

Università degli Studi di Milano – Bicocca  
Facoltà di Medicina e Chirurgia



Dottorato in Tecnologie Biomediche  
XXVI Ciclo

**SAMDI Mass Spectrometry**  
**For High Yield Protein Modification**  
**Reaction Development**

**Studente: Coghi Maria Donata**

**Matricola: 068320**

**Tutor: Prof. Mantegazza Francesco**

to Dado & Giulia

## Abstract

Efficient chemical strategies that attach synthetic molecules to desired positions on protein surfaces are useful tools in the field of chemical biology and represent one major prerequisite for the development of new drugs and materials. Protein modification with polyethylene glycol (PEG) groups is indeed routinely performed on therapeutic proteins to improve serum half-life, or even cytotoxins or imaging agents are efficiently conjugated to cancer-targeting elements. In a typical approach, a synthetic functional group of interest is attached to a uniquely reactive amino acid group introduced by recombinant methods. Most bioconjugation reactions, however, do not reach full conversion. Therefore the development of a straightforward and reliable method to increase the extent of conversion into bioconjugates would be very helpful. In this perspective, we developed a generalizable combinatorial peptide library screening platform suitable for the identification of sequences displaying high levels of reactivity toward a desired bioconjugation reaction. This was achieved by using SAMDI MS technique (Self-Assembled Monolayer and Desorption/Ionization Mass Spectrometry) as a new, efficient and simple method for the evaluation of highly reactive amino acid motifs. The bioconjugation reaction we selected is the oxidative modification of electron-rich tyrosine residues performed using cerium(IV) ammonium nitrate (CAN) as oxidant reagent. The peptides were identified on a 361-member hexapeptide array, wherein the two N- and C-terminal residues to the target residue were varied. The arrays were prepared by immobilizing the peptides to a self-assembled monolayer of alkanethiolates on gold and could therefore be analyzed by mass spectrometry. We found that the most reactive peptides had either a serine N-terminal to the tyrosine residue or another tyrosine in proximity of the reactive site. Conversely, peptides displaying the lowest conversion level contained a positive charged residue: histidine, lysine or arginine, where the lowest relative activity was reached with arginine and leucine as C- and N-terminal residues, respectively. This study provides an important example of how synthetic peptide libraries can accelerate the discovery and optimization of protein bioconjugation strategies.

## Abbreviations

<b>aaRSs</b>	aminoacyl-tRNA synthetases
<b>ADCs</b>	antibody-drug conjugates
<b>Aha</b>	azidohomoalanine
<b>Anisidine derivative</b>	4-methoxy- <i>N,N</i> -dimethylaniline
<b>AUCs</b>	area under curves
<b>CAN</b>	cerium(IV) ammonium nitrate
<b>Codon UAG</b>	amber non sense stop codon
<b>CuAAC</b>	Copper-catalyzed azide-alkyne cycloaddition
<b>DCM</b>	dichloromethane
<b>DE</b>	delayed extraction in mass spectrometry
<b>DHAP</b>	2,6-dihydroxyacetophenone
<b>DIC</b>	diisopropylcarbodiimide
<b>DIUF</b>	deionized ultrafiltered water
<b>DMF</b>	dimethylformamide
<b>EDT</b>	ethanedithiol
<b>ESI</b>	electrospray ionization
<b>Fmoc</b>	9-fluorenylmethoxy- carbonyl
<b>Hag</b>	homoallylglycine
<b>HOBt</b>	hydroxybenzotriazole
<b>Hpg</b>	homopropargylglycine
<b>KE</b>	kinetic energy
<b>MALDI</b>	matrix assisted laser desorption/ionization
<b>MS</b>	mass spectrometry
<b>NMM</b>	N-methylmorpholine
<b><i>p</i>AcF</b>	<i>p</i> -acetylphenylalanine
<b>PEG</b>	polyethylene glycol
<b>PyBop</b>	benzotriazol-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate
<b>SAMDl MS</b>	Self-Assembled Monolayer and Desorption/Ionization Mass Spectrometry
<b>SAMs</b>	self-assembled monolayers
<b>TES</b>	triethylsilane
<b>TFA</b>	trifluoroacetic acid
<b>THAP</b>	2,4,6-trihydroxyacetophenone

<b>TOF</b>	time-of-flight
<b>UAAs</b>	unnatural amino acids

## Table of Contents

Abstract. ....	i
Abbreviations. ....	ii
Table of Contents. ....	iv
List of Figures and Tables .....	vi
1. Introduction. ....	1
1.1. Protein Modification. ....	2
1.1.1. Nonspecific labeling. ....	3
1.1.2. Site selective modification. ....	4
1.1.2.1. Cysteine residues. ....	4
1.1.2.2. Artificial amino acids. ....	5
1.2. SAMDI MS technique. ....	14
1.2.1. Matrix Assisted Laser Desorption/Ionization Time-Of-Flight Mass Spectrometry..	14
1.2.1.1. MALDI MS principles .....	14
1.2.1.2. MALDI-TOF-MS applications .....	17
1.2.2. Self Assembled Monolayers. ....	20
1.2.3. SAMDI MS applications. ....	23
1.2.3.1. Enzyme activity. ....	24
1.2.3.2. Chemical reaction discovery .....	27
1.2.3.3. High Throughput Assays and drug discovery .....	28
1.3. Aim of the work. ....	31
2. Materials and Methods. ....	32
2.1. Ac-GYGC, Ac-GYGGC and Ac-GGYGGC peptides. ....	33
2.2. Ac-GGYXGC and Ac-GXYGGC peptide libraries .....	36
2.3. Ac-GXYZGC peptide library. ....	38
3. Results. ....	40
3.1. Development of reactions for protein bioconjugation. ....	41
3.2. Ac-GYGC. ....	43
3.2.1. Solid phase synthesis and immobilization. ....	43
3.2.2. Immobilized Ac-GYGC oxidative modification. ....	43
3.3. Ac-GYGGC and Ac-GGYGGC synthesis, immobilization and modification. ....	45

3.3.1. Ac-GGYGGC biotinylation, oxidative modification and subsequent immobilization. .....	47
3.3.2. Biotinylated Ac-GRKFGC and Ac-GRK <sup>Ac</sup> FGC control. ....	49
3.3.3. Revised modification procedure for Ac-GYGGC and Ac-GGYGGC modification. .....	50
3.4. Ac-GGYXGC and Ac-GXYGGC peptide library synthesis and modification.....	53
3.5. Ac-GXYZGC library synthesis and screening .....	55
4. Discussion. ....	57
References.....	63
Supplementary Materials. ....	76
Acknowledgments.....	95

## List of Figures and Tables

Figure 1.1. Classic bioconjugation reactions for the modification of Lys residues. ....	3
Figure 1.2. Fluorescein isothiocyanate (FITC) conjugation reaction scheme. ....	4
Figure 1.3. Classic and modern method to modify Cys residues. ....	6
Figure 1.4. Incorporation of unnatural amino acids (UAAs) into proteins. ....	9
Figure 1.5. Selected amino acids site-specifically incorporated into proteins in vivo.....	10
Figure 1.6. Protein conjugates synthesized with unnatural amino acids and examples of their applications. ....	13
Table 1.1. Commonly used matrices and their main properties and applications. ....	15
Figure 1.7. Principle of matrix-assisted laser desorption/ionization mass spectrometry with TOF and TOF reflectron.....	17
Figure 1.8. MALDI-MS for protein analysis. ....	19
Figure 1.9. Strategy for the analysis of purified proteins by MALDI-MS. ....	19
Figure 1.10. SAMs formation and functionalization. ....	22
Figure 1.11. SAMDI MS principle. ....	23
Figure 1.12. Examples of enzyme activity assays performed with self-assembled monolayers for matrix-assisted laser desorption/ionization mass spectrometry (SAMDI MS). ....	25
Figure 1.13. Chemical reaction study by SAMDI MS. ....	28
Figure 1.14. High Throughput Assays and drug discovery with SAMDI MS.....	30
Figure 3.1. Tyrosine modification reaction scheme.....	41
Figure 3.2. Proposed mechanisms for the formation of the observed products.....	42
Figure 3.3 Immobilized Ac-GYGC oxidative modification scheme. ....	43
Figure 3.4. Representative SAMDI spectra of immobilized Ac-GGYGGC.....	44
Figure 3.5. Immobilized Ac-GGYGGC oxidative modification scheme. ....	45
Figure 3.6. Representative SAMDI spectra of immobilized Ac-GGYGGC.....	46
Figure 3.7. MALDI spectra of 0.1 mM biotinylated Ac-GGYGGC peptide.....	48
Figure 3.8. SAMDI spectra obtained after the reacted peptide immobilization. ....	49
Figure 3.9. MALDI spectra of biotinylated Ac-GRKFGC and Ac-GRKAcFGC. ....	49
Figure 3.10. MALDI spectra of Ac-GYGGC. ....	50-51
Figure 3.11. MALDI spectra of Ac-GGYGGC. ....	52
Figure 3.12. Relative conversion activities of Ac-GGYXGC and Ac-GXYGGC libraries. ....	53
Figure 3.13. MALDI spectra of Ac-GGYSGC and Ac-GSYGGC.....	54



Figure 3.14. Sequence dependence determined by SAMDI for CAN-mediated tyrosine modification. .....	55
Table 3.1. Relative modification activities for Ac-GXYZGC library..	56
Figure S.1. MALDI spectra of Ac-GGYAGC and Ac-GAYGGC.	77
Figure S.2. MALDI spectra of Ac-GGYDGC and Ac-GDYGGC.	78
Figure S.3. MALDI spectra of Ac-GGYEGC and Ac-GEYGGC.	79
Figure S.4. MALDI spectra of Ac-GGYFGC and Ac-GFYGGC.	80
Figure S.5. MALDI spectra of Ac-GGYGGC and Ac-GGYGGC.	81
Figure S.6. MALDI spectra of Ac-GGYHGC and Ac-GHYGGC.	82
Figure S.7. MALDI spectra of Ac-GGYIGC and Ac-GIYGGC.	83
Figure S.8. MALDI spectra of Ac-GGYKGC and Ac-GKYGGC.	84
Figure S.9. MALDI spectra of Ac-GGYLGC and Ac-GLYGGC.	85
Figure S.10. MALDI spectra of Ac-GGYMGC and Ac-GMYGGC.	86
Figure S.11. MALDI spectra of Ac-GGYNGC and Ac-GNYGGC.	87
Figure S.12. MALDI spectra of Ac-GGYPGC and Ac-GPYGGC.	88
Figure S.13. MALDI spectra of Ac-GGYQGC and Ac-GQYGGC.	89
Figure S.14. MALDI spectra of Ac-GGYRGC and Ac-GRYGGC.	90
Figure S.15. MALDI spectra of Ac-GGYTGC and Ac-GTYGGC.	91
Figure S.16. MALDI spectra of Ac-GGYVGC and Ac-GVYGGC.	92
Figure S.17. MALDI spectra of Ac-GGYWGC and Ac-GWYGGC.	93
Figure S.18. MALDI spectra of Ac-GGYYGc and Ac-GYYGGC.	94

# 1. Introduction

The need to track protein localization, study their behavior in a variety of contexts and develop new drugs and materials, strongly demands for efficient chemical strategies to attach synthetic molecules to desired positions on protein surfaces.

Protein bioconjugates show varying levels of complexity: they can involve the simple and nonspecific attachment of a fluorescent dye to a protein or even tethering of highly functionalized drug molecules to specific antibodies (Stephanopoulos and Francis, 2011).

In a typical approach, when site selectivity is required, proteins are expressed by recombinant methods as they allow the introduction of a uniquely reactive amino acid group. Then, a synthetic functional group of interest can be attached to a specific amino acid side chain using one out of a relatively limited set of chemical reactions that can modify unprotected biomolecules in aqueous solution (Witus et al., 2010). Most bioconjugation reactions, however, do not reach full conversion even when this can be achieved, and the addition of excess reagent (in order to try to optimize the conversion level) can lead to overmodification and erosion of chemospecificity (Stephanopoulos and Francis, 2011). It stands to reason that finding a straightforward and reliable method to increase the extent of conversion into bioconjugates would be very useful.

In this perspective we developed a generalizable combinatorial peptide library screening platform using SAMDI MS technique (Self-Assembled Monolayer and Desorption/Ionization Mass Spectrometry) as a new, efficient and simple method for the identification of sequences that display high levels of reactivity toward a desired bioconjugation reaction.

This work was done in collaboration with Dr. Milan Mrksich's group at Northwestern University.

### **1.1. Protein Modification**

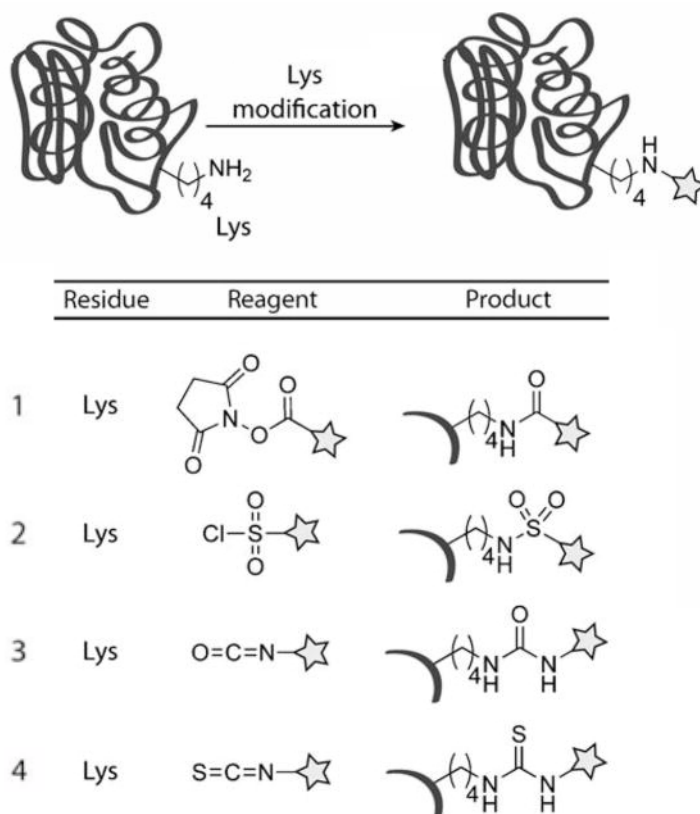
During the past decades biologists and chemists have combined and shared their efforts to develop protein modification methods in order to study proteins in their native settings. The invention of genetically encoded fluorescent proteins such as the green fluorescent protein, GFP, likely represents the clearest example of innovation that this cooperation can bring and the importance it can have (its widespread use and impact was indeed recognized with the 2008 Nobel Prize in Chemistry).

Proteins were initially selected as the best biomolecules to be targeted because of their numerous side-chain functionalities, complex tertiary structures, and diverse biological functions. More recently these methods have become important also for the biotechnology industry: for instance, the modification of proteins with polyethylene glycol (PEG) groups (PEGylation) is

routinely performed on therapeutic proteins to improve serum half-life, the conjugation of cytotoxins or imaging agents to cancer-targeting elements, such as monoclonal antibodies has become usual as well (Sletten and Bertozzi, 2009).

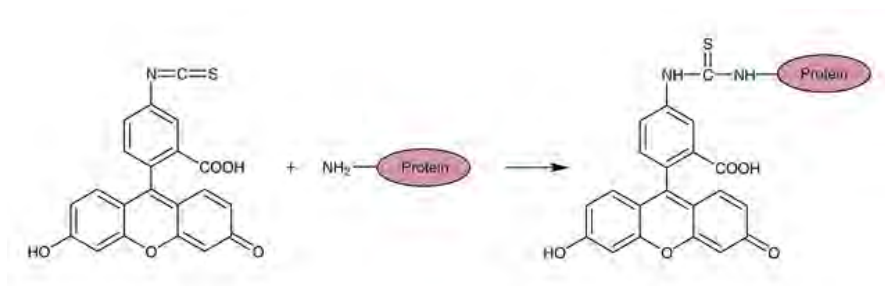
### 1.1.1. Nonspecific labeling

Nonspecific labeling exploits protein bioconjugation strategies that primarily target the functionalities present in the abundant amino acid side chains such as lysine, aspartic and glutamic acids. Although glutamate and aspartate residues have been used for bioconjugation through coupling with amines via carbodiimides, the potential cross-linking of proteins represents the limit of this technique. Lysine residues on the other hand are certainly the most popular targets because of the abundance of methods available to selectively modify primary amines. Lysine can react with activated esters, sulfonyl chlorides, isocyanates, or isothiocyanates to afford amides, sulfonamides, ureas, or thioureas, respectively (Figures 1.1, entries 1-4); moreover, it can undergo reductive amination reactions with aldehydes and reducing agents (Sletten and Bertozzi, 2009). In addition, a large number of *N*-hydroxysuccinimide (NHS) esters and isothiocyanates are available from commercial suppliers (Stephanopoulos and Francis, 2011). Furthermore all the reactions that can affect lysine residues can also be used to modify the N termini of proteins. One of the most



**Figure 1.1.** Classic bioconjugation reactions for the modification of Lys residues. Lys residues can be modified through amide, sulfonamide, urea, and thiourea formation with *N*-hydroxysuccinimide-activated esters, sulfonyl chlorides, isocyanates, and isothiocyanates (entries 1-4, respectively).

common applications that take advantage of this binding capability is the FITC-labeling process. Fluorescein isothiocyanate (FITC), is a widely used fluorophore, popular because of its high quantum efficiency and stability when conjugated. The conjugation occurs through free amino groups of proteins (or peptides) forming a stable thiourea bond (Figure 1.2). FITC conjugates of antibodies, lectins, hormones, and growth factors are regularly used in a variety of immunohistochemical and flow cytometry applications.



**Figure 1.2.** Fluorescein isothiocyanate (FITC) conjugation reaction scheme.

These nonspecific labeling approaches are particularly general because of the very high prevalence of lysine groups on nearly all protein surfaces, which often offer several sites for the attachment of dyes, biotin, polymers, nanoparticles and surfaces. However, with this labeling generality it is often difficult to control the number and locations of the modifications introduced (Tilley et al., 2008).

### 1.1.2. Site selective modification

When site specificity is required, it is often difficult and time consuming to find a method that can install a single copy of a new functional group in a desired location on a wild-type protein surface (Stephanopoulos and Francis, 2011). For this reason, the best choice is usually to introduce a uniquely reactive amino acid group through site-directed mutagenesis during protein expression.

#### 1.1.2.1. Cysteine residues

The most commonly modified functionality is the thiol group of cysteine residues whose modification is particularly effective due to the low abundance and nucleophilic nature of reduced thiolates on protein surfaces (Seim et al., 2011). The sulfhydryl group can undergo disulfide exchange to form mixed disulfides (Figure 1.3, entry 1) as well as alkylation with alkyl halides or Michael addition with maleimides to yield thioethers (Figure 1.3, entries 2 and 3). Furthermore, over the past few years some next-generation cysteine modification reactions have been developed. For instance, Davis and co-workers developed a two-step method for cysteine modification (Figure 1.3, entry 4). In the first step the cysteine is treated with O-mesitylenesulfonylhydroxylamine under

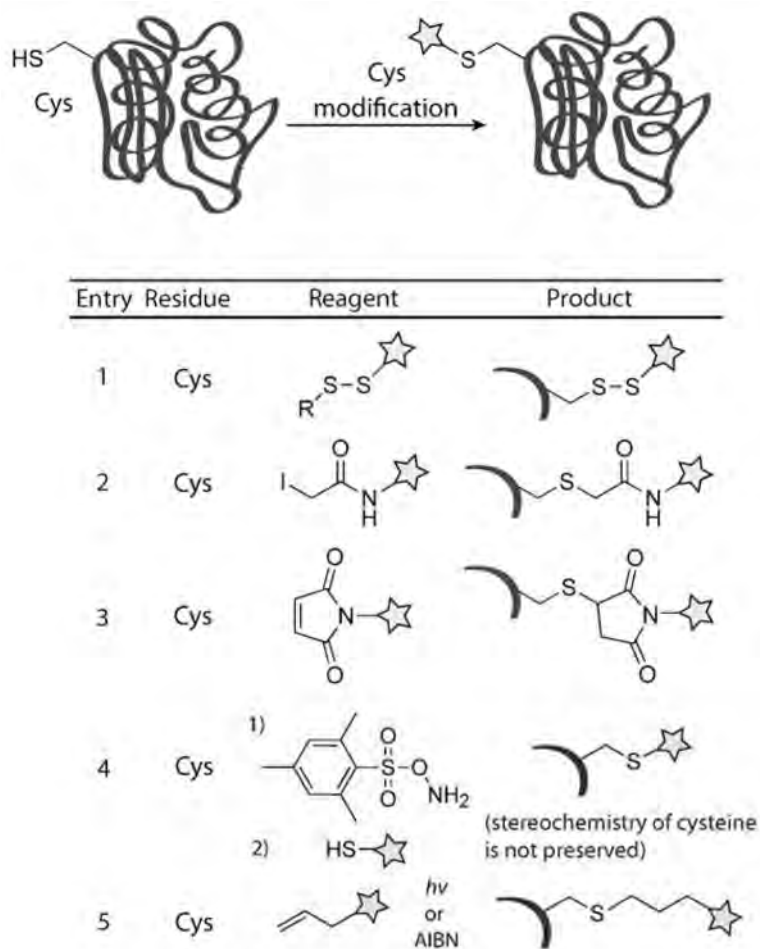
basic conditions and it turns into dehydroalanine. The dehydroalanine residues then undergo a Michael addition with thiol reagents to yield a thioether linkage. The Michael addition is not stereospecific and, thus, a diastereomeric mixture of modified proteins is produced (Bernardes et al., 2008). Another emerging technique for the modification of cysteine that yields a thioether linkage is thiol-ene chemistry (Dondoni, 2008), which consists of a radical-based mechanism that leads to the addition of a thiol across an alkene (Figure 1.3, entry 5) (Triola et al., 2008). The radical species can be generated by standard radical initiators or by irradiation with light (Dondoni, 2008), with the latter method displaying greater functional-group tolerance and shorter reaction times (Campos et al., 2008). Thiolene reactions have been performed on proteins functionalized with either thiols (Wittrock et al., 2007) or alkenes (Jonkheijm et al., 2008), although their use in the direct modification of cysteine residues has yet to be reported, possibly because generation of a protein-associated thiyl radical could lead to unwanted side reactions.

#### 1.1.2.2. Artificial amino acids

The site-specific (Xie and Schultz, 2006) or **residue-specific** (Link et al., 2003) incorporation of artificial amino acids provides a particularly compelling way to solve bioconjugation challenges, as these methods allow uniquely reactive chemical functional groups to be introduced into the protein sequence (Stephanopoulos and Francis, 2011).

The straightforward method of introducing chemical reporters into cellular proteins, pioneered by Tirrell and co-workers, consists in simply administering to cells an unnatural amino acid that is tolerated by the translational machinery, particularly the aminoacyl-tRNA synthetases (aaRSs; Figure 1.4 A) (Sletten and Bertozzi, 2009). The concept dates back to the 1950s, when methionine residues were shown to be replaced by their selenium analogues after the addition of selenomethionine to methionine-depleted growth media (Cowie and Cohen, 1957). We now know that surrogates for methionine, leucine, tryptophan, or phenylalanine can be incorporated into proteins expressed in *E. coli*, although reports can be found for replacement of almost any amino acid with an unnatural derivative. The yields are optimal when the *E. coli* strain is rendered auxotrophic for the amino acid being targeted for replacement, and over-expression of the required aaRS can be helpful as well.

By using this method, a variety of bioorthogonal functional groups have been incorporated into proteins, both in *E. coli* and in mammalian cells. For example, the methionine surrogates



**Figure 1.3.** Classic and modern method to modify Cys residues. Cys residues can be modified through disulfide exchange, alkylation with iodoacetamide reagents, and Michael addition with maleimides (entries 1–3, respectively). More recently Cys has been modified also through a two-step labeling procedure which involves formation of dehydroalanine and subsequent Michael addition of a thiol (entry 4), or the photochemically promoted thiol-ene reaction (entry 5; AIBN = 2,2'-azobisisobutyronitrile).

homopropargylglycine (Figure 1.5, 42, Hpg), homoallylglycine (43, Hag), and azidohomoalanine (44, Aha) were used to introduce alkynes, alkenes, and azides, respectively, in proteins with good efficiency (Figure 1.5) (van Hest et al., 2000; Link et al., 2004). Hpg was employed to label newly synthesized proteins with an azidocoumarin dye by Copper-catalyzed azide-alkyne cycloaddition (CuAAC) in bacterial (Beatty et al., 2005) and mammalian (Beatty et al., 2006) systems. Similarly, Aha has been used to interrogate newly synthesized proteins by labeling with CuAAC, (Link and Tirrell, 2003) Staudinger ligation, (Kiick et al., 2002) and, more recently, with cyclooctyne probes (Link et al., 2006). The method has also been applied to proteomic analysis of newly synthesized proteins, a process termed bioorthogonal noncanonical amino acid tagging (BONCAT) (Dieterich et al., 2006). Moreover, Aha and Hpg have been used together to image two distinct protein populations simultaneously (Beatty and Tirrell, 2008). Azides and alkynes have also been installed within viruslike capsids by replacing methionine residues with Aha and Hpg. (Strable et al., 2008)

Some unnatural amino acids are too structurally dissimilar from their native relatives for recognition by natural aaRS enzymes. In such cases, the aaRSs can be mutated to accept the unnatural substrate, either by rational structure-based (Ibba et al., 1994; Kast and Hennecke, 1991) or selection-based methods (Kothakota et al., 1995). Such strategies have been used to obtain a mutant phenylalanine tRNA synthetase (ePheRS) (Kast and Hennecke, 1991) and incorporate a ketone-bearing amino acid para-acetyl-phenylalanine (Figure 1.5, 45) into proteins (Datta et al., 2002; Kast and Hennecke, 1991). The same ePheRS has also been used to incorporate para-azido-, para-bromo-, and para-iodophenyl- alanines into proteins (Kirshenbaum et al., 2002; Sharma et al., 2000). The azide allowed for modification with phosphines or alkynes, while the halogen derivatives were modified by using palladium-catalyzed cross-coupling methods (Kodama et al., 2006, 2007).

**Site-specific** introduction of unnatural amino acids into proteins was first reported more than 20 years ago as an *in vitro* technique (Bain et al., 1989; Noren et al., 1989) and it has since been extended, primarily through the work of Schultz and co-workers, to *in vivo* applications. (Wang and Schultz, 2005) This method utilizes the codon UAG (the amber nonsense stop codon), which normally directs termination of protein synthesis, to instead encode an unnatural amino acid loaded onto a complementary tRNA. The tolerance of the ribosome for unnatural amino acids allows for incorporation into proteins during normal translation (Figure 1.4 B, C).

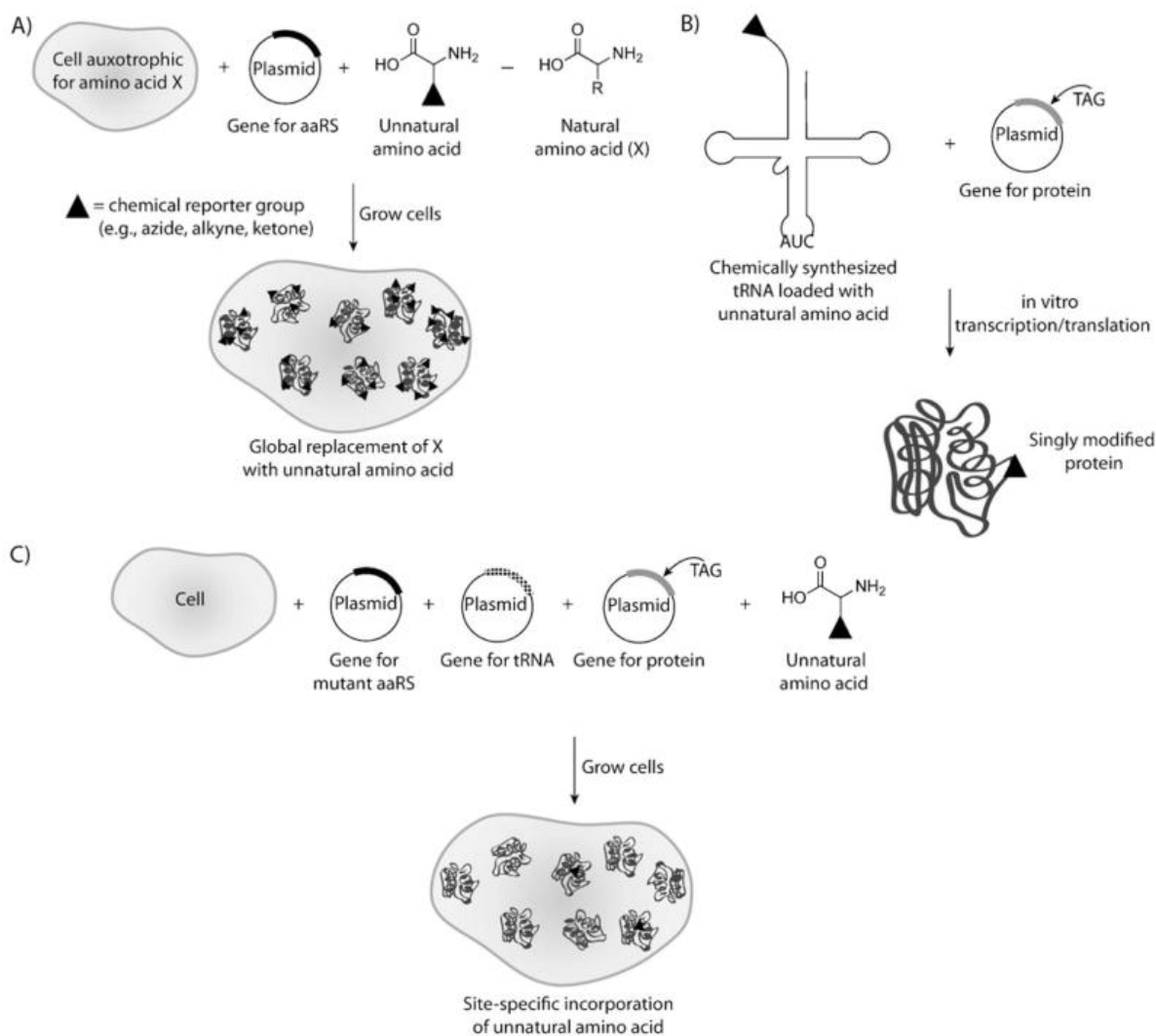
To integrate unnatural amino acids into proteins in an *in vitro* setting (Figure 1.4 B), the gene encoding the protein of interest is first mutated so that the amber stop codon is situated at the



desired modification site, and all other amber stop codons are removed from the sequence. A tRNA is synthesized that contains a complementary anticodon as well as a covalently ligated unnatural amino acid at the 3' end. The addition of this artificial tRNA and the gene encoding the mutated protein to an *E. coli* in vitro transcription/translation system generates the modified protein (Noren et al., 1989). Many unnatural amino acids have been incorporated into proteins by using this technique; however, it suffers from low yields, and considerable labor is involved in synthesizing the aminoacylated tRNA (Ellman et al., 1991). Despite these limitations, the in vitro technique has been used to study a variety of proteins. Residues in  $\alpha$  helices and  $\beta$  sheets were replaced with ester analogues to explore the contributions of backbone amide bonds in protein structure (Chapman et al., 1997; Ellman et al., 1992; Koh et al., 1997). In addition, the contribution of cation- $\pi$  interactions to protein stability was analyzed by using unnatural amino acids, (Ting et al., 1998). A notable extension of the system has been microinjection of engineered mRNA and aminoacylated tRNA into *Xenopus* oocytes (Dougherty, 2000, 2008). Dougherty, Lester, and co-workers have used this approach to study the mechanism of neuroreceptors through electrophysiology (Lummiss et al., 2005; Xiu et al., 2009), and to this end more than 100 unnatural amino acids (UAAs) have been incorporated into *Xenopus* ion channels. The in vitro method coupled with microinjection is ideally suited for these applications because of the very low levels of protein that can be detected using electrophysiology.

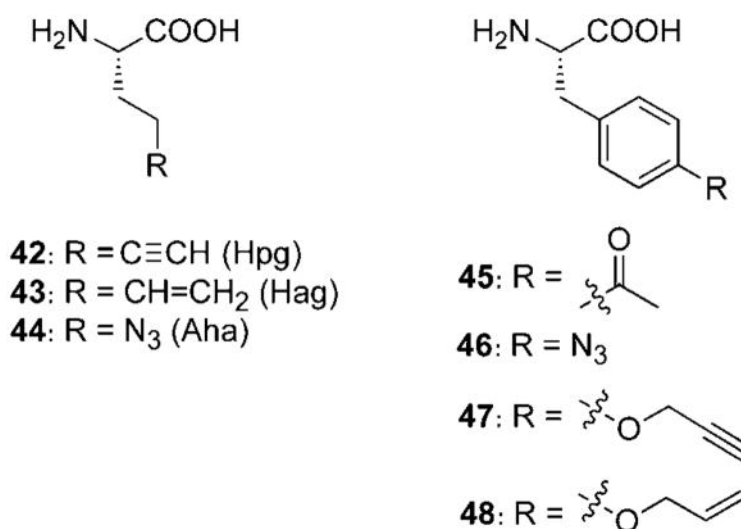
To overcome the limitations of the in vitro method, Schultz and co-workers have created an in vivo system for site-specific mutagenesis of unnatural amino acids (Figure 1.4 C). The breakthrough was the selection of orthogonal tRNA and aaRSs that recognized the amber stop codon and unnatural amino acid, respectively (Wang and Schultz, 2005). Expression of the corresponding genes in a heterologous host together with the gene encoding the desired protein with the amber mutation produced the modified protein. Typically, the cell-culture media were supplemented with the unnatural amino acid, but an *E. coli* strain engineered to produce para-aminophenylalanine was able to incorporate this "21st amino acid" by total biosynthesis (Mehl et al., 2003). The extension of in vivo unnatural amino acid mutagenesis to yeast (Chin et al., 2003; Young et al., 2009) and mammalian cells (Köhler et al., 2001; Liu et al., 2007; Ye et al., 2008) has also been achieved.

Dozens of unnatural amino acids have been incorporated into proteins by using the in vivo amber stop codon method; most are based on an aromatic (tyrosine or phenylalanine) core (Wang and Schultz, 2005). As depicted in Figure 1.5, amino acids containing azides (44, 46)



**Figure 1.4.** Incorporation of unnatural amino acids (UAAs) into proteins. A) Global incorporation of an unnatural amino acid using auxotrophic cell lines and promiscuous aminoacyl tRNA synthetases (aaRSs). B) Site-specific incorporation of an unnatural amino acid in vitro by using a transcription/translation system, a chemically synthesized tRNA loaded with the unnatural amino acid, and the gene for the protein of interest mutated at the site of modification. C) Site-specific incorporation of an unnatural amino acid in vivo using mutant tRNA and aaRS as well as the gene for the protein of interest mutated at the site of modification.

(Chin et al., 2002a) alkynes (42, 47) (Deiters and Schultz, 2005), ketones (45) (Wang et al., 2003a; Zhang et al., 2003) and alkenes (43, 48) (Zhang et al., 2002) have all been incorporated for further reaction with bioorthogonal chemical reporters. Other functional groups installed into proteins using this approach include anilines (Carrico et al., 2008; Santoro et al., 2002), aryl halides (Wang et al., 2003b), boronic acids (Brustad et al., 2008a), photoisomerizable and cross-linking groups



**Figure 1.5.** Selected amino acids site-specifically incorporated into proteins in vivo.

(Bose et al., 2006; Chin et al., 2002b), sometimes coupled with post-translational modifications (Liu and Schultz, 2006; Zhang et al., 2004) caged versions of amino acids to allow the masking of putatively important residues, (Deiters et al., 2006; Lemke et al., 2007; Wu et al., 2004) and even whole fluorophores (Summerer et al., 2006; Wang et al., 2006).

**Unnatural amino acid applications to protein therapeutics.** The ability to engineer proteins using UAAs provides for the first time medicinal chemistry-like control over the pharmacological and physical properties of protein therapeutics. More traditional methods like lysine or cysteine conjugation often form heterogeneous products with limited control over conjugation site and stoichiometry. Furthermore, it is likely that individual product species have distinct efficacy, safety, and pharmacokinetic properties that complicate the optimization of a therapeutic candidate. Some recent applications of genetically incorporated UAAs to the development of protein therapeutics are highlighted below (Figure 1.6).

One of the earliest applications involved site-specific polyethylene glycol (PEG) conjugation to extend the serum half-life of proteins (Figure 1.6a). The orthogonal chemical handle allows for the control of conjugation site and stoichiometry to maximize biological activity, serum half-life and

stability/solubility. For example, human growth hormone (Cho et al., 2011) and fibroblast growth factor 21 (Mu et al., 2012) were expressed with pAcF at defined positions and conjugated to an alkoxyamine derivatized PEG in 80–97% yield. The resulting conjugate showed excellent pharmacokinetics, minimal loss of biological activity, and no apparent immunogenicity in vivo.

Genetically encoded UAAs have also been used to synthesize antibody–drug conjugates (ADCs). ADCs preferentially deliver cytotoxic drugs to cells presenting tumor-associated antigens to achieve improved drug efficacy and safety. For example, brentuximab vedotin (Adcetris/ SGN-35; an anti-CD30 antibody conjugated with the microtubule-disrupting agent monomethyl auristatin E) has been recently approved for Hodgkin’s lymphoma (Doronina et al., 2006), and several other ADCs are currently in late-stage clinical trials. However, the first approved ADC, gemtuzumab ozogamicin (Mylotarg; an anti-CD33 antibody conjugated with the DNA cleaving agent calicheamicin) (Hamann et al., 2002), was pulled from the market in 2010 because of toxicity and lack of efficacy in larger trials, showing the need for further optimization of this class of drugs.

Typical synthetic methods for most ADCs involve modification of lysine or cysteine residues, which leads to a distribution of zero to eight toxins per antibody at various sites. The individual species within this mixture are likely to have distinct affinities, stabilities, pharmacokinetics, efficacies and safety profiles. A site-specific ADC with only two conjugation sites (generated by THIOMAB technology in which additional cysteines are introduced into the antibody for maleimide conjugation (Junutula et al., 2008)) showed similar efficacy to randomly labeled ADCs, but had an improved therapeutic index and better pharmacokinetic properties in rodents. This methodology, however, requires complicated reduction/oxidization steps and still has inherent cysteine–maleimide instability (Shen et al., 2012). An anti-Her2 IgG expressed with two genetically encoded pAcF residues was conjugated to an auristatin derivative via a stable oxime-linkage in excellent yield (Axup et al., 2012) (Figure 1.6b). This construct showed selective in vitro cytotoxicity ( $EC_{50}$  100–400 pM) against Her2<sup>+</sup> breast cancer cell lines as well as complete regression of Her2<sup>+</sup> mammary fat pad tumors in mouse xenograft models. Full-length IgGs containing pAcF can be expressed in mammalian cells at yields over 1 g/L, making this a commercially viable technology. The ability to synthesize homogeneous ADCs enables their medicinal chemistry-like optimization, and is currently being applied to the synthesis of antibodies conjugated to kinase and phosphatase inhibitors, nuclear hormone receptor agonists and antagonists, and other drug classes for treatment of autoimmune, cardiovascular, and metabolic disease.

Bispecific antibodies that can bind two different antigens simultaneously (on the same or

distinct proteins and/or cells) also have gained considerable attention as next-generation immunotherapeutics for cancer. Bispecific antibodies consisting of an antibody (or fragments thereof) specific for a tumor-associated antigen linked to an anti-human CD3 (cluster of differentiation 3) antibody have demonstrated impressive efficacy in the clinic (Baeuerle and Reinhardt, 2009). These engineered antibodies recruit CD8<sup>+</sup>/CD3<sup>+</sup> cytotoxic T lymphocytes to tumor cells and form a pseudo-immunological synapse that results in the activation of T lymphocytes and subsequent lysis of the target cells. A number of recombinant methods (including single chain variable fragment (scFv) and IgG-based formats) and some chemical crosslinking methods have been developed to generate bispecific antibodies (Graziano and Guptill, 2004; Jäger et al., 2009; Moore et al., 2011; Wu et al., 2007). However, genetically fused scFvs can have short serum half-lives and relatively poor stability, and some IgG-based quadroma formats, which consist of chimeric mouse IgG2a and rat IgG2b, can generate human anti-mouse (HAMA) or human anti-rat (HARA) antibody responses in patients (Chames and Baty, 2009). Moreover, these methods often lead to restricted (and less optimal) orientations of the fusion proteins (e.g., C-terminal to N-terminal fusions). Synthesis of bispecific antibodies using traditional chemical methods is challenging since one generates many different orientations of the two antigen binding sites, which may or may not form productive immunological synapses.

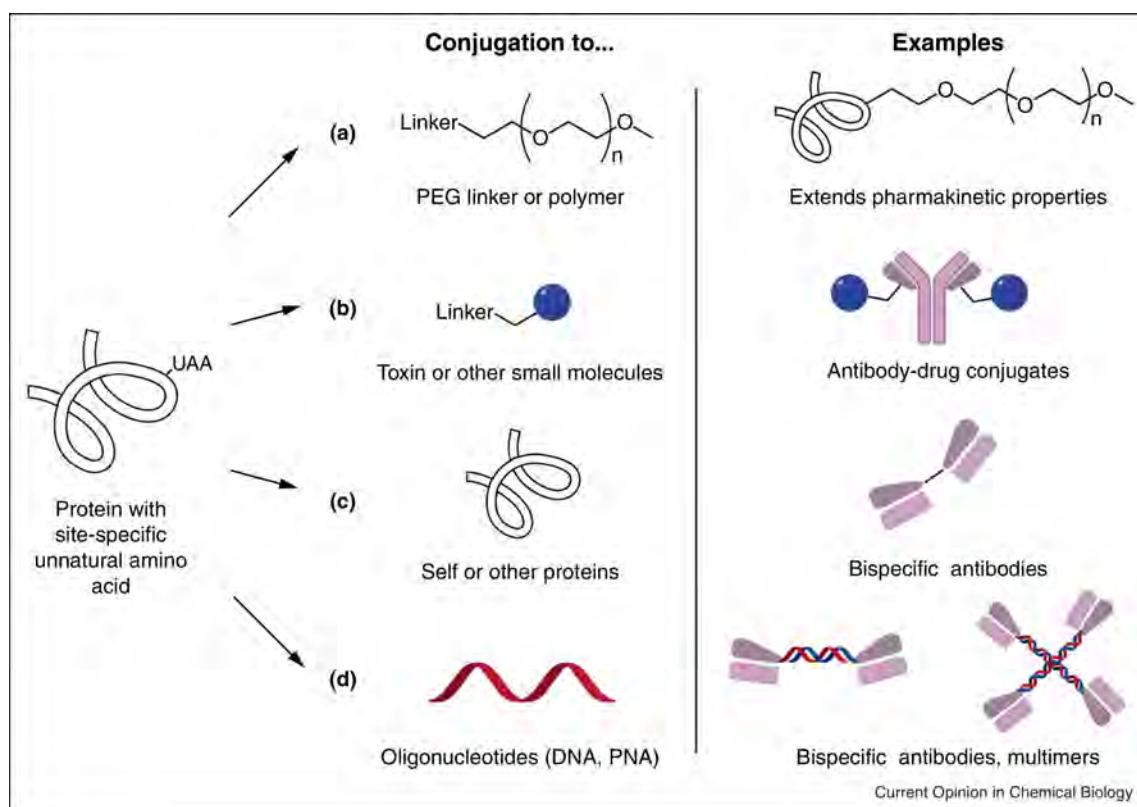
Kim and colleagues developed a two-step method to synthesize homogeneous bispecific antibodies using genetically encoded UAAs and flexible bifunctional linkers (Kim et al., 2012). This approach involves conjugating the two pAcF-containing Fabs to bifunctional linkers (a short PEG with an alkoxyamine on one end and either an azide or cyclooctyne on the other). In a second step, the antibody-linker conjugates are coupled to each other via copper-free [3 + 2] Huisgen cycloaddition in good yields. The affinity of anti-Her2 Fab homodimers generated by this method toward Her2 was comparable to full-length IgG. Moreover, heterodimers of anti-Her2/anti-CD3 Fabs showed subnanomolar killing ( $EC_{50} \sim 20$  pM) of Her2<sup>+</sup> cancer cells in the presence of purified peripheral blood mononuclear cells (PBMCs) in vitro, which is comparable to the most potent bispecific antibodies reported to date. Furthermore, the conjugate has a half-life of  $\sim 7$  hours in rodents and has shown effective tumor suppression in mouse xenograft models. Similar bispecific antibodies against other targets, for example, CD20, CD33, and EGFR, are being developed.

Proteins have also been conjugated to single-stranded DNAs or peptide nucleic acids (PNAs) and can be subsequently hybridized to form homodimers, heterodimers, tetramers or higher order complexes (Kazane et al., 2013) (Figure 1.6d). For example, PNA was conjugated to a pAcF-containing Fab via oxime ligation in excellent yields. The PNA-linked anti-Her2/anti-CD3 Fabs

showed potent and specific cytotoxicity *in vitro*, similar to that of the PEG-linked bispecific Fabs. A tetramer of anti-CD20 Fab was generated utilizing PNAs that form a Holliday junction, and potently induced apoptosis of CD20<sup>+</sup> Ramos cells. Additionally, an anti-Her2 Fab conjugated to single-stranded DNA was used to detect Her2<sup>+</sup> cells by immuno-PCR with exquisite sensitivity (Kazane et al., 2012).

Other applications of selective conjugation with genetically encoded UAAs include the conjugation of fluorescent dyes and other spectroscopic probes to proteins. Fluorescent dyes have been conjugated both *in vitro*, such as for FRET experiments to study protein dynamics (Brustad et al., 2008b) and *in vivo* for specific labeling on cell surfaces (Lang et al., 2012; Plass et al., 2011; Ye et al., 2008).

All these approaches are very promising, but there are still some advantages to using protein modification methods that target native amino acids such as tyrosine and tryptophan introduced without specialized expression systems (Seim et al., 2011). Even in cases in which the introduction of a uniquely reactive amino acid group can be achieved expressing proteins with recombinant methods, most bioconjugation reactions do not reach full conversion and sometimes the addition of excess reagent can be disadvantageous. For this reason we tried to develop a combinatorial



**Figure 1.6.** Protein conjugates synthesized with unnatural amino acids and examples of their applications.

screening platform to identify and optimize the sequence target reactivity using SAMDY Mass Spectrometry (MS).

## 1.2. SAMDY MS technique

SAMDY Mass Spectrometry is an innovative technique that combines self-assembled monolayers (SAMs) of alkanethiolates on gold with matrix assisted laser desorption/ionization (MALDI) time-of-flight (TOF) MS that can be used for a wide range of chemical and biological assays (Gurard-Levin and Mrksich, 2008a; Mrksich, 2008).

### 1.2.1. Matrix Assisted Laser Desorption/Ionization Time-Of-Flight Mass Spectrometry

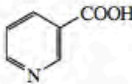
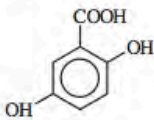
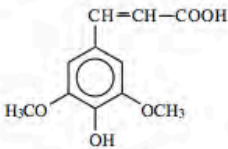
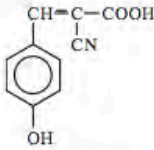
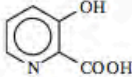
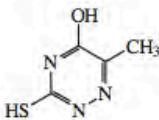
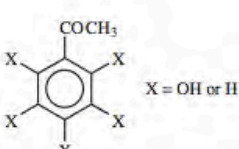
Matrix-assisted laser desorption/ionization (MALDI) is one of the two “soft” ionization techniques besides electrospray ionization (ESI) which allow for the sensitive detection of large, non-volatile and labile molecules by mass spectrometry (MS). Introduced by Hillenkamp and Karas (Karas and Hillenkamp, 1988), MALDI MS has developed over the past 27 years into an indispensable analytical tool for peptides, proteins, and most other biomolecules (oligonucleotides, carbohydrates, natural products, and lipids) (Hillenkamp and Peter-Katalinic, 2013)

#### 1.2.1.1. MALDI MS principles

The general principle of MALDI-TOF-MS revolves around the rapid photo-volatilization of a sample embedded in a UV-absorbing matrix followed by time-of-flight mass spectrum analysis (Marvin et al., 2003). The analyte is first co-crystallized with a large molar excess of a matrix compound, usually a UV-absorbing weak organic acid, then laser radiation of this analyte–matrix mixture results in the vaporization of the matrix which carries the analyte with it. Therefore in MALDI analysis the matrix plays a key role because it serves as a proton donor and receptor (acting to ionize the analyte in both positive and negative ionization modes, respectively) and at the same time absorbs the laser light energy causing, indirectly, the analyte to vaporize.

Several theories have been developed to explain desorption of large molecules by MALDI, as well as to define the ionization process but both these mechanisms are still largely unknown. It is generally thought that ionization occurs through proton transfer or cationization but it seems that a more complex interaction of analyte and matrix, rather than simple acid–base chemistry, is responsible for it. Indeed the ionization is not as significantly dependent on the number of acidic or basic groups of the analyte as it is on the matrix–analyte combination (Lewis et al., 2000; Marvin et al., 2003).

Of critical importance for success in MALDI experiments are both the matrix choice (among the variety shown in Table 1.1) and the sample – matrix preparation. Although many protocols exist, the sample is usually embedded in an excess of a solid matrix, which, upon laser irradiation, assists in the volatilization and ionization of the analytes. The high excess of matrix to sample (from 100:1 to 10000:1) is important, since the matrix serves as the primary (and highly efficient) absorber of the UV laser radiation and breaks down rapidly, expanding into the gas phase.

<b>Matrix</b>	<b>Structure</b>	<b>Wavelength</b>	<b>Major applications</b>
Nicotinic acid		UV 266 nm	Proteins, peptides, adduct formation
2,5-Dihydroxybenzoic acid (plus 10% 2-hydroxy-5-methoxybenzoic acid)		UV 337 nm, 353 nm	Proteins, peptides, carbohydrates, synthetic polymers
Sinapinic acid		UV 337 nm, 353 nm	Proteins, peptides
$\alpha$ -Cyano-4-hydroxycinnamic acid		UV 337 nm, 353 nm	Peptides, fragmentation
3-Hydroxy-picolinic acid		UV 337 nm, 353 nm	Best for nucleic acids
6-Aza-2-thiothymine		UV 337 nm, 353 nm	Proteins, peptides, non-covalent complexes; near-neutral pH
k,m,n-Di(tri)hydroxy-acetophenone		UV 337 nm, 353 nm	Protein, peptides, non-covalent complexes; near-neutral pH
Succinic acid	$\text{HOOC}-\text{CH}_2-\text{CH}_2-\text{COOH}$	IR 2.94 $\mu\text{m}$ , 2.79 $\mu\text{m}$	Proteins, peptides
Glycerol	$\text{H}_2\text{C}-\text{CH}-\text{CH}_2$               OH OH OH	IR 2.94 $\mu\text{m}$ , 2.79 $\mu\text{m}$	Proteins, peptides, liquid matrix

IR = infrared; UV = ultraviolet.

**Table 1.1.** Commonly used matrices and their main properties and applications.

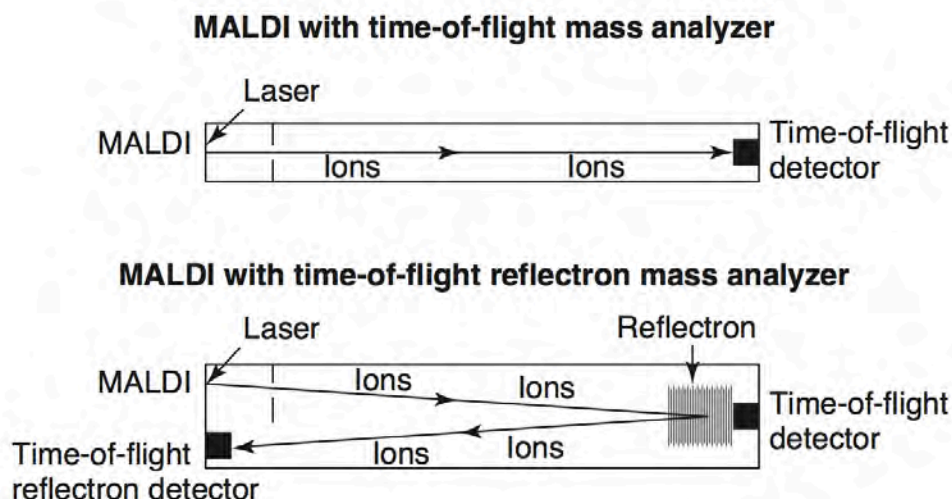


Additionally, the high matrix/sample ratio reduces associations between analyte molecules, and provides protonated and free-radical products that ionize the molecules of interest. With this “soft” ionization technique (the analyte itself does not absorb the laser energy directly) the analysis of complex biomolecules up to several hundred kilodaltons is allowed. Typically, peptides or proteins are generally solubilized in a 0.1% aqueous trifluoroacetic acid (TFA). One microliter of this solution is mixed with a saturated aqueous solution of matrix and the solvent allowed to evaporate to form crystals (Marvin et al., 2003).

After MALDI, generally the desorbed ions can be analyzed using a time-of-flight mass spectrometer (TOF- MS) working in the linear or reflectron mode (Figure 1.7).

In the **linear TOF** mass analyzer, TOF analysis is based on accelerating a set of ions to a detector where all of the ions are given the same amount of energy. Since the ions have same energy but different mass, they reach the detector at different times: the smaller ones reach the detector first because of their greater velocity while the larger take longer owing to their larger mass. Hence, the analyzer is called TOF because the mass is determined from the ions’ time of flight. The arrival time at the detector is dependent upon the mass, charge, and kinetic energy (KE) of the ion. Since KE is equal to  $1/2 mv^2$  or velocity  $v = D (2KE/m)^{1/2}$  ions will travel a given distance,  $d$ , within a time,  $t$ , where  $t$  is dependent upon their mass-to-charge ratio ( $m/z$ ).

The **TOF reflectron** combines TOF technology with an electrostatic analyzer, the reflectron. The reflectron serves to increase the amount of time,  $t$ , ions need to reach the detector while reducing their KE distribution, thereby reducing the temporal distribution  $\Delta t$ . Since resolution is defined by the mass of a peak divided by the width of a peak or  $m/\Delta m$  (or  $t/\Delta t$  since  $m$  is related to  $t$ ), increasing  $t$  and decreasing  $\Delta t$  results in higher resolution. This increased resolution, however, often comes at the expense of sensitivity and a relatively low mass range, typically  $<10\ 000\ m/z$ . One innovation that has had a dramatic effect on increasing the resolving power of MALDI/TOF instruments has been delayed extraction (DE). DE is a relatively simple means of cooling the ions (and possibly focusing them) immediately after the MALDI ionization event. In traditional MALDI instruments the ions were accelerated out of the ionization source as soon as they were formed; however, with DE the ions are allowed to “cool” for  $3/4150\ ns$  before being accelerated to the analyzer. This cooling period generates a set of ions with a much lower KE distribution, ultimately reducing the temporal spread of ions once they enter the TOF analyzer. Overall, this results in increased resolution and accuracy. Incidentally, the benefits of DE significantly diminish with higher-molecular-weight proteins ( $>30\ 000\ Da$ ) (Lewis et al., 2000).



**Figure 1.7.** Principle of matrix-assisted laser desorption/ionization mass spectrometry with TOF and TOF reflectron. The irradiation of the analyte-matrix mixture by the laser induces the ionization of the matrix, desorption and the transfer of protons from photo-excited matrix to analyte to form a protonated molecule.

#### 1.2.1.2. MALDI-TOF-MS applications

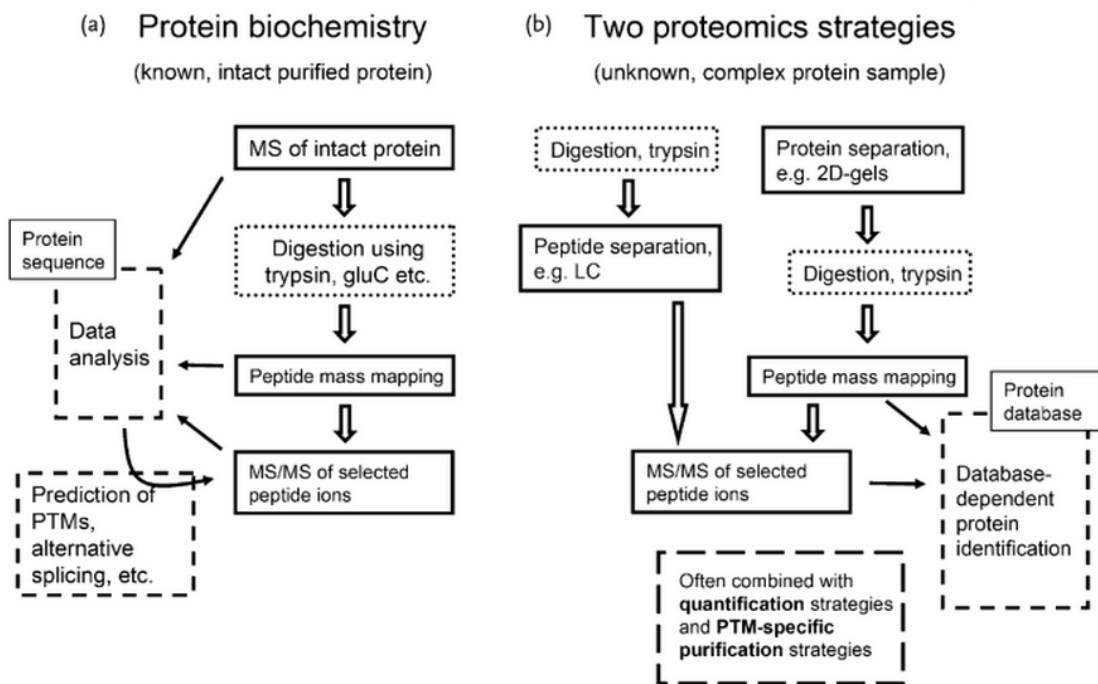
Today, mass spectrometry (MS) is the most useful analytical technique for protein and proteome analysis, as it provides a relatively simple platform for determining the molecular mass, which is one of the fundamental properties of biological molecules. In this view, depending on the complexity of the sample and the information sought, several approaches can be chosen. There are mainly two strategies for MS-based protein analysis, summarized in Figure 1.8: the biochemical characterization of purified proteins; and two parallel strategies for the analysis of proteomes or subproteomes - that is, the multitude of proteins within a given cell or cell organelle at any one time. The first approach (Figure 1.8a) is very useful for the validation of native or recombinant proteins destined for structure analysis by nuclear magnetic resonance (NMR) or X-ray diffraction, as well as in the quality control of proteins for biotechnological applications. The method is based on an ability to obtain pure protein in soluble form, and in a solvent that is suitable for matrix-assisted laser desorption/ionization (MALDI) -MS analysis. In most cases, the identity of the protein is known *a priori*, and the amino acid sequence is available either from protein databases or by translation from the corresponding gene sequences. The mass determination of a purified intact protein will reveal any discrepancies from the molecular mass of a "naked" protein, that is, the calculated molecular mass as determined by the amino acid sequence predicted from the cognate gene. Similarly, MS analysis of peptides derived from a protein by enzymatic digestion will not only confirm the amino acid composition of the individual peptides but also reveal the presence of

any chemical modifications by a mass increment or a mass deficit relative to the expected masses of the unmodified peptides. This type of analysis is referred to as "peptide mass mapping."

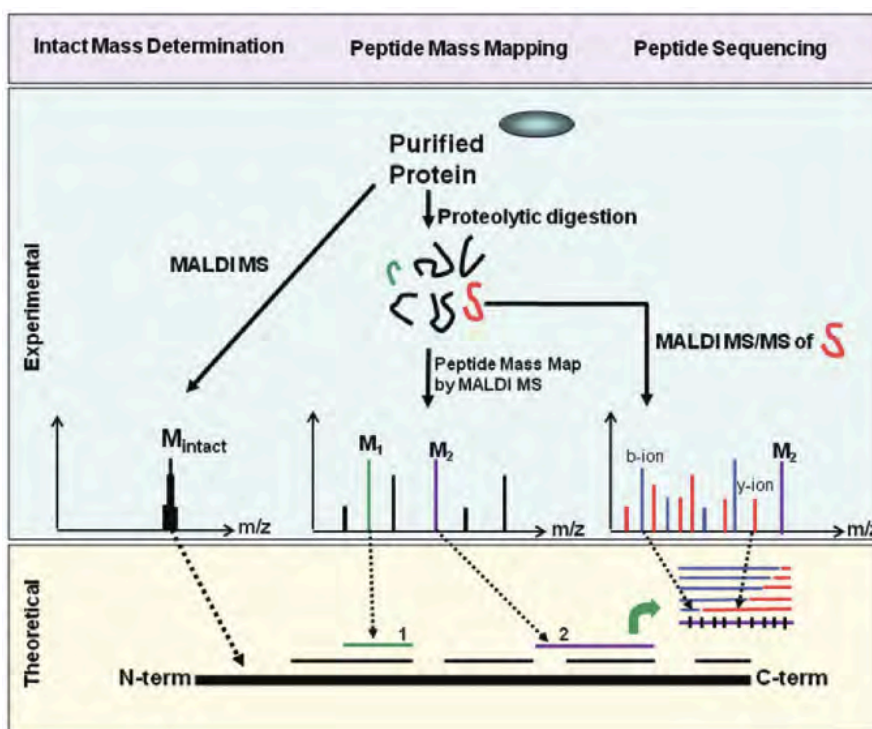
Strategies of MALDI-MS in protein biochemistry can also be used to analyze proteins for which the amino acid sequence is unknown, or is only partially characterized. In that case the MS analysis combined with other technologies, such as cloning, would help to elucidate the complete primary structure of the proteins. Depending on the size and properties of the protein, 2 to 50 pmol of purified protein is needed for a complete and comprehensive molecular analysis by multiple MALDI-MS experiments, including intact mass determination, several peptide mass mapping experiments, and MS/MS analysis of individual peptides. Intact mass determination is a very useful first step towards the full characterization of proteins (Figure 1.9), and the purified protein should therefore be analyzed by MALDI-MS for determination of the molecular mass. Gel electrophoresis will often be used as the final stage of protein purification, depending on the purification procedure used. If not, it is advantageous to monitor the protein purification by SDS-PAGE to establish the purity and integrity of the protein sample and to estimate protein abundance. The intact proteins can sometimes be successfully extracted from SDS-PAGE gels or 2-D electrophoresis gels by electroelution or by detergent-mediated passive elution. However, this requires significant amounts of purified protein (often more than 20 pmol) in the gel. Alternatively, the proteins are analyzed directly from solution. Usually, 1-10 pmol of protein loaded onto the MALDI probe will ensure a good signal, but this will depend on the protein size and sample solubility/complexity and solvent composition (Hillenkamp and Peter-Katalinic, 2013).

When used for proteomics analysis and the analysis of large protein complexes (Figure 1.8b), MS must be combined with protein or peptide separation strategies, in order to deal with the high complexity of such samples. These are typically one-dimensional (1-D) or two-dimensional (2-D) gel electrophoresis or liquid chromatography (LC) approaches. The individual proteins/peptides are then identified by searching of the obtained MS-data against in-silico-digested proteins from databases of known or predicted sequences. Correlations between the experimental and theoretical data are used for scoring of the proteins and thereby determining the identity of the proteins in the sample being analyzed.

Modern TOF-MS are relatively simple and inexpensive, exhibit high sensitivity, and have virtually an unlimited mass range; therefore, TOF-MS is well suited to clinical and biomedical analyses when coupled with MALDI. As a matter of fact, it has already been used as a primary investigative tool to characterize a number of cancer, Alzheimer, arthritis, and allergy protein markers of disease or susceptibility to disease (Marvin et al., 2003).



**Figure 1.8.** MALDI-MS for protein analysis. MALDI MS is a very powerful tool in many types of analysis involving proteins, from in-depth analyses of individual proteins to studies of large complex protein samples, as in proteomic studies. Here, two different strategies are schematized: (a) the analysis of a purified protein for which the sequence is known; and (b) the analysis of all proteins in, for example, a given cell (i.e., the proteome of the cell). Both, the LC- and gel-based strategies are shown, and these can be combined with quantification strategies and studies of modified proteins. As can be seen, many of the steps involved, such as protein digestion, peptide mass mapping and peptide fragmentation, are identical and used in all of the strategies shown. PTM: post translational modification.



**Figure 1.9.** Strategy for the analysis of purified proteins by MALDI-MS. The initial stage is an intact mass determination by MALDI-MS, followed by proteolytic cleavage using one or more sequence-specific enzymes. The resulting peptide mix is analyzed using MALDI-MS. Finally, the individual peptides can be subjected to amino acid sequencing by MALDI-MS/MS. Comparison of the experimental and calculated mass values will reveal discrepancies that might originate from natural processing of the protein, or from chemical artifacts. The theoretical molecular mass of a protein and the derived peptides is calculated from the amino acid sequence using computer tools.

### 1.2.2. Self Assembled Monolayers

Self-assembled monolayers (SAMs) are highly ordered molecular assemblies formed spontaneously by chemisorptions and self-organization of long chain molecules on the surface of appropriate substrates. Nowadays SAMs are broadly used in biochip applications as they allow to control structure and properties of the surface (Whitesides et al., 2005). They provide indeed a broad chemical flexibility for interface engineering and at the same time the use of ethylene glycol molecules prevents nonspecific adsorption of protein. (Gurard-Levin and Mrksich, 2008a) The most extensively studied SAMs are n-alkanethiolates on gold: they were first described nearly 30 years ago and have still remained the most important strategy for preparing structurally well-defined and complex organic surfaces (Finklea et al., 1987; Nuzzo and Allara, 1983).

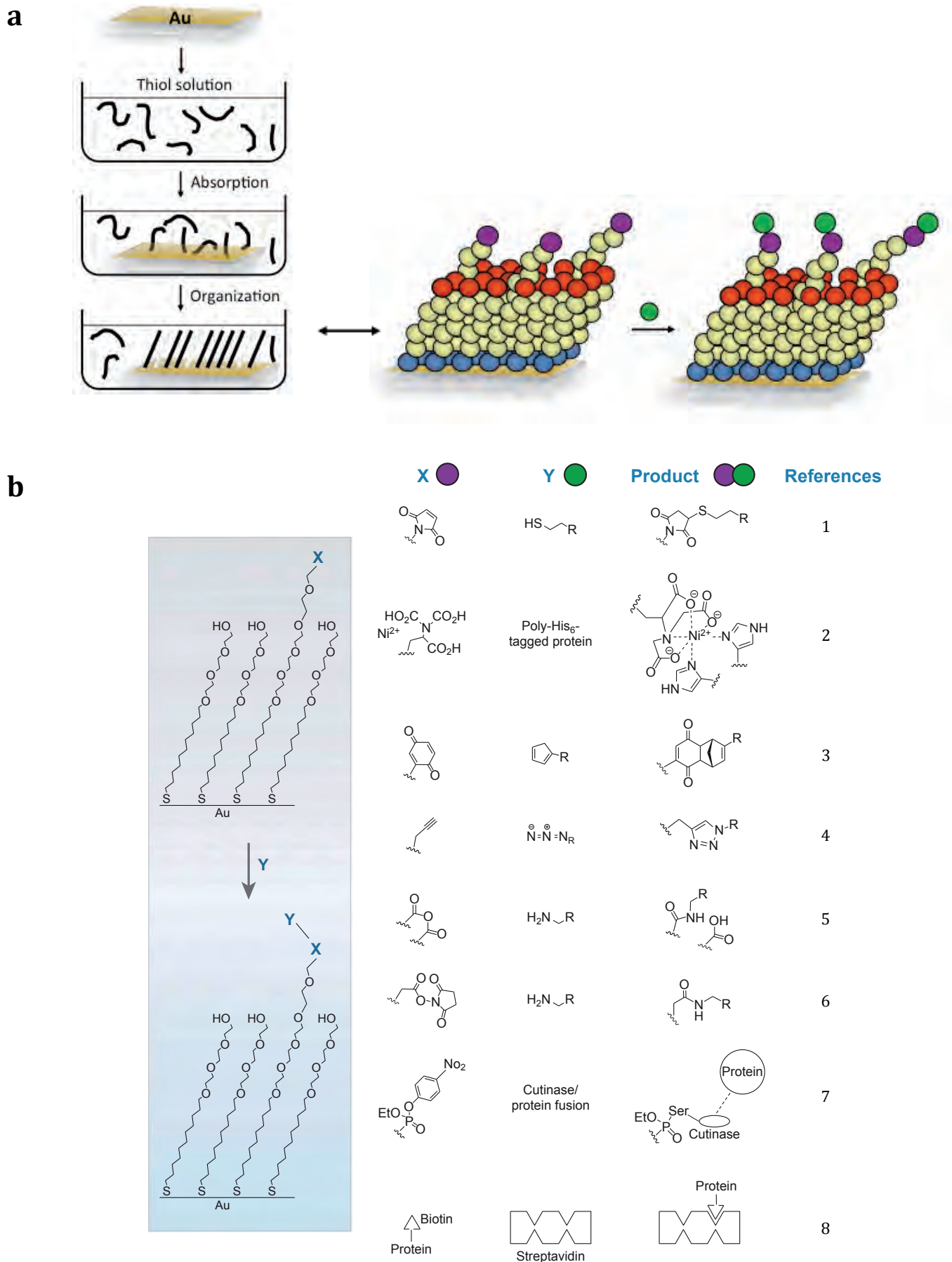
The molecular units of a SAM of alkanethiolate on gold have three components: a chemical moiety or "head" group (generally a thiol group) that binds to a substrate, hydrocarbon spacer group (in our case a tri(ethylene) glycol chain), and a terminus or "tail" group X. The gold-thiolate bond has the highest energy in the system (17.7 kcal) and orients the molecules with their head groups toward the solid surface, leaving the tail groups free. This conformation spaces the sulfur atoms by  $\sim 5\text{\AA}$  and the lattice spacing is slightly larger than the van der Waals diameter of the alkane chains, and the chains tilt  $30^\circ$  to maximize the attractive interactions between chains. The strength of the van der Waals interactions induces the alkane chains to align in an all-trans conformation. (Whitesides et al., 2005)

Gold is an attractive solid support for the thiolated SAMs due to its high affinity towards sulfur, low oxidation capability, and compact packing ability (Ulman, 1996; Vericat et al., 2005). The most accessible and common method for forming monolayers is the absorption of alkanethiols resulting from the immersion of the Au substrate into an ethanolic solution of either alkanethiolates or alkyldisulfides and subsequently letting it self-assemble over a period of several hours (Figure 1.10a). Our SAMs are mixed disulfide-terminated monolayers containing the ligand of interest at a defined percentage against a background of tri(ethylene) glycol groups. The ligand of interest can either be an enzyme or chemical substrate, or it can be a ligand that would allow further immobilization of substrates. The monolayers are compatible with several immobilization chemistries (Figure 1.10b) including maleimide-thiol (Houseman and Mrksich, 2001), copper-assisted Click (Ban et al., 2012), benzoquinone–cyclopentadiene Diels–Alder (Houseman and Mrksich, 2002), trichloroacetimidate–alcohol (Ban and Mrksich, 2008). The maleimide immobilization is particularly useful for creating peptide arrays via a C-terminal cysteine (Min and

Mrksich, 2004). As previously mentioned the tri(ethylene) glycol groups have been proven to prevent non-specific protein adsorption (Prime and Whitesides, 1991) and this is important not only for eliminating false positives for protein binding experiments, but also for preventing reaction sites from being blocked and allowing for accurate activity measurements.

The SAMs have been characterized using an array of methods such as X-ray Photoelectron Spectroscopy (XPS), ellipsometry, scanning tunneling and atomic force microscopy (STM and AFM), etc. (Mrksich, 2008; Ulman, 1996; Whitesides and Laibinis, 1990). These methods have been crucial for characterizing the shapes, mechanics, and physical properties of surfaces; however they are not very useful for fully characterizing the molecular structure of complex biological interfaces.

Inspired by previous studies of SAMs characterization, which investigated laser desorption of alkanethiolates and showed that both the monomers and dimers of the alkanethiolates were observed, along with fragments of these species (Gong et al., 2001; Trevor et al., 1998), the Mrksich group tested a conventional MALDI TOF mass spectrometer on its ability to characterize functionally modified SAMs. It was found that the use of matrix aids in the desorption of intact chains of alkanethiolates and alkyldisulfides and the spectra obtained are clean and straightforward to analyze, as outlined in Figure 1.11 (Su and Mrksich, 2002).

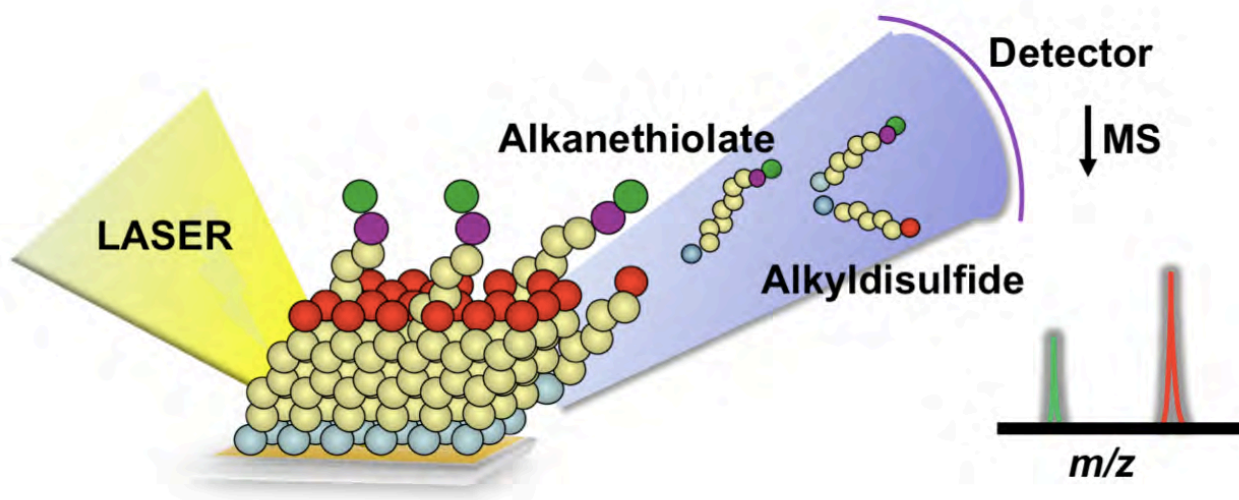


**Figure 1.10.** SAMs formation and functionalization. (a) SAMs formation. SAMs are composed of tri(ethylene) glycol groups (red) and ligand of interest (purple). The substrate of interest (green) can be immobilized using the immobilization strategy of choice. (b) Selected immobilization strategies used to functionalize SAMs. References: (1) Houseman and Mrksich 2001; (2) Sigal et al. 1996; (3) Houseman et al. 2002; (4) Houseman, Gawalt, and Mrksich 2003; (5) Devaraj et al. 2005; (6) Patel et al. 1997; (7) Luk et al. 2004; (8) Kwon et al. 2004; Mrksich, Grunwell, and Whitesides 1995).

### 1.2.3. SAMDI MS applications

Through a variety of chemistries (Figure 1.10b) the sample is immobilized onto SAMs in an array format either before or after treatment with the reaction mixture (Min et al., 2004a). After the incubation time necessary for the reaction to happen, the surfaces are simply rinsed to remove salts and other components in the reactions, treated with matrix, and analyzed with a commercial MALDI-TOF instrument. Upon irradiation with the laser, the monolayers are efficiently desorbed from the surface through cleavage of the thiolate-gold bond and ionized, but undergo little fragmentation, to give relatively simple spectra that reflect the masses of the alkanethiolates and corresponding disulfides (Figure 1.11) (Mrksich, 2008).

Over the past decade SAMDI arrays have been continually developed demonstrating a remarkable versatility towards a broad array of applications including enzyme and protein/protein interaction assays (Mrksich, 2008), chemical reaction discovery (Li et al., 2007) and high-throughput screening (Gurard-Levin et al., 2011). An overview of SAMDI applications is shown below.



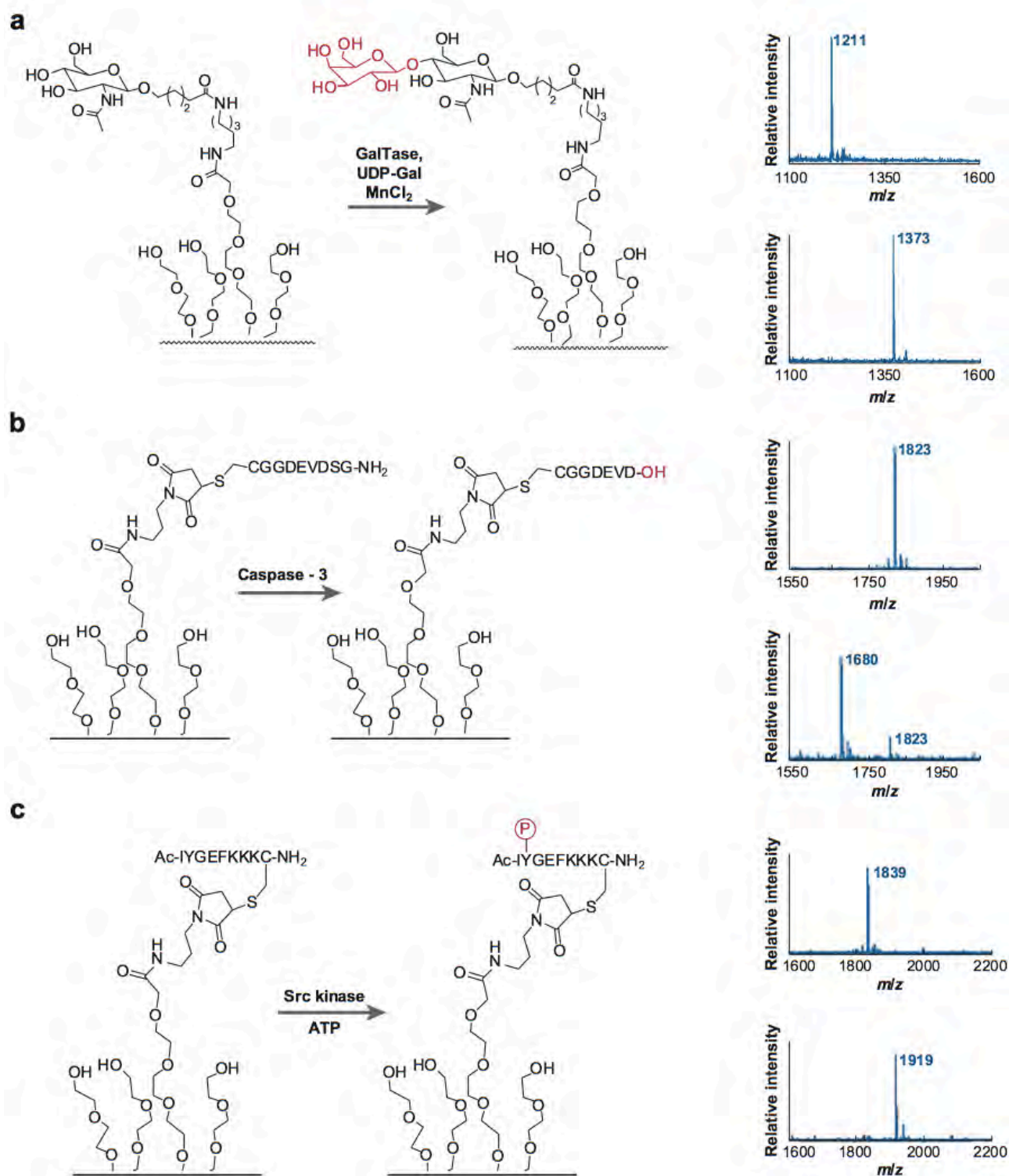
**Figure 1.11.** SAMDI MS principle. SAMDI allows for fast and straightforward mass analysis of SAMs on gold. Upon laser irradiation, molecules are desorbed as either alkanethiolates (green peak) or alkyldisulfides (red peak).



### 1.2.3.1. Enzyme activity

Most enzymes act on substrates and generate products whose masses are distinct from those of the substrates. For example, kinases add phosphate groups to hydroxyl-bearing side chains of proteins, proteases cleave proteins by hydrolyzing an amide bond, and acetylases convert the amino groups of lysine residues to the corresponding acetamides (Walsh, 2006). SAMDI offers the ability to perform label-free assays of enzyme activities, starting with the immobilization of the relevant enzyme substrate to monolayers that are otherwise inert. Treatment of the monolayers with the enzyme and any required cofactors leads to the product, which can then be observed in the mass spectrum.

For instance, in previous works a kinase activity assay was carried out by immobilizing a cysteine-terminated peptide to a monolayer presenting the maleimide group (Min et al., 2004b). SAMDI spectrum showed clear peaks corresponding to the immobilized peptide and to the background tri(ethylene glycol)-terminated alkanethiolates, and also showed a lack of peaks for the initial maleimide-terminated alkanethiolates demonstrating that the immobilization reaction was complete. The surface was then treated with src kinase, a SAMDI spectrum revealed that the peaks representing peptide-alkanethiol conjugates shifted by 80 Da, as expected for the increase in mass following phosphorylation (Figure 1.12c). In this and most other examples, multiple peaks were observed for the anticipated analyte that represented the alkanethiolates in their monomeric and disulfide forms. It is not clear whether the disulfides formed upon desorption of the monolayer from the gold surface, which is well preceded in thermal reactions (Di Valentin et al., 2005), or whether the disulfides formed in the ion cloud following desorption. Mrksich and colleagues also observed adducts of the alkanethiols with multiple counter ions, including proton, sodium, and potassium. In any event, they did not see significant fragmentation of the functionalized alkanethiolates and they consequently obtained spectra that are straightforward to interpret. Other examples have characterized the methylation of arginine by protein arginine methyltransferase 1 (Min et al., 2004a), the protease activity of caspase-3 (Figure 1.12b) (Su et al., 2006), and the modification of carbohydrates by glycosyltransferase (Figure 1.12a) (Ban and Mrksich, 2008; Ban et al., 2012; Guan et al., 2011; Houseman and Mrksich, 2002), lysine acetylases and deacetylases (Gurard-Levin and Mrksich, 2008b; Gurard-Levin et al., 2010). Su and co-workers studied by SAMDI MS endogenous caspase protease activities in apoptotic cell lysates and compared them with the ones obtained with the commonly used fluorogenic assays. Cultured cells were stimulated to initiate apoptosis and lysed at different times. Each lysate was then split into fractions for analysis by the fluorescence assay and the SAMDI-MS assay.



**Figure 1.12.** Examples of enzyme activity assays performed with self-assembled monolayers for matrix-assisted laser desorption/ionization mass spectrometry (SAMDI MS), wherein monolayers presenting carbohydrate or peptide ligands were treated with an enzyme. In each case, mass spectra revealed peaks corresponding to the masses of the substituted alkanethiolates before and after modification of the enzyme. Examples are shown for (a) galactosylation of an immobilized carbohydrate by  $\beta(1,4)$ -galactosyl transferase and (b) proteolysis of a peptide by caspase-3, as well as phosphorylation of a peptide by src kinase (c).

Monolayers presenting specific peptide substrates for either caspase-3 or -8 were treated with the lysates from Jurkat cells (stimulated with staurosporine) and SKW6.4 cells (stimulated with LzCD95L). In both cases, the SAMDI assays reported on the activation of endogenous caspase enzymes with levels of detection similar to those observed with the commonly employed fluorescence assay. Furthermore the use of longer peptide substrates, not compatible with the fluorogenic assays, provided for a better resolution of the two caspase activities. These results clearly show that the SAMDI assay can be used to measure endogenous enzyme activities. It also avoids the loss of activity and specificity that often accompany label-dependent assay formats. In particular, the comparison between caspase assays using the fluorogenic tetrapeptide substrates and the label-free substrates pointed out that SAMDI assay format performs better at resolving caspase activities and, therefore, may provide a new opportunity to understand caspase activation in several biological contexts (Su et al., 2006). Studies on multiple activities have also been performed with SAMDI as the molecular weight shift of every single substrate can be resolved from the others and its specific product as well.

All these examples illustrate several benefits of the SAMDI method in enzyme assays. First, the use of MS provides a label-free method for assaying a broad range of biochemical activities. The lack of labels eliminates several steps in the assay and avoids the need for developing antibodies to label an intended analyte, thereby significantly reducing development time and decreasing the risk that introduction of the label will interfere with the biological activity being assayed. MS also permits simultaneous assay of distinct enzyme activities—which would normally be incompatible because of different labeling strategies—to be performed on the same chip and thus with a single sample.

MS also provides instructive information about the analyte. Whereas surface plasmon resonance spectroscopy (SPR) and related optical methods provide information on the amount of protein that interacts with an immobilized ligand, they do not provide any information regarding the chemical properties of the proteins under investigation. The ability of SAMDI to identify each species at the surface according to its molecular weight allows a straightforward discrimination between signals due to specific analytes and to background. This ability also allows observation of both the products and substrates of an enzyme and therefore can verify that the immobilized molecule was indeed on the chip (thereby identifying possible false negatives); further, SAMDI can provide a better assessment of the yield of the enzymatic reaction. Moreover, the mass resolution of this method also permits multianalyte assays to be performed. In one example, a mixture of four peptides was immobilized on the surface, and each one of them could be clearly resolved in the

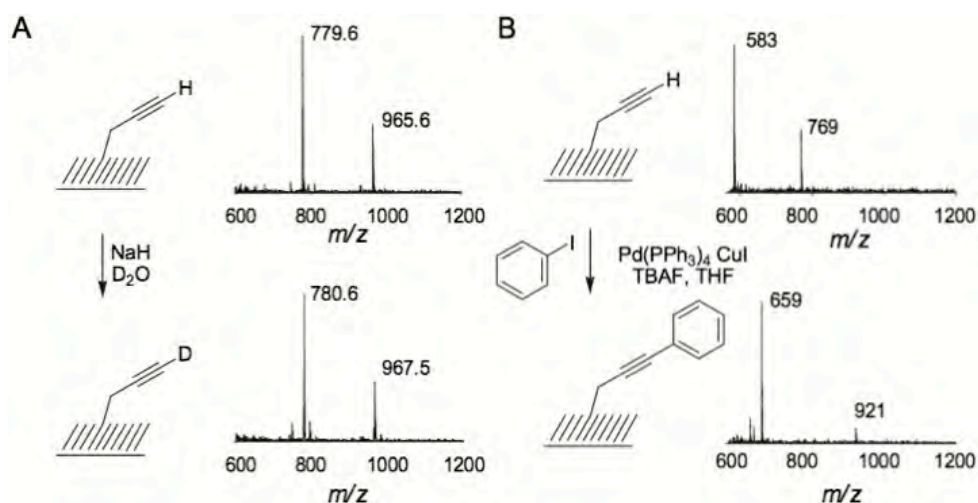
SAMDI spectrum. Treatment of the monolayer with mixtures of kinases resulted in selective phosphorylation reactions, which could be analyzed by SAMDI to determine which kinases were present in the sample (Min et al., 2004b).

### 1.2.3.2. Chemical reaction discovery

The task of reaction discovery in synthetic chemistry is also rooted (in part) in numerous trial-and-error experiments to identify reagents and conditions that promote a desired reaction. Yet, the same chip-based tools that are under intense development for biological applications have not, with one notable exception, been applied to reaction discovery.

The SAMDI method is able to develop reactions of immobilized molecules with soluble reagents and identify unanticipated reactions. Whereas current methods require a combination of several analytical methods to characterize the products of interfacial reactions (with a substantial effort required for each reaction), SAMDI rapidly provides information for the products formed as well as the approximate yields. Li and colleagues screened, with the aid of the SAMDI method, 15 reactions, and for each of them they identified conditions that gave a high-yielding conversion. For instance, treatment of a monolayer presenting terminal alkyne groups with sodium methoxide in deuterated water promoted the exchange of the terminal hydrogen with a deuterium atom (Figure 1.13a). The mass spectrum clearly showed the mass increase associated with the addition of a single neutron to the alkyne. In Figure 1.13 are shown both the replacement of the hydrogen of a terminal alkyne with deuterium, and the palladium-catalyzed coupling of the alkyne with iodobenzene. In both cases, SAMDI showed clear peaks corresponding to the mass of the expected product with a mass resolution better than 1 Da (Mrksich, 2008).

In order to investigate the application of this method to the identification of unanticipated reactions, monolayers presenting a single functional group at low density against a nonreactive background were treated with an array of common reagents. The spots were then analyzed by SAMDI to identify those that gave a high-yielding conversion to a new product whose structure was not readily deduced from the mass. This screen identified a novel reaction, wherein a primary amine reacted with three equivalents of an aldehyde under mild conditions to provide an N-alkylpyridinium product (Evans and Miller, 2002; Li et al., 2007).



**Figure 1.13.** Chemical reaction study by SAMDI MS. Chemical reactions of molecules attached to a monolayer are rapidly characterized with SAMDI MS. Treatment of a monolayer presenting a terminal alkyne group with sodium hydride and deuterated water results in the exchange of the terminal hydrogen for a deuterium atom. SAMDI MS identifies the product with a mass change of 1 Da (at  $m/z$  780.6) and the associated disulfide product (at  $m/z$  967.5) (A). A Sonogashira coupling of iodobenzene to the alkyne also proceeds in high yield to produce the phenylalkyne adduct (B).

### 1.2.3.3. High Throughput Assays and drug discovery

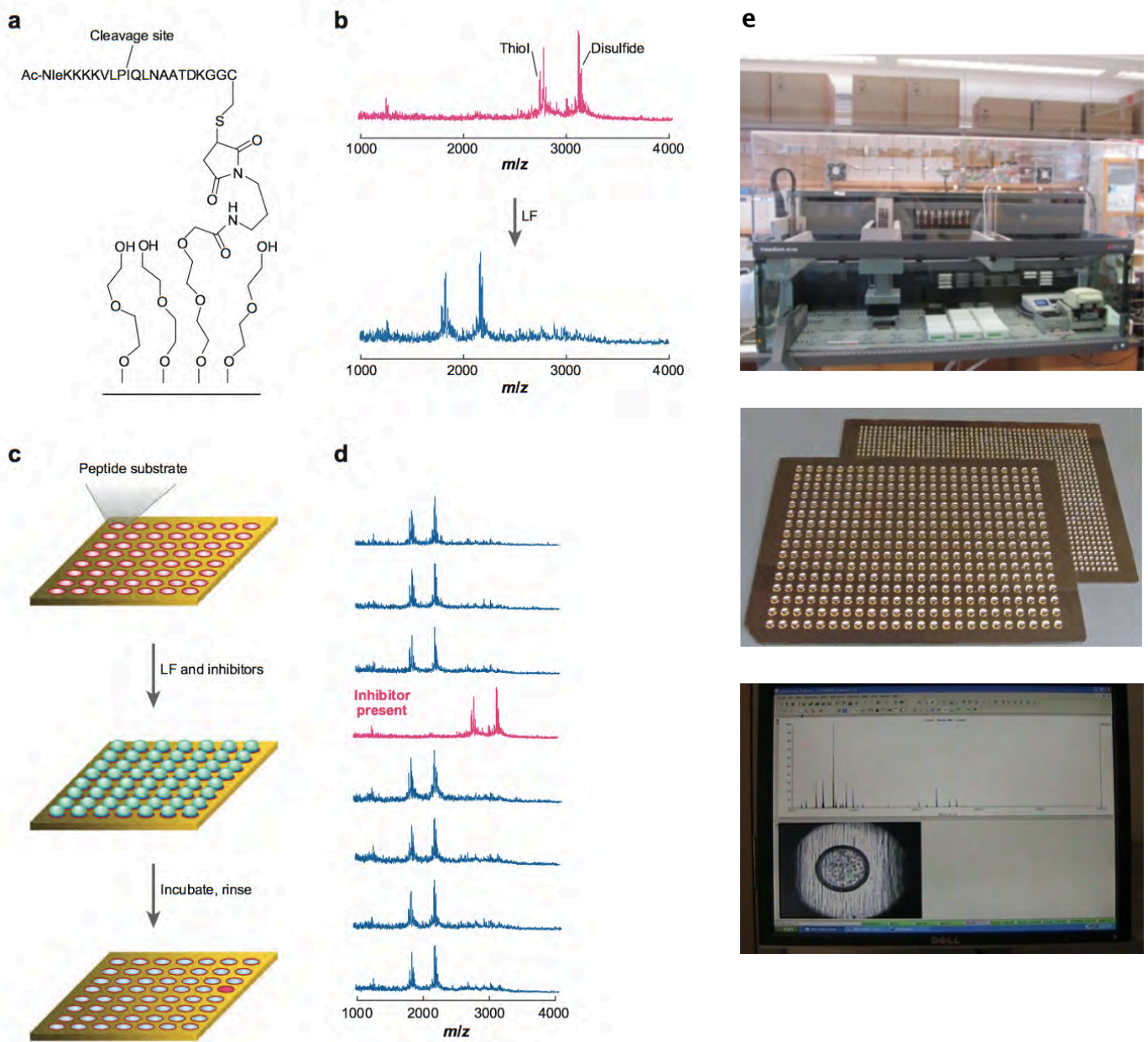
Other SAMDI features that have been studied with always growing interest and effort are the high-throughput and the automation achievable in performing and analyzing the reactions. Nowadays, mass spectrometry is routinely performed on stainless steel ‘array plates’ that are modified with gold circles that are each 2.8 mm in diameter and are arranged in the standard geometry of a 384-well microtiter plate. (Ban et al., 2012; Gurard-Levin et al., 2011) In cases in which the reagents are too costly or hard to prepare, it can be used even a 1536 gold islands fashion. Previous high-throughput assays instead consisted of 5x10 or 10x10 arrays of circular islands. In addition, laboratory automation equipments are currently available: liquid handlers can be used to transfer the reagents onto the metal plates and also data acquisition can be performed in automatic mode using a conventional MALDI TOF MS instrument (Figure 1.14e).

This automation makes it possible to screen 10000s of compounds in a day, which together with its versatility makes SAMDI a very competitive method for high-throughput arrays.

In drug discovery screening it is common to perform enzyme assays in the presence of small molecules from a library, and then to identify those molecules that inhibit (or activate) the enzyme activity. Therefore, SAMDI is an effective approach to also this kind of applications. Min and co-workers developed an assay for the anthrax lethal factor toxin, which has proteolytic activity toward peptide substrates, to identify inhibitors from a library containing 10,000 small molecules (Figure 1.14). Droplets containing the enzyme and eight compounds were applied to a target plate having

immobilized peptide and incubated for 1 h. The plate was then rinsed and analyzed by mass spectrometry to identify those droplets that had no activity. Using this approach, they could evaluate several thousand compounds in a day, and they identified a small molecule that inhibited the lethal factor protease with a micro-molar dissociation constant and that was active in cell culture models. (Gurard-Levin et al., 2011; Min et al., 2004c)

Gurard-Levin and colleagues, on the other hand, screened 100000 compounds to identify inhibitors of lysine deacetylase 8 (KDAC8). KDACs (formerly histone deacetylases or HDACs) remove acetyl moieties from the N<sup>ε</sup>-amino group of lysine residues and play a role in regulating gene expression, diabetes, and cancer and therefore have become popular therapeutic targets. They ran every reaction in a single spot of a 384-well plate by mixing compound, enzyme and KDAC8 peptide substrate and incubating them at 37°C for 1 h. The reaction mixtures were then transferred onto the array plates and kept at room temperature in a humidified chamber for 1 h to allow the substrate/product peptides to immobilize to the maleimide-terminated monolayers. The array plates were then rinsed, treated with matrix and analyzed with a MALDI TOF/TOF spectrometer using an automated protocol (20 kV accelerating voltage, positive reflector mode, 200 laser shots per spot). Then, they evaluated each spectrum obtained and found that 48 compounds showed greater than 50% inhibition, and they also identified one compound, which gave nearly complete inhibition. (Gurard-Levin et al., 2011)



**Figure 1.14.** High Throughput Assays and drug discovery with SAMDI MS. Self-assembled monolayers for matrix-assisted laser desorption/ionization (SAMDI) were used to perform a screen of 10000 small molecules to identify inhibitors of the anthrax lethal factor (LF) protease. A peptide substrate for LF was immobilized to a monolayer presenting maleimide groups (a). Treatment of the monolayer with recombinant protease resulted in cleavage of the peptide, which was then analyzed by SAMDI mass spectrometry (MS) (b). Chemical screens were performed by arraying 100 droplets containing the protease and eight compounds from the library, followed by analysis of the spots with MS (c). The analysis clearly identified the spots that had an inhibitor in the reaction mixture (d). Automation devices used in SAMDI analysis. From the top: liquid handler Tecan EVO used to transfer the reaction mixtures onto individual maleimide functionalized gold islands on 384-plates (center) and software snapshot (bottom right corner) of the MALDI MS automated mode analyzer (e).

### 1.3. Aim of the work

With this project we aim to take advantage of the benefits that come from using SAMDI MS technique in order to develop a generalizable screening platform to evaluate the best target sequence for protein modification.

In protein modification reactions we know that even in the presence of a uniquely reactive amino acid residue, bioconjugation often does not reach full conversion and the addition of excess reagent is not always convenient and advantageous. Taking this into account, we intended to find a straightforward way to optimize the conversion level by simply screening and spanning all the amino acid combinations upstream and downstream of the target residue.

Therefore first of all we selected a bioconjugation reaction model, and among all the protein modification strategies that target native amino acid residue that can be introduced without specialized expression system we selected the one developed by Francis M. B. and co-workers that selectively modifies tyrosine residues using cerium(IV) ammonium nitrate (CAN) as one-electron oxidant.

For the purpose of investigate if upstream and downstream residues are significant in determining the reactivity of proteins toward CAN-mediate tyrosine oxidative modification, we generated a 361-peptide library of hexamers with sequence AcGX<sub>1</sub>YZGC (where X and Z represent all the possible 19 amino acids, cysteine excluded).

To obtain short sequences that lead to optimal conversion levels, an efficient method for the evaluation of all possible amino acid combinations was needed and we considered to exploit SAMDI MS advantageous features such as the throughput, the possibility of using a selective immobilization chemistry to control our peptide density and, in turn, to ensure a uniform density of distinct species across the entire array in order to estimate the reaction yield.

We then efficiently immobilized the 361-peptide library onto a maleimide functionalized Self-Assembled Monolayer (through the specific interaction between the thiol group of the C-terminal cysteine and the maleimide moiety); then we tested it under CAN-mediated oxidative conditions, which allowed us to identify the sequences that display a highly-reactive motif toward the bioconjugation reaction under consideration.

With this study we also provide an important example of how synthetic peptide libraries can accelerate the discovery and optimization of protein bioconjugation strategies.



## 2. Materials and Methods

## 2.1. Ac-GYGC, Ac-GYGGC and Ac-GGYGGC peptides

**Reagents.** Regular Fmoc-amino acids and Fmoc-Rink amide resin were purchased from Anaspec. Trifluoroacetic acid (TFA), Triethylsilane (TES), ethanedithiol (EDT), piperidine, and N-methylmorpholine (NMM) were purchased from Sigma Aldrich. EZ-Link Maleimide-PEG<sub>2</sub>-Biotin was purchased from Thermo Scientific, while N-Biotinyl-N'-cysteinyl Ethylenediamine Trifluoroacetic Acid Salt (BCE) from Toronto Research Chemicals. Streptavidin was purchased from New England BioLabs. All other peptide synthesis reagents were purchased from VWR, unless otherwise noted. SAMDI-TOF mass spectrometry was performed on an Applied Biosystems 4800 TOF/TOF mass spectrometers using automated data acquisition.

**Peptide Synthesis.** Ac-GYGC, Ac-GYGGC and Ac-GGYGGC peptides were synthesized manually using standard Fmoc solid phase peptide synthesis using Fmoc Rink-Amide resin previously reported (Houseman and Mrksich 1998). Briefly, the deprotection steps were performed using 20% piperidine solution in dimethylformamide (DMF) for 10 minutes. The amino acids were coupled using 4-molar equivalents of Fmoc-amino acids, benzotriazol-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate (PyBop) and N-methylmorpholine (NMM) for 20 minutes. Peptides were cleaved and deprotected using a 95% TFA, 2.5% water, 1.25% triethylsilane (TES), and 1.25% ethanedithiol (EDT) solution for 2 hours at room temperature. The resin was filtered and the peptides were precipitated in cold diethylether, followed by centrifugation and an ether wash. Peptides were dissolved in water, and then lyophilized. The lyophilized peptide powder was dissolved in 0.1% aqueous TFA solution and then used for immobilization.

**Preparation of gold substrates.** Metal plates of the same dimensions as the Applied Biosystems 4800/5800 MALDI inserts were cleaned in hexanes, followed by ethanol, water, and then ethanol again, and then dried under nitrogen. Titanium (3nm) and gold (30 nm) were deposited in a 384 gold circle array using aluminum masks by an electron beam evaporator (Thermionics). The gold plates were then immersed in a 10% maleimide/triethylene glycol alkyldisulfide solution for 48 hrs at 4°C for the SAM formation, preparation adapted from previous report (Houseman, Gawalt, and Mrksich 2003). The plates were then rinsed with ethanol, water, ethanol, dried under nitrogen and then used for peptide immobilization followed by the oxidative reaction.

**SAMD1 Mass Spectrometry.** The gold coated plates were treated with 2, 4, 6 trihydroxyacetophenone (THAP) matrix, 30 mg/mL in acetone, air-dried and then fixed to the sample holder for the Applied Biosystems™ 4800 plus MALDI TOF/TOF mass spectrometer. A 355 nm Nd:YAG laser was used as the desorption/ionization source with an accelerating voltage of

20 kV and extraction delay time of 50 ns. All spectra were acquired in positive reflector mode, using an automatic acquisition method and collecting 400 shots/spot.

**Peptide Immobilization.** 2  $\mu\text{L}$  of 0.1 mM peptide solution in 10 mM TRIS buffer pH 7.0, were transferred to the gold array metal plates using a Multidrop<sup>TM</sup> Combi Reagent Dispenser (Thermo Scientific), let immobilize for 30 minutes at room temperature, rinsed with water and ethanol, and then dried. The peptide plates were then spotted with matrix (THAP, 30 mg/mL acetone) analyzed by SAMDI to get an accurate control baseline and then rinsed and dried.

**CAN-mediated Oxidative Reaction.** The assay was performed directly on Ac-GYGC, Ac-GYGGC or Ac-GGYGGC functionalized SAMs. The reaction mixture was prepared with TRIS buffer (10 mM, pH 7.0), 4-methoxy-*N,N*-dimethylaniline (5  $\mu\text{L}$  of a 10 mM solution in DMF, 500  $\mu\text{M}$  final concentration) and fresh cerium(IV) ammonium nitrate (CAN) in ddH<sub>2</sub>O (final concentrations were varied as follows: 0.1, 0.5, 1 and 1.5 mM), briefly vortexed and then transferred onto the plate using a Combi liquid handling robot. The control spots were treated with 10 mM TRIS buffer, pH 7.01. The reaction was allowed to sit at room temperature for 1 h and then stopped by rinsing and drying the plate. The post-reaction plates were analyzed by SAMDI using automatic acquisition mode.

**Ac-GYGGC and Ac-GGYGGC biotinylation.** Ac-GYGGC and Ac-GGYGGC were incubated for 1 h at room temperature with EZ-Link<sup>®</sup> Maleimide-PEG<sub>2</sub>-Biotin (1:2 molar ratio) in 50 mM TRIS buffer, pH 7.01. The peptide/biotinylation agent mixture was then run through an HPLC with a 2 solvent gradient system. (100% DIUF with 0.1% TFA, and 75% Acetonitrile 25% water 0.1% TFA) Once the aliquots with the biotinylated peptide were found, they were lyophilized, weighed and with them peptide stocks were made.

**SAM biotinylation and streptavidination.** A 10 mM stock solution of N-Biotinyl-N'-cysteinyl Ethylenediamine Trifluoroacetic Acid Salt (BCE) was prepared in DIUF, aliquoted into 10  $\mu\text{L}$  stocks and freeze in -80°C. When ready to use BCE was diluted 1:1000 with 50mM TRIS buffer pH 8 (final concentration 10  $\mu\text{M}$ ) and applied to the SAMDI surface. BCE was allowed to react for 25 min. The surface was washed with ethanol, water and ethanol again.

A 1 mg/ml streptavidin solution in 50 mM TRIS buffer pH 8 was diluted 3:100 (30  $\mu\text{L}$  of Streptavidin diluted with 970  $\mu\text{L}$  of 50 mM Tris pH 8) and applied to the SAMDI plate. After 20 min incubation the surface was rinsed with DIUF water only. The biotinylated peptide reacted with the anisidine derivative was applied to the streptavidinated monolayer for 10 minutes and again

washed with water only. DHAP matrix drops were deposited onto each spot and the plate was read in MALDI-TOF instrument.

**Oxidative Reaction in Solution.** The biotinylated peptides were dissolved in 10 mM TRIS buffer pH 7.01, 0.1 mM final concentration. To this solution were added in order: 4-methoxy-*N,N*-dimethylaniline (500  $\mu$ M final concentration) and fresh cerium(IV) ammonium nitrate (CAN) in DIUF H<sub>2</sub>O (final concentrations: 0.1 mM and 1.5 mM), briefly vortexed and incubated for 1h at room temperature. To quench the reaction was then used 3 mM neutralized tris(2-carboxyethyl) phosphine hydrochloride (TCEP) in water. The control spots were treated with 10 mM TRIS buffer, pH 7.01. The post-reaction mixture was desalted, concentrated, and purified with C18 ZipTip® Pipette Tips (Millipore) and then analyzed with MALDI MS. Eventually the solution was spotted onto a biotin-streptavidin functionalized monolayer for the peptide immobilization (10 min), rinsed with water, dried with nitrogen, treated with DHAP matrix and then analyzed with SAMDI MS.

**Oxidative Reaction Control.** 0.1 mM Ac-GRK<sup>Ac</sup>FGC-biotin and Ac-GRK<sup>Ac</sup>FGC-biotin peptides (previously synthesized in Prof. Mrksich lab) were reacted with 0.5 mM 4-methoxy-*N,N*-dimethylaniline and 1.5 mM cerium(IV) ammonium nitrate (CAN), briefly vortexed and incubated for 1h at room temperature. 3 mM neutralized tris(2-carboxyethyl) phosphine hydrochloride (TCEP) in water to quench the reaction. The post-reaction mixture was desalted, concentrated, and purified with C18 ZipTip® Pipette Tips (Millipore) and then analyzed with MALDI MS, after treatment with DHAP matrix.

**Revised CAN-mediated Oxidative Reaction.** The assay was performed directly on Ac-GYGGC and Ac-GGYGGC functionalized SAMs. The reaction mixture was prepared with BIS TRIS buffer (50 mM, pH 6.2), 4-methoxy-*N,N*-dimethylaniline (0.1 mM final concentration) and fresh cerium(IV) ammonium nitrate (CAN) solution in ddH<sub>2</sub>O (final concentration 1 mM), briefly vortexed and then transferred onto the plate using a Combi liquid handling robot. Three different controls were prepared: one with 50 mM BIS TRIS buffer pH 6.2, the second with BIS TRIS buffer and 0.1 mM 4-methoxy-*N,N*-dimethylaniline and the third one with BIS TRIS and 1 mM CAN. The reactions were allowed to sit at room temperature for 10 minutes and then stopped by rinsing and drying the plate. The post-reaction plate was analyzed by SAMDI using automatic acquisition mode.

## 2.2. Ac-GGYXGC and Ac-GXYGGC peptide libraries

**Reagents.** Unless specified otherwise, laboratory chemicals and reagents were purchased from Sigma Aldrich and used without additional purification. Peptide synthesis reagents, including Fmoc amino acids and Rink-amide resin, were purchased from Anaspec. MALDI-TOF and SAMDI-TOF mass spectrometry were performed on an Applied Biosystems 4800 TOF/TOF mass spectrometer by manual and automated protocols.

**Peptide Synthesis.** Peptides were synthesized manually using standard Fmoc solid phase peptide synthesis using Fmoc Rink-Amide MBHA resin as described previously (Houseman and Mrksich 1998). Briefly, the deprotection steps were performed using 20% piperidine solution in dimethylformamide (DMF) for 10 minutes. The amino acids were coupled using 4-molar equivalents of Fmoc-amino acids, benzotriazol-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate (PyBop) and *N*-methylmorpholine (NMM) for 20 minutes. To minimize the synthesis time, the first two amino acids were coupled in a big batch and then the resin was split in 2 portions that would generate our two libraries. One of the 2 portions was used to couple a glycine residue as third amino acid, a tyrosine (as fourth) and then it was split again into another 19 portions for the coupling of the final two amino acids (X and acetylated glycine). The other portion as first step was instead split into 19 parts for the coupling of the X residue, followed by the last three amino acids: tyrosine, glycine and acetylated glycine. The coupling of the X residue in both the libraries was performed using 4-molar equivalents of 1-Hydroxybenzotriazole hydrate (HOBt) and *N,N'*-Diisopropylcarbodiimide (DIC) instead of PyBop and NMM, for 2 hrs. Peptides were cleaved and deprotected using a 95% TFA, 2.5% water, 1.25% TES, and 1.25% EDT solution for 2 hours at room temperature. The resin was filtered and the peptides were precipitated in cold diethylether, followed by centrifugation and an ether wash. Peptides were dissolved in water then lyophilized. The lyophilized peptide powder was dissolved in 0.1% aqueous TFA solution. The two 19-peptides libraries were aliquoted into a 96-well plate, diluted 5 fold in BIS TRIS buffer (50 mM, pH 6.2) and then used for immobilization.

**Preparation of gold substrates.** Metal plates having a 24 by 16 array of gold islands were made in the geometry of a conventional 384-well format used in the Applied Biosystems TM 4800 MALDI plus TOF/TOF spectrometer. The plates were cleaned by soaking in hexanes for 10 min, in absolute ethanol for 10 min, in water, again in ethanol and then dried under a stream of nitrogen. Copper stencils, having the same size as the metal plate, were machined to have a 24 x 16 array of circular holes 2 mm in diameter. The clean metal slides were covered with the stencils to allow the circular

regions of the glass to be exposed to the evaporator. Titanium (5 nm) and gold (50 nm) were evaporated at a rate of 0.05-0.10 nm s<sup>-1</sup> and at a pressure of  $5.0 \times 10^{-7}$  Torr. Monolayers were prepared by immersing the gold-patterned plates in an ethanolic solution of maleimide-terminated disulfides and tri(ethylene glycol)-terminated disulfide in a ratio of 1:4 for 48 h at 4° C (total concentration of disulfide: 1 mM). The substrates were then washed with ethanol and dried under nitrogen.

**SAMDY Mass Spectrometry.** The post reaction plates were treated with 2, 4, 6 trihydroxyacetophenone (THAP) matrix, 30 mg/mL in acetone, air-dried and then fixed to the sample holder for the mass spectrometer. Mass analysis was performed using a 4800 MALDI-TOF/TOF (Applied Biosystems). A 355 nm Nd:YAG laser was used as a desorption/ionization source, and all spectra were acquired with 20 kV accelerating voltage using positive reflector mode. The extraction delay was 50 ns, and 400 shots/spot were applied.

**Peptide Immobilization.** 2 µL of each one of the 38 peptides was transferred to the gold array metal plates using a multichannel pipette, let immobilize for 30 minutes at room temperature, rinsed with ethanol, water and ethanol again, and then dried under nitrogen gas. The peptide plates were then treated with the reaction mixture, spotted with matrix (THAP, 30 mg/mL acetone) and analyzed by SAMDY. The SAMDY data was then exported and analyzed using the Profiler software to obtain the control baseline.

**CAN-mediated Oxidative Reaction.** The assay was performed on SAMs functionalized with the two libraries Ac-GGYXGC and Ac-GXYGGC. The reaction mixture consisted of BIS TRIS buffer (50 mM, pH 6.2), 4-methoxy-*N,N*-dimethylaniline (0.1 mM final concentration) and fresh cerium(IV) ammonium nitrate (CAN) solution in H<sub>2</sub>O (final concentration 1 mM), was briefly vortexed and then transferred onto the plate using a Combi liquid handling robot. The reaction was allowed to sit at room temperature for 10 minutes and then stopped by rinsing and drying the plate. The post-reaction plates were analyzed by SAMDY using automatic acquisition mode. The data was exported and then analyzed using an in-house Profiler software and Microsoft Excel in order to calculate the relative activity of each peptide by inputting the area under curves (AUCs) for the acetylated product peak and the substrate peak into the following formula:  $AUC_{Product}/(AUC_{Product} + AUC_{Substrate})$ . AUCs were calculated using Applied Biosystems Data Explorer software after baseline subtraction.

### 2.3. Ac-GXYZGC peptide library

**Reagents.** SynPhase D-series PA lanterns were purchased from Mimotopes, while regular Fmoc-amino acids and hydroxybenzotriazole (HOBt) from Anaspec. Trifluoroacetic acid (TFA), Triethylsilane (TES), piperidine, and diisopropylcarbodiimide (DIC) were purchased from Sigma Aldrich. SAMDI-TOF mass spectrometry was performed on Applied Biosystems 5800 MALDI TOF/TOF mass spectrometer using automated data acquisition.

**Synthesis of Peptide Library.** Parallel synthesis of 361 peptides was performed in 96-well filter plates with commercially available SynPhase D-series PA lanterns by using standard HOBt/DIC coupling protocols. Briefly, a lantern with 9-fluorenylmethoxy-carbonyl (Fmoc) Rink Amide linker was placed in each well and the Fmoc deprotected with 20% piperidine in DMF (30 min). Lanterns were filtered and rinsed 5 times with DMF on a filtration vacuum manifold (Millipore), activated with HOBt/DIC and coupled to an Fmoc-amino acid. The coupling and deprotection steps were then carried out until the desired sequences were synthesized. For coupling, the lantern in each well was incubated for at least two hours with the activated amino acid solution in DMF (15-molar equivalents of amino acid, 18.8-molar equivalents of hydroxybenzotriazole (HOBt), and 15-molar equivalents of diisopropylcarbodiimide (DIC); amino acid/HOBt/DIC 1:1.2:1, mole ratio) followed by 5 DMF washing steps. After the final coupling the last amino acid (acetylated glycine), the lantern in each well was washed 5 times with DMF, 3 times with dichloromethane (DCM) and dried under vacuum overnight. During the synthesis process, all solutions were drained completely from the filter plate using a Multiscreen vacuum filtration manifold (Millipore). The lanterns were then placed in 96-well plates and the synthesized peptides were cleaved from each lantern using the cleavage solution (TFA/water/TES 95:2.5:2.5, v/v/v) for 2 hrs at room temperature. After removing the cleavage solution under a gentle stream of nitrogen, the cleaved peptides were retrieved by dissolving in 0.1 % TFA, and lyophilized. The remaining solids were then suspended in 0.1% TFA (in deionized ultrafiltered, DIUF), diluted 1:10 in 50 mM BIS TRIS pH 6.2, transferred in a 384 plate and stored at -80°C. The peptide purity was checked by MALDI MS.

**Mass Spectrometry.** Mass analysis was performed using a 5800 MALDI-TOF/TOF (Applied Biosystems). A 355 nm Nd:YAG laser was used as a desorption/ionization source, and all spectra were acquired with 20 kV accelerating voltage using positive reflector mode. The extraction delay was 50 ns, 400 laser shots were applied, and the entire surface of the circle was sampled.

**Preparation of gold substrates and monolayer.** Stainless steel plates (8 X 12.3 cm) were washed in individual baths of hexanes, ethanol, deionized ultra filtered (DIUF) water, ethanol again, and

then dried under nitrogen. Copper stencils, having the same size as the metal plate, were machined to have the geometry of a conventional 384-well format used in the Applied Biosystems TM 5800 MALDI plus TOF/TOF spectrometer (24 x 16 array of circular holes 2 mm in diameter). The clean metal slides were covered with the stencils to allow the circular regions of the glass to be exposed to the electron beam evaporator (Thermionics). Titanium (3 nm) and gold (30 nm) were evaporated at a rate of 0.05-0.10 nm s<sup>-1</sup> and at a pressure of 5.0 × 10<sup>-7</sup> Torr. Self-assembled monolayers of mixed alkylthiolates on gold were prepared and analyzed according to previously reported methods (Gurard-Levin et al. 2011). Briefly, the gold-patterned plates were immersed in an ethanolic solution of maleimide-terminated disulfides and tri(ethylene glycol)-terminated disulfide in a ratio of 1:4 (total concentration of disulfide: 1 mM) for 48 h at 4°C. The maleimide-terminated acyl thiolates density was 10% against an inert background of triethylene glycol. The substrates were then washed with ethanol, water, again ethanol and dried under nitrogen. The library was then immobilized.

**Peptide Immobilization.** 2 µL of each one of the 361 peptides were transferred to the gold array metal plates using a 96-well head liquid handler (Tecan Freedom EVO 200), let immobilize for 30 minutes at room temperature, rinsed with ethanol, water and ethanol again, and then dried under nitrogen gas. The peptide plates were then treated with the reaction mixture, spotted with matrix (THAP, 30 mg/mL acetone) and analyzed by SAMDI.

**CAN-mediated Oxidative Reaction.** The assay was performed on SAMs functionalized with the whole library Ac-GXYZGC. The experiment was run in triplicate. The reaction mixture consisted of BIS TRIS buffer (50 mM, pH 6.2), 4-methoxy-*N,N*-dimethylaniline (0.1 mM final concentration) and fresh cerium(IV) ammonium nitrate (CAN) solution in DIUF H<sub>2</sub>O (final concentration 1 mM), was briefly vortexed and then transferred onto the plate using a Combi liquid handling robot. The reaction was allowed to sit at room temperature for 10 minutes and then stopped by rinsing and drying the plate. The control plates were analyzed with MALDI MS right after the library immobilization and after incubation with 50 mM BIS TRIS buffer pH 6.2 and 1 mM CAN for 10 minutes. The post-reaction plates were analyzed by SAMDI using automatic acquisition mode. The data was exported and then analyzed using an in-house Profiler software and Microsoft Excel in order to calculate the relative activity of each peptide by inputting the area under curves (AUCs) for the acetylated product peak and the substrate peak into the following formula:  $AUC_{Product}/(AUC_{Product} + AUC_{Substrate})$ . AUCs were calculated using Applied Biosystems Data Explorer software after baseline subtraction.



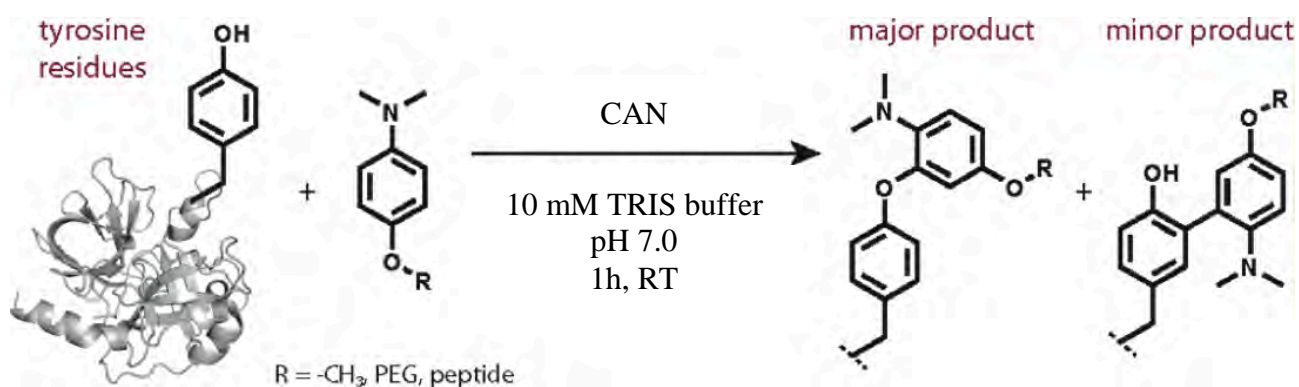
## 3. Results

### 3.1. Development of reactions for protein bioconjugation

This project was aimed at finding a reliable and efficient strategy for the straightforward evaluation and selection of amino acid sequences displaying a preferential reactivity toward a desired bioconjugation reaction. Therefore we first defined the bioconjugation reaction to be used as a model for the screening. We needed a reaction capable of creating precise and well-defined linkages in aqueous solution under mild pH, temperature conditions and at low substrate concentrations.

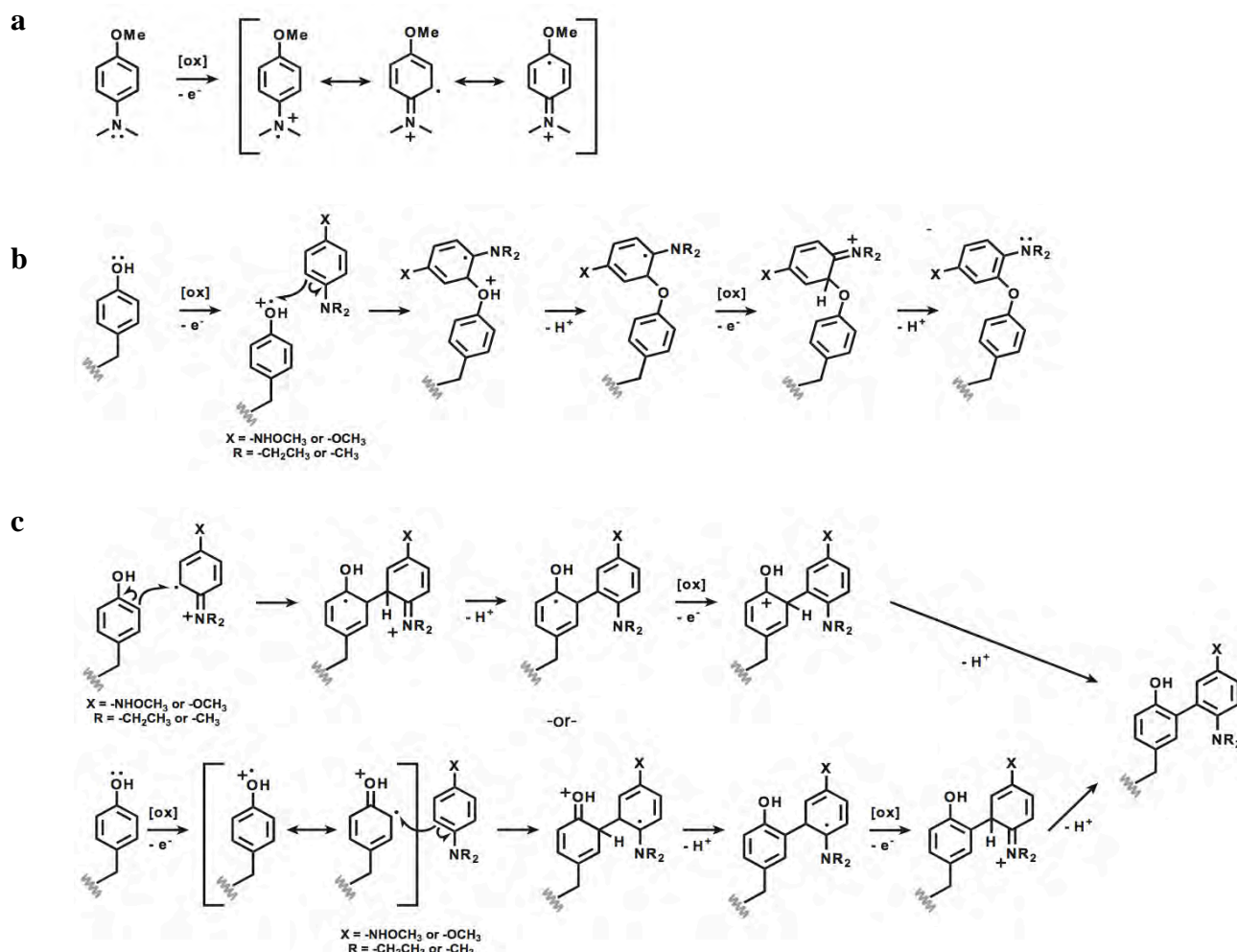
For this purpose, we selected the Cerium(IV) ammonium nitrate (CAN)-mediated oxidative modification reaction that targets electron-rich amino acids, developed by Francis and co-workers. They studied the coupling of tyrosine and tryptophan residues to phenylene diamine and anisidine derivatives using CAN as oxidant reagent (Seim et al., 2011). While *N-acyl* phenylene diamine derivative was found to modify both tryptophan and tyrosine residues, alkylated anisidine selectively modified tyrosines. For this reason, we chose as our model reaction the CAN-mediated coupling between 4-methoxy-*N,N*-dimethylaniline and tyrosine residues (Figure 3.1).

Francis and colleagues also tuned the reaction conditions to make them broadly applicable for protein modification: they tested different buffers also assaying buffers of increasing strength so as ensure constant and mild pH, and defined both the best oxidant and coupling partner concentration. In more detail, the optimized reaction mixture for the modification of tyrosine consisted of 100  $\mu\text{M}$  solution of the tyrosine-containing peptide in 10 mM TRIS buffer (pH 7.0), 500  $\mu\text{M}$  anisidine derivative in dimethylformamide (DMF), 1.5 mM cerium(IV) ammonium nitrate (CAN) in ddH<sub>2</sub>O. The reaction mixture was allowed to sit at room temperature for 1 h; then, a 3 mM neutralized tris(2-carboxyethyl) phosphine hydrochloride (TCEP) solution in ddH<sub>2</sub>O was added to quench it.



**Figure 3.1.** Tyrosine modification reaction scheme.

Although the mechanisms for these modifications are currently under investigation, the formation of the observed products can be rationalized on the basis of radical coupling pathways. In detail, we report the current proposed mechanisms for formation of observed products below in figure 3.2.



**Figure 3.2.** Proposed mechanisms for the formation of the observed products. Either the amino acid side chain or the coupling partner could be oxidized. Resonance considerations can be used to rationalize the observed addition locations. (a) Oxidation of anisidine coupling partner; (b) tyrosine modification through O-C bond formation. To avoid the formation of 9-electron intermediates the *O*-alkylation of tyrosine should proceed through the oxidation of the phenolic group first. (c) Tyrosine modification through C-C bond formation. In this case, *C*-alkylation of tyrosine could proceed through the initial oxidation of either the coupling partner or the phenolic side chain.

After we chose the bioconjugation reaction to test, before generating a library to be screened, we first assessed the reaction on the model tetrapeptide Ac-GYGC (where Ac indicates an Acetyl group) that we synthesized by solid phase synthesis.

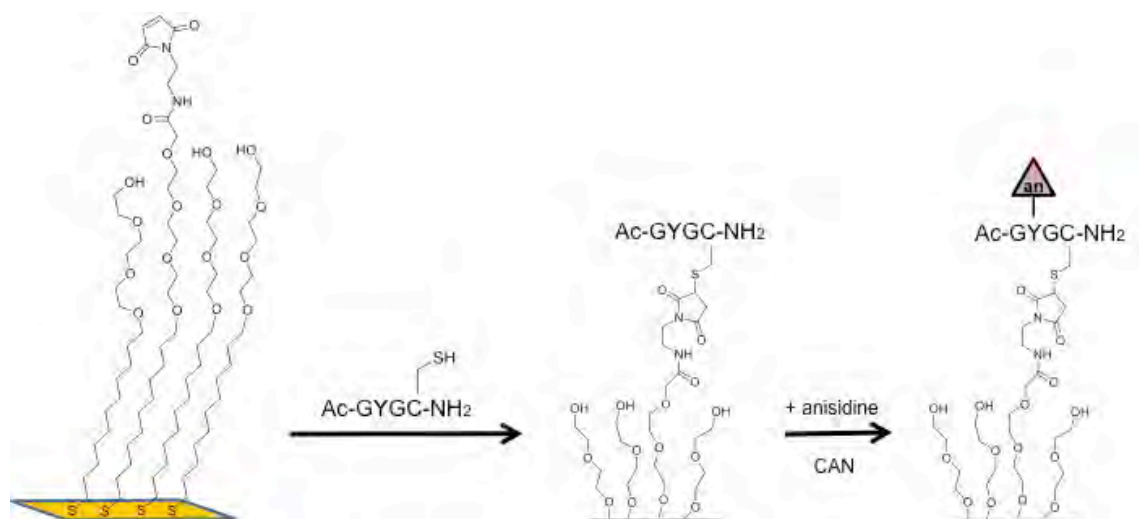
## 3.2. Ac-GYGC

### 3.2.1. Solid phase synthesis and immobilization

In designing our model peptide sequence, we needed a C-terminal cysteine for its immobilization onto the maleimide functionalized monolayer, a tyrosine, which is the reaction target residue and a simple amino acid as spacer, so we selected glycine and its acetylated form as N-terminal residue. The peptide Ac-GYGC was synthesized using standard Fmoc solid phase peptide synthesis protocol and dissolved in 10 mM TRIS buffer pH 7.01. Subsequently, a solution of the thiol-terminated peptide (0.1 mM) was applied to a maleimide-presenting monolayer onto SAMDI array plates for 30 min at room temperature and then rinsed to remove salts and other solution components. The plate was dried under nitrogen, treated with a solution of 2,4,6-trihydroxyacetophenone matrix (THAP) and analyzed by SAMDI mass spectrometry. The SAMDI spectra obtained showed that our peptide Ac-GYGC was efficiently immobilized to the monolayer through the cysteine-maleimide bond formation, as they revealed a peak at  $m/z$  1312 that corresponds to the sodium adduct of the peptide-substituted alkyl disulfide (Figure 3.4a).

### 3.2.2. Immobilized Ac-GYGC oxidative modification

Once we got Ac-GYGC immobilized onto the surface, we treated it with the optimized reaction mixture for tyrosine modification reported by Seim and colleagues, which consists of 500  $\mu$ M 4-methoxy-*N,N*-dimethylaniline and 1.5 mM cerium(IV) ammonium nitrate (CAN) in 10 mM TRIS buffer, pH 7.01 (Seim et al., 2011). The reaction mixture was allowed to sit at room temperature for 1 h and then the gold substrates were rinsed with ethanol, water and ethanol, dried under nitrogen, spotted with matrix (DHAP) and analyzed by SAMDI (Figure 3.3).

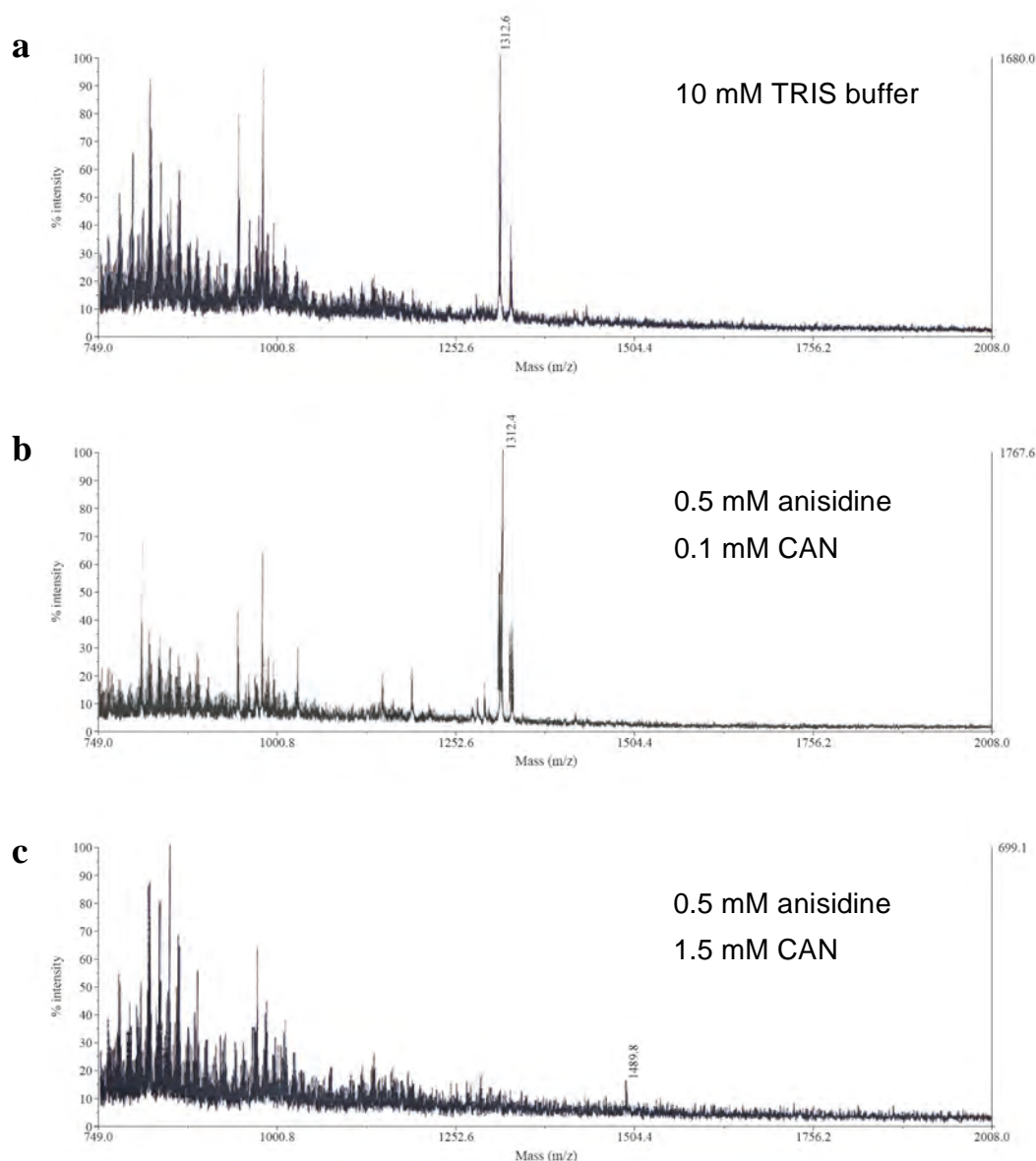


**Figure 3.3.** Immobilized Ac-GYGC oxidative modification scheme. Cysteine-terminate peptides are immobilized onto maleimide terminated SAMs, treated with the reaction mixture and analyzed with SAMDI MS. Treatment of the peptide-SAMs with CAN and anisidine results in the tyrosine-anisidine coupling corresponding to a mass shift of +149.

We expected a product peak corresponding to the peptide-substituted alkyldisulfide with a mass shift of 149 given by the anisidine derivative coupling.

Although the peptide was efficiently immobilized onto the monolayer as figure 3.4a clearly shows, the oxidative conditions seemed to somehow cause the detachment of the Ac-GYGC-substituted alkyl disulfide, resulting in the signal loss for both the peptide and the hypothetical product (Figure 3.4c).

Since SAMDI usually allows the use of low reagent concentrations, different CAN concentrations (0.1 mM, 0.5 mM, 1 mM) were also tested, but with none of them the product peak appearance was noticed.



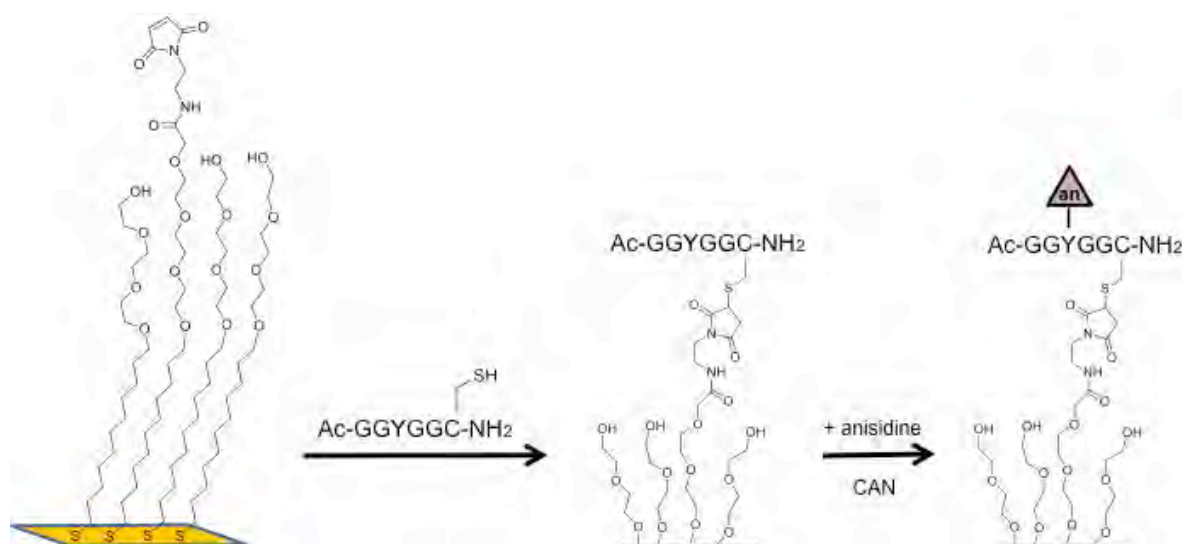
**Figure 3.4.** SAMDI spectra of immobilized Ac-GYGC ( $m/z$  1312) (a) in the presence of 10 mM TRIS buffer pH 7.0, (b) after 1 h treatment with 0.1 mM CAN and 0.5 mM 4-methoxy-*N,N*-dimethylaniline (anisidine derivative) in TRIS buffer and (c) after incubation with the optimized reaction mixture developed by Seim and colleagues: 1.5 mM CAN, 0.5 mM anisidine derivative in TRIS buffer.

For this reason before drawing any conclusion we decided to synthesize two different peptides with additional glycines as spacers: Ac-GGYGGC and Ac-GYGGC.

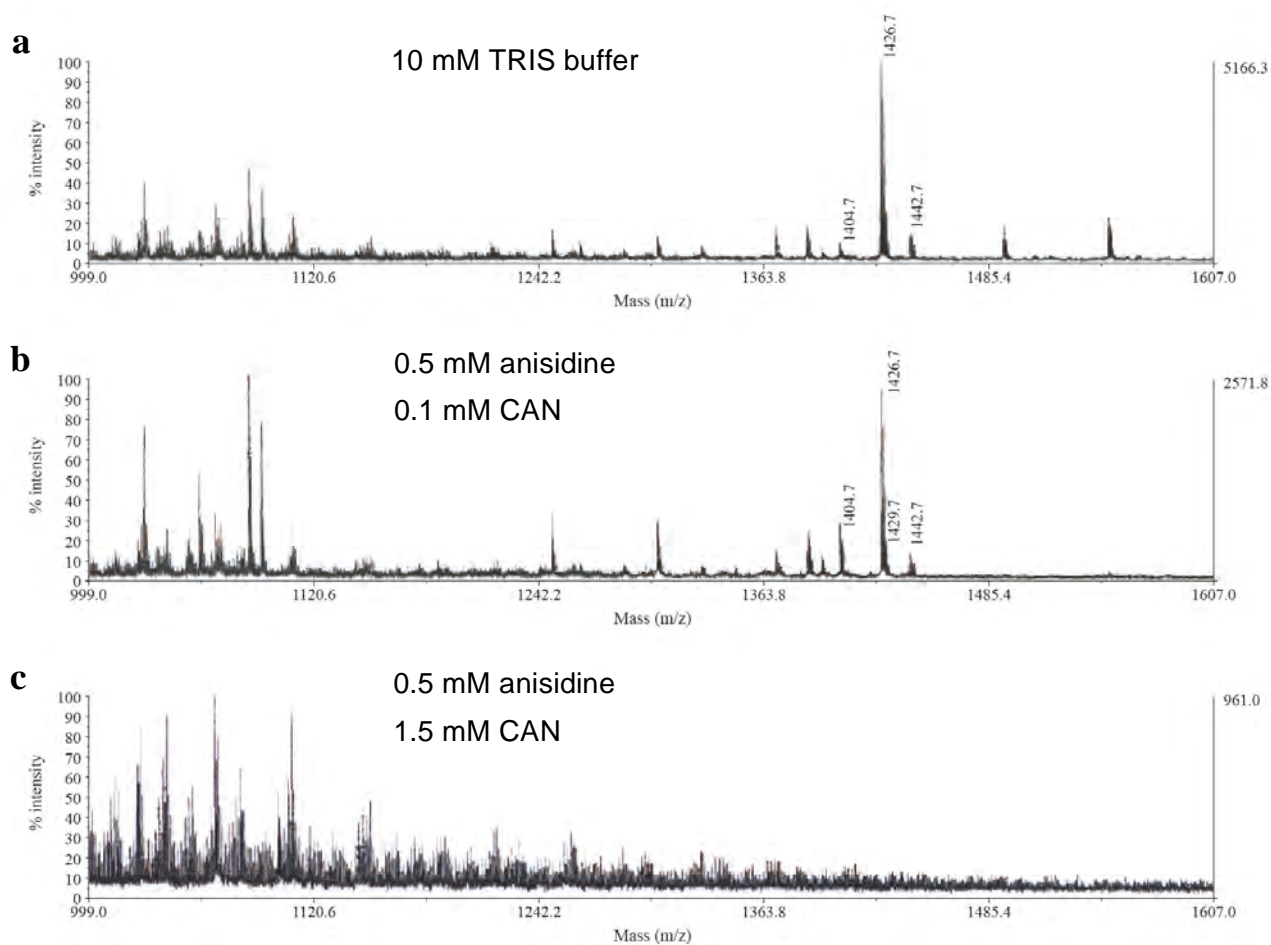
### 3.3. Ac-GYGGC and Ac-GGYGGC synthesis, immobilization and modification

As a model molecule for testing our derivatization reaction, we produced two peptides. In one case glycine was inserted downstream of tyrosine, in the second, both upstream and downstream. Peptide synthesis was performed following again the standard Fmoc solid phase procedure and both peptides Ac-GYGGC and Ac-GGYGGC were then immobilized onto the maleimide-functionalized monolayer. Once they were anchored to the SAM, they were treated with the reaction mixture for 1 h at room temperature. Upon reaction completion, the monolayers were rinsed and SAMDI spectra were acquired to check if the tyrosine modification had occurred as expected (Figure 3.5).

Exactly as observed in the case of Ac-GYGC, both Ac-GYGGC and Ac-GGYGGC could be effectively immobilized onto the monolayer through the maleimide-cysteine bond, but when they were incubated with the reaction mixture containing a high oxidant concentration, they seemed to desorb from the surface and no product peak was displayed. Moreover, when using the CAN concentration allowing the best reaction performance, the monolayer appeared to be clearly damaged, as shown in figure 3.6c, where data regarding Ac-GGYGGC are presented.



**Figure 3.5.** Immobilized Ac-GGYGGC oxidative modification scheme. Cysteine-terminate peptides are immobilized onto maleimide terminated SAMs, treated with the reaction mixture and analyzed with SAMDI MS. Treatment of the peptide-SAMs with CAN and anisidine results in the tyrosine-anisidine coupling corresponding to a mass shift of +149.



**Figure 3.6.** Representative SAMDI spectra of immobilized Ac-GGYGGC ( $m/z$  1426) (a) in the presence of 10 mM TRIS buffer pH 7.0, (b) after 1 h treatment with 0.1 mM CAN and 0.5 mM anisidine derivative and (c) after 1h incubation at room temperature with 1.5 mM CAN and 0.5 mM anisidine derivative in TRIS buffer.

As revealed by spectra in figures 3.4c and 3.6c, besides the loss of the immobilized peptide signal and the absence of any product peak, the treatment with high concentrations of the oxidizing reagent (1.5 mM cerium(IV) ammonium nitrate) seems to affect the integrity of the monolayer. For this reason the experiment had to be re-designed and the oxidative tyrosine modification reaction had to be run in solution to prevent undesired monolayer modifications. Of course, this also implied that the reacted peptide had to be immobilized after completion of the reaction. Furthermore, a new immobilization strategy was necessary to anchor our peptides to the surface, since the oxidative conditions can modify the thiol functionality of the cysteine and therefore the cysteine-maleimide interaction was not usable anymore.

Based on this background we managed to bypass the immobilization issue by biotinylating both the peptide and the monolayer, and using streptavidin as linker between them.

### 3.3.1. Ac-GGYGGC biotinylation, oxidative modification and subsequent immobilization

Ac-GGYGGC was modified through the C-terminal addition of EZ-Link<sup>®</sup> Maleimide-PEG<sub>2</sub>-Biotin, which is a sulfhydryl-reactive biotin labeling reagent. The labeling procedure consists of 1 h-incubation at room temperature of a solution composed of the peptide and the biotin reagent dissolved in a 1:2 ratio in 50 mM TRIS buffer, pH 7.0. After the biotinylation process, the peptide was incubated with the reaction mixture (0.1 mM peptide, 0.5 mM anisidine, 0.1 mM or 1.5 mM CAN in 10 mM TRIS buffer, pH 7.0) and after 1 h the reaction was quenched with a neutralized tris(2-carboxyethyl) phosphine hydrochloride (TCEP) solution in ddH<sub>2</sub>O.

The reaction mixture were then desalted, concentrated, purified with ZipTipC18 and finally analyzed with MALDI MS to check if the product had formed as expected.

Looking at the spectra obtained analyzing the biotinylated peptide (Ac-GGYGGC-biotin) before and after the treatment with the reaction mixture containing the oxidant and the coupling reagent (Figure 3.7), we could clearly see that the biotinylated peptide was present in solution (m/z 1079 peak, Figure 3.7a) and that at low CAN concentration it remained unmodified (Figure 3.7b). Furthermore, only the incubation at high CAN concentration (1.5 mM) led to product formation by coupling the anisidine to the tyrosine residues (m/z 1228 peak, Figure 3.7c).

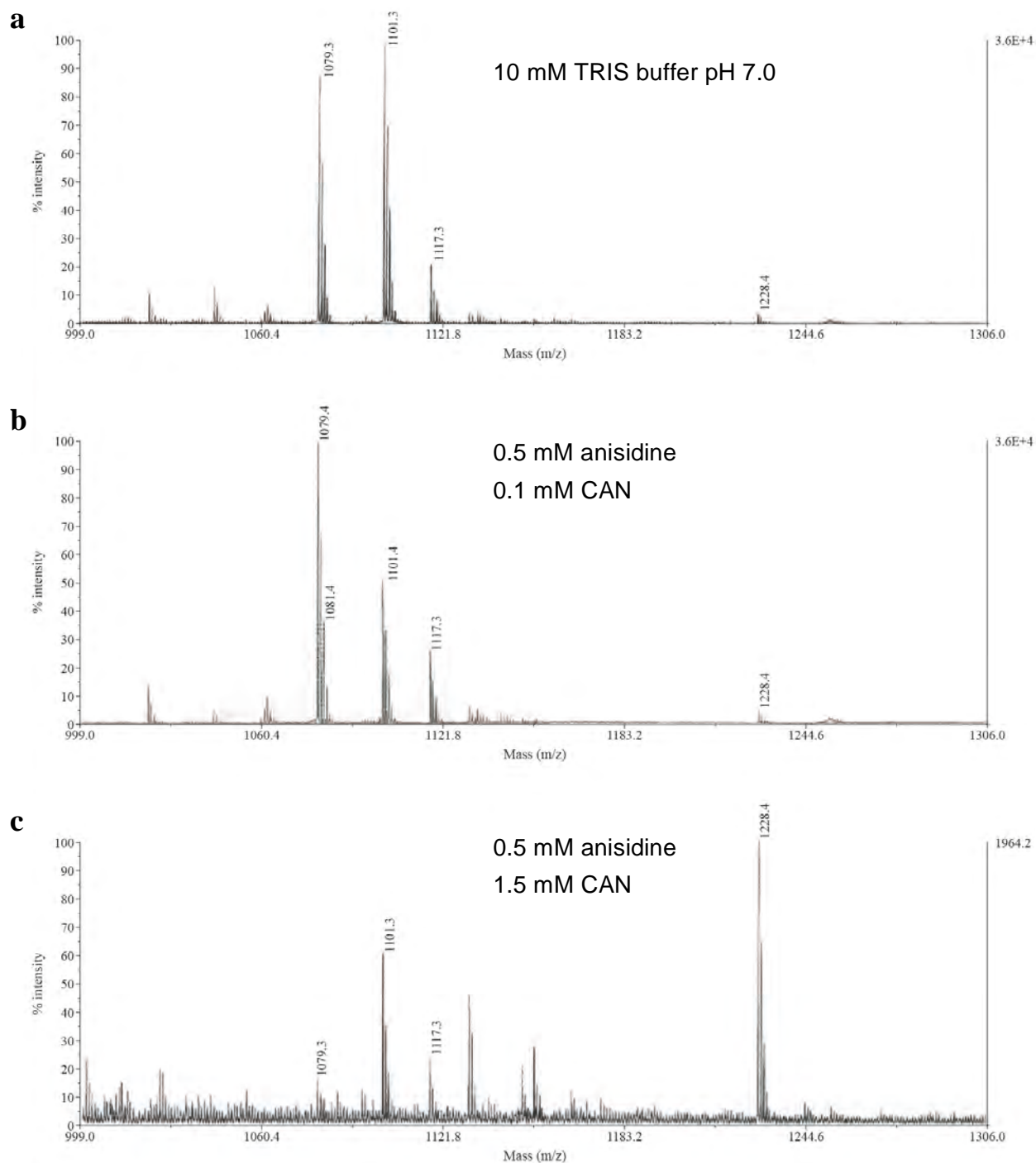
Given that the oxidative modification reaction was actually working on the biotinylated version of Ac-GGYGGC, the next step consisted of reacted peptide immobilization through the specific interaction between biotin and streptavidin.

Monolayer biotinylation was therefore achieved using N-Biotinyl-N'-cysteinyl ethylenediamine that through its sulfhydryl group forms a stable thioether bond with the maleimide groups on the monolayer, resulting in biotin presentation on the surface.

The biotinylated monolayer was then treated with a streptavidin solution for 20 min at room temperature and rinsed with water. Thereafter the reacted peptide was deposited onto the streptavidinated surface and incubated for 10 min. The plate was then rinsed, treated with DHAP matrix and analyzed with MALDI to check if any of the product had been efficiently immobilized (Figure 3.8).

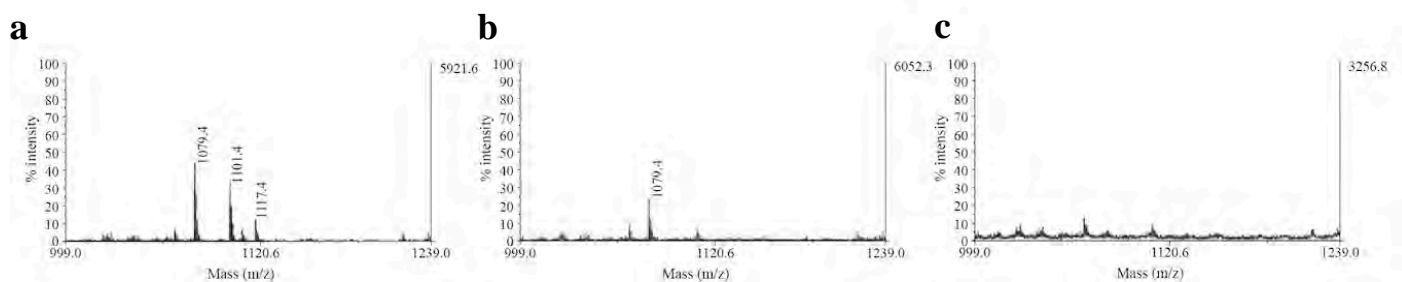
Unexpectedly, even using the same solutions previously analyzed with MALDI MS, in which the product peak had already been detected, no reacted peptide could be immobilized onto the streptavidinated surface. Both the biotinylated Ac-GGYGGC solution in TRIS buffer pH 7.0 and with 0.1 mM CAN and 0.5 mM anisidine, still in TRIS buffer, gave clear peaks for the peptide alone (Figures 3.8a and 3.8b), but incubation with a higher CAN concentration (1.5 mM) resulted in the loss of any signal (Figure 3.8c).





**Figure 3.7.** MALDI spectra of 0.1 mM biotinylated Ac-GGYGGC peptide ( $m/z$  1079) in 10 mM TRIS buffer pH 7.01 (a); after 1 h treatment with the reaction mixture made of 0.5 mM anisidine and 0.1 mM CAN and following reaction quenching with 3 mM TCEP (b); after 1 h incubation at room temperature with the optimized reaction mixture for tyrosine modification: 0.5 mM anisidine, 1.5 mM CAN, 3 mM TCEP added after 1 h to stop the reaction. The reaction product was clearly visible at  $m/z$  1228 (c).

Since apparently the reacted biotinylated peptide Ac-GGYGGC could not be immobilized onto the surface, it was necessary to verify that the mass shift seen in the product peak was actually given by coupling between anisidine and the tyrosine residue present in the peptide sequence and not by non-specific interaction of the anisidine with the C-terminal biotin. For this reason, another biotinylated peptide was reacted with the same solution used previously and then analyzed with MALDI mass spectrometry.

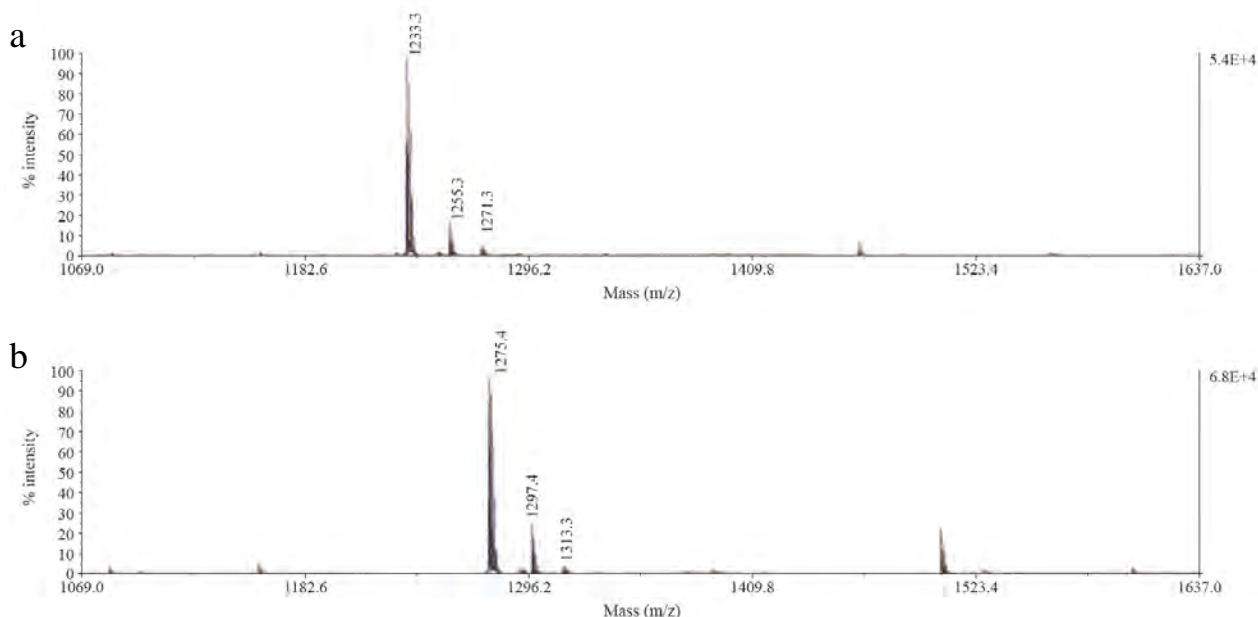


**Figure 3.8.** SAMDI spectra obtained after the reacted peptide immobilization. 0.1 mM biotinylated Ac-GGYGGC (m/z 1079) was immobilized after 1 h treatment with 10 mM TRIS buffer pH 7.0 (a), 0.5 mM anisidine and 0.1 mM CAN with 3 mM TCEP added to quench the reaction (b), and 0.5 mM anisidine, 1.5 mM CAN, 3 mM TCEP (added after 1 h) (c). Even using the solutions already analyzed with MALDI that contained the tyrosine-modified peptides, it wasn't possible to detect any product immobilized onto the surface. There is no peak at m/z 1228.

### 3.3.2. Biotinylated Ac-GRKFGC and Ac-GRK<sup>Ac</sup>FGC control

As a control for the site-specificity of the examined oxidative modification, we tested the optimized reaction conditions on two different biotinylated peptide without any tyrosine residue in their sequence: Ac-GRKFGC and Ac-GRK<sup>Ac</sup>FGC (where in the latter the lysine residue was acetylated). Both peptides were incubated with 0.5 mM anisidine and 1.5 mM CAN in 10 mM TRIS buffer for 1h. Then, the reaction was stopped by TCEP (3 mM), the samples desalted with ZipTipC18 and analyzed with MALDI MS.

None of the analyzed reaction solutions contained a modified peptide with a mass increment of 149 m/z, as shown in figure 3.9. This confirms that the oxidative modification actually occurs on the tyrosine in the peptide sequence but also shows that for some reason it cannot be immobilized onto the monolayer after it has been modified.

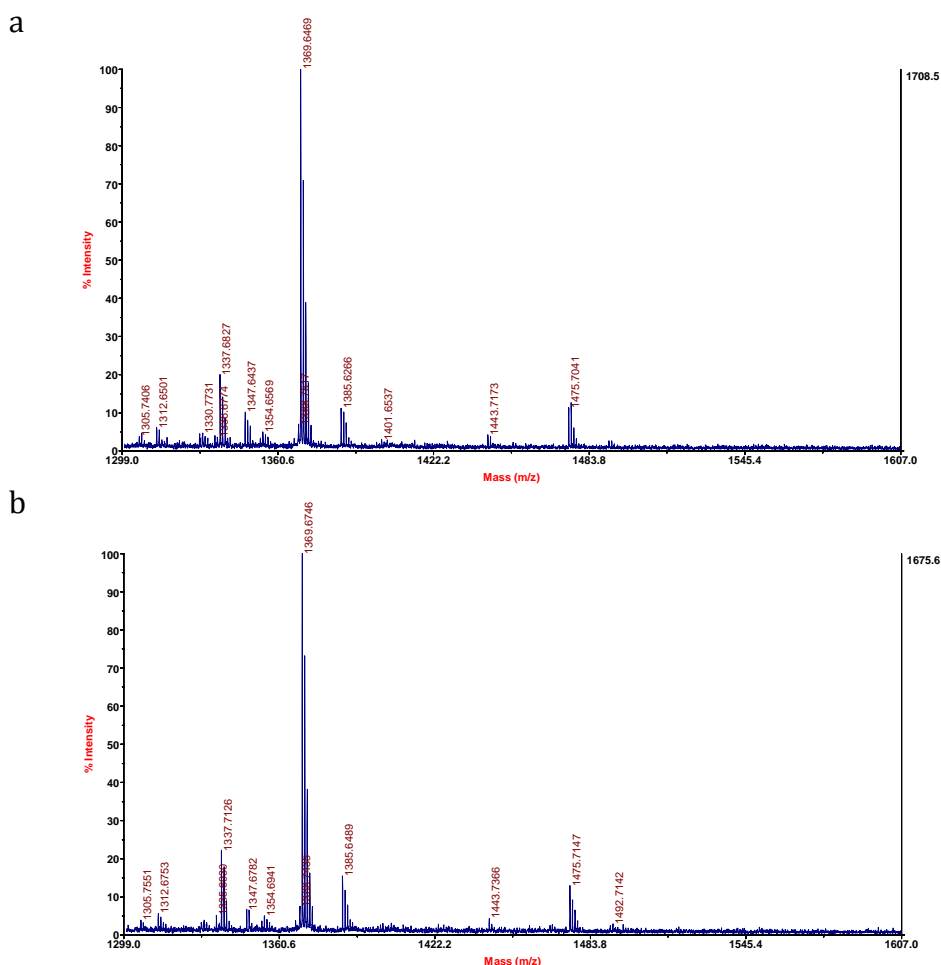


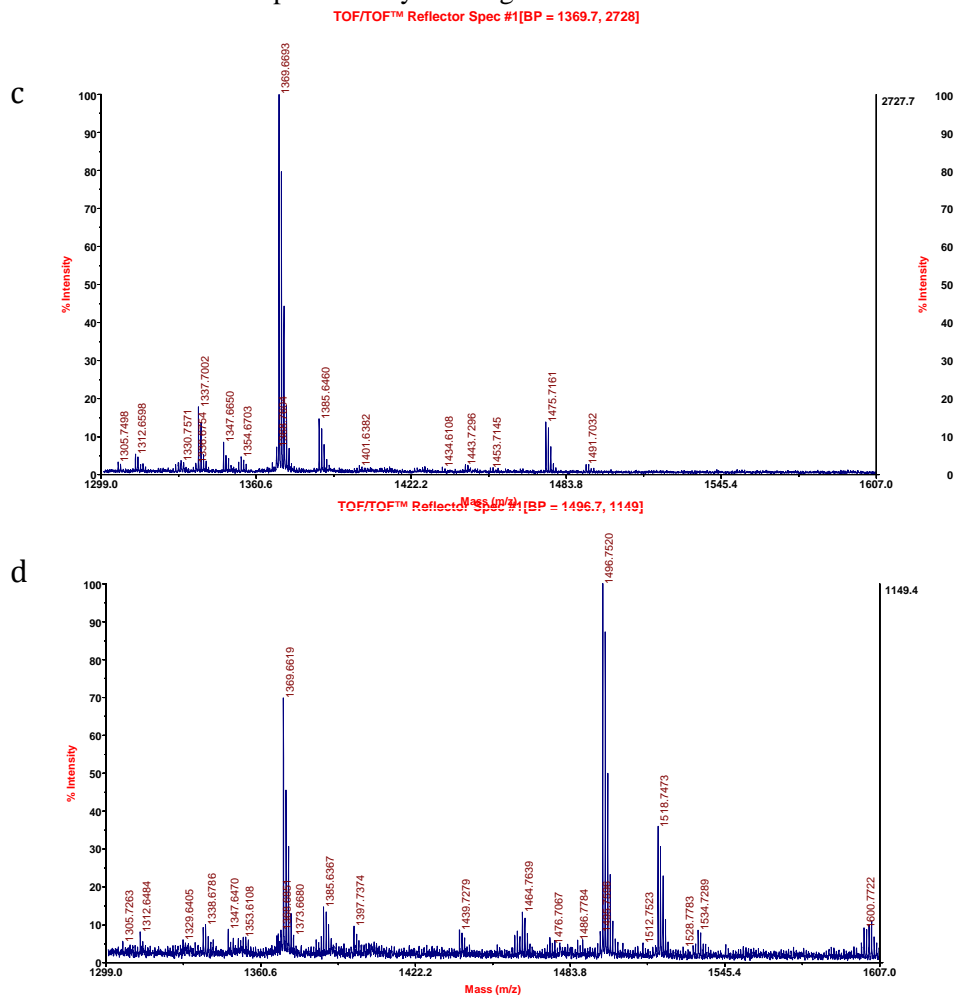
**Figure 3.9.** MALDI spectra of biotinylated Ac-GRKFGC (m/z 1233) (a), and Ac-GRK<sup>Ac</sup>FGC (m/z 1275) (b), after 1 h incubation with 0.5 mM anisidine and 1.5 mM CAN. No peaks were detected either at m/z 1382 or at m/z 1424, which indicates that no non-specific interaction between biotin and anisidine takes place.

### 3.3.3. Revised modification procedure for Ac-GYGGC and Ac-GGYGGC modification

Although the biotinylation step provides evidence of the specific modification of tyrosine residues, it did not lead to any improvement of the immobilization procedure. Therefore we revised the experimental set up of tyrosine modification so as to use the initial immobilization strategy without any biotin linker. As regards the oxidant reagent, the coupling partner ratio was adjusted to 10:1 (CAN: anisidine derivative); furthermore, the buffer was switched to 50 mM BIS-TRIS pH 6.2 and the incubation time was reduced to 10 min.

Both Ac-GYGGC (3.10a) and Ac-GGYGGC peptide (Figure 3.11a) were immobilized onto the monolayer through the sulfur-maleimide bond. A 0.1 mM peptide solution in BIS-TRIS buffer was deposited onto a maleimide functionalized monolayer for 20 min and then rinsed with ethanol, water and again ethanol. The new reaction mixture (0.1 mM anisidine, 1 mM CAN in BIS-TRIS buffer) was applied to the surface and allowed to sit for 10 min, after it was rinsed again with ethanol, water and ethanol, treated with THAP matrix and then analyzed by MALDI MS.

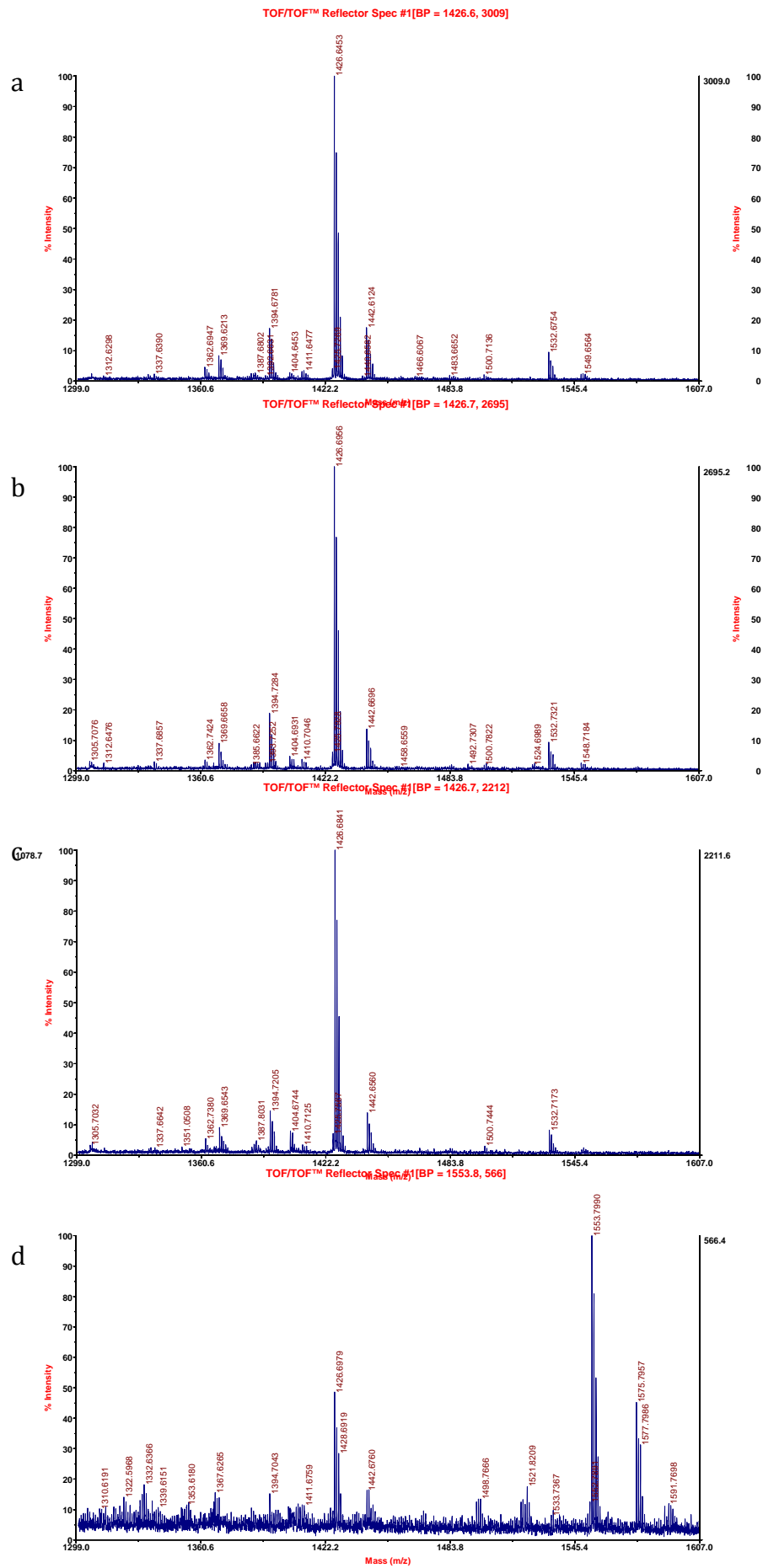




**Figure 3.10.** MALDI spectra of Ac-GYGGC ( $m/z$  1369, sodium adduct) immobilized onto the monolayer in the presence of 50 mM BIS-TRIS buffer (a), after 10 min incubation at room temperature with 1 mM CAN in BIS-TRIS (b), 0.1 mM anisidine in BIS-TRIS (c), and 0.1 mM anisidine and 1 mM CAN in BIS-TRIS (d). The product peak ( $m/z$  1496), given by anisidine-tyrosine coupling, was detected only after the treatment with the reaction mixture in its entirety.

These new oxidative conditions eventually resulted in the selective modification of the tyrosine residue of both Ac-GYGGC (3.10d) and Ac-GGYGGC peptides (Figure 3.11d). So we could eventually develop a procedure to efficiently couple the anisidine derivative to the electron-rich amino acid present in the peptide sequence, which disclosed the possibility to synthesize a large number of peptides containing the target amino acid with also different residues upstream and downstream of it.

As the first stage of this process we decided to synthesize two small peptide libraries in which either the upstream or the downstream amino acid varied while keeping unchanged the rest of the molecule. The two library sequences were respectively Ac-GGYXGC and Ac-GXYGGC, X being the position in which any amino acid is inserted, with the exception of cysteine. This was excluded as it is employed for stable and directional tethering of the molecule to the surface.



**Figure 3.11.** MALDI spectra of Ac-GGYGGC (m/z 1426, sodium adduct) anchored to the monolayer in the presence of 50 mM BIS-TRIS buffer (a), after 10 min incubation with 1 mM CAN in BIS-TRIS (b), 0.1 mM anisidine in BIS-TRIS (c), and 0.1 mM anisidine and 1 mM CAN in BIS-TRIS (d). Also in this case the product peak (m/z 1553) was only detected on the surface that had been treated with the complete reaction mixture.

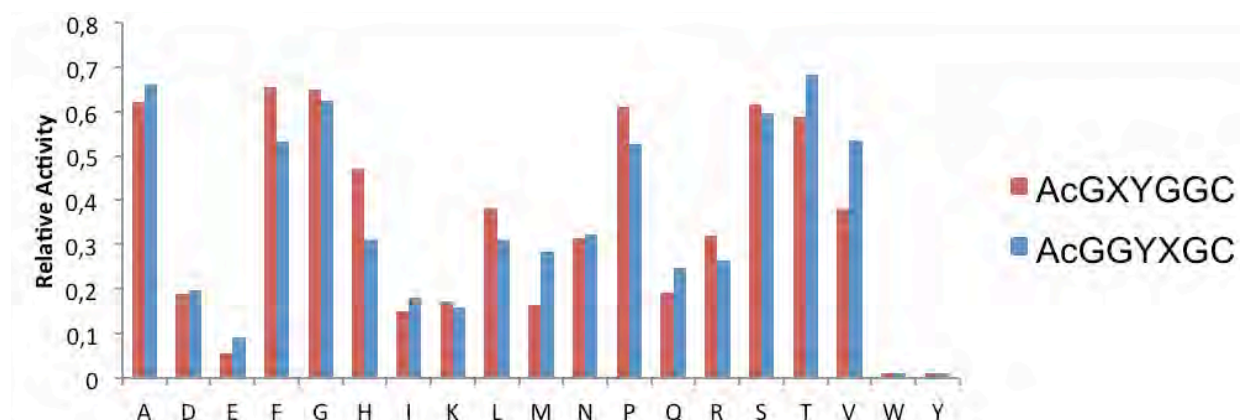
### 3.4. Ac-GGYXGC and Ac-GXYGGC peptide library synthesis and modification

For the synthesis of Ac-GGYXGC and Ac-GXYGGC libraries we adopted again the standard protocol for solid phase synthesis, using all the possible amino acid combinations either in the third or fifth position. Since the amino acid we inserted in the X position were 19, the two libraries counted a total of 38 different peptides.

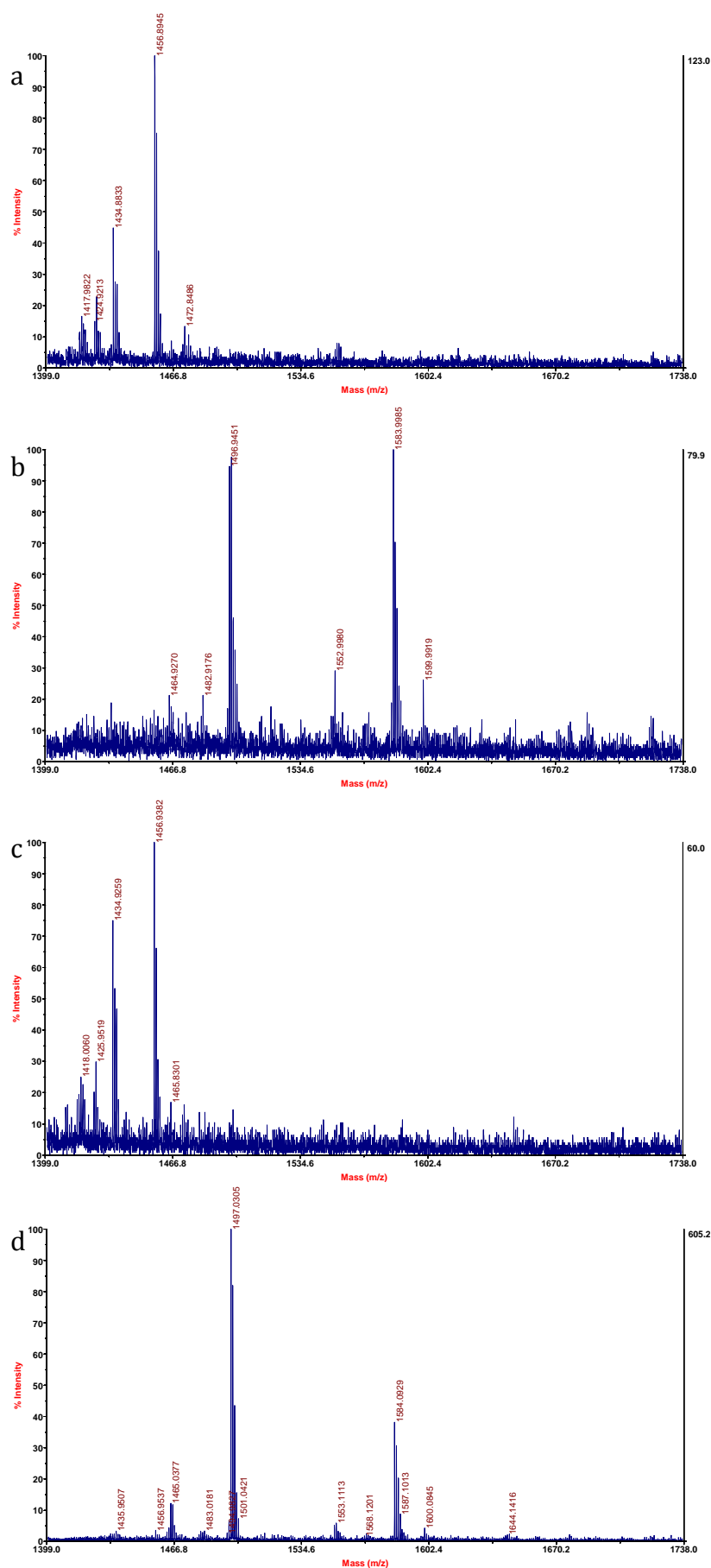
Each peptide was immobilized onto a monolayer through specific bond between the cysteine at the C-terminus and the maleimide functionalities on the gold surface. The whole libraries were then treated for 10 min with the readjusted reaction mixture (0.1 mM anisidine and 1 mM CAN in 50 mM BIS-TRIS buffer), rinsed, treated with THAP matrix, then analyzed with MALDI MS (Representative spectra of Ac-GGYSGC and Ac-GSYGGC peptides are shown in figure 3.13; see Figure S1-S18 in supplementary material for MALDI-TOF MS spectra of the two whole libraries).

The different peptides underwent oxidation to different extents, depending on the amino acid present upstream and downstream of the tyrosine target. In order to compare the particular behavior of each peptide, the reaction yield was calculated dividing the area under the curve (AUC) of the product peak by the sum of the areas of the product and the reactant peak:  $AUC_{\text{Product}} / (AUC_{\text{Product}} + AUC_{\text{Reactant}})$ . AUCs were calculated using Applied Biosystems Data Explorer software and Excel was used for data analysis. After plotting the relative activity of the different peptides, it became evident the dependence of the tyrosine modification on the amino acid present in the X position of the sequence (Figure 3.12).

Although a contaminant peak was detected in some peptide fractions ( $m/z$  1497), the results encouraged us to develop a whole 361-peptide library, in the expectation to achieve a more complete overview of the sequence dependence of the oxidative modification reaction under investigation.



**Figure 3.12.** Relative conversion activities of Ac-GGYXGC and Ac-GXYGGC library. The reaction yield of each peptide was calculated based on the formula:  $AUC_{\text{Product}} / (AUC_{\text{Product}} + AUC_{\text{Reactant}})$ .

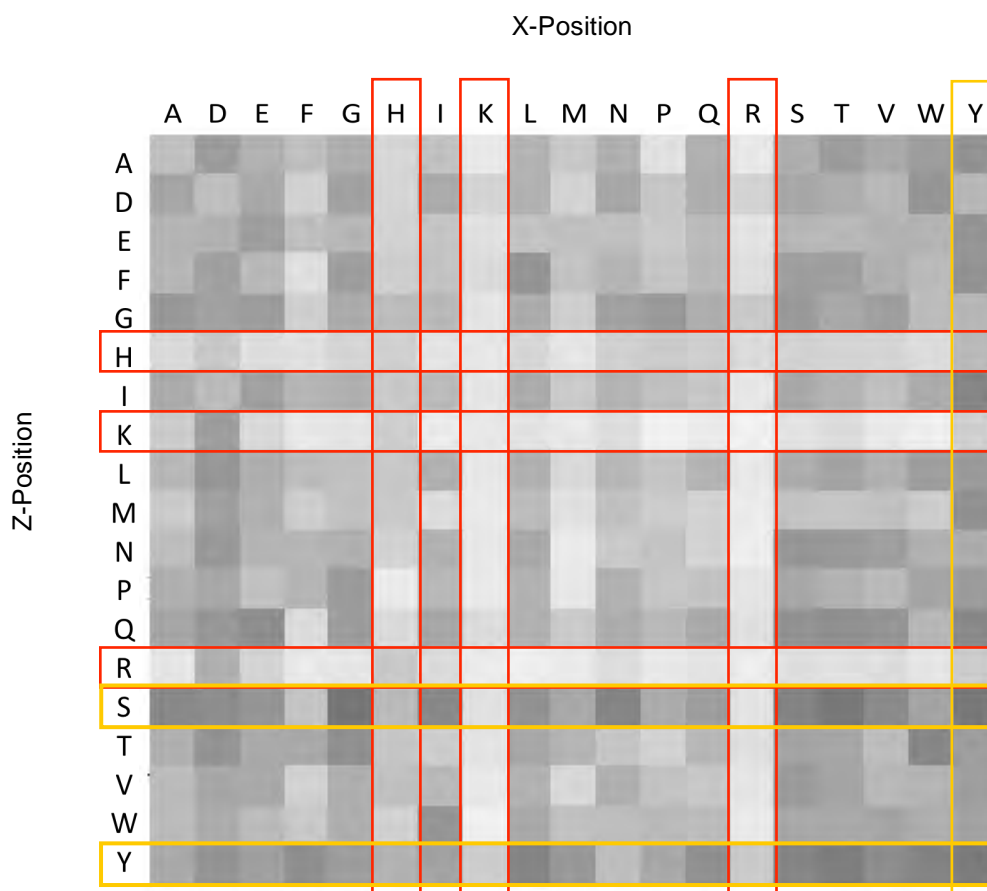


**Figure 3.13.** MALDI spectra of Ac-GGYSGC (a, b) and Ac-GSYGGC (c, d) anchored to the monolayer in the presence of 50 mM BIS-TRIS buffer (a, c) m/z 1456 sodium adduct, and after 10 min incubation with 0.1 mM anisidine and 1 mM CAN in BIS-TRIS (b, d), product m/z 1583 ( $H^+$ ).

### 3.5. Ac-GXYZGC library synthesis and screening

The library was synthesized following the solid phase peptide synthesis protocol but instead of the Rink-amide resin, we used Rink amide lanterns (Mimotopes) housed in 96-well filter plates. The library sequence is Ac-GXYZGC where X and Z are the two positions in which all the possible amino acid combinations were inserted. Also in this case, the cysteine residue was excluded for the aforementioned reasons. Thus, by combining the remaining 19 amino acids at the two positions, we produced a library consisting of 361 peptides.

The whole library was transferred to the gold array metal plates using a 96-well head liquid handler (Tecan EVO), immobilized for 20 min at room temperature, rinsed with ethanol, water and again ethanol, then dried. The peptide plates were then spotted with matrix (THAP), analyzed by SAMDI and the data acquired were then exported and analyzed using the Profiler software to quantify the reaction yield for each peptide by inputting the area under curves (AUCs) for the anisidine-coupled product peak and the peptide peak into the following formula:  $AUC_{Product}/(AUC_{Product} + AUC_{Peptide})$ . AUCs were calculated using Applied Biosystems Data Explorer software. We also used the Profiler output that converts data into a heat map representing the degree of conversion by a grey scale bar: white corresponds to no conversion; black to complete



**Figure 3.14.** Sequence dependence determined by SAMDI for CAN-mediated tyrosine modification. An array of peptides with the sequence Ac-GXYZGC-NH<sub>2</sub> was created, where the X and Z residues are shown on the horizontal and vertical axes, respectively. The degree of modification was determined by SAMDI and is represented by the grey scale bar with white for no modification activity and black the maximum relative activity.



tyrosine modification (Figure 3.14).

		X-Position																		
		A	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
Z-Position	A	25,0%	38,5%	31,2%	25,8%	32,3%	16,6%	23,7%	9,3%	29,7%	22,0%	31,4%	12,1%	31,2%	9,0%	31,3%	38,6%	33,3%	38,0%	42,5%
	D	35,4%	27,2%	34,3%	18,7%	36,6%	17,4%	33,3%	17,0%	30,2%	19,0%	34,5%	23,9%	32,9%	18,3%	35,1%	32,5%	28,3%	41,2%	27,8%
	E	30,0%	30,7%	38,1%	24,9%	29,7%	18,0%	23,4%	12,4%	24,3%	26,0%	25,8%	22,2%	27,2%	12,4%	29,7%	27,4%	30,6%	31,2%	41,9%
	F	29,5%	37,7%	28,3%	13,5%	38,1%	20,3%	23,8%	14,1%	41,5%	24,2%	30,1%	23,2%	27,8%	12,9%	38,2%	37,8%	31,0%	29,3%	42,5%
	G	40,3%	37,2%	39,1%	19,6%	33,2%	26,7%	27,1%	10,3%	32,1%	19,3%	36,3%	38,7%	31,5%	21,0%	39,3%	34,4%	37,1%	26,4%	29,8%
	H	15,2%	23,7%	13,7%	14,3%	16,5%	19,0%	12,4%	10,8%	16,7%	13,1%	20,0%	21,0%	19,6%	14,6%	18,7%	17,9%	16,8%	14,6%	27,5%
	I	32,7%	27,1%	36,4%	29,0%	30,9%	22,0%	28,2%	10,4%	32,9%	21,0%	29,6%	25,1%	22,6%	9,5%	33,9%	29,7%	25,6%	32,4%	47,3%
	K	21,2%	38,3%	15,6%	9,0%	10,7%	20,4%	8,1%	10,4%	11,2%	9,7%	14,7%	7,2%	8,8%	8,3%	11,4%	14,6%	9,3%	7,3%	18,6%
	L	28,3%	41,0%	32,3%	26,4%	24,8%	22,9%	30,9%	9,9%	30,5%	17,0%	29,0%	23,1%	29,9%	9,2%	32,0%	35,6%	30,5%	38,4%	39,5%
	M	20,9%	37,9%	32,7%	20,0%	25,0%	23,3%	14,4%	8,6%	26,5%	15,9%	25,8%	21,3%	18,7%	10,4%	24,1%	22,4%	24,9%	21,6%	43,2%
	N	25,0%	40,1%	29,7%	28,8%	31,1%	18,4%	30,3%	9,4%	33,5%	10,5%	22,7%	23,6%	17,6%	8,1%	40,7%	39,9%	37,1%	31,0%	32,4%
	P	31,1%	35,2%	25,4%	29,5%	39,2%	9,4%	27,5%	8,8%	30,4%	13,0%	30,2%	23,1%	24,1%	11,3%	34,5%	30,6%	27,4%	35,4%	38,8%
	Q	34,9%	40,4%	44,3%	14,4%	39,0%	14,4%	34,7%	13,9%	30,6%	21,2%	32,0%	27,5%	33,4%	13,8%	42,5%	43,4%	41,6%	30,5%	45,8%
	R	10,4%	31,1%	17,9%	9,6%	10,4%	21,5%	14,2%	9,4%	6,9%	10,3%	14,1%	9,2%	12,0%	8,1%	12,9%	11,8%	14,1%	11,4%	20,6%
	S	46,3%	44,8%	41,7%	23,4%	53,7%	27,3%	46,2%	12,1%	42,4%	34,4%	47,7%	32,9%	37,8%	11,2%	48,2%	53,0%	45,6%	36,0%	52,5%
	T	31,2%	41,9%	33,8%	33,9%	43,6%	23,0%	20,1%	11,0%	34,4%	30,3%	23,0%	18,0%	28,2%	13,0%	36,0%	34,8%	26,4%	46,6%	38,4%
	V	27,3%	34,6%	34,2%	19,5%	31,3%	25,1%	28,5%	9,3%	30,6%	16,8%	30,6%	25,1%	22,9%	11,9%	39,4%	35,8%	28,8%	28,2%	37,6%
	W	27,4%	33,3%	29,5%	24,3%	32,5%	19,3%	40,7%	6,3%	33,2%	27,1%	24,6%	28,0%	28,8%	9,5%	34,9%	34,7%	35,4%	34,8%	38,2%
	Y	34,4%	42,0%	37,4%	43,8%	39,0%	29,9%	35,8%	19,3%	48,0%	40,8%	27,7%	34,7%	43,5%	20,9%	47,8%	51,0%	45,1%	48,3%	51,9%

**Table 3.1.** Relative modification activities for Ac-GXYZGC library. The X and Z residues are shown on the horizontal and vertical axes, respectively.

As previously reported, SAMDI is a useful tool for following the oxidative modification of tyrosine-containing peptides immobilized to maleimide-terminated self-assembled monolayers. residues to the tyrosine were varied, it enabled us to identify sequences displaying high reactivity using the examined modification strategy. Indeed, we could observe that the most reactive peptides had either serine N-terminal to the tyrosine or another tyrosine in proximity. On the other hand, the residues displaying low conversion levels contained a positive charged residue, i.e., histidine, lysine or arginine. In particular, the lowest yield was observed with arginine and leucine as Z and X residues, respectively (Table 3.1).

We have clearly demonstrated that this screening strategy can be effectively accomplished by SAMDI MS thanks to several advantages it offers, among which automation, straightforwardness in acquiring, interpreting and analyzing data and, primarily, its high-throughput potential. Indeed, we could easily screen 361 peptides simultaneously, and at the same time identify the specific residues that contribute to increased (or decreased) reactivity. Thus, this new approach provides a new tool for rapidly selecting the best target sequence to be inserted into a protein in order to achieve a specific and efficient bioconjugation.

## 4. Discussion

Currently, protein bioconjugates are increasingly being used as functional and structural components of new drugs and materials. Therefore, there is a strong need for chemical reactions that facilitate the efficient attachment of a synthetic molecule to specific desired positions on protein surfaces. In a typical approach, a synthetic functional group of interest is attached to a specific amino acid side chain using one out of a relatively limited set of chemical reactions that can modify unprotected biomolecules in aqueous solution. Most bioconjugation reactions, however, do not reach full conversion and an excess reagent can cause both overmodification and erosion of chemospecificity (Stephanopoulos and Francis 2011). In this perspective, we wanted to develop a generalizable screening platform to evaluate the best target sequence for a desired protein modification reaction, in order to optimize the bioconjugation strategy.

As a bioconjugation model directed towards natural amino acids, among all protein modification strategies available we chose the selective modification of tyrosine residues using Cerium(IV) ammonium nitrate (CAN) as one-electron oxidant. This reaction has been proven to be selective for tyrosine residues and since they occur with intermediate to low frequency on native protein surfaces and are often partially or completely buried due to the amphiphilic nature of the phenolic group, this bioconjugation strategy is considered particularly useful to also control the number and location of the modifications. Previous work by Francis and colleagues exploited this mechanism to efficiently modify both native and artificially introduced residues on proteins with polyethylene glycol (PEG) and small peptides. As a native, solvent-accessible tyrosine-containing protein, chymotrypsinogen was modified. On the other hand, to evaluate the reaction applicability to a different protein substrate, new tyrosine residues were introduced onto the surface of genome-free MS2 capsids. In both cases, the presence of CAN led to a single modification of tyrosine with high yields and short reaction times. Moreover, the same strategy was used in conjunction with cysteine alkylation to doubly modify viral capsid proteins, aiming at developing both targeting and imaging functionalities. To label proteins with fluorescent dyes, cysteine residues were modified using Oregon Green 488 maleimide, while as potential cancer targeting groups, small cell-targeting peptides (cyclic RGD) were coupled to the tyrosine using the oxidative coupling strategy. The synthesis of useful bioconjugates through this site-selective methodology attracted our interest and made us quite confident in exploring a way to further improve the conversion level of this new reaction.

After choosing a modification strategy to be optimized, we sought a model peptide to set up the experimental conditions. For this purpose, we synthesized a very simple tyrosine-containing tetramer with the sequence Ac-GYGC, where Ac-G is a *N*-Acetylglycine. The cysteine residue at the C-terminus was required for the efficient and directional immobilization of the peptide onto a

maleimide functionalized self-assembled monolayer. While the peptide was efficiently linked in the presence of 10 mM TRIS buffer, pH 7.01, a clear signal loss was noticed after incubation with the oxidative reaction mixture composed of 500  $\mu$ M 4-methoxy-*N,N*-dimethylaniline (anisidine derivative), 1.5 mM cerium(IV) ammonium nitrate (CAN) in TRIS buffer, so that no product peak was eventually detected (no *m/z* shift of 149). In order to check if this effect was non-specific or due to a peculiar behavior of the Ac-GYGC peptide, we then synthesized two different peptides carrying one or two more glycine residues in their sequence: Ac-GYGGC and Ac-GGYGGC. Also in this case, after 30 min incubation at room temperature both peptides were immobilized onto the surface but as in the case of Ac-GYGC after the treatment with the reaction mixture, we did not detect any more either reaction product or the starting signal. Moreover, when observing the SAMDI spectra obtained by analyzing the monolayer immediately after the incubation with the oxidative mixture, we noticed an increased background noise probably due to damage of the monolayer caused by the extreme acidic conditions. Indeed, even though the reaction was performed in TRIS buffer pH 7.01, the simple addition of CAN, leading to significant solution acidification, may have caused monolayer fragmentation and consequently sample desorption. Since the oxidative condition required could not guarantee the integrity of the monolayer, this result led us to judge as inadequate the modification reaction performed directly on the monolayer. Therefore we designed a new experimental procedure so as to run the reaction in solution and immobilize the reacted peptide after the reaction was quenched. For this purpose, we biotinylated the C-termini of our peptides, reacted them in solution, stopped the reaction after 1 h and finally deposited them onto the surface (previously biotinylated and streptavidinated). However, even under these conditions the reaction product could not be anchored to the monolayer, as shown by MALDI MS analyses that revealed product formation but no immobilization. Even though the reasons of this failure are unclear, it is quite likely that the extreme oxidative conditions used, played a key role in the outcome. So we managed to readjust some critical condition, such as incubation time (10 min), buffer (50 mM BIS TRIS, pH 6.2) and the oxidant/coupling partner ratio (10:1). Under these revised conditions we could finally observe the product peak appearance. 10 min of incubation were therefore sufficient for the reaction to take place and at the same time the monolayer was not significantly damaged.

To assess the reaction rates relevant to each substrate, we selected conditions far from reaction completion, so as to compare the different conversion efficiencies and eventually identify those short sequences capable of leading to optimal conversion.

We then synthesized two small libraries in the sequences AcGGYXGC and AcGXYGGC, respectively, by inserting into the X position any amino acid, except for cysteine. Then, we

performed the screening according to the reaction conditions previously set up. At this stage, we also intended to develop a quantitative method to compare the conversion levels among the peptides of the two libraries. For this purpose, we calculated the relative intensity of each peptide. Actually, for some peptides such as those containing histidine, arginine, lysine we obtained very clear spectra with low signal to noise ratio and easy-to-read peaks, but in most cases an unexpected peptide was detected that could be modified and coupled with the anisidine derivative as well. This contaminant was quite likely a truncated version of the peptide. Despite the presence of this undesired peak, as the values obtained are relative intensities, we could compare them with each other, thus obtaining a general overview of the behavior of the two libraries. By plotting the percentage of conversion, we could assign to every peptide its own intensity. Thus, we set up to screen an entire 361-peptide library to exhaustively characterize the reaction pattern and develop a screening platform. For this purpose, the above-mentioned library of 361 hexapeptides, having variable sequence at the N- and C-terminal residues to the target tyrosine (Ac-GXYZGC-NH<sub>2</sub> where X, Z = all amino acids except cysteine) was synthesized and anchored in the format of an array onto a maleimide-terminated monolayer, which was then treated with the oxidative coupling mixture and analyzed by SAMDI-MS to determine the degree of conversion for each peptide. In other similar studies, in which a combinatorial peptide library had been generated on resin beads (Witus et al. 2010), the examination was made under a microscope, the most reactive beads were colorimetrically identified, then manually removed for sequencing. In the present work, the use of metal plates carrying an array of 384 gold-coated islands in the standard microtiter plate geometry, allowed us to handle, react and analyze the whole library at the same time. Even more important, we could immediately establish how each residue contributes to reaction performance, since the sequence of the peptide immobilized on each spot in the array was known in advance. Thus, we found that the amino acid identity at C- and N-terminal position to the target tyrosine had a significant impact on the extent of conversion. The array revealed that it was more strongly affected by the nature of the amino acid rather than by its position. Indeed His, Lys, and Arg-containing peptides displayed the lowest activities, when they were both in the X or Z position. On the other hand, the peptides leading to the best performance had Ser and Tyr in their sequence, in particular when Ser occupied the Z position, whereas in the X position it had little influence on the conversion. In conclusion, the specificity profiles provided a clear view of the sequence characteristics of peptides that are optimally active for the bioconjugation reaction under investigation: the most active substrates presented a C-terminal Ser relative to the target tyrosine, while basic residues at both the N- and C-terminal position drastically reduce the conversion level.

In the high-throughput screening assay here presented we took advantage of many of the SAMDI MS unique properties, mainly the availability of hundreds of functionalized spots on a single array (384 gold-coated islands on each plate). On each spot we immobilized a specific peptide with a uniform density, so we also could quantify the activity of the monolayer moieties. By calculating the relative activity of distinct peptides across the entire array, we could compare their conversion levels and generate a “specificity profile” for the examined reaction.

In detail, by the use of SAMDI we optimized different features of the procedure. In particular, the amount of reagent was much smaller than in previous approaches, as the peptide starting solution was 0.1 mM and the final concentrations of oxidant reagent and coupling partner 1 mM and 0.1 mM, respectively. Moreover, the sample peptides did not require previous purification, since the immobilization strategy involves exclusively the moieties with which the monolayer has been functionalized for the specific interaction of the library peptides. In our case, the C-terminal cysteine residue bound specifically to the maleimide-presenting alkyldisulfide of the monolayer and nonspecific absorption of other reaction components was prevented by the use of polyethylene glycol molecules. In addition, the reaction timing was also drastically reduced: 10 min were enough to detect the product, although the reaction at this stage was far from completion. Furthermore, for the reaction to be quenched, no reagent addition was required (such as TCEP or phosphate buffer necessary to deactivate Cerium(IV) ammonium nitrate when the reaction was run in solution), but this was achieved by simply rinsing off the plate.

Therefore, in the present study we have developed an efficient and generalizable method for the evaluation of short sequences displaying a highly-reactive motif toward the bioconjugation reaction under consideration. In our future perspectives we will identify sequences reactive also in the context of a full-length protein, and to do so we will choose three different peptide sequences that show high, low and average coupling efficiency with the anisidine derivative. Using these three motifs, we will modify a model protein by recombinant DNA technology, by adding the selected hexapeptides to either its N-terminus or in the middle of the protein sequence by site-specific mutagenesis. Then, we will monitor tyrosine conversion levels and will compare them with the results obtained using the peptides in isolation. As an additional proof of the different anisidine-coupling yields, we will subsequently exploit the oxidative reaction as immobilization strategy for nanoparticles (NPs) functionalization. Using proteins modified so as to display highly, on average or poorly-reactive peptides and coupling them with anisidine-presenting NPs, we will measure the specific protein loading onto the NPs. In this way, we will be able to verify if the conversion levels reflect those obtained by reacting anisidine derivatives with the peptides in isolation or the mutant-proteins.

Even though we have tested as a reaction model the oxidative modification of tyrosine residues, our goal was to create a generalizable method that might be used to characterize and identify sequences that allow optimal conversion levels in a variety of bioconjugation reactions. Therefore, a future perspective will be the use of the peptide array technology in the best-target selection of other bioconjugation reactions commonly used in protein selective modification. Actually, this technology offers a universal preliminary analysis tool that can be exploited to apply a site-specific modification reaction to protein substrates of interest for researchers in different areas of investigation. They could identify the best target motif to be inserted in their protein sequence or even predict the conversion efficiency to be expected for the modification. Thus, our results provide an important example of how synthetic peptide libraries combined with SAMDY MS technique generate an efficient and straightforward combinatorial library screening platform that can accelerate the discovery and optimization of new protein bioconjugation reactions.

# References



Axup, J.Y., Bajjuri, K.M., Ritland, M., Hutchins, B.M., Kim, C.H., Kazane, S.A., Halder, R., Forsyth, J.S., Santidrian, A.F., Stafin, K., et al. (2012). Synthesis of site-specific antibody-drug conjugates using unnatural amino acids. *Proc. Natl. Acad. Sci.* *109*, 16101–16106.

Baeuerle, P.A., and Reinhardt, C. (2009). Bispecific T-Cell Engaging Antibodies for Cancer Therapy. *Cancer Res.* *69*, 4941–4944.

Bain, J.D., Diala, E.S., Glabe, C.G., Dix, T.A., and Chamberlin, A.R. (1989). Biosynthetic site-specific incorporation of a non-natural amino acid into a polypeptide. *J. Am. Chem. Soc.* *111*, 8013–8014.

Ban, L., and Mrksich, M. (2008). On-chip synthesis and label-free assays of oligosaccharide arrays. *Angew. Chem. Int. Ed Engl.* *47*, 3396–3399.

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.

Beatty, K.E., and Tirrell, D.A. (2008). Two-color labeling of temporally defined protein populations in mammalian cells. *Bioorg. Med. Chem. Lett.* *18*, 5995–5999.

Beatty, K.E., Xie, F., Wang, Q., and Tirrell, D.A. (2005). Selective Dye-Labeling of Newly Synthesized Proteins in Bacterial Cells. *J. Am. Chem. Soc.* *127*, 14150–14151.

Beatty, K.E., Liu, J.C., Xie, F., Dieterich, D.C., Schuman, E.M., Wang, Q., and Tirrell, D.A. (2006). Fluorescence Visualization of Newly Synthesized Proteins in Mammalian Cells. *Angew. Chem.* *118*, 7524–7527.

Bernardes, G.J.L., Chalker, J.M., Errey, J.C., and Davis, B.G. (2008). Facile conversion of cysteine and alkyl cysteines to dehydroalanine on protein surfaces: versatile and switchable access to functionalized proteins. *J. Am. Chem. Soc.* *130*, 5052–5053.

Bose, M., Groff, D., Xie, J., Brustad, E., and Schultz, P.G. (2006). The Incorporation of a Photoisomerizable Amino Acid into Proteins in *E. coli*. *J. Am. Chem. Soc.* *128*, 388–389.

Brustad, E., Bushey, M.L., Lee, J.W., Groff, D., Liu, W., and Schultz, P.G. (2008a). A Genetically Encoded Boronate-Containing Amino Acid. *Angew. Chem. Int. Ed.* *47*, 8220–8223.

- Brustad, E.M., Lemke, E.A., Schultz, P.G., and Deniz, A.A. (2008b). A General and Efficient Method for the Site-Specific Dual-Labeling of Proteins for Single Molecule Fluorescence Resonance Energy Transfer. *J. Am. Chem. Soc.* *130*, 17664–17665.
- Campos, L.M., Killups, K.L., Sakai, R., Paulusse, J.M.J., Damiron, D., Drockenmuller, E., Messmore, B.W., and Hawker, C.J. (2008). Development of Thermal and Photochemical Strategies for Thiol–Ene Click Polymer Functionalization. *Macromolecules* *41*, 7063–7070.
- Carrico, Z.M., Romanini, D.W., Mehl, R.A., and Francis, M.B. (2008). Oxidative coupling of peptides to a virus capsid containing unnatural amino acids. *Chem. Commun.* 1205–1207.
- Chames, P., and Baty, D. (2009) Bispecific antibodies for cancer therapy: The light at the end of the tunnel? *mAbs* *1*, 539–547.
- Chapman, E., Thorson, J.S., and Schultz, P.G. (1997). Mutational Analysis of Backbone Hydrogen Bonds in Staphylococcal Nuclease. *J. Am. Chem. Soc.* *119*, 7151–7152.
- Chin, J.W., Santoro, S.W., Martin, A.B., King, D.S., Wang, L., and Schultz, P.G. (2002a). Addition of p-Azido-l-phenylalanine to the Genetic Code of *Escherichia coli*. *J. Am. Chem. Soc.* *124*, 9026–9027.
- Chin, J.W., Martin, A.B., King, D.S., Wang, L., and Schultz, P.G. (2002b). Addition of a photocrosslinking amino acid to the genetic code of *Escherichia coli*. *Proc. Natl. Acad. Sci.* *99*, 11020–11024.
- Chin, J.W., Cropp, T.A., Anderson, J.C., Mukherji, M., Zhang, Z., and Schultz, P.G. (2003). An Expanded Eukaryotic Genetic Code. *Science* *301*, 964–967.
- Cho, H., Daniel, T., Buechler, Y.J., Litzinger, D.C., Maio, Z., Putnam, A.-M.H., Kraynov, V.S., Sim, B.-C., Bussell, S., Javahishvili, T., et al. (2011). Optimized clinical performance of growth hormone with an expanded genetic code. *Proc. Natl. Acad. Sci.* *108*, 9060–9065.
- Cowie, D.B., and Cohen, G.N. (1957). Biosynthesis by *Escherichia coli* of active altered proteins containing selenium instead of sulfur. *Biochim. Biophys. Acta* *26*, 252–261.
- Datta, D., Wang, P., Carrico, I.S., Mayo, S.L., and Tirrell, D.A. (2002). A Designed Phenylalanyl-tRNA Synthetase Variant Allows Efficient in Vivo Incorporation of Aryl Ketone Functionality into Proteins. *J. Am. Chem. Soc.* *124*, 5652–5653.

- Deiters, A., and Schultz, P.G. (2005). In vivo incorporation of an alkyne into proteins in *Escherichia coli*. *Bioorg. Med. Chem. Lett.* *15*, 1521–1524.
- Deiters, A., Groff, D., Ryu, Y., Xie, J., and Schultz, P.G. (2006). A Genetically Encoded Photocaged Tyrosine. *Angew. Chem.* *118*, 2794–2797.
- Devaraj, N.K., Miller, G.P., Ebina, W., Kakaradov, B., Collman, J.P., Kool, E.T., and Chidsey, C.E.D. (2005). Chemoselective Covalent Coupling of Oligonucleotide Probes to Self-Assembled Monolayers. *J. Am. Chem. Soc.* *127*, 8600–8601.
- Dieterich, D.C., Link, A.J., Graumann, J., Tirrell, D.A., and Schuman, E.M. (2006). Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (BONCAT). *Proc. Natl. Acad. Sci.* *103*, 9482–9487.
- Di Valentin, C., Scagnelli, A., and Pacchioni, G. (2005). Theory of nanoscale atomic lithography. An ab initio study of the interaction of “cold” Cs atoms with organothiols self-assembled monolayers on Au(111). *J. Phys. Chem. B* *109*, 1815–1821.
- Dondoni, A. (2008). The emergence of thiol-ene coupling as a click process for materials and bioorganic chemistry. *Angew. Chem. Int. Ed Engl.* *47*, 8995–8997.
- Doronina, S.O., Mendelsohn, B.A., Bovee, T.D., Cervený, C.G., Alley, S.C., Meyer, D.L., Oflazoglu, E., Toki, B.E., Sanderson, R.J., Zabinski, R.F., et al. (2006). Enhanced Activity of Monomethylauristatin F through Monoclonal Antibody Delivery: Effects of Linker Technology on Efficacy and Toxicity. *Bioconjug. Chem.* *17*, 114–124.
- Dougherty, D.A. (2000). Unnatural amino acids as probes of protein structure and function. *Curr. Opin. Chem. Biol.* *4*, 645–652.
- Dougherty, D.A. (2008). Cys-Loop Neuroreceptors: Structure to the Rescue? *Chem. Rev.* *108*, 1642–1653.
- Ellman, J., Mendel, D., Anthony-Cahill, S., Noren, C.J., and Schultz, P.G. (1991). [15] Biosynthetic method for introducing unnatural amino acids site-specifically into proteins. In *Methods in Enzymology*, John J. Langone, ed. (Academic Press), pp. 301–336.
- Ellman, J.A., Mendel, D., and Schultz, P.G. (1992). Site-specific incorporation of novel backbone structures into proteins. *Science* *255*, 197–200.

- Evans, C.A., and Miller, S.J. (2002). Proton-activated fluorescence as a tool for simultaneous screening of combinatorial chemical reactions. *Curr. Opin. Chem. Biol.* *6*, 333–338.
- Finklea, H.O., Avery, S., Lynch, M., and Furtch, T. (1987). Blocking oriented monolayers of alkyl mercaptans on gold electrodes. *Langmuir* *3*, 409–413.
- Gong, W., Elitzin, V.I., Janardhanam, S., Wilkins, C.L., and Fritsch, I. (2001). Effect of laser fluence on laser desorption mass spectra of organothiols self-assembled monolayers on gold. *J. Am. Chem. Soc.* *123*, 769–770.
- Graziano, R.F., and Guptill, P. (2004). Chemical production of bispecific antibodies. *Methods Mol. Biol.* Clifton NJ *283*, 71–85.
- Guan, W., Ban, L., Cai, L., Li, L., Chen, W., Liu, X., Mrksich, M., and Wang, P.G. (2011). Combining carbochips and mass spectrometry to study the donor specificity for the *Neisseria meningitidis*  $\beta$ 1,3-N-acetylglucosaminyltransferase LgtA. *Bioorg. Med. Chem. Lett.* *21*, 5025–5028.
- Gurard-Levin, Z.A., and Mrksich, M. (2008a). Combining self-assembled monolayers and mass spectrometry for applications in biochips. *Annu. Rev. Anal. Chem. Palo Alto Calif* *1*, 767–800.
- Gurard-Levin, Z.A., and Mrksich, M. (2008b). The activity of HDAC8 depends on local and distal sequences of its peptide substrates. *Biochemistry (Mosc.)* *47*, 6242–6250.
- 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.
- 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.
- Hamann, P.R., Hinman, L.M., Hollander, I., Beyer, C.F., Lindh, D., Holcomb, R., Hallett, W., Tsou, H.R., Upešlaciš, J., Shochat, D., et al. (2002). Gemtuzumab Ozogamicin, A Potent and Selective Anti-CD33 Antibody–Calicheamicin Conjugate for Treatment of Acute Myeloid Leukemia. *Bioconjug. Chem.* *13*, 47–58.
- Van Hest, J.C.M., Kiick, K.L., and Tirrell, D.A. (2000). Efficient Incorporation of Unsaturated Methionine Analogues into Proteins in Vivo. *J. Am. Chem. Soc.* *122*, 1282–1288.

Hillenkamp, F., and Peter-Katalinic, J. (2013). *MALDI MS: A Practical Guide to Instrumentation, Methods and Applications* (John Wiley & Sons).

Houseman, B.T., and Mrksich, M. (1998). Efficient Solid-Phase Synthesis of Peptide-Substituted Alkanethiols for the Preparation of Substrates That Support the Adhesion of Cells. *J. Org. Chem.* *63*, 7552–7555.

Houseman, B.T., and Mrksich, M. (2001). The microenvironment of immobilized Arg-Gly-Asp peptides is an important determinant of cell adhesion. *Biomaterials* *22*, 943–955.

Houseman, B.T., and Mrksich, M. (2002). Carbohydrate arrays for the evaluation of protein binding and enzymatic modification. *Chem. Biol.* *9*, 443–454.

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.

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.

Ibba, M., Kast, P., and Hennecke, H. (1994). Substrate Specificity Is Determined by Amino Acid Binding Pocket Size in *Escherichia coli* Phenylalanyl-tRNA Synthetase. *Biochemistry (Mosc.)* *33*, 7107–7112.

Jäger, M., Schoberth, A., Ruf, P., Hess, J., and Lindhofer, H. (2009). The Trifunctional Antibody Ertumaxomab Destroys Tumor Cells That Express Low Levels of Human Epidermal Growth Factor Receptor 2. *Cancer Res.* *69*, 4270–4276.

Jonkheijm, P., Weinrich, D., Köhn, M., Engelkamp, H., Christianen, P.C.M., Kuhlmann, J., Maan, J.C., Nüsse, D., Schroeder, H., Wacker, R., et al. (2008). Photochemical surface patterning by the thiol-ene reaction. *Angew. Chem. Int. Ed Engl.* *47*, 4421–4424.

Junutula, J.R., Raab, H., Clark, S., Bhakta, S., Leipold, D.D., Weir, S., Chen, Y., Simpson, M., Tsai, S.P., Dennis, M.S., et al. (2008). Site-specific conjugation of a cytotoxic drug to an antibody improves the therapeutic index. *Nat. Biotechnol.* *26*, 925–932.

Karas, M., and Hillenkamp, F. (1988). Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal. Chem.* *60*, 2299–2301.

- Kast, P., and Hennecke, H. (1991). Amino acid substrate specificity of Escherichia coli phenylalanyl-tRNA synthetase altered by distinct mutations. *J. Mol. Biol.* 222, 99–124.
- Kazane, S.A., Sok, D., Cho, E.H., Uson, M.L., Kuhn, P., Schultz, P.G., and Smider, V.V. (2012). Site-specific DNA-antibody conjugates for specific and sensitive immuno-PCR. *Proc. Natl. Acad. Sci.* 109, 3731–3736.
- Kazane, S.A., Axup, J.Y., Kim, C.H., Ciobanu, M., Wold, E.D., Barluenga, S., Hutchins, B.A., Schultz, P.G., Winssinger, N., and Smider, V.V. (2013). Self-Assembled Antibody Multimers through Peptide Nucleic Acid Conjugation. *J. Am. Chem. Soc.* 135, 340–346.
- Kiick, K.L., Saxon, E., Tirrell, D.A., and Bertozzi, C.R. (2002). Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation. *Proc. Natl. Acad. Sci.* 99, 19–24.
- Kim, C.H., Axup, J.Y., Dubrovskaya, A., Kazane, S.A., Hutchins, B.A., Wold, E.D., Smider, V.V., and Schultz, P.G. (2012). Synthesis of bispecific antibodies using genetically encoded unnatural amino acids. *J. Am. Chem. Soc.* 134, 9918–9921.
- Kirshenbaum, K., Carrico, I.S., and Tirrell, D.A. (2002). Biosynthesis of Proteins Incorporating a Versatile Set of Phenylalanine Analogues. *ChemBioChem* 3, 235–237.
- Kodama, K., Fukuzawa, S., Nakayama, H., Kigawa, T., Sakamoto, K., Yabuki, T., Matsuda, N., Shirouzu, M., Takio, K., Tachibana, K., et al. (2006). Regioselective Carbon–Carbon Bond Formation in Proteins with Palladium Catalysis; New Protein Chemistry by Organometallic Chemistry. *ChemBioChem* 7, 134–139.
- Kodama, K., Fukuzawa, S., Nakayama, H., Sakamoto, K., Kigawa, T., Yabuki, T., Matsuda, N., Shirouzu, M., Takio, K., Yokoyama, S., et al. (2007). Site-Specific Functionalization of Proteins by Organopalladium Reactions. *ChemBioChem* 8, 232–238.
- Koh, J.T., Cornish, V.W., and Schultz, P.G. (1997). An Experimental Approach to Evaluating the Role of Backbone Interactions in Proteins Using Unnatural Amino Acid Mutagenesis†. *Biochemistry (Mosc.)* 36, 11314–11322.
- Köhler, C., Xie, L., Kellerer, S., Varshney, U., and RajBhandary, U.L. (2001). Import of amber and ochre suppressor tRNAs into mammalian cells: A general approach to site-specific insertion of amino acid analogues into proteins. *Proc. Natl. Acad. Sci.* 98, 14310–14315.

- Kothakota, S., Mason, T.L., Tirrell, D.A., and Fournier, M.J. (1995). Biosynthesis of a Periodic Protein Containing 3-Thienylalanine: A Step Toward Genetically Engineered Conducting Polymers. *J. Am. Chem. Soc.* *117*, 536–537.
- Kwon, Y., Han, Z., Karatan, E., Mrksich, M., and Kay, B.K. (2004). Antibody arrays prepared by cutinase-mediated immobilization on self-assembled monolayers. *Anal. Chem.* *76*, 5713–5720.
- Lang, K., Davis, L., Wallace, S., Mahesh, M., Cox, D.J., Blackman, M.L., Fox, J.M., and Chin, J.W. (2012). Genetic Encoding of Bicyclononynes and trans-Cyclooctenes for Site-Specific Protein Labeling in Vitro and in Live Mammalian Cells via Rapid Fluorogenic Diels–Alder Reactions. *J. Am. Chem. Soc.* *134*, 10317–10320.
- Lemke, E.A., Summerer, D., Geierstanger, B.H., Brittain, S.M., and Schultz, P.G. (2007). Control of protein phosphorylation with a genetically encoded photocaged amino acid. *Nat. Chem. Biol.* *3*, 769–772.
- Lewis, J.K., Wei, J., and Siuzdak, G. (2000). Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry in Peptide and Protein Analysis. In *Encyclopedia of Analytical Chemistry*, (John Wiley & Sons, Ltd),.
- Li, J., Thiara, P.S., and Mrksich, M. (2007). Rapid evaluation and screening of interfacial reactions on self-assembled monolayers. *Langmuir ACS J. Surf. Colloids* *23*, 11826–11835.
- Link, A.J., and Tirrell, D.A. (2003). Cell Surface Labeling of *Escherichia coli* via Copper(I)-Catalyzed [3+2] Cycloaddition. *J. Am. Chem. Soc.* *125*, 11164–11165.
- Link, A.J., Mock, M.L., and Tirrell, D.A. (2003). Non-canonical amino acids in protein engineering. *Curr. Opin. Biotechnol.* *14*, 603–609.
- Link, A.J., Vink, M.K.S., and Tirrell, D.A. (2004). Presentation and detection of azide functionality in bacterial cell surface proteins. *J. Am. Chem. Soc.* *126*, 10598–10602.
- Link, A.J., Vink, M.K.S., Agard, N.J., Prescher, J.A., Bertozzi, C.R., and Tirrell, D.A. (2006). Discovery of aminoacyl-tRNA synthetase activity through cell-surface display of noncanonical amino acids. *Proc. Natl. Acad. Sci.* *103*, 10180–10185.
- Liu, C.C., and Schultz, P.G. (2006). Recombinant expression of selectively sulfated proteins in *Escherichia coli*. *Nat. Biotechnol.* *24*, 1436–1440.

- Liu, W., Brock, A., Chen, S., Chen, S., and Schultz, P.G. (2007). Genetic incorporation of unnatural amino acids into proteins in mammalian cells. *Nat. Methods* 4, 239–244.
- Luk, Y.Y., Tingey, M.L., Dickson, K.A., Raines, R.T., and Abbott, N.L. (2004). Imaging the binding ability of proteins immobilized on surfaces with different orientations by using liquid crystals. *J. Am. Chem. Soc.* 126, 9024–9032.
- Lummis, S.C.R., Beene, D.L., Lee, L.W., Lester, H.A., Broadhurst, R.W., and Dougherty, D.A. (2005). Cis–trans isomerization at a proline opens the pore of a neurotransmitter-gated ion channel. *Nature* 438, 248–252.
- Marvin, L.F., Roberts, M.A., and Fay, L.B. (2003). Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry in clinical chemistry. *Clin. Chim. Acta Int. J. Clin. Chem.* 337, 11–21.
- Mehl, R.A., Anderson, J.C., Santoro, S.W., Wang, L., Martin, A.B., King, D.S., Horn, D.M., and Schultz, P.G. (2003). Generation of a Bacterium with a 21 Amino Acid Genetic Code. *J. Am. Chem. Soc.* 125, 935–939.
- Min, D. H., and Mrksich, M. (2004). Peptide arrays: towards routine implementation. *Curr. Opin. Chem. Biol.* 8, 554–558.
- Min, D. H., Yeo, W.S., and Mrksich, M. (2004a). A method for connecting solution-phase enzyme activity assays with immobilized format analysis by mass spectrometry. *Anal. Chem.* 76, 3923–3929.
- Min, D. H., Su, J., and Mrksich, M. (2004b). Profiling kinase activities by using a peptide chip and mass spectrometry. *Angew. Chem. Int. Ed Engl.* 43, 5973–5977.
- Min, D. H., Tang, W.-J., and Mrksich, M. (2004c). Chemical screening by mass spectrometry to identify inhibitors of anthrax lethal factor. *Nat. Biotechnol.* 22, 717–723.
- Moore, P.A., Zhang, W., Rainey, G.J., Burke, S., Li, H., Huang, L., Gorlatov, S., Veri, M.C., Aggarwal, S., Yang, Y., et al. (2011). Application of dual affinity retargeting molecules to achieve optimal redirected T-cell killing of B-cell lymphoma. *Blood* 117, 4542–4551.
- Mrksich, M., Grunwell, J.R., and Whitesides, G.M. (1995). Biospecific Adsorption of Carbonic Anhydrase to Self-Assembled Monolayers of Alkanethiolates That Present Benzenesulfonamide Groups on Gold. *J. Am. Chem. Soc.* 117, 12009–12010.



- Mrksich, M. (2008). Mass spectrometry of self-assembled monolayers: a new tool for molecular surface science. *ACS Nano* 2, 7–18.
- Mu, J., Pinkstaff, J., Li, Z., Skidmore, L., Li, N., Myler, H., Dallas-Yang, Q., Putnam, A.-M., Yao, J., Bussell, S., et al. (2012). FGF21 Analogs of Sustained Action Enabled by Orthogonal Biosynthesis Demonstrate Enhanced Antidiabetic Pharmacology in Rodents. *Diabetes* 61, 505–512.
- 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, 182–188.
- Nuzzo, R.G., and Allara, D.L. (1983). Adsorption of bifunctional organic disulfides on gold surfaces. *J. Am. Chem. Soc.* 105, 4481–4483.
- Patel, N., Davies, M.C., Hartshorne, M., Heaton, R.J., Roberts, C.J., Tendler, S.J.B., and Williams, P.M. (1997). Immobilization of Protein Molecules onto Homogeneous and Mixed Carboxylate-Terminated Self-Assembled Monolayers. *Langmuir* 13, 6485–6490.
- Plass, T., Milles, S., Koehler, C., Schultz, C., and Lemke, E.A. (2011). Genetically Encoded Copper-Free Click Chemistry. *Angew. Chem. Int. Ed.* 50, 3878–3881.
- Prime, K.L., and Whitesides, G.M. (1991). Self-assembled organic monolayers: model systems for studying adsorption of proteins at surfaces. *Science* 252, 1164–1167.
- Santoro, S.W., Wang, L., Herberich, B., King, D.S., and Schultz, P.G. (2002). An efficient system for the evolution of aminoacyl-tRNA synthetase specificity. *Nat. Biotechnol.* 20, 1044–1048.
- Seim, K.L., Obermeyer, A.C., and Francis, M.B. (2011). Oxidative modification of native protein residues using cerium(IV) ammonium nitrate. *J. Am. Chem. Soc.* 133, 16970–16976.
- Sharma, N., Furter, R., Kast, P., and Tirrell, D.A. (2000). Efficient introduction of aryl bromide functionality into proteins in vivo. *FEBS Lett.* 467, 37–40.
- Shen, B.Q., Xu, K., Liu, L., Raab, H., Bhakta, S., Kenrick, M., Parsons-Reponte, K.L., Tien, J., Yu, S.-F., Mai, E., et al. (2012). Conjugation site modulates the in vivo stability and therapeutic activity of antibody-drug conjugates. *Nat. Biotechnol.* 30, 184–189.
- Sigal, G.B., Bamdad, C., Barberis, A., Strominger, J., and Whitesides, G.M. (1996). A self-assembled monolayer for the binding and study of histidine-tagged proteins by surface plasmon resonance. *Anal. Chem.* 68, 490–497.

- Sletten, E.M., and Bertozzi, C.R. (2009). Bioorthogonal chemistry: fishing for selectivity in a sea of functionality. *Angew. Chem. Int. Ed Engl.* *48*, 6974–6998.
- Stephanopoulos, N., and Francis, M.B. (2011). Choosing an effective protein bioconjugation strategy. *Nat. Chem. Biol.* *7*, 876–884.
- Strable, E., Prasuhn, D.E., Udit, A.K., Brown, S., Link, A.J., Ngo, J.T., Lander, G., Quispe, J., Potter, C.S., Carragher, B., et al. (2008). Unnatural Amino Acid Incorporation into Virus-Like Particles. *Bioconjug. Chem.* *19*, 866–875.
- 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.
- Su, J., Rajapaksha, T.W., Peter, M.E., and Mrksich, M. (2006). Assays of endogenous caspase activities: a comparison of mass spectrometry and fluorescence formats. *Anal. Chem.* *78*, 4945–4951.
- Summerer, D., Chen, S., Wu, N., Deiters, A., Chin, J.W., and Schultz, P.G. (2006). A genetically encoded fluorescent amino acid. *Proc. Natl. Acad. Sci.* *103*, 9785–9789.
- Tilley, S.D., Joshi, N.S., and Francis, M.B. (2008). The Chemistry and Chemical Reactivity of Proteins. In *The Wiley Encyclopedia of Chemical Biology*, (Weinheim: Begley T; Wiley-VCH),.
- Ting, A.Y., Shin, I., Lucero, C., and Schultz, P.G. (1998). Energetic Analysis of an Engineered Cation- $\pi$  Interaction in Staphylococcal Nuclease. *J. Am. Chem. Soc.* *120*, 7135–7136.
- 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.
- Triola, G., Brunsveld, L., and Waldmann, H. (2008). Racemization-free synthesis of S-alkylated cysteines via thiol-ene reaction. *J. Org. Chem.* *73*, 3646–3649.
- Ulman, A. (1996). Formation and structure of self-assembled monolayers. *Chem. Rev.* *96*, 1533–1554.
- Vericat, C., Vela, M.E., and Salvarezza, R.C. (2005). Self-assembled monolayers of alkanethiols on Au(111): surface structures, defects and dynamics. *Phys. Chem. Chem. Phys.* *PCCP* *7*, 3258–3268.

- Walsh, C. (2006). *Posttranslational Modification of Proteins: Expanding Nature's Inventory* (Roberts and Company Publishers).
- Wang, L., and Schultz, P.G. (2005). Expanding the Genetic Code. *Angew. Chem. Int. Ed.* *44*, 34–66.
- Wang, L., Xie, J., and Schultz, P.G. (2006). A Genetically Encoded Fluorescent Amino Acid. *J. Am. Chem. Soc.* *128*, 8738–8739.
- Wang, L., Zhang, Z., Brock, A., and Schultz, P.G. (2003a). Addition of the keto functional group to the genetic code of *Escherichia coli*. *Proc. Natl. Acad. Sci.* *100*, 56–61.
- Wang, L., Xie, J., Deniz, A.A., and Schultz, P.G. (2003b). Unnatural Amino Acid Mutagenesis of Green Fluorescent Protein. *J. Org. Chem.* *68*, 174–176.
- Whitesides, G.M., and Laibinis, P.E. (1990). Wet chemical approaches to the characterization of organic surfaces: self-assembled monolayers, wetting, and the physical-organic chemistry of the solid-liquid interface. *Langmuir* *6*, 87–96.
- Whitesides, G.M., Kriebel, J.K., and Love, J.C. (2005). Molecular engineering of surfaces using self-assembled monolayers. *Sci. Prog.* *88*, 17–48.
- Wittrock, S., Becker, T., and Kunz, H. (2007). Synthetic vaccines of tumor-associated glycopeptide antigens by immune-compatible thioether linkage to bovine serum albumin. *Angew. Chem. Int. Ed Engl.* *46*, 5226–5230.
- Witus, L.S., Moore, T., Thuronyi, B.W., Esser-Kahn, A.P., Scheck, R.A., Iavarone, A.T., and Francis, M.B. (2010). Identification of highly reactive sequences for PLP-mediated bioconjugation using a combinatorial peptide library. *J. Am. Chem. Soc.* *132*, 16812–16817.
- Wu, C., Ying, H., Grinnell, C., Bryant, S., Miller, R., Clabbers, A., Bose, S., McCarthy, D., Zhu, R.-R., Santora, L., et al. (2007). Simultaneous targeting of multiple disease mediators by a dual-variable-domain immunoglobulin. *Nat. Biotechnol.* *25*, 1290–1297.
- Wu, N., Deiters, A., Cropp, T.A., King, D., and Schultz, P.G. (2004). A Genetically Encoded Photocaged Amino Acid. *J. Am. Chem. Soc.* *126*, 14306–14307.
- Xie, J., and Schultz, P.G. (2006). A chemical toolkit for proteins--an expanded genetic code. *Nat. Rev. Mol. Cell Biol.* *7*, 775–782.

Xiu, X., Puskar, N.L., Shanata, J.A.P., Lester, H.A., and Dougherty, D.A. (2009). Nicotine binding to brain receptors requires a strong cation– $\pi$  interaction. *Nature* 458, 534–537.

Ye, S., Köhrer, C., Huber, T., Kazmi, M., Sachdev, P., Yan, E.C.Y., Bhagat, A., RajBhandary, U.L., and Sakmar, T.P. (2008). Site-specific Incorporation of Keto Amino Acids into Functional G Protein-coupled Receptors Using Unnatural Amino Acid Mutagenesis. *J. Biol. Chem.* 283, 1525–1533.

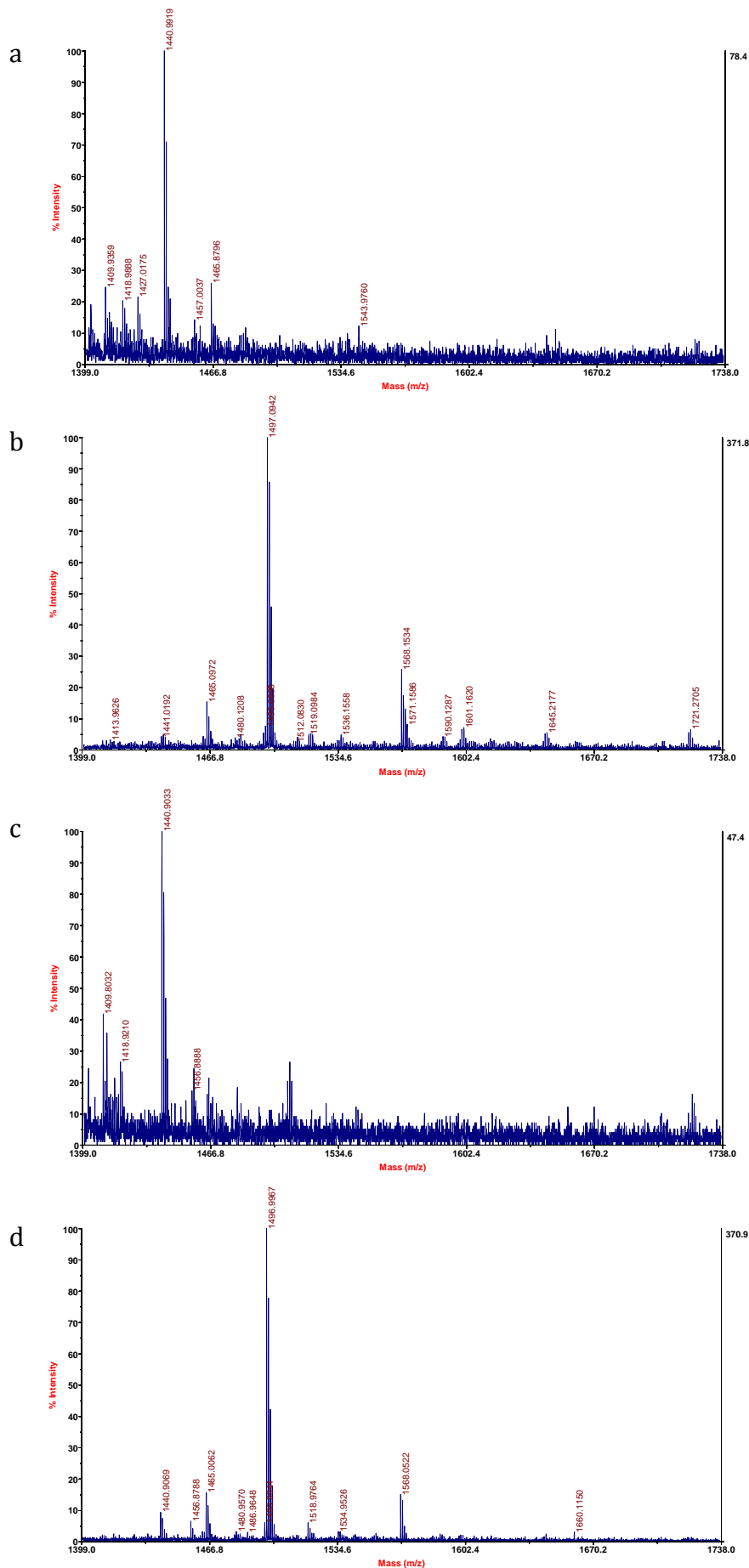
Young, T.S., Ahmad, I., Brock, A., and Schultz, P.G. (2009). Expanding the Genetic Repertoire of the Methylophilic Yeast *Pichia pastoris*†. *Biochemistry (Mosc.)* 48, 2643–2653.

Zhang, Z., Wang, L., Brock, A., and Schultz, P.G. (2002). The Selective Incorporation of Alkenes into Proteins in *Escherichia coli*. *Angew. Chem.* 114, 2964–2966.

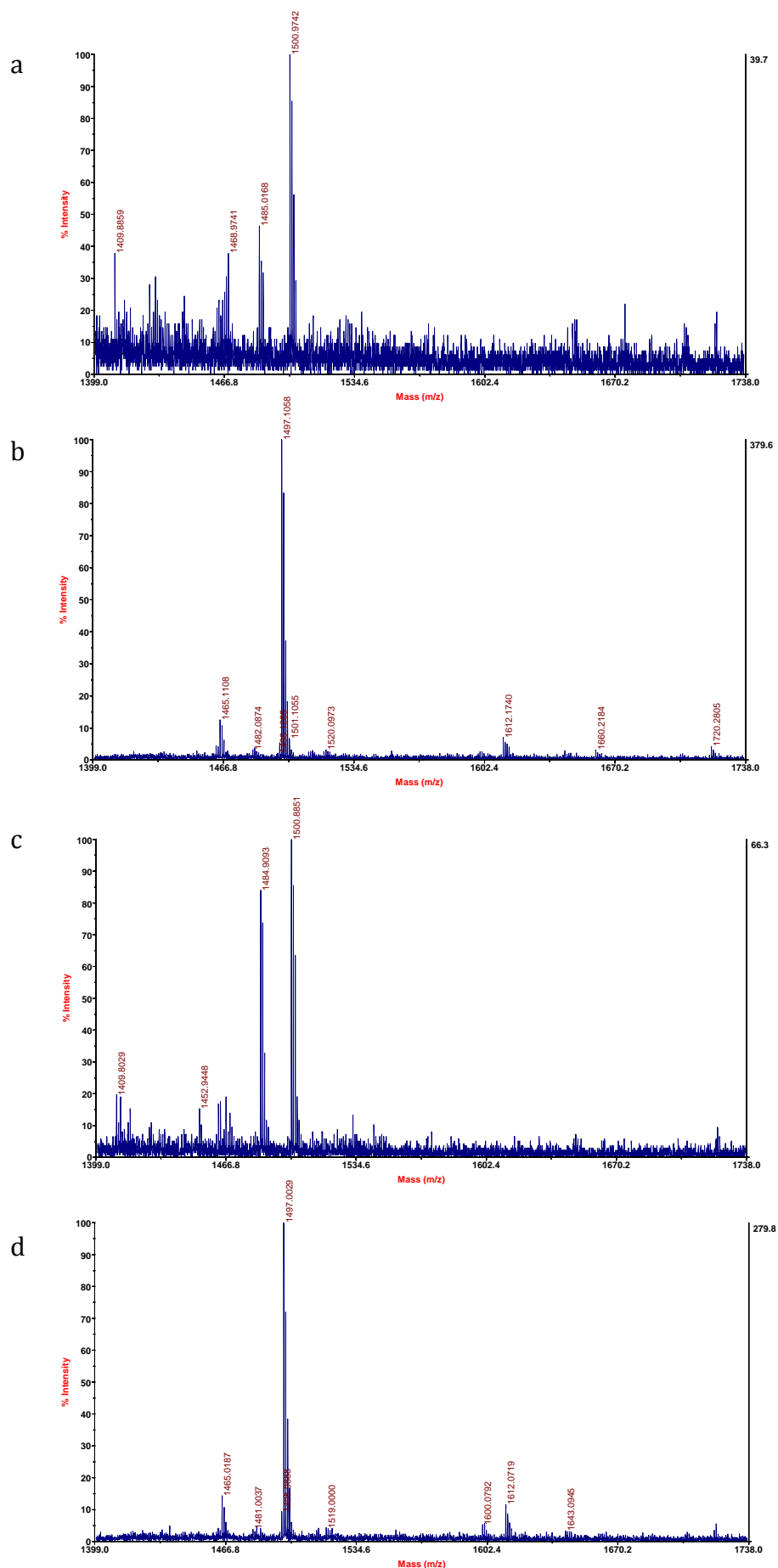
Zhang, Z., Smith, B.A.C., Wang, L., Brock, A., Cho, C., and Schultz, P.G. (2003). A New Strategy for the Site-Specific Modification of Proteins in Vivo†. *Biochemistry (Mosc.)* 42, 6735–6746.

Zhang, Z., Gildersleeve, J., Yang, Y.-Y., Xu, R., Loo, J.A., Uryu, S., Wong, C.-H., and Schultz, P.G. (2004). A New Strategy for the Synthesis of Glycoproteins. *Science* 303, 371–373.

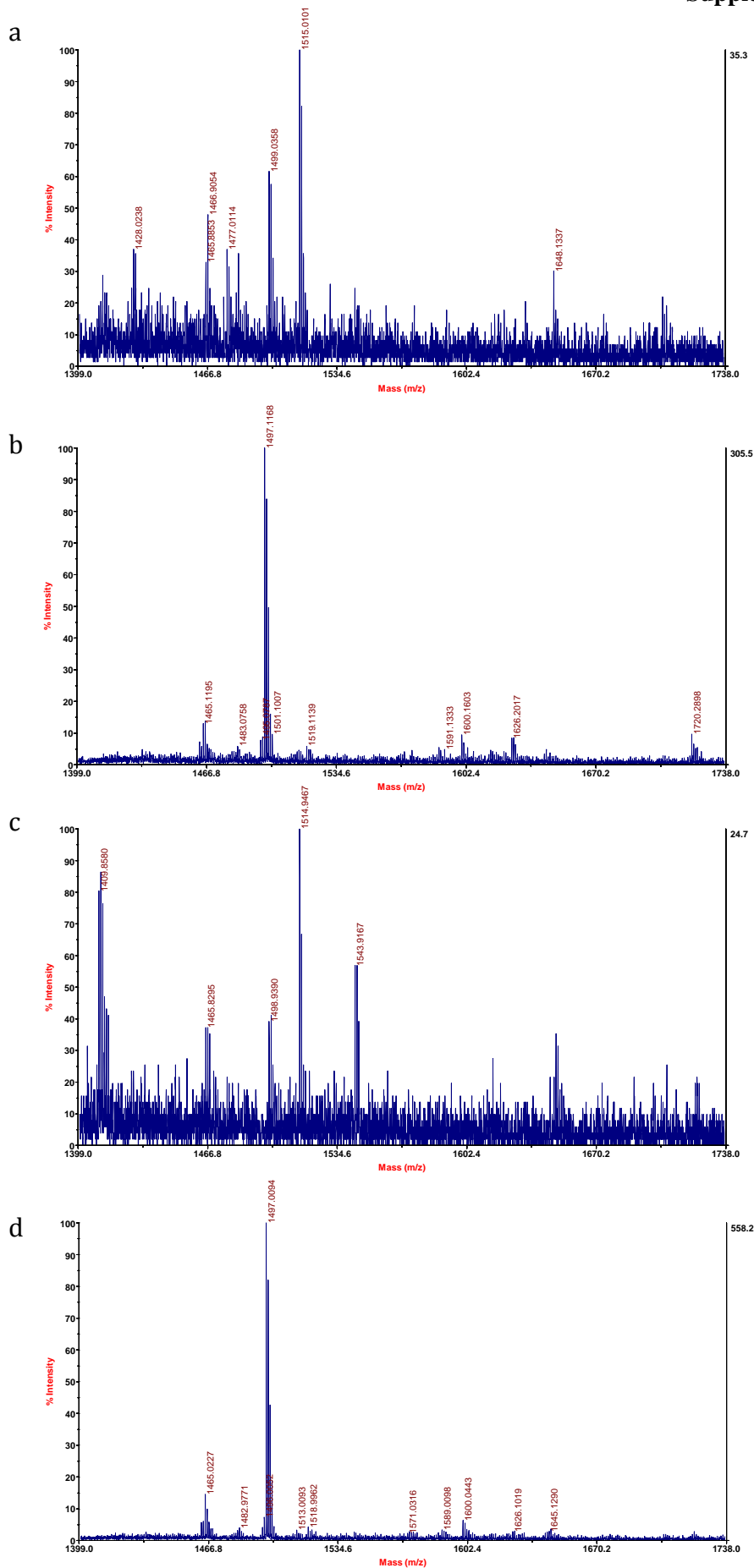
# Supplementary Materials



**Figure S.1.** MALDI spectra of Ac-GGYAGC (a, b) and Ac-GAYGGC (c, d) anchored to the monolayer in the presence of 50 mM BIS-TRIS buffer (a, c) m/z 1440 sodium adduct, and after 10 min incubation with 0.1 mM anisidine and 1 mM CAN in BIS-TRIS (b, d), product m/z 1568 ( $H^+$ ).

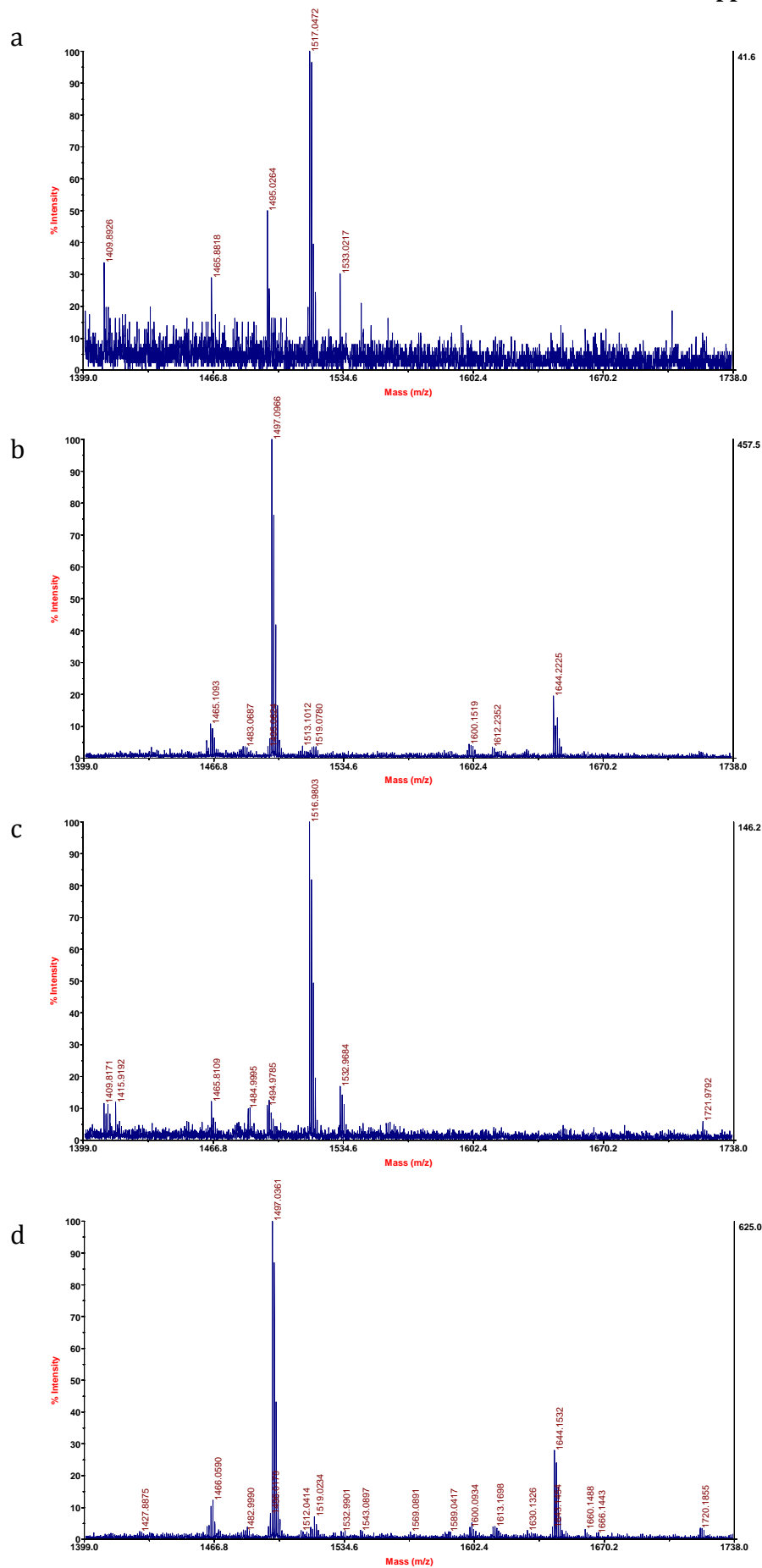


**Figure S.2.** MALDI spectra of Ac-GGYDGC (a, b) and Ac-GDYGGC (c, d) anchored to the monolayer in the presence of 50 mM BIS-TRIS buffer (a, c) m/z 1500 potassium adduct, and after 10 min incubation with 0.1 mM anisidine and 1 mM CAN in BIS-TRIS (b, d), product m/z 1612 ( $H^+$ ).

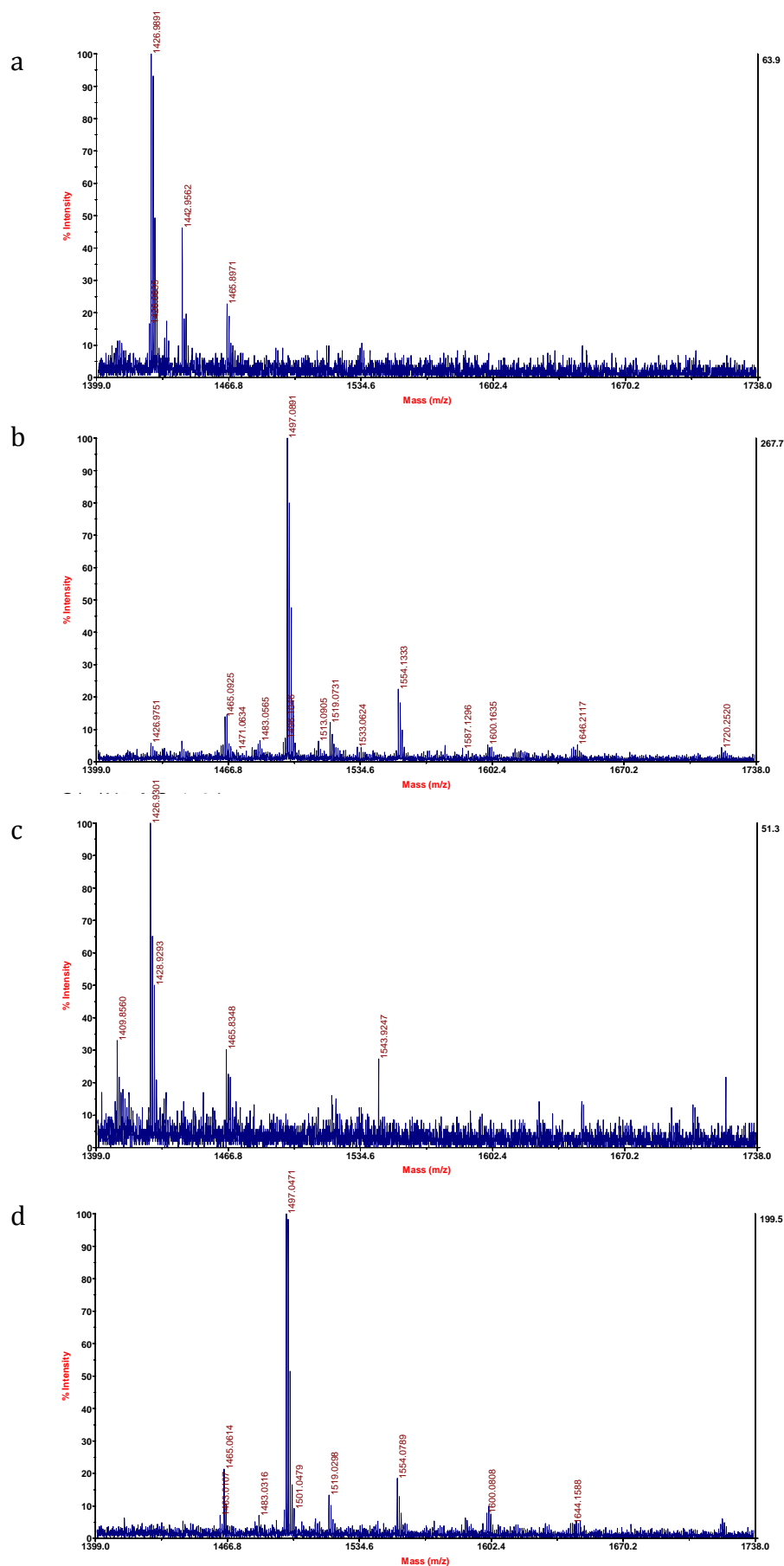


**Figure S.3.** MALDI spectra of Ac-GGYEGC (a, b) and Ac-GEYGGC (c, d) anchored to the monolayer in the presence of 50 mM BIS-TRIS buffer (a, c) m/z 1515 potassium adduct, and after 10 min incubation with 0.1 mM anisidine and 1 mM CAN in BIS-TRIS (b, d), product m/z 1626 ( $H^+$ ).

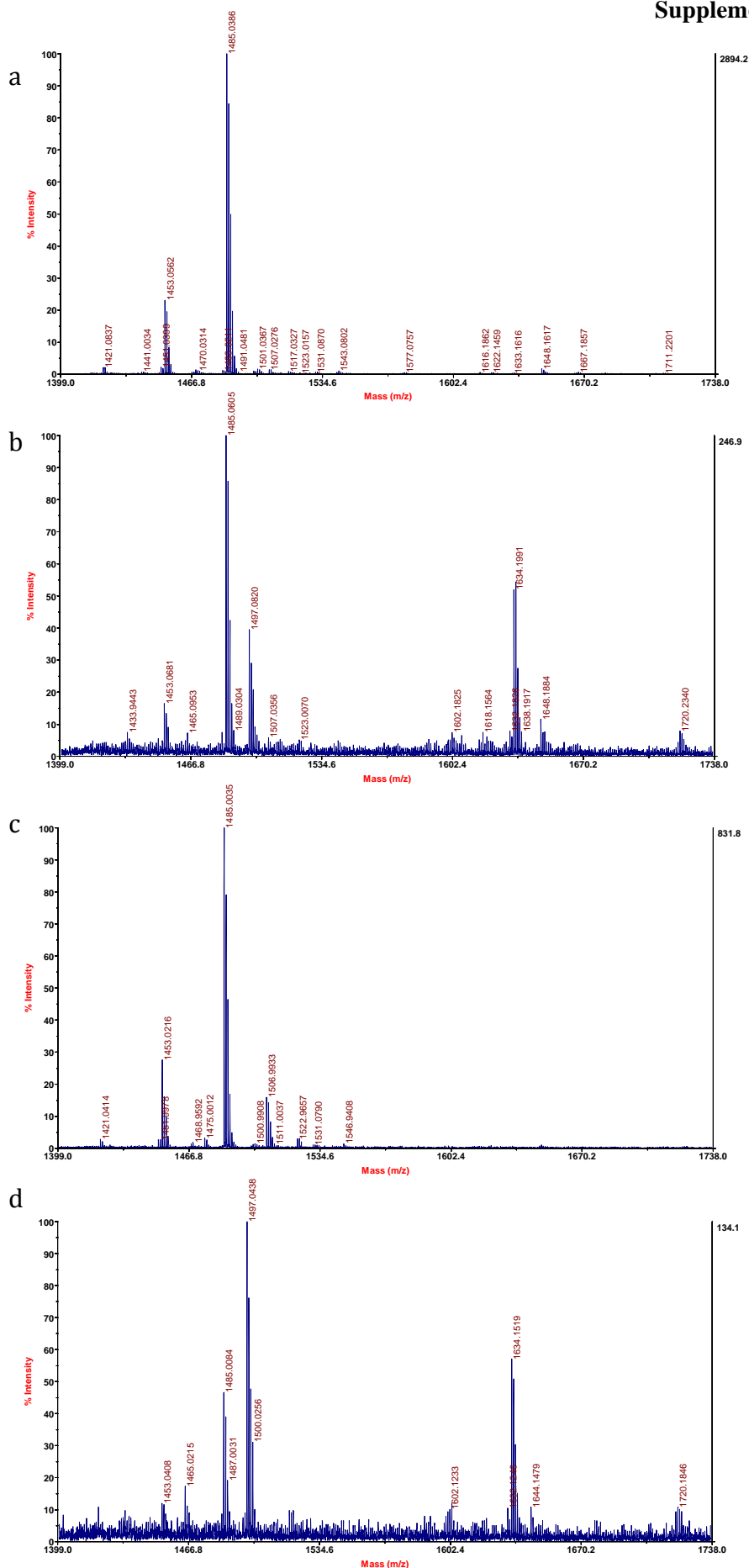




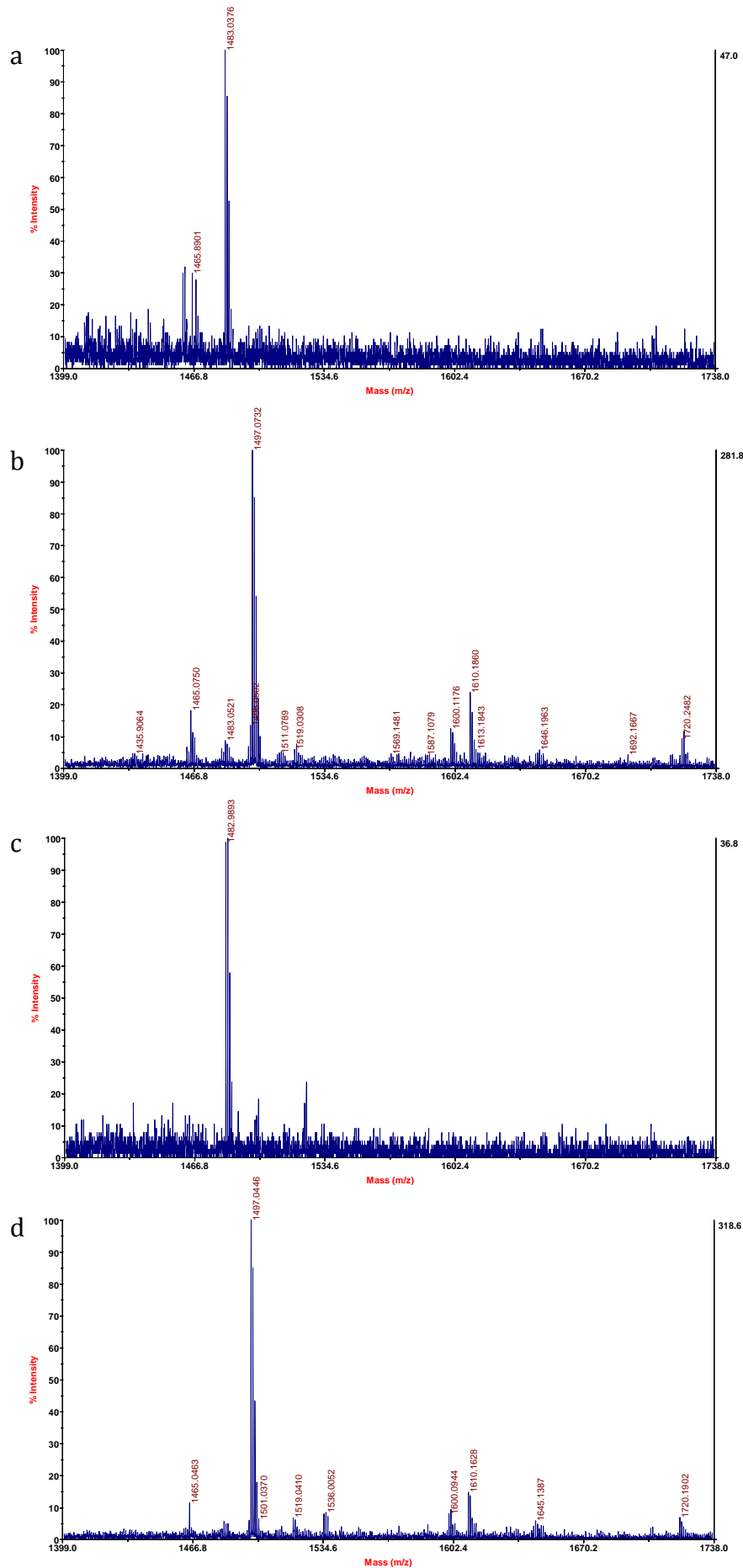
**Figure S.4.** MALDI spectra of Ac-GGYFGC (a, b) and Ac-GFYGGC (c, d) anchored to the monolayer in the presence of 50 mM BIS-TRIS buffer (a, c) m/z 1517 sodium adduct, and after 10 min incubation with 0.1 mM anisidine and 1 mM CAN in BIS-TRIS (b, d), product m/z 1644 ( $H^+$ ).



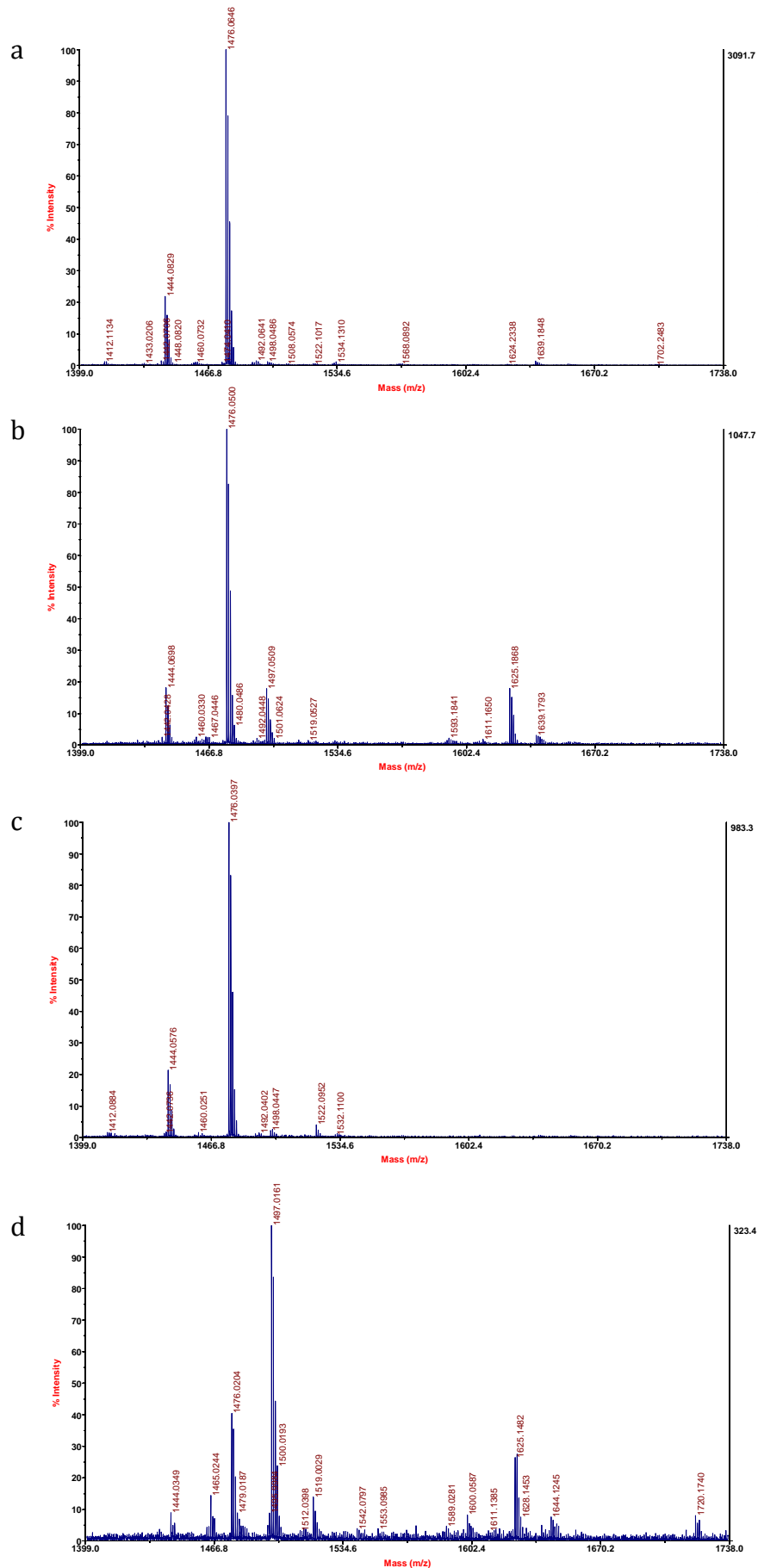
**Figure S.5.** MALDI spectra of Ac-GGYGGC (a, b) and Ac-GGYGGC (c, d) anchored to the monolayer in the presence of 50 mM BIS-TRIS buffer (a, c) m/z 1426 sodium adduct, and after 10 min incubation with 0.1 mM anisidine and 1 mM CAN in BIS-TRIS (b, d), product m/z 1554 ( $H^+$ ).



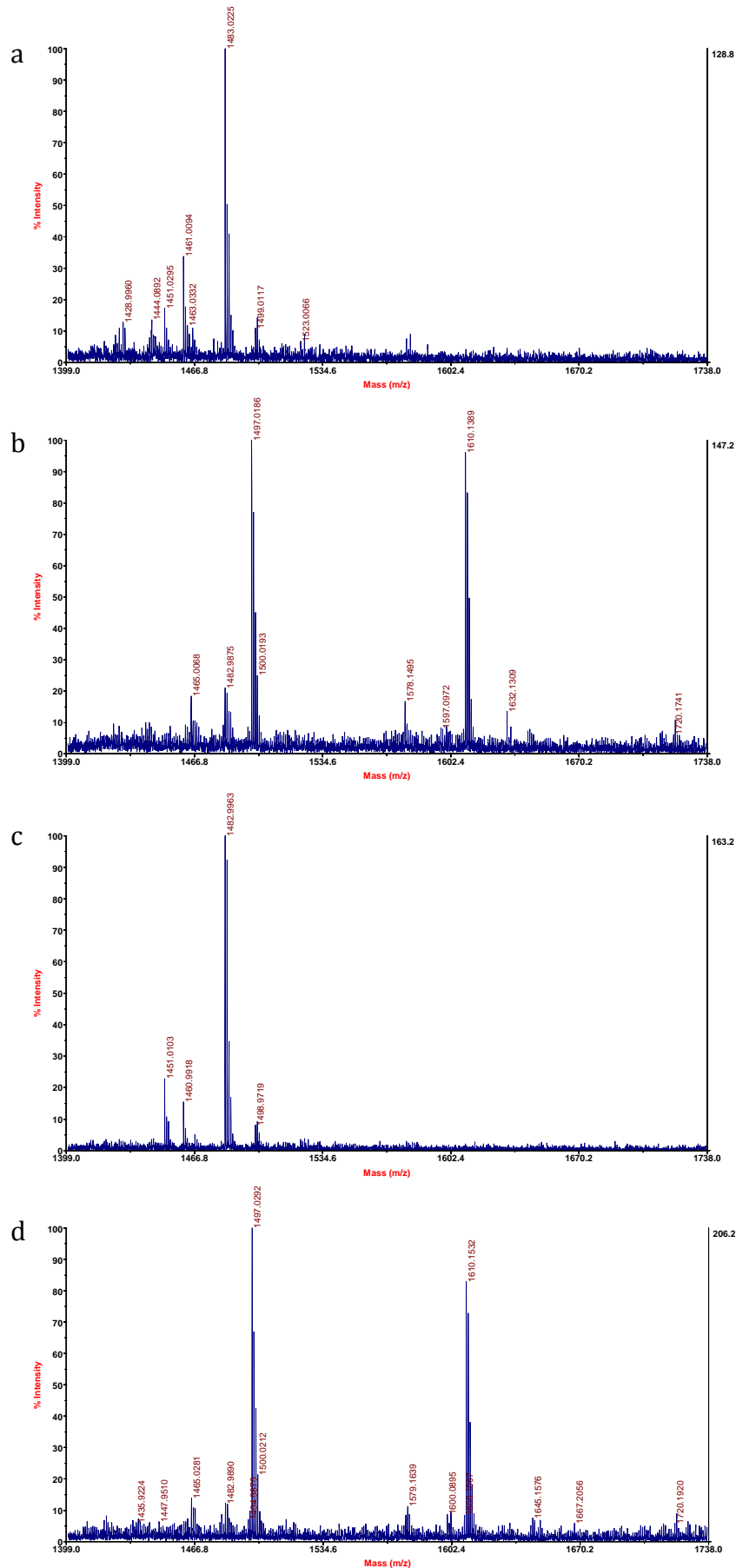
**Figure S.6.** MALDI spectra of Ac-GGYHGC (a, b) and Ac-GHYGGC (c, d) anchored to the monolayer in the presence of 50 mM BIS-TRIS buffer (a, c) m/z 1485, and after 10 min incubation with 0.1 mM anisidine and 1 mM CAN in BIS-TRIS (b, d), product m/z 1634.



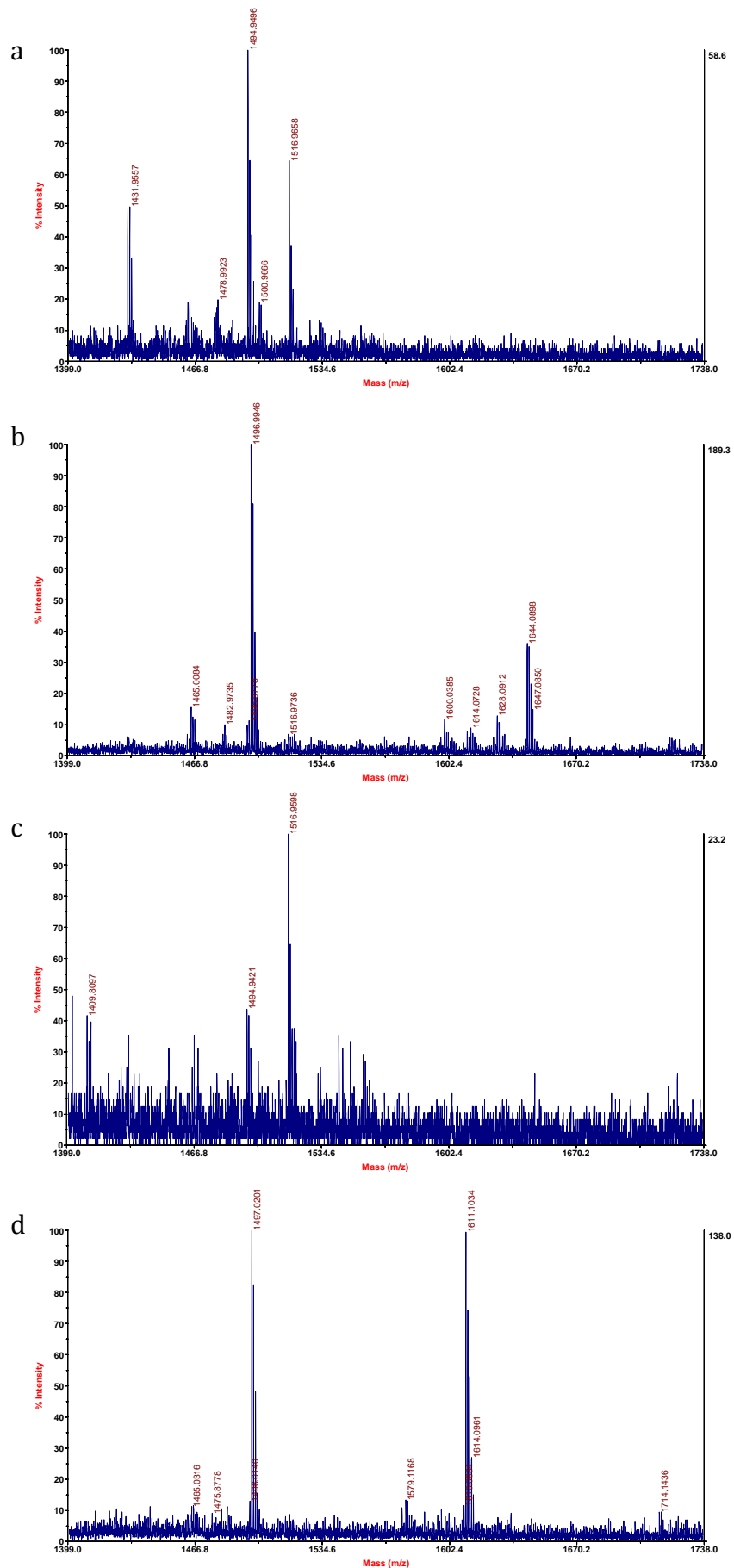
**Figure S.7.** MALDI spectra of Ac-GGYIGC (a, b) and Ac-GIYGGC (c, d) anchored to the monolayer in the presence of 50 mM BIS-TRIS buffer (a, c) m/z 1483 sodium adduct, and after 10 min incubation with 0.1 mM anisidine and 1 mM CAN in BIS-TRIS (b, d), product m/z 1610 ( $H^+$ ).



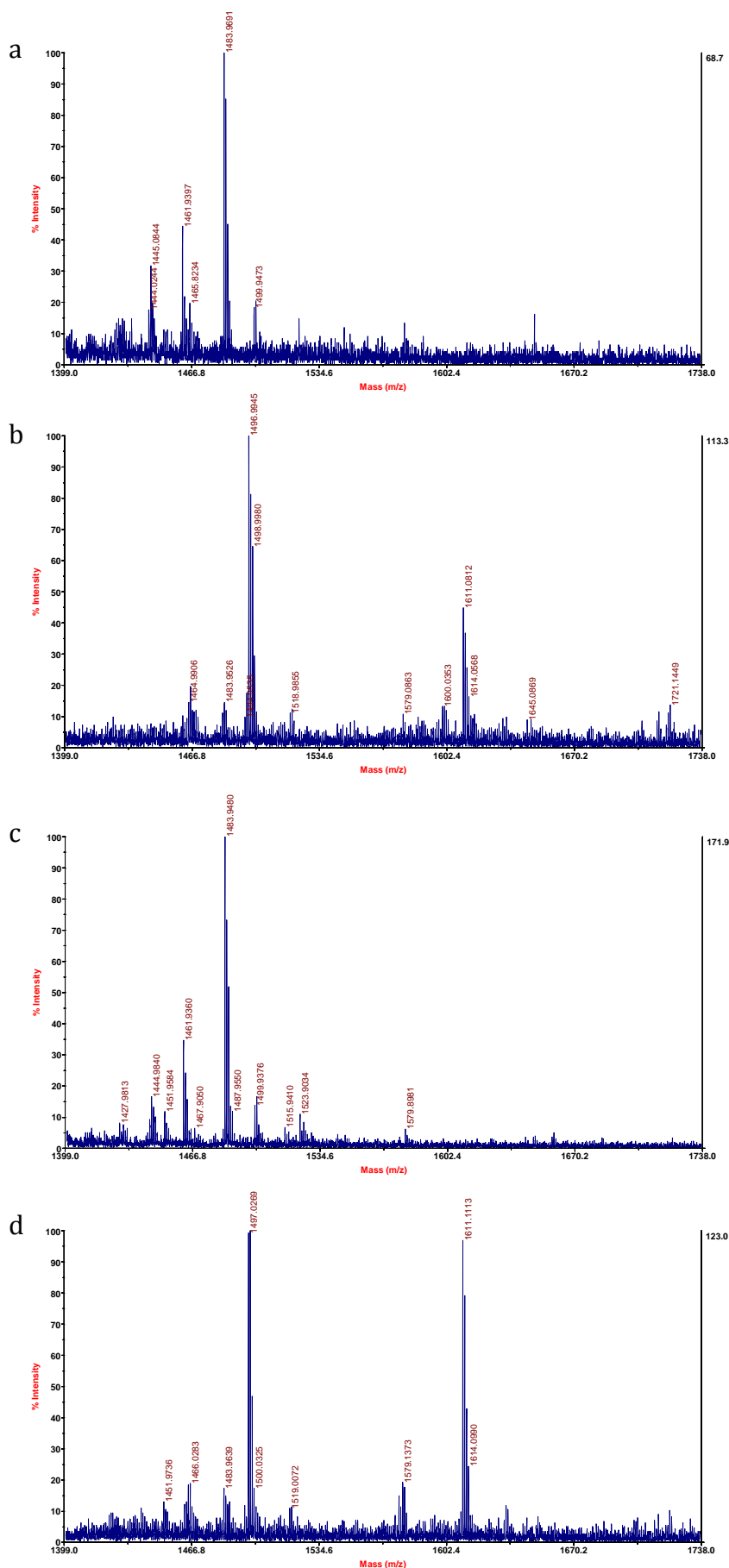
**Figure S.8.** MALDI spectra of Ac-GGYKGC (a, b) and Ac-GKYGGC (c, d) anchored to the monolayer in the presence of 50 mM BIS-TRIS buffer (a, c) m/z 1476, and after 10 min incubation with 0.1 mM anisidine and 1 mM CAN in BIS-TRIS (b, d), product m/z 1625 ( $H^+$ ).



**Figure S.9.** MALDI spectra of Ac-GGYLGC (a, b) and Ac-GLYGGC (c, d) anchored to the monolayer in the presence of 50 mM BIS-TRIS buffer (a, c) m/z 1482 sodium adduct, and after 10 min incubation with 0.1 mM anisidine and 1 mM CAN in BIS-TRIS (b, d), product m/z 1610 ( $H^+$ ).

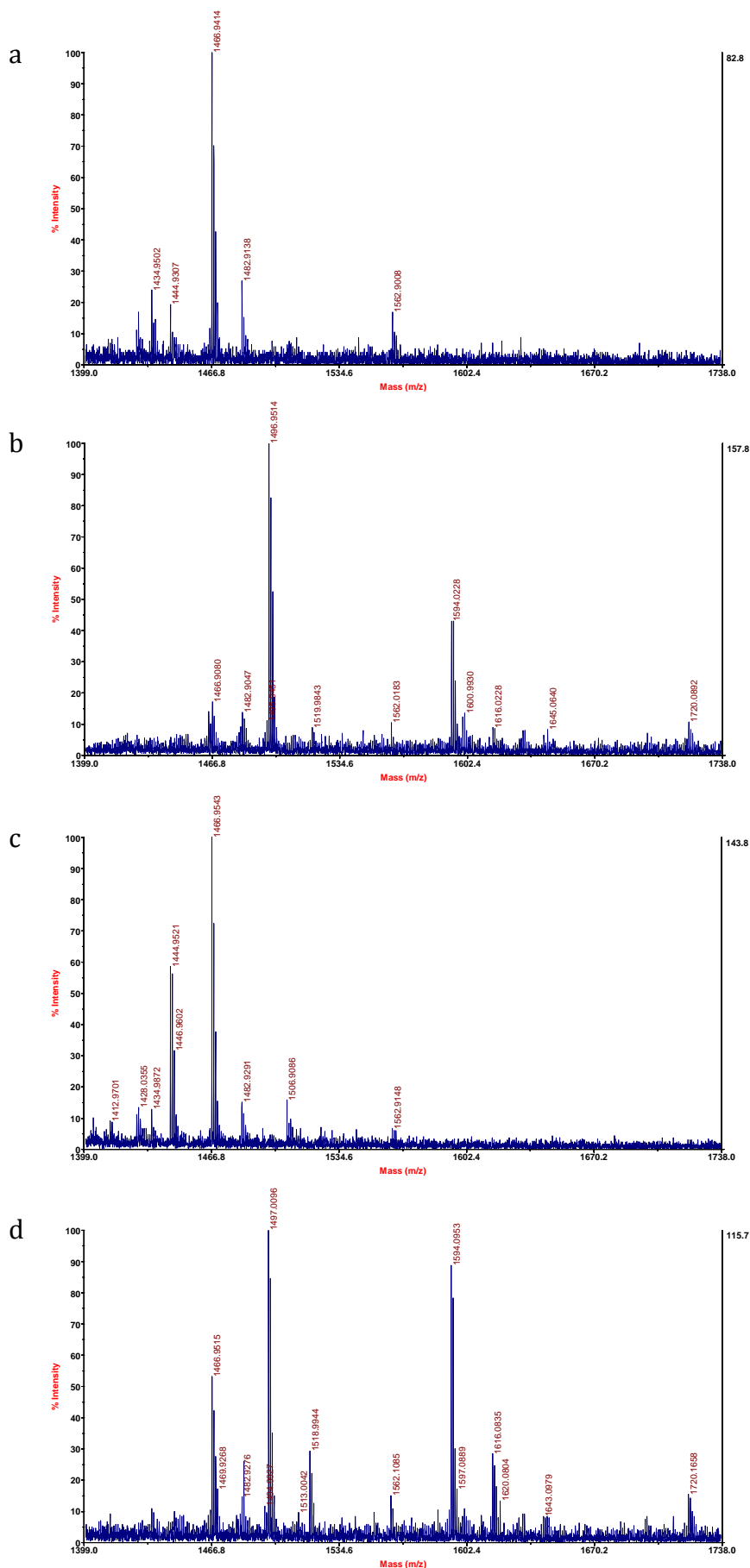


**Figure S.10.** MALDI spectra of Ac-GGYMGC (a, b) and Ac-GMYGGC (c, d) anchored to the monolayer in the presence of 50 mM BIS-TRIS buffer (a, c) m/z 1516 potassium adduct, and after 10 min incubation with 0.1 mM anisidine and 1 mM CAN in BIS-TRIS (b, d), product m/z 1628 ( $H^+$ ).

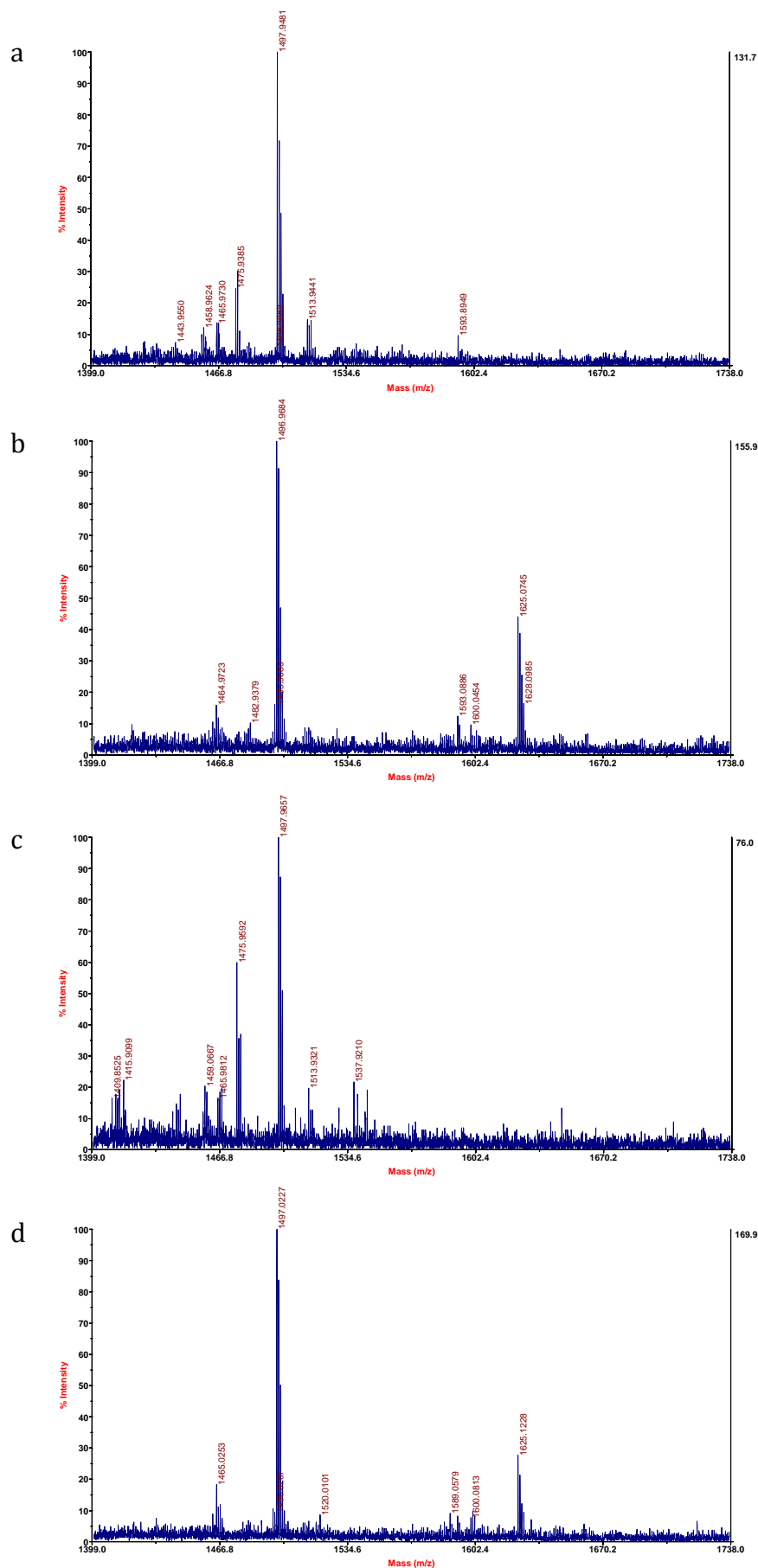


**Figure S.11.** MALDI spectra of Ac-GGYNGC (a, b) and Ac-GNYGGC (c, d) anchored to the monolayer in the presence of 50 mM BIS-TRIS buffer (a, c) m/z 1483 sodium adduct, and after 10 min incubation with 0.1 mM anisidine and 1 mM CAN in BIS-TRIS (b, d), product m/z 1611 ( $H^+$ ).

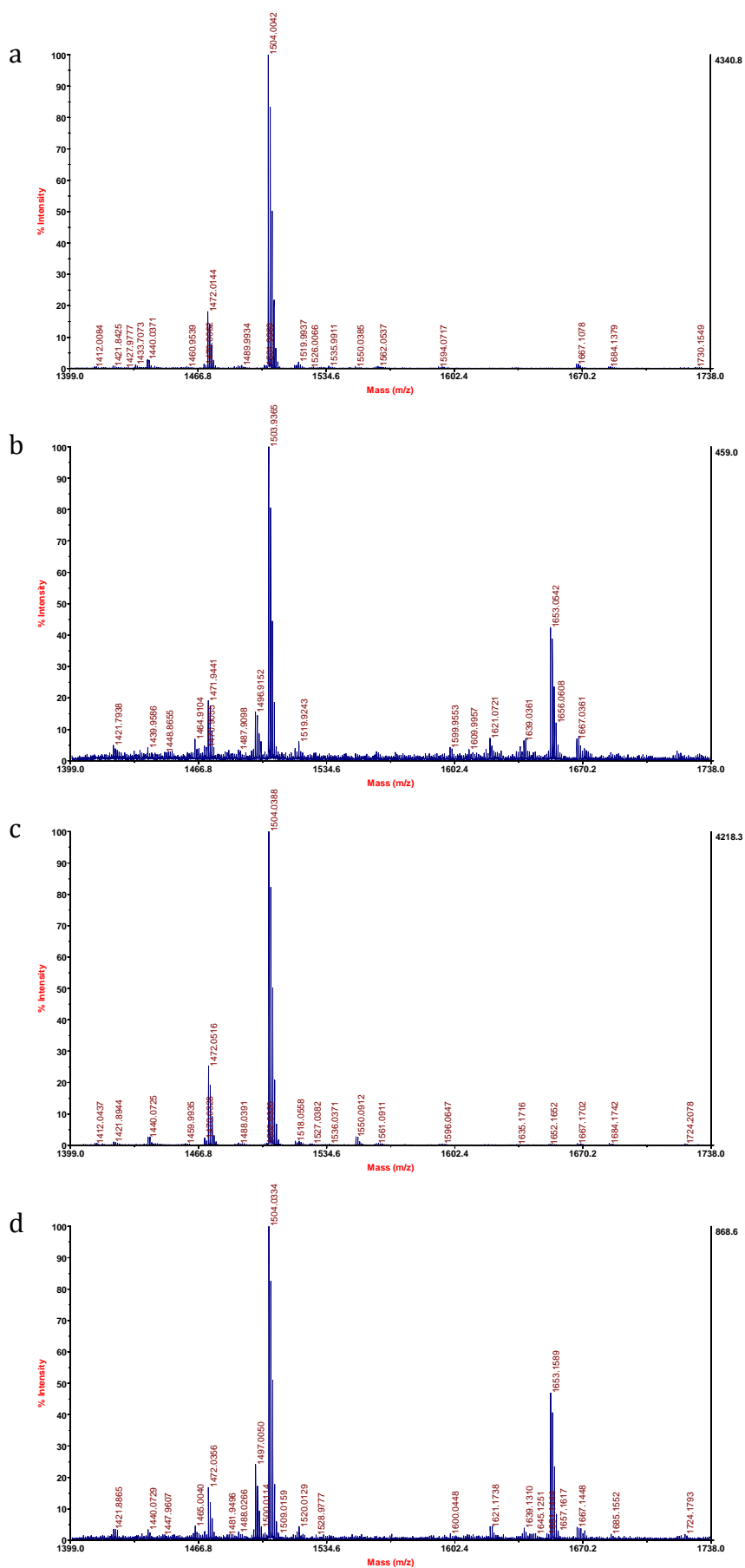




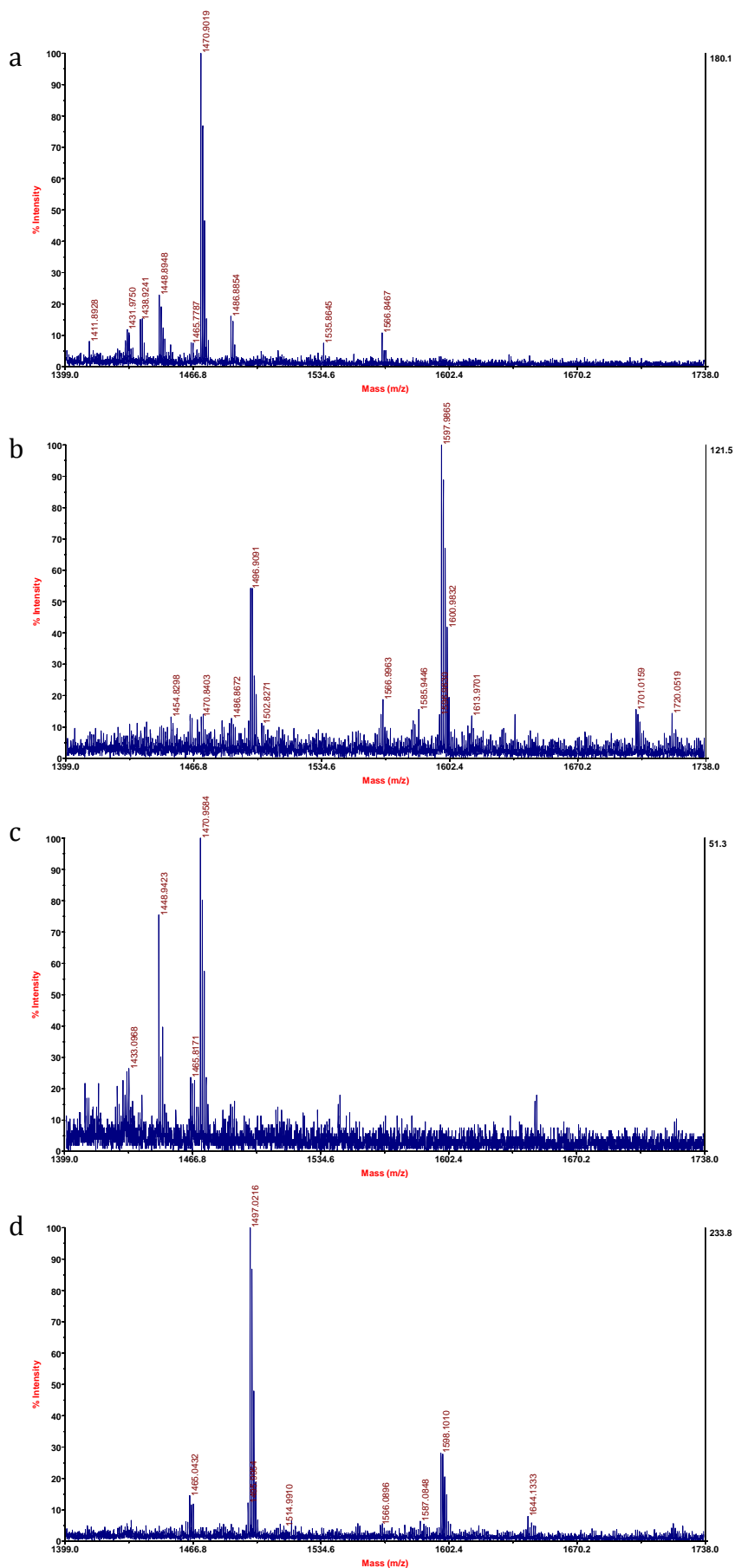
**Figure S.12.** MALDI spectra of Ac-GGYPGC (a, b) and Ac-GPYGGC (c, d) anchored to the monolayer in the presence of 50 mM BIS-TRIS buffer (a, c) m/z 1466 sodium adduct, and after 10 min incubation with 0.1 mM anisidine and 1 mM CAN in BIS-TRIS (b, d), product m/z 1594 ( $H^+$ ).



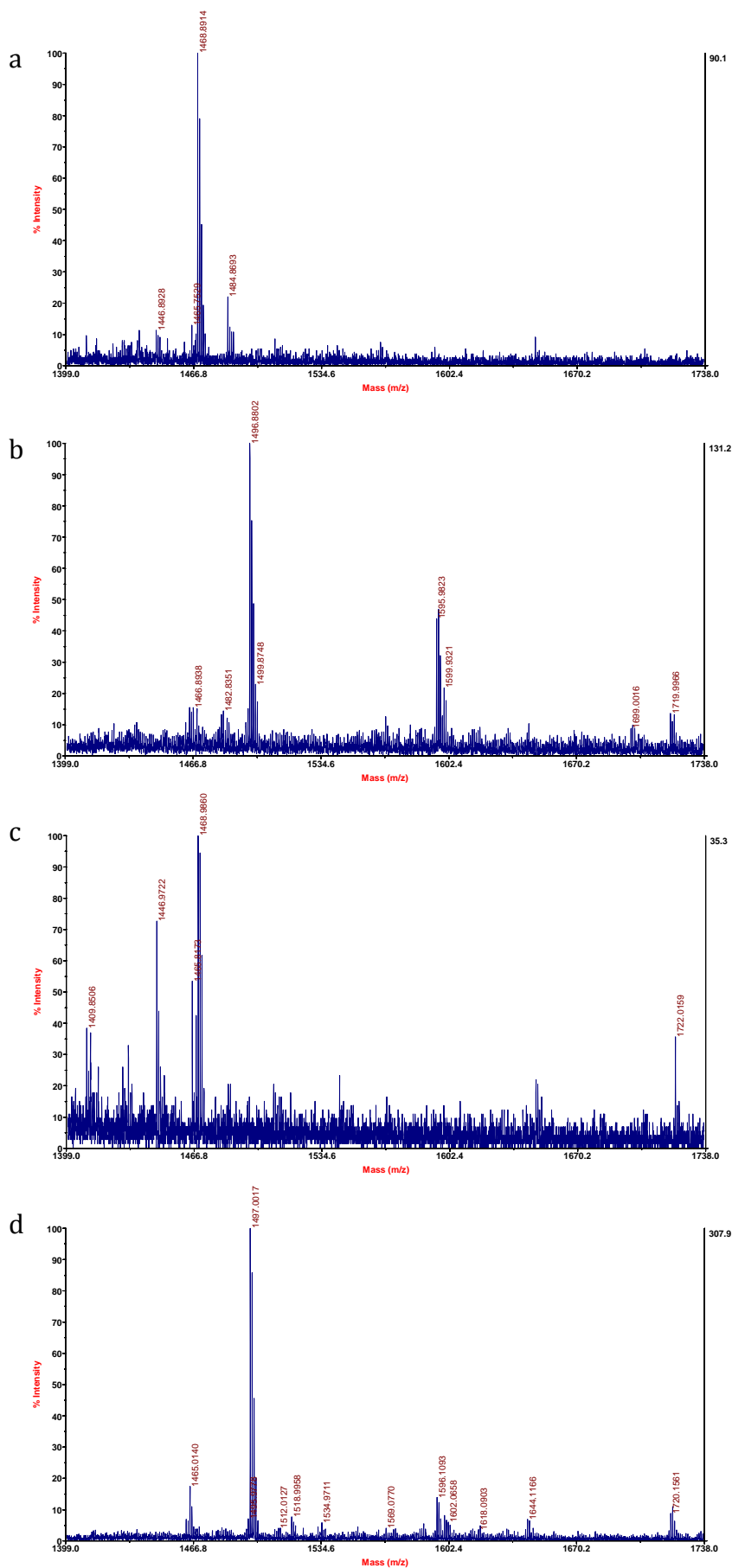
**Figure S.13.** MALDI spectra of Ac-GGYQGC (a, b) and Ac-GQYGGC (c, d) anchored to the monolayer in the presence of 50 mM BIS-TRIS buffer (a, c) m/z 1497 sodium adduct, and after 10 min incubation with 0.1 mM anisidine and 1 mM CAN in BIS-TRIS (b, d), product m/z 1625 ( $H^+$ )



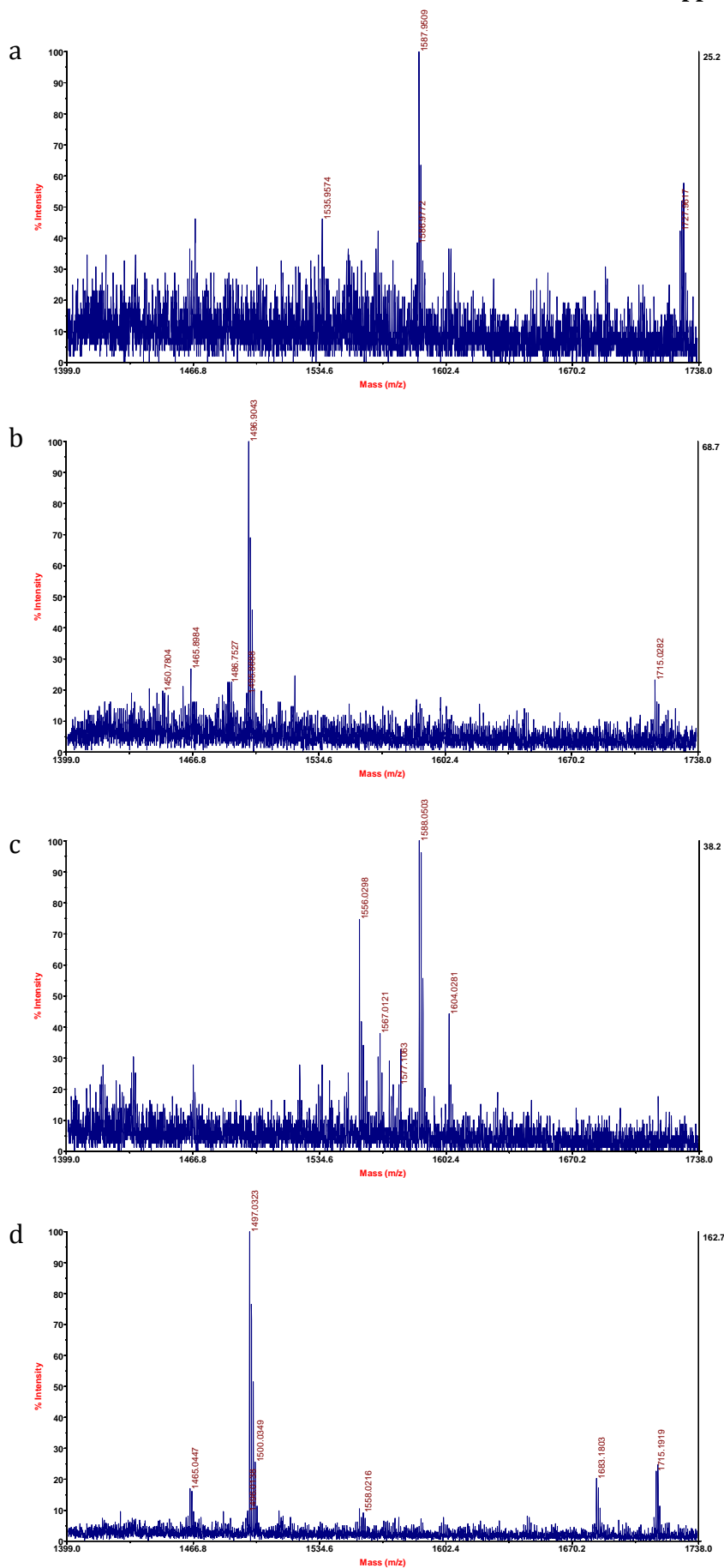
**Figure S.14.** MALDI spectra of Ac-GGYRGC (a, b) and Ac-GRYGGC (c, d) anchored to the monolayer in the presence of 50 mM BIS-TRIS buffer (a, c) m/z 1504, and after 10 min incubation with 0.1 mM anisidine and 1 mM CAN in BIS-TRIS (b, d), product m/z 1653 ( $H^+$ )



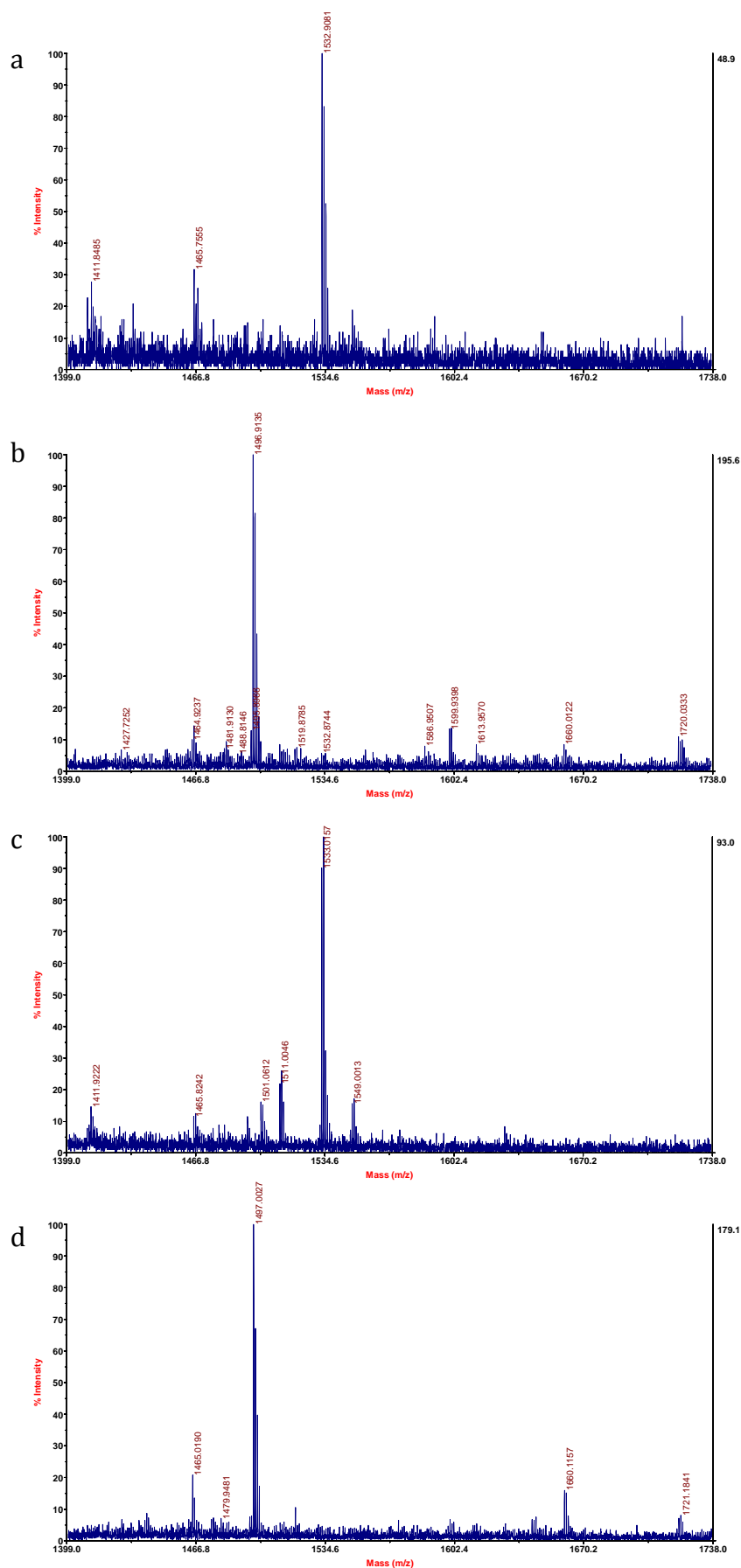
**Figure S.15.** MALDI spectra of Ac-GGYTGC (a, b) and Ac-GTYGGC (c, d) anchored to the monolayer in the presence of 50 mM BIS-TRIS buffer (a, c) m/z 1470 sodium adduct, and after 10 min incubation with 0.1 mM anisidine and 1 mM CAN in BIS-TRIS (b, d), product m/z 1598 (H<sup>+</sup>).



**Figure S.16.** MALDI spectra of Ac-GGYVGC (a, b) and Ac-GVYGGC (c, d) anchored to the monolayer in the presence of 50 mM BIS-TRIS buffer (a, c) m/z 1468 sodium adduct, and after 10 min incubation with 0.1 mM anisidine and 1 mM CAN in BIS-TRIS (b, d), product m/z 1595 ( $H^+$ ).



**Figure S.17.** MALDI spectra of Ac-GGYWGC (a, b) and Ac-GWYGGC (c, d) anchored to the monolayer in the presence of 50 mM BIS-TRIS buffer (a, c) m/z 1556 sodium adduct, and after 10 min incubation with 0.1 mM anisidine and 1 mM CAN in BIS-TRIS (b, d), product m/z 1683 ( $H^+$ ).



**Figure S.18.** MALDI spectra of Ac-GGYYGC (a, b) and Ac-GYYGGC (c, d) anchored to the monolayer in the presence of 50 mM BIS-TRIS buffer (a, c) m/z 1532 sodium adduct, and after 10 min incubation with 0.1 mM anisidine and 1 mM CAN in BIS-TRIS (b, d), product m/z 1660 ( $H^+$ ).

# Acknowledgments



I would first like to thank my tutor, Dr. Francesco Mantegazza, for giving me the opportunity of doing my PhD and working on this project.

I would also like to thank Dr. Milan Mrksich for welcoming me into his research group at Northwestern University and for introducing me to the incredible technique developed in his lab better known as SAMDI MS (and for letting me exploit it in my project). Milan's generosity combined with his extensive knowledge and enthusiasm for science, allowed him to become the great scientist he is.

I would also like to thank Dr. Matthew B. Francis and Dr. Kristen Seim for sharing with me their expertise about the CAN-mediated oxidative reaction they developed.

I am also thankful to Dr. Paolo Tortora and Dr. Davide Prosperi for their help and advice throughout my years as a graduate (and PhD) student. Their professional and personal support has never been lacking for real.

In addition, I have to thank all the labmates I worked with throughout these years, without them all, I would have done almost nothing!

I am grateful for the time I spent in Chicago, it's been a great professional opportunity but above all I have to thank all my friends for making me feel at home. Even my working experience wouldn't have been the same without them.

I cannot describe how thankful I am to my family for all their patience and support through out the years, I know it's not that easy being around me... so, thank you!

And last but not least, I want to thank all my friends: I cannot mention everyone, but I probably wouldn't even be here without your pressing (and loving) presence next to me. Once again, the only thing I have to say is...

THANK YOU!