NORTHWESTERN UNIVERSITY

Applications of SAMDI Mass Spectrometry for High-Throughput Reaction Monitoring on Peptide Arrays

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTORY OF PHILOSOPHY

Field of Chemistry

By

Lindsey C. Szymczak

EVANSTON, ILLINOIS

September 2019

ABSTRACT

Applications of SAMDI Mass Spectrometry for High-Throughput Reaction Monitoring on Peptide Arrays

Lindsey C. Szymczak

This work combines the use of high-throughput mass spectrometry with peptide arrays for to monitor reactions on peptides. The Mrksich lab introduced a high-throughput, label-free, biochemical assay that relies on self-assembled monolayers on gold and matrix-assisted laser desorption/ionization mass spectrometry, termed SAMDI-MS. This dissertation introduces new applications of SAMDI-MS and peptide arrays and illustrates the many advantages of this technique for studying reactions on peptides.

Chapter 2 describes the use of SAMDI-MS and peptide arrays for the identification of new sequence-selective reactions of peptides. Selective reactions on peptides and proteins are important for studying protein function and preparing protein-based materials. Using this technique, we discover and characterize the sequence-selective acetylation of peptides containing histidine and tyrosine by acetic anhydride and demonstrate the advantages of using peptide arrays with a label-free analysis method to discover peptide-modifying reactions.

We then illustrate the use of SAMDI-MS and phosphorylated peptide arrays to gain a systems-level understanding of phosphatase activity in cell lysates. Phosphorylation is the most prominent

post-translational modification on proteins. While phosphorylation plays important roles in many cellular processes, understanding of the regulation of phosphorylation is limited. Kinases and phosphatases work cooperatively to regulate phosphorylation; however, research has primarily focused on kinases due to the technical challenges of studying phosphatase activity. In Chapter 3, we show that phosphatase activity is specific and may play a significant role in the regulation of global phosphorylation.

Lastly, we apply machine learning to high-throughput data acquired from SAMDI-based peptide arrays to improve the efficiency of peptide array design. One of the main challenges of high-throughput experiments is analysis of large data sets. In Chapter 4, we describe the use of machine learning to predict peptide signal-to-noise to design arrays consisting of highly detectable peptides. This allows for higher quality data collection, and therefore more reliable results for future peptide array experiments.

The use of peptide arrays has significantly increased the capabilities of protein research, and SAMDI-MS offer many advantages for the continued expansion of new applications for peptide arrays. This dissertation illustrates the many advantages of SAMDI-MS and peptide arrays for novel biochemical applications.

Acknowledgements

First and foremost, I would like to thank my advisor, Professor Milan Mrksich, for supporting me throughout my PhD research and for encouraging me to become a better thinker, problem-solver, and scientist. I would also like to thank my committee members, Professor Neil Kelleher and Professor Karl Scheidt, for their continued support and guidance throughout my time at Northwestern.

Much of my work was made possible through collaboration, and I am extremely grateful to have had the opportunity to work with Dr. Albert Xue and Professor Neda Bagheri. Our collaboration taught me the value of working with individuals with very different backgrounds. I also had the chance to collaborate with three members in the Mrksich group, Dr. Jennifer Grant, Dr. Hsin-Yu Kuo, and Daniel Sykora, and want to thank them for being great friends, teammates, and motivators.

I am very thankful to the whole Mrksich lab for being a great group of supportive colleagues and friends. I am very appreciative of the mentorship I received from Dr. James Kornacki when I joined the Mrksich lab and want to thank him for helping me transition into the group. I also want to thank Alexei Ten, Dr. Pradeep Bugga, and Dr. Patrick O'Kane for supporting me in lab and in life and for being amazing friends.

I also want to thank my friends from childhood and college, whom I love dearly, for believing in me and constantly reminding me of the world outside of graduate school—Kirsten Kohagen, Doctor Ariel Nehemiah, Jordan Dick, Dr. David Somers, Louis Ryan and Doctor Charlotte Floria.

I would not be where I am today without the love and support of my family. I am extremely and eternally grateful for the love, encouragement, support, and guidance that I have received from my parents throughout my education. I want to thank my dad, Dr. William Szymczak, for always being there for me in any situation, from helping me with math problems and creating PeptidePlot to visualize my data, to helping me move, building shelves, fixing sinks, installing ceiling fans, giving me handyman-type advice and so much more. I want to thank my mom, Dr. Camille D'Annunzio-Szymczak, for always only being a phone call away, proofreading my papers, abstracts, resume, cover letters, etc., and stocking up my freezer and refrigerator during every visit.

I am so grateful for my grandpa, who is no longer with us, for his love and support throughout my life, and I know that he would have been so excited and proud to see me receive my PhD. I want to thank my grandma, for teaching me the value of family and giving me so much love, wisdom and guidance.

I am so thankful for my brother, Andrew, and sister, Sabrina—they are the best siblings a girl could ever ask for, and I love them so very much. I also want to thank my childhood nanny, Lis, for being the best second mother ever, and for always going above and beyond for me.

I am so appreciative of parents-in-law and grandparents-in-law for welcoming me into their family as if I was one of their own and for making it easier to call Chicago home.

Last, but certainly not least, I am most grateful for my husband, Doctor Steven Pearson, for his unwavering love and encouragement. I want to thank him for celebrating my successes and for supporting and believing in me through challenging moments. He is an amazing partner and best friend, and I could not be more blessed to have had him by my side throughout my PhD.

List of Abbreviations

SAM	Self-assembled monolayer
MALDI	Matrix-assisted laser desorption/ionization
MS	Mass spectrometry
SAMDI-MS	Self-assembled monolayers for MALDI-MS
SPPS	Solid Phase Peptide Synthesis
PEG	Poly(ethylene glycol)
SPR	Surface Plasmon Resonance
PI-SAMDI	Protein Interaction SAMDI
HY Dyad/ YH Dyad	Histidine-Tyrosine dipeptide/ Tyrosine-Histidine dipeptide
S/N	Signal-to-noise
NMM	N-methyl morpholine
DMF	Dimethylformamide
TFA	Trifluoroacetic acid
AUC	Area under the curve
Ser	Serine
Thr	Threonine
Tyr	Tyrosine
ELISA	Enzyme-linked immunosorbent assay
K-array	Peptide array with sequence Ac-GRK^{ac}XZC
H-array	Peptide array with sequence Ac-GXZHGC

Table of Contents

Lists of Figures and Tables

Chapter 1. SAMDI-MS and Peptide Arrays

This dissertation combines the use of SAMDI-MS and peptide arrays for important reaction monitoring applications. This chapter will introduce both technologies, describe the advantages of combining SAMDI-MS and peptide arrays, and summarize the research applications that are described in detail in later chapters.

SAMDI-MS

Self-assembled monolayers on gold

Self-assembled monolayers (SAMs) of alkanethiolates on gold are the basis for the robust and widely applicable assays developed in the Mrksich laboratory over the last 15 years. SAMs are designed to form highly organized monolayers with controllable surface structures and compositions. SAMs of alkanethiolates are biocompatible and yield convenient and well-defined interfaces that can be functionalized to present various terminal groups for applications including wetting, nanofabrication, modelling of organic/molecular electronics, and biochemical studies.[1, 2]

The Mrksich lab utilizes monolayers that consist of low densities of alkanethiolates terminated with reactive groups for specific substrate immobilization amongst a background of alkanethiolates terminated with tri(ethylene glycol) groups that resist non-specific protein adsorption.[3, 4] Various reactive groups have been used for substrate immobilization through Diels-Alder reactions, click chemistry, biotin-streptavidin interactions, and nucleophilic addition reactions;[5-9] the Mrksich lab relies mostly on maleimide-terminated SAMs that undergo Michael addition with thiols.[10, 11] Maleimide-terminated SAMs are particularly useful for immobilization of peptides with a terminal cysteine residue as Michael addition of thiols to maleimides occurs

rapidly with high yields at low concentrations and neutral pH. Additionally, the reaction of thiols with maleimides is selective in the presence of common amino acid functional groups, which eliminates the need for post-synthetic modification before immobilization. This work utilizes maleimide-terminated SAMs to create a controlled and consistent presentation of peptide substrates for the investigation of reactions on peptides. (Figure 1-1)

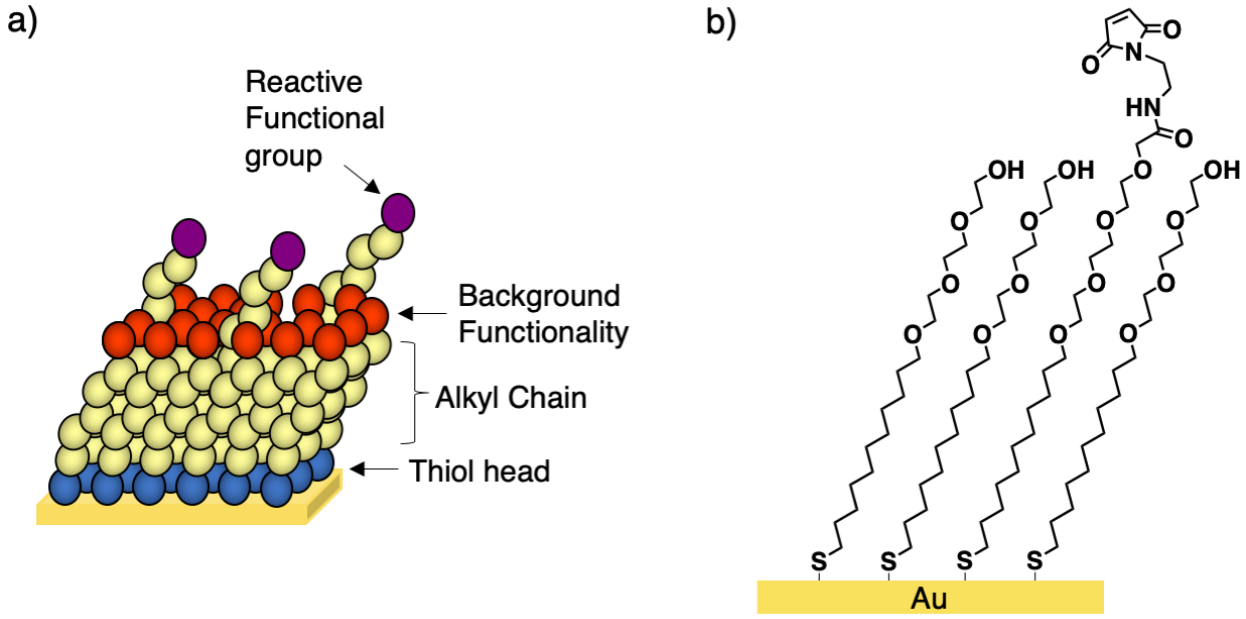


Figure 1-1 Formation of self-assembled monolayers of alkanethiolates on gold. a) Alkyldisulfides self-assemble onto gold and present various reactive functional groups and background functionalities. a) Chemical structure of SAMs presenting maleimide functional groups with a background of tri(ethylene glycol) groups. This figure is adapted with permission from Mrksich, M. (2008) Mass Spectrometry of Self-Assembled Monolayers: A New Tool for Molecular Surface Science, *ACS Nano 2*, 7-18. Copyright 2008 American Chemical Society.

Self-assembled monolayers and mass spectrometry

It was first demonstrated that monolayers of alkanethiolates on gold could be observed by laser desorption mass spectrometry by both Hanley and colleagues as well as Wilkens and colleagues.[12, 13] The Mrksich lab advanced the technique of analyzing SAMs by mass spectrometry, and introduced a method combining SAMs and matrix-assisted laser desorption/ionization (MALDI) mass spectrometry (MS), termed SAMDI-MS. In SAMDI-MS, SAMs are first coated with a matrix solution, and then laser irradiation on the SAM surface breaks the gold-thiol bond causing desorption of the monolayer as a mixture of alkanethiolates and alkyldisulfides. MALDI mass spectrometry generates singly charged cations as well as sodium and potassium adducts of the intact alkanethiolates and alkyldisulfides, which allows for simple interpretation. The SAMDI-MS technique easily resolves changes in mass of immobilized substrates, such as modifications on peptides, making SAMDI a powerful and label-free assay for biochemical reactions that often rely on fluorescent or colorimetric tags. The most common assays for studying protein and peptide modifications typically require multiple steps before final analysis, including purification and labeling, and provide non-quantitative read outs of expected results. SAMDI-MS eliminates the need for tedious sample purification post-treatment, as protein-resistant SAM surfaces can be rinsed, dried and then immediately analyzed using MALDI-MS. With MALDI-MS, both expected and unexpected products can be observed by mass shifts of the substrate. These advantages also make SAMDI adaptable for high-throughput experiments (Figure 1-2).

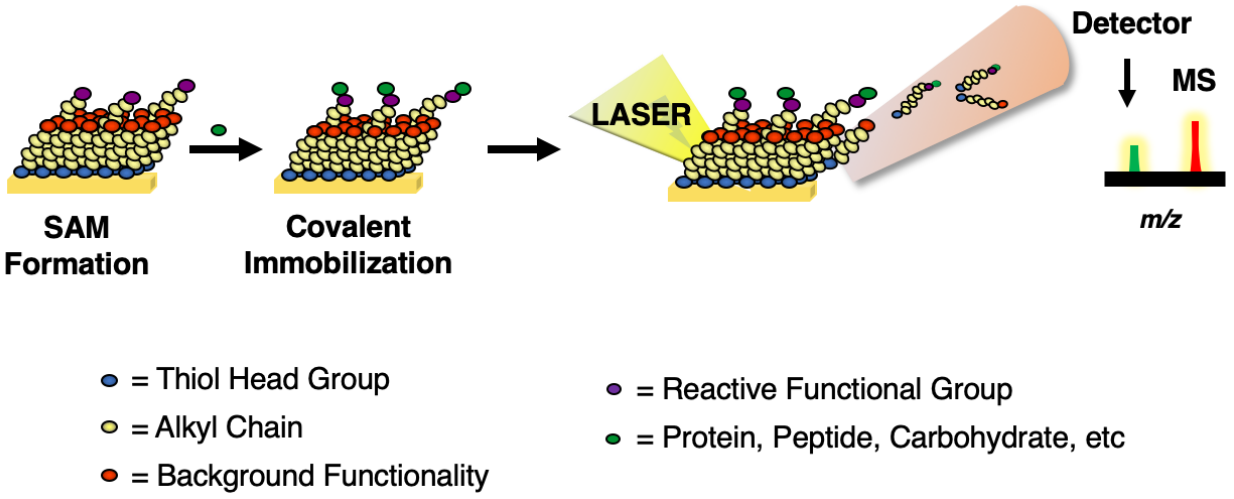


Figure 1-2. Self-assembled monolayers for matrix assisted laser desorption/ionization (SAMDI) mass spectrometry. Monolayers of alkanethiolates self-assemble onto gold and present a low density of reactive functional groups for substrate immobilization. The monolayers are analyzed by MALDI mass spectrometry, where a laser is applied to the surface and breaks the gold-thiol bonds of the alkanethiolate monolayer. The monolayer is desorbed and ionized as alkanethiolates or alkyldisulfides with minimal fragmentation, and the corresponding masses are observed in the SAMDI spectrum. This figure is reprinted with permissions from Mrksich, M. (2008) Mass Spectrometry of Self-Assembled Monolayers: A New Tool for Molecular Surface Science, *ACS Nano 2*, 7-18. Copyright 2008 American Chemical Society.

High-throughput SAMDI-MS

We have demonstrated the use of SAMDI for a broad range of biochemical and chemical assays and have translated the method to operate in high-throughput on 384 and 1,536 spot array plates, which can be analyzed in 15 or 30 minutes, respectively. Each spot on an array represents an island of monolayers on gold, and each monolayer can be treated with unique conditions to allow screening of hundreds of reactions on a single array plate (Figure 1-3). As an example, SAMDI arrays have been used to screen over 1,800 unique reaction conditions to optimize the traceless Petasis reaction.[6] The SAMDI method has also been applied to carbohydrate arrays,

where 14,820 reaction conditions were screened to identify and characterize four new glycosyltransferases.[7]

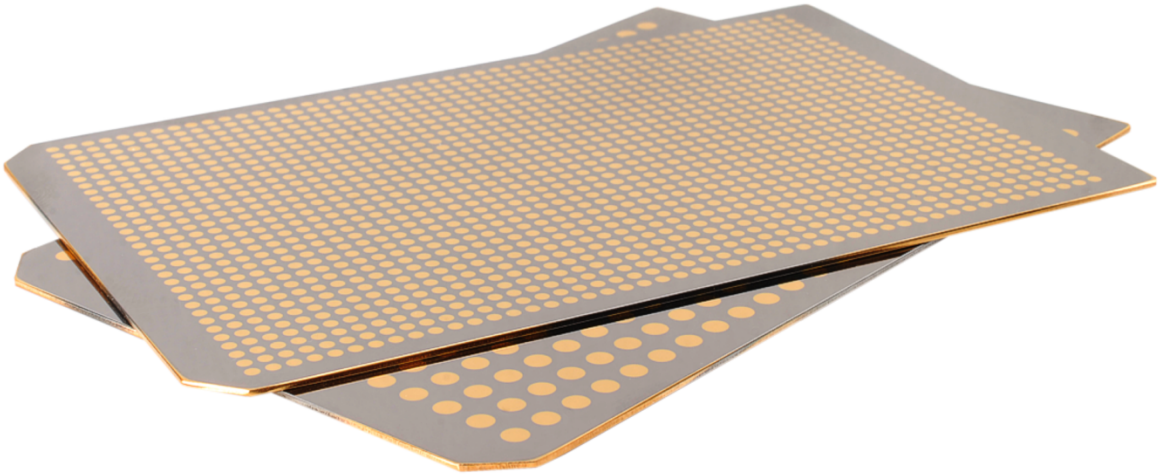


Figure 1-3. SAMDI array plates for high throughput experiments. The arrays consist of 1,536 (top) and 384 (bottom) monolayers on gold that represent unique reactions.

Another important application of SAMDI arrays has been the preparation of peptide arrays to study kinase activity,[14] substrate specificity of deacetylase enzymes,[15] and global deacetylase activities in cell lystes.[16] Peptide arrays are an exciting and rapidly growing technology with a broad range of applications in the basic and applied life sciences, and the SAMDI method offers many advantages in their design and application.

Peptide Arrays

The following descriptions and figures of peptide arrays have been adapted with permission from Szymczak, L. C., Kuo, H.-Y., and Mrksich, M. (2018) Peptide Arrays: Development and Application, *Anal. Chem. 90*, 266-282. Copyright 2018 American Chemical Society.[17] http://dx.doi.org/10.1021/acs.analchem.7b04380 Further permissions related to the material cited should be directed to the ACS.

Biomolecular Array Development

Peptide arrays have been under development for approximately twenty-five years, and commercial versions have been available for the past decade. The arrays typically comprise hundreds to thousands of distinct peptide sequences and have proven important for determining substrate specificities of enzymes, profiling antibodies, mapping epitopes, studying ligand-receptor interactions, and identifying ligands that mediate cell adhesion.[18-23] Despite this, the current applications of peptide arrays are relatively modest—particularly when compared to oligonucleotide arrays—and have not been adapted as a standard method in laboratories and throughout the drug discovery process.

Development of biomolecular arrays—defined as planar substrates having large numbers of molecules immobilized in patterns, where each region of the surface presents one specific molecule—began in the early 1990s with oligonucleotide arrays. While the majority of current biomolecular arrays are prepared with either oligonucleotides, peptides, proteins, or small molecules, oligonucleotide arrays are the most developed. Oligonucleotide array development

began through efforts led by Patrick Brown, Stephen Fodor, and Edwin Southern.[24-27] These researchers used different approaches—based either on the immobilization of pre-synthesized DNA or the *in situ* synthesis of DNA directly on the substrate—and by 1995 were using the arrays to profile gene expression.[27] In 1996, DNA arrays were used to profile the expression of as many as 1000 genes[28] and soon after, the first whole-genome microarray of yeast was reported[29] and a new high-density random array was presented.[30] Now, twenty-five years later, commercially available arrays contain several million oligonucleotides and are routinely used tools in the laboratory.[31, 32] Oligonucleotide arrays have been used in clinical applications to study viruses, human disease, genetic screening, and personalized medicine.[33, 34] The rapid pace of technical advancement of the poorly performing early arrays has led to the robust and inexpensive arrays currently in use. By analogy, peptide arrays are still in the early stage of development and can be expected to see improvements that result in them being an important, if somewhat less ubiquitous, technology for the life sciences.

The development of peptide, protein, and carbohydrate arrays has been motivated in part by the incomplete knowledge provided by genomic information alone. This is because gene expression and mRNA levels do not solely correlate with protein activity, and several additional factors—including post-translational modifications, alternative splicing, allosteric ligands, co-localization, and degradation—serve to regulate protein activity. A systematic understanding of the roles of the approximately 20,000 genes[35] in the human genome—which code for approximately 100,000 transcriptomes[36] and give rise to the expression of more than 1 million

proteoforms[37, 38]—requires systems level information on proteomic activities. Peptide arrays offer one route to this end.

Early work by the Schreiber[39] and Snyder[40] groups has demonstrated the use of protein arrays and their applications to identify protein-protein interactions, substrates for enzymes, and protein targets of small molecules. However, in practice, protein arrays have not had a broader impact because of the inherent challenges in preparing protein arrays, including expressing and purifying large numbers of proteins, immobilizing proteins with control over the orientation, increasing feature density, and maintaining the activities of the immobilized proteins (and preventing their denaturation at the interface). These challenges are avoided when working with arrays of peptides—which are relatively easy to synthesize, chemically stable, and compatible with many immobilization chemistries. While many protein functions cannot be recapitulated at the peptide level due to lack of tertiary structure and sequence truncation, the use of peptides is still appropriate for many applications.

Peptide arrays offer an enormous opportunity to further understand the molecular pathways that underlie normal and pathological functions in cells, guide the drug discovery process, and diagnose and monitor treatment of disease. They also allow for a variety of other studies to understand sequence-dependent reactivity and properties of peptides. Among the early applications for peptide arrays are 1) epitope mapping for antibody binding,[41-45] 2) identifying and characterizing binding interactions between protein and peptide ligands,[46-49] 3) screening for active substrates of enzymes,[50-55] 4) profiling enzyme activity in complex samples,[16, 56, 57] and 5) identifying and studying peptides that mediate cell adhesion.[58-62]

In comparison to DNA arrays, peptide arrays are more challenging to develop. First, the assays performed on DNA arrays are essentially all based on hybridization with fluorescent probes, whereas peptide arrays are used in a wide variety of assay and detection formats. These many applications make it difficult to identify materials and surface chemistries that are optimized for this broad range of applications. Moreover, peptide arrays are often used to assay samples that have high concentrations of protein, and non-specific adsorption of proteins often leads to false positive and negative results. The use of 'inert' surface chemistries can control these unwanted interactions but are still uncommon in peptide arrays. Further, peptides are more heterogeneous in their chemistry than are oligonucleotides. Peptides have greater functional group diversity in side chains, are synthesized in slightly lower yield than oligonucleotides, and have a wider distribution in properties including solubility, stability, and aggregation. Despite the impressive progress made in peptide array research and commercialization over the past 20 years, development has lagged behind that of DNA arrays.

Technical Factors

Peptide arrays are composed of a large number of peptides spatially arranged in an addressable format on a solid support. The array format has the benefit of allowing many experiments to be performed on a single sample. The specific applications for a peptide array will depend on the number of peptides in the array, the compatibility of the array material with samples (for example, those that are not inert to protein adsorption), the control over peptide attachment and density, and the compatibility with different detection methods. Several technical considerations are important in preparing peptide arrays, and among the various methods that have been reported, there are

substantial differences in approach. The reported approaches vary in their use of the different strategies, including the choice of peptide synthesis methodology, solid support and functionalization, immobilization method, patterning strategy, and detection method (Figure 1-4).

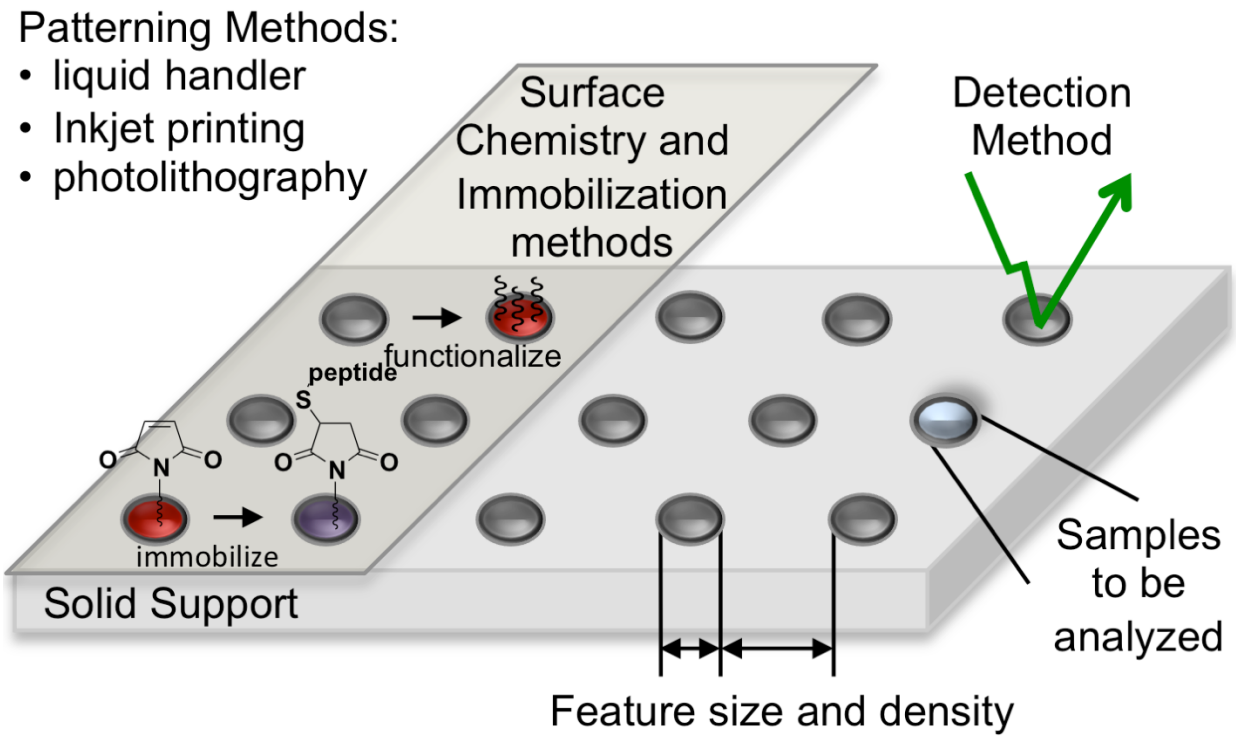


Figure 1-4. Technical components of peptide arrays.

Peptide Synthesis. There are two widely used methods for synthesizing peptides present in arrays: Merrifield solid-phase synthesis and *in situ* synthesis.[63-68] Merrifield solid-phase peptide synthesis (SPPS) is the traditional method for synthesizing peptides and involves the assembly of peptides on a solid support, followed by deprotection of the side chains and cleavage from the support, purification if necessary, and finally immobilization onto an array support.[69] Peptides

prepared in this way are generally of high quality, having fewer impurities resulting from incomplete synthesis. This benefit of SPPS, however, comes with substantial expense and time associated with the synthesis of hundreds or thousands of distinct sequences. This drawback in part motivated the development of *in situ* peptide synthesis, which now uses automated instrumentation for parallel synthesis of the peptides directly on the solid support of the array.[70, 71] *In situ* peptide synthesis has the benefit of using minimal amounts of reagents and eliminates the need or possibility for peptide purification. This approach yields substantial benefits in cost and time requirements to prepare the arrays; however, the purity and quality of peptides in the arrays are difficult to verify.

Solid support and surface functionalization. The choice of material used as the solid support and the surface chemistries for immobilization are critical determinants of how well the array will perform. For example, the availability of well-defined surface chemistries leads to better control over the density of peptide and ensures a regular microenvironment, which allows for the direct comparison of measured activities of different peptides in the array. The solid support is also an important choice when considering the use of different detection methods that have special requirements for the properties of the support. For instance, surface plasmon resonance spectroscopy requires a noble metal film of a specific thickness, and mass spectrometry requires a conductive support. Common choices for solid support and surface functionalization include glass or silicone with polymer surface coatings, nitrocellulose or polyvinylidene difluoride (PVDF) membranes with amino derivatization, and gold on glass or metal with self-assembled monolayers (as used with SAMDI-MS). An often-overlooked problem with the surfaces used in peptide arrays

is that most proteins will adsorb to them non-specifically. In fact, most proteins will adsorb rapidly and strongly to most artificial surfaces. The adsorbed proteins can then contribute to false positive results or obstruct interactions of the immobilized peptide and give false negative results. A common strategy to address this issue is to use blocking proteins that adsorb to the solid support in order to prevent further adsorption of proteins from the sample. However, blocking proteins can likewise obstruct interactions with the immobilized peptides to give false-negative results. A more effective solution is to use surfaces that are functionalized to prevent the non-specific adsorption of protein. Among the most effective options are self-assembled monolayers (SAMs) terminated in oligo(ethylene glycol) groups and poly(ethylene glycol) methacrylate (PEGMA).[72-74] These protein-resistant surfaces have been critical for the development of quantitative assays that are compatible with challenging samples, including complex cell lysates and tissue extracts. The choice of support limits the immobilization strategies, peptide densities, and the detection method available for analyzing the array in an experiment.

Immobilization methods. For the preparation of arrays from pre-synthesized peptides, there are three general methods used to immobilize peptides: physical, chemical, and biological immobilization (Figure 1-5). The simplest way of attaching peptides to solid supports relies on physical adsorption. Most peptides of a suitable length or that contain a complementary property to the support will adsorb rapidly. For example, hydrophobic amino acids will adsorb to a hydrophobic material and acidic amino acids will adsorb onto a positively charged surface. This approach is analogous to the immobilization of proteins in ELISA and Western blotting applications that have been used for many years.

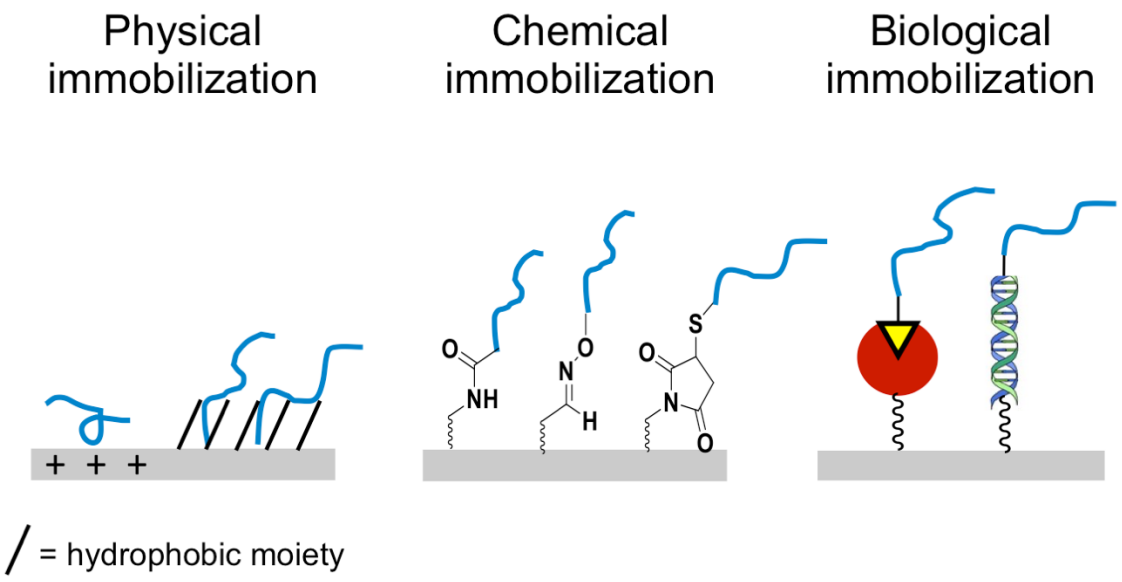


Figure 1-5. Immobilization methods for peptide arrays. Peptides can be immobilized onto array surfaces through physical adsorption, chemical reaction, or biological interactions.

Fabricating peptide arrays by physical adsorption is simple but presents several obvious drawbacks as discussed earlier, including the need for blocking non-specific adsorption sites that can disrupt interactions between proteins of interest and the immobilized peptides. Peptides attached through non-specific adsorption may not be stable, in that the adsorbed peptide can be displaced by proteins in the sample, with the added concern that not all peptides will be subject to replacement at the same rate. It is also difficult to control the densities of the immobilized peptides, which will affect the level of signal that is measured in an assay. Finally, peptides presented on the solid support via physical adsorption are not oriented in a consistent way, and a fraction of the immobilized peptides may not be bioactive in the intended assay, which can lead to substantial variation in otherwise identical arrays.

The most common strategies use selective or non-selective chemical reactions to covalently attach peptides to the support (Figure 1-6). These approaches usually require chemical modification of the surface to install the relevant functional groups for attachment and have the benefit of having the peptides covalently attached without risk of dissociating from the surface during an assay. Chemistries used in the immobilization of large numbers of peptides in an array should be robust, high yielding, and general with respect to the peptide sequence.

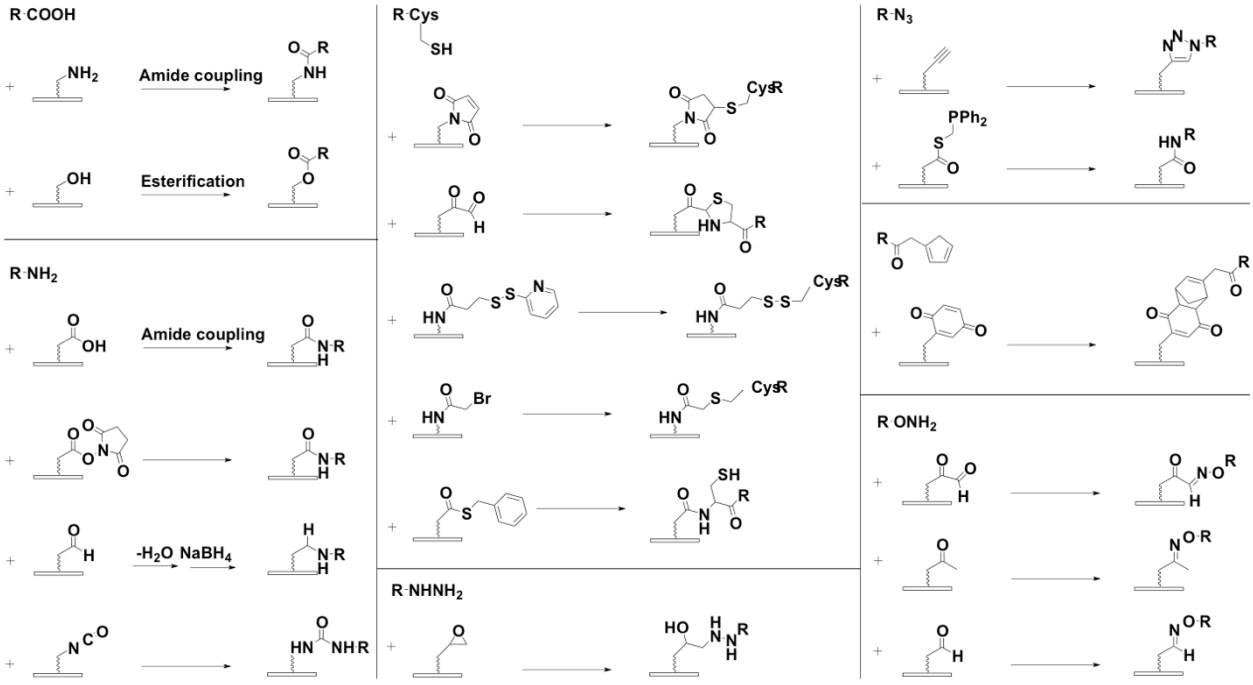


Figure 1-6. Common chemical reactions used to immobilize peptides to materials.

A broad variety of reactions have been used to immobilize peptides, with many using the nucleophilic α-amino group that peptides naturally possess to react with a carboxylate group on the support.[70,75] Other methods utilize aldehyde groups that provide an activated carbonyl for

reaction with primary amines on peptides to form Schiff base linkages.[76] These immobilization schemes do not require additional chemical modification of the peptide reagents, allowing simple routine implementation; however, when more than one of these nucleophilic side chains are present within a peptide, it is difficult to control the regiochemical point of attachment.

The selective reaction of thiols with several electrophilic groups has made the use of cysteine-terminated peptides an important method.[10, 77, 78] Other nucleophilic, bioorthogonal functional groups incorporated into a peptide can also be utilized for chemo-selective immobilization, such as azides which react with alkynes and thioesters, and hydroxylamines and hydrazines which react with aldehydes and ketones.[77, 79-85]

Peptides can also be immobilized using biological strategies, either based on ligand-receptor interactions or enzyme-mediated reactions. Examples of biological strategies for presenting peptide arrays include the use of biotin tags with avidin or streptavidin,[86-89] peptide-oligonucleotide conjugates for hybridization of two complementary oligonucleotides,[90, 91] phage display,[92, 93] yeast surface display,[94] ribosome display,[95] and polysome display.[96]

Patterning Methods. The preparation of a peptide array necessarily requires patterning methods that can direct the immobilization or *in situ* synthesis of each peptide at its designated region of the support. The many different approaches offer contrasting benefits in terms of peptide spot density, speed, cost, reagent compatibility, and array quality. Pre-synthesized peptides are typically patterned onto a functionalized surface using a robotic liquid handling system. The density of peptide spots prepared using this method depends on the minimum dispensing capacity of the robotic liquid handler, the hydrophobicity of the surface (which will prevent spreading of

the peptide solution), and solvent evaporation (rapid evaporation leads to incomplete immobilization). Peptide arrays made with the *in situ* synthesis approaches are patterned by the method used to deposit reagents onto the support. Spot sizes are determined by the dispensed volume of reagent and the physical properties of both the membrane support and the solvents.

Detection Methods. The vast majority of peptide arrays are analyzed with label-dependent assays, though more recent work has developed practical strategies for using 'label-free' methods. The labeled assays are typically rapid and convenient to perform, but always risk interference giving rise to false positive (and negative) results. To take one example, fluorescent labels were responsible for the incorrect finding that resveratrol was a sirtuin activator.[97] The most popular assays used for peptide array studies are based on radioactivity, chemiluminescence, colorimetry, and fluorescence.

The use of radioactive labels is always discouraged for reasons of safety, complicated waste disposal, and expense. However, they remain important in assays that involve the transfer of molecules to the peptides, such as ATP ([γ-$^{32/33}$P]ATP) used in kinase assays[98, 99] and SAM ([methyl-^{3}H]-*S*-adenosyl-L-methionine) used in methyltransferase assays.[100]

Chemiluminescent, colorimetric, and fluorescent detection methods are preferred when antibodies are used to analyze arrays after an experiment. These methods can be tedious as they require multiple blocking, washing, and incubation steps before the array is treated with a primary antibody that recognizes and binds to the reacted peptides. For chemiluminescent and colorimetric detection, the arrays are often treated with a secondary antibody that is conjugated to a reporter enzyme, such as alkaline phosphatase or horseradish peroxidase, to catalyze a

chemiluminescent[101, 102] or colorimetric[103-105] reaction. The arrays are treated with substrates that result in colorimetric changes or chemiluminescence that can be measured using standard equipment. For fluorescent detection, the secondary antibody is fluorescently labeled, and no further steps are required to measure peptide array reactivity.[86] Secondary antibodies are not required if antibodies targeting the desired peptide product with the desired conjugated enzyme or fluorescent label are available. Antibody dependent methods usually give relative rather than absolute quantification of results and can be limited by the availability of antibodies and cross-reactivity.

More recent work has introduced label-free methods to analyze biomolecular arrays, with surface plasmon resonance (SPR) spectroscopy and mass spectrometry (MS) emerging as the most significant. The label-free methods have several advantages: they avoid artifacts stemming from labels, they can be quite general in measuring a broad range of activities, and they have the potential to identify unanticipated activities that would not have been observed with a particular labeling strategy.

SPR is an electromagnetic technique used to measure the interaction of a soluble protein with an immobilized molecule, such as a peptide, in real-time. SPR effectively measures changes in the refractive index of a solution within approximately 100 nm of a metal interface.[106] In SPR, polarized light is reflected from the backside of a metalized glass slide, where it excites a collective motion of electrons in the metal giving rise to an electric field that decays in the solution. The decay characteristics depend on the local refractive index and establish a resonant condition with a particular angle of incident light. This angle can be measured by observing a dip in the intensity

of reflected light, and the angle at which this minimum reflectivity occurs shifts with the refractive index. When soluble proteins bind to ligands that are immobilized at the metal film, there is an increase in the local refractive index that is measured in real-time by monitoring shifts in the angle of minimum intensity of the reflected light. The '*in situ*' nature of the measurement makes SPR important for measuring rates of biomolecular interactions and gives both kinetic and thermodynamic information on binding.[107, 108] While SPR is a label-free, sensitive, and quantitative method that can incorporate many surface chemistries, it requires specialized instrumentation and has a lower throughput than most label-dependent detection methods or mass spectrometry. In addition, SPR is compatible only with peptide arrays that can be prepared on gold (or other metallic films); peptide arrays on cellulose membranes cannot be analyzed using SPR.

Mass spectrometry (MS) is an important method for characterizing peptides and proteins and for identifying post-translational modifications that result in a change in mass.[109] MS methods have the important benefit of providing molecular-level information on peptides and activities since an observed mass change is often consistent with one type of known modification. However, the sample preparation required in MS is usually tedious as it requires removal of salts and other species, often by HPLC or C18 spin columns. Further, these techniques are usually used to process soluble analytes and have not been compatible with arrays. However, we introduced SAMDI-MS to analyze SAMs of alkanethiolates, which offers a mass spectrometry-based technique for high throughput analysis of peptide arrays.[10, 110-112] As described in the previous section, in SAMDI, laser irradiation of the monolayer results in release of the alkanethiolates (or corresponding disulfides) and provides the mass for each peptide-alkanethiolate conjugate in the monolayer

(Figure 1-2). SAMDI is therefore able to identify mass changes—and the associated specific post-translational modification—of peptides modified by an enzyme.

Current Approaches

While there are many technical approaches that are considered in the development of peptide arrays, the current commercially available arrays can be grouped into five classes, which are compared in Table 1-1.

Merrifield synthesis peptide arrays are prepared through immobilization of pre-synthesized peptides onto functionalized glass surfaces.[69, 70, 113-115] As described earlier, there are a variety of methods for functionalizing the glass support and immobilizing the peptides, but glass surfaces with a polymer coating functionalized to capture cysteine-terminated peptides, such as glyoxylyl, are among the most common.[77] Merrifield synthesis peptide arrays have a maximum reported density of 318 spots/cm^2 and can be analyzed with radioactive, chemiluminescent, colorimetric, and fluorescent labels.

SPOT peptide arrays are among the most used.[46, 66, 67, 116-120] The peptides are typically synthesized directly on an amino-functionalized nitrocellulose membrane either using liquid handlers or inkjet printers. With liquid handlers, spot densities of 5 spots/cm^2 have been reported. Higher densities have been achieved through CelluSpot technology, which isolates the peptide-cellulose resin after SPOT synthesis and then spots solutions of the conjugated peptide onto a glass surface, increasing the density by nearly ten-fold with a spot size of ~1 mm. SPOT peptide array density can be significantly increased with the use of inkjet printers to achieve densities of 4,000

spots/cm^2. The SPOT peptide arrays can be analyzed with several radioactive, chemiluminescent, colorimetric, and fluorescent detection methods.

Another type of peptide array commonly used is based on the SPOT arrays, but utilizes laser printers to deliver reagents that are embedded in solid particles.[41, 121-126] The *particle-based peptide arrays* are usually prepared on amino-derivatized polymer-coated glass surfaces. In 2007, Breitling and coworkers reported a maximum density of 40,000 spots/cm^2.[123] The particle-based arrays can consist of up to 60,000 peptides with a spot size of approximately 250 μm, densities of 1,033 spots/cm^2, and can be analyzed with a variety of label-dependent detection methods.

Photolithographic peptide arrays are prepared by using light to direct the synthesis of peptides directly onto functionalized glass or silicon supports.[26, 42, 127-131] The surfaces are usually modified with an amino-derivatized polymer for amino acid immobilization. Peptides can be synthesized either using photolabile-protecting groups on the amino acids or Boc-protected amino acids that are removed on excitation of a photoacid. Light directed patterning can be performed with either a series of physical masks or digitally with a micromirror array. The maximum density reported for mask-based patterning is 694,000 spots/cm^2 and requires approximately 30 hours for preparation. Digital patterning in the mask-less format with micromirrors can achieve densities of 1,000,000 spots/cm^2.

SAMDI peptide arrays are prepared by immobilizing pre-synthesized peptides onto self-assembled monolayers of alkanethiolates on gold-coated glass or metal plates.[132] Peptides are most commonly immobilized by reaction of a terminal cysteine residue with a maleimide group.[10] Many other strategies are also compatible with the use of monolayer surface chemistries, including

biological immobilization through biotin-streptavidin.[5-7] The monolayers are terminated in a tri(ethylene glycol) group that serves to prevent non-specific protein adsorption. This property is critical for reducing false positive and negative results and is compatible with cell lysates and other samples that contain high concentrations of protein. SAMDI arrays can also be analyzed by MALDI mass spectrometry to provide quantitative information on biochemical activities; a SAMDI plate of 1536 peptides (corresponding to a density of 16 spots/cm^2) can be analyzed in less than one hour.

Of the five important approaches that are now practiced and commercially available, it is clear that no single approach offers advantages for all applications. Instead, the variations of peptide arrays enable the study of a wide range of applications. For example, the approaches based on *in situ* synthesis can prepare the highest density arrays and do so at relatively low reagent costs. The approaches that prepare arrays by immobilizing pre-synthesized peptides have the primary benefits of compatibility with a broader range of surface chemistries that may not be sufficiently stable towards the conditions for *in situ* synthesis. The use of self-assembled monolayers brings the important benefits of chemistries that prevent non-specific adsorption of protein, give excellent control over the densities of peptides in the array, and are uniquely compatible with the use of SAMDI mass spectrometry to perform a broad range of label-free assays.

Table 1-1. Current Methods for Preparing and Analyzing Peptide Arrays

		Merrifield Synthesis	**SPOT**		**SAMDI**	**Photolithography**		**Particle-Based**
Technical components	**Peptide Synthesis**	SPPS	*in situ*		SPPS	*in situ*		*in situ*
	Solid Support	glass	nitrocellulose or PVDF membranes		gold on glass/metal	functionalized glass or silicon		functionalized glass
	Surface Chemistry	several methods available- eg. polymer surface coating	amino derivatization, typically with B-alanine or PEG		Self-assembled monolayers- 10% maleimide terminated	amino-derivatized polymer		amino-derivatized polymer (PEGMA)
	Inert layer	Surface blocker with BSA or other blocking agents	Membranes blocked with BSA or other blocking agents		Self-assembled monolayers- 90% tri(ethylene) glycol	Surfaces blocked with BSA or other blocking agents		Surfaces blocked with BSA or other blocking agents
	Immobilization	Several methods- eg. cysteine-terminated peptide on glyoxylyl, bromomethylketone or disulfide surfaces	Coupling of carboxyl group to amino-functionalized surface		Cysteine terminated peptides on maleimide surfaces	Coupling of carboxyl group to amino functionalized surface		Coupling of carboxyl group to amino functionalized surface
	Patterning Method	Liquid Handler	Liquid Handler	Ink-jet printer	Liquid Handler	Masks	Micromirrors/ digital patterning	Laser Printer
	Maximum Feature Density	318 spots/cm^2 (pepscan)	5 spots/cm^2 (intavis)	4000 spots/cm^2 (ArrayJet)	16 spots/cm^2	83,000 spots/cm^2 (Legutki *et al* 2014)	1,000,000 spots/cm^2 (N-Schaffer)	40,000 spots/cm^2 (Beyer *et al* 2007)
	Minimum Feature Size	250 µm (pepscan)	2 - 3 mm (intavis)	90 µm (Arrayjet)	1.25 mm	8 µm	10 x 10 µm	50 µm
Detection methods	**Radioactivity**	x	x		x	x		x
	Chemiluminescence	x	x		x	x		x
	Colorimetry	x	x		x	x		x
	Fluorescence	x	x		x	x		x
	Mass Spectrometry				x			
	SPR	x			x			
	Commercialization	PEPSCAN, New England Peptide, Bio-Synthesis	JPT, Intavis, Arrayit, MultiSyn Tech, Kinexus, Jerini Bio Tools GmbH, Genscript	Arrayjet	SAMDI Tech Inc	---	LC Sciences, Schafer-N	PEPperPRINT

SAMDI-based Peptide Array Applications

SAMDI-based peptide arrays offer many advantages for studying peptide modifications and interactions and have been used for a variety of applications. For example, we recently described a variation of the SAMDI method, termed PI-SAMDI (for Protein Interaction SAMDI), that can characterize weak protein-peptide interactions in an array format.[48] When using peptide arrays to identify ligands for receptor binding, it is necessary that the peptide binds with sufficient affinity, or more precisely, with a sufficiently slow dissociation rate constant so the complex is stable through the washing steps required before imaging of the array. In practice, it is challenging to characterize binding interactions with association constants in the low μM range. Many important regulatory interactions in the cell fall in this range and therefore have not been addressable with peptide arrays. This technique uses SAMs that present an immobilized peptide ligand as well as a peptide substrate for an enzyme. The receptor protein of interest is prepared as a fusion to a reporter enzyme. As the receptor binds to the immobilized ligand, the enzyme is brought to the interface where it can then modify its substrate peptide with up to a twenty-fold greater rate. This results in a 'covalent record' of the binding interaction, and the affinity of the complex corresponds to the amount of product on the monolayer which can be quantified with SAMDI mass spectrometry. PI-SAMDI retains the advantages of a SAMDI assay—quantitative, label-free, and high throughput with inert surfaces to prevent non-specific protein adsorption—and also has the ability to study low-affinity interactions. PI-SAMDI was applied to study the binding of chromodomain proteins to methylated peptides found in the histone amino-terminal tails.

Additionally, our group has used SAMDI peptide arrays to study substrate specificity of enzymes. One example used a SAMDI peptide array to identify active substrates of the sirtuin deacetylase, SIRT3.[53] The active substrate was used to screen SIRT3 inhibitors, and a total of 306 inhibitors were identified from a library of 100,000 molecules. In another example, we used SAMDI peptide arrays to discover an example where one post-translational modification of a histone-derived peptide regulates a second post-translational modification—also known as histone crosstalk.[54] We found that acetylation of histone 3 at lysine 14 by the acetyltransferase PCAF is inhibited when the arginine residue at position 8 is first enzymatically methylated or deaminated. These examples highlight the application of peptide arrays for characterizing the specificity of an enzyme, as it is important to identify sequences that are both active and inactive for a given enzyme.

The monolayer surface chemistries of the SAMDI arrays are among the best available for preventing the unwanted adsorption of protein; therefore, the arrays can be used to profile enzyme activities in complex samples derived from cell lysates and tissue extracts. Our group used SAMDI peptide arrays to profile the sirtuin and KDAC deacetylase families in cell lysates.[16] These studies found that while KDAC activity remained fairly constant, sirtuin activity was significantly decreased during differentiation of CHRF megakaryocytic cells. Peptide arrays have expanded the capabilities of protein research, allowing for a broad range of applications. SAMDI-based arrays offer promise for continuing to study peptide chemistries and interactions.

Dissertation Overview

This dissertation utilizes SAMDI-MS and peptide arrays for several applications. The first application uses SAMDI-based peptide arrays for the discovery and characterization of new and unanticipated sequence-selective reactions of peptides. Reactions that can selectively modify amino acid sequences within peptides and proteins are important for preparing protein reagents, immobilizing proteins, and making antibody-drug conjugates. The development of new reactions often begins with known chemistries and are optimized using only a small set of peptide reactants. Chapter 2 reports the selective acetylation of HY and YH dyads when treated with acetic anhydride in aqueous conditions and illustrates the benefits of using peptide arrays and a label-free analysis method to discover peptide-modifying reactions.

The second application uses SAMDI-MS and three phosphorylated peptide arrays to profile global phosphatase activity in cell lysates. Phosphorylation is an important post-translational modification on proteins involved in many cellular processes; however, understanding of the regulation and mechanisms of global phosphorylation remains limited. Chapter 3 describes the use of SAMDI-based peptide arrays to gain a systems-level understanding of phosphatase activity across five different cell types, and our results show that phosphatase activity may play a much larger role in the regulation of global phosphorylation on proteins than previously believed.

The third application utilizes machine learning with SAMDI-based peptide array data to improve the efficiency of peptide array design. While peptide array development and applications are still at an early stage, it is already becoming difficult to analyze the large data sets that are generated from peptide arrays. As arrays increase in size, it will be more important to develop and

apply complex analytical methods. Chapter 4 describes how machine learning can accurately predict the signal-to-noise (S/N) of a peptide in an array, allowing for the efficient design of arrays through selection of high S/N peptides.

The dissertation concludes with a summary of the described research and a discussion about further research directions for advancing the applications of SAMDI-based peptide arrays.

Chapter 2. Using Peptide Arrays to Discover the Sequence-Specific Acetylation of the Histidine-Tyrosine Dyad

The research and all figures presented in this chapter have been published and are adapted with permission from Szymczak, L. C., and Mrksich, M. (2019) Using Peptide Arrays to Discover the Sequence-Specific Acetylation of the Histidine-Tyrosine Dyad, *Biochemistry 58*, 1810-1817.
https://doi.org/10.1021/acs.biochem.9b00022

Introduction

Reactions of chemical reagents with peptides and proteins are common and remain important for fundamental studies of protein function,[134] developing probes for *in vivo* imaging,[135, 136] introducing motifs for purification,[137] and developing biotherapeutics[138, 139] and protein-based materials.[140, 141] Most current strategies of chemical protein modification target natural amino acids, [142-145] the N- or C- terminus,[146, 147] and native peptide motifs,[137, 148, 149] though others have relied on the incorporation of non-natural amino acids.[150-157] While important research continues to develop new chemistries for the specific reactions on peptides and proteins, current methods typically validate new chemistries on only a small number of peptide sequences, which limits full reaction characterization. Additional challenges include the limited number of functional groups on amino acids as well as the requirement that reactions are compatible with biologically-relevant

conditions. To improve upon current methods to develop new chemical modifications on peptides, we introduce the use of peptide arrays and self-assembled monolayers for matrix-assisted laser desorption/ionization (SAMDI) mass spectrometry (MS) as a method for discovering new acylation reactions on peptides.

Peptide arrays offer great value for the discovery of sequence-specific reactions—a single array can be used to assess reactivity of a reagent with hundreds to thousands of peptide sequences.[17] While many protein functions cannot be translated to peptides because of sequence truncation and the lack of tertiary structure, peptide arrays are very relevant and important for the study of sequence-specific chemical reactivity. The use of a peptide array containing all possible sequences in a set of variable positions allows for an unbiased assessment of specificity that advances the development and validation of new methods for the modification of proteins.

We prepared peptide arrays using SAMDI-MS, a high throughput, label-free technique that provides a quantitative measure of reaction without the need for purification or tedious sample preparation.[112, 132] SAMDI-MS is a technique that relies on self-assembled monolayers (SAMs) of alkanethiolates on gold that present functional maleimide groups for immobilization of thiol-derivatized molecules, including cysteine-terminated peptides, at a density of 10% against a background of tri(ethylene glycol) groups.[112, 132] The functionalized SAMs are treated with reagents of interest and then analyzed using matrix-assisted laser desorption/ionization (MALDI) mass spectrometry to reveal new peaks that correspond to reaction products (Figure 2-1).

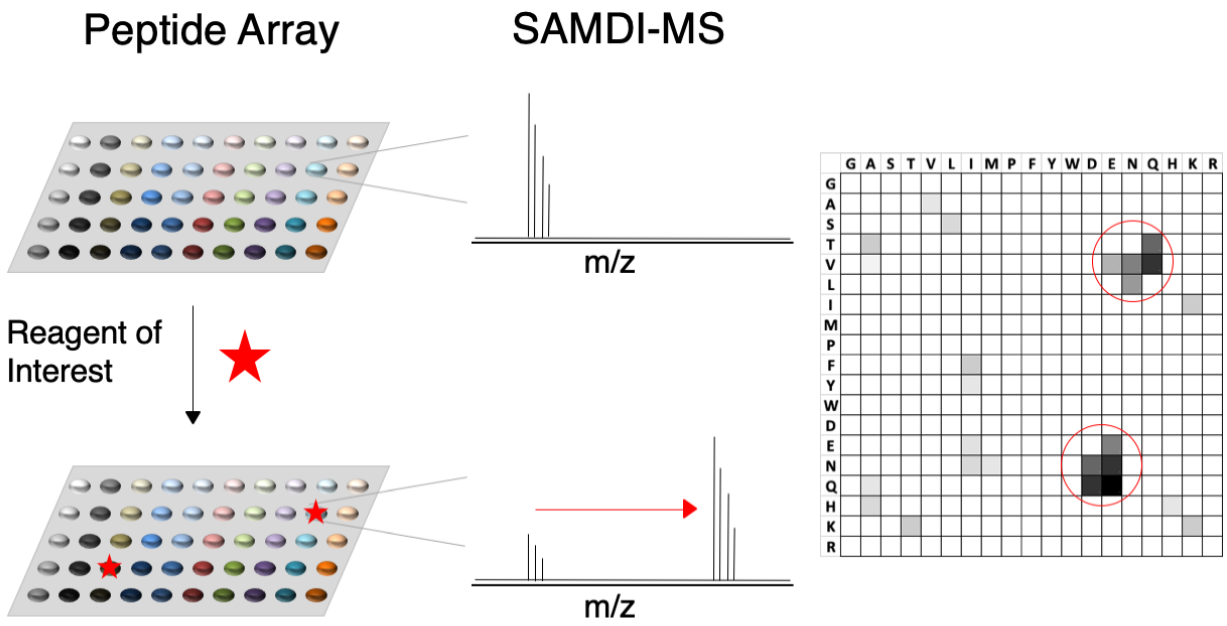


Figure 2-1. SAMDI-based peptide arrays for the discovery of peptide-modifying reactions and characterization their sequence-specificity. SAMDI based peptide arrays are treated with a reagent of interest, and then peptides are analyzed using MALDI mass spectrometry to reveal new peaks corresponding to reaction products.

In comparison to current methods for the development of new peptide chemistries, the combination of peptide arrays and SAMDI-MS offers several advantages, including the abilities to quickly screen hundreds of peptides on a single array and identify unanticipated reaction products. We recently showed the value of SAMDI-MS based peptide arrays for glycosylation sequence characterization and optimization of substrates for various glycosyltransferases.[158] In this chapter, we describe the use of two peptide arrays with SAMDI-MS to discover and characterize the sequence-selective reaction of acetic anhydride with the histidine-tyrosine dyad. We found that acetic anhydride has preferential reactivity with peptides containing both histidine and tyrosine,

and we investigate the mechanism behind this selective reactivity. We show that this reaction proceeds with initial acetylation on histidine, followed by acetyl transfer to tyrosine. We further characterize this reactivity with a second peptide library to examine the effects of amino acids neighboring the HY dyad. Finally, we examine the reactivities of two additional acylating electrophiles, as well as the spatial effects of this acetyl transfer. The combination of peptide arrays and SAMDI-MS provides a powerful platform for reaction discovery on peptides.

Results and Discussion

Experimental design. Acetic anhydride is known to acetylate amines, including the side chain of lysine residues, under basic conditions. We asked whether there were certain peptide sequences that would react with the acylating reagent under neutral conditions, and therefore aimed to map the reactivity of acetic anhydride with a peptide array encompassing all combinations of amino acid dyads. We synthesized a peptide library consisting of 361 peptides with sequences of Ac-GXZRGC, where X and Z are variable and represent all 19 natural amino acids excluding cysteine. The peptides were capped at their N-termini to prevent reaction at this site, and also included an arginine to increase ionization efficiency for MALDI-MS. The cysteine residue provides the thiol functional group for immobilization to SAMs that present the corresponding maleimide group. We prepared a peptide array by immobilizing the peptides onto an array plate having 384 self-assembled monolayers on gold presenting maleimide groups at a density of 10% against a background of tri(ethylene glycol) groups, as described previously.[10] We then treated the peptide array with an aqueous solution of acetic anhydride (100 mM) for 1 hour at room temperature, after

which the plate was rinsed, treated with matrix and analyzed by SAMDI-MS to provide a mass spectrum for each spot (Figure 2-2).

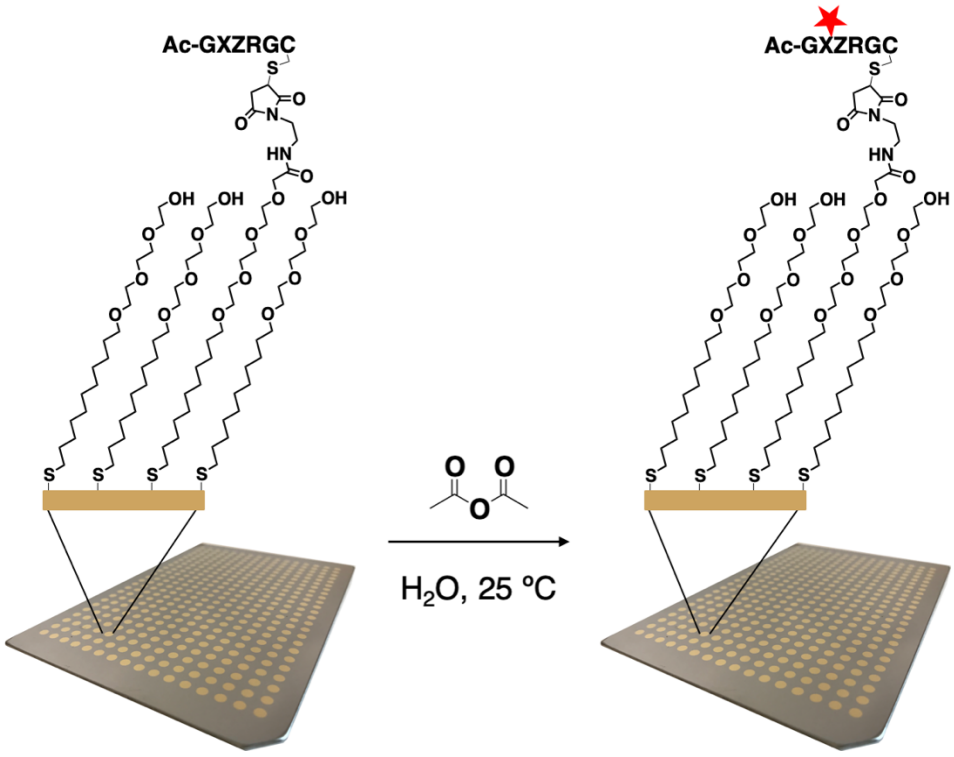


Figure 2-2. A SAMDI based peptide array, consisting of the sequence Ac-GXZRGC, is treated with acetic anhydride and analyzed. The X and Z positions are variable and represent all natural amino acids except for cysteine. This array allows for full reaction characterization because all existing dipeptides are screened.

Selective reactivity of acetic anhydride. The mass spectra showed that the masses of the majority of the peptides were unchanged following treatment with acetic anhydride, and therefore did not react with the anhydride. Surprisingly, only two peptides in the array underwent efficient acetylation, as shown by an increase in mass of 42 Daltons. These peptides, Ac-GHYRGC and Ac-GYHRGC, both contain histidine and tyrosine in adjacent positions (Figure 2-3). We repeated this experiment four times to confirm the relative reaction yields, and we found that Ac-GHYRGC

and Ac-GYHRGC were consistently acetylated with average values of 90% with a standard deviation of 2% and 97% with a standard deviation of 2%, respectively.

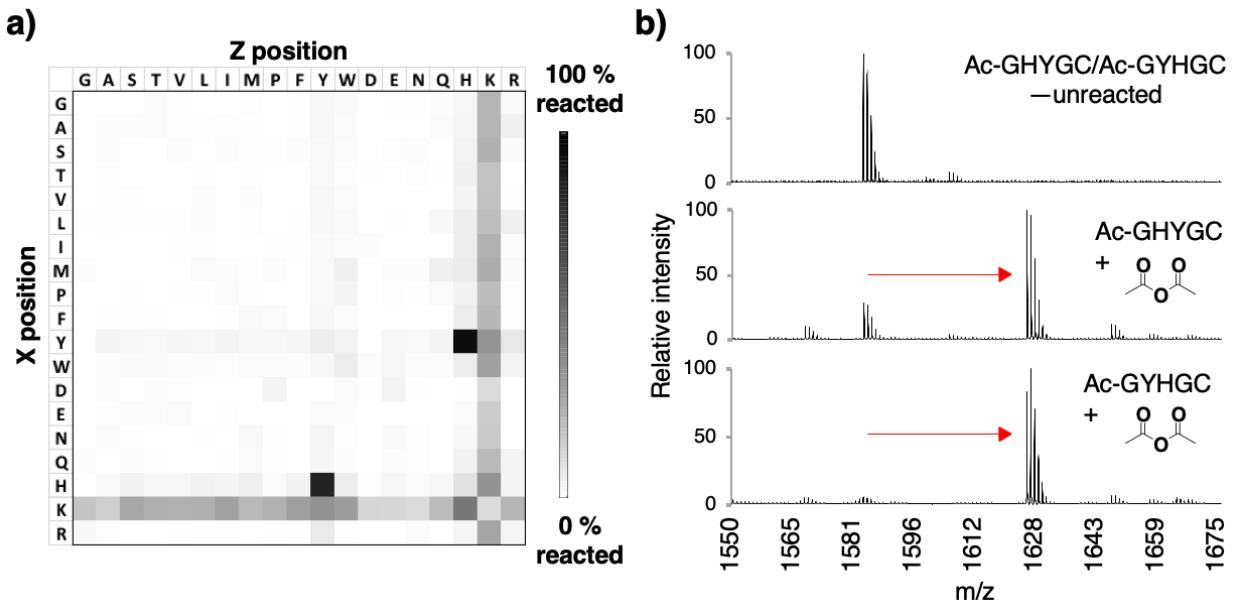


Figure 2-3. Acetic Anhydride selectively acetylates the HY dyad. a) Heatmap showing the average extent of acetylation from four replicates of Ac-GXZRGC peptides with acetic anhydride, where each square represents a peptide having X and Z residues as denoted on the vertical and horizontal axes, respectively. The extent of acetylation is denoted in grayscale, with black representing complete reaction. Acetic anhydride selectively reacted with peptides containing both histidine and tyrosine in the X and Z positions. The proportion of acetylated peptide was calculated by dividing the area under the curve (AUC) of the peak for the acetylated product by the total AUC of the peaks for unreacted substrate and acetylated product. b) Spectra for Ac-GHYRGC and Ac-GYHRGC before and after reaction with acetic anhydride. The top spectrum shows Ac-GYHRGC before treatment with acetic anhydride (Spectrum of unreacted Ac-GHYRGC is similar and not shown). The middle and bottom spectra show that after treatment with acetic anhydride, the mass peaks of Ac-GHYRGC and Ac-GYHRGC are shifted by +42 Da corresponding to acetylation.

The mass spectra also revealed low yields for acetylation of peptides containing lysine residues, which is expected at neutral pH because there is still a small fraction of side chain amino

groups that are not protonated (and therefore nucleophilic). Under basic conditions, in contrast, the acetylation of lysine amino groups with anhydrides is efficient and has been used for sample preparation in bottom-up proteomics.[159] The heatmap also demonstrates that SAMDI has the quantitative resolution to compare reactivities for different sequences. The average yield and standard deviation for acetylation of each peptide is shown in Appendix Figure S-1.

Solution-phase reactivity of acetic anhydride. We next confirm that the acetylation of the HY and YH dyads were not influenced by proximal monolayer and are also observed for solution-phase reactions. Additionally, we repeated the reaction on peptides having a non-capped N-terminal amino group as well as a free cysteine residue to compare relative reactivities. We performed solution-phase reactions with four peptides, each containing only a single site that can react with the anhydride—Ac-G**YH**GRG, Ac-G**K**GRG, Ac-GAGR**C**, and **NH_2**-GAGRG—and three peptides that served as negative controls—Ac-GYGRG, Ac-GHGRG, and Ac-GAGRG. We treated solutions of each peptide (100 μM) with acetic anhydride in water (100 mM) for 1 hour and then analyzed the reactions using MALDI-MS. Figure 2-4 shows the yield and spectra for each reaction. The reactivities of the peptides in solution are consistent with our observations of the reactions directly on the peptide array, though the yields for the former were generally lower. The higher yields for the immobilized peptides likely reflect the different stoichiometry in the solid-phase reaction—where the anhydride is present in far greater excess relative to the immobilized peptide—which is particularly important since hydrolysis of the anhydride competes with acetylation. The HY dyad is acetylated in solution and peptides containing only histidine and tyrosine are not significantly acetylated, eliminating the possibility that the monolayer plays a role

in the reactivity observed on the peptide array. We also observed acetylation of the lysine amino group at a level of about 1/3 of that observed for the peptide containing the HY dyad, which was similar to what we observed in reactions with the peptide array. As anticipated, cysteine and the non-capped N-terminus are more reactive than the HY dyad. The observed reactivity of acetic anhydride with the peptides containing cysteine and a non-capped N-terminus also agrees with previous reports.[160, 161]

Selective reactivity extends to other acylating agents. We also characterized reactivity of the Ac-GYHGC peptide with other electrophilic acylating reagents, including acetyl-CoA, propionic anhydride and m-dPEG-NHS ester. While we did not observe reaction with acetyl-CoA, we did observe acylation of the peptide with propionic anhydride and with the NHS ester; we reason that lower electrophilicity of the acetyl-CoA resulted in poor reactivity under these conditions. We treated the Ac-GXZRGC peptide array with propionic anhydride (Figure 2-5a) and m-dPEG-NHS ester (Figure 2-5b) and found that both reagents gave similar selectivity for the HY dyad, suggesting that this reactivity is broadly applicable to other electrophilic acylating agents.

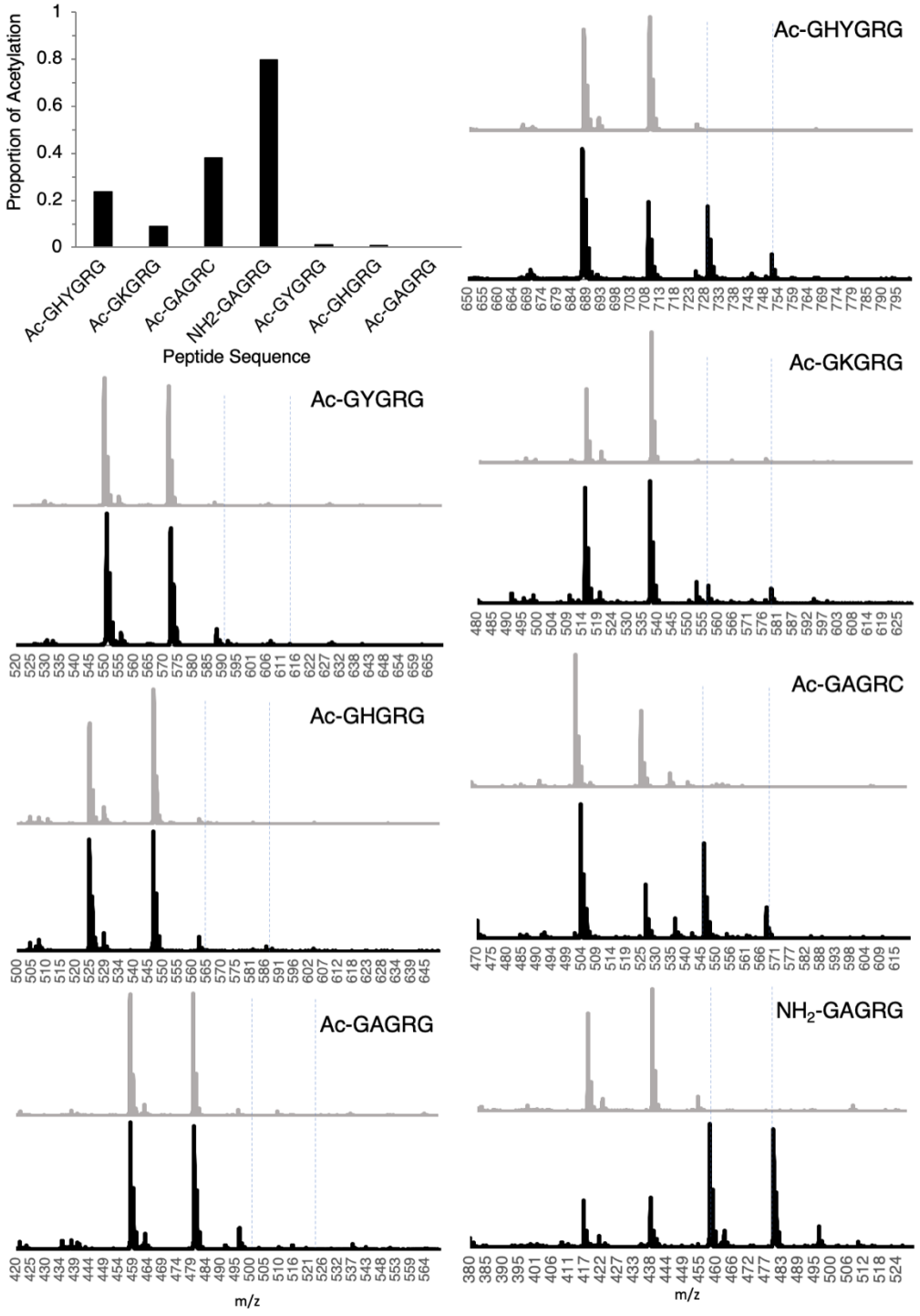


Figure 2-4. Solution-phase acetylation of peptides containing the HY dyad, lysine, cysteine, or a non-capped N-terminus. Seven peptides—Ac-GHYGRG, Ac-GKGRG, Ac-GAGRC, NH_2-GAGRG, Ac-

GYGRG, Ac-GHGRG, and Ac-GAGRG, (100 uM) were treated with of acetic anhydride (100 mM) for 1 hour. The reactions were then lyophilized, resuspended in H2O, and analyzed using MALDI mass spectrometry. The plot shows the yield for acetylation of each peptide under these conditions. The corresponding MALDI-MS spectra of each peptide before (gray) and after (black) in-solution reactions with acetic anhydride are also shown. The blue dotted lines show the expected m/z of H+ and Na+ adducts of acetylated peptide.

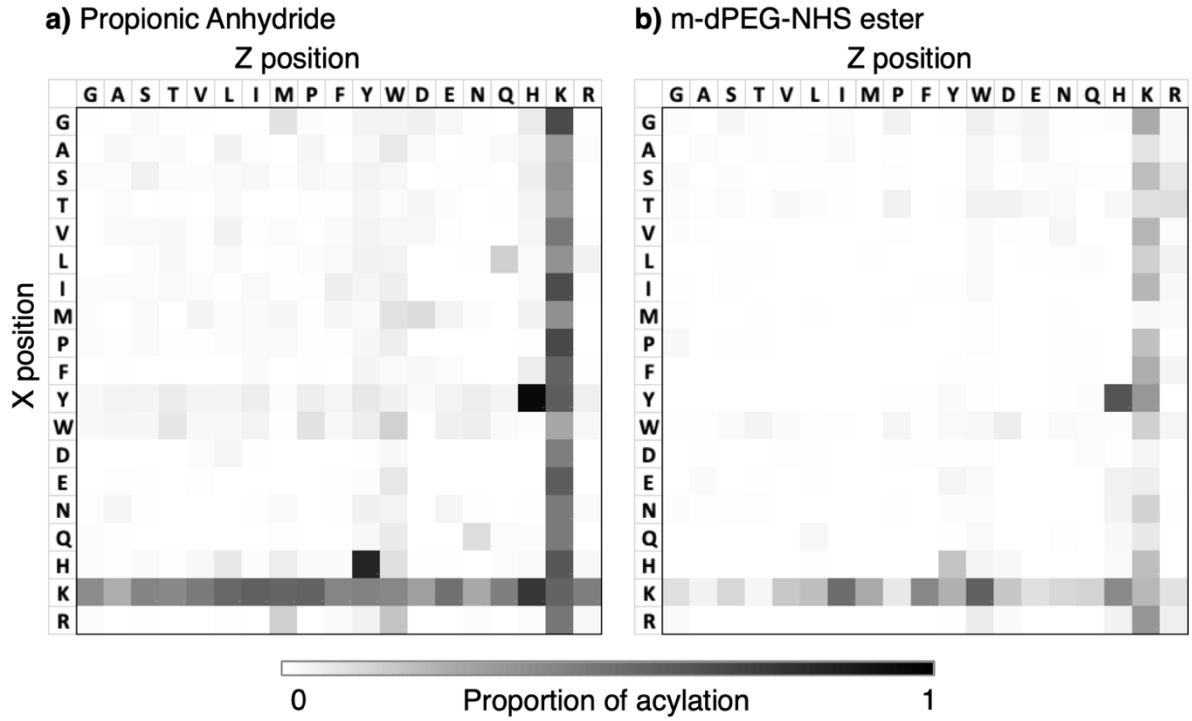


Figure 2-5. Propionic anhydride and m-dPEG-NHS ester also selectively acylate the HY dyad. Heatmaps showing the proportion of acylation of peptides in Ac-GXZRGC peptide array with a) propionic anhydride and b) m-dPEG-NHS ester (100 mM) for 1 hour at room temperature. Each square represents a peptide having X and Z residues as denoted on the vertical and horizontal axes, respectively. The extent of acylation is denoted in grayscale, with black representing complete reactivity.

Mechanistic determination of acetylation of HY dyad. We reasoned that acetylation of the His-Tyr dyad proceeds through initial acetylation of the histidine imidazole group, followed by

intramolecular transfer of the acetyl group to the tyrosine hydroxyl group (Figure 2-6a). This mechanism is consistent with the known reaction of activated esters with imidazole groups (though the acylated adducts undergo rapid hydrolysis) and with the lack of reactivity with phenols (as is also shown by lack of reaction with other tyrosine-containing peptides in the array). To test this mechanism, we performed reactions of acetic anhydride with four additional peptides: Ac-GYHGC, Ac-GY^{OMe}HGC (where the tyrosine hydroxyl group is methylated to remove the nucleophilic character), Ac-GHGC, and Ac-GYGC. We again allowed each peptide to react with acetic anhydride (100 mM) for times ranging from 0 to 30 minutes, in 2- or 4-minute intervals (Figure 2-6b).

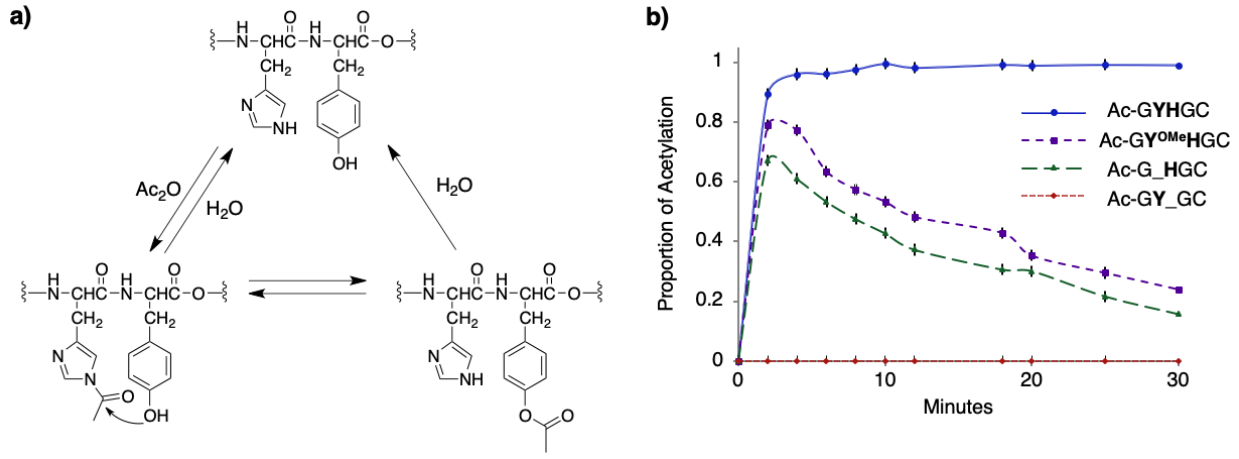


Figure 2-6. Acetylation of the histidine-tyrosine dyad. a) Mechanism of acetylation of the histidine-tyrosine dyad. The histidine residue is first acetylated, followed by intramolecular transfer of the acetyl group to the phenol of the neighboring tyrosine, to give a stable adduct. b) Kinetic profile for acetylation of four additional peptides after reaction with acetic anhydride (100 mM in water) for 30 minutes. The peptide containing both histidine and tyrosine, Ac-GYHGC (blue), was nearly quantitatively acetylated after 2 minutes and remained acetylated. The two peptides, Ac-GY^{OMe}HGC (purple - containing histidine and an O-methylated tyrosine) and Ac-GHGC (green – containing only histidine), were acetylated but showed

significant hydrolysis over 30 minutes. The peptide containing only tyrosine, Ac-GYGC (red), remained unacetylated.

Our results show that the parent peptide containing the YH dyad, Ac-GYHGC, was nearly quantitatively acetylated after 2 minutes and remained acetylated through the thirty-minute reaction. The peptides Ac-GY^{OMe}HGC and Ac-GHGC, which do not include the nucleophilic tyrosine residue, were acetylated in approximately 50% yield after 2 minutes, but the extent of acetylation decreased to approximately 20% after 30 minutes. The peptide containing no histidine, Ac-GYGC, did not react with the anhydride. These results show that the histidine residue is required for acetylation, but acetylation on histidine is not stable and undergoes hydrolysis to give the native peptide. However, a neighboring tyrosine group competes with hydrolysis through a trans-acetylation reaction. We find no previous reports of an acetyl-tyrosine post-translational modification or this His-to-Tyr acylation mechanism, though an early report by Moore and coworkers observed efficient acetylation of the angiotensin II peptide (DRVYIHPF) with acetic anhydride, and suggested that the histidine serves as a base to promote direct acetylation and deacetylation of the tyrosine.[162] Finally, we note that while acetyl transfer from the imidazole side chain of histidine to the hydroxyl group on tyrosine has not been reported in polypeptides, there is significant precedent for imidazole serving as a catalyst in acylation reactions. These examples include the acylation of cellulose through a similar acyl transfer mechanism[163] as well as for mercaptans, amines, azides, and phenols.[164, 165] Similarly, imidazole can catalyze the hydrolysis of esters.[165, 166]

Stability of acetylated tyrosine. We measured the hydrolysis of acetyl-tyrosine on two peptides, Ac-GY^{ac}GC and Ac-GY^{ac}HGC, in both water and PBS buffer pH 7.4 over 96 hours (Figure 2-7). We found that acetyl-tyrosine in a peptide lacking the proximal histidine residue was stable over the 96 hours in both water and PBS buffer, with little hydrolysis. This stability is consistent with microscopic reversibility for the forward reaction and also with prior work showing that acetylated phenols have half-lives of several days.[162, 167] However, the Ac-GY^{ac}HGC peptide was found to hydrolyze more rapidly, with half lives of 157.5 hours in water and 30.4 hours in PBS, similar to the reported values by Moore and coworkers.[162] This lowered stability in buffered solutions could be one of the reasons why this modification has not been observed in proteomic studies, as it may not be stable throughout the sample preparation process.

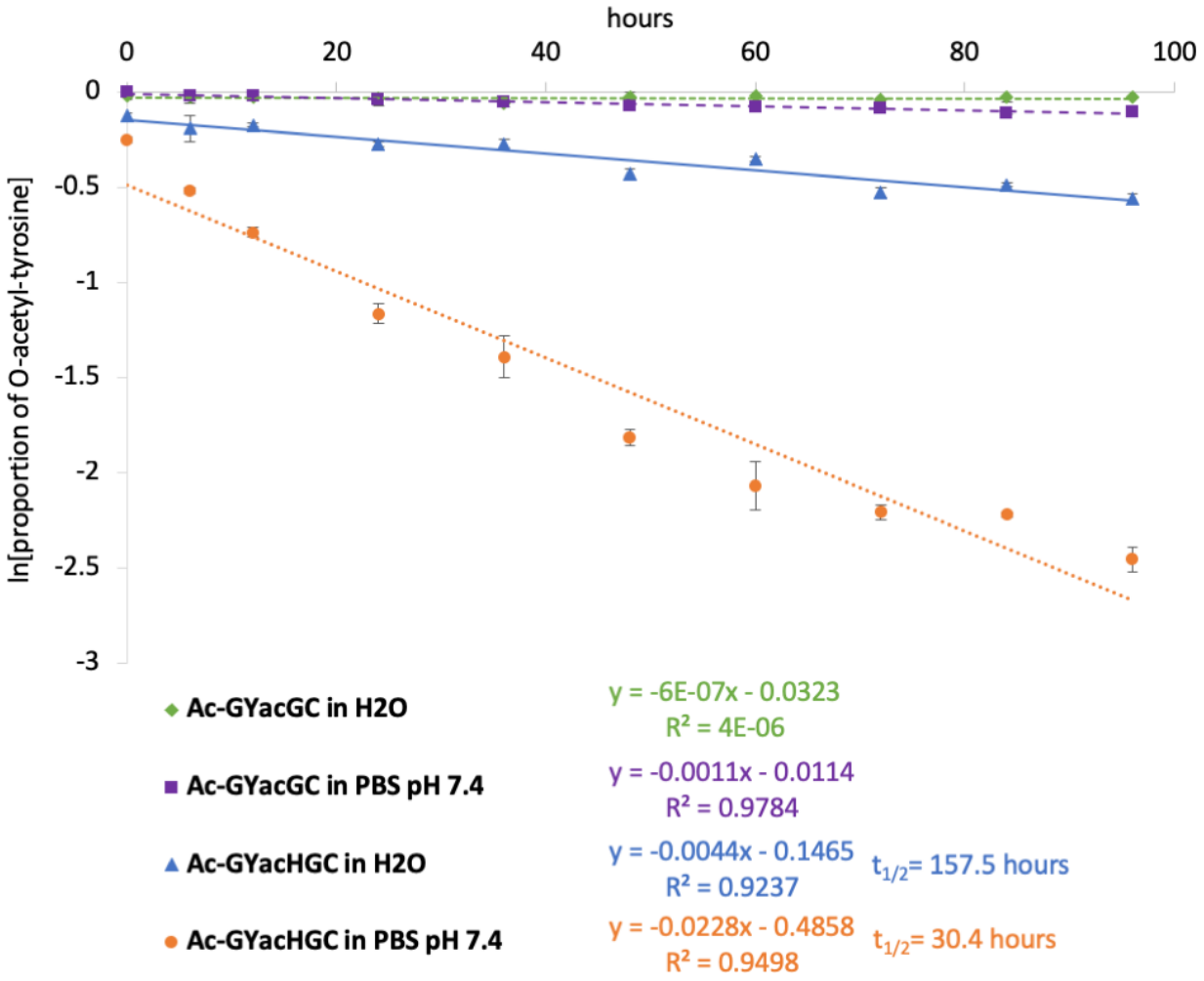


Figure 2-7. Pseudo-first order rate plot of the hydrolysis of O-acetyl-tyrosine. The hydrolysis of O-acetyl-tyrosine on the Ac-GY^{ac}GC peptide in both water and PBS has a rate constant ≈ 0. On the Ac-GY^{ac}HGC peptide, the hydrolysis of acetyl tyrosine has rate constants of 0.004 hour^{-1} and 0.023 hour^{-1} and half-lives of 157.5 hours and 30.4 hours in H_2O and PBS pH 7.4, respectively.

Effects of proximal amino acids on selective reactivity. We next asked whether the acylation of the HY dyad could be further accelerated (or inhibited) by varying the amino acids surrounding this motif. We synthesized a second peptide library having the sequence Ac-GXYHZGRC, where

X and Z are variable positions and again represent all 19 amino acids except for cysteine. We prepared an array comprising these 361 peptides as described above and treated the array with acetic anhydride (100 mM in water) for 1 hour (Figure 2-8a). The average data values and standard deviations for each peptide are shown in Appendix Figure S-2. As expected, we found that all peptides underwent efficient acetylation. However, peptides with two or three tyrosine groups were found to be di- or tri-acetylated, respectively. Further, peptides containing lysine in either one or two variable positions were also acetylated more than once, but to a lesser extent. This result shows that a single histidine residue can catalyze the acetylation of multiple tyrosine residues in the same peptide, as would be expected for its role as a catalyst.

We then repeated this experiment with a lower concentration of acetic anhydride (0.1 mM) to decrease the yield of acetylation and better resolve the relative reactivities of the peptides with the anhydride (Figure 2-8b). The average data values and standard deviations for each peptide are shown in Appendix Figure S-3. We found that peptides containing acidic residues—either aspartic acid or glutamic acid—in both variable positions were not acetylated. Interestingly, a single acidic residue did not significantly affect the reaction rate. A recent report by Denu and colleagues measured the kinetics of non-enzymatic lysine acetylation using acetyl-phosphate and acetyl-CoA on 90 lysine sites within eight purified proteins.[168] They also found that lysines with a glutamic or aspartic acid in the +1 and -1 positions were acetylated at lower rates than were lysines without proximal acidic residues, and suggested this tempered reactivity was due to salt bridge formation. Our similar observation with the acetylation of the HY dyad could similarly be due to a salt bridge, or to a higher pKa value for the protonated imidazole due to an electrostatic interaction.

The array also revealed that peptides containing a histidine in one or both of the variable positions underwent acetylation more rapidly, possibly by increasing the amount of acylated histidine intermediate. Similar reactivity was also found with propionic anhydride (Figure 2-9a) and m-dPEG-NHS ester (Figure 2-9b), however the NHS ester was significantly less reactive across the whole array. This data shows that acetylation of the histidine-tyrosine dyad by acetic anhydride is a reasonably specific reaction that is most favorable in the presence of a neighboring histidine residue and absence of nearby acidic residues.

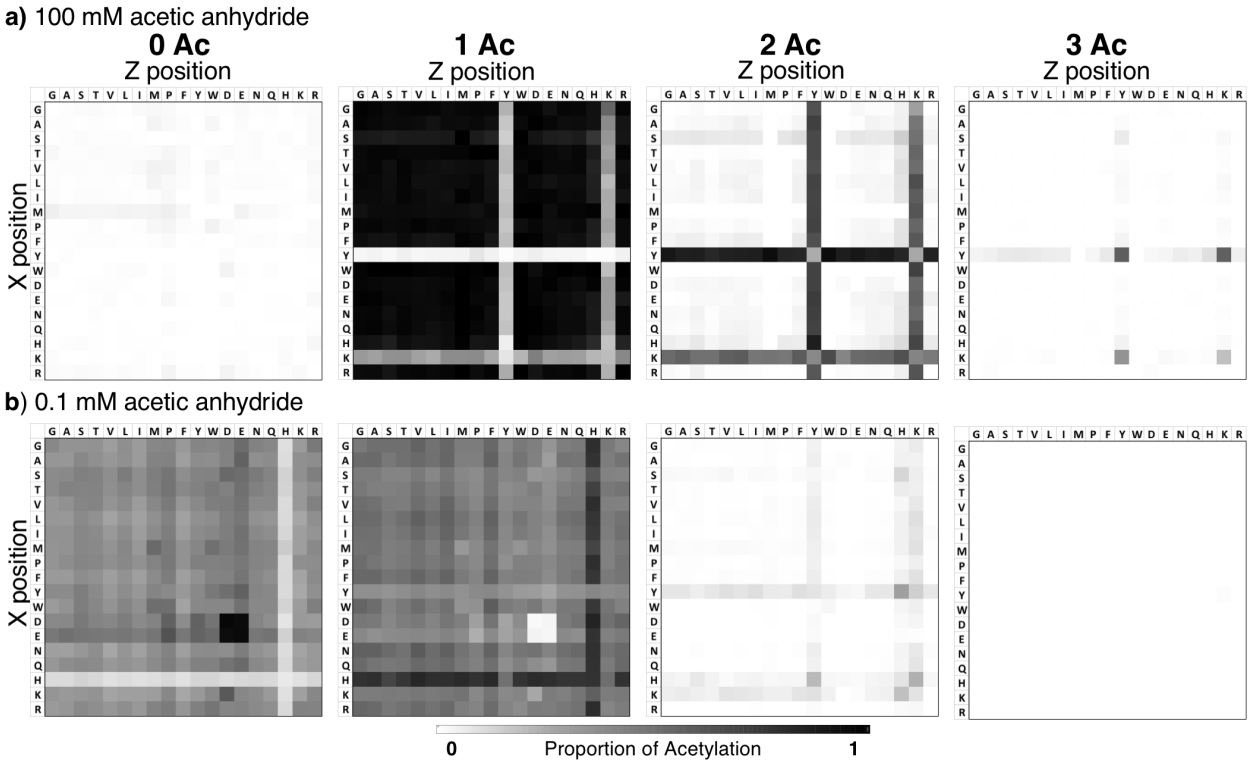


Figure 2-8. Effects of amino acids surrounding HY motif on acetylation. A peptide array based on the sequence Ac-G<u>X</u>YH<u>Z</u>GRC was treated with acetic anhydride in water at either 100 mM (a) or 0.1 mM (b)

and analyzed by SAMDI mass spectrometry. Separate heat maps show the amount of each peptide that did not react (0 Ac), and that underwent acetylation one (1 Ac), two (2 Ac) or three (3 Ac) times.

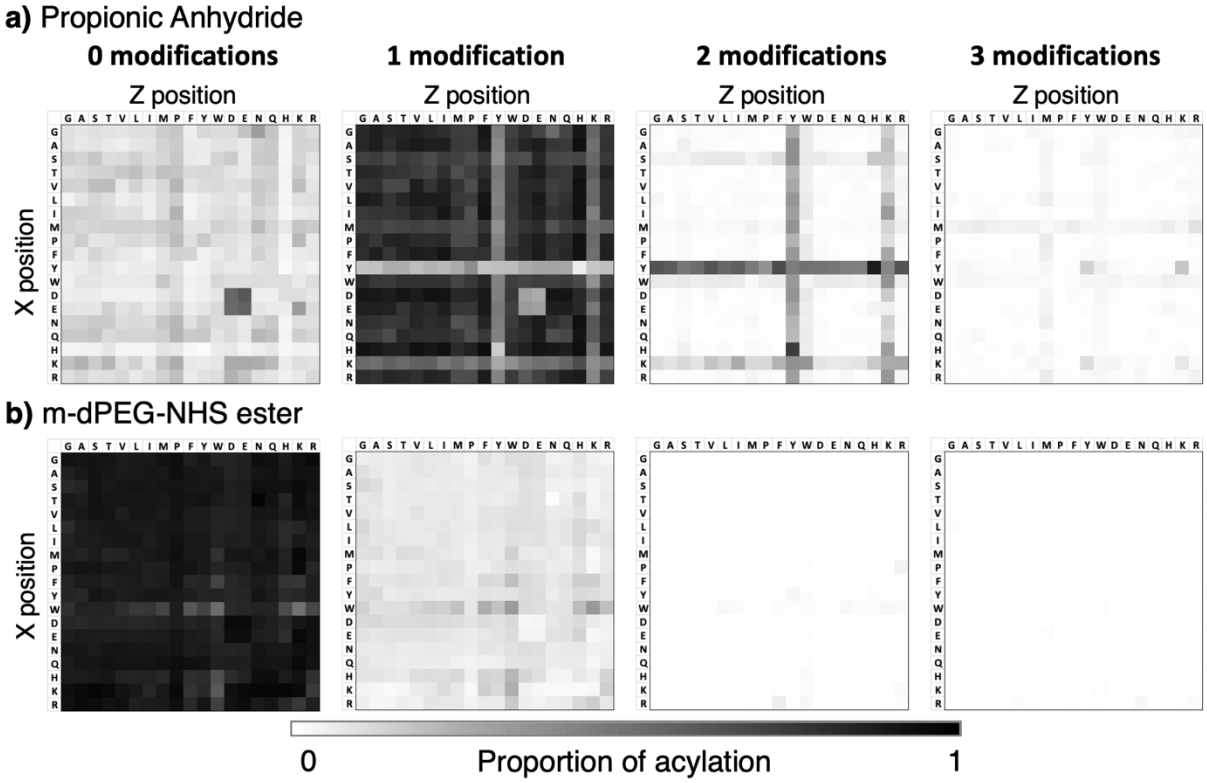


Figure 2-9. Reactivity of propionic anhydride and m-dPEG-NHS ester on Ac-GXYHZGRC peptide array. Heatmaps showing the modification states of peptides in the Ac-GXYHZGRC peptide array after treatment with 0.1 mM **a)** propionic anhydride and **b)** m-dPEG-NHS ester in water for 1 hour.

Distance-dependence of the acetyl transfer from histidine to tyrosine. To examine the intramolecular distance dependence for the transfer of the acetyl group from the histidine to tyrosine, we synthesized 4 peptides having the sequence Ac-GY$(G)_n$HRGC, where n is one, two, three, or four and therefore increases the distance between the histidine and tyrosine residues. We

treated those four peptides as well as Ac-GYHRGC with acetic anhydride (100 mM) for 25 minutes (Figure 2-10a). Each of the peptides reached a maximum yield at about 12 minutes—likely because the acetic anhydride undergoes hydrolysis during the reaction. We observed a slight decrease in the extent of acetylation for the peptides having 1 and 2 glycine residue spacers, and a greater decrease for the peptides having 3 and 4 glycine residue spacers. However, the acetylation did not drop off significantly, even with peptide having four glycine residues, indicating that the acetyl transfer can occur over longer distances (at least for peptides having a conformationally unconstrained oligoglycine linker).

Lastly, we address the possibility that the acetyl transfer from histidine to tyrosine could occur intermolecularly. We co-immobilized equal amounts of two peptides, Ac-GHGC and Ac-GYGC to monolayers with varying densities of maleimide, so that the mean distance between neighboring peptides could be varied. We expected that at high densities of peptide, the former could undergo acetylation and transfer the acetyl group to the latter, but that transfer would not be possible for low densities of peptides. The alkanethiolate chains in a monolayer have an approximate density of 4.5/nm^2.[72, 169] Therefore, the density of immobilized peptides for monolayers having maleimide groups presented at a percentage of the total alkanethiolates are 0.045/nm^2 at 1%, 0.23/nm^2 at 5%, 0.45/nm^2 at 10%, 0.68/nm^2 at 15%, 0.90/nm^2 at 20%, and 1.13/nm^2 at 25%. We treated the monolayers with acetic anhydride (100 mM in water) for 1 hour, and analyzed the reactions using SAMDI-MS (Figure 2-10b). Indeed, we found that the tyrosine-containing peptide showed very little acetylation when the peptides were present at a density of 1 or 5%, but significant levels of acetylation for densities above 10%. Monolayers presenting only the tyrosine-containing peptide,

by contrast, showed no acetylation at any of the densities. This example demonstrates that histidine-mediated acetylation of tyrosine can occur in a distance-dependent intermolecular pathway and by extension suggests that, in proteins, this transfer could occur not only when histidine and tyrosine are close in primary sequence, but also when proximal in space.

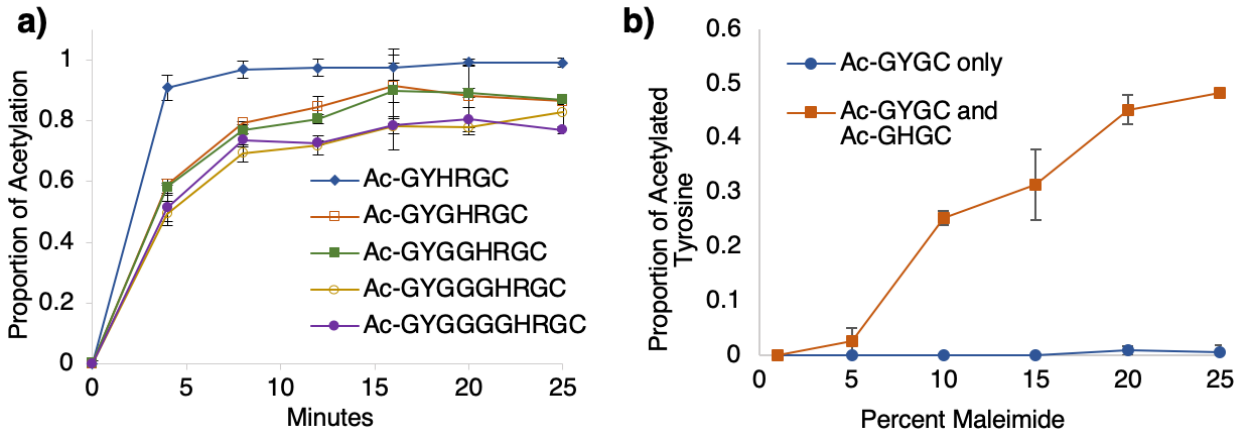


Figure 2-10. Intra and inter molecular distance dependence of the acetyl transfer a) Five peptides where the distance between the tyrosine and histidine residues is varied were each treated with acetic anhydride (100 mM in water) for 25 min. Analysis of the extent of acetylation at several time points shows the relative rates for acetylation. b) Monolayers presenting a mixture of GYGC and GHGC peptides (orange) or only the GYGC peptide (blue) for densities of peptide ranging from 1 to 25% were treated with acetic anhydride in water (100 mM) for 1 hour. The yield for acetylation of the Ac-GYGC peptide was determined by SAMDI mass spectrometry.

Conclusion

This work illustrates how peptide arrays and SAMDI-MS can be used to discover sequence-specific reactions of peptides. We found that acetylation of peptides by acetic anhydride occurs efficiently on tyrosine residues that have a neighboring histidine. It is surprising that this reaction has not been previously reported, even though acylation of proteins has been thoroughly studied

and the mechanistic basis of this reaction is well established. We believe that this approach will be important for the discovery of additional chemistries for site-specific protein modification and conjugation, as well as for optimization of those that have already been reported.[142, 144-146, 153, 154, 170, 171] Current approaches often do not assess the global sequence specificity, and also are limited to identifying anticipated reactivities. Our method uses non-biased peptide arrays comprising of two variable positions with mass spectrometry to characterize all possible sequences and identify unforeseen reactivities, allowing for a complete understanding of the specificity of the reaction. This example demonstrates the benefits of using peptide arrays and SAMDI-MS to discover peptide-modifying reactions and to understand their sequence-specificity.

Methods

Peptide library synthesis. Peptide libraries, Ac-GXZRGC and Ac-GXYHZGRC, were synthesized using standard Fmoc-based solid-phase peptide synthesis[172] on SynPhase polyamide lanterns with Fmoc-Rink amide linkers purchased from Mimotopes. Synthesis was performed in 96-well filter plates purchased from Arctic White with the use of a 96-well plate vacuum manifold purchased from Millipore. Fmoc- and side chain-protected amino acids and N-acetyl-glycine were purchased from either Anaspec or Sigma-Aldrich. Coupling reagents, Pybop and N-methylmorpholine (NMM), as well as dimethylformamide (DMF) were purchased from Fisher or Sigma Aldrich.

For each library, lanterns were swelled for 1 hour in DMF and then separated into four 96-well filter plates. The lanterns were Fmoc-deprotected with a solution of 20% piperidine in DMF

for 20 minutes. Using vacuum filtration, the wells were drained, and the lanterns were rinsed with DMF. The first coupling reaction was performed for 20 minutes with Fmoc-Cys(Trt)-OH, Pybop and NMM in 250 μL DMF at a molar excess of 8:8:16, respectively, to the molar loading capacity of the lanterns. The wells were then drained, and the lanterns were rinsed with DMF. Fmoc-deprotection, washing, and coupling reactions with respective amino acids were repeated until the last amino acid (N-acetyl-glycine) was coupled. After the last coupling reaction, the lanterns were rinsed with DMF and dichloromethane and dried under vacuum for 1 hour. Lanterns were then transferred into new 96-well plates. Amino acid side chain deprotection and peptide cleavage from the resin was performed in tandem for 2 hours using a solution of 2.5% triethylsilane, 2.5% H_2O, and 95% trifluoroacetic acid (TFA). The lanterns were then discarded, and the cleavage solution was evaporated under nitrogen. The peptides were resuspended in 0.1% TFA in H_2O and lyophilized. The peptides were resuspended again in 0.1% TFA in H_2O to a final concentration of 500 μM.

Additional peptide synthesis. Sixteen additional peptides—Ac-GYHGRG, Ac-GKGRG, Ac-GAGRC, NH2-GAGRG, Ac-GHGRG, Ac-GYGRG, Ac-GAGRG, Ac-GYHGC, Ac-GY^{OMe}HGC (O-methylated), Ac-GHGC, Ac-GYGC, Ac-GYHRGC, Ac-GYGHRGC, Ac-GYGGHRGC, Ac-GYGGGHRGC, and Ac-GYGGGGHRGC—were synthesized using standard Fmoc-based solid-phase peptide synthesis on Fmoc-Rink Amide MBHA resin purchased from Anaspec. The O-methylated Fmoc-protected tyrosine was purchased from Advanced Chem Tech. The resin was swelled for 30 minutes in DMF. Fmoc-deprotection and amino acid coupling reactions were performed as described for the peptide library synthesis. The peptides were cleaved, and the amino

acid side chains were deprotected in a solution of 2.5% triethylsilane, 2.5% H_2O, and 95% TFA for 2 hours. The cleaved peptides were filtered to remove the resin and the peptides were precipitated with ethyl ether. The precipitated peptides were resuspended in 0.1% TFA in H_2O and lyophilized. They were then resuspended in 0.1% TFA to a final concentration of 500 μM.

Preparation of SAMDI peptide arrays. Peptide arrays were prepared as described previously.[10, 173-175] Briefly, 384 gold spots were evaporated onto titanium coated steel plates using a Thermionics E-beam evaporator. The plates were soaked in a 1 mM total disulfide monolayer solution of 0.8 mM tri(ethylene glycol) disulfide and 0.2 mM tri(ethylene glycol)-maleimide disulfide in ethanol. The alkanethiolate monolayer self-assembles onto the gold surfaces and presents a functional maleimide group against a background of tri(ethylene glycol).

Peptides were neutralized by dilution in 50 mM Tris pH 7.5 to a concentration of 50 μM and pipetted onto the gold spots using Tecan robotics (one peptide per gold monolayer). The peptides were incubated on the surface in a humidified chamber at room temperature for 1 hour for immobilization. Peptide immobilization occurs through 1,4-Michael addition of the cysteine-thiol side chain to the maleimide. The arrays were then washed with H_2O and then ethanol, dried under nitrogen, and stored in vacuum sealed bags at 4°C until ready for use.

SAMDI peptide array reactions. Acetic anhydride, propionic anhydride, and m-dPEG-NHS ester were purchased from Sigma-Aldrich. Ac-GXZRGC peptide library— Four prepared arrays were treated with 100 mM acetic anhydride in H_2O at room temperature for 1 hour. Two additional prepared arrays were treated with either 100 mM propionic anhydride or 100 mM m-dPEG-NHS ester in H_2O at room temperature for 1 hour. After reactional completion, the plates were rinsed

with H_2O and then ethanol and dried under nitrogen. Ac-GXYHZGRC peptide library— Ten SAMDI peptide arrays were prepared; five were treated with 100 mM acetic anhydride and five were treated with 0.1 mM acetic anhydride in H_2O at room temperature for 1 hour. Two additional prepared arrays were treated with either 0.1 mM propionic anhydride or 0.1 mM m-dPEG-NHS ester in H_2O for 1 hour at room temperature. The plates were then rinsed with H_2O and then ethanol and dried under nitrogen. All arrays were analyzed by SAMDI-MS. The heat maps shown in Figure 2-3 and Figure 2-8 portray average values of acetylation on each peptide.

In-solution peptide reactions. 7 peptides—Ac-GHYGRG, Ac-GKGRG, Ac-GAGRC, NH2-GAGRG, Ac-GYGRG, Ac-GHGRG, and Ac-GAGRG—were treated with acetic anhydride in solution. The in-solution reactions consisted of 100 μM of peptide with 100 mM acetic anhydride in H_2O for 1 hour at room temperature. The reactions were then flash frozen in liquid nitrogen and lyophilized to remove the acetic anhydride. The peptides were then resuspended in H_2O. The resuspended peptide solutions were mixed at a ratio of 1:9 with a matrix solution consisting of 10 mg/mL α-Cyano-4-hydroxycinnamic acid in 50% acetonitrile, 50% H2O, 0.1% TFA, and spotted onto a MALDI plate. The peptides were analyzed using reflector positive mode on an AB Sciex 5800 MALDI TOF/TOF. Spectra were collected from 900 laser shots with a stage velocity of 2500 μm/sec. The proportions of acetylation and the spectra of each peptide are shown in Figure 2-4.

Timed SAMDI reactions on individual peptides. Ac-GYHGC, Ac-GY^{OMe}HGC (O-methylated), Ac-GHGC, and Ac-GYGC peptides were immobilized (as described above) in quintuplet on eleven SAMDI plates. Each plate was treated with 100 mM acetic anhydride in H_2O for 2, 4, 6, 8, 10, 12, 18, 20, 25, or 30 minutes. Each plate was rinsed with water and then ethanol

and dried under nitrogen at the completion of its time point, and then analyzed using SAMDI-MS. The averages and standard deviations were calculated from five replicates of each time point. The averages and standard deviations are plotted in Figure 2-6b.

SAMDI Acetyl-tyrosine hydrolysis experiments. Forty micro SAMDI plates were prepared—Ac-GYHGC was immobilized (as described above) in quintuplet on twenty plates, and Ac-GYGC was immobilized in quintuplet on each of the remaining twenty plates. The twenty Ac-GYHGC-immobilized plates were chemically acetylated by three treatments of 100 mM acetic anhydride in acetonitrile for 1 hour at room temperature. The twenty Ac-GYGC-immobilized plates were chemically acetylated by three treatments of 100 mM acetic anhydride, 50 mM triethylamine in acetonitrile for 1 hour at room temperature. All plates were then rinsed with ethanol, water, and then ethanol and dried under nitrogen. Ten of the Ac-GY^{ac}HGC-immobilized plates, and ten of the Ac-GY^{ac}GC-immobilized plates were soaked in H_2O for 0, 6, 12, 24, 36, 48, 60, 72, 84, or 96 hours. The ten remaining Ac-GY^{ac}HGC-immobilized plates and the ten remaining Ac-GY^{ac}GC-immobilized plates were soaked in PBS pH 7.4 buffer for 0, 6, 12, 24, 36, 48, 60, 72, 84, or 96 hours. The plates were rinsed with water, then ethanol, dried under nitrogen, and analyzed by SAMDI-MS. The average remaining amount of acetylation was measured at each time point, and the pseudo-first order rate plot is shown in Figure 2-7.

SAMDI Intramolecular distance experiments. To examine the effects of increased intramolecular distance between the histidine and tyrosine, five peptides—Ac-GYHRGC, Ac-YGHRGC, Ac-GYGGHRGC, Ac-GYGGGHRGC, and Ac-GYGGGGHRGC—were all immobilized in triplicate onto seven SAMDI plates. Each plate was treated with 100 mM acetic

anhydride in H_2O at room temperature for either 0, 4, 8, 12, 16, 20, or 25 minutes, and rinsed with water and then ethanol and dried under nitrogen. Each plate was analyzed by SAMDI-MS. The proportion of acetylation on each peptide at each time point was averaged and plotted in Figure 2-10a.

SAMDI Intermolecular distance experiments. To examine the effects of increased intermolecular distance between the histidine and tyrosine, six gold patterned plates were soaked in a 1 mM total disulfide monolayer solution with different ratios of tri(ethylene glycol) disulfide and tri(ethylene glycol)-maleimide disulfide in ethanol to present maleimide at a density of 1%, 5%, 10%, 15%, 20% or 25%. Peptides Ac-GYGC and Ac-GHGC were co-immobilized in triplicate on each plate. As a control, peptide Ac-GYGC alone was also spotted in triplicate on each plate. All plates were treated with 100 mM acetic anhydride for 30 minutes a room temperature, rinsed with water and then ethanol, and dried under nitrogen. The plates were analyzed using SAMDI-MS and the proportion of acetylation on the Ac-GYGC peptide (when co-immobilized with Ac-GHGC and when immobilized independently) was plotted in Figure 2-10b.

SAMDI mass spectrometry. A matrix solution of 10 mg/mL 2,4,6-Trihydroxyacetophenone in 0.1% TFA in acetonitrile was applied to all treated peptide array plates. The arrays were analyzed in reflector positive mode on an AB Sciex 5800 MALDI TOF/TOF. Spectra were collected from 900 laser shots with a stage velocity of 3000 μm/sec.

Data analysis. All spectra were analyzed by integrating the area under the curve of the peptide product and substrate peaks. The proportion of acylation on each peptide was calculated by dividing the area under the curve (AUC) of the acylated peptide product divided by the AUC of

both the acylated product and unreacted substrate. All replicates were averaged in presented results. The peptide library results are displayed as heat maps, which are a grayscale graphical representation of the average proportion of acetylation of each peptide in the array, where black represents complete reactivity. Each box represents an individual peptide in the array with the X position amino acid on the vertical axis and the Z position amino acid on the horizontal axis.

Chapter 3. The Role of Phosphatases in Determining the Distribution of Phosphorylation in the Proteome

This work was done in collaboration with Daniel Sykora in the Mrksich Lab.

Introduction

Background: Phosphorylation

Protein phosphorylation is a prominent, reversible post-translational modification that serves as an important regulatory mechanism for various cellular processes such as signaling, migration, proliferation, apoptosis, differentiation, and metabolism.[176-178] The addition and removal of phosphorylation sites on proteins are regulated by kinases and phosphatases, respectively, and improper regulation of phosphorylation has been found to contribute to several diseases such as cancer, diabetes, and neurodegenerative and inflammatory disorders.[179] Phosphorylation occurs on over 30% of cellular proteins with a distribution of 86.4% on serine (Ser), 11.8% on threonine (Thr), and 1.8% on tyrosine (Tyr) residues.[180, 181]

The prevalence of phosphorylation on serine residues remains unclear; however, total kinase activity on serine residues must be relatively greater than that of phosphatases. In fact, most explanations for the regulation and distribution of phosphorylation sites has often emphasized or solely addressed changes in kinase activity, with little consideration for the roles that phosphatases play. This is in part because the number of serine/threonine kinases (428) far outnumber the phosphatases (only 30), [182] leading many researchers to postulate that Ser/Thr phosphatases play a ‘housekeeping’ role. This is in stark contrast to tyrosine phosphorylation, where the number of kinases (90) is of the same order as the number of phosphatases (107), [182] and tyrosine

phosphatases are believed to play a more prominent role in regulating dynamic signaling events. However, in recent years it has been recognized that Ser/Thr phosphatases are important for regulation of phosphorylation.[183] Different from protein kinases, which are derived from a common ancestor, the Ser/Thr and Tyr phosphatase families evolved separately.[184] *In vivo,* the Ser/Thr family acts as holoenzymes that form various complexes with large numbers of regulatory subunits to gain specificity,[183] which in turn introduces new challenges when studying their activity. The inherent challenges of studying Ser/Thr phosphatases and the lack of compatible assays quantifying both phosphatase activity and substrate specificity in complex samples, such as cell lysates, have contributed to the research imbalance between kinases and phosphatases.

While many kinase assays rely on proteomic techniques to identify phosphorylation sites, the same techniques cannot be used to measure the loss of phosphorylation. Current assays used to measure phosphatase activity rely on quantifying the decrease in phospho-substrates by measuring the amount of phosphate released through radioactivity or colorimetry, or through antibody-based detection methods such as enzyme-linked immunosorbent assay (ELISA).[185-188] These techniques have allowed for significant progress in understanding the roles of phosphatases, but they are limited in their use for direct quantification of phosphatase activity in cell lysates. Observing the global roles of the phosphatase families in a cellular environment is important for understanding the mechanisms behind phosphatase specificity and regulation.

SAMDI-MS and peptide arrays to profile phosphatase activity

Here, we combine peptide arrays and self-assembled monolayers for matrix-assisted laser desorption/ionization (SAMDI) mass spectrometry to profile total phosphatase activity in cell lysates and develop a systems-level understanding of phosphatase activity. SAMDI-MS is a mass-spectrometry based technique that uses self-assembled monolayers (SAMs) of alkanethiolates on gold that present functional groups for immobilization of substrates against a background of tri(ethylene glycol) groups. The use of a tri(ethylene glycol) background prevents non-specific protein adsorption to the surface and vastly improves signal to noise.[112, 132] SAMDI is a high throughput, label-free technique that provides a quantitative measure of reaction without the need for purification or tedious sample preparation, making it a powerful platform for large screening experiments with complex samples, such as cell lysates.[57, 112, 132]

In this work, we use SAMs that present a functional maleimide group for immobilization of cysteine-terminated phosphorylated peptides. We synthesized three peptide libraries consisting of a central phosphorylated serine, threonine, or tyrosine residue that is surrounded by two variable positions. We then prepared SAMDI-based peptide arrays that were treated with cell lysate and analyzed using matrix-assisted laser desorption/ionization (MALDI) mass spectrometry (Figure 3-1). Using this technique, we profile the substrate specificities of Ser/Thr and Tyr phosphatases and compare their relative activities to develop a better understanding of global phosphatase activity. This chapter gives a clear example of the sequence-dependent roles played by phosphatases, specifically the Ser/Thr phosphatases, in regulating the global phosphorylation state.

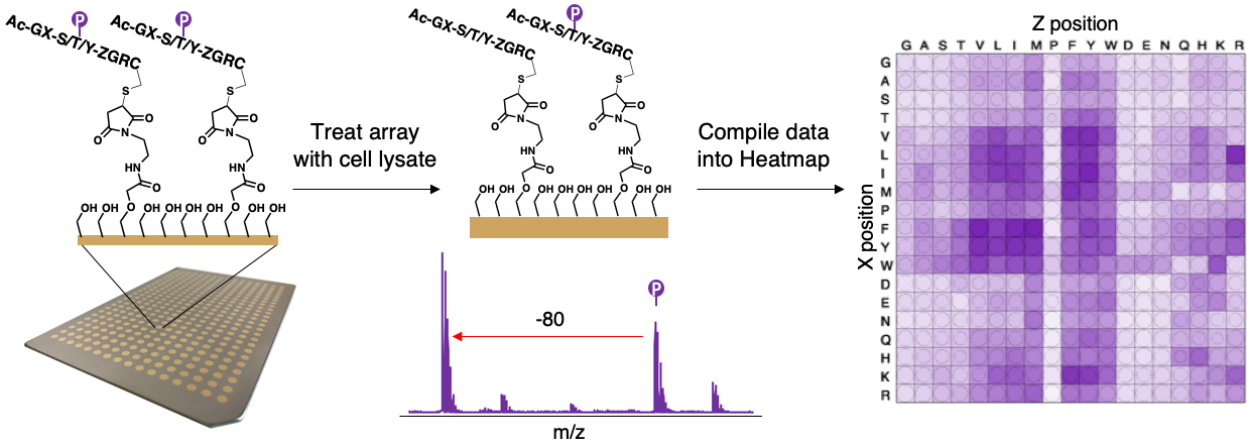


Figure 3-1 Profiling phosphatase activity from cell lysates on SAMDI peptide array. Phosphorylated peptide libraries were immobilized onto SAMDI arrays. The arrays were treated with cell lysate and the extent of dephosphorylation of each peptide was measured using MALDI-MS. The proportion of dephosphorylation on each peptide is shown in heatmaps where the dark purple represents complete dephosphorylation.

Results

Preparation of Peptide Arrays. We used standard Fmoc-based synthesis to prepare a phosphoserine peptide library based on the following sequence: Ac-GXS^{P}ZGRC (where X and Z represent all natural amino acids except for cysteine). In this way, the peptides represent a broad distribution of sequences and can resolve the general specificities of phosphatases. We then prepared SAMDI-based arrays from steel plates containing 384 SAMs on gold islands, arranged in the standard geometry of a microwell plate. Each SAM presented functional maleimide groups at a density of 10% against a background of protein-resistant tri(ethylene glycol) groups, as described previously. The peptides in the phosphoserine library were spotted onto the monolayer array and underwent immobilization through reaction of the cysteine thiol with the maleimide

group. We evaluated the quality of each peptide using MALDI-MS, and the signal-to-noise ratios of each peptide are listed in Appendix Tables S-1, S-2, and S-3.

Previously, we have used SAMDI-based peptide arrays to determine the substrate-specificities of enzymes and to measure deacetylase activities in cell lysates;[16, 189] however, SAMDI-MS is particularly significant for analyzing phosphatase reactions because it can observe the reaction product directly, without the need for labels, and is applicable to all phosphopeptide substrates.

Profiling Phosphatase Activity in Cell Lysate. We used the prepared phosphoserine peptide array to profile phosphatases that were present in lysates derived from cultured NIH/3T3 cells (fibroblasts from mouse embryo). We applied the NIH/3T3 cell lysate to each spot on the array for 15 minutes at 37 °C. After lysate treatment, the peptide array was rinsed and dried. We then applied MALDI matrix and measured the extent of dephosphorylation of each peptide using MALDI mass spectrometry. The extent of dephosphorylation of each peptide was determined by dividing the area under the curve (AUC) of the dephosphorylated product peak by the sum of the AUC of the dephosphorylated product and phosphorylated substrate peaks. The percentages of dephosphorylation of each peptide in the array are displayed in heatmaps, where each square represents an individual peptide with X and Z residues on the vertical and horizontal axes, respectively. The heatmaps display the percent of dephosphorylation on a color scale, where dark purple represents 100% dephosphorylation. The percentages of dephosphorylation shown in the heatmaps are an average from three biological replicates and the standard deviation for each peptide is displayed by circle size within each peptide square—the smaller the circle, the larger the standard deviation.

To assess dephosphorylation on threonine and tyrosine, we similarly prepared arrays based on the sequences: Ac-GXT^{p}ZGRC, and Ac-GXY^{p}ZGRC (where X and Z represent all natural amino acids except for cysteine). We treated the phosphothreonine and phosphotyrosine peptide arrays with the same NIH/3T3 cell lysate, and analyzed the extent of dephosphorylation of each peptide using MALDI-MS. The use of three phosphorylated peptide arrays containing the same basic sequence allowed for a direct assessment and comparison of phosphatase activities on phosphorylated serine, threonine, and tyrosine residues. The phosphatase activity profiles on all three arrays from the NIH/3T3 cell lysate are shown in Figure 3-2.

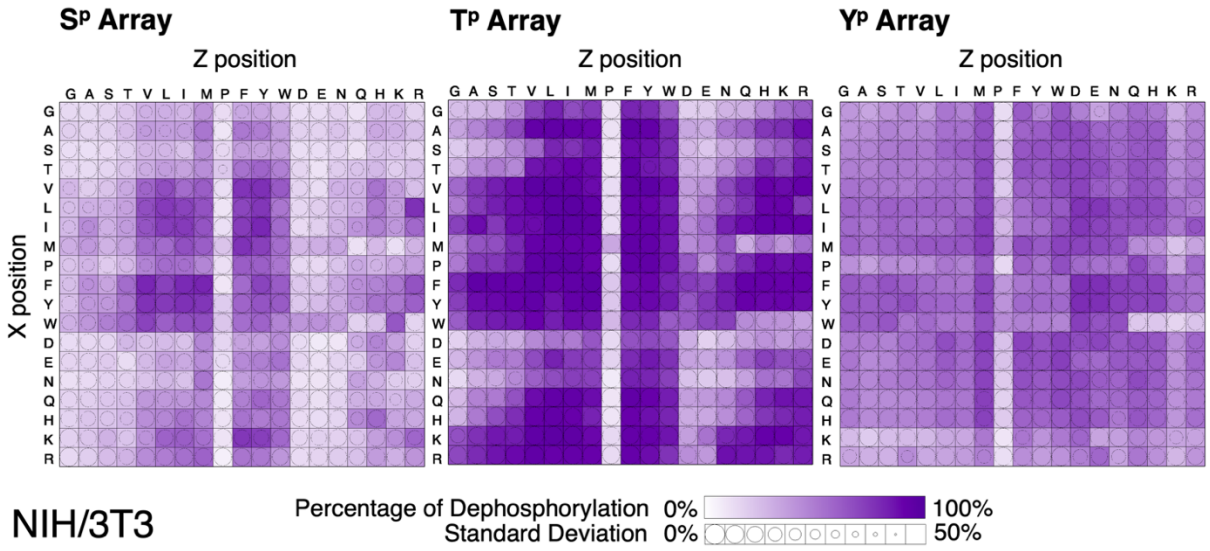


Figure 3-2. Heatmaps showing global phosphatase activity in cell lysate from the NIH/3T3 cell line. The average extent of dephosphorylation of each peptide is plotted in the heatmap, where complete dephosphorylation is dark purple, and each square represents a peptide having X and Z residues as denoted on the vertical and horizontal axes, respectively. This experiment was repeated 3 times and the standard deviation of the dephosphorylation of each peptide is displayed by circle size in each peptide square, with larger standard deviations resulting in smaller circles. The proportion of dephosphorylated peptide was

calculated by dividing the area under the curve (AUC) of the peak for the dephosphorylated product by the total AUC of the peaks for unreacted substrate and dephosphorylated product.

Inspection of the heatmaps first reveals striking differences in activity between the phospho- serine, threonine, and tyrosine arrays. While Ser/Thr phosphatases act on both phospho-serine and threonine substrates, the levels of activity observed on the arrays are very different. The substrate specificities observed on the serine and threonine arrays appear to be similar, but we observed much higher activity on the threonine array than on the serine array. The phospho-serine and threonine arrays show preferential Ser/Thr phosphatase activity on substrates containing hydrophobic and aromatic amino acids in both the X and Z positions (valine, leucine, isoleucine, methionine, proline, phenylalanine, tyrosine and tryptophan). The higher activity in the phosphothreonine array shows disfavored Ser/Thr phosphatase activity on peptides with combinations of glycine, alanine, serine, threonine, aspartic acid, glutamic acid, asparagine and glutamine in both X and Z positions. The phosphotyrosine array revealed low Tyr phosphatase specificity, where most peptides were dephosphorylated between 40-60%; however, we note slightly lower Tyr phosphatase activity on substrates with lysine or arginine in either variable position. Additionally, on all three arrays, we observe very little activity on peptides containing proline in the Z position. While this observation may be related to phosphatase substrate specificity, we speculate that proline in the Z position causes surface-related steric hindrance of the phosphorylated residue which results in our observed low phosphatase activity.

We note that our measurements are not calibrated, and there are differences in ionization efficiency between unphosphorylated and phosphorylated peptides; however, these differences are

consistent between peptides of various sequences and are therefore not critical to our study which focuses on comparative reactivity across all peptides (Appendix Figure S-4). Additionally, we observe no kinase activity in our cell lysates without the addition of ATP, a required co-factor for kinase activity (Appendix Figure S-5).

Following our analysis of global phosphatase activity in NIH/3T3 cell lysates, we extended our study to measure phosphatase activity in cell lysates from four additional mammalian cell lines: HT-1080 (fibrosarcoma cells from human connective tissue), MCF-7 (breast cancer cells: ER^{+}, $PR^{+/-}$, $HER2^{-}$), BT474 (breast cancer cells: ER^{+}, $PR^{+/-}$, $HER2^{+}$), and MDA-MB-231 (breast cancer cells: ER^{-}, PR^{-}, $HER2^{-}$). Each cell lysate was applied to the phospho-serine, threonine, and tyrosine peptide arrays and phosphatase activity on each array was analyzed by MALDI-MS. The corresponding heatmap profiles are shown in Figure 3-3, and larger individual heatmaps for each cell lysate are shown in Appendix Figures S-6, S-7, S-8, and S-9, respectively. All data values for each peptide from all five cell lysates are listed in Appendix Table S-1, Table S-2 and Table S-3.

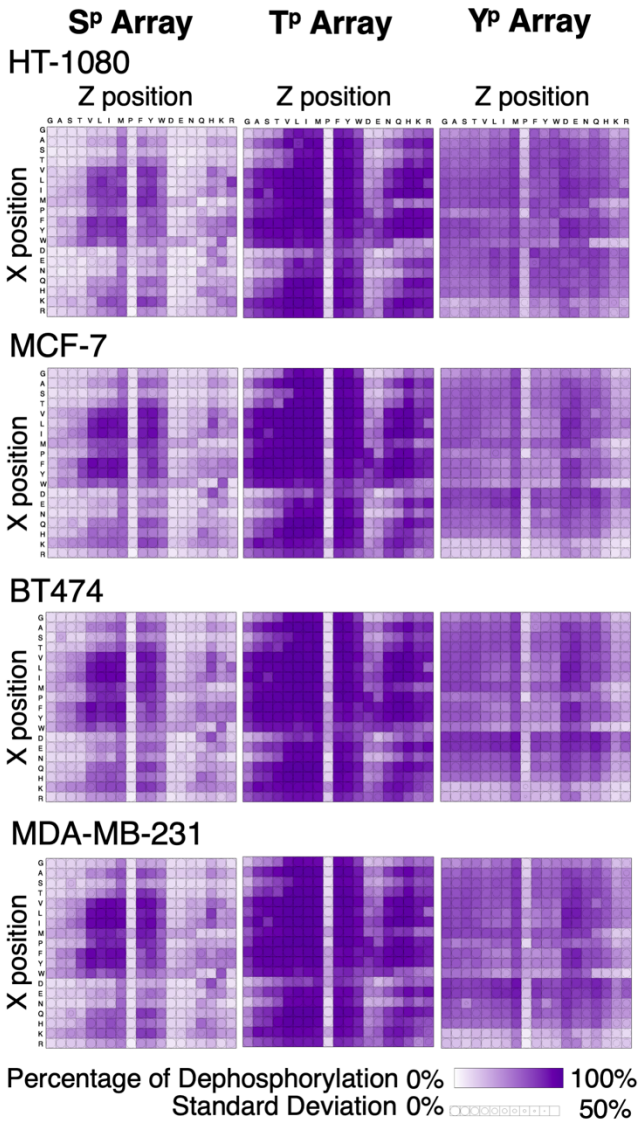


Figure 3-3. Heatmaps showing the global phosphatase activity in cell lysates from HT-1080, MCF-7, BT474, and MDA-MB-231 cell lines. Three replicates were performed for each cell lysate and the average extent of dephosphorylation of each peptide is plotted in the heatmaps, where complete dephosphorylation is dark purple, and each square represents a peptide having X and Z residues as denoted on the vertical and horizontal axes, respectively. The standard deviation of the dephosphorylation of each peptide is displayed by circle size in each peptide square, with larger standard deviations resulting in smaller circles. The proportion of dephosphorylated peptide was calculated by dividing the area under the curve (AUC) of the peak for the dephosphorylated product by the total AUC of the peaks for the unreacted substrate and

dephosphorylated product. Larger heatmaps for each of these cell lines are shown in Appendix Figures S-6, S-7, S-8, and S-9.

We observed consistent differences in activity between the phospho- serine, threonine, and tyrosine arrays across the additional four cell lysates. In fact, across all five cell lysates, the highest variance observed in the dephosphorylation of a single peptide was 20%, with most peptides having variances between 1-10% (Figure 3-4).

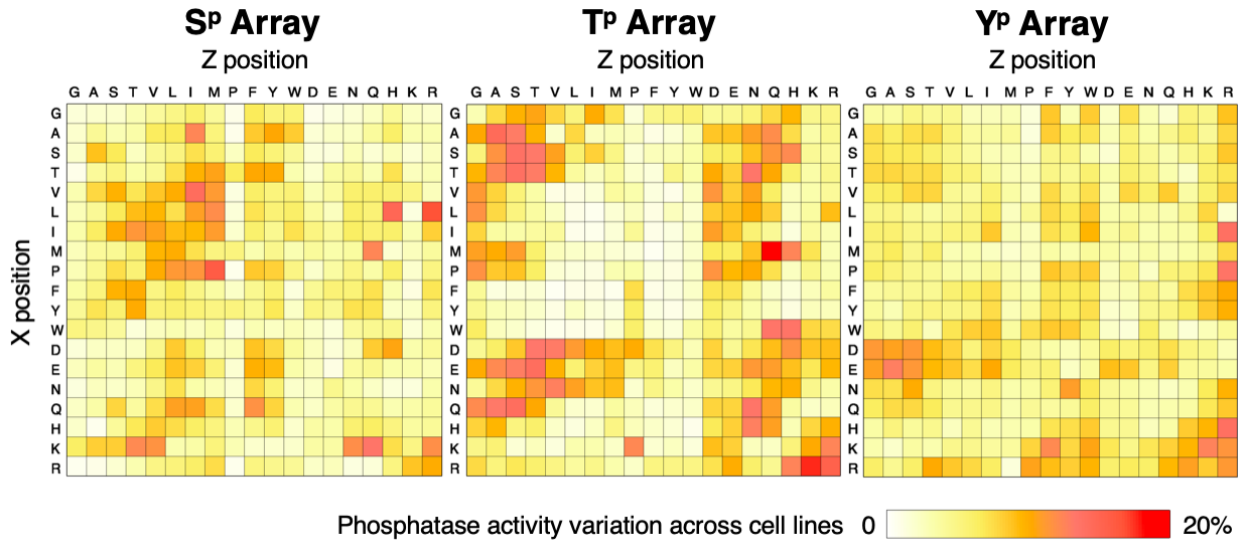


Figure 3-4. Heatmaps displaying the variation of phosphatase activity on each peptide in the phosphorylated peptide arrays. The differences in phosphatase activity on each peptide observed from all five cell lines is plotted in the heatmaps. Each square represents a peptide having X and Z residues as denoted on the vertical and horizontal axes, respectively. The variance in phosphatase activity observed on each peptide is shown on a color scale range of white-yellow-orange-red, where white represents 0% variance and red represents 20% variance.

Discussion

Our global phosphatase profiles show preferential activity on phosphothreonine compared to phosphoserine substrates. As mentioned earlier, several proteomic studies have found that in various cellular conditions, phosphorylation occurs with ~86.4% on serine, ~11.8% on threonine, and ~1.8% on tyrosine residues.[180, 181] Proteomic studies for the detection of phosphorylated sites offer an unbiased view of the *in vivo* proteome; however, they do not provide information on the enzymes responsible for the addition and removal of the phosphorylation modification or the stability and regulatory roles of each phosphorylation site. Previous reports suggest the phosphoproteome distribution and the prevalence of phosphorylation on serine residues is likely due to the higher number of Ser/Thr kinases, with serine being a better phosphoacceptor than threonine, and therefore the preferred site for kinases. However, our results suggest that phosphatases, and more specifically Ser/Thr phosphatases, play a much larger role in the phosphoproteome distribution than previously believed, and that their lowered activity on phosphoserine substrates is likely partially responsible for the high levels of phosphoserine observed.

The high specificity and lower reactivity of phosphatases on the phosphoserine array may indicate that phosphorylation on serine is more regulated than on threonine, and/or that Ser/Thr phosphatases generally have a stronger preference for phosphothreonine over phosphoserine substrates, both of which could explain why proteomic studies report the highest level of phosphorylation on serine. Previous reports have also observed the Ser/Thr phosphatase preference for threonine substrates. For example, Merlevede and coworkers found that members of the PP2A

family strongly preferred phosphothreonine over phosphoserine substrates and that phosphatase activity on phosphoserine could be increased by changing the proximal amino acid sequences surrounding the phosphoserine, while activity on phosphothreonine was less affected by changes in proximal amino acid sequences.[190] Additionally, Burgess and coworkers observed in a proteomic study that phosphorylated threonine sites with proline in the +1 position were dephosphorylated at a higher ratio than phosphoserine-proline motifs.[191] The findings that Ser/Thr phosphatases often have higher preference toward phosphothreonine and that phosphatase activity on phosphoserine substrates is largely affected by surrounding amino acid sequences, is consistent with our results and may indicate that the high proportion of phosphoserine in the phosphoproteome is due to the lack phosphatase activity on phosphoserine substrates.

Our additional observation of minimal variation in phosphatase activity between cell lysates brings into question the number of phosphorylation sites that have important regulatory roles. The vast majority of phosphorylation sites have yet to be examined for functionality and importance; however, several reports have speculated that many phosphorylation sites are non-functional.[192] Many proteomic studies that have monitored global phosphorylation in various cellular processes only found small variations across a myriad of conditions. For instance, proteomic studies that examined global phosphorylation changes in various mouse tissues,[193] and throughout mitotic exit,[191] epidermal growth factor stimulation,[180] and DNA damage response,[194] all found that global phosphorylation was only altered between 10-15%. The finding of only small changes in global phosphorylation throughout various cellular states is in agreement with our observation of little variation in phosphatase activity across five different cell lines. The similar,

but specific phosphatase profiles that we observed across all five cell lysates likely indicate that phosphatase activity is similarly regulated in different cell types and could also suggest that many phosphorylation sites are non-functional. The mechanisms behind phosphatase specificity remain largely unknown; however, it is clear from our results, as well as from others, that phosphatases are highly regulated enzymes.

Conclusions

The combination of peptide arrays and SAMDI-MS is a significant approach to studying phosphatase activity because it allows for direct relative quantification of phosphatase activity in complex samples on a large number of peptide substrates without the need for labels or tedious purification. Using three phosphorylated peptide arrays and SAMDI-MS, we measured phosphatase activity in cell lysates from 5 cell lines on >1,100 peptides and compared the relative activities on serine, threonine, and tyrosine substrates. We found that phosphatase activity is lower on phosphoserine in comparison to phosphothreonine, which may suggest that the phosphorylation distribution across serine, threonine and tyrosine residues is largely impacted by phosphatase activity, rather than being solely determined by kinases. It is clear that phosphatases have significant regulatory roles in the phosphorylation cycle; however, further studies on the substrate specificities and dynamics between kinases and phosphatases are necessary to fully decipher the mechanisms behind phosphorylation and the regulatory roles of each site.

Methods

Peptide Library Synthesis. Peptide libraries, Ac-GXS^{p}ZGRC, Ac-GXT^{p}ZGRC and Ac-GXY^{p}ZGRC, were synthesized using standard Fmoc-based solid-phase peptide synthesis[172] on Fmoc-Rink Amide MBHA resin (Anaspec). Synthesis was performed in 96-well filter plates (Arctic White) with the use of a 96-well plate vacuum manifold (Millipore). Fmoc- and side chain-protected amino acids and N-acetyl-glycine were purchased from either Anaspec or Sigma-Aldrich. Phosphorylated amino acids: Fmoc-Ser(PO(OBzl)OH)-OH, Fmoc-Thr(PO(OBzl)OH)-OH, and Fmoc-Tyr(PO(OBzl)OH)-OH were purchansed from Sigma-Aldrich. Coupling reagents, Pybop and N-methylmorpholine (NMM), as well as dimethylformamide (DMF) were purchased from Fisher or Sigma-Aldrich.

For each library, 15 mg of resin was put into 361 wells in four 96 well plates. Resin was swelled for 1 hour in DMF. The resin was Fmoc-deprotected with a solution of 20% piperidine in DMF for 20 minutes. Using vacuum filtration, the wells were drained, and the resin was rinsed with DMF. The first coupling reaction was performed for 20 minutes with Fmoc-Cys(Trt)-OH, Pybop and NMM in 250 μL DMF at a molar excess of 8:8:16, respectively, to the molar loading capacity of the resin. The wells were then drained, and the resin was rinsed with DMF. The coupling was repeated once more to ensure complete cysteine addition. The wells were again drained, and the resin was rinsed with DMF. Fmoc-deprotection, washing, and double coupling reactions with respective amino acids were repeated until the last amino acid (N-acetyl-glycine) was coupled. We note that β-elimination can occur on protected phosho- threonine and serine amino acids during synthesis. To prevent β-elimination after coupling the protected

phosphorylated serine and threonine, the remaining Fmoc-deprotections and couplings were performed very carefully, and strictly for only 20 minutes. After the last coupling reaction, the resin was rinsed with DMF and dichloromethane and dried under vacuum for 1 hour. Amino acid side chain deprotection and peptide cleavage from the resin was performed in tandem for 2 hours using a solution of 2.5% triethylsilane, 2.5% H_2O, and 95% trifluoroacetic acid (TFA). The cleavage solutions for each peptide were then transferred into new 96-well plates and then evaporated under nitrogen. The peptides were resuspended in 0.1% TFA in H_2O and lyophilized. We found that phosphorylated peptides do not store well in water. After lyophilization the peptides were resuspended again in 0.1% TFA in H_2O to a final concentration of 500 μM, and 3 μL of each peptide were transferred into several 50 μL 384 plates. All peptide library plates were again lyophilized and stored at -80 °C until needed for peptide array preparation.

Additional Peptide Synthesis. Fifty-eight additional peptides were synthesized—one of which was used for ATP-dependence kinase activity control experiments: Ac-IYGEFKKKC. The remaining fifty-seven peptides were used to calculate relative MALDI ionization efficiency ratios between the phosphorylated and unphosphorylated peptide peaks and consisted of the sequence Ac-GX-S/T/Y-GGRC, where X represents all natural amino acids except for cysteine. All peptides were synthesized similarly to the peptide libraries, using standard Fmoc-based solid-phase peptide synthesis on Fmoc-Rink Amide MBHA resin. The resin was swelled for 60 minutes in DMF. Fmoc-deprotection and amino acid coupling reactions were performed as described above for the peptide library synthesis. The peptides were cleaved, and the amino acid side chains were deprotected in a solution of 2.5% triethylsilane, 2.5% H_2O, and 95% TFA for 2 hours. The cleaved

peptides were filtered to remove the resin and then precipitated with ethyl ether. The precipitated peptides were resuspended in 0.1% TFA in H_2O and lyophilized. They were then resuspended in 0.1% TFA to a final concentration of 500 μM and stored at -20 °C until needed.

Preparation of SAMDI Peptide Arrays. Preparation of SAMDI peptide arrays have been described previously.[10, 173-175] Briefly, using a Thermionics E-beam evaporator, 384 titanium spots were evaporated onto steel plates. Gold was then evaporated over the titanium spots, resulting in array plates of 384 gold spots. The plates were soaked in a 1 mM total disulfide monolayer solution of 0.8 mM tri(ethylene glycol) disulfide and 0.2 mM tri(ethylene glycol)-maleimide disulfide in ethanol, allowing for self-assembly of an alkanethiolate monolayer onto the gold surfaces. The monolayer presents a functional maleimide group at a density of 10% against a background of tri(ethylene glycol), which prevents non-specific protein adsorption to the surface.

The lyophilized peptide libraries were resuspended in 50 mM Tris pH 7.5 to a concentration of 50 μM and 3 μL of each peptide solution was pipetted onto the gold spots of array plates. The peptides were incubated on the surface of array plates in a humidified chamber at room temperature for 1 hour for immobilization. Peptide immobilization occurs through 1,4-Michael addition of the cysteine-thiol side chain to the maleimide. The arrays were then washed with H_2O and then ethanol, dried under nitrogen, and stored in vacuum sealed bags at 4°C until ready for use.

Cell Culture. Cell lines were acquired by ATCC. For preparation of lysates, cells were grown and passaged in similar conditions. HT-1080, BT-474, MCF-7, and MDA-MB-231 cells were cultured in Dulbecco's Modified Eagle Medium (Gibco) supplemented with 10% fetal bovine serum (Gibco) and 1% penicillin-streptomycin (Gibco). NIH/3T3 cells were supplemented with

10% bovine calf serum (Gibco) instead. All cells were grown at 37 °C and 5% CO_2. DMEM for all cells came supplemented with 4.5 g/L D-glucose, 4 mM L-glutamine, and 110 mg/L sodium pyruvate.

Cell Lysis. Cells were grown to ~90% confluency for lysis. Cells were resuspended in cold phosphate-buffered saline (Fisher Scientific) on ice following trypsinization. After pelleting this solution, cells were lysed in cold lysis buffer (25 mM HEPES pH 7.4, 50 mM NaCl, 1 mM EDTA, 1 mM DTT, 5% glycerol, 0.5% Triton-X 100, EDTA-free protease inhibitor (Roche, Mini cOmplete™, 1 tablet per 10 mL lysis buffer) on ice for 10 minutes. The DTT and protease inhibitor tablets were added fresh before every lysate collection. The lysate solution was put into a 7mL dounce homogenizer (KONTES), where it was sheared with 25 strokes on ice. Lysate was spun at 1500*g* for 3 minutes, and the supernatant was removed. The supernatant was spun again for 20 minutes at 14,000*g* and 4 °C. The remaining supernatant collected was the final lysate used in our experiments and was stored on ice before use. To estimate lysate protein concentrations for experimentation, a BCA assay (Pierce) was performed using a NanoDrop™ Spectrophotometer (Thermo Scientific). Typical protein concentrations were around 2-4mg/mL from a near-confluent 225cm^2 flask. Cell lysates were flash frozen in liquid nitrogen and stored at -80 °C until needed for experiments.

Lysate Array Experiments. Cell lysates were thawed on ice and then diluted to a total protein concentration of 0.5 mg/mL in lysis buffer. 5 mM $MgCl_2$, 2.5 mM $MnCl_2$, and 1 mM $CaCl_2$ were added to the lysate solution. A multidrop Combi Reagent Dispenser was used to pipette 1.5 µL of lysate onto each peptide spot on the phospho-serine, threonine, and tyrosine SAMDI peptide arrays

and incubated in a humidified chamber for 15 minutes at 37 °C. The arrays were then rinsed with a solution of 0.5% w/v of Alconox detergent (Sigma-Aldrich), then water and dried under nitrogen. The plates were stored in vacuum sealed bags at 4 °C until analyzed by SAMDI-MS. Cell lysate experiments were repeated on all three phosphorylated peptide arrays three times for all five cell lines, and the results are listed in Appendix Tables S-1, S-2, and S-3. Additionally, to make sure the freeze-thaw cycle didn't diminish phosphatase activity, we compared the phosphatase activity in fresh and frozen cell lysate from BT474 cells. We observed negligible variation between fresh vs frozen lysate, indicating that one freeze/thaw cycle did not disrupt phosphatase activity (Appendix Figure S-10).

Control Experiments. Measuring ionization efficiency of phosphorylated vs unphosphorylated peptides: The 57 unphosphorylated peptides consisting of the sequence Ac-GX-S/T/Y-GGRC, where X represents all amino acids except for cysteine, were neutralized by dilution in 50 mM Tris, pH 7.5 to a final concentration of 50 μM. The unphosphorylated peptides and their phosphorylated counter peptides from the arrays were each immobilized onto individual gold spots on a SAMDI array plate as described above. The plate was rinsed with water, and then ethanol, and dried under nitrogen. Each peptide was analyzed by SAMDI-MS as described below. The area under the curve (AUC) of all adducts of the peptide substrate was standardized against the area under the curve of the adducts from the EG3 background peaks. The EG3 peak is a good standard because it is at 90% density on every SAM, and its relative intensity to the immobilized peptide is constant from spectra to spectra. The area under the curve of the peptide was divided by the area under the curve of EG3, and ratios between the phosphorylated and unphosphorylated peptides

were generated. We found that most unphosphorylated peptides had an ionization efficiency about 2-4 times greater than that of the phosphorylated peptide. We plot the standardized ratios for all 19 phosphorylated/unphosphorylated pairs in Appendix Figure S-4.

Demonstrating that ATP is required for kinase activity in cell lysates: We synthesized one peptide that is a known biological substrate of the common tyrosine kinase, Src: Ac-IYGEFKKKC.[195] The peptide was neutralized by dilution in 50 mM Tris, pH 7.5 to a final concentration of 50 μM, and immobilized onto 6 spots on a SAMDI array plate. The peptide was treated with lysate solutions with and without ATP. NIH/3T3 cells were lysed in the same lysis buffer used to measure phosphatase activity with the addition of phosphatase inhibitors: 80 mM sodium fluoride, 100 mM β-glycerol phosphate, 50 mM sodium pyrophosphate, and 5 mM sodium orthovanadate, to prevent all phosphatase activity. The lysate was diluted to a total protein concentration of 0.5 mg/mL. To make sure kinase activity could be observed, we added 100 nM of active Src kinase (Sigma-Aldrich) and 20 mM $MgCl_2$ to the lysate. The immobilized peptide was treated with the lysate and Src solution with and without the addition of 2.5 mM ATP for 1.5 hours at 37 °C. The SAMDI plate was rinsed with a solution of 0.5% w/v of Alconox detergent, then water, and dried under nitrogen. The plate was analyzed by SAMDI-MS as described below, and we only observed kinase activity with the addition of ATP. The spectra showing the yield of phosphorylation of the peptide in both conditions are shown in Appendix Figure S-5.

SAMDI-MS. A matrix solution of 10 mg/mL 2,4,6-Trihydroxyacetophenone, 5 mg/mL ammonium citrate dibasic in 0.1% TFA in 50% acetonitrile and 50% water was prepared fresh. 1 μL of matrix solution was applied to each spot on all treated peptide array plates using a multidrop

Combi Reagent Dispenser, and the matrix crystalized at room temperature and pressure over 15 minutes. The arrays were analyzed in reflector positive mode on an AB Sciex 5800 MALDI TOF/TOF. Spectra were collected from 900 laser shots with a stage velocity of 2000 μm/sec.

Data Analysis. The proportion of dephosphorylation (or phosphorylation for control experiments) on each peptide was calculated by dividing the area under the curve (AUC) of the dephosphorylated peptide product divided by the AUC of both the dephosphorylated product and phosphorylated substrate. All replicates were averaged in presented results. Phosphoserine, phosphothreonine, and phosphotyrosine array data for all five lysates are shown in Appendix Table S-1, Table S-2, and Table S-3, respectively. The peptide array data is also displayed in heat maps, which are a graphical representation of the average and standard deviation of the proportion of dephosphorylation of each peptide in the array. Each square represents an individual peptide in the array with the X and Z position amino acid son the vertical axis and horizontal axis, respectively. Dark purple represents complete dephosphorylation and the circle size inversely represents standard deviation.

Chapter 4. Machine learning on Signal to Noise Ratios Improves Peptide Array Design in SAMDI mass spectrometry

This work was done in collaboration with Albert Xue and Neda Bagheri. The research and all figures presented in this chapter have been published and are adapted with permission from Xue, A. Y., Szymczak, L. C., Mrksich, M., and Bagheri, N. (2017) Machine Learning on Signal-to-Noise Ratios Improves Peptide Array Design in SAMDI Mass Spectrometry, *Anal. Chem. 89*, 9039-9047. Copyright 2017 American Chemical Society. [196] https://pubs.acs.org/doi/10.1021/acs.analchem.7b01728 Further permissions related to the material cited should be directed to the ACS.

Introduction

Peptide arrays have emerged as an enabling tool for identifying biologically relevant peptide substrates and molecular recognition sites, and hold great promise as a new analytical method for basic and translational research in the biomedical sciences.[197, 198] Applications of peptide arrays include screening for active substrates of enzymes, profiling enzyme activity in complex samples, epitope mapping for antibody binding, measuring interactions between proteins and peptide ligands, and identifying substrates for cell adhesion.[17] Many existing methods for peptide array analysis rely on either radioisotopic, colorimetric, or fluorescent labels to detect reaction products.[199, 200] The requirement of labels introduces additional protocol steps and can alter natural biological activity leading to false interpretations, as when resveratrol was erroneously found to enhance deacetylation on a peptide with an attached fluorophore.[97]

We recently introduced SAMDI based peptide arrays, a label-free method that relies on MALDI mass spectrometry to analyze peptides that are immobilized to a self-assembled monolayer of alkanethiolates on gold, and we have demonstrated the use of this method for profiling enzyme specificities,[174] discovering new enzymes,[7] and profiling activities in a lysate.[16] The SAMDI-MS method offers many advantages for peptide arrays, including the use of surface chemistries that are intrinsically inert to the non-specific adsorption of proteins, the availability of a broad range of chemistries for immobilization of peptides, and, most significantly, the compatibility with MALDI mass spectrometry to analyze the masses of the peptide-alkanethiolate conjugates. The ability to directly measure peptide masses allows for a straightforward analysis of peptide modifications by identifying the corresponding mass shifts.[173] This method provides a semi-quantitative measure of the peptides' substrate activity, however the ionization efficiency of a peptide often depends on its amino acid sequence, resulting in different signal-to-noise ratios (S/N) for individual peptides in the array.[16]

Peptides with poor ionization efficiency have low S/N, and therefore often do not provide useful information when relying on mass spectrometry for reaction product detection.[201, 202] To gain better insight into optimizing future peptide array designs, we evaluated the effects of peptide sequence on S/N. We predicted S/N from amino acid sequences measured by SAMDI mass spectrometry, and identified amino acids associated with high S/N peptides in two peptide arrays. We then used machine learning to highlight properties that predict the relationships between amino acids and S/N. While SAMDI-specific results are not generalizable, the method we describe can be adapted and applied to diverse peptide array technologies.

Previous work has explored S/N relationships involving peptide charge (as with arginine residues)[203, 204] or hydrophilicity, where hydrophilic proteins can be preferentially detected in MALDI-MS due to easier co-crystallization with MALDI matrix.[205, 206] In addition to hydrophilicity, many specific and complex peptide-matrix interactions can effect MALDI peptide S/N.[204, 207, 208] Single amino acids have been reported to improve signal strength. For example, Krause and co-workers reported that peptides containing arginine or phenylalanine typically contributed to higher MALDI signal strength.[209] The relationship between S/N and amino acid sequence is also affected by the addition of chemical adducts. For instance, Kolarich and coworkers reported that peptides with attached N-glycans have altered signal strengths depending on the type of glycosylation and MS instrument.[210] Many studies use peptides that may have undergone oxidation[127, 209, 211, 212] which likely also affects peptide signal strength. Peptide modifications, such as glycosylation, oxidation, and phosphorylation introduce difficulties in signal detection and emphasize the need to integrate computational strategies to better understand the relationship between the amino acid sequence of a peptide and the quality of its signal. We selected two peptide libraries that are unbiased in their composition to evaluate differences in S/N due to differing amino acid sequences, and we offer a complete empirical analysis relating amino acid composition and S/N of the peptides.

Using statistical and machine learning strategies, we investigated how amino acid composition affects S/N in SAMDI mass spectrometry and how subtle amino acid differences can give rise to different S/N. We focus on two peptide arrays, each containing two consecutive variable positions, X and Z, that are represented by all 19 amino acids except for cysteine. The amino acids

surrounding the variable positions however are different. The two peptide arrays are Ac-GRK^{ac}XZC (K-array) and Ac-GXZHGC (H-array). We collected peptide spectra by SAMDI mass spectrometry and calculated the S/N of each peptide. Statistical analysis identified amino acids associated with low or high S/N peptides. We trained machine learning models to identify factors that predict S/N from the physical properties of the peptide's amino acids. We then used the model to predict the S/N of peptides that were experimentally evaluated by SAMDI high-throughput data. Accurate prediction of peptide S/N from machine learning models allows for the pre-selection of peptides that are well suited for inclusion in an array without costly experimental screening.

Results and Discussion

Experimental design. We calculated peptide S/N using SAMDI mass spectrometry for two peptide arrays: Ac-GRK^{ac}XZC (K-array) and Ac-GXZHGC (H-array) where X and Z represent all combinations of the 19 natural amino acids excluding cysteine for a total of 361 peptides in each array. To investigate the relationships between specific amino acids and S/N, we conducted statistical tests and machine learning. We then applied the statistical test and machine learning results to a published peptide array data set to reveal how S/N information can serve as a guide for experimental design and analysis. In doing so, we discovered specific amino acid interactions that can explain observed S/N-amino acid relations. Through subsequent machine learning analysis, we identified physical properties and amino acid positions that predict the experimentally observed peptide S/N. Finally, we used our machine learning model to predict the S/N of an unknown array and a partially synthesized array.

Preparation of peptide arrays. We used solid-phase peptide synthesis to synthesize two peptide libraries containing terminal-cysteine residues, Ac-GRK^{ac}XZC (K-array) and Ac-GXZHGC (H-array), where X and Z represent all amino acids except cysteine. Steel plates with 384 gold spots were soaked in a solution of disulfides as described earlier.[9] The monolayers self-assembled onto the gold surfaces and presented a functional maleimide group allowing for the immobilization of thiol-containing molecules. We treated each monolayer surface with a unique peptide, which was immobilized to the surface through the side-chain thiol of the terminal cysteine residue. Eleven identical arrays were printed for the experiments that follow.

S/N is attributable to single amino acids in the K-array. Comprehensive analysis of the K-array revealed general effects of single amino acids in a peptide on the observed S/N. We collected spectra for each immobilized peptide using MALDI mass spectrometry. Noise was quantified as the area under the curve (AUC) of the mass spectrum in a region devoid of signals, and the peptide signal was quantified as AUC of the expected peptide-terminated alkanethiol mass minus the noise AUC. Finally, we calculated S/N as the peptide's signal AUC divided by the noise AUC and calculated the mean for each peptide over the eleven plates in each array.

We used the Fischer exact test (Bonferroni corrected $p < 10^{-4}$) to determine whether peptides with low or high S/N were enriched with specific amino acids. The corresponding *p*-values reflect the probability that the observed number of amino acids is within either the low or high S/N regions by random chance (gray regions in Figure 4-1A). All *p*-values are reported in Appendix Figure S-11. We found that the majority of peptides containing tryptophan and leucine in the X-position were in the low S/N region and most peptides with Z-position glycine to have high S/N (Figure

4-1B). These results indicate that single amino acids can have a strong effect on a peptide's detectability in MALDI-MS. The exceptionally low S/N of tryptophan and leucine-containing peptides suggests that the design of future arrays similar to the K-array should consider disregarding tryptophan and leucine.

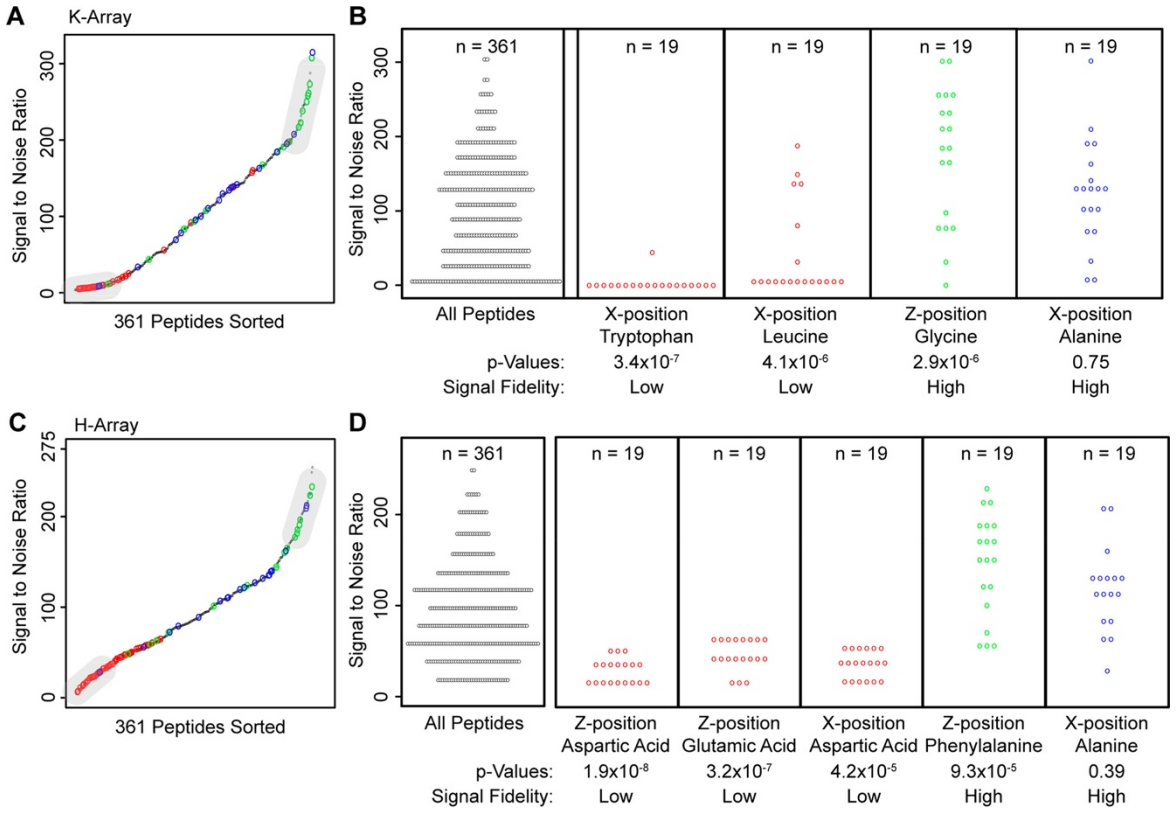


Figure 4-1. General effects of specific amino acids on peptide S/N in the K- and H- arrays. Low peptide S/N is observed in peptides containing tryptophan and leucine in the K-array and aspartic acid and glutamic acid in the H-array. Peptide S/N was averaged over 11 control plates. (A) Peptides in the K-array were sorted according to S/N. Low/high S/N regions are identified (see Methods). S/N ranges from 3.8 to 313.7, demonstrating that peptides vary greatly in S/N. (B) Amino acids found in the low/high regions were found to be statistically significant (Bonferroni corrected $p < 10–4$) using a Fischer exact test. The reported p-value is the chance the observed number of amino acids is within the low or high region by random chance. Peptides with X-position tryptophan and leucine have statistically low S/N, and peptides with Z-position glycine have statistically high S/N. Peptides that have X-position alanine are not statistically

significant and are representative of other amino acids. Panels C and D describe the same methods for the H-array. S/N has a similarly large range for both arrays, but the differences in amino acids observations suggest that dissimilar mechanisms are responsible for S/N.

Peptides in the K-array display a wide range of S/N—from 3.8 to 313.7 (S/N is unitless)—demonstrating a wide range of poorly-detectable to detectable peptides (Figure 4-1A). In MS-based proteomics experiments, proteins are commonly digested, and the fragments are detected using mass spectrometry. It is rare for complete detection of all peptide sequences following digestion,[213, 214] and incorporation of known poorly-detectable peptide information could increase confidence of protein observation. Our observation of widely varying S/N across a peptide array may explain differences in MS-detectable peptide fragments after protein digestion.[215, 216] Our statistical analysis of peptide S/N suggests that poorly-detectable peptides can be predicted by their sequence. As we demonstrate, characterization of a MALDI-MS experimental pipeline with known peptide sequences can inform subsequent protein quantification experiments.

A machine learning model predicts SAMDI-MS S/N as a function of amino acid sequence. We developed a machine learning model to predict the S/N of peptide-terminated alkanethiolates in the SAMDI spectrum based on amino acid sequence with high accuracy suggesting that amino acid composition drives S/N observations in a predictive manner. We trained a random forest machine learning model[217] to predict S/N based on the hydrophilic, steric, and electronic physical properties of amino acids.[218] The training data contained 361 peptides (rows) and 39 associated physical properties for each of the X or Z position amino acids (resulting in 78 columns). The response vector, or predicted variable, defined the mean S/N from the 11 control plates. We used

cross-validation, where a data sample (randomly selected rows) is left out for model testing, to calculate the predictive power Q^2 statistic[219]:

$$Q^2 = 1 - \frac{\sum_i^n (y_i - \hat{y}_i)^2}{\sum_i^n (y_i - \bar{y}_{train})^2}$$

In this formulation, y_i is the true S/N for the omitted test peptide i, $\hat{y}_i$ is the predicted S/N of the test peptide, $\bar{y}_{train}$ is the sample mean of S/N in the training set, and n is the number of cross-validated test peptides. The Q^2 statistic can take on values from -∞ to 1, where 1 represents perfect prediction and 0 is equivalent to random performance. We create an explicit null model for each case by randomizing the data values prior to model training; the average Q^2 value of the null case was about 0.

The analysis of the K-array resulted in a Q^2 of 0.59 using both X- and Z-positions and all 39 amino acid physical properties. The high Q^2 value confirms our hypothesis that S/N values can be reliably predicted from amino acid sequences. This performance further suggests that S/N of new amino acids, such as non-natural amino acids, can be predicted using their known physical properties. Together these results strongly indicate that amino acid sequence influences S/N in MALDI-MS. However, the inability to acquire a Q^2 value closer to 1 suggests that hidden variables—such as chemical interactions with amino acids outside the X- and Z-positions—play an important role in the overall response. These interactions are challenging to take into account, as they cannot be characterized with physical properties alone.

A bubble chart illustrates the S/N as an experimental design parameter. As a measure of data quality, the S/N becomes another experimental design parameter. When studying enzyme activity on a SAMDI peptide array, we measure the extent of peptide conversion by the enzyme.[16] Enzyme-

treated peptides can be sorted into four categories: (i) high enzyme activity and high S/N, (ii) high enzyme activity and low S/N, (iii) low enzyme activity and high S/N, and (iv) low enzyme activity and low S/N. In the past, the SAMDI peptide array data was compiled into heat-maps that portrayed only enzymatic activity. We wanted to incorporate a metric into the output of SAMDI array data to differentiate between peptides that offer reliable and valuable information (category i) from those of lesser importance.

To this end, we include S/N information to complement a previously published experiment.[16] We construct a bubble chart where each peptide is represented by a circle, whose color represents the extent of enzyme activity on that peptide, and whose size represents normalized S/N of the peptide before enzyme treatment. Previous approaches that use a color-only heatmap give the impression that each data point is equally valid in an analysis of the array data; however, some of the peptides contribute information that is more reliable because they have smaller errors. Therefore, observed enzyme activity on a peptide does not always correlate to significance. By incorporating S/N in bubble size, we rule out low performance signals and focus the analysis on high S/N results. We illustrate this approach by replotting the heatmaps from Kuo *et al.* to include S/N (Figure 4-2).[16]

In Kuo *et al*, the K-array was exposed to cell lysates, and endogenous deacetylase activity was quantified by measuring the fraction of deacetylated peptides with MALDI mass spectrometry.[16] Deacetylation activity was quantified as the AUC of the modified (deacetylated) peptide divided by the AUC of both modified and unmodified peptides. AUC of each peptide is the sum of the three background-subtracted ion peaks in MALDI-MS: H+, Na+, and K+.

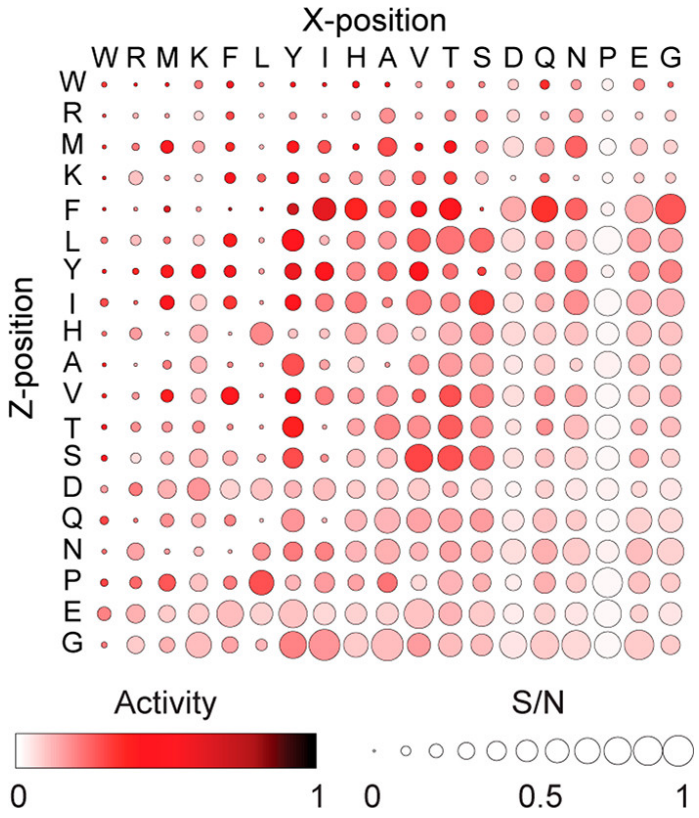


Figure 4-2. Heatmap including S/N and enzyme activity data highlights peptides that provide valuable information. This heatmap shows the S/N and cell lysate deacetylase activity of each peptide from Kuo *et al.*[16] Bubble color is based on deacetylase activities from lysate treated ac-GRK^{ac}XZKC peptide arrays. Bubble area represents peptide S/N before lysate treatment, normalized by max S/N. Amino acids are sorted by their general trend in peptide S/N when in either X- or Z-position. Peptides containing tryptophan (W), arginine (R), methionine (M), and lysine (K) have consistently low S/N, regardless of position. This illustration emphasizes peptides that are both active in terms of enzymatic activity and reliable in terms of S/N. In contrast, the highest activity peptides (darkest in color) do not necessarily give the highest S/N (largest bubble).

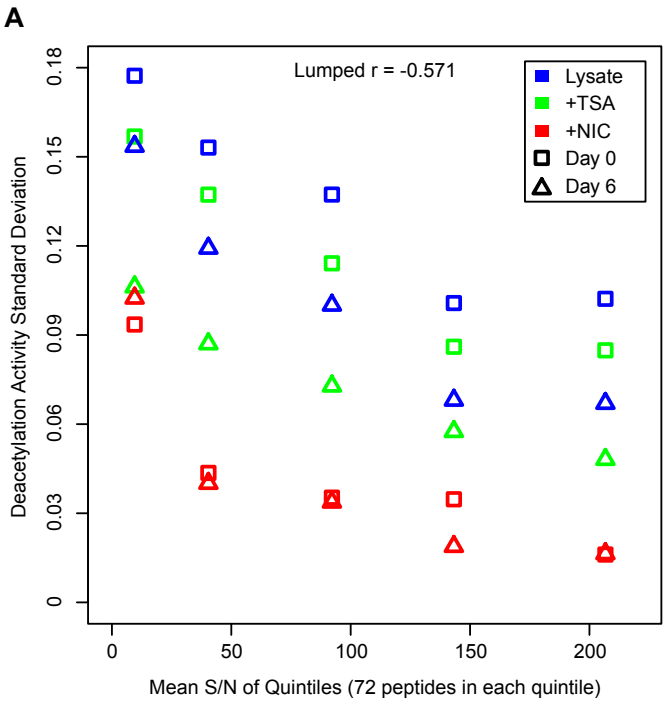
This new analysis revealed additional insights into the previous data. Peptides containing amino acids tryptophan, leucine, arginine, methionine, and lysine reflect low S/N, suggesting that

their activity profiles are less informative. Conversely, peptides containing proline, glutamic acid, and glycine reflect high S/N, suggesting that their activity profiles are more dependable. Peptides containing leucine exhibit low S/N exclusively in the X-position, demonstrating that amino acid position can affect peptide S/N. Though amino acid composition can largely explain a peptide's S/N, we also find that some peptides have inexplicably low S/N—such as KAA, KIT, and KIQ—despite general trends suggesting that these peptides should have high S/N. This peculiarity highlights the complexities in S/N and reinforces the utility of machine learning strategies to predict S/N, which can be a critical design factor for future arrays.

Testing for S/N does not supplant tightly controlled and validated peptide array experiments. Instead, we suggest that accounting for unknown influences that lower confidence of a signal's true value—such as peptide synthesis inefficiencies, side reactions, peptide loss from washing, or ionization efficiencies—can better guide experimental design and data analysis. These influences are especially complicated with peptide species, where it is not clear how different amino acid sequences affect S/N. Machine learning can be used to account for such effects.

Low S/N peptides offer unrepresentative signals. The experiments by Kuo *et al.* demonstrated that peptides with low S/N have higher variance across replicates. The same K-array measurements as describe above were carried out on two time points and across three different cellular conditions. We compared the variance in replicates of peptides with S/N in the top 20% to those with S/N in the bottom 20%. A one-sided F-test verified that the data acquired from peptides with S/N in the top 20% have lower replicate variance than peptides with S/N the lower 20% across all conditions and time points ($p < 10^{-10}$ for all cases). This finding suggests that peptides with low S/N may have

unrepresentative (or possibly random) signals, and they should be weighted less during analysis to avoid misled conclusions. To investigate further and confirm the relation between variance and S/N, we calculated the standard deviation of deacetylase activity on each peptide and plotted it against S/N (Figure 4-3A); peptides with lower S/N have a higher variance in deacetylase activity. This trend was consistent for all timepoints and experimental conditions, with a high anti-correlation coefficient (ranging from -0.814 to -0.975, Figure 4-3B), demonstrating that peptides with low S/N can give unrepresentative measurements.



B

Lysate Condition	Day	r
Lysate	Day 0	-0.948
Lysate	Day 6	-0.957
+TSA	Day 0	-0.843
+TSA	Day 6	-0.942
+NIC	Day 0	-0.975
+NIC	Day 6	-0.814

Figure 4-3. Peptide S/N is anti-correlated with deacetylation activity standard deviation. Data is drawn from Kuo *et al.*[8] Peptides are grouped into five quintiles (72 peptides each) based on S/N, and the standard deviation across replicates for each quintile is calculated. (A) The standard deviation of deacetylase activity versus mean S/N of each quintile is shown for the three cellular conditions of untreated lysate (blue), +TSA (green), and +NIC (red), and for day 0 (square) and day 6 (triangle). This consistent trend suggests that low S/N peptides have higher variations in deacetylase activity, which indicate that those peptides give

unreliable data. (B) The lumped correlation (r) is the Pearson correlation over all conditions and days, and data with randomized signal S/N gives zero correlation. Within each condition and day, signal standard deviation and mean S/N is strongly anti-correlated.

S/N is attributable to single amino acids in the H-array. We investigate the H-array, Ac-GXZHGC, to analyze the generalizability of our findings. More specifically, we examined if the S/N characteristics of peptides from the K-array also apply to other peptide arrays. Similar to the K-array, the H-array has a wide range of S/N values, ranging from 5.5 to 255 (Figure 4-1C and Figure 4-1D), reinforcing the fact that peptides can span a wide range of non-detectable to detectable signals in MALDI. All *p*-values are reported in Appendix Figure S-12. Peptides with statistically low S/N contain aspartic acid and/or glutamic acid, with peptide containing phenylalanine were found to have statistically high S/N. These findings may be important to consider when designing future arrays.

Context matters: S/N characteristics are inconsistent between the K-array and H-array. In addition to the variable composition of amino acids, we found that the surrounding amino acids (those not in the X- or Z-position) affect overall S/N. Within the K- or H-array, the total amino acid composition appears to have a larger effect on peptide S/N, than the specific order of amino acids. In general, we find that a specific amino acid in either the X or Z positions have similar effects on the S/N (Figure 4-4).

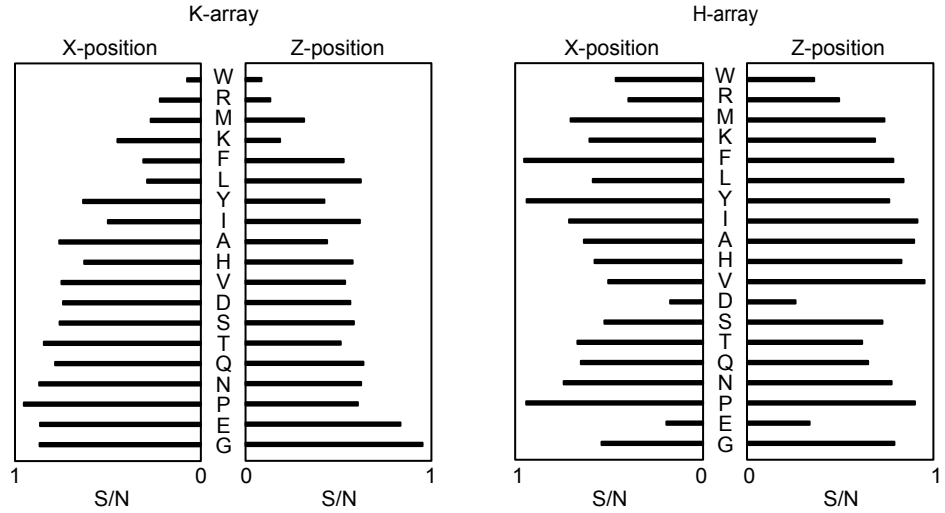


Figure 4-4. Peptide S/N stays consistent between positions, but not between peptide arrays. Peptides with specific amino acids are sorted by their mean S/N when in either X- or Z-positions in the K-array. The two rows of bar plots are the K-array and H- array, and both have the same amino acid order. Within the K-array, S/N is correlated between the X and Z position amino acids. However, the correlation disappears between the H- and K-arrays, demonstrating that the surrounding amino acids have an influence on peptide S/N. Conversely, peptides with certain amino acids, such as proline, have similar S/N values on both peptide arrays which suggests that some amino acids have a consistent effect regardless of surrounding amino acids.

However, general trends in the K-array did not match those for the H-array. Peptides with the lowest S/N in the K-array contained tryptophan, arginine and methionine, while in the H-array, peptides with the lowest S/N contained aspartic acid and glutamic acid. This disparity demonstrates that S/N characteristics in one array can be contextual and are not always generalizable to other arrays (Figure 4-5). These results also suggest that the outside amino acids—arginine and lysine in the K-array and histidine in the H-array—strongly influence SAMDI peptide detection. In other words, the amino acids surrounding the X- and Z-positions influence overall

peptide detection, and partial knowledge or evaluation of amino acid sequence is insufficient in understanding S/N values.

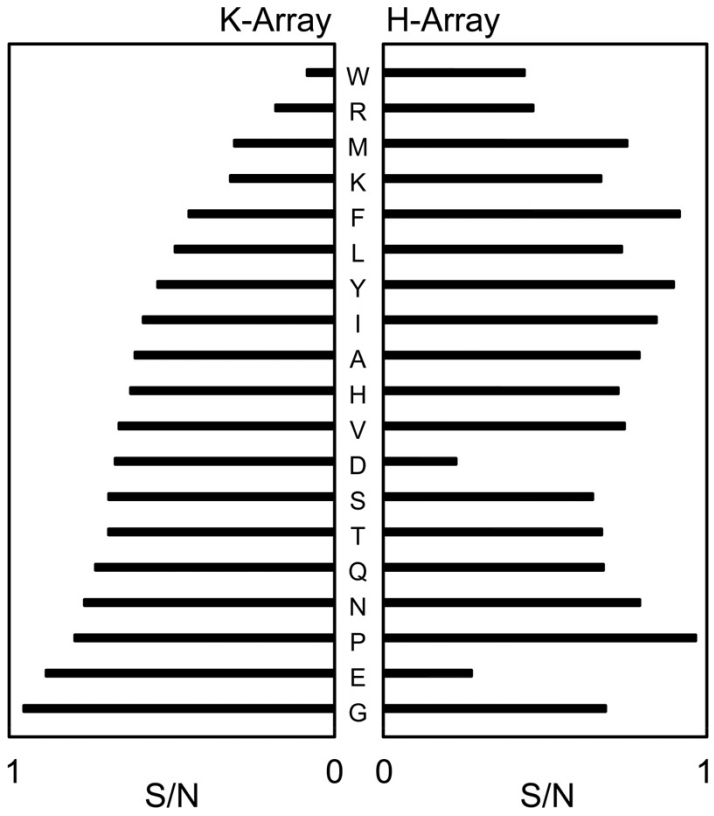


Figure 4-5. Amino acid influence on S/N is context dependent. Amino acids are sorted by their correlated mean peptide S/N in the K-array when in either X- or Z-positions. Bars represent S/N normalized by the highest value within each array. The S/N values from the H-array are very different from those in the K-array, suggesting that the surrounding amino acids strongly influence S/N. The two negatively charged amino acids, aspartic and glutamic acid, have the largest difference between the two arrays, suggesting a relation between charge and S/N, but only within the H-array.

Physical interactions help inform S/N differences. S/N differences can arise from a variety of sources, including synthesis inefficiencies, side reactions, and poor MALDI-MS ionization. Peptides with both methionine and tryptophan have low S/N in the K-array. These results were

unsurprising because methionine and tryptophan are both sensitive to oxidation,[127, 211] which sequesters relevant peaks and lowers signal strength. However, in the H-array, peptides containing glutamic acid and aspartic acid generally have the lowest S/N (Figure 4-5). We speculate that the oxidation on methionine and tryptophan in the K-array could be increased by the hydrogen bonding capabilities of lysine when in close vicinity.[220] Tryptophan-containing peptides have statistically low S/N specifically when in the X-position (Figure 4-1), which is directly adjacent to the lysine and further supports this hypothesis. If hydrogen bonding stabilization is required for methionine or tryptophan oxidation, then the presence of carboxylic acid groups on surrounding amino acids may be unfavorable for oxidation. To explore this concept further, we compared the S/N of methionine and tryptophan containing peptides in the K-array with and without glutamic and aspartic acid. We applied a Mann-Whitney U test and found that peptides with either methionine or tryptophan and glutamic or aspartic acid had higher S/N values (p = 0.0050) than peptides with methionine or tryptophan without glutamic or aspartic acid. This indicated that acidic amino acids may slow the oxidation of methionine and tryptophan residues.

High S/N peptides in the K-array commonly contain hydrophilic amino acids such as glutamic acid, asparagine, and glutamine, potentially due to more efficient crystallization within the matrix. This finding is in agreement with a report by Fenselau and coworkers, where hydrophilic proteins were preferentially detected in MALDI-MS due to differences in the co-crystallization.[205] However, the H-array has high S/N associated with hydrophobic amino acids: proline, tyrosine, phenylalanine, and isoleucine. The divergence in S/N of hydrophobic and hydrophilic amino acids suggests that hydrophilicity does not alone determine S/N.

A two sided Mann-Whitney U test (Bonferroni corrected $p < 2.6\text{x}10^{-3}$) reveals statically different S/N values between the K- and H-arrays for peptides that have at least one of eight specific amino acids in the X or Z position: glutamic acid, tryptophan, aspartic acid, methionine, arginine, lysine, glycine, and phenylalanine. The two sided Mann-Whitney U test directly compares differences between the two arrays rather than within the array, which offers slightly different S/N characteristics for amino acids than from the Fischer exact test in Figure 4-1. Notably, the effects on S/N from arginine and phenylalanine differ from the results shown in Figure 4-1, and both amino acids have lower S/N in the K-array. This result contrasts with those of Krause and coworkers where peptides with higher numbers of arginine or phenylalanine typically contributed to higher MALDI signal strength (the K-array has an additional arginine).[209] The unusual observation may be due to unknown interactions with surrounding amino acids, indicating again that peptide S/N should be tested for each peptide array.

Machine learning performance across positions and physical properties help explain S/N observations. We trained random forest models with individual physical properties to assess the impact each property has on S/N. Highly predictive properties (namely those with highest Q^2) suggest that the associated physical property is highly relevant and predictive of SAMDI-MS S/N. In addition, we independently evaluated the X and Z positions to see if one position reflected more predictive power.

Using both X and Z positions and all physical properties, the K-array and H-array had a Q^2 of 0.59 and 0.61, respectively (Figure 4-6). The similar Q^2 values suggest that the models reached an upper limit to predictive performance from amino acid sequence. Predictions based on the

amino acids in both X- and Z- positions consistently performed better than predictions based solely on one position: Q^2=0.22 and 0.20 for the X- and Z-positions in the K-array, respectively, and Q^2=0.16 and 0.28 in the X- and Z-positions of the H-array, respectively. As expected, more complete amino acid information results in better prediction. However, the higher Q^2 for the Z-position in the H-array suggests that positions can have varied influence on S/N. In addition, the highest single property Q^2 values—0.57 and 0.54 in the K- and H-array, respectively—are close to the Q^2 value of all properties. This observation indicates that only a few properties are necessary to predict S/N and that many physical properties are redundant.

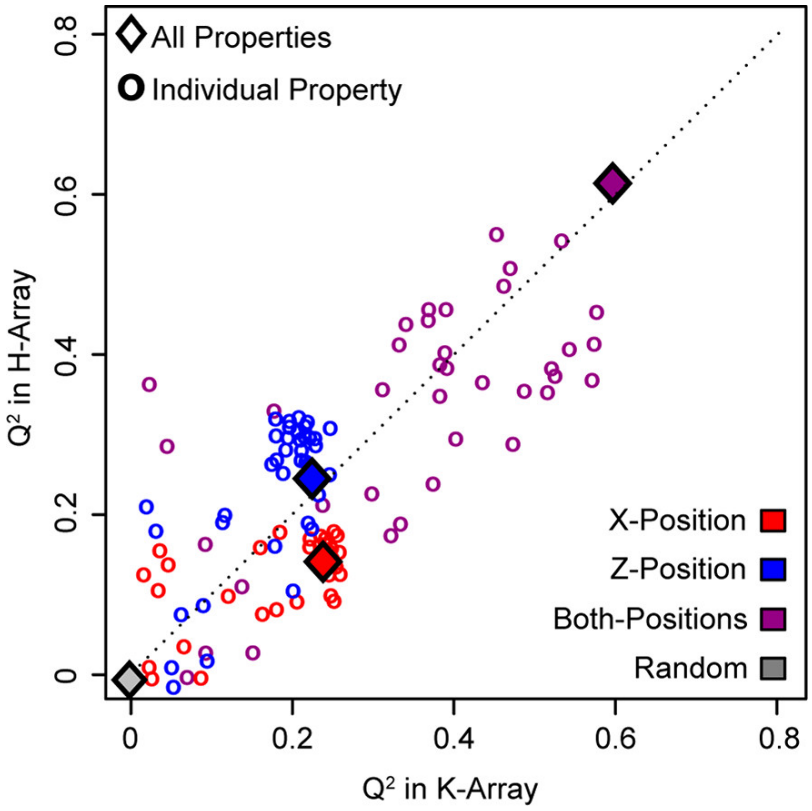


Figure 4-6. Peptide S/N is predicted as a function of amino acid properties. Peptide S/N was predicted using a random forest machine learning model based on 39 amino acid physical properties, shown in diamonds, of the amino acid in either the X-position (red), Z-position (blue), or both (purple). Additionally, models

were fit on individual properties to identify their predictive power, shown in circles. Random forest models contained 1000 trees and the predictive power, quantified by the Q^2 metric, was calculated based on 5-fold cross-validation. All Q^2 values are listed in Appendix Tables S-4 and S-5. Consistent for both peptide arrays, the highest Q^2 values were attained when using both positions with all physical properties (purple diamond). Z-position Q^2 values (blue) are higher in the H-array, which suggests that positions have varied predictive power on S/N. In addition, most properties lie near the diagonal indicating that they have similar predictive power between peptide arrays; the amino acid disagreement in Figure 4-5, however, suggests that those properties may be predictive for different reasons.

In terms of physical properties (39 total), we find both consistent and inconsistent trends between the two arrays. Electronic properties (15 total) tended to be less predictive for both arrays than steric (16 properties) or hydrophilic properties (8 properties), that were found to be highly predictive in the K- and H-arrays, respectively. (Appendix Tables S-4 and S-5). We speculate that hydrophilic properties were found to be highly predictive in the H-array because of the hydrophilicity of glutamic and aspartic acid and their association with low S/N, and the association of hydrophobic amino acids such as proline, tyrosine, phenylalanine, and isoleucine with high S/N. However, it remains unclear why electronic properties were found to be relatively less predictive and steric properties were found to be more predictive in the K-array.

Despite these differences, the physical properties examined were found to be similarly predictive between the K-array and H-arrays, as evident in their closeness to the diagonal in Figure 4-6. The diagonal line indicated that property, whether predictive or not, has similar predictive power between arrays. While the physical properties examined were found to have similar predictive power, we observe differences in S/N characteristics from single amino acids in the two arrays (Figure 4-5). These results altogether suggest that S/N values manifest from different

mechanisms between the arrays, and that the different mechanisms likely result from differences in the surrounding amino acids.

Machine learning cannot predict S/N on completely unknown peptide arrays. We trained various machine learning models on the K-array and then tested them on the H-array, and vice versa, to assess the feasibility of predicting S/N on a *de novo* peptide array. We trained models for every positional combination to interrogate exhaustively the entire space. For example, we trained a model on X-position data in the K-array, then tested the model on the Z-position in the H-array. We continued to train and test models with all combinations of positions. We also trained several types of models based on random forest,[217] deep learning,[221] nearest neighbor regression,[222] and partial least squares regression.[223] The models had the following model-specific parameters: random forest had 1000 trees, deep learning consisted of two layers of 200 nodes with feed-forward connections; nearest neighbors regression used 10 neighbors; and partial least squares regression used one component, or loading vector.

All models trained on the K-array failed to predict S/N in the H-array, and vice versa (Q^2<0.1). This failure is attributable to S/N disagreement between peptide arrays for each amino acid (Figure 4-5), which likely arises from the unique outside amino acids in the two arrays (GRK^{ac}XZC and GXZHGC). This finding reinforces the idea that context matters: interactions with outside amino acids influence S/N.

Only 1/3 of peptides in an array are required for machine learning model prediction of peptide S/N. We investigated the minimum number of peptides in an array needed to train a model that could accurately predict S/N of the full array. We simulated a partially synthesized array by

randomly selecting training peptides to predict S/N of the non-selected peptides. The number of training peptides ranged from 5 to 350, and each training size contained 200 repetitions of selecting random peptides. We trained a random forest model with all 39 physical properties in both amino acid positions. We then identified the point of diminishing returns, which balances minimum training size with maximum predictive power, by normalizing the number of training peptides and finding the sample size closest to training size 1 and $Q^2 = 1$. The point of diminishing returns was found to be 87 and 111 peptides for the K-array and H-array, respectively, both of which had a Q^2=0.48 (shown with arrows in Figure 4-7). The null model that used randomized data performed consistently around Q^2=0. This result shows that we can partially screen future peptide arrays by synthesizing about 100 of the peptides in the planned 361-sized array, or roughly one-third, reducing the use of resources and time. Though we cannot generalize this specific ratio to larger array sizes, these results suggest that only a fraction or an array needs to be synthesized to evaluate which peptides in the remaining array will contain high S/N. This machine learning technique can prevent costly experimental screens and allow researchers to focus on predicted *a priori* high S/N peptides.

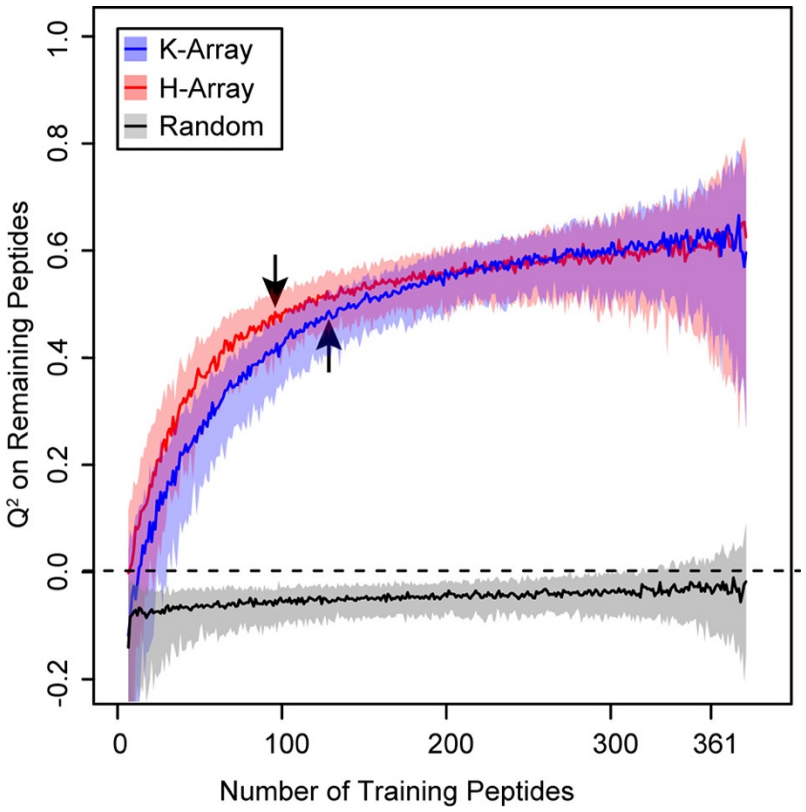


Figure 4-7. Peptide array S/N can be predicted from a minimal peptide subsample. A specified number of peptides were randomly selected for training to predict S/N of the remaining peptides using all physical properties of both X and Z position amino acids. Because of computational constraints, random forest was used with 100 trees for training set sizes from 5 to 350. The median Q^2 and 80% confidence intervals are shown for 200 random training sets. For both peptide arrays, predictive power increases with training size and levels out around 100 peptides. The optimal trade-off was identified by normalizing the number of training peptides and finding the sample size closest to training size 1 and $Q^2 = 1$. The trade-off is shown with arrows: 87 training peptides for K-array and 111 for H-array, which demonstrates that machine learning can predict S/N for future peptide arrays, avoiding costly experiments that screen for high S/N peptides. A randomized data set performed consistently around $Q^2 = 0$.

Conclusion

There are significant variations in the intensities of peaks in SAMDI mass spectrometry that can arise from different peptide sequences. SAMDI analysis of peptide arrays demonstrates that

peptide signals can have a wide range of S/N, where many of the peptides are nearly undetectable. We find that S/N is attributable to single amino acids, offering design choices to increase information content. The underlying basis of properties that determine S/N remains unclear but may be due to complex interactions among amino acids, matrix, crystallization, or ionization efficiencies. Additionally, we find that the two arrays used in this work exhibited different S/N values for different amino acids, demonstrating that the whole amino acid sequence of a peptide can affect S/N values in MALDI-MS. Machine learning identified physical properties that predict S/N with high accuracy. We found that machine learning models can be trained on a fraction of peptides in an array to describe the full set of sequences, which allows for more efficient and higher quality peptide array design and synthesis. Accounting for S/N as a design choice can ensure higher quality results and prevent inaccurate conclusions drawn from poor peptide measurements.

This work significantly improves and simplifies high-throughput data analysis by factoring in data quality. The statistical and machine learning methods presented here allow us to discover the most valuable information from peptide arrays and plan future experiments with more confidence. As demonstrated, these methods can inform the design of new peptide arrays using a small set of peptides. The presented methodology and applications of S/N have been adapted to maximize the information learned from peptide array experiments and can improve peptide design across a wide range of technologies.

Methods

Solid phase peptide synthesis. Data was collected from K- and H-peptide array experiments. The K-peptide array synthesis and methods were published previously[16] and contains peptides of the form Ac-GRK^{ac}XZC, where X and Z represent all combinations of 19 amino acids (cysteine omitted) for a total of 361 peptides. We synthesized another 361 membered unmodified histidine peptide array with the sequence Ac-GXZHGC, referred to as the H-array. The constant amino acids (everything except X or Z) are referred to as the outside amino acids. Peptides were synthesized using standard solid phase peptide synthesis on Fmoc-Rink Amide MBHA resin purchased from Anaspec. Fmoc-protected amino acids were purchased from either Anaspec or Sigma-Aldrich. The Fmoc-Rink Amide resin was swelled in dimethylformamide (DMF) for 30 min and treated with 20% piperidine in DMF for 20 minutes to remove the Fmoc protecting group. The first Fmoc-protected amino acid was coupled to the resin with PyBOP and N-methylmorpholine at a 4:4:8 ratio, which was repeated until all the amino acids were coupled to the resin. Once the Fmoc protecting group was removed from the final amino acid, the resin was treated with 10% acetic anhydride in DMF for 30 minutes to acetylate the N-terminus. The peptide was cleaved from the resin with a solution of 95% trifluoroacetic acid (TFA), 2.5% triethylsilane, and 2.5% milli-q water for 2 hours. To remove the resin, the solution was filtered and precipitated with peptides with ethyl ether. The peptides were re-suspended in 0.1% TFA, lyophilized and re-suspended in 0.1% TFA again. The peptides are neutralized by dilution into 50 mM Tris buffer pH 7.5 before immobilization.

Preparing peptide arrays. Peptide arrays were prepared as described previously.[173, 174] Briefly, steel plates were evaporated with 384 gold spots. The plates were soaked in an alkanethiolate solution that self-assembles onto the gold surfaces. The alkanethiolate monolayers presented a functional maleimide group against a background of tri(ethylene glycol). Peptides were transferred onto the gold spots using Tecan robotics and incubated at room temperature for 1 hour for immobilization. Peptide immobilization occurs through conjugate addition of the thiol on the terminal cysteine residue to the maleimide.

SAMDI Mass Spectrometry. The SAMDI peptide array plates were coated with a 10 mg/mL 2′,4′,6′-Trihydroxyacetophenone (THAP) MALDI matrix in acetonitrile. Each immobilized peptide was analyzed in the reflector positive mode with 900 shots on an AB Sciex TOF/TOF 5800 MALDI mass spectrometer.

Statistical testing to identify amino acids associated with high or low S/N. The S/N of all peptides were calculated by dividing the integrated product (area under the curve) of the expected peptide peaks (the signal) by the integrated product of a region in the spectrum devoid of peaks (the noise). The S/N for each peptide were sorted and ranked from lowest to highest. The S/N increase for consecutive peptides was calculated, and the low region boundary was defined as when a large change in S/N increase occurs. Similarly, a high S/N region was identified with the same process. This method allows different sizes for low and high regions. High levels or a specific amino acid in either region was determined using the Fischer exact test, which calculated the probability to observe at least as many amino acids in the region. Since there were 19 amino acids, the significance threshold was determined by a Bonferroni corrected p-value cutoff of 10^{-4}; all

reported *p*-values define the likelihood that the observed number of amino acids is within the low or high regions by random chance.

Chapter 5. Conclusions and Perspectives

This dissertation introduces new applications for reaction monitoring on peptide arrays using SAMDI-MS. This platform has previously been utilized to study substrate specificity of isolated and purified enzymes, including lysine deacetylases, lysine acetyltransferases, proteases, kinases, and glycosyltransferases.[14, 15, 54, 158, 189] More recently, SAMDI-based peptide arrays were used to measure lysine deacetylase and sirtuin activity in cell lysates.[16] The research described in this dissertation expands upon previous applications such as measuring global enzyme activity in cell lysates, and introduces reaction discovery on peptides as a new application. Further, this work introduces statistical and machine learning analysis on peptide array data for the optimization of peptide array design and synthesis.

Chapter 2 illustrates the use of peptide arrays and SAMDI-MS for reaction discovery on peptides. We show that acetic anhydride selectively acetylates peptides containing histidine and tyrosine and that this reactivity occurs through initial acetylation of the imidazole side chain of histidine, followed by acetyl-transfer to the hydroxyl group of tyrosine. With a second peptide array, we show that amino acids surrounding the HY dyad can affect the reactivity of acetic anhydride. Additionally, we measure inter- and intramolecular distance dependent effects, the stability of acetylation on the HY dyad, and the solution-phase reactivity for comparison to surface-phase. One of the main advantages of SAMDI-based peptide arrays that permitted the discovery of this reaction is the ability to detect unanticipated products, which is a limitation of other forms of peptide arrays that rely on labels for detection. SAMDI-MS offers a convenient method for reaction optimization for current and future protein bioconjugation reactions.

In Chapter 3, we used SAMDI-MS and phosphorylated peptide arrays to profile global phosphatase activity in cell lysates derived from 5 different mammalian cell lines. We utilized peptide arrays consisting of a central phosphorylated serine, threonine, or tyrosine residue surrounded by two variable positions. While these peptides are not biologically relevant, they are unbiased towards any one specific phosphatase and screen a broad range of substrates to allow a full picture of global phosphatase activity. Phosphatases remain an understudied class of enzymes due to a lack of compatible assays and the past belief that phosphatases serve as housekeeping enzymes without regulation or specificity. Our results show that phosphatases are specific and likely play important roles in the regulation of phosphorylation. Further research on phosphatases could greatly impact our understanding of phosphorylation and have important implications for the advancement of treatments for disease.

Chapter 4 addresses the need for data analysis on high-throughput data acquired from peptide arrays. We first evaluate peptide quality using SAMDI-MS signal-to-noise ratios. We then use the corresponding signal-to-noise as well as properties of each peptide sequence to train a machine learning model for optimization of future array design. We show that only 1/3 of an array is needed to predict the quality of the remaining peptides in that array. Using this technique, peptide arrays can be designed to contain only high-quality peptides, which increases the efficiency of peptide library synthesis. While the application of machine learning to biochemical systems offers promise for experimental design, few examples have been reported.[224-226] Machine learning on peptide array data will likely continue to be utilized for enzyme-substrate optimization and the design of bioactive peptides for drug applications.

The use of SAMDI-MS for peptide array experiments offers several advantages over other peptide array technologies, many of which derive from the use of self-assembled monolayers (SAMs) on gold and the compatibility of the technique with MALDI mass spectrometry. The use of SAMs of alkanethiolates on gold allows for a well-defined structure with density control and uniform orientation of peptide substrates. Further, the incorporation of tri(ethylene glycol) groups resist non-specific protein adsorption to the surface, which eliminates the need for blocking steps and significantly reduces background noise. The compatibility of SAMs with MALDI mass spectrometry allows for label-free detection. Peptide arrays that rely on colorimetric or fluorescent labels for detection are limited by label availability and low signal-to-noise ratios and are often prone to perturb enzyme activity. Label-free detection methods can detect unanticipated reaction products, which is an important advantage for studying systems that are not well understood or unpredictable. MALDI-MS also provides quantitative results, allowing for higher resolution data to be obtained from studies comparing reactivity on large numbers of peptides.

SAMDI-MS and peptide arrays are an efficient high-throughput screening technology, with current arrays consisting of 384 and 1,536 features that can be analyzed in 15 and 30 minutes, respectively. However, this technique has two main limitations. First, other forms of peptide arrays, such as SPOT or particle-based arrays, can offer much higher feature density. Second, SAMDI-based peptide arrays rely on immobilization of pre-synthesized peptides. Solid-phase peptide synthesis (SPPS) is slow and tedious for large numbers of peptides. The development of more efficient peptide synthesis methods and higher density SAMDI arrays will continue to

expand the possible applications and allow for higher-throughput screening. In fact, the Mrksich lab is currently working on developing arrays that contain 6,144 and 24,576 features.

Peptide arrays have expanded the capabilities of protein research, allowing for a broad range of applications. They will continue to play a central role in applications such as epitope mapping for antibody binding, identification and characterization of binding interactions between protein and peptide ligands, screening for active substrates of enzymes, and profiling enzyme activity in complex samples. The full potential of peptide arrays is not yet fully appreciated, and they will likely play important roles in applications with new purposes. Some examples may include identifying and optimizing reactions for sequence-specific modification of peptide tags,[227] providing patterns of enzyme activities to understand cell function at a systems level, and identifying peptides with biological activity for drug conjugation. Peptide arrays may also find application in recognizing peptides with nonbiological properties, including affinity for nanoparticle surfaces[228] and development of other peptide-based materials.

References

(1) Whitesides, G. M., Kriebel, J. K., and Love, J. C. (2005) Molecular engineering of surfaces using self-assembled monolayers, *Sci. Prog. 88*, 17-48.

(2) Poirier, G. E., and Pylant, E. D. (1996) The Self-Assembly Mechanism of Alkanethiols on Au(111), *Science 272*, 1145-1148.

(3) Mrksich, M., and Whitesides, G. M., Using Self-Assembled Monolayers That Present Oligo(ethylene glycol) Groups To Control the Interactions of Proteins with Surfaces. In *Poly(ethylene glycol)*, American Chemical Society: 1997; Vol. 680, pp 361-373.

(4) Herrwerth, S., Eck, W., Reinhardt, S., and Grunze, M. (2003) Factors that Determine the Protein Resistance of Oligoether Self-Assembled Monolayers − Internal Hydrophilicity, Terminal Hydrophilicity, and Lateral Packing Density, *J. Am. Chem. Soc. 125*, 9359-9366.

(5) Anderson, L. L., Berns, E. J., Bugga, P., George, A. L., and Mrksich, M. (2016) Measuring Drug Metabolism Kinetics and Drug–Drug Interactions Using Self-Assembled Monolayers for Matrix-Assisted Laser Desorption-Ionization Mass Spectrometry, *Anal. Chem. 88*, 8604-8609.

(6) Diagne, A. B., Li, S., Perkowski, G. A., Mrksich, M., and Thomson, R. J. (2015) SAMDI Mass Spectrometry-Enabled High-Throughput Optimization of a Traceless Petasis Reaction, *ACS Comb. Sci. 17*, 658-662.

(7) Ban, L., Pettit, N., Li, L., Stuparu, A. D., Cai, L., Chen, W., Guan, W., Han, W., Wang, P. G., and Mrksich, M. (2012) Discovery of glycosyltransferases using carbohydrate arrays and mass spectrometry, *Nat Chem Biol 8*, 769-773.

(8) Li, J., Thiara, P. S., and Mrksich, M. (2007) Rapid Evaluation and Screening of Interfacial Reactions on Self-Assembled Monolayers, *Langmuir 23*, 11826-11835.

(9) Kim, J., and Mrksich, M. (2009) Profiling the selectivity of DNA ligases in an array format with mass spectrometry, *Nucleic Acids Res. 38*, e2-e2.

(10) Houseman, B. T., Gawalt, E. S., and Mrksich, M. (2003) Maleimide-Functionalized Self-Assembled Monolayers for the Preparation of Peptide and Carbohydrate Biochips, *Langmuir 19*, 1522-1531.

(11) Grant, J., O'Kane, P. T., Kimmel, B. R., and Mrksich, M. (2019) Using Microfluidics and Imaging SAMDI-MS To Characterize Reaction Kinetics, *ACS Cent. Sci. 5*, 486-493.

(12) Gong, W., Elitzin, V. I., Janardhanam, S., Wilkins, C. L., and Fritsch, I. (2001) Effect of Laser Fluence on Laser Desorption Mass Spectra of Organothiol Self-Assembled Monolayers on Gold, *J. Am. Chem. Soc. 123*, 769-770.

(13) Trevor, J. L., Lykke, K. R., Pellin, M. J., and Hanley, L. (1998) Two-Laser Mass Spectrometry of Thiolate, Disulfide, and Sulfide Self-Assembled Monolayers, *Langmuir 14*, 1664-1673.

(14) Min, D.-H., Su, J., and Mrksich, M. (2004) Profiling Kinase Activities by Using a Peptide Chip and Mass Spectrometry, *Angew. Chem. Int. Ed. 43*, 5973-5977.

(15) Gurard-Levin, Z. A., Kim, J., and Mrksich, M. (2009) Combining Mass Spectrometry and Peptide Arrays to Profile the Specificities of the Histone Deacetylases, *Chembiochem 10*, 2159-2161.

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

(17) Szymczak, L. C., Kuo, H.-Y., and Mrksich, M. (2018) Peptide Arrays: Development and Application, *Anal. Chem. 90*, 266-282.

(18) Katz, C., Levy-Beladev, L., Rotem-Bamberger, S., Rito, T., Rudiger, S. G. D., and Friedler, A. (2011) Studying protein-protein interactions using peptide arrays, *Chem. Soc. Rev. 40*, 2131-2145.

(19) Arsenault, R., Griebel, P., and Napper, S. (2011) Peptide arrays for kinome analysis: New opportunities and remaining challenges, *Proteomics 11*, 4595-4609.

(20) Thiele, A., Stangl, G. I., and Schutkowski, M. (2011) Deciphering Enzyme Function Using Peptide Arrays, *Molecular Biotechnology 49*, 283-305.

(21) Foong, Y. M., Fu, J., Yao, S. Q., and Uttamchandani, M. (2012) Current advances in peptide and small molecule microarray technologies, *Curr. Opin. Chem. Biol. 16*, 234-242.

(22) Chiari, M., Cretich, M., Corti, A., Damin, F., Pirri, G., and Longhi, R. (2005) Peptide microarrays for the characterization of antigenic regions of human chromogranin A, *Proteomics 5*, 3600-3603.

(23) Cretich, M., Damin, F., Pirri, G., and Chiari, M. (2006) Protein and peptide arrays: Recent trends and new directions, *Biomol. Eng 23*, 77-88.

(24) Maskos, U., and Southern, E. M. (1992) Oligonucleotide hybridisations on glass supports: a novel linker for oligonucleotide synthesis and hybridisation properties of oligonucleotides synthesised in situ, *Nucleic Acids Res. 20*, 1679-1684.

(25) Southern, E., Method and apparatus for analysing polynucleotide sequences. Google Patents: 1994.

(26) Fodor, S., Read, J., Pirrung, M., Stryer, L., Lu, A., and Solas, D. (1991) Light-directed, spatially addressable parallel chemical synthesis, *Science 251*, 767-773.

(27) Schena, M., Shalon, D., Davis, R. W., and Brown, P. O. (1995) Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray, *Science 270*, 467-470.

(28) Schena, M., Shalon, D., Heller, R., Chai, A., Brown, P. O., and Davis, R. W. (1996) Parallel human genome analysis: microarray-based expression monitoring of 1000 genes, *Proc. Natl. Acad. Sci. U.S.A. 93*, 10614-10619.

(29) Lashkari, D. A., DeRisi, J. L., McCusker, J. H., Namath, A. F., Gentile, C., Hwang, S. Y., Brown, P. O., and Davis, R. W. (1997) Yeast microarrays for genome wide parallel genetic and gene expression analysis, *Proc. Natl. Acad. Sci. U.S.A. 94*, 13057-13062.

(30) Michael, K. L., Taylor, L. C., Schultz, S. L., and Walt, D. R. (1998) Randomly Ordered Addressable High-Density Optical Sensor Arrays, *Anal. Chem. 70*, 1242-1248.

(31) Davies, R. W., Wells, G. A., Stewart, A. F. R., Erdmann, J., Shah, S. H., Ferguson, J. F., Hall, A. S., Anand, S. S., Burnett, M. S., Epstein, S. E., Dandona, S., Chen, L., Nahrstaedt, J., Loley, C., König, I. R., Kraus, W. E., Granger, C. B., Engert, J. C., Hengstenberg, C., Wichmann, H. E., Schreiber, S., Tang, W. H. W., Ellis, S. G., Rader, D. J., Hazen, S. L., Reilly, M. P., Samani, N. J., Schunkert, H., Roberts, R., and McPherson, R. (2012) A Genome Wide Association Study for Coronary Artery Disease Identifies a Novel Susceptibility Locus in the Major Histocompatibility Complex, *Circ Cardiovasc Genet. 5*, 217-225.

(32) Telenti, A., Pierce, L. C. T., Biggs, W. H., di Iulio, J., Wong, E. H. M., Fabani, M. M., Kirkness, E. F., Moustafa, A., Shah, N., Xie, C., Brewerton, S. C., Bulsara, N., Garner, C., Metzker, G., Sandoval, E., Perkins, B. A., Och, F. J., Turpaz, Y., and Venter, J. C. (2016) Deep sequencing of 10,000 human genomes, *Proc. Natl. Acad. Sci. U.S.A. 113*, 11901-11906.

(33) Abegglen, L. M., Caulin, A. F., Chan, A., Lee, K., Robinson, R., Campbell, M. S., Kiso, W. K., Schmitt, D. L., Waddell, P. J., Bhaskara, S., Jensen, S. T., Maley, C. C., and Schiffman, J. D. (2015) Potential Mechanisms for Cancer Resistance in Elephants and Comparative Cellular Response to DNA Damage in Humans, *JAMA 314*, 1850-1860.

(34) Jabara, H. H., Boyden, S. E., Chou, J., Ramesh, N., Massaad, M. J., Benson, H., Bainter, W., Fraulino, D., Rahimov, F., Sieff, C., Liu, Z.-J., Alshemmari, S. H., Al-Ramadi, B. K., Al-

Dhekri, H., Arnaout, R., Abu-Shukair, M., Vatsayan, A., Silver, E., Ahuja, S., Davies, E. G., Sola-Visner, M., Ohsumi, T. K., Andrews, N. C., Notarangelo, L. D., Fleming, M. D., Al-Herz, W., Kunkel, L. M., and Geha, R. S. (2016) A missense mutation in TFRC, encoding transferrin receptor 1, causes combined immunodeficiency, *Nat. Genet. 48*, 74-78.

(35) Human Genome Sequencing, C. (2004) Finishing the euchromatic sequence of the human genome, *Nature 431*, 931-945.

(36) Pan, Q., Shai, O., Lee, L. J., Frey, B. J., and Blencowe, B. J. (2008) Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing, *Nat. Genet. 40*, 1413-1415.

(37) Harper, J. W., and Bennett, E. J. (2016) Proteome complexity and the forces that drive proteome imbalance, *Nature 537*, 328-338.

(38) Jensen, O. N. (2004) Modification-specific proteomics: characterization of post-translational modifications by mass spectrometry, *Curr. Opin. Chem. Biol. 8*, 33-41.

(39) MacBeath, G., and Schreiber, S. L. (2000) Printing Proteins as Microarrays for High-Throughput Function Determination, *Science 289*, 1760-1763.

(40) Zhu, H., Klemic, J. F., Chang, S., Bertone, P., Casamayor, A., Klemic, K. G., Smith, D., Gerstein, M., Reed, M. A., and Snyder, M. (2000) Analysis of yeast protein kinases using protein chips, *Nat. Genet. 26*, 283-289.

(41) Weber, L. K., Isse, A., Rentschler, S., Kneusel, R. E., Palermo, A., Hubbuch, J., Nesterov-Mueller, A., Breitling, F., and Loeffler, F. F. (2017) Antibody fingerprints in lyme disease deciphered with high density peptide arrays, *Eng. Life Sci. 17*, 1078-1087.

(42) Hansen, L. B., Buus, S., and Schafer-Nielsen, C. (2013) Identification and Mapping of Linear Antibody Epitopes in Human Serum Albumin Using High-Density Peptide Arrays, *PLOS ONE 8*, e68902.

(43) Müller, A.-M., Bockstahler, M., Hristov, G., Weiß, C., Fischer, A., Korkmaz-Icöz, S., Giannitsis, E., Poller, W., Schultheiss, H.-P., Katus, H. A., and Kaya, Z. (2016) Identification of novel antigens contributing to autoimmunity in cardiovascular diseases, *Clinical Immunology 173*, 64-75.

(44) Zandian, A., Forsström, B., Häggmark-Månberg, A., Schwenk, J. M., Uhlén, M., Nilsson, P., and Ayoglu, B. (2017) Whole-Proteome Peptide Microarrays for Profiling Autoantibody Repertoires within Multiple Sclerosis and Narcolepsy, *J. Proteome Res. 16*, 1300-1314.

(45) Navalkar, K. A., Johnston, S. A., Woodbury, N., Galgiani, J. N., Magee, D. M., Chicacz, Z., and Stafford, P. (2014) Application of Immunosignatures for Diagnosis of Valley Fever, *Clin. Vaccine Immunol. 21*, 1169-1177.

(46) Kume, A., Kawai, S., Kato, R., Iwata, S., Shimizu, K., and Honda, H. (2017) Exploring high-affinity binding properties of octamer peptides by principal component analysis of tetramer peptides, *J. Biosci. Bioeng. 123*, 230-238.

(47) Rouka, E., Simister, P. C., Janning, M., Kumbrink, J., Konstantinou, T., Muniz, J. R. C., Joshi, D., O'Reilly, N., Volkmer, R., Ritter, B., Knapp, S., von Delft, F., Kirsch, K. H., and Feller, S. M. (2015) Differential Recognition Preferences of the Three Src Homology 3 (SH3) Domains from the Adaptor CD2-associated Protein (CD2AP) and Direct Association with Ras and Rab Interactor 3 (RIN3), *J. Biol. Chem. 290*, 25275-25292.

(48) O'Kane, P. T., and Mrksich, M. (2017) An Assay Based on SAMDI Mass Spectrometry for Profiling Protein Interaction Domains, *J. Am. Chem. Soc. 139*, 10320-10327.

(49) Zhao, S., Yang, M., Zhou, W., Zhang, B., Cheng, Z., Huang, J., Zhang, M., Wang, Z., Wang, R., Chen, Z., Zhu, J., and Li, H. (2017) Kinetic and high-throughput profiling of epigenetic interactions by 3D-carbene chip-based surface plasmon resonance imaging technology, *Proc. Natl. Acad. Sci. U.S.A. 114*, E7245-E7254.

(50) Kusevic, D., Kudithipudi, S., and Jeltsch, A. (2016) Substrate Specificity of the HEMK2 Protein Glutamine Methyltransferase and Identification of Novel Substrates, *J. Biol. Chem. 291*, 6124-6133.

(51) Shishkova, E., Zeng, H., Liu, F., Kwiecien, N. W., Hebert, A. S., Coon, J. J., and Xu, W. (2017) Global mapping of CARM1 substrates defines enzyme specificity and substrate recognition, *Nat. Commun. 8*, 15571.

(52) Palma, A., Tinti, M., Paoluzi, S., Santonico, E., Brandt, B. W., Hooft van Huijsduijnen, R., Masch, A., Heringa, J., Schutkowski, M., Castagnoli, L., and Cesareni, G. (2017) Both Intrinsic Substrate Preference and Network Context Contribute to Substrate Selection of Classical Tyrosine Phosphatases, *J. Biol. Chem. 292*, 4942-4952.

(53) Patel, K., Sherrill, J., Mrksich, M., and Scholle, M. D. (2015) Discovery of SIRT3 Inhibitors Using SAMDI Mass Spectrometry, *J. Biomol. Screening 20*, 842-848.

(54) Kornacki, J. R., Stuparu, A. D., and Mrksich, M. (2015) Acetyltransferase p300/CBP Associated Factor (PCAF) Regulates Crosstalk-Dependent Acetylation of Histone H3 by Distal Site Recognition, *ACS Chem. Biol. 10*, 157-164.

(55) Castaneda, C. A., Lopez, J. E., Joseph, C. G., Scholle, M. D., Mrksich, M., and Fierke, C. A. (2017) Active Site Metal Identity Alters Histone Deacetylase 8 Substrate Selectivity: A Potential Novel Regulatory Mechanism, *Biochemistry 56*, 5663-5670.

(56) Sharif, S., Shi, J., Bourakba, M., Ruijtenbeek, R., and Pieters, R. J. (2017) Measuring O-GlcNAc cleavage by OGA and cell lysates on a peptide microarray, *Anal. Biochem. 532*, 12-18.

(57) Berns, E. J., Cabezas, M. D., and Mrksich, M. (2016) Cellular Assays with a Molecular Endpoint Measured by SAMDI Mass Spectrometry, *Small 12*, 3811-3818.

(58) Kato, R., Kaga, C., Kunimatsu, M., Kobayashi, T., and Honda, H. (2006) Peptide array-based interaction assay of solid-bound peptides and anchorage-dependant cells and its effectiveness in cell-adhesive peptide design, *J. Biosci. Bioeng. 101*, 485-495.

(59) Ahmed, S., Mathews, A. S., Byeon, N., Lavasanifar, A., and Kaur, K. (2010) Peptide Arrays for Screening Cancer Specific Peptides, *Anal. Chem. 82*, 7533-7541.

(60) Ruoslahti, E., and Pierschbacher, M. (1987) New perspectives in cell adhesion: RGD and integrins, *Science 238*, 491-497.

(61) Sánchez-Cortés, J., and Mrksich, M. (2009) The Platelet Integrin αIIbβ3 Binds to the RGD and AGD Motifs in Fibrinogen, *Chem. Biol. 16*, 990-1000.

(62) Sánchez-Cortés, J., and Mrksich, M. (2011) Using Self-Assembled Monolayers To Understand α8β1-Mediated Cell Adhesion to RGD and FEI Motifs in Nephronectin, *ACS Chem. Biol. 6*, 1078-1086.

(63) Hilpert, K., Winkler, D. F., and Hancock, R. E. (2007) Cellulose-bound peptide arrays: preparation and applications., *Biotechnol. Genet. Eng. Rev. 24*, 31-106.

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

(65) Frank, R., and Dübel, S. (2006) Analysis of Protein Interactions with Immobilized Peptide Arrays Synthesized on Membrane Supports, *Cold Spring Harb. Protoc. 2006*, pdb.prot4566.

(66) Frank, R. (2002) The SPOT-synthesis technique: Synthetic peptide arrays on membrane supports—principles and applications, *J. Immunol. Methods 267*, 13-26.

(67) Reineke, U., Volkmer-Engert, R., and Schneider-Mergener, J. (2001) Applications of peptide arrays prepared by the SPOT-technology, *Curr. Opin. Biotechnol. 12*, 59-64.

(68) Volkmer, R., Tapia, V., and Landgraf, C. (2012) Synthetic peptide arrays for investigating protein interaction domains, *FEBS Lett. 586*, 2780-2786.

(69) Merrifield, R. B. (1963) Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide, *J. Am. Chem. Soc. 85*, 2149-2154.

(70) Geysen, H. M., Meloen, R. H., and Barteling, S. J. (1984) Use of peptide synthesis to probe viral antigens for epitopes to a resolution of a single amino acid, *Proc Natl Acad Sci U S A 81*.

(71) Houghten, R. A. (1985) General method for the rapid solid-phase synthesis of large numbers of peptides: specificity of antigen-antibody interaction at the level of individual amino acids, *Proc. Natl. Acad. Sci. U.S.A. 82*, 5131-5135.

(72) Love, J. C., Estroff, L. A., Kriebel, J. K., Nuzzo, R. G., and Whitesides, G. M. (2005) Self-Assembled Monolayers of Thiolates on Metals as a Form of Nanotechnology, *Chem. Rev. 105*, 1103-1170.

(73) Prime, K. L., and Whitesides, G. M. (1993) Adsorption of proteins onto surfaces containing end-attached oligo(ethylene oxide): a model system using self-assembled monolayers, *J. Am. Chem. Soc. 115*, 10714-10721.

(74) Beyer, M., Felgenhauer, T., Ralf Bischoff, F., Breitling, F., and Stadler, V. (2006) A novel glass slide-based peptide array support with high functionality resisting non-specific protein adsorption, *Biomaterials 27*, 3505-3514.

(75) Saxinger, C., Conrads, T. P., Goldstein, D. J., and Veenstra, T. D. (2005) Fully automated synthesis of (phospho)peptide arrays in microtiter plate wells provides efficient access to protein tyrosine kinase characterization, *BMC Immunology 6*, 1.

(76) Schutkowski, M., Reimer, U., Panse, S., Dong, L., Lizcano, J. M., Alessi, D. R., and Schneider-Mergener, J. (2004) High-Content Peptide Microarrays for Deciphering Kinase Specificity and Biology, *Angew. Chem. Int. Ed. 43*, 2671-2674.

(77) Falsey, J. R., Renil, M., Park, S., Li, S., and Lam, K. S. (2001) Peptide and Small Molecule Microarray for High Throughput Cell Adhesion and Functional Assays, *Bioconjugate Chem. 12*, 346-353.

(78) Han, X., Shigaki, S., Yamaji, T., Yamanouchi, G., Mori, T., Niidome, T., and Katayama, Y. (2008) A quantitative peptide array for evaluation of protein kinase activity, *Anal. Biochem. 372*, 106-115.

(79) Xu, Q., Miyamoto, S., and Lam, K. S. (2004) A novel approach to chemical microarray using ketone-modified macromolecular scaffolds: Application in micro cell-adhesion assay, *Mol. Diversity 8*, 301-310.

(80) Chan, E. W. L., and Yousaf, M. N. (2006) Immobilization of Ligands with Precise Control of Density to Electroactive Surfaces, *J. Am. Chem. Soc. 128*, 15542-15546.

(81) Salisbury, C. M., Maly, D. J., and Ellman, J. A. (2002) Peptide Microarrays for the Determination of Protease Substrate Specificity, *J. Am. Chem. Soc. 124*, 14868-14870.

(82) Pokholok, D. K., Harbison, C. T., Levine, S., Cole, M., Hannett, N. M., Lee, T. I., Bell, G. W., Walker, K., Rolfe, P. A., Herbolsheimer, E., Zeitlinger, J., Lewitter, F., Gifford, D. K., and Young, R. A. (2005) Genome-wide Map of Nucleosome Acetylation and Methylation in Yeast, *Cell 122*, 517-527.

(83) Wammes, A. E. M., Fischer, M. J. E., de Mol, N. J., van Eldijk, M. B., Rutjes, F. P. J. T., van Hest, J. C. M., and van Delft, F. L. (2013) Site-specific peptide and protein immobilization on surface plasmon resonance chips via strain-promoted cycloaddition, *Lab Chip 13*, 1863-1867.

(84) Shamsi, F., Coster, H., and Jolliffe, K. A. (2011) Characterization of peptide immobilization on an acetylene terminated surface via click chemistry, *Surf. Sci. 605*, 1763-1770.

(85) Soellner, M. B., Dickson, K. A., Nilsson, B. L., and Raines, R. T. (2003) Site-Specific Protein Immobilization by Staudinger Ligation, *J. Am. Chem. Soc. 125*, 11790-11791.

(86) Lesaicherre, M.-L., Uttamchandani, M., Chen, G. Y. J., and Yao, S. Q. (2002) Developing site-Specific immobilization strategies of peptides in a microarray, *Bioorg. Med. Chem. Lett. 12*, 2079-2083.

(87) Aina, O. H., Sroka, T. C., Chen, M.-L., and Lam, K. S. (2002) Therapeutic cancer targeting peptides, *Pept. Sci. 66*, 184-199.

(88) De Keersmaecker, K., Versele, M., Cools, J., Superti-Furga, G., and Hantschel, O. (2008) Intrinsic differences between the catalytic properties of the oncogenic NUP214-ABL1 and BCR-ABL1 fusion protein kinases, *Leukemia 22*, 2208-2216.

(89) Rothbart, S. B., Krajewski, K., Strahl, B. D., and Fuchs, S. M., Chapter Six - Peptide Microarrays to Interrogate the "Histone Code". In *Methods Enzymol.*, Carl, W.; Allis, C. D., Eds. Academic Press: 2012; Vol. Volume 512, pp 107-135.

(90) Winssinger, N., Damoiseaux, R., Tully, D. C., Geierstanger, B. H., Burdick, K., and Harris, J. L. (2004) PNA-Encoded Protease Substrate Microarrays, *Chem. Biol. 11*, 1351-1360.

(91) Winssinger, N., Ficarro, S., Schultz, P. G., and Harris, J. L. (2002) Profiling protein function with small molecule microarrays, *Proc. Natl. Acad. Sci. U.S.A. 99*, 11139.

(92) Scott, J., and Smith, G. (1990) Searching for peptide ligands with an epitope library, *Science 249*, 386-390.

(93) Smith, G. P., and Petrenko, V. A. (1997) Phage Display, *Chem. Rev. 97*, 391-410.

(94) Gai, S. A., and Wittrup, K. D. (2007) Yeast surface display for protein engineering and characterization, *Curr. Opin. Struct. Biol. 17*, 467-473.

(95) Zahnd, C., Amstutz, P., and Plückthun, A. (2007) Ribosome display: selecting and evolving proteins in vitro that specifically bind to a target, *Nat. Methods 4*, 269.

(96) Mattheakis, L. C., Bhatt, R. R., and Dower, W. J. (1994) An in vitro polysome display system for identifying ligands from very large peptide libraries, *Proc. Natl. Acad. Sci. U.S.A. 91*, 9022.

(97) Kaeberlein, M., McDonagh, T., Heltweg, B., Hixon, J., Westman, E. A., Caldwell, S. D., Napper, A., Curtis, R., DiStefano, P. S., Fields, S., Bedalov, A., and Kennedy, B. K. (2005) Substrate-specific Activation of Sirtuins by Resveratrol, *J. Biol. Chem. 280*, 17038-17045.

(98) Dostmann, W. R. G., Taylor, M. S., Nickl, C. K., Brayden, J. E., Frank, R., and Tegge, W. J. (2000) Highly specific, membrane-permeant peptide blockers of cGMP-dependent protein kinase Iα inhibit NO-induced cerebral dilation, *Proc. Natl. Acad. Sci. U.S.A. 97*, 14772-14777.

(99) Smith, F. D., Samelson, B. K., and Scott, J. D. (2011) Discovery of cellular substrates for protein kinase A using a peptide array screening protocol, *Biochem. J 438*, 103-110.

(100) Rathert, P., Dhayalan, A., Murakami, M., Zhang, X., Tamas, R., Jurkowska, R., Komatsu, Y., Shinkai, Y., Xiaodong, C., and Jeltsch, A. (2008) Protein lysine methyltransferase G9a acts on non-histone targets, *Nat. Chem. Biol. 4*, 344-346.

(101) Dürauer, A., Kopecky, E., Berger, E., Seifert, M., Hahn, R., and Jungbauer, A. (2006) Evaluation of a sensitive detection method for peptide arrays prepared by SPOT synthesis, *J. Biochem. Bioph. Methods 66*, 45-57.

(102) Landgraf, C., Panni, S., Montecchi-Palazzi, L., Castagnoli, L., Schneider-Mergener, J., Volkmer-Engert, R., and Cesareni, G. (2004) Protein Interaction Networks by Proteome Peptide Scanning, *PLOS Biology 2*, e14.

(103) Blake, M. S., Johnston, K. H., Russell-Jones, G. J., and Gotschlich, E. C. (1984) A rapid, sensitive method for detection of alkaline phosphatase-conjugated anti-antibody on Western blots, *Anal. Biochem. 136*, 175-179.

(104) Hawkes, R., Niday, E., and Gordon, J. (1982) A dot-immunobinding assay for monoclonal and other antibodies, *Anal. Biochem. 119*, 142-147.

(105) Verastegui, M., Moro, P., Guevara, A., Rodriguez, T., Miranda, E., and Gilman, R. H. (1992) Enzyme-linked immunoelectrotransfer blot test for diagnosis of human hydatid disease, *J. Clin. Microbiol. 30*, 1557-1561.

(106) Schuck, P. (1997) Use of Surface Plasmon Resonance to Probe the Equilibrium and Dynamic aspects of Interactions between Biological Macromolecules, *Annu. Rev. Biophys. Biomol. Struct. 26*, 541-566.

(107) Inoue, Y., Mori, T., Yamanouchi, G., Han, X., Sonoda, T., Niidome, T., and Katayama, Y. (2008) Surface plasmon resonance imaging measurements of caspase reactions on peptide microarrays, *Anal. Biochem. 375*, 147-149.

(108) Inamori, K., Kyo, M., Nishiya, Y., Inoue, Y., Sonoda, T., Kinoshita, E., Koike, T., and Katayama, Y. (2005) Detection and Quantification of On-Chip Phosphorylated Peptides by Surface Plasmon Resonance Imaging Techniques Using a Phosphate Capture Molecule, *Anal. Chem. 77*, 3979-3985.

(109) de Rond, T., Danielewicz, M., and Northen, T. (2015) High throughput screening of enzyme activity with mass spectrometry imaging, *Curr. Opin. Biotechnol. 31*, 1-9.

(110) Su, J., and Mrksich, M. (2002) Using Mass Spectrometry to Characterize Self-Assembled Monolayers Presenting Peptides, Proteins, and Carbohydrates, *Angew. Chem. Int. Ed. 41*, 4715-4718.

(111) Min, D.-H., Yeo, W.-S., and Mrksich, M. (2004) A Method for Connecting Solution-Phase Enzyme Activity Assays with Immobilized Format Analysis by Mass Spectrometry, *Anal. Chem. 76*, 3923-3929.

(112) Mrksich, M. (2008) Mass Spectrometry of Self-Assembled Monolayers: A New Tool for Molecular Surface Science, *ACS Nano 2*, 7-18.

(113) Gemini-Piperni, S., Milani, R., Bertazzo, S., Peppelenbosch, M., Takamori, E. R., Granjeiro, J. M., Ferreira, C. V., Teti, A., and Zambuzzi, W. (2014) Kinome profiling of osteoblasts on hydroxyapatite opens new avenues on biomaterial cell signaling, *Biotechnol. Bioeng. 111*, 1900-1905.

(114) de Bourayne, M., Gallais, Y., El Ali, Z., Rousseau, P., Damiens, M.-H., Cochet, C., Filhol, O., Chollet-Martin, S., Pallardy, M., and Kerdine-Römer, S. (2017) Protein kinase CK2 controls T-cell polarization through dendritic cell activation in response to contact sensitizers, *Journal of Leukocyte Biology 101*, 703-715.

(115) Tsang, S. H., Chan, L., Tsai, Y.-T., Wu, W.-H., Hsu, C.-W., Yang, J., Tosi, J., Wert, K. J., Davis, R. J., and Mahajan, V. B. (2014) Silencing of Tuberin Enhances Photoreceptor Survival and Function in a Preclinical Model of Retinitis Pigmentosa (An American Ophthalmological Society Thesis), *Transactions of the American Ophthalmological Society 112*, 103-115.

(116) Kramer, A., and Schneider-Mergener, J., Synthesis and Screening of Peptide Libraries on Continuous Cellulose Membrane Supports. In *Combinatorial Peptide Library Protocols*, Cabilly, S., Ed. Humana Press: Totowa, NJ, 1998; pp 25-39.

(117) Deschamps, T., Bazot, Q., Leske, D. M., MacLeod, R., Mompelat, D., Tafforeau, L., Lotteau, V., Maréchal, V., Baillie, G. S., Gruffat, H., Wilson, J. B., and Manet, E. (2017) Epstein–Barr virus nuclear antigen 1 interacts with regulator of chromosome condensation 1 dynamically throughout the cell cycle, *J. Gen. Virol. 98*, 251-265.

(118) López-Pérez, P. M., Grimsey, E., Bourne, L., Mikut, R., and Hilpert, K. (2017) Screening and Optimizing Antimicrobial Peptides by Using SPOT-Synthesis, *Front. Chem. 5*, 25.

(119) Tanaka, M., Hikiba, S., Yamashita, K., Muto, M., and Okochi, M. (2017) Array-based functional peptide screening and characterization of gold nanoparticle synthesis, *Acta Biomater. 49*, 495-506.

(120) Webster, C. G., Thillier, M., Pirolles, E., Cayrol, B., Blanc, S., and Uzest, M. (2017) Proteomic composition of the acrostyle: novel approaches to identify cuticular proteins involved in virus–insect interactions, *Insect Science*, n/a-n/a.

(121) Vernet, T., Choulier, L., Nominé, Y., Bellard, L., Baltzinger, M., Travé, G., and Altschuh, D. (2015) Spot peptide arrays and SPR measurements: throughput and quantification in antibody selectivity studies, *J. Mol. Recognit. 28*, 635-644.

(122) Loeffler, F. F., Foertsch, T. C., Popov, R., Mattes, D. S., Schlageter, M., Sedlmayr, M., Ridder, B., Dang, F.-X., von Bojničić-Kninski, C., Weber, L. K., Fischer, A., Greifenstein, J., Bykovskaya, V., Buliev, I., Bischoff, F. R., Hahn, L., Meier, M. A. R., Bräse, S., Powell, A. K., Balaban, T. S., Breitling, F., and Nesterov-Mueller, A. (2016) High-flexibility combinatorial peptide synthesis with laser-based transfer of monomers in solid matrix material, *Nat. Commun. 7*, 11844.

(123) Beyer, M., Nesterov, A., Block, I., König, K., Felgenhauer, T., Fernandez, S., Leibe, K., Torralba, G., Hausmann, M., Trunk, U., Lindenstruth, V., Bischoff, F. R., Stadler, V., and Breitling, F. (2007) Combinatorial Synthesis of Peptide Arrays onto a Microchip, *Science 318*, 1888-1888.

(124) Breitling, F., Felgenhauer, T., Nesterov, A., Lindenstruth, V., Stadler, V., and Bischoff, F. R. (2009) Particle-Based Synthesis of Peptide Arrays, *ChemBioChem 10*, 803-808.

(125) Stadler, V., Felgenhauer, T., Beyer, M., Fernandez, S., Leibe, K., Güttler, S., Gröning, M., König, K., Torralba, G., Hausmann, M., Lindenstruth, V., Nesterov, A., Block, I., Pipkorn, R., Poustka, A., Bischoff, F. R., and Breitling, F. (2008) Combinatorial Synthesis of Peptide Arrays with a Laser Printer, *Angew. Chem. Int. Ed. 47*, 7132-7135.

(126) Nesterov-Mueller, A., Maerkle, F., Hahn, L., Foertsch, T., Schillo, S., Bykovskaya, V., Sedlmayr, M., Weber, L. K., Ridder, B., Soehindrijo, M., Muenster, B., Striffler, J., Bischoff, F. R., Breitling, F., and Loeffler, F. F. (2014) Particle-Based Microarrays of Oligonucleotides and Oligopeptides, *Microarrays 3*, 245-262.

(127) Legutki, J. B., Zhao, Z.-G., Greving, M., Woodbury, N., Johnston, S. A., and Stafford, P. (2014) Scalable high-density peptide arrays for comprehensive health monitoring, *Nat. Commun. 5*, 4785.

(128) Price, J. V., Tangsombatvisit, S., Xu, G., Levy, D., Baechler, E. C., Gozani, O., Varma, M., Utz, P. J., and Liu, C. L. (2012) "On silico" peptide microarrays for high-resolution mapping of antibody epitopes and diverse protein-protein interactions, *Nat. Med. 18*, 1434-1440.

(129) Mortensen, R., Nissen, T. N., Fredslund, S., Rosenkrands, I., Christensen, J. P., Andersen, P., and Dietrich, J. (2016) Identifying protective Streptococcus pyogenes vaccine antigens recognized by both B and T cells in human adults and children, *Sci. Rep. 6*, 22030.

(130) Forsström, B., Axnäs, B. B., Stengele, K.-P., Bühler, J., Albert, T. J., Richmond, T. A., Hu, F. J., Nilsson, P., Hudson, E. P., Rockberg, J., and Uhlen, M. (2014) Proteome-wide Epitope Mapping of Antibodies Using Ultra-dense Peptide Arrays, *Mol. Cell. Proteomics 13*, 1585-1597.

(131) Buus, S., Rockberg, J., Forsström, B., Nilsson, P., Uhlen, M., and Schafer-Nielsen, C. (2012) High-resolution Mapping of Linear Antibody Epitopes Using Ultrahigh-density Peptide Microarrays, *Mol. Cell. Proteomics 11*, 1790-1800.

(132) Gurard-Levin, Z. A., and Mrksich, M. (2008) Combining Self-Assembled Monolayers and Mass Spectrometry for Applications in Biochips, *Annu. Rev. Anal. Chem. 1*, 767-800.

(133) Szymczak, L. C., and Mrksich, M. (2019) Using Peptide Arrays To Discover the Sequence-Specific Acetylation of the Histidine-Tyrosine Dyad, *Biochemistry 58*, 1810-1817.

(134) Speers, A. E., and Cravatt, B. F. (2004) Profiling Enzyme Activities In Vivo Using Click Chemistry Methods, *Chem. Biol. 11*, 535-546.

(135) Xue, L., Karpenko, I. A., Hiblot, J., and Johnsson, K. (2015) Imaging and manipulating proteins in live cells through covalent labeling, *Nat. Chem. Biol. 11*, 917.

(136) Griffin, B. A., Adams, S. R., and Tsien, R. Y. (1998) Specific Covalent Labeling of Recombinant Protein Molecules Inside Live Cells, *Science 281*, 269.

(137) Hochuli, E., Bannwarth, W., Döbeli, H., Gentz, R., and Stüber, D. (1988) Genetic Approach to Facilitate Purification of Recombinant Proteins with a Novel Metal Chelate Adsorbent, *Nat. Biotechnol. 6*, 1321.

(138) de Goeij, B. E. C. G., and Lambert, J. M. (2016) New developments for antibody-drug conjugate-based therapeutic approaches, *Curr. Opin. Immunol. 40*, 14-23.

(139) Dozier, J. K., and Distefano, M. D. (2015) Site-Specific PEGylation of Therapeutic Proteins, *Int. J. Mol. Sci. 16*, 25831-25864.

(140) Mackenzie, K. J., and Francis, M. B. (2013) Recyclable Thermoresponsive Polymer–Cellulase Bioconjugates for Biomass Depolymerization, *J. Am. Chem. Soc. 135*, 293-300.

(141) Witus, L. S., and Francis, M. B. (2011) Using Synthetically Modified Proteins to Make New Materials, *Acc. Chem. Res. 44*, 774-783.

(142) Joshi, N. S., Whitaker, L. R., and Francis, M. B. (2004) A Three-Component Mannich-Type Reaction for Selective Tyrosine Bioconjugation, *J. Am. Chem. Soc. 126*, 15942-15943.

(143) Shannon, D. A., Banerjee, R., Webster, E. R., Bak, D. W., Wang, C., and Weerapana, E. (2014) Investigating the Proteome Reactivity and Selectivity of Aryl Halides, *J. Am. Chem. Soc. 136*, 3330-3333.

(144) Zhang, C., Welborn, M., Zhu, T., Yang, N. J., Santos, M. S., Van Voorhis, T., and Pentelute, B. L. (2015) π-Clamp-mediated cysteine conjugation, *Nat. Chem. 8*, 120.

(145) Lee Hong, G., Lautrette, G., Pentelute, B. L., and Buchwald, S. L. (2017) Palladium-Mediated Arylation of Lysine in Unprotected Peptides, *Angew. Chem. Int. Ed. 56*, 3177-3181.

(146) MacDonald, J. I., Munch, H. K., Moore, T., and Francis, M. B. (2015) One-step site-specific modification of native proteins with 2-pyridinecarboxyaldehydes, *Nat. Chem. Biol. 11*, 326.

(147) Dawson, P. E., Muir, T. W., Clark-Lewis, I., and Kent, S. B. (1994) Synthesis of proteins by native chemical ligation, *Science 266*, 776.

(148) Los, G. V., Encell, L. P., McDougall, M. G., Hartzell, D. D., Karassina, N., Zimprich, C., Wood, M. G., Learish, R., Ohana, R. F., Urh, M., Simpson, D., Mendez, J., Zimmerman, K., Otto, P., Vidugiris, G., Zhu, J., Darzins, A., Klaubert, D. H., Bulleit, R. F., and Wood, K. V. (2008) HaloTag: A Novel Protein Labeling Technology for Cell Imaging and Protein Analysis, *ACS Chem. Biol. 3*, 373-382.

(149) Cravatt, B. F., Wright, A. T., and Kozarich, J. W. (2008) Activity-Based Protein Profiling: From Enzyme Chemistry to Proteomic Chemistry, *Annu. Rev. Biochem 77*, 383-414.

(150) Lang, K., and Chin, J. W. (2014) Cellular Incorporation of Unnatural Amino Acids and Bioorthogonal Labeling of Proteins, *Chem. Rev. 114*, 4764-4806.

(151) Noren, C. J., Anthony-Cahill, S. J., Griffith, M. C., and Schultz, P. G. (1989) A general method for site-specific incorporation of unnatural amino acids into proteins, *Science 244*, 182.

(152) Wang, L., Zhang, Z., Brock, A., and Schultz, P. G. (2003) Addition of the keto functional group to the genetic code of Escherichia coli, *Proc. Natl. Acad. Sci. U.S.A. 100*, 56.

(153) Beatty, K. E., Fisk, J. D., Smart, B. P., Lu, Y. Y., Szychowski, J., Hangauer, M. J., Baskin, J. M., Bertozzi, C. R., and Tirrell, D. A. (2010) Live-Cell Imaging of Cellular Proteins by a Strain-Promoted Azide–Alkyne Cycloaddition, *ChemBioChem 11*, 2092-2095.

(154) Saxon, E., and Bertozzi, C. R. (2000) Cell Surface Engineering by a Modified Staudinger Reaction, *Science 287*, 2007.

(155) Nilsson, B. L., Kiessling, L. L., and Raines, R. T. (2000) Staudinger Ligation: A Peptide from a Thioester and Azide, *Org. Lett. 2*, 1939-1941.

(156) Lang, K., Davis, L., Torres-Kolbus, J., Chou, C., Deiters, A., and Chin, J. W. (2012) Genetically encoded norbornene directs site-specific cellular protein labelling via a rapid bioorthogonal reaction, *Nat. Chem. 4*, 298.

(157) Wang, J., Zhang, W., Song, W., Wang, Y., Yu, Z., Li, J., Wu, M., Wang, L., Zang, J., and Lin, Q. (2010) A Biosynthetic Route to Photoclick Chemistry on Proteins, *J. Am. Chem. Soc. 132*, 14812-14818.

(158) Kightlinger, W., Lin, L., Rosztoczy, M., Li, W., DeLisa, M. P., Mrksich, M., and Jewett, M. C. (2018) Design of glycosylation sites by rapid synthesis and analysis of glycosyltransferases, *Nat. Chem. Biol. 14*, 627-635.

(159) Garcia, B. A., Mollah, S., Ueberheide, B. M., Busby, S. A., Muratore, T. L., Shabanowitz, J., and Hunt, D. F. (2007) Chemical derivatization of histones for facilitated analysis by mass spectrometry, *Nat. Protoc. 2*, 933.

(160) Saravanan, P., and Singh, V. K. (1999) An efficient method for acylation reactions, *Tetrahedron Lett. 40*, 2611-2614.

(161) Dmuchovsky, B., Vineyard, B. D., and Zienty, F. B. (1964) The Mechanism of the Base-Catalyzed Addition of Thiols to Maleic Anhydride, *J. Am. Chem. Soc. 86*, 2874-2877.

(162) Moore, G. J. (1985) Kinetics of acetylation-deacetylation of angiotensin II, *Int. J. Pept. Protein Res. 26*, 469-481.

(163) Nawaz, H., Pires, P. A. R., and El Seoud, O. A. (2013) Kinetics and mechanism of imidazole-catalyzed acylation of cellulose in LiCl/N,N-dimethylacetamide, *Carbohydr. Polym. 92*, 997-1005.

(164) Jencks, W. P., and Carriuolo, J. (1959) Imidazole Catalysis: III. General base catalysis and the reactions of acetyl imidazole with thiols and amines, *J. Biol. Chem. 234*, 1280-1285.

(165) Jencks, W. P., and Carriuolo, J. (1960) Reactivity of nucleophilic reagents toward esters, *J. Am. Chem. Soc. 82*, 1778-1786.

(166) Bender, M. L., and Turnquest, B. W. (1957) General Basic Catalysis of Ester Hydrolysis and Its Relationship to Enzymatic Hydrolysis1, *J. Am. Chem. Soc. 79*, 1656-1662.

(167) Kirsch, J. F., and Jencks, W. P. (1964) Nonlinear Structure-Reactivity Correlations. The Imidazole-Catalyzed Hydrolysis of Esters, *J. Am. Chem. Soc. 86*, 837-846.

(168) Baeza, J., Smallegan, M. J., and Denu, J. M. (2015) Site-Specific Reactivity of Nonenzymatic Lysine Acetylation, *ACS Chem. Biol. 10*, 122-128.

(169) Dubois, L. H., and Nuzzo, R. G. (1992) Synthesis, Structure, and Properties of Model Organic Surfaces, *Annu. Rev. Phys. Chem. 43*, 437-463.

(170) Boutureira, O., and Bernardes, G. J. L. (2015) Advances in Chemical Protein Modification, *Chem. Rev. 115*, 2174-2195.

(171) deGruyter, J. N., Malins, L. R., and Baran, P. S. (2017) Residue-Specific Peptide Modification: A Chemist's Guide, *Biochemistry 56*, 3863-3873.

(172) Høeg-Jensen, T., Havsteen Jakobsen, M., Olsen, C. E., and Holm, A. (1991) Formation of peptide thioamides by use of Fmoc amino monothioacids and PyBOP, *Tetrahedron Lett. 32*, 7617-7620.

(173) Gurard-Levin, Z. A., Scholle, M. D., Eisenberg, A. H., and Mrksich, M. (2011) High-Throughput Screening of Small Molecule Libraries using SAMDI Mass Spectrometry, *ACS Comb. Sci. 13*, 347-350.

(174) Gurard-Levin, Z. A., Kilian, K. A., Kim, J., Bähr, K., and Mrksich, M. (2010) Peptide Arrays Identify Isoform-Selective Substrates for Profiling Endogenous Lysine Deacetylase Activity, *ACS Chem. Biol. 5*, 863-873.

(175) Szymczak, L. C., Huang, C.-F., Berns, E. J., and Mrksich, M., Chapter Fourteen - Combining SAMDI Mass Spectrometry and Peptide Arrays to Profile Phosphatase Activities. In *Methods Enzymol.*, Allen, K. N., Ed. Academic Press: 2018; Vol. 607, pp 389-403.

(176) Manning, G., Whyte, D. B., Martinez, R., Hunter, T., and Sudarsanam, S. (2002) The Protein Kinase Complement of the Human Genome, *Science 298*, 1912.

(177) Manning, G., Plowman, G. D., Hunter, T., and Sudarsanam, S. (2002) Evolution of protein kinase signaling from yeast to man, *Trends Biochem. Sci 27*, 514-520.

(178) Ubersax, J. A., and Ferrell Jr, J. E. (2007) Mechanisms of specificity in protein phosphorylation, *Nat. Rev. Mol. Cell Biol. 8*, 530.

(179) Tan, C. S. H., Bodenmiller, B., Pasculescu, A., Jovanovic, M., Hengartner, M. O., Jørgensen, C., Bader, G. D., Aebersold, R., Pawson, T., and Linding, R. (2009) Comparative Analysis Reveals Conserved Protein Phosphorylation Networks Implicated in Multiple Diseases, *Sci. Signaling 2*, ra39-ra39.

(180) Olsen, J. V., Blagoev, B., Gnad, F., Macek, B., Kumar, C., Mortensen, P., and Mann, M. (2006) Global, In Vivo, and Site-Specific Phosphorylation Dynamics in Signaling Networks, *Cell 127*, 635-648.

(181) Ficarro, S. B., McCleland, M. L., Stukenberg, P. T., Burke, D. J., Ross, M. M., Shabanowitz, J., Hunt, D. F., and White, F. M. (2002) Phosphoproteome analysis by mass spectrometry and its application to Saccharomyces cerevisiae, *Nat. Biotechnol. 20*, 301.

(182) Shi, Y. (2009) Serine/Threonine Phosphatases: Mechanism through Structure, *Cell 139*, 468-484.

(183) Virshup, D. M., and Shenolikar, S. (2009) From Promiscuity to Precision: Protein Phosphatases Get a Makeover, *Molecular Cell 33*, 537-545.

(184) Moorhead, Greg B. G., De Wever, V., Templeton, G., and Kerk, D. (2009) Evolution of protein phosphatases in plants and animals, *Biochem. J 417*, 401.

(185) Bose, A. K., and Janes, K. A. (2013) A High-throughput Assay for Phosphoprotein-specific Phosphatase Activity in Cellular Extracts, *Mol. Cell. Proteomics 12*, 797-806.

(186) Killilea, S. D., Cheng, Q., and Wang, Z.-X., Protein Phosphatase Type 1 and Type 2A Assays. In *Protein Phosphatase Protocols*, Ludlow, J., Ed. Humana Press: 1998; Vol. 93, pp 23-33.

(187) McAvoy, T., and Nairn, A. C. (2010) Serine/Threonine Protein Phosphatase Assays, *Curr. Protoc. Mol. Biol. 92*, 18.18.11-18.18.11.

(188) Espanel, X., Huguenin-Reggiani, M., and van Huijsduijnen, R. H. (2002) The SPOT technique as a tool for studying protein tyrosine phosphatase substrate specificities, *Protein Sci. 11*, 2326-2334.

(189) Wood, S. E., Sinsinbar, G., Gudlur, S., Nallani, M., Huang, C.-F., Liedberg, B., and Mrksich, M. (2017) A Bottom-Up Proteomic Approach to Identify Substrate Specificity of Outer-Membrane Protease OmpT, *Angew. Chem. Int. Ed. 56*, 16531-16535.

(190) Agostinis, P., Goris, J., Pinna, L. A., Marchiori, F., Perich, J. W., Meyer, H. E., and Merlevede, W. (1990) Synthetic peptides as model substrates for the study of the specificity of the polycation-stimulated protein phosphatases, *Eur. J. Biochem. 189*, 235-241.

(191) McCloy, R. A., Parker, B. L., Rogers, S., Chaudhuri, R., Gayevskiy, V., Hoffman, N. J., Ali, N., Watkins, D. N., Daly, R. J., James, D. E., Lorca, T., Castro, A., and Burgess, A. (2015) Global Phosphoproteomic Mapping of Early Mitotic Exit in Human Cells Identifies Novel Substrate Dephosphorylation Motifs, *Mol. Cell. Proteomics 14*, 2194-2212.

(192) Lienhard, G. E. (2008) Non-functional phosphorylations?, *Trends Biochem. Sci 33*, 351-352.

(193) Huttlin, E. L., Jedrychowski, M. P., Elias, J. E., Goswami, T., Rad, R., Beausoleil, S. A., Villén, J., Haas, W., Sowa, M. E., and Gygi, S. P. (2010) A Tissue-Specific Atlas of Mouse Protein Phosphorylation and Expression, *Cell 143*, 1174-1189.

(194) Bennetzen, M. V., Larsen, D. H., Bunkenborg, J., Bartek, J., Lukas, J., and Andersen, J. S. (2010) Site-specific Phosphorylation Dynamics of the Nuclear Proteome during the DNA Damage Response, *Mol. Cell. Proteomics 9*, 1314.

(195) Houseman, B. T., Huh, J. H., Kron, S. J., and Mrksich, M. (2002) Peptide chips for the quantitative evaluation of protein kinase activity, *Nat. Biotechnol. 20*, 270-274.

(196) Xue, A. Y., Szymczak, L. C., Mrksich, M., and Bagheri, N. (2017) Machine Learning on Signal-to-Noise Ratios Improves Peptide Array Design in SAMDI Mass Spectrometry, *Anal. Chem. 89*, 9039-9047.

(197) Aarts, J. M. M. J. G., Wang, S., Houtman, R., van Beuningen, R. M. G. J., Westerink, W. M. A., Van De Waart, B. J., Rietjens, I. M. C. M., and Bovee, T. F. H. (2013) Robust Array-Based Coregulator Binding Assay Predicting ERα-Agonist Potency and Generating Binding Profiles Reflecting Ligand Structure, *Chem. Res. Toxicol. 26*, 336-346.

(198) Murayama, N., van Beuningen, R., Suemizu, H., Guguen-Guillouzo, C., Shibata, N., Yajima, K., Utoh, M., Shimizu, M., Chesne, C., Nakamura, M., Guengerich, F. P., Houtman, R., and Yamazaki, H. (2014) Thalidomide Increases Human Hepatic Cytochrome P450 3A Enzymes by Direct Activation of the Pregnane X Receptor, *Chem. Res. Toxicol. 27*, 304-308.

(199) Lesaicherre, M.-L., Uttamchandani, M., Chen, G. Y. J., and Yao, S. Q. (2002) Antibody-Based fluorescence detection of kinase activity on a peptide array, *Bioorg. Med. Chem. Lett. 12*, 2085-2088.

(200) Takahashi, M., Nokihara, K., and Mihara, H. (2003) Construction of a Protein-Detection System Using a Loop Peptide Library with a Fluorescence Label, *Chem. Biol. 10*, 53-60.

(201) Medzihradszky, K. F., Campbell, J. M., Baldwin, M. A., Falick, A. M., Juhasz, P., Vestal, M. L., and Burlingame, A. L. (2000) The Characteristics of Peptide Collision-Induced Dissociation Using a High-Performance MALDI-TOF/TOF Tandem Mass Spectrometer, *Anal. Chem. 72*, 552-558.

(202) Zenobi, R., and Knochenmuss, R. (1999) Ion formation in MALDI mass spectrometry, *Mass Spectrom. Rev. 17*, 337-366.

(203) Krause, E., Wenschuh, H., and Jungblut, P. R. (1999) The Dominance of Arginine-Containing Peptides in MALDI-Derived Tryptic Mass Fingerprints of Proteins, *Anal. Chem. 71*, 4160-4165.

(204) Valero, M.-L., Giralt, E., and Andreu, D. (1999) An investigation of residue-specific contributions to peptide desorption in MALDI-TOF mass spectrometry, *Lett. Pept. Sci. 6*, 109-115.

(205) Ryzhov, V., and Fenselau, C. (2001) Characterization of the protein subset desorbed by MALDI from whole bacterial cells, *Anal. Chem. 73*, 746-750.

(206) Pashkova, A., Moskovets, E., and Karger, B. L. (2004) Coumarin Tags for Improved Analysis of Peptides by MALDI-TOF MS and MS/MS. 1. Enhancement in MALDI MS Signal Intensities, *Anal. Chem. 76*, 4550-4557.

(207) Beavis, R. C., and Bridson, J. N. (1993) Epitaxial protein inclusion in sinapic acid crystals, *J. Phys. D: Appl. Phys. 26*, 442-7.

(208) Amado, F. M. L., Domingues, P., Santana-Marques, M. G., Ferrer-Correia, A. J., and Tomer, K. B. (1997) Discrimination effects and sensitivity variations in matrix-assisted laser desorption/ionization, *Rapid Commun. Mass Spectrom. 11*, 1347-1352.

(209) Baumgart, S., Lindner, Y., Kuehne, R., Oberemm, A., Wenschuh, H., and Krause, E. (2004) The contributions of specific amino acid side chains to signal intensities of peptides in matrix-assisted laser desorption/ionization mass spectrometry, *Rapid Commun. Mass Spectrom. 18*, 863-868.

(210) Stavenhagen, K., Hinneburg, H., Thaysen-Andersen, M., Hartmann, L., Silva, D. V., Fuchser, J., Kaspar, S., Rapp, E., Seeberger, P. H., and Kolarich, D. (2013) Quantitative mapping of glycoprotein micro-heterogeneity and macro-heterogeneity: an evaluation of mass spectrometry signal strengths using synthetic peptides and glycopeptides, *J. Mass Spectrom. 48*, 627-639.

(211) Stadtman, E. R., and Levine, R. L. (2003) Free radical-mediated oxidation of free amino acids and amino acid residues in proteins, *Amino Acids 25*, 207-218.

(212) Gay, S., Binz, P.-A., Hochstrasser, D. F., and Appel, R. D. (2002) Peptide mass fingerprinting peak intensity prediction: extracting knowledge from spectra, *Proteomics 2*, 1374-1391.

(213) Scheler, C., Lamer, S., Pan, Z., Li, X. P., Salnikow, J., and Jungblut, P. (1998) Peptide mass fingerprint sequence coverage from differently stained proteins on two-dimensional electrophoresis patterns by matrix assisted laser desorption/ionization-mass spectrometry (MALDI-MS), *Electrophoresis 19*, 918-927.

(214) Silva, J. C., Gorenstein, M. V., Li, G.-z., Vissers, J. P. C., and Geromanos, S. J. (2006) Absolute quantification of proteins by LCMS. A virtue of parallel MS acquisition, *Mol. Cell. Proteomics 5*, 144-156.

(215) Fusaro, V. A., Mani, D. R., Mesirov, J. P., and Carr, S. A. (2009) Prediction of high-responding peptides for targeted protein assays by mass spectrometry, *Nat. Biotechnol. 27*, 190-198.

(216) Sanders, W. S., Bridges, S. M., McCarthy, F. M., Nanduri, B., and Burgess, S. C. (2007) Prediction of peptides observable by mass spectrometry applied at the experimental set level, *BMC Bioinf. 8*, No pp. given.

(217) Svetnik, V., Liaw, A., Tong, C., Culberson, J. C., Sheridan, R. P., and Feuston, B. P. (2003) Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling, *J. Chem. Inf. Comput. Sci. 43*, 1947-1958.

(218) Mei, H., Liao, Z. H., Zhou, Y., and Li, S. Z. (2005) A new set of amino acid descriptors and its application in peptide QSARs, *Biopolymers 80*, 775-786.

(219) Schueuermann, G., Ebert, R.-U., Chen, J., Wang, B., and Kuehne, R. (2008) External Validation and Prediction Employing the Predictive Squared Correlation Coefficient - Test Set Activity Mean vs. Training Set Activity Mean, *J. Chem. Inf. Model. 48*, 2140-2145.

(220) Bobrowski, K., and Schoneich, C. (1993) Hydroxyl radical adducts at sulfur in substituted organic sulfides stabilized by an internal hydrogen bond, *J. Chem. Soc., Chem. Commun.*, 795-7.

(221) Lusci, A., Pollastri, G., and Baldi, P. (2013) Deep Architectures and Deep Learning in Chemoinformatics: The Prediction of Aqueous Solubility for Drug-Like Molecules, *J. Chem. Inf. Model. 53*, 1563-1575.

(222) Cedeno, W., and Agrafiotis, D. K. (2003) Using particle swarms for the development of QSAR models based on K-nearest neighbor and kernel regression, *J. Comput. Aided Mol. Des. 17*, 255-63.

(223) Janes, K. A., Kelly, J. R., Gaudet, S., Albeck, J. G., Sorger, P. K., and Lauffenburger, D. A. (2004) Cue-Signal-Response Analysis of TNF-Induced Apoptosis by Partial Least Squares Regression of Dynamic Multivariate Data, *J. Comput. Biol. 11*, 544-561.

(224) Smith, B. C., Settles, B., Hallows, W. C., Craven, M. W., and Denu, J. M. (2011) SIRT3 Substrate Specificity Determined by Peptide Arrays and Machine Learning, *ACS Chem. Biol. 6*, 146-157.

(225) Yoshida, M., Hinkley, T., Tsuda, S., Abul-Haija, Y. M., McBurney, R. T., Kulikov, V., Mathieson, J. S., Galiñanes Reyes, S., Castro, M. D., and Cronin, L. (2018) Using Evolutionary Algorithms and Machine Learning to Explore Sequence Space for the Discovery of Antimicrobial Peptides, *Chem 4*, 533-543.

(226) Tallorin, L., Wang, J., Kim, W. E., Sahu, S., Kosa, N. M., Yang, P., Thompson, M., Gilson, M. K., Frazier, P. I., Burkart, M. D., and Gianneschi, N. C. (2018) Discovering de novo peptide substrates for enzymes using machine learning, *Nat. Commun. 9*, 5253.

(227) Ramil, C. P., An, P., Yu, Z., and Lin, Q. (2016) Sequence-Specific 2-Cyanobenzothiazole Ligation, *J. Am. Chem. Soc. 138*, 5499-5502.

(228) Shao, Q., and Hall, C. K. (2016) Binding Preferences of Amino Acids for Gold Nanoparticles: A Molecular Simulation Study, *Langmuir 32*, 7888-7896.

Appendix

Supporting Information

Figures

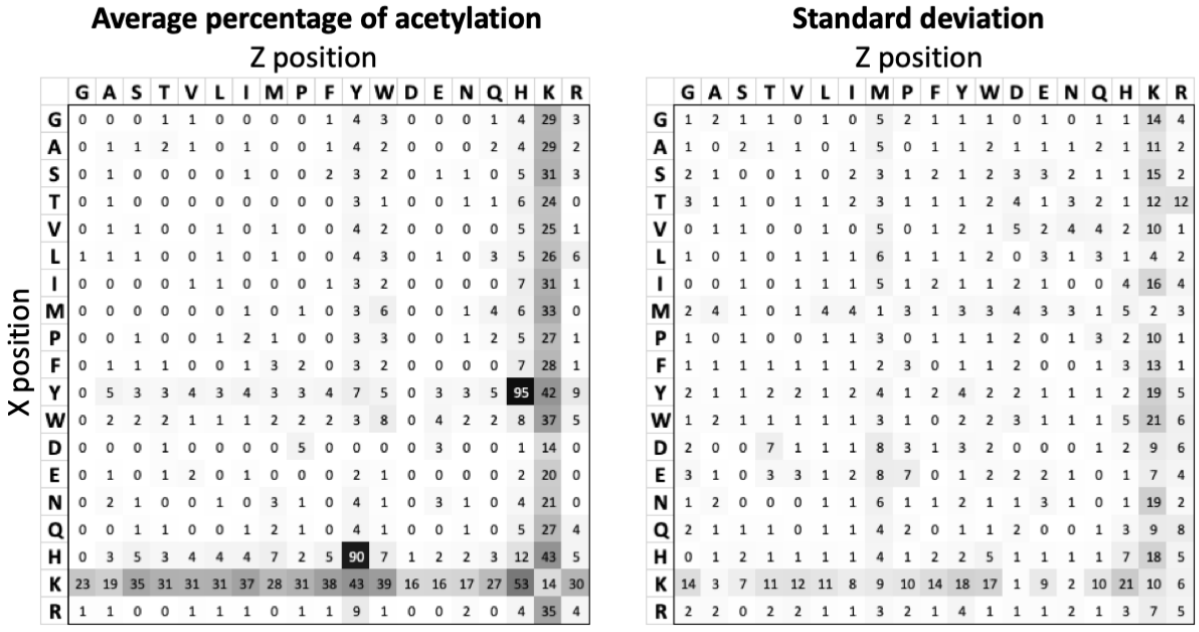


Figure S-1. Heatmaps showing the average percent yield and standard deviation of the acetylation of each peptide in the Ac-GXZRGC peptide array after treatment with 100 mM acetic anhydride at room temperature for 1 hour. These values were obtained from four replicates and correlate to the heatmap shown in Figure 2-3a.

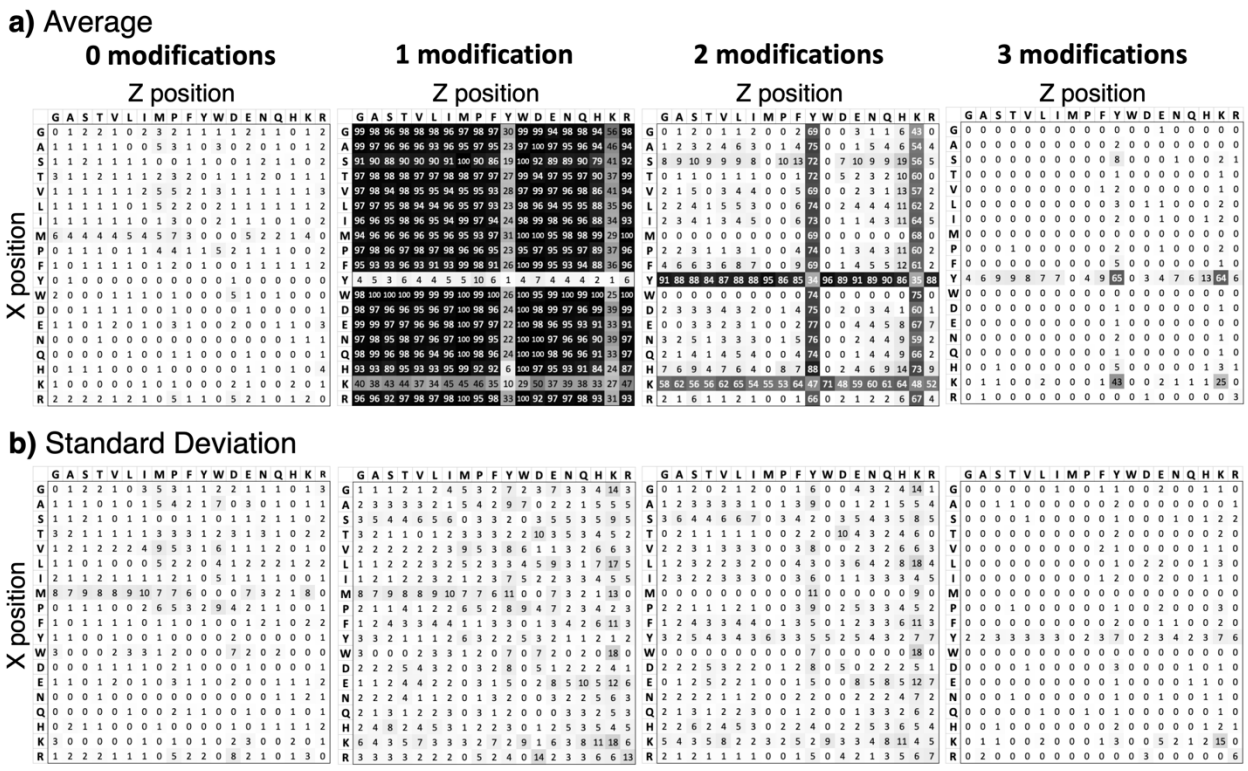


Figure S-2. Average percent yield and standard deviation of the acetylation of each peptide in the Ac-GXYHZRGC peptide array after treatment with 100 mM acetic anhydride at room temperature for 1 hour. The heatmaps above show the **a)** averaged percentage and **b)** standard deviation of the acetylation states of each peptide in the peptide array. These values were obtained from five replicates and correlate to the heatmap shown in Figure 2-8a.

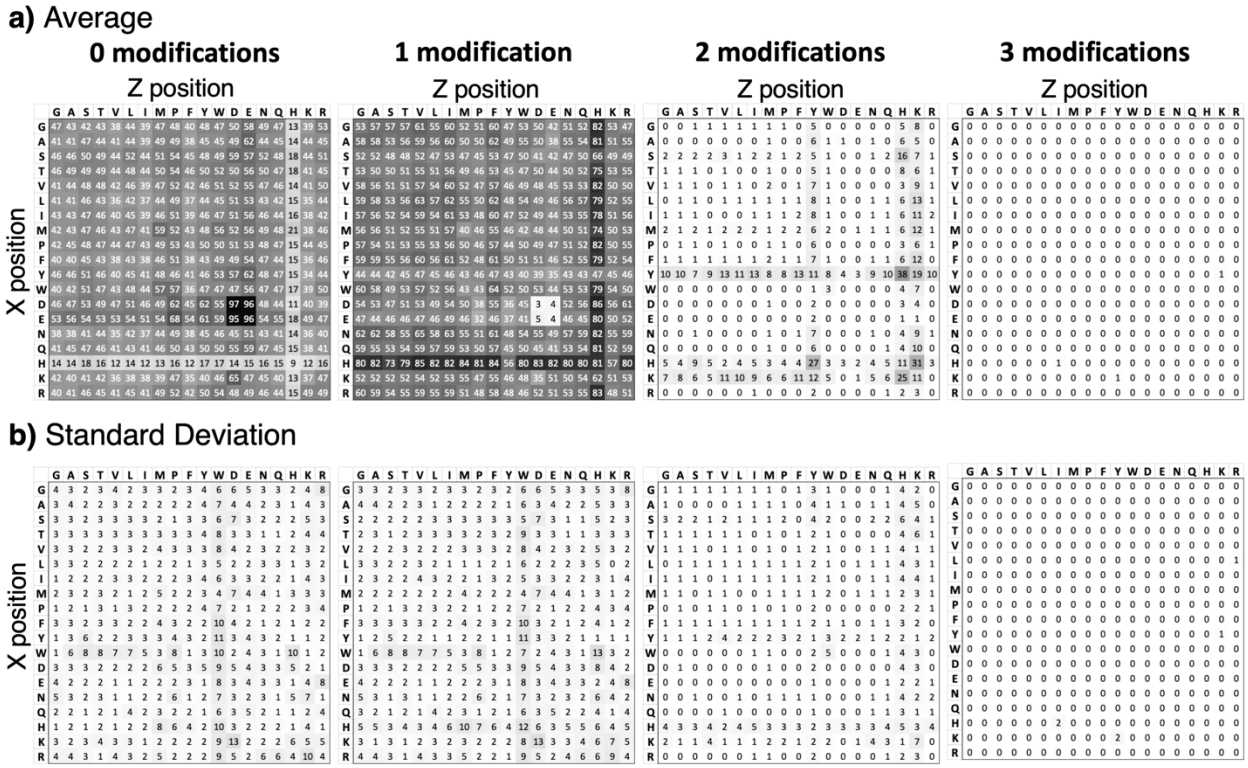


Figure S-3. Average percent yield and standard deviation of the acetylation of each peptide in the Ac-GXYHZRGC peptide array after treatment with 0.1 mM acetic anhydride at room temperature for 1 hour. The heatmaps above show the a) averaged percentage and b) standard deviation of the acetylation states of each peptide in the peptide array. These values were obtained from five replicates and correlate to the heatmap shown in Figure 2-8b.

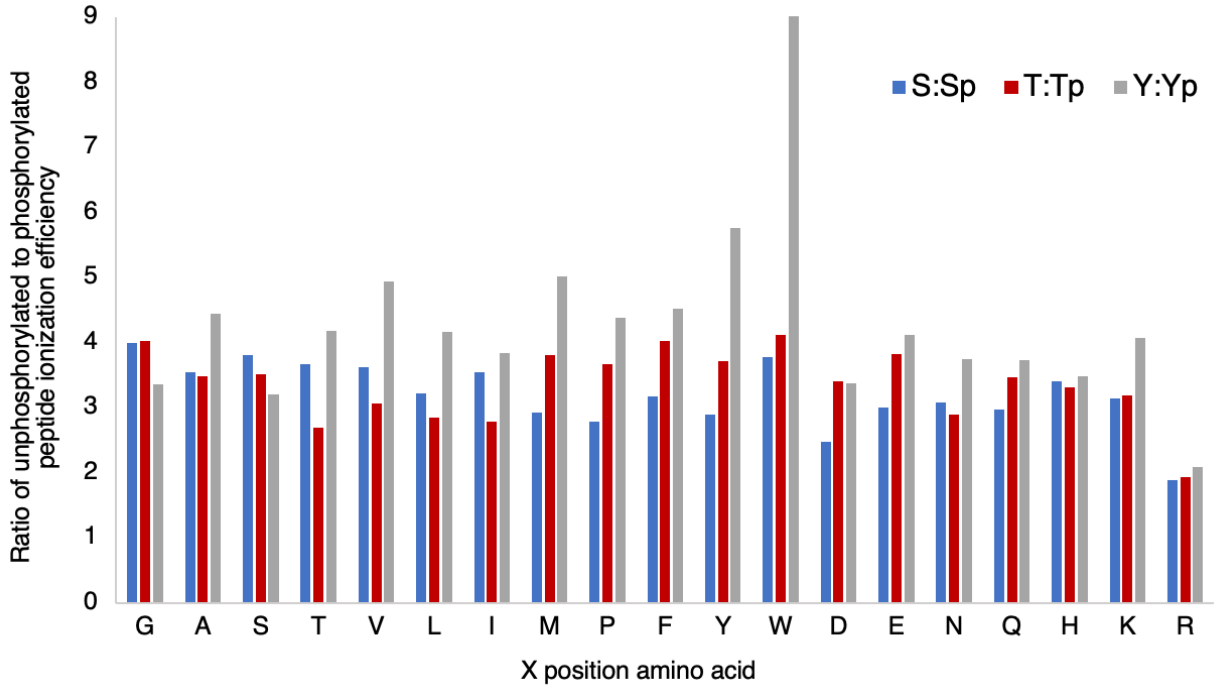


Figure S-4. Relative ratios of unphosphorylated to phosphorylated peptide ionization efficiency for 19 unphosphorylated/phosphorylated peptide pairs with the sequences Ac-GX-S/S^{p}-GGRC, Ac-GX-T/T^{p}-GGRC, and Ac-GX-Y/Y^{p}-GGRC. The unphosphorylated peptides generally have ionization efficiencies between 2-4 times greater than phosphorylated peptides of the same sequence. The relative ratios between S:Sp, T:Tp, and Y:Yp are relatively consistent, indicating that the central serine, threonine, or tyrosine residues do not have large effects on ionization efficiency.

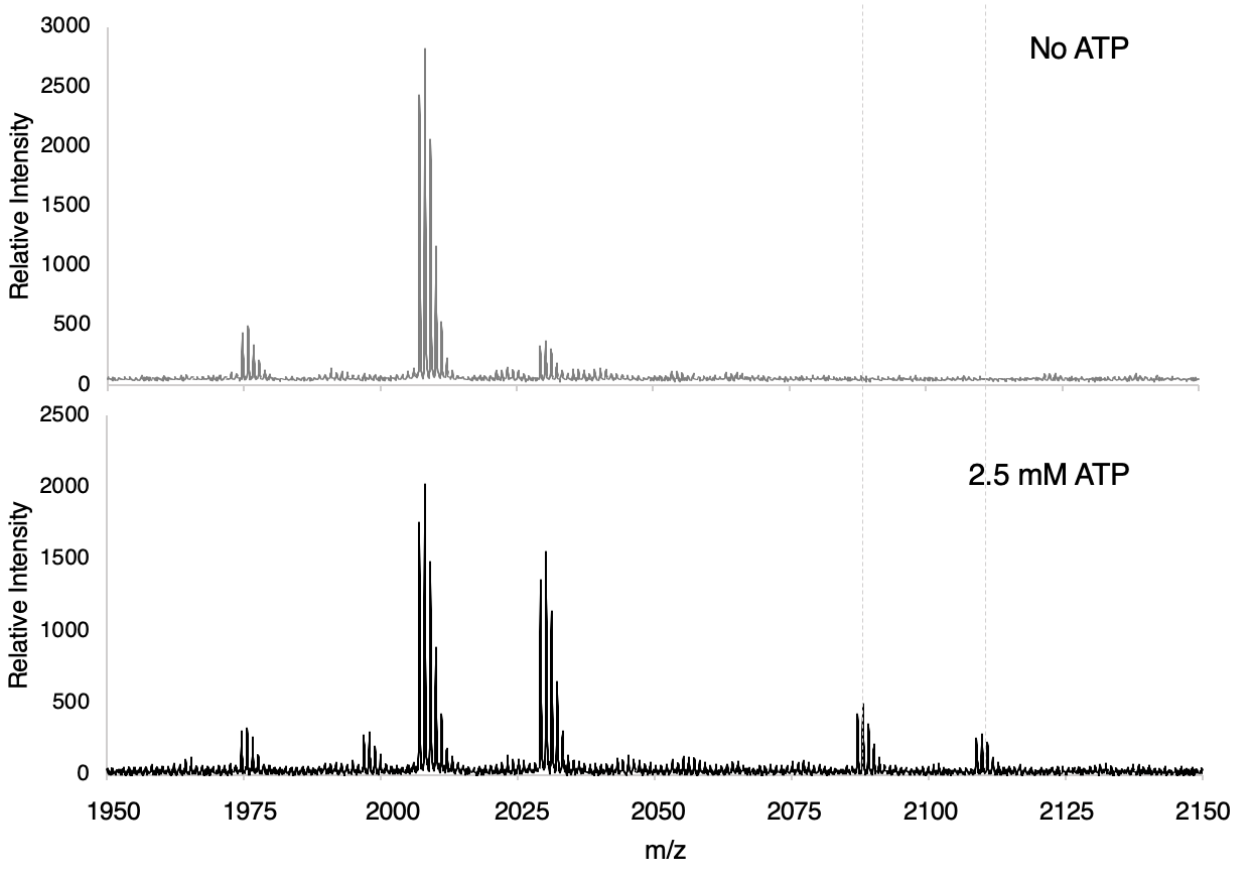


Figure S-5. ATP is required for kinase activity. MALDI-MS spectra of Ac-IYGEFKKKC peptide, a known substrate for the Src kinase, were treated with 100 nM active Src kinase in NIH/3T3 cell lysate with 20 mM MgCl2 and either 0 mM (top, gray spectrum) or 2.5 mM ATP (bottom, black spectrum) for 1.5 hours at 37 °C. Phosphorylation results in a peak shift of 80 Da. The expected H+ and Na+ adduct product masses are shown by the dashed gray lines. Kinase activity is only observed with the addition of ATP, indicating that ATP is required for kinase activity in cell lysates.

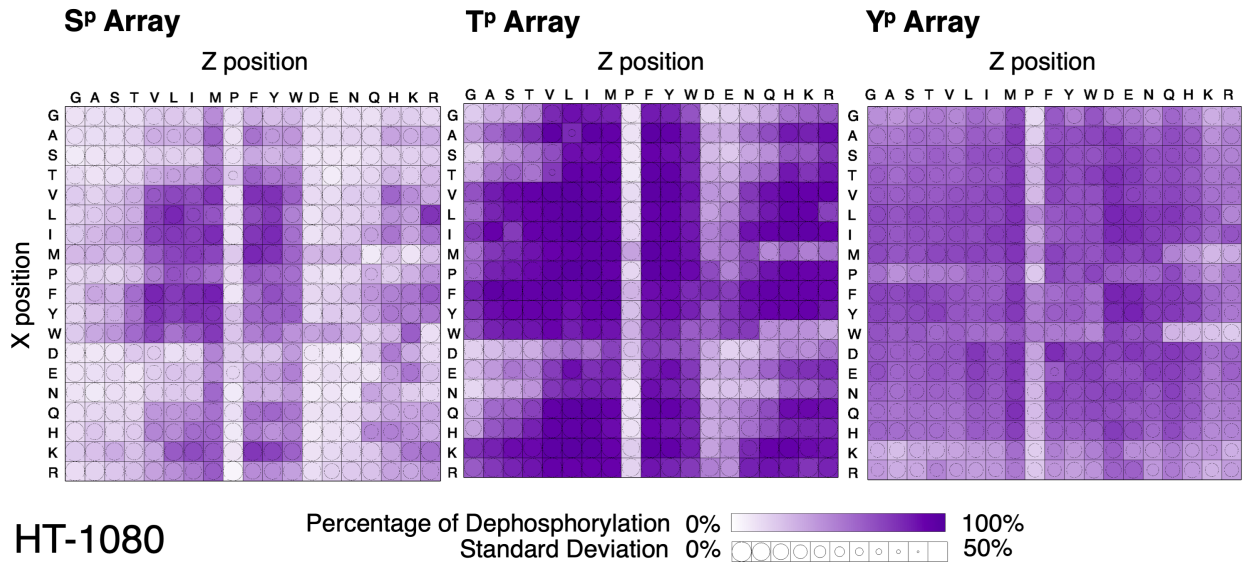


Figure S-6. Heatmaps showing global phosphatase activity in cell lysate from the HT-1080 cell line. The average extent of dephosphorylation of each peptide is plotted in the heatmaps, where complete dephosphorylation is dark purple, and each square represents a peptide having X and Z residues as denoted on the vertical and horizontal axes, respectively. The standard deviation of the dephosphorylation of each peptide is displayed by circle size in each peptide square, with larger standard deviations resulting in smaller circles. The proportion of dephosphorylated peptide was calculated by dividing the area under the curve (AUC) of the peak for the dephosphorylated product by the total AUC of the peaks for unreacted substrate and dephosphorylated product.

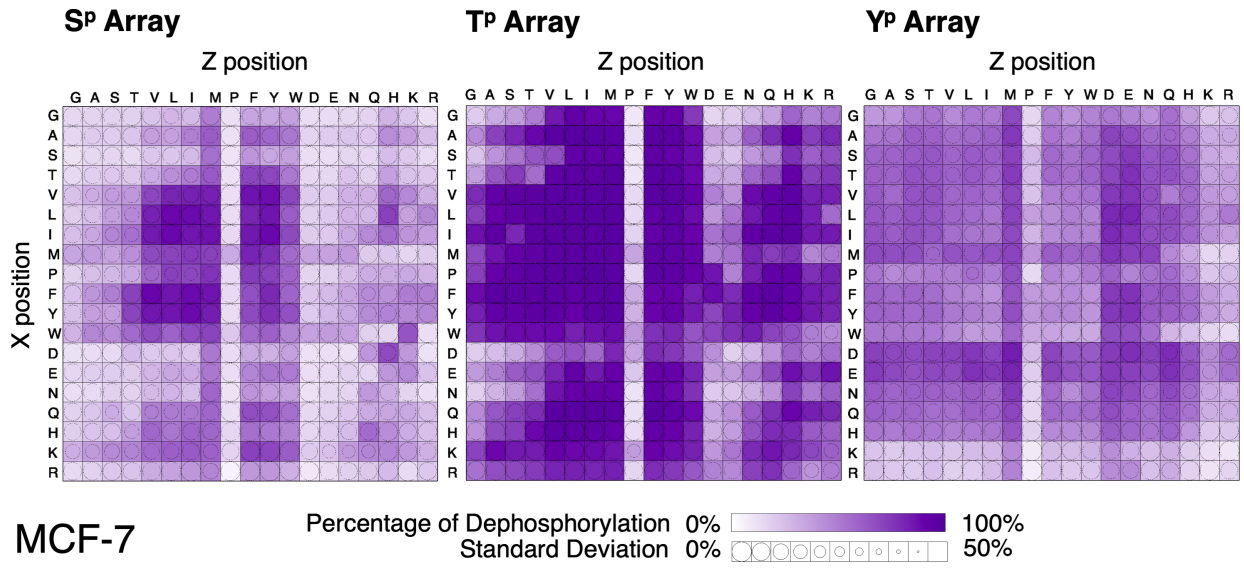


Figure S-7. Heatmaps showing global phosphatase activity in cell lysate from the MCF-7 cell line. The average extent of dephosphorylation of each peptide is plotted in the heatmaps, where complete dephosphorylation is dark purple, and each square represents a peptide having X and Z residues as denoted on the vertical and horizontal axes, respectively. The standard deviation of the dephosphorylation of each peptide is displayed by circle size in each peptide square, with larger standard deviations resulting in smaller circles. The proportion of dephosphorylated peptide was calculated by dividing the area under the curve (AUC) of the peak for the dephosphorylated product by the total AUC of the peaks for unreacted substrate and dephosphorylated product.

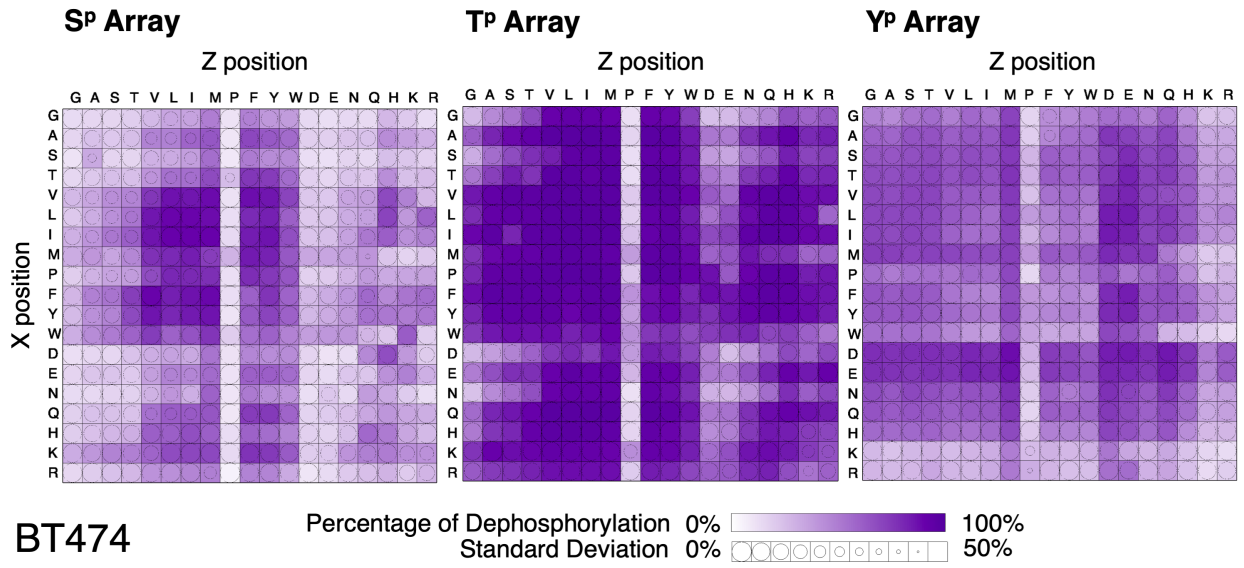


Figure S-8. Heatmaps showing global phosphatase activity in cell lysate from the BT474 cell line. The average extent of dephosphorylation of each peptide is plotted in the heatmaps, where complete dephosphorylation is dark purple, and each square represents a peptide having X and Z residues as denoted on the vertical and horizontal axes, respectively. The standard deviation of the dephosphorylation of each peptide is displayed by circle size in each peptide square, with larger standard deviations resulting in smaller circles. The proportion of dephosphorylated peptide was calculated by dividing the area under the curve (AUC) of the peak for the dephosphorylated product by the total AUC of the peaks for unreacted substrate and dephosphorylated product.

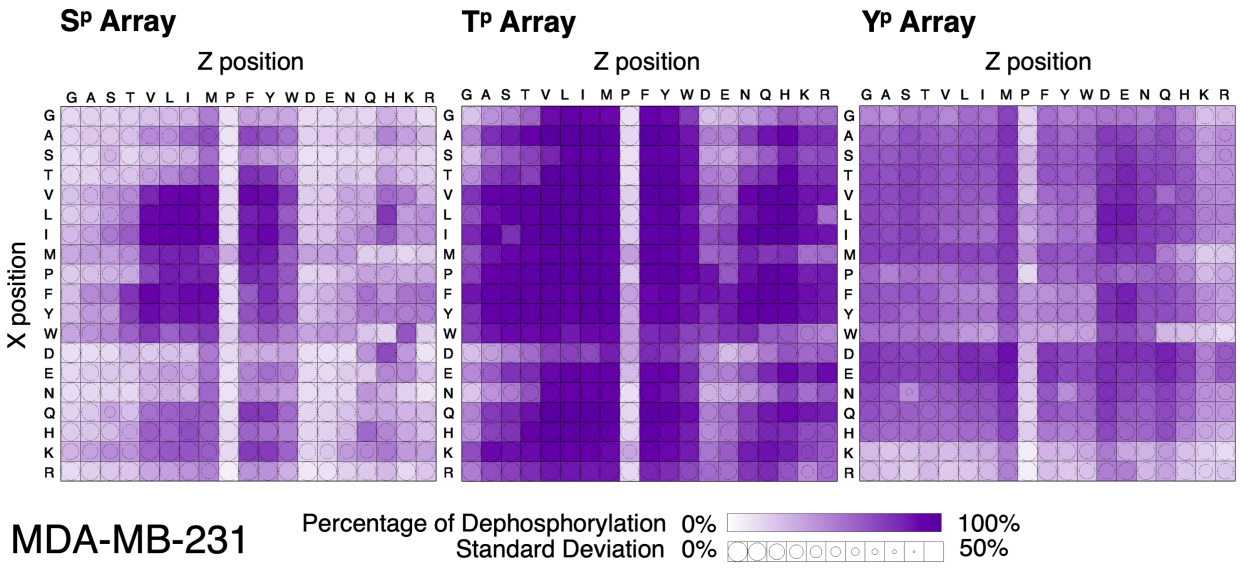


Figure S-9. Heatmaps showing global phosphatase activity in cell lysate from the MDA-MB-231 cell line. The average extent of dephosphorylation of each peptide is plotted in the heatmaps, where complete dephosphorylation is dark purple, and each square represents a peptide having X and Z residues as denoted on the vertical and horizontal axes, respectively. The standard deviation of the dephosphorylation of each peptide is displayed by circle size in each peptide square, with larger standard deviations resulting in smaller circles. The proportion of dephosphorylated peptide was calculated by dividing the area under the curve (AUC) of the peak for the dephosphorylated product by the total AUC of the peaks for unreacted substrate and dephosphorylated product.

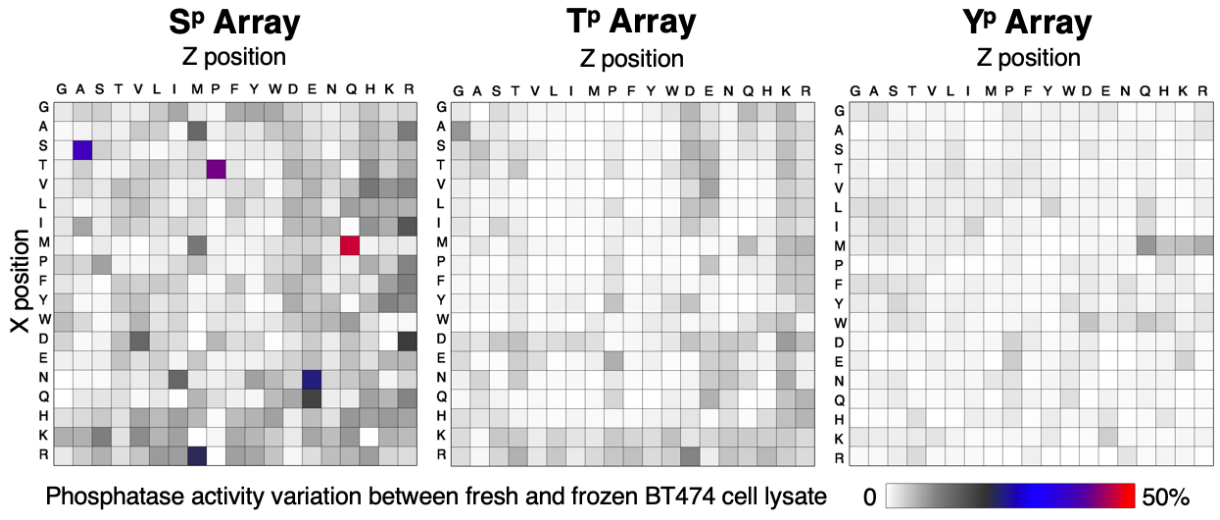


Figure S-10. Heatmaps displaying the variation of phosphatase activity from fresh vs frozen BT474 lysate on each peptide in the phosphorylated peptide arrays. Each square represents a peptide having X and Z residues as denoted on the vertical and horizontal axes, respectively. Peptides with a variance of less than 12.5% are white or light gray, between 12.5-25% are dark gray, 25-37.5% are blue and 37.5-50% are red.

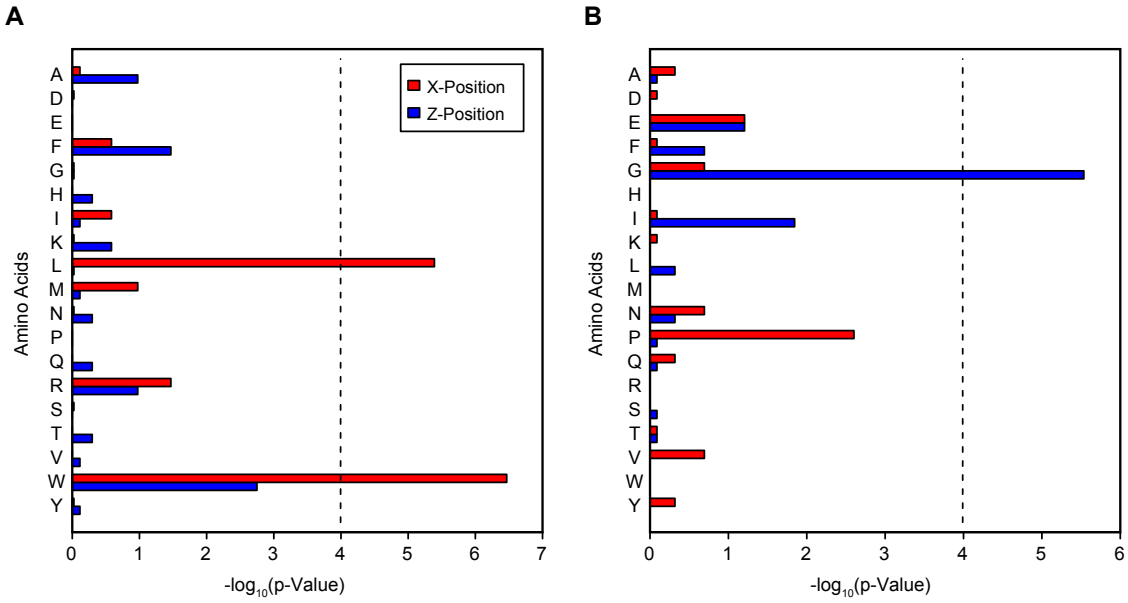


Figure S-11. Peptide S/N is affected most by tryptophan, leucine, and glycine in K-array peptides. The peptide's S/N were calculated from the AUC of the peptide peaks in each mass spectrum for 11 control plates containing Ac-GRK^{ac}XZC peptides. Peptides were identified to have low or high S/N as shown in shaded region in Figure 2 with 50 low and 30 high S/N peptides (See Methods in Chapter 4). A Fischer exact test (Bonferroni corrected p<10-4) was performed to determine general trends in peptides containing specific amino acids. The reported p- value is the chance the observed number of amino acids is within the bottom 50 or top 30 by random chance. (A) The low S/N region was enriched with peptides having X-position tryptophan, W, and leucine, L, and (B) the high S/N region was enriched with peptides having Z-position glycine, G.

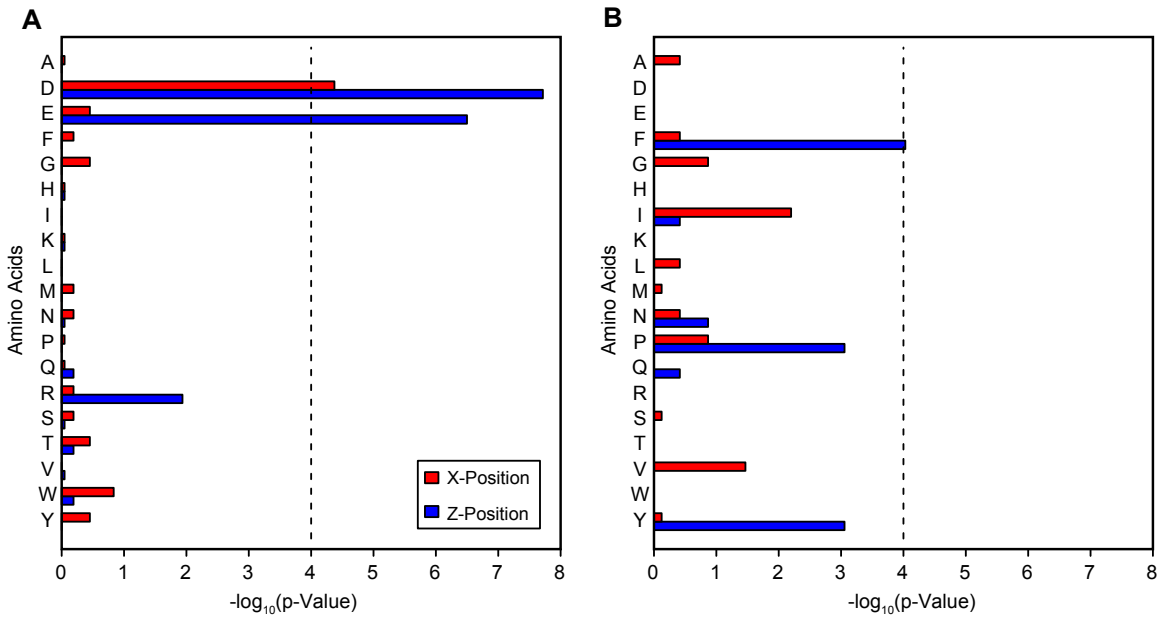


Figure S-12. Peptide S/N is affected most by aspartic acid, glutamic acid, and phenylalanine in H-array peptides. The peptide's S/N were calculated from the AUC of the peptide peaks in each mass spectrum for 11 control plates containing Ac-GXZHGC peptides. Peptides were identified to have low or high S/N as shown in shaded region in Figure 2 with 40 low and 25 high S/N peptides. A Fischer exact test (Bonferroni corrected $p<10\text{-}4$) was performed to determine general trends in peptides containing specific amino acids. The reported p- value is the chance the observed number of amino acids is within the bottom 40 or top 25 by random chance. (A) The low S/N region was enriched with peptides containing X- or Z-position aspartic acid, D, and Z-position glutamic acid, E, and (B) the high S/N region was enriched with peptides having Z-position phenylalanine, F.

Tables

Table S-1. S^p Array Data Summary

Peptide Sequence: Ac-GXS^{p}ZGRC		Peptide Signal-to-Noise		% Dephosphorylation: NIH/3T3		% Dephosphorylation: HT-1080		% Dephosphorylation: MCF-7		% Dephosphorylation: BT474		% Dephosphorylation: MDA-MB-231	
X position	Z position	Average	Std. Dev.	Average	Std. Dev.	Average	Std. Dev.	Average	Std. Dev.	Average	Std. Dev.	Average	Std. Dev.
G	G	224	24	14.3	6.4	11.4	2.9	15.2	3.7	13.7	1.0	15.2	4.7
A	G	245	9	18.3	13.2	12.8	7.6	15.6	3.5	16.2	3.0	17.0	6.3
S	G	271	60	13.1	9.2	8.4	2.7	11.3	2.6	11.0	1.0	13.9	4.4
T	G	222	102	15.0	10.9	13.4	7.2	14.5	3.5	15.4	2.3	15.7	4.7
V	G	221	81	27.6	18.3	19.9	4.7	23.4	7.2	26.9	9.0	23.4	8.4
L	G	205	45	25.3	19.1	18.1	3.5	19.2	4.8	21.2	5.2	19.5	6.2
I	G	187	6	26.6	16.9	21.0	5.8	25.1	7.8	30.2	9.5	26.9	7.2
M	G	54	0	30.9	9.3	27.0	2.4	34.4	4.4	31.9	1.5	32.9	4.1
P	G	242	56	23.1	15.5	15.1	5.0	18.2	4.6	19.2	4.1	20.2	5.0
F	G	223	60	25.1	15.3	18.2	2.1	23.9	6.7	28.1	5.5	25.0	3.3
Y	G	182	37	24.2	11.8	15.6	3.0	23.5	7.3	26.2	4.3	25.5	4.1
W	G	48	39	30.1	12.7	20.4	4.4	29.0	6.7	31.1	1.5	34.4	4.8
D	G	178	37	13.1	6.6	9.3	3.1	11.8	3.4	12.5	0.7	12.6	1.9
E	G	162	12	16.4	9.9	10.1	3.9	15.6	5.7	16.4	3.0	15.1	3.4
N	G	170	22	11.9	5.2	10.5	6.2	12.4	4.2	13.9	1.6	13.8	3.5
Q	G	139	72	16.7	8.2	12.5	2.4	17.7	6.3	18.2	3.6	17.7	6.3
H	G	172	24	17.1	7.9	14.7	3.7	19.0	6.9	19.3	4.1	20.0	5.2
K	G	165	17	18.1	3.2	20.5	0.6	30.8	13.1	31.7	2.9	29.1	9.0
R	G	157	11	15.0	2.0	13.6	1.4	14.5	5.1	15.3	1.8	14.2	4.1
G	A	154	48	14.3	4.6	12.7	4.9	16.5	4.7	15.8	2.6	17.7	4.5

A	A	190	29	15.9	6.5	13.1	3.9	19.8	9.2	23.8	9.9	21.2	7.2
S	A	129	42	12.2	5.1	10.5	1.3	14.9	6.4	32.9	28.6	15.8	5.5
T	A	144	48	13.8	5.3	10.8	3.3	18.6	8.1	21.1	9.7	17.0	4.4
V	A	145	57	19.2	9.5	17.3	6.6	32.2	16.6	34.0	16.9	30.6	12.4
L	A	199	23	30.2	21.4	21.0	11.8	24.2	9.8	31.3	13.3	25.6	9.0
I	A	142	48	44.0	18.6	27.3	10.3	34.1	13.6	40.7	17.9	35.6	10.6
M	A	53	14	38.6	10.9	29.7	4.1	38.0	5.6	39.9	7.2	38.1	4.3
P	A	201	85	28.7	18.8	18.2	6.2	24.2	9.5	29.4	13.3	24.6	8.9
F	A	189	79	42.6	20.9	34.2	12.3	40.0	12.6	48.3	12.9	44.9	6.1
Y	A	182	63	39.0	15.2	24.2	2.1	37.0	10.0	42.4	12.7	38.9	7.0
W	A	32	29	39.8	12.3	33.3	7.3	45.8	10.1	43.9	7.9	41.2	8.7
D	A	123	24	15.1	8.0	9.6	3.5	13.2	5.7	15.7	4.5	13.8	3.7
E	A	142	69	17.8	9.2	12.1	6.2	18.8	7.9	22.2	7.5	19.8	7.5
N	A	223	95	14.1	7.5	8.8	2.3	14.0	5.6	14.9	4.6	14.9	3.8
Q	A	182	33	18.4	10.1	14.9	6.4	21.7	8.5	23.8	10.5	21.5	6.8
H	A	153	80	22.5	11.7	21.8	9.1	22.9	6.7	23.7	7.7	22.4	6.1
K	A	184	78	22.5	8.5	23.7	3.6	38.4	14.5	39.5	7.0	36.0	11.5
R	A	193	93	18.8	4.0	17.3	1.5	17.4	4.5	19.6	1.3	18.5	3.8
G	S	164	7	16.5	6.8	12.7	2.8	17.8	6.6	16.2	2.2	17.3	5.3
A	S	121	16	17.7	7.0	13.9	3.8	19.5	5.8	21.0	6.7	21.1	5.6
S	S	157	1	14.0	7.1	11.8	4.2	15.9	5.8	16.3	4.6	29.1	20.6
T	S	130	8	16.2	7.2	12.9	2.8	21.5	9.7	24.7	11.6	23.9	8.5
V	S	154	48	23.2	9.0	22.7	6.0	37.9	15.0	43.1	18.0	40.2	10.9
L	S	141	68	30.1	12.1	25.4	7.9	36.4	14.2	42.4	17.6	38.3	13.4
I	S	123	13	30.4	9.4	33.1	7.1	47.1	14.3	51.9	17.9	50.6	10.8
M	S	41	5	29.5	6.1	26.7	5.0	35.3	10.2	34.5	10.2	36.9	7.9

P	S	132	14	20.3	9.8	16.5	3.3	27.2	11.2	35.2	21.0	29.2	10.2
F	S	91	13	34.9	12.3	30.7	1.9	48.5	11.4	53.1	14.6	48.8	8.8
Y	S	139	54	39.3	16.9	30.7	11.5	37.0	12.3	47.0	11.0	38.5	7.5
W	S	18	13	48.7	12.8	40.2	11.7	43.7	7.5	50.4	8.8	45.9	5.8
D	S	125	59	16.2	9.0	10.1	2.2	13.9	2.1	15.6	0.9	14.3	3.3
E	S	108	52	20.4	12.7	15.0	8.0	19.4	4.9	22.8	7.1	21.1	6.6
N	S	105	31	17.1	9.4	12.4	6.0	17.9	5.5	16.9	2.2	16.9	5.7
Q	S	90	0	22.0	12.7	15.4	3.9	31.7	16.7	24.0	9.2	35.2	23.0
H	S	131	50	22.7	12.8	16.9	3.8	24.5	6.7	27.8	7.0	25.4	7.9
K	S	87	3	29.4	5.9	31.6	6.1	44.7	12.0	47.6	1.9	43.6	12.1
R	S	112	23	25.4	5.1	21.9	3.1	21.4	3.9	23.9	1.5	21.2	3.6
G	T	144	79	21.4	10.8	13.1	2.6	17.8	3.5	18.8	3.7	18.9	4.6
A	T	147	96	25.4	12.4	17.1	5.0	22.3	4.9	25.3	8.9	24.5	8.2
S	T	155	62	18.9	9.2	12.4	2.4	14.1	4.4	16.2	3.6	16.4	4.8
T	T	127	21	21.9	11.1	15.6	6.6	20.8	6.4	24.8	8.6	22.7	9.9
V	T	152	40	33.0	14.8	31.1	10.3	39.2	11.3	47.2	20.3	44.5	13.0
L	T	108	28	36.0	15.7	31.5	10.7	45.1	13.8	52.4	20.1	50.8	10.5
I	T	83	4	38.6	15.8	38.9	13.0	54.3	14.1	61.2	17.3	61.0	8.7
M	T	34	6	35.5	6.8	31.9	4.6	42.7	10.3	45.7	10.9	43.3	9.2
P	T	145	87	26.9	13.4	23.3	9.8	32.8	11.3	38.1	15.7	32.4	9.3
F	T	105	21	55.8	7.9	55.1	9.8	71.4	3.8	73.1	8.0	76.1	3.4
Y	T	132	75	52.5	4.7	51.6	10.8	70.4	5.7	68.6	9.5	72.3	3.7
W	T	18	8	56.2	7.0	56.3	7.9	57.0	8.5	59.3	6.8	59.1	4.7
D	T	49	20	27.2	14.5	20.8	7.7	25.8	4.9	23.6	5.2	24.1	3.2
E	T	89	12	13.5	4.6	9.7	2.6	19.5	7.8	20.9	8.8	21.5	3.8
N	T	97	9	17.2	11.0	10.9	3.4	17.5	6.8	17.6	6.0	15.8	4.0

Q	T	144	53	23.9	14.2	15.3	6.5	21.6	8.6	23.5	9.3	24.8	6.5
H	T	95	7	26.8	17.1	20.2	10.9	37.1	13.9	31.0	8.4	35.5	13.8
K	T	135	77	22.9	11.0	24.1	7.0	46.7	14.0	48.2	8.4	40.8	11.3
R	T	105	7	29.2	10.5	25.1	5.9	22.7	6.5	27.6	2.5	22.0	3.5
G	V	121	19	30.3	17.3	19.9	8.1	22.6	5.2	26.8	8.8	23.6	7.0
A	V	105	2	38.5	22.1	31.4	13.0	37.4	10.9	48.0	20.4	45.1	10.1
S	V	116	9	27.5	15.4	18.6	6.2	21.8	7.2	29.8	13.4	24.9	9.5
T	V	98	25	31.9	17.4	30.6	13.3	32.6	12.1	41.0	19.5	36.8	14.5
V	V	125	8	58.6	18.7	63.5	15.4	69.4	15.1	75.2	15.3	80.3	4.7
L	V	130	79	66.6	11.4	71.1	10.2	82.2	8.3	85.5	5.8	88.7	1.8
I	V	85	20	64.6	13.0	71.5	10.1	84.2	7.2	85.9	6.2	89.8	1.1
M	V	22	3	56.9	9.7	61.9	4.1	68.3	11.2	77.3	7.9	79.2	1.8
P	V	102	17	47.5	18.5	48.6	13.2	58.1	14.5	67.8	19.1	68.5	8.1
F	V	113	23	80.7	3.6	81.2	5.0	89.4	2.6	90.0	1.2	91.3	0.7
Y	V	67	4	76.7	2.3	75.9	7.8	85.4	4.6	86.0	3.9	88.1	1.5
W	V	15	9	71.0	6.6	69.4	8.5	67.9	5.4	73.5	3.9	73.7	3.6
D	V	106	32	23.6	15.4	20.2	16.7	21.0	8.9	31.4	13.2	20.4	6.4
E	V	89	26	26.3	13.4	18.7	12.9	28.1	10.6	37.0	15.2	31.7	4.8
N	V	103	4	22.6	10.7	17.9	4.8	27.0	12.2	29.5	14.5	28.7	8.7
Q	V	90	1	39.3	6.3	39.7	10.2	49.6	15.4	54.3	20.6	54.0	9.5
H	V	95	2	43.2	6.0	46.4	6.3	57.8	9.0	60.8	3.7	60.9	4.7
K	V	105	32	41.4	14.2	32.8	4.1	56.4	10.7	58.8	12.5	57.5	7.2
R	V	211	54	43.9	4.7	35.9	4.4	31.9	5.1	41.5	3.7	34.4	0.9
G	L	108	37	30.8	14.9	21.4	3.7	28.1	11.4	34.7	14.6	26.4	7.8
A	L	212	105	36.7	20.0	30.6	15.9	36.0	12.5	48.4	20.6	42.3	13.9
S	L	203	84	27.8	15.3	18.4	9.8	22.2	6.7	32.1	13.6	25.7	11.0

T	L	143	27	37.7	15.0	36.4	16.8	42.7	13.0	54.5	19.9	45.0	4.7
V	L	191	116	67.3	9.3	69.2	8.1	81.2	13.3	86.5	5.1	89.2	1.7
L	L	131	3	74.0	6.4	80.5	7.0	89.1	3.0	89.5	2.3	91.1	1.2
I	L	130	0	72.9	6.7	74.6	9.6	88.7	4.2	90.4	1.4	91.9	1.2
M	L	35	5	56.4	3.9	70.3	11.2	70.3	13.5	79.8	3.9	81.5	3.9
P	L	164	25	55.9	13.6	65.3	15.7	72.1	16.7	80.5	9.0	83.6	3.5
F	L	123	11	71.9	2.4	73.7	5.5	85.4	2.2	83.0	2.0	85.5	1.6
Y	L	172	20	74.1	1.6	71.5	6.7	85.4	3.5	80.3	2.3	83.8	0.4
W	L	17	13	60.7	3.7	52.9	4.2	59.3	5.3	59.8	3.5	59.1	4.6
D	L	160	55	25.8	8.3	12.2	2.5	21.9	7.6	35.7	12.4	24.3	7.9
E	L	131	17	35.1	15.2	23.0	10.1	34.1	9.9	47.8	17.5	37.5	10.6
N	L	129	18	27.2	8.6	20.6	9.1	28.8	8.8	40.5	15.0	29.9	7.9
Q	L	124	28	39.2	10.8	39.4	12.0	50.2	14.0	61.9	20.2	59.7	10.0
H	L	134	49	46.0	9.3	43.7	10.1	51.9	6.7	61.0	5.9	58.3	3.8
K	L	112	7	59.5	8.0	62.4	4.0	60.8	11.3	67.7	1.9	63.7	1.9
R	L	139	28	49.3	6.3	42.3	7.5	35.5	2.2	42.5	2.9	37.0	3.3
G	I	153	11	27.7	12.7	18.8	6.2	26.1	6.9	35.4	12.9	29.0	8.9
A	I	176	4	36.7	12.9	32.0	7.9	49.2	14.4	58.8	22.9	58.9	13.2
S	I	168	62	23.6	8.0	18.1	6.0	26.2	10.5	37.3	18.6	30.8	9.8
T	I	117	8	35.9	8.3	36.5	5.5	50.3	14.3	55.2	21.6	55.4	9.6
V	I	159	41	56.1	5.5	66.3	8.8	81.9	10.6	85.5	7.4	90.5	2.9
L	I	125	23	70.6	7.7	69.6	16.0	88.2	3.6	89.2	2.8	91.8	0.8
I	I	174	45	73.7	9.0	72.2	10.6	85.6	9.0	90.5	3.2	92.3	1.5
M	I	53	24	66.3	6.3	72.9	10.0	74.3	14.0	83.4	3.9	83.1	4.9
P	I	203	109	54.9	18.9	60.8	13.7	69.6	18.0	80.1	12.5	80.8	6.8
F	I	98	20	79.7	2.8	78.2	1.5	88.2	3.6	86.8	1.6	88.8	1.5

Y	I	62	28	77.4	1.5	76.0	6.8	88.1	3.8	84.3	2.2	87.2	1.0
W	I	16	11	65.9	1.8	61.5	8.0	67.6	4.9	65.1	5.2	67.7	3.7
D	I	93	16	23.5	10.4	15.5	8.4	21.4	7.0	33.1	11.3	23.9	7.3
E	I	128	38	32.6	11.7	23.7	6.3	32.0	9.9	46.2	17.4	34.9	12.4
N	I	159	27	23.7	9.9	21.5	7.6	24.3	6.8	30.5	8.7	26.3	5.1
Q	I	117	17	39.5	10.5	43.0	11.2	51.0	14.6	62.6	19.3	62.4	11.5
H	I	125	44	53.3	10.7	55.7	11.3	58.4	8.2	66.6	2.2	67.1	8.5
K	I	147	33	61.3	8.5	68.0	5.0	62.3	7.0	69.4	1.5	64.1	1.0
R	I	147	21	54.5	6.0	50.7	9.3	39.5	0.6	46.6	3.7	41.6	5.1
G	M	28	1	40.9	5.7	43.8	6.3	49.7	5.4	49.2	4.9	49.6	2.7
A	M	27	8	52.3	3.6	59.6	4.1	61.4	13.8	64.0	11.3	66.3	12.9
S	M	29	0	42.3	5.0	48.1	2.1	54.1	6.2	55.4	6.5	55.1	1.8
T	M	27	1	43.3	5.3	53.2	9.1	62.8	11.0	68.3	9.8	67.5	0.6
V	M	32	1	61.4	7.1	75.7	3.7	82.8	8.7	88.4	1.3	87.8	2.8
L	M	26	2	65.4	5.7	64.8	12.5	85.2	4.7	87.8	0.5	87.8	3.2
I	M	40	9	64.7	6.8	78.7	6.0	86.9	6.2	90.7	0.9	91.2	3.4
M	M	16	7	61.2	5.4	69.6	2.8	72.1	16.2	72.9	14.7	74.3	15.4
P	M	42	17	46.1	6.4	53.6	6.7	76.0	11.2	78.0	4.3	81.0	3.0
F	M	43	17	81.0	7.2	82.8	7.8	85.9	5.9	89.1	1.9	89.2	5.4
Y	M	32	10	75.8	6.1	76.0	11.4	83.4	3.1	86.2	0.7	85.5	4.6
W	M	7	2	67.0	3.0	74.2	5.9	70.3	6.0	75.6	1.4	73.9	5.1
D	M	21	6	46.2	4.8	47.0	2.9	51.5	1.7	52.9	4.6	50.7	2.7
E	M	29	7	43.6	3.7	44.5	4.1	51.5	6.0	56.6	12.3	53.2	5.8
N	M	25	3	52.3	2.3	57.3	5.1	58.5	4.1	59.7	3.4	57.8	0.5
Q	M	25	1	45.1	4.2	53.7	7.5	58.2	8.4	64.4	6.1	61.4	3.5
H	M	37	13	47.2	7.0	57.2	10.9	57.8	9.5	60.7	2.8	59.4	3.8

K	M	28	17	54.4	10.5	63.2	7.4	59.8	12.4	69.1	3.1	61.3	10.1
R	M	24	7	59.5	5.3	62.1	8.1	46.1	10.3	49.2	19.2	46.9	11.9
G	P	171	47	13.7	8.2	12.0	5.1	12.3	1.1	11.1	1.6	11.8	2.0
A	P	140	29	10.0	4.7	11.6	6.2	11.6	2.8	9.9	0.5	11.3	1.9
S	P	150	20	15.3	16.1	13.1	10.2	10.8	2.3	9.1	0.6	10.1	2.3
T	P	117	0	21.7	28.2	21.7	27.7	11.4	3.1	25.7	27.4	11.2	2.2
V	P	125	34	12.6	5.3	10.9	4.2	12.8	1.5	11.5	0.6	14.2	2.4
L	P	116	5	10.4	3.3	12.6	4.4	13.4	3.1	12.9	0.9	14.6	2.8
I	P	107	5	12.2	5.6	11.2	1.6	14.7	2.3	14.6	0.5	16.0	3.8
M	P	30	0	23.2	6.6	21.3	1.8	32.7	4.1	29.4	3.0	32.7	1.9
P	P	91	2	12.5	4.8	13.6	5.3	14.2	2.3	12.5	1.2	13.9	2.8
F	P	101	6	9.7	4.0	10.0	2.0	14.2	3.5	12.9	1.3	14.2	1.6
Y	P	122	3	20.0	5.1	23.4	10.7	13.4	2.7	10.8	0.8	13.4	2.4
W	P	18	11	22.2	4.5	22.6	6.5	21.2	3.2	19.5	1.2	19.7	1.5
D	P	61	35	16.1	5.6	18.8	3.4	13.8	1.2	11.3	2.2	14.4	3.1
E	P	91	7	13.8	14.3	16.0	17.4	12.4	2.9	11.3	1.1	12.2	2.2
N	P	205	88	7.3	2.5	8.2	2.2	11.4	2.9	10.1	1.7	11.0	2.0
Q	P	177	67	8.7	3.7	12.1	7.1	12.3	3.1	9.5	1.5	13.5	1.8
H	P	103	49	10.4	8.1	8.7	6.4	15.8	3.4	13.2	2.4	14.9	2.4
K	P	73	5	10.5	4.1	10.3	3.2	17.1	6.8	14.4	2.5	15.2	3.4
R	P	127	20	6.3	4.5	4.4	0.7	5.3	1.9	5.3	0.5	6.1	1.1
G	F	127	56	33.5	15.8	30.7	2.5	34.4	8.8	48.0	16.5	36.4	7.2
A	F	108	44	51.3	9.4	53.5	11.2	61.7	10.6	69.3	7.4	69.3	7.2
S	F	122	11	37.1	11.6	33.3	4.6	40.3	9.4	50.3	10.5	45.4	8.0
T	F	113	2	53.9	8.5	59.0	9.3	69.8	10.3	75.3	7.2	77.6	3.8
V	F	113	18	75.9	5.1	77.3	5.0	86.2	4.1	87.5	1.6	89.5	1.0

L	F	101	5	69.7	5.8	67.2	5.9	81.2	3.9	79.8	2.0	82.6	2.0
I	F	101	0	73.4	1.6	75.2	3.8	84.9	3.1	83.7	1.9	86.4	2.5
M	F	35	14	76.9	4.2	80.4	4.0	81.5	5.7	84.1	1.9	84.8	7.0
P	F	115	7	61.6	3.9	62.6	8.3	73.5	2.8	78.3	3.4	79.5	2.6
F	F	70	22	56.3	1.1	54.5	5.3	65.9	9.4	60.1	5.0	61.7	5.3
Y	F	91	2	58.8	1.6	57.1	4.9	68.1	6.7	62.9	2.5	65.6	5.3
W	F	13	3	55.1	2.5	49.0	6.5	57.0	6.5	54.9	3.5	53.2	3.2
D	F	76	14	35.8	10.8	21.7	4.2	32.6	5.9	46.1	8.4	31.0	8.5
E	F	84	3	44.5	8.6	32.5	4.7	46.5	9.5	60.3	13.0	46.8	9.9
N	F	115	21	37.3	8.1	29.8	2.2	40.9	10.7	50.6	9.9	44.3	7.9
Q	F	81	16	48.8	10.7	53.3	10.3	66.9	11.5	73.4	10.2	74.6	6.8
H	F	85	37	53.4	8.7	48.1	6.6	54.5	2.6	59.6	3.9	59.2	3.4
K	F	90	40	75.3	8.8	75.2	4.9	71.2	1.6	76.5	4.8	73.5	5.6
R	F	117	17	52.5	5.1	43.8	9.7	39.4	1.7	47.4	5.7	41.7	5.2
G	Y	106	57	34.2	9.4	24.3	2.5	33.4	4.7	39.3	8.4	34.1	5.7
A	Y	97	27	49.4	9.0	38.9	14.5	57.8	9.6	63.2	11.1	63.4	4.7
S	Y	154	1	35.9	13.0	28.3	10.0	34.2	13.5	43.9	9.9	36.0	9.2
T	Y	143	23	58.9	9.2	50.5	8.2	69.2	8.9	73.2	5.6	74.2	1.9
V	Y	160	6	77.8	3.6	77.2	6.0	87.3	1.8	85.5	2.7	88.3	1.6
L	Y	106	6	75.2	3.8	74.3	8.7	85.8	2.4	82.6	1.6	85.9	3.0
I	Y	147	41	80.4	2.3	79.7	6.3	88.6	2.2	86.4	1.3	89.3	0.6
M	Y	37	1	72.3	2.3	76.3	5.4	78.0	5.5	82.3	1.4	83.7	3.4
P	Y	113	29	61.0	6.2	58.8	8.5	71.3	8.2	74.4	3.5	76.9	2.4
F	Y	129	44	70.7	3.1	66.4	5.2	80.0	3.6	76.1	1.6	78.8	1.7
Y	Y	91	3	64.9	1.9	60.2	2.9	77.2	4.2	70.7	2.2	74.1	4.0
W	Y	14	5	59.2	4.1	54.3	4.4	60.5	4.4	59.6	2.3	59.4	0.3

D	Y	101	30	36.3	5.6	22.9	4.5	29.7	6.5	43.6	9.2	31.9	7.1
E	Y	94	17	48.6	8.5	36.3	7.6	51.8	8.2	61.0	12.0	55.3	11.3
N	Y	93	24	41.3	10.6	28.8	5.7	36.3	8.0	43.3	8.2	40.1	5.6
Q	Y	89	15	58.3	8.3	57.7	6.9	65.4	8.1	74.1	4.1	73.7	2.6
H	Y	115	18	49.9	4.2	41.4	9.4	51.3	2.3	55.8	1.9	54.9	2.0
K	Y	67	36	72.4	10.0	70.5	6.0	69.3	1.5	74.3	5.7	73.6	4.2
R	Y	94	4	51.4	4.6	43.6	8.8	39.8	1.5	47.1	5.2	41.7	3.4
G	W	8	1	25.8	4.1	26.5	6.5	36.5	5.1	37.9	5.3	33.1	4.9
A	W	9	3	38.2	5.9	39.9	5.1	51.4	9.8	56.5	6.2	54.0	7.1
S	W	8	3	35.9	6.5	32.6	5.8	36.9	7.4	44.0	6.8	36.0	9.2
T	W	9	3	48.2	8.1	50.0	8.5	51.1	9.4	56.2	11.6	53.8	11.3
V	W	12	3	62.6	5.3	60.0	6.4	69.8	7.0	70.6	5.2	73.3	4.4
L	W	16	7	52.0	6.4	48.4	10.9	61.3	8.3	60.4	6.5	62.0	3.8
I	W	15	4	56.7	3.2	54.6	11.3	68.5	7.3	66.5	7.5	70.6	2.9
M	W	7	2	55.6	8.9	58.6	8.9	56.6	10.2	63.4	1.9	59.7	1.5
P	W	10	4	52.2	10.4	54.8	8.5	58.7	10.1	63.9	5.0	62.9	6.4
F	W	13	6	61.8	3.3	58.0	5.7	58.9	2.8	59.2	1.5	60.2	1.5
Y	W	13	9	62.5	4.3	61.3	8.5	65.9	2.1	60.0	3.4	63.9	3.1
W	W	8	3	47.0	5.0	41.5	5.1	47.0	1.7	47.7	4.3	48.9	3.8
D	W	8	2	43.3	8.3	38.3	7.2	36.5	7.2	43.9	6.2	35.3	2.3
E	W	11	7	53.9	6.8	48.3	8.0	50.4	8.1	55.9	11.9	48.7	10.7
N	W	10	6	38.9	5.6	38.0	5.3	35.4	5.1	40.6	7.1	36.1	4.0
Q	W	10	5	51.5	9.8	51.5	9.9	54.8	7.3	58.8	6.2	55.9	7.1
H	W	12	6	51.8	7.6	47.6	5.2	47.2	4.3	46.8	2.1	47.4	1.1
K	W	11	5	57.2	6.6	59.5	8.0	61.5	1.6	62.4	5.6	58.8	4.7
R	W	13	7	40.2	7.8	36.6	4.8	30.1	4.2	32.5	6.1	30.7	4.6

G	D	113	6	12.9	6.0	12.1	3.7	13.3	1.7	14.4	0.9	13.6	1.6
A	D	113	21	17.6	8.5	14.1	3.6	16.8	3.4	18.4	3.5	18.0	1.6
S	D	122	13	12.7	6.3	9.3	3.0	12.4	3.0	12.9	2.6	12.1	0.9
T	D	112	19	16.2	8.4	12.8	3.8	15.3	2.7	17.9	3.5	15.4	1.2
V	D	110	1	16.9	8.2	14.0	3.2	19.9	3.5	21.5	4.6	19.3	3.7
L	D	81	1	15.6	7.0	11.1	1.5	18.1	4.2	20.7	2.7	20.2	0.7
I	D	100	5	14.2	5.9	10.9	2.3	19.8	4.0	22.6	5.0	22.2	2.4
M	D	30	10	27.5	5.2	28.2	3.9	35.3	3.5	34.1	2.0	35.8	2.0
P	D	99	8	15.1	6.6	12.5	4.4	16.4	2.7	18.0	4.6	16.8	2.8
F	D	102	8	22.0	8.5	15.3	2.4	21.8	4.5	26.8	7.0	24.6	5.7
Y	D	144	27	21.7	7.5	13.2	2.6	20.4	3.7	27.2	6.7	23.4	4.8
W	D	12	4	40.2	5.3	35.0	9.6	35.6	4.8	37.1	6.3	34.6	5.9
D	D	107	53	13.8	12.0	9.4	10.0	11.1	5.7	16.4	4.6	13.0	2.1
E	D	110	35	15.8	7.2	9.7	4.6	13.0	2.6	16.3	0.9	15.1	2.2
N	D	129	9	10.9	5.3	7.5	2.1	11.9	2.6	12.7	0.8	13.4	0.5
Q	D	112	19	15.8	6.2	11.3	2.7	14.9	2.6	17.0	2.2	16.5	2.5
H	D	99	16	13.6	7.9	8.5	2.2	15.5	4.5	17.9	4.5	16.5	2.1
K	D	92	18	15.8	9.3	11.3	2.4	20.5	9.1	19.4	6.8	18.2	8.3
R	D	96	6	14.0	4.8	12.8	4.8	9.1	3.2	9.8	2.9	9.2	1.8
G	E	98	9	13.3	6.9	11.4	4.2	13.8	0.8	12.7	0.8	14.9	0.8
A	E	90	13	17.3	8.3	14.4	2.7	20.0	3.7	18.4	2.2	19.8	1.7
S	E	128	41	12.8	8.0	9.7	3.0	13.1	1.7	12.2	1.0	12.9	1.8
T	E	98	10	10.9	5.6	7.5	2.9	14.9	4.6	15.6	2.8	16.3	2.2
V	E	105	10	13.4	5.4	13.3	2.2	19.7	1.7	21.0	3.0	20.2	3.4
L	E	98	20	19.1	7.9	15.6	3.4	21.0	3.9	22.1	3.4	21.1	3.8
I	E	96	29	18.4	6.8	15.3	3.6	20.7	3.6	24.7	3.7	24.8	4.1

M	E	32	5	30.9	4.9	28.5	2.5	36.0	2.5	34.7	1.9	36.0	1.9
P	E	91	16	14.8	6.0	11.1	2.6	17.7	1.5	20.4	3.5	19.0	1.7
F	E	112	36	21.6	6.7	17.9	3.2	26.1	5.2	31.6	7.6	29.4	6.7
Y	E	81	8	25.9	5.3	20.9	2.3	26.7	5.5	32.1	7.3	30.9	5.5
W	E	11	4	41.9	10.6	38.3	5.8	37.9	4.6	39.6	4.9	39.6	6.6
D	E	92	1	8.7	4.7	9.2	3.9	12.2	2.2	10.9	1.5	12.2	1.0
E	E	85	9	11.8	5.8	13.2	4.7	14.1	2.6	13.8	1.6	14.7	3.3
N	E	120	0	10.6	5.0	10.2	2.9	15.1	3.2	21.6	14.8	14.4	3.3
Q	E	131	40	12.3	4.2	12.0	3.0	17.7	3.4	20.0	7.4	18.2	2.1
H	E	109	28	13.2	3.5	11.9	4.0	20.0	3.9	16.6	3.0	17.9	2.7
K	E	70	3	18.2	10.3	14.4	5.2	23.7	16.4	21.5	10.3	21.7	12.9
R	E	124	16	16.4	5.7	18.5	11.9	13.6	6.7	13.7	4.4	11.6	3.9
G	N	134	16	16.8	8.3	13.0	3.9	15.4	4.7	14.8	3.4	17.4	5.7
A	N	129	46	23.1	13.4	17.4	4.4	22.1	6.8	23.1	7.8	22.8	6.2
S	N	115	14	16.0	8.6	9.5	2.6	15.4	4.5	16.7	5.1	14.5	4.0
T	N	138	17	19.7	10.7	12.6	2.5	22.6	7.0	22.7	8.9	21.7	6.7
V	N	126	32	29.7	16.0	20.0	3.5	29.9	11.7	34.8	14.4	34.4	11.0
L	N	111	18	26.4	15.2	16.8	2.9	28.7	11.7	31.8	12.5	31.5	9.8
I	N	130	41	29.6	13.6	21.9	3.5	33.4	11.8	35.7	10.7	38.0	11.5
M	N	30	1	34.5	8.2	30.2	3.1	38.9	6.2	37.4	6.4	39.3	3.9
P	N	143	4	22.9	11.7	15.7	3.3	24.1	9.0	28.7	10.2	27.3	7.7
F	N	123	2	32.5	12.3	26.0	2.5	34.7	11.2	40.5	11.9	37.8	9.1
Y	N	93	2	32.0	9.9	21.3	2.3	32.8	9.2	39.8	7.7	37.8	8.2
W	N	12	3	37.6	8.1	29.8	3.6	35.5	8.7	40.1	5.3	38.8	6.8
D	N	109	11	9.6	4.9	6.5	2.0	11.6	2.8	13.4	3.6	12.4	3.4
E	N	94	19	16.5	7.8	11.6	3.6	17.8	5.6	21.9	7.2	19.4	6.7

N	N	123	10	13.5	7.7	7.7	1.8	16.0	6.5	14.6	2.2	16.0	4.1
Q	N	138	18	17.1	7.0	13.5	1.1	22.0	6.5	24.1	9.2	23.0	8.4
H	N	115	5	15.9	6.5	13.0	5.2	25.0	9.6	24.3	8.8	25.5	9.1
K	N	114	34	17.6	6.0	17.6	3.0	38.8	17.6	41.0	10.9	38.4	13.3
R	N	192	32	19.7	4.3	20.7	3.9	18.2	7.0	20.7	4.5	18.1	4.6
G	Q	129	56	11.1	4.1	12.3	3.5	16.7	5.9	14.2	3.4	16.2	2.9
A	Q	185	5	16.6	6.7	16.5	3.0	21.5	8.7	23.9	11.0	25.0	8.2
S	Q	178	3	11.3	4.2	12.8	5.6	15.7	5.0	14.5	4.7	15.0	3.8
T	Q	126	3	23.2	15.9	15.7	7.7	17.8	6.8	21.5	8.5	18.3	6.0
V	Q	148	31	30.3	17.9	20.4	9.0	29.0	9.6	36.9	14.3	31.9	9.0
L	Q	115	10	31.6	21.6	22.0	7.9	26.7	8.4	36.6	13.7	32.2	10.3
I	Q	98	4	40.6	23.8	36.3	15.8	44.6	9.4	51.5	13.3	46.6	10.1
M	Q	26	3	12.0	5.0	8.3	2.8	20.1	4.4	40.5	37.9	19.0	4.7
P	Q	92	2	33.5	21.3	30.2	19.8	31.5	8.6	39.0	11.7	35.9	7.8
F	Q	137	25	43.3	22.6	40.7	13.9	44.7	15.5	57.3	15.9	53.9	12.9
Y	Q	120	35	44.0	16.4	35.6	9.3	44.0	11.2	54.0	12.4	50.6	8.4
W	Q	12	6	17.6	3.4	18.0	5.2	18.5	5.7	28.2	14.5	22.9	7.5
D	Q	35	7	31.6	19.0	24.0	4.5	39.8	8.7	45.2	7.2	40.7	4.0
E	Q	91	18	18.7	13.6	10.5	5.9	16.6	6.3	22.2	8.7	20.3	5.6
N	Q	31	0	35.6	12.9	34.9	7.5	39.0	6.1	38.2	1.8	40.4	3.8
Q	Q	87	20	29.9	12.8	23.0	3.0	32.0	8.8	33.6	8.4	32.9	6.3
H	Q	52	0	44.7	16.8	46.0	14.7	56.4	6.8	57.1	1.7	57.2	1.1
K	Q	106	4	23.1	11.6	21.4	5.1	46.5	16.1	48.6	8.8	45.7	16.4
R	Q	104	12	23.8	7.9	23.4	5.9	21.3	6.4	24.2	4.2	20.8	4.9
G	H	73	4	26.3	11.0	19.4	5.1	28.8	3.9	27.5	4.3	29.7	5.9
A	H	61	1	33.2	15.0	35.3	4.6	43.6	5.2	43.8	7.5	47.7	2.7

S	H	81	8	22.9	11.3	18.4	3.7	24.9	5.4	24.9	4.3	24.7	7.0
T	H	123	23	26.8	9.9	27.2	7.3	36.2	5.6	43.1	5.7	38.3	1.4
V	H	167	28	52.9	9.6	57.9	6.1	57.6	11.2	68.7	4.1	57.4	16.9
L	H	16	0	51.0	5.7	41.7	7.9	72.2	10.7	70.6	2.3	75.3	2.6
I	H	83	7	48.7	5.3	52.5	7.1	56.4	7.4	61.1	3.2	63.5	7.9
M	H	20	1	26.5	5.1	22.6	8.5	21.7	5.6	22.5	2.7	22.8	3.9
P	H	80	27	29.5	20.2	18.9	8.3	31.4	9.8	32.4	5.2	34.6	10.9
F	H	28	3	47.2	13.1	47.9	9.3	42.0	5.6	44.9	4.3	41.0	1.8
Y	H	80	17	53.4	6.8	51.5	10.9	47.4	9.1	51.4	2.6	49.9	6.9
W	H	9	4	23.5	6.1	27.5	8.2	15.9	1.6	21.4	5.5	17.2	1.0
D	H	6	0	46.9	17.9	50.2	8.7	68.2	11.6	66.9	7.2	67.8	7.4
E	H	11	1	40.2	15.8	36.9	15.2	39.5	7.3	42.0	7.0	45.9	10.5
N	H	60	3	28.9	7.4	29.5	6.0	30.5	8.4	32.5	3.5	31.6	2.5
Q	H	73	10	33.0	13.9	32.6	8.9	34.2	9.5	40.1	2.2	33.7	4.6
H	H	57	4	57.4	6.5	48.2	12.9	43.7	3.6	50.0	2.9	46.0	7.7
K	H	9	0	37.0	5.6	35.0	3.9	46.5	5.4	49.8	3.8	51.1	6.6
R	H	53	8	37.8	5.7	35.2	7.3	26.2	1.5	29.0	3.4	27.1	3.0
G	K	69	9	25.5	15.0	18.8	6.2	21.6	6.8	20.9	4.4	21.9	2.5
A	K	46	0	30.3	13.4	31.2	12.5	36.9	6.5	36.5	4.0	36.1	5.4
S	K	63	15	21.9	13.7	17.9	7.7	20.7	6.1	22.1	4.7	21.2	2.4
T	K	93	14	19.1	11.9	19.0	8.0	23.4	8.2	24.5	6.9	23.6	7.3
V	K	59	13	37.1	14.5	43.5	12.1	45.5	6.8	48.0	3.3	50.3	6.7
L	K	88	1	38.0	15.5	36.7	9.2	35.9	9.1	39.3	6.7	37.5	8.5
I	K	95	0	39.4	14.5	34.9	8.5	40.7	8.7	42.3	9.7	42.2	7.0
M	K	18	7	12.7	6.5	10.6	1.5	16.0	6.0	16.8	6.1	16.2	4.9
P	K	89	7	33.7	10.2	29.5	4.1	32.8	9.0	34.4	7.1	31.2	4.8

F	K	141	14	51.4	5.5	52.1	3.9	49.7	8.3	52.5	4.0	52.7	8.3
Y	K	120	6	51.5	3.0	49.1	4.1	45.3	4.4	49.7	3.0	49.5	7.5
W	K	7	1	64.0	3.0	61.2	1.4	64.7	0.6	62.1	1.9	64.8	2.9
D	K	72	2	35.4	5.8	31.8	3.4	39.4	6.0	39.0	6.2	38.9	2.4
E	K	42	10	45.8	4.8	52.4	9.8	45.2	5.6	48.3	3.8	44.4	4.1
N	K	52	1	19.7	14.6	19.1	6.8	16.5	5.5	20.0	3.8	18.1	0.9
Q	K	107	33	26.8	20.2	34.4	15.1	32.8	9.2	33.9	6.0	28.4	2.5
H	K	66	3	34.6	7.2	41.6	8.8	31.6	6.8	34.0	4.3	33.6	1.1
K	K	77	27	42.8	8.5	51.4	5.1	39.3	5.2	43.3	13.0	38.7	2.1
R	K	118	53	30.4	5.7	33.5	12.0	14.3	0.7	20.3	10.8	15.3	3.7
G	R	85	23	17.2	5.1	15.7	4.9	12.6	2.6	13.2	0.9	12.8	1.5
A	R	81	7	24.6	4.0	31.8	5.4	25.0	4.6	29.7	4.1	26.1	1.2
S	R	35	14	17.9	3.0	20.7	5.8	13.8	6.1	17.9	2.9	15.9	0.7
T	R	127	24	23.5	3.6	26.9	4.4	20.6	4.7	25.3	2.5	20.0	3.6
V	R	97	7	24.5	7.5	31.4	5.3	30.4	6.0	27.0	2.8	28.7	2.5
L	R	6	0	77.4	2.0	76.2	7.4	45.9	10.2	59.5	4.9	41.2	11.6
I	R	85	0	57.2	6.7	54.1	6.8	37.4	2.8	47.8	8.0	42.3	4.5
M	R	16	3	29.2	10.0	25.3	6.8	19.8	2.6	21.1	1.3	20.6	1.9
P	R	130	1	37.7	5.5	42.4	8.2	28.7	1.6	35.8	3.4	33.3	3.9
F	R	133	44	65.3	1.8	62.7	7.5	48.1	8.4	55.8	10.9	53.8	12.9
Y	R	107	39	60.2	3.4	52.6	4.8	46.0	5.6	51.3	7.6	49.0	8.0
W	R	7	1	16.6	7.6	12.6	3.5	12.6	1.7	18.9	3.0	18.5	4.8
D	R	90	30	18.1	10.6	15.9	5.9	11.2	2.5	20.1	12.0	10.8	2.8
E	R	99	7	23.8	12.3	24.3	7.5	24.0	4.3	30.5	8.8	23.9	6.0
N	R	117	22	18.5	13.0	16.1	7.4	12.1	2.9	12.6	2.7	10.3	2.3
Q	R	123	2	32.1	6.4	34.4	7.1	25.1	5.1	31.1	3.1	26.0	1.6

H	R	44	10	33.0	7.3	31.8	5.4	24.3	3.6	26.5	0.8	24.2	3.2
K	R	71	14	58.1	12.1	54.0	3.2	30.1	7.2	42.0	14.7	35.7	3.8
R	R	51	9	40.8	4.5	42.0	7.4	21.0	5.6	28.5	10.4	21.9	8.3

Table S-2. T^p Array Data Summary

Peptide Sequence: Ac-GXT^{p}ZGRC		**Peptide Signal-to-Noise**		**% Dephosphorylation: NIH/3T3**		**% Dephosphorylation: HT-1080**		**% Dephosphorylation: MCF-7**		**% Dephosphorylation: BT474**		**% Dephosphorylation: MDA-MB-231**	
X position	**Z position**	**Average**	**Std. Dev.**	**Average**	**Std. Dev.**	**Average**	**Std. Dev.**	**Average**	**Std. Dev.**	**Average**	**Std. Dev.**	**Average**	**Std. Dev.**
G	G	140	5	25.1	10.0	20.1	5.7	20.0	7.1	26.8	4.2	23.3	5.8
A	G	142	30	35.0	7.9	37.9	5.3	44.8	9.4	61.6	12.7	47.5	4.7
S	G	192	0	20.5	5.9	18.5	4.2	23.0	6.6	31.8	8.6	27.6	4.6
T	G	192	19	28.3	7.0	30.3	5.1	37.3	8.8	51.9	10.5	42.9	9.4
V	G	185	22	56.7	5.4	65.4	6.9	78.4	9.7	85.2	5.4	75.6	8.2
L	G	142	31	50.0	8.6	55.2	8.0	72.2	8.5	78.4	6.1	65.5	9.6
I	G	167	16	71.3	7.6	74.2	10.4	87.1	6.3	89.5	2.5	80.4	5.4
M	G	110	16	49.1	6.3	54.3	7.8	68.3	6.3	76.1	6.9	63.5	6.5
P	G	163	26	50.4	8.1	59.3	11.6	72.4	6.8	80.1	4.0	71.0	8.3
F	G	170	13	78.7	6.1	81.3	2.8	88.3	2.3	87.7	1.3	86.0	4.3
Y	G	137	34	72.1	7.4	74.1	2.2	82.6	2.0	81.2	2.3	76.4	2.6
W	G	80	11	61.2	7.8	64.3	5.2	76.2	1.5	73.6	3.8	71.9	4.7
D	G	107	3	17.5	3.5	18.1	4.4	19.1	4.2	27.1	4.7	24.9	7.3
E	G	94	21	27.7	5.0	26.9	5.2	38.6	7.2	49.3	6.6	45.1	4.7
N	G	161	15	14.5	1.9	15.4	4.0	18.7	5.0	24.0	3.9	22.2	5.0
Q	G	142	26	26.5	4.3	31.8	7.9	44.8	5.0	56.5	8.9	47.6	8.2
H	G	133	48	28.8	3.3	36.0	9.9	41.9	6.8	50.5	5.1	42.2	10.0
K	G	131	17	66.1	12.0	76.8	3.5	70.9	3.9	75.9	5.7	67.3	3.1
R	G	152	11	68.9	4.3	62.5	4.5	53.7	4.4	62.7	7.8	49.8	3.1

G	A	195	8	26.8	4.3	31.2	6.0	32.6	6.0	44.8	10.8	41.8	12.3
A	A	161	27	46.4	6.1	57.8	6.6	71.4	5.6	81.2	7.3	77.7	7.2
S	A	150	14	29.0	3.4	34.9	4.4	41.3	10.1	55.7	8.9	53.8	12.4
T	A	134	32	43.6	6.0	53.1	6.2	63.4	8.0	72.3	10.3	72.1	9.6
V	A	154	24	75.0	7.5	84.1	5.0	93.3	2.8	93.3	1.9	92.1	3.3
L	A	110	31	73.6	7.8	81.2	7.1	90.1	5.3	92.8	2.0	87.3	6.8
I	A	119	7	87.7	3.7	91.2	2.3	95.9	1.4	96.3	0.4	95.5	1.7
M	A	93	33	64.7	3.3	75.2	4.1	86.3	4.4	89.1	4.3	85.0	6.6
P	A	151	13	73.8	3.3	82.0	4.2	90.9	2.7	94.6	1.4	91.7	3.6
F	A	129	1	94.3	1.8	94.8	2.1	97.0	0.7	96.4	1.1	94.0	1.5
Y	A	92	3	87.7	2.0	88.3	3.3	94.1	1.9	93.2	1.1	92.2	1.9
W	A	52	16	81.2	1.3	82.1	1.8	86.8	2.5	85.2	0.8	86.4	1.6
D	A	89	15	28.9	5.9	24.9	2.2	24.9	5.2	37.1	7.4	37.4	11.2
E	A	83	13	40.7	6.7	39.1	3.1	49.0	10.6	66.2	8.2	61.7	8.4
N	A	122	0	29.3	5.0	28.5	3.3	29.1	9.3	44.1	8.2	41.5	10.7
Q	A	86	16	47.7	4.6	56.9	6.1	63.4	7.5	80.8	7.1	73.9	6.9
H	A	111	3	50.6	5.7	64.0	4.9	67.3	3.8	75.8	1.9	71.3	6.4
K	A	55	5	77.2	6.5	82.3	4.5	89.0	2.4	90.6	0.9	88.8	1.0
R	A	103	6	79.3	4.0	74.9	3.2	66.5	2.1	73.6	7.7	67.5	4.0
G	S	106	16	29.6	3.5	33.7	5.3	37.9	7.5	52.7	6.0	48.4	8.9
A	S	143	13	56.6	6.2	68.2	7.5	80.3	1.8	88.5	3.8	84.3	6.2
S	S	130	40	34.7	4.7	43.2	6.8	52.7	4.9	67.9	7.0	63.8	7.4
T	S	112	10	49.8	4.7	59.5	11.0	72.8	3.9	83.0	5.6	79.7	7.0
V	S	95	21	83.0	5.1	88.4	4.3	95.3	1.4	95.9	1.1	94.1	1.5
L	S	95	8	83.1	6.2	86.9	3.8	94.9	1.5	94.6	1.1	93.1	1.5
I	S	61	21	74.4	2.6	73.6	1.4	80.5	1.7	81.3	2.2	77.7	3.6

M	S	85	32	66.6	6.1	79.8	5.5	92.0	1.9	92.4	2.3	89.3	2.3
P	S	114	11	74.9	5.7	84.6	5.2	95.2	1.0	94.8	1.4	93.7	2.3
F	S	128	9	93.1	1.9	93.5	1.5	97.0	1.1	95.8	0.4	95.7	1.3
Y	S	87	6	89.1	2.5	90.2	1.8	94.5	1.2	94.5	0.7	94.2	0.6
W	S	46	1	83.9	1.7	84.9	0.7	88.1	1.9	87.8	1.3	88.5	1.7
D	S	66	12	32.7	9.9	30.6	6.1	32.3	6.5	47.6	9.5	47.4	6.5
E	S	71	6	50.5	9.8	51.8	8.1	62.0	5.5	77.5	6.5	76.3	3.7
N	S	114	2	33.1	9.0	33.7	5.4	38.5	8.1	54.4	10.1	50.1	10.3
Q	S	92	8	50.4	5.6	60.0	3.9	71.8	11.0	85.7	4.9	79.0	6.8
H	S	47	0	67.1	3.4	69.8	1.8	72.7	0.6	80.4	2.9	76.5	0.9
K	S	61	5	85.0	2.8	88.1	2.7	84.2	1.6	88.0	5.6	87.0	2.3
R	S	114	18	85.8	2.0	81.7	3.7	68.8	2.6	76.9	5.8	72.1	3.8
G	T	99	3	43.2	14.1	43.7	6.9	50.0	6.7	67.9	7.4	58.9	9.1
A	T	105	8	69.5	6.7	78.1	4.0	89.2	1.6	92.5	2.5	90.2	4.6
S	T	114	20	45.1	5.7	53.4	3.8	64.3	6.8	77.8	7.4	71.2	9.9
T	T	100	17	45.6	7.2	53.0	8.5	55.6	6.5	76.0	4.4	72.1	5.8
V	T	63	15	87.9	0.8	92.5	2.0	96.2	0.1	95.8	1.1	95.9	1.8
L	T	77	13	87.3	0.9	89.8	2.3	95.6	1.0	95.1	0.8	94.7	1.4
I	T	94	6	87.6	3.8	91.3	2.7	96.4	0.3	96.0	0.7	95.7	1.4
M	T	65	10	80.4	5.5	86.2	4.5	94.9	0.7	93.5	1.6	91.3	3.0
P	T	87	1	82.0	4.6	88.6	5.7	96.4	0.4	96.2	0.4	95.5	2.0
F	T	85	2	92.0	2.1	93.4	0.9	96.2	1.1	95.3	0.6	95.2	0.2
Y	T	52	4	91.1	1.8	91.7	1.3	96.1	0.5	94.7	0.6	94.7	0.6
W	T	39	2	85.6	2.4	86.0	4.0	88.3	2.7	87.1	0.4	88.5	1.5
D	T	55	5	31.3	5.7	31.0	4.7	43.8	6.4	59.1	10.4	56.3	8.7
E	T	69	2	49.7	6.4	49.5	6.9	66.2	4.8	77.7	5.8	78.3	3.3

N	T	87	6	45.1	15.8	41.0	7.2	54.3	7.4	66.3	9.7	65.1	9.5
Q	T	90	12	65.6	9.0	71.7	9.4	78.4	8.3	90.5	3.6	86.8	3.8
H	T	106	24	77.9	4.0	82.3	4.1	89.1	2.7	92.3	0.9	90.7	2.4
K	T	80	1	84.1	6.5	88.6	4.8	83.8	3.2	89.7	5.5	88.1	3.6
R	T	104	33	88.6	0.6	86.0	3.3	73.8	2.7	82.0	7.9	78.3	4.1
G	V	85	16	70.4	7.3	79.9	8.1	84.9	4.6	90.3	2.3	87.7	6.4
A	V	107	2	93.0	1.3	95.1	1.4	97.4	1.0	97.9	0.1	97.6	0.5
S	V	104	12	57.2	4.6	67.0	8.2	67.1	7.8	82.9	7.0	81.0	4.7
T	V	87	3	86.0	3.5	74.2	34.4	95.4	1.4	96.5	1.0	95.9	2.2
V	V	91	21	96.0	0.9	95.9	1.0	98.1	0.5	97.9	0.3	97.9	0.7
L	V	110	14	92.2	5.4	95.2	3.6	97.7	0.3	97.3	0.7	97.5	1.0
I	V	68	16	88.1	12.8	94.5	2.6	97.7	0.3	97.1	0.6	97.2	0.8
M	V	57	4	91.9	0.2	90.4	2.9	96.4	0.9	96.0	1.3	95.8	0.8
P	V	87	15	94.2	2.3	94.7	0.7	98.0	0.5	97.3	0.7	97.5	0.8
F	V	75	4	95.3	1.1	95.8	0.5	96.7	0.6	95.6	0.3	96.2	0.3
Y	V	72	6	93.6	1.0	93.0	1.3	95.6	0.7	94.8	0.9	95.2	0.4
W	V	43	1	89.6	1.7	89.8	1.5	88.6	1.5	88.1	1.3	89.9	2.1
D	V	72	19	44.6	2.6	47.1	5.4	61.7	8.6	73.6	7.2	69.8	6.9
E	V	68	10	67.8	6.0	72.9	5.2	80.6	2.9	90.3	2.9	88.7	2.6
N	V	90	27	60.5	8.1	75.3	5.7	86.8	2.8	90.3	3.7	89.9	6.4
Q	V	65	3	87.9	3.5	93.4	3.5	97.4	0.6	97.7	0.4	97.4	0.9
H	V	55	1	91.5	2.1	94.0	1.4	96.5	0.5	96.2	0.4	96.4	0.9
K	V	52	3	94.1	2.4	95.3	1.5	91.4	1.8	93.3	3.1	93.2	3.4
R	V	88	6	93.9	2.0	92.1	1.4	81.6	1.7	86.9	6.3	84.9	6.9
G	L	65	14	81.5	3.9	87.1	3.7	94.0	1.8	93.5	1.9	92.0	2.3
A	L	67	29	93.8	1.5	79.8	32.2	97.5	0.8	97.9	0.2	97.8	0.4

S	L	71	31	81.1	4.9	87.3	2.6	93.4	2.0	95.2	1.5	94.0	2.3
T	L	58	27	88.0	6.5	93.2	1.3	96.7	1.9	97.0	0.2	97.0	0.5
V	L	75	38	95.8	2.5	96.6	1.0	98.4	0.6	98.1	0.8	98.5	0.4
L	L	77	12	96.0	0.7	97.2	0.4	98.4	0.4	98.1	0.6	98.2	0.4
I	L	62	26	96.4	0.9	97.5	0.8	98.6	0.2	98.3	0.4	98.1	0.3
M	L	44	12	92.9	0.7	92.8	3.0	97.0	0.9	97.3	0.6	96.1	0.4
P	L	63	16	95.3	1.9	97.0	0.7	98.6	0.2	98.5	0.5	98.2	0.6
F	L	57	14	92.9	0.4	92.9	0.6	95.0	0.7	94.2	0.9	94.8	1.0
Y	L	23	4	86.8	1.7	87.3	2.7	91.5	1.3	89.8	1.8	92.2	1.0
W	L	31	16	83.2	2.4	83.1	4.6	82.2	2.4	82.0	1.6	84.6	1.4
D	L	54	19	52.3	4.2	55.6	3.2	64.7	4.8	79.1	6.9	72.2	9.0
E	L	36	19	82.0	4.4	87.6	4.3	94.0	0.2	95.6	2.7	94.3	2.6
N	L	59	27	67.4	8.7	80.4	4.9	91.1	1.4	93.1	3.8	90.7	5.6
Q	L	72	40	93.0	2.0	96.2	1.1	97.9	0.2	97.9	0.8	97.2	1.2
H	L	74	43	93.1	1.5	94.5	0.8	96.1	1.0	96.6	0.4	95.3	1.4
K	L	52	25	95.8	2.4	96.9	0.6	89.1	4.1	92.5	4.7	91.5	4.0
R	L	83	60	91.3	3.1	91.0	2.0	82.1	2.8	87.8	5.6	85.2	4.9
G	I	71	45	67.5	8.8	83.4	4.9	89.8	3.0	92.5	3.0	89.4	5.7
A	I	68	52	92.3	3.4	95.5	1.4	98.5	0.2	98.1	0.2	97.6	0.7
S	I	62	45	76.2	9.3	87.3	4.3	94.5	2.2	95.3	1.5	94.3	2.2
T	I	52	14	90.4	4.1	92.6	2.2	96.4	0.7	94.9	0.5	96.6	0.9
V	I	60	40	94.6	4.2	95.5	1.6	98.4	0.2	97.0	0.3	97.9	0.6
L	I	74	24	96.7	0.9	97.1	0.6	98.5	0.2	97.7	0.5	97.5	0.8
I	I	62	30	96.1	1.6	97.1	0.6	98.2	0.3	97.6	0.1	97.7	0.3
M	I	36	19	92.2	2.8	95.2	0.8	97.2	0.9	96.6	0.8	96.5	0.8
P	I	56	30	95.8	1.2	97.0	0.8	98.7	0.4	98.6	0.3	98.4	0.1

F	I	49	18	94.2	0.6	94.9	0.9	96.7	0.4	96.0	1.0	96.7	0.7
Y	I	46	18	92.6	2.8	92.8	0.7	95.7	0.4	94.9	1.8	95.9	1.1
W	I	30	14	87.8	1.9	88.0	2.3	86.8	2.3	86.3	2.1	88.8	0.8
D	I	42	8	49.6	4.1	51.5	5.3	56.4	4.3	72.4	7.3	68.3	6.8
E	I	56	37	73.7	3.7	77.8	4.4	84.5	3.0	92.5	3.7	90.7	3.9
N	I	70	35	72.4	7.5	82.5	3.9	91.5	2.1	93.4	3.7	91.7	5.5
Q	I	59	27	93.2	1.5	96.1	1.2	97.9	0.1	97.7	1.3	97.5	1.0
H	I	49	22	90.2	6.7	94.7	1.0	95.4	0.9	95.9	0.6	96.0	0.8
K	I	53	27	95.1	2.1	96.2	0.3	88.2	5.2	91.9	4.3	91.2	4.8
R	I	65	43	92.3	1.4	91.1	1.4	80.8	5.7	87.6	3.2	84.4	6.0
G	M	37	23	73.6	4.2	81.5	3.5	89.2	5.4	90.5	1.1	89.0	0.9
A	M	43	29	89.5	3.5	92.9	3.0	96.0	2.3	96.2	1.1	96.0	0.4
S	M	42	15	83.9	5.4	89.7	4.2	94.5	1.7	95.8	0.1	94.4	1.3
T	M	37	17	88.4	4.3	94.5	1.9	97.7	1.5	97.5	0.7	96.6	0.7
V	M	52	19	88.2	10.0	93.5	2.1	97.4	1.1	97.4	0.6	96.4	0.7
L	M	46	29	93.3	2.8	94.8	1.2	97.8	0.5	96.9	1.2	96.2	1.8
I	M	45	36	94.8	1.6	94.5	2.7	98.1	0.4	97.5	0.8	97.0	0.9
M	M	20	9	91.5	2.3	95.7	0.1	95.9	1.0	94.6	1.6	93.7	0.9
P	M	38	19	89.8	3.4	92.3	2.1	96.8	0.6	96.2	0.8	96.1	1.1
F	M	40	29	95.3	1.6	95.9	1.6	97.4	0.8	96.7	0.2	96.5	1.1
Y	M	42	21	94.3	1.0	93.6	2.0	96.9	0.3	94.7	2.1	95.8	1.2
W	M	18	9	86.1	0.8	86.2	2.6	87.8	6.8	90.7	0.5	88.8	4.0
D	M	20	11	63.9	4.8	71.4	6.7	80.3	5.7	85.6	7.9	83.3	3.8
E	M	25	13	72.6	4.7	82.1	3.1	91.3	3.2	94.0	4.0	92.6	1.9
N	M	38	28	71.7	8.1	82.1	1.8	91.1	4.1	93.4	1.3	91.7	0.7
Q	M	35	16	88.0	3.1	92.1	0.8	96.3	1.6	96.9	1.1	96.0	0.9

H	M	31	8	81.8	5.4	88.3	1.2	92.2	4.0	94.3	1.2	93.1	1.2
K	M	23	6	89.0	2.7	91.3	1.1	85.3	6.3	89.8	4.5	88.3	4.9
R	M	27	17	90.4	1.6	90.3	1.8	82.8	6.1	87.4	3.3	85.4	4.9
G	P	92	63	10.5	3.1	9.1	3.3	9.2	1.3	12.5	1.1	11.8	0.8
A	P	81	59	11.5	3.9	11.1	4.3	12.1	2.1	16.5	2.3	15.6	1.7
S	P	89	75	10.1	4.4	9.3	3.8	9.5	1.7	13.0	2.6	12.0	2.4
T	P	75	44	9.9	2.4	8.7	2.8	10.2	1.9	12.9	2.2	13.3	1.7
V	P	75	50	13.2	2.6	13.7	4.5	14.7	3.5	20.7	3.1	18.4	2.0
L	P	80	57	13.0	3.1	13.5	3.3	14.7	1.4	18.5	3.5	17.7	2.8
I	P	97	59	12.9	3.1	15.2	6.3	16.9	3.8	23.8	5.9	20.1	2.7
M	P	59	39	33.7	1.7	35.7	4.6	35.1	2.3	37.9	3.1	37.2	2.1
P	P	82	51	13.9	3.3	15.7	6.1	15.8	2.0	21.4	4.5	20.9	2.4
F	P	80	47	21.6	2.7	29.2	9.0	31.4	5.2	41.5	10.2	36.1	8.3
Y	P	82	55	21.8	4.0	24.4	6.4	30.0	6.0	37.1	5.8	33.8	6.4
W	P	42	13	27.5	1.7	30.2	4.7	39.2	6.0	42.2	8.0	35.2	4.9
D	P	41	26	14.3	1.5	24.3	12.3	34.9	6.3	38.4	10.0	36.0	6.5
E	P	54	31	16.8	2.4	17.3	6.0	21.7	3.6	28.0	6.1	26.1	4.5
N	P	84	49	8.3	1.5	9.2	3.0	11.2	2.4	11.8	2.4	13.1	1.9
Q	P	84	35	10.3	2.5	12.2	5.3	12.2	1.6	16.7	2.4	16.0	1.1
H	P	81	61	10.1	3.9	11.6	6.0	13.8	3.0	16.4	4.1	16.0	4.3
K	P	62	41	14.8	4.6	19.3	7.2	38.0	17.0	43.3	16.6	34.6	14.9
R	P	75	41	14.1	4.5	17.4	5.1	15.9	5.1	19.8	5.4	15.3	5.0
G	F	53	27	81.1	5.7	84.1	2.7	90.9	1.3	91.2	0.3	90.5	2.0
A	F	68	37	95.2	0.9	95.8	1.2	97.7	0.2	97.4	0.1	97.3	0.6
S	F	56	25	88.4	3.9	90.9	1.6	94.8	0.4	94.4	0.8	93.9	1.6
T	F	54	30	94.0	1.1	94.6	0.5	97.4	0.2	96.7	0.3	96.6	0.9

V	F	67	26	94.6	2.0	95.8	0.4	97.8	0.7	97.5	0.5	97.5	0.9
L	F	65	42	92.0	1.8	92.5	1.0	94.6	1.2	93.3	1.5	94.5	1.3
I	F	41	44	91.9	0.9	92.8	1.3	93.5	1.7	92.6	1.8	94.4	0.7
M	F	32	29	94.1	1.2	95.4	0.6	95.4	1.5	95.4	0.4	95.1	0.7
P	F	87	64	93.3	2.0	95.6	0.6	96.9	0.7	96.4	1.1	96.3	1.3
F	F	67	42	87.9	1.2	87.8	2.3	89.8	2.2	87.4	1.6	89.1	1.9
Y	F	45	25	88.6	1.5	86.8	1.9	90.0	2.4	88.0	1.4	89.7	1.5
W	F	37	15	79.6	2.7	79.0	4.0	76.9	4.2	75.1	2.9	79.7	0.2
D	F	60	35	65.3	7.9	71.8	5.5	78.4	4.2	83.8	6.1	79.3	7.6
E	F	54	27	79.4	4.8	84.0	4.5	91.8	1.3	93.0	2.6	90.0	3.8
N	F	58	31	82.5	7.1	87.4	3.8	94.5	0.9	93.4	0.7	91.2	3.7
Q	F	71	37	93.8	2.0	95.1	0.8	97.8	0.1	96.9	0.4	96.3	0.9
H	F	64	34	90.2	1.8	89.7	1.8	94.0	1.0	93.9	1.3	93.8	1.5
K	F	51	32	92.4	5.5	94.7	1.0	89.7	4.7	92.0	4.6	91.2	5.1
R	F	63	27	91.5	1.5	85.9	2.9	77.9	5.2	84.0	7.4	81.0	7.4
G	Y	54	21	78.3	4.4	81.3	3.1	90.1	3.4	88.9	2.0	88.4	1.3
A	Y	61	2	92.9	0.5	93.4	1.1	97.1	0.6	95.6	1.1	95.5	1.0
S	Y	69	33	87.2	2.0	87.4	2.0	92.6	0.9	92.9	0.7	92.2	0.8
T	Y	68	13	85.2	18.2	93.6	0.4	96.7	0.1	96.4	0.7	96.2	0.9
V	Y	85	10	95.4	1.0	95.5	0.7	97.4	0.1	96.9	0.3	97.2	0.6
L	Y	75	8	89.9	6.0	92.6	0.3	95.8	0.5	95.2	0.9	96.1	0.8
I	Y	80	23	93.4	1.5	93.5	0.5	95.9	0.3	95.6	1.1	96.0	0.5
M	Y	47	4	93.1	0.2	92.9	2.1	95.4	0.8	95.8	0.5	94.6	0.9
P	Y	69	6	93.8	0.9	94.1	0.4	96.3	0.6	96.3	0.2	96.0	0.3
F	Y	70	29	89.9	1.7	88.8	2.9	90.6	2.0	90.7	2.0	92.3	1.0
Y	Y	61	18	89.2	2.9	89.1	1.5	91.0	1.8	89.6	1.7	91.1	0.8

W	Y	41	3	80.7	3.4	80.1	3.1	80.1	2.5	77.8	3.3	82.0	0.1
D	Y	47	6	68.8	5.3	67.6	1.7	75.0	0.9	80.4	6.1	75.0	5.5
E	Y	43	13	84.9	2.9	87.6	2.5	90.8	2.0	90.4	4.2	90.2	2.4
N	Y	69	11	81.2	3.1	82.5	2.6	88.5	2.0	88.1	3.5	87.4	3.4
Q	Y	57	17	90.5	3.2	92.5	0.8	95.4	1.1	94.2	1.7	94.8	0.7
H	Y	48	7	87.0	0.9	86.0	1.6	90.1	1.9	89.4	1.5	89.5	1.1
K	Y	65	8	91.7	6.2	93.5	1.0	87.3	4.4	89.9	4.7	88.4	5.6
R	Y	79	11	87.8	0.5	81.3	3.8	74.2	3.0	80.7	5.6	77.9	5.7
G	W	34	7	61.6	3.9	64.7	4.6	74.1	1.7	72.6	3.4	72.8	4.5
A	W	28	10	80.0	3.1	82.8	2.1	87.9	1.7	84.5	2.4	85.6	0.5
S	W	25	7	79.3	3.9	83.3	2.6	87.7	0.2	85.2	1.8	87.1	0.7
T	W	25	8	80.9	3.4	82.3	1.7	89.1	0.2	85.7	3.3	87.3	0.1
V	W	28	0	87.9	1.1	86.4	3.6	91.9	0.4	90.1	0.2	91.8	1.4
L	W	31	3	84.7	2.0	82.9	1.1	87.8	1.5	84.1	0.4	86.4	1.1
I	W	32	1	87.2	1.6	85.4	3.3	90.7	0.5	86.7	4.2	90.2	0.9
M	W	23	6	85.7	1.5	86.4	3.9	90.9	3.0	89.1	2.8	88.8	2.8
P	W	26	3	86.3	2.4	88.6	0.5	91.3	0.6	91.0	0.9	91.3	2.4
F	W	25	5	81.7	4.9	78.9	6.1	82.7	1.6	80.7	4.6	84.6	2.1
Y	W	27	1	84.3	2.7	82.4	4.5	84.4	1.6	82.9	4.0	86.9	1.2
W	W	12	5	62.1	4.9	61.6	3.7	64.4	3.2	67.2	5.6	70.4	3.4
D	W	11	3	63.5	5.2	62.5	4.4	63.8	2.0	69.8	3.1	65.7	3.3
E	W	18	1	77.0	3.4	74.7	2.5	81.6	1.0	83.3	3.1	80.9	1.8
N	W	17	5	72.3	3.1	74.9	2.1	77.0	1.4	77.4	4.0	76.8	5.1
Q	W	25	3	83.5	1.3	84.4	3.8	88.5	0.8	86.6	4.7	87.8	2.5
H	W	19	2	81.0	1.4	81.0	2.6	81.9	0.9	81.7	2.0	83.1	3.0
K	W	21	3	85.9	3.3	83.5	3.7	81.9	5.8	83.1	3.1	83.8	5.4

R	W	21	3	76.5	1.0	74.2	1.3	63.5	4.5	69.3	3.5	68.8	6.9
G	D	75	22	18.4	3.8	17.1	1.1	16.1	3.3	24.9	2.0	24.5	9.2
A	D	46	7	33.9	5.8	34.6	4.1	36.3	4.1	51.8	3.4	49.6	9.7
S	D	50	16	27.1	7.3	26.3	2.0	25.4	5.5	38.3	3.1	36.7	9.3
T	D	48	16	31.2	8.6	33.5	3.9	32.5	5.5	52.0	6.7	49.7	9.2
V	D	35	3	38.2	3.2	40.5	5.8	48.5	4.7	62.6	7.2	61.7	5.1
L	D	45	19	36.0	6.9	36.7	6.1	39.6	2.0	52.7	8.4	52.4	6.4
I	D	45	16	44.4	5.3	47.9	7.9	53.2	2.4	68.4	5.2	66.5	4.6
M	D	25	10	43.9	1.7	49.0	3.5	49.0	4.3	59.3	7.3	58.4	6.3
P	D	19	6	63.1	4.0	68.2	6.8	85.8	3.3	86.9	4.5	86.8	2.8
F	D	29	11	73.0	6.6	78.0	3.3	88.3	3.1	88.2	3.9	87.1	3.5
Y	D	52	18	64.0	6.5	64.1	2.4	69.5	3.9	77.0	4.1	75.6	5.3
W	D	27	5	61.4	5.3	60.4	5.0	70.0	4.7	72.9	7.4	69.4	3.7
D	D	16	4	30.5	0.8	32.2	6.5	42.6	9.9	48.8	4.0	44.8	7.6
E	D	36	8	35.5	6.5	33.1	2.4	36.8	3.8	50.7	5.6	47.5	7.4
N	D	48	6	23.9	3.9	22.6	2.2	20.4	3.5	30.2	6.1	29.9	8.8
Q	D	48	26	36.8	7.4	34.5	3.3	34.5	4.4	51.0	7.5	48.1	7.9
H	D	47	17	32.0	5.6	32.5	4.3	30.2	4.5	43.4	7.3	44.2	12.2
K	D	44	4	37.0	5.9	35.8	3.3	39.9	4.9	55.2	10.0	52.4	7.0
R	D	62	24	43.2	7.2	48.9	6.5	47.9	8.0	61.3	7.7	56.1	10.9
G	E	54	22	19.5	6.6	18.8	2.4	19.2	3.0	26.4	3.2	28.2	7.5
A	E	55	38	31.8	6.6	32.7	2.7	34.3	4.4	50.4	4.4	47.3	8.1
S	E	56	41	22.1	5.4	21.2	4.3	22.6	4.0	34.6	3.1	33.8	7.9
T	E	41	21	30.9	5.5	31.5	1.6	33.0	4.2	47.3	3.1	45.9	8.8
V	E	51	27	41.8	6.9	39.4	3.4	40.8	3.7	56.4	1.1	54.7	7.1
L	E	50	25	48.8	5.7	45.8	3.5	52.2	3.1	65.7	3.9	63.2	6.4

I	E	41	15	55.3	8.5	55.5	4.9	59.7	4.1	72.7	5.8	72.0	2.8
M	E	23	5	50.4	3.1	52.2	2.8	53.3	3.2	64.2	5.9	62.7	3.3
P	E	45	16	43.2	5.2	47.2	4.4	47.6	4.7	62.6	4.7	63.3	5.7
F	E	56	30	69.7	6.0	72.5	5.6	78.8	2.9	85.9	4.1	84.1	3.9
Y	E	44	15	74.9	4.5	75.9	3.1	82.1	2.3	85.7	4.1	84.3	2.6
W	E	24	8	69.6	5.8	70.6	4.0	77.0	1.6	77.2	4.3	76.2	3.3
D	E	40	20	17.6	1.9	17.1	1.9	18.2	3.2	25.2	5.8	26.0	7.5
E	E	39	18	33.3	4.2	33.3	2.5	38.2	5.6	50.2	8.5	48.2	9.1
N	E	49	22	20.4	1.5	23.0	4.2	22.2	3.8	28.1	5.0	32.5	7.7
Q	E	50	19	38.8	4.5	39.5	5.2	42.1	4.0	53.8	6.9	55.2	5.9
H	E	41	11	33.6	2.8	37.7	6.0	37.7	4.7	49.0	10.2	52.4	8.6
K	E	23	9	53.3	5.8	45.8	3.4	51.6	8.7	66.1	7.1	62.0	7.3
R	E	44	17	39.3	6.4	42.0	5.1	45.2	8.0	63.5	14.5	54.6	9.0
G	N	62	14	29.0	8.5	26.0	3.9	26.6	6.4	38.3	7.4	34.5	7.5
A	N	67	34	51.2	9.3	54.7	4.9	60.6	6.6	76.0	8.0	72.4	4.2
S	N	52	20	33.2	5.5	34.7	6.4	38.9	6.8	53.4	9.3	49.1	7.4
T	N	56	20	43.7	7.2	48.4	2.6	57.2	6.3	74.2	8.8	71.1	6.4
V	N	62	14	68.7	7.4	75.0	5.2	88.1	2.8	92.9	4.1	90.5	4.4
L	N	47	19	65.4	10.1	71.3	4.6	84.7	4.0	88.1	6.8	86.4	7.5
I	N	47	26	81.8	3.9	86.5	4.6	94.3	0.6	94.1	4.0	94.0	3.4
M	N	33	18	65.2	4.3	69.5	4.1	81.1	3.9	85.9	6.5	81.5	4.9
P	N	49	20	65.6	6.0	72.4	6.6	82.3	3.4	89.6	5.2	87.1	4.7
F	N	46	12	84.5	4.0	88.1	2.5	93.5	0.7	92.6	3.4	92.6	3.2
Y	N	46	14	84.0	2.2	86.2	3.4	91.1	2.2	90.8	3.7	90.8	0.5
W	N	26	9	72.3	3.3	74.1	4.5	83.2	2.4	80.5	5.5	79.6	3.6
D	N	44	15	24.6	5.8	22.9	4.9	26.5	3.7	38.9	3.7	35.2	6.9

E	N	41	23	41.7	7.3	43.5	2.6	50.6	4.1	65.7	5.0	64.1	4.4
N	N	51	14	24.3	0.9	26.3	6.0	29.5	6.5	39.3	7.3	38.1	7.3
Q	N	51	21	45.5	5.8	53.1	7.0	61.9	6.5	77.2	8.5	74.4	3.1
H	N	54	17	43.3	5.9	53.7	9.3	64.5	6.7	74.3	8.1	71.0	8.4
K	N	44	5	80.4	9.7	79.9	11.6	79.7	2.9	85.7	5.0	83.5	3.3
R	N	43	14	87.1	2.7	82.5	2.2	71.6	0.6	78.7	7.2	74.9	6.3
G	Q	63	7	30.0	4.1	35.6	5.1	35.8	4.0	46.8	8.3	47.5	7.2
A	Q	58	18	58.6	4.8	67.1	7.5	78.1	3.8	85.9	6.7	85.9	3.6
S	Q	65	25	37.3	3.8	43.1	5.1	48.4	4.6	63.8	10.0	61.4	6.4
T	Q	43	7	58.7	9.3	60.5	5.3	66.9	6.1	81.7	6.9	78.3	2.8
V	Q	48	36	81.3	9.8	87.1	3.7	93.3	1.5	96.1	1.4	95.3	1.3
L	Q	68	47	76.9	11.2	84.8	5.0	94.2	2.0	95.9	1.4	94.6	2.1
I	Q	46	16	87.7	7.5	92.4	2.6	96.8	0.5	97.6	0.2	97.1	0.7
M	Q	31	12	31.9	6.7	44.5	6.9	72.5	11.5	78.4	11.4	74.7	4.9
P	Q	45	13	78.7	8.2	86.3	4.4	93.2	1.3	96.1	1.9	94.8	2.0
F	Q	51	16	92.2	1.3	93.0	1.1	97.3	0.2	96.6	0.7	95.8	1.7
Y	Q	44	27	90.7	1.9	90.3	3.1	95.7	0.7	95.0	1.9	93.7	0.8
W	Q	22	9	44.0	7.9	42.6	7.4	70.2	9.4	69.3	14.9	62.9	3.9
D	Q	36	23	35.4	6.3	35.3	3.5	40.6	3.5	55.3	3.2	50.5	4.1
E	Q	44	19	58.7	5.7	59.1	2.4	68.1	4.5	81.7	4.1	80.0	2.1
N	Q	41	17	35.6	7.8	33.5	3.6	40.8	4.8	55.8	5.6	50.0	4.3
Q	Q	42	18	62.7	8.2	67.4	6.0	77.6	3.1	88.3	5.4	84.6	2.1
H	Q	48	25	58.8	6.1	67.2	6.7	78.6	2.6	85.0	6.6	80.8	5.1
K	Q	38	18	84.6	6.5	90.8	1.6	85.7	1.1	88.2	4.1	86.7	3.1
R	Q	55	21	86.4	2.6	86.4	1.3	76.8	3.4	82.2	6.7	78.3	4.4
G	H	49	27	42.5	3.8	53.4	9.3	58.3	4.7	67.3	2.9	62.7	4.7

A	H	51	2	75.0	6.0	84.6	4.5	91.1	1.8	93.1	1.1	91.6	0.1
S	H	50	28	54.3	6.2	64.4	8.0	77.5	2.7	83.3	2.9	80.4	4.6
T	H	46	12	79.4	6.1	85.2	2.0	91.7	0.8	93.3	0.6	92.4	1.1
V	H	50	21	90.1	4.6	93.5	1.0	96.2	0.7	95.9	0.8	96.2	1.2
L	H	57	18	89.7	3.1	93.3	0.8	95.6	0.4	95.6	0.4	95.7	1.1
I	H	44	8	93.4	0.7	95.0	0.4	96.8	0.1	95.7	0.6	96.7	1.0
M	H	35	1	51.1	8.1	58.6	2.9	76.8	12.0	77.3	4.9	77.4	7.7
P	H	48	4	88.5	5.0	91.9	2.0	93.4	1.7	94.8	1.2	93.9	1.6
F	H	50	19	91.8	2.5	92.3	0.6	94.6	1.2	94.7	1.2	94.8	1.3
Y	H	43	11	89.1	3.3	91.1	0.8	90.8	2.8	92.5	3.2	91.5	3.4
W	H	29	3	41.9	7.0	43.4	2.8	65.5	11.0	72.1	5.6	62.8	8.5
D	H	39	11	41.5	8.1	45.7	7.6	57.8	6.3	68.8	8.0	64.3	5.8
E	H	29	10	71.9	6.2	74.7	4.9	87.0	2.3	91.7	3.6	89.1	1.4
N	H	37	13	54.3	4.6	58.0	6.3	71.3	3.4	77.9	3.2	70.4	3.2
Q	H	40	17	81.8	2.9	87.7	2.4	89.4	2.7	91.6	1.0	88.5	2.0
H	H	33	13	75.3	4.1	80.1	4.5	79.3	6.3	83.4	1.4	78.4	4.6
K	H	15	2	91.4	2.0	92.5	1.4	80.0	8.0	84.9	7.0	83.0	8.3
R	H	25	0	86.3	2.0	84.2	3.4	57.8	8.3	67.1	8.6	64.7	10.9
G	K	31	12	55.3	5.4	59.8	3.8	48.0	2.8	58.7	6.9	53.9	5.9
A	K	34	14	78.1	4.3	82.0	3.6	72.7	1.1	79.7	7.1	78.4	6.2
S	K	33	16	60.7	5.9	63.4	5.2	60.2	3.7	70.6	5.6	65.8	6.7
T	K	29	10	71.5	6.9	75.6	5.7	71.3	2.8	79.7	6.2	76.3	5.7
V	K	40	10	88.3	3.9	91.5	1.5	82.5	1.2	87.7	5.8	85.9	6.2
L	K	42	19	90.0	5.8	92.0	0.8	82.6	3.9	87.3	6.5	85.2	6.8
I	K	31	4	92.1	3.9	94.5	0.2	84.5	3.9	89.1	6.1	87.7	6.3
M	K	31	0	40.7	9.4	49.2	4.0	45.9	11.1	58.5	10.7	54.9	11.5

P	K	51	20	87.1	3.6	89.9	0.9	77.5	3.6	83.3	8.6	82.2	8.7
F	K	45	7	90.6	3.9	91.5	1.4	84.7	6.1	87.6	6.5	87.4	7.4
Y	K	39	4	89.4	5.5	91.6	1.6	84.9	5.4	88.1	5.4	85.8	7.5
W	K	23	1	40.9	6.5	42.4	3.1	49.5	15.6	60.2	6.3	51.5	13.2
D	K	47	7	37.7	3.7	43.0	8.0	48.0	5.6	58.0	3.8	57.1	5.4
E	K	30	5	66.3	7.3	77.4	3.8	76.7	4.2	84.7	2.0	81.0	2.9
N	K	34	16	52.3	8.4	60.0	3.8	50.5	4.6	61.2	4.8	54.9	4.3
Q	K	20	1	86.8	3.2	88.7	3.3	83.0	6.2	87.0	7.2	84.6	7.5
H	K	26	1	80.1	5.7	80.1	6.1	62.2	4.1	73.1	13.9	67.5	11.9
K	K	26	0	91.0	1.7	88.1	5.6	69.1	13.4	73.7	19.0	71.0	17.5
R	K	29	7	80.0	4.5	75.6	2.3	41.4	14.4	53.2	25.4	45.1	19.7
G	R	47	2	61.9	6.2	63.5	5.3	52.3	6.4	63.3	8.1	55.8	9.9
A	R	45	10	87.8	2.6	87.5	1.9	79.7	7.2	82.6	6.7	77.6	10.2
S	R	44	11	75.5	4.6	76.1	3.2	67.1	2.2	73.1	3.8	68.6	6.8
T	R	30	3	85.3	2.4	85.1	2.6	75.1	2.6	80.6	2.3	76.5	4.1
V	R	46	29	92.2	1.9	91.5	0.7	82.4	5.1	87.0	3.1	85.5	4.4
L	R	2	0	74.0	3.0	71.4	5.8	55.9	3.2	57.9	1.7	55.0	2.7
I	R	37	20	91.7	1.6	90.4	1.8	84.3	5.1	87.3	2.9	86.0	5.8
M	R	16	5	55.2	6.1	54.6	2.8	52.7	12.3	60.6	3.2	55.8	10.2
P	R	44	16	88.8	2.8	88.9	2.0	82.0	6.8	85.2	6.2	83.8	7.2
F	R	49	8	90.8	0.7	89.7	0.6	82.1	6.2	85.4	5.2	83.9	7.1
Y	R	38	2	87.9	1.8	85.4	2.7	80.9	5.9	83.4	5.1	82.2	7.8
W	R	17	2	35.6	3.6	33.9	1.9	45.2	15.9	51.6	2.3	46.4	11.4
D	R	34	3	43.1	6.7	52.9	7.3	54.9	3.6	67.7	7.5	60.3	7.5
E	R	36	4	67.2	10.0	78.8	3.4	86.3	1.6	90.8	3.3	88.2	2.1
N	R	48	8	68.8	6.9	67.4	2.6	61.2	2.0	69.6	5.4	64.5	6.1

Q	R	43	2	85.6	1.7	86.9	1.4	80.0	2.9	83.9	5.7	81.9	7.6
H	R	32	8	83.4	2.8	80.7	1.9	60.3	3.6	69.9	9.2	65.7	12.4
K	R	32	5	85.0	6.9	86.3	2.7	60.5	13.7	66.3	19.2	62.8	19.0
R	R	21	3	83.8	5.0	81.7	2.1	50.2	10.4	59.9	12.5	55.9	16.6

Table S-3. Y^p Array Data Summary

Peptide Sequence: Ac-GXY^{p}ZGRC		**Peptide Signal-to-Noise**		**% Dephosphorylation: NIH/3T3**		**% Dephosphorylation: HT-1080**		**% Dephosphorylation: MCF-7**		**% Dephosphorylation: BT474**		**% Dephosphorylation: MDA-MB-231**	
X position	**Z position**	**Average**	**Std. Dev.**	**Average**	**Std. Dev.**	**Average**	**Std. Dev.**	**Average**	**Std. Dev.**	**Average**	**Std. Dev.**	**Average**	**Std. Dev.**
G	G	143	63	39.5	4.3	42.7	4.8	39.4	0.5	47.5	4.9	46.0	5.1
A	G	118	22	42.1	4.0	50.8	9.4	52.2	3.9	59.2	6.8	60.2	4.9
S	G	125	32	46.6	4.5	56.0	9.2	57.9	7.5	64.5	6.5	64.1	5.4
T	G	116	61	48.9	5.4	61.7	12.6	60.5	7.9	67.6	7.1	68.0	5.5
V	G	91	45	53.7	1.4	63.6	13.0	62.5	7.3	69.1	5.3	69.5	5.3
L	G	150	80	58.7	2.2	68.1	10.1	63.6	5.9	69.8	5.1	70.3	4.3
I	G	138	40	56.8	2.0	67.6	9.3	65.1	7.9	70.4	5.0	71.5	4.2
M	G	90	33	59.3	2.8	68.6	4.4	68.4	4.3	73.1	3.8	73.1	3.0
P	G	139	98	42.3	2.7	54.5	6.5	52.7	5.6	55.7	5.5	59.2	3.7
F	G	124	62	63.3	1.7	70.4	8.7	59.7	2.8	64.6	4.3	65.3	2.1
Y	G	75	28	63.5	3.0	70.3	8.1	59.1	2.7	64.3	4.0	65.1	0.8
W	G	55	25	60.9	2.9	65.2	7.3	51.9	0.2	55.4	1.8	55.4	2.6
D	G	133	86	50.2	4.1	63.6	10.2	71.7	4.9	77.9	6.5	75.8	3.8
E	G	103	13	52.0	4.1	65.8	11.8	73.5	3.5	77.6	6.7	77.5	3.9
N	G	131	73	49.5	2.4	59.2	11.3	62.1	6.8	67.4	4.9	67.5	5.8
Q	G	129	32	50.3	3.7	55.9	14.4	60.1	5.9	65.0	5.0	66.1	5.4

H	G	149	40	50.1	5.0	58.2	7.1	52.4	3.1	56.7	4.0	57.9	5.7
K	G	121	49	26.0	5.6	31.6	13.9	22.7	3.6	26.7	3.1	26.1	2.0
R	G	92	19	34.9	14.1	38.7	17.3	24.1	1.1	28.2	2.1	26.0	2.9
G	A	135	58	34.4	7.3	40.2	10.2	40.5	5.6	44.3	5.5	42.4	4.1
A	A	148	59	42.7	6.8	50.4	12.5	52.9	6.2	58.1	7.1	58.1	4.9
S	A	91	31	47.1	5.7	55.8	13.2	59.1	8.9	64.0	7.9	63.3	4.4
T	A	141	56	47.7	5.5	55.9	9.4	59.5	9.1	66.8	8.2	65.9	4.9
V	A	152	58	52.7	5.1	60.6	12.1	56.4	5.6	67.6	5.4	67.8	4.8
L	A	145	55	56.0	5.3	64.3	10.0	61.2	4.0	69.1	6.2	66.4	8.0
I	A	136	41	53.7	4.6	63.5	10.8	61.6	8.0	68.2	6.8	67.4	7.7
M	A	78	26	55.6	2.4	64.0	5.7	66.8	6.1	72.3	5.3	70.8	4.4
P	A	102	39	36.1	5.4	42.5	7.7	44.6	8.5	50.1	7.3	48.5	3.5
F	A	107	40	60.3	5.5	68.4	11.3	56.7	4.1	62.4	5.1	62.9	4.3
Y	A	83	35	61.9	1.4	66.4	7.1	58.4	3.9	63.7	3.7	63.8	3.5
W	A	50	15	61.0	1.2	61.3	5.5	51.6	2.6	54.0	1.1	54.3	2.2
D	A	83	27	49.4	4.0	60.3	13.9	68.9	9.7	76.5	9.3	71.8	4.9
E	A	94	31	44.5	2.9	58.0	13.5	68.8	7.7	76.3	8.2	72.5	4.1
N	A	124	61	45.5	3.1	54.6	12.6	58.3	9.6	63.6	8.0	62.3	7.1
Q	A	155	99	43.6	4.3	52.7	12.6	57.5	7.9	62.2	8.2	61.0	6.4
H	A	122	68	44.2	3.0	55.1	9.9	50.3	3.1	56.4	4.4	56.6	7.3
K	A	106	49	19.9	2.4	25.1	10.7	20.5	4.9	23.7	3.7	22.8	2.4
R	A	108	32	30.0	3.0	31.3	8.8	20.6	1.6	26.2	1.9	23.8	4.2
G	S	140	82	40.5	3.3	46.6	11.0	49.6	5.4	52.7	5.2	51.8	5.3
A	S	142	49	47.8	2.6	56.1	12.9	60.9	8.4	64.6	6.5	65.8	5.7
S	S	112	40	50.7	2.3	56.2	13.1	62.3	7.8	66.2	6.8	66.7	6.3
T	S	105	19	51.2	1.2	61.1	12.3	65.0	8.0	68.8	7.6	69.6	6.6

V	S	137	45	51.4	6.2	62.5	9.3	65.0	8.7	70.1	5.9	70.8	6.4
L	S	154	48	60.0	3.3	68.4	10.0	65.2	7.0	69.9	5.3	70.6	5.7
I	S	117	47	58.9	6.2	66.3	11.2	65.5	7.1	70.6	5.7	71.3	5.2
M	S	73	12	58.1	4.3	64.9	6.1	66.4	4.7	70.1	5.0	69.5	3.8
P	S	146	32	43.9	4.1	48.1	10.1	47.7	6.5	54.2	6.1	54.9	4.6
F	S	161	45	64.7	4.4	71.3	8.8	58.6	4.3	64.3	3.6	63.9	4.1
Y	S	198	251	62.9	2.1	68.5	9.1	56.9	0.7	64.9	4.1	57.3	8.5
W	S	28	4	63.8	1.4	62.9	3.5	51.0	0.7	55.2	0.6	52.8	3.8
D	S	104	70	47.9	3.1	60.8	11.7	67.9	8.8	77.8	8.9	70.1	5.2
E	S	58	28	46.1	4.0	59.9	12.7	65.5	11.5	75.4	6.9	70.7	5.6
N	S	78	36	49.3	3.3	58.8	14.4	60.2	12.4	69.2	6.6	43.2	33.5
Q	S	135	76	43.9	3.9	54.6	12.4	57.8	10.4	64.1	7.2	60.8	6.8
H	S	78	34	46.7	4.3	54.5	10.0	52.5	7.0	57.3	4.1	54.6	5.8
K	S	83	42	24.9	4.8	32.3	12.9	22.8	1.8	27.2	1.6	24.5	2.2
R	S	64	25	29.7	2.3	33.6	14.9	19.9	3.5	27.5	5.9	19.8	1.7
G	T	163	93	44.8	3.1	49.7	10.5	52.5	9.2	56.2	5.6	53.4	7.2
A	T	145	77	46.0	3.1	54.6	11.9	59.7	9.7	65.3	6.1	62.6	7.2
S	T	86	38	52.7	1.1	60.9	12.7	63.6	12.2	68.3	5.5	66.4	8.2
T	T	117	45	50.7	3.7	59.9	13.7	63.7	8.3	70.0	7.1	67.6	7.5
V	T	100	27	50.0	7.1	62.8	11.7	65.3	7.8	70.2	5.0	67.8	6.4
L	T	103	40	59.1	2.3	66.7	9.5	64.1	7.8	69.0	3.7	64.4	3.5
I	T	99	50	56.8	2.4	66.0	8.9	64.0	7.7	68.7	4.7	68.1	6.4
M	T	58	25	61.0	5.6	69.2	4.8	58.0	16.9	71.5	4.3	70.0	4.0
P	T	107	40	43.2	2.1	49.5	7.1	49.7	6.2	54.3	6.1	53.0	4.1
F	T	101	37	61.5	0.6	67.7	8.4	55.4	2.3	62.5	3.4	62.2	3.2
Y	T	95	57	66.7	8.0	71.8	7.7	57.7	3.2	62.4	3.9	63.1	3.6

W	T	46	14	51.4	8.8	57.4	13.3	50.2	3.8	52.9	0.7	52.6	3.5
D	T	81	24	54.1	4.6	63.6	10.5	70.2	6.2	77.9	9.2	72.7	5.8
E	T	96	47	52.6	5.9	64.1	9.5	68.6	6.3	77.4	9.0	72.1	6.6
N	T	136	74	51.9	2.8	59.6	14.2	60.0	9.6	66.3	7.3	61.2	8.7
Q	T	119	59	45.4	4.2	54.1	11.3	54.9	11.7	63.3	7.7	59.0	10.3
H	T	90	55	47.1	5.0	55.2	12.6	51.6	6.9	58.9	5.9	56.2	5.8
K	T	81	50	26.3	6.7	32.9	10.1	24.5	0.7	29.3	2.0	28.2	3.1
R	T	63	41	44.0	18.6	47.1	16.8	24.3	1.5	33.7	3.5	26.6	4.2
G	V	126	50	36.7	4.7	45.6	12.9	44.1	7.1	49.0	5.3	45.0	4.3
A	V	251	264	46.8	4.4	52.2	8.9	55.5	9.5	60.8	6.0	58.7	6.5
S	V	119	50	51.1	4.4	56.7	10.3	58.5	11.7	63.8	6.1	61.8	7.6
T	V	129	73	50.1	4.8	59.2	12.7	59.5	13.1	65.0	6.9	62.8	6.8
V	V	123	50	54.2	3.2	65.2	12.5	57.7	10.2	62.6	4.3	61.9	6.6
L	V	86	54	59.3	1.8	69.1	9.9	58.4	6.4	63.0	1.5	62.5	3.7
I	V	159	125	55.3	2.3	68.3	11.2	55.1	4.8	59.7	1.3	58.8	3.7
M	V	69	32	63.4	1.7	69.2	6.3	65.2	4.0	69.8	2.9	69.6	4.6
P	V	163	90	44.1	6.2	49.9	9.1	48.7	7.1	52.6	5.2	52.9	7.3
F	V	102	47	56.8	2.5	63.7	8.8	48.5	3.2	52.4	0.9	52.3	2.5
Y	V	104	32	59.3	2.8	64.4	7.5	51.0	2.3	57.0	1.0	56.1	3.2
W	V	72	47	56.5	6.1	59.7	7.2	44.4	3.3	46.0	2.8	46.1	3.7
D	V	126	38	56.5	5.2	64.1	10.5	70.7	7.4	76.5	7.8	73.7	7.0
E	V	90	56	53.8	5.0	62.2	8.5	67.9	6.2	76.0	6.9	72.2	5.8
N	V	73	35	53.1	3.7	58.1	10.8	58.4	8.6	63.5	6.8	62.7	7.3
Q	V	130	42	48.6	1.8	54.3	9.4	54.8	8.5	60.2	7.4	58.7	6.4
H	V	127	52	49.3	4.0	51.8	10.2	48.7	6.2	54.5	3.5	53.9	5.9
K	V	93	40	27.2	2.5	34.8	10.7	22.2	1.2	25.6	1.7	25.5	2.6

R	V	82	47	32.8	3.8	39.3	16.9	18.4	2.1	27.9	5.0	20.8	0.7
G	L	101	38	52.3	0.7	55.9	10.0	50.7	6.0	57.1	3.0	55.9	5.4
A	L	154	45	54.4	4.5	61.4	6.9	56.9	6.3	63.5	4.7	62.0	11.1
S	L	129	66	58.2	6.2	65.8	9.6	62.2	6.9	67.6	5.4	66.7	9.1
T	L	72	37	53.8	6.5	64.3	10.3	62.7	9.0	68.4	5.7	65.3	11.0
V	L	135	89	56.2	5.5	66.5	7.8	58.9	7.8	62.9	3.2	63.3	7.5
L	L	116	40	59.4	1.6	69.3	7.3	56.3	6.0	59.6	0.8	60.9	3.4
I	L	130	71	56.0	3.4	66.9	7.6	52.3	6.7	55.4	1.4	57.3	3.0
M	L	78	15	63.5	3.1	70.4	6.9	63.6	6.7	67.8	2.6	69.4	4.4
P	L	106	43	51.8	7.4	58.8	9.1	55.2	19.0	57.0	7.8	60.8	9.0
F	L	187	125	53.4	1.8	61.4	11.1	45.4	4.1	47.9	1.6	47.8	2.3
Y	L	114	37	58.0	0.2	62.9	9.9	49.6	3.5	52.5	1.1	53.3	2.9
W	L	45	14	51.4	4.1	55.6	9.5	38.1	5.0	39.2	2.2	39.2	4.6
D	L	97	27	67.1	3.9	75.2	7.8	75.0	5.8	80.7	3.3	79.6	4.0
E	L	91	66	62.5	5.1	72.1	7.9	76.5	5.9	80.9	3.2	79.8	4.9
N	L	124	71	58.4	6.8	67.3	11.5	61.6	8.8	65.6	3.8	66.9	7.3
Q	L	125	72	52.1	5.5	62.0	11.8	59.6	8.0	62.6	4.7	63.8	7.2
H	L	118	75	55.6	3.7	61.8	9.1	50.1	6.4	53.3	2.6	56.8	5.9
K	L	129	36	32.9	5.9	38.0	12.1	25.9	0.4	28.7	2.4	29.5	3.9
R	L	84	33	35.0	4.1	37.4	12.7	20.1	0.2	27.2	4.0	21.9	2.9
G	I	129	107	47.2	4.6	51.1	8.8	50.3	7.5	53.4	4.4	54.1	6.5
A	I	126	70	54.2	4.7	60.4	11.1	59.5	8.2	63.1	4.7	63.6	5.6
S	I	126	64	61.1	3.7	67.1	12.9	60.8	9.3	66.0	4.2	65.4	6.0
T	I	138	80	55.6	3.3	64.2	12.9	58.9	8.5	64.6	5.7	63.4	6.6
V	I	140	87	55.1	0.7	69.7	6.8	56.7	6.7	61.5	1.9	62.8	5.2
L	I	154	81	57.4	1.6	68.6	10.6	52.0	2.9	55.4	1.2	55.0	4.1

I	I	139	92	62.8	3.9	73.1	8.2	51.7	3.2	55.6	2.8	54.9	5.4
M	I	65	25	63.8	3.3	69.8	4.6	65.7	5.1	67.5	2.5	69.1	5.5
P	I	140	18	48.0	7.6	57.8	9.6	49.4	8.0	55.2	4.5	55.6	4.4
F	I	109	49	52.2	3.4	60.7	11.4	41.6	0.7	46.3	0.3	45.2	4.0
Y	I	126	60	56.3	2.7	65.9	9.2	45.7	0.5	51.0	1.3	52.8	2.4
W	I	43	7	51.6	1.9	54.4	5.9	38.8	4.4	37.4	2.8	35.7	12.9
D	I	99	49	57.7	5.9	66.5	9.7	69.3	8.8	77.1	5.7	75.6	7.5
E	I	140	49	56.1	5.5	66.2	10.3	74.1	13.1	80.1	4.7	80.2	9.9
N	I	151	44	54.1	4.8	62.8	12.0	58.6	9.5	65.2	4.3	66.6	10.5
Q	I	279	200	51.9	4.1	60.7	12.4	58.6	10.7	63.7	5.3	63.7	8.0
H	I	141	79	53.9	3.9	59.6	10.2	51.9	7.8	56.6	2.9	56.9	7.6
K	I	112	82	33.6	4.6	38.3	11.8	27.5	1.5	31.4	2.7	32.3	6.8
R	I	96	63	38.1	6.7	39.6	14.8	25.4	4.2	32.3	1.9	27.0	6.8
G	M	83	31	63.8	5.9	67.7	6.4	68.7	6.6	70.8	3.1	71.1	6.5
A	M	62	32	66.0	4.5	73.3	6.4	74.7	6.0	75.1	2.9	76.7	6.8
S	M	72	33	66.7	3.0	71.4	8.4	71.1	6.8	74.0	3.1	74.0	6.9
T	M	75	28	69.6	4.2	74.7	4.5	73.8	7.0	76.1	3.9	76.7	6.6
V	M	81	28	68.6	5.0	75.3	5.9	69.5	5.7	74.6	2.6	76.4	7.7
L	M	93	36	71.3	4.3	76.8	6.3	70.4	7.1	72.6	3.0	74.6	5.5
I	M	88	35	73.7	4.5	75.8	5.6	68.4	7.4	70.8	1.9	74.6	4.4
M	M	66	18	73.4	3.3	75.5	3.7	70.6	6.3	72.3	3.2	75.1	5.5
P	M	110	38	63.7	4.9	68.2	5.6	66.1	8.0	68.9	3.2	71.7	3.8
F	M	114	50	70.8	0.9	74.5	5.5	64.4	5.9	67.2	1.2	67.7	5.1
Y	M	93	40	70.6	2.1	74.2	4.9	66.4	3.6	68.9	1.3	69.5	3.7
W	M	49	21	65.8	1.6	63.9	7.4	57.1	6.3	58.7	4.5	58.7	7.6
D	M	62	33	73.0	3.6	76.3	6.3	83.6	6.4	87.9	4.0	86.9	4.8

E	M	83	28	70.2	3.8	75.6	5.6	80.9	8.6	86.2	5.1	84.6	6.1
N	M	53	18	70.9	1.9	73.3	7.0	74.3	9.0	77.7	3.4	78.9	6.9
Q	M	84	25	72.3	10.7	77.2	10.2	69.3	9.2	75.1	3.1	75.2	7.3
H	M	71	17	71.5	9.4	72.2	8.0	64.4	7.3	68.7	2.9	70.2	7.6
K	M	78	25	51.5	5.8	55.4	7.1	44.9	1.9	49.3	2.8	48.6	6.9
R	M	56	28	48.8	14.8	50.7	12.6	47.4	3.2	50.9	2.3	48.7	9.0
G	P	234	206	13.3	1.8	13.8	1.9	15.4	3.0	16.6	2.5	16.4	2.7
A	P	125	94	15.9	1.5	18.8	3.8	19.0	4.4	19.1	3.3	19.5	3.6
S	P	140	82	23.6	2.1	27.5	8.0	26.7	4.8	28.5	4.3	28.7	5.1
T	P	119	48	22.5	3.5	27.2	7.5	24.9	1.9	29.8	3.9	28.5	3.0
V	P	138	68	20.0	4.3	27.1	7.0	23.7	4.2	24.1	4.0	26.0	4.9
L	P	117	55	33.2	7.3	44.8	10.1	36.7	5.4	38.5	4.1	39.5	6.0
I	P	119	74	27.5	6.8	34.1	6.7	28.2	4.1	29.9	3.5	31.5	4.5
M	P	88	38	36.3	1.1	42.9	5.7	40.7	2.2	39.9	4.1	42.1	2.5
P	P	93	45	14.9	2.4	21.0	7.4	14.5	0.1	15.4	2.1	15.7	1.2
F	P	104	56	39.5	4.1	46.8	5.4	38.3	5.5	39.9	2.2	40.6	5.2
Y	P	109	75	41.2	4.2	50.1	5.4	39.5	3.1	42.5	2.2	42.9	5.4
W	P	56	25	43.9	4.3	50.4	3.8	33.2	2.4	35.3	1.4	35.0	4.7
D	P	115	53	23.9	5.4	35.1	8.9	25.9	2.7	29.9	3.3	28.8	3.5
E	P	78	28	23.3	7.3	33.4	9.3	19.0	2.0	26.0	3.0	24.2	1.6
N	P	92	45	20.3	5.1	27.4	11.7	17.1	0.9	20.2	2.7	20.0	1.6
Q	P	133	95	17.7	3.4	25.9	12.3	15.0	0.5	16.6	2.3	16.9	2.1
H	P	122	51	20.4	3.2	27.1	10.5	18.2	1.8	21.4	2.8	19.8	3.0
K	P	87	28	10.1	2.0	19.0	15.7	8.5	0.2	25.9	26.4	7.8	0.3
R	P	105	16	12.6	3.7	21.7	13.0	6.9	0.1	31.5	37.7	6.6	1.4
G	F	103	38	61.9	2.6	62.9	6.0	47.2	6.0	43.7	15.1	54.2	4.9

A	F	124	37	58.6	6.4	66.1	10.4	52.7	7.1	45.7	22.1	63.6	5.7
S	F	128	44	64.0	3.2	69.9	6.8	56.5	8.1	60.9	3.4	63.6	6.4
T	F	145	82	65.2	3.8	70.6	7.6	57.0	9.2	62.5	2.6	64.4	6.8
V	F	134	57	59.5	2.7	69.2	7.4	48.5	3.3	54.1	1.6	57.0	4.7
L	F	106	27	54.8	1.8	63.5	10.4	43.6	0.4	47.8	1.5	49.5	4.1
I	F	90	34	54.0	1.4	60.9	8.9	43.6	4.1	44.6	2.2	47.3	4.1
M	F	77	17	67.4	1.0	70.7	5.6	61.5	4.9	63.0	3.0	65.5	5.4
P	F	100	36	59.5	1.4	67.9	3.3	46.1	5.4	49.7	2.5	58.9	1.8
F	F	98	37	52.3	1.5	56.0	10.9	40.8	5.0	41.4	3.5	42.5	4.3
Y	F	107	51	57.3	1.2	61.3	10.4	45.5	4.7	46.2	2.1	48.2	4.7
W	F	35	6	48.7	4.0	50.7	10.3	34.9	5.2	32.9	5.7	34.7	7.2
D	F	99	18	70.9	2.3	78.4	5.3	69.9	4.1	73.3	1.4	75.1	4.6
E	F	90	40	67.1	6.5	61.6	29.7	70.2	3.2	74.3	2.4	75.9	4.2
N	F	110	38	66.2	1.9	70.1	6.3	55.9	6.4	59.6	2.9	63.6	6.7
Q	F	124	52	63.3	5.5	66.8	8.7	56.4	6.9	61.4	2.8	63.9	6.2
H	F	126	39	59.4	2.7	61.1	7.1	44.9	4.8	48.1	2.5	50.8	5.9
K	F	92	29	52.0	13.2	48.7	15.7	26.1	0.8	29.9	2.5	28.7	4.9
R	F	67	28	43.2	3.8	40.9	15.9	23.7	2.7	28.2	1.5	25.3	4.3
G	Y	130	39	46.2	14.3	48.8	12.8	40.7	3.6	44.8	2.5	46.3	4.7
A	Y	109	18	61.1	3.7	66.0	8.1	48.8	4.5	53.6	3.3	56.7	4.8
S	Y	108	38	65.3	1.4	65.3	6.0	52.9	6.7	59.3	2.4	60.2	6.5
T	Y	109	122	63.4	3.6	65.9	6.7	54.4	7.5	60.0	2.5	60.9	5.8
V	Y	106	45	59.9	3.2	64.8	5.4	50.1	6.5	56.0	2.2	58.1	4.5
L	Y	89	28	60.8	1.9	66.2	8.3	48.1	6.1	54.1	2.6	56.1	4.5
I	Y	98	37	55.5	1.6	64.2	8.1	46.9	3.6	51.6	0.9	54.2	4.2
M	Y	50	5	62.9	1.7	66.2	5.2	57.9	6.3	60.3	2.0	62.1	4.9

P	Y	186	216	53.4	3.3	62.9	4.2	42.6	6.1	46.8	3.1	54.7	0.8
F	Y	101	27	55.2	2.6	60.3	10.5	44.5	5.5	46.4	2.5	46.9	4.0
Y	Y	111	65	56.8	1.2	61.5	8.7	45.5	3.5	49.2	2.2	49.1	3.6
W	Y	42	7	49.1	0.2	50.8	6.0	33.0	1.5	36.0	2.9	35.8	7.3
D	Y	132	125	65.0	3.8	69.4	7.1	63.2	1.5	69.5	2.7	68.5	2.8
E	Y	90	18	66.1	2.1	70.4	6.9	64.7	5.0	69.8	2.7	68.0	3.6
N	Y	81	16	62.8	2.4	66.5	9.8	43.3	7.8	49.9	12.3	43.2	6.5
Q	Y	80	15	62.2	4.0	65.4	8.6	54.6	5.2	58.3	2.5	57.9	5.4
H	Y	68	7	58.3	6.2	60.4	8.4	44.0	3.9	47.3	1.9	47.2	5.8
K	Y	79	35	41.2	4.5	40.7	11.4	25.6	4.7	27.4	3.0	27.5	6.4
R	Y	55	21	37.5	2.5	36.8	10.8	20.7	3.1	24.8	1.8	21.3	4.7
G	W	33	8	63.4	2.5	64.3	4.3	46.9	7.2	49.2	3.0	50.0	7.5
A	W	39	13	68.8	1.1	69.4	4.2	54.6	5.3	55.9	3.3	59.8	7.4
S	W	27	6	64.9	6.4	65.2	9.7	58.1	6.0	59.5	2.4	60.7	6.9
T	W	41	6	70.7	0.3	72.0	4.1	59.0	6.1	61.2	1.9	62.4	7.6
V	W	32	6	67.6	1.3	66.2	2.7	53.4	6.9	53.1	2.4	54.4	7.6
L	W	37	13	62.9	2.2	64.9	5.5	47.1	5.7	48.9	2.7	49.9	7.4
I	W	37	8	66.5	3.6	66.8	6.6	47.5	5.6	49.4	4.6	50.7	9.1
M	W	42	16	70.9	1.4	64.5	9.0	60.6	7.5	62.6	4.5	62.3	7.9
P	W	46	13	66.8	1.2	69.6	7.8	49.8	5.4	54.1	1.6	60.3	7.2
F	W	47	3	54.7	1.8	55.0	3.3	40.6	3.1	43.2	4.3	43.7	8.1
Y	W	44	12	57.3	3.7	56.3	5.8	40.4	0.5	45.3	3.3	47.0	7.2
W	W	16	6	46.8	3.1	45.3	3.8	33.0	2.3	35.3	3.9	36.2	10.4
D	W	22	4	71.3	2.4	71.9	4.5	62.6	1.1	64.4	2.4	67.2	6.1
E	W	24	8	72.5	2.6	71.5	6.1	63.6	0.6	66.1	1.2	69.3	7.9
N	W	28	5	68.2	1.5	69.2	3.8	53.7	2.7	57.3	1.3	59.8	8.1

Q	W	38	10	69.1	1.0	68.1	5.2	55.9	7.0	57.9	2.5	58.2	8.4
H	W	41	12	62.2	0.8	61.9	3.6	45.5	5.4	46.2	3.5	48.7	8.7
K	W	27	7	51.0	3.3	48.5	6.1	29.4	6.7	30.2	5.3	32.4	12.2
R	W	30	8	41.6	1.9	41.1	8.8	22.3	5.9	25.3	4.6	23.8	10.6
G	D	69	16	48.4	4.1	52.0	6.5	51.0	9.6	54.7	5.4	52.2	8.5
A	D	67	11	67.7	6.5	73.4	6.7	70.3	8.4	75.2	4.8	73.2	7.3
S	D	42	14	68.9	2.1	71.7	8.3	68.0	6.3	71.5	2.3	71.8	5.1
T	D	26	11	70.6	2.3	77.5	5.0	69.2	5.0	74.2	0.9	74.9	5.8
V	D	39	12	71.2	2.3	76.9	6.7	74.6	3.8	78.1	0.9	78.8	4.9
L	D	46	15	76.1	5.0	80.9	5.9	80.2	6.2	83.8	1.2	84.2	4.6
I	D	83	7	75.0	3.2	79.6	7.2	79.6	7.9	84.4	2.4	83.4	5.3
M	D	51	20	70.5	4.4	72.9	6.2	75.7	5.3	79.7	3.5	76.8	4.9
P	D	60	29	56.7	8.2	61.7	8.1	59.5	8.0	63.5	5.2	62.7	8.3
F	D	83	39	78.2	1.2	81.3	5.7	73.6	3.5	78.9	2.1	77.8	5.6
Y	D	64	30	78.4	3.0	79.8	4.2	74.6	4.4	79.1	0.7	77.1	5.2
W	D	30	6	72.7	0.9	69.7	4.6	66.8	2.6	69.5	0.6	68.3	5.2
D	D	54	17	67.4	12.5	70.4	6.3	73.6	1.6	83.4	3.7	77.6	3.7
E	D	57	14	59.1	7.7	65.5	7.6	73.7	2.2	81.0	4.1	78.2	3.2
N	D	41	20	68.5	5.8	70.7	7.1	70.4	12.2	77.3	4.7	77.5	6.7
Q	D	35	10	65.5	4.7	66.7	7.6	67.1	9.9	74.2	6.1	74.3	7.6
H	D	58	37	67.5	4.5	73.5	5.5	61.2	7.7	70.2	3.5	69.1	7.6
K	D	41	20	57.9	9.0	61.1	10.6	43.3	6.7	51.9	6.0	50.6	9.8
R	D	87	97	49.1	12.9	59.6	10.8	43.7	8.4	54.4	3.1	49.5	9.4
G	E	96	46	35.6	22.9	49.6	10.7	47.2	4.2	54.3	4.5	51.5	7.0
A	E	118	23	57.9	7.4	64.7	10.2	67.1	9.7	71.3	5.3	71.3	6.9
S	E	66	29	63.7	10.8	71.7	10.0	73.4	8.5	78.4	4.3	76.8	5.6

T	E	73	26	67.0	9.2	73.5	10.9	76.9	8.0	81.6	3.1	80.9	5.7
V	E	51	12	62.0	11.4	70.8	14.5	75.7	5.8	81.0	3.6	80.5	5.7
L	E	75	16	75.3	5.5	78.4	8.3	80.6	8.0	84.1	2.6	84.2	4.8
I	E	84	25	70.6	6.8	78.5	8.4	76.1	3.8	82.0	2.7	81.3	5.0
M	E	41	12	68.5	2.1	69.4	4.1	72.5	4.4	75.6	4.6	74.4	4.8
P	E	37	10	54.0	5.2	58.5	10.1	47.0	6.2	51.9	3.6	55.6	6.3
F	E	33	3	77.7	3.9	80.0	6.8	76.3	0.1	83.3	2.1	82.7	3.3
Y	E	47	9	77.7	3.8	81.4	4.7	76.5	2.9	81.2	2.0	81.7	4.0
W	E	26	8	64.7	8.5	64.4	11.9	65.9	3.6	64.8	2.8	68.3	4.6
D	E	72	36	69.8	7.1	74.9	8.8	78.0	0.4	84.3	5.3	79.1	4.6
E	E	58	38	56.9	6.0	65.5	9.8	72.0	1.4	79.8	6.2	73.4	5.5
N	E	74	37	61.6	10.4	69.1	9.6	67.6	4.9	70.4	14.2	70.6	7.6
Q	E	57	37	57.0	7.1	62.8	8.0	60.2	9.3	68.3	5.6	63.3	8.9
H	E	107	59	62.3	1.7	67.3	6.3	57.5	3.9	65.6	6.8	57.8	9.5
K	E	67	19	34.9	2.8	43.1	13.2	30.2	7.5	34.2	3.5	33.7	6.1
R	E	47	18	57.7	7.7	59.9	8.3	43.9	10.2	55.8	6.7	49.1	7.4
G	N	70	17	47.5	4.4	42.0	12.6	41.5	8.0	47.7	5.8	46.4	7.2
A	N	81	23	56.3	6.7	59.3	8.2	56.9	12.4	64.1	5.4	62.3	9.2
S	N	76	22	57.3	13.3	65.5	12.9	63.0	12.6	68.0	6.0	66.9	9.8
T	N	66	27	59.3	10.2	64.9	10.8	65.1	12.9	70.4	4.3	71.8	10.1
V	N	94	31	57.0	9.1	67.5	10.9	66.6	13.7	70.2	4.5	72.5	9.0
L	N	92	20	67.5	7.4	73.3	12.2	66.0	8.2	71.9	3.3	74.1	9.0
I	N	110	45	66.2	6.7	72.4	9.7	66.9	12.1	70.3	3.5	73.0	9.2
M	N	72	17	65.2	8.5	71.9	6.0	70.9	6.6	73.4	3.6	73.9	6.6
P	N	102	41	53.2	8.6	58.1	12.9	53.4	11.6	56.5	5.5	57.6	9.4
F	N	100	26	72.4	5.3	73.2	7.6	61.8	9.0	65.0	4.2	68.0	8.5

Y	N	99	46	72.0	1.6	73.3	7.4	62.4	7.1	64.8	2.8	66.6	7.8
W	N	36	16	66.2	2.8	66.9	4.4	53.6	3.5	55.5	4.0	57.9	7.0
D	N	67	15	63.1	5.4	70.7	8.1	71.4	8.3	78.4	6.1	77.0	9.6
E	N	85	20	62.9	4.2	72.0	10.4	67.7	6.3	75.4	8.4	73.5	10.6
N	N	90	20	58.3	8.3	62.9	12.0	57.1	13.1	62.6	4.9	63.9	10.1
Q	N	96	17	56.6	7.8	62.6	13.5	55.5	11.4	60.1	5.9	61.0	10.0
H	N	102	32	51.1	15.5	56.5	15.9	52.7	10.6	55.3	3.4	57.5	11.4
K	N	71	30	36.7	2.7	41.0	11.8	24.2	0.5	27.4	4.0	28.8	6.6
R	N	50	19	35.7	13.3	39.5	20.0	24.4	4.0	29.8	3.2	27.3	7.0
G	Q	81	33	55.8	7.1	58.0	12.1	54.2	10.1	58.7	4.7	57.4	9.8
A	Q	77	29	62.3	6.8	67.4	12.1	58.7	20.6	69.8	4.8	68.0	10.6
S	Q	91	28	63.2	13.7	67.3	13.6	64.7	13.1	71.5	5.0	67.6	11.5
T	Q	90	50	65.4	6.3	67.7	9.4	65.7	19.4	73.1	4.8	70.2	10.8
V	Q	96	43	65.5	7.5	69.7	8.5	49.5	4.0	66.6	13.4	55.9	2.8
L	Q	67	14	69.2	5.7	75.4	9.5	67.9	15.4	74.0	4.0	71.9	10.6
I	Q	117	68	65.6	6.8	72.2	8.4	66.3	16.8	73.1	5.1	70.5	10.6
M	Q	46	17	43.9	6.5	47.6	7.9	44.7	13.7	50.7	8.3	49.7	11.1
P	Q	86	24	68.9	9.2	68.6	11.5	56.2	16.2	61.2	6.4	60.0	10.3
F	Q	82	17	70.8	4.5	73.6	9.5	63.4	9.4	66.9	3.2	67.9	6.9
Y	Q	135	110	70.5	5.0	74.3	7.6	60.7	4.7	65.3	3.2	66.8	5.7
W	Q	42	24	21.5	7.9	28.5	11.6	28.2	4.5	28.8	2.2	27.7	3.7
D	Q	102	32	66.3	7.8	74.3	11.3	78.9	13.3	85.2	5.7	81.8	7.7
E	Q	77	16	62.8	6.8	72.3	9.5	77.0	13.4	83.9	6.1	80.4	8.6
N	Q	79	26	68.4	7.7	71.8	11.9	68.7	15.1	74.2	5.8	71.8	12.4
Q	Q	68	29	61.3	4.4	67.4	11.5	63.0	14.2	68.1	6.4	68.4	13.5
H	Q	74	20	64.3	8.8	68.4	11.0	59.0	9.0	63.8	4.2	66.5	11.3

K	Q	57	23	46.3	6.2	49.4	11.8	30.5	5.5	33.4	5.0	38.0	12.1
R	Q	54	22	52.0	3.6	50.1	12.9	31.9	7.6	34.9	4.7	35.1	11.2
G	H	92	28	51.1	8.8	47.5	12.6	37.4	8.6	41.2	3.6	42.4	10.3
A	H	137	126	62.0	5.0	62.3	8.6	49.0	8.6	51.8	3.4	54.4	10.8
S	H	89	45	64.3	4.1	63.8	11.6	53.5	9.2	56.8	3.5	59.2	12.2
T	H	123	93	61.8	7.1	62.4	9.2	57.4	10.0	60.5	3.6	63.0	11.5
V	H	108	41	63.2	6.0	65.2	8.7	55.9	9.7	61.0	4.0	64.0	11.0
L	H	77	28	67.7	5.7	69.0	6.6	57.5	9.7	62.3	3.4	63.8	11.4
I	H	83	10	63.8	10.0	67.0	9.1	57.3	8.8	62.4	4.1	63.7	11.3
M	H	55	6	41.0	8.6	39.2	8.6	32.1	10.2	36.4	6.7	36.6	7.7
P	H	92	33	55.1	9.2	52.4	11.3	41.7	13.4	43.8	4.1	47.5	7.4
F	H	78	17	70.8	1.0	70.0	4.5	54.3	10.6	58.0	4.3	59.4	7.5
Y	H	62	7	70.7	1.3	69.9	3.6	55.6	8.1	60.6	2.9	61.6	7.3
W	H	40	19	21.8	2.4	25.9	7.9	23.8	3.5	25.7	3.0	25.0	4.0
D	H	57	10	64.3	4.7	70.8	7.3	63.6	7.9	75.4	5.1	72.1	8.2
E	H	65	41	62.6	4.5	67.6	9.6	65.6	12.1	73.5	5.1	69.7	5.5
N	H	143	89	61.2	5.8	63.3	10.7	55.3	14.9	58.1	6.6	59.1	12.3
Q	H	119	62	58.0	6.4	63.1	9.6	53.3	14.1	55.8	6.5	56.7	10.9
H	H	67	23	58.2	6.0	57.3	6.9	39.2	4.2	43.5	7.3	46.3	13.8
K	H	76	15	39.9	5.7	39.1	14.9	19.4	2.0	23.4	4.9	23.7	5.4
R	H	73	17	41.9	6.6	38.7	10.5	17.5	1.3	23.9	4.0	22.1	7.3
G	K	80	28	31.3	4.6	31.5	8.9	21.2	2.9	23.1	5.2	24.7	8.1
A	K	56	16	38.2	3.9	42.0	10.0	31.4	7.4	33.1	7.3	37.1	11.3
S	K	97	46	42.5	6.0	48.1	11.2	33.6	6.3	36.0	4.9	38.2	10.4
T	K	72	17	44.4	4.7	51.1	10.6	35.5	6.6	38.0	6.0	42.5	11.6
V	K	115	95	47.1	6.4	54.3	11.7	39.2	7.8	40.6	5.9	45.9	13.2

L	K	105	53	55.5	3.8	60.2	10.9	42.0	9.1	43.2	7.5	48.9	13.9
I	K	73	15	51.2	6.1	56.8	10.6	41.7	8.3	42.5	6.5	47.4	12.9
M	K	48	11	23.4	3.0	27.7	5.2	15.6	2.3	20.7	5.8	20.1	5.9
P	K	87	11	33.6	3.3	38.8	9.9	22.3	2.5	23.7	3.2	26.5	7.3
F	K	93	33	55.8	4.8	56.9	12.0	38.1	8.7	41.0	6.1	43.4	12.4
Y	K	71	27	56.5	5.6	56.9	10.6	38.7	6.5	42.0	6.5	44.8	11.9
W	K	29	10	16.4	5.1	23.0	11.4	16.6	2.9	18.7	1.6	21.1	7.6
D	K	106	25	43.1	7.1	54.5	16.4	43.9	10.3	49.4	4.3	48.0	10.8
E	K	88	29	40.8	6.3	49.7	8.8	42.9	14.4	46.2	4.5	49.5	11.3
N	K	78	8	42.8	5.8	49.9	10.4	34.0	9.8	34.3	7.2	42.8	15.5
Q	K	118	45	41.4	6.5	47.9	10.8	30.9	10.3	32.8	7.3	40.4	15.0
H	K	76	21	40.0	4.7	45.1	11.0	21.2	0.9	27.1	6.7	34.0	15.3
K	K	75	25	35.0	15.6	41.0	23.7	12.6	0.2	16.0	3.8	18.4	8.6
R	K	76	27	28.0	4.1	29.6	16.2	10.7	1.8	13.9	5.2	19.7	15.4
G	R	79	21	43.8	5.4	37.4	7.4	23.4	9.1	24.3	8.1	27.9	12.0
A	R	99	30	49.1	5.6	48.7	9.1	33.4	13.5	33.5	7.2	43.8	20.5
S	R	74	45	50.8	7.3	48.0	12.0	30.4	7.9	35.5	6.3	44.2	18.8
T	R	65	13	50.6	8.2	53.1	8.8	36.5	11.6	40.2	6.6	43.2	13.6
V	R	78	17	53.9	8.2	53.2	10.9	36.7	10.2	39.2	6.3	43.4	14.4
L	R	3	2	50.4	16.2	46.7	9.0	50.8	4.2	47.0	4.7	48.4	4.0
I	R	80	35	64.6	15.4	64.8	12.3	36.5	11.5	39.0	7.3	43.1	13.6
M	R	50	22	35.3	6.5	32.5	7.2	17.5	2.7	22.3	5.3	21.6	6.3
P	R	70	21	56.3	7.3	50.1	9.6	25.9	8.0	28.4	6.2	32.1	11.3
F	R	87	35	53.5	3.6	52.2	11.8	31.7	8.1	35.1	5.1	37.2	10.7
Y	R	76	12	53.4	3.9	51.2	13.8	32.0	6.0	34.7	4.7	37.3	10.6
W	R	31	12	22.8	9.0	20.4	10.2	13.7	3.1	14.7	1.4	14.2	4.6

D	R	79	53	58.0	10.1	57.8	13.8	56.4	10.4	62.2	4.7	62.5	12.7
E	R	77	23	57.7	6.8	60.0	9.0	54.2	11.4	61.7	4.9	62.7	13.1
N	R	91	29	52.9	7.6	53.6	12.4	32.1	6.8	36.7	5.3	39.5	12.0
Q	R	82	37	51.0	0.5	50.6	11.9	31.8	8.9	36.5	6.5	40.2	13.4
H	R	85	32	55.0	6.5	47.3	14.2	22.1	4.1	28.0	5.3	29.5	11.3
K	R	69	33	35.3	6.6	30.8	16.5	9.9	1.2	13.7	3.4	13.8	5.8
R	R	45	18	40.3	7.3	38.1	14.2	16.6	6.6	18.8	5.9	21.6	10.4

Table S-4. Predictive power of amino acid physical properties on K-array. 39 amino acid physical properties are used to predict pep- tide S/N in Ac-GRK^{ac}XZC SAMDI peptide arrays. Properties are sorted based on Q^2 predictive power of both X- and Z-position amino acids. The individual positions have correlated Q^2's of r=0.924 (0 for randomized data). Random forest with 1000 trees and 5-folds cross validation was used. Hydrophilic, steric, and electronic properties are shown in green, red, and blue, respectively.

		Q^2 of individual properties on K-array		
Property type	Physical properties	X-position	Z-position	Both positions
Steric	Substituent van der Waals volume	0.224	0.204	0.568
Steric	Average volume of buried residue	0.251	0.189	0.565
Steric	STERIMOL length of the side chain	0.234	0.204	0.563
Steric	Radius of gyration of side chain	0.238	0.214	0.535
Hydrophobic	Solvation free energy	0.228	0.173	0.525
Steric	Refractivity	0.234	0.209	0.517
Steric	Distance between C-alpha and centroid of side chain	0.239	0.203	0.513
Steric	Normalized van der Waals volume	0.238	0.199	0.508
Steric	STERIMOL maximum width of the side chain	0.252	0.239	0.479
Steric	van der Waals parameter epsilon	0.244	0.171	0.465
Steric	Average accessible surface area	0.248	0.208	0.462
Electronic	Isoelectric point	0.242	0.209	0.454
Hydrophobic	Retention coefficient in HPLC pH 2.1	0.227	0.221	0.445
Steric	Residue accessible surface area in tripeptide	0.233	0.211	0.427
Steric	side chain torsion angle phi	0.237	0.174	0.395
Electronic	Electron-ion interaction potential values	0.236	0.187	0.383
Hydrophobic	Hydration number	0.227	0.225	0.383
Hydrophobic	Partition coefficient in thin-layer chromatography	0.215	0.172	0.381
Electronic	aNH chemical shifts	0.244	0.221	0.375
Hydrophobic	Melting point	0.233	0.213	0.375
Steric	Graph shape index	0.240	0.212	0.366
Electronic	pKCOOH(COOH on C_alpha)	0.212	0.162	0.363
Hydrophobic	Free energy of solution in water	0.214	0.189	0.361
Electronic	Polarity	0.177	0.167	0.360
Hydrophobic	Retention coefficient in HPLC pH 7.4	0.153	0.185	0.333
Steric	Side-chain angle theta	0.173	0.217	0.326
Electronic	Nuclear magnetic resonance (NMR) chemical shift of αcarbon	0.241	0.201	0.325
Electronic	Amphiphilicity index	0.114	0.194	0.314
Electronic	aCH chemical shifts	0.247	0.239	0.304
Steric	van der Waals parameter Ro	0.156	0.107	0.291
Electronic	pKNH2(NH2 on C_alpha)	0.198	0.181	0.230
Electronic	Net charge	0.040	0.110	0.170
Electronic	A parameter of charge transfer donor capability	0.080	0.088	0.144
Electronic	A parameter of charge transfer capability	0.059	0.083	0.131
Steric	STERIMOL minimum width of the side chain	0.016	0.046	0.086
Hydrophobic	Number of hydrogen-bond donors	0.027	0.056	0.085
Electronic	Positive charge	0.019	0.044	0.063
Electronic	Negative charge	0.009	0.024	0.038
Electronic	Localized electrical effect	0.029	0.013	0.016

Table S-5. Predictive power of amino acid physical properties on H-array. 39 amino acid physical properties are used to predict mean peptide S/N in the Ac-GXZHC SAMDI peptide array. Properties are sorted based on Q^2 predictive power of both X- and Z-position amino acids. The individual positions have correlated Q^2's of r=0.942 (0 for randomized data). Random forest with 1000 trees and 5-folds cross validation was used. Hydrophilic, steric, and electronic properties are shown in green, red, and blue, respectively.

		Q^2 of individual properties on H-array		
Property type	Physical properties	X-position	Z-position	Both positions
Hydrophobic	Retention coefficient in HPLC pH 2.1	0.156	0.280	0.544
Hydrophobic	Solvation free energy	0.167	0.292	0.535
Steric	Average accessible surface area	0.168	0.291	0.501
Electronic	Isoelectric point	0.139	0.304	0.479
Hydrophobic	Hydration number	0.132	0.219	0.450
Hydrophobic	Free energy of solution in water	0.154	0.303	0.450
Steric	Substituent van der Waals volume	0.139	0.274	0.446
Electronic	Polarity	0.172	0.257	0.436
Hydrophobic	Retention coefficient in HPLC pH 7.4	0.153	0.275	0.431
Steric	Average volume of buried residue	0.147	0.310	0.407
Electronic	Nuclear magnetic resonance (NMR) chemical shift of αcarbon	0.151	0.315	0.406
Steric	Radius of gyration of side chain	0.119	0.290	0.400
Hydrophobic	Partition coefficient in thin-layer chromatography	0.164	0.313	0.396
Hydrophobic	Melting point	0.146	0.256	0.381
Electronic	Electron-ion interaction potential values	0.134	0.290	0.377
Steric	Distance between C-alpha and centroid of side chain	0.136	0.287	0.376
Electronic	pKCOOH(COOH on C_alpha)	0.144	0.287	0.370
Steric	Refractivity	0.143	0.260	0.367
Steric	STERIMOL length of the side chain	0.164	0.262	0.361
Steric	Residue accessible surface area in tripeptide	0.157	0.309	0.359
Electronic	Localized electrical effect	0.149	0.204	0.356
Electronic	αCH chemical shifts	0.129	0.244	0.350
Steric	STERIMOL maximum width of the side chain	0.119	0.302	0.348
Steric	Normalized van der Waals volume	0.148	0.297	0.346
Electronic	aNH chemical shifts	0.173	0.289	0.342
Electronic	Net charge	0.131	0.193	0.323
Steric	side chain torsion angle phi	0.135	0.262	0.288
Steric	van der Waals parameter epsilon	0.086	0.155	0.282
Electronic	Negative charge	0.119	0.173	0.279
Steric	Graph shape index	0.093	0.183	0.232
Steric	van der Waals parameter Ro	0.070	0.184	0.220
Electronic	pKNH2(NH2 on C_alpha)	0.085	0.245	0.206
Steric	Side-chain angle theta	0.075	0.176	0.182
Electronic	Amphiphilicity index	0.092	0.099	0.168
Hydrophobic	Number of hydrogen-bond donors	0.099	0.069	0.157
Electronic	A parameter of charge transfer capability	0.029	0.081	0.104
Electronic	A parameter of charge transfer donor capability	-0.010	0.011	0.022
Steric	STERIMOL minimum width of the side chain	0.003	-0.021	0.021
Electronic	Positive charge	-0.011	0.003	-0.009