MASS SPECTROMETRIC CHARACTERIZATION OF EXTENSIN CROSS-LINKAGES IN ARABIDOPSIS TISSUE CULTURE CELL WALLS

By

LAWRIE VEALE GAINEY

Bachelor of Science Biochemistry and Molecular Biology
Oklahoma State University
Stillwater, OK
2014

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
May, 2020

MASS SPECTROMETRIC CHARACTERIZATION OF EXTENSIN CROSS-LINKAGES IN ARABIDOPSIS TISSUE CULTURE CELL WALLS

Dissertation Approved:

Dr. Andrew J. Mort

Dissertation Adviser

Dr. Steven D. Hartson

Dr. Robert L. Matts

Dr. Junpeng Deng

Dr. Gabriel Cook

ACKNOWLEDGEMENTS

For my family, friends, and mentors that I've been fortunate to gain along my journey.

Name: LAWRIE VEALE GAINEY

Date of Degree: MAY, 2020

Title of Study: MASS SPECTROMETRIC CHARACTERIZATION OF EXTENSIN CROSS-LINKAGES IN ARABIDOPSIS TISSUE CULTURE CELL WALLS

Major Field: BIOCHEMISTRY AND MOLECULAR BIOLOGY

Abstract: In this work we characterize and identify tyrosine-tyrosine cross-linkages occurring in the Arabidopsis extensin scaffold. Cross-linked extensin peptides were solubilized out of Arabidopsis tissue culture cell walls using chemical and enzymatic methods adapted from Qi et al. The resulting peptides were analyzed by liquid chromatography mass spectrometry (LC-MS/MS). Peptides derived from four hydroxyproline-rich extensins and one leucine-rich extensin were identified. Four different types of tyrosine cross-linkages were identified and characterized. One intramolecular linkage was identified: isodityrosine. This cross-linkage occurred at the amino acid sequence yXy (where y is a crosslinked tyrosine, and X is any amino acid). Three intermolecular linkages between two peptides were also identified: dityrosine, pulcherosine, and di-isodityrosine made up of two, three, and four tyrosines respectively. This is the first identification and characterization of cross-linkages formed in the extensin scaffold in *vivo.* These tyrosine containing cross-linkages corroborate earlier suggestions that the extensin scaffold is a highly cross-linked network *in muro*.

TABLE OF CONTENTS

Chapter Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

CHAPTER I

LITERATURE REVIEW

The Primary Plant Cell Wall

Responsible for the first line of defense and growth, is a thin region of the plant cell wall. This region is named, the primary cell wall. All plants have primary cell walls. Growth of new cells start with the formation of the primary cell wall. When a cell divides, a cell plate forms at the location at which the cell will divide. From there, the primary wall continually expands and encompasses the growing plant cell.[1] There are two major types of primary cell walls in flowering plants: type I and type II.[2] Both types of primary cell walls consist of highly cross-linked networks of polymers that continually self-assemble to form the cells creating a growing plant. Type I cell walls exist in dicots and most monocot plants, and are the focus of this dissertation. The type I cell wall consists of three self-assembling networks of interpenetrating carbohydrate polymers: cellulose, hemicellulose, and pectin, along with hydroxyproline-rich glycoproteins, extensins.[3,4] The self-assembly of the cell wall is dependent on the correct self-assembly of the extensin scaffold.[5]

Structural Polysaccharides

Cellulose

Cellulose is the most abundant carbohydrate on earth. Cellulose is a linear

polymer consisting of repeating units of β-(1, 4)-D-glucose that create glucan chains. After synthesis, thirty-six glucan chains aggregate parallel to each other through hydrogen bonds to form crystalline bundles called microfibrils. Microfibrils are formed in the plasma membrane by the cellulose synthase complex.[1,6,7] Cellulose microfibrils are the major load bearing component of cell walls. They serve as a framework that gives the cell wall mechanical strength and rigidity.

Hemicellulose

Hemicelluloses strengthen the cell wall through extensive cross-linking with the cellulose microfibrils.[1] The hemicelluloses are a group of polysaccharides that generally have backbones composed of repeating β-(1, 4)-D-glucose, xylose, or mannose units. This backbone can be extensively decorated with several types of monosaccharides such as arabinose, galactose, fucose, and/or glucuronic acids. Non-sugar components such as methyl esters, ethers, acetyl esters, and/or feruloyl esters may also decorate the hemicelluloses. The degree of substitution of both sugar and non-sugar components affects the structural properties of the hemicelluloses. Hemicelluloses include the polysaccharides xyloglucan, glucuronoxylan, arabinoxylan, glucomannan, galactomannan, and galactoglucomannan. Composition and type of hemicelluloses vary across species and cell wall type.[1]

Pectins

Pectins form highly cross-linked networks that embed the other cell wall polymers. Dynamic to change, the pectin networks can serve as an adhesive while being strong and flexible.[8,9] In some plant organs such as leaves, it can make up to 50% of the primary cell wall.[10] Three major classes of the pectins have been defined: homogalacturonans, rhamnogalacturonan-I and rhamnogalacturonan-II.[11,12]

Homogalacturonan is the most abundant type of pectins. Known as the "smooth" region of pectins, it is composed of unbranched α-(1, 4)-linked galacturonic acids. Decoration of this chain with methyl esters and/or acetyl groups allows dynamic changes to properties to occur.

Rhamnogalacturonan-I is known as the "hairy" region of pectins. It has a backbone composed of repeating units of α-(1, 4)-linked galacturonic acids and α-(1, 2)-linked rhamnoses. This backbone has long arabinan and arabinogalactan branching that have the ability to be interlinked.

Rhamnogalacturonan-II is composed of the same unbranched α-(1, 4)-linked galacturonic acids that create homogalacturonan. However, rhamnogalacturonan-II's backbone is decorated with four types of oligosaccharide side chains with differing degrees of compositional complexity.

Evidence supports that correct extensin self-assembly is key in the organizational assembly of the pectin network, but very little is known about the interaction between extensin and pectin. There is also evidence of covalent extensin-pectin cross-linkages, however inability to extract the assembled matrix nanostructure has prevented characterization.[13,14]

Structural Proteins

Cell wall proteins are important for the structure, function, and defense of cell walls. Hydroxyproline-rich glycoproteins (HRGPs), namely extensins, are the most characterized group of this superfamily.[4,15,16] In Arabidopsis, it is believed there are up to 59 different extensins.[17,18] However, the function of each extensin remains unknown.[4]

Extensins

Extensins are a family of hydroxyproline-rich glycoproteins (HRGP) that self-assemble into a nanoscaffolding network. It is believed that the scaffolding organizes other cell wall polymer during wall assembly.[5] Assembly of the scaffold is dependent on extensive post-translational modifications occurring on their highly conserved, repetitive peptide motifs (Fig.1.1).[5] The repetitive motifs are

composed of stretches of alternating hydrophilic and hydrophobic regions. Charged amino acids, in particular Lysine or Histidine, are regularly placed within each repeat. These amino acids give extensins a basic pI, and an evenly distributed positive net charge in the physiological pH in the cell wall.[19,20] Two main peptide motifs are characteristic of "classical extensins".[4]

EXT1:HYSOOOVYKSOOOOVKHYSOOOVYKSOOOOVKYYSOOOVYKSOOOVYKSOOOOVKHYSOOOVYK
EXT3:HYEYKHYSOOOVYHSOOOOKKHYVYKSOOOOVKHYSOOOVYHSOOOOKKHYVYKSOOOOVK
EXT8:SOOOOYYYHSOOOOVKSOOOOYYYHSOOOOVKSOOOOYLYSSOOOOVKSOOOOVYIYASOOOOTHY
EXT10:VYYYKSOOOOYVYSSOOOOYYSOSOKVHYKSOOOOYYAPTPKVHYKSOOOOYVYSSOOOOYYSOSOK

Figure 1.1: Examples of Repetitive Sequences Found in the Extensins. Shown above are segments of amino acid sequences taken from the four extensins identified in this study. This is to demonstrate the repetitive nature of the extensin sequences and motifs. Potential cross-linkage sites for the tyrosines are noted in red. The green Ks are to denote the regularly occurring tryptic cleavage sites.

The two major peptide motifs are:

Hydroxyproline-Rich Glycosylation Motif

X-$(Hyp)_n$: [where X is usually a serine, and n is 3 to 6 hydroxyprolines (Hyp)][13,21] Blocks of this highly glycosylated motif are found in the plant kingdom from higher plants down to green algae.[17,18] About 60% of extensins' molecular weight is due to a high level of glycosylation at these motifs.[4] Proline residues are post-translationally modified to trans-4-hydroxyprolines by prolyl 4-hydroxylases.[20] Hydroxyprolines are then O-glycosylated with up to five arabinofuranoses by a

family of glycosyltransferases.[22] Additional O-glycosylation of serines with a single galactose monosaccharides also occurs.[23] O-arabinosylation of the hydroxyprolines is not dependent on the presence of the presence of Ser-O-galactosylation. However, both types of glycosylation are important for protein functionality and structure.[23] Glycosylation of this motif is well documented, and occurs in an amino acid sequence dependent manner following the Hyp contiguity hypothesis, or the O-Hyp glycosylation code.[24,25]

Loss of function mutants in genes encoding the glycosyltransferases responsible for both types of O-glycosylation (Hyp-O-arabinosylation and Ser-O-galactosylation) in extensins support that correct glycosylation is essential for both vegetative and reproductive development in Arabidopsis.[22,26,27] It is believed this is due to structural differences that occur as a result of the lack of these sugars. Further evidence supports the high degree of glycosylation stabilizes a polyproline II left-handed helix that is integral/fundamental to self-assembly.[12,25,28,29] Correct Hyp-O-arabinosylation is also important for cross-linkage formation through creation of a bend around the tyrosine-rich cross-linking motifs.[23]

Tyrosine-Rich Cross-linking Motif

Extensins are secreted as soluble proteins which become insolubilized into the plant cell wall though formation of intermolecular cross-linkages on adjacent monomers.[30] Extensin scaffold formation is believed to be achieved through peroxidase formed tyrosine-tyrosine derivatives that cross-link its repetitive hydrophobic motifs.[31-34] Additional evidence from previous studies demonstrated that deglycosylation does not solubilize the extensin scaffold, and suggests a non-polysaccharide linkage.[13,35,36]

Tyrosine residues in the hydrophobic motifs are involved in both intramolecular and intermolecular cross-linking of the extensin scaffold (Fig. 1.2). The intramolecular linkage, isodityrosine, occurs at the highly conserved motif, “yXy” (where y is a cross-linked tyrosine and

X can be any amino acid).[31] Formation of this diphenyl ether linkage occurs when the two tyrosine residues are oxidatively crosslinked by a class III peroxidase and H_2O_2.[33,37] Intramolecular isodityrosine cross-links cause "kinks" in the rod shaped extensin monomers without disrupting the polyproline II helical structure.[38]

Three different types of intermolecular cross-linkages at the tyrosine-containing motifs have also been documented to occur.[34] These intermolecular cross-linkages give rise to the extensin nanoscaffold network.[23,32,39] The first two of these linkages that are formed involve IDT. Once intramolecular IDT is formed, it acts as a stabilizer to locally rigidify the yXy motif.[30] This resulting IDT may then link to a tyrosine or IDT in another extensin molecule to form pulcherosine or di-isodityrosine (Di-IDT) (Fig.1.2). These two different types of linkages are thought to be dictated by monomer sequences, and by different configurations of their alignment during scaffold assembly. Meaning, if two extensin monomers align in a staggered configuration, the pulcherosine cross-linkage is favored. Likewise, if they align in a parallel configuration where two IDTs can align, Di-IDT is favored.

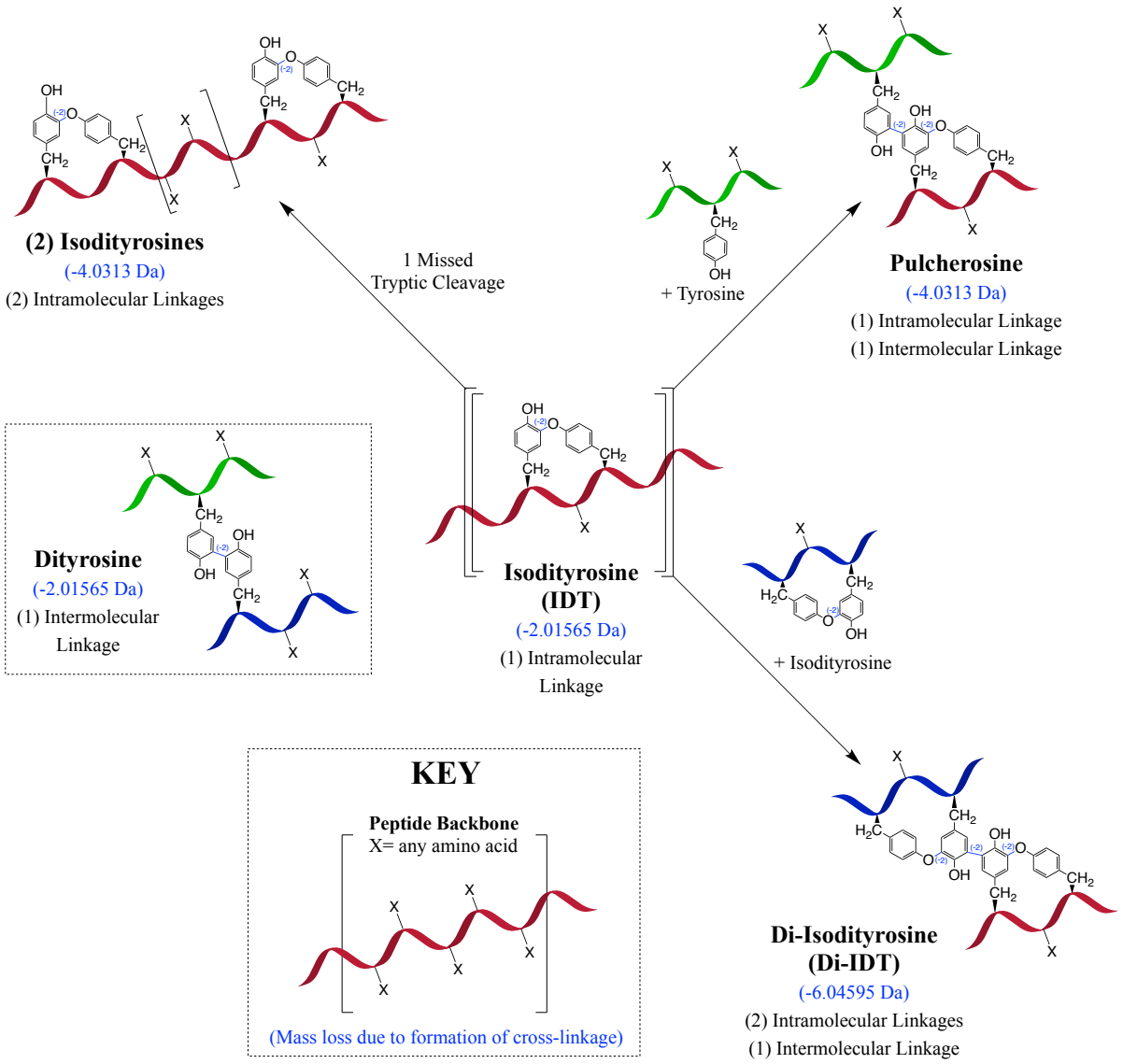

Figure 1.2: Tyrosine Cross-linkages Occurring in the Extensin Scaffold: Schematic depiction of the four types of cross-linkages suspected to arise from peroxidase activity. Note: It is unknown if the intermolecular "dityrosine" (-2.01565 Da) linkage drawn is through dityrosine or is from an intermolecular IDT. However, to date no evidence of intermolecular IDT has been found.[34]

Each repetitive motif component, either singly, or in combination has the information encoded within its structure necessary to self-assemble into a functional nanostructure.[20] In vitro studies involving the cross-linking of extensin-extensin monomers into their nanoscaffold network have made visualization through atomic force microscopy (AFM) possible.[5] However, due to the insolubility of the extensins after network formation, how wall properties arise from correct self-assembly remains a mystery.

Leucine-Rich Repeat Extensins (LRXs)

LRX extensins are related family members with the hydroxyproline-rich extensins. Similar to the HRGP extensins, LRXs are also important for cell wall development and pathogen defense.[40] However, the exact function of each of the different LRX proteins is yet to be determined.[41]

LRXs are chimeric extensins. The C-termini of the LRX proteins contain classical extensin motifs as described above. These motifs are thought to provide anchoring functions similar to the HRGP extensins. The N-termini of the LRXs contain leucine-rich repetitive (LRR) domains. These domains are thought to have roles in protein-protein interactions.[42-44] However, direct evidence of interaction though cross-link formation with the extensin scaffold has not previously been obtained.

CHAPTER II

SEQUENTIAL DEGRADATION OF THE ARABIDOPSIS CELL WALL FOR ISOLATION OF THE EXTENSIN SCAFFOLD

Introduction

The plant cell wall is composed of cross-linked networks mainly of cellulose, hemicellulose, pectins, and scaffold network of glycoproteins, extensins. Isolation of the extensin scaffold has proved challenging due to the high degree of cross-links between and within these networks.[3,4]

The workflow required for isolation and partial characterization of the extensin scaffold from cotton suspension cultures was previously established in methods developed by Qi et al.[13] To achieve soluble fragments (tryptic peptides) of the extensin scaffold from Arabidopsis suspension cultures, we adapted the workflow as described below. The goal of this chapter is to obtain solubilized extensin scaffold fragments from Arabidopsis tissue cultures using the Qi workflow. Changes in polysaccharide composition of the Arabidopsis cell walls as a result of these treatments will be reported and discussed. Subsequent analysis of the cross-linked extensin tryptic peptides occurred using methodologies described in Chapter III.

Methods

Preparation of Cell Walls from Arabidopsis Suspension Cultures

Source of Cell Walls

Arabidopsis suspension cell walls were prepared from suspension cultured Arabidopsis cells (Arabidopsis thaliana T87 originally from ecotype Columbia).

Growth of Arabidopsis Suspension Cultured Cells

Suspension cultures were started by adding 3-4 g of callus from NT-1 phytagel medium to 250 ml culture flasks containing 75 ml of NT-1 liquid medium. Cultures were grown for two weeks at room temperature with 120 rpm rotary shaking and constant light.

Collection of Arabidopsis Suspension Cultured Cells

Cells were collected after two weeks (late logarithmic phase) using vacuum filtration with a 20 micron Sefar Nitex 03-37/24 nylon mesh. Cell walls were prepared similarly to the method described by Komalavilas and Mort [45]. Cells were suspended in nanopure water and broken using a Poyltron Homogenizer on ice for 16 minutes at top speed with 1 minute breaks every 4 minutes to prevent heating of cells. Breakage of cells was monitored under the microscope for complete disruption. The homogenized walls were then washed and vacuum filtered as described above using 5 volumes of 500 mM potassium phosphate buffer pH=7.0, 10 volumes of nanopure water, and 7 volumes of 1:1 chloroform-methanol. After the chloroform-methanol all residual green color was removed from cell walls. The cell walls were then rinsed with acetone to remove the remaining water and air dried overnight. Dried cell walls were white and fluffy upon collection the next day.

Cell Wall Compositional Analysis

Determination of Carbohydrate Composition Using Gas–Liquid Chromatography (GLC)

Sugar compositions were determined by GLC analysis of trimethylsilyl methyl glycosides. Methanolysis and derivatization were performed using the protocol detailed by Komalavilas and Mort. [45]

Cell wall samples (100 μg dry weight) and/or fractions were placed in screw-cap glass vials fitted with Teflon-lined lids containing 100 nmoles of myo-inositol as an internal standard. After speed vacuuming the vial contents to dryness, 200 μl of MeOH- 1.5 M HCL and 50 μl of methyl acetate was added to the vial. The Teflon-lined lids were tightly fastened and the samples were placed in an 80°C heating block for at least 3 hours. Samples were then allowed to cool down to room temperature, and 4 drops of t-butanol were added. Samples were then dried to complete dryness under a stream of nitrogen with 40°C heat. Derivatization with 50 μl of freshly prepared trimethylsiylating reagent ((5ml pyridine, 1ml chlorotrimethylsilane, 1ml hexamethylsilazane) was then added to each dried sample for derivatization. Derivatized samples were then evaporated under a stream of nitrogen until just dry, and were redissolved in 50 μl of isooctane. Aliquots (2 μl) of the trimethyl sugar derivatives with 1 μl of isooctane were separated using a Durabond-I liquid phase (J&W Scientific Inc.) fused silica capillary column (30m x 0.25-mm i.d., installed in a Varian 3300 gas chromatograph.

Amount of each sugar in sample= (area of sugar peak in sample/ area of inositol peak in sample) / (area of sugar peak in standards/ area of inositol in standards) X 100=> # of mmoles in sample.

Preparation of Recombinant Enzymes from *Pichia pastoris*

Source

Recombinant Endopolygalacturonase (AN8327.2) (EPG) from *A. nidulans* expressed in *P. pastoris* was cloned as described previously. [46] Supernatant of the expressed EPG was then centrifuged at 3000 rpm, collected, aliquoted, and stored at -80°C for further use.

Purification of Pectinase Enzyme EPG

Aliquots of EPG (1.5 ml) were pumped through a TSK-GEL Toyopearl DEAE-650 S Anion exchange column purchased from Supleco (Bellefonte, PA, USA) pretreated with running buffer (50 mM Tris buffer: pH 7.0) at 5 ml/min. The adherent EPG was eluted from the column with an increasing gradient of elution buffer (50 mM Tris buffer containing 1M NaCl: pH 7.0). Fractions were assayed with methods described below, and fractions containing highest activity were pooled and buffer exchanged with 20 mM sodium acetate pH 4.8.

Determination of EPG Enzymatic Activity

EPG concentration was determined to be 342 ug/ml using the Bradford Protein Assay Kit (BioRad) using bovine serum album as the reference standard at 596 nm. Activity was assayed using DNS reagent to detect the formation of the reducing end product galacturonic acid, with galacturonic acid used as a reference standard according to methods of Miller. [47]

The reaction mixture contained 1% pectic acid with 1mM $CaCl_2$ in 20 mM sodium acetate buffer (pH 4.8). Enzymatic digestions were carried out at 48°C with 200 rpm shaking. Reactions were terminated by the addition of DNS reagent followed by heating at 100°C for 5 minutes. Absorbance readings were taken in 96-well flat bottom polystyrene plates (Corning New York, USA) at 575 nm using the Infinite® 200 series plate reader (Tecan, San Jose, CA, USA). One unit of EPG activity is defined as the amount of enzyme liberating one umol of product per

minute per milliliter (U/ml). Assays were performed in triplicate with absorbance readings of just substrate and just enzyme controls subtracted from reaction absorbance.

Sequential Methods for Degradation on Arabidopsis Cell Wall

The following methods were adapted from the work of Qi et al. [13]

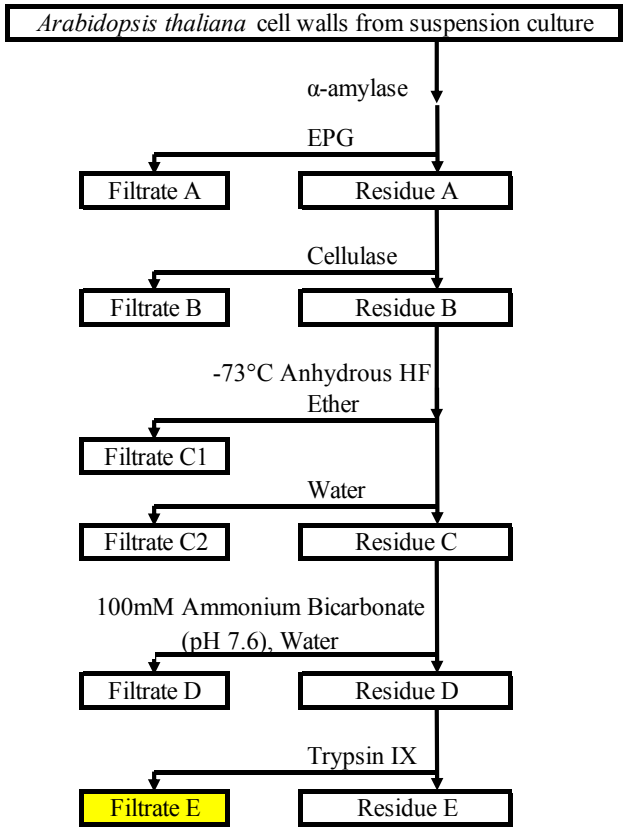

Figure 2.1: Flow Chart of Sequential Treatments for Solubilization of Cross-Linked Extensin Peptides from Arabidopsis Tissue Cultures: Arabidopsis cell walls were subjected to sequential degradation of their polysaccharide networks using enzymatic and chemical techniques to isolate cross-linked fragments of the extensin scaffold. Digestions of the polysaccharide networks in the cell wall included: α-amylase for digestion of starches; EPG for digestion of pectins; cellulase for digestion of cellulose and hemicelluloses. Chemical deglycosylation of extensin's arabinofuranoses occurred at -73 °C using anhydrous hydrogen fluoride (HF). Subsequent washes with ammonium bicarbonate buffer were to disrupt ionically associated species. Trypsin was used to solubilize fragments of the cross-linked extensin network (Filtrate E).

Enzymatic Digestions of the Cell Walls

Starch Removal (α-Amylase) Pre-Treatment

Before proceeding with the subsequent degradative methods of Qi et al, 20 g of dried cell walls were subjected to residual starch removal using α-amylase (*Aspergillus oryzae*, Sigma). The cell walls (10mg/ml) were suspended in 100mM potassium phosphate buffer pH 6.0 and placed in vacuum oven for 20 minutes to hydrate. α-Amylase (2 g) at a final concentration of 30 U/ml (3 U/1mg cell wall) was added. One unit corresponds to the amount of enzyme to liberate 1 μmol of maltose per minute at 25°C. The suspension was stirred for 24 hours at 25°C. The enzyme treated cell walls were collected using vacuum filtration with a 37 micron Sefar Nitex 03-37/24 nylon mesh, and washed with water.

Endopolygalacturonase Pectinase Treatment: (Residue/Filtrate A)

Cell walls underwent enzymatic digestion of the pectin backbone using endopolygalacturonase **(EPG)** (AN8327.2, P. *pastoris*).[46] The cell walls were suspended in 50 mM sodium acetate buffer pH 4.8 at 10 mg/ml PCW (2000ml for a final 10 mg/ml). EPG (55.5 ml) at 3 U/ml were added to the mixture. One unit corresponds to the amount of enzyme to liberate 1 μmol of galacturonic acid per minute at 48°C. A couple of drops of toluene were added, and the sample was placed in incubator at 48°C, at 105 RPM, overnight. The resulting insoluble material and solubilized filtrate were named Residue A and Filtrate A.

Worthington Cellulase Cellulose/Hemicellulose Treatment: (Residue/Filtrate B)

Pectinase digested cell walls (Residue A) were washed with water, then collected using vacuum filtration with a 37 micron Sefar Nitex 03-37/24 nylon mesh. The cell walls then underwent enzymatic digestion utilizing the commercial blend Worthington Cellulase. The cellulase blend is known to have enzymatic activity targeting the cellulose and hemicellulose

components of the cell wall. The cell walls were suspended in 150 ml of 50 mM ammonium acetate buffer pH 5.0 at 10 mg/ml cell walls. 2% (w/w) of cellulase was added to the mixture. A couple of drops of toluene were added, and the sample was placed in incubator at 45°C, at 130 RPM, overnight. After treatment, cell walls were collected using vacuum filtration with a 37 micron Sefar Nitex 03-37/24 nylon mesh, washed with several volumes of buffer, then water, and freeze dried. The resulting insoluble material and solubilized filtrate were named Residue B and Filtrate B.

Selective Hydrogen Fluoride Solvolysis at Different Temperatures

-73°C Anhydrous Hydrogen Fluoride (HF) Treatment (Residue/Filtrate C1)

Residue B (2100 mg dry wt) was placed in custom made Teflon vessel. A lid was placed on the vessel, and the closed vessel was attached to the HF apparatus containment system as previously described.[48] The apparatus is shown in Figure 2.2.[45,49]

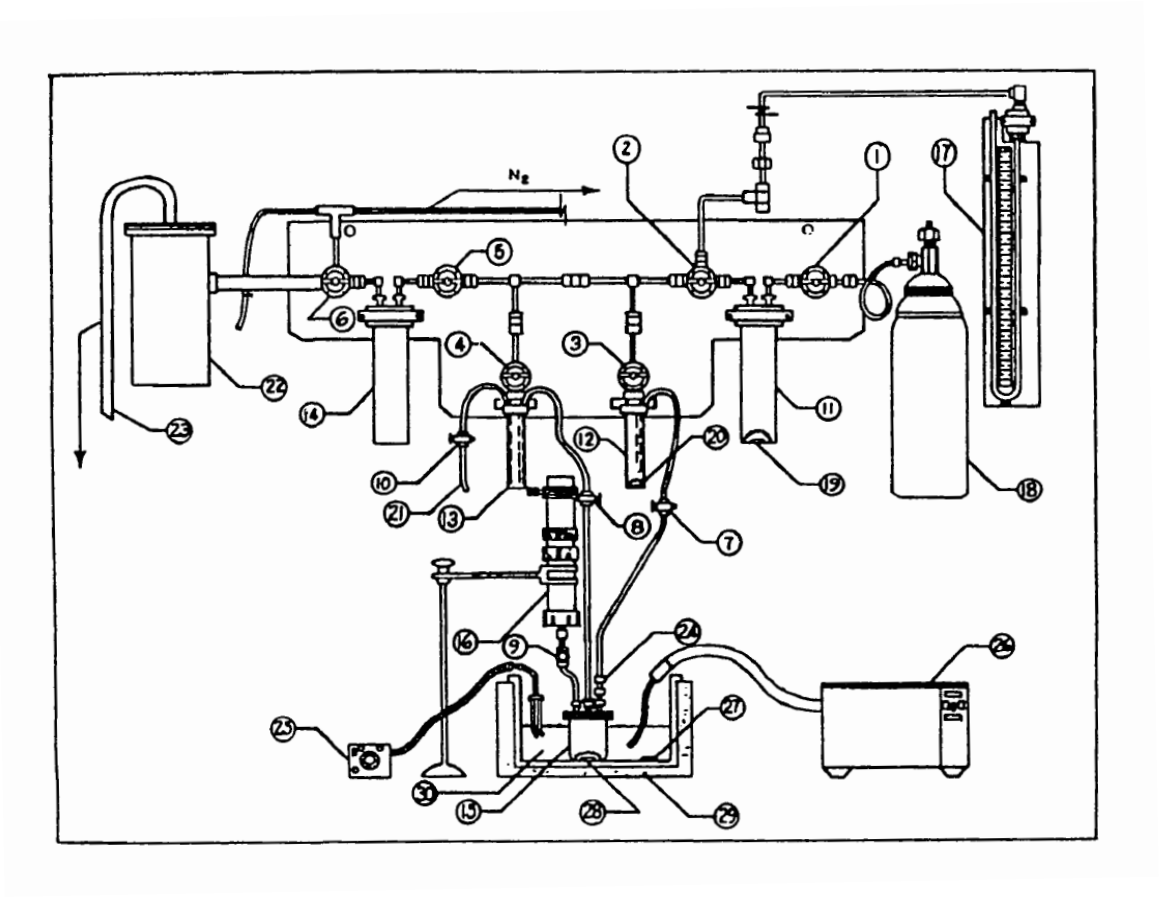

Figure 2.2: Schematic Representation of the Hydrogen Fluoride Solvolysis Apparatus: 1-8, 10, stopcocks; 9, teflon needle valve; 11-16, Teflon and Teflon vessels; 17, manometer; 18, hydrogen fluoride tank; 19, 20, stirrer bars; 21, exit to the sink for pressure release, if necessary; 22, calcium oxide trap; 23, connection to the vacuum pump; 24, 3 mm to 6 mm adaptor; 25, heater/regulator; 26, immersion cooler; 27, 28, stirrer bars; 29, insulated container; 30, 95% ethanol.[49]

Anhydrous HF (20 ml) was allowed to enter the enclosed vessel containing the sample and the reaction proceeded at -73°C for 30 min as previously described.[36] After 30 min, the vessel was placed into a liquid nitrogen bath to stop the reaction. After allowing the reaction to freeze, 100 ml of diethyl ether was added. The vessel was then allowed to reach room temperature, and dried by evaporation under vacuum in the fume hood. The dried sample was transferred to a weighed storage container using 0.1 M ammonium bicarbonate buffer (ABC) (pH=7.6), shell frozen with liquid nitrogen, and freeze dried overnight.

Subsequent Washes of Cell Wall Residues

Water (Residue/Filtrate C2)

Due to excessive amounts of salt in the resulting C1 sample after -73°C HF treatment, the residue and filtrate were then diafiltered using a 200 ml Millipore stirred cell using an ultrafiltration membrane with a 10 kDa cutoff. Water was added 4 times to remove excess salt. The remaining sample was then freeze-dried overnight. Resulting material was named Residue/Filtrate C2, respectively.

Ammonium Bicarbonate Buffer (0.1M) (Residue/Filtrate D)

Residue C2 was placed in 100 mM ammonium bicarbonate buffer pH 7.6 overnight at 4°C. This was done to ensure removal of any residual ionically bound species, and to place the cell walls into a suitable buffer for trypsinization.

Proteolytic Digestions of the Extensin Scaffold

Trypsin IX Proteolysis: (Residue/Filtrate E)

Remnants of cell walls (Residue D) were subjected to overnight proteolysis at 26°C using Trypsin IX (Sigma) (2% w/w) for cleavage of the extensin nanoscaffold. The resulting insoluble

material and solubilized filtrate were named Residue E and Filtrate E. Characterization of the cross-linked extensin peptides are discussed in Chapter III of this dissertation.

Results and Discussion

Cell Wall Compositional Analysis

Sample Analyzed (sample + treatment)	mol % Sugar Composition[a]								%Wt fraction[b]
	Ara	Rha	Fuc	Xyl	Gal A	Man	Gal	Glc	
Starting Cell Wall (SCW) (homogenized)	4.5	1.5	0.3	3.0	73.3	1.3	8.6	7.4	100
Filtrate A (pretreated SCW + amylase + EPG)	11.5	4.5	0.4	2.6	63.4	4.6	10.5	2.5	79.9
Filtrate B (Residue A+ Cellulase)	12.7	5.3	1.3	10.4	26.6	2.9	13.1	27.7	9.6
Filtrate C1 (Residue B+ -73°C HF/Ether)	73.7	8.1	2.0	6.4	2.5	0.3	3.4	3.6	1.2
Filtrate C2 (Residue C1+H_2O)	11.4	1.0	0.6	6.7	12.8	3.2	10.3	53.9	1.3
Filtrate D (Residue C2+ABC)	2.9	9.2	0.2	4.9	49.2	Trace	17.1	16.4	1.4
Filtrate E (Residue D + Trypsin IX)	14.6	9.9	2.4	4.8	46.2	5.2	26.0	17.5	5.6
Residue E (Residue D+Trypsin IX)	6.4	5.6	0.2	3.4	28.8	3.0	18.0	34.7	1.7

Table 2.1: Mole Percent of Sugar Composition of Arabidopsis Cell Walls and in the Subsequent Degradative Fractions After Each Enzymatic and Chemical Treatment Outlined in Figure 2.1: Filtrate is the solubilized fraction recovered after each treatment; Residue is the insoluble fraction recovered and used for next step of treatment. a: mol% was calculated from the moles of sugar detected by GLC divided by the moles of all sugars present in

the sample, multiplied by 100.; b: The percent fraction of dry weight remaining of the original wall present.

Enzymatic Digestions of the Arabidopsis Cell Walls Removed 89.5% Total of Polysaccharides by Weight

Removal of the polysaccharide components of the cell wall is required to access the extensin scaffold. To disrupt and remove the major polysaccharide networks in the Arabidopsis cell walls, we sequentially targeted each network using enzymes. Networks of the pectin polysaccharides fill in the plant cell wall and largely dictate pore size, or accessibility to all other components of the cell wall. To digest the pectin backbone, we used EPG.[46] EPG is known to randomly cleave chains of galacturonic acids that compose the backbone of homogalacturonan and to solubilize fragments of rhamnogalacturonan II.[50] Treatment with EPG resulted in a 79.9% decrease in weight from original walls present (Fig. 2.1, Table 2.1). The major sugar component resulting from EPG digestion was Galacturonic Acid (GalA). This is agrees with the anticipated release of the pectins homogalacturonan and rhamnogalacturonan-II. After this digestion we expect the majority of remaining cell wall pectin will be rhamnogalacturonan-I and xylogalacturonan.

To increase our access to the extensin scaffold, we used a commercial enzymatic blend of Cellulases to target and degrade the networks of cellulose and hemicellulose. Treatment with Cellulase resulted in an additional 9.6% decrease in weight from original walls present. The major sugar component resulting from this digestion was glucose, followed by galacturonic acid (Fig. 2.1, Table 2.1). Presence of these sugars are consistent with the removal of celluloses and hemicelluloses, in addition to remaining pectins. After both enzymatic digestions of the pectin

and cellulose/hemicellulose networks, the remaining residue's (Residue B) dry weight was 10.5% of the original starting cell wall.

Minus 73°C Hydrogen Fluoride Deglycosylation Removes O-Arabinose Glycosylations (Residue/Filtrate C1)

The Hyp-continuity hypothesis describes the arabinosylation of the contiguous hydroxyproline residues that lock extensin's repetitive hydrophilic motifs into an extended polyproline II helical conformation.[24,51] The high degree of glycosylation also serves a defensive role by allowing extensins to be proteolytic resistant. To remove these arabinofuranoses, cell wall resides (Residue B) were subjected to 30 minutes of mild HF treatment at -73°C using methods as described in Qi et al.[13] The -73°C HF treatment resulted in an additional 2.5% decrease in weight from original walls present. The major sugar component resulting from this deglycosylation was arabinose (Fig. 2.1, Table 2.1). These results support our expectations that -73°C HF preferentially removes arabinofuranoses.

The Extensin Scaffold is Solubilized Using Trypsin (Residue/Filtrate E)

Extensin scaffold formation is believed to be achieved through peroxidase formed tyrosine-tyrosine derivatives that cross-link the repetitive hydrophobic motifs.[31-34] Additional evidence from previous studies demonstrate that deglycosylation does not solubilize the extensin scaffold.[13,35]

To solubilize the scaffold into cross-linked peptides, we used trypsin to digest the peptides at their regularly occurring lysine residues (Fig. 1.1). The tryptic digest created the solubilized fraction, Filtrate E. Filtrate E accounted for about 5.6% by weight of the original walls. Interestingly, GLC carbohydrate compositional analysis of the tryptic digest (Filtrate E)

resulted in the identification of a sugar composition expected for a region of pectin, designated as Rhamnogalacturonan-I (Fig. 2.1, Table 2.1).[45,52]

A small amount of the Arabidopsis cell walls remained insoluble after tryptic digestion of the extensin scaffold (Residue E). This insoluble material was a shade of dark brown. The insoluble residue E accounted for about 1.7% by weight of the starting cell wall material. The major sugar component of this material was glucose, followed by galacturonic acid (Fig. 2.1, Table 2.1). These results support our expectations that other cross-linkages are present in the cell wall. Speculation as to possible compositions of these cross-linkages could be of a polysaccharide-phenolic, or polysaccharide-phenolic-protein nature.[13,53]

Conclusion

In this study, we demonstrated that after targeted removal of the cell wall polysaccharides, proteolytic digestion results in soluble fraction rich in sugars expected for a region of pectin, Rhamnogalacturonan-I. This suggests a covalent linkage between rhamnogalacturonan-I and extensins is present. Evidence of a covalent linkage between rhamnogalacturonan I and extensins is consistent with previous results found with cotton cell walls.[13] More work is needed to fully characterize this linkage.

The same methodology utilized in this chapter was previously used to isolate soluble peptides containing extensin-extensin cross-linkages in cotton cell walls. Due to the extensive degrative approach utilized, we anticipate that the peptides obtained through tryptic digestion are due to the presence of covalent cross-linkage in the extensin scaffold network. Characterization and identification of these cross-linked peptides are in chapter III.

CHAPTER III

MASS SPECTROMETRIC CHARACTERIZATION OF EXTENSIN CROSS-LINKAGES IN ARABIDOPIS TISSUE CULTURE CELL WALLS

Introduction

Plants are self-assembling sky scrapers of stacked plant cells encompassed in cell walls. Plant cell walls are cross-linked networks of complex polysaccharides organized by a scaffold of structural glycoproteins, extensins.[5,13,14,16,5,15,54-56] However, direct characterization of the assembled extensin scaffold has proven to be quite challenging.

Extensins are defined by their highly periodic hydrophilic and hydrophobic motifs. Hydrophilic motifs in extensins consist of short, rigid blocks of contiguously glycosylated hydroxyprolines [Ser-$(Hyp)_n$ (where n is 4 to 6 hydroxyprolines)].[13,57,24,51] These are interspaced with hydrophobic tyrosine-rich motifs. Self-assembly of the extensin scaffold is believed to be achieved though molecular alignment of soluble extensin monomers subsequently locked into place though peroxidase mediated intermolecular and intramolecular tyrosine cross-links.[31,33]

Much of what is known about the tyrosine cross-linkages has been achieved through studies of resulting tyrosine derivatives obtained from cell wall hydrolysates, or through *in vitro* characterization of cross-linked extensin precursors.[2,33,58,5,32,34,59,60,61] Two cross-linkage motifs are suspected to occur: (i**) yXy**, (where y is cross-linked tyrosine, and X is any amino acid) forms

intramolecular isodityrosine **(IDT)**, a dimer of two tyrosines. Two IDT motifs from neighboring monomers may then cross-link to form intermolecular di-isodityrosine **(Di-IDT),** a tetramer of four tyrosines. (ii) **VyK,** an additional intermolecular linkage point attaching to IDT to yield a trimer of three tyrosines, pulcherosine (Fig 1.2).[31-34]

Due to the insolubility of the extensin scaffold, characterization of the cross-linkages in vivo has remained a challenge. However, development of a workflow to obtain solubilized extensin scaffold fragments along with recent advances in cross-link mass spectrometry have made it possible to finally characterize this scaffold in vivo.[13] This study's aim is to characterize extensin cross-linked peptides in Arabidopsis. To do this, we first extracted cross-linked extensin scaffold fragments. We then used liquid chromatography mass-spectrometry (LC-MS/MS) to analyze these fragments. Finally, we established a data analysis workflow using Byos (Protein Metrics) to interpret the complex cross-linked product ion spectra.[62] We then used this workflow to determine which extensins are cross-linked and how these cross-links are arranged in the protein scaffold. Our data reported here, indicates that the extensin scaffold is built through four different types of tyrosine cross-linkages. These linkages are formed largely through sequence dependent manners occurring in at least four extensins. Required considerations for mass spectrometry data analysis of the cross-linkages in extensins is also discussed.

Methods

Sample Preparation: Solubilization of Extensin Scaffold from Cell Walls

An overview of this workflow is depicted in (Fig. 1.2). Arabidopsis cell walls were prepared from suspension cultured cells of *Arabidopsis thaliana* T87 ecotype Columbia as described by Komalavilas and Mort.[45] Dried cell walls (20 g) were treated overnight with α-amylase (Aspergillus oryzae, Sigma) to remove residual starch residue at (3U/1mg cell wall). After this pretreatment, the cell walls were subjected to subsequent degradative methods adapted from the work of Qi et al.[13] Changes to the methods of Qi et al, include endopolygalacturonase (EPG) (AN8327.2, P. pastoris) [46]; and a wash of the cell walls with 8 M urea prior to proteolytic treatment with Trypsin IX (Sigma), yielding a protein rich solubilized fraction. Ninety mg of this fraction was then subjected to complete deglycosylation using 0°C anhydrous Hydrogen Fluoride (HF) as previously described, [36,48] and freeze dried overnight.

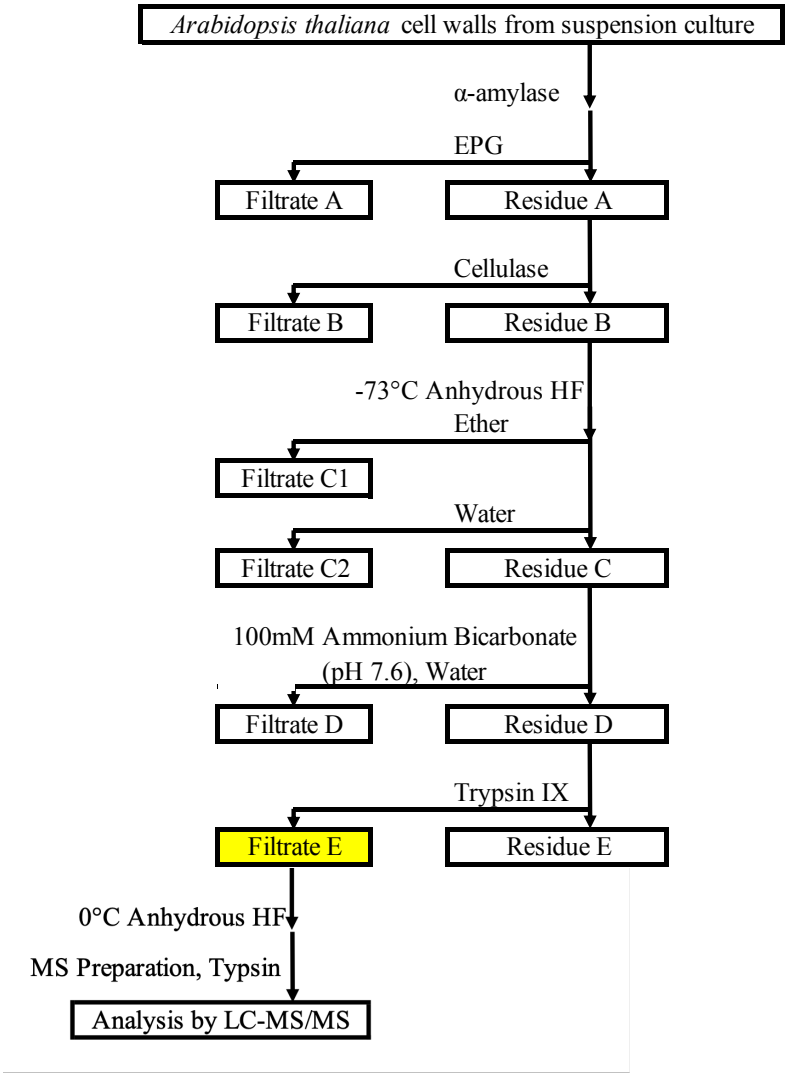

Figure 3.1: Flow Chart of Preparation of Solubilized Cross-Linked Extensin Peptides from Arabidopsis Tissue Cultures for LC-MS/MS Analysis: Figure 3.1 is a continuation of the workflow depicted in Figure 2.1. Solubilized cross-linked extensin peptides (Filtrate E) were subjected to a second full chemical deglycosylation at 0°C using anhydrous hydrogen fluoride (HF) to remove hexose sugars as described in Chapter II.[36] The resulting sample was then washed using 0.1 M ammonium bicarbonate buffer (ABC) (pH=7.6), and freeze dried. Preparation for mass spectrometry included dissolving the material into buffered urea reduction solution (BURS), followed by alkylation, and a second tryptic digestion.

Preparation of Samples for Mass Spectrometry

HF deglycosylated peptides (100 μg dry weight) were dissolved in freshly prepared Buffered Urea Reducing Solution (BURS) containing 8 M urea/100 mM Tris Buffer (pH=8.5), reduced with 2.5 mM TCEP, 30 min at room temperature, and alkylated with 5 mM iodoacetamide, 20 min at room temperature in the dark. Samples were diluted with 0.8 M Tris Buffer (pH=8.5) and water for a 2 M final urea concentration. Trypsin/Lys-C (Promega) was added at a 4 μg/ml final concentration, and the mixture was incubated for 18h at 37°C in an incubator with shaking at 120 rpm. Digestion was quenched with 100% TFA for a final concentration of 1% TFA to reaction volume. After acidification with TFA, samples were desalted using Pierce C18 tips following the manufacturer's suggested methods. Resulting sample was then dissolved into 50 μl of Mass Spectrometry grade water (Burdick & Jackson) containing 0.1% formic acid for an anticipated concentration of 1 μg/μl.

NanoLC-Mass Spectrometric Analysis

Resulting desalted peptides were analyzed with an Orbitrap Fusion Tribrid Mass Spectrometer (Thermo Scientific) coupled to a Proceon nano-electrospray ion source, and an Easy-nLC 1200 system (Thermo Scientific) in a vented trap configuration. Ten μl of peptides were loaded and separated using a 75 μm x 2 cm Acclaim PepMap 100 (Thermo Scientific 164943) trap column and a 75 μm ID x 50 cm Acclaim PepMap RSLC column (Thermo Scientific 164942). A binary solvent system composed of Solvent A: water containing 0.1% formic acid, and Solvent B: 80% ACN in water containing 0.1% formic acid was used for all analyses. All solvents were Burdick & Jackson of mass spectrometry grade. The 173 min gradient at 250 nl/min was as follows: 0% to 30% solvent B in 120 minutes, an increase to 60% B over 18.5 minutes, followed by an immediate increase to 100% B. 100% B was then held for 18.5

minutes, followed by a return to initial conditions for 12 minutes. Eluted peptides were sprayed directly into the mass spectrometer and analyzed using the following methods.

Data Dependent Acquisition Methods

Cross-Link Mass Spectrometry

In each Top 20 or Top 40 acquisition cycle, a survey full scan MS1 (from m/z 375 to 1800) was acquired in the Orbitrap with a resolution of 120,000, (AGC 4e5, max injection time 50ms). The intensity threshold was set to 5e4. A mass exclusion list created from a blank control run was used to exclude background ions of noninterest. MS2 acquisition of precursors with charge states from 2+ to 9+ were selected for fragmentation with selection priority set to the highest charge states in order of intensity. Selected ions were isolated using the quadrupole with an isolation width of 1.6m/z. Each selected precursor was fragmented using 9 consecutive fragmentation events (Table 3.1) and subsequently excluded from fragmentation for 60 seconds with an exclusion mass tolerance of 10ppm. Detection occurred in the Orbitrap (30,000 resolution, 1e5 AGC, 150ms max injection time). A second experiment utilizing the same settings but with electron-transfer/higher-energy collision dissociation **(EThcD)** was also carried out, but we found EThcD did not increase MS2 identifications (Data not shown).

Fragmentation Method	**Normalized Collision Energy (NCE)%**
HCD	39, 36, 33, 30, 27, 24, 21, 18, 15

Table 3.1: Higher Collisional Dissociation Mass Spectrometry Settings

Data Analysis

Analysis of all spectra was performed using Byos (Preview, Byonic, and Byologic) (Protein Metrics, Cupertino, CA, USA)[62] against the corresponding UniProt Arabidopsis thaliana proteome (retrieved: 04/07/18) and a resulting focused database of 800 proteins.

Identification of Cross-Linked Candidates

A well-documented challenge with mass spectrometry characterization of cross-linked peptides is the "n^2 problem". The n^2 problem states that data complexity increases as a square of the number of peptides and their potential cross-linking sites. To mitigate this challenge, a focused approach for cross-link characterization searches is required.[63,64]

To first identify intramolecular isodityrosine candidates, and to create a focused database of identified proteins for subsequent searches, we ran a Byos search with the following parameters: (i) Trypsin was selected with 1 missed cleavage allowed. (ii) A precursor ion tolerance of 10ppm and a product ion mass tolerance of 10ppm. (iii) Static modifications were: carbamidomethylation of C, and oxidation of P (for hydroxyproline). (iv) Dynamic modifications included: Oxidation of M, (De)oxidation of P, Q or E ->pyro-Glu Ammonia-loss from N-terminal C, and N-terminal Acetylation. (v) Custom modifications were: Didehydro (-2.01565Da) and Quatrodehydro (-4.0313Da) of Y, and (De)Oxidation of P (max 2). (vi) Results were filtered at 1% FDR, a confidence threshold of Byonic score $\geq$150, and further validation was performed manually. A focused database was then created for subsequent cross-link searches.

Identification of Cross-Linked Peptides: (XLINKS search)

To identify and characterize the cross-linked peptides the same search parameters were used as described above with an additional cross-linking parameter: Dityro (-2.015650) at Y. This additional parameter allows us accommodate our search for cross-linked candidates carrying mass losses suspected for extensins. These include: intramolecular (1 peptide) (-2.015650 Da), and intermolecular cross-links connecting 2 peptides with mass losses of (-2.015650 Da, -4.0313 Da and -6.04595 Da). Cross-linked candidates were selected from our "shotgun" search data analysis as described above (Data not shown). A maximum of 1 common modification, and 1 rare

modification for each peptide was allowed. Results were filtered at 1% FDR, a confidence threshold of Byonic score ≥150, and further validation was performed manually.

Manual Validation of Cross-linked Extensin Peptides

Interpretation of the mass spectral results of extensin peptides is challenging due to several features specific to the extensins. The features that must be taken into consideration include:

(1) There is a large family of extensins in Arabidopsis. Their sequences are very repetitive and similar to each other; thus, some peptides are not unique to one family member.

(2) Some tryptic peptides have the same amino acid composition but slightly different sequence. Thus, an exact match of the m/z of the intact parent ion does not guarantee identity. Extensive fragmentation coverage is required for unambiguous identification of peptides. This particularly becomes a challenge in cases of two peptides crosslinked together due fragmentation preferentially occurring on one peptide, while leaving the other peptide intact.[65]

(3) There are multiple possibilities for which tyrosine cross-links are present.

(4) Multiply charged internal fragment ions (fragmentation from both the N- and C- termini of the peptide) are formed from peptides containing Hydroxyproline and Serine which often dominate the resulting spectra. (See Section: "Extensin Fragment Ion Assignment")

(5) Cross-link searches are very time consuming. This is known as the "n^2 problem" described in the XLINKS section above.

To overcome the challenges, each PSM identified in Byos was manually validated taking into consideration several criteria at both the MS1 and MS2 levels. Criteria that were used for evaluation included:

(1) Peptide sequence(s) carrying the mass loss identified for the linkage.

   a. Agreement of the mass loss with the number of tyrosines that were identified to be cross-linked in the peptide sequence(s).[66]

(2) MS1 Parent Ion:

   a. Correct monoisotopic assignment: MS1 isotopic profiles were inspected to insure correct identification of the first isotope as the monoisotopic peak.[67]

   b. An experimental isotopic distribution that appears to be one cross-linked species and is not overly contaminated with ions arising from co-eluting species.

   c. Comparison of the experimental isotopic distribution to the theoretically calculated isotopic profile.[68]

(3) MS2 Spectra:

   a. A rich fragment series of the cross-linked peptides was required for confident identification (as described above).

   b. Presence of large cross-linked ions added significant evidence that cross-linkage was real.[66]

   c. Internal fragment ions formed were taken into consideration, as needed, when they dominated the spectra (See Section: "Extensin Fragment Ion Assignment").[68]

Extensin Fragment Ion Assignment

Close inspection of the HCD data revealed a high number of unassigned fragment ions dominating the spectra. These ions were typically a serine or a hydroxyproline residue mass apart and occurred in several charge states. To assign fragment ion identity, theoretical internal fragment ion masses occurring on both peptide chains were calculated using Protein Prospector's MS-Product.[68] Unassigned ions were matched to theoretically calculated internal fragment ions with a 10ppm mass tolerance. These assignments indicated that simultaneous losses occur at both N- and C-termini of either or both cross-linked peptides. To label the internal fragment ions, we used the Roepstorff and Fohlman nomenclature.[69] Several ions contained both a y-type fragment at their N-terminus, and a b-type fragment at their C-terminus. We denoted these fragments to be "B|Y", where "|" refers to the fact they are internal (Fig.3.2). Assignment of these ions improve sequence coverage, and add additional evidence for location of the cross-linkage site.

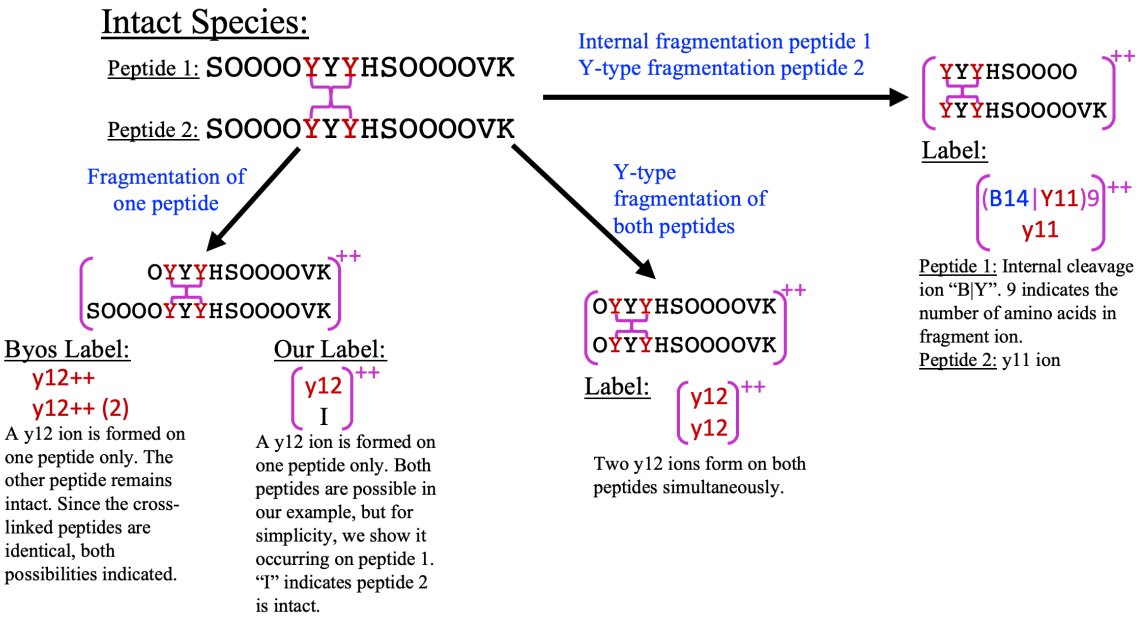

Figure 3.2: MS2 labeling scheme of three types of possible fragments of Di-IDT cross-linked peptides. All fragment ions are in charge state z=2, indicated by ++. Counterclockwise around figure: (Left, top) Intact Di-IDT cross-linked tryptic peptides. (Left, bottom) Fragmentation occurring on one chain of the homodimer: Since this fragmentation is possible on either chain, we created a labeling scheme to accompany the scheme depicted in Byos. Byos labeling indicates this mass loss may occur on either peptide. The equivalent species is shown using our labelling scheme by indicating fragmentation occurs on one peptide chain only. (Right, bottom) y-type fragmentation occurs on both peptide chains, resulting in mass losses in both. (Right, top) Two types of fragmentation occur in species: Peptide 1 is an internal "B|Y" species: (B14|Y11)9. Peptide 2 is standard y-type fragmentation.

Results

The Arabidopsis cell wall is composed of a cross-linked network of at least four hydroxyproline-rich extensins (EXT) and one leucine-rich extensin (LRX).

To characterize the extensin scaffold, we utilized the workflow adopted from Qi et al to enrich soluble cross-linked extensin peptides (Fig. 3.1).[13] Degradation of the cell wall material was monitored throughout the Qi workflow using GLC analysis as described in chapter II. Arabidopsis cell wall carbohydrate composition results were consistent with the previous results found with the cotton cell walls (Table 2.1).[13] Trypsin solubilized peptides obtained from the degraded material were characterized using LC-MS/MS, followed by data analysis in Byos (Protein Metrics) as described in our methods section (Fig. 3.1). Peptide spectrum matches (PSMs) identified four hydroxyproline-rich extensins: EXT1, EXT3, EXT8, EXT10.[17] PSMs corresponding to one leucine-rich extensin, LRX4, were also identified (Table 1S). The extensive degradative approach utilized removed polysaccharide-polysaccharide polysaccharide-protein cross-linkage and ionic associations. Therefore, we anticipate that the presence of these corresponding PSMs are due to a covalent protein-protein cross-linkage into the EXT scaffold network.

Four different types of cross-linkages are formed during extensin scaffold assembly.

To characterize the different cross-linkages suspected to occur in the extensin scaffold, we customized our XLINKS search to identify intra- and intermolecular tyrosine-tyrosine cross-linkages (Fig. 1.2).[5,32,33,41] The XLINKS function considered the possibility of cross-linkages to occur between every tyrosine in our EXT and LRX candidates identified above. Following manual validation, we identified cross-linkages in peptides containing three different mass losses. These mass losses are a result of two, three, and four cross-linked tyrosines.

Sequence analysis of the cross-linked peptides identified four different types of linkages: one intramolecular linkage joining two tyrosines on one peptide: IDT (-2.01565 Da); and three intermolecular linkages joining tyrosines on two peptides resulting in mass losses of -2.01565 Da (dityrosine); -4.0313 Da (pulcherosine); and -6.04595 Da (Di-IDT).

Our findings support that the scaffolding network is cross-linked in a mixture of these four different tyrosine derivatives. Further details of the characterization of these four types of linkages is as follows:

Isodityrosine (IDT) Forms the EXT Scaffold's Intramolecular Linkages.

IDT is a diphenyl ether cross-linkage joining two tyrosines on a single peptide containing the sequence yXy (where y is a crosslinked tyrosine and X is any amino acid). Formation of the IDT bond results in a total mass loss of (-2.01565 Da), or two hydrogens lost in the intramolecular cross-linkage formation between two tyrosines (Fig. 1.2). To determine if an IDT linkage was present in the EXT scaffold, a -2.01565 Da mass difference joining two tyrosines in a single peptide was investigated.

Mass Spectrometry Characterization of IDT Cross-linked Peptides

An example of mass spectrometric evidence used for characterizing the IDT cross-linkage from the major repetitive peptide in EXT8 is given below (Fig. 3.3). The MS1 isotopic profile depicts an intact parent ion at m/z=974.932, z=2 that was isolated for fragmentation. This parent ion mass is -2.01565 Da less than the calculated mass of a peptide containing the same sequence without the cross-link. This supports that an IDT linkage occurs in this species. MS2 fragmentation gives further evidence for the IDT linkage. The MS2 spectra give a rich fragment ladder of a/b- and y-ions. These ions alone give the peptide sequence and localization of the IDT cross-linkage at the yXy motif. The fragment ions containing the mass loss of the IDT cross-linkage are: a/b6 to a/b15 and y11 to y15. Fragment ions that do not contain the yXy motif, do not

have the mass loss for IDT. Additional fragmentation creating the B|Y internal fragments give additional evidence. The B|Y ions occurring above 600 m/z contain the yXy motif, and also carry the mass loss for IDT.

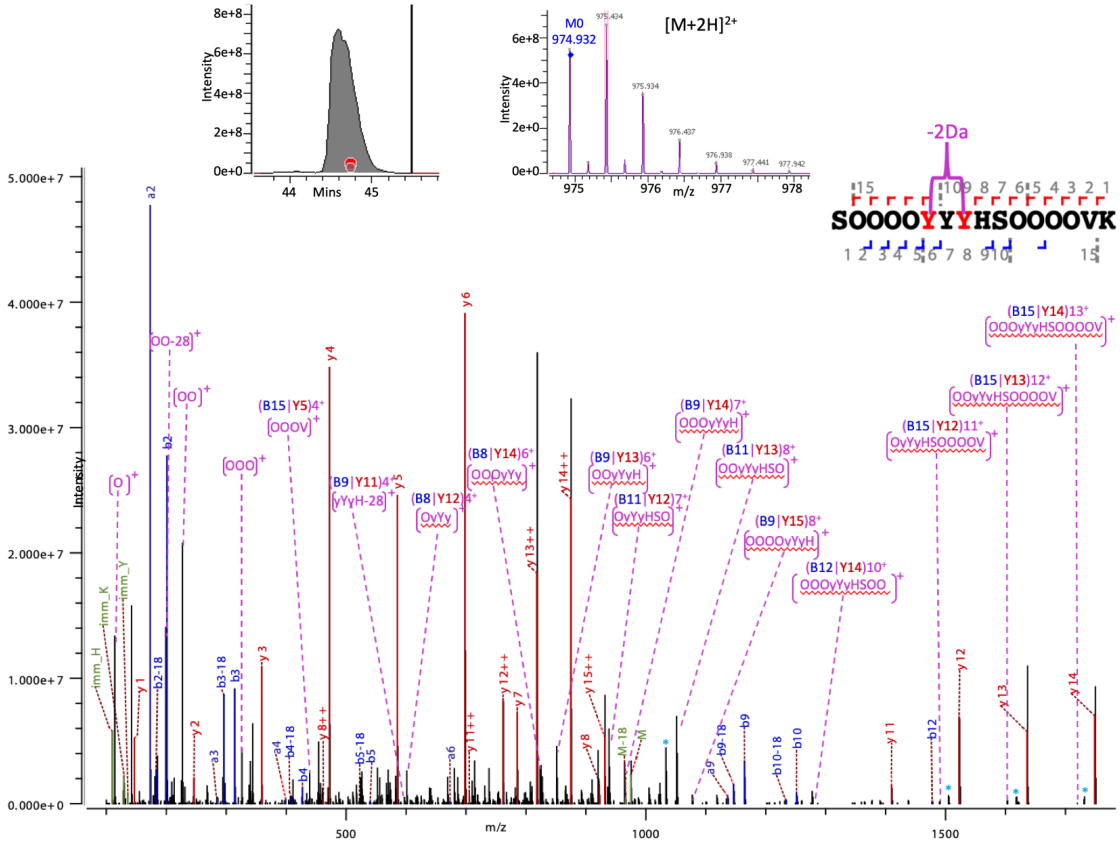

Figure 3.3: MS Characterization of SOOOOyYyHSOOOVK IDT Cross-Linked Peptide: The species above is intramolecularly IDT cross-linked at yYy, (where lowercase "y" indicates a cross-linked tyrosine). (top, left) Extracted ion chromatogram: The red dots indicate the times of several MS2s of this species collected. (top, middle) The isotopic profile of the [M+2H]2+ precursor ion at m/z 974.932 that was isolated for fragmentation. The blue dot indicates correct monoisotopic assignment based on calculated mass. (top right) a-, b-, & y-fragment ion coverage: Cross-linked tyrosines carrying the mass loss of (2.0156 Da) are denoted in red. Positions of y-type ions are denoted by the red bracket above the peptide sequence, a-/b-type ions are denoted in blue below the peptide sequence. (bottom) The annotated HCD MS2 spectrum: Internal fragment

ions are indicated in pink as depicted in (Fig. 3.2) with corresponding sequence below. Light blue asterisk indicates ions due to a loss of water from right neighboring ion. Fragment ions containing the mass loss of the IDT cross-link are a/b6 to a/b15; y11 to y15; and B|Y internal fragments above 600m/z.

Identified IDT Cross-Linked Peptides

Sixteen peptide species containing a mass loss of (-2.01565 Da) at their IDT yXy motifs were identified (Table 3.2). These peptides corresponded to two out of the four hydroxyproline-rich EXTs identified in this study: EXT3 and EXT8.

Fifteen of these identifications corresponded to EXT8 peptides. Of the six possible yXy motifs in EXT8, four were found to form IDT linkages (Table 3.2). Notably, the most repetitive motif, yYy, was identified with many different sites of one or two non-hydroxylated prolines, and with missed tryptic cleavages. Out of the five potential sites for IDT cross-linking in EXT3, only was identified to have the IDT linkage (Table 3.2).

Table 3.2: List of Identified IDT Cross-Linked Peptides

SFig	Peptide	Protein Name	Neut. Mass (Da)	Obs. z	XIC Area Sum	Pro Sites
S1.1	SOOOOVyIyASOOOOTHY	EXT8	2104.9266	2	1.36E+9	--
S1.2	SOOOOyLySSOOOOVK	EXT8	1847.8465	3	4.04E+8	--
S1.3	SOOOOOVHHYSOOHHPyLyK	EXT3	2456.1186	4; 5; 6	2.90E+8	P16
S1.4	SOOOOyVySSOOOOVK	EXT8	1833.8309	3	2.62E+8	--
S1.5	SOOOOyYyHSOOOOVK	EXT8	1947.8527	2; 3; 4	8.05E+11	--
S1.6	SPOOOyYyHSOOOOVK	EXT8	1931.8578	3	1.71E+9	P2
S1.7	SOPOOyYyHSOOOOVK	EXT8	1931.8578	3	5.47E+9	P3
S1.8	SOOPOyYyHSOOOOVK	EXT8	1931.8578	3	2.10E+9	P4; P5
S1.9	SOOOPyYyHSOOOOVK	EXT8	1931.8578	3	2.35E+9	P5
S1.10	SOOOOyYyHSOPOOVK	EXT8	1931.8578	3	1.02E+9	P12
S1.11	SOOOOyYyHSOOPOVK	EXT8	1931.8578	3	1.36E+9	P13
S1.12	SOPPOyYyHSOOOOVK	EXT8	1915.8628	3	3.22E+8	P3; P4
S1.13	SOPOPyYyHSOOOOVK	EXT8	1915.8628	3	4.52E+8	P3; P5
S1.14	SOOPPyYyHSOOOOVK	EXT8	1915.8628	3	6.45E+7	P4; P5
S1.15	SOOOOVKSOOOOyYyHSOOOOVK	EXT8	2714.2388	4	1.16E+7	--
S1.16	SOOOOYYYHSOOOOVKSOOOOyYyHSOOOOVK	EXT8	3879.7104	5	4.64E+7	--

Table 3.2: List of Identified IDT Cross-Linked Peptides: Corresponding mass spectrometry evidence for each identification is in the supplemental figure indicated. The peptide sequence is shown with the tyrosine involved in the linkage noted by the lowercase y. The calculated neutral mass of the precursor is given along with the observed charge states (z). XIC Area Sum: intensity of the monoisotopic precursor summed for all charge states observed. Area is calculated as the area under the curve (ion count intensities x seconds). Pro Sites: Position of a proline residue in the sequence.

An Intermolecular Cross-Linkages of Two Tyrosines Was Identified in Small Quantities.

A dityrosine cross-linkage joining two tyrosines on two peptides was identified. Formation of the dityrosine bond results in a total mass loss of (-2.01565 Da), or two hydrogens lost in the intermolecular cross-linkage formation between the two tyrosines (Fig. 1.2). To determine if a dityrosine linkage was present in the EXT scaffold, a -2.01565 Da mass difference joining two peptides at their tyrosine residues was also investigated.

Mass Spectrometry Characterization of a Dityrosine Cross-Linked Peptide

An example of mass spectrometric evidence used for characterizing the dityrosine cross-linkage is given below (Fig. 3.4). In this example, we have two EXT1 peptides intermolecularly cross-linked to each other through their tyrosines. The MS1 isotopic profile depicts an intact parent ion at m/z=567.7707, z=4 that was isolated for fragmentation. The isotopic profile is free of contaminating peaks, supporting one cross-linked species. The parent ion mass is -2.01565 Da less than the calculated mass of two non-crosslinked peptides containing the same sequences.

This supports the existence of dityrosine linkage occurring in this species based on MS1 evidence.

MS2 fragmentation gives further evidence for the dityrosine linkage. The MS2 spectra give a rich fragment ladder of a/b- and y-ions. These ions alone give the peptide sequence and localization of the dityrosine cross-linkage. Additionally, the large cross-linked ions contain the mass loss of the dityrosine linkage. The fragment ions that do not contain the dityrosine linked motifs, do not have the mass loss for dityrosine. Internal fragment ion formation does occur in this species, but not to the extent of the peptides with the "SOOOO" motifs flanking the tyrosine cross-linking motifs demonstrated in this chapter.

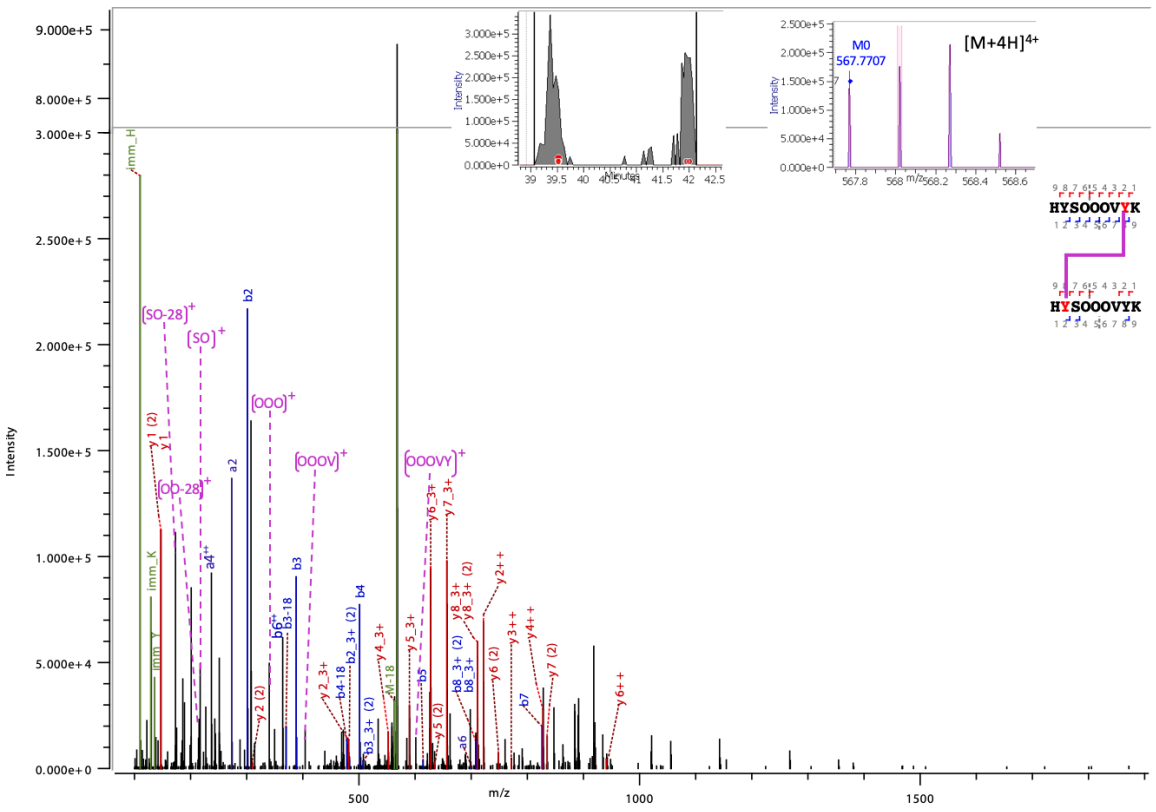

Figure 3.4: MS Characterization of HYSOOOOVyK Dityrosine Cross-Linked to HySOOOVYK: Peptide 1 is intermolecularly cross-linked at VyK to Peptide 2's HyS motif (where lowercase "y" indicates a cross-linked tyrosine). (top, left) Extracted ion chromatogram: The red dots indicate the times of several MS2s of this species collected. Interestingly, this species eluted in two peaks. (top, middle) The isotopic profile of the [M+4H]4+ precursor ion at m/z 567.7707 that was isolated for fragmentation. The blue dot indicates correct monoisotopic assignment based on calculated mass. (top right) a-, b-, & y-fragment ion coverage: Cross-linked tyrosines carrying the mass loss of (2.01565 Da) are denoted in red. Positions of y-type ions are denoted by the red bracket above the peptide sequence, a- and b-type ions are denoted in blue below the peptide sequence. (bottom) The annotated HCD MS2 spectrum: Internal fragment ions are indicated in pink as depicted in (Fig.3.2) with corresponding sequence below. Light blue asterisk indicates ions due to a loss of water from right neighboring ion. Fragment ions containing

the mass loss of the dityrosine cross-link are (peptide 1): a/b8; y8 to y2; and (peptide 2): a/b2(2) to a/b8(2); y8(2).

Identification of a Dityrosine Cross-Linked Peptide

One dityrosine linkage was identified between the two identical repetitive EXT1 peptides: HySOOOVyK. This linkage was found to occur through the second tyrosine on the first peptide and the first tyrosine on the second peptide (Fig.3.4).

SFig.	Peptide 1	Peptide 2	Prot. 1	Prot. 2	Neutral Mass (Da)	Obs z	XIC Area Sum
S2.1	HYSOOOVyK	HySOOOVYK	EXT 1	EXT 1	2267.0535	4	2.55E+07

Table 3.3: The Identified Dityrosine Cross-Linked Peptide: Corresponding mass spectrometry evidence for this identification is given in the supplemental figure indicated. The peptide sequence is shown with the tyrosine involved in the linkage noted by the lowercase y. The calculated neutral mass of the precursor is given along with the observed charge states (z). XIC Area Sum: intensity of the monoisotopic precursor summed for all charge states observed. Area is calculated as the area under the curve (ion count intensities x seconds).

Pulcherosine Intermolecular Linkages are Formed in All Four Hydroxyproline-Rich EXTs.

A pulcherosine linkage is the intermolecular linkage joining one intramolecular IDT linkage on one peptide with a third tyrosine on a second peptide. Formation of the pulcherosine bond results in a total mass loss of (4.0313 Da), or four hydrogens lost in the cross-linkage formation between three tyrosines (Fig. 1.2). To determine if a pulcherosine linkage was present in the EXT scaffold, a -4.0313 Da mass difference joining two peptides at their three tyrosine residues was investigated.

Mass Spectrometry Characterization of Pulcherosine Cross-linked Peptides

An example of mass spectrometry evidence used for characterizing the pulcherosine cross-linkage is given below (Fig. 3.5). In this example, we have an IDT linkage at EXT8's yYy motif forming the intermolecular cross-linkage to the VyK in an EXT1 peptide. The MS1 isotopic profile depicts an intact parent ion at m/z=965.4396, z=3 that was isolated for fragmentation. The isotopic profile is free of contaminating peaks, supporting the presence of one cross-linked species. The parent ion mass is -4.0313 Da less than the calculated mass of two non-crosslinked peptides containing the same sequences. This supports that the pulcherosine linkage occurs in this species based on MS1 evidence. MS2 fragmentation gives further evidence for the pulcherosine linkage. The MS2 spectra give a reasonably rich fragment ladder of a/b- and y-ions. These ions alone give the peptide sequence and localization of the pulcherosine cross-linkage. Additionally, the large cross-linked ions contain the mass loss of the pulcherosine linkage. The fragment ions that do not contain the pulcherosine-linked motifs, do not have a mass loss for pulcherosine. Internal fragment ions existing in the +2 and +3 charge states were also formed. Both +2 and +3 charge states of the internal fragment ions depict mass losses from both N-termini of both of the cross-linked peptides.

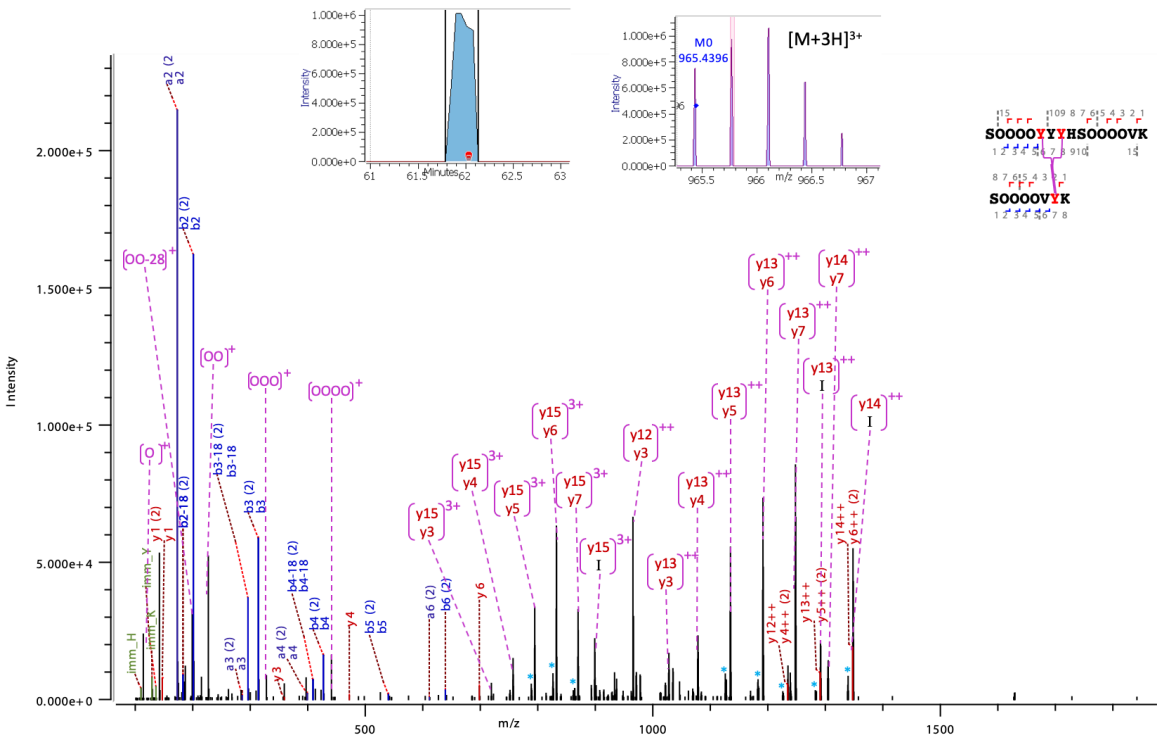

Figure 3.5: MS Characterization of SOOOOyYyHSOOOVK Pulcherosine Cross-Linked to SOOOOVyK: Peptide 1 is intramolecularly IDT cross-linked at yYy, which is then intermolecularly cross-linked to Peptide 2's VyK motif (where lowercase "y" indicates a cross-linked tyrosine). (top, left) Extracted ion chromatogram: The red dots indicate the times of several MS2s of this species collected. (top, middle) The isotopic profile of the [M+3H]3+ precursor ion at m/z 965.4396 that was isolated for fragmentation. The blue dot indicates correct monoisotopic assignment based on calculated mass. (top right) a-, b-, & y-fragment ion coverage: Cross-linked tyrosines carrying the mass loss of (4.0313 Da) are denoted in red. Positions of y-type ions are denoted by the red bracket above the peptide sequence, a-&b-type ions are denoted in blue below the peptide sequence. (bottom) The annotated HCD MS2 spectrum: Internal fragment ions are indicated in pink as depicted in (Fig.3.2) with corresponding sequence below. Light blue asterisk indicates ions due to a loss of water from right neighboring ion. Fragment ions containing the mass loss of the pulcherosine cross-link are (peptide 1): y11 to y15; and (peptide 2): y6(2) to y2(2).

Identified Pulcherosine Cross-Linked Peptides

Five different pulcherosine cross-linked species were identified (Table 3.4). Of the five species, two pulcherosine linkages were found involving a cross-linkage to an IDT in EXT8. One of these EXT8 IDT's were found to form the pulcherosine cross-linkage to the tyrosine in the EXT1 peptide SOOOOVyK (Fig. 3.5, Table 3.4). One example of the EXT1's SOOOOVyK peptide characterized in this study was found to cross-link with the IDT in the EXT 10 SOOOOyVyKSOOOOSYSOSOK peptide (Table 3.4).

Pulcherosine cross-linkages were also identified in EXT3. These linkages were found joining the first tyrosine in the EXT3 HySOOOVYHSOOOOK peptide with the IDT containing EXT3 peptide HyVyK with no missed tryptic cleavage site, and HyVyKSOOOOVK containing one missed tryptic site (Table 3.4). Pulcherosine cross-linkages were identified joining the first tyrosine in the EXT3 HySOOOVYHSOOOOK peptide with the IDT containing EXT3 peptide HyVyK with no missed tryptic cleavage site, and HyVyKSOOOOVK containing one missed tryptic site (Table 3.4).

Two additional pulcherosine linkages identified in this study occurred with two different EXT 8 IDT containing peptides. The first joined the IDT in the major repetitive EXT8 (SOOOPyYyHSOOOOVK) peptide cross-linked to one of the tyrosines in the EXT10 peptide (SOOOOYYAOTOK). Due to lack of fragmentation between these tyrosines, we could not identify the exact site of linkage (Table 3.4).

S Fig	Peptide 1	Peptide 2	Prot. 1	Prot. 2	Neutral Mass (Da)	Obs z	XIC area	Pro Site
S3.1	SOOOOyYyHSOOOOVK	SOOOOVyK	EXT 8	EXT 1	2893.2970	3;4	1.09E+08	--
S3.2	SOOOOyVyKSOOOOSYSOSOK	SOOOOVyK	EXT 10	EXT 1	3371.5245	4;5	3.02E+07	--
S3.3	HySOOOVYHSOOOOK	HyVyK	EXT 3	EXT 3	2515.1444	4;5	2.68E+08	--
S3.4	SOOOOyVyK	SOOOOyHY	EXT 10	EXT 3	2126.9109	3;4	2.74E+08	--
S3.5	SOOOPyYyHSOOOOVK	SOOOOYYAOTOK	EXT 8	EXT 10	3339.4772	4	1.74E+07	P5

Table 3.4: List of Identified Pulcherosine Cross-Linked Peptides: Corresponding mass spectrometry evidence for each identification is in the supplemental figure indicated. The peptide sequence is shown with the tyrosine involved in the linkage noted by the lowercase y. The calculated neutral mass of the precursor is given along with the observed charge states (z). XIC Area Sum: intensity of the monoisotopic precursor summed for all charge states observed. Area is calculated as the area under the curve (ion count intensities x seconds). Proline Sites: Position of a proline residue in the sequence.

Intermolecular Di-isodityrosine (Di-IDT) Cross-Links Join Two Peptides in All Four Hydroxyproline-Rich EXTs.

A Di-IDT linkage is the intermolecular linkage joining of two intramolecular IDTs linkages occurring on two peptides. Formation of the di-isodityrosine bond results in a total mass loss of (6.04595 Da), or six hydrogens (Fig. 1.2). To determine if a di-isodityrosine linkage was present in the EXT scaffold, a -6.04595 Da mass difference joining two peptides at their tyrosine motifs was investigated.

Mass Spectrometry Characterization of Di-IDT Cross-linked Peptides

An example of mass spectrometric evidence used for characterizing the Di-IDT cross-linkage is given below (Fig.3.6). In this example, we have two EXT8 peptides intramolecularly IDT cross-linked at yYy. These peptides are then subsequently intermolecularly cross-linked to one another to create Di-IDT. The MS1 isotopic profile depicts an intact parent ion at m/z=974.4297, z=4 that was isolated for fragmentation. The isotopic profile is free of contaminating peaks, supporting the presence of one cross-linked species. The parent ion mass is -6.04595 Da less than the calculated mass of two non-crosslinked peptides containing the same sequences. This supports that the Di-IDT linkage occurs in this species based on MS1 evidence.

MS2 fragmentation gives further evidence for the Di-IDT linkage. The MS2 spectra give a reasonably rich fragment ladder of a/b- and y-ions. These ions alone give the peptide sequence and localization of the Di-IDT cross-linkage. Additionally, the large cross-linked ions contain the mass loss of the Di-IDT linkage. The fragment ions that do not contain the Di-IDT linked motifs, do not have mass loss for Di-IDT. Extensive internal fragment ion formation occurs. These ions exist in the +2, +3, and +4 charge states and arise from fragmentation of both N- and C-termini.

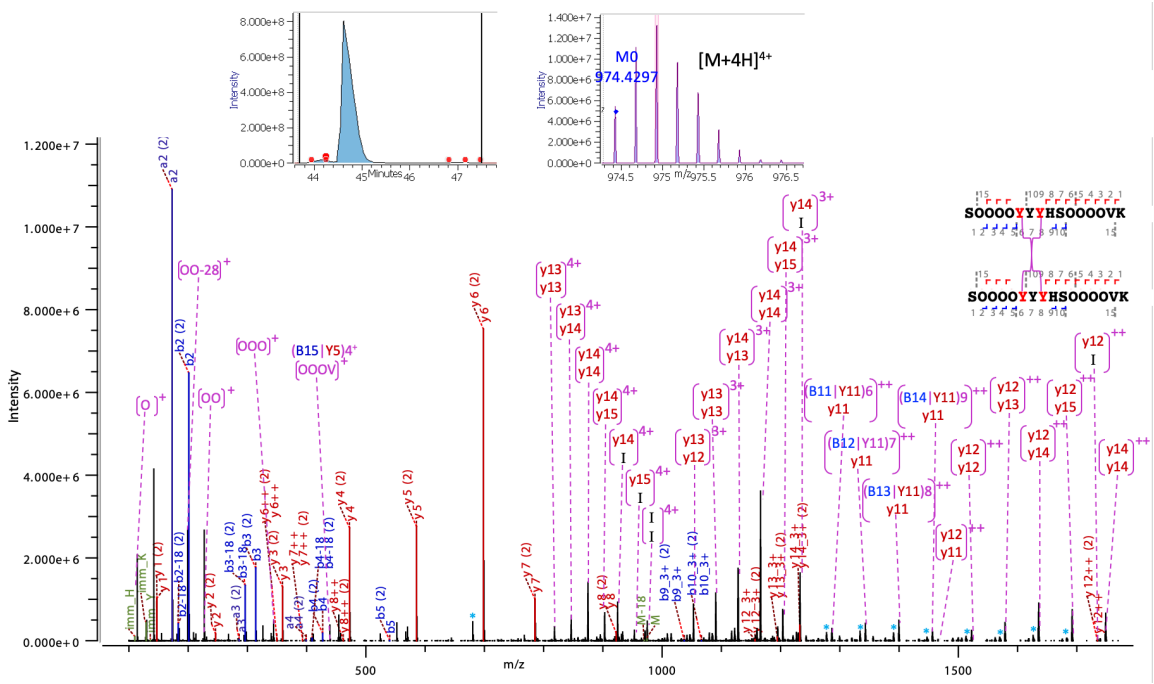

Figure 3.5: MS Characterization of SOOOOyYyHSOOOVK Di-IDT Cross-Linked to a second SOOOOyYyHSOOOVK: Peptide 1 and 2 are both intramolecularly IDT cross-linked at yYy, and subsequently intermolecularly cross-linked to one another to create Di-IDT (where lowercase "y" indicates a cross-linked tyrosine). (top, left) Extracted ion chromatogram: The red dots indicate the times of several MS2s of this species collected. (top, middle) The isotopic profile of the [M+3H]3+ precursor ion at m/z 965.4396 that was isolated for fragmentation. The blue dot indicates correct monoisotopic assignment based on calculated mass. (top right) a-, b-, & y-fragment ion coverage: Cross-linked tyrosines carrying the mass loss of (6.04595 Da) are denoted in red. Positions of y-type ions are denoted by the red bracket above the peptide sequence, a-&b-type ions are denoted in blue below the peptide sequence. (bottom) The annotated HCD MS2 spectrum: Internal fragment ions are indicated in pink as depicted in (Fig.3.2) with corresponding sequence below. Light blue asterisk indicates ions due to a loss of water from right neighboring ion. Fragment ions containing the mass loss of the Di-IDT cross-link are (peptide 1): y11 to y15; and (peptide 2): y6(2) to y2(2).

Identified Di-IDT Cross-Linked Peptides

A total of sixteen different cross-linked species carrying the -6.04595 Da were identified. (Table 3.5). Fifteen of the sixteen species identified were in EXT3 or EXT8 peptides. EXT3 is a major point of cross-linkage for this -6.04595 Da type of linkage with eight different species identified. For the EXT3 cross-linkages identified, four cross-links occurred with other EXT3 peptides, and four with EXT8 peptides (Table 3.5). All but one of these cross-linkages included the HyVyK isodityrosine with and without one missed tryptic cleavage site. EXT8 is also a major point of cross-linkage for this -6.04595 Da type of linkage with eleven different species identified. For the EXT8 cross-linkages identified, four were to other EXT8 peptides, four were to EXT3 peptides, and one was to an EXT10 peptides. Ten of these cross-linked peptides occurred on the major repetitive EXT8 peptide: SOOOOyYyHSOOOOVK (Table 3.5). Two of the Di-IDT cross-linkages identified occurred through the tryptic peptide HyEyK. Since this peptide occurs in both EXT1 and EXT3, we cannot be certain of which EXT it originated from (Table 3.5)

SFig	Peptide 1	Peptide 2	Prot 1	Prot 2	Neutral Mass (Da)	Obs z	XIC Area Sum
S4.1	HyVyKSOOOOVK	HyVyK	EXT 3	EXT 3	2177.0582	4;5	2.91 E+08
S4.2	HyVyKSOOOOVK	HyEyK	EXT 3	EXT 3 (1)	2207.0324	4;5	1.16 E+07
S4.3	HyVyKSOOOOVK	HyVyKSOOOOVK	EXT 3	EXT 3	2943.4443	4;5;6	7.15 E+07
S4.4	SOOOOyYyHSOOOOVK	yVyK	EXT 8	EXT 3	2515.122	4;5	3.13 E+08
S4.5	SOOOOyYyHSOOOOVK	VyYK	EXT 8	EXT 10	2515.122	4;5	9.51 E+08
S4.6	SOOOOyYyHSOOOOVK	HyVyK	EXT 8	EXT 3	2652.1809	4;5	1.70 E+09
S4.7	SOOOOyYyHSOOOOVK	HyEyK	EXT 8	EXT 3 (1)	2682.1551	4;5	5.90 E+07
S4.8	SOOOOyYyHSOOOOVK	HyVyKSOOOOVK	EXT 3	EXT 8	3418.567	5;6	2.57 E+08
S4.9	SOOOOyYyHSOOOOVK	SOOOOyVySSOOOOVK	EXT 8	EXT 8	3779.6679	4;5	3.49 E+09
S4.10	SOOOOyYyHSOOOOVK	SOOOOyEyKSOOOOVK	EXT 8	EXT 8	3850.705	4;5;6	5.01 E+08
S4.11	SOOOOyYyHSOOOOVK	SOOOOyLySSOOOOVK	EXT 8	EXT 8	3793.6836	3;4;5	3.58 E+09
S4.12	SOOOOyYyHSOOOOVK	SOOOOVyIyASOOOOTH	EXT 8	EXT 8	3887.7003	3;4;5	3.62 E+08
S4.13	SOOOOyYyHSOOOOVK	SOOOOyYyHSOOOOVK	EXT 8	EXT 8	3893.6897	4;5;6	3.77 E+11
S4.14	SOOOOOVHHYSOOHHPyLyK	HyVyK	EXT 3	EXT 3	3160.4468	4;5;6	5.45 E+08

Table 3.5: List of Identified Di-IDT Cross-Linked Peptides: Corresponding mass spectrometry evidence for each identification is in the supplemental figure indicated. The peptide sequence is shown with the tyrosine involved in the linkage noted by the lowercase y. The calculated neutral mass of the precursor is given along with the observed charge states (z). XIC Area Sum: intensity of the monoisotopic precursor summed for all charge states observed. Area is calculated as the area under the curve (ion count intensities x seconds). Proline Sites: Position of a proline residue in the sequence.

The Major Repetitive Peptides of EXT8 and EXT3 Contain Non-Hydroxylated Prolines at Different Sites.

Glycosylation of extensins' repetitive hydroxyprolines is suspected to be a requirement for cross-linking.[23,28] Thus, mapping site-specific missed proline hydroxylation may give insight to the effects of glycosylation on scaffold assembly.

To characterize sites of missed hydroxylation in these motifs, Byos searches were set to include up to two non-oxygenated prolines in their sequences. Several examples of the two major repetitive peptides were identified containing prolines instead of the expected hydroxyproline at different sites in the S(O)n repetitive motifs:

EXT8: SOOOOYYYHSOOOOVK: The major repetitive peptide was identified to contain sequences with one non- oxygenated proline occurring in different positions in both IDT and non-IDT cross-linked forms. The non-IDT cross-linked form contained one proline at 5 different positions in the sequence, and two prolines at 5 different positions in the sequence (Table S2a). The IDT cross-linked form, contained one proline at 6 different positions in the sequence, and two prolines at 3 different positions in the sequences (Table S2a).

EXT3: HYSOOOVYHSOOOOK: This repetitive peptide contained sites of missed hydroxylation of proline. Four different positions in the sequence contained proline, and one example was found to contain two prolines (Table S2b).

Discussion

Establishment of the rules outlining the self-assembly of the extensin scaffold has been a process of continually evolving for nearly three decades.[4,20,70] These rules first described the ability of extensins' highly periodic, repetitive, motifs to drive the alignment of extensin monomers for assembly of a scaffold stabilized by enzymatic catalyzed cross-links.[20,59-61] Notable progress in extensin research has allowed for several updates to these rules.[5,32-34,41] However, in vivo characterization of cross-linked species has proven to be a challenging undertaking. Our data, presented here, significantly expands upon this set of rules by reporting the first *in vivo* characterization and identification of the tyrosine cross-linkages in the extensin scaffold in Arabidopsis tissue culture cell walls. Furthermore, our assignment of the cross-linkages to peptide identity allows significant progress to the understanding of the assembled extensin network as a whole. Insights gained from our characterization, and the implications for each are discussed as follows:

The extensin scaffold network contains intra- and intermolecular cross-linkages arising from at least four extensins: EXT3, EXT8, and to a lesser extent EXT1 and EXT10 (Tables 3.2-5).

Simultaneous determination of cross-linkage identity and extensin identity is a requirement for full understanding the self-assembly of the extensin scaffold.[32-34,41] Our data corroborate that the extensin scaffold is cross-linked through its tyrosine residues (Tables 3.2-4). Using our methodology, we identified four extensins containing cross-linkages arising from the creation of one intramolecular and three intermolecular linkages composed of tyrosine dimers, trimers, and tetramers. This information in addition to extensin identity obtained allows for a better understanding the roles of these cross-linked extensins.

High degree of IDT formation in EXT8 supports that intramolecular IDT-rich extensins are critically important for cell wall assembly and defense.

Extensins are known to form intramolecular IDT cross-links at yXy *in vitro*.[60,61,5,71] Our data supports that proteins containing the yXy motif also have the ability to form IDT crosslinks *in vivo* (Table 3.2). Of the sixteen IDT species we identified, fifteen occurred in EXT8 peptides and one occurred in EXT3. Intramolecular IDT is thought to act as a stabilizer that rigidifies the polypeptide chain while not disrupting its left handed helical conformation.[31] It has also been suspected to act as starting point for intermolecular cross-linkages to form for scaffold growth. Indeed, our data overwhelmingly demonstrates that IDTs are incorporated into pulcherosine and Di-IDT intermolecular linkages (Tables 3.4 & 3.5).

Pulcherosine and Di-IDT intermolecular cross-linkages occurring within the same extensin and between two different extensins describes a highly interacting network exhibiting two types of dimensional growth.

Previous *in vivo* characterization of cross-linked extensin peptides demonstrated that two configurations of monomeric alignment (staggered vs parallel) result in the formation of the intermolecular linkage pulcherosine (staggered) or Di-IDT (parallel).[72] Our data supports that both alignments of extensin monomers occur though the identification of pulcherosine, and to a greater extent, Di-IDT linkages that readily form in the cell wall (Tables 3.4 & 3.5).

Pulcherosine cross-linkages are suspected to drive assembly of the extensin scaffold outwards both linearly and laterally in two dimensions.[5] Our data suggests that EXT1 and EXT3 are important for two dimensional cell wall assembly. Previous *in vitro* characterization of EXT3, identified that pulcherosine cross-linkages form readily at its staggered monomers containing the repetitive sequences SOOOOKKHyVyKSOOOOVKHySOOOVyH.[5] Our data shows that pulcherosine linkages in fact do occur *in vivo* at all possible tryptic peptides arising from this

sequence (Table 3.4). The role of extensins low in IDT motifs has been a point of discussion for some time.[4] We identified five instances of pulcherosine cross-linkages involving EXT1's VyK motif to an IDT in EXT8 or EXT10. These findings demonstrate that EXT1 is important for cell wall assembly though intermolecular cross-linkage formation of its repetitive motifs lacking IDT (Table 3.4).[60,61]

Formation of the Di-IDT linkages are thought to drive assembly of the extensin scaffold outwards linearly or laterally in one dimension.[5] For the formation of Di-IDT to occur, monomers containing IDT must be aligned parallel.[4,5,73] Our data suggest that EXT3 and EXT8 are particularly important for this type of cell wall assembly. We identified four Di-IDTs arising from EXT3 IDTs cross-linked to other EXT3 IDTs. Examples of IDTs from EXT8 cross-linking to IDTs in EXT3 were also identified (Table 3.5).

Taken together, the findings above describe a multidimensional scaffolding network assembled as a result of different combinations of cross-linkages forming between different combinations of extensins. Additionally, characterization of the dual utility of EXT3 in formation of both types of intermolecular linkages adds to the explanation of why mutations of this extensin are embryonic lethal.[72]

Identification of the previously uncharacterized intermolecular dityrosine occurring at the suspected P1 motif VyK with the uncharacterized HyK motif, supports that intermolecular linkages do not require IDT (Figure 3.4).

Previous research investigating the **P1:**SOOOOTOVyK extensin monomer describes its ability insolubilize into the extensin network through suspected intermolecular cross-linking. Thus, the VyK motif has been a suspected point of intermolecular linkages. However, direct evidence of this monomer forming a cross-linkage with another tyrosine, creating a di-tyrosine or an intermolecular IDT has been elusive.[60,61]

Our data suggests that the formation of a dityrosine linkage is possible in low quantities. A dityrosine linkage was found between the two identical repetitive EXT1 peptides: HySOOOVyK. Interestingly, this linkage was found to occur through the first tyrosine on the first peptide at the HyS motif and the second tyrosine on the second peptide at the VyK motif (Fig. 3.4).

Hydrophobic interaction of multiple tyrosines occurring in the C-termini may initiate end-on adhesion.

It has been suggested that hydrophobic interaction of multiple tyrosine residues at the C-terminus may initiate end-on adhesion of extensin monomers.[5] The ubiquity of the C-terminal tyrosine is a notable feature of 18 out of the 20 Arabidopsis hydroxyproline-rich extensins.[5]

Cross-linkages involving both of the C-terminal peptides of EXT3 and EXT8 were identified (Tables 3.4 & 3.5). These peptides contain two and three tyrosines respectively. Notably, we did not find cross-linkages in the C-terminal peptides of EXT1 or EXT10. These peptides only contain one tyrosine each (Tables S3a & S3d). Our evidence adds further support that hydrophobic interaction may initiate end-on adhesion.

Amino acid (similarity or differences) contained in and flanking the different types cross-linked motifs may shed light on the utility of extensin peroxidase to create different linkages, or open the possibility of the existence of another peroxidase.

Several cross-linking studies have documented the ability of a highly specific extensin peroxidase to form IDT, pulcherosine, and Di-IDT cross-linkages.[4,34,59,72] An additional extensin peroxidase has been documented, but more characterization is needed to determine its exact cross-linkage specificity.[74]

In vitro cross-linking studies involving the mutation of a flanking lysine to leucine in the motif OyLyK resulted in a significant decrease in cross-linking. This suggested a role for lysine in enzyme recognition.[34] Additional cross-linkages have been identified in peptides containing OyLyK, OyYyK., and HyVyK. This adds support the role of lysine for enzymatic recognition.[5,60,61] Our data agrees that IDTs form at these motifs with the flanking lysine. However, we found intramolecular IDT cross-linkages formed in peptides that do not contain a lysine (Tables 2-4). Intermolecular cross-linkages were also identified occurring at motifs without a flanking lysine (Tables 3.4 & 3.5). Additionally, we also identified unanticipated intermolecular linkages that occur at motifs containing a YY (Tables 3.4 & 3.5). Due to these findings, we suspect that either the flanking lysine is not a requirement for cross-linking by the extensin peroxidase, or an additional activity of other uncharacterized peroxidases is at work.

Localization of non-hydroxylated prolines in different positions does not prevent IDT formation in the yYyK motif, but it may decrease the formation of intermolecular linkages.

Continuous stretches of O-arabinosylated hydroxyprolines are a hallmark of extensins.[20] It is believed that correct Hyp-O-arabinosylation is integral for scaffold assembly by: (i) locking the extensins into their rod shaped structure; and (ii) creating bends around tyrosine crosslinking motifs, exposing them to create a better scenario for both intra- and intermolecular cross-linking.[5,23] Therefore, it could be hypothesized that missed sites of hydroxylation of prolines could negatively affect cross-link formation.

Several examples of one or two non-hydroxylated prolines were identified in EXT8's major repetitive peptide: SOOOOyYyHSOOOOVK (Table S2a). Several key points were noticed:

(1) The most abundant forms (based on XIC intensity) of this peptide contain no missed sites of hydroxylation for both non-IDT and IDT cross-linked forms. The IDT cross-linked peptides

occurred in a slightly higher abundance. Therefore, we suspect the fully glycosylated form of this peptide is favored in the assembled scaffold.

(2) Both non-IDT and IDT cross-linked forms of the peptide contain sites of one missed hydroxylation of proline before or after the YYY motif. Since both IDT and non-IDT peptides are of relatively similar intensities, we suspect one missed hydroxylation site does not affect the ability to create intramolecular cross-linkages.

(3) Both non-IDT and IDT peptides contain sites of two missed hydroxylations of prolines. Peptides containing two prolines are the least abundant form of these species. Additionally, both non-IDT and IDT peptides are also of reasonably similar intensities. However, a distinct trend is observed with the location of these missed hydroxylated prolines. IDT was only formed if both prolines occur before the yYy motif. No IDT was formed in peptides containing: (i) a proline before the YYY and a proline after; or (ii) both prolines occur after the YYY motif. Based on these observations, we conclude that structural changes due to missed hydroxylation (or in other words loss Hyp-O-arabinosylation) occurring before the yYy is more forgiving.

(4) Finally, intermolecular linkages were only identified if the SOOOOyYyHSOOOOVK was in a fully hydroxylated state (Table S2a). We conclude that while the formation of intramolecular IDT seems more adaptable to some degree to structural changes caused by missed hydroxylation, intermolecular linkages require the ridged rod-shaped motifs to increase tyrosine accessibility.

LRX4 extensin peptides may cross-link into the scaffolding network.

Although suspected, direct evidence of a cross-linked LRX extensins with the scaffold has remained a mystery. LRX proteins are chimeric extensins. The C-termini contains classical hydroxyproline-rich extensin motifs that are thought to have roles in anchoring. The N-termini of

the LRXs contain leucine-rich repetitive (LRR) domains. These domains are thought to have roles in protein-protein interactions.[42-44]

We identified tryptic peptides originating from LRX4 extensin in our sample (Table S1). However, we could not confidently identify a peptide containing a direct tyrosine cross-linkage in our data. Identification of peptides corresponding to LRX4 suggest that it is in some way cross-linked into the scaffold network. LRX4 is suspected to behave as a cell wall salt sensor and it is thought to have a role in controlling cell expansion in root cells.[75,76] Therefore, we can imagine the utility of having sensors decorate the assembled scaffold. Future directions to characterize this cross-linkage will give better understanding to the role of LRXs in the scaffold.

Conclusion

In conclusion, our evidence presented here provides significant progress in updating the understanding of the assembled extensin scaffold. Cross-link mass spectrometry is a powerful tool for characterizing the extensin scaffold due to its ability to simultaneously characterize linkage and identity of the extensins. However, data analysis is challenging. Future directions, including the incorporation of these fragmentation trends into the algorithms for data analysis, would provide an exciting step forward for full characterization of this scaffolding network.

REFERENCES

1 Cosgrove, D. J. Wall structure and wall loosening. A look backwards and forwards. *Plant Physiol* **125**, doi:10.1104/pp.125.1.131 (2001).

2 Carpita, N. C. & Gibeaut, D. M. Structural models of primary cell walls in flowering plants: consistency of molecular structure with the physical properties of the walls during growth. *Plant J* **3**, doi:10.1111/j.1365-313X.1993.tb00007.x (1993).

3 McCann, M. *et al.* in *Abstr Pap Am Chem S.* U259-U259 (AMER CHEMICAL SOC 1155 16TH ST, NW, WASHINGTON, DC 20036 USA).

4 Lamport, D. T. A., Kieliszewski, M. J., Chen, Y. & Cannon, M. C. Role of the Extensin Superfamily in Primary Cell Wall Architecture. *Plant Physiol* **156**, 11-19, doi:10.1104/pp.110.169011 (2011).

5 Cannon, M. C. *et al.* Self-assembly of the plant cell wall requires an extensin scaffold. *Proceedings of the National Academy of Sciences* **105**, 2226-2231 (2008).

6 Somerville, C. Cellulose synthesis in higher plants. *Annu. Rev. Cell Dev. Biol.* **22**, 53-78 (2006).

7 Kimura, S. *et al.* Immunogold labeling of rosette terminal cellulose-synthesizing complexes in the vascular plant Vigna angularis. *The Plant Cell* **11**, 2075-2085 (1999).

8 Parker, C. C., Parker, M. L., Smith, A. C. & Waldron, K. W. Pectin distribution at the surface of potato parenchyma cells in relation to cell-cell adhesion. *J Agric Food Chem* **49**, doi:10.1021/jf0104228 (2001).

9 Daher, F. B. & Braybrook, S. A. How to let go: pectin and plant cell adhesion. *Front Plant Sci* **6**, doi:ARTN 523 10.3389/fpls.2015.00523 (2015).

10 Harholt, J., Suttangkakul, A. & Vibe Scheller, H. Biosynthesis of Pectin. *Plant Physiol* **153**, 384-395, doi:10.1104/pp.110.156588 (2010).

11 Ochoa-Villarreal, M., Aispuro-Hernández, E., Vargas-Arispuro, I. & Martínez-Téllez, M. Á. in *Polymerization* (ed Ailton De Souza Gomes) Ch. 04 (InTech, 2012).

12 Buchanan, B. B. *Biochemistry and molecular biology of plants*. (John Wiley & Sons, 2015).

13 Qi, X., Behrens, B. X., West, P. R. & Mort, A. J. Solubilization and partial characterisation of extensin fragments from cell walls of cotton suspension cultures. *Plant Physiol* **108**, doi:10.1104/pp.108.4.1691 (1995).

14 Nunez, A., Fishman, M. L., Fortis, L. L., Cooke, P. H. & Hotchkiss Jr, A. T. Identification of extensin protein associated with sugar beet pectin. *J Agr Food Chem* **57**, 10951-10958 (2009).

15 Lamport, D. T. A. & Northcote, D. H. Hydroxyproline in Primary Cell Walls of Higher Plants. *Nature* **188**, 665-666, doi:10.1038/188665b0 (1960).

16 Lamport, D. T. A. in *Advances in Botanical Research* Vol. 2 (ed R. D. Preston) 151-218 (Academic Press, 1966).
17 Showalter, A. M., Keppler, B., Lichtenberg, J., Gu, D. & Welch, L. R. A Bioinformatics Approach to the Identification, Classification, and Analysis of Hydroxyproline-Rich Glycoproteins. *Plant Physiol* **153**, 485-513, doi:10.1104/pp.110.156554 (2010).
18 Johnson, K. L. *et al.* Insights into the Evolution of Hydroxyproline-Rich Glycoproteins from 1000 Plant Transcriptomes. *Plant Physiol* **174**, 904 (2017).
19 Kieliszewski, M. J., Lamport, D. T., Tan, L. & Cannon, M. C. Hydroxyproline-Rich Glycoproteins: Form and Function. *Annual Plant Reviews: Plant Polysaccharides, Biosynthesis and Bioengineering, Volume 41*, 321-342 (2010).
20 Kieliszewski, M. J. & Lamport, D. T. A. Extensin: Repetitive motifs, functional sites, post-translational codes, and phylogeny. *Plant J* **5**, doi:10.1046/j.1365-313X.1994.05020157.x (1994).
21 Campargue, C., Lafitte, C., Esquerré-Tugayé, M.-T. & Mazau, D. Analysis of Hydroxyproline and Hydroxyproline-Arabinosides of Plant Origin by High-Performance Anion-Exchange Chromatography–Pulsed Amperometric Detection. *Analytical Biochemistry* **257**, 20-25, doi:https://doi.org/10.1006/abio.1997.2526 (1998).
22 Ogawa-Ohnishi, M., Matsushita, W. & Matsubayashi, Y. Identification of three hydroxyproline O-arabinosyltransferases in Arabidopsis thaliana. *Nature Chemical Biology* **9**, 726, doi:10.1038/nchembio.1351 https://www.nature.com/articles/nchembio.1351#supplementary-information (2013).
23 Velasquez, S. M. *et al.* Low sugar is not always good: Impact of specific O-glycan defects on tip growth in Arabidopsis. *Plant Physiol* **168**, 808-813 (2015).
24 Kieliszewski, M. J. The latest hype on Hyp-O-glycosylation codes. *Phytochemistry* **57**, 319-323, doi:https://doi.org/10.1016/S0031-9422(01)00029-2 (2001).
25 Shpak, E., Barbar, E., Leykam, J. F. & Kieliszewski, M. J. Contiguous Hydroxyproline Residues Direct Hydroxyproline Arabinosylation in Nicotiana tabacum. *J Biol Chem* **276**, 11272-11278, doi:10.1074/jbc.M011323200 (2001).
26 Velasquez, S. M. *et al.* Low Sugar Is Not Always Good: Impact of Specific O-Glycan Defects on Tip Growth in Arabidopsis. *Plant Physiol* **168**, 808 (2015).
27 Saito, F. *et al.* Identification of Novel Peptidyl Serine α-Galactosyltransferase Gene Family in Plants. *The Journal of Biological Chemistry* **289**, 20405-20420, doi:10.1074/jbc.M114.553933 (2014).
28 Shinohara, H. & Matsubayashi, Y. Chemical Synthesis of Arabidopsis CLV3 Glycopeptide Reveals the Impact of Hydroxyproline Arabinosylation on Peptide Conformation and Activity. *Plant Cell Physiol* **54**, 369-374, doi:10.1093/pcp/pcs174 (2013).
29 Ishiwata, A., Kaeothip, S., Takeda, Y. & Ito, Y. Synthesis of the highly glycosylated hydrophilic motif of extensins. *Angewandte Chemie* **126**, 9970-9974 (2014).
30 Yoshiba, Y. *et al.* Characterization of Four Extensin Genes in Arabidopsis thaliana by Differential Gene Expression under Stress and Non-Stress Conditions. *DNA Research* **8**, 115-122, doi:10.1093/dnares/8.3.115 (2001).
31 Epstein, L. & Lamport, D. T. An intramolecular linkage involving isodityrosine in extensin. *Phytochemistry* **23**, 1241-1246 (1984).
32 Brady, J., Sadler, I. & Fry, S. Pulcherosine, an oxidatively coupled trimer of tyrosine in plant cell walls: Its role in cross-link formation. *Phytochemistry* **47**, doi:10.1016/S0031-9422(97)00592-X (1998).
33 Fry, S. C. Isodytyrosine, a new cross-linking amino acid from plant cell-wall glycoprotein. *Biochem J* **204**, doi:10.1042/bj2040449 (1982).

34 Held, M. A. *et al.* Di-isodityrosine is the intermolecular cross-link of isodityrosine-rich extensin analogs cross-linked in vitro. *J Biol Chem* **279**, doi:10.1074/jbc.M408396200 (2004).
35 Chen, Y. *et al.* Identification of the Abundant Hydroxyproline-Rich Glycoproteins in the Root Walls of Wild-Type Arabidopsis, an ext3 Mutant Line, and Its Phenotypic Revertant. *Plants* **4**, doi:10.3390/plants4010085 (2015).
36 Mort, A. J. & Lamport, D. T. A. Anhydrous hydrogen fluoride deglycosylates glycoproteins. *Analytical Biochemistry* **82**, 289-309, doi:http://dx.doi.org/10.1016/0003-2697(77)90165-8 (1977).
37 Andrews, J., Adams, S., Burton, K. & Edmondson, R. Partial purification of tomato fruit peroxidase and its effect on the mechanical properties of tomato fruit skin. *J Exp Bot* **53**, 2393-2399 (2002).
38 Showalter, A. M. & Basu, D. Extensin and Arabinogalactan-Protein Biosynthesis: Glycosyltransferases, Research Challenges, and Biosensors. *Front Plant Sci* **7**, doi:10.3389/fpls.2016.00814 (2016).
39 Fry, S. C. Primary cell wall metabolism: tracking the careers of wall polymers in living plant cells. *New Phytol* **161**, 641-675 (2004).
40 Boller, T. & Felix, G. A renaissance of elicitors: perception of microbe-associated molecular patterns and danger signals by pattern-recognition receptors. *Annual review of plant biology* **60**, 379-406 (2009).
41 D. Brady, J., Sadler, I. & Fry, S. *Di-isodityrosine, a novel tetrameric derivative of tyrosine in plant cell wall proteins: A new potential cross-link*. Vol. 315 (Pt 1) (1996).
42 Baumberger, N. *et al.* Whole-Genome Comparison of Leucine-Rich Repeat Extensins in Arabidopsis and Rice. A Conserved Family of Cell Wall Proteins Form a Vegetative and a Reproductive Clade. *Plant Physiol* **131**, 1313-1326, doi:10.1104/pp.102.014928 (2003).
43 Baumberger, N., Ringli, C. & Keller, B. The chimeric leucine-rich repeat/extensin cell wall protein lrx1 is required for root hair morphogenesis in Arabidopsis thaliana. *Genes Dev* **15**, doi:10.1101/gad.200201 (2001).
44 Ringli, C. The role of extracellular LRR-extensin (LRX) proteins in cell wall formation. *Plant Biosystems-An International Journal Dealing with all Aspects of Plant Biology* **139**, 32-35 (2005).
45 Komalavilas, P. & Mort, A. J. The acetylation of O-3 of galacturonic acid in the rhamnose-rich portion of pectins. *Carbohyd Res* **189**, 261-272 (1989).
46 Bauer, S., Vasu, P., Persson, S., Mort, A. J. & Somerville, C. R. Development and application of a suite of polysaccharide-degrading enzymes for analyzing plant cell walls. *Proceedings of the National Academy of Sciences* **103**, 11417-11422, doi:10.1073/pnas.0604632103 (2006).
47 Miller, G. L. Use of Dinitrosalicylic Acid Reagent for Determination of Reducing Sugar. *Analytical Chemistry* **31**, 426-428 (1959).
48 Mort, A. J. An apparatus for safe and convenient handling of anhydrous, liquid hydrogen fluoride at controlled temperatures and reaction times. Application to the generation of oligosaccharides from polysaccharides. *Carbohyd Res* **122**, 315-321 (1983).
49 Mort, A., Komalavilas, P., Rorrer, G. & Lamport, D. in *Plant Fibers* 37-69 (Springer, 1989).
50 Latarullo, M. B. G., Tavares, E. Q. P., Maldonado, G. P., Leite, D. C. C. & Buckeridge, M. S. Pectins, Endopolygalacturonases, and Bioenergy. *Front Plant Sci* **7**, doi:ARTN 1401
10.3389/fpls.2016.01401 (2016).
51 van Holst, G.-J. & Varner, J. E. Reinforced Polyproline II Conformation in a Hydroxyproline-Rich Cell Wall Glycoprotein from Carrot Root. *Plant Physiol* **74**, 247-251 (1984).

52 Gainey, L. & Mort, A. J. in *14th Annual Research Symposium in Biological Sciences*.
53 Albersheim, P., Talmadge, K., Keegstra, K. & Bauer, W. The structure of plant cell walls (1). *Plant Physiol* **51**, 158-178 (1973).
54 Hall, Q. & Cannon, M. C. The cell wall hydroxyproline-rich glycoprotein rsh is essential for normal embryo development in Arabidopsis. *Plant Cell* **14**, doi:10.1105/tpc.010477 (2002).
55 Hyams, J. & Davies, D. R. The induction and characterisation of cell wall mutants of Chlamydomonas reinhardi. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **14**, 381-389 (1972).
56 Lamport, D. T. Oxygen fixation into hydroxyproline of plant cell wall protein. *J Biol Chem* **238**, 1438-1440 (1963).
57 LAMPORT, D. T. in *Carbohydrates: structure and function* 501-541 (Elsevier, 1980).
58 Lamport, D. T. & Clark, L. Isolation and partial characterization of hydroxyproline-rich glycopeptides obtained by enzymic degradation of primary cell walls. *Biochemistry* **8**, 1155-1163 (1969).
59 Schnabelrauch, L. S., Kieliszewski, M., Upham, B. L., Alizedeh, H. & Lamport, D. T. A. Isolation of pl 4.6 extensin peroxidase from tomato cell suspension cultures and identification of val-tyr-lys as putative intermolecular cross-link site. *Plant J* **9**, doi:10.1046/j.1365-313X.1996.09040477.x (1996).
60 Smith, J. J., Muldoon, E. P. & Lamport, D. T. Isolation of extensin precursors by direct elution of intact tomato cell suspension cultures. *Phytochemistry* **23**, 1233-1239 (1984).
61 Smith, J. J., Muldoon, E. P., Willard, J. J. & Lamport, D. T. Tomato extensin precursors P1 and P2 are highly periodic structures. *Phytochemistry* **25**, 1021-1030 (1986).
62 Bern, M., Kil, Y. J. & Becker, C. Byonic: Advanced Peptide and Protein Identification Software. *Current protocols in bioinformatics / editoral board, Andreas D. Baxevanis ... [et al.]* **CHAPTER**, Unit13.20-Unit13.20, doi:10.1002/0471250953.bi1320s40 (2012).
63 Rappsilber, J. The Beginning of a Beautiful Friendship: Cross-Linking/mass Spectrometry and Modelling of Proteins and Multi-Protein Complexes. *J. Struct. Biol.* **173**, 530 (2011).
64 Mukherjee, S. *et al.* Characterization and Identification of Dityrosine Cross-Linked Peptides Using Tandem Mass Spectrometry. *Analytical Chemistry* **89**, 6136-6145, doi:10.1021/acs.analchem.7b00941 (2017).
65 Iacobucci, C. & Sinz, A. To Be or Not to Be? Five Guidelines to Avoid Misassignments in Cross-Linking/Mass Spectrometry. *Anal. Chem.* **89**, 7832 (2017).
66 Giese, S. H., Belsom, A., Sinn, L., Fischer, L. & Rappsilber, J. Noncovalently Associated Peptides Observed during Liquid Chromatography-Mass Spectrometry and Their Effect on Cross-Link Analyses. *Analytical Chemistry* **91**, 2678-2685, doi:10.1021/acs.analchem.8b04037 (2019).
67 Lenz, S., Giese, S. H., Fischer, L. & Rappsilber, J. In-Search Assignment of Monoisotopic Peaks Improves the Identification of Cross-Linked Peptides. *Journal of Proteome Research* **17**, 3923-3931, doi:10.1021/acs.jproteome.8b00600 (2018).
68 Baker, P. R. a. C., K.R. *Protein Prospector*, <http://prospector.ucsf.edu> (
69 Roepstorff, P. & Fohlman, J. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomedical mass spectrometry* (1984).
70 Lamport, D. Life behind cell walls: paradigm lost, paradigm regained. *Cellular and Molecular Life Sciences* **58**, 1363-1385 (2001).
71 Stafstrom, J. P. & Staehelin, L. A. Cross-linking patterns in salt-extractable extensin from carrot cell walls. *Plant Physiol* **81**, 234-241 (1986).
72 Cannon, M. C. *et al.* Self-assembly of the plant cell wall requires an extensin scaffold. *Proc Natl Acad Sci USA* **105**, doi:10.1073/pnas.0711980105 (2008).

73 Chen, Y., Dong, W., Tan, L., Held, M. A. & Kieliszewski, M. J. Arabinosylation Plays a Crucial Role in Extensin Cross-linking In Vitro. *Biochemistry Insights* **8**, 1-13, doi:10.4137/BCI.S31353 (2015).
74 Price, N. J. *et al.* A biochemical and molecular characterization of LEP1, an extensin peroxidase from lupin. *J Biol Chem* **278**, 41389-41399 (2003).
75 Zhao, C. *et al.* Leucine-rich repeat extensin proteins regulate plant salt tolerance in Arabidopsis. *Proceedings of the National Academy of Sciences* **115**, 13123-13128, doi:10.1073/pnas.1816991115 (2018).
76 Draeger, C. *et al.* Arabidopsis leucine-rich repeat extensin (LRX) proteins modify cell wall composition and influence plant growth. *Bmc Plant Biol* **15**, 155, doi:10.1186/s12870-015-0548-8 (2015).

APPENDICES

APPENDIX A:

Supplemental Table S1: List of Non-Cross-Linked Peptide Identifications

Supplemental Table S2a: Missed Hydroxylated Prolines of the EXT8 Major Repetitive Peptide: SOOOOYYYHSOOOOVK

Supplemental Table S2b: Missed Hydroxylated Prolines of the EXT3 Major Repetitive Peptide: HYSOOOOVYHSOOOOK

APPENDIX B:

Supplemental Figures S1.1 to S1.16: MS Evidence for IDT Cross-Linked Peptides

Supplemental Figures S2.1: MS Evidence for the Dityrosine Cross-Linked Peptide

Supplemental Figures S3.1 to S3.5: MS Evidence for Pulcherosine Cross-Linked Peptides

Supplemental Figures S4.1 to S4.14: MS Evidence for Di-IDT Cross-Linked Peptides

Supplemental Figures S5.1 to S5.33: MS Evidence for Non Cross-Linked Peptides

APPENDIX C:

Supplemental Tables S3a-e: Full Protein Sequences of EXTs and LRX Identified with Repetitive Tryptic Peptides Numbered

APPENDIX A

Supplemental Table S1: List of Non Cross-Linked Peptide Identifications

SFig.	Peptide	Prot Name	Neut. Mass (Da)	Obs z	XIC Area Sum	Pro Sites
S5.1	HYTOOVK	EXT 3	872.4392	2	5.36E+05	--
S5.2	LSEFDDR	LRX 4	880.3927	2	1.33E+09	--
S5.3	IPASICQLPK	LRX 4	1125.6216	2	3.84E+09	P2; P9
S5.4	YYSOOOVYK	EXT 1	1160.539	2	3.80E+09	--
S5.5	FNEFEGTVPK	LRX 4	1166.5608	2	2.73E+09	P9
S5.6	FNeFEGTVPK	LRX 4	1180.5764	2	7.32E+08	P9
S5.7	NCLPGRPAQR	LRX 4	1167.5931	3	8.42E+07	P4; P7
S5.8	PSVDCGSFGCGR	LRX 4	1297.518	2	1.41E+09	P1
S5.9	LLFELDLSNNR	LRX 4	1332.7038	2	4.86E+08	--
S5.10	FPTVVLHLPSLK	LRX 4	1349.8071	3	9.04E+08	P2; P9
S5.11	SOOOPYYAOTOK	EXT 10	1393.6402	2	1.21E+09	P5
S5.12	SOOOOYYAPTOK	EXT 10	1393.6402	2	2.37E+09	P9
S5.13	HYSOOOVYHSOOOOK	EXT 3	1810.8162	3; 4	1.09E+09	--
S5.14	HYSOPOVYHSOOOOK	EXT 3	1794.8213	3; 4	1.60E+07	P5
S5.15	HYSOOOVYHSOPOOK	EXT 3	1794.8213	3; 4	1.07E+08	P12
S5.16	HYSOOOVYHSOOPOK	EXT 3	1794.8213	4	2.35E+07	P13
S5.17	HYSOOOVYHSOOOPK	EXT 3	1794.8213	3; 4	4.44E+08	P14

S5.18	HYSOOOVYHSOOPPK	EXT 3	1778.8264	3	3.15E+07	P13; P14
S5.19	HYSOOOVYHSOOOOKK	EXT 3	1938.9112	3; 4	3.94E+07	--
S5.20	HYSOOOVYHSOOPOKK	EXT 3	1922.9163	4	9.10E+06	P13
S5.21	HYSOOOVYHSOOOPKK	EXT 3	1922.9163	4	9.10E+06	P14
S5.22	HYSOOOVYHSOOOOKEK	EXT 3	2067.9538	4	3.94E+06	--
S5.23	SOOOOYYYHSOOOOVK	EXT 8	1949.8683	2; 3	1.01E+10	--
S5.24	SOPOOYYYHSOOOOVK	EXT 8	1933.8734	3	3.02E+08	P3
S5.25	SOOPOYYYHSOOOOVK	EXT 8	1933.8734	3	3.02E+08	P4
S5.26	SOOOOYYYHSOPOOVK	EXT 8	1933.8734	3	2.02E+09	P12
S5.27	SOOOOYYYHSOOPOVK	EXT 8	1933.8734	3	2.02E+09	P13
S5.28	SOOOOYYYHSOOOPVK	EXT 8	1933.8734	3	2.02E+09	P14
S5.29	SOPOOYYYHSOPOOVK	EXT 8	1917.8785	3	2.07E+08	P3; P12
S5.30	SOPOOYYYHSOOPOVK	EXT 8	1917.8785	3	2.07E+08	P3; P13
S5.31	SOOPOYYYHSOOPOVK	EXT 8	1917.8785	3	1.02E+08	P4; P13
S5.32	SOOOOYYYHSOPPOVK	EXT 8	1917.8785	3	1.40E+08	P12; P13
S5.33	SOOOOYYYHSOPOPVK	EXT 8	1917.8785	3	1.40E+08	P12; P14

Table S1: List of Identified Non-Cross-Linked Peptides: Corresponding mass spectrometry evidence for each identification is in the supplemental figure indicated. The peptide sequence is shown. The lower case "e" in S5.6 denotes methylation (+14.0157) of E. The calculated neutral mass of the precursor is given along with the observed charge states (z). XIC Area Sum: intensity of the monoisotopic precursor summed for all charge states observed. Area is calculated as the area under the curve (ion count intensities x seconds). Pro Sites: Position of a proline residue in the sequence.

Supplemental Table S2a: Missed Hydroxylated Prolines of the EXT8 Major Repetitive Peptide: SOOOOYYYHSOOOOVK

SFig	Non-Cross-Linked Peptides	Neut. Mass (Da)	Obs z	XIC Area Sum	Pro Sites
S5.23	SOOOOYYYHSOOOOVK	1949.8683	2; 3	1.01E+10	--
S5.24	SOPOOYYYHSOOOOVK	1933.8734	3	3.02E+08	P3
S5.25	SOOPOYYYHSOOOOVK	1933.8734	3	3.02E+08	P4
S5.26	SOOOOYYYHSOPOOVK	1933.8734	3	2.02E+09	P12
S5.27	SOOOOYYYHSOOPOVK	1933.8734	3	2.02E+09	P13
S5.28	SOOOOYYYHSOOOPVK	1933.8734	3	2.02E+09	P14
S5.29	SOPOOYYYHSOPOOVK	1917.8785	3	2.07E+08	P3; P12
S5.30	SOPOOYYYHSOOPOVK	1917.8785	3	2.07E+08	P3; P13
S5.31	SOOPOYYYHSOOPOVK	1917.8785	3	1.02E+08	P4; P13
S5.32	SOOOOYYYHSOPPOVK	1917.8785	3	1.40E+08	P12; P13
S5.33	SOOOOYYYHSOPOPVK	1917.8785	3	1.40E+08	P12; P14
SFig	**Cross-Linked Peptides**	**Neut. Mass (Da)**	**Obs z**	**XIC Area Sum**	**Pro Sites**
S1.5	SOOOOyYyHSOOOOVK	1947.8527	2; 3; 4	8.05E+11	--
S1.6	SPOOOyYyHSOOOOVK	1931.8578	3	1.71E+09	P2
S1.7	SOPOOyYyHSOOOOVK	1931.8578	3	5.47E+09	P3
S1.8	SOOPOyYyHSOOOOVK	1931.8578	3	2.10E+09	P4
S1.9	SOOOPyYyHSOOOOVK	1931.8578	3	2.35E+09	P5
S1.10	SOOOOyYyHSOPOOVK	1931.8578	3	1.02E+09	P12
S1.11	SOOOOyYyHSOOPOVK	1931.8578	3	1.36E+09	P13
S1.12	SOPPOyYyHSOOOOVK	1915.8628	3	3.22E+08	P3; P4
S1.13	SOPOPyYyHSOOOOVK	1915.8628	3	4.52E+08	P3; P5
S1.14	SOOPPyYyHSOOOOVK	1915.8628	3	6.45E+07	P4; P5
S1.16	SOOOOYYYHSOOOOVKSOOOOyYyHSOOOOVK	3879.7104	5	4.64E+07	--
S3.9	SOOOOyYyHSOOOOVKSOOOOyYyHSOOOOVK	3877.6948	4; 5	5.71E+08	--
S1.15	SOOOOVKSOOOOyYyHSOOOOVK	2714.2388	4	1.16E+07	--

Table S2a: List of Identified Missed Hydroxylated Prolines of the EXT8 Major Repetitive Peptide: Corresponding mass spectrometry evidence for each identification is in the

supplemental figure indicated. The peptide sequence is shown with the tyrosine involved in the linkage noted by the lowercase y. The calculated neutral mass of the precursor is given along with the observed charge states (z). XIC Area Sum: intensity of the monoisotopic precursor summed for all charge states observed. Area is calculated as the area under the curve (ion count intensities x seconds). Proline Sites: Position of a proline residue in the sequence.

Supplemental Table S2b: Missed Hydroxylated Prolines of the EXT3 Major Repetitive Peptide: HYSOOOOVYHSOOOOK

SFig.	Peptide	Neut. Mass (Da)	Obs z	XIC Area Sum	Pro Sites
S5.13	HYSOOOVYHSOOOOK	1810.8162	3; 4	1.09E+09	
S5.14	HYSOPOVYHSOOOOK	1794.8213	3; 4	1.60E+07	P5
S5.15	HYSOOOVYHSOPOOK	1794.8213	3; 4	1.07E+08	P12
S5.16	HYSOOOVYHSOOPOK	1794.8213	4	2.35E+07	P13
S5.17	HYSOOOVYHSOOOPK	1794.8213	3; 4	4.44E+08	P14
S5.18	HYSOOOVYHSOOPPK	1778.8264	3	3.15E+07	P13; P14
S5.19	HYSOOOVYHSOOOOKK	1938.9112	3; 4	3.94E+07	
S5.20	HYSOOOVYHSOOPOKK	1922.9163	4	9.10E+06	P13
S5.21	HYSOOOVYHSOOOPKK	1922.9163	4	9.10E+06	P14
S5.22	HYSOOOVYHSOOOOKEK	2067.9538	4	3.94E+06	

Table S2b: List of Identified Missed Hydroxylated Prolines of the EXT3 Major Repetitive Peptide: Corresponding mass spectrometry evidence for each identification is in the supplemental figure indicated. The peptide sequence is shown with the tyrosine involved in the linkage noted by the lowercase y. The calculated neutral mass of the precursor is given along with the observed charge states (z). XIC Area Sum: intensity of the monoisotopic precursor summed for all charge states observed. Area is calculated as the area under the curve (ion count intensities x seconds). Proline Sites: Position of a proline residue in the sequence.

APPENDIX B

Supplemental Figures S1.1 to S1.16: MS Evidence for IDT Cross-Linked Peptides

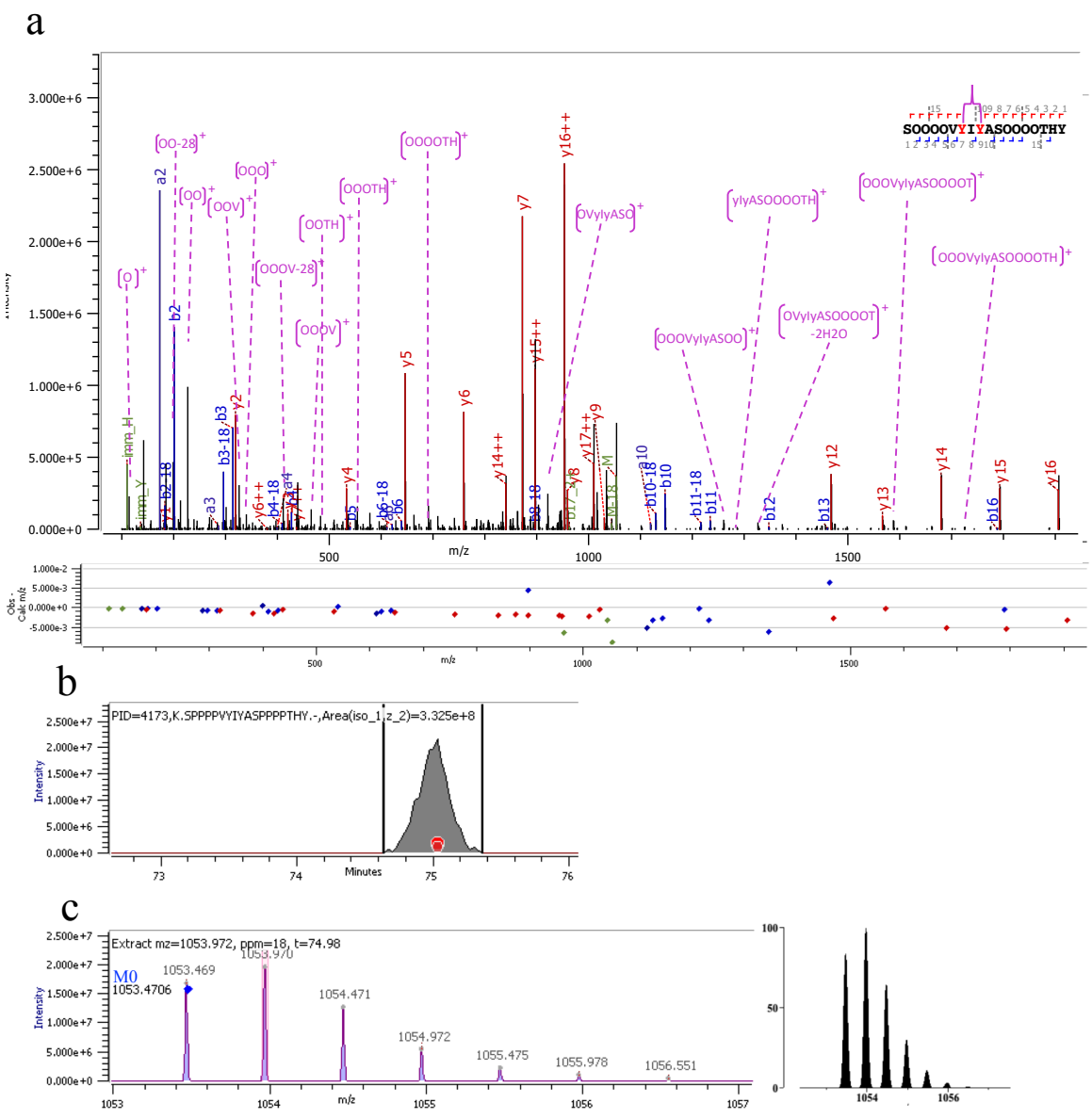

Figure S1.1: Characterization of the intramolecular IDT cross-linked peptide from EXT 8. (a) The annotated HCD MS/MS spectrum of the [M+2H]2+ ion at m/z 1053.4706 with m/z errors of mass fragments at a 10ppm tolerance denoted below; and (inset, top right) a-, b-, & y-fragment ion coverage. Cross-linked tyrosines carrying the mass loss of (2.0156 Da) are denoted in red. (b) Extracted ion chromatogram. The red dots indicate several MS2 collected. (c) (Left) The isotopic profile of the [M+2H]2+ precursor ion at m/z 1053.4706 that was isolated for fragmentation. The blue dot indicates correct monoisotopic assignment based on calculated mass. (Right) The theoretically calculated isotopic profile of the [M+2H]2+ precursor ion at m/z 1053.4706 from MS-Isotope.[68]

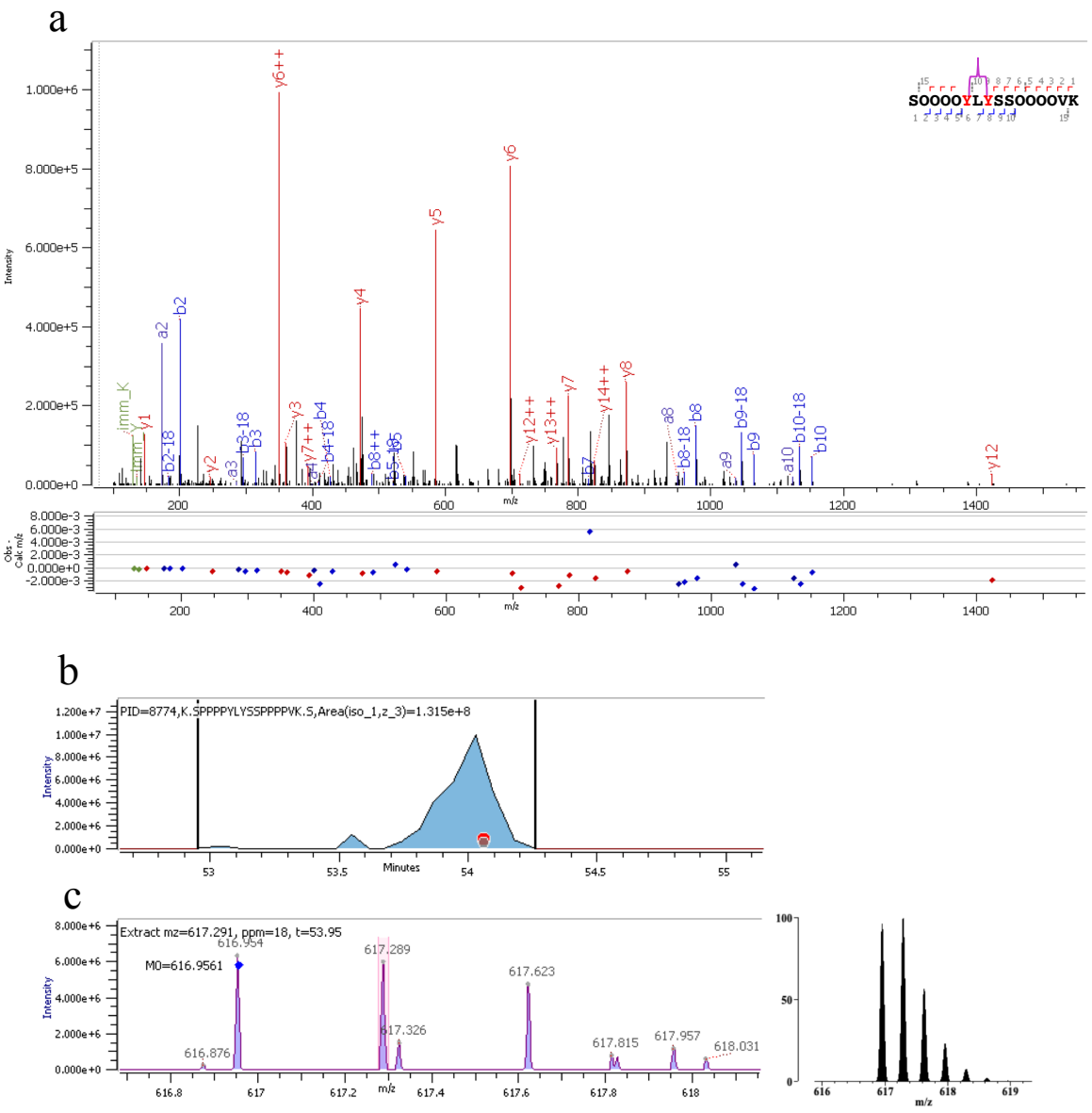

Figure S1.2: Characterization of the intramolecular IDT cross-linked peptide from EXT 8. (a) The annotated HCD MS/MS spectrum of the [M+3H]3+ ion at m/z 616.9561 with m/z errors of mass fragments at a 10ppm tolerance denoted below; and (inset, top right) a-, b-, & y-fragment ion coverage. Cross-linked tyrosines carrying the mass loss of (2.0156 Da) are denoted in red. (b) Extracted ion chromatogram. The red dots indicate several MS2 collected. (c) (Left) The isotopic profile of the [M+3H]3+ precursor ion at m/z 616.9561 that was isolated for fragmentation. The blue dot indicates correct monoisotopic assignment based on calculated mass. (Right) The

theoretically calculated isotopic profile of the [M+3H]3+ precursor ion at m/z 616.9561 from MS-Isotope.[68]

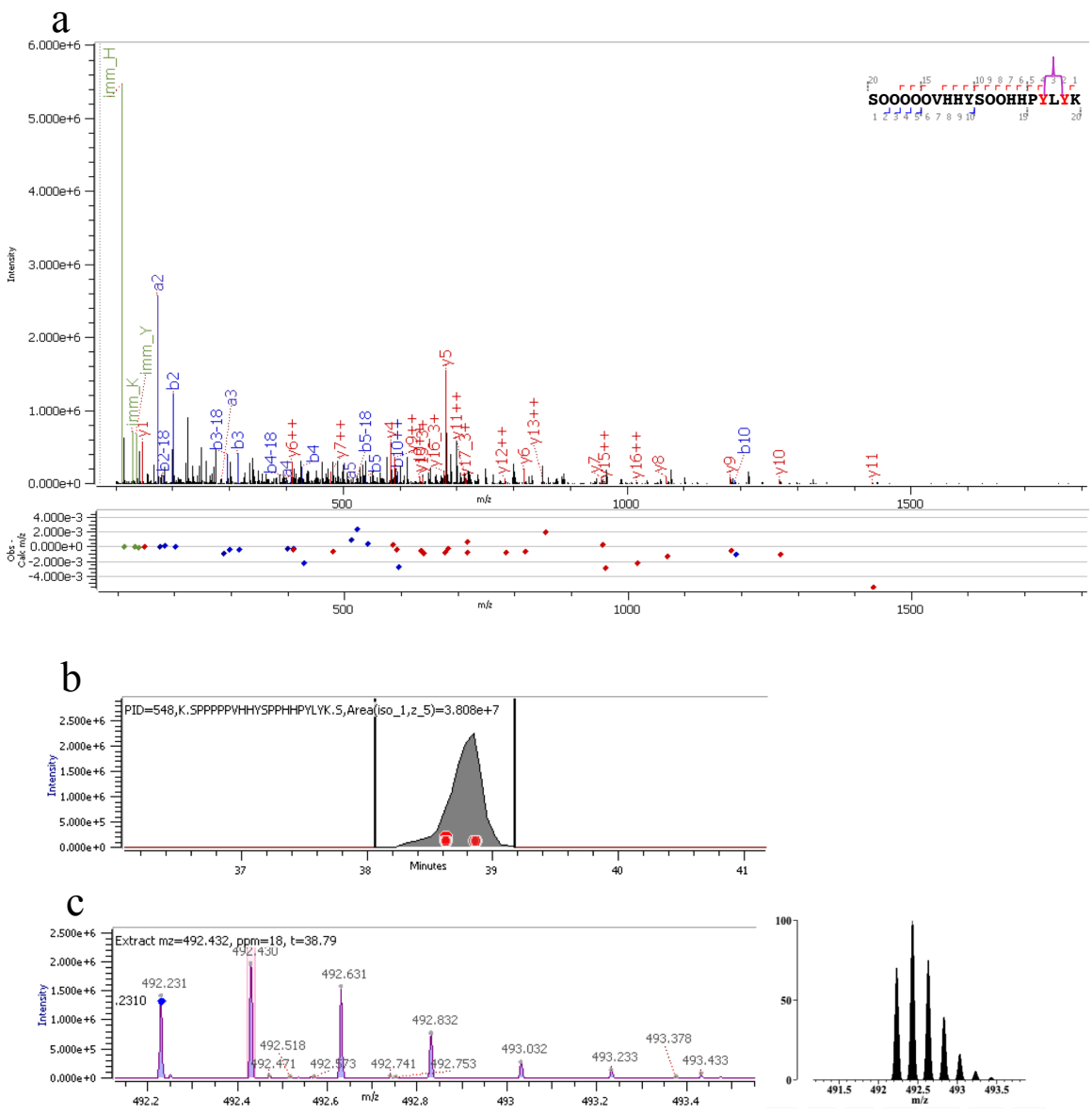

Figure S1.3: Characterization of the intramolecular IDT cross-linked peptide from EXT 3. (a) The annotated HCD MS/MS spectrum of the [M+5H]5+ ion at m/z 492.231 with m/z errors of mass fragments at a 10ppm tolerance denoted below; and (inset, top right) a-, b-, & y-fragment ion coverage. Cross-linked tyrosines carrying the mass loss of (2.0156 Da) are denoted in red. (b) Extracted ion chromatogram. The red dots indicate several MS2 collected. (c) (Left) The isotopic profile of the [M+5H]5+ precursor ion at m/z 492.231 that was isolated for fragmentation. The blue dot indicates correct monoisotopic assignment based on calculated mass. (Right) The

theoretically calculated isotopic profile of the [M+5H]5+ precursor ion at m/z 492.231 from MS-Isotope.[68]

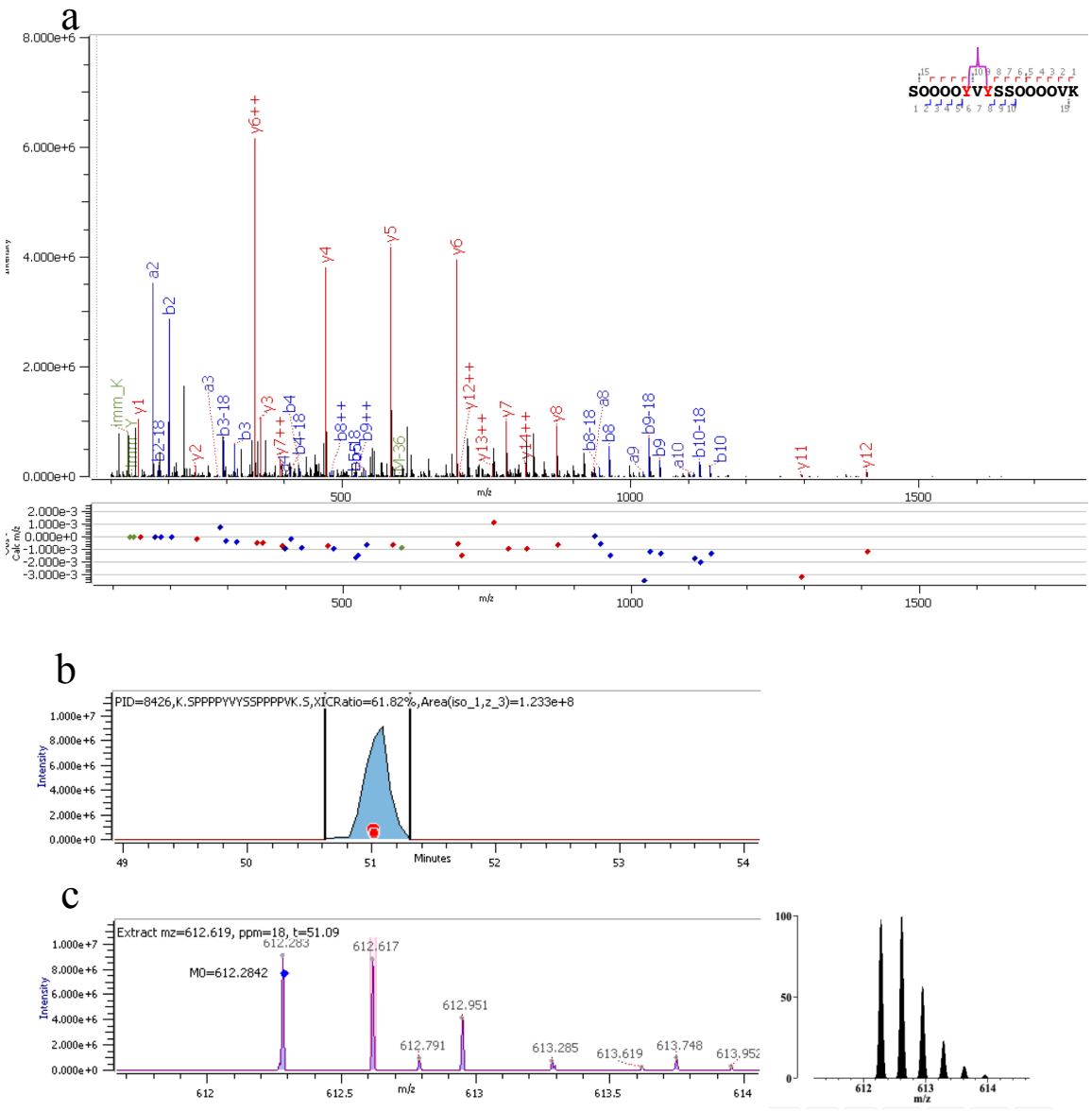

Figure S1.4: Characterization of the intramolecular IDT cross-linked peptide from EXT 8. (a) The annotated HCD MS/MS spectrum of the [M+3H]3+ ion at m/z 612.2842 with m/z errors of mass fragments at a 10ppm tolerance denoted below; and (inset, top right) a-, b-, & y-fragment ion coverage. Cross-linked tyrosines carrying the mass loss of (2.0156 Da) are denoted in red. (b) Extracted ion chromatogram. The red dots indicate several MS2 collected. (c) (Left) The isotopic profile of the [M+3H]3+ precursor ion at m/z 612.2842 that was isolated for fragmentation. The blue dot indicates correct monoisotopic assignment based on calculated mass. (Right) The

theoretically calculated isotopic profile of the [M+3H]3+ precursor ion at m/z 612.2842 from MS-Isotope.[68]

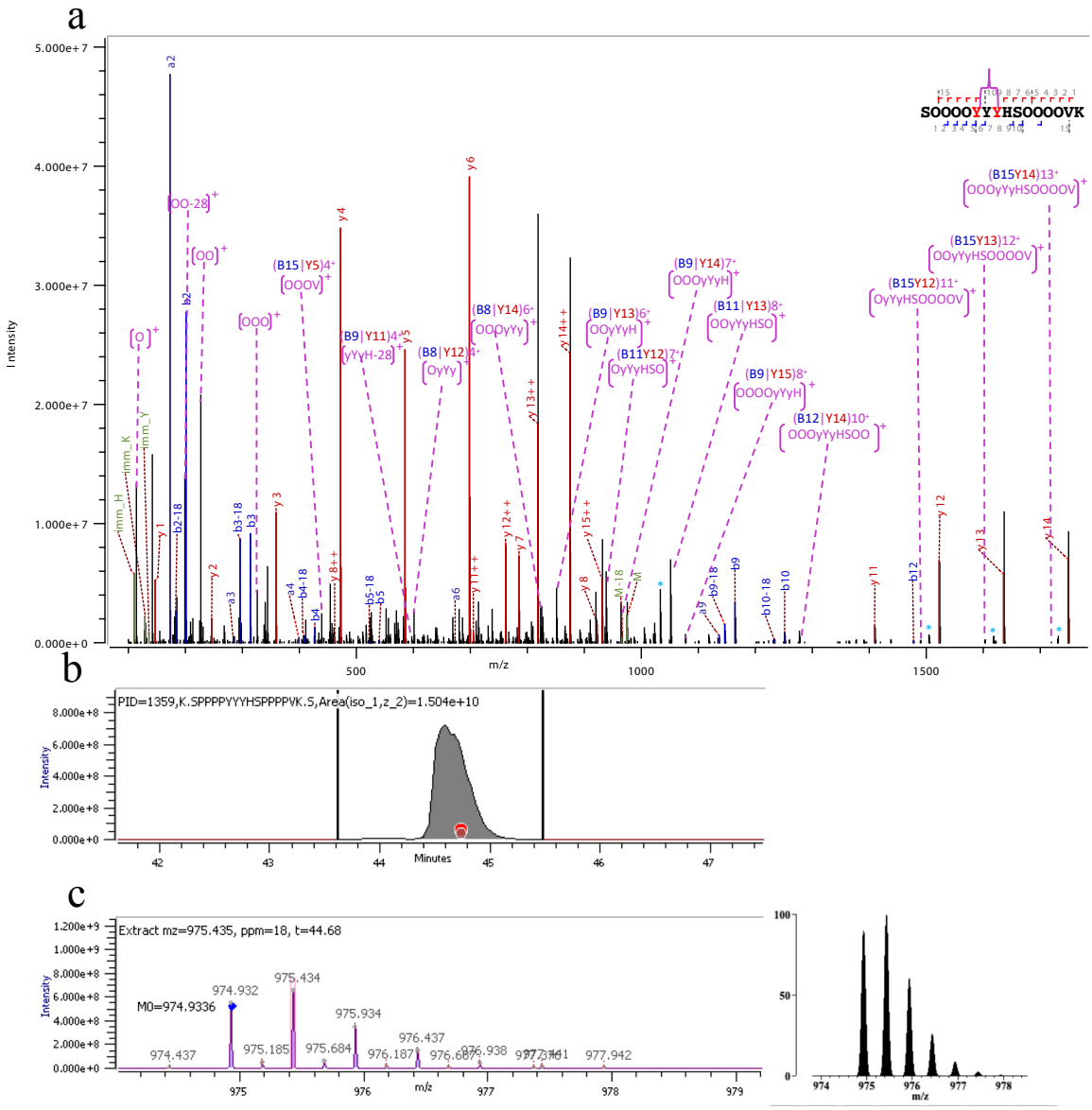

Figure S1.5: Characterization of the intramolecular IDT cross-linked peptide from EXT 8. (a) The annotated HCD MS/MS spectrum of the [M+2H]2+ ion at m/z 974.9336 with m/z errors of mass fragments at a 10ppm tolerance denoted below; and (inset, top right) a-, b-, & y-fragment ion coverage. Cross-linked tyrosines carrying the mass loss of (2.0156 Da) are denoted in red. (b) Extracted ion chromatogram. The red dots indicate several MS2 collected. (c) (Left) The isotopic profile of the [M+2H]2+ precursor ion at m/z 974.9336 that was isolated for fragmentation. The blue dot indicates correct monoisotopic assignment based on calculated mass. (Right) The

theoretically calculated isotopic profile of the [M+2H]2+ precursor ion at m/z 974.9336 from MS-Isotope.[68]

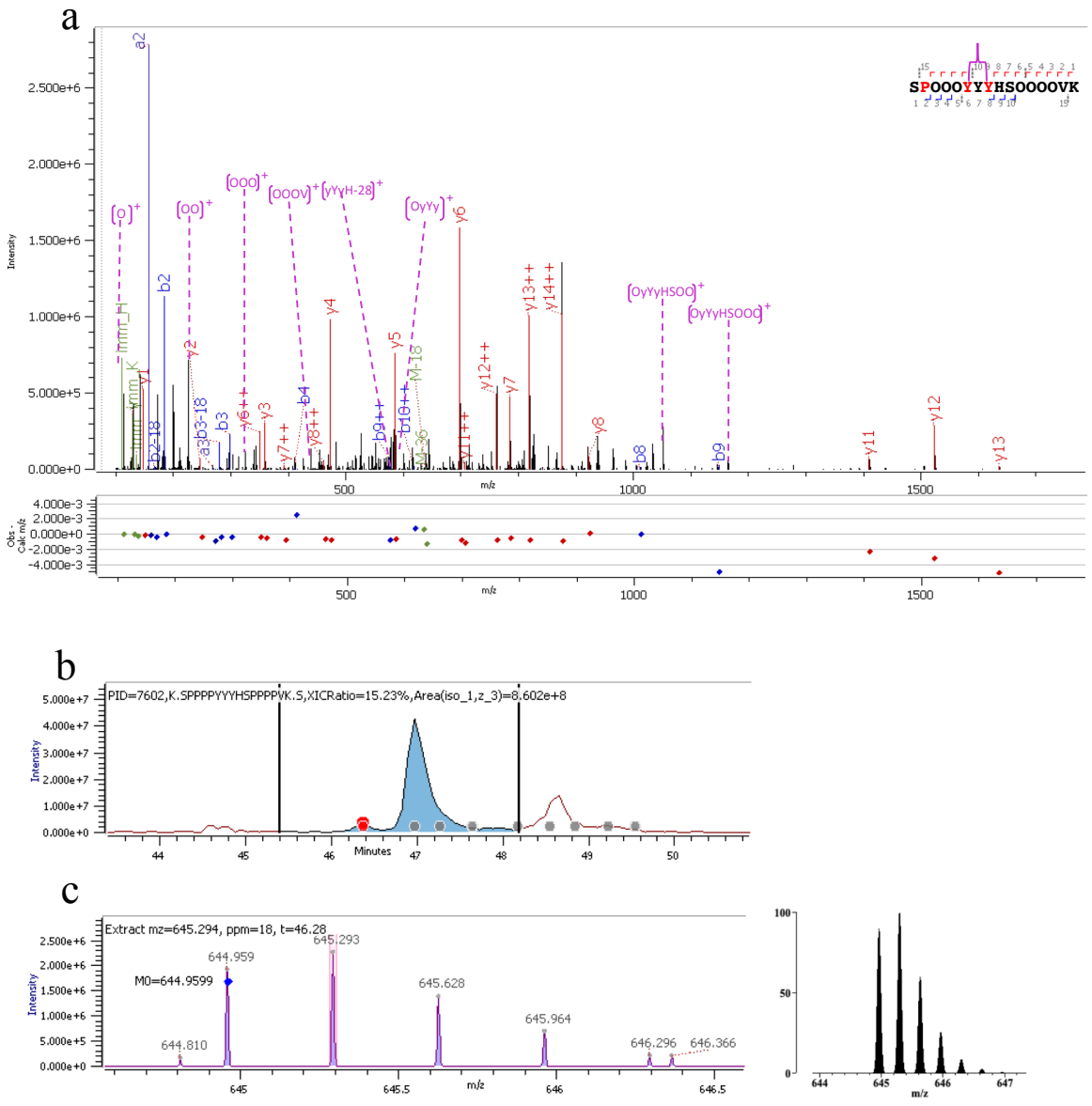

Figure S1.6: Characterization of the intramolecular IDT cross-linked peptide from EXT 8 with one missed hydroxylation at residue P2 as indicated by the b2 ion. (a) The annotated HCD MS/MS spectrum of the [M+3H]3+ ion at m/z 644.9599 with m/z errors of mass fragments at a 10ppm tolerance denoted below; and (inset, top right) a-, b-, & y-fragment ion coverage. Cross-linked tyrosines carrying the mass loss of (2.0156 Da) are denoted in red. (b) Extracted ion chromatogram. The red dots indicate several MS2 with the same identification were collected. The grey dots indicate the time point of a MS2 of the same mass, but with a different ID such as a

different site of proline. (c) (Left) The isotopic profile of the [M+3H]3+ precursor ion at m/z 644.9599 that was isolated for fragmentation. The blue dot indicates correct monoisotopic assignment based on calculated mass. (Right) The theoretically calculated isotopic profile of the [M+3H]3+ precursor ion at m/z 644.9599 from MS-Isotope.[68]

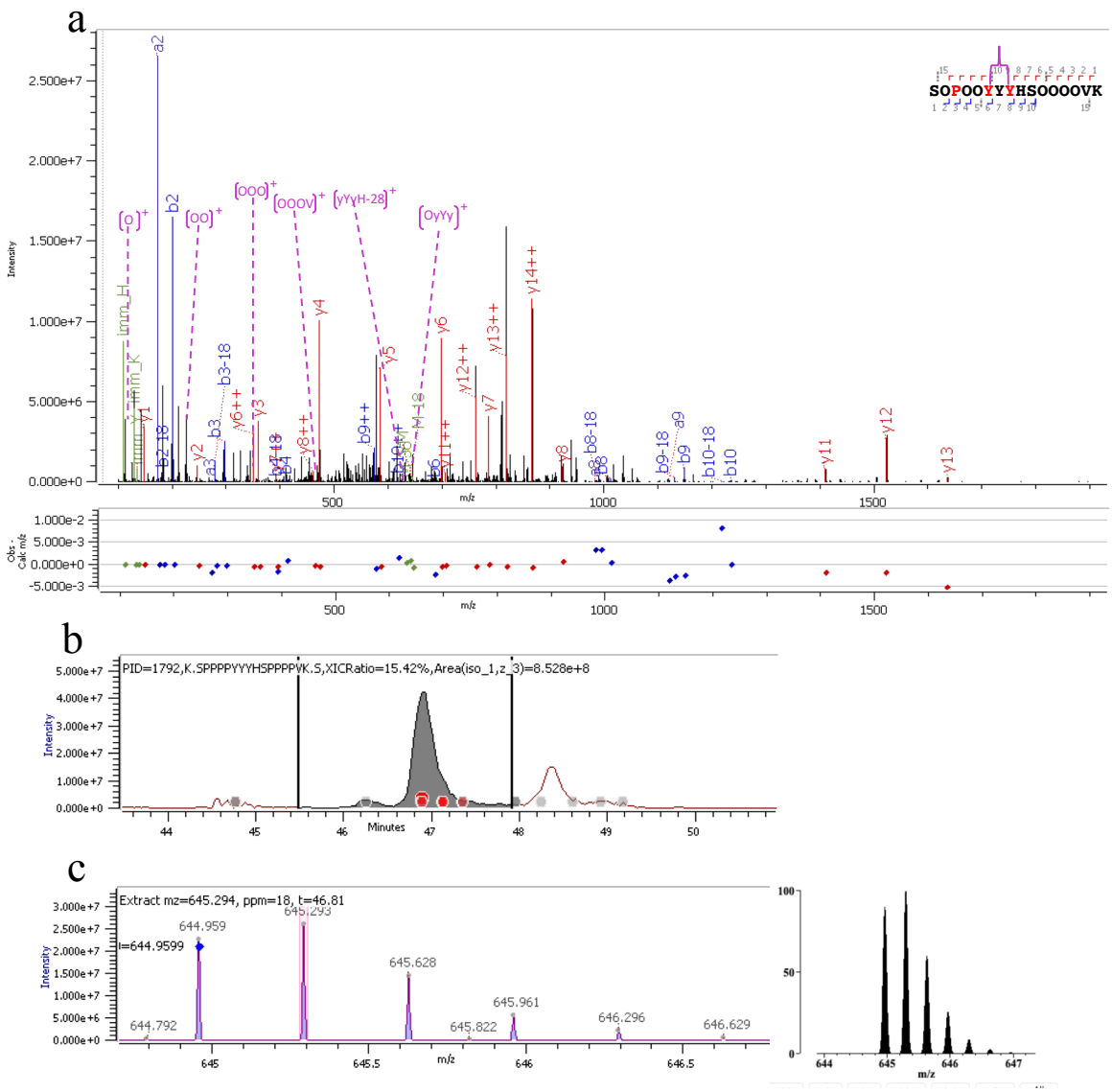

Figure S1.7: Characterization of the intramolecular IDT cross-linked peptide from EXT 8 with one missed hydroxylation at residue P3 as indicated by the b3 ion. (a) The annotated HCD MS/MS spectrum of the [M+3H]3+ ion at m/z 644.9599 with m/z errors of mass fragments at a 10ppm tolerance denoted below; and (inset, top right) a-, b-, & y-fragment ion coverage. Cross-linked tyrosines carrying the mass loss of (2.0156 Da) are denoted in red. (b) Extracted ion chromatogram. The red dots indicate several MS2 with the same identification were collected. The grey dots indicate the time point of a MS2 of the same mass, but with a different ID such as a

different site of proline. (c) (Left) The isotopic profile of the [M+3H]3+ precursor ion at m/z 644.9599 that was isolated for fragmentation. The blue dot indicates correct monoisotopic assignment based on calculated mass. (Right) The theoretically calculated isotopic profile of the [M+3H]3+ precursor ion at m/z 644.9599 from MS-Isotope.[68]

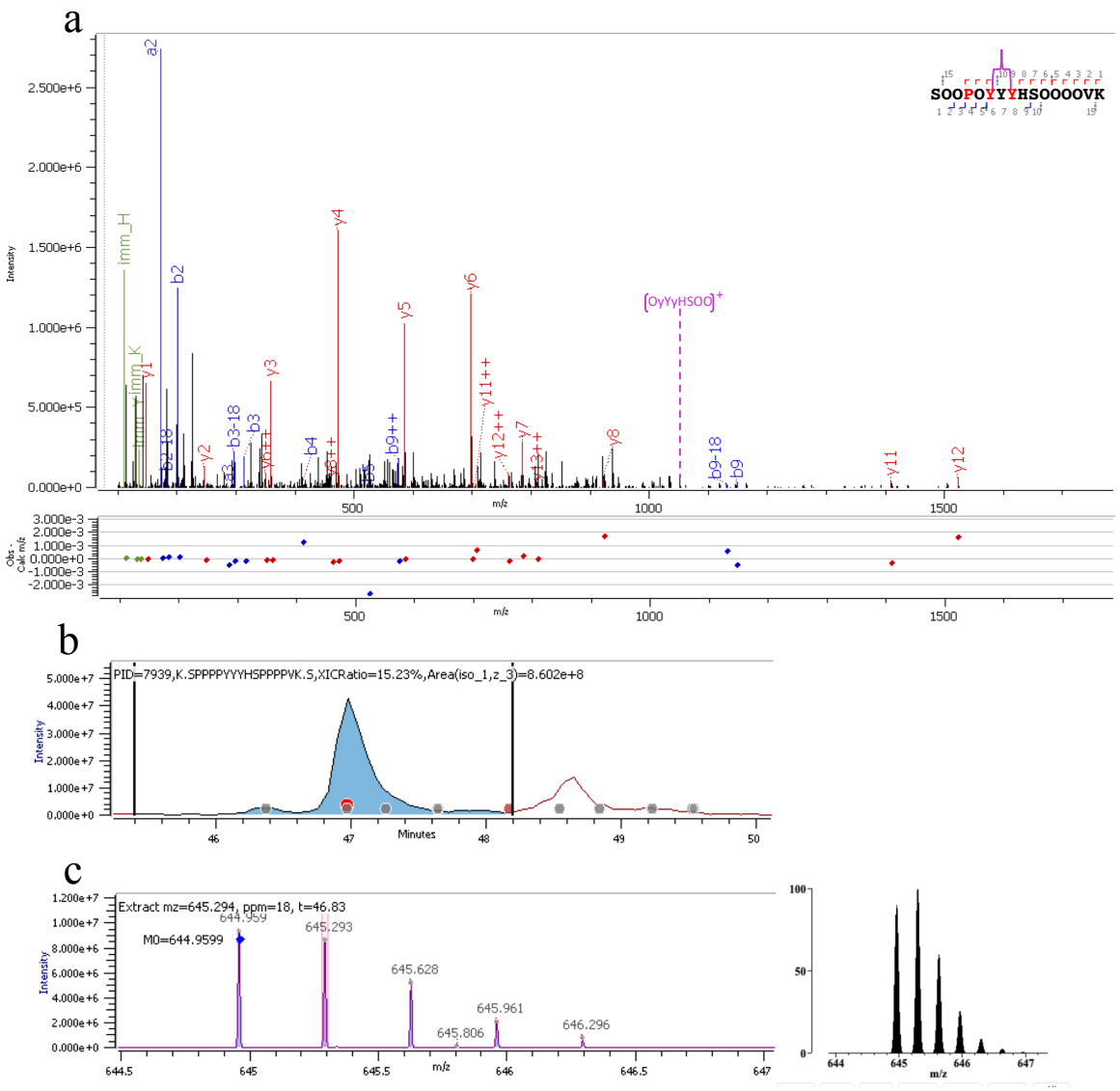

Figure S1.8: Characterization of the intramolecular IDT cross-linked peptide from EXT 8 with one missed hydroxylation at residue P4 as indicated by the b4 ion. (a) The annotated HCD MS/MS spectrum of the [M+3H]3+ ion at m/z 644.9599 with m/z errors of mass fragments at a 10ppm tolerance denoted below; and (inset, top right) a-, b-, & y-fragment ion coverage. Cross-linked tyrosines carrying the mass loss of (2.0156 Da) are denoted in red. (b) Extracted ion chromatogram. The red dots indicate several MS2 with the same identification were collected. The grey dots indicate the time point of a MS2 of the same mass, but with a different ID such as a

different site of proline. (c) (Left) The isotopic profile of the [M+3H]3+ precursor ion at m/z 644.9599 that was isolated for fragmentation. The blue dot indicates correct monoisotopic assignment based on calculated mass. (Right) The theoretically calculated isotopic profile of the [M+3H]3+ precursor ion at m/z 644.9599 from MS-Isotope.[68]

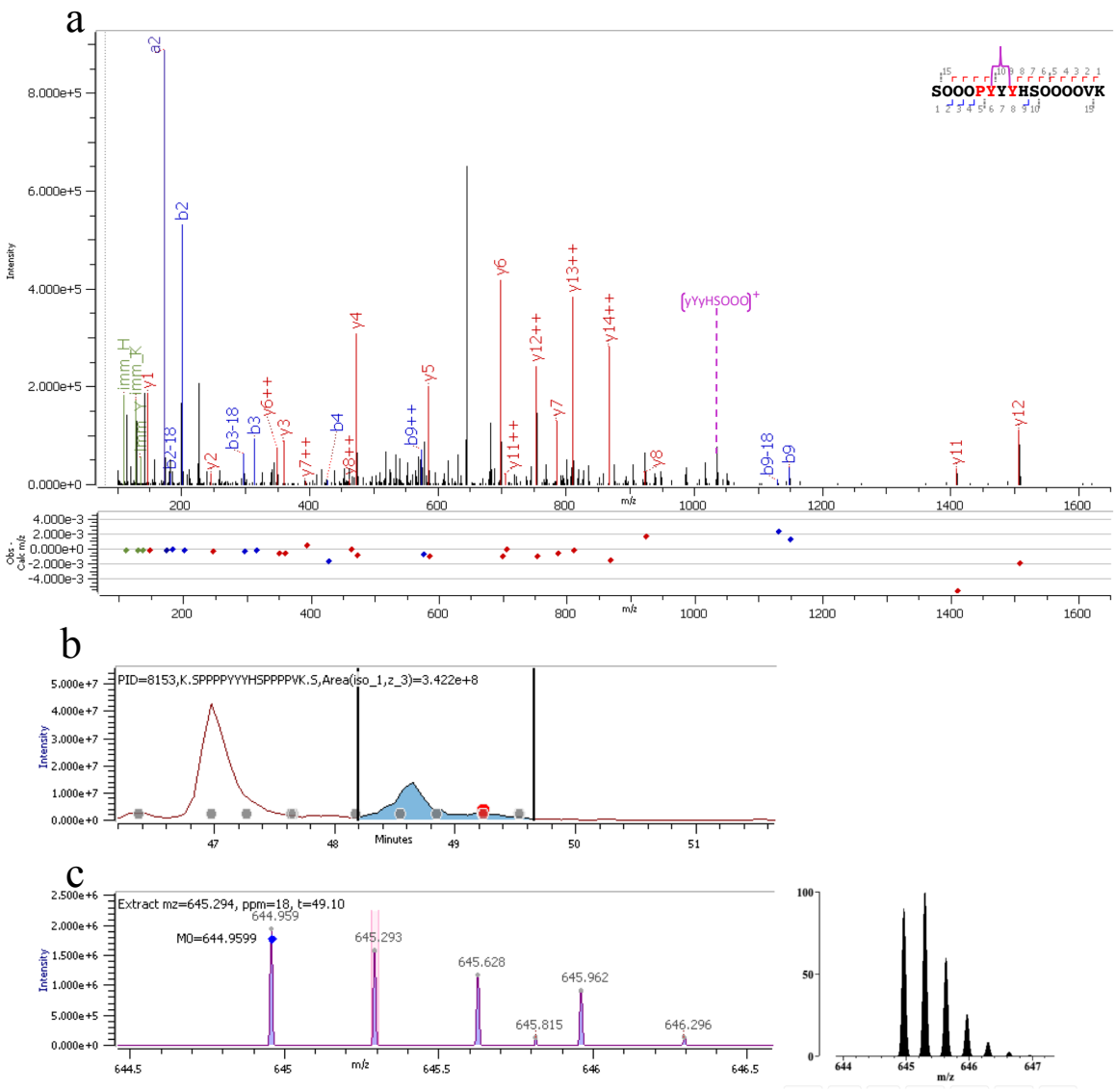

Figure S1.9: Characterization of the intramolecular IDT cross-linked peptide from EXT 8 with one missed hydroxylation at residue P5 as indicated by the y12 and y12++ ions. (a) The annotated HCD MS/MS spectrum of the [M+3H]3+ ion at m/z 644.9599 with m/z errors of mass fragments at a 10ppm tolerance denoted below; and (inset, top right) a-, b-, & y-fragment ion coverage. Cross-linked tyrosines carrying the mass loss of (2.0156 Da) are denoted in red. (b) Extracted ion chromatogram. The red dots indicate several MS2 with the same identification were collected. The grey dots indicate the time point of a MS2 of the same mass, but with a different ID such as a

different site of proline. (c) (Left) The isotopic profile of the [M+3H]3+ precursor ion at m/z 644.9599 that was isolated for fragmentation. The blue dot indicates correct monoisotopic assignment based on calculated mass. (Right) The theoretically calculated isotopic profile of the [M+3H]3+ precursor ion at m/z 644.9599 from MS-Isotope.[68]

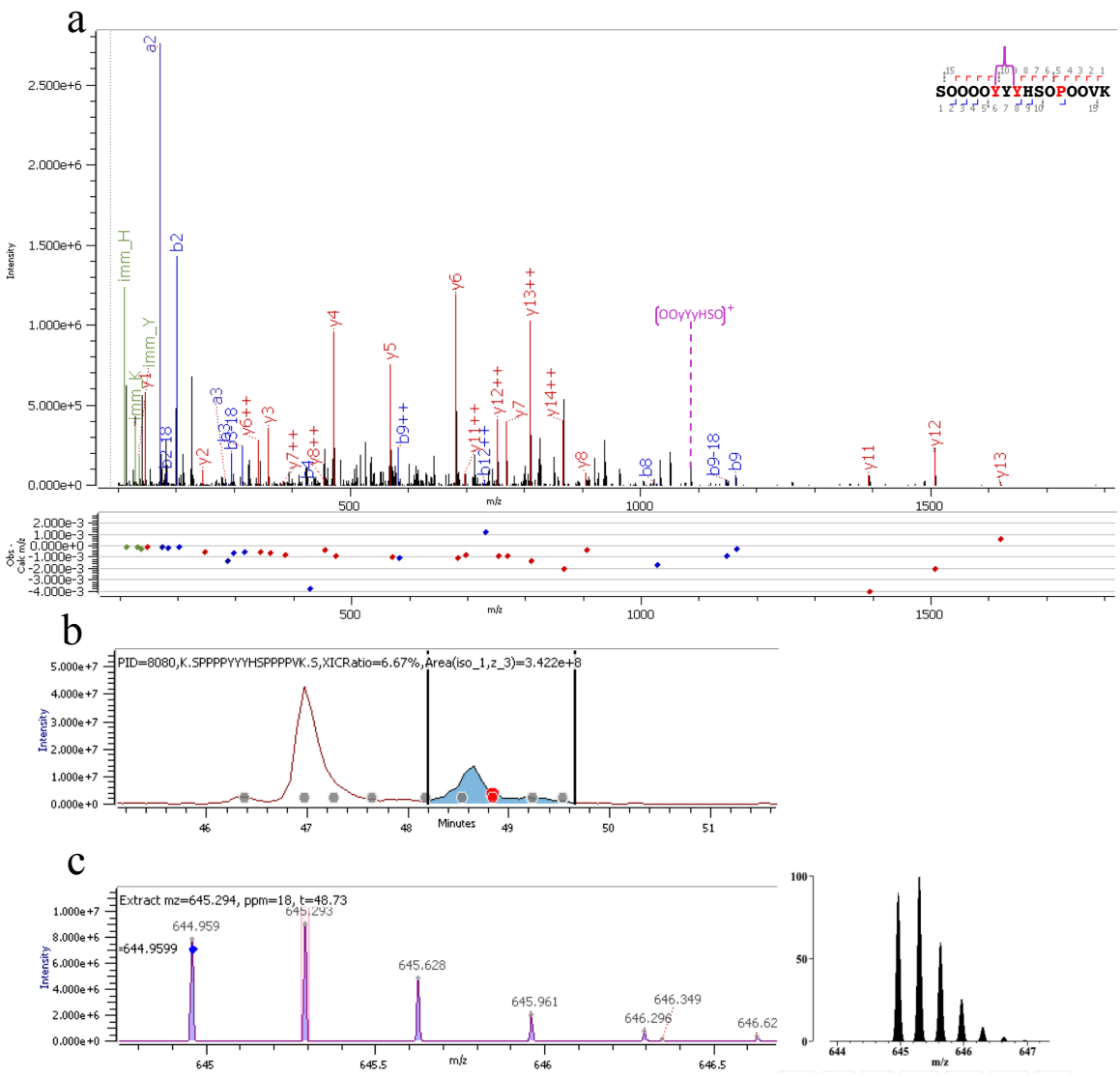

Figure S1.10: Characterization of the intramolecular IDT cross-linked peptide from EXT 8 with one missed hydroxylation at residue P12 as indicated by the b12++ and y5 ions. (a) The annotated HCD MS/MS spectrum of the [M+3H]3+ ion at m/z 644.9599 with m/z errors of mass fragments at a 10ppm tolerance denoted below; and (inset, top right) a-, b-, & y-fragment ion coverage. Cross-linked tyrosines carrying the mass loss of (2.0156 Da) are denoted in red. (b) Extracted ion chromatogram. The red dots indicate several MS2 with the same identification were collected. The grey dots indicate the time point of a MS2 of the same mass, but with a different

ID such as a different site of proline. (c) (Left) The isotopic profile of the [M+3H]3+ precursor ion at m/z 644.9599 that was isolated for fragmentation. The blue dot indicates correct monoisotopic assignment based on calculated mass. (Right) The theoretically calculated isotopic profile of the [M+3H]3+ precursor ion at m/z 644.9599 from MS-Isotope.[68]

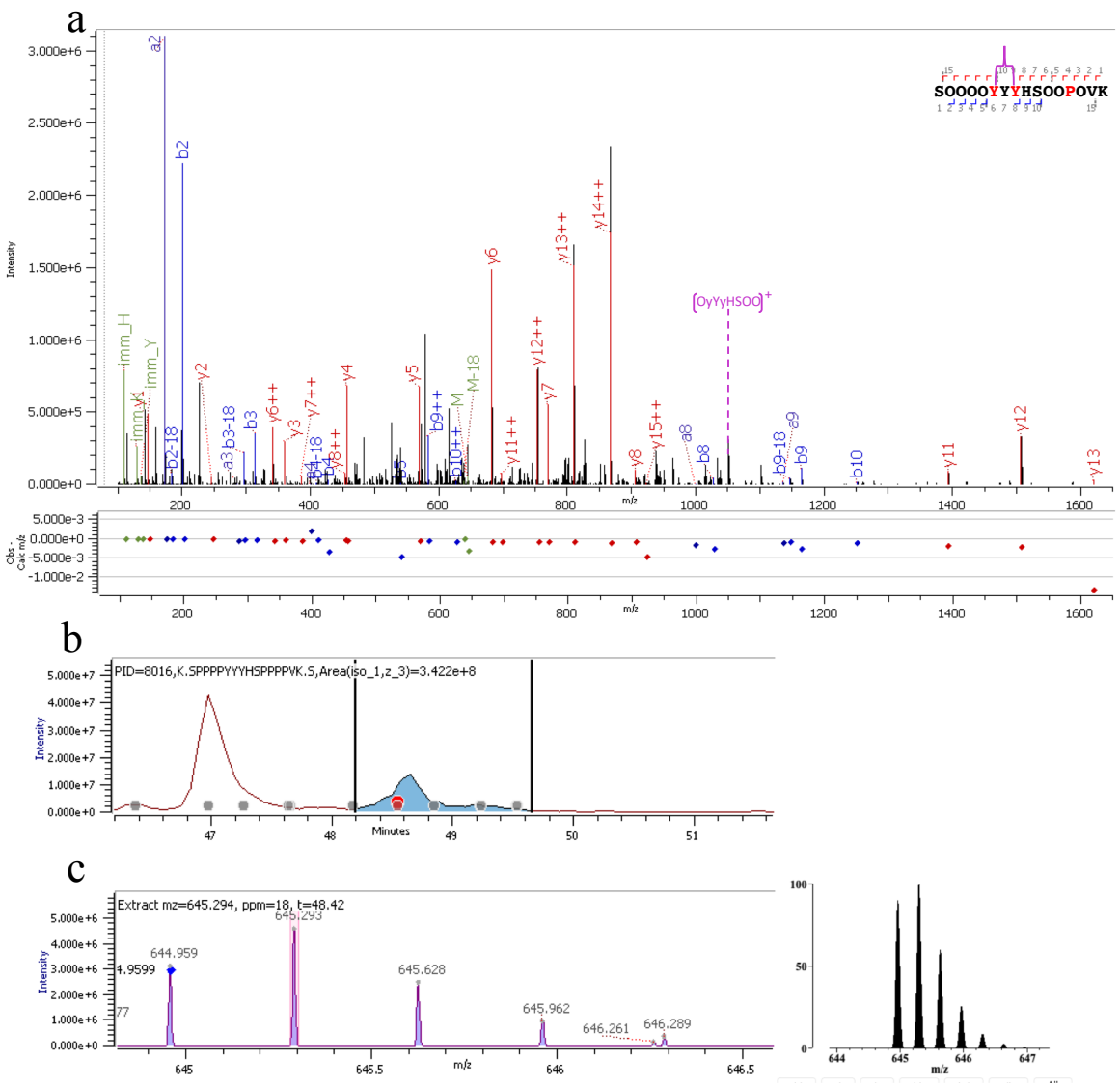

Figure S1.11: Characterization of the intramolecular IDT cross-linked peptide from EXT 8 with one missed hydroxylation at residue P13 as indicated by the y4 ion. (a) The annotated HCD MS/MS spectrum of the [M+3H]3+ ion at m/z 644.9599 with m/z errors of mass fragments at a 10ppm tolerance denoted below; and (inset, top right) a-, b-, & y-fragment ion coverage. Cross-linked tyrosines carrying the mass loss of (2.0156 Da) are denoted in red. (b) Extracted ion chromatogram. The red dots indicate several MS2 with the same identification were collected. The grey dots indicate the time point of a MS2 of the same mass, but with a different ID such as a

different site of proline. (c) (Left) The isotopic profile of the [M+3H]3+ precursor ion at m/z 644.9599 that was isolated for fragmentation. The blue dot indicates correct monoisotopic assignment based on calculated mass. (Right) The theoretically calculated isotopic profile of the [M+3H]3+ precursor ion at m/z 644.9599 from MS-Isotope.[68]

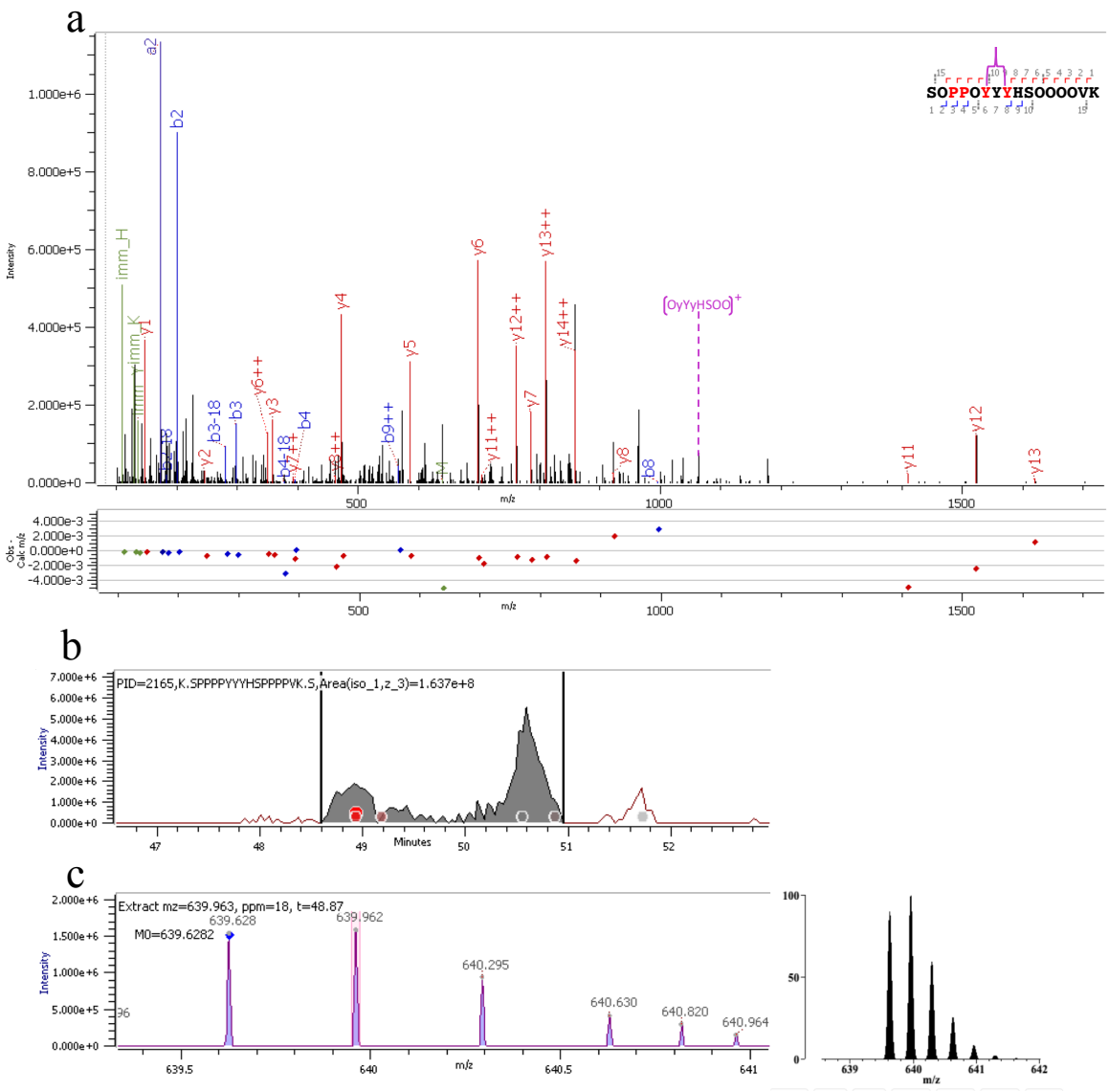

Figure S1.12: Characterization of the intramolecular IDT cross-linked peptide from EXT 8 with two missed hydroxylations at residues P3 and P4 as indicated by the y12, y13, y14, b3, and b4 ions. (a) The annotated HCD MS/MS spectrum of the [M+3H]3+ ion at m/z 639.6282 with m/z errors of mass fragments at a 10ppm tolerance denoted below; and (inset, top right) a-, b-, & y-fragment ion coverage. Cross-linked tyrosines carrying the mass loss of (2.0156 Da) are denoted in red. (b) Extracted ion chromatogram. The red dots indicate several MS2 collected. The grey dots indicate the time point of a MS2 of the same mass, but with a different ID such as a different

site of proline. (c) (Left) The isotopic profile of the [M+3H]3+ precursor ion at m/z 639.6282 that was isolated for fragmentation. The blue dot indicates correct monoisotopic assignment based on calculated mass. (Right) The theoretically calculated isotopic profile of the [M+3H]3+ precursor ion at m/z 639.6282 from MS-Isotope.[68]

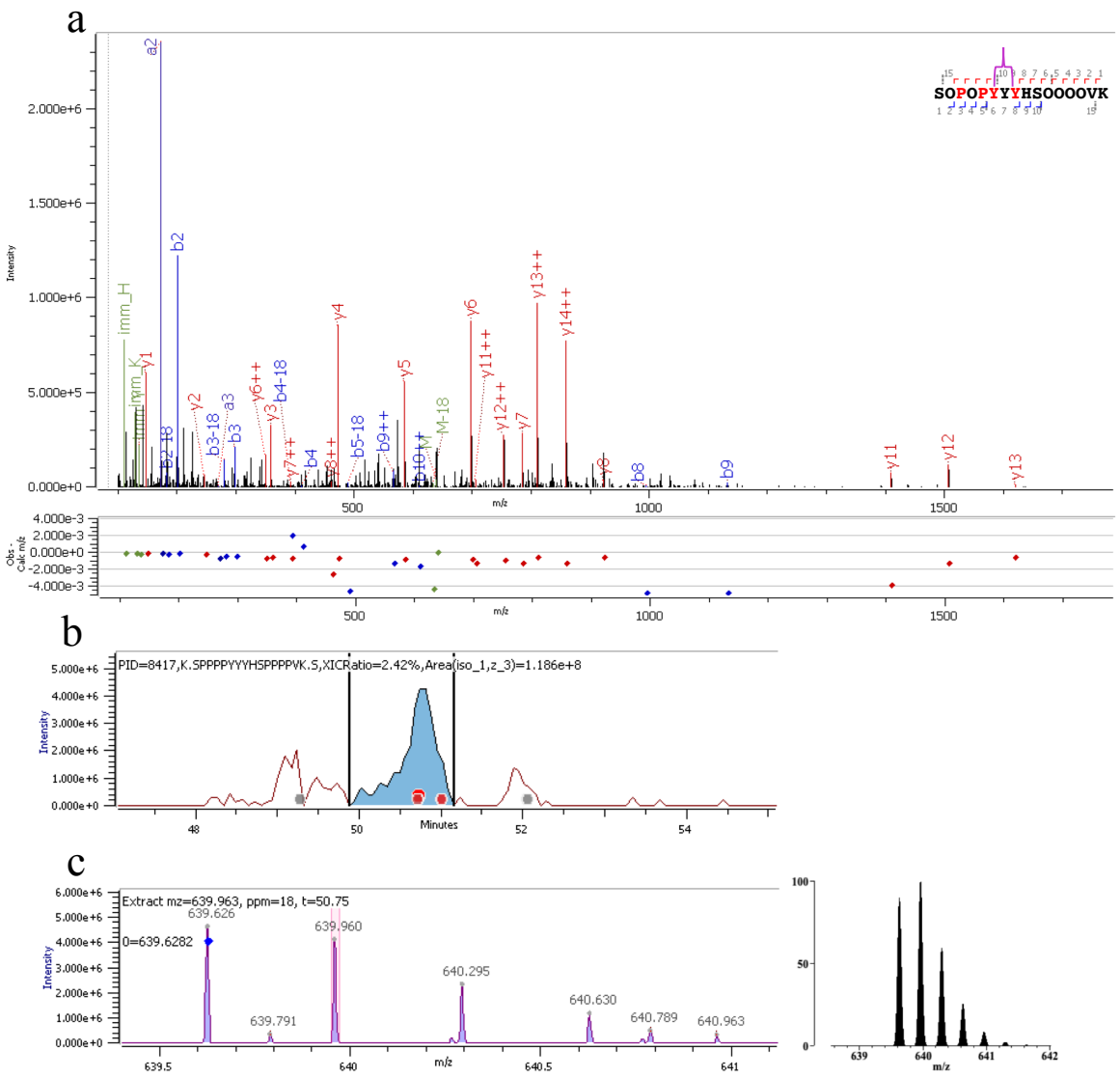

Figure S1.13: Characterization of the intramolecular IDT cross-linked peptide from EXT 8 with two missed hydroxylations at residues P3 and P5 as indicated by the y12, y13, y14, b3, b4 and b5 ions. (a) The annotated HCD MS/MS spectrum of the [M+3H]3+ ion at m/z 639.6282 with m/z errors of mass fragments at a 10ppm tolerance denoted below; and (inset, top right) a-, b-, & y-fragment ion coverage. Cross-linked tyrosines carrying the mass loss of (2.0156 Da) are denoted in red. (b) Extracted ion chromatogram. The red dots indicate several MS2 collected. The grey dots indicate the time point of a MS2 of the same mass, but with a different ID such as a different

site of proline. (c) (Left) The isotopic profile of the [M+3H]3+ precursor ion at m/z 639.6282 that was isolated for fragmentation. The blue dot indicates correct monoisotopic assignment based on calculated mass. (Right) The theoretically calculated isotopic profile of the [M+3H]3+ precursor ion at m/z 639.6282 from MS-Isotope.[68]

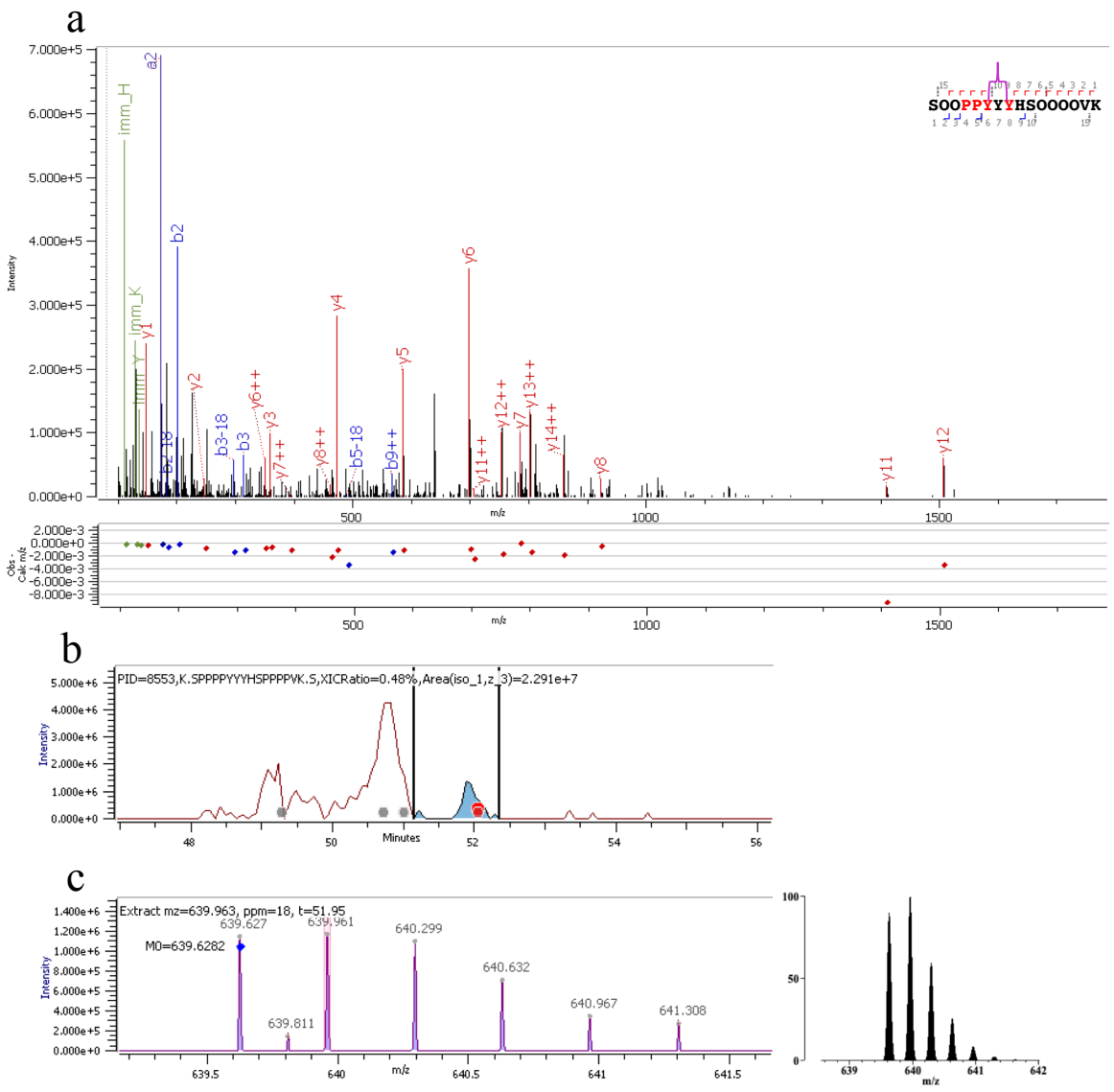

Figure S1.14: Characterization of the intramolecular IDT cross-linked peptide from EXT 8 with two missed hydroxylations at residues P4 and P5 as indicated by the y11, y12, y13, b2, b3, and b5 ions. (a) The annotated HCD MS/MS spectrum of the [M+3H]3+ ion at m/z 639.6282 with m/z errors of mass fragments at a 10ppm tolerance denoted below; and (inset, top right) a-, b-, & y-fragment ion coverage. Cross-linked tyrosines carrying the mass loss of (2.0156 Da) are denoted in red. (b) Extracted ion chromatogram. The red dots indicate several MS2 collected. The grey dots indicate the time point of a MS2 of the same mass, but with a different ID such as a different

site of proline. (c) (Left) The isotopic profile of the [M+3H]3+ precursor ion at m/z 639.6282 that was isolated for fragmentation. The blue dot indicates correct monoisotopic assignment based on calculated mass. (Right) The theoretically calculated isotopic profile of the [M+3H]3+ precursor ion at m/z 639.6282 from MS-Isotope.[68]

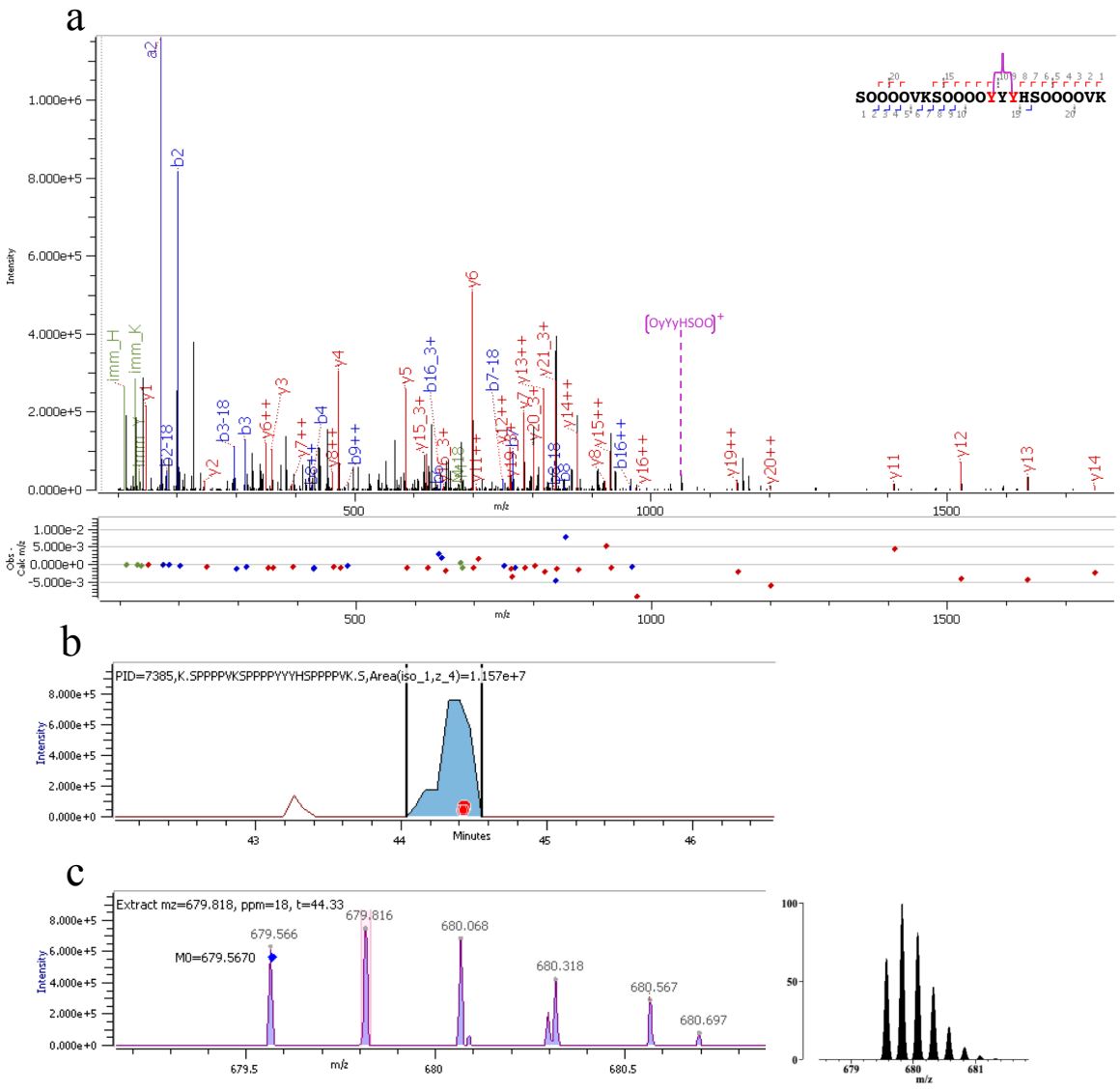

Figure S1.15: Characterization of the intramolecular IDT cross-linked peptide from EXT 8 with one missed tryptic cleavage. (a) The annotated HCD MS/MS spectrum of the [M+4H]4+ ion at m/z 679.567 with m/z errors of mass fragments at a 10ppm tolerance denoted below; and (inset, top right) a-, b-, & y-fragment ion coverage. Cross-linked tyrosines carrying the mass loss of (2.0156 Da) are denoted in red. (b) Extracted ion chromatogram. The red dots indicate several MS2 collected. (c) (Left) The isotopic profile of the [M+4H]4+ precursor ion at m/z 679.567 that was isolated for fragmentation. The blue dot indicates correct monoisotopic assignment based on

calculated mass. (Right) The theoretically calculated isotopic profile of the [M+4H]4+ precursor ion at m/z 679.567 from MS-Isotope.[68]

Supplemental Figures S2.1: MS Evidence for the Dityrosine Cross-Linked Peptide

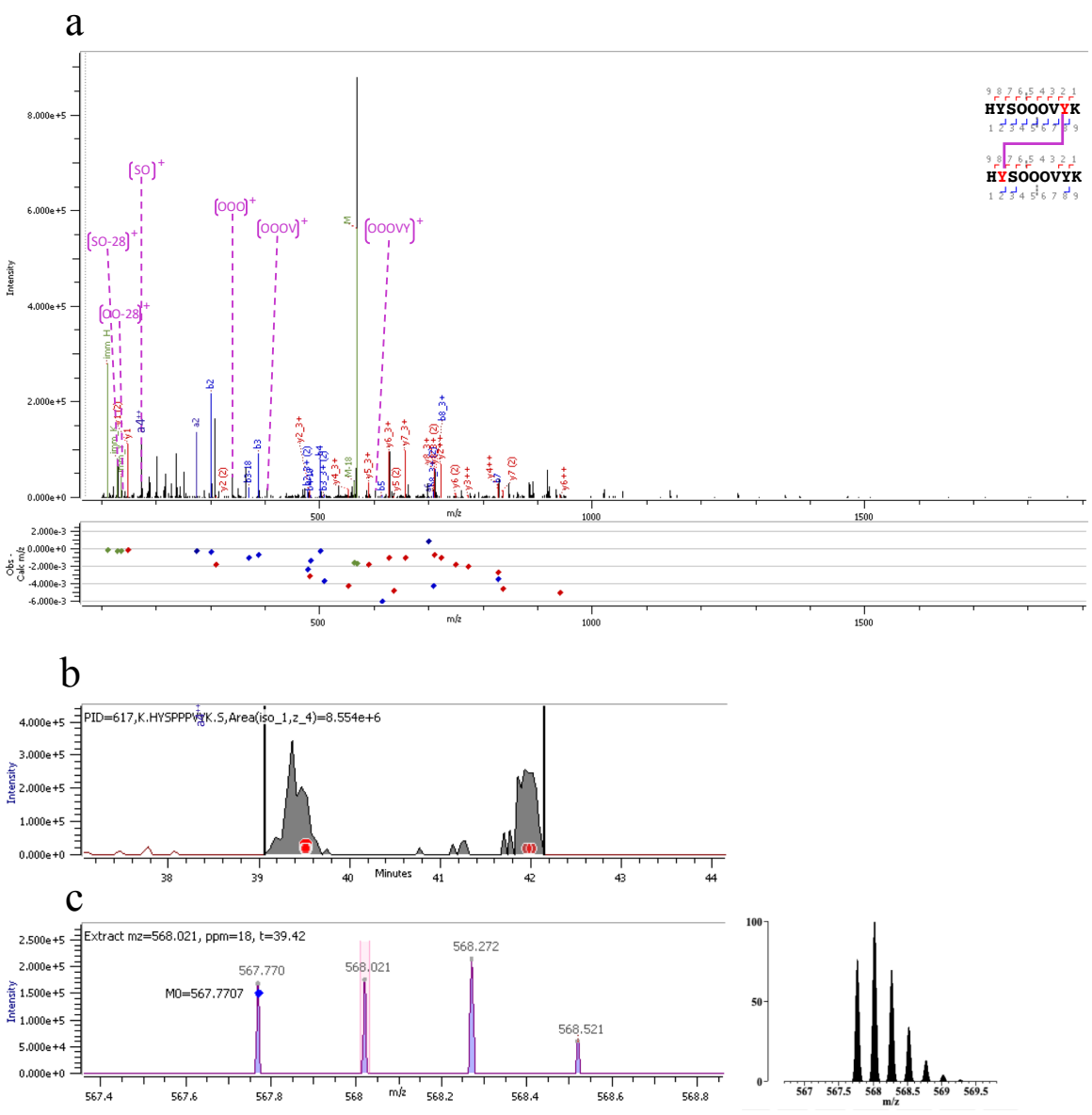

Figure S2.1: Characterization of the intermolecular dityrosine cross-linkage of two peptides from EXT1. (a) The annotated HCD MS/MS spectrum of the [M+4H]4+ ion at m/z 567.7707 with m/z errors of mass fragments at a 10ppm tolerance denoted below; and (inset, top right) a-, b-, & y-fragment ion coverage. Cross-linked tyrosines carrying the mass loss of (2.0156 Da) are denoted in red. (b) Extracted ion chromatogram. The red dots indicate several MS2 collected. (c) (Left)

The isotopic profile of the [M+4H]4+ precursor ion at m/z 567.7707 that was isolated for fragmentation. The blue dot indicates correct monoisotopic assignment based on calculated mass. (Right) The theoretically calculated isotopic profile of the [M+4H]4+ precursor ion at m/z 567.7707 from MS-Isotope.[68]

Supplemental Figures S3.1 to S3.5: MS Evidence for Pulcherosine Cross-Linked Peptides

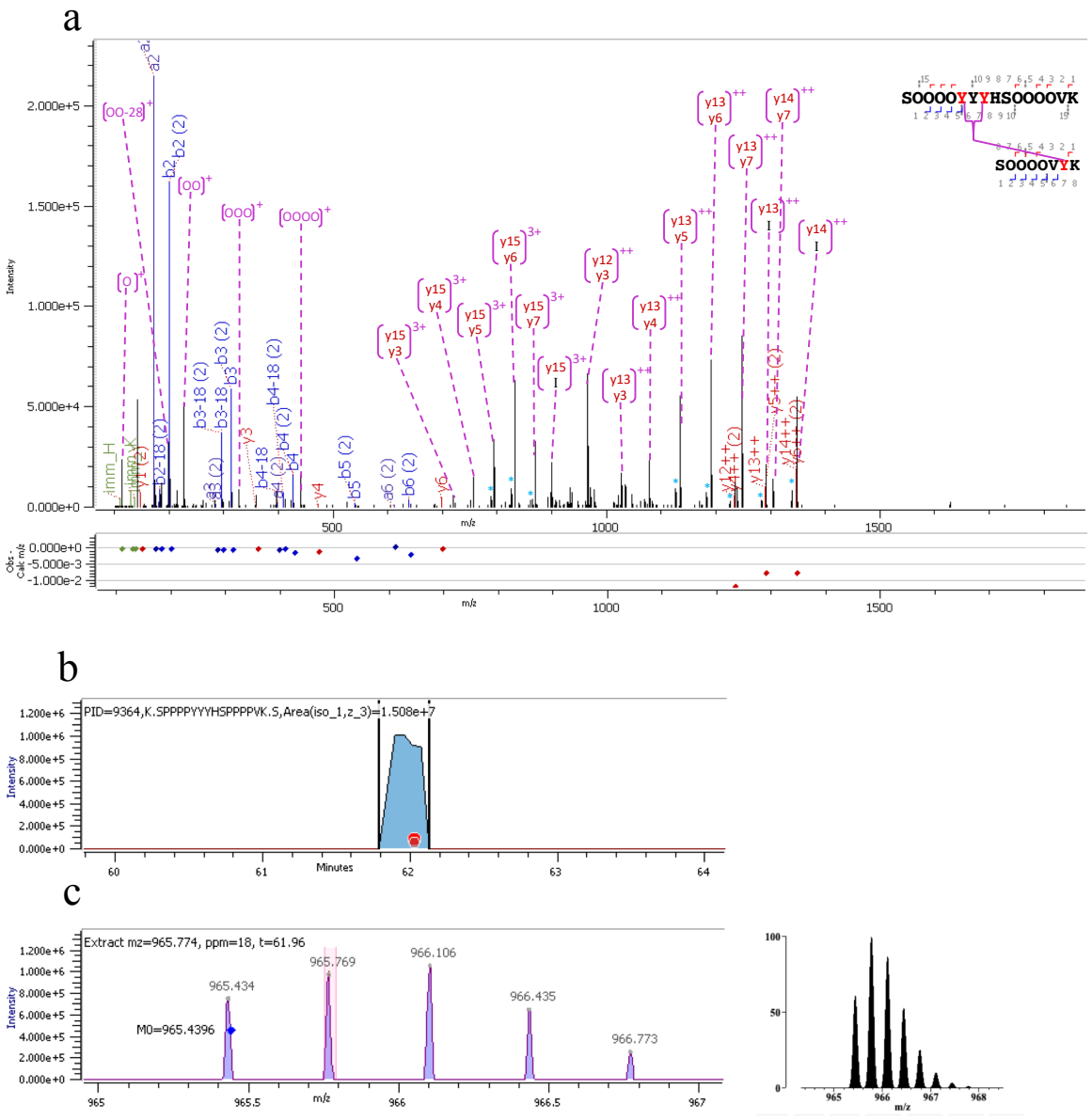

Figure S3.1: Characterization of the intermolecular pulcherosine cross-linkage of two peptides from EXT8 and EXT1. (a) The annotated HCD MS/MS spectrum of the [M+3H]3+ ion at m/z 965.4396 with m/z errors of mass fragments at a 10ppm tolerance denoted below; and (inset, top right) a-, b-, & y-fragment ion coverage. Cross-linked tyrosines carrying the mass loss of (4.0313 Da) are denoted in red. (b) Extracted ion chromatogram. The red dots indicate several MS2

collected. (c) (Left) The isotopic profile of the [M+3H]3+ precursor ion at m/z 965.4396 that was isolated for fragmentation. The blue dot indicates correct monoisotopic assignment based on calculated mass. (Right) The theoretically calculated isotopic profile of the [M+3H]3+ precursor ion at m/z 965.4396 from MS-Isotope.[68]

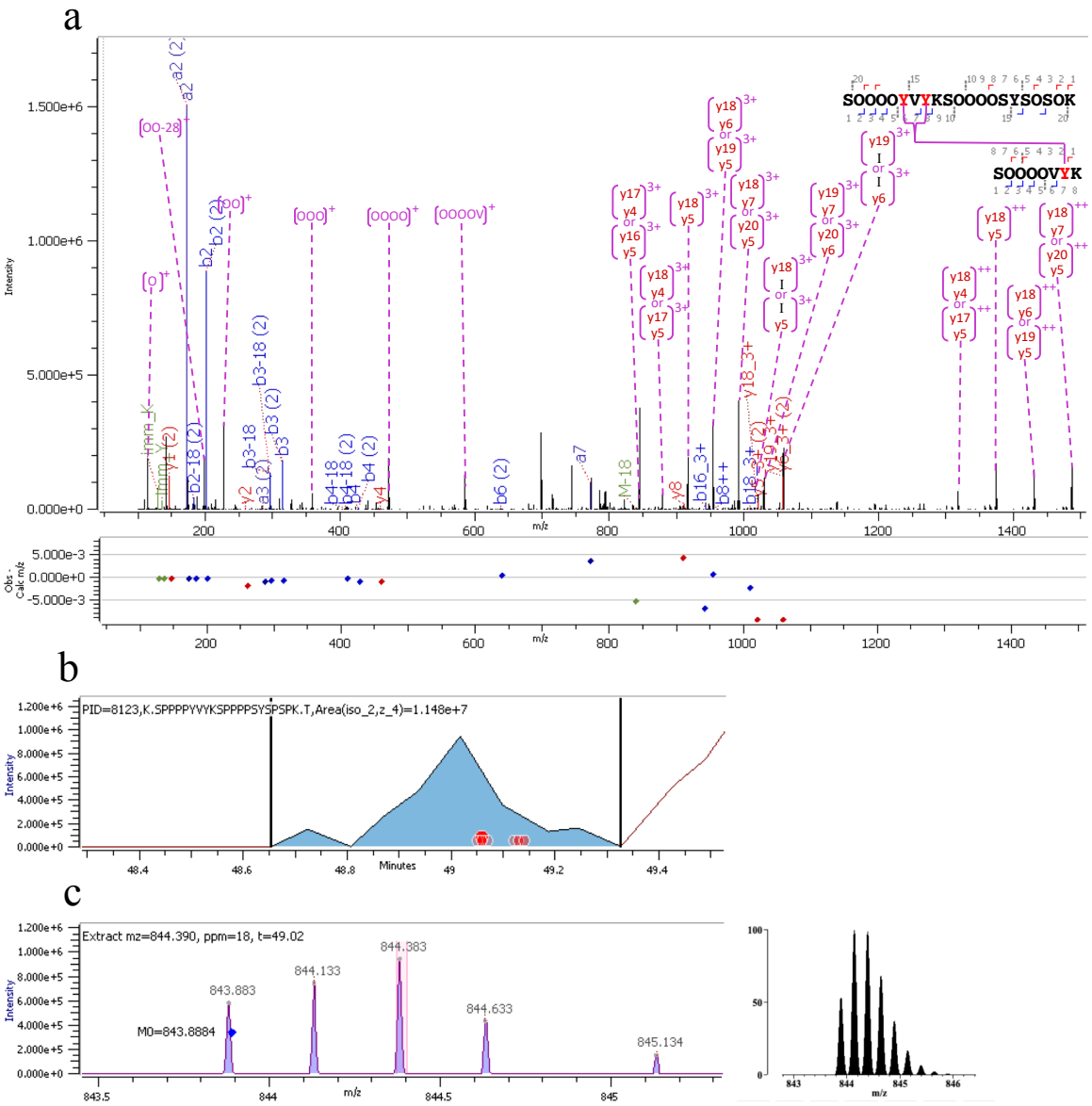

Figure S3.2: Characterization of the intermolecular pulcherosine cross-linkage of two peptides from EXT10 and EXT1. (a) The annotated HCD MS/MS spectrum of the [M+4H]4+ ion at m/z 843.8884 with m/z errors of mass fragments at a 10ppm tolerance denoted below; and (inset, top right) a-, b-, & y-fragment ion coverage. Cross-linked tyrosines carrying the mass loss of (4.0313 Da) are denoted in red. (b) Extracted ion chromatogram. The red dots indicate several MS2 with the same identification were collected. The grey dots indicate the time point of a MS2 of the same mass, but with a different ID such as a different site of nonhydroxylated proline. (c) (Left) The

isotopic profile of the [M+4H]4+ precursor ion at m/z 843.8884 that was isolated for fragmentation. The blue dot indicates correct monoisotopic assignment based on calculated mass. (Right) The theoretically calculated isotopic profile of the [M+4H]4+ precursor ion at m/z 843.8884 from MS-Isotope.[68]

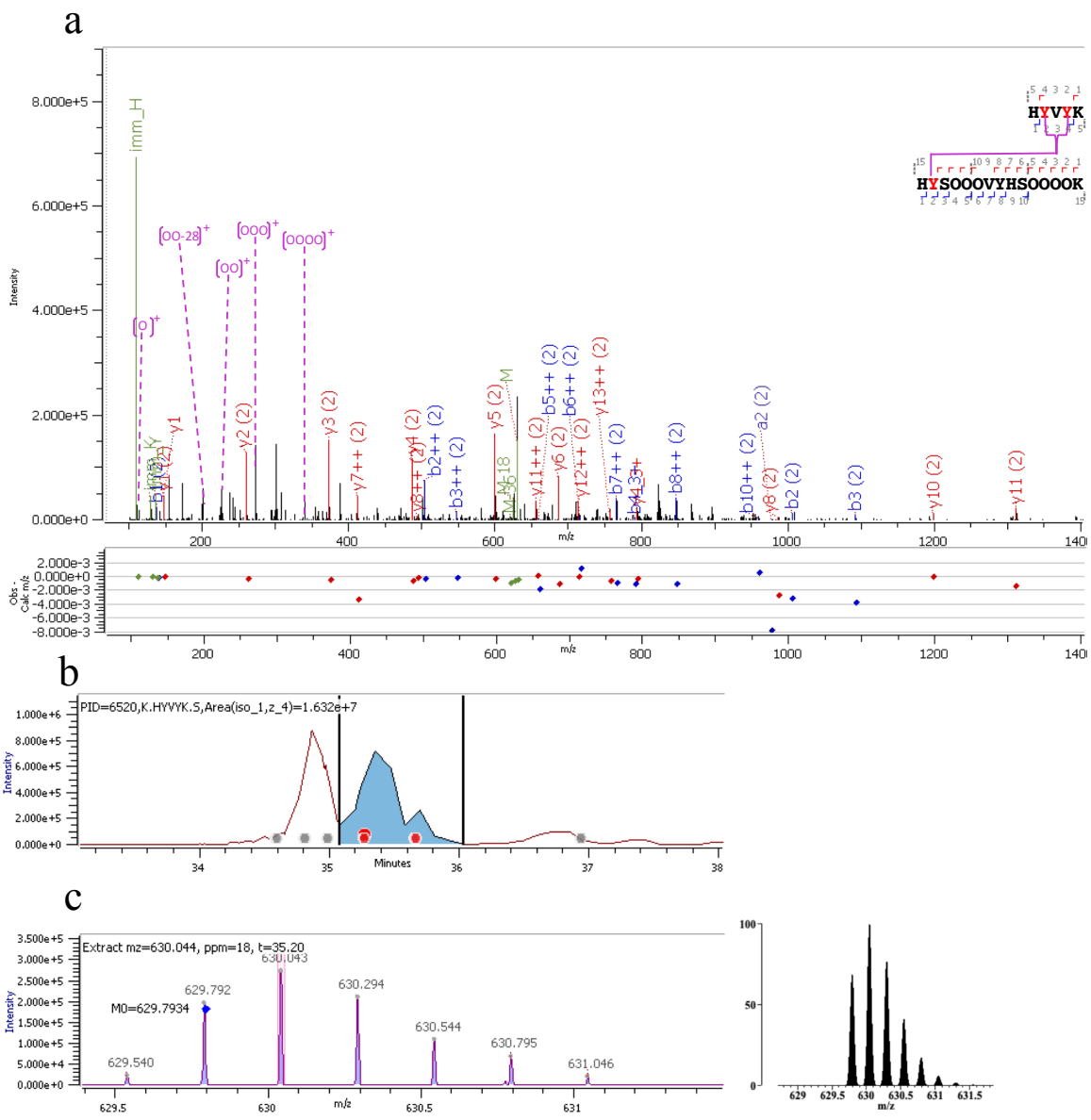

Figure S3.3: Characterization of the intermolecular pulcherosine cross-linkage of two peptides from EXT3. (a) The annotated HCD MS/MS spectrum of the [M+4H]4+ ion at m/z 629.7934 with m/z errors of mass fragments at a 10ppm tolerance denoted below; and (inset, top right) a-, b-, & y-fragment ion coverage. Cross-linked tyrosines carrying the mass loss of (4.0313 Da) are denoted in red. (b) Extracted ion chromatogram. The red dots indicate several MS2 collected. (c) (Left) The isotopic profile of the [M+4H]4+ precursor ion at m/z 629.7934 that was isolated for fragmentation. The blue dot indicates correct monoisotopic assignment based on calculated mass.

(Right) The theoretically calculated isotopic profile of the [M+4H]4+ precursor ion at m/z 629.7934 from MS-Isotope.[68]

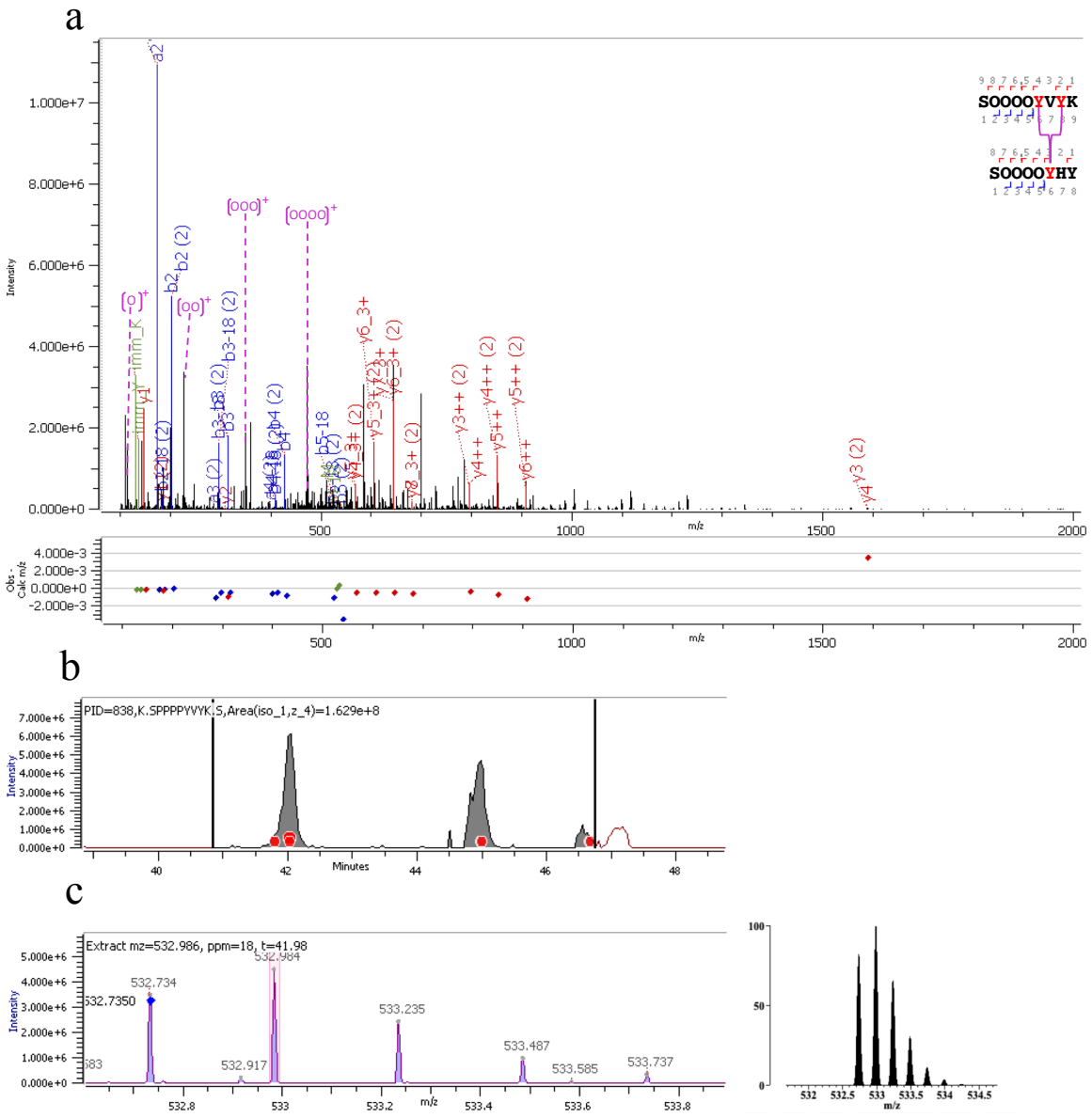

Figure S3.4: Characterization of the intermolecular pulcherosine cross-linkage of two peptides from EXT10 and EXT3. (a) The annotated HCD MS/MS spectrum of the [M+4H]4+ ion at m/z 532.735 with m/z errors of mass fragments at a 10ppm tolerance denoted below; and (inset, top right) a-, b-, & y-fragment ion coverage. Cross-linked tyrosines carrying the mass loss of (4.0313 Da) are denoted in red. (b) Extracted ion chromatogram. The red dots indicate several MS2 collected. (c) (Left) The isotopic profile of the [M+4H]4+ precursor ion at m/z 532.735 that was isolated for fragmentation. The blue dot indicates correct monoisotopic assignment based on

calculated mass. (Right) The theoretically calculated isotopic profile of the [M+4H]4+ precursor ion at m/z 532.735 from MS-Isotope.[68]

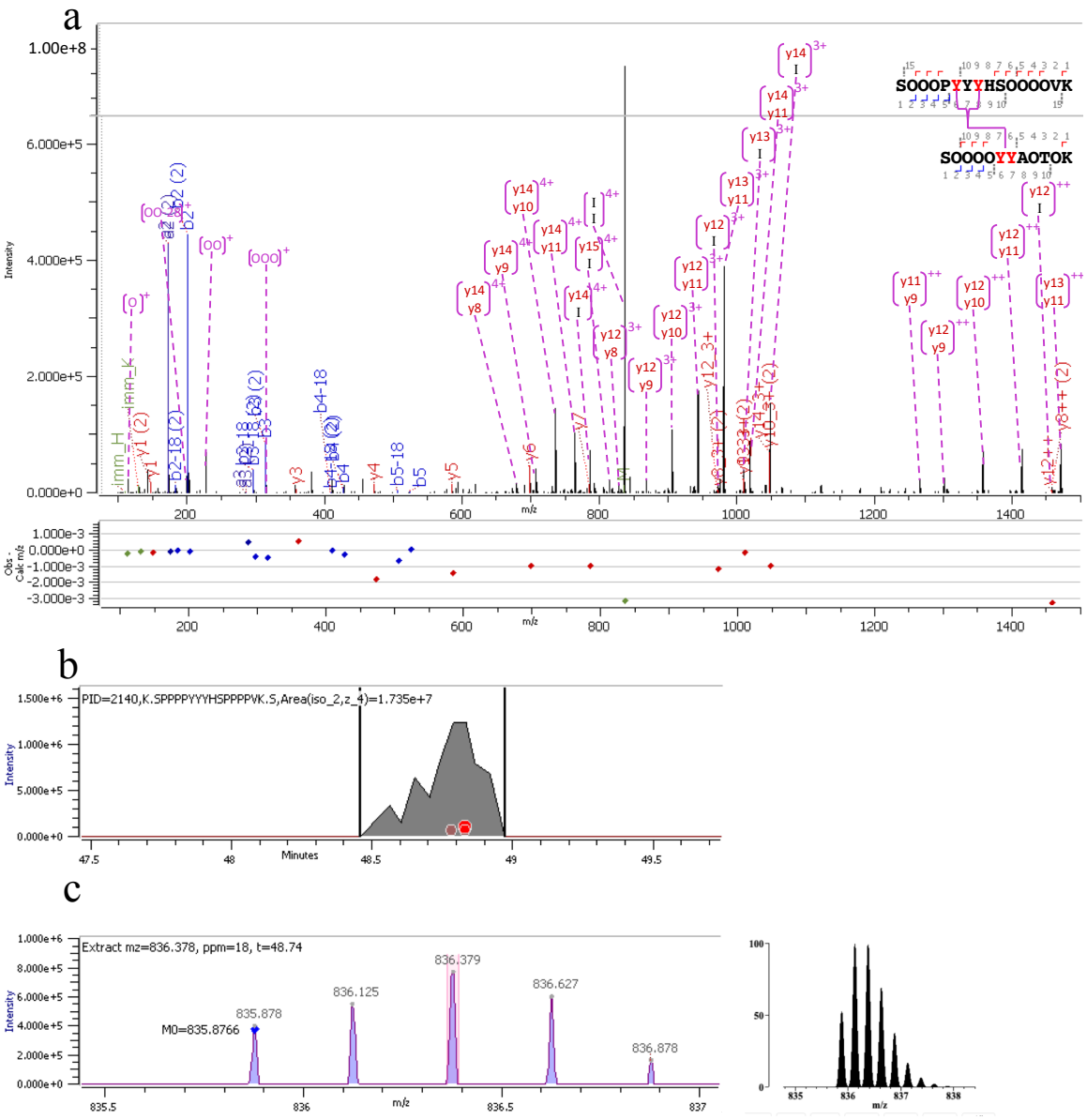

Figure S3.5: Characterization of the intermolecular pulcherosine cross-linkage of two peptides from EXT8 and EXT10. (a) The annotated HCD MS/MS spectrum of the [M+4H]4+ ion at m/z 835.8766 with m/z errors of mass fragments at a 10ppm tolerance denoted below; and (inset, top right) a-, b-, & y-fragment ion coverage. Cross-linked tyrosines carrying the mass loss of (4.0313 Da) are denoted in red. (b) Extracted ion chromatogram. The red dots indicate several MS2 collected. (c) (Left) The isotopic profile of the [M+4H]4+ precursor ion at m/z 835.8766 that was isolated for fragmentation. The blue dot indicates correct monoisotopic assignment based on

calculated mass. (Right) The theoretically calculated isotopic profile of the [M+4H]4+ precursor ion at m/z 835.8766 from MS-Isotope.[68]

Supplemental Figures 4.1 to 4.14: MS Evidence for Di-IDT Cross-Linked Peptides

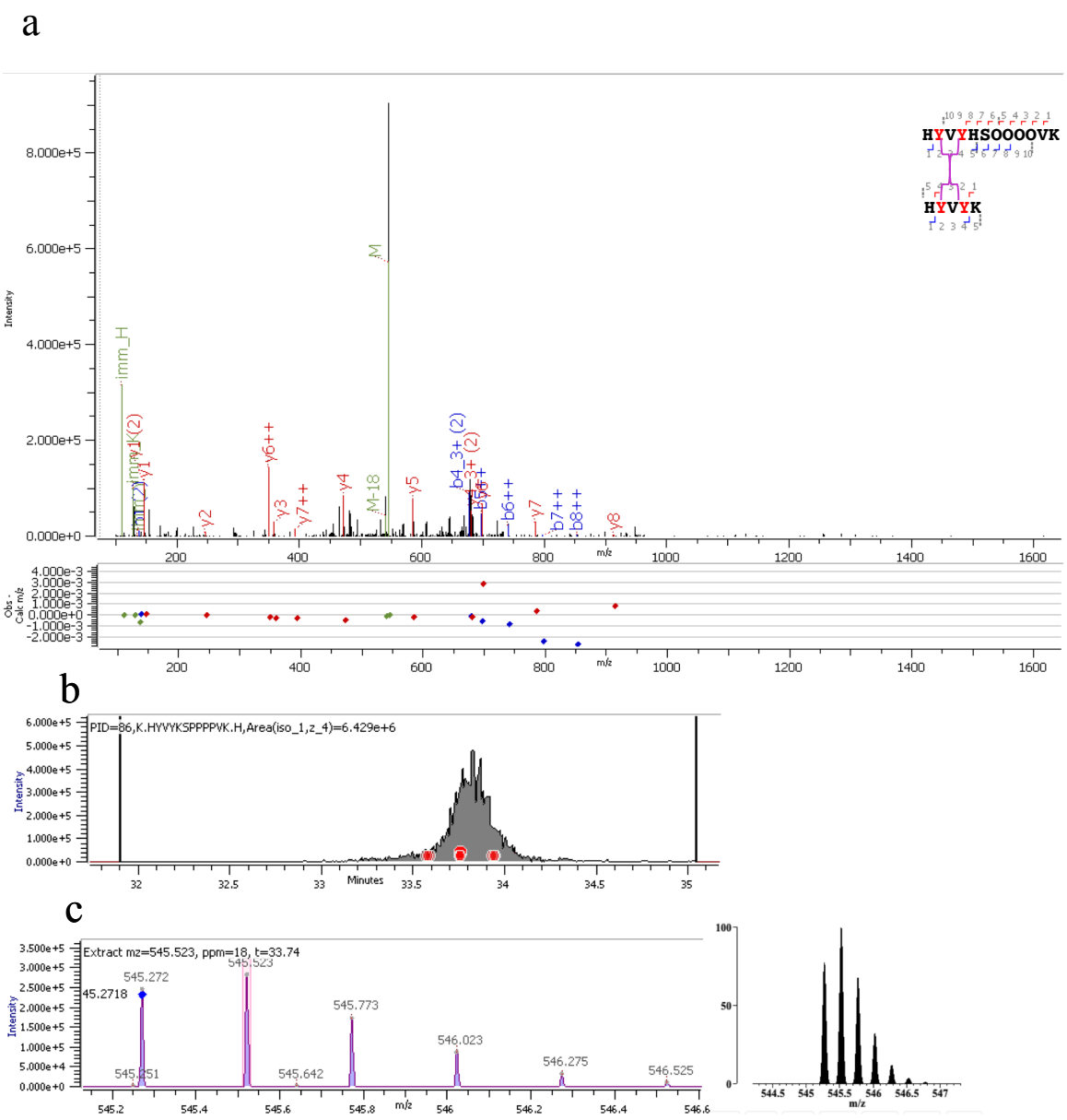

Figure S4.1: Characterization of the intermolecular Di-IDT cross-linkage of two peptides from EXT3. (a) The annotated HCD MS/MS spectrum of the [M+4H]4+ ion at m/z 545.2718 with m/z errors of mass fragments at a 10ppm tolerance denoted below; and (inset, top right) a-, b-, & y-fragment ion coverage. Cross-linked tyrosines carrying the mass loss of (6.040595 Da) are denoted in red. (b) Extracted ion chromatogram. The red dots indicate several MS2 collected. (c)

(Left) The isotopic profile of the [M+4H]4+ precursor ion at m/z 545.2718 that was isolated for fragmentation. The blue dot indicates correct monoisotopic assignment based on calculated mass. (Right) The theoretically calculated isotopic profile of the [M+4H]4+ precursor ion at m/z 545.2718 from MS-Isotope.[68]

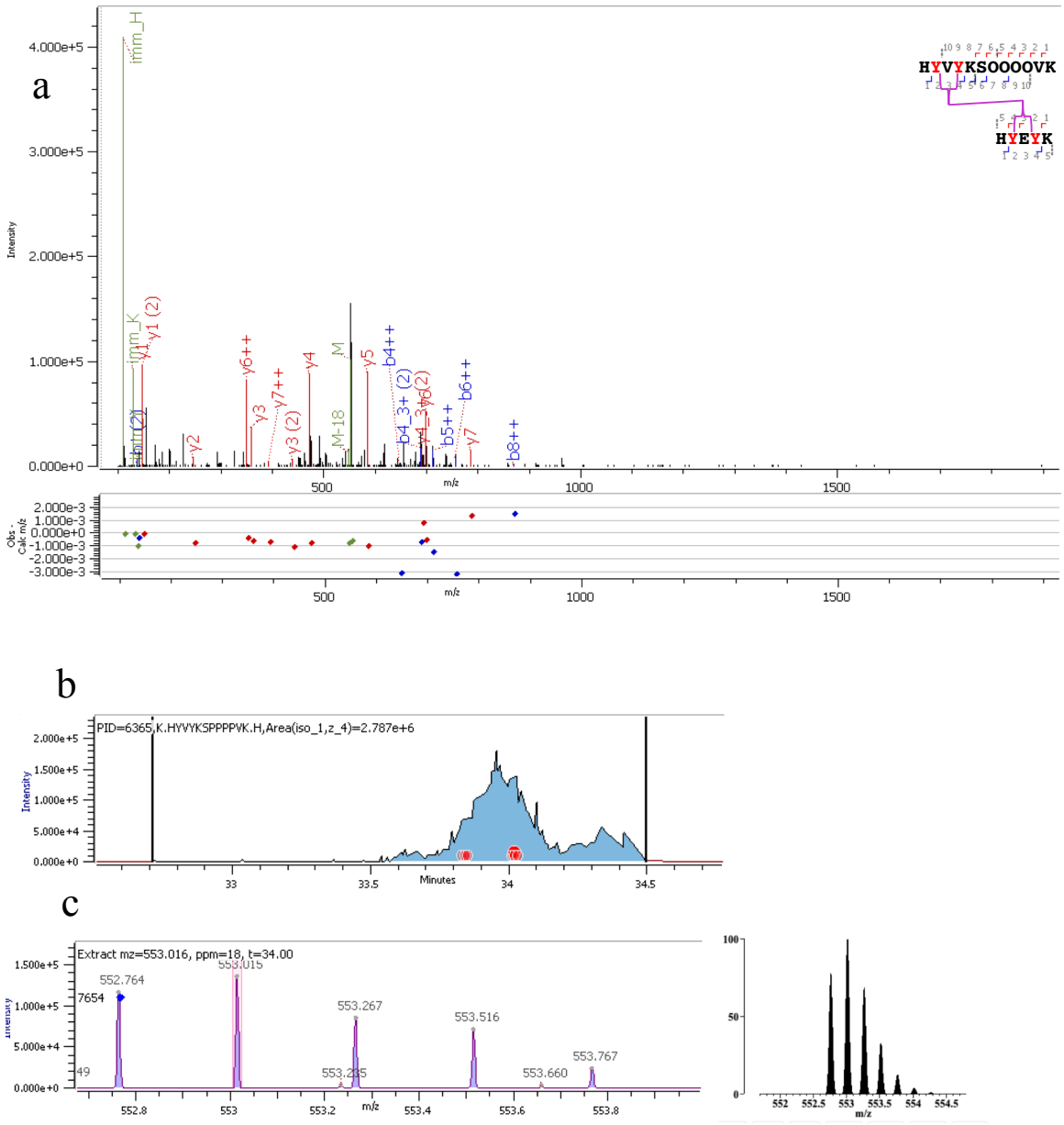

Figure S4.2: Characterization of the intermolecular Di-IDT cross-linkage of two peptides from EXT3 and EXT3 (or EXT1). (a) The annotated HCD MS/MS spectrum of the [M+4H]4+ ion at m/z 552.7654 with m/z errors of mass fragments at a 10ppm tolerance denoted below; and (inset, top right) a-, b-, & y-fragment ion coverage. Cross-linked tyrosines carrying the mass loss of (6.040595 Da) are denoted in red. (b) Extracted ion chromatogram. The red dots indicate several MS2 collected. (c) (Left) The isotopic profile of the [M+4H]4+ precursor ion at m/z 552.7654 that was isolated for fragmentation. The blue dot indicates correct monoisotopic assignment

based on calculated mass. (Right) The theoretically calculated isotopic profile of the [M+4H]4+ precursor ion at m/z 552.7654 from MS-Isotope.[68]

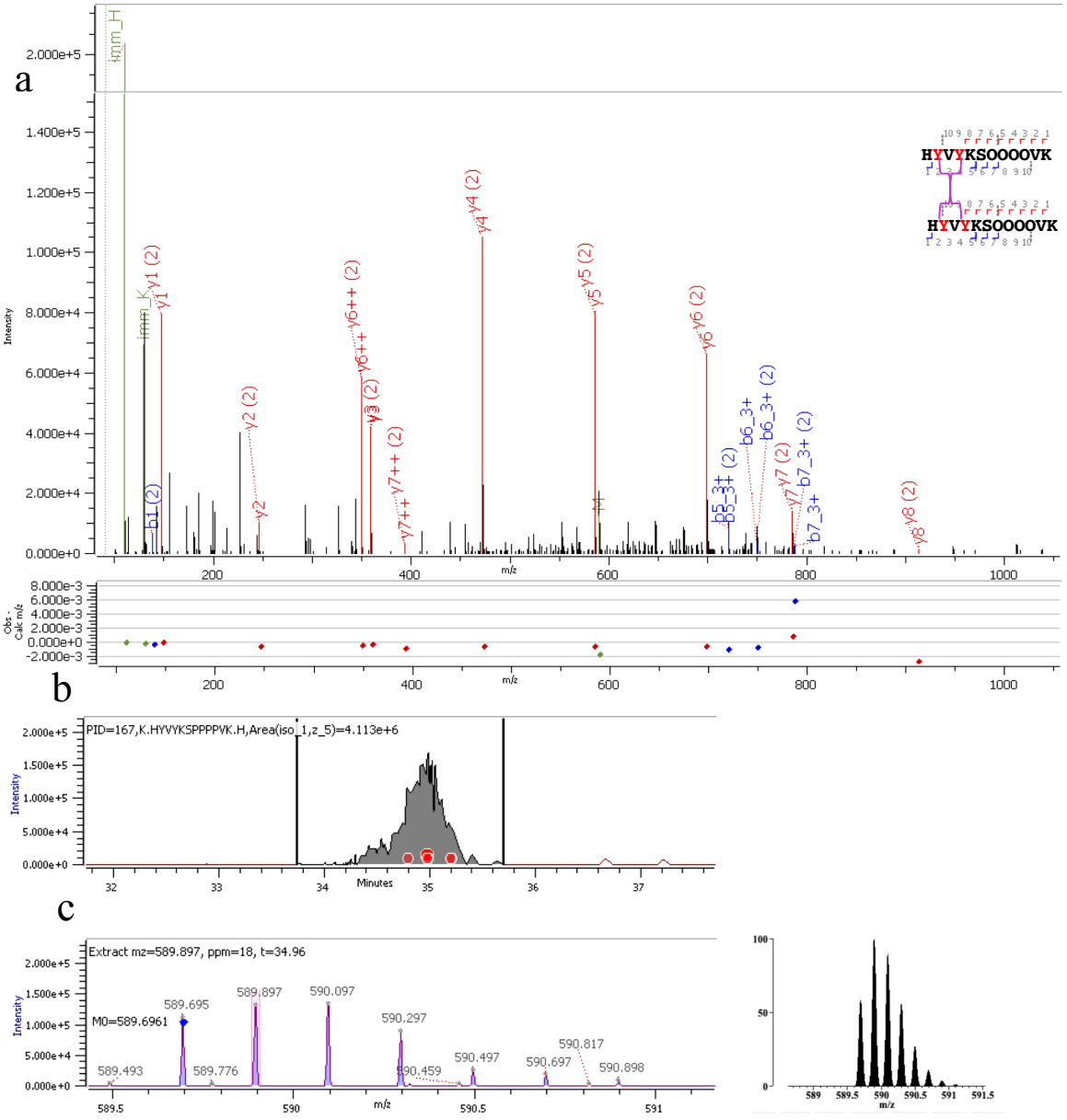

Figure S4.3: Characterization of the intermolecular Di-IDT cross-linkage of two peptides from EXT3. (a) The annotated HCD MS/MS spectrum of the [M+5H]5+ ion at m/z 589.6961 with m/z errors of mass fragments at a 10ppm tolerance denoted below; and (inset, top right) a-, b-, & y-fragment ion coverage. Cross-linked tyrosines carrying the mass loss of (6.040595 Da) are denoted in red. (b) Extracted ion chromatogram. The red dots indicate several MS2 collected. (c) (Left) The isotopic profile of the [M+5H]5+ precursor ion at m/z 589.6961 that was isolated for fragmentation. The blue dot indicates correct monoisotopic assignment based on calculated mass.

(Right) The theoretically calculated isotopic profile of the [M+5H]5+ precursor ion at m/z 589.6961 from MS-Isotope.[68]

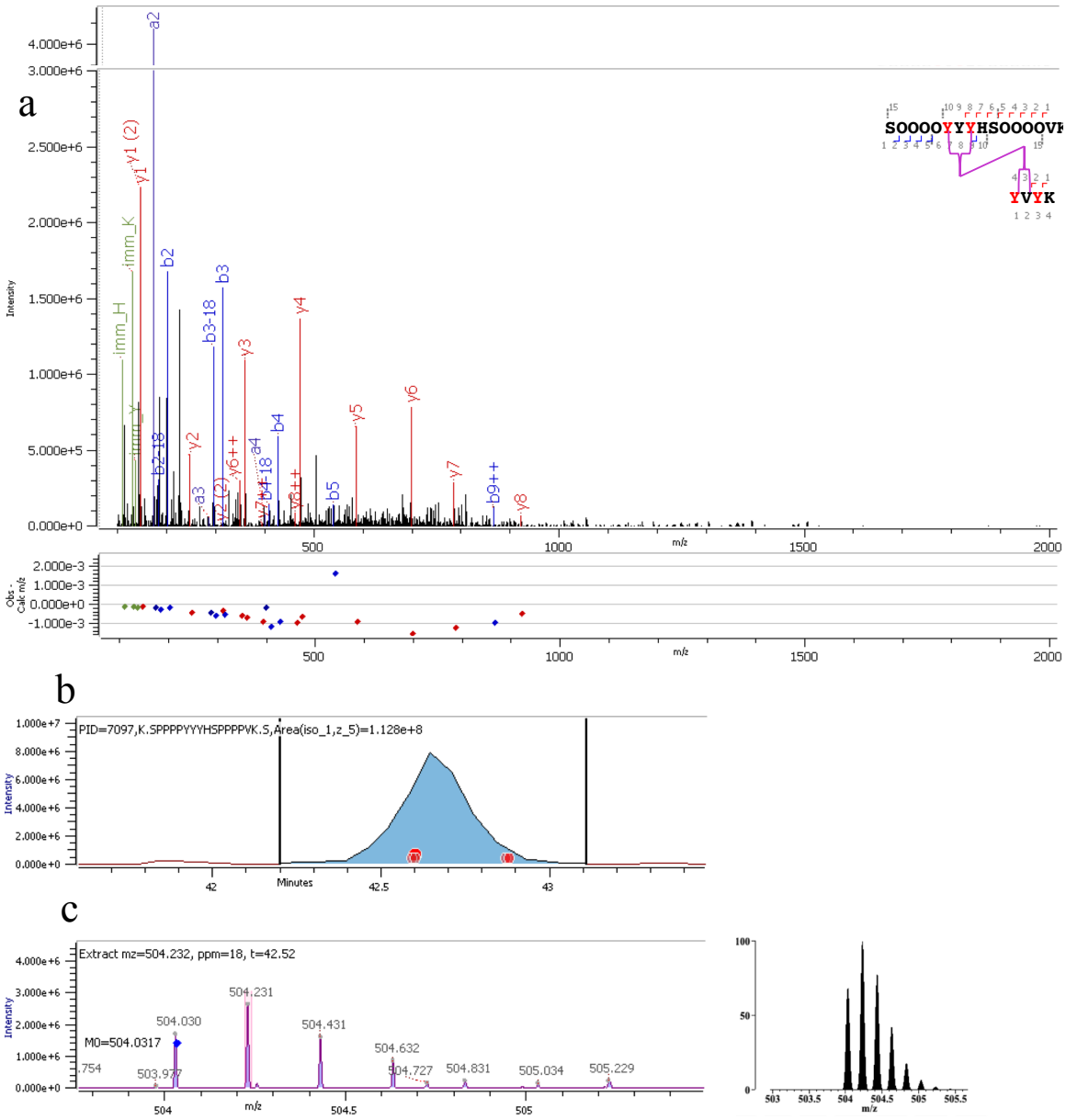

Figure S4.4: Characterization of the intermolecular Di-IDT cross-linkage of two peptides from EXT8 and EXT3. (a) The annotated HCD MS/MS spectrum of the [M+5H]5+ ion at m/z 504.0317 with m/z errors of mass fragments at a 10ppm tolerance denoted below; and (inset, top right) a-, b-, & y-fragment ion coverage. Cross-linked tyrosines carrying the mass loss of (6.040595 Da) are denoted in red. (b) Extracted ion chromatogram. The red dots indicate several MS2 collected. (c) (Left) The isotopic profile of the [M+5H]5+ precursor ion at m/z 504.0317 that was isolated for fragmentation. The blue dot indicates correct monoisotopic assignment

based on calculated mass. (Right) The theoretically calculated isotopic profile of the [M+5H]5+ precursor ion at m/z 504.0317 from MS-Isotope.[68]

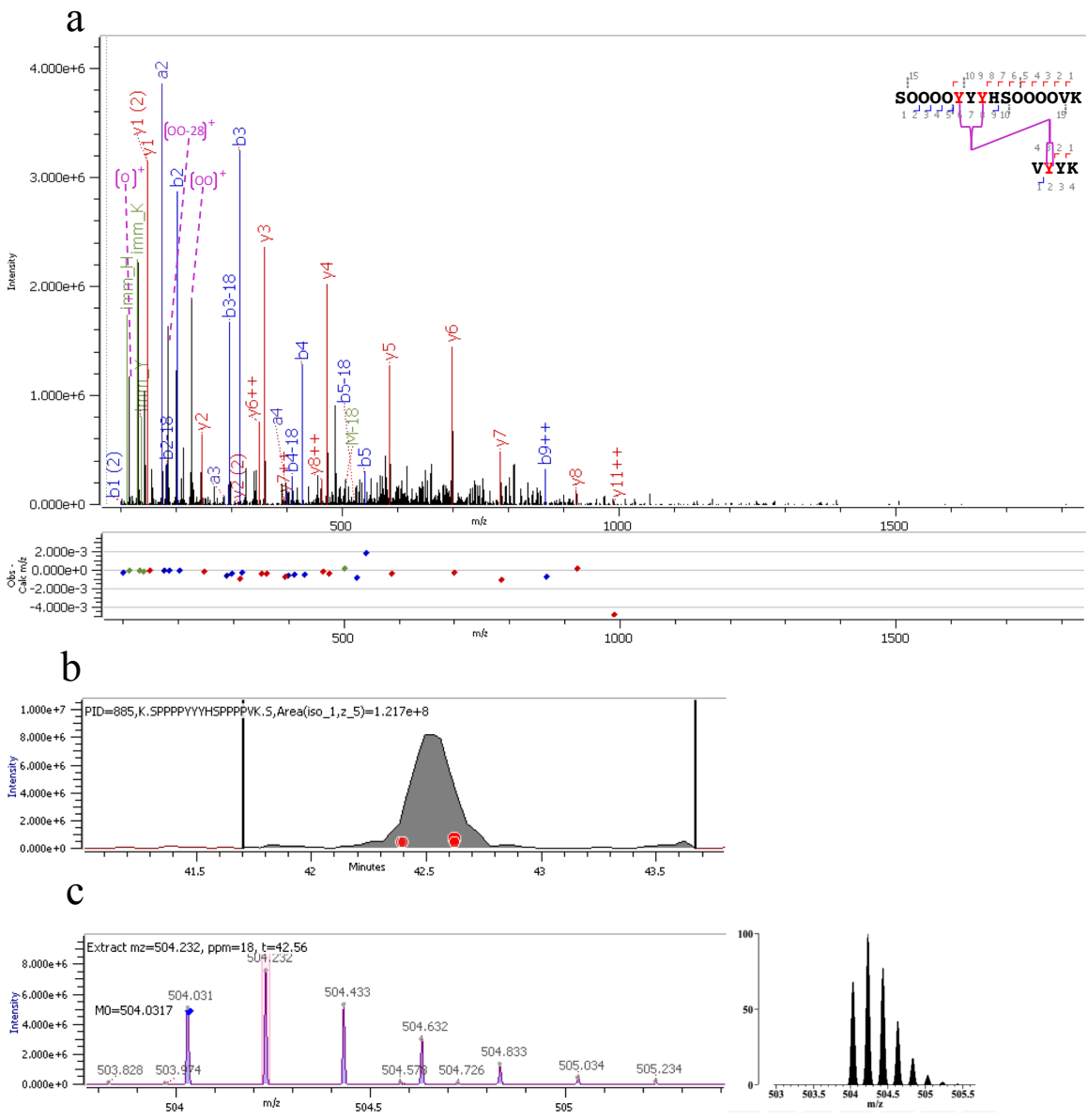

Figure S4.5: Characterization of the intermolecular Di-IDT cross-linkage of two peptides from EXT8 and EXT10. (a) The annotated HCD MS/MS spectrum of the [M+5H]5+ ion at m/z 504.0317 with m/z errors of mass fragments at a 10ppm tolerance denoted below; and (inset, top right) a-, b-, & y-fragment ion coverage. Cross-linked tyrosines carrying the mass loss of (6.040595 Da) are denoted in red. (b) Extracted ion chromatogram. The red dots indicate several MS2 collected. (c) (Left) The isotopic profile of the [M+5H]5+ precursor ion at m/z 504.0317 that was isolated for fragmentation. The blue dot indicates correct monoisotopic assignment

based on calculated mass. (Right) The theoretically calculated isotopic profile of the [M+5H]5+ precursor ion at m/z 504.0317 from MS-Isotope.[68]

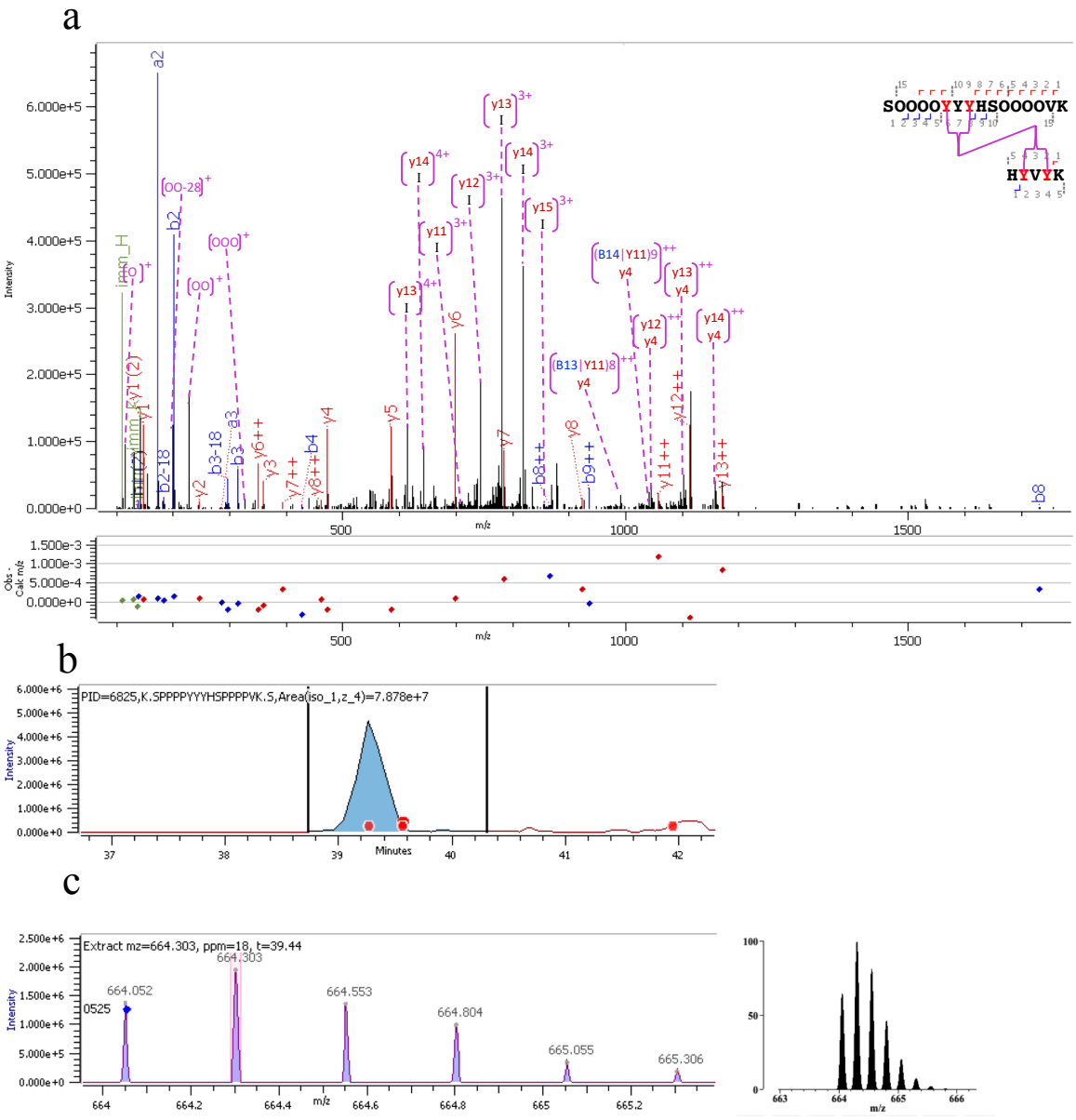

Figure S4.6: Characterization of the intermolecular Di-IDT cross-linkage of two peptides from EXT8 and EXT3. (a) The annotated HCD MS/MS spectrum of the [M+4H]4+ ion at m/z 644.0525 with m/z errors of mass fragments at a 10ppm tolerance denoted below; and (inset, top right) a-, b-, & y-fragment ion coverage. Cross-linked tyrosines carrying the mass loss of (6.040595 Da) are denoted in red. (b) Extracted ion chromatogram. The red dots indicate several MS2 collected. (c) (Left) The isotopic profile of the [M+4H]4+ precursor ion at m/z 644.0525 that was isolated for fragmentation. The blue dot indicates correct monoisotopic assignment

based on calculated mass. (Right) The theoretically calculated isotopic profile of the [M+4H]4+ precursor ion at m/z 644.0525 from MS-Isotope.[68]

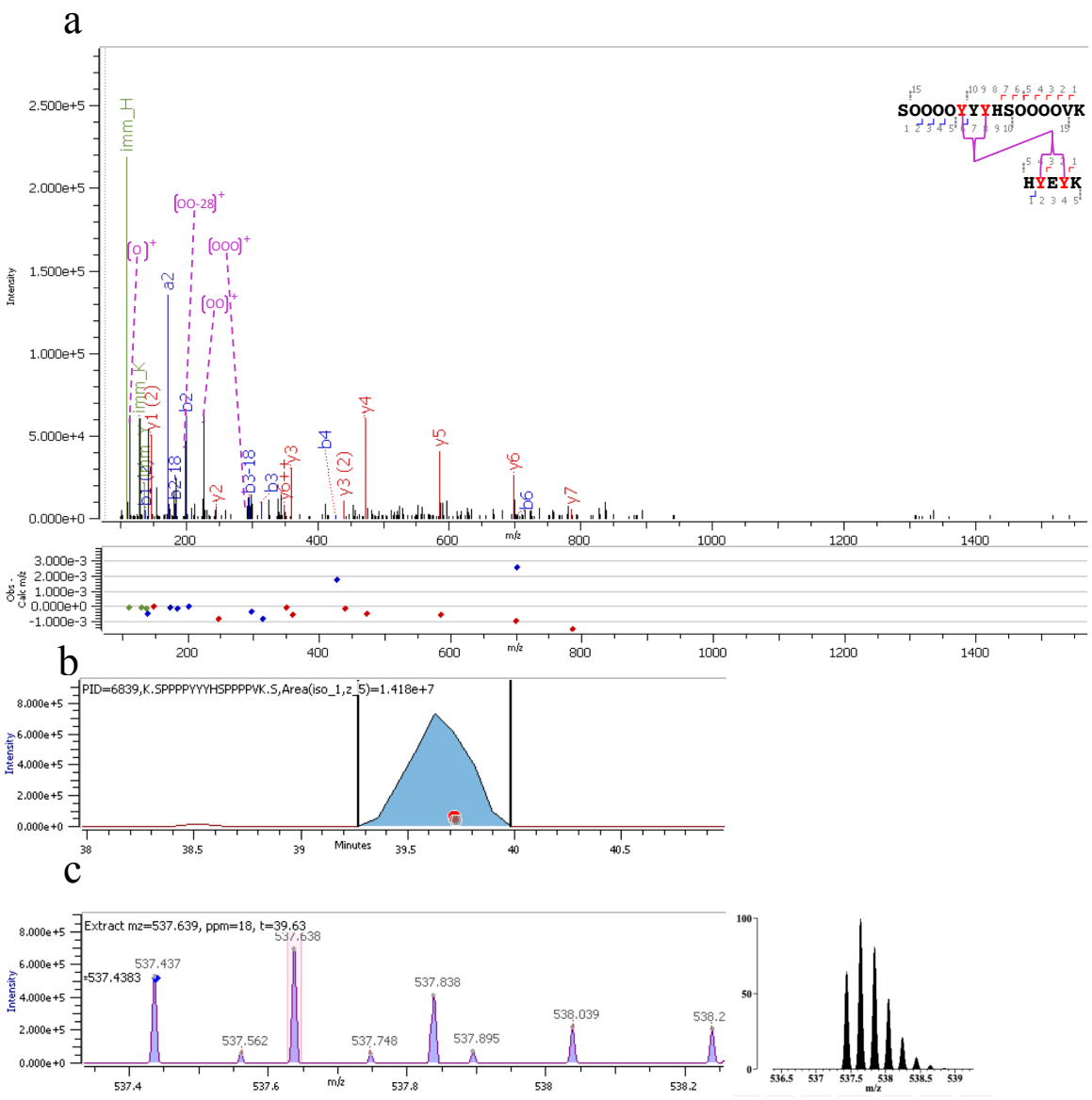

Figure S4.7: Characterization of the intermolecular Di-IDT cross-linkage of two peptides from EXT3 and EXT3 (or EXT1). (a) The annotated HCD MS/MS spectrum of the [M+5H]5+ ion at m/z 537.4383 with m/z errors of mass fragments at a 10ppm tolerance denoted below; and (inset, top right) a-, b-, & y-fragment ion coverage. Cross-linked tyrosines carrying the mass loss of (6.040595 Da) are denoted in red. (b) Extracted ion chromatogram. The red and grey dots indicate several MS2 with the same identification were collected. (The grey dots usually indicate the time point of a MS2 of the same mass, but with a different ID. However, these are a Byos artifact

caused by lack of fragmentation at the yXy motif.) (c) (Left) The isotopic profile of the [M+5H]5+ precursor ion at m/z 537.4383 that was isolated for fragmentation. The blue dot indicates correct monoisotopic assignment based on calculated mass. (Right) The theoretically calculated isotopic profile of the [M+5H]5+ precursor ion at m/z 537.4383 from MS-Isotope.[68]

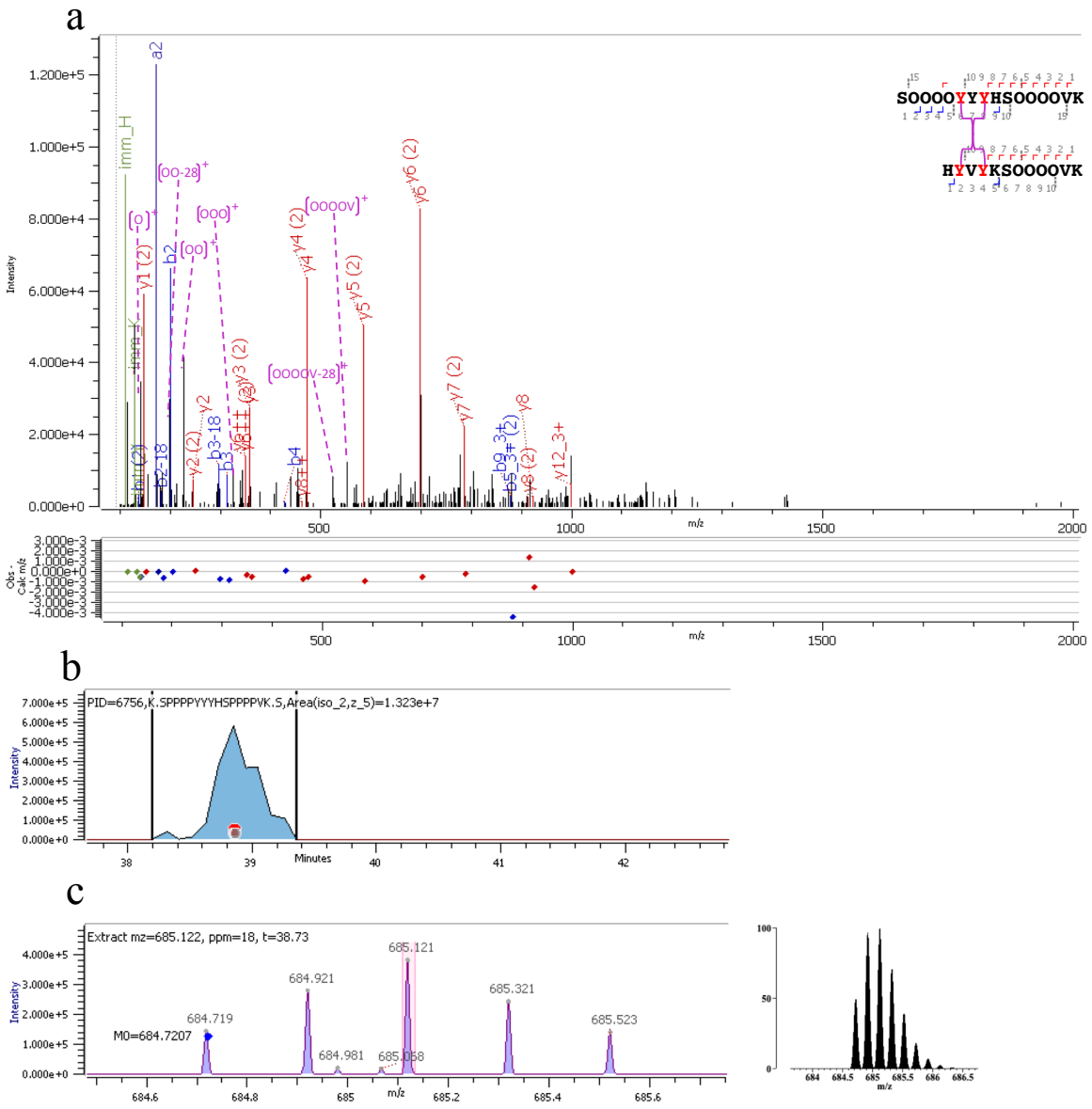

Figure S4.8: Characterization of the intermolecular Di-IDT cross-linkage of two peptides from EXT3 and EXT8. (a) The annotated HCD MS/MS spectrum of the [M+5H]5+ ion at m/z 684.7207 with m/z errors of mass fragments at a 10ppm tolerance denoted below; and (inset, top right) a-, b-, & y-fragment ion coverage. Cross-linked tyrosines carrying the mass loss of (6.040595 Da) are denoted in red. (b) Extracted ion chromatogram. The red and grey dots indicate several MS2 with the same identification were collected. (The grey dots usually indicate the time point of a MS2 of the same mass, but with a different ID. However, these are a Byos artifact

caused by lack of fragmentation at the yXy motif.) (c) (Left) The isotopic profile of the [M+5H]5+ precursor ion at m/z 684.7207 that was isolated for fragmentation. The blue dot indicates correct monoisotopic assignment based on calculated mass. (Right) The theoretically calculated isotopic profile of the [M+5H]5+ precursor ion at m/z 684.7207 from MS-Isotope.[68]

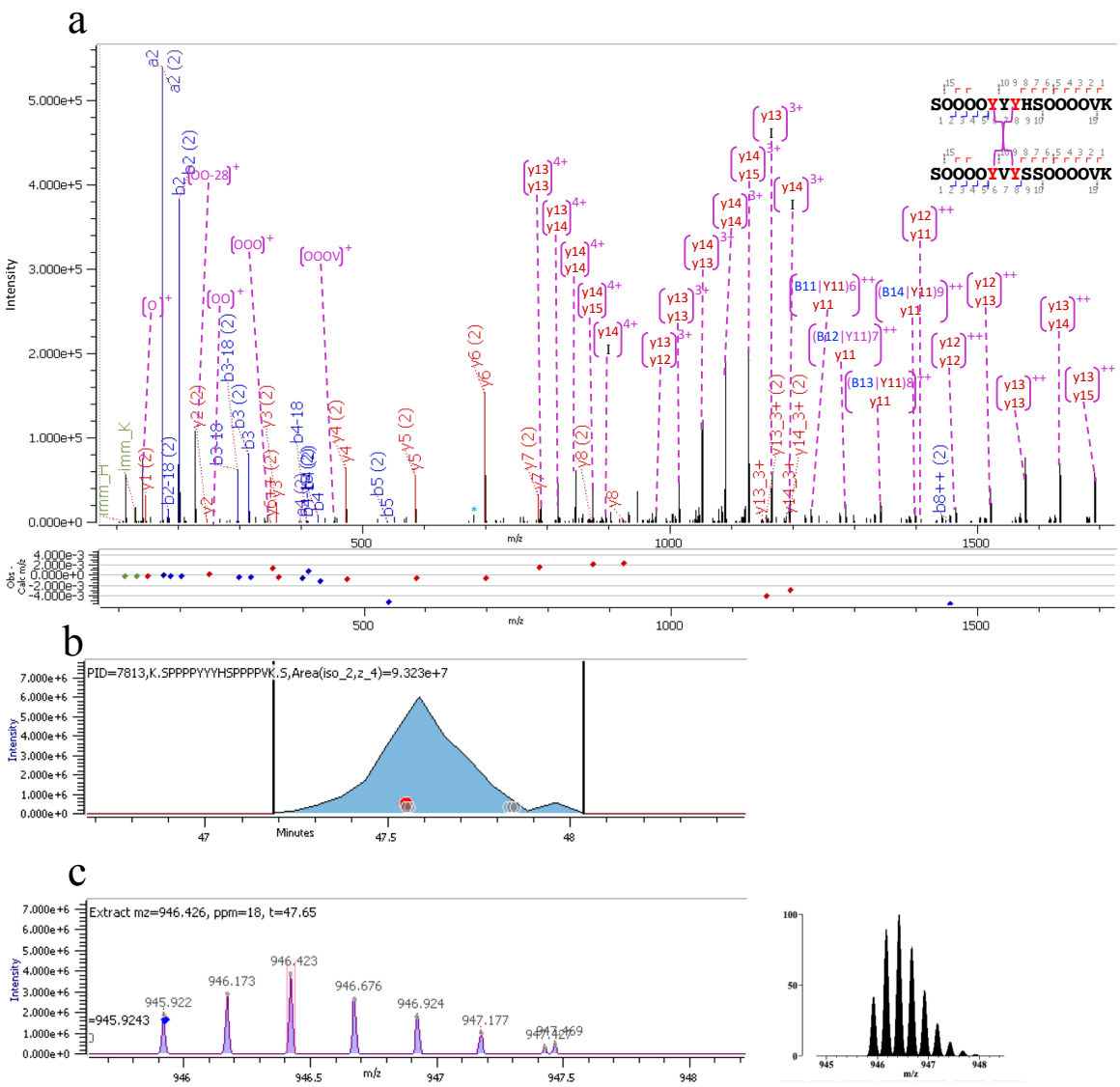

Figure S4.9: Characterization of the intermolecular Di-IDT cross-linkage of two peptides from EXT8. (a) The annotated HCD MS/MS spectrum of the [M+4H]4+ ion at m/z 945.9243 with m/z errors of mass fragments at a 10ppm tolerance denoted below; and (inset, top right) a-, b-, & y-fragment ion coverage. Cross-linked tyrosines carrying the mass loss of (6.040595 Da) are denoted in red. (b) Extracted ion chromatogram. The red and grey dots indicate several MS2 with the same identification were collected. (The grey dots usually indicate the time point of a MS2 of the same mass, but with a different ID. However, these are a Byos artifact caused by lack of

fragmentation at the yXy motif.) (c) (Left) The isotopic profile of the [M+4H]4+ precursor ion at m/z 945.9243 that was isolated for fragmentation. The blue dot indicates correct monoisotopic assignment based on calculated mass. (Right) The theoretically calculated isotopic profile of the [M+4H]4+ precursor ion at m/z 945.9243 from MS-Isotope.[68]

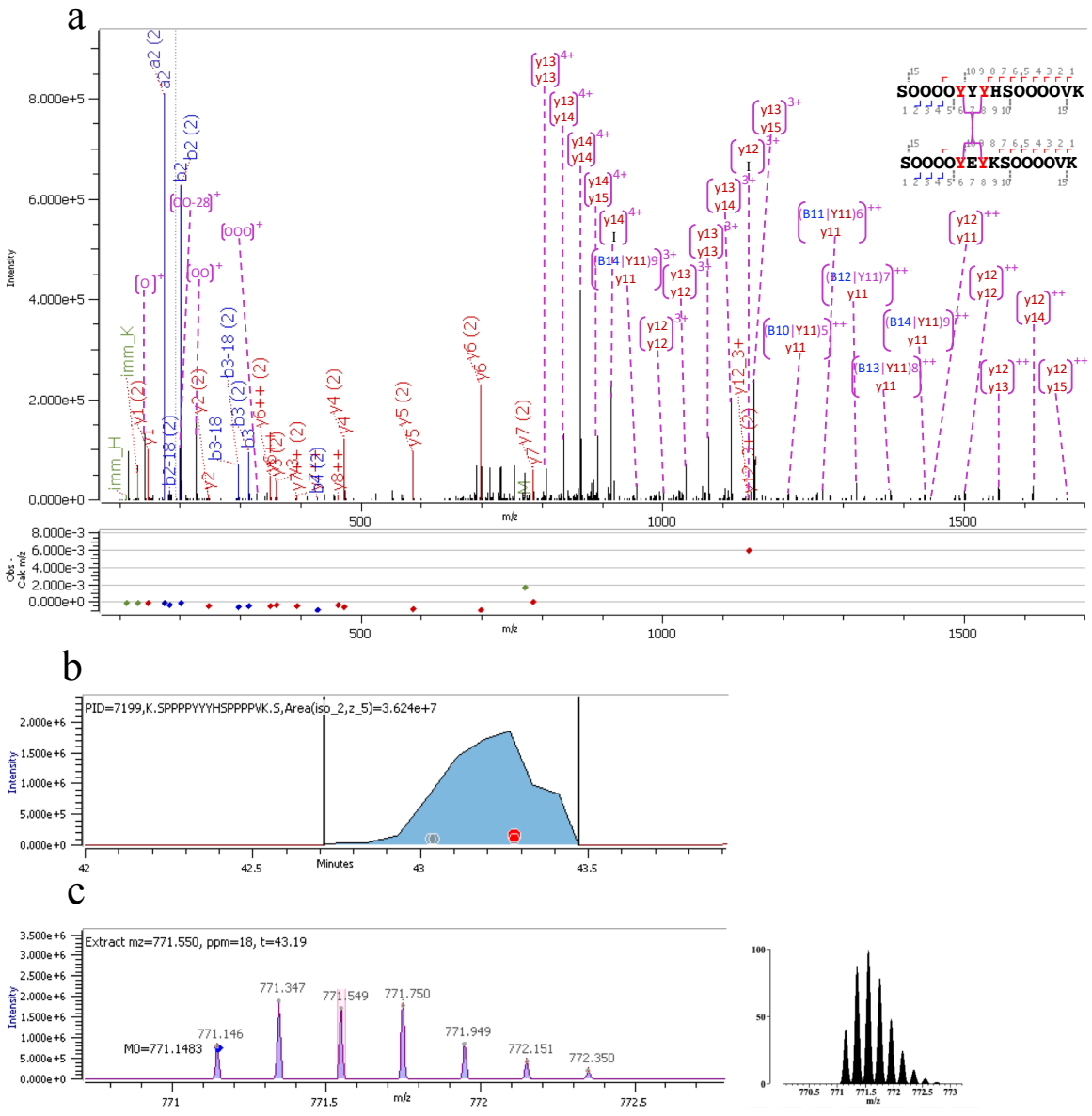

Figure S4.10: Characterization of the intermolecular Di-IDT cross-linkage of two peptides from EXT8. (a) The annotated HCD MS/MS spectrum of the [M+5H]5+ ion at m/z 711.1483 with m/z errors of mass fragments at a 10ppm tolerance denoted below; and (inset, top right) a-, b-, & y-fragment ion coverage. Cross-linked tyrosines carrying the mass loss of (6.040595 Da) are denoted in red. (b) Extracted ion chromatogram. The red and grey dots indicate several MS2 with the same identification were collected. (The grey dots usually indicate the time point of a MS2 of the same mass, but with a different ID. However, these are a Byos artifact caused by lack of

fragmentation at the yXy motif.) (c) (Left) The isotopic profile of the [M+5H]5+ precursor ion at m/z 711.1483 that was isolated for fragmentation. The blue dot indicates correct monoisotopic assignment based on calculated mass. (Right) The theoretically calculated isotopic profile of the [M+5H]5+ precursor ion at m/z 711.1483 from MS-Isotope.[68]

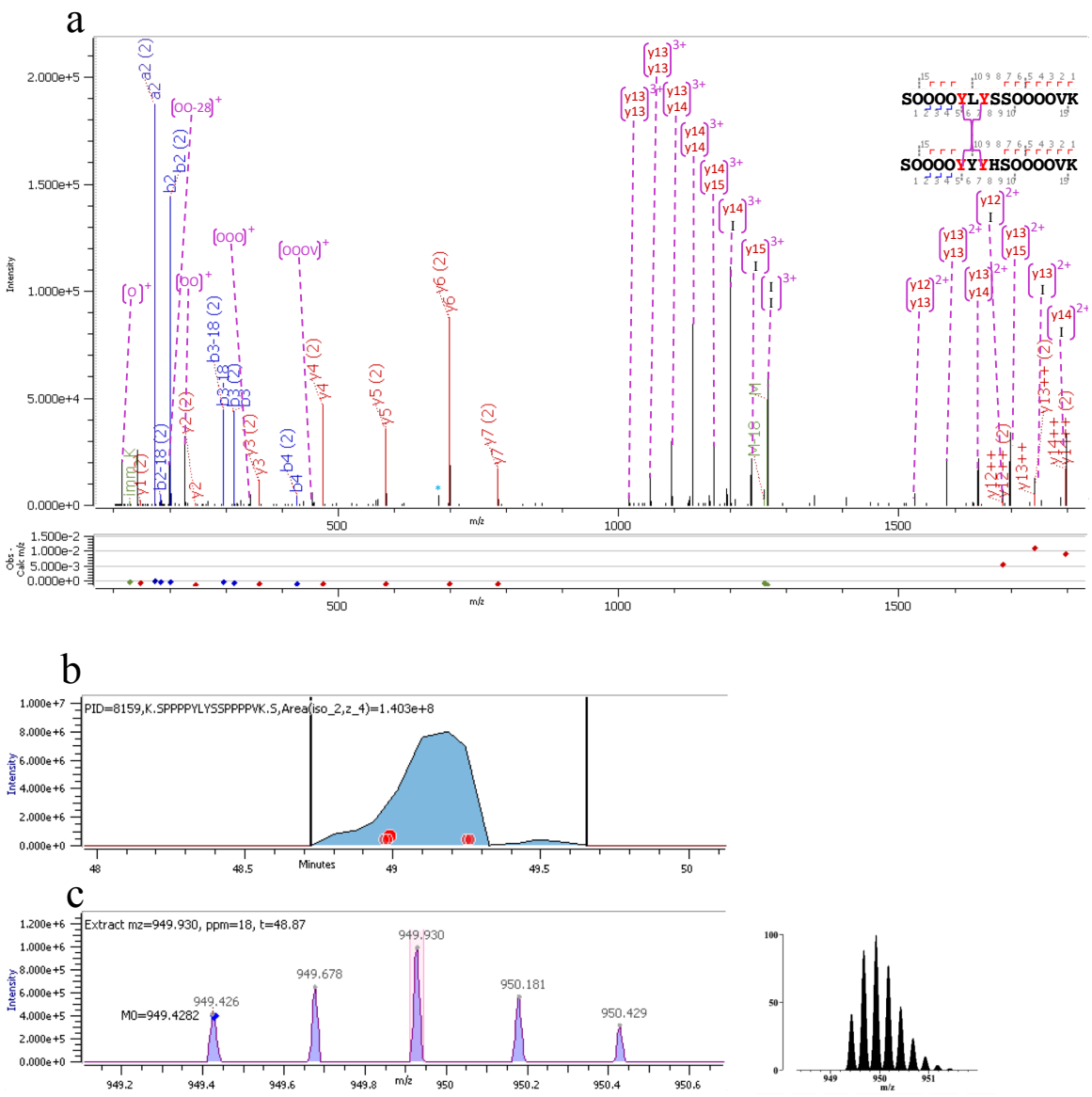

Figure S4.11: Characterization of the intermolecular Di-IDT cross-linkage of two peptides from EXT8. (a) The annotated HCD MS/MS spectrum of the [M+3H]3+ ion at m/z 949.4282 with m/z errors of mass fragments at a 10ppm tolerance denoted below; and (inset, top right) a-, b-, & y-fragment ion coverage. Cross-linked tyrosines carrying the mass loss of (6.040595 Da) are denoted in red. (b) Extracted ion chromatogram. The red dots indicate several MS2 collected. (c) (Left) The isotopic profile of the [M+3H]3+ precursor ion at m/z 949.4282 that was isolated for fragmentation. The blue dot indicates correct monoisotopic assignment based on calculated mass.

(Right) The theoretically calculated isotopic profile of the [M+3H]3+ precursor ion at m/z 949.4282 from MS-Isotope.[68]

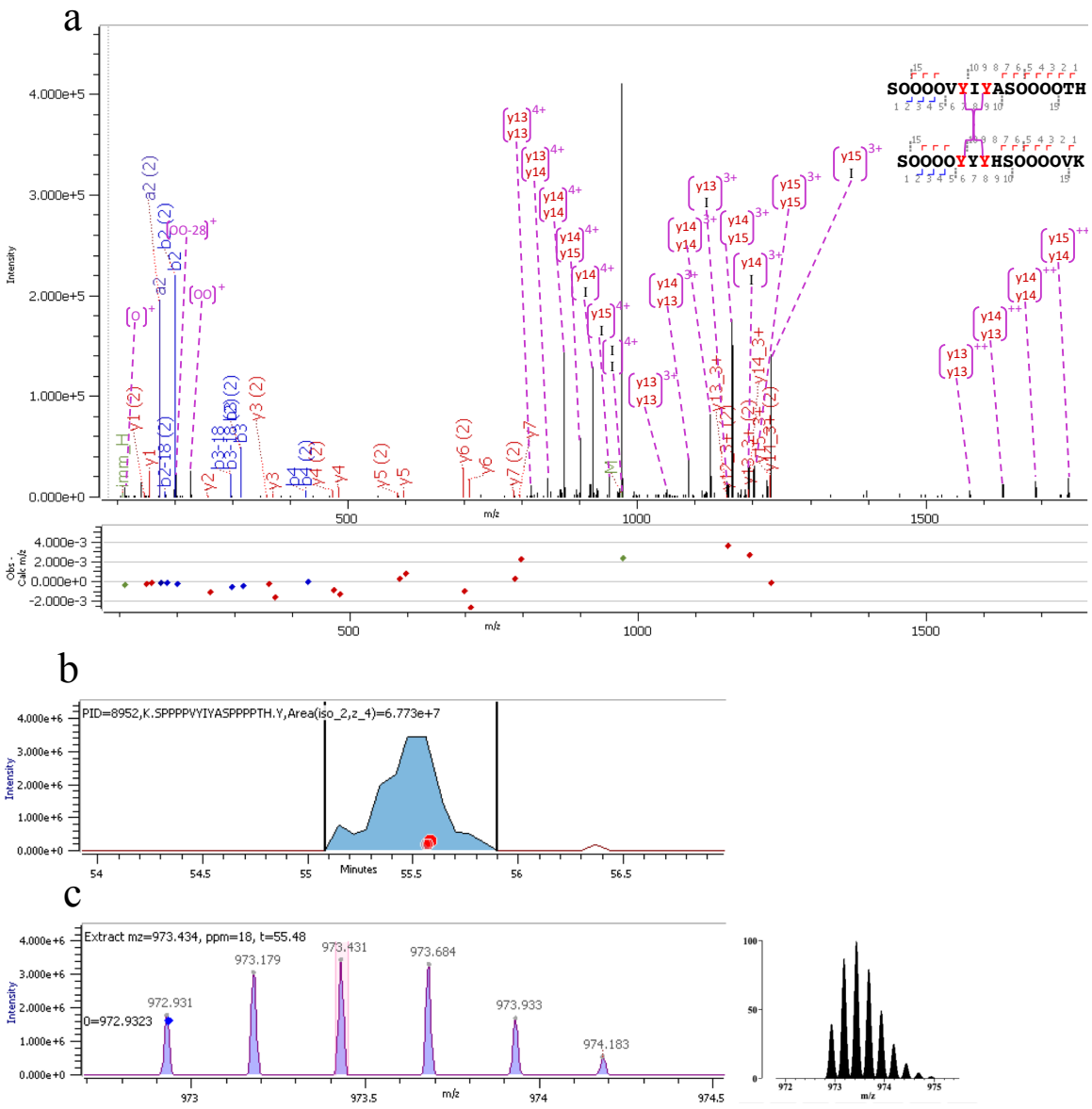

Figure S4.12: Characterization of the intermolecular Di-IDT cross-linkage of two peptides from EXT8. (a) The annotated HCD MS/MS spectrum of the [M+4H]4+ ion at m/z 972.9323 with m/z errors of mass fragments at a 10ppm tolerance denoted below; and (inset, top right) a-, b-, & y-fragment ion coverage. Cross-linked tyrosines carrying the mass loss of (6.040595 Da) are denoted in red. (b) Extracted ion chromatogram. The red dots indicate several MS2 collected. (c) (Left) The isotopic profile of the [M+4H]4+ precursor ion at m/z 972.9323 that was isolated for fragmentation. The blue dot indicates correct monoisotopic assignment based on calculated mass.

(Right) The theoretically calculated isotopic profile of the [M+4H]4+ precursor ion at m/z 972.9323 from MS-Isotope.[68]

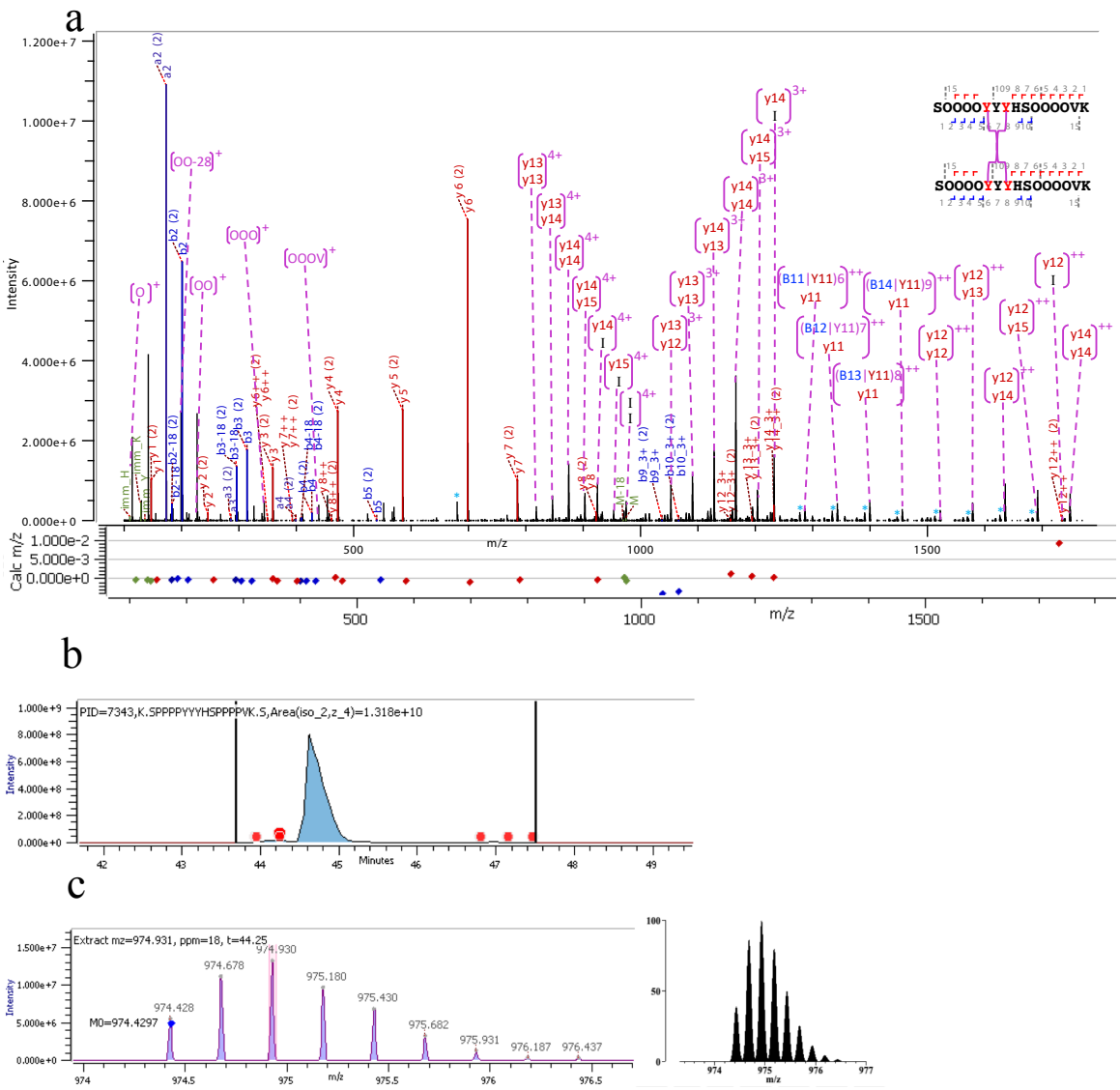

Figure S4.13: Characterization of the intermolecular Di-IDT cross-linkage of two peptides from EXT8. (a) The annotated HCD MS/MS spectrum of the [M+4H]4+ ion at m/z 974.4297 with m/z errors of mass fragments at a 10ppm tolerance denoted below; and (inset, top right) a-, b-, & y-fragment ion coverage. Cross-linked tyrosines carrying the mass loss of (6.040595 Da) are denoted in red. (b) Extracted ion chromatogram. The red dots indicate several MS2 collected. (c) (Left) The isotopic profile of the [M+4H]4+ precursor ion at m/z 974.4297 that was isolated for fragmentation. The blue dot indicates correct monoisotopic assignment based on calculated mass.

(Right) The theoretically calculated isotopic profile of the [M+4H]4+ precursor ion at m/z 974.4297 from MS-Isotope.[68]

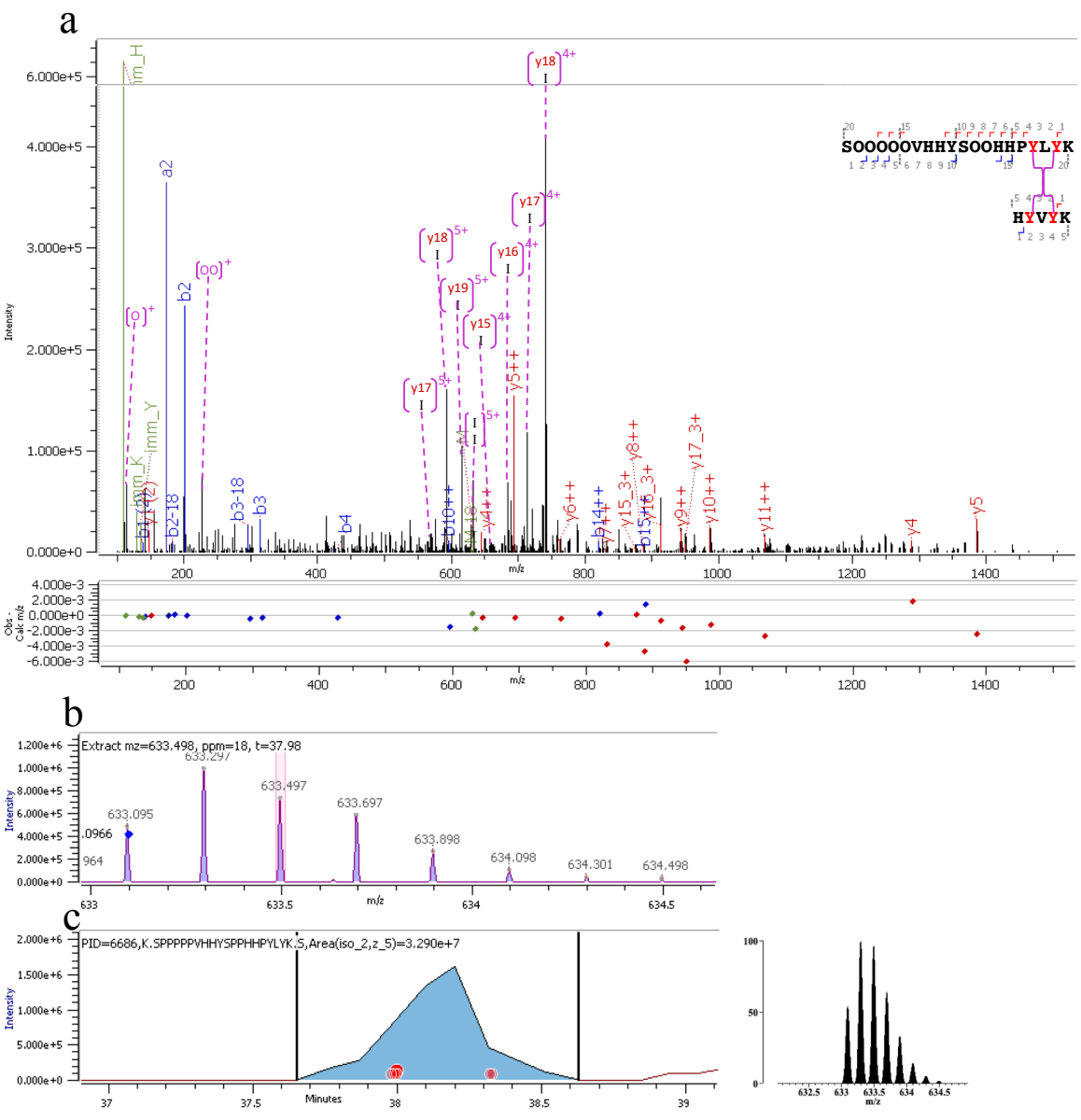

Figure S4.14: Characterization of the intermolecular Di-IDT cross-linkage of two peptides from EXT3. (a) The annotated HCD MS/MS spectrum of the [M+5H]5+ ion at m/z 633.095.0966 with m/z errors of mass fragments at a 10ppm tolerance denoted below; and (inset, top right) a-, b-, & y-fragment ion coverage. Cross-linked tyrosines carrying the mass loss of (6.040595 Da) are denoted in red. (b) Extracted ion chromatogram. The red dots indicate several MS2 collected. (c) (Left) The isotopic profile of the [M+5H]5+ precursor ion at m/z 633.095.0966 that was isolated for fragmentation. The blue dot indicates correct monoisotopic assignment based on calculated

mass. (Right) The theoretically calculated isotopic profile of the [M+5H]5+ precursor ion at m/z 633.095.0966 from MS-Isotope.[68]

Supplemental Figures S5.1 to S5.33: MS Evidence for Non-Cross-Linked Peptides

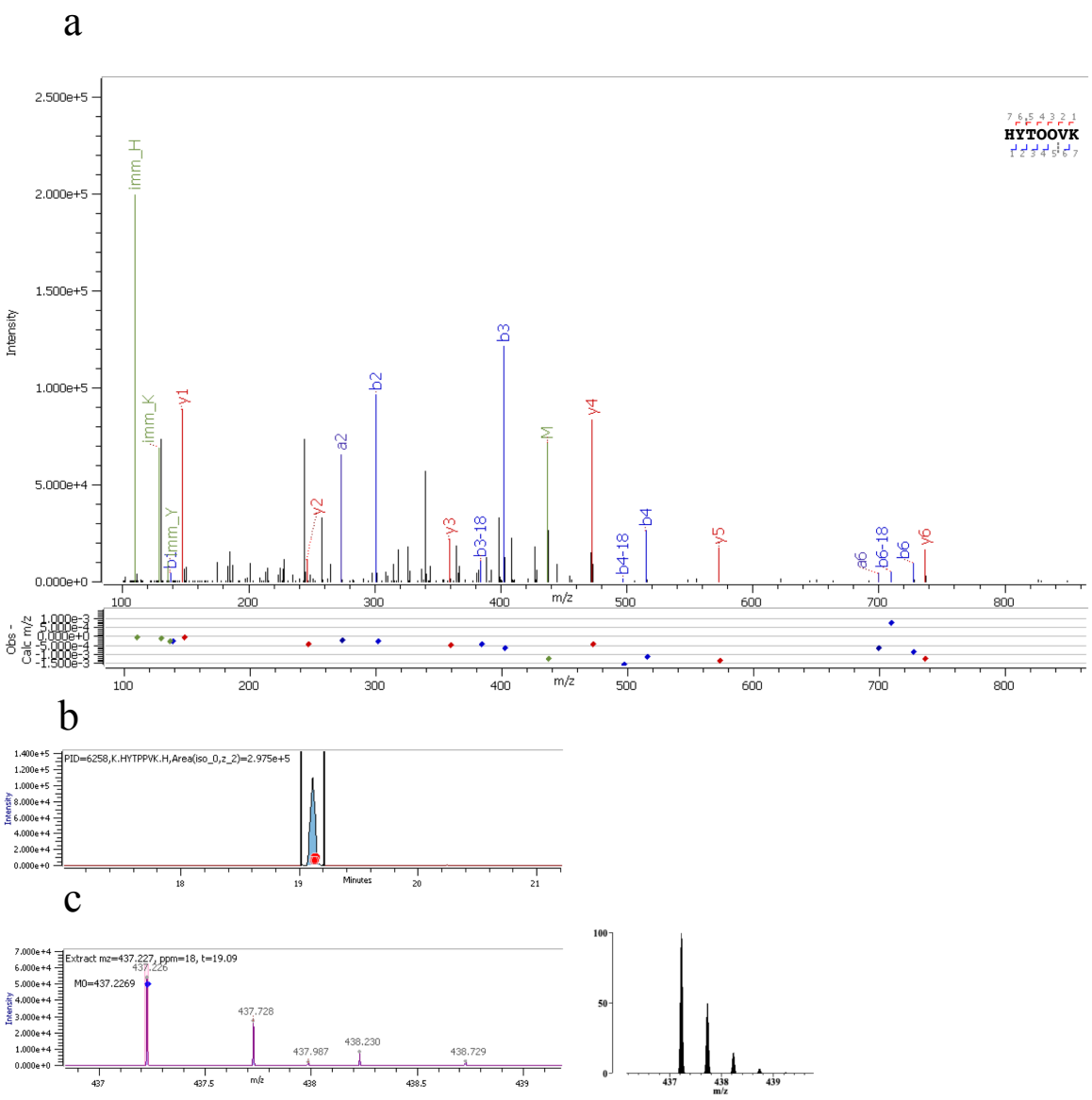

Figure S5.1: Characterization of the non-cross-linked peptide from EXT3. (a) The annotated HCD MS/MS spectrum of the [M+2H]2+ ion at m/z 437.2269 with m/z errors of mass fragments at a 10ppm tolerance denoted below; and (inset, top right) a-, b-, & y-fragment ion coverage. Tyrosines are not cross-linked, resulting in no mass loss. (b) Extracted ion chromatogram. The red dots indicate several MS2 collected. (c) (Left) The isotopic profile of the [M+2H]2+

precursor ion at m/z 437.2269 that was isolated for fragmentation. The blue dot indicates correct monoisotopic assignment based on calculated mass. (Right) The theoretically calculated isotopic profile of the [M+2H]2+ precursor ion at m/z 437.2269 from MS-Isotope.[68]

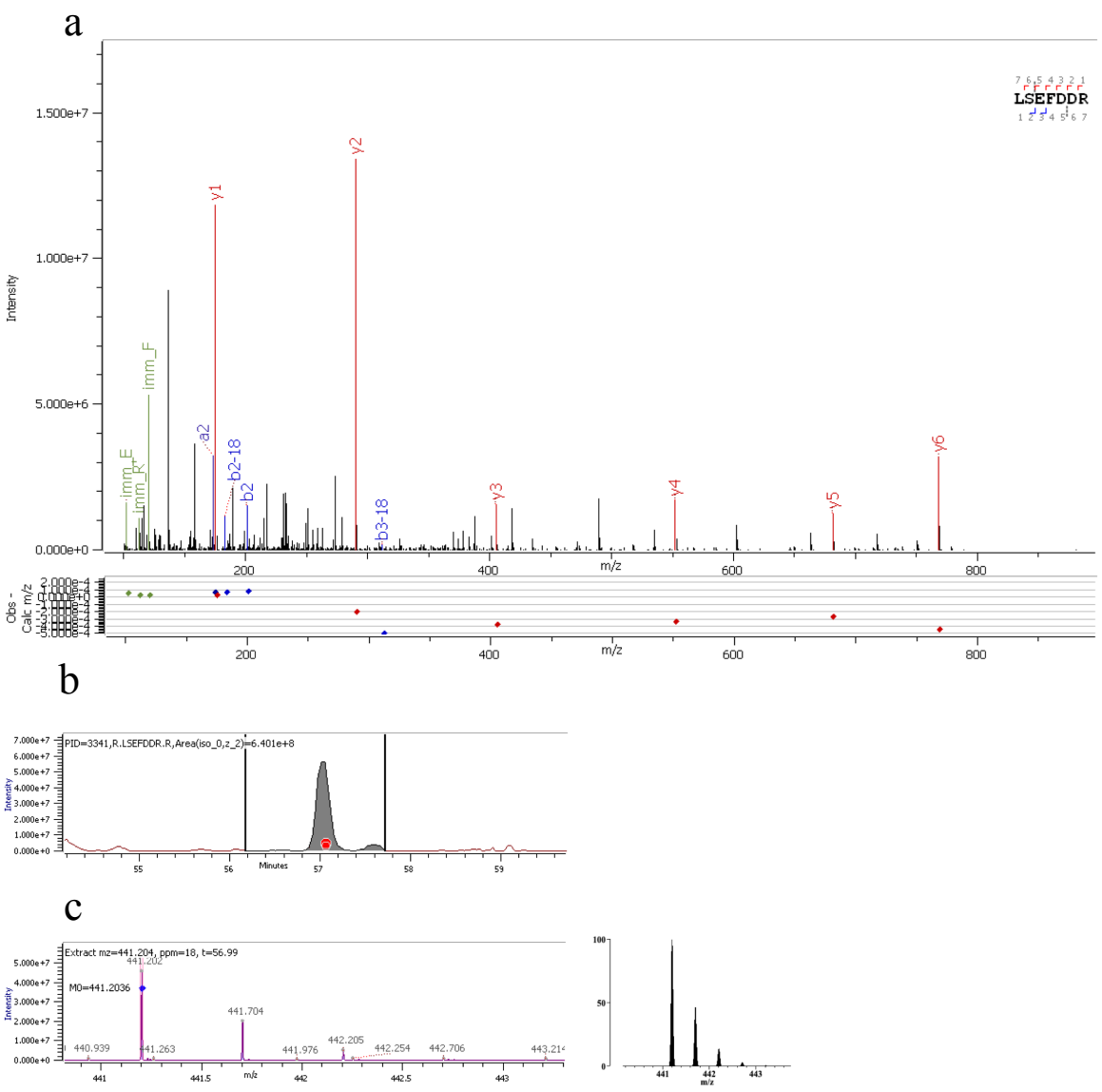

Figure S5.2: Characterization of the non-cross-linked peptide from LRX4. (a) The annotated HCD MS/MS spectrum of the [M+2H]2+ ion at m/z 441.2036 with m/z errors of mass fragments at a 10ppm tolerance denoted below; and (inset, top right) a-, b-, & y-fragment ion coverage. Tyrosines are not cross-linked, resulting in no mass loss. (b) Extracted ion chromatogram. The red dots indicate several MS2 collected. (c) (Left) The isotopic profile of the [M+2H]2+ precursor ion at m/z 441.2036 that was isolated for fragmentation. The blue dot indicates correct

monoisotopic assignment based on calculated mass. (Right) The theoretically calculated isotopic profile of the [M+2H]2+ precursor ion at m/z 441.2036 from MS-Isotope.[68]

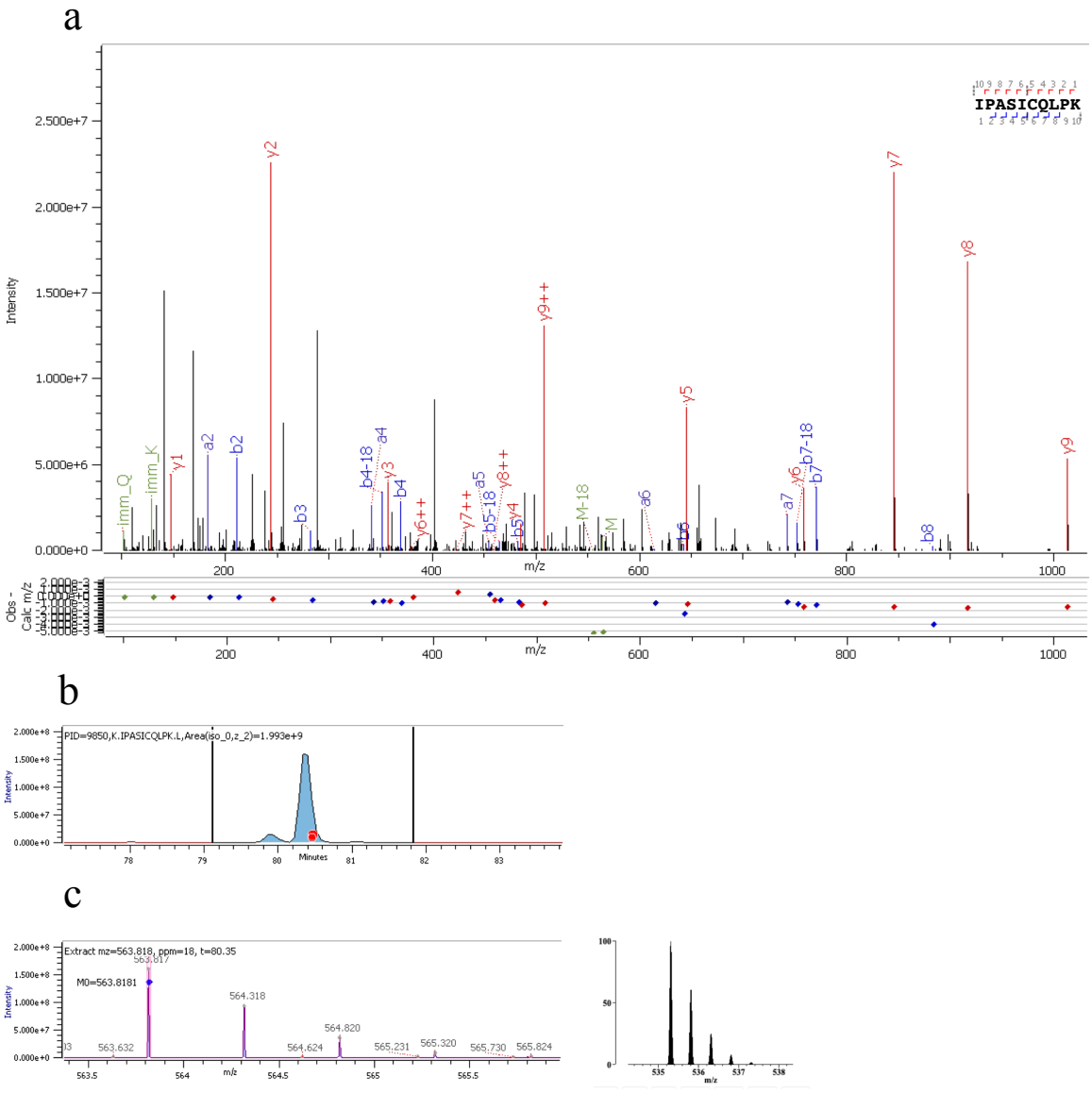

Figure S5.3: Characterization of the non-cross-linked peptide from LRX4. (a) The annotated HCD MS/MS spectrum of the [M+2H]2+ ion at m/z 563.8181 with m/z errors of mass fragments at a 10ppm tolerance denoted below; and (inset, top right) a-, b-, & y-fragment ion coverage. Tyrosines are not cross-linked, resulting in no mass loss. (b) Extracted ion chromatogram. The red dots indicate several MS2 collected. (c) (Left) The isotopic profile of the [M+2H]2+ precursor ion at m/z 563.8181 that was isolated for fragmentation. The blue dot indicates correct

monoisotopic assignment based on calculated mass. (Right) The theoretically calculated isotopic profile of the [M+2H]2+ precursor ion at m/z 563.8181 from MS-Isotope.[68]

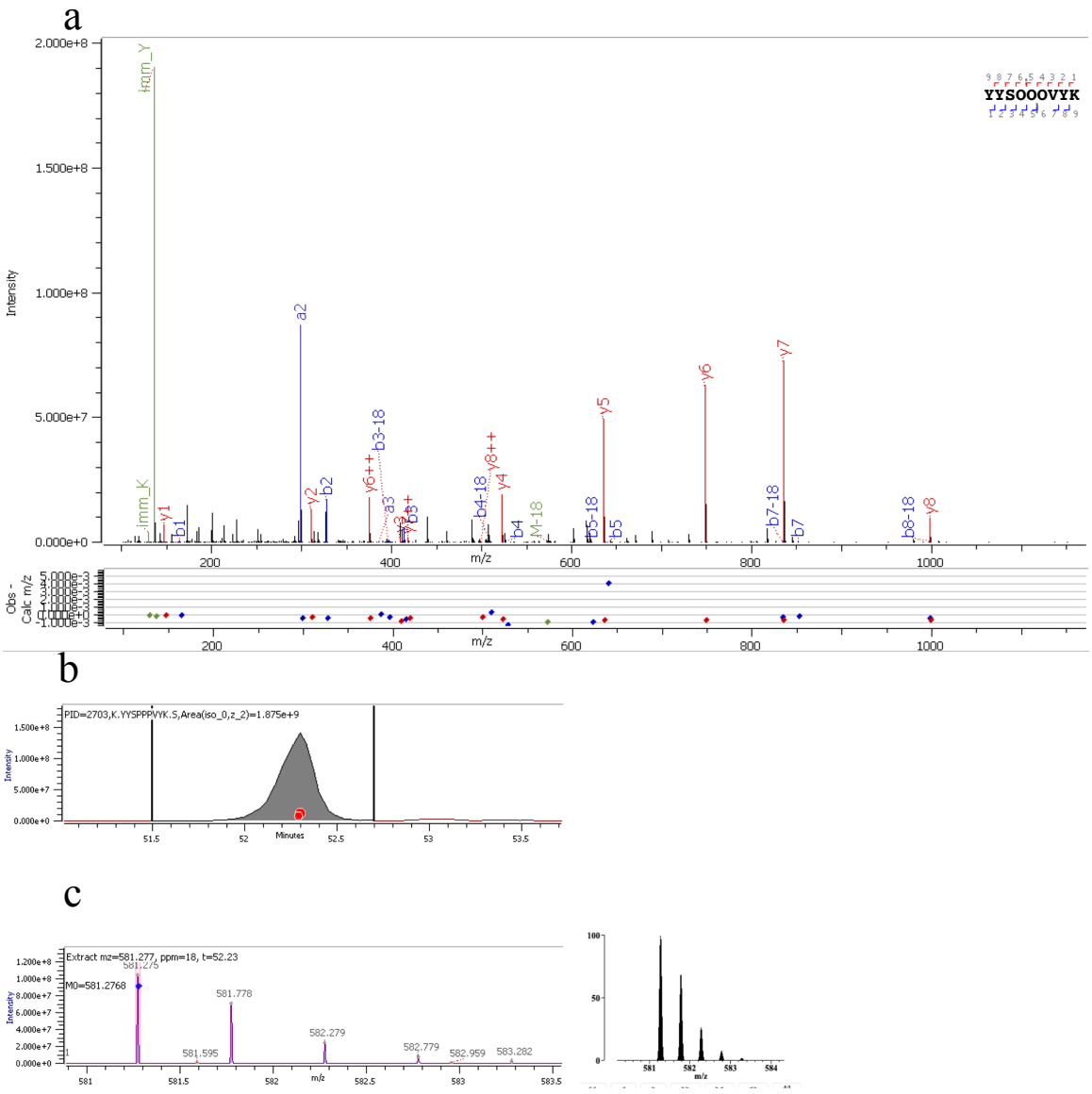

Figure S5.4: Characterization of the non-cross-linked peptide from EXT1. (a) The annotated HCD MS/MS spectrum of the [M+2H]2+ ion at m/z 581.2768 with m/z errors of mass fragments at a 10ppm tolerance denoted below; and (inset, top right) a-, b-, & y-fragment ion coverage. Tyrosines are not cross-linked, resulting in no mass loss. (b) Extracted ion chromatogram. The red dots indicate several MS2 collected. (c) (Left) The isotopic profile of the [M+2H]2+ precursor ion at m/z 581.2768 that was isolated for fragmentation. The blue dot indicates correct

monoisotopic assignment based on calculated mass. (Right) The theoretically calculated isotopic profile of the [M+2H]2+ precursor ion at m/z 581.2768 from MS-Isotope.[68]

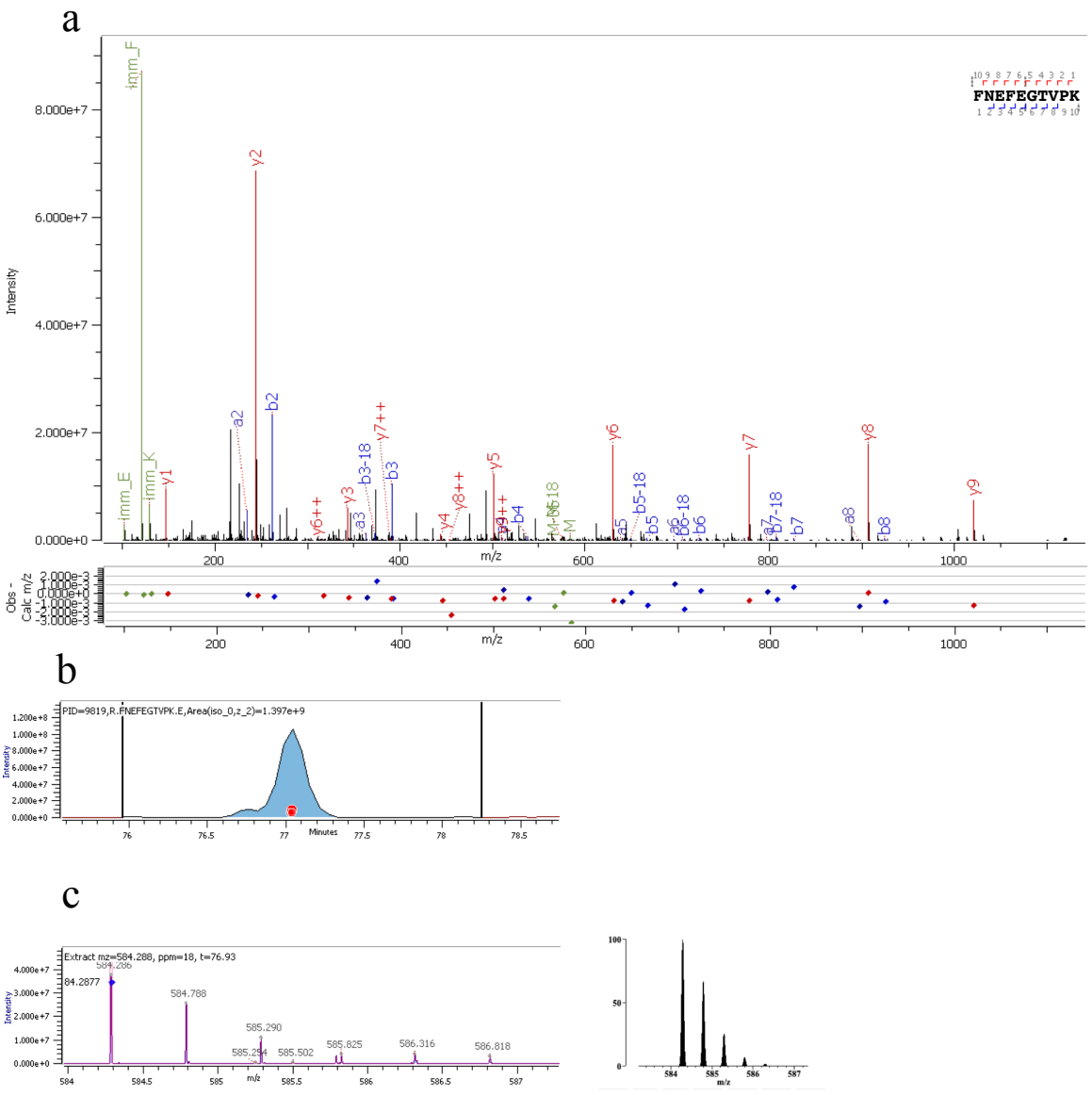

Figure S5.5: Characterization of the non-cross-linked peptide from LRX4. (a) The annotated HCD MS/MS spectrum of the [M+2H]2+ ion at m/z 584.2877 with m/z errors of mass fragments at a 10ppm tolerance denoted below; and (inset, top right) a-, b-, & y-fragment ion coverage. Tyrosines are not cross-linked, resulting in no mass loss. (b) Extracted ion chromatogram. The red dots indicate several MS2 collected. (c) (Left) The isotopic profile of the [M+2H]2+ precursor ion at m/z 584.2877 that was isolated for fragmentation. The blue dot indicates correct

monoisotopic assignment based on calculated mass. (Right) The theoretically calculated isotopic profile of the [M+2H]2+ precursor ion at m/z 584.2877 from MS-Isotope.[68]

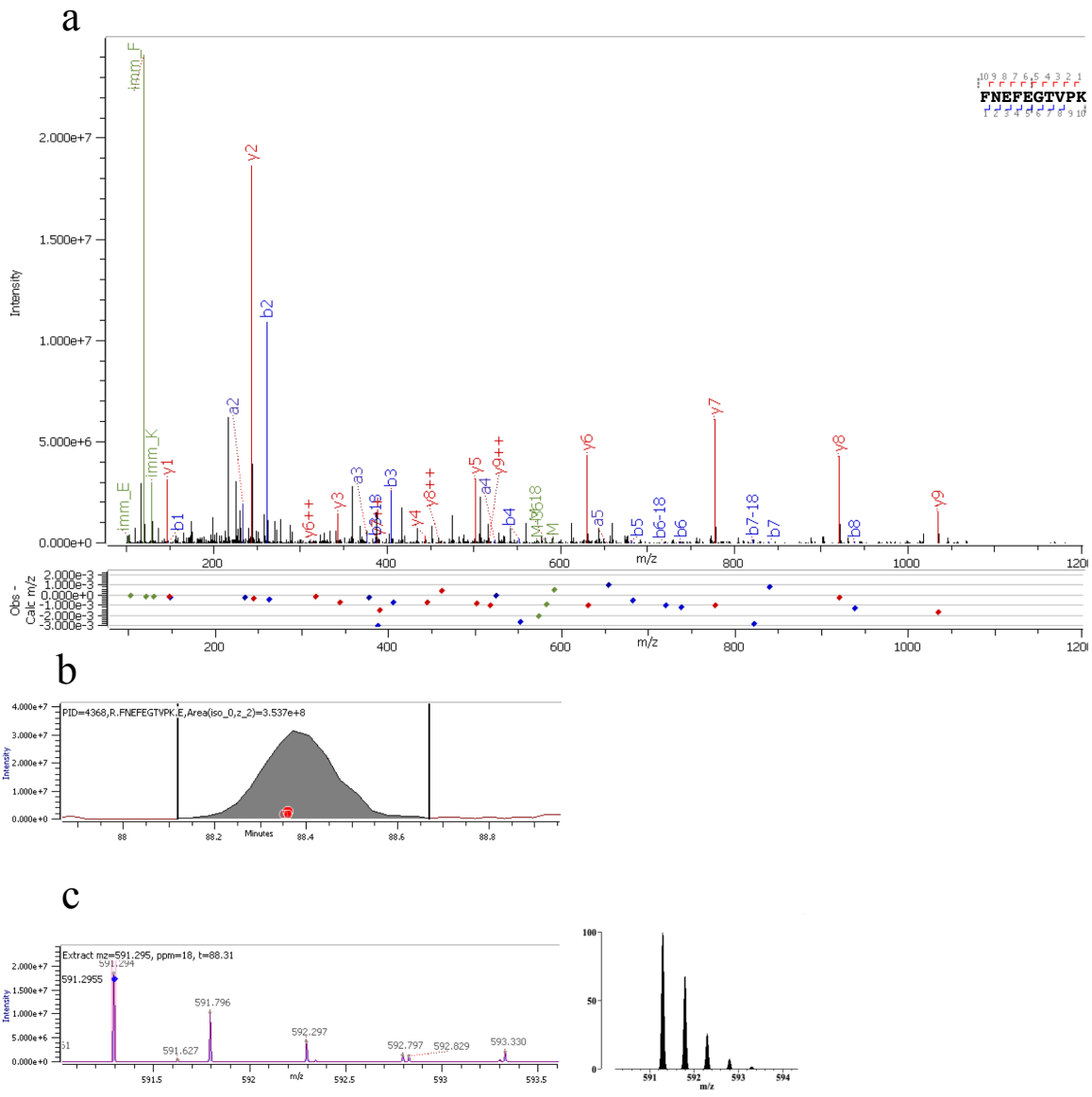

Figure S5.6: Characterization of the non-cross-linked peptide from LRX4. (a) The annotated HCD MS/MS spectrum of the [M+2H]2+ ion at m/z 591.2955 with m/z errors of mass fragments at a 10ppm tolerance denoted below; and (inset, top right) a-, b-, & y-fragment ion coverage. Tyrosines are not cross-linked, resulting in no mass loss. (b) Extracted ion chromatogram. The red dots indicate several MS2 collected. (c) (Left) The isotopic profile of the [M+2H]2+ precursor ion at m/z 591.2955 that was isolated for fragmentation. The blue dot indicates correct

monoisotopic assignment based on calculated mass. (Right) The theoretically calculated isotopic profile of the [M+2H]2+ precursor ion at m/z 591.2955 from MS-Isotope.[68]

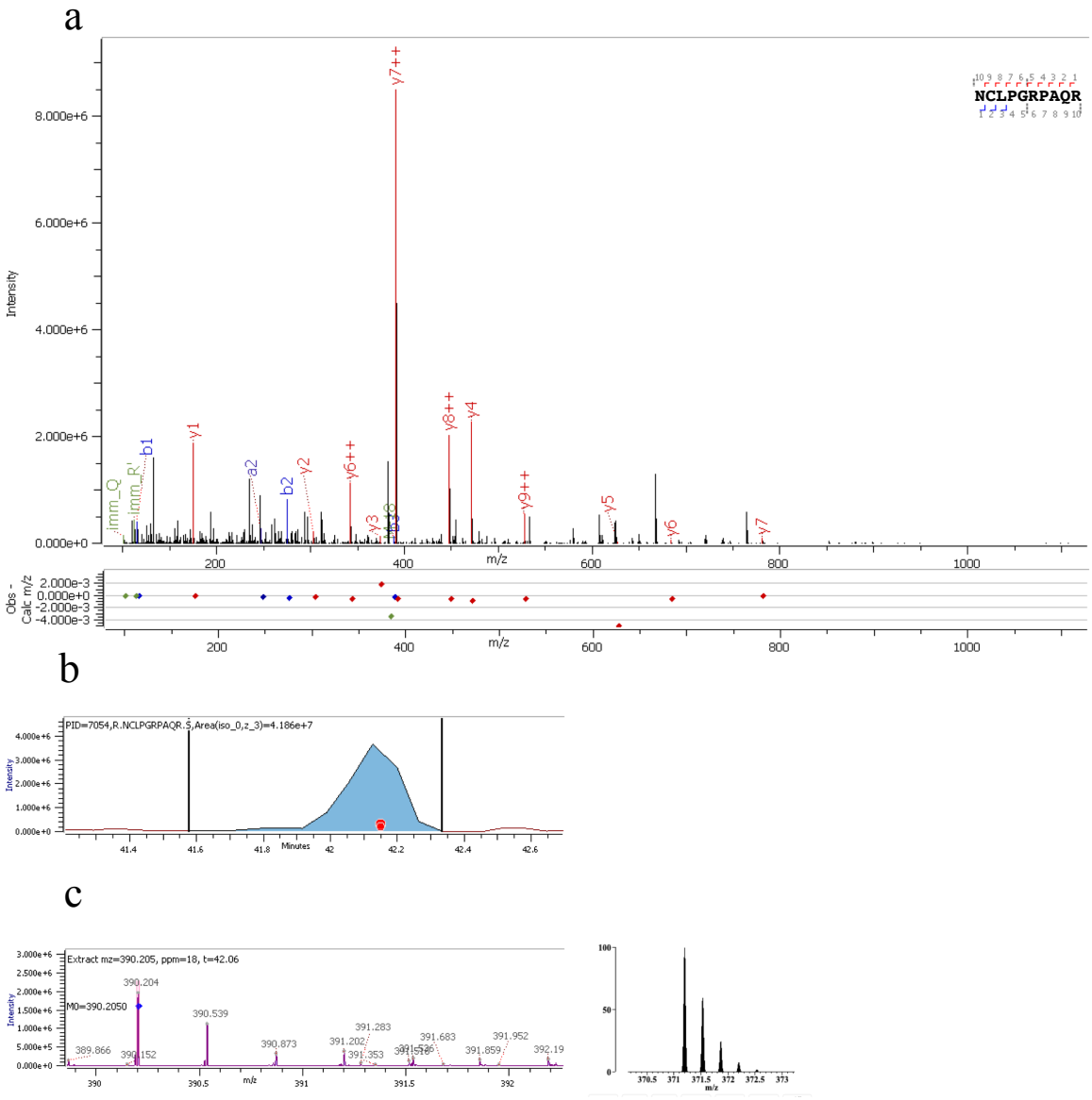

Figure S5.7: Characterization of the non-cross-linked peptide from LRX4. (a) The annotated HCD MS/MS spectrum of the [M+3H]3+ ion at m/z 390.205 with m/z errors of mass fragments at a 10ppm tolerance denoted below; and (inset, top right) a-, b-, & y-fragment ion coverage. Tyrosines are not cross-linked, resulting in no mass loss. (b) Extracted ion chromatogram. The red dots indicate several MS2 collected. (c) (Left) The isotopic profile of the [M+3H]3+ precursor ion at m/z 390.205 that was isolated for fragmentation. The blue dot indicates correct

monoisotopic assignment based on calculated mass. (Right) The theoretically calculated isotopic profile of the [M+3H]3+ precursor ion at m/z 390.205 from MS-Isotope.[68]

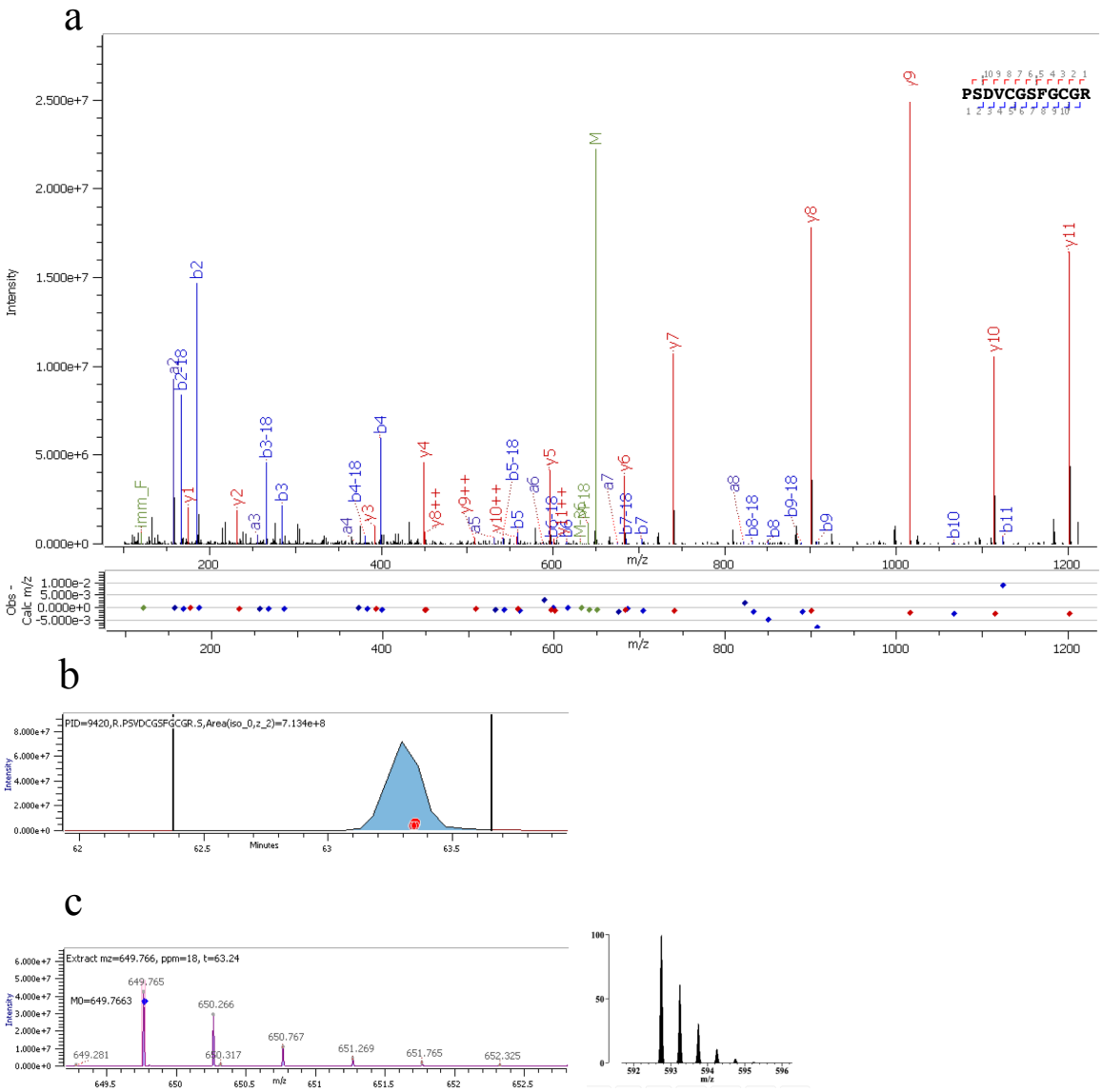

Figure S5.8: Characterization of the non-cross-linked peptide from LRX4. (a) The annotated HCD MS/MS spectrum of the [M+2H]2+ ion at m/z 649.7663 with m/z errors of mass fragments at a 10ppm tolerance denoted below; and (inset, top right) a-, b-, & y-fragment ion coverage. Tyrosines are not cross-linked, resulting in no mass loss. (b) Extracted ion chromatogram. The red dots indicate several MS2 collected. (c) (Left) The isotopic profile of the [M+2H]2+ precursor ion at m/z 649.7663 that was isolated for fragmentation. The blue dot indicates correct

monoisotopic assignment based on calculated mass. (Right) The theoretically calculated isotopic profile of the [M+2H]2+ precursor ion at m/z 649.7663 from MS-Isotope.[68]

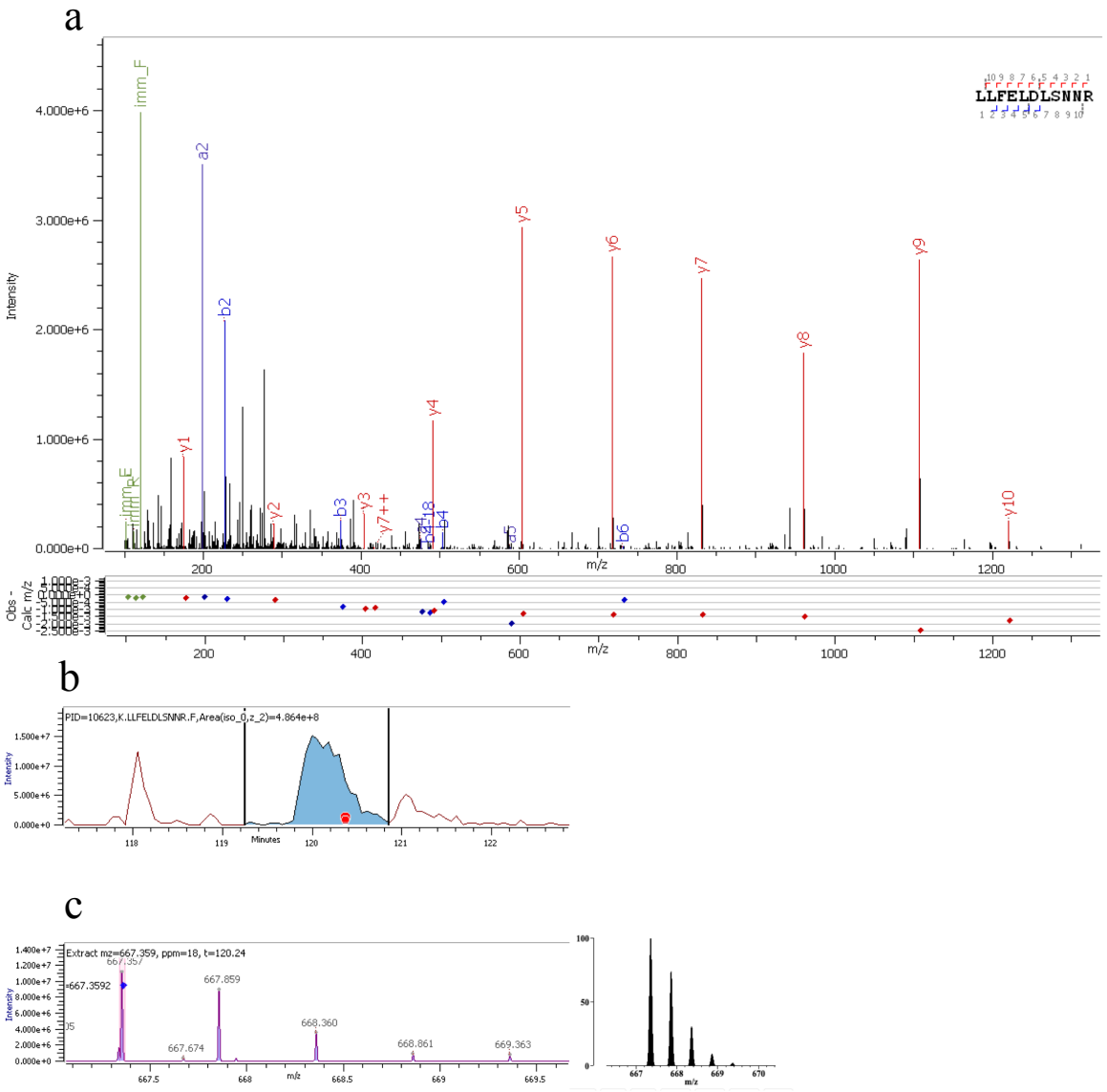

Figure S5.9: Characterization of the non-cross-linked peptide from LRX4. (a) The annotated HCD MS/MS spectrum of the [M+2H]2+ ion at m/z 667.3592 with m/z errors of mass fragments at a 10ppm tolerance denoted below; and (inset, top right) a-, b-, & y-fragment ion coverage. Tyrosines are not cross-linked, resulting in no mass loss. (b) Extracted ion chromatogram. The red dots indicate several MS2 collected. (c) (Left) The isotopic profile of the [M+2H]2+ precursor ion at m/z 667.3592 that was isolated for fragmentation. The blue dot indicates correct

monoisotopic assignment based on calculated mass. (Right) The theoretically calculated isotopic profile of the [M+2H]2+ precursor ion at m/z 667.3592 from MS-Isotope.[68]

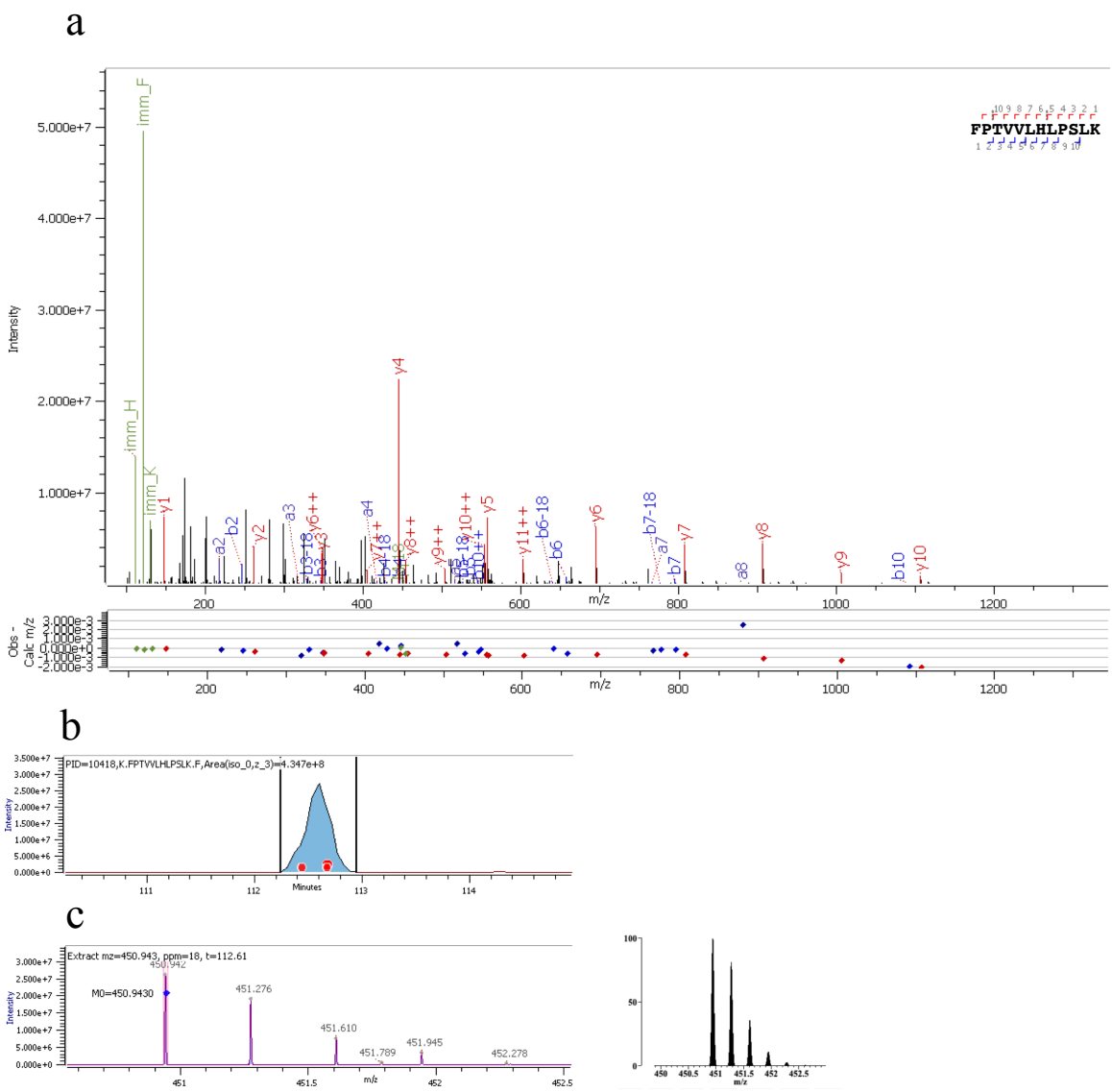

Figure S5.10: Characterization of the non-cross-linked peptide from LRX4. (a) The annotated HCD MS/MS spectrum of the [M+3H]3+ ion at m/z 450.943 with m/z errors of mass fragments at a 10ppm tolerance denoted below; and (inset, top right) a-, b-, & y-fragment ion coverage. Tyrosines are not cross-linked, resulting in no mass loss. (b) Extracted ion chromatogram. The red dots indicate several MS2 collected. (c) (Left) The isotopic profile of the [M+3H]3+ precursor ion at m/z 450.943 that was isolated for fragmentation. The blue dot indicates correct

monoisotopic assignment based on calculated mass. (Right) The theoretically calculated isotopic profile of the [M+3H]3+ precursor ion at m/z 450.943 from MS-Isotope.[68]

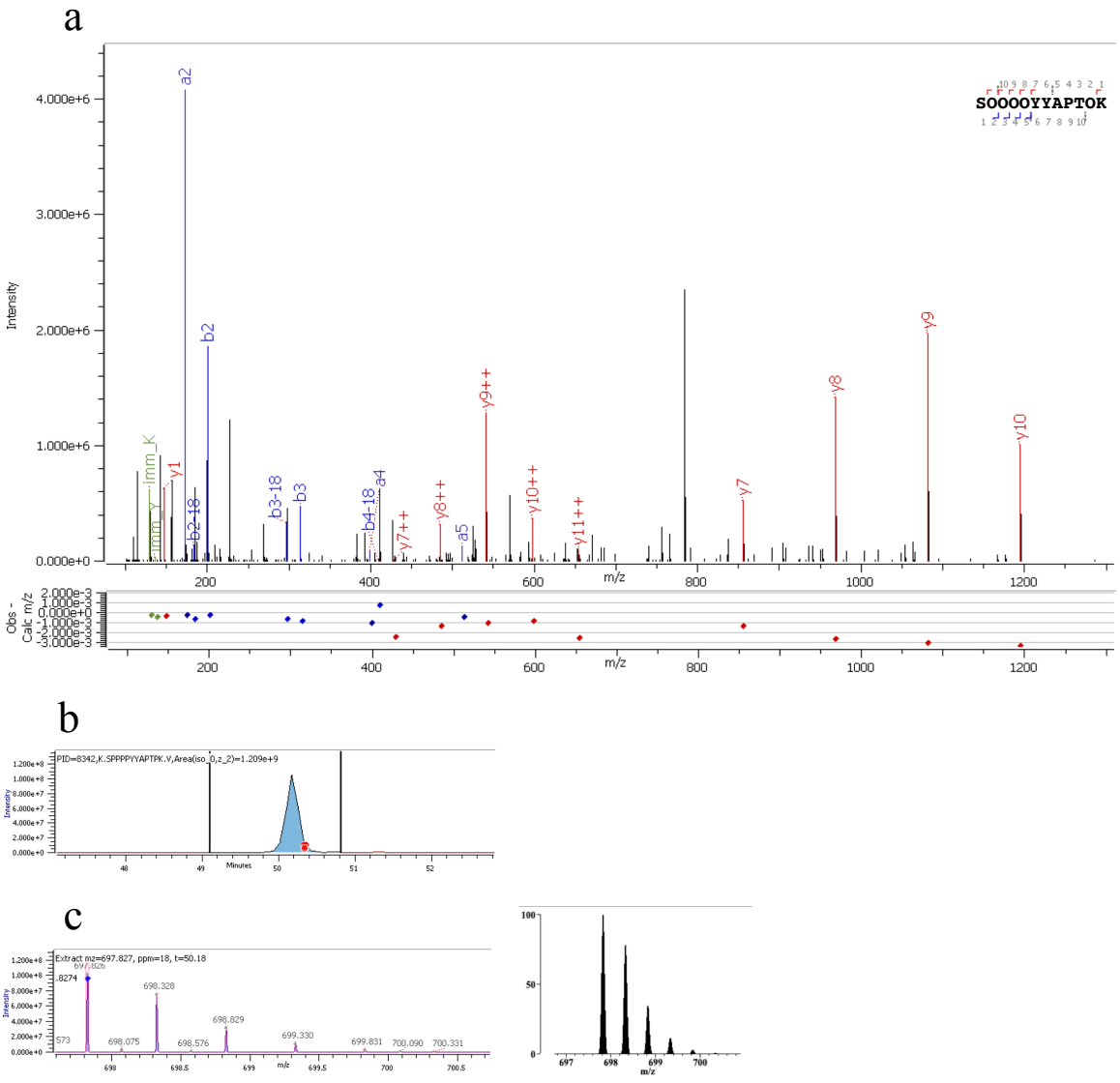

Figure S5.11: Characterization of the non-cross-linked peptide from EXT10. (a) The annotated HCD MS/MS spectrum of the [M+2H]2+ ion at m/z 697.8274 with m/z errors of mass fragments at a 10ppm tolerance denoted below; and (inset, top right) a-, b-, & y-fragment ion coverage. Tyrosines are not cross-linked, resulting in no mass loss. (b) Extracted ion chromatogram. The red dots indicate several MS2 collected. (c) (Left) The isotopic profile of the [M+2H]2+ precursor ion at m/z 697.8274 that was isolated for fragmentation. The blue dot indicates correct

monoisotopic assignment based on calculated mass. (Right) The theoretically calculated isotopic profile of the [M+2H]2+ precursor ion at m/z 697.8274 from MS-Isotope.[68]

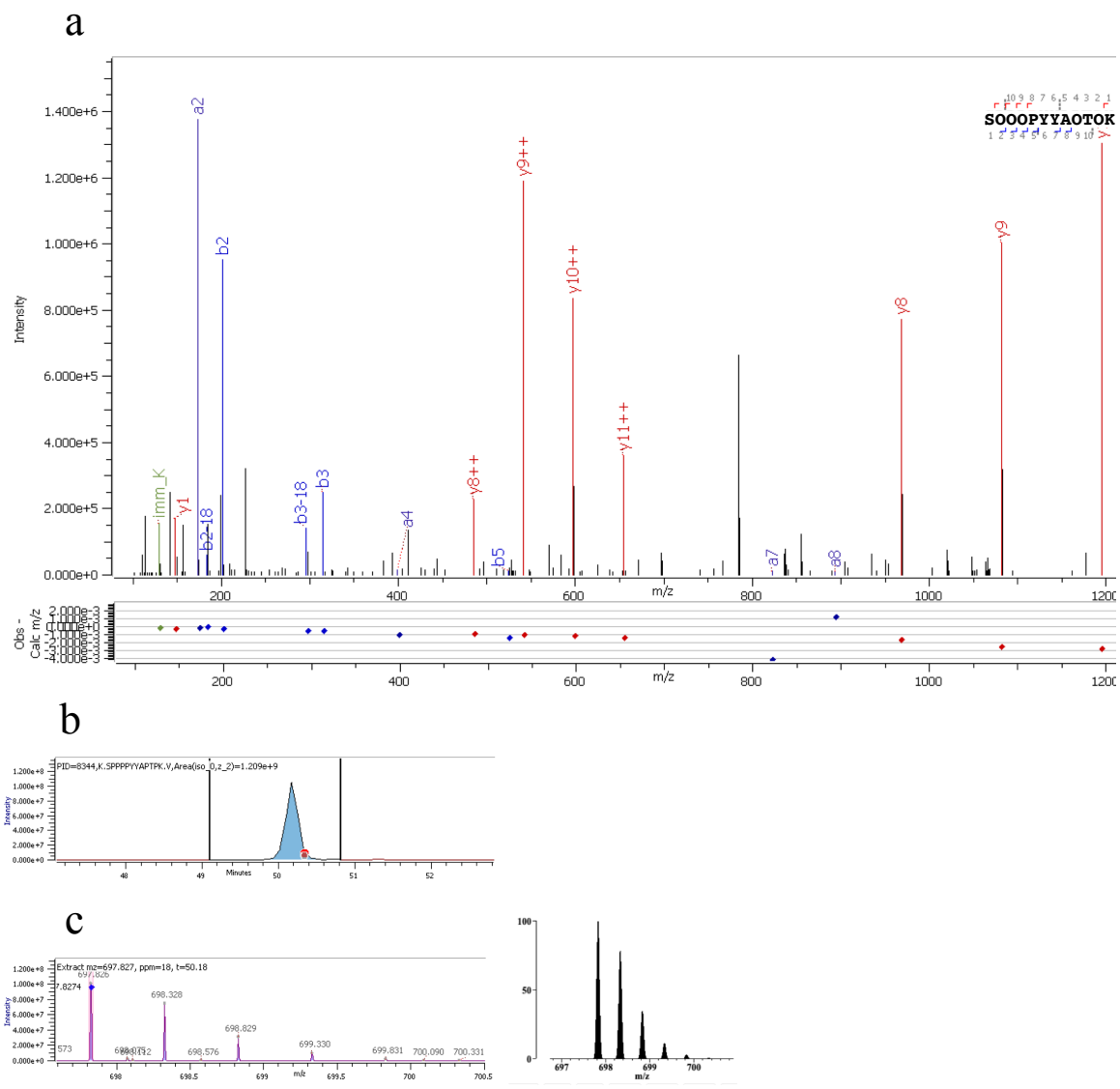

Figure S5.12: Characterization of the non-cross-linked peptide from EXT10. (a) The annotated HCD MS/MS spectrum of the [M+2H]2+ ion at m/z 697.8274 with m/z errors of mass fragments at a 10ppm tolerance denoted below; and (inset, top right) a-, b-, & y-fragment ion coverage. Tyrosines are not cross-linked, resulting in no mass loss. (b) Extracted ion chromatogram. The red dots indicate several MS2 collected. The grey dots indicate the time point of a MS2 of the same mass, but with a different ID such as a different site of nonhydroxylated proline. (c) (Left) The isotopic profile of the [M+2H]2+ precursor ion at m/z 697.8274 that was isolated for

fragmentation. The blue dot indicates correct monoisotopic assignment based on calculated mass. (Right) The theoretically calculated isotopic profile of the [M+2H]2+ precursor ion at m/z 697.8274 from MS-Isotope.[68]

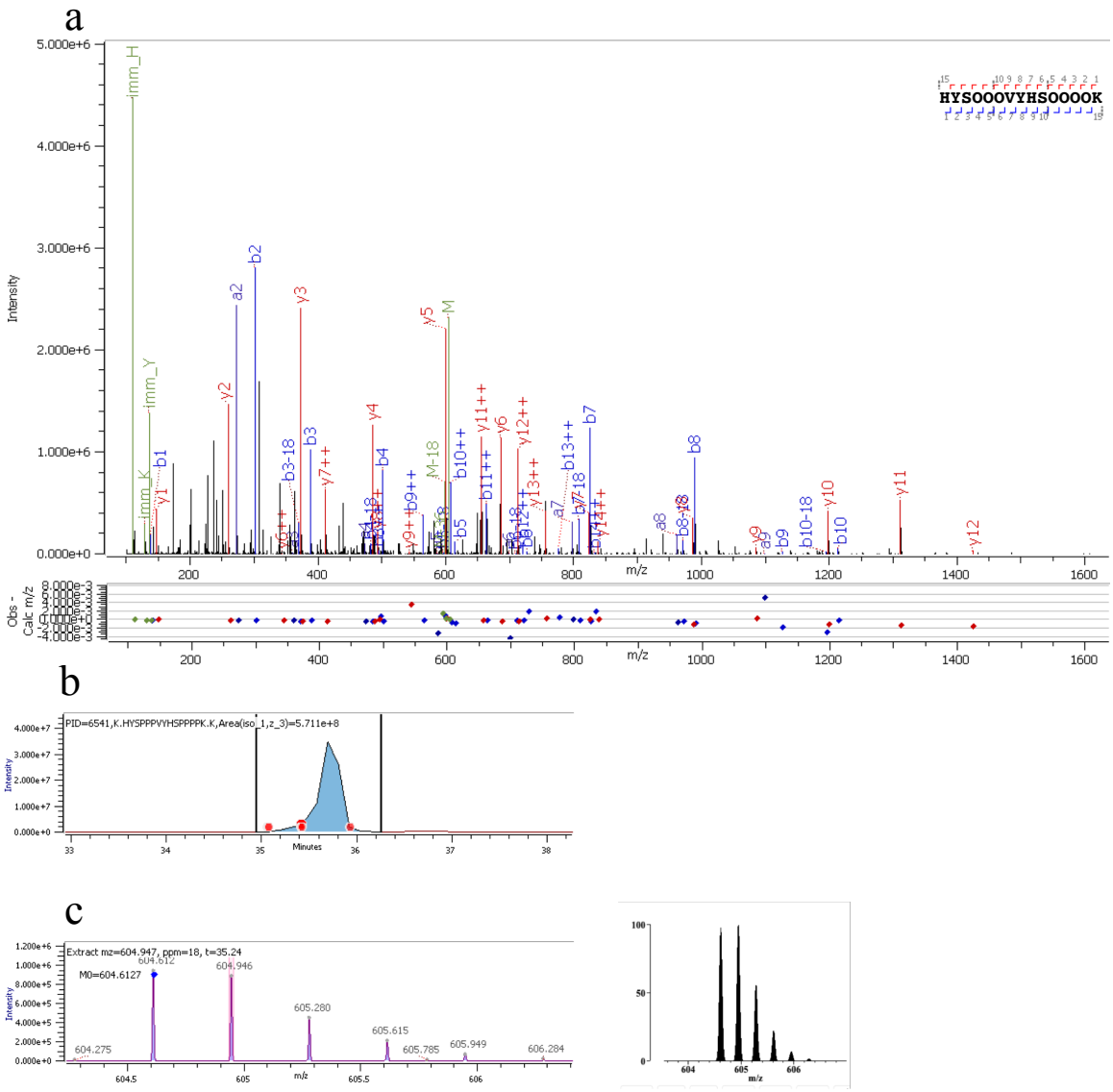

Figure S5.13: Characterization of the non-cross-linked peptide from EXT3. (a) The annotated HCD MS/MS spectrum of the [M+3H]3+ ion at m/z 604.6127 with m/z errors of mass fragments at a 10ppm tolerance denoted below; and (inset, top right) a-, b-, & y-fragment ion coverage. Tyrosines are not cross-linked, resulting in no mass loss. (b) Extracted ion chromatogram. The red dots indicate several MS2 collected. (c) (Left) The isotopic profile of the [M+3H]3+ precursor ion at m/z 604.6127 that was isolated for fragmentation. The blue dot indicates correct

monoisotopic assignment based on calculated mass. (Right) The theoretically calculated isotopic profile of the [M+3H]3+ precursor ion at m/z 604.6127 from MS-Isotope.[68]

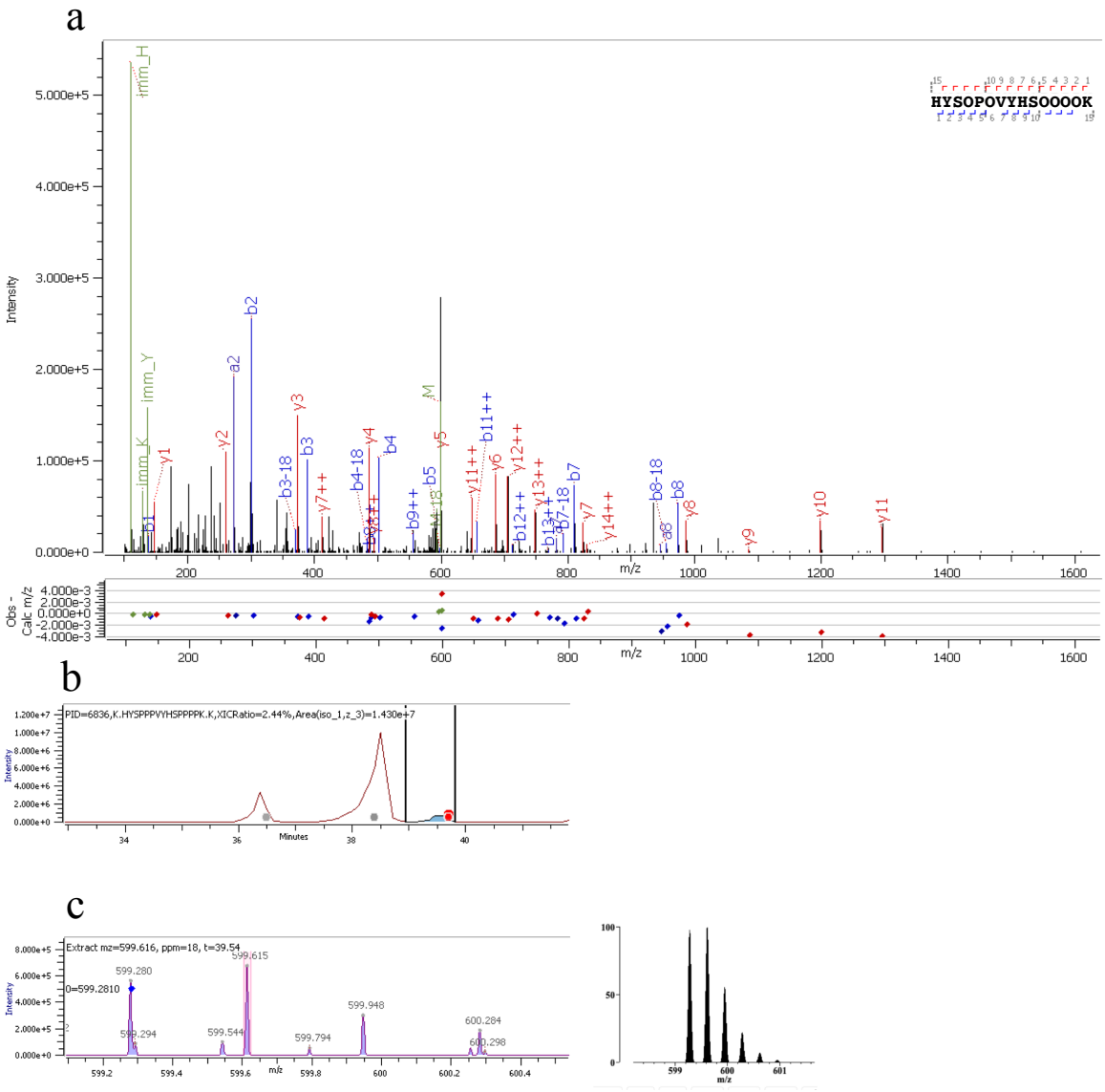

Figure S5.14: Characterization of the non-cross-linked peptide from EXT3. (a) The annotated HCD MS/MS spectrum of the [M+3H]3+ ion at m/z 599.281 with m/z errors of mass fragments at a 10ppm tolerance denoted below; and (inset, top right) a-, b-, & y-fragment ion coverage. Tyrosines are not cross-linked, resulting in no mass loss. (b) Extracted ion chromatogram. The red dots indicate several MS2 collected. The grey dots indicate the time point of a MS2 of the same mass, but with a different ID such as a different site of nonhydroxylated proline. (c) (Left) The isotopic profile of the [M+3H]3+ precursor ion at m/z 599.281 that was isolated for

fragmentation. The blue dot indicates correct monoisotopic assignment based on calculated mass. (Right) The theoretically calculated isotopic profile of the [M+3H]3+ precursor ion at m/z 599.281 from MS-Isotope.[68]

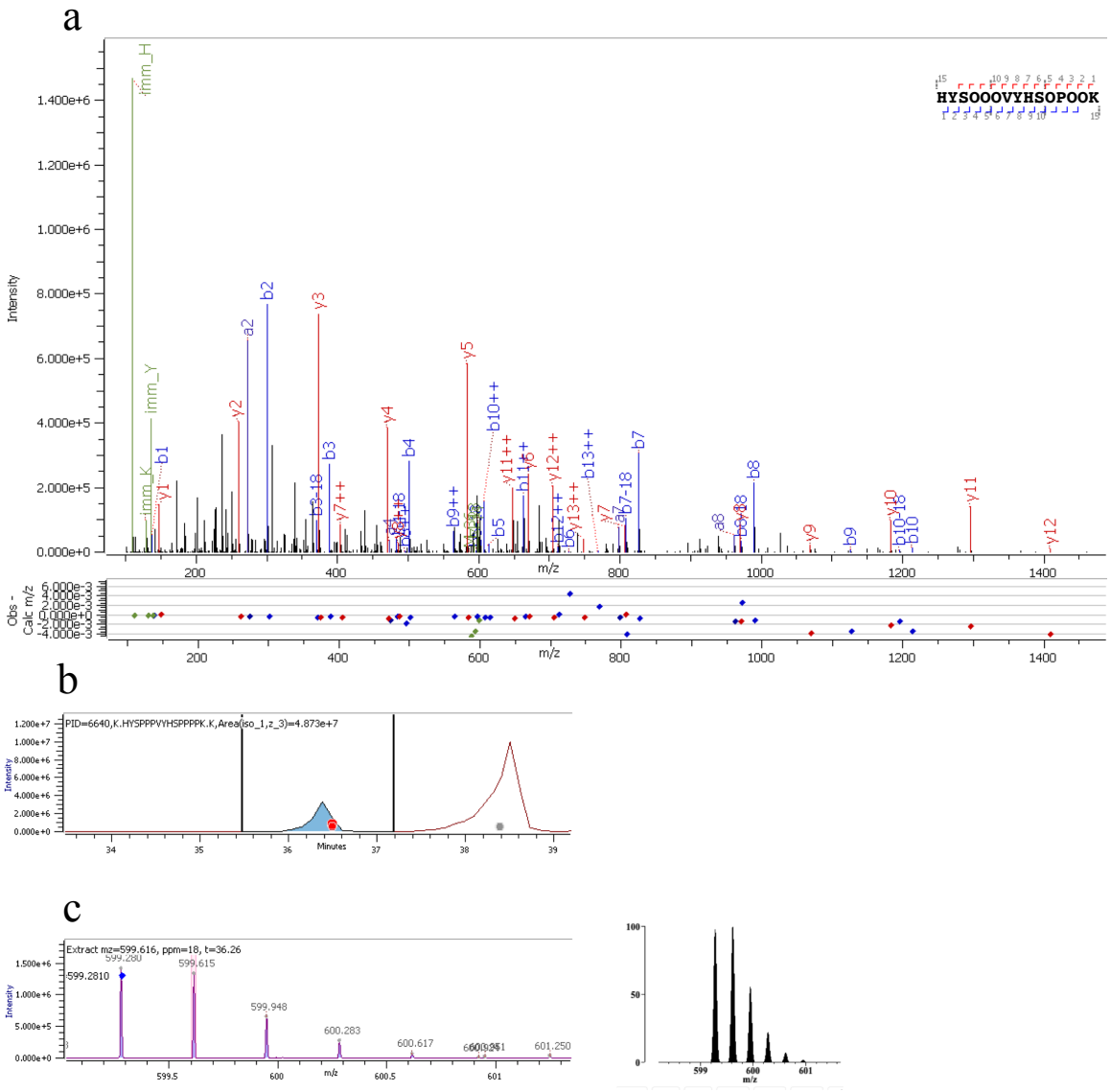

Figure S5.15: Characterization of the non-cross-linked peptide from EXT3. (a) The annotated HCD MS/MS spectrum of the [M+3H]3+ ion at m/z 599.281 with m/z errors of mass fragments at a 10ppm tolerance denoted below; and (inset, top right) a-, b-, & y-fragment ion coverage. Tyrosines are not cross-linked, resulting in no mass loss. (b) Extracted ion chromatogram. The red dots indicate several MS2 collected. The grey dots indicate the time point of a MS2 of the same mass, but with a different ID such as a different site of nonhydroxylated proline. (c) (Left) The isotopic profile of the [M+3H]3+ precursor ion at m/z 599.281 that was isolated for

fragmentation. The blue dot indicates correct monoisotopic assignment based on calculated mass. (Right) The theoretically calculated isotopic profile of the [M+3H]3+ precursor ion at m/z 599.281 from MS-Isotope.[68]

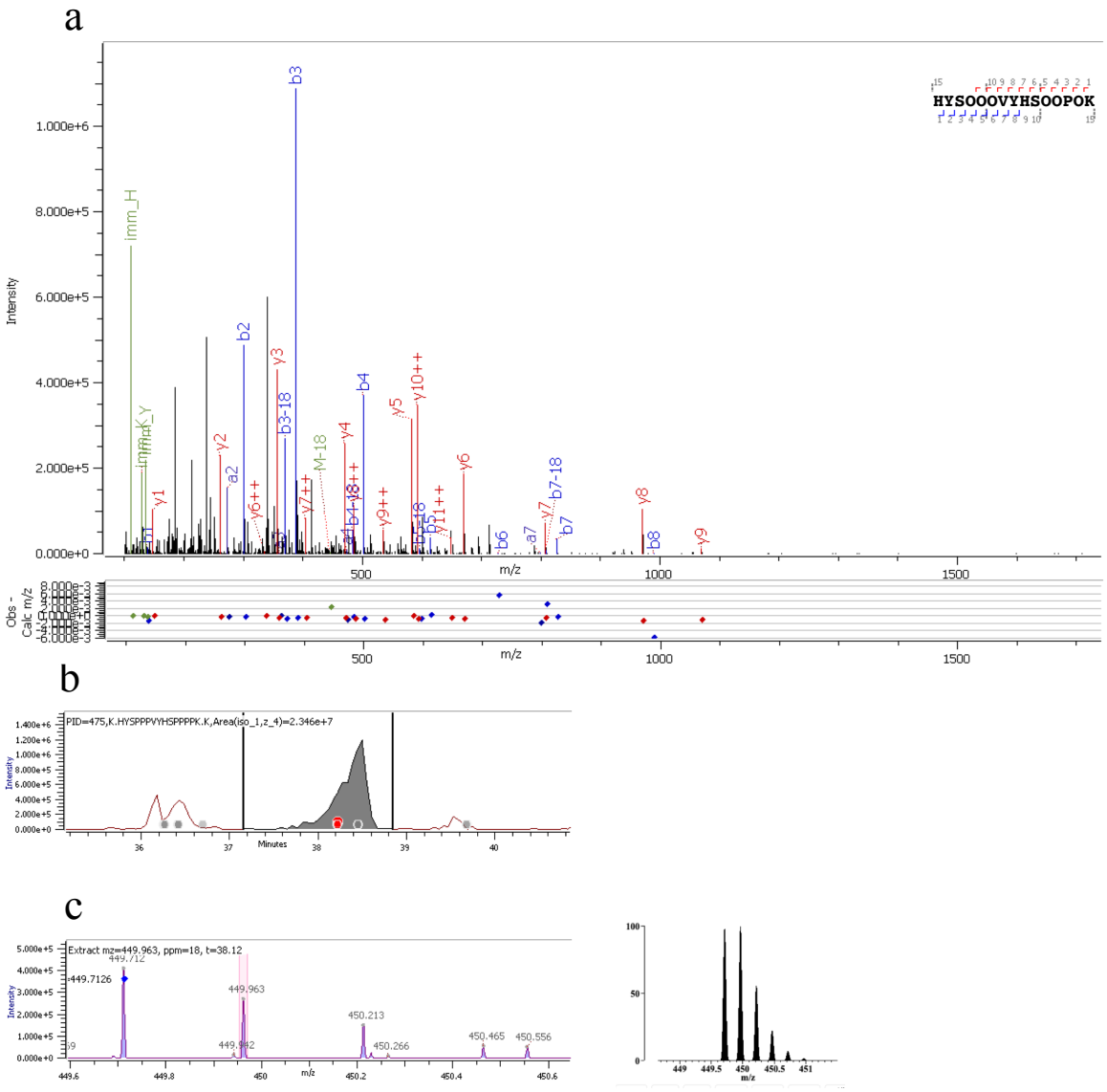

Figure S5.16: Characterization of the non-cross-linked peptide from EXT3. (a) The annotated HCD MS/MS spectrum of the [M+4H]4+ ion at m/z 499.7126 with m/z errors of mass fragments at a 10ppm tolerance denoted below; and (inset, top right) a-, b-, & y-fragment ion coverage. Tyrosines are not cross-linked, resulting in no mass loss. (b) Extracted ion chromatogram. The red dots indicate several MS2 collected. The grey dots indicate the time point of a MS2 of the same mass, but with a different ID such as a different site of nonhydroxylated proline. (c) (Left) The isotopic profile of the [M+4H]4+ precursor ion at m/z 499.7126 that was isolated for

fragmentation. The blue dot indicates correct monoisotopic assignment based on calculated mass. (Right) The theoretically calculated isotopic profile of the [M+4H]4+ precursor ion at m/z 499.7126 from MS-Isotope.[68]

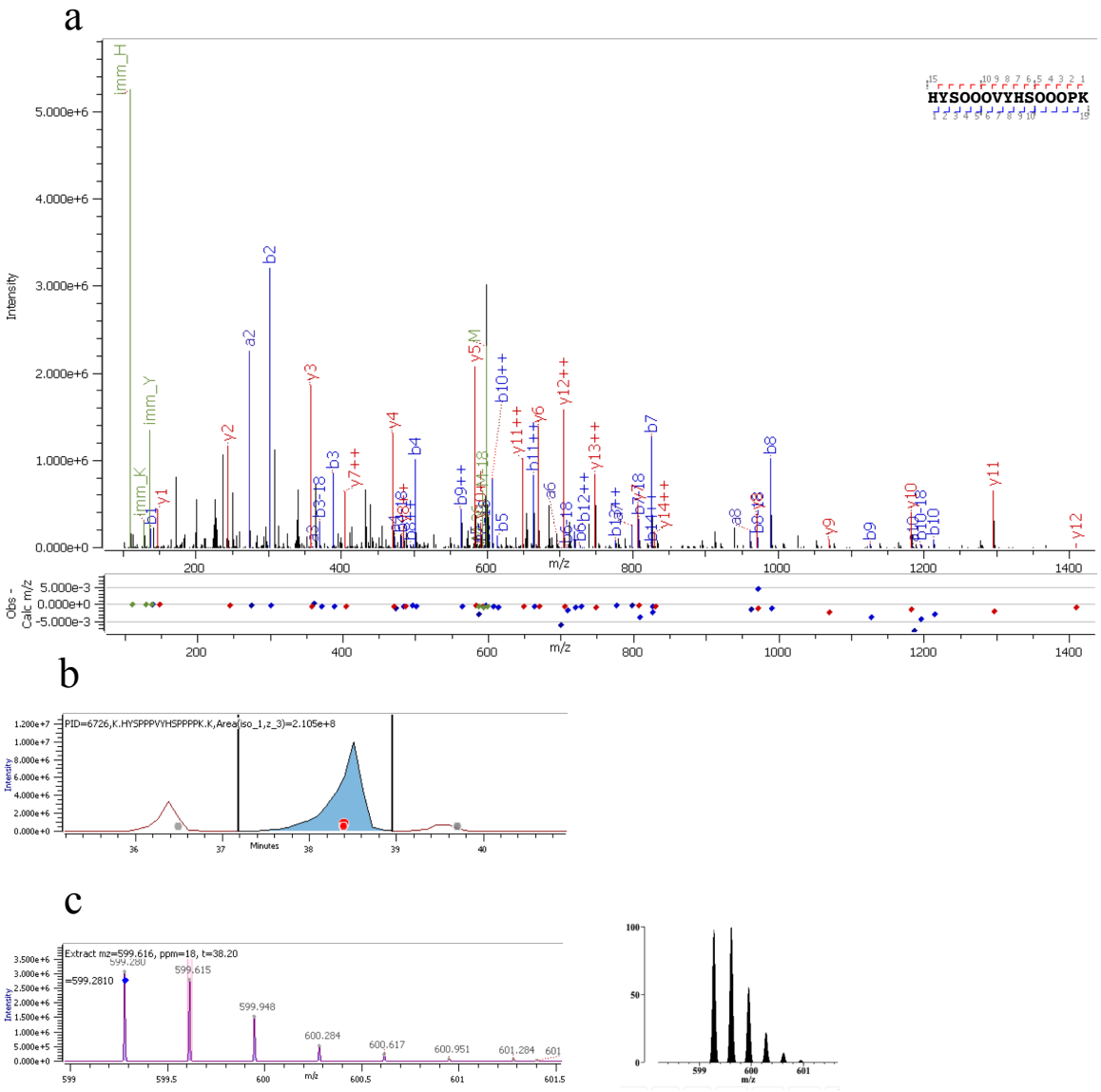

Figure S5.17: Characterization of the non-cross-linked peptide from EXT3. (a) The annotated HCD MS/MS spectrum of the [M+3H]3+ ion at m/z 599.281 with m/z errors of mass fragments at a 10ppm tolerance denoted below; and (inset, top right) a-, b-, & y-fragment ion coverage. Tyrosines are not cross-linked, resulting in no mass loss. (b) Extracted ion chromatogram. The red dots indicate several MS2 collected. The grey dots indicate the time point of a MS2 of the same mass, but with a different ID such as a different site of nonhydroxylated proline. (c) (Left) The isotopic profile of the [M+3H]3+ precursor ion at m/z 599.281 that was isolated for

fragmentation. The blue dot indicates correct monoisotopic assignment based on calculated mass. (Right) The theoretically calculated isotopic profile of the [M+3H]3+ precursor ion at m/z 599.281 from MS-Isotope.[68]

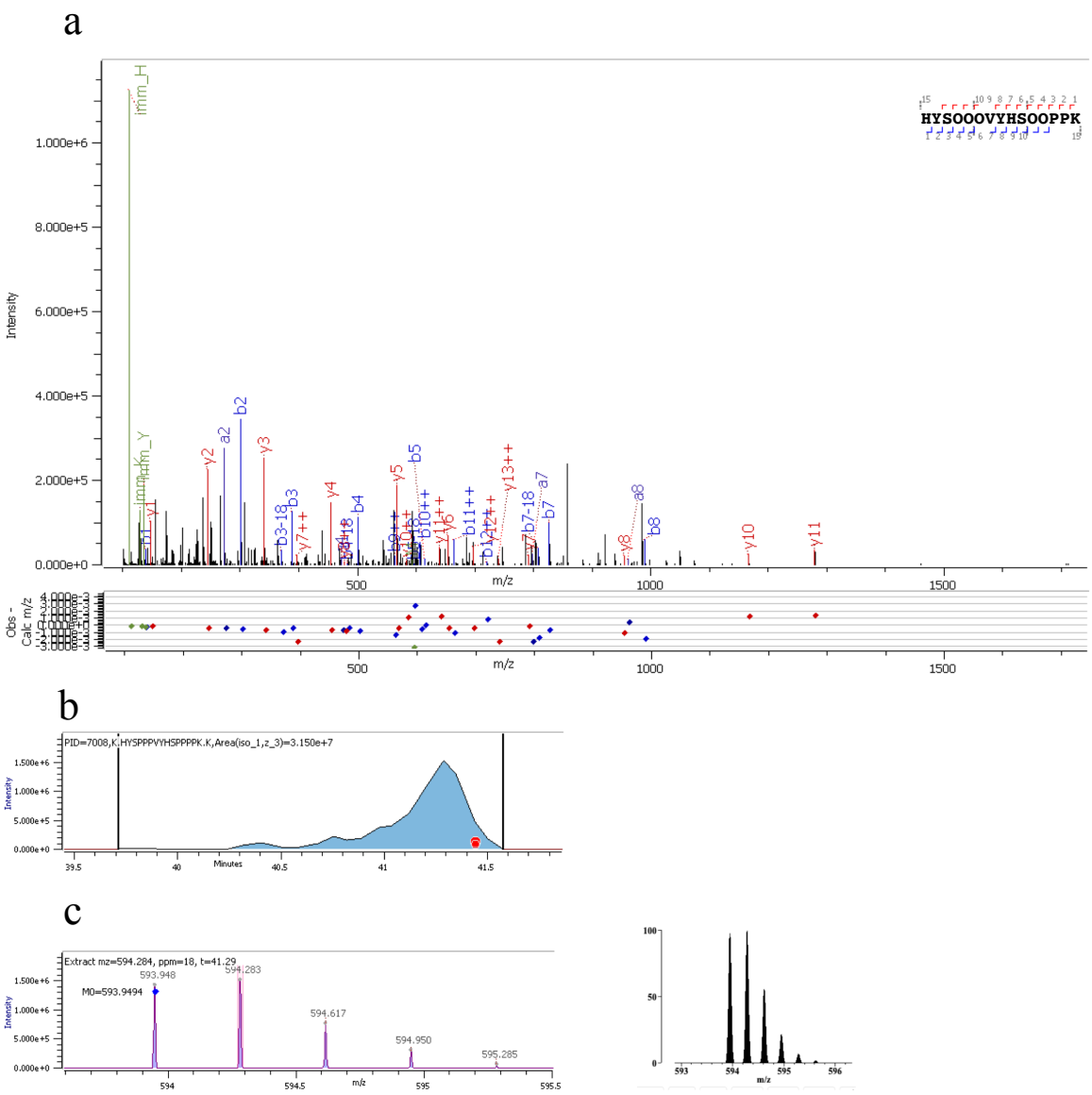

Figure S5.18: Characterization of the non-cross-linked peptide from EXT3. (a) The annotated HCD MS/MS spectrum of the [M+3H]3+ ion at m/z 593.9494 with m/z errors of mass fragments at a 10ppm tolerance denoted below; and (inset, top right) a-, b-, & y-fragment ion coverage. Tyrosines are not cross-linked, resulting in no mass loss. (b) Extracted ion chromatogram. The red dots indicate several MS2 collected. (c) (Left) The isotopic profile of the [M+3H]3+ precursor ion at m/z 593.9494 that was isolated for fragmentation. The blue dot indicates correct

monoisotopic assignment based on calculated mass. (Right) The theoretically calculated isotopic profile of the [M+3H]3+ precursor ion at m/z 593.9494 from MS-Isotope.[68]

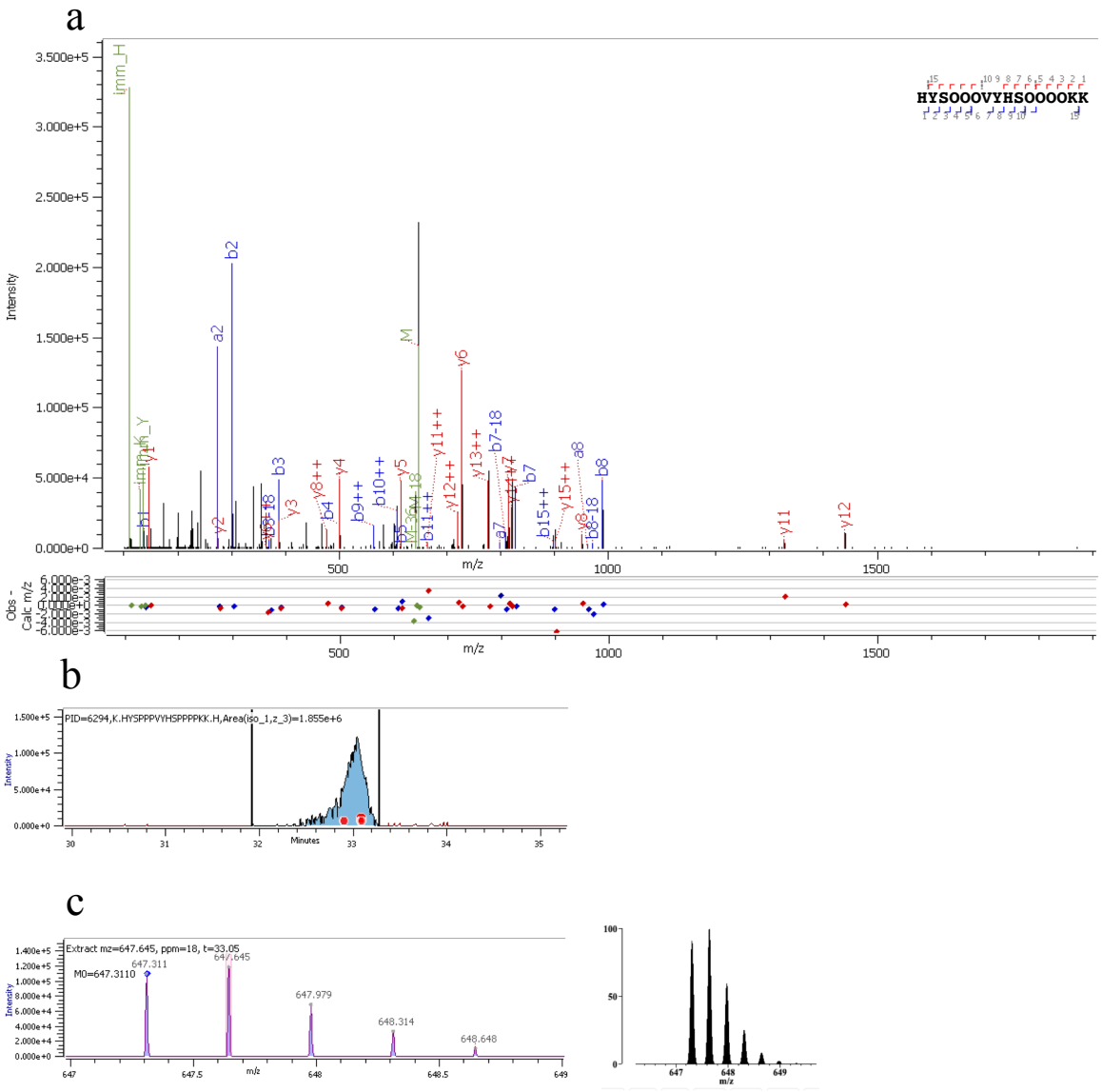

Figure S5.19: Characterization of the non-cross-linked peptide from EXT3. (a) The annotated HCD MS/MS spectrum of the [M+3H]3+ ion at m/z 647.311 with m/z errors of mass fragments at a 10ppm tolerance denoted below; and (inset, top right) a-, b-, & y-fragment ion coverage. Tyrosines are not cross-linked, resulting in no mass loss. (b) Extracted ion chromatogram. The red dots indicate several MS2 collected. (c) (Left) The isotopic profile of the [M+3H]3+ precursor ion at m/z 647.311 that was isolated for fragmentation. The blue dot indicates correct

monoisotopic assignment based on calculated mass. (Right) The theoretically calculated isotopic profile of the [M+3H]3+ precursor ion at m/z 647.311 from MS-Isotope.[68]

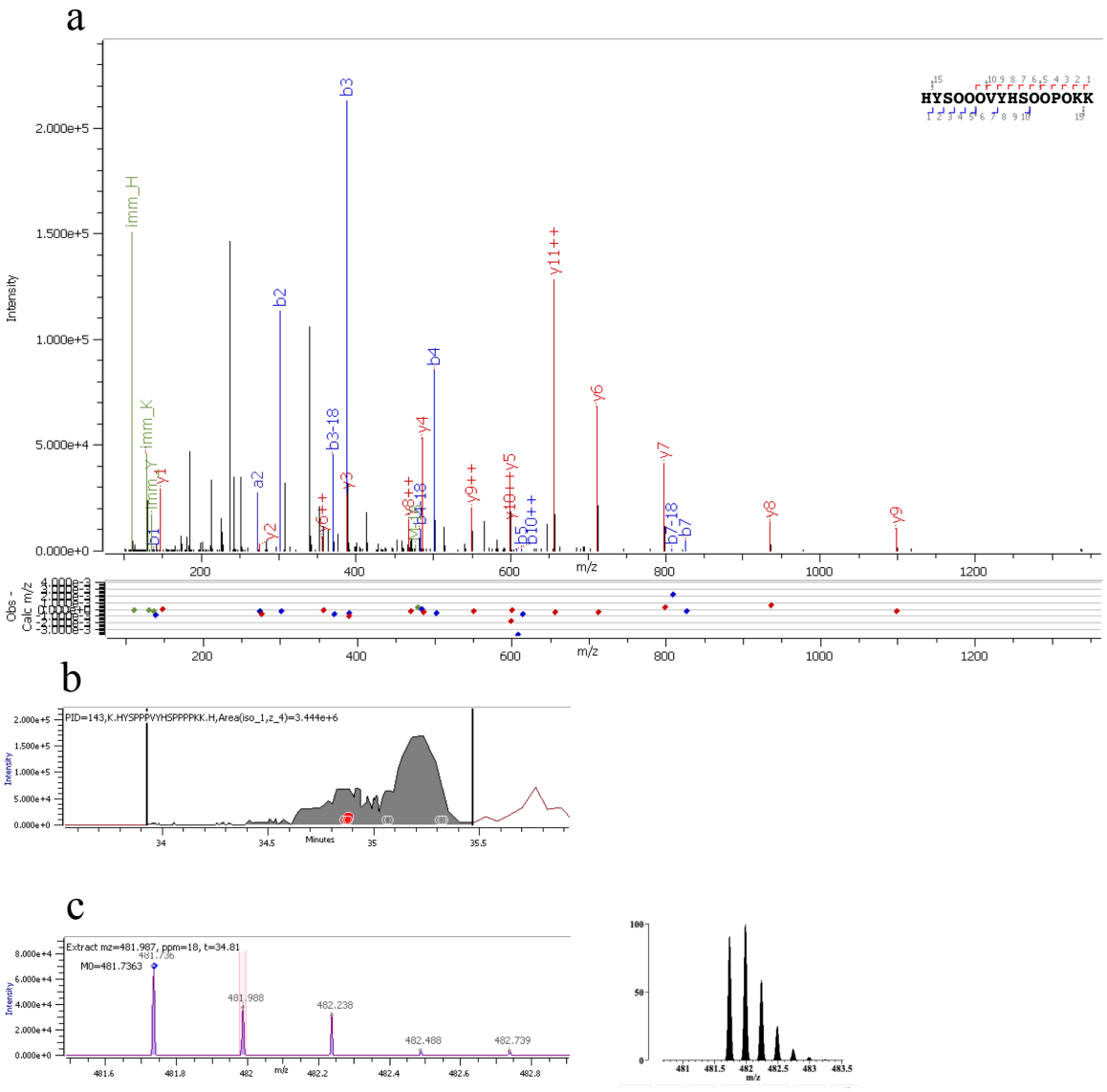

Figure S5.20: Characterization of the non-cross-linked peptide from EXT3. (a) The annotated HCD MS/MS spectrum of the [M+4H]4+ ion at m/z 481.7363 with m/z errors of mass fragments at a 10ppm tolerance denoted below; and (inset, top right) a-, b-, & y-fragment ion coverage. Tyrosines are not cross-linked, resulting in no mass loss. (b) Extracted ion chromatogram. The red dots indicate several MS2 collected. The grey dots indicate the time point of a MS2 of the same mass, but with a different ID such as a different site of nonhydroxylated proline. (c) (Left) The isotopic profile of the [M+4H]4+ precursor ion at m/z 481.7363 that was isolated for

fragmentation. The blue dot indicates correct monoisotopic assignment based on calculated mass. (Right) The theoretically calculated isotopic profile of the [M+4H]4+ precursor ion at m/z 481.7363 from MS-Isotope.[68]

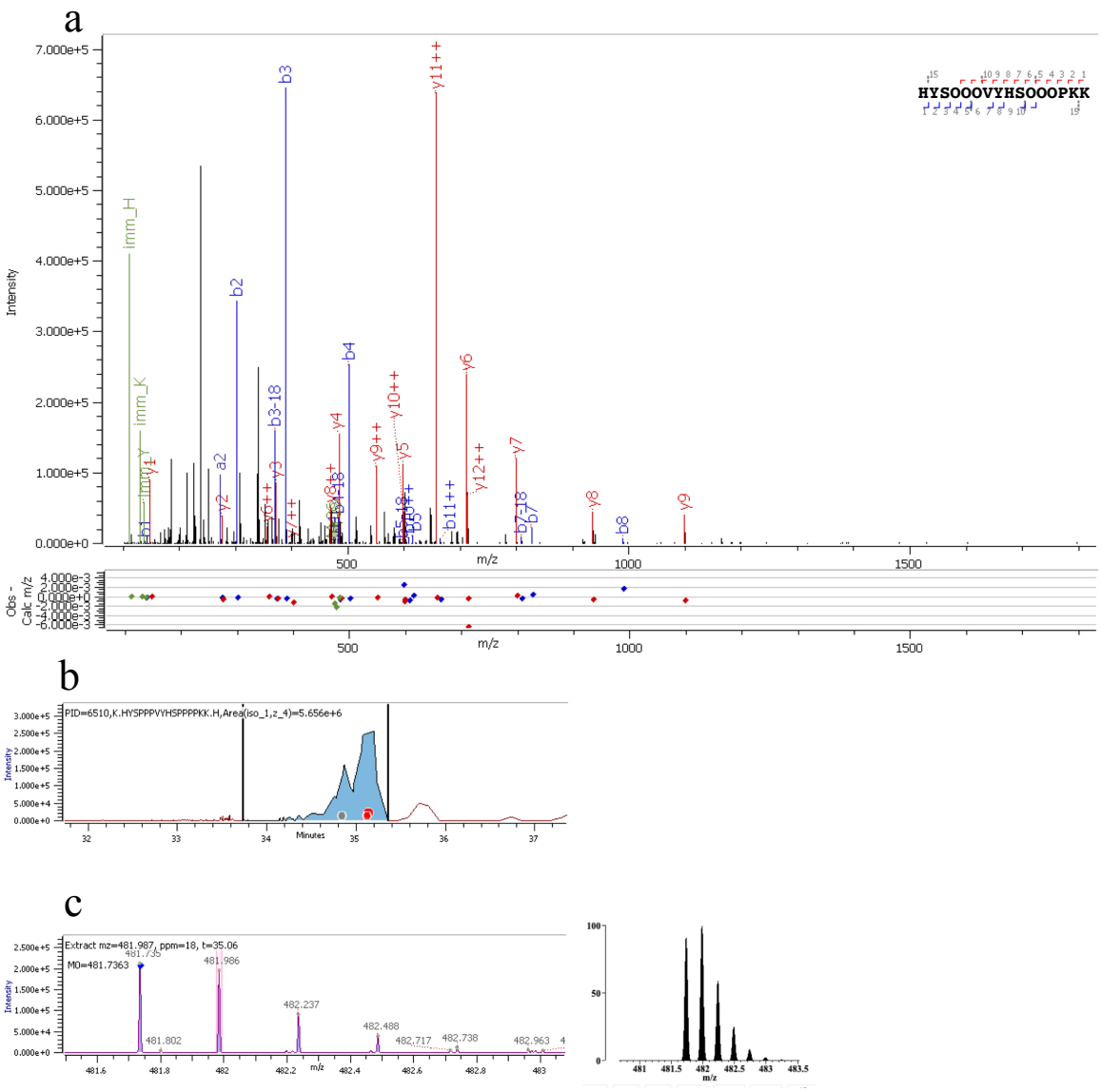

Figure S5.21: Characterization of the non-cross-linked peptide from EXT3. (a) The annotated HCD MS/MS spectrum of the [M+4H]4+ ion at m/z 481.7363 with m/z errors of mass fragments at a 10ppm tolerance denoted below; and (inset, top right) a-, b-, & y-fragment ion coverage. Tyrosines are not cross-linked, resulting in no mass loss. (b) Extracted ion chromatogram. The red dots indicate several MS2 collected. The grey dots indicate the time point of a MS2 of the same mass, but with a different ID such as a different site of nonhydroxylated proline. (c) (Left) The isotopic profile of the [M+4H]4+ precursor ion at m/z 481.7363 that was isolated for

fragmentation. The blue dot indicates correct monoisotopic assignment based on calculated mass. (Right) The theoretically calculated isotopic profile of the [M+4H]4+ precursor ion at m/z 481.7363 from MS-Isotope.[68]

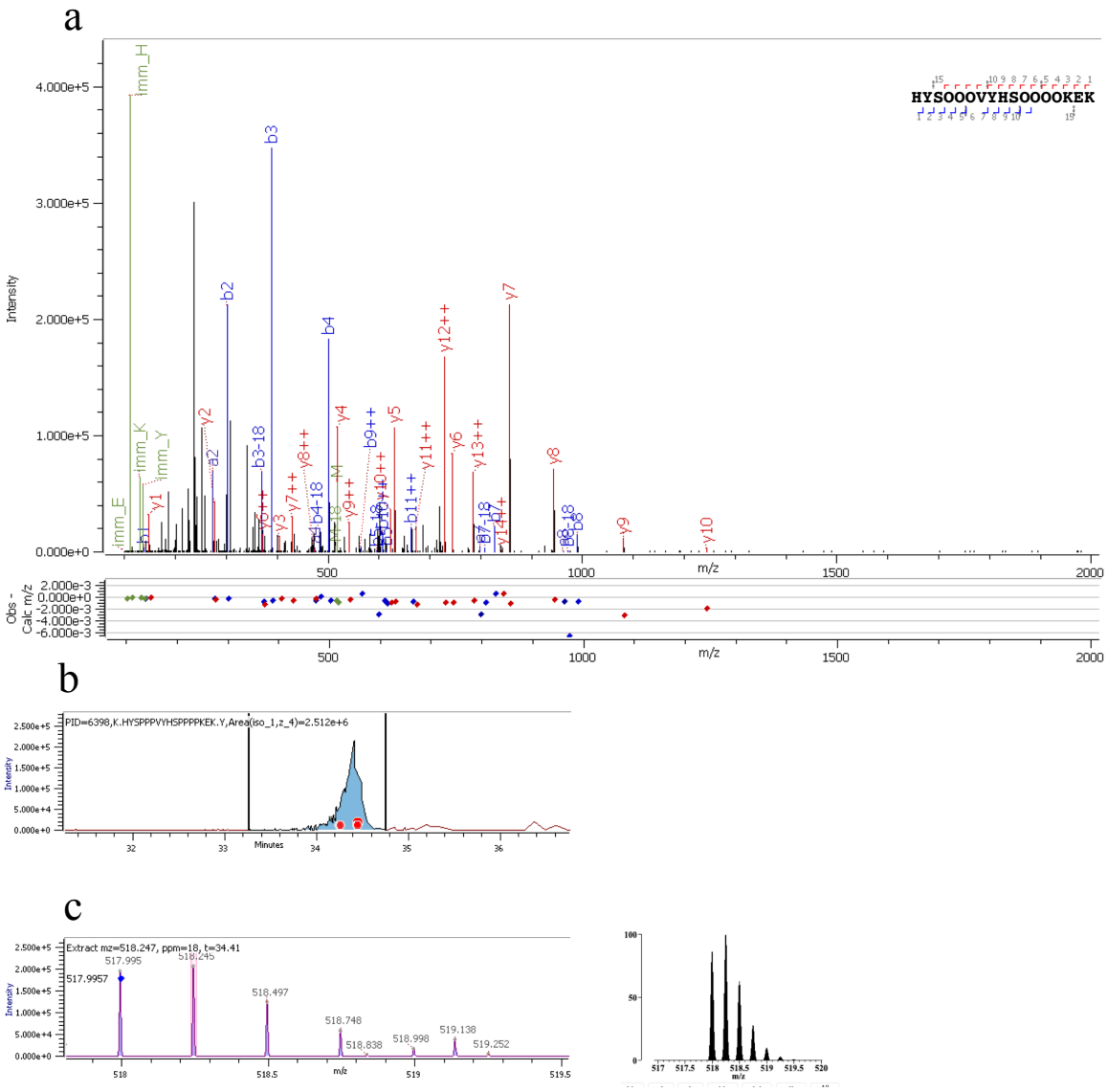

Figure S5.22: Characterization of the non-cross-linked peptide from EXT3. (a) The annotated HCD MS/MS spectrum of the [M+4H]4+ ion at m/z 517.9957 with m/z errors of mass fragments at a 10ppm tolerance denoted below; and (inset, top right) a-, b-, & y-fragment ion coverage. Tyrosines are not cross-linked, resulting in no mass loss. (b) Extracted ion chromatogram. The red dots indicate several MS2 collected. (c) (Left) The isotopic profile of the [M+4H]4+ precursor ion at m/z 517.9957 that was isolated for fragmentation. The blue dot indicates correct

monoisotopic assignment based on calculated mass. (Right) The theoretically calculated isotopic profile of the [M+4H]4+ precursor ion at m/z 517.9957 from MS-Isotope.[68]

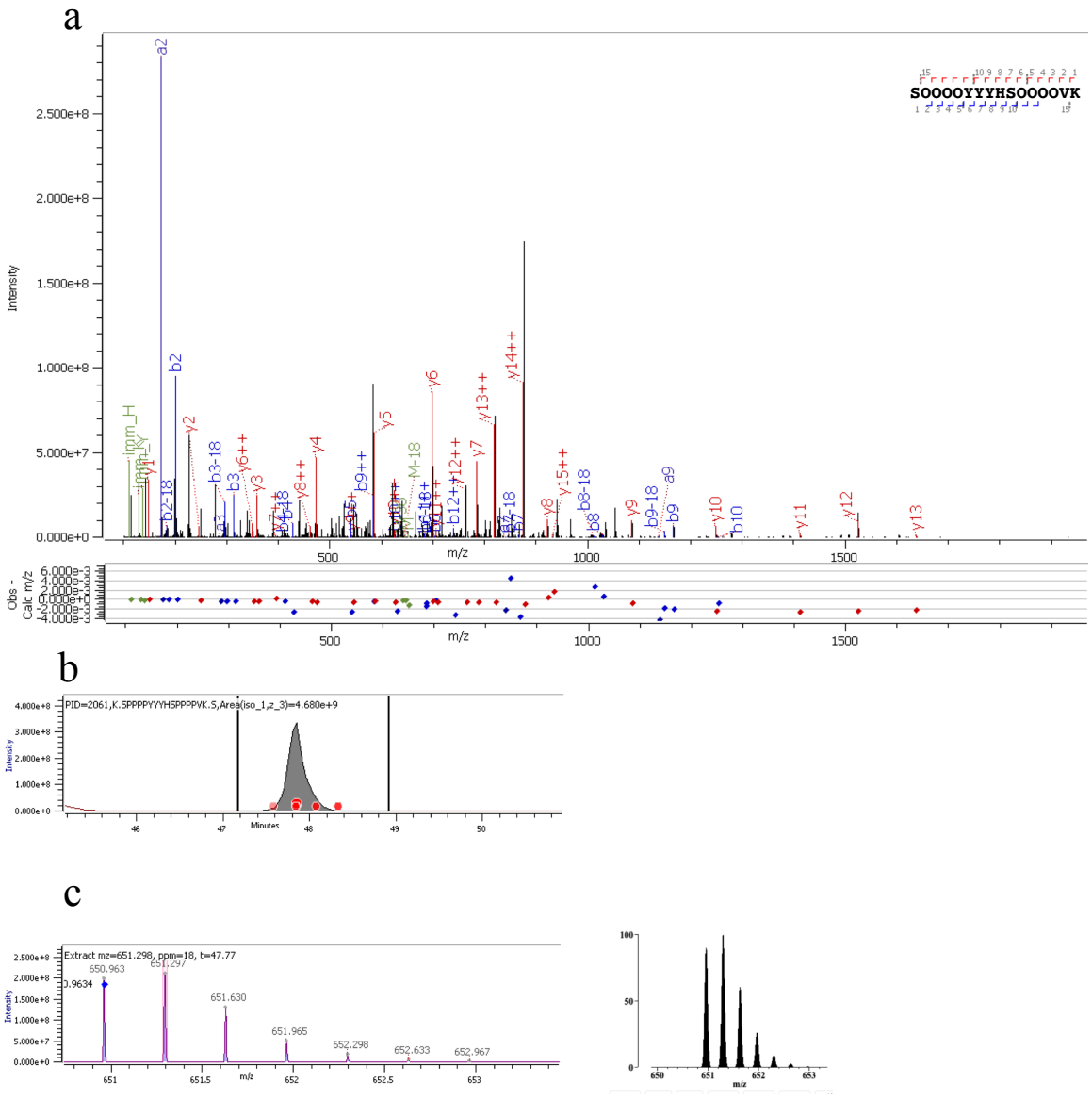

Figure S5.23: Characterization of the non-cross-linked peptide from EXT8. (a) The annotated HCD MS/MS spectrum of the [M+3H]3+ ion at m/z 650.9634 with m/z errors of mass fragments at a 10ppm tolerance denoted below; and (inset, top right) a-, b-, & y-fragment ion coverage. Tyrosines are not cross-linked, resulting in no mass loss. (b) Extracted ion chromatogram. The red dots indicate several MS2 collected. The grey dots indicate the time point of a MS2 of the same mass, but with a different ID such as a different site of nonhydroxylated proline. (c) (Left) The isotopic profile of the [M+3H]3+ precursor ion at m/z 650.9634 that was isolated for

fragmentation. The blue dot indicates correct monoisotopic assignment based on calculated mass. (Right) The theoretically calculated isotopic profile of the [M+3H]3+ precursor ion at m/z 650.9634 from MS-Isotope.[68]

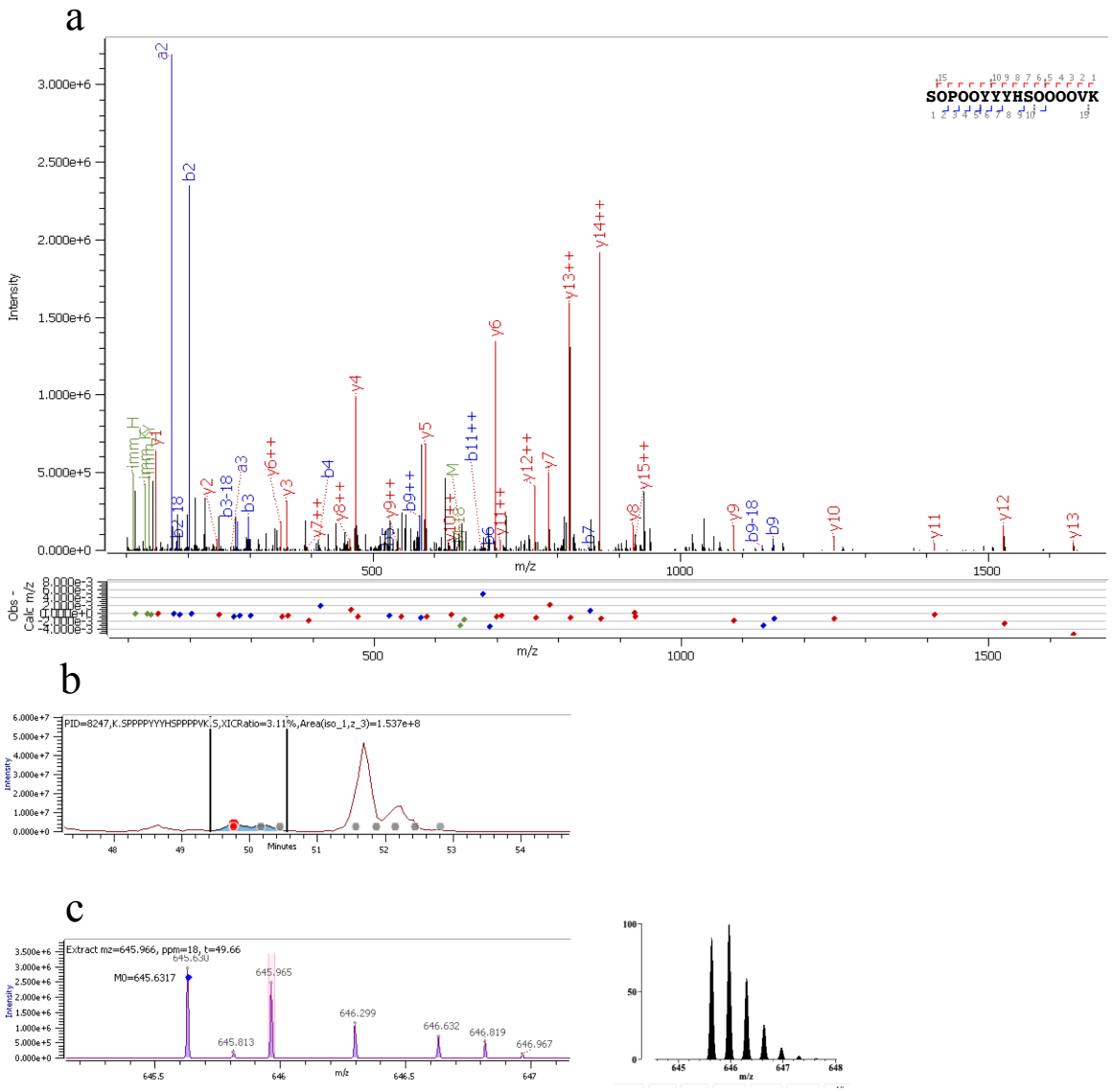

Figure S5.24: Characterization of the non-cross-linked peptide from EXT8. (a) The annotated HCD MS/MS spectrum of the [M+3H]3+ ion at m/z 645.6317 with m/z errors of mass fragments at a 10ppm tolerance denoted below; and (inset, top right) a-, b-, & y-fragment ion coverage. Tyrosines are not cross-linked, resulting in no mass loss. (b) Extracted ion chromatogram. The red dots indicate several MS2 collected. The grey dots indicate the time point of a MS2 of the same mass, but with a different ID such as a different site of nonhydroxylated proline. (c) (Left) The isotopic profile of the [M+3H]3+ precursor ion at m/z 645.6317 that was isolated for

fragmentation. The blue dot indicates correct monoisotopic assignment based on calculated mass. (Right) The theoretically calculated isotopic profile of the [M+3H]3+ precursor ion at m/z 645.6317 from MS-Isotope.[68]

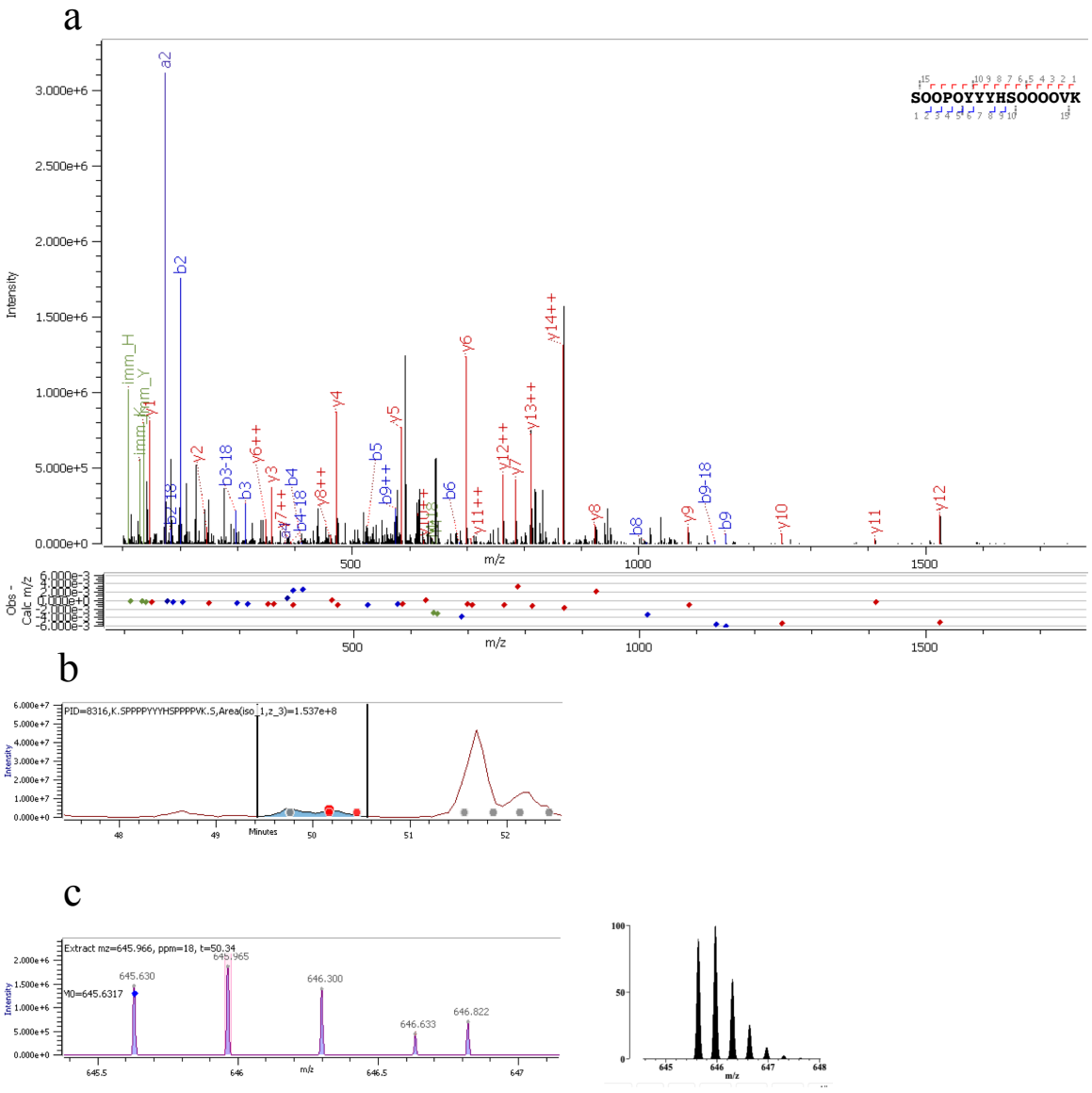

Figure S5.25: Characterization of the non-cross-linked peptide from EXT8. (a) The annotated HCD MS/MS spectrum of the [M+3H]3+ ion at m/z 645.6317 with m/z errors of mass fragments at a 10ppm tolerance denoted below; and (inset, top right) a-, b-, & y-fragment ion coverage. Tyrosines are not cross-linked, resulting in no mass loss. (b) Extracted ion chromatogram. The red dots indicate several MS2 collected. The grey dots indicate the time point of a MS2 of the same mass, but with a different ID such as a different site of nonhydroxylated proline. (c) (Left) The isotopic profile of the [M+3H]3+ precursor ion at m/z 645.6317 that was isolated for

fragmentation. The blue dot indicates correct monoisotopic assignment based on calculated mass. (Right) The theoretically calculated isotopic profile of the [M+3H]3+ precursor ion at m/z 645.6317 from MS-Isotope.[68]

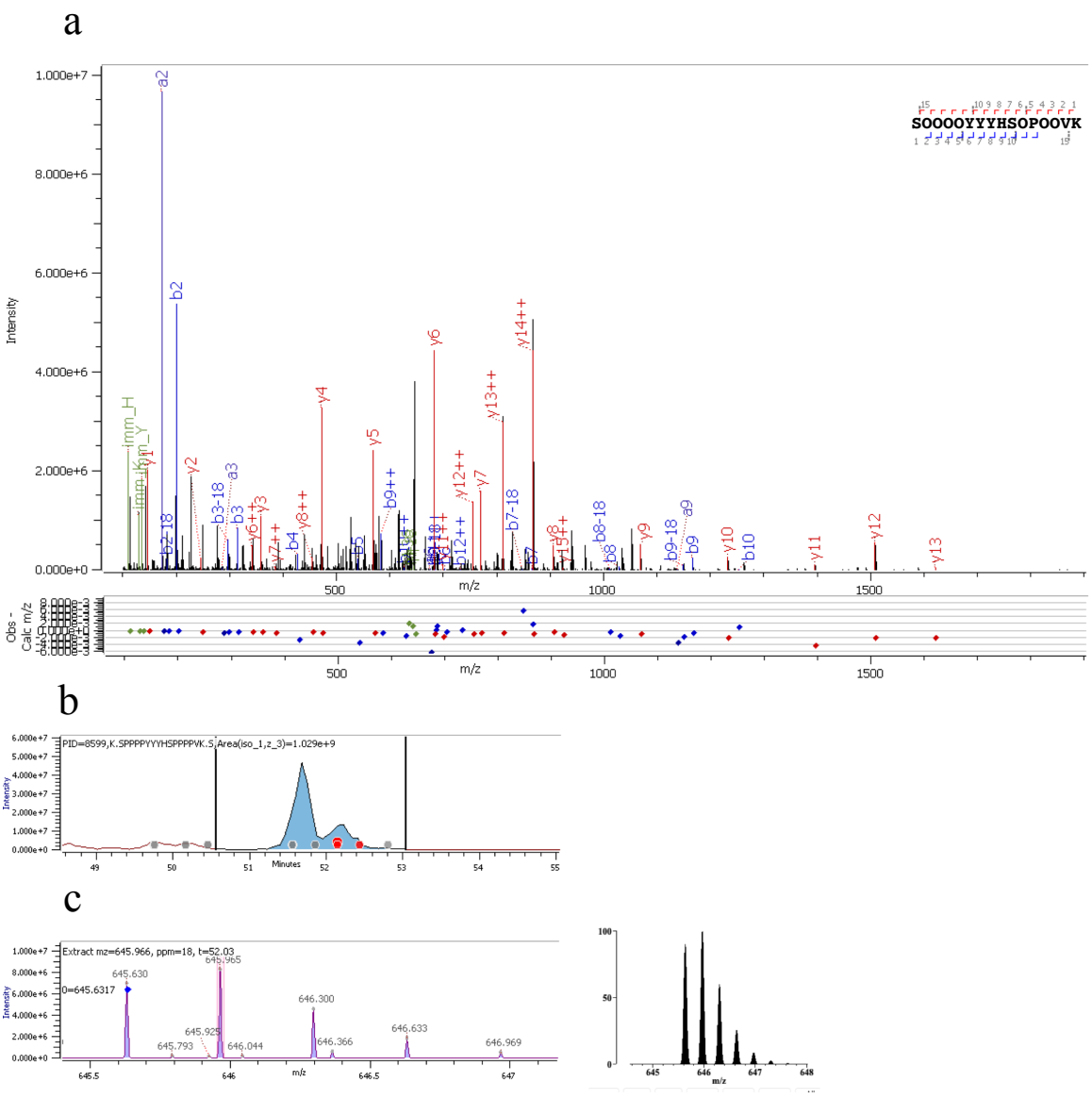

Figure S5.26: Characterization of the non-cross-linked peptide from EXT8. (a) The annotated HCD MS/MS spectrum of the [M+3H]3+ ion at m/z 645.6317 with m/z errors of mass fragments at a 10ppm tolerance denoted below; and (inset, top right) a-, b-, & y-fragment ion coverage. Tyrosines are not cross-linked, resulting in no mass loss. (b) Extracted ion chromatogram. The red dots indicate several MS2 collected. The grey dots indicate the time point of a MS2 of the same mass, but with a different ID such as a different site of nonhydroxylated proline. (c) (Left) The isotopic profile of the [M+3H]3+ precursor ion at m/z 645.6317 that was isolated for

fragmentation. The blue dot indicates correct monoisotopic assignment based on calculated mass. (Right) The theoretically calculated isotopic profile of the [M+3H]3+ precursor ion at m/z 645.6317 from MS-Isotope.[68]

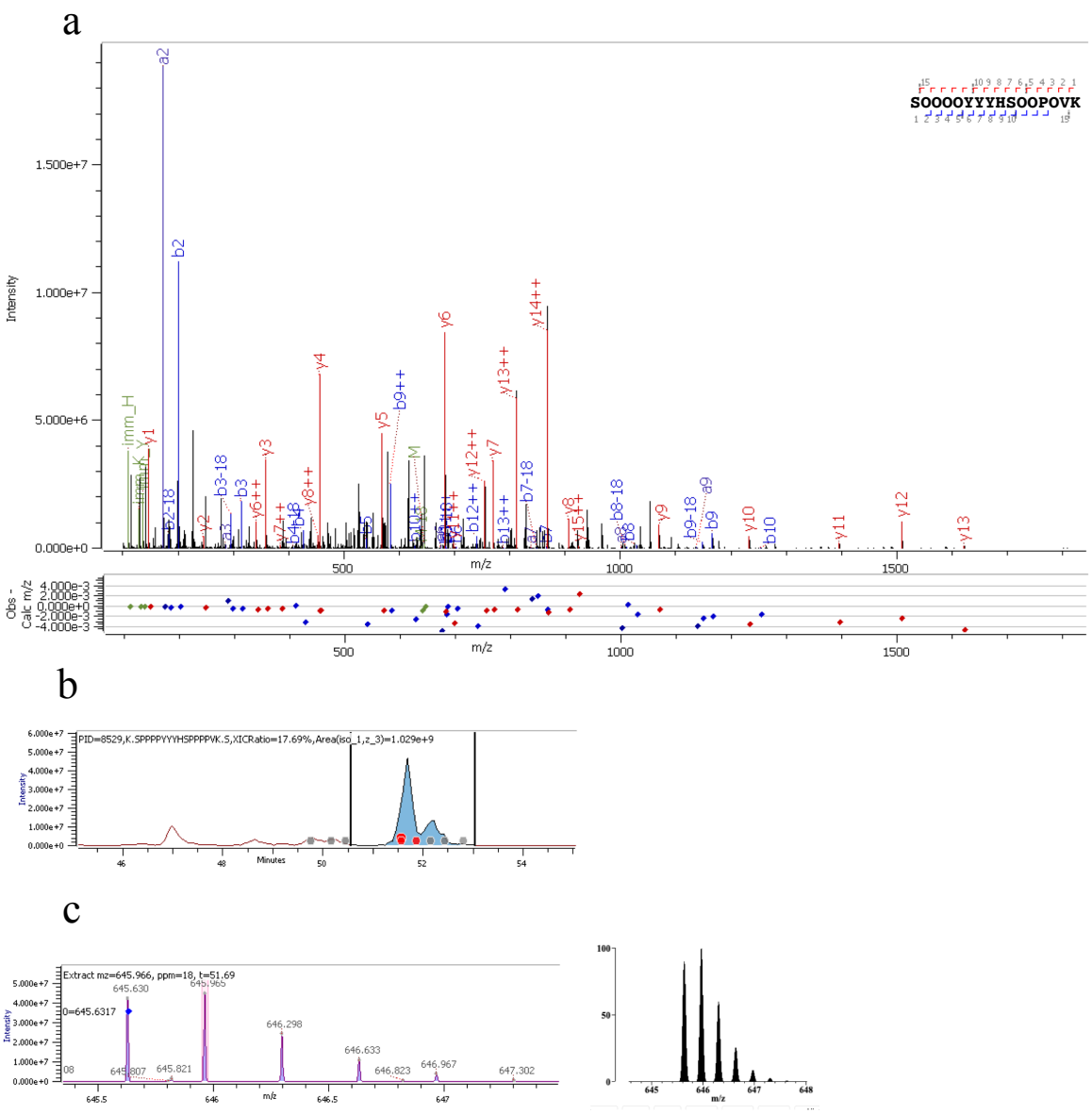

Figure S5.27: Characterization of the non-cross-linked peptide from EXT8. (a) The annotated HCD MS/MS spectrum of the [M+3H]3+ ion at m/z 645.6317 with m/z errors of mass fragments at a 10ppm tolerance denoted below; and (inset, top right) a-, b-, & y-fragment ion coverage. Tyrosines are not cross-linked, resulting in no mass loss. (b) Extracted ion chromatogram. The red dots indicate several MS2 collected. The grey dots indicate the time point of a MS2 of the same mass, but with a different ID such as a different site of nonhydroxylated proline. (c) (Left) The isotopic profile of the [M+3H]3+ precursor ion at m/z 645.6317 that was isolated for

fragmentation. The blue dot indicates correct monoisotopic assignment based on calculated mass. (Right) The theoretically calculated isotopic profile of the [M+3H]3+ precursor ion at m/z 645.6317 from MS-Isotope.[68]

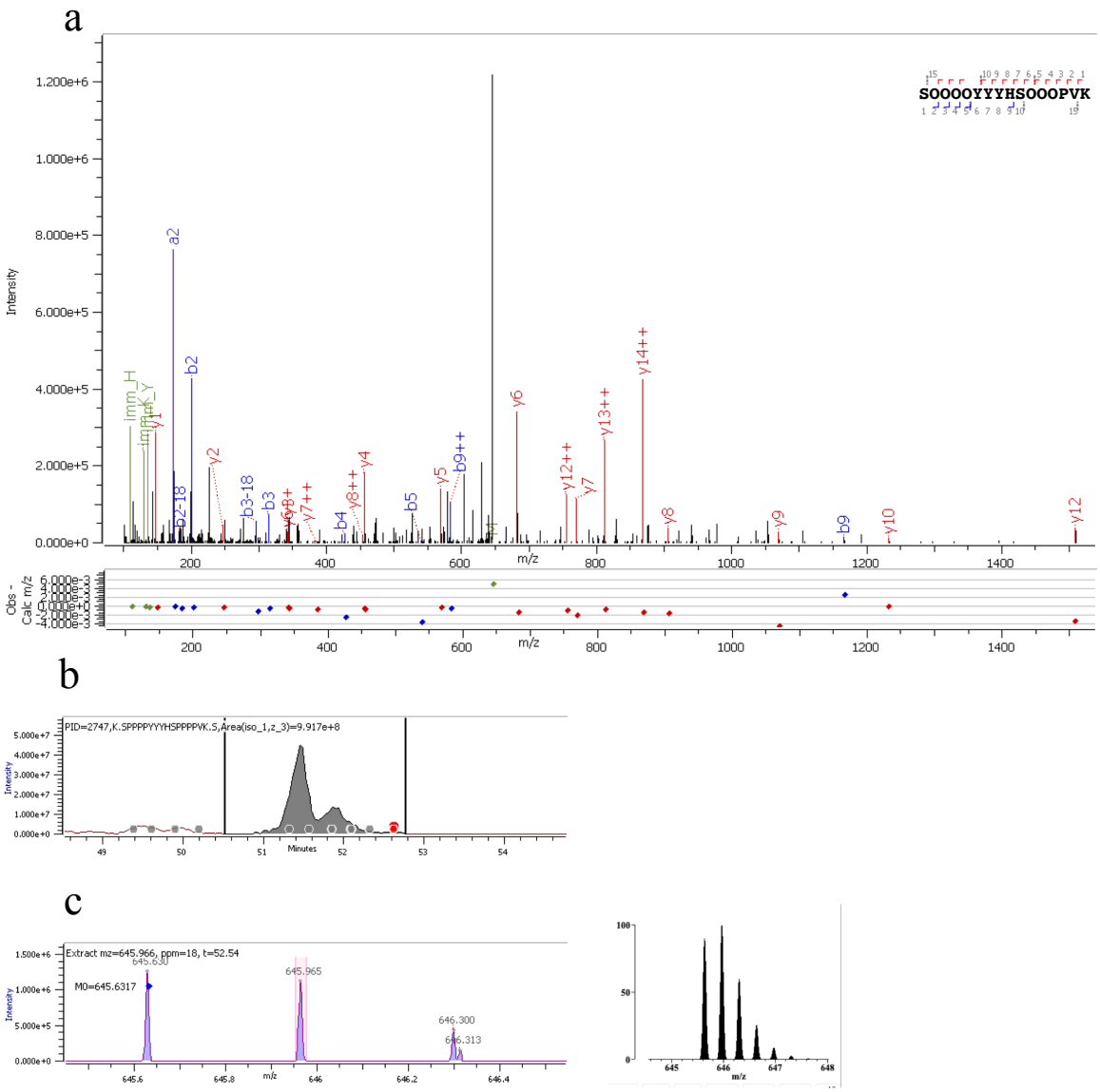

Figure S5.28: Characterization of the non-cross-linked peptide from EXT8. (a) The annotated HCD MS/MS spectrum of the [M+3H]3+ ion at m/z 645.6317 with m/z errors of mass fragments at a 10ppm tolerance denoted below; and (inset, top right) a-, b-, & y-fragment ion coverage. Tyrosines are not cross-linked, resulting in no mass loss. (b) Extracted ion chromatogram. The red dots indicate several MS2 collected. The grey dots indicate the time point of a MS2 of the same mass, but with a different ID such as a different site of nonhydroxylated proline. (c) (Left) The isotopic profile of the [M+3H]3+ precursor ion at m/z 645.6317 that was isolated for

fragmentation. The blue dot indicates correct monoisotopic assignment based on calculated mass. (Right) The theoretically calculated isotopic profile of the [M+3H]3+ precursor ion at m/z 645.6317 from MS-Isotope.[68]

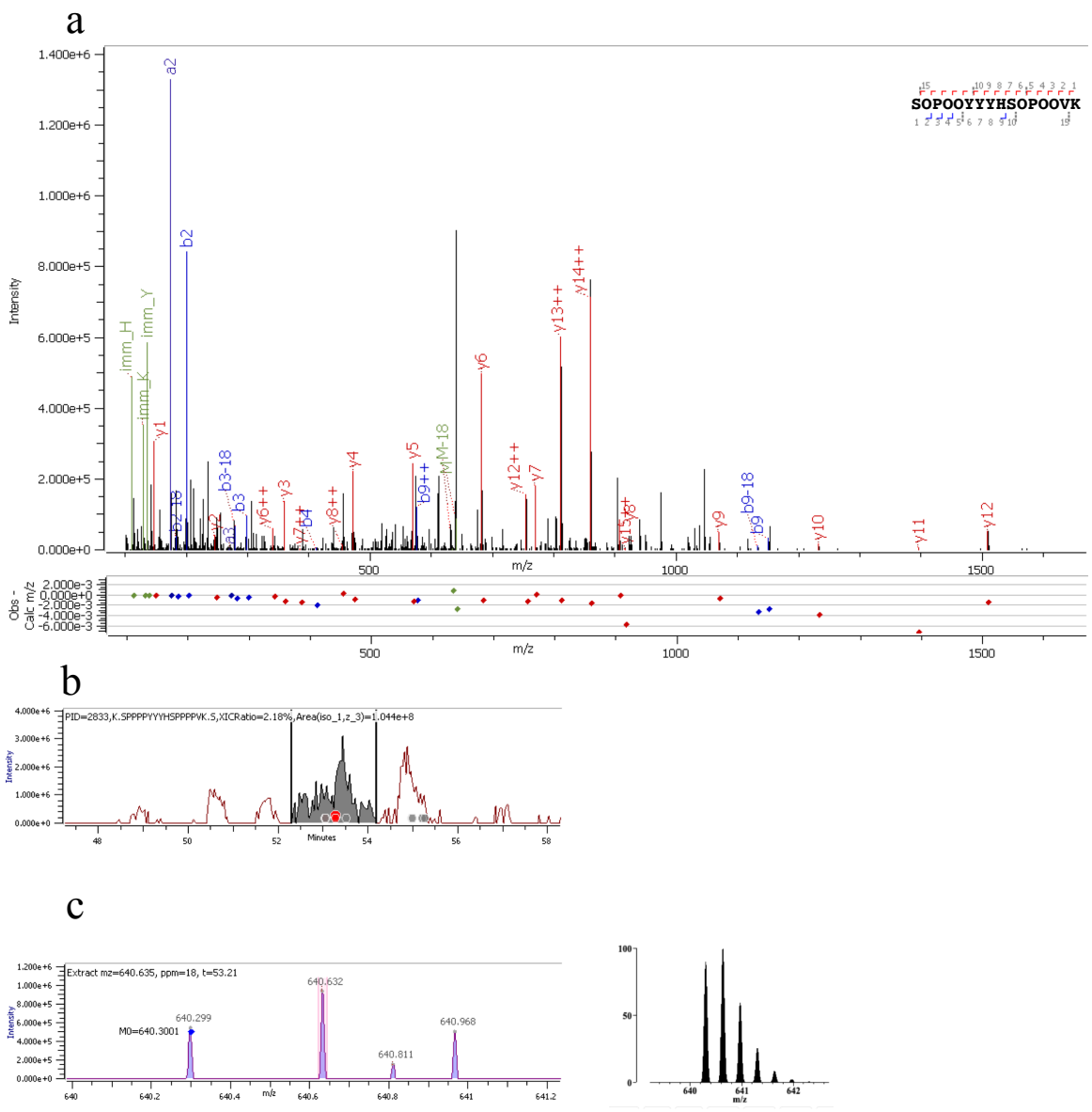

Figure S5.29: Characterization of the non-cross-linked peptide from EXT8. (a) The annotated HCD MS/MS spectrum of the [M+3H]3+ ion at m/z 640.3001 with m/z errors of mass fragments at a 10ppm tolerance denoted below; and (inset, top right) a-, b-, & y-fragment ion coverage. Tyrosines are not cross-linked, resulting in no mass loss. (b) Extracted ion chromatogram. The red dots indicate several MS2 collected. The grey dots indicate the time point of a MS2 of the same mass, but with a different ID such as a different site of nonhydroxylated proline. (c) (Left) The isotopic profile of the [M+3H]3+ precursor ion at m/z 640.3001 that was isolated for

fragmentation. The blue dot indicates correct monoisotopic assignment based on calculated mass. (Right) The theoretically calculated isotopic profile of the [M+3H]3+ precursor ion at m/z 640.3001 from MS-Isotope.[68]

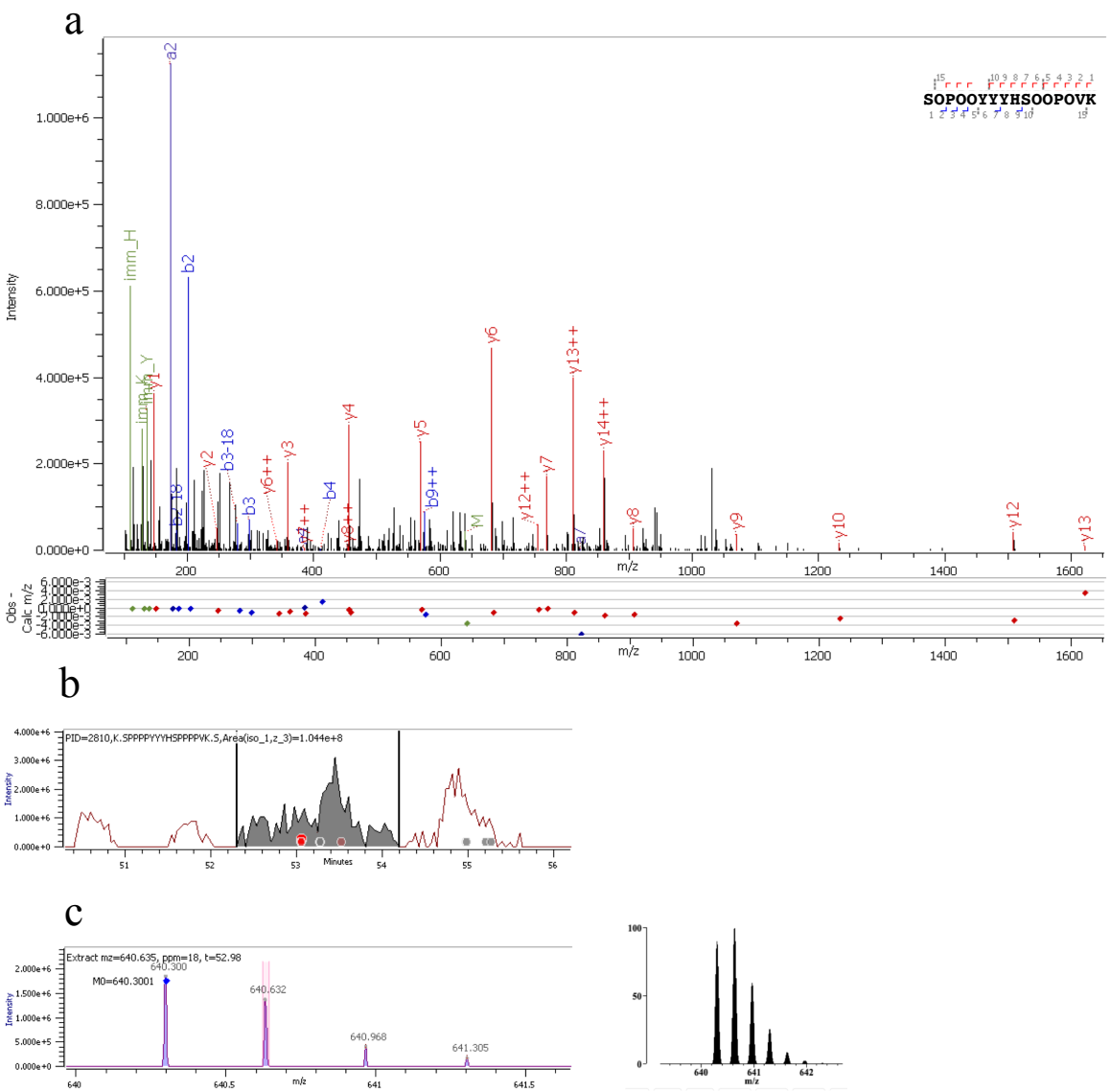

Figure S5.30: Characterization of the non-cross-linked peptide from EXT8. (a) The annotated HCD MS/MS spectrum of the [M+3H]3+ ion at m/z 640.3001 with m/z errors of mass fragments at a 10ppm tolerance denoted below; and (inset, top right) a-, b-, & y-fragment ion coverage. Tyrosines are not cross-linked, resulting in no mass loss. (b) Extracted ion chromatogram. The red dots indicate several MS2 collected. The grey dots indicate the time point of a MS2 of the same mass, but with a different ID such as a different site of nonhydroxylated proline. (c) (Left) The isotopic profile of the [M+3H]3+ precursor ion at m/z 640.3001 that was isolated for

fragmentation. The blue dot indicates correct monoisotopic assignment based on calculated mass. (Right) The theoretically calculated isotopic profile of the [M+3H]3+ precursor ion at m/z 640.3001 from MS-Isotope.[68]

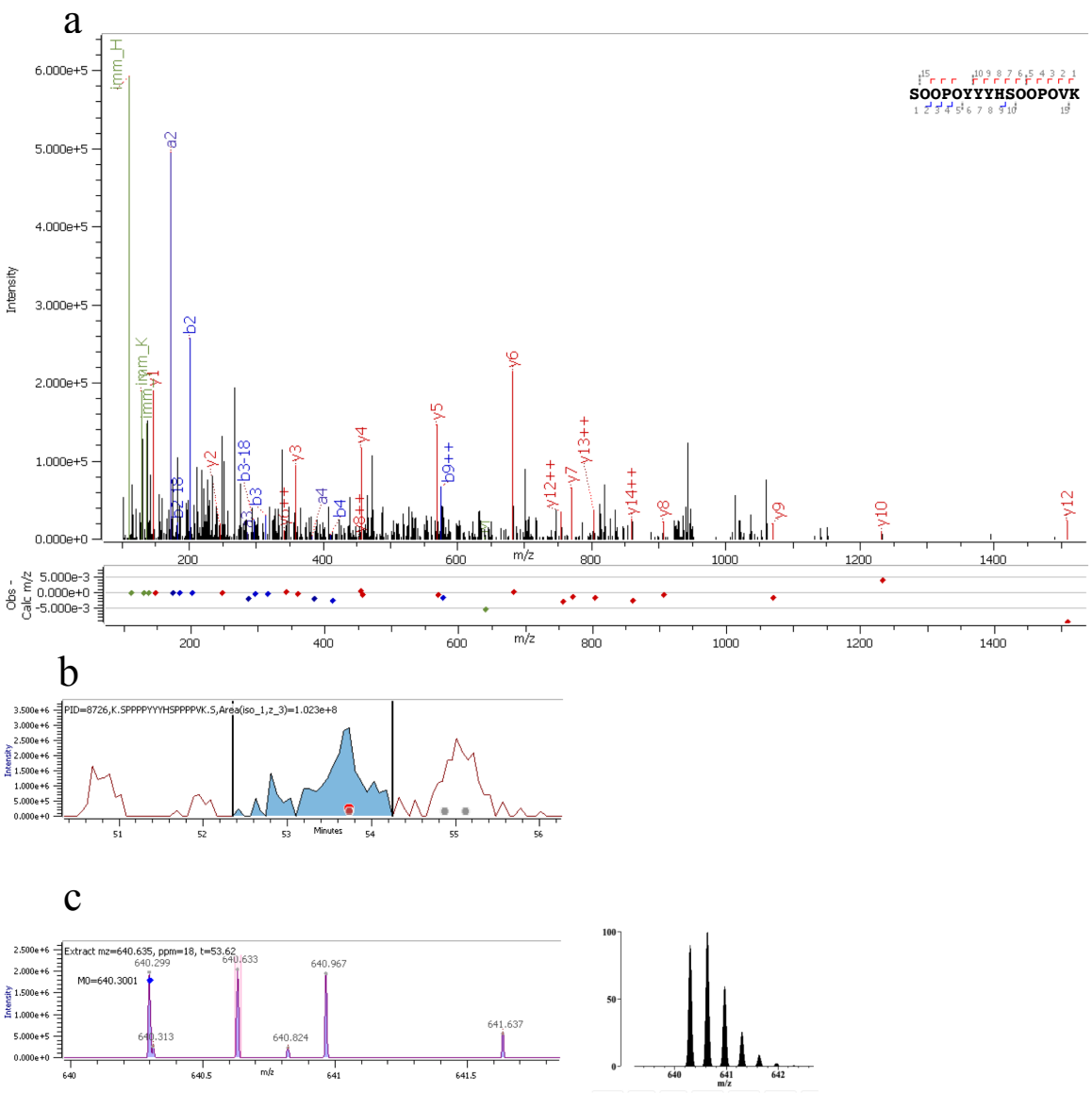

Figure S5.31: Characterization of the non-cross-linked peptide from EXT8. (a) The annotated HCD MS/MS spectrum of the [M+3H]3+ ion at m/z 640.3001 with m/z errors of mass fragments at a 10ppm tolerance denoted below; and (inset, top right) a-, b-, & y-fragment ion coverage. Tyrosines are not cross-linked, resulting in no mass loss. (b) Extracted ion chromatogram. The red dots indicate several MS2 collected. The grey dots indicate the time point of a MS2 of the same mass, but with a different ID such as a different site of nonhydroxylated proline. (c) (Left) The isotopic profile of the [M+3H]3+ precursor ion at m/z 640.3001 that was isolated for

fragmentation. The blue dot indicates correct monoisotopic assignment based on calculated mass. (Right) The theoretically calculated isotopic profile of the [M+3H]3+ precursor ion at m/z 640.3001 from MS-Isotope.[68]

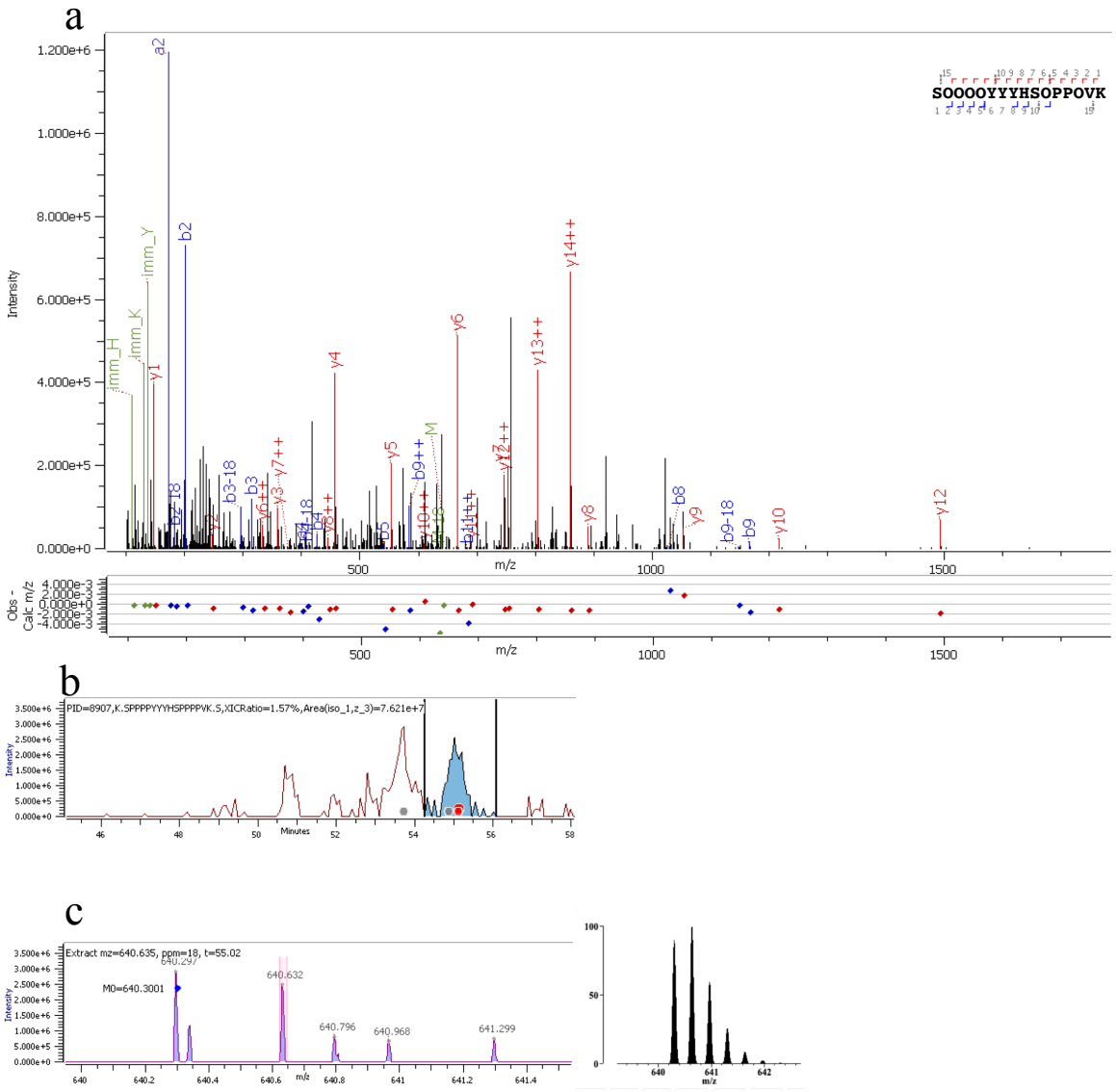

Figure S5.32: Characterization of the non-cross-linked peptide from EXT8. (a) The annotated HCD MS/MS spectrum of the [M+3H]3+ ion at m/z 640.3001 with m/z errors of mass fragments at a 10ppm tolerance denoted below; and (inset, top right) a-, b-, & y-fragment ion coverage. Tyrosines are not cross-linked, resulting in no mass loss. (b) Extracted ion chromatogram. The red dots indicate several MS2 collected. The grey dots indicate the time point of a MS2 of the same mass, but with a different ID such as a different site of nonhydroxylated proline. (c) (Left) The isotopic profile of the [M+3H]3+ precursor ion at m/z 640.3001 that was isolated for

fragmentation. The blue dot indicates correct monoisotopic assignment based on calculated mass. (Right) The theoretically calculated isotopic profile of the [M+3H]3+ precursor ion at m/z 640.3001 from MS-Isotope.[68]

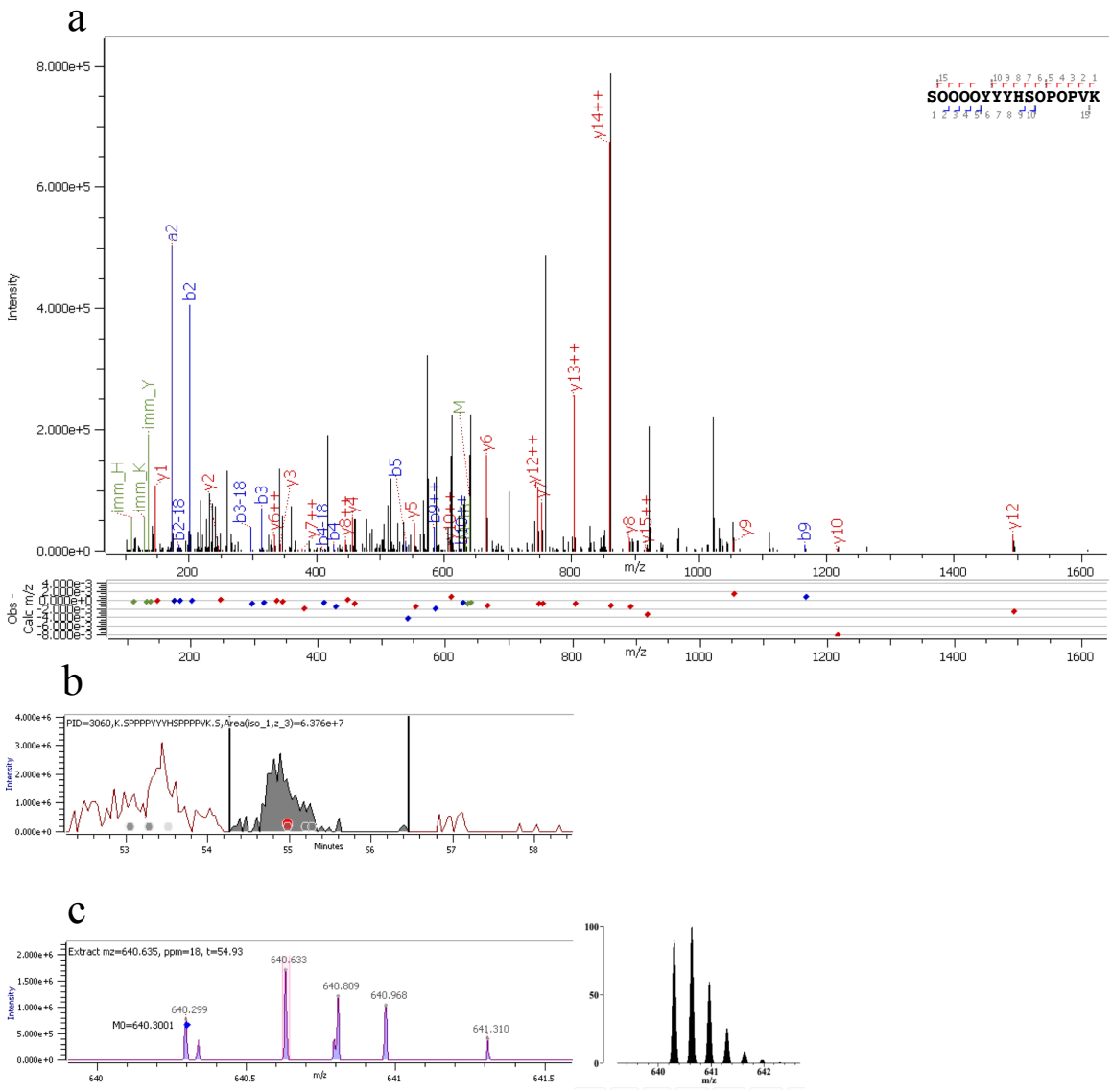

Figure S5.33: Characterization of the non-cross-linked peptide from EXT8. (a) The annotated HCD MS/MS spectrum of the [M+3H]3+ ion at m/z 640.3001 with m/z errors of mass fragments at a 10ppm tolerance denoted below; and (inset, top right) a-, b-, & y-fragment ion coverage. Tyrosines are not cross-linked, resulting in no mass loss. (b) Extracted ion chromatogram. The red dots indicate several MS2 collected. The grey dots indicate the time point of a MS2 of the same mass, but with a different ID such as a different site of nonhydroxylated proline. (c) (Left) The isotopic profile of the [M+3H]3+ precursor ion at m/z 640.3001 that was isolated for

fragmentation. The blue dot indicates correct monoisotopic assignment based on calculated mass. (Right) The theoretically calculated isotopic profile of the [M+3H]3+ precursor ion at m/z 640.3001 from MS-Isotope.[68]

APPENDIX C

Supplemental Tables 3a-e: Full Protein Sequences of EXTs and LRX Identified with Repetitive Tryptic Peptides Numbered

Full protein sequences of EXT 1, 3, 8, 10, and LRX 4 identified in this study. Extensin numbers are designated by work of Showalter et al[17]. Uniprot designation is included in each title. Protein sequences are split into tryptic peptides to illustrate the repetitive nature of the extensin sequences. Each peptide is numbered based on the first occurrence of the repetitive peptide in the protein's sequence. Coloration of peptide sequence is as indicated in above key. Potential tyrosine cross-linkage sites are indicated by a lowercase tyrosine. S(P)n glycosylation motifs of serines and hydroxyprolines follows the O-Hyp glycosylation code as described by Kieliszewski, et al.[24]

Key: Extensin Motifs

Non-Crosslinking Motifs	Potential for Cross-Linkage Motifs
Signal Peptide	yXy: Isodityrosine
Contiguous O-Arabinofuranosylated Hydroxyprolines	yy
Mono-Galactosylated Serine	VyK
SPSP Motif	y
	Cysteine Disulfide

Key for Supplemental Tables ST3a-e

Table S3a **Extensin 1: Q38913**

Signal Peptide:	MASFLVLAFSLAFVSQTTAN
Peptide 1: (1)	**y**F**y**SSPPPPVK
Peptide 2: (9)	HySPPP**V**y**K**
Peptide 3: (12)	SPPPPVK
Peptide 2:	HySPPP**V**y**K**
Peptide 3:	SPPPPVK
Peptide 2:	HySPPP**V**y**K**
Peptide 3:	SPPPPVK
Peptide 4: (4)	**yy**SPPP**V**y**K**
Peptide 5: (1)	SPPPP**V**y**K**
Peptide 3:	SPPPPVK
Peptide 2:	HySPPP**V**y**K**
Peptide 3:	SPPPPVK
Peptide 2:	HySPPP**V**y**K**
Peptide 3:	SPPPPVK
Peptide 2:	HySPPP**V**y**K**
Peptide 3:	SPPPPVK
Peptide 2:	HySPPP**V**y**K**
Peptide 3:	SPPPPVK
Peptide 4:	**yy**SPPP**V**y**K**
Peptide 3:	SPPPPVK
Peptide 2:	HySPPP**V**y**K**
Peptide 3:	SPPPPVK
Peptide 4:	**yy**SPPP**V**y**K**
Peptide 3:	SPPPPVK
Peptide 2:	HySPPP**V**y**K**
Peptide 3:	SPPPPVK
Peptide 4:	**yy**SPPP**V**y**K**
Peptide 6: (1)	SPPPPVHySPPPVVyHSPPPPVHySPPPVVyHSPPPPVHySPPP
	VVyHSPPPPVHySPPPVVyHSPPPPVHySPPPVVyHSPPPPK(K)
Peptide 7: (1)	H**y**E**y**K
Peptide 8: (1)	SPPPPVHySPPTVyHSPPPPVHHySPPHQP**y**L**y**K
Peptide 9: (1)	SPPPPH**y**

Table S3a: Full Protein Sequence of EXT1: Tryptic digest of EXT1 results in 9 unique peptides. Several peptides are repetitive. The first occurrence of a repetitive peptide is denoted in

bold along with the number of times it is repeated in the protein sequence. Neighboring residues of tryptic cleavage sites that would result in a peptide of AA≤2 are indicated by a parenthesis containing that residue.

Table S3b **Extensin 3: Q9FS16**

Signal Peptide:	MGSPMASLVATLLVLTISLTFVSQSTA
Peptide 1: (1)	NyFySSPPPPVK
Peptide 2: (1)	HyTPPVK
Peptide 3: (12)	HySPPPVyHSPPPPK(K)
Peptide 4: (1)	HyEyK
Peptide 5: (12)	SPPPPVK
Peptide 3:	HySPPPVyHSPPPPK(K)
Peptide 6: (11)	HyVyK
Peptide 5:	SPPPPVK
Peptide 3:	HySPPPVyHSPPPPK(K)
Peptide 6:	HyVyK
Peptide 5:	SPPPPVK
Peptide 3:	HySPPPVyHSPPPPK(K)
Peptide 6:	HyVyK
Peptide 5:	SPPPPVK
Peptide 3:	HySPPPVyHSPPPPK(K)
Peptide 6:	HyVyK
Peptide 5:	SPPPPVK
Peptide 3:	HySPPPVyHSPPPPK(K)
Peptide 6:	HyVyK
Peptide 5:	SPPPPVK
Peptide 3:	HySPPPVyHSPPPPK(K)
Peptide 6:	HyVyK
Peptide 5:	SPPPPVK
Peptide 3:	HySPPPVyHSPPPPK(K)
Peptide 6:	HyVyK
Peptide 5:	SPPPPVK
Peptide 3:	HySPPPVyHSPPPPK(K)
Peptide 6:	HyVyK
Peptide 5:	SPPPPVK
Peptide 3:	HySPPPVyHSPPPPK(K)

Peptide 6:	H**y**V**y**K
Peptide 5:	SPPPPVK
Peptide 3:	H**y**SPPPV**y**HSPPPPK(K)
Peptide 6:	H**y**V**y**K
Peptide 5:	SPPPPVK
Peptide 3:	H**y**SPPPV**y**HSPPPPK(K)
Peptide 6:	H**y**V**y**K
Peptide 5:	SPPPPVK
Peptide 7: (1)	H**y**SPPPV**y**HSPPPPK(EK)
Peptide 8: (1)	**y**V**y**K
Peptide 9: (1)	SPPPPPVHH**y**SPPHHP**y**L**y**K
Peptide 10: (1)	SPPPP**y**H**y**

Table S3b: Full Protein Sequence of EXT3: Tryptic digest of EXT3 results in 10 unique peptides. Several peptides are repetitive. The first occurrence of a repetitive peptide is denoted in bold along with the number of times it is repeated in the protein sequence. Neighboring residues of tryptic cleavage sites that would result in a peptide of AA≤2 are indicated by a parenthesis containing that residue.

Table S3c **Extensin 8: Q9ZW80**

Signal Peptide:	MATPAWSHAKAQWVVAMLALLVGSAMA
Peptide 1: (1)	TEP**y**Y**y**SSPPPP**y**E**y**K
Peptide 2: (2)	SPPPPVK
Peptide 3: (1)	SPPPP**y**E**y**K
Peptide 2:	SPPPPVK
Peptide 4: (6)	SPPPP**y**Y**y**HSPPPPVK
Peptide 5: (1)	SPPPP**y**V**y**SSPPPPVK
Peptide 4:	SPPPP**y**Y**y**HSPPPPVK
Peptide 4:	SPPPP**y**Y**y**HSPPPPVK
Peptide 4:	SPPPP**y**Y**y**HSPPPPVK
Peptide 4:	SPPPP**y**Y**y**HSPPPPVK
Peptide 4:	SPPPP**y**Y**y**HSPPPPVK
Peptide 6: (1)	SPPPP**y**L**y**SSPPPPVK
Peptide 7: (1)	SPPPPV**y**I**y**ASPPPPTH**y**

Table S3c: Full Protein Sequence of EXT8: Tryptic digest of EXT8 results in 7 unique peptides. Several peptides are repetitive. The first occurrence of a repetitive peptide is denoted in bold along with the number of times it is repeated in the protein sequence.

Table S3d Extensin 10: F4JCZ1

Signal Peptide:	MRSSSMMGPSTHLICALGVVIMATMVAA
Peptide 1: (1)	yDPyTDSSPPPLySSPLPK
Peptide 2: (1)	IEyK
Peptide 3: (1)	TPPLPyIDSSPPPTySPAPEVEyK
Peptide 4: (12)	SPPPPyVySSPPPPTySPSPK
Peptide 5: (14)	VEyK
Peptide 4:	SPPPPyVySSPPPPTySPSPK
Peptide 5:	VEyK
Peptide 4:	SPPPPyVySSPPPPTySPSPK
Peptide 5:	VEyK
Peptide 4:	SPPPPyVySSPPPPTySPSPK
Peptide 5:	VEyK
Peptide 4:	SPPPPyVySSPPPPTySPSPK
Peptide 5:	VEyK
Peptide 6: (17)	SPPPPyVySSPPPPyySPSPK
Peptide 5:	VEyK
Peptide 4:	SPPPPyVySSPPPPTySPSPK
Peptide 7: (5)	VDYK
Peptide 6:	SPPPPyVySSPPPPyySPSPK
Peptide 5:	VEyK
Peptide 4:	SPPPPyVySSPPPPTySPSPK
Peptide 7:	VDYK
Peptide 6:	SPPPPyVySSPPPPyySPSPK
Peptide 7:	VDYK
Peptide 4:	SPPPPyVySSPPPPTySPSPK
Peptide 5:	VEyK
Peptide 4:	SPPPPyVySSPPPPTySPSPK
Peptide 5:	VEyK
Peptide 4:	SPPPPyVySSPPPPTySPSPK
Peptide 5:	VEyK

Peptide 4:	SPPPPyVySSPPPPTySPSPK
Peptide 5:	VEyK
Peptide 6:	SPPPPyVySSPPPPyySPSPK
Peptide 5:	VEyK
Peptide 4:	SPPPPyVySSPPPPTySPSPK
Peptide 7:	VDYK
Peptide 6:	SPPPPyVySSPPPPyySPSPK
Peptide 8: (9)	VyyK
Peptide 6:	SPPPPyVySSPPPPyySPSPK
Peptide 8:	VyyK
Peptide 6:	SPPPPyVySSPPPPyySPSPK
Peptide 8:	VyyK
Peptide 9: (1)	SPPPPyySPSPK
Peptide 8:	VyyK
Peptide 6:	SPPPPyVySSPPPPyySPSPK
Peptide 8:	VyyK
Peptide 10: (1)	SPPPPyVySSPPPPyHSPSPK
Peptide 11: (1)	VQyK
Peptide 6:	SPPPPyVySSPPPPyySPSPK
Peptide 8:	VyyK
Peptide 6:	SPPPPyVySSPPPPyySPSPK
Peptide 8:	VyyK
Peptide 6:	SPPPPyVySSPPPPyySPSPK
Peptide 8:	VyyK
Peptide 12: (1)	SPPHPHVCVCPPPPPCySPSPK
Peptide 13: (1)	VVyK
Peptide 14: (1)	SPPPPyVySSPPPPHySPSPK
Peptide 8:	VyyK
Peptide 6:	SPPPPyVySSPPPPyySPSPK
Peptide 15: (4)	VHyK
Peptide 16: (1)	SPPPPyyAPTPK
Peptide 15:	VHyK
Peptide 6:	SPPPPyVySSPPPPyySPSPK
Peptide 15:	VHyK
Peptide 6:	SPPPPyVySSPPPPyySPSPK
Peptide 5:	VEyK
Peptide 6:	SPPPPyVySSPPPPyySPSPK
Peptide 15:	VHyK
Peptide 6:	SPPPPyVySSPPPPyySPSPK

Peptide 7:	VDYK
Peptide 17: (1)	SPPPPyVySSPPPPyySPSPVVDyK
Peptide 6:	SPPPPyVySSPPPPyySPSPK
Peptide 5:	VEyK
Peptide 18: (1)	SPPPPyVyK
Peptide 19: (1)	SPPPPSySPSPK
Peptide 20: (1)	TEy

Table S3d: Full Protein Sequence of EXT10: Tryptic digest of EXT10 results in 20 unique peptides. Several peptides are repetitive. The first occurrence of a repetitive peptide is denoted in bold along with the number of times it is repeated in the protein sequence.

Table S3e **LRX 4: Q9LHF1**

Signal Peptide:	MKNNTTQSLLLLLLFFFFFFEISHS
Peptide 1:	LSISSNAPLSDTEVR
Peptide 2:	FIQR(R)
Peptide 3:	QLLyyR
Peptide 4:	DEFGDR
Peptide 5:	GENVTVDPSLIFENPR
Peptide 6:	LR
Peptide 7:	SAyIALQAWK
Peptide 8:	QAILSDPNNITVNWIGSNVCNYTGVFCSK
Peptide 9:	ALDNR(K)
Peptide 10:	IR
Peptide 11:	TVAGIDLNHADIAGyLPEELGLLTDLALFHVNSNR
Peptide 12:	FCGTVPHK
Peptide 13:	FK
Peptide 14:	QLK
Peptide 15:	LLFELDLSNNR
Peptide 16:	FAGK
Peptide 17:	FPTVVLHLPSLK
Peptide 18:	FLDLR
Peptide 19:	FNEFEGTVPK
Peptide 20:	ELFSK
Peptide 21:	NLDAIFINHNR

Peptide 22:	FR
Peptide 23:	FELPENFGDSPVSVIVLANNHFHGCIPTSLVEMK
Peptide 24:	NLNEIIFMNNGLNSCLPADIGR
Peptide 25:	LKNVTVFDVSFNELVGPLPESVGGMVEVEQLNVAHNLLSGK
Peptide 26:	IPASICQLPK
Peptide 27:	LENFTySSyNFFTGEAPVCLR
Peptide 28:	LSEFDDR(R)
Peptide 29:	NCLPGR
Peptide 30:	PAQR
Peptide 31:	SSR
Peptide 32:	QCSAFLSR
Peptide 33:	PSVDCGSFGCGR
Peptide 34:	SVVK
Peptide 35:	PSPPIVALPPPPPPSPPLPPPVySPPPSPPVFSPPPSPPVySPPPPPSIHySS PPPPPVHHSSPPPPSPEFEGPLPPVIGVSyASPPPPPFy

Table S3e: Full Protein Sequence of LRX4: Tryptic digest of LRX4 results in 35 peptides. Neighboring residues of tryptic cleavage sites that would result in a peptide of AA≤2 are indicated by a parenthesis containing that residue. Four of these peptides contain sites of potential tyrosine cross-linkage. However, peptide 35 would be too large to find using our mass spectrometry methods. There are no repetitive peptides that arise from tryptic digestion of LRX4.

VITA

Lawrie Veale Gainey

Candidate for the Degree of

Doctor of Philosophy

Dissertation: MASS SPECTROMETRIC CHARACTERIZATION OF EXTENSIN CROSS-LINKAGES IN ARABIDOPSIS TISSUE CULTURE CELL WALLS

Major Field: Biochemistry and Molecular Biology

Biographical:

Education:

Completed the requirements for the Doctor of Philosophy in Biochemistry and Molecular Biology at Oklahoma State University, Stillwater, Oklahoma in May, 2020.

Completed the requirements for the Bachelor of Science in Biochemistry and Molecular Biology at Oklahoma State University, Stillwater, Oklahoma in 2014.