1630	1562	1433	1304
FAQ	1631	1427	1356	1290
FQA	1631	1427	1361	1228
AFQ	1631	1503	1356	1290
AQF	1631	1503	1437	1228
QFA	1631	1508	1299	1228
QAF	1631	1508	1375	1228
FLS	1632	1428	1315	1304

FAE	1632	1428	1357	1304
FEA	1632	1428	1375	1228
FIS	1632	1428	1391	1304
FSL	1632	1428	1417	1228
FSI	1632	1428	1417	1304
LFS	1632	1462	1315	1304
LSF	1632	1462	1451	1228
AFE	1632	1504	1357	1304
AEF	1632	1504	1451	1228
EFA	1632	1522	1299	1228
EAF	1632	1522	1375	1228
IFS	1632	1538	1315	1304
ISF	1632	1538	1451	1228
SFL	1632	1564	1341	1228
SFI	1632	1564	1341	1304
SLF	1632	1564	1375	1228
SIF	1632	1564	1451	1228
FFG	1636	1432	1285	1228
FGF	1636	1432	1375	1228
GFF	1636	1522	1375	1228
PQQ	1638	1498	1418	1290
QPQ	1638	1515	1370	1290
QQP	1638	1515	1387	1242
LLQ	1639	1469	1356	1290
LQL	1639	1469	1403	1228
LQI	1639	1469	1403	1304
LIQ	1639	1469	1432	1290
PQE	1639	1499	1419	1304
PEQ	1639	1499	1432	1290
QLL	1639	1516	1341	1228
QLI	1639	1516	1341	1304
QPE	1639	1516	1371	1304
QEP	1639	1516	1401	1242
QIL	1639	1516	1417	1228
QII	1639	1516	1417	1304

EPQ	1639	1529	1370	1290
EQP	1639	1529	1387	1242
ILQ	1639	1545	1356	1290
IQL	1639	1545	1403	1228
IQI	1639	1545	1403	1304
IIQ	1639	1545	1432	1290
LLE	1640	1470	1357	1304
LEL	1640	1470	1417	1228
LEI	1640	1470	1417	1304
LIE	1640	1470	1433	1304
PEE	1640	1500	1433	1304
QQV	1640	1517	1389	1304
QVQ	1640	1517	1432	1290
ELL	1640	1530	1341	1228
ELI	1640	1530	1341	1304
EPE	1640	1530	1371	1304
EEP	1640	1530	1401	1242
EIL	1640	1530	1417	1228
EII	1640	1530	1417	1304
ILE	1640	1546	1357	1304
IEL	1640	1546	1417	1228
IEI	1640	1546	1417	1304
IIE	1640	1546	1433	1304
VQQ	1640	1560	1418	1290
QEV	1641	1518	1403	1304
QVE	1641	1518	1433	1304
EQV	1641	1531	1389	1304
EVQ	1641	1531	1432	1290
VQE	1641	1561	1419	1304
VEQ	1641	1561	1432	1290
FLP	1642	1438	1325	1242
FPL	1642	1438	1355	1228
FPI	1642	1438	1355	1304
FIP	1642	1438	1401	1242
LFP	1642	1472	1325	1242

LPF	1642	1472	1389	1228
PFL	1642	1502	1341	1228
PFI	1642	1502	1341	1304
PLF	1642	1502	1375	1228
PIF	1642	1502	1451	1228
EEV	1642	1532	1403	1304
EVE	1642	1532	1433	1304
IFP	1642	1548	1325	1242
IPF	1642	1548	1389	1228
VEE	1642	1562	1433	1304
FLV	1644	1440	1327	1304
FIV	1644	1440	1403	1304
FVL	1644	1440	1417	1228
FVI	1644	1440	1417	1304
LFV	1644	1474	1327	1304
LVF	1644	1474	1451	1228
IFV	1644	1550	1327	1304
IVF	1644	1550	1451	1228
VFL	1644	1564	1341	1228
VFI	1644	1564	1341	1304
VLF	1644	1564	1375	1228
VIF	1644	1564	1451	1228
FQS	1647	1443	1377	1304
FSQ	1647	1443	1432	1290
QFS	1647	1524	1315	1304
QSF	1647	1524	1451	1228
SFQ	1647	1579	1356	1290
SQF	1647	1579	1437	1228
FES	1648	1444	1391	1304
FSE	1648	1444	1433	1304
EFS	1648	1538	1315	1304
ESF	1648	1538	1451	1228
SFE	1648	1580	1357	1304
SEF	1648	1580	1451	1228
FFA	1650	1446	1299	1228

FAF	1650	1446	1375	1228
AFF	1650	1522	1375	1228
LQQ	1654	1484	1418	1290
QLQ	1654	1531	1356	1290
QQL	1654	1531	1403	1228
QQI	1654	1531	1403	1304
QIQ	1654	1531	1432	1290
IQQ	1654	1560	1418	1290
LQE	1655	1485	1419	1304
LEQ	1655	1485	1432	1290
QLE	1655	1532	1357	1304
QEL	1655	1532	1417	1228
QEI	1655	1532	1417	1304
QIE	1655	1532	1433	1304
ELQ	1655	1545	1356	1290
EQL	1655	1545	1403	1228
EQI	1655	1545	1403	1304
EIQ	1655	1545	1432	1290
IQE	1655	1561	1419	1304
IEQ	1655	1561	1432	1290
LEE	1656	1486	1433	1304
ELE	1656	1546	1357	1304
EEL	1656	1546	1417	1228
EEI	1656	1546	1417	1304
EIE	1656	1546	1433	1304
IEE	1656	1562	1433	1304
FPQ	1657	1453	1370	1290
FQP	1657	1453	1387	1242
PFQ	1657	1517	1356	1290
PQF	1657	1517	1437	1228
QFP	1657	1534	1325	1242
QPF	1657	1534	1389	1228
FLL	1658	1454	1341	1228
FLI	1658	1454	1341	1304
FPE	1658	1454	1371	1304

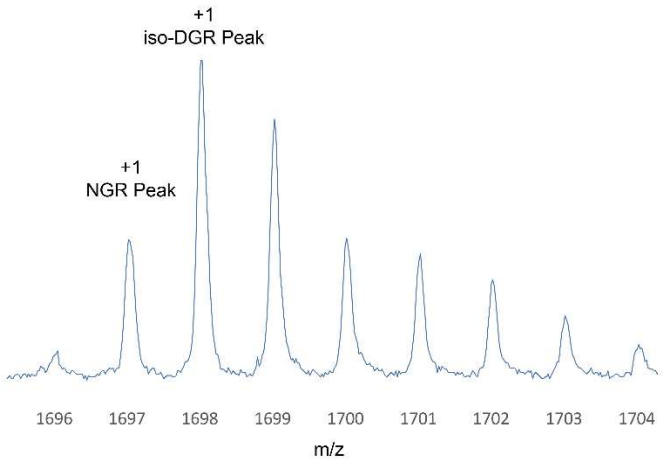
FEP	1658	1454	1401	1242
FIL	1658	1454	1417	1228
FII	1658	1454	1417	1304
LFL	1658	1488	1341	1228
LFI	1658	1488	1341	1304
LLF	1658	1488	1375	1228
LIF	1658	1488	1451	1228
PFE	1658	1518	1357	1304
PEF	1658	1518	1451	1228
EFP	1658	1548	1325	1242
EPF	1658	1548	1389	1228
IFL	1658	1564	1341	1228
IFI	1658	1564	1341	1304
ILF	1658	1564	1375	1228
IIF	1658	1564	1451	1228
FQV	1659	1455	1389	1304
FVQ	1659	1455	1432	1290
QFV	1659	1536	1327	1304
QVF	1659	1536	1451	1228
VFQ	1659	1579	1356	1290
VQF	1659	1579	1437	1228
FEV	1660	1456	1403	1304
FVE	1660	1456	1433	1304
EFV	1660	1550	1327	1304
EVF	1660	1550	1451	1228
VFE	1660	1580	1357	1304
VEF	1660	1580	1451	1228
FFS	1666	1462	1315	1304
FSF	1666	1462	1451	1228
SFF	1666	1598	1375	1228
QQQ	1669	1546	1418	1290
QQE	1670	1547	1419	1304
QEQ	1670	1547	1432	1290
EQQ	1670	1560	1418	1290
QEE	1671	1548	1433	1304

EQE	1671	1561	1419	1304
EEQ	1671	1561	1432	1290
EEE	1672	1562	1433	1304
FLQ	1673	1469	1356	1290
FQL	1673	1469	1403	1228
FQI	1673	1469	1403	1304
FIQ	1673	1469	1432	1290
LFQ	1673	1503	1356	1290
LQF	1673	1503	1437	1228
QFL	1673	1550	1341	1228
QFI	1673	1550	1341	1304
QLF	1673	1550	1375	1228
QIF	1673	1550	1451	1228
IFQ	1673	1579	1356	1290
IQF	1673	1579	1437	1228
FLE	1674	1470	1357	1304
FEL	1674	1470	1417	1228
FEI	1674	1470	1417	1304
FIE	1674	1470	1433	1304
LFE	1674	1504	1357	1304
LEF	1674	1504	1451	1228
EFL	1674	1564	1341	1228
EFI	1674	1564	1341	1304
ELF	1674	1564	1375	1228
EIF	1674	1564	1451	1228
IFE	1674	1580	1357	1304
IEF	1674	1580	1451	1228
FFP	1676	1472	1325	1242
FPF	1676	1472	1389	1228
PFF	1676	1536	1375	1228
FFV	1678	1474	1327	1304
FVF	1678	1474	1451	1228
VFF	1678	1598	1375	1228
FQQ	1688	1484	1418	1290
QFQ	1688	1565	1356	1290

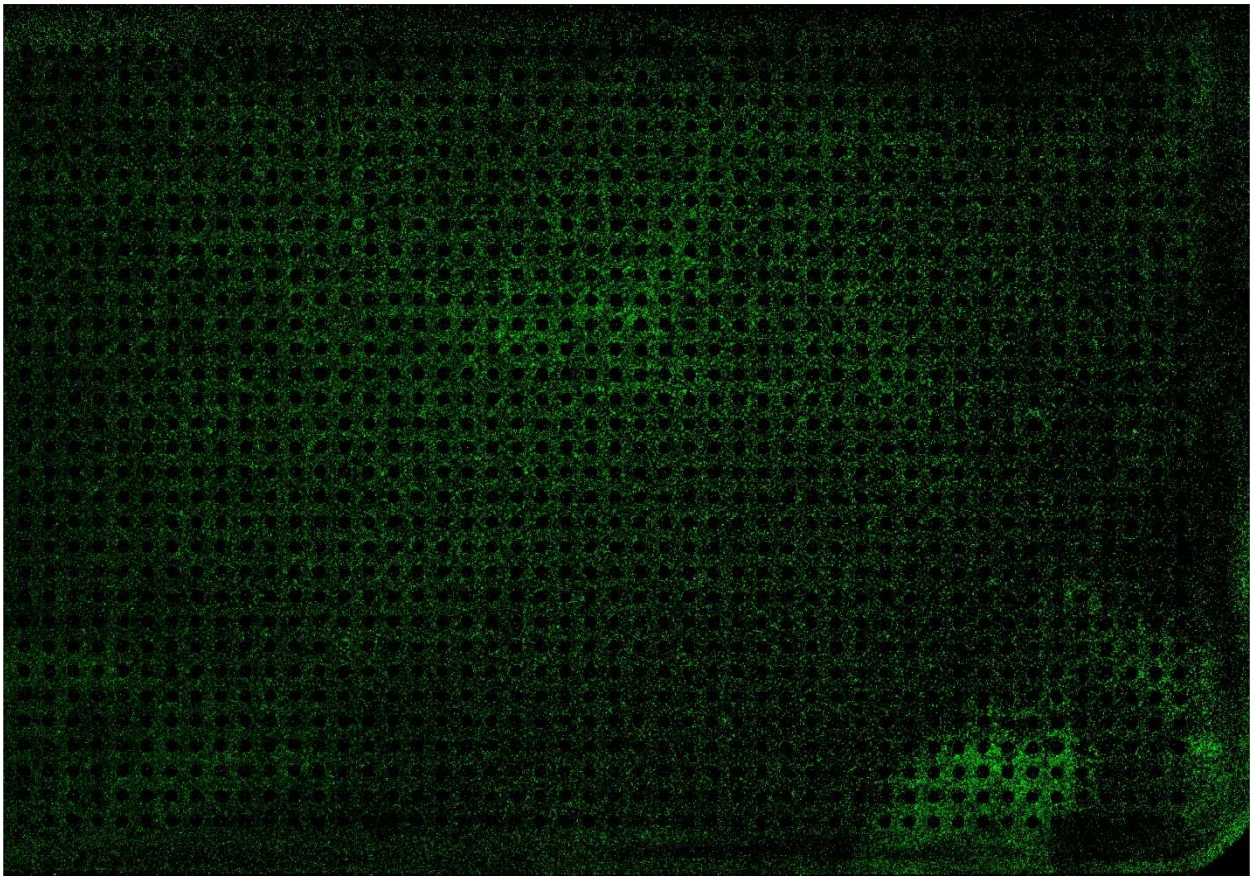
QQF	1688	1565	1437	1228
FQE	1689	1485	1419	1304
FEQ	1689	1485	1432	1290
QFE	1689	1566	1357	1304
QEF	1689	1566	1451	1228
EFQ	1689	1579	1356	1290
EQF	1689	1579	1437	1228
FEE	1690	1486	1433	1304
EFE	1690	1580	1357	1304
EEF	1690	1580	1451	1228
FFL	1692	1488	1341	1228
FFI	1692	1488	1341	1304
FLF	1692	1488	1375	1228
FIF	1692	1488	1451	1228
LFF	1692	1522	1375	1228
IFF	1692	1598	1375	1228
FFQ	1707	1503	1356	1290
FQF	1707	1503	1437	1228
QFF	1707	1584	1375	1228
FFE	1708	1504	1357	1304
FEF	1708	1504	1451	1228
EFF	1708	1598	1375	1228
FFF	1726	1522	1375	1228

GGWNGRC	GGWNGRC
GHRNGRC	GHRNGRC
GVRNGRC	GVRNGRC
GRANGRC	GRANGRC
GSFNGRC	GSFNGRC
GKGNGRC	GKGNGRC
GTRNGRC	GTRNGRC
GQWNGRC	GQWNGRC
GQRNGRC	GQRNGRC
GLGNGRC	GLGNGRC
GLKNGRC	GLKNGRC
GRNNGRC	GRNNGRC
GFGNGRC	GFGNGRC
GRRNGRC	GRRNGRC
GYNNGRC	GYNNGRC
GEHNGRC	GEHNGRC
GYKNGRC	GYKNGRC
GELNGRC	GELNGRC

Appendix Figure 7.1: 19 possible adhesion peptides from first 5,832-member library. After running the first 1536 plate with HUVEC cells to look for adhesion peptides we discovered 24 locations with possible adhesion ligand hits. These hits included 1 RGD sequence, and 19 other unique sequences. Initially it was believed that all the sequences in the left column were possible hits in some capacity. However, upon further consideration, it was noticed that all the sequences contained an 'NGR' sequence.



Appendix Figure 7.2: Isotopic splitting pattern of a NGR peptide that underwent deamidation to iso-DGR. Here the +1 peak of the NGR peptide is lower than it should be, based on isotopic distribution, as the conversion of asparagine to iso-aspartate causes an increase in mass of 1.



Appendix Figure 7.3: A representative image of an entire 1,536 SAMDI plate used for cell adhesion studies with HUVEC cells. The background of the plate has been functionalized with hexadecylphosphonic acid, which causes the cells to stick to the areas between the gold spots on the plate. The dark circles are the individual gold islands with immobilized peptides. This image was obtained on the Nikon® Eclipse Ti-E microscope with a 1X objective using NIS-Elements software to automatically take and stich multiple images into a single large image.