

Identification and Aspects of Commercial Production of Anti-
Diabetic Peptide(s) from Salmon Protein Hydrolysates

by

Jonathan Rolin

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

at

Dalhousie University
Halifax, Nova Scotia
December 2019

© Copyright by Jonathan Rolin, 2019

To my parents and family for their unconditional love and support

TABLE OF CONTENTS

List of Tables	vii
List of Figures.....	xi
Abstract.....	xv
List of Abbreviations and Symbols Used.....	xvi
Acknowledgements.....	xviii
Chapter 1: Introduction.....	1
Chapter 2: Literature Review	3
2.1 Bioactive Peptides	3
2.2 Diabetes Mellitus.....	7
2.2.1 Physiology and Pathology of Type 2 Diabetes	7
2.2.2 Small Molecular Modulators of Glucose Uptake.....	8
2.2.3 Anti-Diabetic Compounds in Foods and Natural Sources.....	10
2.3 Marine Muscle Proteins.....	15
2.3.1 Effect of pH and Ionic Strength on Protein Solubility	15
2.3.2 Myofibrillar Proteins	17
2.3.3 Sarcoplasmic Proteins.....	18
2.3.4 Stromal Proteins.....	18
2.4 Marine Teleost Protein Hydrolysates as a Source of Bioactive Peptides..	19
2.5 Methodology for the Purification and Identification of Low Molecular Weight Peptides.....	21
2.5.1 Chromatographic Fractionation of LMW Peptides	21
2.5.2 Non-Chromatographic Fractionation of Proteins and Peptides.....	26
2.5.3 Protein and Peptide Detection and Identification	29
Chapter 3: Materials and Methods	36
3.1 Preparation of SPF	36

3.2 Fractionation of the SPF using Strong Anion Exchange and Gel Filtration Chromatography.....	37
3.3 <i>In vitro</i> Screening Assay for SPF Action on Metabolism	41
3.4 Protein Identification and Peptide Detection: LC-MS Method 1	42
3.5 Amino Acid Analysis	43
3.6 Peptide Sequence Alignment.....	44
3.7 Optimization of Protein Recovery from Atlantic Salmon using Alkaline Solubilization and Isoelectric Precipitation	44
3.8 Solubility Curves	45
3.9 Preparation of Salmon Protein Isoelectric Precipitates	47
3.10 Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis	48
3.11 In-Gel Digestion	50
3.12 Protein Identification.....	51
3.13 Peptide Detection: LC-MS Method 2.....	51
3.14 Sequence Logos	53
3.15 Peptide Identification: <i>In silico</i> Digestion of SPF Progenitor Proteins	54
3.16 Peptide Identification: <i>De novo</i> Sequencing.....	55
3.17 Validation of Potential Bioactive Peptides	56
Chapter 4: Objectives.....	57
Chapter 5: Results and Discussion	59
5.1 SAX Separation 1 of SPF.....	59
5.1.1 Objective	59
5.1.2 Chromatography of SAX Separation 1	59
5.1.3 Glucose Uptake Analysis of Fractions of SAX Separation 1.....	61
5.1.4 Peptide Identification	62
5.1.5 Protein Identification from SAX Separation 1	70
5.1.6 Conclusion.....	73
5.1.7 Future Work.....	73
5.2 SAX Separation 2 and GF Separation 3 of SPF	74
5.2.1 Objectives.....	74
5.2.2 Chromatography of SAX Separation 2 and GF Separation 3	75

5.2.3	Glucose Uptake Analysis of Fractions from SAX Separation 2 and GF Separation 3 of SPF	79
5.2.4	Peptide Identification in Fractions from SAX Separation 2 and GF Separation 3	82
5.2.5	Conclusion.....	83
5.2.6	Future Work.....	83
5.3	Alkaline Solubilization of Proteins from Salmon Muscle Tissue	84
5.3.1	Objective	84
5.3.2	Evaluation of Protein Recovery from Salmon Muscle.....	85
5.3.3	Protein Identification	99
5.3.4	Glucose Uptake of Low Molecular Weight Peptides Derived from Proteins Precipitated Directly from Salmon Muscle Tissues	101
5.3.5	Peptide Identification	103
5.3.6	Conclusion.....	107
5.3.7	Future Work.....	108
5.4	GF Separation 4 and SAX Separation 5 of SPF	109
5.4.1	Objective	109
5.4.2	Chromatography of GF Separation 4 and SAX Separation 5	110
5.4.3	Glucose Uptake Analysis of GF Separation 4 and SAX Separation 5 of SPF	114
5.4.4	Identification and Validation of Peptides in Bioactive Fractions	117
5.4.5	Characteristics of Validated Peptides in Bioactive Fractions	124
5.4.6	Peptide Compositions of Bioactive Fractions	132
5.4.7	Consensus Residue Positioning in Peptides from each Fraction	154
5.4.8	Previously Identified Bioactive Peptides	159
5.4.9	<i>In Silico</i> Digestion of a Curated Protein Library	164
5.4.10	Conclusion.....	176
5.4.11	Future Work.....	177
5.5	Validation of Putative Bioactive Peptide Sequences.....	178
5.5.1	Objective	179
5.5.2	Peptide Selection	180
5.5.3	Glucose Uptake Analysis of Synthetic Peptides	191

5.5.4 Prediction of Peptide Interactions using Structural Homology	194
5.5.5 Conclusion.....	203
5.5.6 Future Work.....	203
References	205
Appendix I	223

LIST OF TABLES

Table 1	Common anti-diabetic pharmaceutical drugs.....	11
Table 2	Foods identified as hosts to anti-diabetic compounds and their characteristics.....	12
Table 3	Hydrolysate from <i>Salmo salar</i> muscle protein investigated for glucose uptake stimulation.....	20
Table 4	Non-chromatographic methods of protein and peptide fractionation in proteomic and bioactive peptide research.....	28
Table 5	Operational characteristics for strong anion exchange and gel filtration separations of SPF.....	39
Table 6	Elution strategies prepared for strong anion exchange and gel filtration separation of SPF.....	40
Table 7	Glucose uptake modulation of SPF and its fractions from SAX Separation 1.....	62
Table 8	Composition of bioactive fractions from SAX Separation 1.....	63
Table 9	Amino acid analysis of fractions from SAX Separation 1.....	66
Table 10	Parent proteins of identified peptide in fractions from SAX Separation 1.....	71
Table 11	Glucose uptake modulation of SPF and its fractions from SAX Separation 2.....	80
Table 12	Glucose uptake modulation of SPF and its fractions from GF Separation 3.....	81
Table 13	Glucose uptake modulation of hydrolysate samples processed following Jin (2012) protocol 2 using high- or low-alkali solubilized salmon muscle.....	90
Table 14	Putative identities of the major bands identified in the low-alkali solubilized isoelectrically precipitate salmon muscle protein.....	100
Table 15	Glucose uptake modulation of hydrolysates produced from isolated <i>Salmo salar</i> proteins.....	102
Table 16	Glucose uptake modulation of fractions from GF Separation 4.....	115

Table 17	Glucose uptake modulation of fractions from SAX Separation 5 ...	116
Table 18	Comparison of various validation criteria for the identification of peptide in bioactive fractions from Separations 4 and 5	120
Table 19	Candidate bioactive peptides identified in glucose uptake modulating fractions from GF Separation 4 and SAX Separation 5	148
Table 20	Previously identified bioactive peptides reported in the bioactive peptide database, BIOPEP	162
Table 21	Peptide identification of dominant ions in the base peak chromatograms of fractions from GF Separation 4 and SAX Separation 5	171
Table 22	Glucose uptake modulation of chemically synthesized peptides representing the Ile-X-Ile and Ile-X-Tyr motifs	192
Table 23	Swiss Target Prediction of chemically synthesized peptides	197
Table 24	Swiss Target Prediction of dipeptides containing one branched chain amino acid and one aromatic amino acid.	198
Table 25	Peptides identified in bioactive SPF fractions resembling opioid receptor binding motifs.	202
Table A1	Peptides identified by LC-MS in fraction 1 from SAX Separation 1	223
Table A2	Peptides identified by LC-MS in fraction 2 from SAX Separation 1	230
Table A3	Peptides validated using the modified criteria in fraction 2 from GF Separation 4	235
Table A4	Peptides validated using the modified criteria in fraction 3 from GF Separation 4	236
Table A5	Peptides validated using the modified criteria in fraction 6 from GF Separation 4	238
Table A6	Peptides validated using the modified criteria in fraction 7 from GF Separation 4	239

Table A7	Peptides validated using the modified criteria in fraction 9 from GF Separation 4	239
Table A8	Peptides validated using the modified criteria in fraction 6 from SAX Separation 5	240
Table A9	Peptides validated using the modified criteria in fraction 7 from SAX Separation 5	241
Table A10	Primary sequences of identified progenitor proteins from the low-alkali solubilized salmon muscle	244
Table A11	Compound list generated using the Find by algorithm with Molecular Feature Extraction using Agilent MassHunter Software on hydrolysates generated from the extracted proteins from band 2.....	247
Table A12	Compound list generated using the Find by algorithm with Molecular Feature Extraction using Agilent MassHunter Software on hydrolysates generated from the extracted proteins from band 4.....	249
Table A13	Compound list generated using the Find by algorithm with Molecular Feature Extraction using Agilent MassHunter Software on hydrolysates generated from the extracted proteins from band 6.....	252
Table A14	Compound list generated using the Find by algorithm with Molecular Feature Extraction using Agilent MassHunter Software for fraction 2 from GF Separation 4.	254
Table A15	Compound list generated using the Find by algorithm with Molecular Feature Extraction using Agilent MassHunter Software for fraction 3 from GF Separation 4	266
Table A16	Compound list generated using the Find by algorithm with Molecular Feature Extraction using Agilent MassHunter Software for fraction 6 from GF Separation 4	282
Table A17	Compound list generated using the Find by algorithm with Molecular Feature Extraction using Agilent MassHunter Software for fraction 7 from GF Separation 4	286
Table A18	Compound list generated using the Find by algorithm with Molecular Feature Extraction using Agilent MassHunter Software for fraction 9 from GF Separation 4	287

Table A19	Compound list generated using the Find by algorithm with Molecular Feature Extraction using Agilent MassHunter Software for fraction 6 from SAX Separation 5	289
Table A20	Compound list generated using the Find by algorithm with Molecular Feature Extraction using Agilent MassHunter Software for fraction 7 from SAX Separation 5	307

LIST OF FIGURES

Figure 1	Experimental approaches for the discovery of bioactive peptides	4
Figure 2	Peptide identification by tandem mass spectrometry combined with database searching	30
Figure 3	Simulated fragmentation of a tripeptide precursor ion and the product ions used for its identification by <i>de novo</i> sequencing	35
Figure 4	SAX Separation 1 of SPF	60
Figure 5	The precursor ion mass and isoelectric point distribution of peptides identified by LC-MS in bioactive fractions from SAX Separation 1	64
Figure 6	Multiple sequence alignment using Clustal Omega of sequences identified in fractions 1 and 2 from SAX Separation 1.....	69
Figure 7	SAX Separation 2 of SPF	77
Figure 8	GF Separation 3 of SPF.....	78
Figure 9	Protein recovery (%) of alkali-solubilized <i>Salmo salar</i> muscle precipitated by IEP using various pH combinations	86
Figure 10	Solubility curves of the (A) high-alkali solubilized salmon muscle and (B) low-alkali solubilized salmon muscle	88
Figure 11	SDS-PAGE showing the protein distributions of (A) high- and low-alkali solubilized salmon muscle proteins and B) high- and low-alkali solubilized salmon prepared from whole loin or mechanically deboned frame meat	92
Figure 12	SDS-PAGE showing the protein distributions of (A) insoluble and (B) soluble muscle proteins, in isoelectric precipitates from Atlantic salmon muscle tissue solubilized using NaOH concentrations between 1.0 and 0.1 M	95
Figure 13	SDS-PAGE showing the protein distribution of TCA-insoluble protein from the (A) high-alkali and (B) low-alkali solubilized salmon muscle protein as a function of solubilization time	98

Figure 14	The predicted compositions of bioactive hydrolysates produced from isolated <i>Salmo salar</i> proteins using the Molecular Feature Extraction algorithm with comparisons to <i>in silico</i> protein digestion libraries.	106
Figure 15	GF Separation 4 of SPF.....	111
Figure 16	SAX Separation 5 of SPF	113
Figure 17	Molecular weight distribution of identified peptides in bioactive fractions of Separations 4 and 5 analyzed by LC-MS following the application of various validation criteria	121
Figure 18	Abundance (%) of individual amino acids in all fractions following each validation criteria	123
Figure 19	The precursor ion mass and isoelectric point distribution of peptides identified in the bioactive fractions of Separation 4 and 5.....	125
Figure 20	Molecular weight distribution of (A) database search hits (B) molecular feature extraction compound lists, of each bioactive fraction from Separations 4 and 5.....	126
Figure 21	Abundance (%) of individual amino acids as components of validated peptides in fractions of Separations 4 and 5.....	128
Figure 22	LC-MS analysis of fraction 2 of GF Separation 4 (F2_S4), represented using the (A) base peak chromatogram and (B) product ion spectra (MS/MS)	134
Figure 23	LC-MS analysis of fraction 3 of GF Separation 4 (F3_S4), represented using the (A) base peak chromatogram and (B) product ion spectra (MS/MS)	137
Figure 24	LC-MS analysis of fraction 6 of GF Separation 4 (F6_S4), represented using the (A) base peak chromatogram and (B) product ion spectra (MS/MS)	139
Figure 25	LC-MS analysis of fraction 7 of GF Separation 4 (F7_S4), represented using the (A) base peak chromatogram and (B) product ion spectra (MS/MS)	141
Figure 26	LC-MS analysis of fraction 9 of GF Separation 4 (F9_S4), represented using the (A) base peak chromatogram and (B) product ion spectra (MS/MS)	143

Figure 27	LC-MS analysis of fraction 6 of SAX Separation 5 (F6_S5), represented using the (A) base peak chromatogram and (B) product ion spectra (MS/MS)	146
Figure 28	LC-MS analysis of fraction 7 of SAX Separation 4 (F7_S5), represented using the (A) base peak chromatogram and (B) product ion spectra (MS/MS)	147
Figure 29	Sequence logos of peptides in glucose uptake stimulating fractions	156
Figure 30	Sequence logo of tripeptides in the glucose uptake inhibiting fraction 6 of GF Separation 4.....	157
Figure 31	(A) The extracted ion chromatogram at m/z 356.5 – 360.0 from fraction 3 of GF Separation 4 targeting the Ile-Ile-Ile peptide ion and (B) the superimposed MS/MS spectra of each indicated precursor ion	185
Figure 32	(A) The extracted ion chromatogram at m/z 314.5 – 318.0 from fraction 3 of GF Separation 4 targeting the Ile-Ala-Ile peptide ion and (B) the superimposed MS/MS spectra of each indicated precursor ion	186
Figure 33	(A) The extracted ion chromatogram at m/z 300.5 – 303.9 from fraction 3 of GF Separation 4 targeting the Ile-Gly-Ile peptide ion and (B) the superimposed MS/MS spectra of each precursor ion.....	187
Figure 34	(A) The extracted ion chromatogram at m/z 408.1 – 408.3 from fraction 6 of GF Separation 4 targeting the Ile-Ile-Tyr peptide ion and (B) the superimposed MS/MS spectra of each indicated precursor ion	188
Figure 35	(A) The extracted ion chromatogram at m/z 366.1 – 366.3 from fraction 6 of GF Separation 4 targeting the Ile-Ala-Tyr peptide ion and (B) the superimposed MS/MS spectra of each indicated precursor ion	189
Figure 36	(A) The extracted ion chromatogram at m/z 352.1 – 352.3 from fraction 6 of GF Separation 4 targeting the Ile-Gly-Tyr peptide ion and (B) the superimposed MS/MS spectra of each indicated precursor ion	190

Figure A1	Simulated fragmentation of the Ile-Ile-Ile peptide during LC-MS analysis to facilitate its <i>de novo</i> sequencing	333
Figure A2	Simulated fragmentation of the Ile-Ala-Ile peptide during LC-MS analysis to facilitate its <i>de novo</i> sequencing	334
Figure A3	Simulated fragmentation of the Ile-Gly-Ile peptide during LC-MS analysis to facilitate its <i>de novo</i> sequencing	335
Figure A4	Simulated fragmentation of the Ile-Ile-Tyr peptide during LC-MS analysis to facilitate its <i>de novo</i> sequencing	336
Figure A5	Simulated fragmentation of the Ile-Ala-Tyr peptide during LC-MS analysis to facilitate its <i>de novo</i> sequencing	337
Figure A6	Simulated fragmentation of the Ile-Gly-Tyr peptide during LC-MS analysis to facilitate its <i>de novo</i> sequencing	338

ABSTRACT

A low molecular weight (< 1 kDa) hydrolysate of Atlantic salmon protein has previously been shown to have anti-diabetic effects *in vitro* and in mouse models. This salmon peptide fraction (SPF) produced using the enzymes pepsin, trypsin and chymotrypsin, contains hundreds of potential bioactive compounds and the identification the functional modulators of this activity may realize novel therapeutic compounds or targets as treatments for type 2 diabetes. The aim of this study was therefore to identify the sequences and characteristics of potential bioactive peptides from the SPF with a functional effect in cultured L6 myotubes. Separation of progenitor proteins by electrophoresis and of SPF by column chromatography with gel filtration and strong anion exchange formats were successful at concentrating bioactive peptides into multiple subfractions. Tandem mass spectrometry of subfractions suggested that di- and tripeptides composed of Ile/Leu, Val, Asp, Glu, Trp, Tyr were common among peptides identified in bioactive fractions. The importance of functional peptide concentration and sequence motifs on bioactivity was tested using the synthetic sequences Ile-Ala-Tyr and Ile-Gly-Tyr and exhibited a stimulating effect at 2.8 nM, but an inhibiting effect at 2.8 pM. Swiss Target Prediction suggested these sequences are peptidomimetics of agonists to mu-type, delta-type and kappa-type opioid receptors. When considering the large total number of peptides and abundance of sequences containing similar motifs in SPF, its bioactivity is likely the result of complex interactions of many peptides.

LIST OF ABBREVIATIONS AND SYMBOLS USED

2DG	2-deoxy-D-[³ H] glucose
AA	Amino Acid
AAA	Aromatic Amino Acid
ACE	Angiotensin Converting Enzyme
AH	Alkaline Hydrolysis
ALDOA	Fructose Bisphosphate Aldolase
AMPK	5' AMP-Activated Protein Kinase
ANOVA	Analysis of Variance
BCA	Bicinchoninic Acid
BCAA	Branched Chain Amino Acid
BPC	Base Peak Chromatogram
BSA	Bovine Serum Albumin
CF	Chromatofocusing
CID	Collision Induced Dissociation
CK	Creatine Kinase
CV	Column Volume
CVD	Cardiovascular Disease
Da	Daltons
DM	Diabetes Mellitus
DPM	Disintegrations Per Minute
DPP-IV	Dipeptidyl Peptidase IV
E:S	Enzyme-To-Substrate Ratio
EAA	Essential Amino Acid
EC ₅₀	Half Maximal Effective Concentration
EDFM	Electrodialysis With Filtration Membranes
EIC	Extracted Ion Chromatogram
ESI	Electrospray Ionization
FDR	False Discovery Rate
FPLC	Fast Protein Liquid Chromatography
G3P Dh	Glyceraldehyde-3-Phosphate Dehydrogenase
GF	Gel Filtration
GI	Gastrointestinal
GIP	Gastric Inhibitory Polypeptide
GLP-1	Glucagon-Like Peptide-1
HCD	High-Energy Collision Dissociation
HIC	Hydrophobic Interaction Chromatography
HILIC	Hydrophilic-Lipophilic Interaction Chromatography
HMW	High Molecular Weight
HPLC	High Performance Liquid Chromatography
IC ₅₀	Half Maximal Inhibitor Concentration
IEF	Isoelectric Focusing
IEP	Isoelectric Precipitation
IEX	Ion Exchange Chromatography
IR	Insulin Receptor

IRS-1	Insulin Receptor Substrate 1
IS	Ionic Strength
ISP	Isoelectric Solubilization/Precipitation
LC	Liquid Chromatography
LMW	Low Molecular Weight
LOD	Limit of Detection
LPLC	Low Pressure Liquid Chromatography
LSD	Least Squared Difference
MetS	Metabolic Syndrome
MFE	Molecular Feature Extraction
MHC	Myosin Heavy Chain
MLC-2	Myosin Light Chain-2
MS	Mass Spectrometry
MS ¹	Stage 1 Ion Separation by MS – Precursor Ions
MS ²	Stage 2 Ion Separation by MS – Fragment Ions
MS/MS	Tandem Mass Spectrometry
mTOR1	Mechanistic Target of Rapamycin 1
MW	Molecular Weight
NHP	Natural Health Product
ODS	Octadecylsilane
PEG	Polyethylene Glycol
PGM	Phosphoglycerate Mutase
pI	Isoelectric Point
PPAR _γ	Peroxisome Proliferator-Activated Receptor Gamma
PPG	Polypropylene Glycol
PPI	Protein-Protein Interactions
Q-TOF	Quadrupole-Time-of-Flight
QSAR	Quality Structure-Activity Relationship
RP	Reversed Phase
RT	Retention Time
SAX	Strong Anion Exchange
SDS-PAGE	Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis
SEM	Standard Error of the Mean
SLiM	Short Linear Motif
SPF	Salmon Peptide Fraction
T2DM	Type 2 Diabetes Mellitus
TCA	Trichloroacetic Acid
TFF	Tangential Flow Filtration
TICC	Total Ion Current Chromatogram
TM	Tropomyosin
TPIS	Triosephosphate Isomerase
UPLC	Ultra Performance Liquid Chromatography
UV	Ultraviolet

ACKNOWLEDGEMENTS

I would like to extend deep gratitude to Drs. Tom Gill, Allan Paulson, André Marette, Laurent Bazinet, and David Byers for their contribution, assistance and patience throughout this journey. I want to thank Dr. Alejandro Cohen of the Dalhousie Biological Mass Spectrometry Core Facility for generously offering his consultation and assistance with data analysis and interpretation. I also want to thank both Bruno Marcotte, Marion Valle and Jacinthe Thibodeau for their kind assistance with glucose uptake assay, LC-MS, and assisting with the data interpretation. As well, I am thankful for financial support from the Nova Scotia Graduate Scholarship and the Mitacs Accelerate Program (Project: Salmon peptide identification and purification: insulin modulation).

CHAPTER 1: INTRODUCTION

Farmed Atlantic salmon (*Salmo salar*) is a sustainable and efficient model for supplying an increasing consumer protein demand and all processing by-products including meat, oil, skin, and bones can be recovered and further processed into value-added products. The global production of aquaculture Atlantic salmon totaled 2.25 M tonnes in 2016 with Canada contributing 5.5 % (FAO, 2018), and post-harvest processing can divert up to 50 % of this mass as by-product. Traditionally, by-products are converted into low-value fish feed, fish silage or pet foods, but the affordability and quality of plant proteins have reduced the demand of these products. Proteins from marine by-products are high in essential amino acids, display broad functionalities and are nutritionally equivalent to those sold for human consumption, representing a suitable target for value-added processing (Atef and Mahdi Ojagh, 2017). Marine proteins have also been identified as sources of bio(logically)active peptides; physiologically relevant products of protein digestion. Hydrolyzed protein from Atlantic salmon has been demonstrated to exhibit anti-inflammatory activity (Pilon et al., 2011), antioxidative activity (Girgih et al., 2013), renin and angiotensin I-converting enzyme (ACE) inhibition (Girgih et al., 2016; Neves, Harnedy, O'Keeffe, and FitzGerald, 2017), and dipeptidyl peptidase IV (DPP-IV) inhibition (Neves, Harnedy, O'Keeffe, Alashi, et al., 2017). A low molecular weight (LMW) (MW < 1000 Daltons, Da) salmon peptide fraction (SPF) has been shown to protect against Type 2 Diabetes Mellitus (T2DM) and the development of several pathologies associated with metabolic syndrome (MetS, Chevrier et al., 2015). The conversion of Atlantic salmon protein by-product

into a commercialized natural health product (NHP) or pharmaceutical can be desirable, but the identification of bioactive peptides mediating these functions is critical for its development.

Fractionation of the SPF by reversed-phase high-performance liquid chromatography (RP-HPLC) was unable to concentrate its glucose uptake modulating activity (Girgih et al., 2013), but electro dialysis with filtration membranes (EDFM) generated fractions consisting of anionic, cationic or neutral peptides that exhibited variable modulation of the anti-diabetic effects (Henaux et al., 2019; Roblet et al., 2016). Results from both studies did not identify candidate bioactive peptide sequences. Therefore, the objective of this study was to identify peptide sequences found within the SPF that are involved in the mediation of its bioactivity.

In the initial chapters of this thesis, a review of the literature, followed by methods and materials, and the main objectives of this study will be presented. Then, the chromatographic fractionation of SPF, and identification of LMW peptides are optimized to provide a comprehensive analysis of the SPF composition. The functional SPF fractions are evaluated qualitatively and used to select prospective bioactive sequences for chemical synthesis. The validation of their glucose uptake modulating peptides and a proposed mechanism of action will conclude this thesis.

CHAPTER 2: LITERATURE REVIEW

2.1 Bioactive Peptides

Bioactive peptides are functional molecules that interact with the cardiovascular (Erdmann et al., 2008), nervous (Saavedra et al., 2013), gastrointestinal (GI) and immune systems (Schanbacher et al., 1997). The typical size of bioactive peptides range between 2 and 30 amino acids in length (Pihlanto and Korhonen, 2003), and are most frequently investigated from terrestrial animals and their by-products (Schanbacher et al., 1997), marine organisms (Giri and Ohshima, 2012), plants such as corn, soy and wheat (Pihlanto and Korhonen, 2003), and microorganisms (Bahar and Ren, 2013) as a common value-added approach for process by-product handling. *Ex vivo* generation of protein hydrolysates under non-gastric hydrolytic conditions enables the generation of novel peptides and the opportunity to identify novel therapeutic compounds.

Methodologies that optimize hydrolysate processing conditions for a desired biological functions and the assays to screen for these functions are constantly under development and improving (Nongonierma and FitzGerald, 2017). Following the empirical approach to bioactive peptide identification, hydrolysates are prepared directly from food proteins extracts and are subjected to additional fractionation to minimize sample complexity and concentrate potential bioactive peptides (Daliri et al., 2018) (Figure 1). Bioactive peptide sequences and their functionalities are stored in bioactive peptide databases such as BIOPEP (<http://www.uwm.edu.pl/biochemia/index.php/en/biopep>; (Minkiewicz et al., 2008),

EROP-Moscow (<http://erop.inbi.ras.ru>; (Zamyatnin, 2006), and PepBank (<http://pepbank.mgh.harvard.edu>; (Shtatland et al., 2007). BIOPEP sequences are derived from food sources, providing the opportunity to screen various food commodities as sources of bioactive peptides.

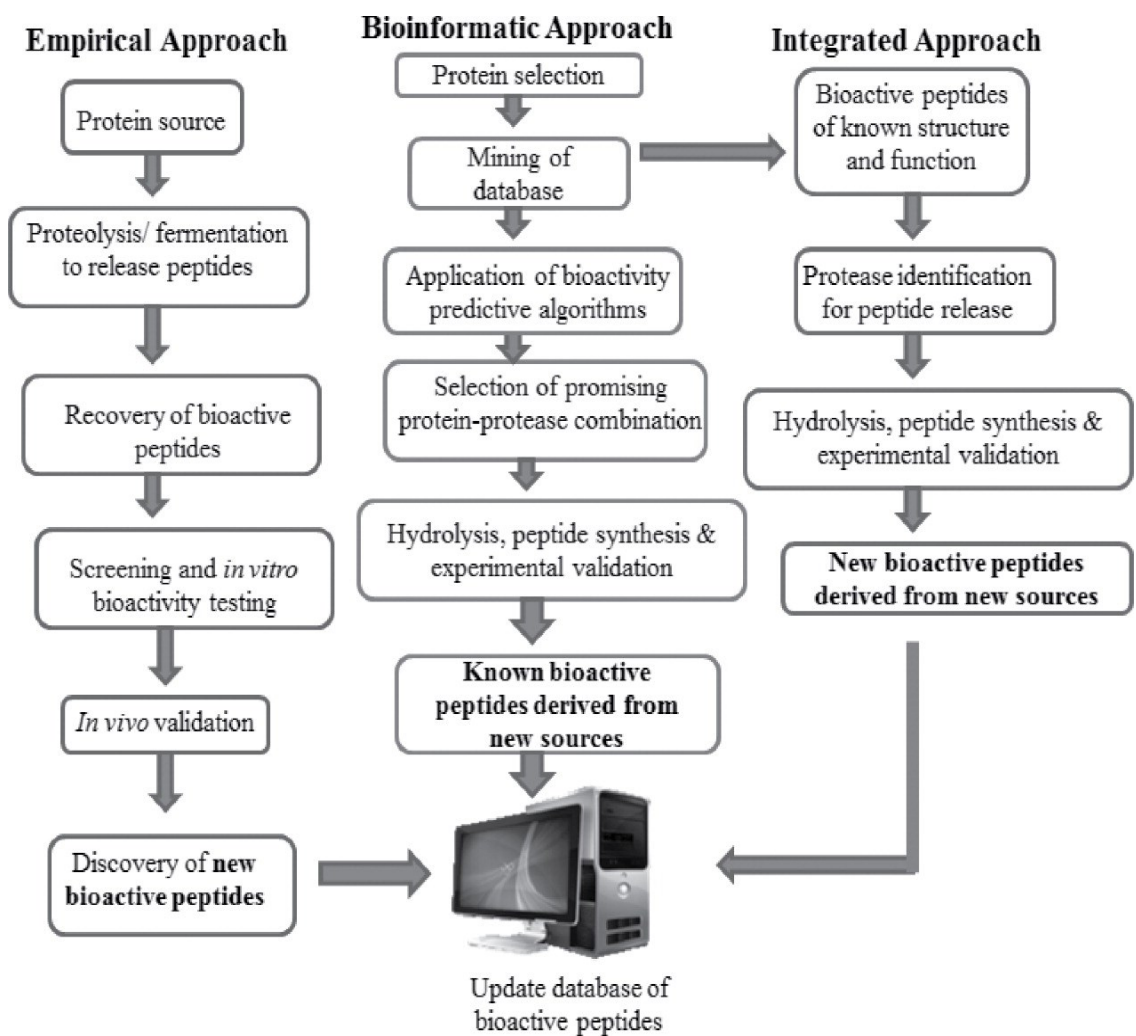


Figure 1. Experimental approaches for the discovery of bioactive peptides. Reprinted with permission from Daliri et al., 2018.

Following the empirical approach, protein extracts are hydrolyzed by enzymatic digestion with plant or microbial proteases, chemical digestion by alkaline or acidic hydrolysis, or bacterial fermentation (Ryder et al., 2016), where the ideal methodology can vary depending on the protein source and subsequent extraction or selection methods. Food-derived protein hydrolysates are commonly prepared using gastric enzymes to generate peptides resembling the products of *in vivo* gastric digestion to provide an indication of the functional properties of whole-food consumption (Korhonen and Pihlanto, 2006). Bioactivity screening tests are then used to assess whether hydrolysates exhibit the desired functionality *in vitro*. Functional hydrolysates may be subject to further fractionation to concentrate the bioactivity and minimize sample complexity prior to peptide identification. The validation of putative bioactive sequences is completed by testing purified synthetic peptides using the same *in vitro* assays at relevant concentrations.

The bioinformatic approach relies on *in silico* predictions of peptide sequences generated by enzyme-mediated hydrolysis and relates them to previously identified functional sequences from *in vitro* measurements using synthetic peptides or sequences predicted to exhibit functionality from computational models. Many online tools are freely available to facilitate *in silico* digestions, but BIOPEP has integrated a tool into their web-platform that directly compares *in silico* predicted peptides to the > 3,700 bioactive peptides stored in the database. The integration of empirical methodology with bioinformatic tools has led to the development of a more robust workflow for identifying bioactive peptides from

novel protein sources (Figure 1). The reliability of the bioinformatic and integrated approaches improves as more bioactive peptides are identified following the empirical approach.

Other integrated approaches have been developed to identify potentially functional sequences, such as quality-structure-activity-relationship (QSAR) methods that aim to describe structural features common to compounds with similar functions. Some research groups have specifically investigated the importance of secondary structures for mediating the specificity of cell signaling pathways (Watt, 2006), where primary sequences that adopt α -helical, β -sheet or β -turn structures have been described as essential for mediating protein-protein interactions (PPI) (Jesus Perez de Vega et al., 2007). PPIs are facilitated by the recognition of short linear motifs (SLiMs) within an unstructured domain of one protein by a conserved protein-binding domain on another protein. Peptide motifs have been described to coordinate the complete protein life cycle (transcription, localization and degradation), such that > 100,000 unique primary motifs are estimated to participate in maintaining cellular homeostasis (Tompa et al., 2014). It has even been demonstrated that these motifs can be exploited to activate non-physiological pathways (Howard et al., 2003), and when it is considered that intracellular peptides regulate numerous cellular functions (Ferro et al., 2004), the mimicry of peptides within protein hydrolysates to known SLiMs could represent an alternative approach for the identification of functional peptides, independent of the common bioactive peptide methodologies.

2.2 Diabetes Mellitus

The global incidence of Diabetes Mellitus (DM) has nearly doubled to 8.5 % since 1980, and is expected to be the seventh leading cause of mortality by 2030 (Mathers and Loncar, 2006). DM is a metabolic disorder characterized by chronic hyperglycemia, resulting from impairments to insulin signaling, insulin secretion, or both (Punthakee et al., 2018). Etiological classifications of DM emphasize the complexity of whole-body carbohydrate metabolism, specifically Type 2 diabetes (T2DM) can be caused by hyperinsulinemia and characterized by peripheral insulin resistance with only a minor deficiency to insulin secretion (Punthakee et al., 2018). T2DM accounts for greater than 90% of all DM and may develop concurrently with MetS, of which T2DM is a significant risk factor. As a major cause of global mortality, the discovery of novel treatments for these diseases has significant public health interest. This section intends to describe the complex metabolic dysregulation of T2DM and the diversity of chemical structures, including those from natural compounds, that elicit anti-diabetic effects and could potentially provide clues towards anti-diabetic bioactive peptide sequences.

2.2.1 Physiology and Pathology of Type 2 Diabetes

Insulin is a peptide hormone released by pancreatic β -islet cells in response to elevations in circulating blood glucose that acts on peripheral tissues to stimulate glucose absorption. Insulin mediates its effects by binding to the insulin receptor (IR) that triggers a cascade of signaling events collectively promoting energy storage by the internalization of glucose. Insulin-signaling increases the rate of

glycogen synthesis and suppresses gluconeogenesis in muscle and liver and stimulates fatty acid biosynthesis and suppresses beta-oxidation in adipose tissue; both processes require the translocation of glucose transporter channels to the cell membranes to increase the cellular capacity to absorb glucose. Downstream signaling from the IR has been extensively reviewed and interventions *via* these signaling pathways are significant opportunities for drug development in the treatment of T2DM (Huang and Czech, 2007; Saltiel and Kahn, 2001; Wilcox, 2005). Insulin resistance is characterized by the attenuation of IR-autophosphorylation upon insulin binding, reducing the rate of peripheral glucose uptake from the bloodstream into the muscle. Sustaining a hyperglycemic state increases the pancreatic demand for insulin secretion, leading to hyperinsulinemia and the progressive loss of β -islet cell function. DM is clinically diagnosed using multiple metabolic biomarkers (Punthakee et al., 2018), and treatments depend on both the progression and pathology of T2DM.

2.2.2 Small Molecular Modulators of Glucose Uptake

Anti-diabetic drugs and their derivatives are broadly classified according to their structures, activities, and functions. Several biomarkers involved in whole-body carbohydrate metabolism have been identified as suitable targets for pharmacological intervention, however the signaling pathways compromised in T2DM have received the most interest (Liu et al., 2010). The major anti-diabetic drug classifications are reviewed below to understand the structures and mechanisms behind the established pharmacological agents used to regulate

these metabolic benefits (Table 1). Alpha-glucosidase inhibitors function by limiting the total amount of glucose available in the gut for absorption in response to a meal. The thiazolidinedione class inhibits peroxisome proliferator-activated receptor-gamma (PPAR γ) causing a suppression in β -oxidation in adipose tissue and improvements to insulin sensitivity in muscle and liver (Tahrani et al., 2011). The biguanide drug Metformin is a naturally-derived compound from French lilac (*Galega officinalis*) that reduces hepatic gluconeogenesis and improves insulin-dependent and insulin-independent glucose uptake in muscle (Wiernsperger and Bailey, 1999). The guanide monomer and its derivatives have also been shown to exhibit therapeutic functions (Bailey and Day, 2004). Sulphonylureases are the most prescribed class of oral anti-diabetic drugs, binding sulphonylurea receptor 1 stimulating calcium influx into β -islet cells and the secretion of insulin. DPP-IV, a cell surface enzyme that cleaves incretins glucagon-like peptide 1 (GLP-1) and glucose-dependent insulinotropic peptide (GIP), aids in the mediation of pancreatic insulin secretion. Inhibitors to DPP-IV, including members of the gliptin family of compounds, are common anti-diabetic compounds that sustain the insulin secretion stimulation activity of incretins.

Peptide analogs to incretin hormones, DPP-IV inhibitors and insulin can be broadly classified as peptidomimetics. Peptidomimetics are peptide-like molecules having similar structures to the endogenous peptide hormones and through amino acid substitutions and/or chemical modifications, affect the biological half-life or receptor binding energies of peptides to cellular binding partners. Insulin mimetics

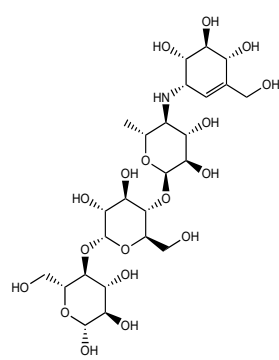
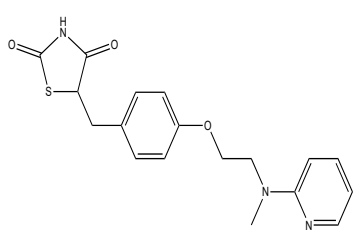
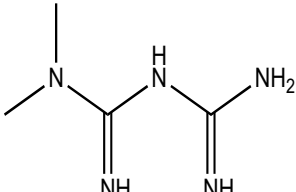
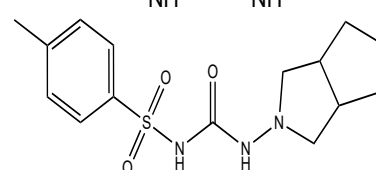
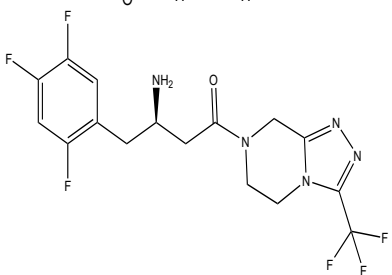
derived from the venoms of cone snails have been shown to retain both critical residues to adopt the necessary secondary structures for interacting with the IR-binding pocket with high specificity (Menting et al., 2016). DPP-IV has broad specificity to peptide ligands, where > 250 of 400 possible dipeptide sequences based on the standard 20 amino acids (20^2) act as inhibitors to DPP-IV with reported IC_{50} concentrations in the range of 0.001 to 1.0 mM, according to BIOPEP. Therefore, the production of novel low molecular weight hydrolysates may also generate novel peptidomimetics of endogenous signaling molecules.

2.2.3 Anti-Diabetic Compounds in Foods and Natural Sources

Various *in vivo* activities such as the reduction of gastric glucose uptake, enhancement of insulin sensitivity, action and secretion, enhancement of incretin action and the alleviation of oxidative stress are broadly classified as anti-diabetic properties, such that many foods have been described as protective against DM based on these characteristics (Lacroix and Li-Chan, 2014). Incretin peptide hormones stimulate insulin secretion and therefore assist in glucose uptake from the blood, and where peptides from numerous food sources have been shown to enhance this effect. However, most reported anti-diabetic foods initially identified as traditional medicines contain flavonoids, polyphenols, chlorogenic acids and/or fibers, hence plant extracts are commonly investigated to identify the bioactive compounds mediating these functions. Gooseberry, fenugreek, green tea, bitter melon, turmeric and cinnamon are foods and spices whose apparent bioactivities have been suggested, but lack the clinical proof to support their potential

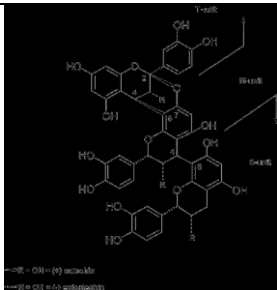
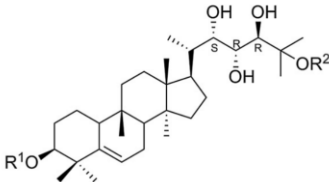
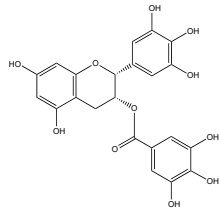
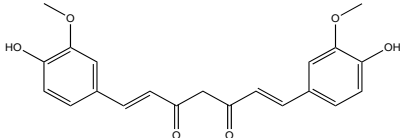
pharmacological applications (Deng, 2012; Marles and Farnsworth, 1995), and summarized in Table 2.

Table 1. Common anti-diabetic pharmaceutical drugs.

Drug Class	Drug names	Cellular Function	Structure ^a
α -glucosidase Inhibitor	Acarbose	Inhibitor of intestinal glucose absorption	
Thiazolidinediones	Rosiglitazone	Inhibitor of PPAR γ	
Biguanide	Metformin	Suppression of hepatic glucose production	
Sulfonylurease	Gliclazide	Stimulate intracellular Ca ²⁺ release	
Gliptin	Sitagliptin	Inhibitor of DPP-IV	

^aAll structures retrieved from ChemACX database

Table 2. Foods identified as hosts to anti-diabetic compounds and their characteristics. Data summarized from Lacroix and Li-Chan, 2014.

Food	Chemical Class	Putative anti-diabetic functions	Structure	Reference
Cinnamon ^a	Doubly linked procyanidin A-type catechin/epicatechin oligomers	Enhance insulin action & sensitivity		Anderson et al., 2004
Bitter Melon ^b	Cucurbitane triterpenoid	Enhance insulin secretion		Tan et al., 2008
Green Tea	Epigallocatechin gallate	Enhance insulin action & inhibition of intestinal glucose transport		Kobayashi et al., 2000
Turmeric	Curcumin	Reduction of hepatic glucose production		Ghorbani et al., 2014

^aReprinted with permission from Anderson et al., 2004. Isolation and characterization of polyphenol type-A polymers from cinnamon with insulin-like biological activity. *Journal of Agricultural and Food Chemistry*, 52(1), 65-70, Copyright 2019 American Chemical Society; ^bReprinted from Chemistry & Biology, 15, Tan et al., Antidiabetic activities of triterpenoids isolated from bitter melon associated with activation of the AMPK pathway, 263-273, 2008, with permission from Elsevier.

Metabolomic analyses of obese, insulin-resistant subjects versus lean, insulin-sensitive subjects revealed that numerous protein-derived compounds were among circulating metabolites that could effectively distinguish subject groups and directly related to insulin sensitivity, including the branched chain amino acids (BCAA), aromatic amino acids, Glu and Gln, Ala and acylcarnitines (Newgard et al., 2009). Importantly, BCAA-activation of mechanistic target of rapamycin 1 (mTOR1) inhibits insulin signaling by its phosphorylation of insulin receptor substrate 1 (IRS-1) that is demonstrated to improve certain metabolic parameters associated with T2DM (Yoon, 2016). There is no consensus on the signaling role of BCAAs due to the numerous *in vitro* and *in vivo* models that collectively suggest a multifunctional role. According to Tremblay and Marette (2001), micromolar concentrations of the free amino acids Leu, His, Met, Cys, Thr, and Tyr can inhibit insulin-dependent glucose uptake in cultured L6 myotubes when screened using cell culture media free of amino acids, but were less potent than the stimulating effects of SPF (Chevrier et al., 2015) and its subfractions (Roblet et al., 2016) screened in the same cell line. The amino acid content of cell culture media used to screen SPF and its fractions in these studies are orders of magnitude larger than their evaluated concentrations, indicating that the influence of free amino acids in these samples may not substantially vary from controls.

BCAA have also been implicated as components of glucose uptake modulating dipeptides due to their apparent insulin-independent stimulating activity in L6 myotubes at 1 mM (Morifuji et al., 2009). The hydrophobic character of these

residues has been suggested as an essential element for mediating active site interactions with many enzymes (de Castro and Sato, 2015), but other peptides found to modulate glucose uptake *in vitro* and *in vivo* have more diverse compositions: Trp-His stimulated insulin-independent uptake to a greater extent than insulin-stimulation alone in L6 myotubes at 300 μM (Soga et al., 2014); Gly-Glu-Tyr and Gly-Tyr-Gly from silk fibroin stimulated insulin-independent and insulin-dependent uptake in 3T3-L1 adipocytes at 250 μM (Han et al., 2016); Ile-Ala-Val-Pro-Glu-Gly-Val-Ala, Ile-Ala-Val-Pro-Thr-Gly-Val-Ala and Leu-Pro-Tyr-Pro activated both AMPK and Akt at 500 μM in HepG2 cells (Lammi et al., 2015); Diapin (Gly-Gly-Leu) has a blood-glucose lowering-effect following oral administration in mice when administered at 2 mg/g body weight by stimulating GLP-1 secretion from L-cells (Zhang et al., 2013); Asp-Ile-Tyr-Glu-Thr derived from the insulin receptor (IR) alpha-loop mediates IR-autophosphorylation by competing with native IR-alpha-loop for binding to the IR-catalytic domain at concentrations as low as 40 μM , reducing the rate of insulin-independent activation (Kurian et al., 2014). Peptide-based modulators of glucose uptake have diverse compositions and interact with various mechanisms over broad concentrations. Therefore, identification of novel bioactive peptides from the SPF must consider these variable structure/function relationships when considering the bioactivity of SPF and its fractions. Specifically, anti-diabetic peptides are most commonly classified as DPP-IV inhibitors, but the stimulating effects of SPF in cultured L6 myotubes cannot result from DPP-IV inhibition due to the absence of pancreatic GLP-1 and GIP and their inability to stimulate insulin secretion by this model.

2.3 Marine Muscle Proteins

Marine teleost fishes, such as Atlantic salmon, contain a higher proportion of myofibrillar and sarcoplasmic proteins than terrestrial animals, and BCAA make up a larger component of their sequences (Comerford and Pasin, 2016). The contractile unit of skeletal muscle function identically in both terrestrial and marine species, where all contractile elements (muscle fibers) are surrounded by a membrane (sarcolemma) and bathed in cytoplasm (sarcoplasm) containing water soluble proteins (Hultin, 1984). The unique organization of muscle fibers in marine teleosts results from the presence of myotomes that are organized perpendicular to the muscle fibers and separated from adjacent myotomes by a connective tissue membrane called the myosepta. The proteins composing the contractile units of skeletal muscles from marine and terrestrial animals are identical, but the sequences of these proteins still vary considerably (Chaijan et al., 2007). Marine teleosts have lower proportions of connective tissues, and therefore have relatively higher proportions of myofibrillar and sarcoplasmic proteins in the muscle compared to terrestrial mammals. Therefore, unique characteristics of *Salmo salar* skeletal muscle proteins may be important factors when considering the selection of proteins from deboned fish mince for identifying bioactive peptides in the SPF.

2.3.1 Effect of pH and Ionic Strength on Protein Solubility

Protein solubility occurs when protein-water interactions exceed the forces promoting protein-protein interactions; in dilute solutions, protein precipitation occurs when this equilibrium is reversed. Protein solubility is affected by

adjustments to pH and/or the ionic strength (IS) of a solution, and these characteristics can be utilized for the selective recovery of proteins based on their surface charge or isoelectric point (pI) (Matak et al., 2015). The process of selectively recovering proteins at their pI is known as isoelectric precipitation (IEP), and when combined with a protein solubilization step also using pH adjustments, is termed isoelectric solubilization/precipitation (ISP).

Solubility can be accomplished at both high (alkaline) and low pH (acid), where both function by elevating the protein net surface charge, promoting their electrostatic interaction with water (Chen and Jaczynski, 2007). Acidification leads to protonation of protein side-chains when the solution pH is lower than the pK_a of the side-chains, causing an increase to its positive surface charges. Likewise, an increase in pH causes the deprotonation of side chain when the solution pH exceeds the pK_a of side-chains, increasing the negative surface charge of proteins. Proteins achieve minimum solubility (precipitation) when the net surface charge is zero, causing them to self-aggregate due to van der Waals attraction and hydrophobic interactions. The major protein classes in marine muscle are easily recoverable based on their difference in solubility, indicating that ISP and IEP represent the first opportunity for introducing selectivity into the methodology for recovering marine proteins to produce bioactive peptides.

Due to the ionic nature of interactions mediating protein solubility, the strengths of these interactions are amenable to IS adjustments. The IS can be modified through

the addition or removal of ionic compounds. Salts interact with ionic amino acid residues, modifying their net surface charge and the strength of their interaction with water. Protein solubility curves are useful tools to identify the optimal conditions to extract protein by ISP methods (Chen and Jaczynski, 2007), where protein from different sources may show preference to each method. Ionic strength adjustments alone are sufficient to fractionate protein isolates by class (Tahergorabi et al., 2012).

2.3.2 Myofibrillar Proteins

Myofibrillar proteins represent the functional contractile units of muscle tissues, represented predominantly by α -actin, the heavy and light myosin chains, troponin and tropomyosin. These proteins comprise approximately 65 – 80 % of total striated muscle and are soluble at both high (0.6) and very low (0.0002) IS (Matak et al., 2015; Stefansson and Hultin, 1994). Fish myofibrillar proteins are considered to have a pI of 5.5. Marine myofibrillar proteins also exhibit high functionality *in vitro*, providing water- and oil-holding, foaming, emulsion, and gelling capacities (Halim et al., 2016); the surimi industry exists entirely due to the physical properties of marine myofibrillar proteins. These proteins are of great interest because they are large in both size and abundance and therefore contribute disproportionately to protein hydrolysates prepared from marine muscle. Myofibrillar protein hydrolysates have successfully produced bioactive peptides with ACE-inhibitory (Ghassem et al., 2011), antimicrobial (Capriotti et al., 2015), and antioxidative activities (Saiga et al., 2003).

2.3.3 Sarcoplasmic Proteins

Sarcoplasmic proteins are found in the intracellular fluid between myofibrils, within the muscle fibers (Matak et al., 2015). These proteins facilitate glucose catabolism and include all enzymes involved in glycolysis, as well as oxygen carriers such as myoglobin and hemoglobin. Sarcoplasmic proteins represent approximately 18 – 20 % of total muscle protein and display almost no *ex vivo* functionality. These proteins are exclusively water-soluble and can be extracted by the physical compression of striated marine muscle. Sarcoplasmic protein hydrolysates have also successfully produced bioactive peptides with antimicrobial (Capriotti et al., 2015) and antioxidative activities (Najafian and Babji, 2014).

2.3.4 Stromal Proteins

Stromal proteins compose the connective tissues of teleost fish that maintain muscle structure and represent approximately 3 – 5 % of total muscle protein. Collagen and elastin represent the single most abundant proteins in this category, and function to maintain the structural integrity of tissues. These proteins are completely insoluble in water, acid or alkaline, except under high heat or high salt concentrations (Tahergorabi et al., 2012). Stromal proteins exist at far lower proportions in marine sources compared to terrestrial sources. Fish skins are the largest source of these proteins and are frequently investigated for sources of fish gelatin as well as bioactive peptides (Gu et al., 2011). The stromal protein content of fish processing by-products is often greater than skeletal muscle alone (Wasswa et al., 2007), and represents an appealing target for value-added processing.

2.4 Marine Teleost Protein Hydrolysates as a Source of Bioactive Peptides

The physiological benefits of seafood consumption can originate from the effects of omega-3 polyunsaturated fatty acids that reduce risk for cardiovascular disease (CVD), but more recently have been shown to be host to many non-lipid anti-diabetic and anti-obesity compounds (Hu et al., 2016). Marine proteins have been identified as suitable candidates for producing bioactive peptides (Giri and Ohshima, 2012; Pangestuti and Kim, 2017), and their enzymatic hydrolysis has been demonstrated to increase their biological potencies (Nongonierma and FitzGerald, 2017). A comparison of hydrolyzed seafood proteins from various sources for anti-diabetic activity highlighted *Salmo salar* as a potential source of bioactive peptides (Pilon et al., 2011). A series of functional protein hydrolysates was developed under various conditions (Table 3), but the measured glucose uptake stimulation in cultured L6 myotubes was greatest following protocol 2 (SPF), which involved protein solubilization in 1.0 M NaOH, followed by isoelectric precipitation at pH 4.5, then enzymatic digestion with pepsin followed by trypsin/chymotrypsin. In these experiments, the active fraction was recovered as ultrafiltration permeate through a 1000 Da membrane (Jin, 2012).

Table 3. Hydrolysates from *Salmo salar* muscle protein investigated for glucose uptake stimulation (Jin, 2012).

Protocol	Dissolution Solution	Isoelectric Precipitation	Enzymatic Digestion	MWCO ^a (Da)
1	0.1 M NaOH	-	Pancreatin or Pepsin	1000
2	1.0 M NaOH	pH 4.5	Pepsin, Trypsin and Chymotrypsin	1000
3	dH ₂ O	-	Pepsin, Trypsin and Chymotrypsin	1000

^aMolecular weight cutoff

Other research groups have produced functional hydrolysates from Atlantic salmon gelatin (Neves, Harnedy, O’Keeffe, Alashi, et al., 2017), processing by-products (Neves, Harnedy, O’Keeffe, and FitzGerald, 2017; Vik et al., 2015), and skin (Gu et al., 2011) that have identified peptides with antihypertensive, antioxidant, DPP-IV inhibition activities. The SPF produced by protocol 2 has also been shown to exhibit antioxidative activity (Girgih et al., 2013), antihypertensive activity (ACE-inhibition) and renin-inhibition (Girgih et al., 2016), and protection against the obesity-linked features of MetS (Chevrier et al., 2015). The fractionation of SPF by preparative RP-HPLC generated subfractions that inhibited the rate of glucose uptake in cultured L6 myotubes compared to the unfractionated hydrolysate (Jin, 2012). Additional separations of the SPF by electro dialysis with filtration membranes (EDFM) demonstrated that anionic, cationic and/or neutral peptide fractions may enhance insulin-independent or insulin-dependent glucose uptake (Henaux et al., 2019; Roblet et al., 2016). Both studies suggested the

possibility that unfractionated SPF contains peptides with both stimulating and inhibiting functions.

2.5 Methodology for the Purification and Identification of Low Molecular Weight Peptides

The peptide composition and functionality of the SPF are poorly understood, partially because of the challenges posed by the separation and detection of small peptides (< 7 amino acids, AA) (Panchaud et al., 2012). Peptide separations generate compositionally distinct subfractions and provide the opportunity to fractionate the functional peptides from non-functional peptides. A single method is often insufficient to completely resolve all functional components from low molecular weight hydrolysates, because free amino acids, di- and tripeptides and larger oligopeptides are collectively poorly resolved by a single chromatographic method. Multidimensional or orthogonal fractionation approaches for peptide separation present the greatest opportunity to resolve complex mixtures (Liu et al., 2002), and the methods to facilitate this are discussed below.

2.5.1 Chromatographic Fractionation of LMW Peptides

The technological platform for identifying medium and large peptides (> 7 AA) is established because of the generation of these peptides in biomarker identification (Panchaud et al., 2012). Small, or LMW peptides (< 7 AA) are difficult to separate and identify following the biomarker methodology because they are often highly hydrophilic, poorly retained on reversed-phase columns, and incompatible with database searching algorithms. Furthermore, high-resolution fractionation of LMW

protein hydrolysates is challenged by the broad molecular characteristics of peptides in these samples. The strong relationship between peptide size and biological potency highlights the importance of effective fractionation strategies targeted at LMW bioactive peptides (Saavedra et al., 2013).

Separations by liquid chromatography (LC) are capable of refining complex samples with high resolution, and therefore represent an essential tool for studying LMW protein hydrolysates. The fractionation of bioactive peptides by low-pressure or fast-protein liquid chromatography (LPLC/FPLC), high-performance liquid chromatography (HPLC) and ultra-performance liquid chromatography (UPLC) have each been applied to the empirical method of identifying bioactive peptides, but do not exhibit equivalent separation efficiency for LMW peptide samples. Most LC techniques use packed columns of porous particles under pressure to facilitate sample partitioning. LPLC methods use columns with particle sizes $> 5.0 \mu\text{m}$ and operate at low pressures ($< 3 \text{ MPa}$), conventional HPLC uses $2.5 - 5.0 \mu\text{m}$ particles operating under 40 MPa , and UPLC methods use sub- $2 \mu\text{m}$ particles at higher pressures ($> 60 \text{ MPa}$), where $1 \text{ MPa} = 10 \text{ bar}$. These parameters, together with the column dimensions and buffer composition, establish the chromatographic separation efficiency, retention, and selectivity for all separations.

When studying protein hydrolysates, HPLC coupled with ultraviolet (UV) detection is useful for evaluating the compositions of peptide fractions and to assess sample purity, but peptide identification requires detection by mass spectrometry

(MS). UPLC and HPLC formats provide sufficient resolution to pre-fractionate hydrolysates when coupled to MS detection but LPLC operates at higher solvent flow rates and pressures incompatible with MS. UV-detection coupled to LPLC with preparative columns can tolerate higher sample concentrations and larger injection volumes that make these formats ideal for fractionation steps. All chromatographic fractions must be compatible with *in vitro* screening tests such that they are free from buffer and solute residues (volatile buffers) and can be accurately quantified for normalization. Colourimetric methods are often used to measure the 'protein' concentration of protein hydrolysates solutions, but due to the presence of free amino acids, di- and tripeptides, and larger oligopeptides that do not lead to dye-complex formation with equivalent kinetics as polypeptides or proteins, these methods may not be as reliable as measuring total nitrogen.

2.5.1.1 Analytical Stationary Phase Chemistries

The suitability of stationary phase chemistries for the purification and identification of bioactive peptides has been extensively reviewed (see Table 3 in de Castro & Sato, 2015). The reversed-phase (RP) functionalization is a widely adopted chromatographic separation technique to mediate the separation of analytes on the basis of their hydrophobicity. RP is suitable for most biological samples because their predominantly aqueous nature enables their high retention on RP columns. RP columns commonly use octadecylsilane (ODS, C₁₈) functional groups, whereas hydrophobic interaction chromatography (HIC) use aromatic or less hydrophobic aliphatic moieties n-octyl (C₈) or n-butyl (C₄) ligands to improve

the kinetic efficiency and retention of proteins (Fekete et al., 2012). LMW protein hydrolysates are poorly separated on RP columns due of the poor retention of short peptides and their tendency to co-elute.

Hydrophilic-lipophilic interaction (HILIC) or normal-phase stationary phase chemistries are targeted toward polar analytes that are not retained well on RP columns and have been adopted in mass spectrometry-based proteomics and metabolomics analyses, without the need for derivatization (Schlichtherle-Cerny et al., 2003). A comparison of RP and HILIC separation of dipeptides found that all dipeptides eluted before the gradient began on a RP column, but were well retained and easily resolved using a HILIC column (Tang et al., 2014). Separations using ion exchange (IEX), gel filtration (GF), hydrophobic interaction (HIC), and chromatofocusing (CF) formats do not achieve comparable resolution to RP or HILIC chemistries at pressures corresponding to HPLC or UHPLC, making them less suitable for coupling to MS (Conlon, 2007).

2.5.1.2 Preparative Stationary Phase Chemistries

Ion exchange chromatography (IEX) is mediated by a reversible interaction of an ion with the charged stationary phase, with selectivity for negatively charged ions (anion exchange) or positively charged ions (cation exchange) and optimized over a broad (strong) or narrow (weak) pH range. IEX separations are typically applied as a pre-fractionation step prior to a more selective separation method and suitable for use with both proteins and peptides (GE Healthcare, 2007). Retention on the

column is mediated by the net surface charge of ions and can be modified by changing the ionic strength and/or pH of the eluant. When using non-volatile elution buffers, fractions require additional processing steps to eliminate the mineral content that may produce interferences during the *in vitro* assays. However, volatile elution buffers have been used for the IEX fractionation of bioactive peptides from shrimp for subsequent *in vitro* bioactivity testing (Ma et al., 2006; Mekata et al., 2017).

Gel filtration chromatography (GF) is a separation mode capable of resolving analytes based on molecular weight (MW) and shape. The MW range that can be resolved is nominally defined by the pore dimensions on stationary phase particles. Fractionation by GF is facilitated by the selective inclusion or exclusion of peptides into these pores, whereby the inclusion of a peptide into particle pores effectively increases its path length to the detector and therefore elution time from the column. Over the duration of the separation, each analyte has a unique path length that is approximately proportional to its log MW. GF is widely used to characterize MW profiles of protein and peptide samples, to fractionate protein hydrolysates, and to desalt purified proteins (Chalamaiah et al., 2013; Silvestre, 1997; Van Der Ven et al., 2001). GF has also been performed with volatile buffers (facilitating their removal following fractionation) to identify peptides (Rodríguez et al., 2012). IEX and GF stationary phase chemistries do not yield good resolution on HPLC yet are used frequently by LPLC to separate bioactive peptides (GE Healthcare, 2007). The lower resolution by these methods can be overcome using a multidimensional

or consecutive chromatography approach, that couple complementary methods in series to offer the orthogonality required for complex samples (Rodriguez, et al. 2012).

2.5.2 Non-Chromatographic Fractionation of Proteins and Peptides

Chromatography is an excellent tool for analytical and preparative separations of protein hydrolysates but to focus exclusively on these methods can limit the efficacy of the fractionation. Ultrafiltration and nanofiltration are commonly used to fractionate bioactive peptides (Picot et al., 2010), and are well suited for pilot- or industrial-scale applications. Non-chromatographic fractionation can be applied to both proteins and peptides produced from protein hydrolysates. Strategies that focus on the biological requirement for peptides to cross gastrointestinal membranes have been developed, separating molecules that have the ability to internalize from those that don't using Caco-2 cell monolayer cultures (derived from cells of the small intestinal membrane) to evaluate, identify and validate candidate bioactive peptides (Satake et al., 2002; Stevenson et al., 1999). Fractionation of protein isolates are common performed to minimize the complexity of proteome analysis (Righetti and Boschetti, 2013).

Electrophoretic methods are common approaches to fractionate protein mixtures but are less commonly used to fractionate peptide mixtures because of their poor electrophoretic mobility. Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) increases the dynamic range by which to study

protein samples by expanding the proteome analysis to multiple samples (Shevchenko et al., 2007). Ramos et al. (2008) evaluated a two-dimensional electrophoretic approach for the characterization of low abundance proteins from whole-cell lysates using both protein and peptide separations. Proteins were first separated using SDS-PAGE, then defined segments were trypsinized in-gel, followed by the separation of each digest onto SDS-free gels, resulting in a 2.5-fold greater number of proteins observed than by the first dimension alone. Mojica and de Meija (2015) performed an in-gel digestion on proteins from bean cultivars and used bioactive peptide databases to identify potential bioactive peptides, indicating that prefractionation of protein isolates can be a valuable step to identify bioactive peptides following a bioinformatic approach. Electrophoretic peptide separations, however, are impractical when chromatographic methods provide the opportunity to fractionate peptide mixtures with greater resolution and selectivity. However, EDFM has been demonstrated to concentrate bioactive peptides from the SPF, as well as flaxseed and soybeans hydrolysates, amongst others (Doyen et al., 2014; Henaux et al., 2019; Roblet et al., 2016, 2014). The electrophoretic applications for protein and peptide separations have been extensively reviewed (Righetti et al., 2013) and are summarized in Table 4.

Table 4. Non-chromatographic methods of protein and peptide fractionation in proteomic and bioactive peptide research.

Method	Protein/Peptide	Basis of Separation	Reference
SDS-PAGE with in-gel Digestion	Protein	Electrophoretic Mobility	Mojica and de Mejía, 2015
SDS-free PAGE	Peptide	Electrophoretic Mobility	Ramos et al., 2008
IEF	Protein or Peptide	Electrophoretic Mobility	Pergande and Cologna, 2017
Rotofor Cells	Protein or Peptide	Electrophoretic Mobility	Xiao et al., 2004
OFF-Gel Electrophoresis	Protein	Electrophoretic Mobility	Chenau et al., 2008
Capillary Electrophoresis	Protein or Peptide	Electrophoretic Mobility	Ibáñez et al., 2013
EDFM	Protein or Peptide	Electrophoretic Mobility	Roblet et al., 2016
Caco-2 Cell Monolayers	Peptide	Membrane Permeation	Stevenson, 1999
Ionic Strength	Protein	Solubility	Capriotti et al., 2015
Ultra-/Nanofiltration	Protein or Peptide	Molecular Size	Picot et al., 2010

2.5.3 Protein and Peptide Detection and Identification

The high mass accuracy and sensitivity of mass spectrometry detection has led to its universal adoption for high-throughput protein identification methodologies. In lieu of these technological developments, the gold standard for peptide identification still involves the use of a synthetic purified peptide standard to confirm the mass and retention time of experimentally determined peptide sequences. The use of liquid chromatography (LC) coupled to electrospray ionization (ESI) and tandem mass (MS/MS) spectrometry detection (LC-MS) represents the most widely adopted technological platform for protein and peptide analysis because prior knowledge of sample compositions is not required. Protein identification combines mass spectrometry detection with software-assisted database searching protocols (Figure 2): proteins are first hydrolyzed using trypsin, then are resolved by LC and ionized by ESI as they elute from the column to transition peptides from the liquid to gas phase and impart a positive charge (Figure 2A). Upon entry into the mass spectrometer, ionized peptides known as precursor ions, are captured in the first stage of mass detection (MS¹). Various chromatograms (detection intensity vs. time) can be generated to visualize the composition of measured samples: the total ion current chromatogram (TICC) represents the sum intensity of all precursor ions detected at any given retention time, whereas the base peak chromatogram (BPC) represents only the precursor with the highest intensity of detection at any given retention time, and the extracted ion chromatogram (EIC) represents the detector intensity of ions within a selected *m/z* range at any given retention time (Murray et al., 2013).

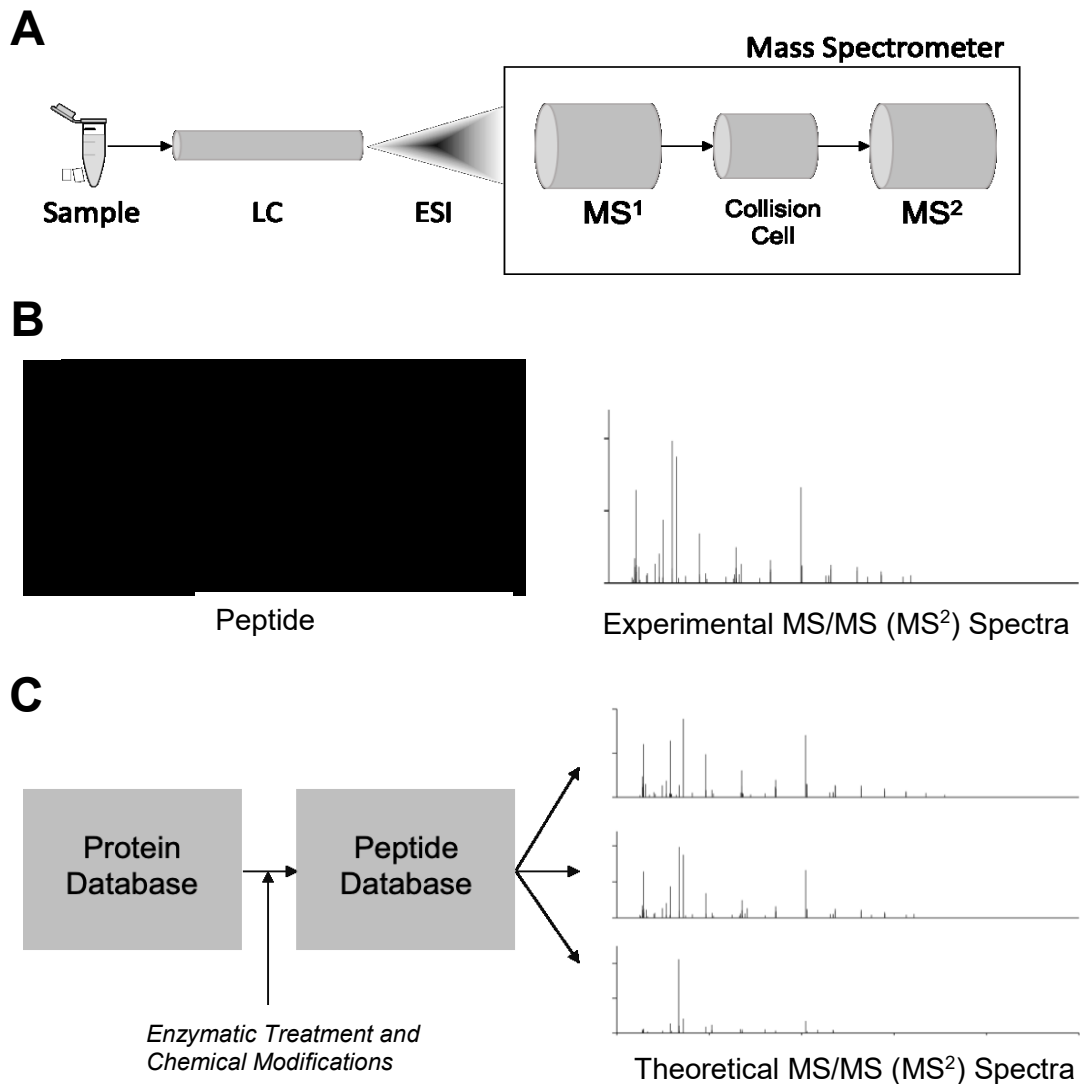


Figure 2. Peptide identification by tandem mass spectrometry combined with database searching. (A) LC-ESI-MS/MS schematic typical for proteomic analysis. (B) The fragmentation of each peptide generates a spectrum characteristic of its primary sequence. (C) Database searching compares experimental spectra to theoretical spectra then scores and ranks peptides matching these masses.

Each detected ion is individually fragmented in a collision cell and the fragments known as product ions are then captured in a second stage of mass detection (MS^2). The detected product ions from each precursor are represented using the MS/MS spectrum, where product ion masses are directly related to the masses of the individual amino acids and their derivatives that compose the captured precursor ion (Figure 2B). Product ions are classified as a-, b-, or c-series ions if they contain the original N-terminal from the precursor ion, or x-, y-, and z-series ions if they contain the original C-terminal from the precursor ion, and depending which bond is cleaved; b- and y-series ions are most common, produced from the fragmentation of the peptide bond. Following data acquisition by LC-MS, the masses of precursor ions and their associated product ions are recorded, and software-assisted database searching is required to assign sequences to process the data (Figure 2C).

Detection of protein and peptides by software-assisted database searching relies on knowledge of which species the protein or peptide samples were derived from. These processes use primary sequences available in online databases such as PubMed or UniProtKB and replicate enzymatic treatments and/or known chemical modifications to protein samples during experimentation or in preparation for LC-MS analysis. These tools are essential to ensure accurate mass prediction of theoretical peptides sequences for comparison to experimental data. Samples containing unknown proteins are commonly hydrolyzed with trypsin prior to LC-MS, but protein hydrolysates are often prepared with more complex enzymatic

treatments making the prediction of sequences more challenging (Panchaud et al., 2012). Additionally, the occurrence of oxidation, deamidation, alkylation, and/or phosphorylation modifications change the masses of theoretical peptides and therefore must be considered using software tools. For each predicted peptide sequence, a theoretical precursor ion mass is calculated, and a theoretical MS/MS spectrum of its product ions generated. The masses of theoretical precursor ions are compared directly to the masses of experimental precursor ions to identify matches within acceptable thresholds to select for further comparison using the MS/MS spectra (Figure 2C). Each match representing a putative identification is scored based on the similarity of the experimental spectrum to the theoretical spectrum. Finally, thresholds of these scores are set to validate putative identifications, minimize the occurrence of false positives, and increase the quality (accuracy with respect to peptide identification) of the results (Panchaud et al., 2012).

In the case of SPF enzymatic processing, pepsin hydrolysis occurs at N- and C-terminal positions of Phe, Tyr, Trp and Leu; trypsin hydrolysis occurs at C-terminal positions of Arg and Lys, and chymotrypsin hydrolysis occurs at C-terminal positions of Phe, Tyr and Trp (Keil, 1992). Numerous *in silico* (computer-simulated) digestion tools are freely available online, including within the BIOPEP bioactive peptide database (Darewicz et al., 2016), and are commonly used to screen known protein sources as hosts to bioactive peptides or to predict the functionality of established hydrolysates based on their identified peptide compositions.

The analysis of samples concentrated in LMW peptides by LC-MS is associated with numerous limitations. Advancements in modern MS have addressed some of the challenges associated with peptide separation and detection, but the non-standard fragmentation patterns of short peptides (Panchaud et al., 2012) and biases for certain residues (Gehrig et al., 2004; Harrison, 2001) still persist that challenge the software-assisted interpretation of product ion spectra. The presence of isobaric (identical mass) amino acids, the co-elution of peptides, and the high frequency of short (< 5 AA) primary sequences within multiple proteins present from entire species proteomes collectively contribute to increase the likelihood of reporting false positives in samples concentrated in LMW peptides (Lahrichi et al., 2013). As well, LMW hydrolysates often contain greater numbers of total peptides per sample analysis and are focused to a mass range where background contamination is simultaneously detected, reducing the quality of identification.

Various methods focused on improving LMW peptide identification are under investigation (Le Maux et al., 2015). The methodology to identify peptide sequences by interpretation of the MS/MS spectra independent of database searching is termed *de novo* sequencing, and is supported by high-throughput tools for comparison to high-quality MS/MS spectral databases such as the Metlin database (Guijas et al., 2018), or manually based on identifying masses common to each of the 20 standard amino acids. A limitation preventing accurate interpretations of MS/MS spectra of LMW peptides is the finding that they do not

reliably generate the same b- and y-series ions produced by the fragmentation of larger oligopeptides (Figure 3): fragmentation of LMW peptides is also associated with the formation of [a₂] ions, [immonium] ions, along with other internal peptide fragments (Palzs and Suhal, 2005) that add to the complexity of *de novo* peptide identifications. Each detected product ion contributes information that are collectively used to indicate the composing peptide residues and their positionings. For instance, immonium ions are commonly generated from the N-terminal residue but are generated less frequently as a result of secondary fragmentation of either the b₂ or y₂ ions. Using these guidelines and with assistance from bioinformatic tools, the manual interpretation of MS/MS spectra is always permissible to identify a precursor ion sequence and represents a reasonable approach for LMW peptide identification in samples containing few peptides due to its low-throughput nature. With the consistent development of high resolution peptide spectral libraries and databases, peptide sequences can be predicted using MS/MS spectra exclusively (Shao and Lam, 2017). Methodologies that combined database searching and *de novo* sequencing approaches have also been established (Wang and Wilson, 2013), and provide additional validation in the reported sequences by either method.

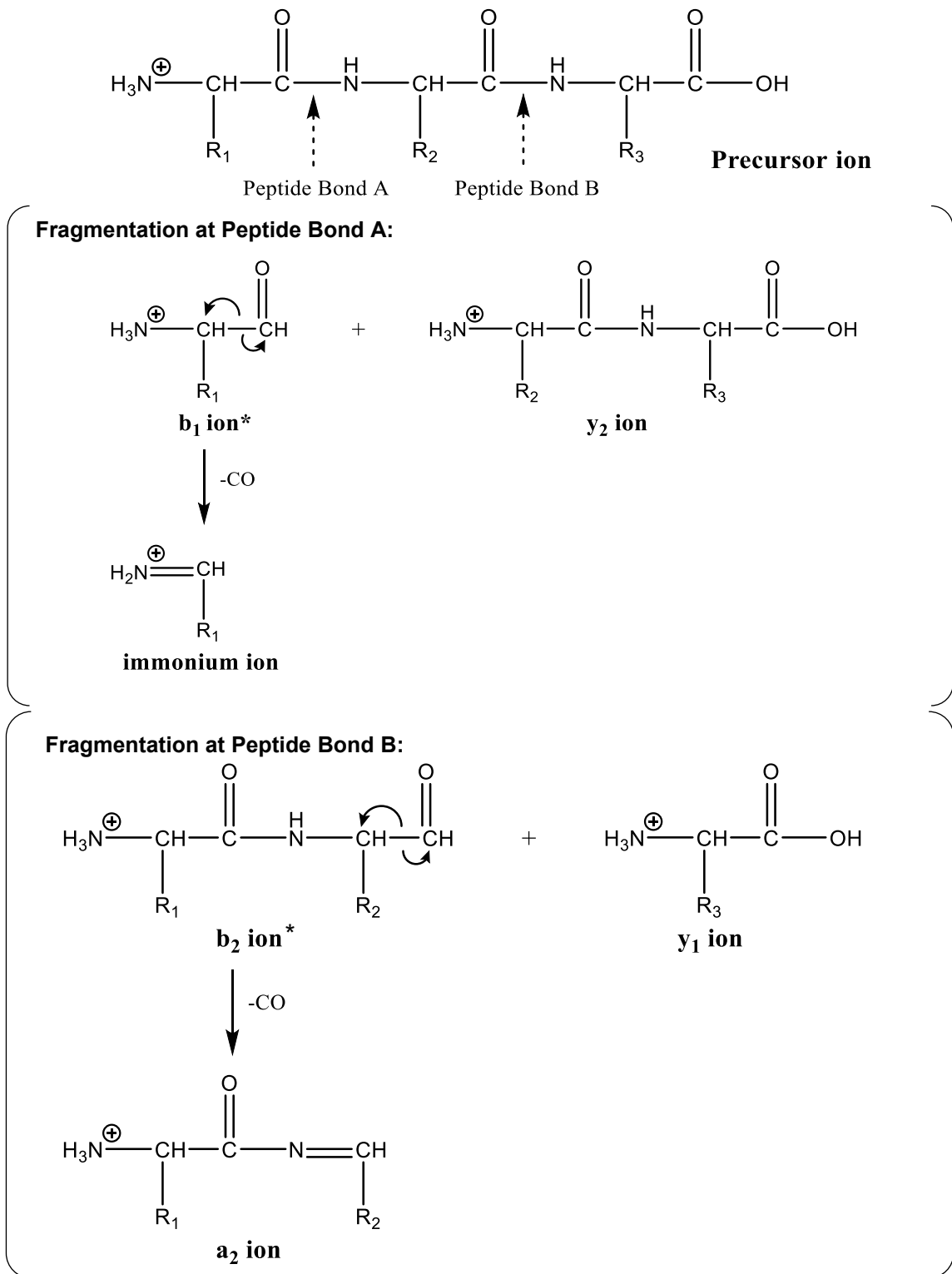


Figure 3. Simulated fragmentation of a tripeptide precursor ion and the product ions used for its identification by *de novo* sequencing. * ions susceptible to further fragmentation.

CHAPTER 3: MATERIALS AND METHODS

3.1 Preparation of SPF

The SPF was prepared following protocol 2, as previously described in Jin (2012), with modifications. Briefly, mechanically deboned salmon mince was homogenized with 1.0 M NaOH at a 1:4 (w/v) ratio in a blender on high speed for 2 min, then stirred for 2 h. The pH was then adjusted to 4.5 using 2.0 M HCl to isoelectrically precipitate the protein. The isoelectric precipitate was kept at 4 °C overnight to make sure the precipitation was complete. The precipitated proteins were pelleted by centrifugation at 5200 x g for 20 min at 4 °C in a Sorvall RC-3 refrigerated centrifuge (Sorvall Instruments Div., Dupont Co., Newtown, CT, USA). The supernatant was discarded, and the pellet was resuspended in the same volume H₂O as NaOH during homogenization with vigorous shaking. The pH of this protein dispersion was adjusted to 2.0 using 2.0 M HCl in preparation for pepsin digestion. Pepsin (EC 3.4.23.1, Millipore Sigma Cat. No.: P7125, 600-1800 units/mg protein) was added at an enzyme:substrate (E:S) ratio of 1:100, assuming the protein content in salmon muscle was 15 % of the wet weight, and the average pepsin activity was 1000 units/mg protein. Enzyme-substrate mixtures were continuously stirred overnight at 37 °C. The pH of the pepsin digest was then adjusted to 7.8 using 2.0 M NaOH to irreversibly deactivate the pepsin, and to prepare for trypsin and chymotrypsin digestion. Trypsin and chymotrypsin (Enzeco Trypsin:Chymotrypsin® 1:1, trypsin: 1000 units/mg protein, chymotrypsin: 1000 units/mg protein, Enzyme Development Corporation, NY) were also added at an E:S ratio of 1:100 and stirred continuously for 4 h at 37 °C. Reactions were

terminated by heating to 100 °C for 10 min. The digests were centrifuged at 5200 x *g* for 30 min at 4 °C and the supernatant was filtered using Whatman #1 filter paper through a Celite cake to remove any insoluble material. The filtrate was subsequently filtered using a Prep / Scale Tangential Flow Filtration (TFF) 2.5 ft² (0.232 m²) cartridge with a 1 kDa exclusion limit (Millipore Corporation, Bedford, MA, USA). The permeate fraction was collected, demineralized by electro dialysis in Dr. Laurent Bazinet laboratory at Laval University, lyophilized, and stored at -30 °C until further use.

Manufacturing of Enzeco Trypsin:Chymotrypsin 1:1 stopped in 2014, so new enzymes were procured for the SPF processing. Pepsin was also replaced because of its relatively low specific activity. Pepsin (EC 3.4.23.1, Millipore Sigma Cat. No.: P6887, 3,200-4,500 units/mg protein) was purchased from Millipore Sigma (St. Louis, MO, USA) and a Trypsin (EC 3.4.21.4) and Chymotrypsin (EC 3.4.21.1) mixture (1:1) was purchased from Creative Enzymes (Shirley NY, USA) (Cat. No.: PHAM-378, 1000 units/mg protein). Otherwise all steps for SPF production were performed as described by Jin (2012).

3.2 Fractionation of the SPF using Strong Anion Exchange and Gel Filtration Chromatography

Three strong anion exchange (SAX) and two gel filtration (GF) chromatographic separations of the SPF were performed using an ÄKTAexplorer 10 XT FPLC system, equipped with a Frac-950 fraction collector (GE Healthcare, Chicago, IL, USA) to fractionate potential bioactive peptides. The operating conditions and the

elution strategy (Tables 5 and 6, respectively) for each separation varied depending on the separation technique and the results of previous separations. Tricorn columns (GE Healthcare, Chicago, IL, USA) were packed in-house with UNOsphere™ Q SAX media (quaternary ammonia, NH₄⁺) for Separations 1, 2 and 5, or Bio-Gel P-2 GF media (1800 Da - 100 Da exclusion limit) (both from Bio-Rad Laboratories, Hercules, CA, USA) for Separations 3 and 4. The difference between SAX Separations 1 and 2 were column length (100 mm vs. 300 mm) and buffer compositions: Separation 1 contained 20 mM ammonium formate and elution carried out on a non-volatile salt gradient; Separation 2 contained 25 mM ammonium formate and elution was accomplished with a gradient of ammonium formate that was volatile. Freeze-dried SPF was suspended in eluant A, then filtered using 0.2 µm Whatman syringe filters before loading the sample loop. UV absorbance was measured at 214, 254, and/or 280 nm wavelengths. Repeated fractionations of n = 8 for Separation 1, n = 6 for Separations 2 and 3, n = 8 for Separation 4 and n = 7 for Separation 5 were performed. Eluted peptides from repeated fractionations were pooled, then freeze-dried, weighed and stored at -30 °C until *in vitro* screening.

Table 5. Operational characteristics for strong anion exchange and gel filtration separations of SPF. Separations were performed using an ÄKTA Explorer 10 XT system (GE Healthcare, Chicago, IL, USA) and Unicorn V4.12 Software, at ambient temperature. Strong anion exchange was performed using UNOsphere Q media and gel filtration was performed using Bio-Gel P-2 media (both from Bio-Rad Laboratories, Hercules, CA, USA). Column packing media were prepared into Tricorn columns (GE Healthcare, Chicago, IL, USA).

Separation (No.) and Format	Column Dimension (mm)	Flow Rate (mL/min)	Injection Volume (μ L)	Protein Concentration (mg/mL)	Collected Subfractions (No.)	V _{Fraction} (mL)
1 – SAX	10 x 100	0.3	500	100	2	14.5 & 21.5
2 – SAX	10 x 300	0.5	500	50	8	20 & 10
3 – GF	10 x 300	0.5	500	50	7	5
4 – GF	10 x 300	0.1	100	40	13	2
5 – SAX	10 x 300	1.0	500	40	8	15

SAX: strong anion exchange; GF: gel filtration; V_{TColumn} = total column volume; V_{Fraction} = fraction volumes collected

Table 6. Elution strategies prepared for strong anion exchange and gel filtration separations of SPF on an ÄKTA Explorer 10 XT system (GE Healthcare, Chicago, IL, USA) processed with Unicorn V4.12 Software. Dual buffer systems were used for strong anion exchange to facilitate a gradient elution strategy, and gel filtration separations used a single buffer system with an isocratic elution strategy. Following elution, columns were washed using 100 % eluant B solutions until baseline absorbance was achieved, then regenerated using eluant A.

Separation (No.) and Format	pH	Eluant A	Eluant B	Elution Duration (CV)	Elution Gradient (% B)
1 – SAX	8.0	20 mM AF	20 mM AF + 1.0 M NaCl	3.0	0 – 30
2 – SAX	9.0	25 mM AF	1.0 M AF	2.0	0 – 70
3 – GF	8.5	100 mM AF + 0.5 % FA	-	1.5	-
4 – GF	6.0	50 mM AF	-	1.75	-
5 – SAX	8.0	25 mM AF	1.0 M AF	3.0	0 – 50

SAX: strong anion exchange; GF: gel filtration; AF: Ammonium Formate; FA: Formic Acid; CV: Column Volumes; % B: Proportion of eluant B in mobile phase

3.3 *In vitro* Screening Assay for SPF Action on Metabolism

Screening assays were performed in Dr. André Marette's laboratory at Laval University (Quebec, QC, Canada). L6 rat myoblasts (courtesy of Dr. Amira Klip, Hospital for Sick Children, Toronto, Canada) were grown and maintained in monolayer culture in alpha-MEM containing 2 % (v/v) fetal bovine serum in an atmosphere of 5% CO₂ at 37 °C. L6 myoblasts were plated in 12-well or 24-well plates at 20,000 cells/mL, replacing the media every two days until complete differentiation into myotubes (7 days post-plating).

Once completely differentiated, myotubes were serum deprived (alpha-MEM with 0% FBS) for 3 h and treated or not with SPF (1 ng/mL or 1 µg/mL) for 2 h without (insulin-independent) or with 100 nM insulin (insulin-dependent) during the last 45 min. Cells were rinsed once with glucose-free HEPES-buffered saline solution pH 7.4 (140 mM NaCl, 20 mM HEPES/Na, 5 mM KCl, 2.5 mM MgSO₄ and 1 mM CaCl₂), and were subsequently incubated for 8 min with 10 µM 2-deoxy-D-glucose containing 0.3 µCi/mL 2-deoxy-D-[³H]glucose in the same buffer. After incubation in transport medium, cells were rinsed three times with ice-cold saline solution (0.9% NaCl) and stored at -20 °C. Cells were disrupted by adding 1 mL of 50 mM NaOH to plates with agitation for 15 min. Cell-incorporated radioactivity was measured using a Perkin Elmer Tricarb liquid scintillation counter. Protein concentrations were determined by the micro bicinchoninic acid method (BCA) using a bovine serum albumin (BSA) standard curve (ThermoScientific, Waltham,

Mass, USA), and the results expressed in pM/min/mg protein, calculated using the following equation:

$$\frac{\text{DPM (sample)}}{C \times \text{DPM (2DG)} \times t}$$

where DPM (sample) is the number of disintegrations per minute (DPM) measured for the tested sample, C is the concentration of protein (mg), DPM (2DG) is the number of DPM measure for the solution of radioactive 2-deoxy-D-[³H] glucose for 1 pmol and equal to 72.2025 dpm/pmol, and t is the incubation time with 2-deoxy-D-[³H] glucose, and reported in terms of relative activity to the control sample in the absence of insulin. Statistical analysis was performed in Microsoft Excel using a two-tailed Student's t-test assuming equal variance.

3.4 Protein Identification and Peptide Detection: LC-MS Method 1

Bioactive fractions from strong anion exchange chromatography of SPF (Separation 1) were subject to analysis by LC-MS/MS at the Dalhousie University Biological Mass Spectrometry Core Facility on a VelosPRO orbitrap mass spectrometer (ThermoFisher Scientific, Waltham, Mass, USA) equipped with an UltiMate 3000 Nano-LC system (ThermoFisher Scientific, Waltham, Mass, USA). Chromatographic separation of the digests was performed on a PicoFRIT C₁₈ self-packed 75 mm x 60 cm capillary column (New Objective, Woburn, MA) at a flow rate of 300 nL/min. MS and MS/MS data was acquired using a data-dependent acquisition method in which a full scan was obtained at a resolution of 30,000, followed by ten consecutive MS/MS spectra in both higher-energy collisional dissociation (HCD) and collision-induced dissociation (CID) mode (normalized

collision energy 36%). Internal calibration was performed using the ion signal of polysiloxane at m/z 445.120025 as a lock mass.

Raw MS data were analyzed using Proteome Discoverer Version 2.1 (ThermoFisher Scientific, Waltham, Mass, USA). The Sequest HT program was used to compare peak lists to the *Salmonideae* UniprotKB protein database as well as the cRAP database of common contaminants (Global Proteome Machine Organization), based on their tryptic cleavage for a minimum peptide length of 6, with tolerance for two missed cleavages. Cysteine carbamidomethylation was set as a fixed modification, while methionine (Met) oxidation, N-terminal Met loss, and phosphorylation on serine, threonine, and tyrosine were included as variable modifications. A mass accuracy tolerance of 10 ppm was used for precursor ions, while 0.02 Da for HCD fragmentation was used for product ions. The Percolator program was used to determine confident peptide identifications using a 0.1% false discovery rate (FDR). Parent proteins reporting Sequest HT scores > 10 were accepted.

3.5 Amino Acid Analysis

The amino acid analysis of bioactive fractions from SAX Separation 1 was performed at the SickKids Proteomics, Analytics, Robotics & Chemical Biology Centre (SPARC Biocentre) (Toronto, ON, CA) using the Waters Pico-Tag method. Lyophilized fractions were completely hydrolyzed in formic acid for 24 h, and derivatized with phenylisothiocyanate (PITC). Samples were separated on a

Waters Acquity UPLC BEH C18 column (2.1 mm x 10 cm) with a column temperature of 48 °C and the relative abundance of individual amino acids determined by calibration with Pierce™ Amino Acid Standard (ThermoFisher Scientific, Waltham, Mass, USA). Each sample was measured in triplicate.

3.6 Peptide Sequence Alignment

Sequence alignments were accomplished using the Clustal Omega web tool (<https://www.ebi.ac.uk/services>) (Sievers et al., 2014).

3.7 Optimization of Protein Recovery from Atlantic Salmon using Alkaline Solubilization and Isoelectric Precipitation

Protein recovery from Atlantic salmon muscle tissue was investigated following a method as previously described (Chen and Jaczynski, 2007), with modifications. Previously frozen, whole Atlantic salmon fillets were ground using a Moulinex household meat grinder and homogenized in dH₂O at a 1:4 (w/v) ratio in a blender for 30 s at high speed. The pH was measured using a Fisher Scientific™ Accumet™ AR15 benchtop pH meter with 3-point calibration at pH 4.0, 7.0 and 10.0. The pH was adjusted to target pH's at 12.0, 12.5 and 13.0 ± 0.05 with 2.0 M NaOH then adjusted to a final volume of 50 mL and stirred for 2 h. Each solution was centrifuged at 5,200 x *g* for 20 min at 4 °C using an IEC Centra MP 4R centrifuge (International Equipment Company, Chattanooga, TN, USA) and then an aliquot was collected from the supernatant and its protein concentration was measured using the BCA method with a BSA standard curve to represent the total available protein in each solution (ThermoFisher Scientific, Waltham, Mass, USA).

The supernatants were decanted into a clean beaker and adjusted to pH 4.5 or 5.5 \pm 0.05 with 6.0 M HCl to isoelectrically precipitate soluble protein. Precipitates were left overnight at 4 °C and the following day centrifuged at 5200 x *g* for 20 min at 4 °C to separate the soluble protein from insoluble protein. An aliquot was collected from the supernatant and its protein concentration measured to represent to amount of soluble protein remaining at each suggested isoelectric point. The protein recovery following each pH combination was estimated following Formula (1).

$$\text{Protein Recovery (\%)} = 1 - \left(\frac{(\text{Total Protein}) - (\text{Soluble Protein})}{\text{Total Protein}} \right) * 100 \quad (1)$$

Statistical analyses were performed by analysis of variance (ANOVA) at a confidence level of 95 % for significance. *Post hoc* analyses were performed using the least squared difference (LSD) test. Statistical analyses were performed using the IBM SPSS Statistics Premium software Version 25.

3.8 Solubility Curves

Solubility curves describing the solubility kinetics of alkali-solubilized, isoelectrically precipitated Atlantic salmon muscle tissue were prepared as previously described (Chen and Jaczynski, 2007), with modifications. An alkali-solubilized Atlantic salmon muscle protein precipitate prepared following the SPF methodology (hereafter termed the high-alkali solubilized salmon muscle) was prepared using 1.0 M NaOH and compared to a second Atlantic salmon muscle

protein precipitate solubilized using a solution with lower alkalinity (0.1 M NaOH) (hereafter termed the low-alkali solubilized salmon muscle). Following the 2 h alkaline-solubilization, solutions were adjusted to the acidic range using 6.0 M HCl, and 1 mL aliquots from the high-alkali solubilized salmon muscle solution were collected at pH 13.0, 11.0, 9.0, 8.0, 7.0, 6.0, 5.0, 4.5, 4.0, 3.0, 2.0, and 1.0, and aliquots from the low-alkali solubilized salmon muscle solution were collected at pH 12.5, 11.5, 10.5, 9.5, 8.5, 7.5, 6.5, 5.5, 5.0, 4.5, 3.5, 2.5, 1.5. Samples were immediately mixed with 1 mL of 5 % trichloroacetic acid (TCA), vortexed, and left overnight at 4 °C. At each pH interval, the conductivity was also measured using a Mettler Toledo SevenEasy S30 conductivity meter equipped with an InLab® 731 conductivity probe and calibrated using a 1,413 µS/cm conductivity standard solution (Mettler Toledo, Columbus, OH, USA). The ionic strength (IS) of each solution was estimated by comparing conductivity to a standard curve of NaCl prepared between 0.0 and 1.0 M, based on formula (2).

$$\text{Ionic Strength} = \frac{1}{2} \sum C_i Z_i^2 \quad (2)$$

where C represents the molarity of the solute and Z represents the ion valency, and assuming that NaCl represents the primary salt in each solution. The following day, samples were centrifuged at 4,800 x g for 10 min at 4 °C using an IEC Centra MP 4R centrifuge (International Equipment Company, Chattanooga, TN, USA). The soluble protein concentrations of each sample were measured using the BCA

method with a BSA standard curve (ThermoScientific, Waltham, Mass, USA), and protein solubility was calculated following formula (3).

$$\text{Protein Solubility (\%)} = \frac{[\text{Sample Protein}]}{[\text{Starting Protein}]} * 100 \quad (3)$$

Statistical analysis was performed by ANOVA at a confidence level of 95 % for significance. *Post hoc* analyses were performed using the least squared difference (LSD) or Bonferroni tests and performed using the IBM SPSS Statistics Premium software Version 25.

3.9 Preparation of Salmon Protein Isoelectric Precipitates

Isoelectric precipitates of salmon protein were prepared from previously frozen, Atlantic salmon loin muscle, ground using a Moulinex household meat grinder and homogenized with 1.0 M, 0.5 M, 0.25 M or 0.1 M NaOH at a 1:4 (w/v) ratio in a standard blender on high for 30 s at high speed, then stirred for 2 h, or as indicated. Salmon muscle solutions were centrifuged at 3,500 x *g* for 10 min at 4 °C using an IEC Centra MP 4R centrifuge (International Equipment Company, Chattanooga, TN, USA) and supernatants were transferred to a clean beaker. Proteins were precipitated by adjustment to pH 4.5 using 6.0 M HCl and were then centrifuged at 3,500 x *g* for 10 min at 4 °C. The pellet was lyophilized then kept at -30 °C until later use but the supernatant protein was precipitated by using ice cold acetone at a 1:4 (v/v) ratio and incubated under freezing conditions for 30 min. The

precipitated proteins from supernatants were lyophilized and kept at -30 °C until later use.

3.10 Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) was performed using an LKB 2001 vertical electrophoresis unit coupled to an LKB 2297 Macro Drive 5 and an LKB 2219 Multitemp II Thermostatic Circulator (LKB-Produkter AB, Bromma, Sweden). Gels were poured in-house into assembled cassettes (16 x 16 x 0.15 cm) with a 38.4 mL volume. Three stock solutions were used to prepare the gels: solution A (30% acrylamide (w/v), 0.8% (w/v) bisacrylamide in dH₂O, 37.5:1), solution B (0.5 M Tris-HCl, pH 6.8), and solution C (3.0 M Tris-HCl, pH 8.8) (Bio Rad Laboratories, Hercules, CA), as well as a 10 % SDS (Millipore Sigma, St. Louis, MO) and 1 % ammonium persulfate (AMMO) solutions (Bio Rad Laboratories, Hercules, CA). Gels were prepared following a variation of the manufacturer's recipe to a volume of 30 mL for resolving gels, and 10 mL for stacking gels, per cassette. Gradient resolving gels (5 – 15 %) were prepared using an LKB gradient gel former with light (2.5 mL Solution A, 2.0 mL Solution B, 10.0 mL dH₂O, 0.15 mL SDS, 0.35 mL AMMO, 0.04 mL TEMED) and heavy (7.5 mL Solution A, 2.0 mL Solution B, 5.0 mL dH₂O, 0.15 mL SDS, 0.35 mL AMMO, 0.02 mL TEMED) acrylamide solutions, while the stacking gel (1.25 mL Solution A, 2.5 mL Solution C, 6.25 mL dH₂O, 0.1 mL SDS, 0.5 mL AMMO, 0.04 mL TEMED) was poured isocratically. Alternatively, 15 % resolving gels were prepared using 15.0 mL Solution A, 4.0 mL Solution B, 10.0 mL dH₂O, 0.15 mL

SDS, 0.35 mL AMMO, and 0.04 mL TEMED. All solutions were degassed for five min before the addition of the SDS, AMMO and TEMED to prevent foaming and premature polymerization. The resolving gel was left to polymerize for 1 h, after which the stacking gel solution was poured with an inserted 10-well comb and left to polymerize overnight. Reservoir buffer was prepared from a 10x stock solution (144 g glycine, 30.3 g tris, 10 g SDS/L) to a final volume of 5.5 L.

Protein samples were prepared to 2 mg/mL and diluted 1.25-fold with sample buffer (0.1 M Tris-HCl, 10 % β -mercaptoethanol (v/v), 8 % SDS (w/v), 33 % glycerol (v/v), 0.05 % bromophenol blue (w/v)), heated to 70 °C for 15 min, then centrifuged at 5,250 x g for 10 min at 4 °C using an IEC Centra MP 4R centrifuge (International Equipment Company, Chattanooga, TN, USA), and 30 μ L were loaded into each well. Precision Plus Protein™ All Blue Prestained Protein Standards were used for the MW ladder (Cat No. 1610373, Bio Rad Laboratories, Hercules, CA) and 20 μ L were loaded directly to the well. Each run was operated at a constant temperature of 15 °C and followed a constant-voltage electrophoretic treatment, starting at 100 V for 1 h or until the dye reached the stacking/resolving gel interface, and then increased to 300 V and terminated when the dye front reached 1 cm from the end of the gel. Following the completion of the run, the gel was transferred to a glass dish and washed three times with enough dH₂O to cover the gel surface, for 5 min each. Fixing solution (45:10:55, methanol:acetic acid:water) was added to cover the gel, and left overnight at 4 °C. The gel was washed once with dH₂O and stained for 1 h with staining solution (10% (v/v) acetic

acid, 0.025% (w/v) Coomassie Brilliant Blue R-250 (Bio Rad Laboratories, Hercules, CA, USA) with shaking for 1 h. The gel was destained with water and heating to 60 °C for 2 h, replacing the solution after one hour. Destaining continued until sufficient minimization of background was achieved. Gel images were recorded using a ChemiDoc XRS+ System and processed using the Image Lab Software (Bio Rad Laboratories, Hercules, CA).

3.11 In-Gel Digestion

Excised gel slices stained with Coomassie Brilliant Blue R-250 were processed for in-gel digestion as previously described, with slight modifications (Shevchenko et al., 2007). Briefly, gel slices were washed for 2 h in dH₂O and then cut into ~1 mm cubes and rinsed twice with 200 µL of dH₂O. Gel cubes were reduced with 10 mM dithiothreitol (DTT) at 56 °C for 30 min, then alkylated with 55 mM iodoacetamide for 30 min at room temperature in the dark, and then dehydrated with 200 µL acetonitrile (ACN). Dried gel cubes were saturated with a pH 2.0 pepsin solution (EC 3.4.23.1, Millipore Sigma Cat. No.: P6887, 3,200-4,500 units/mg protein) and incubated overnight at 37 °C. The following day, each sample was dried under a stream of nitrogen, and then resuspended in a pH 7.8 trypsin:chymotrypsin solution (10,000 U/g fish protein; PHAM-378, Creative Enzyme, Shirley, NY), and incubated for 4 h at 37 °C. Gel cubes were washed with 50 µL of 4 % formic acid for 10 min, then peptide extracts were transferred to a new tube. Gel cubes were washed once more with 100 % ACN until completely dried and then combined the peptide extracts. Peptide extracts were dried in a vacuum centrifuge, then

resuspended in dH₂O and loaded onto Sartorius Vivaspin 2 Hydrosart 2000 Da MWCO ultrafiltration centrifugal concentrators and centrifuged at 3,250 x g for 10 min at 4 °C using an IEC Centra MP 4R centrifuge (International Equipment Company, Chattanooga, TN, USA). Permeates were collected and transferred to clean, pre-weighed microcentrifuge tubes, lyophilized and stored at -30 °C until later use.

3.12 Protein Identification

Proteins were identified as outlined in Section 3.4 with modifications. Excised gel bands were processed as outlined in Section 3.11, but following reduction and alkylation steps, dried gel cubes were saturated with 20 µg/mL of trypsin protease (Cat. No.: 90057, Pierce™ ThermoFisher Scientific, Waltham, Mass, USA) for 2 h, then 20 µL of 50 mM ammonium bicarbonate was added and the samples were incubated overnight at 37 °C. Digested peptides were extracted from the gel cubes by treatment with 100 µL of 50 % ACN in 5 % FA. The peptide-containing solution was dried to a pellet in a vacuum centrifuge and subsequently resuspended in 20 µL of a 3 % ACN, 0.5 % formic acid solution, and processed as outlined.

3.13 Peptide Detection: LC-MS Method 2

LC-MS/MS analyses were performed by the team of Dr. L. Bazinet at Laval University using a 1290 Infinity II UPLC (Agilent Technologies, Santa Clara, CA, USA) consisting of a binary pump (G7120A), a multisampler (G7167B), an in-line degasser and a variable wavelength detector (G7114B) adjusted to 214 nm. The

sample was loaded (10 μ L) onto an Acquity UPLC CSH 130 1.7 μ m C18 column (2.1 \times 150 mm i.d., Waters Corporation, Milford, MA, USA). The column was operated at a flow rate of 400 μ L/min at 45 $^{\circ}$ C. A linear gradient consisting of solvent A (LC-MS grade water with 0.1 % formic acid) and solvent B (LC-MS grade ACN with 0.1 % formic acid) was applied, with solvent B going from 2 % to 25 % in 50 min, holding until 53 min, ramping to 90 % and holding until 57 min, then back to initial conditions.

A hybrid ion mobility quadrupole TOF mass spectrometer (6560 high definition mass spectrometry (IM-Q-TOF), Agilent, Santa Clara, USA) was used to identify and quantify the relative abundances of the peptides. Signals were recorded in positive mode at Extended Dynamic Range, 2 Ghz, 3200 m/z with a scan range between 100 – 3200 m/z . Nitrogen was used as the drying gas at 13.0 L/min and 150 $^{\circ}$ C, and as nebulizer gas at 30 psig (0.207 MPa). The capillary voltage was set at 3500 V, the nozzle voltage at 300 V and the fragmentor at 400 V. The instrument was calibrated using an ESI-L low concentration tuning mix (G1969-85000, Agilent Technologies, Santa Clara, CA, USA). Data acquisition and analysis was performed using the Agilent MassHunter Software package (LC-MS/MS Data Acquisition, Version B.07.00 and Qualitative Analysis for IM-MS, Version B.07.00 with BioConfirm Software) to compare detected ions to the NCBI *Salmo salar* protein database, based on a no enzyme, pepsin, and/or trypsin:chymotrypsin cleavage with a minimum peptide length of three and tolerance for two missed cleavages. Variable modifications for oxidized methionine, pyroglutamic acid, deamination of Asp, and phosphorylation to Ser,

Thr, and Tyr, were tolerated by the database searching algorithm. Precursor ions between 100 – 3,200 Da were selected for MS/MS. A mass accuracy tolerance of 20 ppm was used for precursor ions, while 50 ppm was used for product ions. The Agilent MassHunter Find by Molecular Feature algorithm was performed with Molecular Feature Extraction (MFE) as a pre-processing step to identify features (peptides), ions distinct from background noise, from within the raw MS spectral data. Compound lists for each bioactive fraction were calculated, reporting the mass, retention time, and relative intensity of significant precursor ions detected during LC-MS, but contained no sequence information. The validation of database searching was performed using modified validation criteria; peptides with scores > 7, or % SPI > 70, or minimum spectrum intensity of 1.0×10^6 , and simultaneously identified by the MFE algorithm, were considered valid peptide sequences.

3.14 Sequence Logos

Sequence logos (Schneider and Stephens, 1990) were prepared using the WebLogo 3 web-application (<http://weblogo.threeplusone.com/>) (Crooks et al., 2004). Branched chain amino acids (Ile, Leu, Val) were coloured blue, anionic residues (Asp, Glu) were coloured green, cationic residues (Arg, His, Lys) were coloured red, aromatic amino acids (Phe, Trp, Tyr) were coloured purple and all others black. The frequency was expressed as the fraction of the indicated residue at each position, and peptides were reported according to their length using single letter amino acid codes.

3.15 Peptide Identification: *In silico* Digestion of SPF Progenitor Proteins

The high-quality prediction of compositions for each bioactive fraction using the Agilent MassHunter Find by algorithm with Molecular Feature Extraction (MFE) (Section 3.13) was used to support the identification of peptide sequences. An in-house peptide database was developed resulting from *in silico* digestions of ten potential SPF progenitor proteins: myosin heavy chain, alpha actin, tropomyosin, creatine kinase, glyceraldehyde-3-phosphate dehydrogenase, fructose-bisphosphate aldolase A, triosephosphate isomerase, phosphoglycerate mutase, myosin light chain, and beta enolase (Appendix A, Table A10), and selected due to their identification as progenitors to peptides identified in SAX Separation 1 and as components of the low-alkali salmon muscle precipitate. Each primary sequence was processed by specific or non-specific hydrolysis rules explained below to enable the annotation of each compound list generated for bioactive SPF fractions by the MFE with putative peptide identities (Appendix A, Table A14 – A20), independent of software-assisted database searching.

Specific. The first in-house peptide database was generated using the Peptide Mass tool (SIB, Swiss Institute of Bioinformatics; Artimo et al., 2012), and represents peptide sequences generated through the activities of pepsin or trypsin and chymotrypsin, and allowing for up to 3 missed cleavages. The masses for each peptide sequence predicted by the *in silico* digestion of SPF progenitor proteins were directly compared to the masses calculated by the MFE algorithm for each

bioactive fraction, where matching masses represented a putative identification corresponding to the sequence from the in-house database.

Non-Specific. A second in-house database was generated using the FindPep Tool (SIB, Swiss Institute of Bioinformatics; Artimo et al., 2012) and represents any peptide sequence from within SPF progenitor primary sequences that match the masses of MFE-calculated ions in bioactive fractions. In contrast to the 'specific' database, the FindPep Tool searches the complete primary sequence of progenitor proteins for matching peptide sequences and is not limited to only those peptides generated by the specificity of enzymatic activity. Theoretical sequences were matched to experimental precursor ions with a mass tolerance of 10 ppm.

3.16 Peptide Identification: *De novo* Sequencing

De novo sequencing was facilitated using Arcadiate software Version 4.5 and mMass software Version 5.5 (Strohalm et al., 2008). Arcadiate was used to generate base peak chromatograms of each bioactive fraction and to identify the precursor ions represented by dominant peaks. The MS/MS spectra of each precursor ion generated by fragmentation were visualized using mMass and its interpretation represented a putative identification. Sequences for each MS/MS spectra were informed by identifying masses unique to each of the 20 standard amino acids, with assistance from theoretical calculations performed by the mMass software. The MS/MS spectra of each putative peptide identification by this

approach were exported to Sigma Plot Version 11 (San Jose, CA, USA) for additional annotation of the product ions.

3.17 Validation of Potential Bioactive Peptides

Synthetic peptides Ile-Ala-Ile (4.2 mg; 99.59 % purity), Ile-Gly-Ile (4.3 mg; 99.47 % purity), Ile-Ile-Ile (4.1 mg; 98.56 % purity), Ile-Ala-Tyr (4.3 mg; 98.48 % purity), Ile-Gly-Tyr (4.1 mg; 98.73 % purity) and Ile-Ile-Tyr (4.2 mg; 99.19 % purity) were purchased from Bio Basic Canada Inc. (Markham, ON, Canada) and were validated by the manufacturer by HPLC-MS/MS.

CHAPTER 4: OBJECTIVES

The SPF was derived from Atlantic salmon (*Salmo salar*) protein digested with a combination of enzymes followed by the concentration of the LMW peptide fraction. The identification of glucose uptake stimulating peptides, and others targeting systemic metabolic abnormalities with high therapeutic potential, present an opportunity for the development of naturally-derived bioactive peptide extracts. Functional LMW hydrolysates are frequently developed, but analyzing their compositions presents a significant challenge using the established proteomic methodologies of peptide separation and detection, often leading to a limited assessment of the peptide composition in each fraction (Panchaud et al., 2012). Chromatographic separations based on size (gel filtration), charge (ion exchange, chromatofocusing), and surface hydrophobicity (reverse phase, normal phase) and coupled to tandem mass spectrometry are formats capable of providing suitable selection of low molecular weight peptides. A combination of these methods is consistent with the peptidomic approach for peptide identification to overcome the diverse molecular characteristics of LMW hydrolysates (Arroume et al., 2016).

In the following chapters, methodologies for evaluating the SPF to identify bioactive peptides are described. First, a chromatographic approach to fractionate SPF is performed. Then SPF fractions are screened for their ability to modulate glucose uptake in cultured L6 myotubes and bioactive fractions are analyzed to identify the peptide sequences. Various computational approaches targeting both

the identification of di- and tripeptides and consistent qualitative attributes are investigated, then putative bioactive sequences are synthesized and screened for their ability to affect glucose uptake under purified conditions. It is hypothesized that complementary strategies to each the fractionation and identification will demonstrate the need for improvements to the methodologies for screening di-, tri- and oligopeptides in SPF and from other food-derived protein hydrolysates for identifying putative sequences and/or characteristics of bioactive peptides.

CHAPTER 5: RESULTS AND DISCUSSION

5.1 SAX Separation 1 of SPF

5.1.1 Objective

A previous separation of the SPF by Roblet et al. (2016) and other LMW hydrolysates (Doyen et al., 2014) using electro dialysis with filtration membranes have suggested that charged and neutral peptides are each capable of stimulating glucose uptake in cultured L6 myotubes, but the identification of peptide sequences was not performed. In this section, SPF fractions are prepared by strong anion exchange (SAX) chromatography to confirm the glucose uptake stimulation of charged peptides, and to identify the potential bioactive peptides in these fractions.

5.1.2 Chromatography of SAX Separation 1

SAX chromatography of SPF with UV detection at 214 nm and 254 nm is shown in Figure 4. UV absorbance peaks observed during the isocratic and gradient elution steps correspond to fractions 1 and 2, respectively. Detection at 214 nm (Figure 4A) targets amide bonds indicating peptide elution by peaks at ~ 30 min and 160 min. Detection at 254 nm (Figure 4B) targets phenylalanine and nucleotides that suggest the presence of Phe in peptides from both fractions.

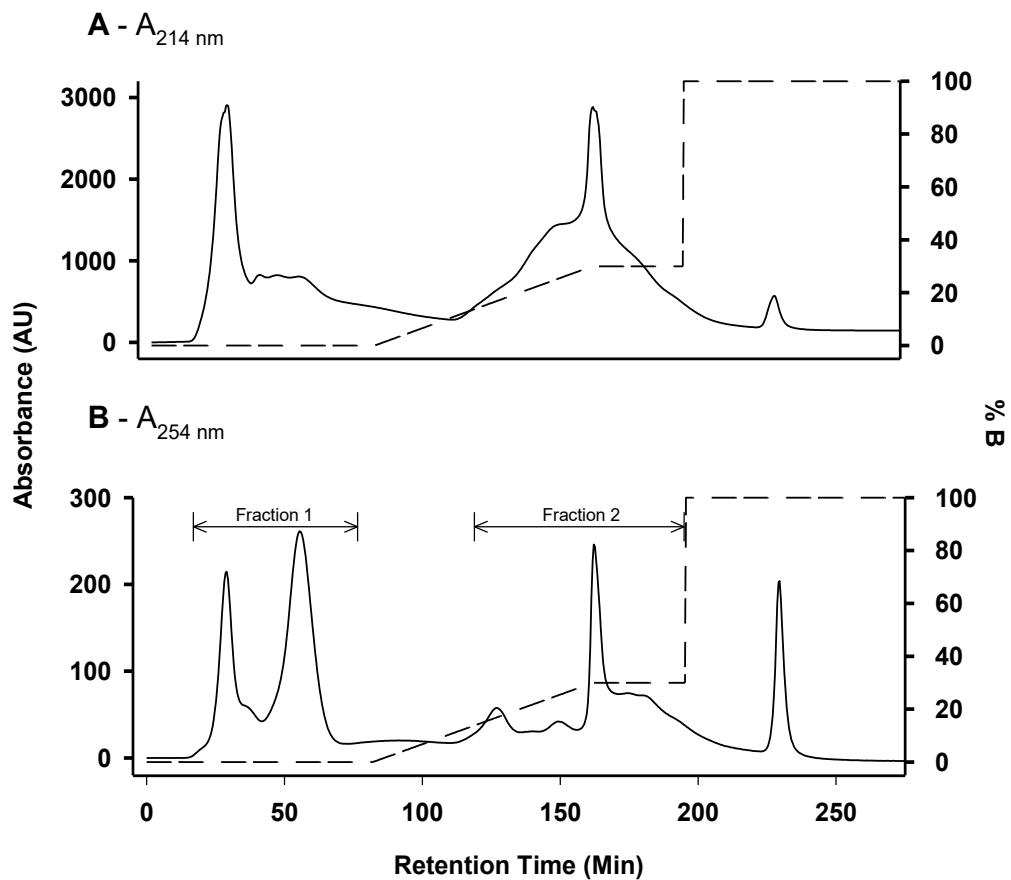


Figure 4. SAX Separation 1 of SPF on UNOsphere Q media; 10 mm x 100 mm column; flow rate, 0.3 mL/min; eluant A, 20 mM ammonium formate pH 8.0; eluant B, 20 mM ammonium formate plus 1.0 M NaCl, pH 8.0; gradient, 0 – 30 %, hold for 40 min, detection at (A) 214 nm and (B) 254 nm; sample concentration, 100 mg/mL; injection volume, 500 μ L. Solid lines represent the UV absorbance and dotted lines represent the percent of elution buffer. Fractions were collected within the indicated boundaries and pooled from $n = 8$ separations.

In the initial isocratic portion of SPF elution, two distinct peaks were observed at elution times of ~ 30 and 55 min as observed at 254 nm (Figure 4B), whereas only the first peak (elution time ~ 30 min) appeared when monitored at 214 nm (Figure 4A). This suggests that perhaps the peak eluting at ~ 55 min contained few amide bonds and perhaps could have been a nucleotide or a LMW peptide containing Phe. During the column regeneration step, the peak at ~ 240 min at both wavelengths indicates that the elution strategy may have been insufficient to remove all peptides bound to the column.

5.1.3 Glucose Uptake Analysis of Fractions of SAX Separation 1

The activity of SPF fractions 1 and 2 from SAX Separation 1 was tested using the *in vitro* screening assay outline in Section 4.4 (Table 7). The unfractionated SPF stimulated glucose uptake at 1 µg/mL (p-value = 0.044) only in the absence of insulin but was not evaluated at 1 ng/mL. Fraction 1 had no modulating effect on glucose uptake in either the absence or presence of insulin at 1 µg/mL but had a stimulating effect in both conditions at 1 ng/mL (p-values = 0.006 and 0.002, respectively).

In contrast, fraction 2 exhibited a glucose uptake stimulating effect in both the presence and absence of insulin at 1 µg/mL (p-value = 0.018 and 0.009, respectively), and in the absence of insulin at 1 ng/mL (p-value = 0.012). Bioactivity in the absence of insulin by fractions 1 and 2 at 1 ng/mL is consistent to the unfractionated SPF control of the present study, as well as others (Chevrier et al., 2015), indicating that both fractions contain potential bioactive peptide sequences.

Interestingly, the dilution of fraction 1 increased its glucose uptake stimulation, but the same treatment to fraction 2 had the reverse effect and could reflect the importance of peptide concentration or the balance of stimulating and inhibiting peptides that others have proposed (Roblet et al., 2016). The peptide compositions of both fractions are evaluated in the following section to identify the sequences that compose both fractions and potential bioactive peptide sequences.

Table 7. Glucose uptake modulation, of SPF and its fractions from SAX Separation 1, following the methodology outlined in Section 3.3. The screening assay was performed with a 2 h incubation time at 1 $\mu\text{g}/\text{mL}$ and 1 ng/mL , in the (-) absence and (+) presence of insulin.

Concentration	Glucose Uptake (fold activity to control)			
	1 $\mu\text{g}/\text{mL}$		1 ng/mL	
Sample	Insulin (-)	Insulin (+)	Insulin (-)	Insulin (+)
Control	1.00 \pm 0.00	1.68 \pm 0.03	1.00 \pm 0.00	1.68 \pm 0.03
SPF	1.14 \pm 0.07*	1.76 \pm 0.09	-	-
Fraction 1	0.98 \pm 0.04	1.66 \pm 0.07	1.15 \pm 0.05*	2.02 \pm 0.08*
Fraction 2	1.21 \pm 0.07*	1.99 \pm 0.11*	1.16 \pm 0.06*	1.83 \pm 0.09

Values are means \pm standard error of the mean, n = 8 individual experiments, performed in triplicate; * p < 0.05 vs. control.

5.1.4 Peptide Identification

Peptide identification was performed following the methodology outlined in Section 4.5.1. Over 500 unique peptides were identified in SAX Separation 1 fractions 1 and 2 (Appendix A, Tables A1 and A2) and are summarized in Table 8. These results exclude peptides < 6 amino acids omitted by the database searching

software to eliminate false positive identifications. The average masses of identified peptides from each fraction were both concentrated between 1 and 2 kDa (Figure 5) that was inconsistent with previous studies that characterized bioactive SPF fractions using a methodology that instead set a minimum peptide length of 3 amino acids, and reported that 50 % of peptides in the SPF were < 500 Da and fewer than 2 % of peptides were > 1000 Da (Roblet et al., 2016), although different SPF batches were used. Clearly the limitation of only listing peptides with > 6 amino acids as set in the software is a major reason for the discrepancy between this and previous studies. Peptides from both fractions 1 and 2 in the present study were clustered around isoelectric points at pH 4.0, 7.0 and 10.0 (Figure 5), therefore peptides in fraction 1 were found from both neutral and cationic clusters and peptides in fraction 2 were found exclusively in the anionic cluster.

Table 8. Composition of bioactive fractions from the SAX Separation 1 determined by LC-MS analysis.

Fraction	Peptides (No.)	Average Molecular Weight (Da)	Average Isoelectric Point
1	293	1183.013 ± 23.111	8.25 ± 0.10
2	209	1353.816 ± 26.630	4.21 ± 0.05

Values are means ± standard error of the mean.

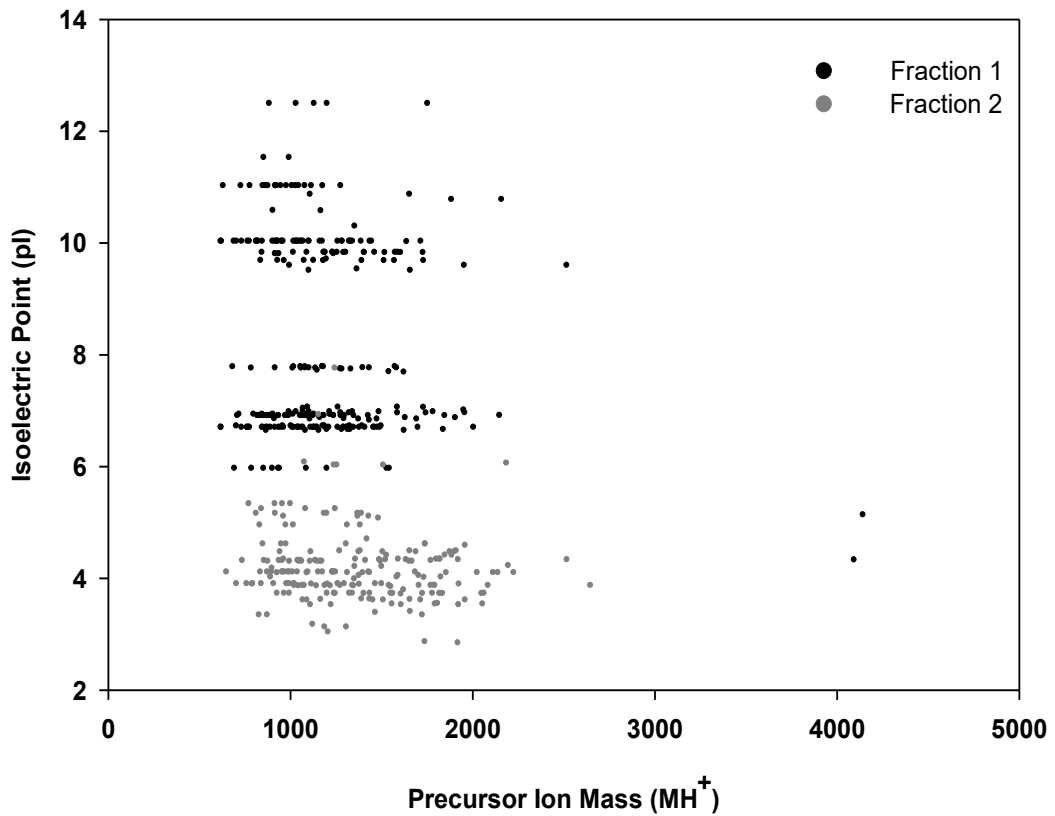


Figure 5. The precursor ion mass and isoelectric point distribution of peptides with > 6 amino acid residues identified by LC-MS in bioactive fractions from SAX Separation 1. Each dot represents a single peptide identified by LC-MS in: (Black) fraction 1, (Grey) fraction 2.

The isoelectric clustering of identified peptides can be explained by the amino acid compositions of each fraction. Anionic residues in fraction 2 were 4.85-fold more abundant than cationic residues without differences to other major amino acid classes and cluster at pH 4.0 (Table 9). In contrast, cationic residues in fraction 1 were only 1.26-fold more abundant than anionic residues, and the peptide clusters at pH 7.0 and 10.0 are consistent with both neutral and cationic peptides in this fraction, respectively. Regarding the glucose uptake activity of fraction 1, the absence of activity at 1 $\mu\text{g/mL}$, but stimulation at 1 ng/mL is an indication that the individual peptide concentrations could be important for the observed activity of each fraction. Roblet et al. (2016) indicated the possibility that anionic peptides inhibit glucose uptake stimulation, but the absence of activity in fraction 1 compared to the SPF control, and the recovery of its activity when diluted suggests that cationic or neutral peptides exhibit concentration-dependent inhibition. The activity of fraction 2 was slightly reduced when measured at 1 ng/mL but maintained its effect in the absence of insulin, as previously observed when recovered at alkaline pH (Roblet et al., 2016) and could indicate the concentration-dependent stimulation of glucose uptake by anionic peptides.

Regarding the amino acid composition of each fraction, database searching was performed targeting tryptic-peptides, i.e., contain Lys or Arg on the C-terminal, and as a result of the SAX stationary phase, these residues were found at double the abundance in fraction 1 than in fraction 2 (Table 9). Rather, His residues (weakly positive) were more concentrated in fraction 2 and may be an indication of the

Table 9. Total amino acid composition of fractions from SAX Separation 1.

	Abundance (%)	
	Fraction 1	Fraction 2
Gly	14.53 ± 0.21	10.63 ± 0.40
Ala	10.73 ± 0.21	9.20 ± 0.35
Met	2.83 ± 0.06	2.67 ± 0.12
Phe	4.97 ± 0.15	3.13 ± 0.12
Tyr	1.57 ± 0.00	2.63 ± 0.12
Trp	0.00 ± 0.00	0.00 ± 0.00
Val	6.80 ± 0.00	6.00 ± 0.20
Ile	5.43 ± 0.06	4.87 ± 0.15
Leu	9.30 ± 0.00	8.40 ± 0.10
Asx ^a	4.43 ± 0.23	12.13 ± 1.25
Glx ^b	6.07 ± 0.12	16.00 ± 1.04
Ser	5.47 ± 0.06	6.00 ± 0.10
Pro	8.40 ± 0.17	5.53 ± 0.32
Thr	4.50 ± 0.26	4.57 ± 0.23
Cys	0.17 ± 0.06	0.40 ± 0.00
Lys	7.97 ± 0.21	3.77 ± 0.12
His	1.67 ± 0.06	2.10 ± 0.10
Arg	5.33 ± 0.31	2.03 ± 0.15
Cationic AA ^c	13.30 ± 0.46	5.80 ± 0.26
Anionic AA ^d	10.50 ± 0.35	28.13 ± 2.29
EAA ^e	43.47 ± 0.35	35.50 ± 0.87
BCAA ^f	21.53 ± 0.06	19.27 ± 0.45
AAA ^g	6.53 ± 0.12	5.77 ± 0.23

Values are mean ± standard deviation, evaluated in triplicate; ^a Asx: D + N; ^b Glx: E + Q; ^c cationic amino acids: K, R; ^d anionic amino acids: D, E; ^e essential amino acids

(EAA): F, H, I, K, L, M, T, W, V; ^f branched chain amino acids (BCAA): I, L, V; ^g aromatic amino acids (AAA): F, W, Y.

close association of His with anionic residues in peptides generated from *Salmo salar* proteins. Residues with weakly anionic side-chains (Tyr, Thr, Cys, Ser) were slightly concentrated in fraction 2, and Asx (Asp + Asn) and Glx (Glu + Gln) abundances in fraction 2 measured over double their abundance in fraction 2. The presence of cationic and anionic amino acids in all fractions was also observed by EDFM separation of the SPF (Roblet et al., 2016), but the abundance of the cationic residue Arg was found in a higher abundance than Glu in the fraction concentrated in fractions with anionic charges, indicating that charge-based separations of peptides do not necessarily exclude residues containing opposite charges. Neither study found that the net charge of a peptide mixture could be used to predict the stimulation of glucose uptake. From both studies, anionic and cationic residues were identified in peptides of all fractions and therefore their individual importance as related to the stimulation of glucose uptake could not be evaluated, however it is also possible that the identification of peptides with fewer than 6 amino acid residues in these fractions affect these conclusions.

The extent of bioactivity exhibited by the unfractionated SPF was also different from previous SPF batches that either stimulated glucose uptake in both the presence and absence of insulin at 1 µg/mL (Chevrier et al., 2015), or had no activity at all (Roblet et al., 2016), where amino acid analyses also measured inconsistencies between SPF samples. Peptides tentatively identified in both fractions show evidence of incomplete hydrolysis during SPF production by the

identification of common peptide motifs within unique peptides in both fractions (Figure 6). Incomplete hydrolysis could also explain how hydrolysates from *Salmo salar* protein processed identically could produce both unique amino acid compositions and *in vitro* activities. Even though no peptides found in fractions 1 and 2 shared identical primary sequences, incomplete hydrolysis could represent a basis for how each fraction can exhibit similar activity, where short primary sequence motifs are fundamental for mediating the protein-protein interactions required for the propagation of cell signaling, and designing peptidomimetics of these interactions is commonly investigated for therapeutic interventions, including T2DM (Kurian et al., 2014). Additionally, the extent of proteolysis by endogenous proteases in salmon muscle or by other non-enzymatic mechanisms could also contribute to the compositional differences of SPF. Therefore, when evaluating identified peptides in each fraction, the total amino acid compositions provide only limited information, and that complete evaluation of the primary peptide sequences is necessary to identify the relationships between the individual peptide sequences of a fraction and its bioactivity.

```

Fraction_2a    AGFAGDDAPRAVF----- 13
Fraction_2b    AGFAGDDAPR----- 10
Fraction_2c    -GFAGDDAPR----- 9
Fraction_1a    AGFAGDDAPRAVFPSIVGR-- 19
Fraction_1b    AGFAGDDAPRAVFPSIVGRPR 21
Fraction_2d    ---AGDDAPR----- 7
Fraction_1c    ---AGDDAPRAVFPSIVGRPR 18
Fraction_1d    ---AGDDAPRAVFPSIVGR-- 16
Fraction_2e    ---AGDDAPRAVF----- 10
                *****

Fraction_2a    VAPEEHPTLLTEAPLN-- 16
Fraction_2b    VAPEEHPTLLTEAPLNPK 18
Fraction_1a    -----TEAPLNPK 8
Fraction_1b    -----LTEAPLNPK 9
Fraction_1c    -----LLTEAPLNPK 10
Fraction_1d    -----PTLLTEAPLNPK 12
Fraction_1e    -----HPTLLTEAPLNPK 13
                *****

Fraction_2a    DFENEMATAASSSSLEK----- 17
Fraction_1a    -----MATAASSSSLEK----- 12
Fraction_1b    -----ATAASSSSLEK----- 11
Fraction_2b    -----AASSSSLEKSYELPDGQVIT 20
Fraction_2c    -----AASSSSLEKSYELPDGQ--- 17
Fraction_1c    -----AASSSSLEK----- 9
                *****

```

Figure 6. Multiple sequence alignment using Clustal Omega software (Sievers et al., 2011), of sequences identified in fractions 1 and 2 of SAX Separation 1 containing the identical actin-derived primary sequences AGFFAPR, TEAPLN, and AASSSSLEK. Peptides containing these peptide motifs are derived from the indicated fraction.

5.1.5 Protein Identification from SAX Separation 1

The identification of progenitor proteins of SPF peptides enables the application of *in silico* tools for predicting potential bioactive peptides insofar that the proteolytic cleavage sites for proteolytic enzymes are well known. Although no identical peptides were identified in both fractions, many peptides were derived from the same salmon myofibrillar proteins and cytosolic enzymes involved in carbohydrate metabolism (Table 10). In both fractions, actin possessed the highest Sequest HT score and produced ~ 20 % of all identified peptides. Tropomyosin, beta-enolase, fructose-bisphosphate aldolase and glyceraldehyde-3-phosphate dehydrogenase were also highly represented by identified sequences.

Only two peptides from a salmon myosin heavy chain (MHC) fragment (20 kDa) were identified, even though myofibrillar proteins are the most abundant protein class in *Salmo salar* muscle tissues and where the MHC represents a significant proportion (Tahergorabi et al., 2012). The low sequence coverage as reported for each identified protein could be reflected by the omission of peptides < 6 AA by database searching that are expected to be the most abundant in the SPF based on its processing methodology. However, due to the theoretical abundance of the MHC in fish muscle tissue, it would be expected that a greater representation of this sequence would be detected from peptides identified in the SPF. At this point in time, it is unknown whether the MHC is recovered during isoelectric precipitation during SPF processing.

Table 10. Parent proteins of identified peptides in fractions of SAX Separation 1, using Proteome Discoverer Version 2.1 and Sequest HT scoring.

Fraction	Accession	Protein	Sequest HT Score	Sequence Coverage (%)	Unique Peptides (No.)
1	Q78BU2	Actin alpha 1-1	399.35	46.15	44
	B5DGM7	Fructose-bisphosphate aldolase A	369.55	23.42	17
	Q91472	Fast myotomal muscle tropomyosin	138.04	51.76	22
	B5DGP0	Creatine kinase-2	97.34	23.62	6
	B5DGQ7	Beta-enolase	92.70	30.18	16
	B5DGG5	Creatine kinase	75.35	13.42	1
	B5DGR5	Glyceraldehyde-3-phosphate dehydrogenase	45.67	27.69	12
	cRAP_P00766	Chymotrypsinogen A	37.90	18.37	9
	B5DG55	Phosphorylase	34.39	11.97	11
	Q7ZZN0	Myosin regulatory light chain 2	32.59	23.53	4
	B5DFX8	Phosphoglycerate kinase	23.61	6.24	2
	B5XH68	Triosephosphate isomerase	21.51	8.94	2
	B9EP57	Troponin C, skeletal muscle	16.51	24.38	3
	B5DGU1	Pyruvate kinase	14.94	10.38	5
	B5DG72	Phosphoglucomutase 1	10.55	2.67	2
	B5DH15	Parvalbumin beta 1	10.55	30.28	3
	2	B5DG40	Fast myotomal muscle actin 2	814.67	36.07
B5DGR2		Glyceraldehyde-3-phosphate dehydrogenase	586.06	42.22	32
B5DGM7		Fructose-bisphosphate aldolase A	137.53	31.68	6
Q2HXU3		Slow myosin heavy chain 1 (Fragment)	129.32	3.95	2
I0J1J3		Fructose-bisphosphate aldolase	99.29	30.85	3
B5DGQ7		Beta-enolase	89.87	13.82	13
B5DGT1		Myosin, light polypeptide 3-1	78.73	23.83	9

Fraction	Accession	Protein	Sequest HT Score	Sequence Coverage (%)	Unique Peptides (No.)
2, con't	B5DGT2	Myosin light chain 3, skeletal muscle isoform	65.95	20.50	7
	B5DGU1	Pyruvate kinase	59.10	9.25	11
	cRAP_P00766	Chymotrypsinogen A	54.66	19.18	11
	B5XH68	Triosephosphate isomerase	41.16	24.26	9
	Q7ZZN0	Myosin regulatory light chain 2	38.83	24.71	8
	cRAP_P00760	Cationic trypsin	37.80	17.48	10
	B5DGP1	Creatine kinase-2	30.54	17.05	1
	B5DGG5	Creatine kinase	29.39	18.16	2
	B5DG39	L-lactate dehydrogenase	28.57	20.18	7
	B8XA43	Collagen 1a (Fragment)	20.95	10.80	4
	cRAP_P06732	Creatine kinase M-type	19.97	8.66	1
	B5DGZ9	Fast myotomal muscle troponin-T-2	19.49	8.66	3
	B5DG55	Phosphorylase	16.81	5.33	4
	B5DG72	Phosphoglucomutase 1	14.74	6.77	3
	B5DG45	Myosin binding protein H-like	14.37	9.09	4
	B5X293	Lumican	13.10	3.22	3
	B5XAW0	Nucleoside diphosphate kinase	11.70	6.62	1
	B5DG78	ATP synthase subunit alpha	10.37	5.11	3

5.1.6 Conclusion

In Separation 1, the SPF was separated into fractions by SAX that contained anionic peptide or cationic and neutral peptides and stimulated glucose uptake in the absence of insulin at 1 ng/mL. The activity of anionic peptides selected under alkaline elution conditions supported previous studies, but the stimulating activity from a fraction containing cationic and neutral peptides suggested that inhibiting effects at higher concentrations may mask the stimulating effects exhibited at lower concentrations. Only peptides > 6 amino acids in length were identified resulting from limitations of the LC-MS methodology, preventing the identification of the short peptides thought to be bioactive. These results support how the bioactivities of peptides in SPF, and likely in other hydrolysate subfractions, originate from dynamic interactions of various peptide components and collectively influence the rate of glucose uptake in cultured L6 myotubes.

5.1.7 Future Work

The separation of SPF by SAX column chromatography produced active fractions. At present, it is not known whether the observed bioactivities were due to peptides with more or less than 6 amino acid residues, however sequence determinations of the larger peptides suggested highly complex peptide compositions. These fractions represent ideal candidates for subsequent orthogonal separations, however the potential for identifying bioactive peptides remains more advantageous when separations in the first dimension generate fractions with fewer peptides than observed in the present study. Additional SPF separations to

improve the resolution of individual peptide recovery during their elution from the column and the collection of additional fractions should be applied to future separations to ensure that prospective bioactive peptides can be selected from within the fraction they are identified.

5.2 SAX Separation 2 and GF Separation 3 of SPF

5.2.1 Objectives

Separation 1 (SAX, 10 x 100 mm column) demonstrated that SPF peptides were generated from many identifiable salmon proteins. LC-MS was used to further identify at least 500 peptides from the fractions collected from the SAX column. The column resolution and operational parameters of SAX Separation 1 were deemed unsuitable given this complexity, in addition to the finding that peptide identification could not reveal the complete composition of small peptides (< 500 Da) previously described as abundant components of the SPF (Roblet et al., 2016).

SPF separation by SAX was therefore repeated, but under conditions designed to increase the resolution of peptides eluting during the gradient elution, so that the analysis and designation of potential bioactive peptides in each fraction would perhaps be more accurate.

Additionally, SPF separation by GF was also investigated to provide an experimental indication for the presence and importance of peptides < 6 amino

acids and evaluate the importance of peptide chain length on glucose uptake stimulating ability.

5.2.2 Chromatography of SAX Separation 2 and GF Separation 3

SAX. Separation 2 of the SPF was performed using UNOsphere Q media and a 10 x 300 mm column with UV detection at 214 nm, 254 nm and 280 nm as shown in Figure 7. The SPF used for these separations was produced from a different batch than in Separation 1, and the longer column enabled greater partitioning of peptides, as indicated by the peaks detected at 254 nm and 280 nm during the gradient elution thus allowing for more peptide fractions to be collected. In addition, Separation 2 was carried out using a totally volatile buffer system consisting of only 25 – 1000 mM ammonium formate with no added NaCl for gradient elution.

For Separation 2, a broad peak at ~ 50 min was detected at 215 nm, but a rising baseline from eluant B beginning at an elution time of about 60 min occluded the peptide peaks. Detection at higher wavelengths (254 nm and 280 nm) was not impacted by the gradient elution buffer and the relative absorbance of the 50 min peak to peaks at 125 min, 145 min, and 160 min suggest an unequal distribution of aromatic amino acids, purines or pyrimidines between isocratic and gradient stages, a result that was not observed in Separation 1. Absorbance intensity began to decline at ~ 175 min at all wavelengths at an eluant B concentration of 400 mM ammonium formate.

GF. Separation 3 of the SPF was performed on Bio-Gel P-2 media (100 – 1800 Da exclusion limit) with UV detection at 215 nm, 254 nm and 280 nm as shown in Figure 8. The detection artifact at 215 nm between 15 and 25 min indicates overloading of the UV detector flow cell but was not observed when separated by SAX Separation 2 using identical sample loading parameters. The large peaks at 254 nm and 280 nm eluting with the void volume at 15 min also reached an absorbance nearly 2-fold the intensity produced at the same wavelengths when separated by SAX Separation 2. Together, these findings suggest that the SPF was concentrated in peptides at or above the maximum exclusion limit of Bio-Gel P-2 media of 1800 Da. Throughout the separation, each wavelength exhibited a unique absorbance pattern during both the flowthrough and gradient elution stages, and fractions were recovered from broad peaks.

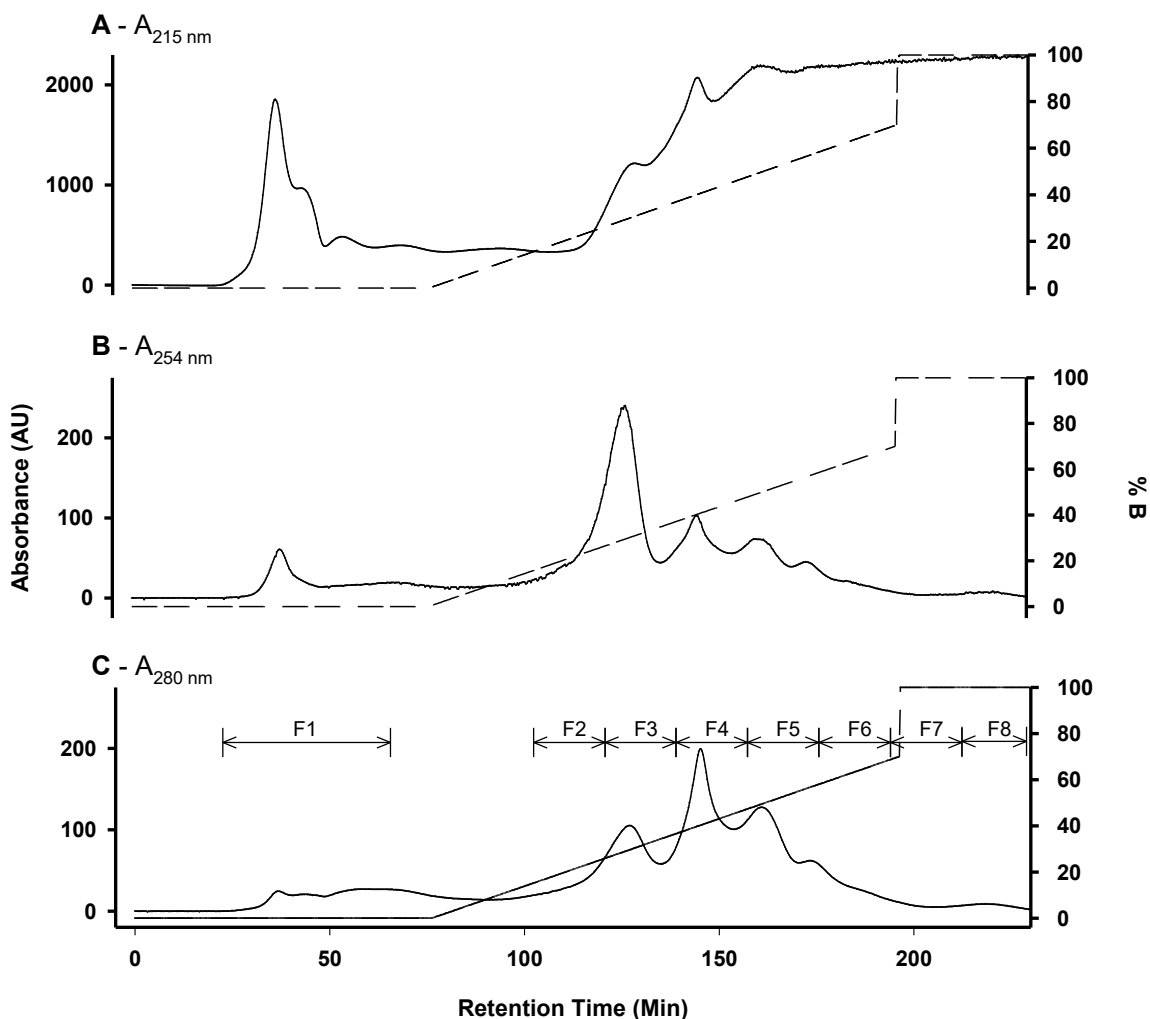


Figure 7. SAX Separation 2 of SPF on UNOsphere Q media; 10 mm x 300 mm column; flow rate, 0.5 mL/min; eluant A, 25 mM ammonium formate pH 9.0; eluant B, 1.0 M ammonium formate; gradient, 0 – 70 %; detection at (A) 215 nm, (B) 254 nm and (C) 280 nm; sample concentration, 50 mg/mL, injection volume, 500 μ L. Solid lines represent the UV absorbance and dotted lines represent the percent of elution buffer. Fractions were collected within the indicated boundaries and pooled from n = 6 separations.

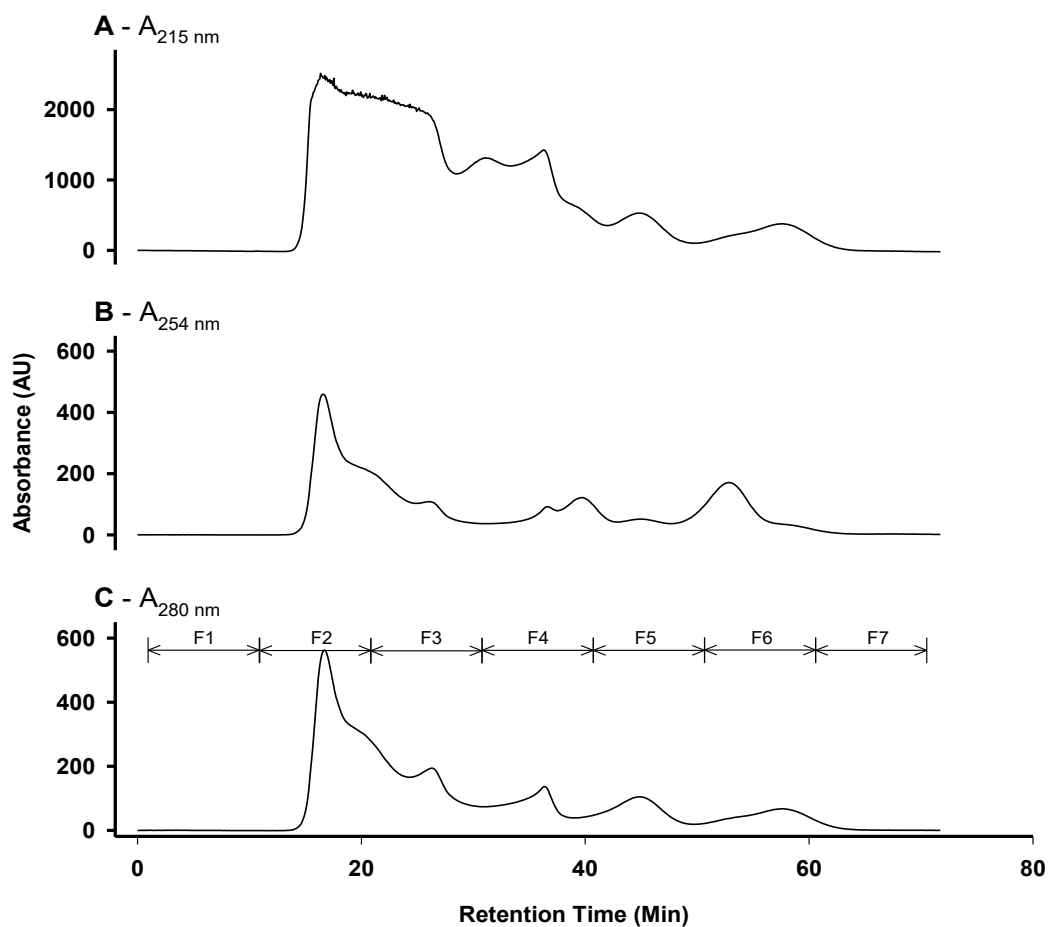


Figure 8. GF Separation 3 of SPF on Bio-Gel P-2 media; 10 mm x 300 mm column; flow rate, 0.5 mL/min; eluant, 100 mM ammonium formate, 0.5 % formic acid, pH 8.5; detection at (A) 215 nm, (B) 254 nm and (C) 280 nm; sample concentration, 50 mg/mL; injection volume, 500 μ L. Lines represent the UV absorbance at the corresponding wavelength. Fractions were collected within the indicated boundaries and pooled from n = 6 separations.

5.2.3 Glucose Uptake Analysis of Fractions from SAX Separation 2 and GF Separation 3 of SPF

SAX. The glucose uptake modulation of fractions from SAX Separation 2 of SPF was screened in cultured L6 myotubes using a peptide concentration of 1 ng/mL (Table 11). Only the unfractionated SPF and fraction 8 had stimulating effects on glucose uptake in the absence of insulin (p -value = 0.021, p -value = 0.0290, respectively). Fraction 8 was collected during the final column wash phase when the desorption of highly anionic compounds is promoted, but after UV absorbance at 254 nm and 280 nm already recovered to baseline suggesting that the peptide concentration in this fraction is low. Major peaks at 125, 145 and 160 min in Separation 2 correspond to fractions 3, 4 and 5, respectively, but were not bioactive having no significant effects on glucose uptake. In fraction 1, low $A_{254\text{ nm}}$ by peptides eluting during the isocratic elution and its absence of glucose uptake stimulating activity is in contrast to observations of fraction 1 from SAX Separation 1 that recorded both high absorbance at 254 nm and glucose uptake stimulation. Therefore, the lack of activity in fraction 1 from SAX Separation 2 could be related to either the paucity of peptides in this fraction with cationic and/or neutral peptides containing Phe, or the presence of purine and pyrimidine-containing compounds, which may mediate the functionality of this fraction in Separation 1.

GF. No fractions from Separation 3 were able to affect glucose uptake (Table 12). Individual fractions appeared to weakly modulate glucose uptake in the absence and presence of insulin, but was not concentrated significantly to any one fraction, and therefore the influence of peptide MW on bioactivity could not be evaluated.

Table 11. Glucose uptake modulation of SPF and its fractions from SAX Separation 2 following the methodology outlined in Section 3.3. Fractions were each tested at 1 ng/mL in the (-) absence and (+) presence of insulin.

Sample	Glucose Uptake (fold activity to control)	
	Insulin (-)	Insulin (+)
Control	1.00 ± 0.00	2.00 ± 0.08
SPF	1.07 ± 0.03*	1.86 ± 0.10
Fraction 1	1.04 ± 0.06	1.97 ± 0.16
Fraction 2	1.16 ± 0.09	2.19 ± 0.15
Fraction 3	1.03 ± 0.06	1.99 ± 0.14
Fraction 4	1.14 ± 0.12	2.03 ± 0.15
Fraction 5	1.09 ± 0.14	2.19 ± 0.20
Fraction 6	0.90 ± 0.05	1.86 ± 0.08
Fraction 7	1.04 ± 0.04	1.99 ± 0.11
Fraction 8	1.19 ± 0.07*	2.16 ± 0.17

Values are means ± SEM, from n = 6 individual experiments, performed in triplicate;
 * p-value < 0.05 vs. control.

Table 12. Glucose uptake modulation of SPF and its fractions from GF Separation 3 following the methodology outlined in Section 3.3. Fractions were each tested at 1 ng/mL in the (-) absence and (+) presence of insulin.

Sample	Glucose Uptake (fold activity to control)	
	Insulin (-)	Insulin (+)
Control	1.00 ± 0.00	2.00 ± 0.08
SPF	1.07 ± 0.03*	1.86 ± 0.10
Fraction 1	1.07 ± 0.07	1.99 ± 0.19
Fraction 2	1.04 ± 0.07	2.12 ± 0.24
Fraction 3	0.98 ± 0.03	1.87 ± 0.14
Fraction 4	0.99 ± 0.09	1.92 ± 0.17
Fraction 5	1.20 ± 0.09	2.20 ± 0.20
Fraction 6	1.08 ± 0.06	2.02 ± 0.20
Fraction 7	0.99 ± 0.01	1.76 ± 0.11

Values are means ± SEM, from n = 6 individual experiments, performed in triplicate;

* p-value < 0.05 vs. control.

5.2.4 Peptide Identification in Fractions from SAX Separation 2 and GF Separation 3

One fraction from SAX Separation 2 exhibited modulating effects on glucose uptake, while fractions from the GF Separation 3 did not exhibit modulating effects. Although different column lengths were used for the SAX separations (1 and 2) and perhaps because an NaCl gradient was used only in Separation 1 but not Separation 2, the elution patterns were quite different; there was a lack of consistency regarding the pattern and intensity of $A_{254 \text{ nm}}$ during the isocratic elution indicating a variation in the compositions of the SPF used for each separation. Given that the SPFs were prepared from different batches, and similar inconsistency regarding the SPF composition was suggested by the apparent concentration of peptides at or above the maximum exclusion limit of 1800 Da by Bio-Gel P-2 media in Separation 3, it was assumed that the unfractionated SPF used for both Separations 2 and 3 may not have been representative of the reference SPF outlined by Jin (2012). Upon further investigation, it was revealed that both pepsin and trypsin:chymotrypsin enzymes used for enzymatic hydrolysis were past their expiration dates and therefore SPF was considered to have not been hydrolyzed to the same extent had fresh enzymes been used, although the specific activities of these enzymes were not directly evaluated to verify. Reduced enzymatic efficiency could explain the apparent high concentration of high MW, incompletely digested peptides eluting during the void volume of Separation 3 and therefore could also mediate the dissimilar absorbance patterns in the isocratic stages of SAX Separations 1 and 2. As a result of these inconsistencies, it was

decided not to analyze fractions from Separations 2 and 3 by LC-MS, but to prepare a new SPF and repeat these separations.

5.2.5 Conclusion

The absorbance pattern for SAX Separations 2 demonstrated an improvement to peptide resolution compared to SAX Separation 1 and therefore increased the opportunity to recover peptide fractions with lower complexity. A similar opportunity to recover peptide fractions of distinct MW was presented during GF Separation 3, but a relationship between peptide mass or charge and glucose uptake stimulating activity could not be determined. However, reduced enzymatic efficiency from expired enzymes may have contributed to the apparent compositional difference of the SPF from Separations 2 and 3 to Separation 1, and suggests that the SPF was likely not processed to the appropriate standard.

5.2.6 Future Work

These results reveal the importance of establishing critical control points such as digestion conditions and time/temperature factors during SPF processing to ensure consistency from batch-to-batch, especially if the process is intended for pilot- or commercial-scale processing.

5.3 Alkaline Solubilization of Proteins from Salmon Muscle Tissue

Protein extractions from fish muscle tissues are commonly performed using alkaline solubilization followed by isoelectric precipitation (IEP) methods. The conditions of IEP dictate the composition of the final digest after proteolytic digestion since the progenitor proteins to generated peptides are diverse with many different isoelectric points. Fish myofibrillar proteins are considered to have a pI of 5.5 on average (Kristinsson and Rasco, 2000), but at high ionic strength, the average pI shifts in the acidic direction (Chen and Jaczynski, 2007). Exposure of proteins to highly alkaline environments is also associated with numerous secondary reactions that can modify their primary sequences (Whitaker et al., 1983). In the development of the SPF methodology, the finding of variable bioactivities from hydrolysates developed from protocols 2 and 3 (Table 3) using identical enzymatic treatments but different IEP methods could indicate that the unique composition of proteins recovered by each IEP method represents a factor mediating these bioactivity variations. Proteins precipitated during IEP are progenitors to potential bioactive peptides. A further understanding of progenitor protein precipitated during SPF processing may therefore affect the final SPF peptide composition and may well be expected to affect the bioactivity of the final product.

5.3.1 Objective

The objective of this chapter was to investigate the protein composition of precipitates recovered from Atlantic salmon muscle by conditions of alkaline

solubilization and IEP, and to reveal how their differences affect the generation and identification of potential bioactive peptides.

5.3.2 Evaluation of Protein Recovery from Salmon Muscle

The original methodology for SPF processing (Jin, 2012) utilizes a 1.0 M NaOH solution to facilitate salmon protein solubilization, but alkaline solubilizations are typically controlled by pH adjustment. Therefore, protein extracts prepared by IEP were used to determine the optimal recovery of Atlantic salmon protein by pH-controlled alkaline solubilization (as compared to alkali concentration-dependent solubilization) as outlined in Section 3.7 (Figure 9). The pH adjustment of solutions with 2.0 M NaOH to pH 12, 12.5 and 13, corresponded to final NaOH concentrations of ~ 0.05 M, 0.075 M and 0.25 M, respectively. Protocol 2 from Jin (2012) used 1.0 M NaOH solution to solubilize the proteins in minced salmon tissue at a v/v ratio of 1:4 (mince:NaOH) and the pH of the final mixture was ~ 13.1. The highest protein recovery calculated at > 95 % was achieved when solubilized in either pH 12 or 12.5 solutions and precipitated to pH 4.5 (Figure 9). Both conditions had greater recovery (5 – 10 %) when precipitated at pH 4.5 compared to 5.5, but not when solubilized from the pH 13 solution.

The final conductivities of each isoelectric precipitate solution prepared above were not measured, but the stoichiometric proportionality of NaOH neutralization by HCl could suggest that protein recovery from pH 13 solutions was influenced by the formation of much more NaCl, increasing the ionic strength that also affects

the protein solubility in addition to pH differences. Total nitrogen contents of each solution were not measured so that overall protein yields by each pH combination were not accurately determined, but estimates of protein concentration using the colourimetric BCA assay suggest that factors other than pH mediate protein recovery from highly alkaline solutions. The protein isolate produced using the 12.5 / 4.5 pH combination demonstrated the highest recovery of soluble salmon muscle proteins when the influence of IS is relatively low, suggesting that perhaps the SPF process could be improved to reduce the need for as much de-salting currently being carried out by conventional electrodialysis or nanofiltration.

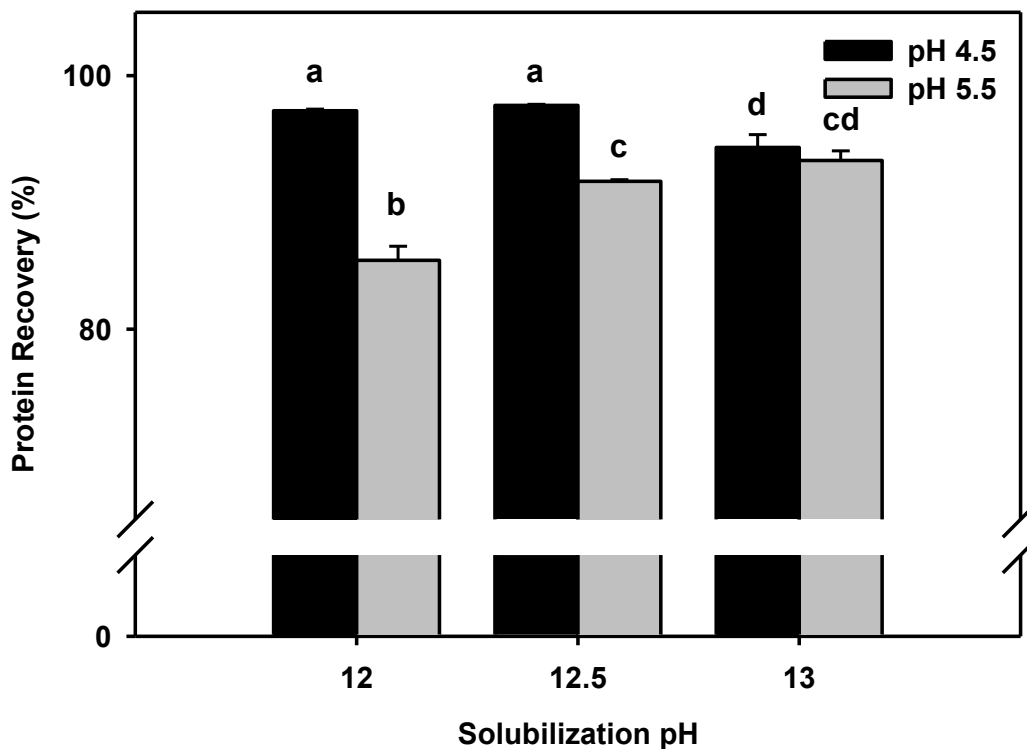


Figure 9. Protein recovery (%) of alkali-solubilized *Salmo salar* muscle precipitated by IEP using various pH combinations, as outlined in Section 3.7. Different lowercase letters indicate significant differences (least squared difference test, p -value < 0.05, $n = 2$, measured in duplicate), between means.

Solubility curves of a high-alkali solubilized salmon muscle (Figure 10A) prepared using a 1.0 M NaOH solution as a representation of the SPF processing methodology, and a low-alkali solubilized salmon muscle (Figure 10B) prepared using a 0.1 M NaOH solution as an approximation of the pH 12.5 / 4.5 solution seen in Figure 9, were compared. Solubility curves were prepared to evaluate the protein recoveries by each solution during the progression of IEP, after their TCA precipitation and centrifugation. The isoelectrically precipitated proteins from the high-alkali solubilized salmon muscle recovered a maximum of ~ 71 % of the available protein solubilized after 2 h agitation in 1.0 M NaOH compared to the low-alkali solubilized salmon muscle that recovered ~ 97.5 %. *Post hoc* LSD tests of the high-alkali condition revealed that protein precipitation to pH 4.5 was more effective ($p < 0.05$) than precipitation to pH 5.0, but not to pH 4.0 or lower ($p > 0.05$). This demonstrated the optimization of IEP following the SPF processing methodology involving alkaline solubilization in 1.0 M NaOH. LSD tests of the low-alkali condition reveal that precipitation to pH 4.5 was equally efficient ($p > 0.05$) to precipitation within the range of pH 3.5 - 5.5.

Protein recovery using high-alkaline solubilization with IEP to pH 4.5 was associated with ~ 29 % protein loss, however Jin (2012) compared the bioactivity of samples prepared from solubilization using both 0.1 M (protocol 1) and 1.0 M (protocol 2) NaOH and found that high-alkali solubilization was superior, although different enzyme treatments were used.

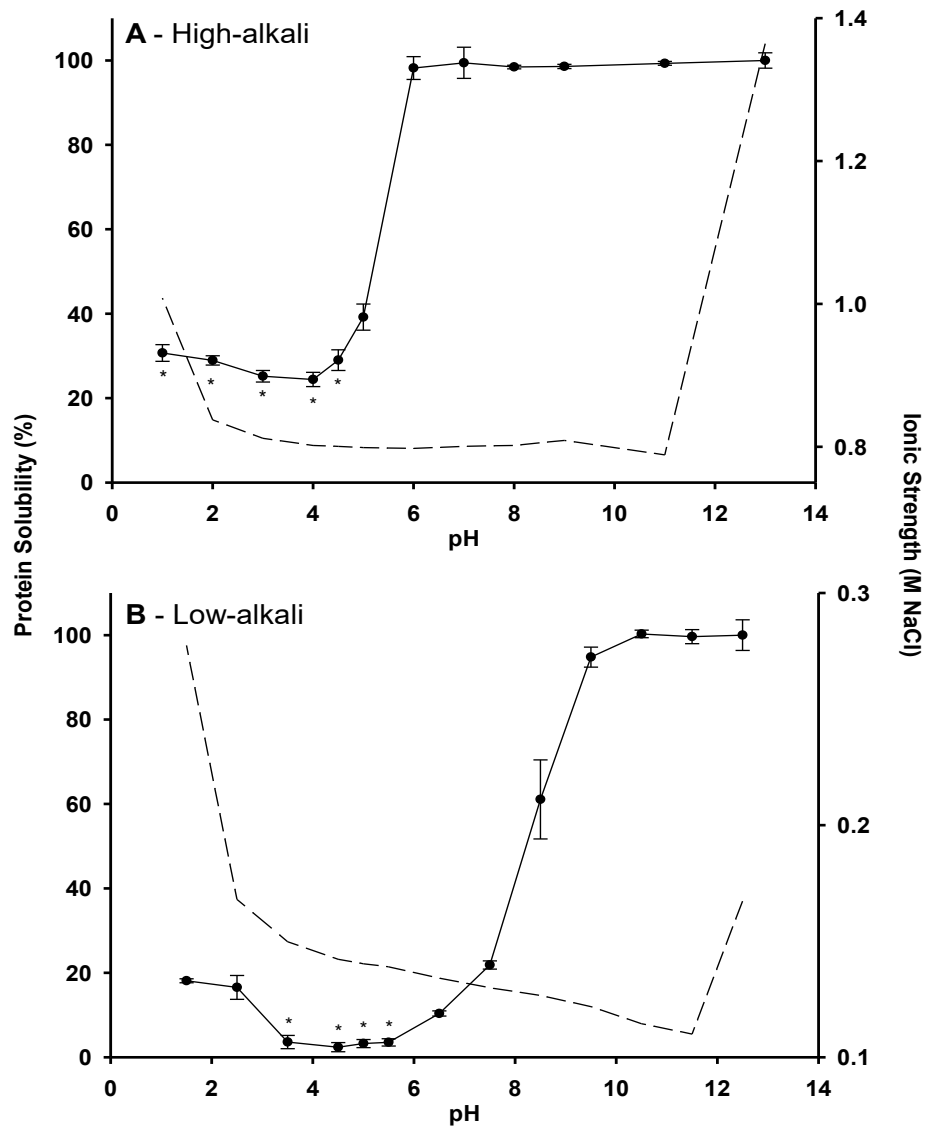


Figure 10. Solubility curves of the (A) high-alkali solubilized salmon muscle and (B) low-alkali solubilized salmon muscle, outlined in Section 3.8. Solid lines represent protein solubility (left y-axis), and dashed lines represent ionic strength calculations based on the measured conductivity of each solution (right y-axis). * samples not significantly different (least squared difference test, $p > 0.05$, $n = 3$).

In the present study, the bioactivity of samples prepared from high (1.0 M NaOH) and low (0.1 M NaOH)-alkali solubilization and processed identically following Jin (2012) protocol 2 confirmed the superiority of high-alkali solubilization of salmon muscle tissue on bioactivity (Table 13). The high-alkali derived peptides are representative of the SPF, and stimulated glucose uptake in the absence of insulin at both 1 $\mu\text{g}/\text{mL}$ (p-value < 0.001) and 1 ng/mL (p-value < 0.001) concentrations, as well as in the presence of insulin at 1 $\mu\text{g}/\text{mL}$ (p-value = 0.0014) and 1 ng/mL (p-value = 0.0017) concentrations. The sample produced from the low-alkali solubilized salmon muscle precipitate had no modulating effect on glucose uptake at 1 ng/mL , and even inhibited glucose uptake at 1 $\mu\text{g}/\text{mL}$ in both the absence and presence of insulin (p-value < 0.001, p-value = 0.0113, respectively, Table 13).

The bioactivities exhibited by LMW peptides from each sample, when considered with findings from the previous section, suggest that perhaps differences between methodologies could be factors leading to their unique activities. Certainly, it is clear that salt removal at the end of SPF processing would be far less challenging at the pilot or industrial scale if less NaOH were used at the protein solubilization step. Alkali-mediated chemical reactions of proteins have been previously described including their hydrolysis, deamination, the conversion of arginine to ornithine and/or citrulline, the removal of post-translational modifications, racemization and beta-elimination reactions targeting cysteine, threonine and serine (Whitaker et al., 1983). These reactions, and others leading to structural modifications such as cyclization (Goodman et al., 2001), have been demonstrated

to contribute enhancements to the bioactivity of peptidomimetics, therefore evaluating the consequences of NaOH exposure on salmon proteins during alkaline solubilization followed by the extent and potential impact of these modifications on the functionality of LMW hydrolysates may realize novel bioactive compounds.

Table 13. Glucose uptake modulation hydrolysate samples processed following Jin (2012) protocol 2 using high- or low-alkali solubilized salmon muscle precipitates, outlined in Section 3.3. Fractions were each tested in the (-) absence and (+) presence of insulin.

Sample	Concentration	Glucose Uptake (fold activity to control)	
		Insulin (-)	Insulin (+)
Control	-	1.00 ± 0.00	1.80 ± 0.04
High-alkali SPF	1 ng/mL	1.27 ± 0.04*	2.20 ± 0.10*
	1 µg/mL	1.27 ± 0.03*	2.23 ± 0.10*
Control	-	1.00 ± 0.00	1.99 ± 0.03
Low-alkali SPF	1 ng/mL	0.94 ± 0.03	2.03 ± 0.13
	1 µg/mL	0.84 ± 0.03*	1.69 ± 0.10*

Values are means ± SEM, from n = 6 - 8 individual experiments, performed in triplicate;
* p-value < 0.05 vs. control.

Various fish myofibrillar proteins can be selectively solubilized at ionic strengths (IS) above or below the value of 0.6 (Stefansson and Hultin, 1994). To this end, Jin (2012) and the present study both attempted to selectively solubilize salmon proteins at low IS but produced peptides with less bioactivity. Solutions of high-alkali solubilized salmon proteins precipitated to pH 4.5 were calculated with IS of 0.8, compared to solutions of low-alkali salmon protein precipitated to pH 4.5 that were measured with an IS of 0.14, and distributing either methodology on both sides of the boundary for IS-mediated solubility of myofibrillar proteins, suggesting that conductivity-mediated protein selection could lead to variations of the protein content in each precipitate. In addition to facilitating protein solubilization from muscle tissues, high molarities of NaOH could also promote the non-specific hydrolysis of salmon proteins through an alkaline hydrolysis (AH) mechanism (Whitaker et al., 1983), so these possible variations in protein composition from each protein precipitate were evaluated qualitatively using SDS-PAGE.

The high-alkali solubilized proteins recovered by isoelectric precipitation to pH 4.5 were resolved on stained SDS gels as a large smear in the low mass range without individual band resolution (Figure 12A, Lane 2). The smearing could be due to partial alkaline hydrolysis (AH) of the constituent proteins to produce low molecular weight constituents, whereas the low-alkali solubilized protein was clearly resolved as individual bands (Figure 11A, Lane 3). Although AH has not been explicitly described in the literature to result from alkaline solubilization of fish proteins in 1.0 M NaOH conditions, the concentrations of NaOH used in each sample could also

have affected the binding interaction of SDS with salmon proteins during the sample preparation.

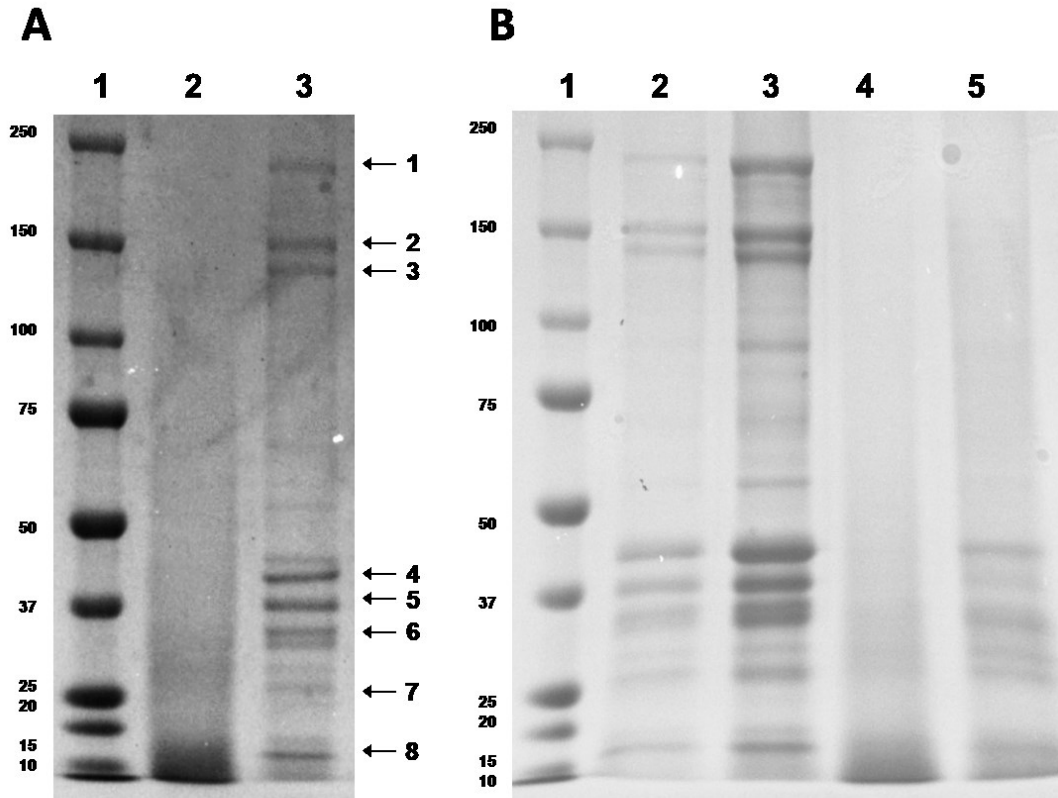


Figure 11. SDS-PAGE showing the protein distributions of (A) high- and low-alkali solubilized salmon muscle proteins. Lane 1: MW ladder (kDa), Lane 2: high-alkali solubilized salmon protein precipitate, Lane 3: low-alkali solubilized salmon protein precipitate; numbered bands refer to samples for experiments in Section 5.3.3. (B) high- and low-alkali solubilized salmon prepared from whole loin or mechanically deboned frame meat, following the methodology in Section 3.10. Lane 1: MW ladder, Lane 2: low-alkali solubilized protein from deboned tissue, Lane 3: low-alkali solubilized protein from loin, Lane 4: high-alkali solubilized protein from deboned tissue, Lane 5: high-alkali solubilized protein from loin. Gels were prepared as a gradient from 5 – 15 % acrylamide.

The protein distributions of low-alkali solubilized muscle protein were nearly identical when prepared from both frozen deboned salmon mince (Figure 11B, Lane 2) and whole salmon loin (Figure 11B, Lane 3), whereas the high-alkali solubilized muscle protein maintained strong background absorbance from deboned mince (Figure 11B, Lane 4), and modestly recovered the resolution of individual protein bands in the low mass range (< 40 kDa) when using whole salmon loin (Figure 11B, Lane 5).

High molecular weight (HMW) proteins (bands at 220, 150, 130 kDa) were not detected in salmon proteins solubilized in 1.0 M NaOH prepared from either intact salmon loin muscle or mechanically deboned salmon frame mince (Figure 11B – Lane 4 and 5) but were detected in both protein precipitates prepared using the low-alkali solubilization methodology (Figure 11B – Lane 2 and 3), suggesting that high-alkali solubilization extensively breaks down high MW proteins and to a lesser extent with low-alkali solubilization. The protein bands at 220 kDa from fish muscle are tentatively identified as the myosin heavy chain that represents > 40 % of total fish protein. Therefore, determining its presence or absence as a progenitor to SPF peptides is important for the peptide identification process. Within each method, the protein distributions prepared from deboned mince were of noticeably lower intensity than from whole salmon loin that may be indicative of non-specific proteolysis by endogenous proteases during the deboning process or in subsequent frozen storage. Furthermore, intact proteins bands < 50 kDa were detected when the salmon loins were processed following low-alkali methodology

(Figure 11B – Lane 5), but absent when processed using the high-alkali methodology. These results suggest that differences of each methodology affect the breakdown of intact HMW protein, and where the activity of endogenous proteases may contribute to these results but were not directly evaluated.

The influence of alkaline hydrolysis on the recovery of HMW proteins during protein solubilization was investigated by preparing both the high and low-alkali solubilized salmon protein precipitates, as well as others solubilized using intermediate NaOH concentrations, from whole salmon fillets. Following IEP, the TCA-soluble (supernatant) and insoluble (pellet) fractions of each preparation were obtained by centrifugation, and SDS-PAGE was used to separate the proteins within each fraction (Figure 12). Both the soluble (Figure 12A – Lane 2) and insoluble (Figure 12B – Lane 2) samples produced strong background staining in the presence of Coomassie Blue R-250 but individual protein bands < 50 kDa were also detected; similar bands were observed in Lane 5 of Figure 11B. Specifically, the loss of HMW protein staining and the gradual gel smearing suggests that protein breakdown was related to the molarity of NaOH used to solubilize the salmon muscle. Furthermore, HMW proteins were identified in both soluble and insoluble protein fractions of the low-alkali solubilized salmon muscle (Figure 12A,B, Lane 5), revealing that HMW proteins may exhibit partial solubility even at the lowest evaluated conductivity, and that differences in conductivity between high- and low-alkali salmon protein precipitate solutions are unlikely to explain the apparent absence of HMW proteins in high-alkali solubilized muscle protein precipitates.

Rather, evidence suggesting the progressive loss (Figure 12A – Box 1 and 2) and development (Figure 12B – Box 3) of band intensity in precipitates from 1.0 M to 0.1 M NaOH solutions could support a hydrolytic mechanism through alkaline hydrolysis, or through selective solubilization or precipitation of proteins with

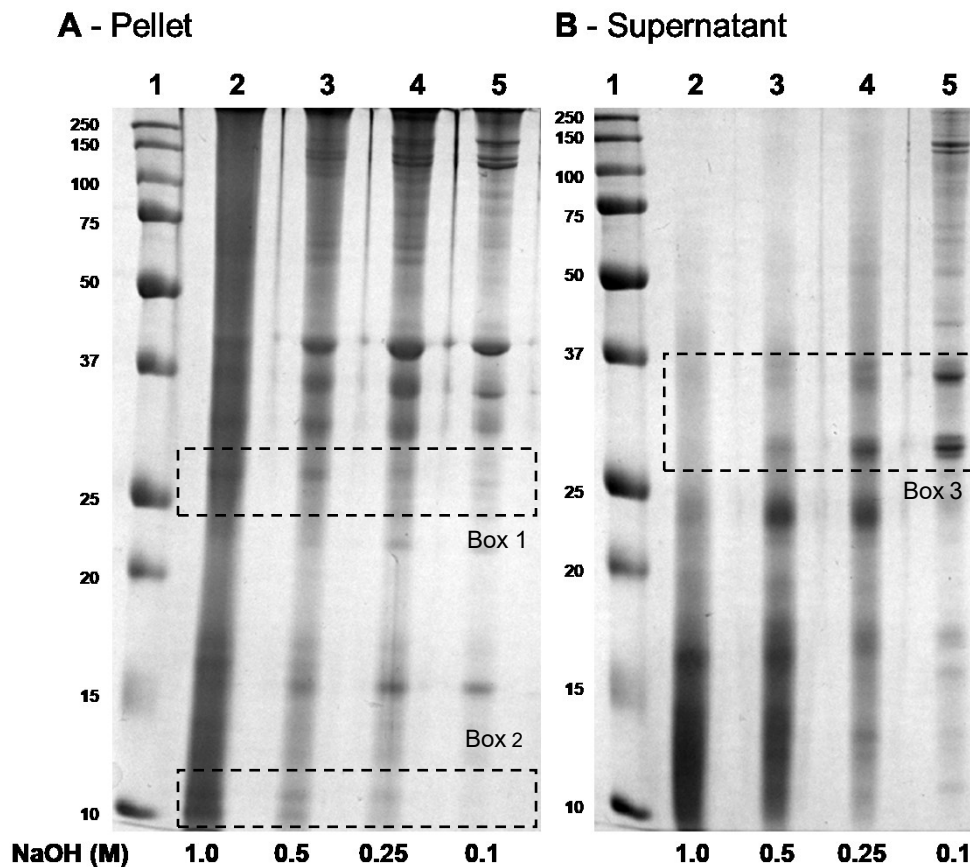


Figure 12. SDS-PAGE showing the protein distributions of (A) insoluble and (B) soluble muscle proteins, in isoelectric precipitates from Atlantic salmon muscle tissue solubilized using NaOH concentrations between 1.0 and 0.1 M, following the methodology outlined in Section 3.10. Gels were prepared to a final concentration of 15% acrylamide. Lane 1 of panel A and B: MW ladder (kDa). Boxes 1, 2 and 3: proteins showing progressive changes in protein distribution.

unique characteristics, but the identification of these bands was not performed to confirm their generation by these mechanisms.

The influence of alkaline (chemical) hydrolysis was further investigated by evaluating the band distribution from protein extracts recovered by both high- and low-alkali solubilized salmon muscle methodologies at time intervals throughout the 2 h solubilization period (Figure 13). Changes to the band distribution and staining intensity of the high-alkali solubilized salmon muscle precipitates were noticeable within the first 0.5 h (Figure 13A), while no changes in the extent of background staining or band intensity were observed in the low-alkali samples throughout the duration of the 2 h solubilization period (Figure 13B). The band distribution and presence of HMW proteins of the high-alkali solubilized salmon muscle at 0 h were identical to low-alkali samples from all intervals, confirming that conductivity-mediated differences in protein selection do not affect the recovery of HMW in the high-alkali solubilized salmon muscle protein extract. Rather, the progressive change in the high-alkali samples support a time-dependent mechanism. Evidence for the concentration- and time-dependent loss of band resolution and increase to background absorbance with high-alkali solubilized salmon muscle by isoelectric precipitation strongly supports the occurrence of alkaline hydrolysis during SPF processing.

The preparation of Atlantic salmon protein precipitates using increasing concentrations of NaOH for tissue solubilization appears to be associated with secondary reactions that affect the primary sequences of target substrates, but the

extent and type of reactions have not been evaluated. Importantly, the influence of these primary sequence modifications on the bioactivity of SPF remains unknown but introduces significant challenges for applying *in silico* digestion tools to facilitate peptide identification. These tools are essential for both conventional and bioinformatic approaches of bioactive peptide identification, where a requirement for accurately translating the processing of primary protein sequences is necessary for a non-biased peptide identification. Furthermore, the hydroxide ion (OH^-) has been associated with influencing many other chemical modifications to primary protein sequence that would also affect the function and identification of peptides in the SPF and could be important to characterize in future work.

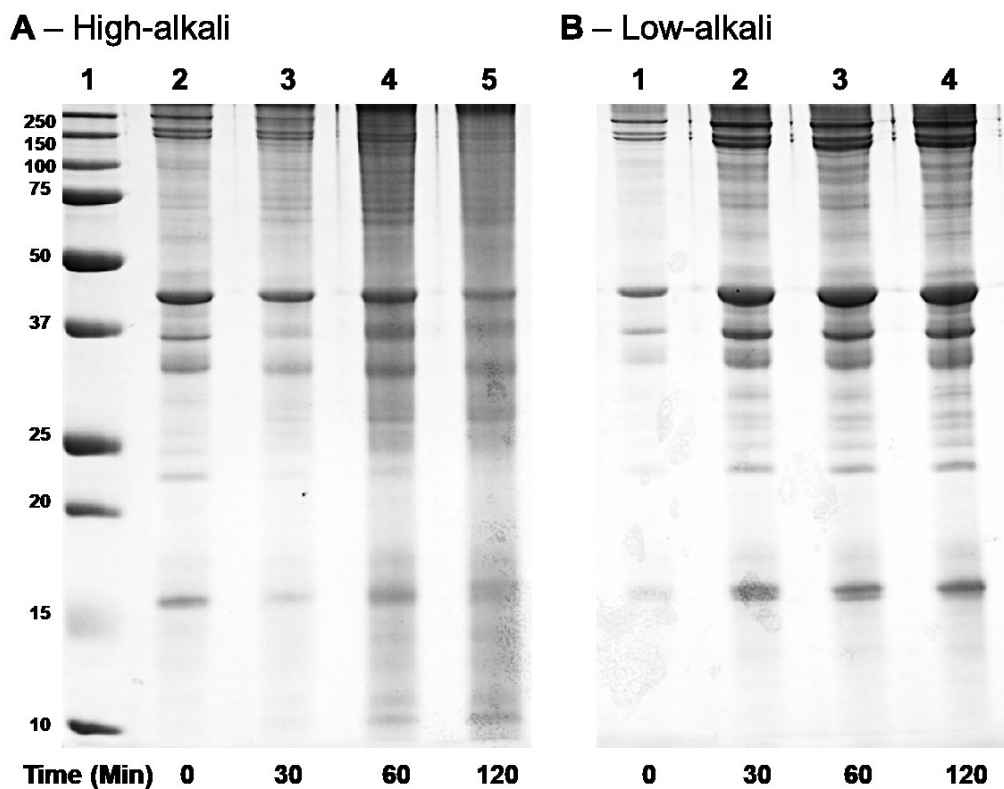


Figure 13. SDS-PAGE showing the protein distribution of TCA-insoluble protein from the (A) high-alkali and (B) low-alkali solubilized salmon muscle protein as a function of solubilization time, following the methodology outlined in Section 3.10. Gels were prepared to a final concentration of 15% acrylamide.

5.3.3 Protein Identification

Protein identification of the SPF was evaluated by LC-MS as outlined in Section 3.4. Electrophoretic bands from the low-alkali solubilized salmon muscle precipitate were selected for this analysis because of their similar distribution to the high-alkali solubilized salmon muscle precipitate observed after immediate mixing (0 h). Numbered bands in Figure 11A - Lane 3 were selected and excised, hydrolyzed using trypsin prior to LC-MS analysis, then all detected peptides were processed using database searching software to identify their potential progenitors from the *Salmo salar* proteome in the UniProtKB protein database. Protein identities (Table 14) of 8 excised bands were consistent with those reported in Separation 1 (Table 10), corresponding to myofibrillar and sarcoplasmic proteins alike. Bands 1 – 3 each corresponded to the myosin heavy chain (MHC, ~ 220 kDa), and their relative migration rates estimated that bands 2 and 3 were ~ 70 kDa and ~ 90 kDa smaller than band 1, respectively. An identical fragmentation pattern was observed in a similar investigation of the cod (*Gadus morhua*) MHC when studying the impacts of post-mortem protein degradation during frozen storage and was attributed to the activity of cathepsin D and acid cysteine proteases (Wang, et al., 2011). Bands 1 – 3 were therefore considered fragments of the identical progenitor sequence, rather than natural isoforms of the MHC with distinct primary sequences. The UniProtKB database consists of the SwissProt section (sequences confirmed at the protein level) and the TrEMBL section (sequences determined at the genomic level). The *Salmo salar* MHC sequence was not reported in the SwissProt section, but various isoforms were frequently

reported in the TrEMBL section, and therefore it was decided to use a consensus sequence to represent the primary amino acid sequence of this protein. Multiple proteins were detected within bands 5, 6 and 7, and collectively represented both myofibrillar and sarcoplasmic protein families.

Table 14. Putative identities of the major electrophoretic bands identified in the low-alkali solubilized isoelectrically precipitated salmon muscle protein resolved by SDS-PAGE.

Band	Protein ID	UniProtKB Accession #	MW (kDa)	pI	Relative Intensity
1	MHC	A0A1S3QJI3	222.01	5.48	16.95
2	MHC	A0A1S3QJI3	222.01	5.48	20.01
3	MHC	A0A1S3QJI3	222.01	5.48	26.97
4	Actin	Q78BU2	41.93	5.22	195.44
5	TM	Q91472	32.49	4.63	159.08
	CK	B5DGG5	42.72	6.44	
6	G3P Dh	B5DGR3	35.95	8.62	72.47
	ALDOA	B5DGM7	39.55	8.61	
7	TPIS	B5XH68	25.11	6.23	23.09
	PGM	B5DGT9	28.64	8.95	
8	MLC-2	Q7ZZN0	18.99	4.96	65.90

MHC: myosin heavy chain; TM: tropomyosin; CK: creatine kinase; G3P Dh: glyceraldehyde-3-phosphate dehydrogenase; ALDOA: fructose biphosphate aldolase; TPIS: triosephosphate isomerase; PGM: phosphoglycerate mutase; MLC-2: myosin light chain-2.

5.3.4 Glucose Uptake of Low Molecular Weight Peptides Derived from Proteins Precipitated Directly from Salmon Muscle Tissues

The electrophoretic separation of protein extracts used to generate protein hydrolysates presents the opportunity to reveal the identity and relative abundance of progenitor proteins. Importantly, the use of whole species proteomes for software-assisted peptide identification considers theoretical peptides that may not be relevant to the SPF composition, when only a few proteins contribute to SPF peptides as seen in Figure 11A – Lane 3. The identities and abundances of major progenitors could therefore aid the prediction of bioactive peptides from whole cell extracts based on a rational use of *in silico* tools. The relative intensity of bands was used to approximate the individual protein abundances in Atlantic salmon muscle (Table 14) and to indicate the likely progenitors of peptides found within SPF than proteins found at lower intensities. Therefore, excised bands of each indicated progenitor were also treated with pepsin, trypsin and chymotrypsin and further processed following the methodology outlined in Section 3.12 then screened in cultured L6 myotubes (Table 15) to evaluate the glucose uptake modulating ability of LMW hydrolysates from isolated Atlantic salmon muscle proteins.

The MHC bands excised from electrophoretic gels and recovered using low-alkali solubilization each exhibited distinct modulation of glucose uptake: band 1 had no effect, band 2 inhibited glucose uptake in the absence (p-value = 0.0128) and presence (p-value = 0.001) of insulin, and band 3 also inhibited glucose uptake but only in the absence of insulin (p-value = 0.0004). Inhibition of glucose uptake was

also observed by hydrolysates from bands 4 and 5 in the absence of insulin, whereas hydrolysates from bands 6 and 7 had stimulating effects in the presence and absence of insulin, respectively (Table 15). It is interesting that intact SPF derived from low-alkali-solubilized proteins exhibited no bioactivity, whereas the individual electrophoretic bands from the same sample did show some glucose uptake modulation.

Table 15. Glucose uptake modulation of hydrolysates produced from isolated *Salmo salar* proteins outlined in Section 3.3. Samples evaluated to different controls are reported separately. Fractions were each tested at 1 ng/mL in the (-) absence and (+) presence of insulin.

Sample	Glucose Uptake (fold activity to control)	
	Insulin (-)	Insulin (+)
Control	1.00 ± 0.00	1.69 ± 0.05
Band 1 - MHC	0.89 ± 0.07	1.58 ± 0.07
Band 2 - MHC	0.86 ± 0.05*	1.41 ± 0.05*
Band 3 - MHC	0.87 ± 0.03*	1.73 ± 0.15
Band 4 - Actin	0.75 ± 0.03*	1.46 ± 0.12
Band 5 -TM/CK	0.87 ± 0.02*	1.45 ± 0.12
Control	1.00 ± 0.00	1.57 ± 0.04
Band 6 – G3P Dh/ALDOA	1.01 ± 0.05	1.73 ± 0.04*
Band 7 – TPIS/PGM	1.11 ± 0.05*	1.69 ± 0.06
Band 8 – MLC-2	1.05 ± 0.06	1.63 ± 0.06

Values are means ± SEM, from n = 5 - 8 individual experiments, performed in triplicate; * p-value < 0.05 vs. control; MHC: myosin heavy chain; TM: tropomyosin; CK: creatine kinase; G3P Dh: glyceraldehyde-3-phosphate dehydrogenase; ALDOA: fructose biphosphate aldolase; TPIS: triosephosphate isomerase; PGM: phosphoglycerate mutase; MLC-2: myosin light chain-2.

Using the previously detected tryptic-peptides for protein identification in Section 5.3.3, sequences from band 3 were localized exclusively to the tail region of the MHC (AA's: 846 – 1927), suggesting that peptides produced from the MHC head domain (Band 1) are able to mask the inhibitory activity exhibited by peptides generated from the tail domain exclusively. Depending on the MHC isoform, head and tail domains contain unique characteristics (Sellers et al., 1996); the myosin head includes both ATP- and actin-binding domains, while the tail is commonly represented by coiled coil secondary helical structures that contain predominantly anionic residues (> 23 %; UniProt Accession No.: A0A1S3QJ13). However, precise data regarding the unique characteristics of the dominant MHC in *Salmo salar* muscle tissues was not determined. Inhibition to glucose uptake was measured for the other myofibrillar proteins, specifically actin, tropomyosin and the sarcoplasmic protein creatine kinase. Rather, stimulating effects were observed from samples derived from the sarcoplasmic progenitor proteins fructose biphosphate aldolase and glyceraldehyde-3-phosphate dehydrogenase or triosephosphate isomerase and phosphoglycerate mutase. These findings suggest that multiple progenitors may contribute bioactive peptides to the SPF and further supports previous findings that SPF activity results from the combined effect of stimulating and inhibiting peptides.

5.3.5 Peptide Identification

As a result of their glucose uptake modulating activities, functional hydrolysate samples prepared from electrophoresed low alkali-solubilized proteins (Figure

11A, Lane 3) were analyzed by LC-MS following the methodology outlined in Section 3.13. Software-assisted database searching of each sample following their LC-MS analysis could not identify any peptide sequences, even though the MFE algorithm confirmed numerous ions were in fact detected in all samples (Appendix A – Table A11 – A13). The average ion mass for each purified protein hydrolysate calculated by the MFE algorithm was 322.4 Da, 308.1 Da, and 286.6 Da, for bands 2, 4 and 6 respectively, suggesting that components of each peptide extract were not targeted by database searching methodology due to their small MWs. Furthermore, the total quantity of available protein in each extracted band was also measured by relative densitometry and was confirmed to be above the limit of detection for the LC-MS instrument. The low peptide identification therefore is unlikely due to their low concentrations, but rather may be caused by losses occurring during the processing of each gel slice.

Although software-assisted database searching was unsuccessful at assigning peptide sequences to the ions detected during LC-MS of each peptide extract, the precursor ion masses reported for each sample by the MFE algorithm enables the prediction of peptide identities with the use of *in silico* digestion tools. Progenitor proteins (Table 13) were hydrolyzed *in silico* using specific digestion criteria reflecting its enzymatic treatment with pepsin, trypsin and chymotrypsin, as well as using non-specific digestion criteria reflecting the formation of all possible peptides from within the progenitor's primary protein sequence that match to experimental ion masses, following the methodology outlined in Section 3.15. Ions

corresponding to Gln-Arg-Arg, Ser-Ser-Arg and Phe sequences were putatively identified in band 2 (Figure 14), and Ala-Val-Phe, Val-Ile-Thr-Ile-Gly and Phe were putatively identified in band 4 using this approach, but further validation of these sequences by *de novo* confirmation using their MS/MS spectra was not performed due to the high degree of background contamination detected. No peptides were putatively detected in band 6 using this approach. The m/z 316.2123 and m/z 430.2437 ions were suspected contaminants because of their common detection within each of the hydrolysates from purified proteins of Atlantic salmon muscle and could also not be identified. Rather, the consistent detection within independent samples prompted their comparison to databases of common MS contaminants (Fisher Scientific, 2017), and many detected ions were indeed consistent with polypropylene glycol (PPG) and polyethylene glycol (PEG) contamination (Appendix A, Table A11 – A13). These ions were not previously observed in the MFE of fractions from Separations 4 and 5, suggesting that these suspected contaminants were likely generated during experimentation rather than introduced during the LC-MS/MS analysis.

Peptide extracts resulting from pepsin, trypsin and chymotrypsin digestion of purified and partially purified proteins from Atlantic salmon muscle were demonstrated to have a modulating effect on glucose uptake in cultured L6 myotubes, but the few peptides that could be identified in these extracts indicates limitations of the methodology. The in-gel digestion of purified proteins mimicking protein hydrolysate processing combined with micro-centrifugal UF techniques for

the identification of bioactive peptides has not been previously described in the literature, and it remains to be determined how peptide losses and sample contamination occur during the experimental methodology. Further validation of putatively identified sequences using MS/MS spectra and of the methodology to generate the hydrolysates from purified proteins could each improve the identification of peptides from these samples.

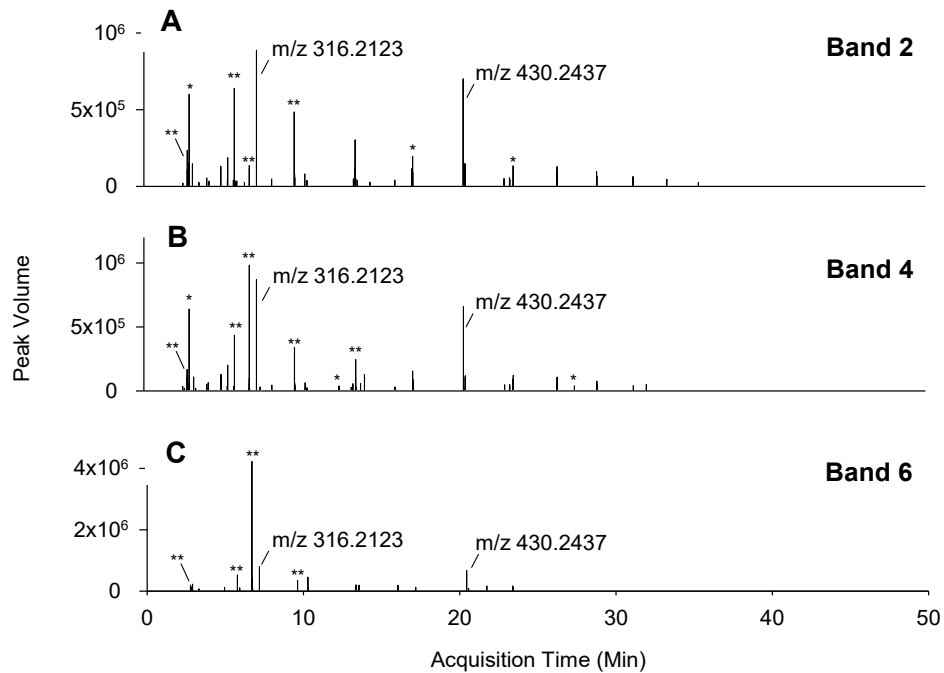


Figure 14. The predicted compositions of bioactive hydrolysates produced from isolated *Salmo salar* proteins using the Molecular Feature Extraction algorithm with comparisons to *in silico* protein digestion libraries. (A) Band 2 – myosin heavy chain, (B) band 4 – actin, (C) band 6 – glyceraldehyde-3-phosphate dehydrogenase and fructose bisphosphate aldolase. * ion is predicted by *in silico* digestions of the curated protein library. ** suspected contaminant ions. Predicted sequences corresponding to detected ions are summarized in Appendix A – Table A11 – A13.

Importantly, most primary protein sequences were identified within UniProtKB's TrEMBL protein database where primary protein sequences are interpretations of genomic data. The primary protein sequences of Atlantic salmon muscle used in this study could deviate from the sequences reported in databases considering the various post-transcriptional and post-translational changes that challenge genomic interpretations, and because farmed Atlantic salmon producers control the genetics of their populations leading to the possibility of intraspecies variation. Validation of each protein sequence from the raw material used for SPF processing could improve the accuracy of this experiment.

5.3.6 Conclusion

The alkaline solubilization with isoelectric precipitation methodology for the production of protein hydrolysates is an important processing step that can influence the generation of bioactive peptides and their functionalities, as well as the formation of salt during IEP. Protein isolation from *Salmo salar* was demonstrated to be sensitive to the NaOH molarity exhibiting both time- and concentration-dependent deterioration of resolution to protein bands and consistent with alkaline hydrolysis-mediated proteolysis affecting total protein recovery and selection processes mediated by differences in conductivity. The inhibiting activity exhibited by a hydrolysate derived from the low-alkali solubilized salmon muscle protein suggested the reactions associated with high NaOH-molarity could be favourable for the glucose uptake stimulating activity of the SPF

and that attempts to reduce salt formation during SPF had a detrimental effect on product quality.

The recovery of high alkali-solubilized and isoelectrically precipitated protein from Atlantic salmon muscle was investigated and its composition revealed that abundant proteins were represented by both myofibrillar and sarcoplasmic protein families. Importantly, it was determined that a high-alkali solubilizing solution was essential for bioactivity in the finished SPF, and that the enzymatic hydrolysis of isolated proteins produced peptide extracts that exhibited variable modulation of glucose uptake in cultured L6 myotubes. Validation of *in silico* digestions to each investigated protein was precluded by challenges to peptide identification, but still demonstrated a practical approach for the identification of potential bioactive peptides from hydrolysates of whole tissue extracts, independent of separations using column chromatography, by combining both conventional and bioinformatic approaches to bioactive peptide identification.

5.3.7 Future Work

Glucose uptake stimulation by Atlantic salmon peptides required protein recovery with highly alkaline conditions that promote primary protein sequence modifications. Protein modifications that occur in high molar NaOH environments are often favourable to the function of peptidomimetics, and it could be plausible that protein modification occurring during salmon protein recovery enhance the agonist and/or antagonistic activity of functional peptides in SPF. Targeted

identification of modified residues and structures consistent with known modifications could support the influence of NaOH-mediated reactions on bioactivity.

Additionally, the use of SDS-PAGE to examine the progenitor proteins prior to enzymatic digestion provides valuable insight with respect to potential bioactive peptides prepared from whole tissue extracts. Specifically, it could be interesting to assess the bioactivity of hydrolysates derived from sarcoplasmic protein extracts of Atlantic salmon and prepared following protocol 2 (Jin, 2012). However, the experimental methodology outlined in Section 3.11 to process each band into single protein hydrolysates could have inadvertently introduced contamination and therefore further validation under controlled conditions is necessary. Alternatively, molecular biology techniques can be used to identify the exact primary protein sequences of the Atlantic salmon muscle raw material to ensure the highest accuracy of peptide identification rather than relying on databases containing primary sequences that represent interpretations of genomic analyses.

5.4 GF Separation 4 and SAX Separation 5 of SPF

5.4.1 Objective

The separation and identification of peptides from an anti-diabetic SPF are essential to deduce the individual modulators of the anti-diabetic effect. SAX Separation 2 and GF Separation 3 of SPF were previously shown unable to concentrate its bioactivity, and inconsistencies in the results suggested that the

unfractionated SPF was not representative of the reference SPF processing methodology (Jin, 2012). Therefore, it was decided to repeat both separations, but with minor modifications to both operational and elution strategies. Furthermore, the identification of LMW peptides poses extreme technical challenges. Bioinformatic tools are resources that can enhance these analyses but are not completely reliable. In this section, both conventional and bioinformatic approaches are used to evaluate amino acid sequences of peptides identified in bioactive fractions of the SPF mixture.

5.4.2 Chromatography of GF Separation 4 and SAX Separation 5

GF. Subsequent chromatographic separations were prepared with SPF from a batch previously demonstrated to exhibit glucose uptake stimulation (Table 13). In Separation 4, the same Bio-Gel P-2 column used as in GF Separation 3 but instead employed a slower elution rate of 0.1 mL/min, and alternate buffer composition of 50 mM ammonium formate pH 6.0 compared to the 100 mM AF with 0.5 % formic acid pH 8.5 eluant A in Separation 3 (Figure 15) were used to improve the resolution of SPF separation. A total of 13 fractions were collected, herein referred to using the FX_{SY} notation, where X is the fraction number and Y is the separation number (F1_{S4} = fraction 1 of Separation 4). UV detection patterns at 215 nm, 254 nm and 280 nm were similar to GF Separation 3 (Figure 8), but without the artifacts that followed the void volume at all wavelengths. The declining intensity of detection at 215 nm until 225 min indicates that peptides eluting beyond this point are LMW containing few peptide bonds (Figure 15A). The elution patterns

at 254 nm and 280 nm throughout the entire separation were quantitatively similar, without a large peak eluting following the void volume as seen in Separation 3, suggesting that the SPF contained a higher proportion of peptides in the low mass range compared to the SPF used for Separations 2 and 3.

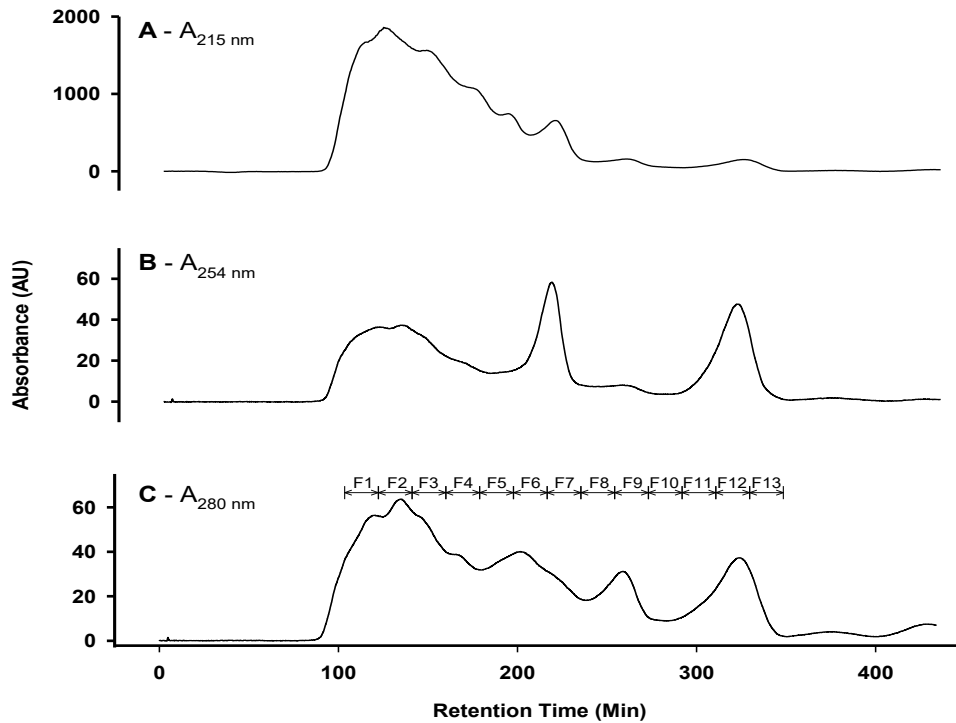


Figure 15. GF Separation 4 of SPF on Bio-Gel P-2 media; 10 mm x 300 mm column; flow rate, 0.1 mL; eluant, 50 mM ammonium formate pH 6.0; detection at (A) 214 nm, (B) 254 nm and (C) 280 nm; sample concentration, 40 mg/mL; injection volume, 100 μ L. Lines represent the UV absorbance at the corresponding wavelength. Fractions were collected within the indicated boundaries and pooled from n = 8 separations.

SAX. Separation 5 was performed on UNOsphere Q media using the same column as SAX Separation 2 but instead also used an increased flow rate of 1.0 mL/min and a shallower, protracted elution gradient; Separation 2 used a gradient of 0 – 70 % eluant B over 2.0 column volumes. Separation 5 used a gradient of 0 – 50 % eluant B over 3.0 column volumes and an alternate buffer pH at 8.0 compared to that of Separation 2 at pH 9.0 (Figure 16).

In Separation 5, two fractions were collected during the isocratic elution, and six collected during the gradient elution. Detection at both 254 nm and 280 nm show two broadly separated peaks of low resolution throughout the gradient elution. The reduction in absolute absorbance intensity compared to previous separations resulted from a malfunction in the UV detector, even though a lower sample concentration was used. Nonetheless, in the initial isocratic step, three distinct peaks were observed at 254 nm at elution times of ~ 15, 22 and 42 min that are consistent with the absorbance pattern observed in the isocratic step in Separation 1 using the same SAX media, but shorter column and slower flow rate. Detection at 280 nm that targets tyrosine and tryptophan did not have the same absorbance pattern as 254 nm in the isocratic elution, suggesting that peptides lacking these residues could be separated from others containing them.

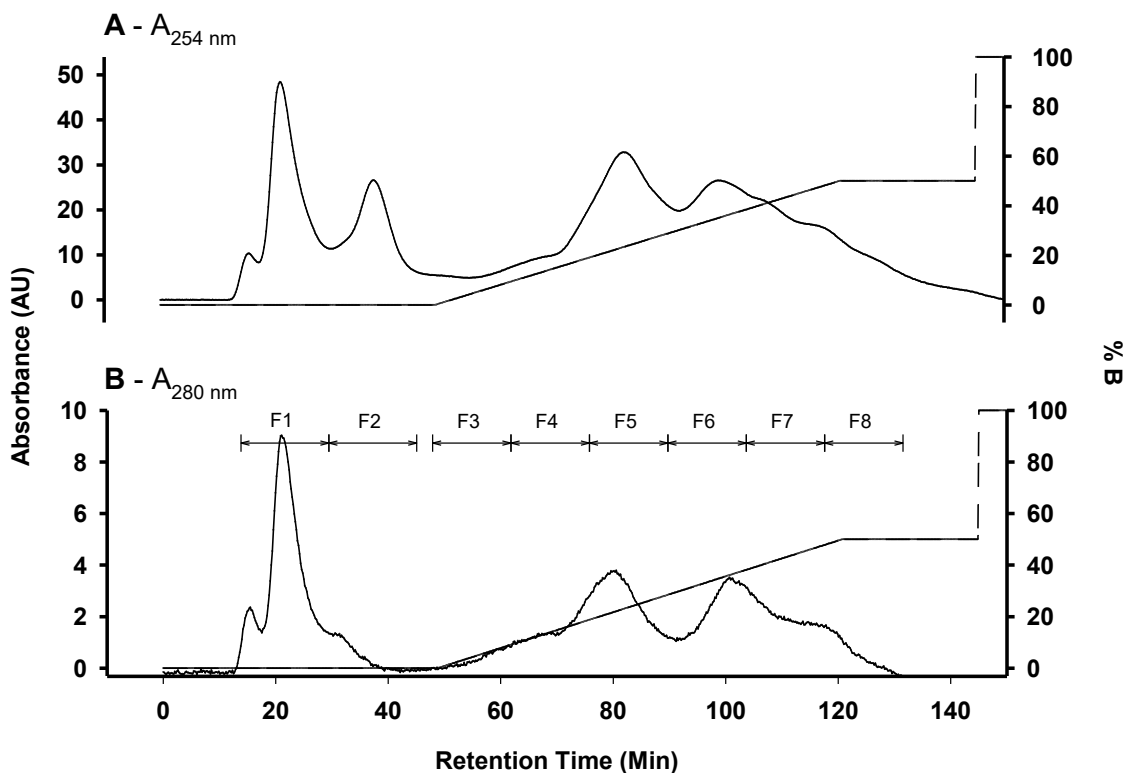


Figure 16. SAX Separation 5 of SPF on UNOsphere Q media; 10 mm x 300 mm column; flow rate, 1.0 mL/min; eluant A, 25 mM ammonium formate pH 8.0; eluant B, 1.0 M ammonium formate pH 8.0; gradient, 0 – 50 % eluant B; detection at (A) 214 nm and (B) 254 nm; sample concentration, 40 mg/mL; injection volume, 500 μ L. Solid lines represent the UV absorbance at the corresponding wavelength and dotted lines represent the percent of elution buffer. Fractions were collected within the indicated boundaries and pooled from n = 7 separations.

5.4.3 Glucose Uptake Analysis of GF Separation 4 and SAX Separation 5 of SPF

GF. The glucose uptake modulating activities from fractions of Separation 4 (Figure 15) are summarized in Table 16. Fractions 2 (p-value = 0.0026) and 3 (p-value = 0.0173) had stimulating effects on glucose uptake only in the presence of insulin. In contrast, fraction 6 inhibited glucose uptake in the absence (p-value = 0.0462) and presence (p-value = 0.0296) of insulin, while fraction 9 also inhibited glucose uptake (p-value = 0.0008) in the presence of insulin. Fraction 7 stimulated glucose uptake in the absence (p-value = 0.0220) and presence (p-value = 0.0399) of insulin, while fraction 8 stimulated glucose uptake only in the absence of insulin (p-value = 0.0011).

SAX. The glucose uptake modulating activity from fractions of Separation 5 (Figure 16) are summarized in Table 17. Fractions 6 and 7 from Separation 5 stimulated glucose uptake by 1.25-fold (p-values = 0.0154) and 1.24-fold (p-value = 0.0054) in the presence of insulin, respectively, while fractions 7 and 8 stimulated glucose uptake by 1.27-fold (p-values = 0.0389) and 1.30-fold (p-value = 0.0111) versus the control in the absence of insulin, respectively. The activity of fraction 6 also stimulated the insulin-dependent condition like fractions 2 and 3 from Separation 4. All functional fractions were collected during an isocratic step at approximately 500 mM ammonium formate following the termination of the gradient stage, whereas active fractions were recovered during column regeneration phase after complete column desorption at ~ 1.0 M during Separation 2.

Table 16. Glucose uptake modulation of fractions from GF Separation 4 on Bio-Gel P-2 media, outlined in Section 3.3. Fractions were each tested at 1 ng/mL in the (-) absence and (+) presence of insulin.

Sample	Glucose Uptake (fold activity to control)	
	Insulin (-)	Insulin (+)
Control	1.00 ± 0.00	1.99 ± 0.03
Fraction 1	1.08 ± 0.08	2.18 ± 0.13
Fraction 2	1.10 ± 0.05	2.38 ± 0.11*
Fraction 3	1.06 ± 0.04	2.24 ± 0.09*
Fraction 4	0.99 ± 0.05	2.11 ± 0.07
Fraction 5	1.00 ± 0.05	2.16 ± 0.11
Fraction 6	0.91 ± 0.04*	1.77 ± 0.08*
Fraction 7	1.08 ± 0.03*	2.19 ± 0.08*
Fraction 8	1.08 ± 0.02*	2.11 ± 0.06
Fraction 9	0.94 ± 0.04	1.71 ± 0.06*
Fraction 10	0.98 ± 0.06	2.08 ± 0.13
Fraction 11	1.00 ± 0.08	2.13 ± 0.14
Fraction 12	0.95 ± 0.04	1.89 ± 0.09
Fraction 13	1.05 ± 0.06	2.15 ± 0.15

Values are means ± SEM, from n = 8 individual experiments, performed in triplicate;
* p-value < 0.05 vs. control.

Table 17. Glucose uptake modulation of fractions from SAX Separation 5 on UNOsphere Q media, outlined in Section 3.3. Fractions were each tested at 1 ng/mL in the (-) absence and (+) presence of insulin.

Sample	Glucose Uptake (fold activity to control)	
	Insulin (-)	Insulin (+)
Control	1.00 ± 0.00	1.69 ± 0.05
Fraction 1	1.00 ± 0.04	1.86 ± 0.10
Fraction 2	0.97 ± 0.05	1.79 ± 0.12
Fraction 3	1.01 ± 0.06	1.76 ± 0.16
Fraction 4	1.05 ± 0.04	1.86 ± 0.09
Fraction 5	1.09 ± 0.06	1.68 ± 0.05
Fraction 6	1.14 ± 0.08	2.11 ± 0.14*
Fraction 7	1.27 ± 0.11*	2.10 ± 0.11*
Fraction 8	1.30 ± 0.10*	1.83 ± 0.11

Values are means ± SEM, from n = 8 individual experiments, performed in triplicate;
 * p-value < 0.05 vs. control.

5.4.4 Identification and Validation of Peptides in Bioactive Fractions

Fractions 2, 3, 6, 7 and 9 from GF Separations 4 (Figure 15, Table 16) and fractions 6 and 7 from SAX Separation 5 (Figure 16, Table 17) with significant modulating effects on glucose uptake in cultured L6 myotubes were selected for further analysis by LC-MS, following the methodology in Section 3.14. In contrast to the software-assisted database searching methodology in Separation 1 that followed a typical proteomics workflow targeting tryptic peptides with a minimum peptide length of 6 residues, a 'no enzyme' digestion parameter targeting all peptides capable of being generated from the primary sequences of the *Salmo salar* proteome and a lower minimum peptide length of 3 residues was used to permit the identification of LMW peptides that were not previously identified in Separation 1. These changes drastically increase the likelihood of reporting false positives because of the increase in the total number of theoretical sequences for comparison to experimental spectra (Panchaud et al., 2012). The changes, in addition to supporting enzymatic cleavage at non-standard sites (Chen et al., 2009), are necessary to tolerate the characteristics of small peptides concentrated in the SPF. In a typical proteomics approach, validation of software-assisted database searching is then performed to ensure that reported peptides from these matches are indeed strong, or valid representations of the experimental data (Nesvizhskii et al., 2007). However, no such validation protocol has been outlined specifically for LMW peptide samples. Therefore, the implementation of a validation step to screen unprocessed software-assisted database search hits

could improve the accuracy of identified sequences in each fraction and was investigated below.

Agilent's MassHunter software for processing LC-MS data includes an advanced algorithmic tool termed the Molecular Feature Extraction (MFE) to generate compound lists for each evaluated fraction (Appendix A, Table A14 – A20) reflecting the composition of samples based exclusively on the detection of patterns in spectral data. The MFE therefore is capable of identifying all detected ions during sample analysis that truly represent the molecules in each fraction distinct from background noise and is a valuable tool for reliably predicting their chemical composition. Unlike software-assisted database searching, no compositional data is generated for each hit, but confirms the retention time and mass of ions with greater reliability than by following the proteomic methodology. Validation in the current study was therefore accomplished by accepting matching ions from filtered database search hits and to the compound lists predicted by the MFE algorithm for each fraction.

When applying the Agilent-recommended validation criteria for good quality database searching data (score > 9, % SPI \geq 60), > 90 % of search hits were eliminated (Table 18) and excluded all putatively identified peptides < 500 Da (Figure 17). These results were dissimilar to both the unprocessed data, as well as the algorithmically predicted chemical composition using the MFE that collectively indicated peptides < 500 Da were abundant components of bioactive

fractions. Considering that the short length and relatively low abundance of most peptides in LMW hydrolysates result in poor scoring and % SPI metrics by database searching (Agilent, 2018), the recommended validation criteria may be too strict to apply to the SPF (Table 18). Therefore, in the present study, modified criteria were developed as there is currently no accepted MS-based validation criteria specific to LMW peptides.

When the modified validation criteria were applied to unprocessed LC-MS database search hits, the total number of validated peptides across all samples increased ~ 4-fold (Table 18) and showed an elevated abundance in the LMW range that resembled both MW distributions of the unfiltered database search hits and all compounds predicted using the MFE (Figure 17), compared to the recommended validation criteria. The consideration of MFE compound lists and database searching hits provides additional confidence for the validation of low abundance and short peptides with poor scores, due to the sophistication of modern algorithmic predictions of spectral data. Compositionally, no tripeptides were identified using the recommended validation criteria, but nearly half of all validated peptides using the modified criteria were tripeptides, demonstrating that the former criteria are also associated with substantial false negative reporting.

Table 18. Comparison of various validation criteria for the identification of peptides in bioactive fractions from Separations 4 and 5. Peptide identification was facilitated by LC-MS with database searching performed using Agilent MassHunter software and the Spectrum Mill MS Workbench, as outlined in Section 3.13.

Sample	Validation Criteria (No. Peptides)			
	No Validation ^a	Recommended ^b	Modified ^c	MFE ^d
F2_S4	190	9	47	142
F3_S4	352	10	82	203
F6_S4	90	2	17	52
F7_S4	29	1	2	13
F9_S4	2	-	1	29
F6_S5	201	31	64	227
F7_S5	167	18	58	394
Total	1031	71	271	1060

^aUnprocessed data; ^bscore > 9 and % SPI ≥ 60; ^cscore > 7, or % SPI > 70, or min. spectrum intensity of 1.0 x 10⁶ and predicted by the MFE; ^dproprietary Agilent algorithm.

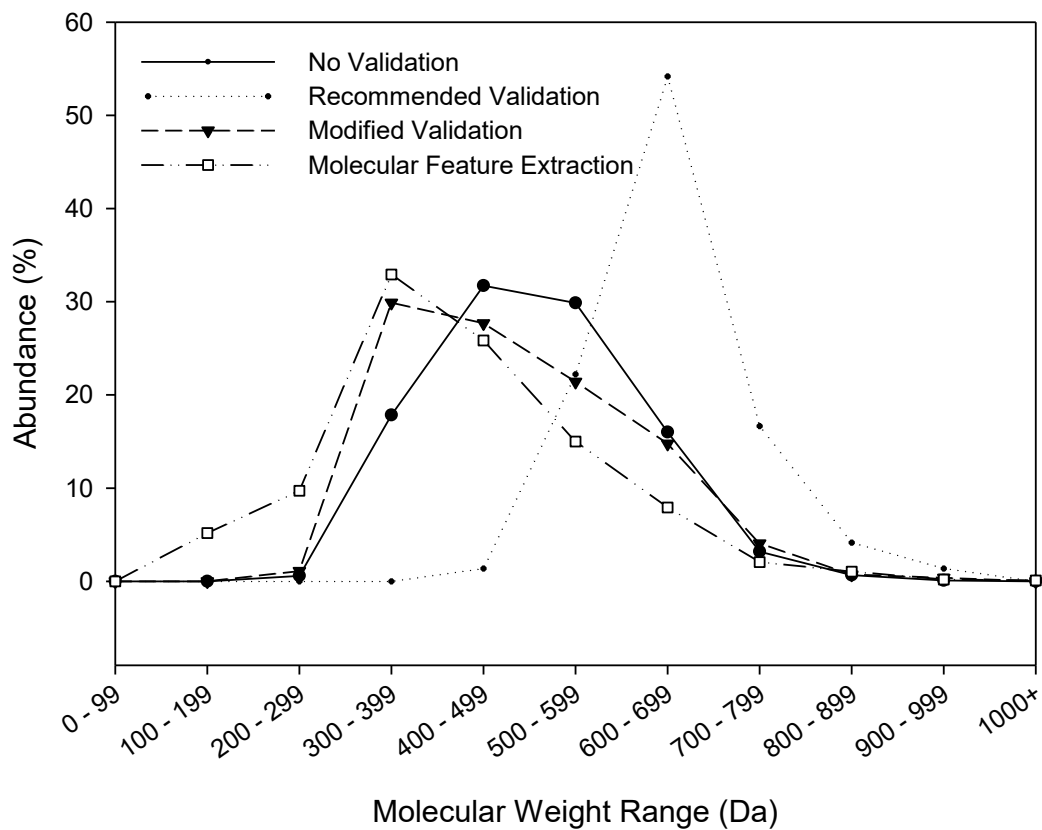


Figure 17. Molecular weight distribution of identified peptides in bioactive fractions of Separations 4 and 5 analyzed by LC-MS following the application of various validation criteria. Peptide identification was accomplished using LC-MS with database searching by the Agilent MassHunter software and the Spectrum Mill MS Workbench as outlined in Section 3.13.

The relative abundances of individual amino acids representing peptides accepted by each validation criterion were also compared to ensure that the modified validation criteria did not introduce any compositional biases (Figure 18). Most residues were reasonably consistent among the three validation criteria indicating that the modified criteria were acceptable in this aspect, but the amino acids Cys, Leu, Lys, His and Arg were each found at very low frequency even when no validation criteria were used, and well below previously reported abundances of each residues in the SPF (Chevrier et al., 2015). The selective omission of peptides containing these residues in bioactive fractions is unlikely to mediate these differences, but these calculated frequencies distinct from abundances calculated using the mole fraction (Chevrier et al., 2015), suggests that the LC-MS acquisition methodology, including the parameters for database searching, may be improved to more accurately determine the SPF composition. All peptides validated using the modified criteria are reported in Appendix A (Table A3 – A9).

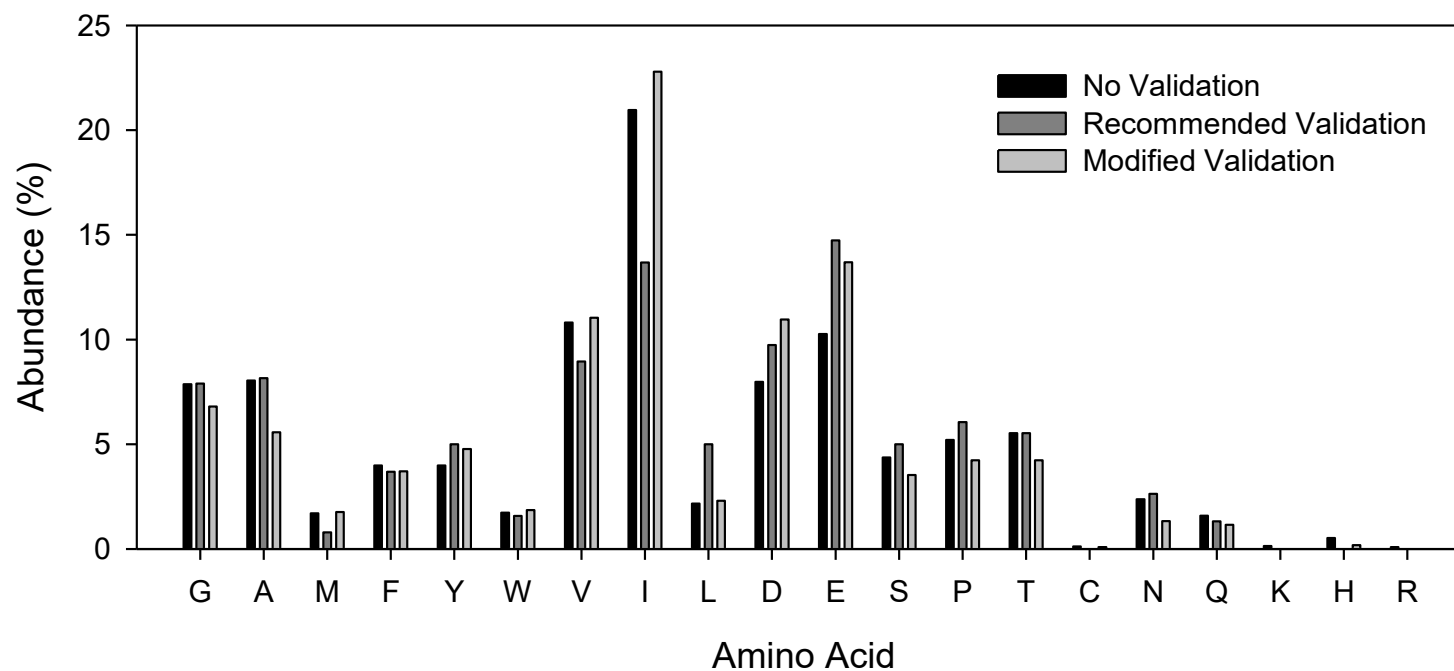


Figure 18. Abundance (%) of individual amino acids in all fractions following each validation criteria. (Black) no validation criteria, (dark grey) Agilent recommended validation criteria for good quality data using high resolution-MS, and (light grey) modified validation criteria. Peptide identification was facilitated by LC-MS with database searching using Agilent MassHunter software and the Spectrum Mill MS Workbench, as outlined in Section 3.13.

5.4.5 Characteristics of Validated Peptides in Bioactive Fractions

Validated peptides from all fractions were collectively evaluated to identify the characteristics of SPF-derived peptides in bioactive fractions. Peptides derived from analyzed fractions were clustered around isoelectric points of pH 3.0 and 6.0 (Figure 19). Peptides containing anionic residues (D, Asp; E, Glu) were identified within the pH 3.0 cluster, whereas all peptides that did not contain these residues were found at the pH 6.0 cluster. In contrast, the identified peptides in fractions from SAX Separation 1 (Figure 4) were clustered broadly at pH 4.0, 7.0 and 10.0, but these differences can in part be attributed to the absence of cationic residues among peptides identified using the methodology targeting LMW peptides. Importantly, peptides from both stimulating and inhibiting groupings were also detected in each cluster indicating that peptide charge cannot be used to reliably predict glucose uptake modulating activity.

The MW distributions of validated peptides (Figure 20A), represented using the abundances of peptides found within each MW range, are concentrated between 300 and 500 Da for the bioactive fractions recovered from SPF separation by gel filtration chromatography, compared to their concentration between 400 and 700 Da for the bioactive fractions recovered from SPF separation by strong anion exchange chromatography. Fractions 7 and 9 of GF Separation 4 however were omitted from this distribution because of the paucity of peptides validated in these fractions (Table 18), but when comparing the spectrally-derived compound lists using the molecular feature extraction algorithm (Figure 20B), each are

concentrated in compounds < 400 Da. Specifically, the high abundances measured for compounds between 0 and 200 Da in fractions 7 and 9 of Separation 4 is a predictable result considering the sieving nature of gel filtration chromatography, and the identification of tripeptides within fractions 1 – 6. Therefore, the high proportion of compounds found in this mass range could reflect that the compositions of these fractions include dipeptides and/or free amino acids that are not targeted by the methodology used for peptide identification.

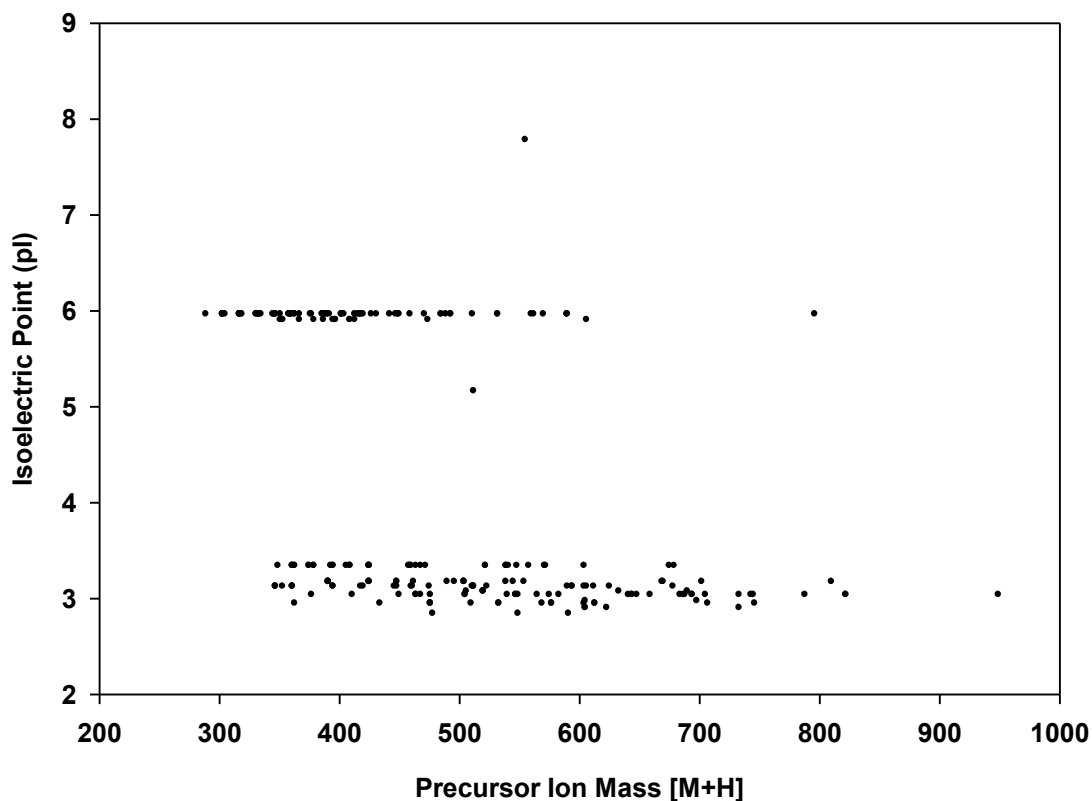
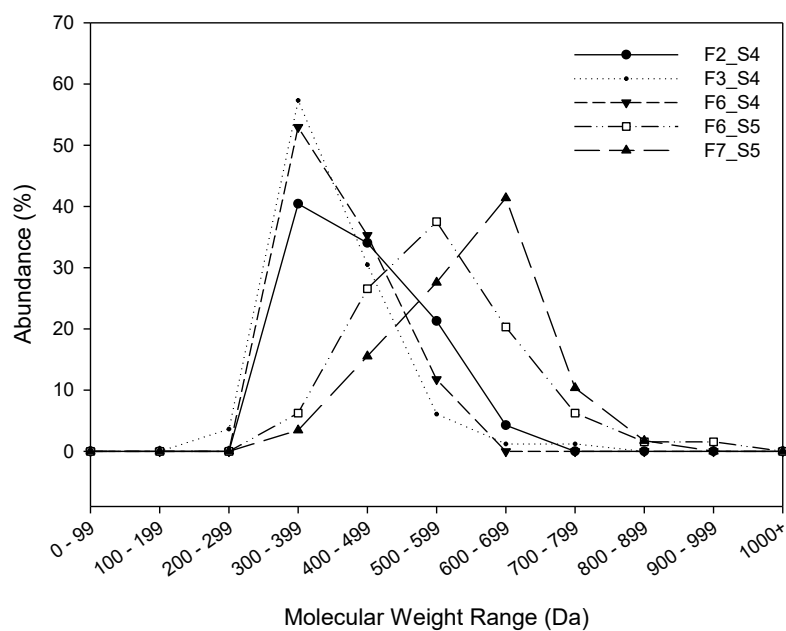


Figure 19. The precursor ion mass and isoelectric point distribution of peptides identified in the bioactive fractions of Separation 4 and 5. Peptide identification was facilitated by LC-MS/MS with database searching performed using Agilent MassHunter software and the Spectrum Mill MS Workbench, as outlined in Section 3.13.

A – Validated Database Search Hits



B – Molecular Feature Extraction Compound List

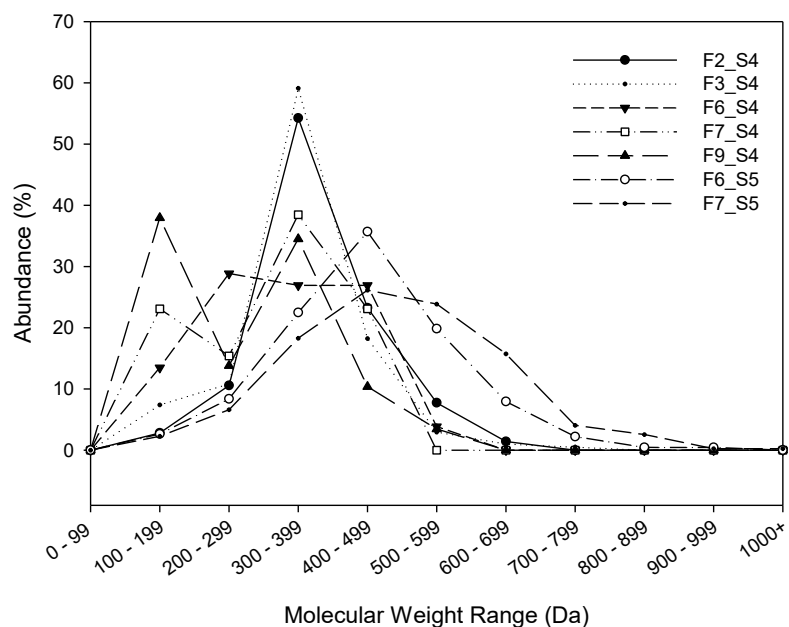


Figure 20. Molecular weight distribution of (A) database search hits (B) molecular feature extraction compound lists, of each bioactive fraction from Separations 4 and 5. Peptide identification was facilitated by LC-MS/MS with database searching and molecular feature extractions performed using Agilent MassHunter software and the Spectrum Mill MS Workbench, as outlined in Section 3.13.

The proportions of amino acids in bioactive fractions from SPF separations (Figure 21) were determined by calculating the frequency of each of the 20 standard amino acids found composing the sequences identified by LC-MS and validated using the modified criteria. In bioactive fractions from the strong anion exchange separation of SPF (FX_S5), high proportions of anionic residues (D, Asp and E, Glu) in these fractions is characteristic of their recovery at the end of gradient elution (Figure 16); anionic residues represented 35 % of all amino acids composing identified peptides in fraction 6 (F6_S5) and 45 % of all amino acids composing identified peptides in fraction 7 (F7_S5). In the glucose uptake stimulating fractions from SPF separation on gel filtration media (FX_S4), anionic amino acids represented 24 % of residues identified in peptide sequences from fraction 2 (F2_S4) but just 2.6 % of residues in fraction 3 (F3_S4), suggesting that these residues may not in fact be necessary components of peptides to exhibit this stimulating activity. However, all of the anionic residue-containing peptides identified in F3_S4 (Ile-Asp-Phe, Ile-Asp-Ile and Ile-Glu-Phe) were each also identified in fraction 2 and considering that the role of anionic peptides on glucose uptake stimulation in cultured L6 myotubes has been previously demonstrated (Roblet et al., 2016), these sequences could represent putative bioactive peptide sequences for further analysis under purified conditions using synthetic standards. In contrast to the anionic amino acids, cationic amino acids (R, Arg; K, Lys) were completely absent in all bioactive fractions, even though unfractionated SPF has previously been shown to contain 6.61 % Arg and 9.87 % Lys (Chevrier et al., 2015) as components of both the peptide and free amino acid composition

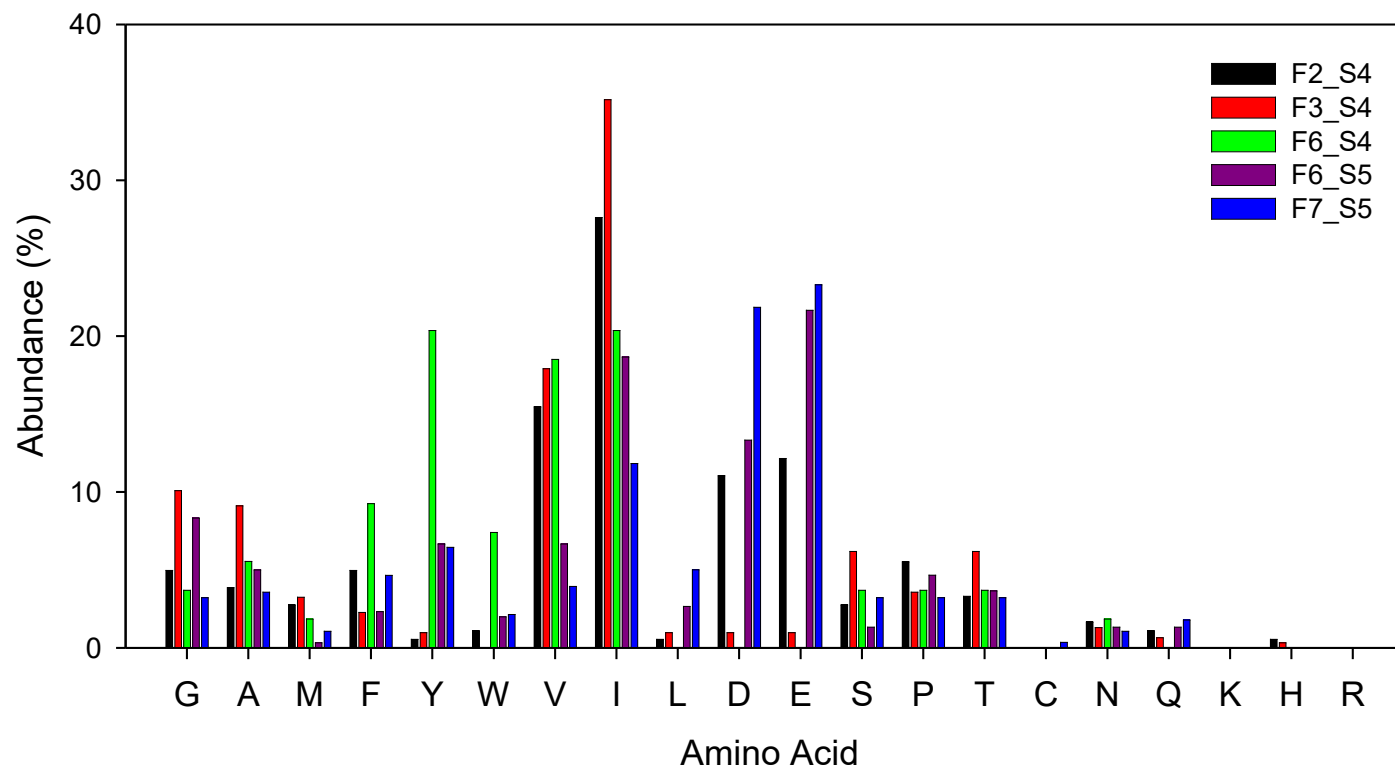


Figure 21. Abundance (%) of individual amino acids as components of validated peptides in fractions of Separations 4 and 5. Abundances are presented as the frequency of each residue from within all identified sequences in bioactive fractions. Peptide identification was facilitated by LC-MS with database searching performed using Agilent MassHunter software and the Spectrum Mill MS Workbench, as outlined in Section 3.13.

(Roblet et al., 2016). Furthermore, neither anionic nor cationic residues were identified in relevant proportions in fraction 6 of Separation 4 (F6_S4) that exhibited potent inhibition of glucose uptake.

Unlike the amino acid compositions of bioactive fractions from Separations 4 and 5 (Figure 21), cationic amino acids did compose many of the identified peptides in fractions from SAX Separation 1 (Appendix A, Table A1) where fractions were analyzed using a different LC-MS methodology, and suggest that these variations could be related to different analytical methodologies than truly reflective of the characteristics of peptides in these fractions.

The abundances of branched chain amino acids (BCAA; I, Ile; L, Leu; V, Val) (Figure 21) in bioactive fractions of Separation 4 (FX_S4) were 20 – 30 % larger than in the those from Separation 5 (FX_S5). The frequency of Ile residues in these fractions ranged between 13 % and 32 %, and Val residues ranged between 2 % and 19 %, exceeding their previously reported frequencies in the unfractionated SPF composition at 5 % and 6 %, respectively (Roblet et al., 2016). The total BCAA frequency in stimulating and inhibiting fractions were also compared, but Leu in fractions from Separation 4 (FX_S4) was rarely identified, and although fractions of Separation 5 (FX_S5) had some representation, neither fraction measured close to the 8 % calculated for the unfractionated SPF composition (Roblet et al., 2016). The low frequency of Leu residues among the identified peptides may be related to the irregular high frequency of Ile residues in bioactive fractions reported by the

current methodology when considering their identical masses and difficulty to differentiate without selective methodology (Xiao et al., 2016).

The aromatic amino acid (AAA; F, Phe; Y, Tyr; W, Trp) contents of bioactive fractions (Figure 21) were also variable, ranging from 4 – 37 %. Fraction 6 of Separation 4 (F6_S4) with inhibitory activity contained a six- and nine-fold larger frequency of AA than F2 and F3 of the same separation, respectively, and over 3-fold larger than F6 and F7 of Separation 5. In total, Tyr, Phe and Trp represented frequencies of 19.8 %, 9.8 % and 7.8 % of identified residues in F6_S4, respectively.

Minor variations in frequency were observed for Met, Gly, Ala, Pro and Ser between bioactive fractions from Separations 4 and 5, but all were close approximations to previous estimates by amino acid analysis in SAX Separation 1 calculated using their percent abundance by their molecular weight (Table 9). The concentration of aromatic residues in F6_S4 was a distinguishing characteristic, and therefore could be related to its inhibitory effect on glucose uptake in cultured L6 myotubes.

In all fractions, Ile, Val, Glu, Asp, Gly, Ala and Tyr represented over 70% of all residues from identified peptides. Importantly, identifying the frequency of residues within bioactive fractions does not reveal candidate bioactive peptide sequences, but simply indicates those residues that may be components of them. Furthermore,

these results represent a qualitative interpretation of complex peptide mixtures and unlike the amino acid analysis, does not account for the quantitative variations of each peptide.

Regarding the absence of cationic residues, various analytical obstacles can prevent their identification using software-assisted database searching. Peptides containing cationic residues often require higher collision energies to achieve comparable fragmentation efficiency than to peptides lacking them (Wysocki et al., 2000). Increasing collision energies simultaneously increases the likelihood of cleaving amino acid side chains (Dongré et al., 1996; Gehrig et al., 2004), such that the product ions of peptides containing basic residues rarely produce the expected b- and y-series ions (Wee et al., 2006). Furthermore, evidence of protein hydrolysis in Atlantic salmon muscle tissue during alkaline solubilization in 1.0 M NaOH could suggest that other NaOH-associated protein modifications, such as deamination that uniquely affects Arg and Lys, occurs prior to their LC-MS analysis. The frequent identification of both residues in fractions of Separation 1 facilitated using Proteome Discoverer software with Sequest HT scoring, demonstrates the limitations of the current LC-MS methodology for identifying LMW peptides. Similar to the cationic residues measured at near negligible frequencies within identified peptides, Leu was reported at a fraction of the Ile abundance, even though these residues have been previously described to be found at equivalent proportions in *Salmo salar* muscle tissues. Leu and Ile are isobaric amino acids, but their differentiation without consideration of retention

times requires the formation of product ions resulting from side-chain cleavage (d- or w-series ions) that were not specifically generated or targeted by the analytical methodology (Johnson et al., 1987). The disproportionate assignment of Ile over Leu residues in database searching hits may result from the challenges associated with the interpretation of MS/MS spectra of LMW peptides containing isobaric residues by Agilent MassHunter software, but the rationale regarding the preference for Ile reporting over Leu is unknown. Importantly, the apparent frequency of BCAAs composing identified peptide sequences in bioactive fractions of SPF indicates that overcoming these analytical challenges could greatly improve the analysis and interpretation of their potential influence over mediating glucose uptake stimulation.

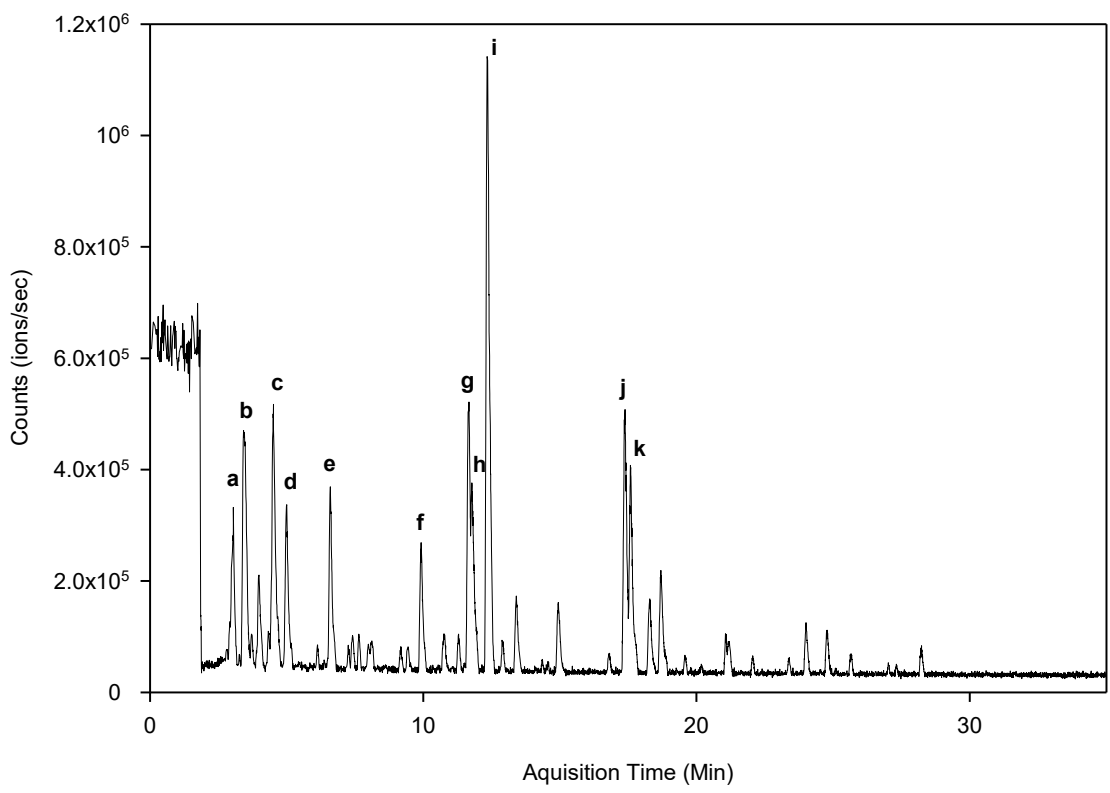
5.4.6 Peptide Compositions of Bioactive Fractions

The compositions of bioactive fractions following their LC-MS analysis can be visually represented using various spectral and chromatographic formats (Section 2.5.3). Mass detection and the compositions of each bioactive fraction in the present study were reported using base peak chromatograms (BPC) to simultaneously identify precursor ion masses and estimate their relative proportions (Figures 22 – 28, Panel A). In this section, putative peptide sequences of each precursor ion were assigned following either software-assisted database searching (Section 3.13), or by manual *de novo* sequencing (Section 3.16) (Figures 22 – 28, Panel B) approaches, as many precursor ions with small *m/z* values were either prevented from being identified with the parameters of

database searching or weren't accepted using the modified validation criteria. In the present study, candidate bioactive peptides in SPF were fractionated by SAX chromatography with UNOsphere Q media or GF chromatography with Bio-Gel P-2 media (Separations 4 and 5, respectively) and those peptides found at high relative proportions to others in glucose uptake modulating fractions are summarized in Table 19.

In fraction 2 from SPF separation by GF Separation 4 (F2_S4, Figure 22A), the most abundant peptide, Ile-Ile-Ala-Pro-Pro-Glu-Arg (*ion i*) though requiring identification by a *de novo* approach, was generated from the apparent tryptic-cleavage of actin (location: 331 – 337) and also previously identified by software-assisted database searching in fraction 1 of SPF Separation 1, although a different LC-MS methodology was used. No other Arg-containing peptides were identified in this fraction and other dominant ions in this fraction were each found at ~ 20 – 40 % the intensity of Ile-Ile-Ala-Pro-Pro-Glu-Arg and were primarily tripeptides containing anionic residues and BCAAs. The dominant peptide sequences identified in this fraction (Table 19) were composed primarily of those same anionic and BCAA residues previously measured at high frequencies in this fraction. The MS/MS spectra for sequences identified by a *de novo* approach (Figure 22B) frequently relied upon the N-terminal immonium ions, a_2 , b_2 , y_1 and y_2 ions for diagnostic purposes, but many other background ions were also detected that could not be directly related to the predicted peptide sequences.

A



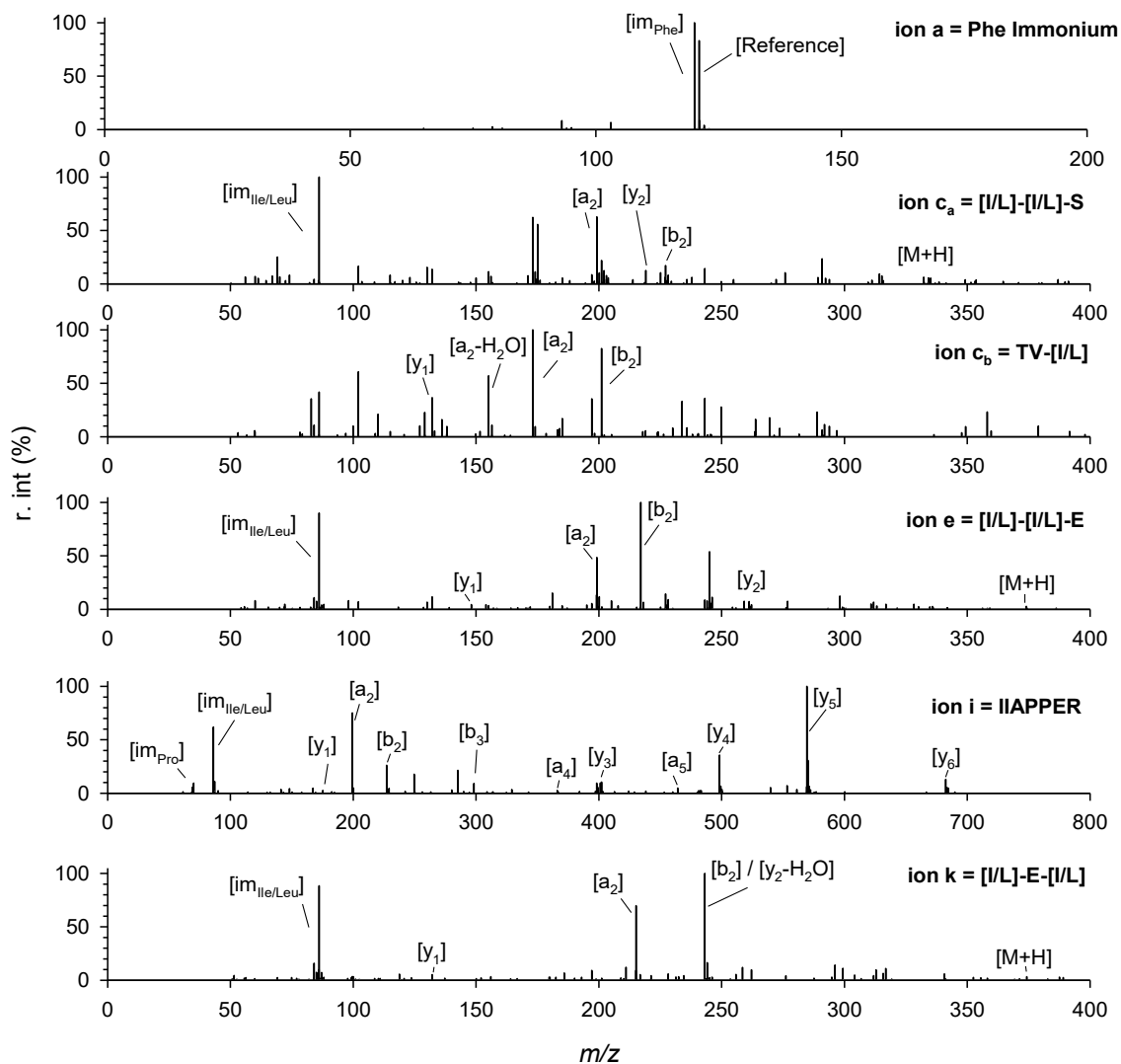
B

Figure 22. LC-MS analysis of fraction 2 of GF Separation 4 (F2_S4), represented using the (A) base peak chromatogram. The mass and putative identity of lettered peaks are summarized in Table 19. (B) The product ion spectra used for manual *de novo* sequencing of peaks in part (A), annotated with the detected fragment ion matching to the identified peptide sequence, with assistance from mMass software Version 5.5.

In fraction 3 from SPF separation by GF Separation 4 (F3_S4; Figure 23A), the compositions and arrangement of residues in identified peptides were similar to those identified in fraction 2. The low frequency of anionic residues found in peptides identified in this fraction (Figure 21) were replaced by the polar but uncharged Ser, Thr and Gln, or by neutral residues such as Gly, Ala and the BCAAs. The dominant ion in this fraction (*ion k*) was initially identified as Ile-Val-Ile by database searching, but *ion l* had the identical mass, a similar retention time, and an MS/MS spectrum supporting an Ile/Leu-Val-Ile/Leu peptide composition (Figure 23B).

The identification of the 86 Da immonium ion is characteristic of both Ile and Leu residues at the N-terminal, such that insufficient information is available to differentiate these residues by interpretation of the MS/MS spectra alone, demonstrating the limitation of both database searching and *de novo* sequencing for LMW peptide identification. Fractions 2 and 3 of Separation 4 both stimulated glucose uptake exclusively in the presence of insulin by nearly the same extent but did not contain common dominant ions.

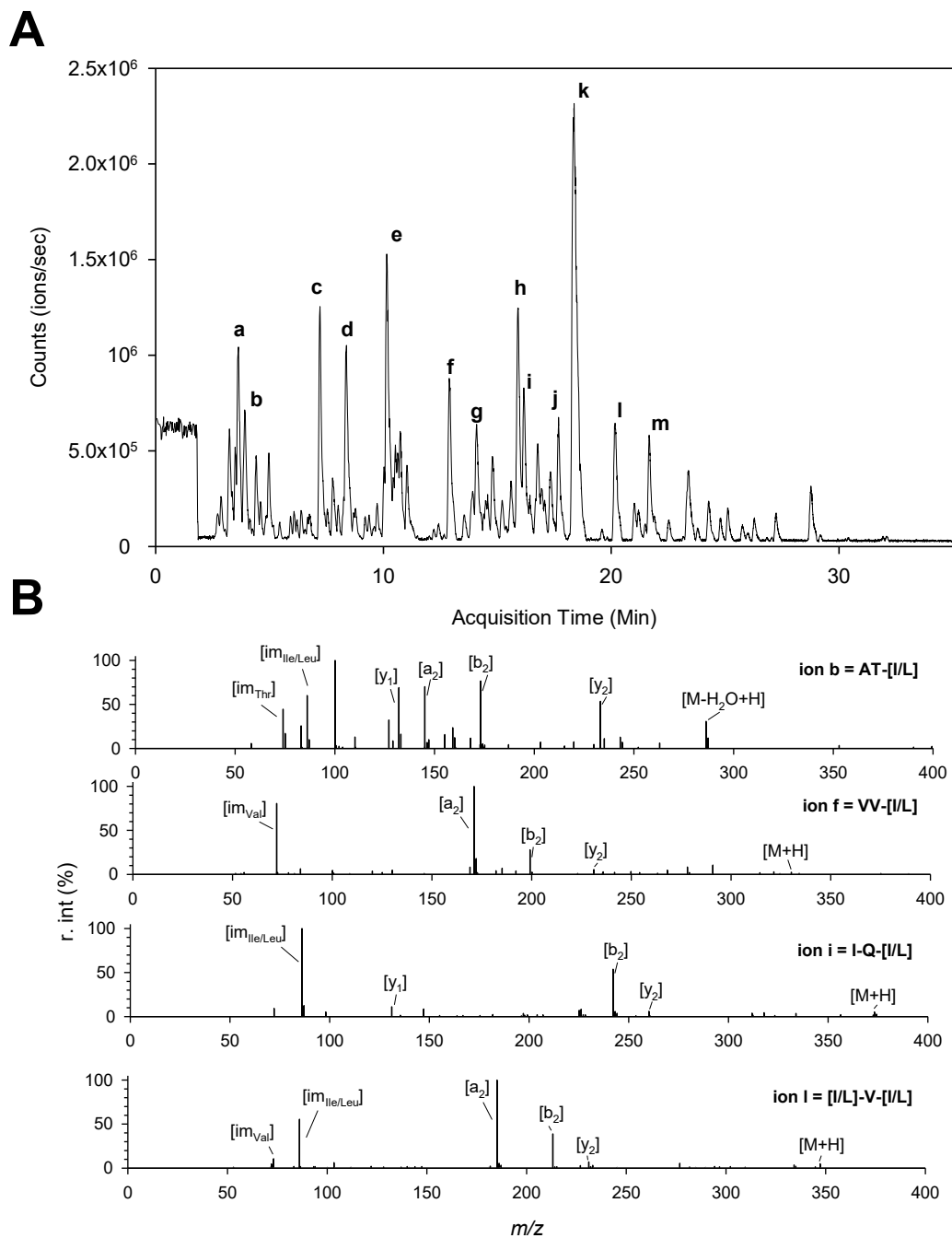
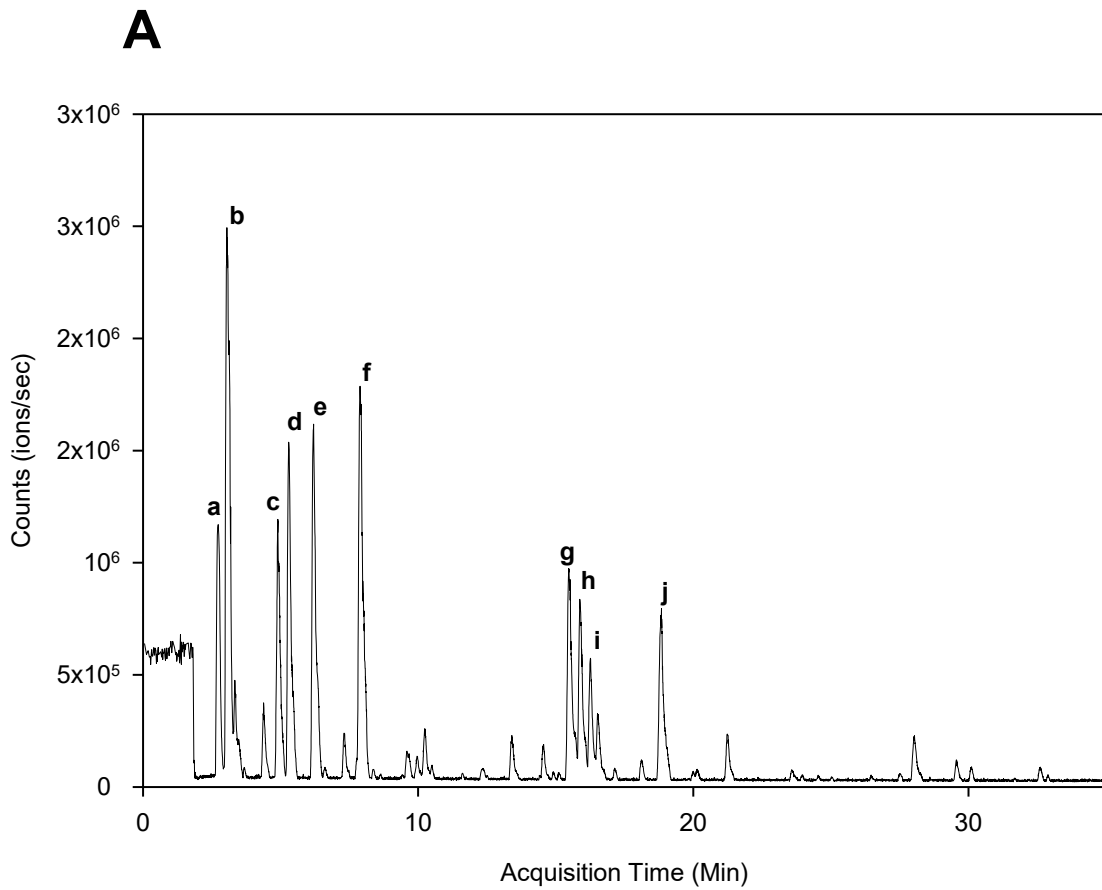


Figure 23. LC-MS analysis of fraction 3 of GF Separation 4 (F3_S4), represented using the (A) base peak chromatogram. The mass and putative identity of lettered peaks are summarized in Table 19. (B) The product ion spectra used for manual *de novo* sequencing of peaks in part (A), annotated with the detected fragment ion matching to the identified peptide sequence, with assistance from mMass software Version 5.5.

In fraction 6 from SPF separation by GF Separation 4 (F6_S4; Figure 24A), the most dominant ions could not be identified by database searching because of their small size. Dipeptides were identified by manual *de novo* sequencing and were predominantly composed of aromatic residues, supporting the idea that similar high frequencies were measured in tripeptides identified by database searching. The most dominant ion in this fraction (*ion b*) had a mass consistent with the Phe immonium ion (m/z 120.0804), and this identity is further supported by the simultaneous detection of the intact Phe y_1 ion (m/z 166.085), albeit at a lower intensity. In MS-analysis, immonium ions are internal peptide fragments with unique masses characteristic to each of the 20 standard amino acids, generated during peptide fragmentation during LC-MS and used for diagnostic purposes. Unlike their traditional detection in MS/MS spectra following precursor ion fragmentation, the identification of immonium ions in precursor ions lists would indicate that free Phe is found abundantly in this fraction. However, the identification of immonium ions as a precursor ion could be an indication that peptide fragmentation, or other changes to the composition of SPF fractions, may be occurring prior to entering the mass spectrometer and therefore modifying its composition.

In-source fragmentation refers to peptide fragmentation within the ionization source during MS-analysis and is a phenomenon extensively described in the literature, but not to the apparent extent observed in the present study for free amino acids. Unlike the results of database searching, the use of BPC to select

precursor ions for *de novo* sequencing demonstrated that dipeptides were in fact abundant components of this fraction. Furthermore, the duplicate identification of Ile/Leu-Tyr dipeptides, *ion d* and *e* (Figure 24B), supports the presence of both Ile- and Leu-isoforms of this peptide in this sample.



B

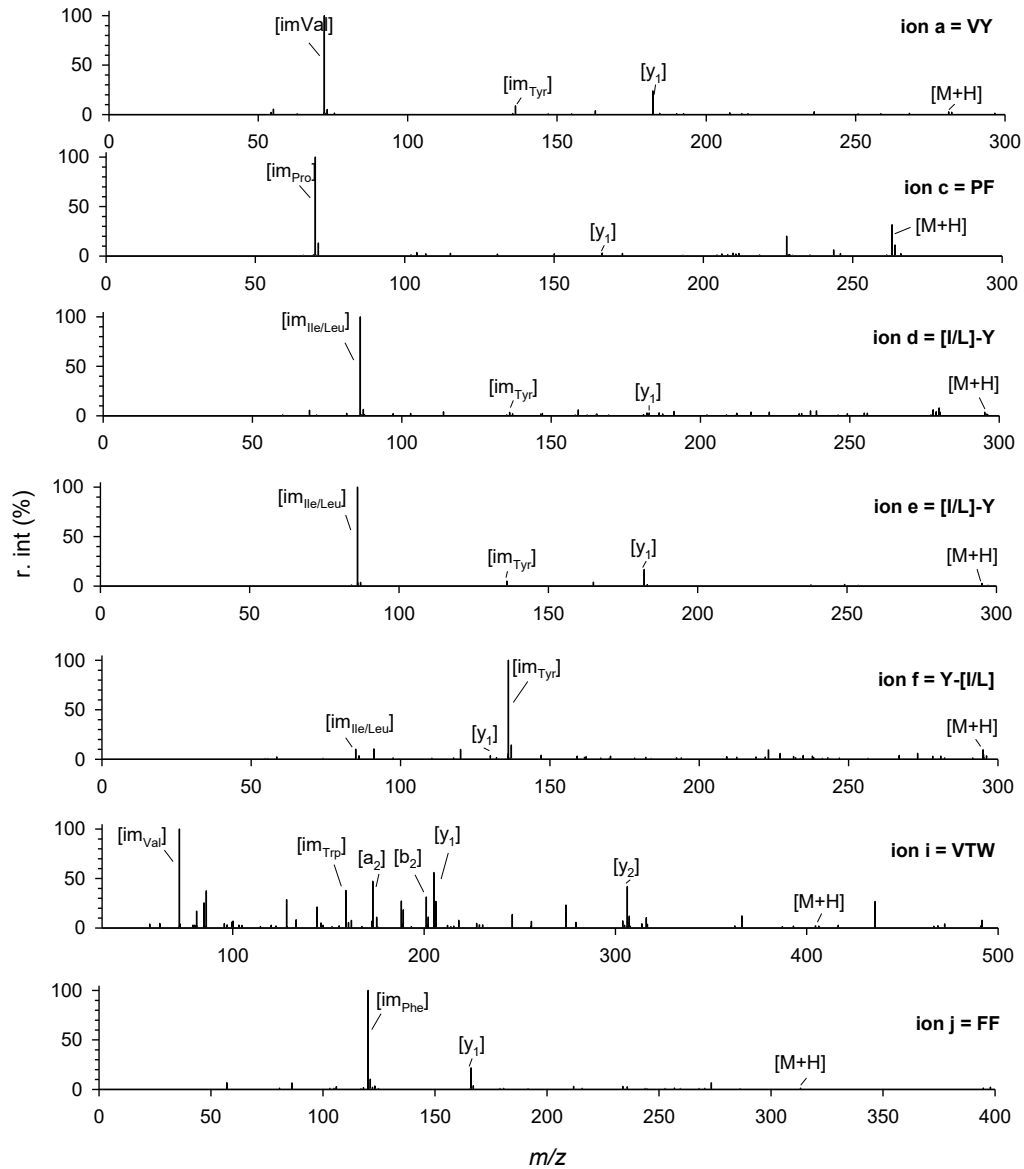
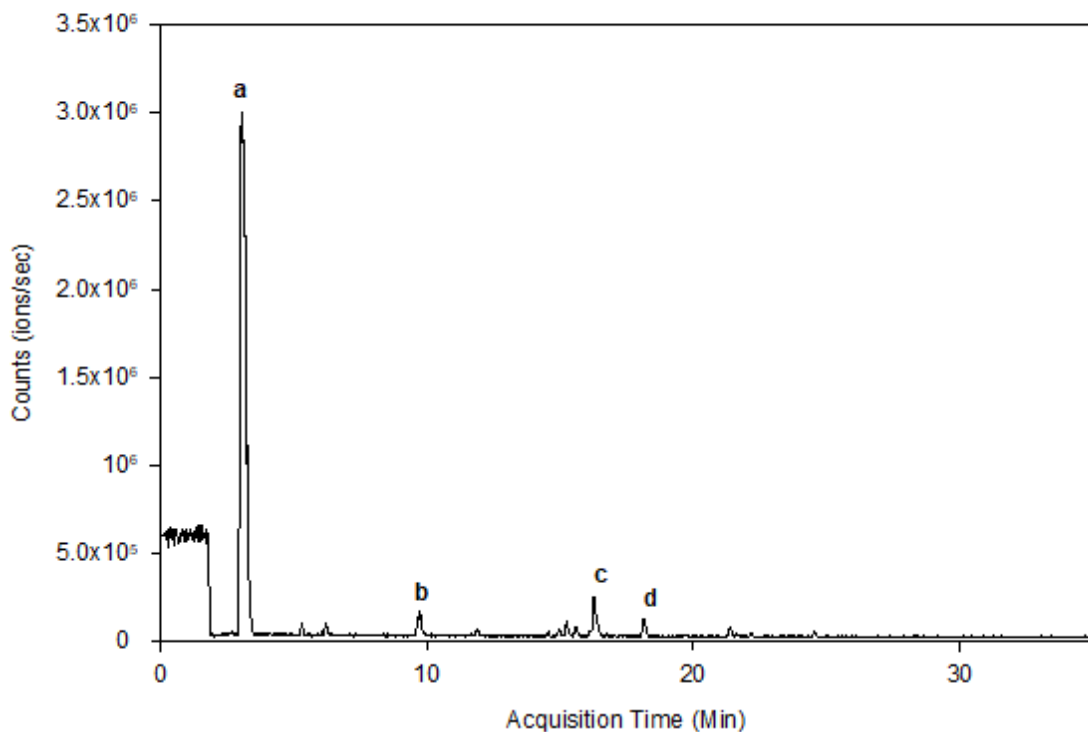


Figure 24. LC-MS analysis of fraction 6 of GF Separation 4 (F6_S4), represented using the (A) base peak chromatogram. The mass and putative identity of lettered peaks are summarized in Table 19. (B) The product ion spectra used for manual *de novo* sequencing of peaks in part (A), annotated with the detected fragment ion matching to the identified peptide sequence, with assistance from mMass software Version 5.5.

The same Phe immonium ion was repeatedly identified as a precursor ion in other fractions, and its consistent identification in various fractions from SPF separation by GF Separation 4 rather supports its formation during their MS-analysis rather than independently eluting in multiple chromatographic fractions from the same separation. In fraction 7 of SPF separation by GF Separation 4 (F7_S4; Figure 25A), *ion a* was also detected with the m/z 120.0809 and measured the highest absolute intensity of any ion recorded in all analyzed fractions of SPF samples and was detected at 10 – 20-fold greater intensity than other peptides identified by database searching in the same fraction, suggesting that Phe as a free amino acid could represent a large proportion of this fraction.

A



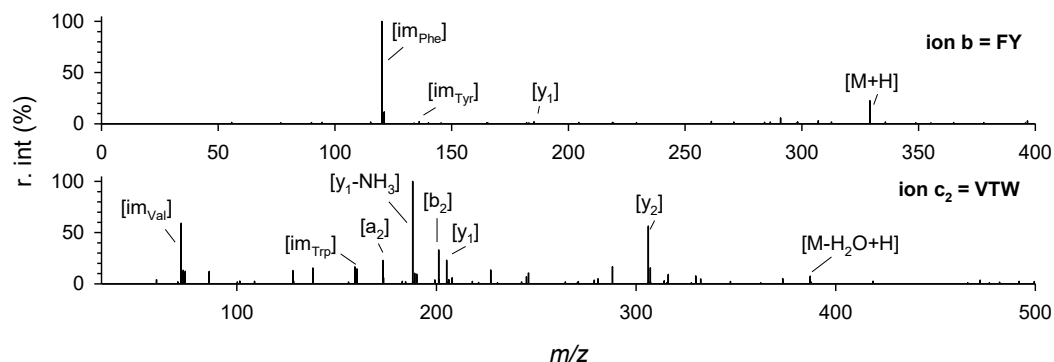
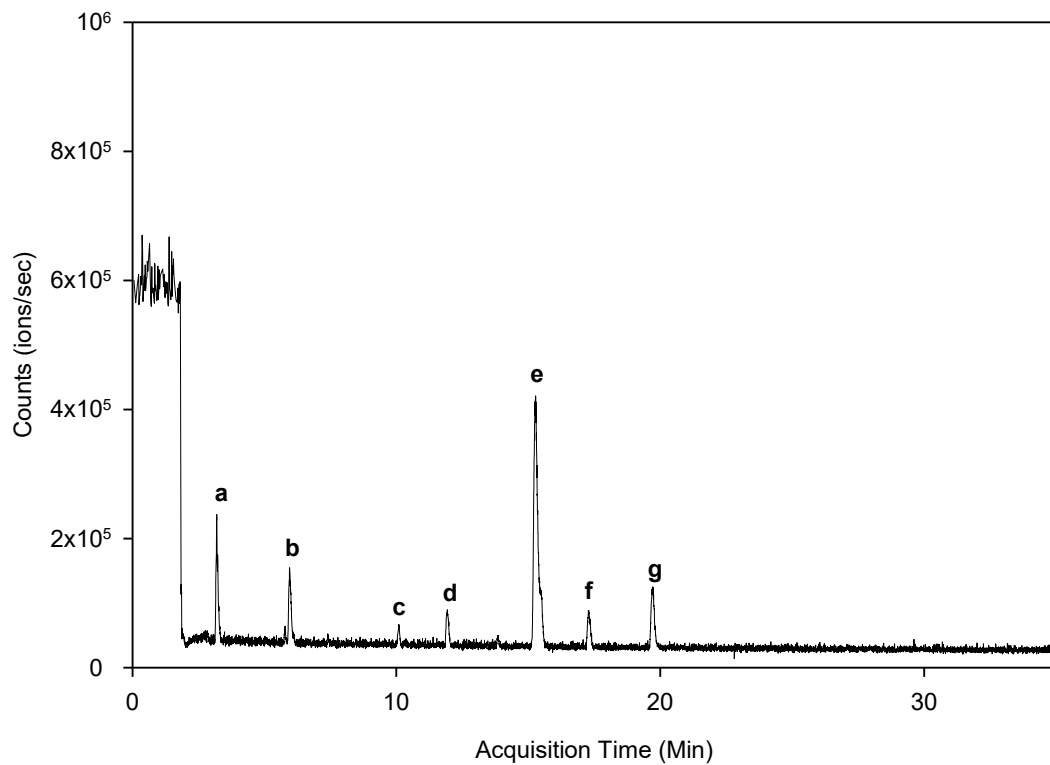
B

Figure 25. LC-MS analysis of fraction 7 of GF Separation 4 (F7_S4), represented using the (A) base peak chromatogram. The mass and putative identity of lettered peaks are summarized in Table 19. (B) The product ion spectra used for manual *de novo* sequencing of peaks in part (A), annotated with the detected fragment ion matching to the identified peptide sequence, with assistance from mMass software Version 5.5.

In fraction 9 of SPF separation by GF Separation 4 (F9_S4; Figure 26A), *de novo* sequencing determined that Trp-containing dipeptides represented the primary composition of this fraction. These peptides were characteristically similar to the Tyr-containing dipeptides identified in fraction 6, and their common inhibition of glucose uptake could be an indication that dipeptides containing aromatic residues are mediators of this inhibition. The recovery of these peptides is reflected by the peaks detected at 280 nm for both fractions (Figure 15). A third major peak at 320 min consistent with the elution of fraction 12 was also observed but did not significantly inhibit glucose uptake ($p > 0.05$). The consistent identification of BCAA with Tyr and Trp among dipeptides identified in fractions 6 and 9 of Separation 4, respectively, were collected at elution times separated by ~ 60 min (6 mL

difference out of the 23.5 mL total column volume). Considering that Tyr and Trp have a mass difference of only 23.03 Da, and that intermediate fractions contained both di- and tri-peptides, it could be possible that factors other than size influence peptide retention when separated on Bio-Gel P-2 media.

A



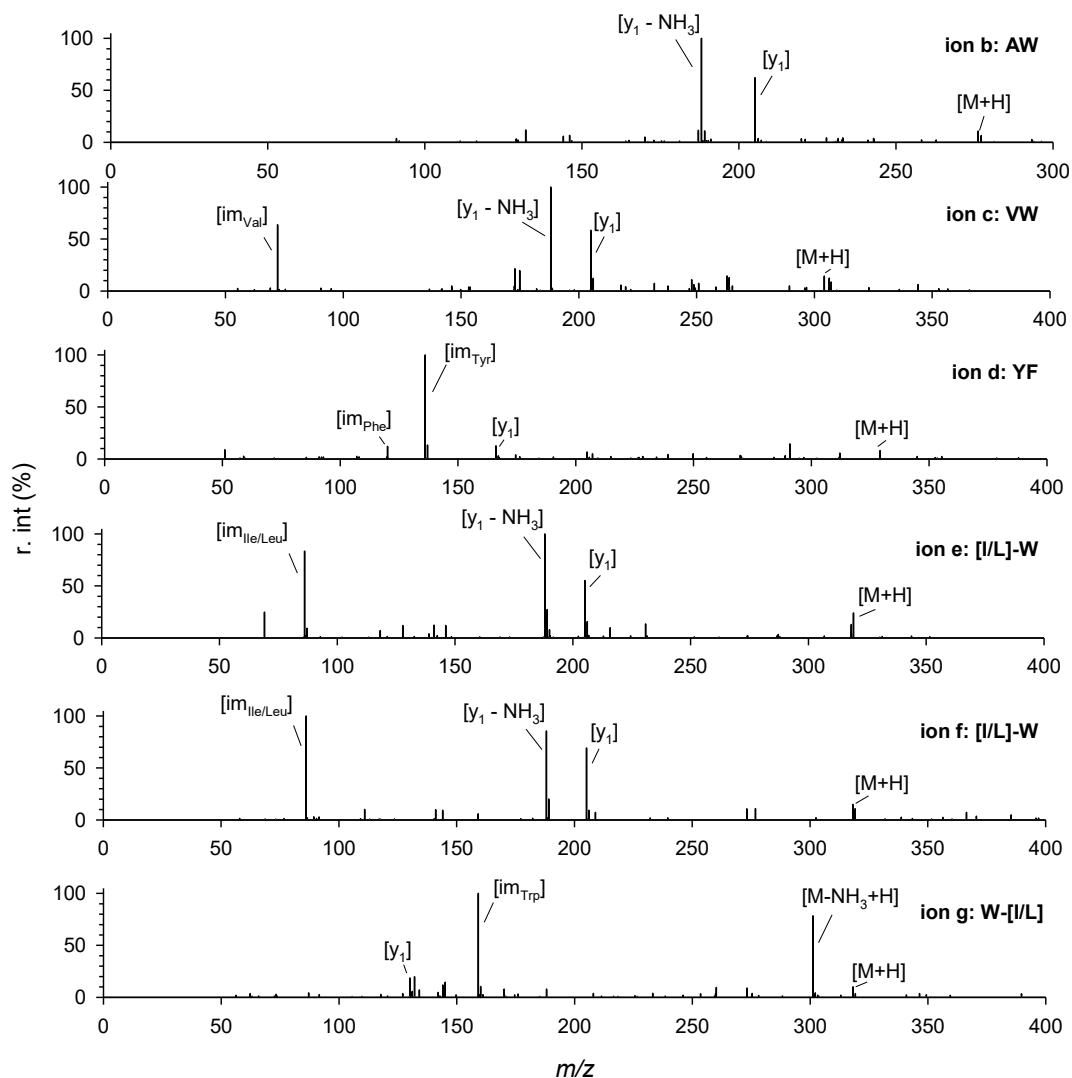
B

Figure 26. LC-MS analysis of fraction 9 of GF Separation 4 (F9_S4), represented using the (A) base peak chromatogram. The mass and putative identity of lettered peaks are summarized in Table 19. (B) The product ion spectra used for manual *de novo* sequencing of peaks in part (A), annotated with the detected fragment ion matching to the identified peptide sequence, with assistance from mMass software Version 5.5.

In fraction 6 of SPF separation by SAX Separation 5 (F6_S5; Figure 27A), multiple peptides were found at a similar intensity to the most dominant ion (Ile-Glu-Glu). All tri- and tetrapeptides were composed of at least one anionic residue, but peptides containing adjacent anionic residues were common to dominant peptides in this fraction. In fraction 7 of SPF by SAX Separation 5 (F7_S5; Figure 28), the BPC shows one dominant precursor ion with the sequence Pro-Thr-Cys-Pro-Asp-Ala generated from the protein *integrin beta-5 like isoform X2* as predicted using database searching, but *de novo* sequencing (Figure 28B) suggested this sequence was Asp-Trp-Pro-Asp-Ala and derived from creatine kinase – an abundant protein identified within the high-alkali solubilized salmon muscle protein precipitate, and a progenitor to peptides identified in fractions from SAX Separation 1.

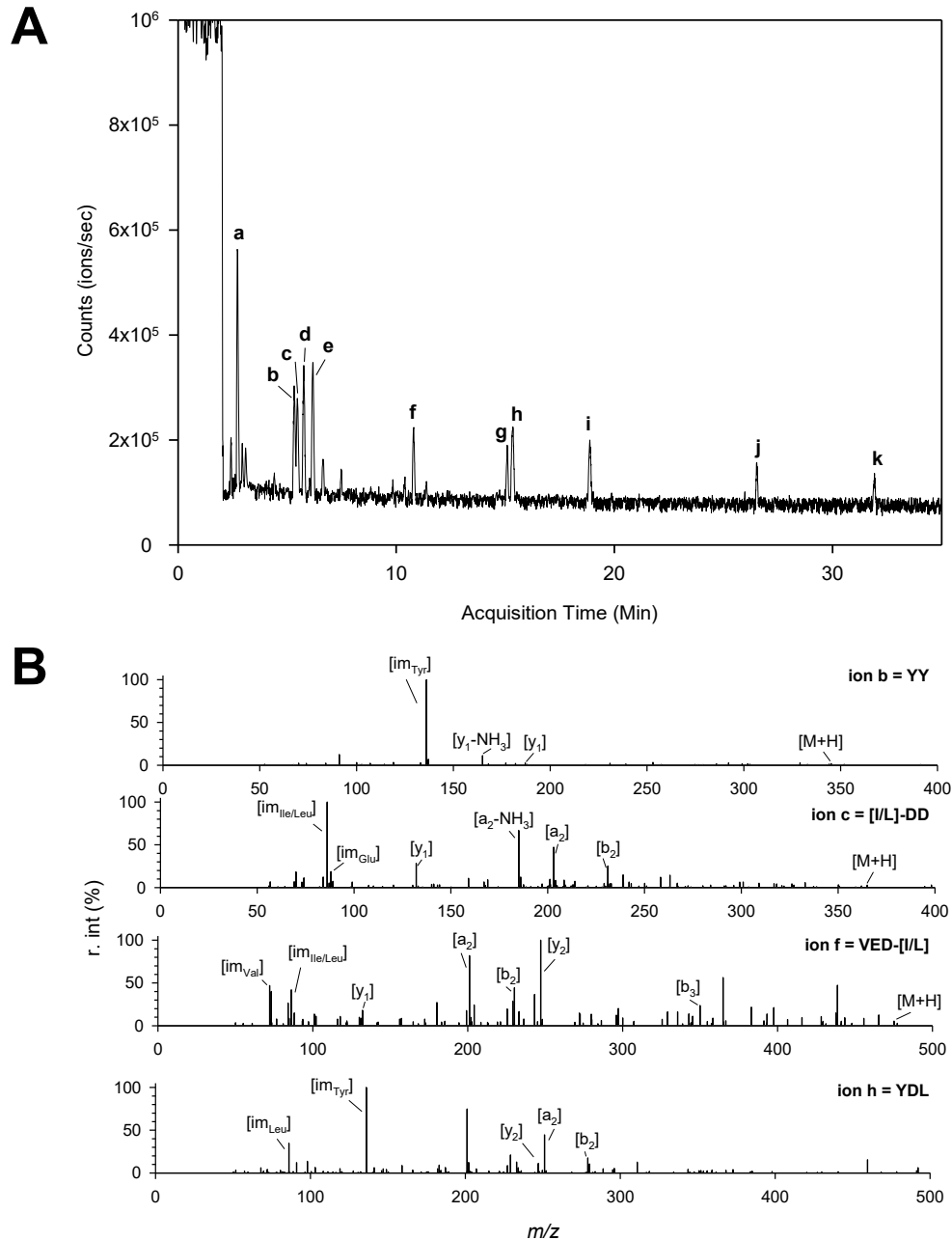


Figure 27. LC-MS analysis of fraction 6 of SAX Separation 5 (F6_S5), represented using the (A) base peak chromatogram. The mass and putative identity of lettered peaks are summarized in Table 19. (B) The product ion spectra used for manual *de novo* sequencing of peaks in part (A), annotated with the detected fragment ion matching to the identified peptide sequence, with assistance from mMass software Version 5.5.

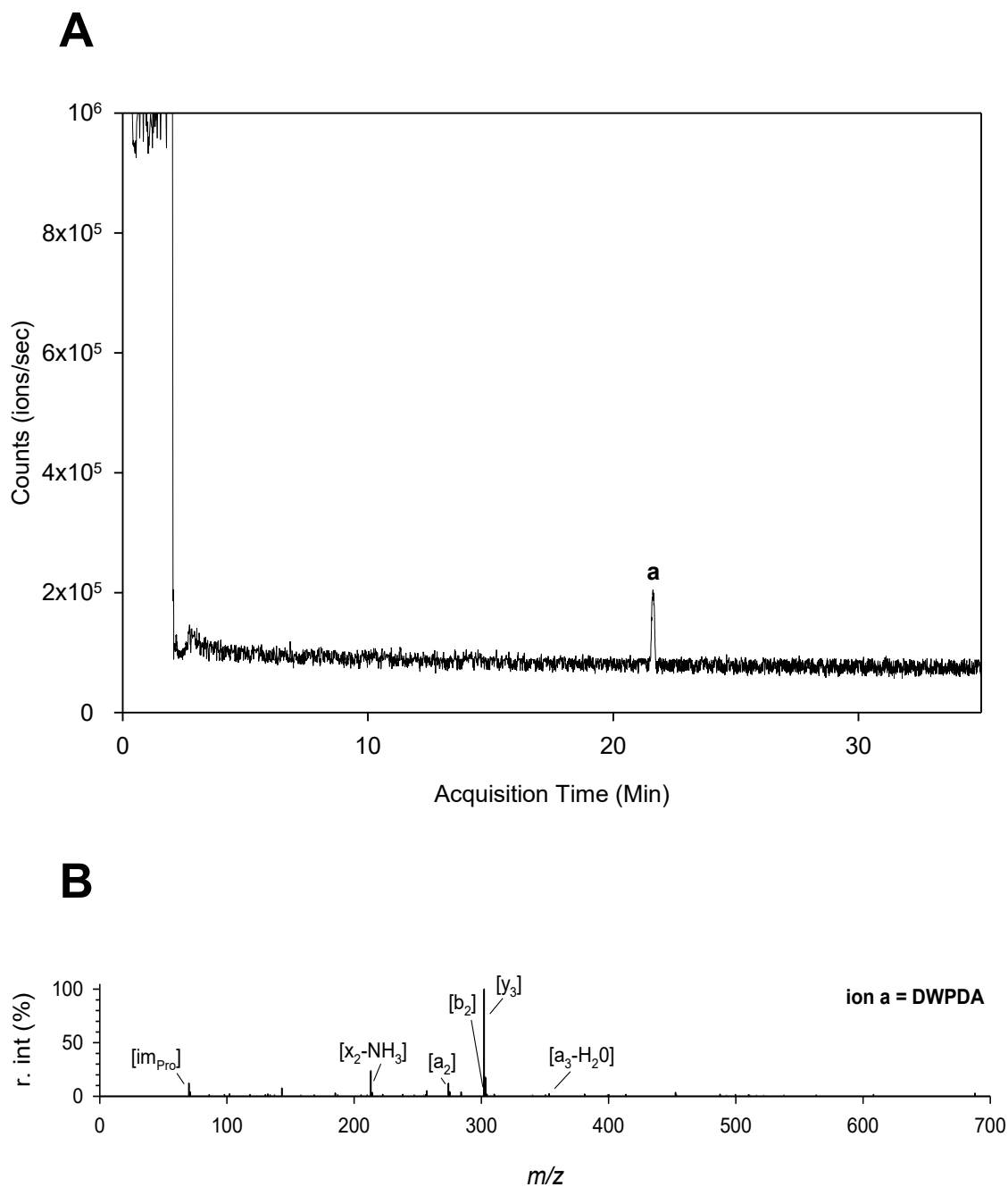


Figure 28. LC-MS analysis of fraction 7 of SAX Separation 5 (F7_S5), represented using the (A) base peak chromatogram. The mass and putative identity of lettered peaks are summarized in Table 19. (B) The product ion spectra used for manual *de novo* sequencing of peaks in part (A), annotated with the detected fragment ion matching to the identified peptide sequence, with assistance from mMass software Version 5.5.

Table 19. Candidate bioactive peptides identified in glucose uptake modulating fractions from GF Separation 4 and SAX Separation 5. The ions chosen from each fraction for further analysis were selected from the dominant peaks in base peak chromatograms (Figure 22A – 28A). Sequences for each ion were generated by using software-assisted database searching using the Agilent MassHunter software Version B.07.00, or by manual *de novo* identification with assistance from mMass software Version 5.5. The stimulating or inhibiting activity is summarized under each fraction label.

Sample	Ion Label	RT (Min)	Precursor Ion Mass (M+H)	Rel. Peak Vol. (%)	Peptide Identification Method	
					Database Searching	<i>De novo</i> Sequencing
F2_S4 Stimulating	a	3.1	120.080	14.73	No Match	Phe Immonium
	b	3.4	360.213	37.12	IVE	[I/L]-VE
	c	4.5	332.182	33.61	No Match	a) [I/L]-[I/L]-S b) TV-[I/L]
	d	5.0	362.192	24.33	ETI	ET-[I/L]
	e	6.6	374.229	24.53	No Match	[I/L]-[I/L]-E
	f	9.9	457.265	19.02	PIVE	P[I/L]VE
	g	11.7	346.198	37.64	VDI	VD-[I/L]
	h	11.8	360.213	29.91	VEI	VE-[I/L]
	i	12.3	795.472	100.00	No Match	IIAPPER
	j	17.4	360.213	36.88	IDI	[I/L]-D-[I/L]
	k	17.6	374.229	28.70	No Match	[I/L]-E-[I/L]
F3_S4 Stimulating	a	3.6	302.207	27.04	IVA	[I/L]-VA
	b	3.9	304.187	21.03	No Match	AT-[I/L]

Sample	Ion Label	RT (Min)	Precursor Ion Mass (M+H)	Rel. Peak Vol. (%)	Peptide Identification Method	
					Database Searching	<i>De novo</i> Sequencing
F3_S4 Stimulating	c	7.2	430.266	32.88	IGIGA	-
	d	8.4	302.208	31.42	AVI	AV-[I/L]
	e	10.2	302.208	47.83	VAI	VA-[I/L]
	f	12.9	330.239	26.96	No Match	VV-[I/L]
	g	14.1	332.218	21.12	SII	S-[I/L]-[I/L]
	h	15.9	316.223	37.96	IAI	[I/L]-A-[I/L]
	i	16.2	373.245	24.16	No Match	[I/L]Q-[I/L]
	j	17.7	346.234	19.65	ITI	[I/L]-T-[I/L]
	k	18.4	344.255	100.00	IVI	[I/L]-V-[I/L]
	l	20.2	344.255	21.25	No Match	[I/L]-V-[I/L]
	m	21.7	401.276	17.88	VIGI	V-[I/L]-G-[I/L]
F6_S4 Inhibiting	a	2.8	281.150	43.05	No Match	VY
	b	3.1	120.080	100.00	No Match	Phe Immonium
	c	4.9	263.139	43.84	No Match	PF
	d	5.3	295.165	-	No Match	[I/L]-Y
	e	6.2	295.165	64.52	No Match	[I/L]-Y
	f	7.9	295.165	82.74	No Match	Y-[I/L]
	g	15.5	391.198	48.53	VSW	SVW
	h	15.9	408.249	38.77	IYY	[I/L]-[I/L]-Y
	i	16.3	405.213	25.93	No Match	VTW

Sample	Ion Label	RT (Min)	Precursor Ion Mass (M+H)	Rel. Peak Vol. (%)	Peptide Identification Method	
					Database Searching	<i>De novo</i> Sequencing
F6_S4 Inhibiting	j	18.8	313.155	38.72	No Match	FF
F7_S4	a	3.1	120.081	100.00	No Match	Phe Immonium
Stimulating	b	9.7	329.150	4.74	No Match	FY
	c ₁	16.4	473.203	3.81	GYSF	GYSF
	c ₂	16.3	405.213	7.52	No Match	VTW
	d	18.2	401.219	3.95	PVW	PVW
F9_S4	a	3.2	120.081	23.19	No Match	Phe Immonium
Inhibiting	b	6.0	276.135	24.19	No Match	AW
	c	10.1	304.166	11.88	No Match	VW
	d	11.9	329.150	16.59	No Match	YF
	e	15.3	318.181	88.62	No Match	[I/L]-W
	f	17.3	318.182	19.27	No Match	[I/L]-W
	g	19.7	318.182	26.67	No Match	W-[I/L]
F6_S5	a	2.7	390.187	100.00	IEE	[I/L]-EE
Stimulating	b	5.3	345.145	72.70	No Match	YY
	c	5.5	362.156	69.23	No Match	[I/L]-DD
	d	5.8	390.187	72.98	E EI	EE-[I/L]
	e	6.2	424.207	84.81	IYE	[I/L]-YE

Sample	Ion Label	RT (Min)	Precursor Ion Mass (M+H)	Rel. Peak Vol. (%)	Peptide Identification Method	
					Database Searching	<i>De novo</i> Sequencing
F6_S5 Stimulating	f	10.8	475.239	53.90	No Match	VED-[I/L]
	g	15.1	521.260	52.85	YPIE	YP-[I/L]-E
	h	15.4	410.192	70.70	No Match	YD[I/L]
	i	18.9	503.271	61.58	EEII	EE-[I/L]-[I/L]
	j	26.5	415.706	49.54	No Match	-
	k	31.9	603.314	47.84	EGIVW	EG-[I/L]-VW
F7_S5 Stimulating	a	21.7	603.240	100.00	PTCPDA	DWPDA

Another strategy to identify putative bioactive peptide sequences involves comparing all bioactive fractions for the identification of common sequences and this was investigated using the peptide identities from SPF fractions reported using software-assisted database searching methodology and validated using modified criteria (Appendix A, Table A3 – A9). The peptides Val-Gly-Val-Pro-Ile, Ile-Asp-Ile, Ile-Asp-Phe, Ile-Glu-Phe, and Ile-Val-Ile were common to the adjacent bioactive fractions 2 and 3 of Separation 4 (Figure 15), and the peptides Glu-Tyr-Leu-Pro-Asp-Glu-Gln, Glu-Tyr-Ile-Pro-Asp-Glu-Gln, Ile-Asp-Asp-Ile, Ile-Asp-Val-Glu, Ser-Ile-Glu-Asp, Thr-Trp-Pro-Trp, Glu-Glu-Phe, Glu-Glu-Ile, Ile-Glu-Glu and Tyr-Glu-Phe were common to the adjacent fractions 6 and 7 of Separation 5 (Figure 16). Alternatively, peptides Tyr-Pro-Ile-Glu, Ile-Thr-Asp-Tyr, Ile-Tyr-Glu and Thr-Trp-Pro-Trp were each separately identified in fractions of both GF Separation 4 and SAX Separation 5. The few duplicate peptides identified in bioactive fractions of these separations demonstrates the benefits of using multiple separation formats for the comprehensive peptide detection of LMW protein hydrolysates, where using a single chromatographic format would have limited the identification of potential bioactive peptides. No common peptide sequences were identified within the glucose uptake inhibiting fractions 6 and 9 of Separation 4 (Figure 15), but their compositional similarity could indicate their activities are mediated by the characteristics of residues at each position in an amino acid sequence rather than any one particular sequence, as others have described for ACE-inhibition (Wu et al., 2006), DPP-IV inhibition (Lan et al., 2015), and antioxidant activity (Saito et al.,

2003) that each also highlight the importance of C-terminal aromatic residues for potent bioactivity.

The duplicate identification of peptide sequences by software-assisted database searching also occurred within individual bioactive fractions of the SPF; ions with sequences reported as Ile-Val-Glu, Val-Gly-Val-Asp-Gly-Phe, Val-Ile-Thr, Ser-Ile, Ile, Ile-Thr-Ile, Ile-Ile-Ile, Ile-Asn-Ile, Ile-Ile-Tyr, Ile-Gly-Glu-Glu, Ile-Glu-Gu-Glu, Ile-Glu-Asp-Asp-Ile, Tyr-Asp-Asp-Ser-Leu were each detected at multiple retention times and the identical *m/z*. With the exception of Val-Gly-Val-Asp-Gly-Phe, all other peptides contained at least one isobaric (Ile/Leu) residue. When considered with the low abundance of Leu residues reported in SPF peptides (Figure 22) and the greater hydrophobicity of Leu compared to Ile (Wimley and White, 1996), duplicate peptide sequence reporting at different RTs is likely indicative of both Ile and Leu peptide variants in samples where this duplicate identification was observed to occur. According to Lahrichi et al. (2012), the LC-MS identification of di-, tri- and tetrapeptides containing entirely BCAA cannot be performed successfully without the consideration of their retention times because the incorporation of Leu increases peptide hydrophobicity thereby delaying its elution from the reversed phase column used for LC-MS analysis. The assignment of an accurate sequence to each identified ion requires the evaluation of their expected retention times using synthetic peptide standards. Insofar that peptides containing Ile or Leu cannot be differentiated by their MS/MS spectra and many duplicates

included more than one of these residues, the reported identities from database searching may therefore represent false positives.

5.4.7 Consensus Residue Positioning in Peptides from each Fraction

The presence of a consensus motif that could describe peptide sequences within bioactive fractions was investigated using the peptides identified by software-assisted database searching and validated using the modified criteria (Appendix A, Table A3 – A9). The WebLogo tool (Crooks et al., 2004) was used to express the frequency of residues and amino acid classes found at positions in SPF peptides of different lengths from GF Separation 4 (Figure 15) and SAX Separation 5 (Figure 16) with either stimulating (Figure 29) and inhibiting (Figure 30) bioactivity. In fraction 2 of GF Separation 4 (Figure 29A), BCAAs were found at all peptide positions, but Ile (or Leu) was positioned at both N- and C-terminals with the equivalent frequency across all peptide lengths; Val was the next most abundant residue at the N-terminal. Glu and Asp residues were found with high frequency at each position and were least frequently identified at the N-terminal. In fraction 3 of GF Separation 4 (Figure 29B), both Ile and Val were found at all positions in tri- and tetrapeptides more frequently than in fraction 2 but were absent from the C-terminals in pentapeptides, indicative of non-specific protein hydrolysis rather than by cleavage with pepsin, trypsin or chymotrypsin. However, due to the generation of these sequences by database searching and its apparent limitation towards differentiating Ile and Leu residues described above (Figure 18), the Ile residues at the C-terminal may in fact be a Leu residue. A higher frequency of

anionic residues was measured in fraction 2 compared to fraction 3, while peptides in fraction 3 reported two-fold increases of their Gly and Ala contents compared to fraction 2. Aromatic amino acids present in fractions 2 and 3 were most frequently localized at the C-terminal, and their detection at non-C-terminal positions are indicative of a missed enzyme cleavage because of the specificity of both pepsin and chymotrypsin for aromatic residues.

In fractions 6 and 7 of SAX Separation 5 (Figure 29C,D), anionic residues were positioned with high frequency at all positions. Glu was frequently positioned at the middle residue position in tripeptides and together with Asp, they were consistently located adjacent to the N-terminus residue in peptides of all lengths. Aromatic amino acids were found at all positions at all peptide lengths, and their elevated frequencies in fractions from SAX Separation 5 (Figure 29C,D) compared to fractions 2 and 3 (Figure 29A,B) from GF Separation 4 (11.5 - 14.3 % to 4.0 – 6.6 %) could be an indication that anionic residues are associated with aromatic residues more frequently than BCAAs in SPF peptides. Bioactive fractions from SPF separations from gel filtration and strong anion exchange chromatography that exhibit stimulating activity had different amino acid compositions, sequence logos, and lacked a universal common characteristic that could describe all stimulating fractions from both Separations 4 and 5, suggesting the possibility that multiple peptide sequences are capable of modulating the bioactivity of SPF.

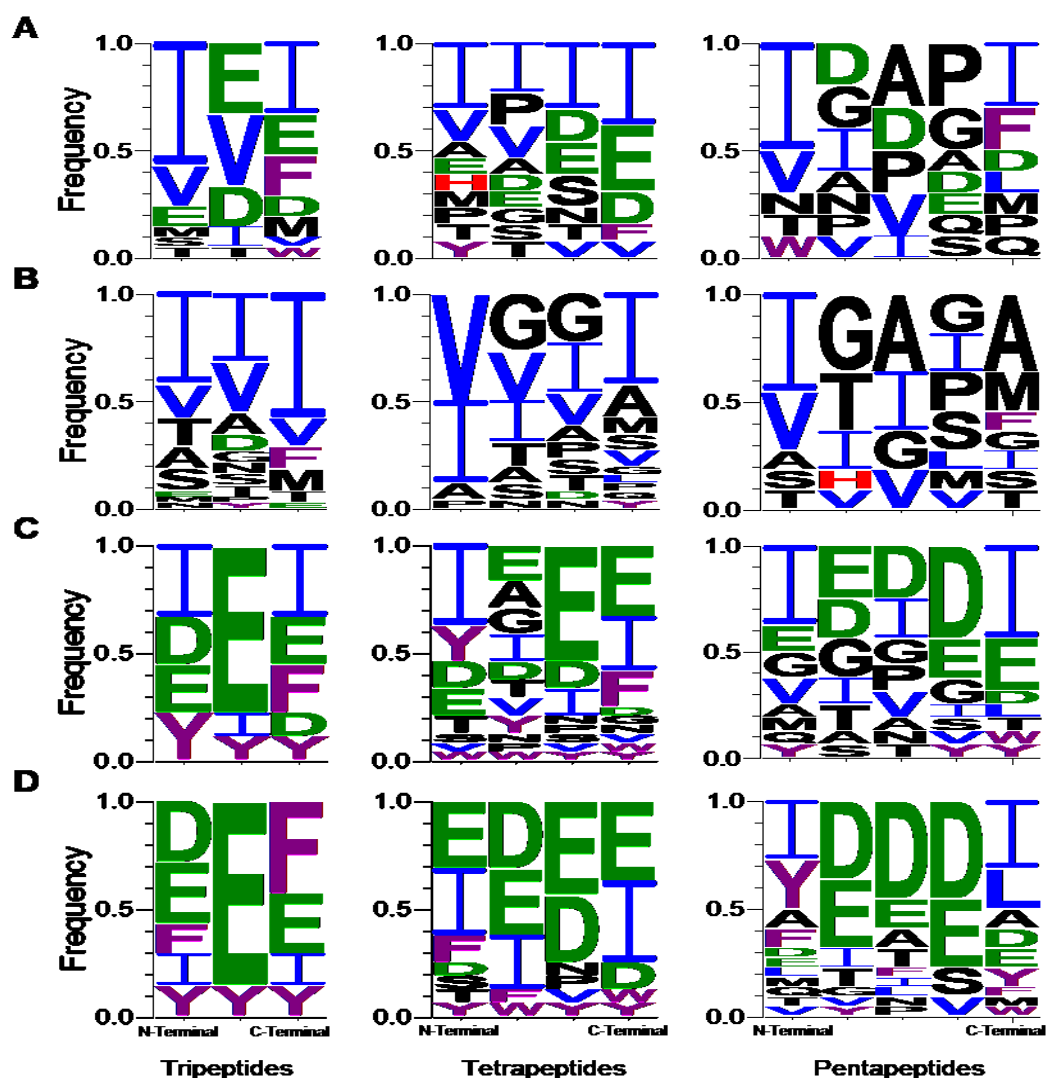


Figure 29. Sequence logos of peptides in glucose uptake stimulating (A) fraction 2, and (B) fraction 3 of GF Separation 4 and (C) fraction 6, and (D) fraction 7 of SAX Separation 5. Logos were generated with WebLogo 3 (outlined in Section 3.14), using peptide sequences identified by software-assisted database searching and validated using the modified criteria. Each logo represents the frequency of residues identified at each position for peptides of the indicated lengths. Amino acid sequences are reported using single letter amino acid codes, and colour coding identify cationic residues (red), anionic residues (green), branched chain amino acids (blue), neutral amino acids (black) and aromatic amino acids (purple).

In fraction 6 of GF Separation 4 (Figure 30) where aromatic residues represented > 35 % of all residues among the identified peptides in this fraction (Figure 21), these residues were positioned at the C-terminal in all but one identified peptide and with disproportionate preference to Tyr than to other aromatic residues. BCAAs were also frequently located at the N-terminals and the middle residue positions of tripeptides could be occupied by any of the most abundant residues identified in SPF. Therefore Ile-X-Tyr motif, where X represents any amino acid, can describe a large proportion of the tripeptide composition of this fraction. When considering that hydrophobic forces were suspected to affect the retention times of dipeptides during GF Separation 4, the sequence logos of tripeptides also

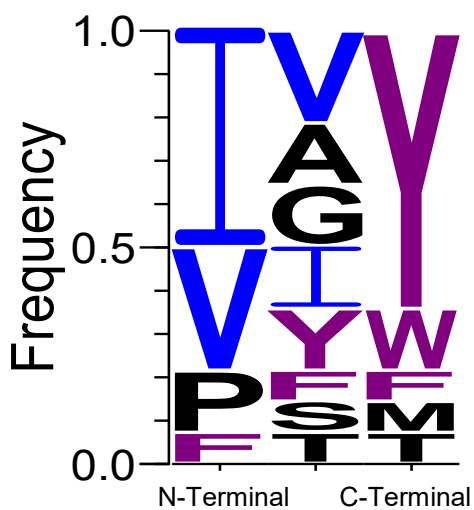


Figure 30. Sequence logo of tripeptides in the glucose uptake inhibiting fraction 6 of GF Separation 4. Logos represent the frequency that indicated residues will be found located at that position. Logos were generated with WebLogo 3 (outlined in Section 3.14), using peptide sequences identified by software-assisted database searching and the modified validation criteria. Amino acid sequences are reported using single letter amino acid codes, and colour coding identify branched chain amino acids (blue), aromatic amino acids (purple), and neutral amino acids (black).

suggests similar inconsistencies in peptide retention; the consensus sequences of tripeptides from all bioactive fractions from GF Separation 4 are represented as Ile-Glu-Ile (373.45 Da), Ile-Ile-Ile (357.49 Da), and Ile-Val-Tyr (393.48 Da) for fractions 2, 3 and 6, recovered at RT's of 120 min, 140 min and 200 min, respectively (Figures 29 and 30). These results demonstrate that although peptides in fraction 6 are larger than those in fraction 2 and 3, their column RT's on Bio-Gel P-2 media predicted they are in fact smaller in size, and this inconsistency supports previous findings that characteristics other than MW affect peptide retention during the GF Separation 4 described in Section 5.4.6.

Sequence logos also reveal the specificity of enzymatic hydrolysis during SPF production. The absence of Arg and Lys residues at the C-terminal position would also suggest that trypsin activity was low, but the absence of these residues at internal peptide positions also suggests this may occur as a result of their data acquisition and analysis by LC-MS rather than by enzymatic inefficiencies or true absence from the SPF. The higher frequencies of Trp, Tyr, Phe positioned at the C-terminal suggests that pepsin and chymotrypsin activity were not impacted. Leu is another classical target of both enzymes that was infrequently identified at the C-terminus position, except among larger peptides that are compatible with peptide identification by a database searching approach (Panchaud et al., 2012). This finding and the disproportionate abundance of Ile highlights the challenges of assigning these residues to detected LMW peptides by a database searching approach. The identification of non-peptic, non-tryptic and non-chymotryptic

enzymatic targets positioned to the C-terminal (Gly, Ala, Ser, Val) are perhaps indicative of the presence of naturally-occurring salmon muscle proteases or alkaline hydrolysis products resulting from high-alkali solubilization of salmon muscle proteins. This could be tested by screening for peptides in the protein extract recovered following their isoelectric precipitation from Atlantic salmon mince.

5.4.8 Previously Identified Bioactive Peptides

Peptides reported by the various identification methodologies used to analyze bioactive fractions from SPF separations by gel filtration (Separation 4, Figure 15) and strong anion exchange (Separation 5, Figure 16) chromatography were searched in the BIOPEP bioactive peptide database to reveal if any of the sequences were previously reported to exhibit biological activities (Table 20). Most functional peptides exhibit angiotensin-II converting enzyme (ACE) inhibition, dipeptidyl peptidase IV (DPP-IV) inhibition and antioxidative activities, with EC_{50} values that range between 1.0 μ M and 1.0 mM. These activities may be desirable characteristics of natural health products due to the antihypertensive effects of ACE-inhibitors and anti-diabetic effects of DPP-IV inhibitors (Dziuba and Darewicz, 2007). However, the EC_{50} values of peptides to these targets are orders of magnitude higher than individual peptide concentrations within the SPF during the benchmark (glucose uptake in L6 mouse myotubes) activity determined at 1 ng/mL (Table 14A), and both ACE and DPP-IV proteins are minimally expressed in skeletal muscle tissues compared to their expression in the gut or pancreatic

tissues, respectively. An EC₅₀ value of 110 µM was measured for ACE inhibition of the SPF (Girgih et al., 2016) that measures 6 – 15-fold lower (more potent) than hydrolysates of *Salmo salar* produced using various microbial protease extracts; in other functional salmon protein hydrolysates, this bioactivity was attributed to peptides Gly-Pro-Ala-Val, Val-Pro, Val-Cys, Phe-Phe, Pro-Pro, Glu-Pro, Trp, Phe, and Tyr (Neves, Harnedy, O’Keeffe, and FitzGerald, 2017). The enzyme digestion strategy for the SPF could be an important factor mediating its potent bioactivity because other researchers have identified bioactive peptides with potent antioxidant (Saito et al., 2003), antihypertensive (Wu et al., 2006) and DPP-IV inhibition (Lan et al., 2015) functionalities when peptides contain C-terminal aromatic residues and that are generated by both pepsin and chymotrypsin in the SPF methodology.

Unfortunately, the concentration of individual peptides found within the SPF when assayed at 1 ng/mL were not directly quantified, but an estimated minimum peptide concentration of ~ 1 µg/mL required to observe DPP-IV and ACE inhibition suggests both proteins were not the targets of SPF that led to the stimulation of glucose uptake in cultured L6 myotubes. In contrast to this model of skeletal muscle, the anti-diabetic activity of SPF resulting from *in vivo* human or animal trials are quite capable of interacting with endogenous ACE and DPP-IV proteins and these interactions could potentially contribute to the improvements of whole-body glucose metabolism as previously measured following SPF consumption in mice by intraperitoneal glucose tolerance tests (Chevrier et al., 2015). However,

lab-to-lab variations of the identical assay and the use of different assays to express identical functions makes it difficult to compare functional peptides to those identified by other research groups. Glucose uptake stimulating peptides reported in BIOPEP includes Val-Leu, Leu-Val, Ile-Val, Ile-Leu, Leu-Ile, Ile-Ile, and Leu-Leu require a minimum peptide concentration of 1 mM to exhibit a stimulating effect (Morifuji et al., 2009), suggesting these peptides do not modulate the effects of SPF in cultured L6 myotubes successfully tested at 1 ng/mL.

Table 20. Previously identified bioactive peptides reported in the bioactive peptide database, BIOPEP. Peptides are reported using single letter amino acid codes. n.d. – not determined.

Fraction	Peptide	BIOPEP ID	Activity	EC ₅₀ (μM)	Protein Origin	Reference
F2_S4	VIPEL	8202	ACE Inhibitor	799.24	Pork	Escudero et al., 2010
	IIAPPER	9436	Antioxidative	1010.00	<i>Gryllodes sigillatus</i>	Zielińska et al., 2018
	LVE	7746	ACE inhibitor	14.20	Pearl Oyster	Qian et al., 2007
F3_S4	VVV	8318	Dvl protein binding	n.d.	Synthetic	Lee et al., 2009
	AVL	9060	ACE inhibitor	7.08	Synthetic	Wu et al., 2006
	LVL	3421	ACE inhibitor	12.30	Porcine Plasma	Hazato and Kase, 1986
F6_S4	IAY	9071	ACE inhibitor	12.59	Synthetic	Wu et al., 2006
	VSW	9066	ACE inhibitor	23.44	Synthetic	Wu et al., 2006
	IVY	7541	ACE inhibitor	0.48	Wheat Germ	Matsui et al., 2000
	VY	3492	ACE Inhibitor	7.10	Sake	Saito et al., 1994
			Antioxidative	n.d.	Potato	Cheng et al., 2010
			DPP IV inhibitor	n.d.	Synthetic	Lan et al., 2015
	PF	8854	DPP IV inhibitor	n.d.	Synthetic	Lan et al., 2015
	IY	3383	ACE inhibitor	2.10	Sardine	Matsufuji et al., 1994
			Antioxidative	n.d.	Soybean	Beermann et al., 2009
			ACE inhibitor	18.00	Sardine	Matsufuji et al., 1994
	LY	3381	Antioxidative	n.d.	Soybean	Beermann et al., 2009
			DPP IV inhibitor	n.d.	Synthetic	Lan et al., 2015
	YL	3550	ACE inhibitor	122.00	Beta-lactoglobulin	Mullally et al., 1997
			Anxiolytic	n.d.	Synthetic	Kanegawa et al., 2010
			DPP IV inhibitor	n.d.	Synthetic	Lan et al., 2015

Fraction	Peptide	BIOPEP ID	Activity	EC ₅₀ (μM)	Protein Origin	Reference
F7_S4	FY	3556	ACE inhibitor	25.00	Corn	Yako et al., 1996
F9_S4	AW	7543	ACE inhibitor	10.00	Synthetic	Cheung et al., 1980
		8460	Antioxidant	830.40	<i>Mactra veneriformis</i>	Liu et al., 2015
	VW	8695	DPP IV inhibitor	n.d.	Milk	Nongonierma and FitzGerald, 2013
		3486	ACE inhibitor	1.40	Sake lees	Saito et al., 1994
		8461	Antioxidant	3588.00	<i>Mactra veneriformis</i>	Liu et al., 2015
		8928	DPP IV inhibitor	n.d.	Synthetic	Lan et al., 2015
		9387	Alpha-glucosidase inhibitor	22.60	Sardine	Matsui et al., 1999
	YF	8935	DPP IV inhibitor	n.d.	Synthetic	Lan et al., 2015
	IW	7544	ACE inhibitor	4.7	Chum Salmon	Ono et al., 2003
		8807	DPP IV inhibitor	n.d.	Synthetic	Lan et al., 2015
	LW	2832	Renin inhibitor	n.d.	Synthetic	Udenigwe et al., 2012
		3389	ACE inhibitor	50.00	Casein	Maruyama et al., 1987
		8462	Antioxidant	3639.00	<i>Mactra veneriformis</i>	Liu et al., 2015
	WI	8688	DPP IV inhibitor	993.40	Milk	Nongonierma and FitzGerald, 2013
		8679	DPP IV inhibitor	138.70	Milk	Nongonierma and FitzGerald, 2013
WL		8677	DPP IV inhibitor	43.60	Milk	Nongonierma and FitzGerald, 2013
	9107	ACE inhibitor	41.40	Synthetic	Martin, 2015	
F6_S5	IEY	8227	ACE inhibitor	182.00	Peanut	Jimsheena and Gowda, 2011
	LEE	9031	ACE inhibitor	100.00	Synthetic	Wu et al., 2006
	YY	8948	DPP IV inhibitor	n.d.	Synthetic	Lan et al., 2015
F7_S5	DYY	7924	Antioxidative	n.d.	Synthetic	Saito et al., 2003

5.4.9 *In silico* Digestion of a Curated Protein Library

In Section 5.4.4, the peptide compositions of bioactive fractions from GF Separation 4 and SAX Separation 5 were determined by software-assisted database searching. Then reported datasets were validated using the modified criteria to report peptide sequences with enhanced confidence. However, even after applying these validation criteria, the reported datasets still included a 10-fold greater abundances of Ile residues than Leu residues and that were frequently localized to the C-terminal. These findings were inconsistent with previous amino acid analyses of unfractionated SPF (Roblet et al., 2016), as well as with the expected cleavage specificity based on the activities of pepsin, trypsin and chymotrypsin enzymes. This indicates that many of the sequences accepted following the modified validation criteria still represented false positive hits and that this strategy was not entirely sufficient for the validation of LMW hydrolysates. Additionally, peptides detected by LC-MS but excluded by the validation criteria may also represent false negatives because scoring parameters applied to database searching are not specifically designed for the identification of short (< 6 AA) peptides (Panchaud et al., 2012). Furthermore, the possible activity of endogenous proteases or by alkaline hydrolysis on salmon protein could indicate that enzymatic specificity does not account for the generation of all peptides generated during SPF manufacturing.

In an attempt to improve the accuracy of assigning amino acid sequences to detected ions in bioactive fractions of SPF, a modified *in silico* digestion (computer-

assisted protein hydrolysis) approach, outlined in Section 3.13, was developed in the present study and was based on the formation of peptides from either enzymatic or non-enzymatic mechanisms from the primary sequences of 10 previously identified progenitor proteins (termed “curated protein library”) to the SPF (Table 21).

The reported compound lists of each bioactive SPF fraction, generated by the interpretation of the mass spectral patterns alone using the Agilent Molecular Feature Extraction (MFE) algorithm (Section 3.13), were completely annotated combining the results of software-assisted database searching, manual *de novo* sequencing and *in silico* digestions of the curated protein library applying both specific (enzyme-mediated) and non-specific proteolytic activities (Appendix A, Table A14 – Table A20). By the latter approach, peptide sequences predicted from the hydrolysis of the abundant progenitors composing the high-alkali solubilized salmon muscle protein extract preferentially reveal the identities of the reported compounds using the MFE, then by peptides derived from the complete *Salmo salar* proteome. The annotation of MFE compound lists was recently performed by searching the Metlin database (Zhou et al., 2018) to identify and quantify di- and tripeptides from chicken muscle, demonstrating that methodologies that incorporate both *de novo* sequencing and a high-quality prediction of sample components using the MFE, are sufficient to confidently identify LMW peptide sequences. However, no distinction was made for the differentiation of Ile and Leu, indicating that this methodology still cannot accurately differentiate isobaric

residues. Furthermore, atypical fragmentation patterns of short precursor ions during MS-analysis and high background noise reduces the quality (accuracy with respect to peptide identification) of MS/MS spectra by introducing peaks that cannot be calculated from the detected peptide and could prevent the suitability of peptide identifications by exclusively *de novo* means. Therefore, the compositions of bioactive SPF fractions were re-evaluated using an *in silico* digestion approach and the dominant peaks in BPC of each were re-evaluated (Table 21).

The common m/z 120.0809 Da precursor ion, putatively identified as the Phe immonium ion, was detected in most bioactive fractions with high intensity, but its identity could not be reliably determined by any of the methodologies investigated thus far. As previously described in Section 5.4.6, its identification as a precursor ion in MS¹ could indicate the presence of the free amino acids in these fractions, but these compounds and dipeptides were not targeted by software-assisted database searching. Importantly, *in silico* digestions of the curated protein library putatively detected the presence of other free amino acids including Ile or Leu and Trp as intact precursor ions based on the detection of their y_1 ion in MS¹, but the diagnostic ions to Tyr, Arg and Lys as free amino acids that would be expected to form as a result of the specific activities of pepsin, trypsin and chymotrypsin, were not detected. The polar nature of Tyr, Arg, and Lys side-chains could make them susceptible to modifications and/or to the addition of adducts that modify their masses and prevent their identification.

In silico digestions also supported the identification of dipeptides within bioactive fractions 6, 7 and 9 of GF Separation 4 (Figure 15) and fraction 6 of SAX Separation 5 (Figure 16). The predicted dipeptide sequences were largely representative of the most abundant residues of identified peptides in SPF fractions considering the enzymatic procedure; the X₁X₂ dipeptide motif where X₁ represents Ile/Leu, Gly, Ala, Asp, Glu, Tyr, Trp, and/or Phe, and X₂ represents Tyr, Trp, Phe and/or Leu could represent potential dipeptide mediators of glucose uptake due to both their presence and abundance within bioactive SPF fractions. The dipeptides previously identified using the low throughput manual *de novo* sequencing approach (Table 19) were each predicted by the specific *in silico* digestion of the curated protein library, along with others that were previously not identified by a *de novo* approach. These findings demonstrate the potential value of the use of selective *in silico* digestions in high throughput methodology for valid LMW peptide identification, independent of database searching. These findings also highlight the potential value of this approach for the characterization of all proteinaceous material (free amino acids, dipeptides and larger oligopeptides) within protein hydrolysate in a single analysis compared to software-assisted database searching that cannot accomplish this task.

As previously states, peptides that contain a C-terminal Ile residue could potentially be generated during SPF processing due to mechanisms of non-specific protein hydrolysis that occurs during alkaline-solubilization of Atlantic salmon muscle (Section 5.3). However, the specificity of pepsin for the peptide bond C-terminal of

Leu residues but not Ile, and considered together with the frequent prediction of peptides containing Leu positioned at the C-terminal following the *in silico* digestion of the curated protein library (Table 21), suggests that identified peptides containing C-terminal Ile residues are representative of false positives and where SPF peptides produced by non-specific mechanisms are expected to be found at lower proportions than those produced by the specificity of enzymatic hydrolysis. Following the *in silico* digestion approach, multiple sequences were commonly predicted for each detected ion following both the specific and non-specific digestion criteria, so the interpretation of each MS/MS spectra provided by database searching contributes strong evidence for the assignment of any of the *in silico* predicted peptide sequences. The comparison of RT and MS/MS spectra of synthetic peptides still represents the gold standard for the validation of peptide identity and therefore the distinction of isobaric residues among LMW peptides is not be achievable when performing data-independent acquisitions (Lahrichi et al., 2013). Nonetheless, this approach was demonstrated to improve the predictions of sequences from low mass ions detected within SPF fractions but still could not conclusively assure the identities of peptide containing isobaric residues.

Software-assisted database searching and *de novo* approaches to peptide identification both also rely upon automated interpretations of MS/MS spectra. Atypical fragmentation patterns of short peptides (formation of non-b- and y-series ions) commonly lead to the misinterpretation of residue position by database searching, where the common formation and detection of stable a₂ ions and

dominant immonium ions in MS/MS spectra were often insufficient to determine the peptide sequence. The low quality of MS/MS spectra from short peptides represents the largest challenge to high-throughput peptide identification of LMW peptides, where it has been argued that a true peptide identification should explain all major peaks in the MS/MS spectrum (Chen et al., 2009). *In silico* digestion using the curated protein library was able to provide the additional confidence required to overcome positioning errors reported by database searching and identify peptides not previously validated by database searching. Identified peptides and the progenitor proteins determined by database searching include: Glu-Tyr-Ile-Pro-Asp-Gly (EYIPDG) derived from the *serine-threonine kinase receptor-associated protein* (Accession No. XP_014008296), Glu-Tyr-Leu-Pro-Asp-Gly-Gln (EYLPDGQ) from the *kinesin-like protein KIF9 isoform X1* (Accession No. XP_014036373), and Glu-Ile-Pro-Asp-Gly-Gln (EIPDGQ) from the *RNA-binding protein 33 isoform X7* (Accession No. XP_014036153). These three peptides were identified within fractions of SAX Separation 5 (Figure 16), when in fact the actin-derived peptide Tyr-Glu-Leu-Pro-Asp-Gly-Gln (YELPDGQ), and its derivatives, also match the detected ion masses of the above sequences, demonstrating that when whole species proteomes are used, poor interpretations of MS/MS spectra can still lead to database matches and false positive identifications.

Ultimately, *in silico* digestion simulations cannot support the actual process of digestion of salmon protein with pepsin, trypsin and chymotrypsin used in SPF processing, i.e., the products of pepsin digestion, including those generated from

missed cleavages are not selected as progenitors to trypsin and chymotrypsin cleavage. As a result, *in silico* digestions of *Salmo salar* proteins in this study are hydrolyzed by pepsin, or a combination of trypsin and chymotrypsin. In addition, there are many endogenous proteases in salmon muscle that could be active in the salmon mince most likely upon thawing. Therefore, the theoretical peptide sequences generated by proteomic software (Section 3.15) may not be representative of detected peptides, but their unique characteristics may still be used to broadly describe peptide mixtures with potential stimulating or inhibiting bioactivities.

Table 21. Peptide identification of dominant ions in the base peak chromatograms of fractions from GF Separation 4 and SAX Separation 5 and processed using database searching, *de novo* sequencing and *in silico* digestion approaches to peptide identification. *De novo* sequencing based on the manual interpretation of MS/MS spectra, database searching using MassHunter software and *in silico* digestions using non-specific and specific hydrolysis of the curated protein library were collectively considered to evaluate peptide identity. No Match indicates the ion was not identified by the indicated peptide identification methodology.

Sample	Ion Label	Peptide Identification Methodology			
		<i>De novo</i> Sequencing	Database Searching	<i>In silico</i> Digestions Specific	Non-Specific
F2_S4	a	Phe Immonium	No match	No Match	No Match
	b	[I/L]-VE	I/V E	VLE, LVE, IVE, IDL	EIV, VLE, ELV, LEV, LID, IVE, LDI, DLL, LDL, IDI, VIE, LVE, VEI, IDL, DII, IID, ILD
	c	a) [I/L]-[I/L]-S	No match	SIL	LVT, SIL, TLV, LLS, SLL, ITV, VIT, LSL, VLT, IIS, LSI, LTV
		b) TV-[I/L]	No match	LEA, IEA, EIA, EAL, ALE, AEL	IEA, LEA, ELA, AEL, EAL, VDV, LAE, ALE, IAE, VVD, AEI, EAI, EIA, AIE
	d	ET-[I/L]	E/T I	TIE, TEL, LTE, ETL	IET, ELT, LTE, TLE, ETI, TIE, TEL, LET, ETL, EIT, ITE
	e	[I/L]-[I/L]-E	No match	LLE, LIE, LEL, ELL, EIL	LLE, IEL, ELI, LIE, ILE, LEL, IIE, EIL, ELL
	f	P-[I/L]-VE	P/I V E	No Match	PEIV, VIPE, DPII, PIVE, IVEP, ILPD
	g	VD-[I/L]	V/D I	VVE, VDL	EVV, VLD, IDV, LVD, DLV, DIV, VID, VVE, VDL, VDI, DVI, DVL, VEV

Peptide Identification Methodology

Sample	Ion Label	De novo Sequencing	Database Searching	In silico Digestions	
				Specific	Non-Specific
F2_S4	h	VE-[I/L]	V/E I	VLE, LVE, IVE, IDL	EIV, VLE, ELV, LEV, LID, IVE, LDI, DLL, LDL, IDI, VIE, LVE, VEI, IDL, DII, IID, ILD
	i	IIAPPER	No match	IIAPPER	No Match
	j	[I/L]-D-[I/L]	I/D I	VLE, LVE, IVE, IDL	EIV, VLE, ELV, LEV, LID, IVE, LDI, DLL, LDL, IDI, VIE, LVE, VEI, IDL, DII, IID, ILD
	k	[I/L]-E-[I/L]	No match	LLE, LIE, LEL, ELL, EIL	LLE, IEL, ELI, LIE, ILE, LEL, IIE, EIL, ELL
F3_S4	a	[I/L]-VA	I VA	VAL, LGL, IVA, GLL, AVL	VLA, IAV, ILG, AVL, LLG, LGL, LGI, GLL, LVA, VAL, GLI, ALV, GIL, GII, VAI, IVA, VIA, LIG, LAV, IIG
	b	AT-[I/L]	No match	TLA, TIA, TAL, LTA, ITA, ATL	ATI, ATL, TAL, TIA, AIT, LAT, ALT, LTA, TLA, ITA, VVS, VSV, SVV, TAI
	c	-	I G I G A	VALQ, VAIQ, NAIL, LQVA, LIGQ	AGLLG, VQAL, ALQV, LQVA, VAIQ, IQAV, QAVL, VALQ, KPVS, IALN, IGVA, IINA, LIGQ, NAIL
	d	AV-[I/L]	A/V I	VAL, LGL, IVA, GLL, AVL	VLA, IAV, ILG, AVL, LLG, LGL, LGI, GLL, LVA, VAL, GLI, ALV, GIL, GII, VAI, IVA, VIA, LIG, LAV, IIG
	e	VA-[I/L]	No match	VAL, LGL, IVA, GLL, AVL	VLA, IAV, ILG, AVL, LLG, LGL, LGI, GLL, LVA, VAL, GLI, ALV, GIL, GII, VAI, IVA, VIA, LIG, LAV, IIG
	f	VV-[I/L]	V/V I	No Match	VVI, VVL, VIV, LVV
	g	S-[I/L]-[I/L]	S/I I	SIL	LVT, SIL, TLV, LLS, SLL, ITV, VIT, LSL, VLT, IIS, LSI, LTV
	h	[I/L]-A-[I/L]	I/A I	LIA, IIA	ALI, LAL, IIA, LLA, ILA, LIA, IAL, LAI, AIL
	i	[I/L]Q-[I/L]	No match	IQL, GIAL	AGLL, QII, LLQ, IQL, QLL, QLI, GIAL, VLAA, GILA, AILG

Peptide Identification Methodology					
Sample	Ion Label	<i>De novo</i> Sequencing	Database Searching	<i>In silico</i> Digestions	
				Specific	Non-Specific
F3_S4	j	[I/L]-T-[I/L]	I/T I	LTL	LIT, ITI, TIL, LTL, LLT, ILT, IIT, TLL, TII
	k	[I/L]-V-[I/L]	No match	VIL	VLI, LVI, IVL, VII, VIL
	l	[I/L]-V-[I/L]	No match	VIL	VLI, LVI, IVL, VII, VIL
	m	V-[I/L]-G-[I/L]	V/I G I	LGVL, IGVL	IGVL, LLGV, LGVL, VGLL, GIVL, VILG, LGVI, VVLA, ILGV
F6_S4	a	VY	No match	VY	YV, VY
	b	Phe Immonium	No match	No Match	No Match
	c	PF	No match	PF	PF, FP
	d	[I/L]-Y	-	-	-
	e	[I/L]-Y	No match	YL, LY, IY	IY, YL, LY, YI
	f	Y-[I/L]	No match	YL, LY, IY	IY, YL, LY, YI
	g	SVW	V/S W	SVW	AAMV, MQI, LQM, QIM, QML, CVGL, VGLC, QLM
	h	[I/L]-[I/L]-Y	I/I Y	LYL, IYL, IYI	IYL, LYL, YLL, IYI
	i	VTW	No match	VTW	VTW, AAFP, TER, RTE, ERT, ETR, VGMV, VMVG, ALVC, GVMV, VMGV, LAAM, IAMA, VIAC, AAIM, LAVC
	j	FF	No match	FF	FF
F7_S4	a	Phe Immonium	No match	No Match	No Match
	b	FY	No match	YF, FY	YF, FY
	c1	GYSF	G Y S F	GYSF	PPME, GYSF, FGYS, FSYG
	c2	VTW	No match	VTW	VTW, AAFP, TER, RTE, ERT, ETR, VGMV, VMVG, ALVC, GVMV, VMGV, LAAM, IAMA, VIAC, AAIM, LAVC

Peptide Identification Methodology

Sample	Ion Label	<i>De novo</i> Sequencing	Database Searching	<i>In silico</i> Digestions	
				Specific	Non-Specific
F7_S4	d	PVW	No match	PVW	PVW
F9_S4	a	Phe Immonium	No match	No Match	No Match
	b	AW	No match	WA, AW	AW, WA
	c	VW	No match	VW	VW, WV
	d	YF	No match	YF, FY	YF, FY
	e	[I/L]-W	No match	LW, IW	WL, LW, IW, WI
	f	[I/L]-W	No Match	LW, IW	WL, LW, IW, WI
	g	W-[I/L]	No Match	LW, IW	WL, LW, IW, WI
F6_S5	a	[I/L]-EE	I/E E	LEE, IEE, ELE, EIE, EEL	HPH, EEL, LEE, EEI, IEE, ELE, EIE
	b	YY	No match	No Match	YY
	c	[I/L]-DD	No match	DDL	DID, IDD, DDL, LDD, VED, DVE, EVD, DDI, DLD, EDV
	d	EE-[I/L]	E E I	LEE, IEE, ELE, EIE, EEL	HPH, EEL, LEE, EEI, IEE, ELE, EIE
	e	[I/L]-YE	I/Y E	ELY	ELY, IYE, YEL, LYE, YLE
	f	VED-[I/L]	No match	VEDL, IDDL	RHY, YRH, ELVD, IDDL, LDDL, VEDL, DIDI, IDLD, EDVI, EVDL
	g	YP-[I/L]-E	Y/P I E	No Match	PIYE, YPIE, YELP, ETGKS, QTVSS, TVSSQ, STGSVA, SVTGGT
	h	YD[I/L]	No match	EVY, DLY, DIY	DIY, EYV, DYL, DLY, YEV, LDY, VYE, EVY, YDL
	i	EE-[I/L]-[I/L]	E E I I	IEEL, EEIL	RDGR, ELIE, ILEE, IEEL, LEEI, EEIL
	j	-	No match	No Match	No Match

Peptide Identification Methodology					
Sample	Ion Label	<i>De novo</i> Sequencing	Database Searching	<i>In silico</i> Digestions	
				Specific	Non-Specific
F6_S5	k	EG-[I/L]-VW	E G I V W	EGIVW	EGIVW, GAQKEA, RLDEA, ERADI, QSLGAQ, EAQKQ, NNVLSG, LDRAE, ELDRA, EEVAR, NLKGGD, GKAAQE, NVKNE, QSQIQ, AGAAEKG
F7_S5	a	DWPDA	No match	No Match	DWPDA, GDTHSS

5.4.10 Conclusion

Potential bioactive peptides in the SPF were investigated using gel filtration and strong anion exchange chromatography formats to identify mediators of glucose uptake in the SPF. Fractions generated from these separations were demonstrated to exhibit both stimulating and inhibiting modulation of glucose uptake in cultured L6 myotubes, and the absence of compositional similarities between functional fractions strongly suggested that SPF activity was mediated by distinct peptides. Glucose uptake stimulation was observed in fractions abundant in tripeptides and/or larger oligopeptides, containing both C- and N-terminal BCAAs, and/or anionic amino acids that were frequently positioned as consecutive pairs and often at the penultimate position (adjacent to the C-terminal amino acid). These characteristics could represent common motifs responsible for glucose uptake.

In contrast, glucose uptake inhibition was observed in SPF fractions containing di- and tripeptides with aromatic amino acids positioned at the C-terminus. Peptide identification by database searching was fraught with difficulties possibly associated with extensive false positive and false negative reporting. However, *in silico* digestions of the curated protein library combined with database searching was able to mitigate some of the challenges associated with the identification of short peptides. Comparing putative bioactive sequences from SPF fractions to previously reported bioactive peptides was unsuccessful, particularly considering that glucose uptake was mediated by SPF at concentrations as low as 1 ng/mL. Evaluating the peptide compositions of each fraction still enabled the identification

of compositional features that could be used to uniquely describe aspects of functional peptide fractions.

5.4.11 Future Work

The low quality of short peptide fragment ion mass spectral data may perhaps be improved by incorporating hydrophilic-lipophilic interaction (HILIC) or other normal stationary phase chromatographic chemistries into the separation phase during LC-MS. This may improve the column retention of polar compounds by reducing the co-elution of peptides with similar characteristics and minimize the detection of background noise during MS-analysis that challenge the interpretation and identification of peptide sequences.

Furthermore, systematic strategies that improve the accuracy of LMW peptide identification through bioinformatic means should be investigated without relying on proprietary tools. A systematic quality structure-activity relation (QSAR) study to investigate glucose uptake modulation could identify the compositional similarities required for peptide-mediated stimulation and inhibition of glucose uptake and test to see if individual peptides or motifs are able to influence bioactivities. The binding of peptide ligands to enzyme active sites represent the most common peptide interactions, however peptide motifs, or short linear motifs (SLiMs) that consist of short conserved primary sequences that mediate protein-protein interactions, have been proposed (Neduva and Russell, 2006) as mechanism of peptide-mediated dysregulation of cell signaling, and peptides

based on these primary sequences could act as regulators of these protein-protein interactions.

5.5 Validation of Putative Bioactive Peptide Sequences

Putative bioactive peptide sequences as mediators of glucose uptake require independent confirmation in cultured L6 myotubes to determine the extent of their modulating activity. In previous sections of this thesis, efforts to identify the all the peptide sequences in the bioactive fractions pooled from the various chromatographic separations of SPF were attempted, and these sequences were compared to each other, to bioactive peptide databases, and to previous studies to select the most likely bioactive peptide sequences. The screening of purified peptides is essential to identify how individual components of peptide mixtures influence the activity of unfractionated hydrolysates. Despite the numerous peptides identified in bioactive fractions, it was the unique characteristics of peptides identified from stimulating or inhibiting fractions to glucose uptake that were most interesting.

Tripeptides represented the majority of the peptide content in bioactive fractions determined by database searching and are frequently investigated as bioactive peptides due to their resistance to gastrointestinal enzymatic degradation and ability to cross biological membranes by receptor-mediated processes (Segura-Campos et al., 2011). Furthermore, mimicry of protein and peptide motifs is commonly exploited to investigate the regulation of protein-protein interactions

mediating cell signaling (Tompa et al., 2014), thus the motifs representing the tripeptide content of bioactive fractions could confirm candidates having a modulating effect to glucose uptake. The Ile-X-Ile motif was derived from fraction 3 of GF Separation 4 (Figure 30) that exhibited stimulating activity on glucose uptake, while the Ile-X-Tyr motif was derived from fraction 6 of GF Separation 4 (Figure 31) that exhibited inhibiting activity on glucose uptake. A decision was made to procure chemically synthesized peptides following these potential bioactivity-mediating criteria to validate their individual glucose uptake modulating activities and the importance of each motif. In both fractions, peptide sequences representing these motifs were identified containing Ala, Gly, and Ile located at the center amino acid position. The peptides Ile-Ala-Ile (m/z 316.223), Ile-Gly-Ile (m/z 302.207), Ile-Ile-Ile (m/z 358.270), Ile-Ala-Tyr (m/z 366.202), Ile-Gly-Tyr (m/z 352.187), and Ile-Ile-Tyr (m/z 408.249) were therefore selected for chemical synthesis and their glucose uptake modulating activities are evaluated below.

5.5.1 Objective

The primary objective of this chapter was to evaluate the glucose uptake modulating activity of putative bioactive peptide sequences identified from within bioactive chromatographic fractions of the SPF. It is hypothesized that common characteristics of peptide sequences identified within bioactive SPF fractions can be used to predict for putative bioactive peptides. The comparison of distinct tripeptide motifs are investigated to reveal the influence of Ala, Gly and Ile residues at the middle position, and of Ile and Tyr at the C-terminal position, for their

influence over a stimulating or inhibiting effect on glucose uptake by peptides in cultured L6 myotubes. The bioactivity of peptides was determined under purified and mixed peptide conditions to assess the complexity of glucose uptake modulation by LMW protein hydrolysates.

5.5.2 Peptide Selection

The putative bioactive peptide sequences selected for chemical synthesis were identified by software-assisted database searching and validated using the modified criteria. Peptides identified by this methodology have since been demonstrated to be of low quality (accuracy with respect to peptide identification) as a result of the challenges met by software to differentiate the sequences of precursor ions with identical masses but different RT's, and to accurately interpret the sequences of peptides thought to contain isobaric residues, such as Ile and Leu, or that do not fragment into the typical b- and y-series ions. Therefore, to evaluate whether the selected peptides for chemical synthesis were in fact present in bioactive fractions of SPF, extracted ion chromatograms (EIC) that represent the MS detector response within a narrow m/z range, were generated corresponding to each putative bioactive peptide sequence (Figure 31A – 36A). For all peaks on EICs, a comparison of their MS/MS spectra revealed which peaks were consistent to the target peptide sequence, and these spectra are overlaid in panel B (Figures 31B – 36B). The high intensity peaks in MS/MS spectra are considered diagnostic and are preferentially used to determine peptide sequences

following a *de novo* sequencing approach, rather than peaks detected at lower intensities that may represent background noise.

In general, the EIC of peptides based on the Ile-X-Ile motif contained more peaks than those based on the Ile-X-Tyr motif, findings that perhaps are related to the additional isobaric residue compared to the Ile-X-Tyr motif and indicates the high likelihood that both Ile- and Leu-variants were present in these fractions. The EIC representing the mass of the Ile-Ile-Ile peptide (m/z 356.5 – 360.0, Figure 31A) shows detected ions repeatedly between 5 and 15 min, then again between 21 and 26 min. The annotated compound lists generated in Section 5.4.9 using database search results and *in silico* digestion of the curated protein library (Appendix A – Table A15) showed that ions detected between 5 and 15 min had sequences consistent with Gly-Val-Ile-Ala or Leu-Asn-Leu with the m/z 359.229, while the four ions detected between 21 and 26 min had sequences consistent with Ile-Ile-Ile or Leu-Leu-Leu with the m/z 358.270.

The comparison of the MS/MS spectra (Figure 31B) of each of these four detected ions consistent with Ile-Ile-Ile showed the same diagnostic product ions at m/z 86 and m/z 199. A simulated fragmentation of the Ile-Ile-Ile peptide (Appendix A – Figure A1) demonstrates these 86 and 199 Da ions correspond to the Ile/Leu immonium ion and the Ile/Leu-Ile/Leu a_2 ion, respectively. Another ion repeatedly detected in each spectrum was the m/z 227 ion consistent to the Ile/Leu-Ile/Leu y_2 ion, that when considering the identical MS/MS spectrum of all four distinct

precursor ions, suggests each are representative of the Ile/Leu-Ile/Leu-Ile/Leu sequence.

The EIC representing the mass of the Ile-Ala-Ile peptide (m/z 314.5 – 318.0, Figure 32A) also showed four sharp peaks between 13.5 and 17.1 min. Annotated compound lists (Appendix A – Table A15) showed these sequences correspond to Ile-Ala-Ile as predicted by database searching, Leu-Ile-Ala or Ile-Ile-Ala as predicted by *in silico* digestion with specific enzyme hydrolysis, or 9 permutations of tripeptides containing Ile, Leu and Ala that were predicted by *in silico* digestion with non-specific protein hydrolysis. The comparison of four overlaid MS/MS spectra (Figure 32B) showed diagnostic product ions at m/z 86, m/z 157 and m/z 203 that are consistent with the Ile/Leu immonium ion, the Ile-Ala a_2 ion, and the Ala-Ile y_2 ion (Appendix A – Figure A2).

The EIC representing the mass of the Ile-Gly-Ile peptide (m/z 300.5 – 303.9, Figure 33A) shows peaks grouped between 3 and 11 min, then again group between 14 and 17 min. Annotated compound lists (Appendix A – Table A15) showed that the peptides with RT's between 3 and 11 min were Ile/Leu-Val-Ala or Ala-Val-Ile/Leu as predicted by database searching and *in silico* digestions, whereas the four peptides with RT's between 14 and 17 min were Gly-Ile-Ile, Leu-Gly-Ile, or Leu-Gly-Leu, as predicted by *in silico* digestions. The peptide Ile-Gly-Ile could not be identified directly by database searching, but as a component of the peptides Val-Ile-Gly-Ile and Ile-Gly-Ile-Gly-Ala. The MS/MS spectra (Figure 33B) of the four

peaks with RT's between 14 and 17 min each showed diagnostic peaks at m/z 86, m/z 132, m/z 143, and m/z 189, consistent to the Ile/Leu immonium ion, the Ile/Leu y_1 ion, the Ile/Leu-Gly a_2 ion, and the Gly-Ile/Leu y_2 ion, respectively (Appendix A – Figure A3).

The EIC representing the mass of the Ile-Ile-Tyr peptide (m/z 408.1 – 408.3, Figure 34A) shows one dominant peak at 15.91 min, followed by several other smaller peaks. Annotated compound lists (Appendix A – Table A15) showed this peak and the adjacent peak following at 17.15 min are both predicted with the sequence Ile-Ile-Tyr by database searching, but *in silico* digestions using specific enzyme hydrolysis suggests that they could either be represented by Leu-Tyr-Leu, Ile-Tyr-Leu, or Ile-Ile-Tyr. The comparison of the MS/MS spectra detected for in each peak (Figure 34B) show diagnostic ions at m/z 86, m/z 199, m/z 227, m/z 295, consistent to the Ile/Leu immonium ion, the Ile/Leu-Ile/Leu a_2 ion, the Ile/Leu-Ile/Leu b_2 ion, and the Ile/Leu-Ile/Leu y_2 ion, respectively (Appendix A – Figure A4).

The EIC representing the mass of the Ile-Ala-Tyr peptide (m/z 366.1 – 366.3, Figure 35A) shows one dominant peak at 10.26 min also followed by other smaller peaks, but sequences for these small peaks were not predicted by database searching. For the precursor ion detected at 10.26 min, database searching predicted the sequence Ile-Ala-Tyr, but *in silico* digestions of the curated protein library indicates that Leu-Tyr-Ala, Ala-Leu-Tyr, and Ala-Ile-Tyr could also represent this peptide sequence. Its MS/MS spectrum (Figure 35B) shows diagnostic ions at

m/z 86, m/z 136, m/z 157, m/z 182, m/z 185 and m/z 253 that are consistent with the Ile/Leu immonium ion, the Tyr immonium ion, the Ile-Ala a_2 ion, the Tyr y_1 ion, the Ile-Ala b_2 ion, and Ala-Tyr y_2 ion, respectively (Appendix A – Figure A5).

Finally, the EIC representing the mass of the Ile-Gly-Tyr peptide (m/z 352.1 – 352.3, Figure 36A) shows numerous peaks detected within the desired m/z range. The peak at 9.96 min was represented by the m/z 352.187 ion as calculated by the MFE algorithm, but the other peaks were not processed by mass spectral patterns alone. In fact, the other dominant peak at 9.8 min however was also consistent to the Ile-Gly-Tyr peptide sequence according to database searching. The MS/MS spectrum of only the precursor detected at 9.96 min was evaluated (Figure 36B), where diagnostic ions at m/z 86, m/z 136, m/z 171, m/z 182 and m/z 239 were consistent with the Ile/Leu immonium ion, the Tyr immonium ion, the Ile/Leu-Gly b_2 ion, the Tyr y_1 ion, and the Gly-Tyr y_2 ion, respectively (Appendix A – Figure A6).

The Ile-variants of each motif were ultimately selected for chemical synthesis and evaluated for their bioactivity in cultured L6 myotubes because of their identification by an established database search methodology. *De novo* sequencing of precursor ions with the identical precursor ion mass further revealed their identical product ion spectra. The presence of Ile in each sequence suggests that Leu-variants of each sequence were also present in most fractions and that the bioactivities of the selected peptides in cultured L6 myotubes may not directly

reflect peptides identified in the SPF or even within the fractions from which each were derived.

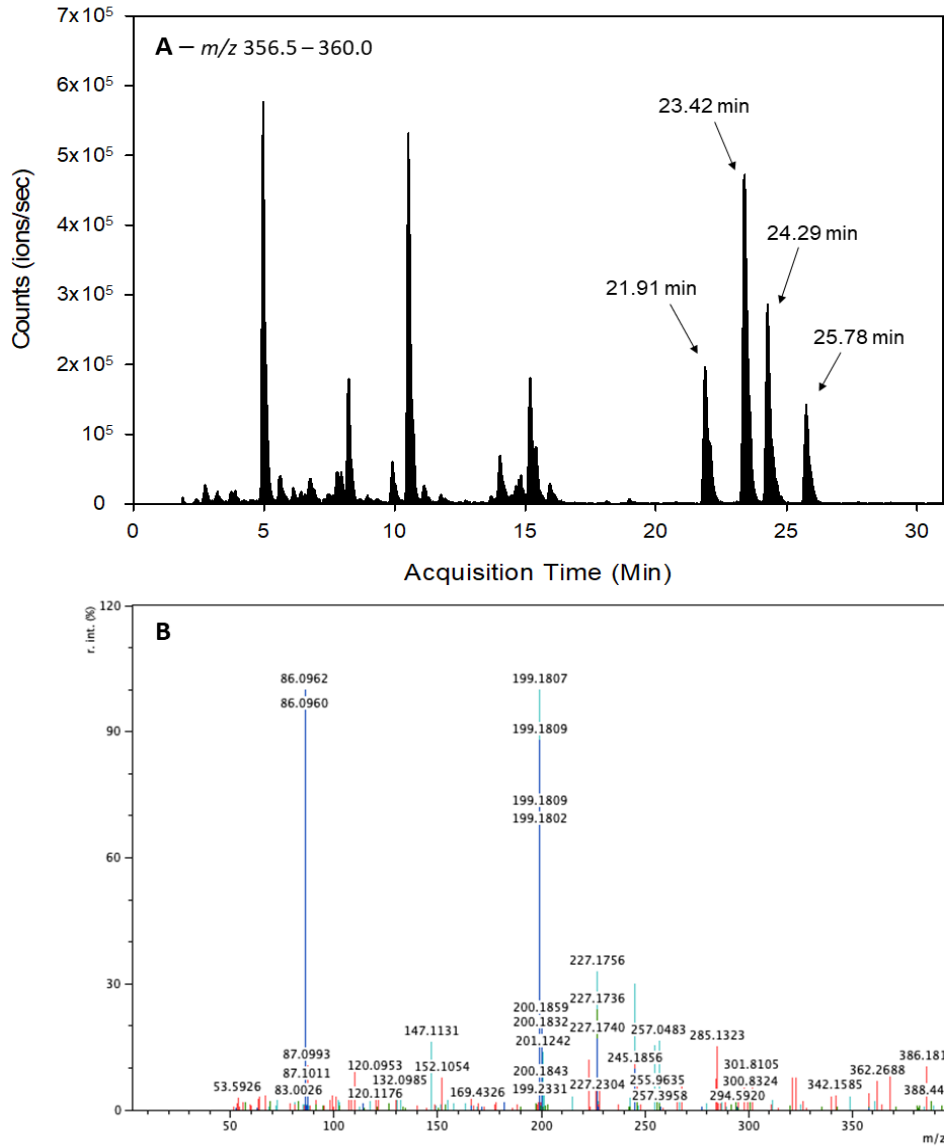


Figure 31. (A) The extracted ion chromatogram at m/z 356.5 – 360.0 from fraction 3 of GF Separation 4 (Figure 16) targeting the Ile-Ile-Ile peptide ion and (B) the superimposed MS/MS spectra of each indicated precursor ion in panel A that resemble this peptide sequence. EIC were generated using Arcadiate software Version 4.5, and MS/MS spectra were generated using mMass software Version 5.5.

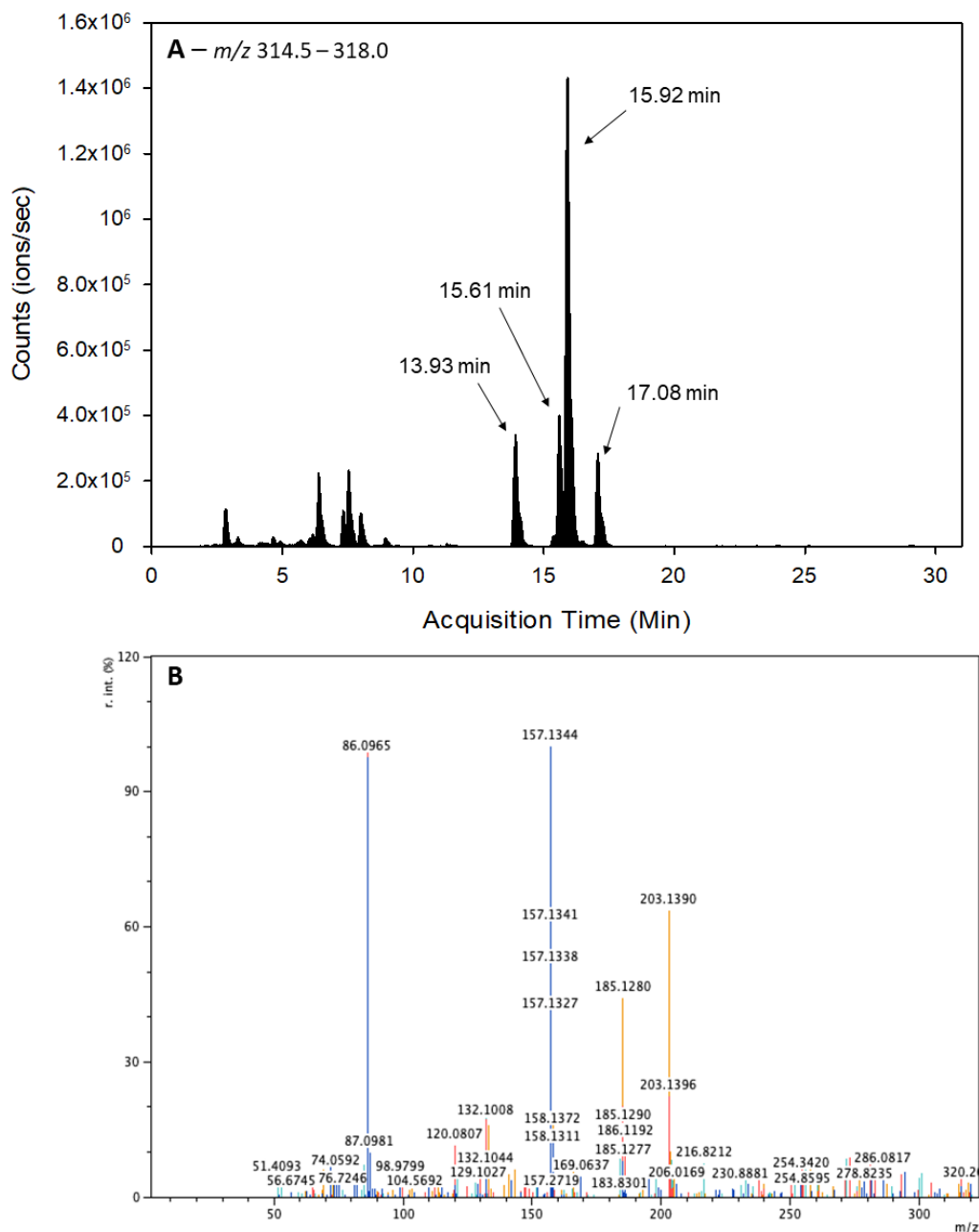


Figure 32. (A) The extracted ion chromatogram at m/z 314.5 – 318.0 from fraction 3 of GF Separation 4 (Figure 16) targeting the Ile-Ala-Ile peptide ion and (B) the superimposed MS/MS spectra of each indicated precursor ion in panel A that resemble this peptide sequence. EIC were generated using Arcadiate software Version 4.5, and MS/MS spectra were generated using mMass software Version 5.5.

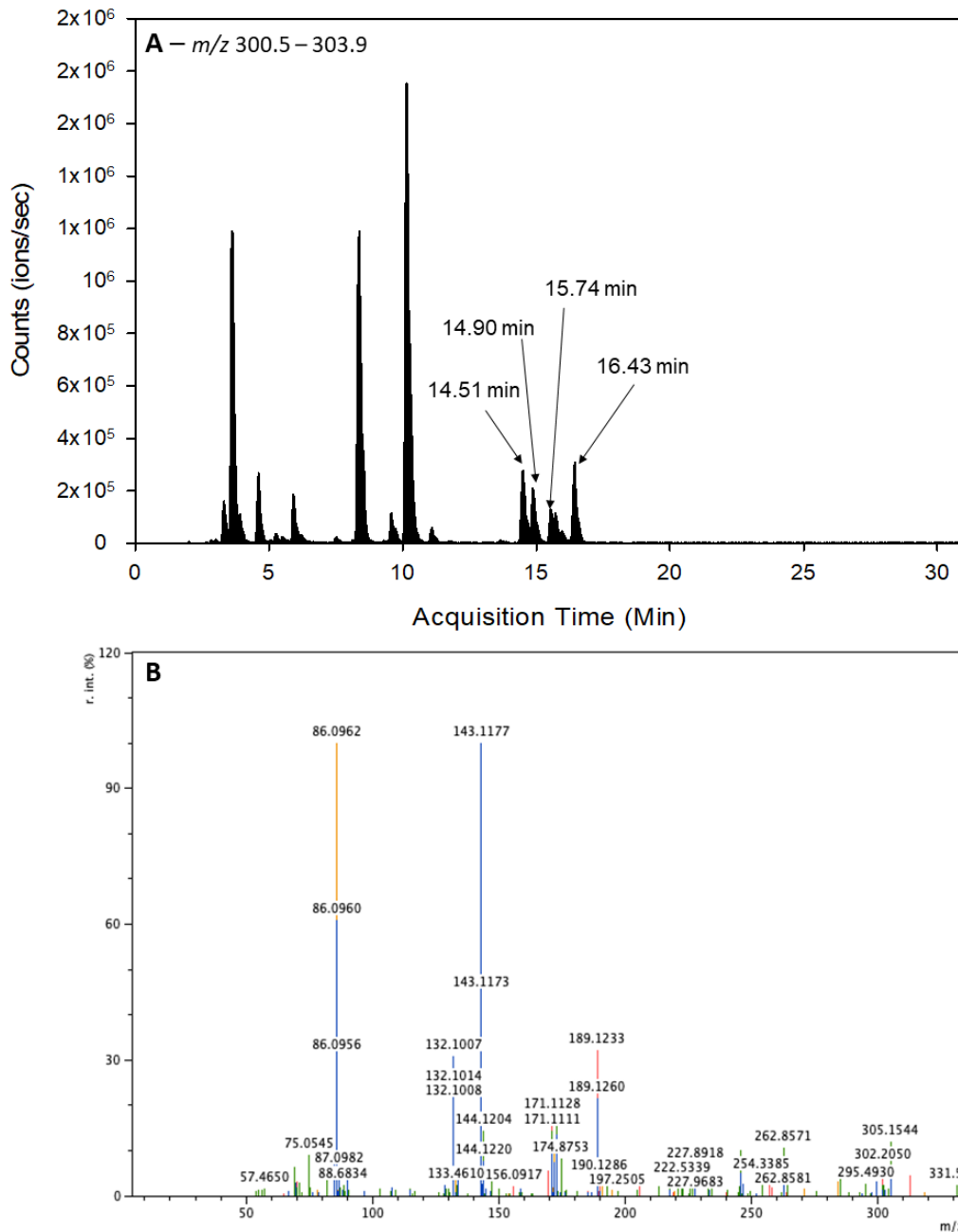


Figure 33. (A) The extracted ion chromatogram at m/z 300.5 – 303.9 from fraction 3 of GF Separation 4 (Figure 16) targeting the Ile-Gly-Ile peptide ion and (B) the superimposed MS/MS spectra of each indicated precursor ion in panel A that resemble this peptide sequence. EIC were generated using Arcadiate software Version 4.5, and MS/MS spectra were generated using mMass software Version 5.5.

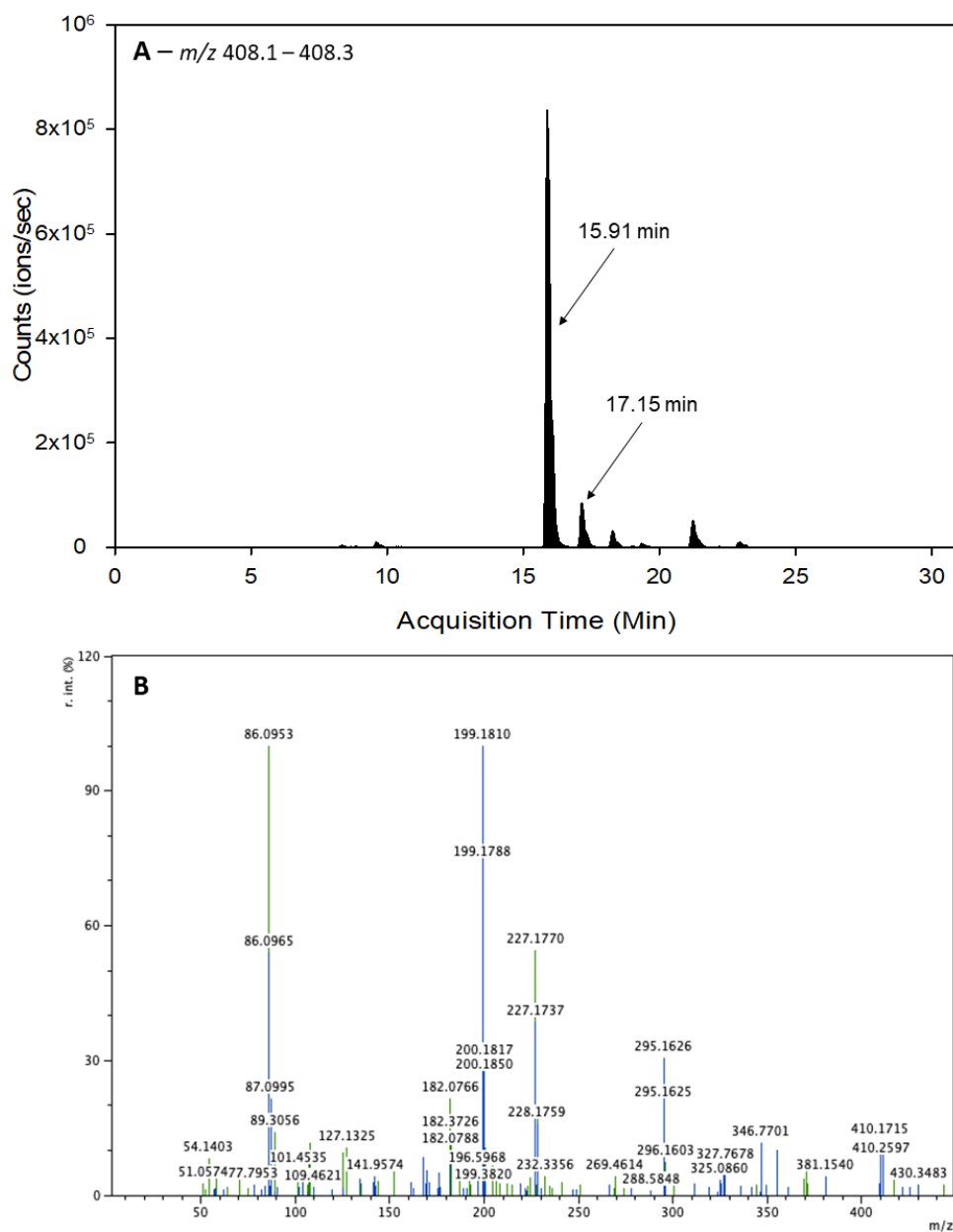


Figure 34. (A) The extracted ion chromatogram at m/z 408.1 – 408.3 from fraction 6 of GF Separation 4 (Figure 16) targeting the Ile-Ile-Tyr peptide ion and (B) the superimposed MS/MS spectra of each indicated precursor ion in panel A that resemble this peptide sequence. EIC were generated using Arcadiate software Version 4.5, and MS/MS spectra were generated using mMass software Version 5.5.

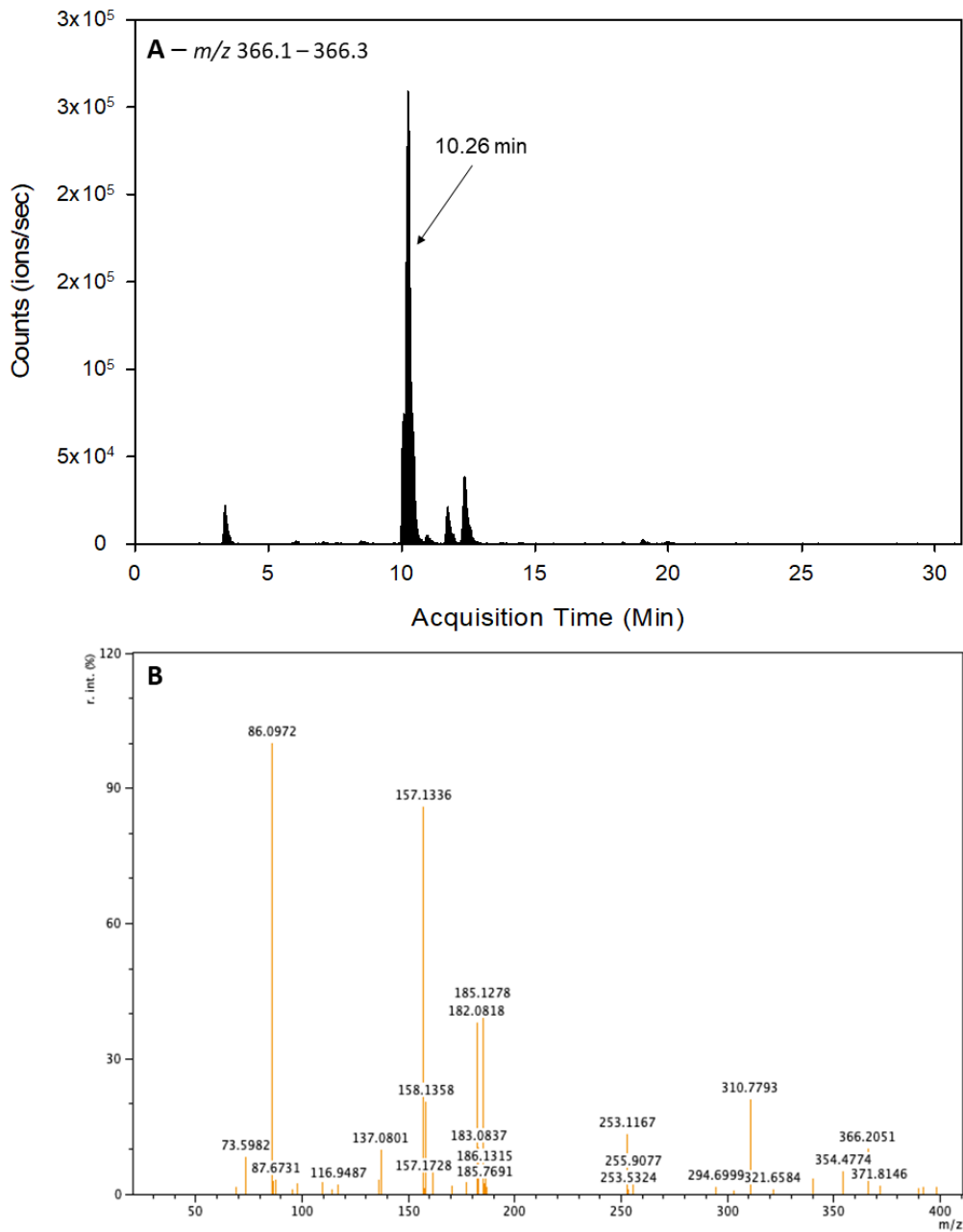


Figure 35. (A) The extracted ion chromatogram at m/z 366.1 – 366.3 from fraction 6 of GF Separation 4 (Figure 16) targeting the Ile-Ala-Tyr peptide ion and (B) the superimposed MS/MS spectra of each indicated precursor ion in panel A that resemble this peptide sequence. EIC were generated using Arcadiate software Version 4.5, and MS/MS spectra were generated using mMass software Version 5.5.

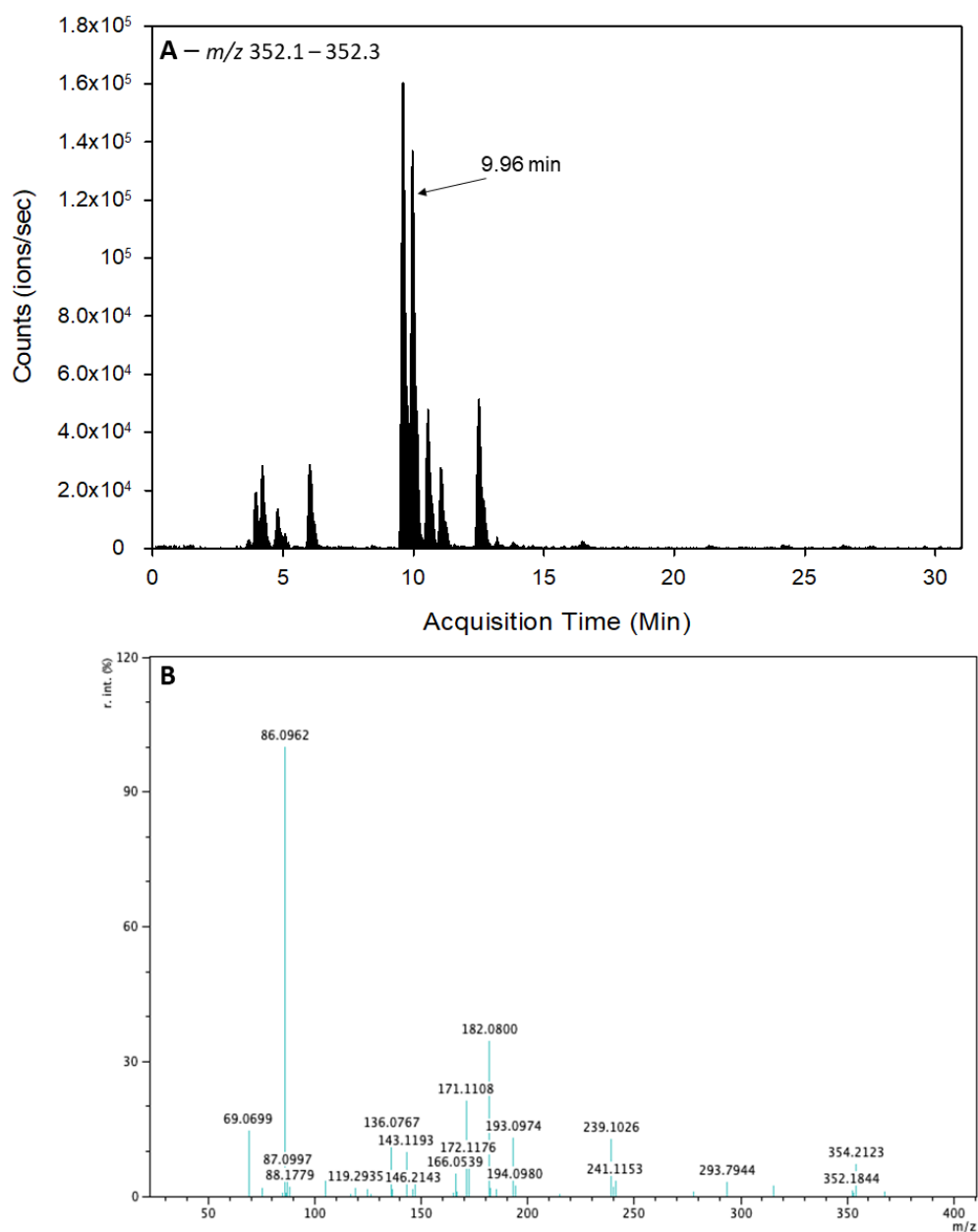


Figure 36. (A) The extracted ion chromatogram at m/z 352.1 – 352.3 from fraction 6 of GF Separation 4 (Figure 16) targeting the Ile-Gly-Tyr peptide ion and (B) the superimposed MS/MS spectra of each indicated precursor ion in panel A that resemble this peptide sequence. EIC were generated using Arcadiate software Version 4.5, and MS/MS spectra were generated using mMass software Version 5.5.

5.5.3 Glucose Uptake Analysis of Synthetic Peptides

The unfractionated SPF from which these peptides were generated exhibited potent glucose uptake stimulating activity at 1 ng/mL (Table 14A) in cultured L6 myotubes. Chemically synthesized peptides Ile-Ala-Ile, Ile-Gly-Ile, and Ile-Ile-Ile based on the Ile-X-Ile motif were evaluated at 1 μ g/mL and/or 1 ng/mL. No significant effect on glucose uptake was observed for any of these synthetic peptides screened individually at 1 ng/mL, but when all peptides were combined at 1 μ g/mL each (3 μ g/mL total), inhibition of glucose uptake in the presence of insulin was observed (p-value = 0.0473).

Chemically synthesized peptides Ile-Ala-Tyr, Ile-Gly-Tyr, and Ile-Ile-Tyr based on the Ile-X-Tyr motif were tested (Table 22) and it was found that Ile-Ala-Tyr peptide stimulated glucose uptake in the absence (p-value = 0.0279) and presence (p-value = 0.0256) of insulin at 1 ng/mL, whereas Ile-Gly-Tyr only stimulated glucose uptake in the presence of insulin at 1 ng/mL (p-values = 0.001).

Ile-Ile-Tyr did not significantly affect (p-value > 0.05) glucose uptake activity in the presence or absence of insulin, when tested at 1 ng/mL. A mixture of the three tripeptides containing 1 ng/mL of each peptide (3 ng/mL total) inhibited glucose uptake (p-value = 0.0170) in the absence of insulin, while testing at 1 μ g/mL each (3 μ g/mL total) had an inhibitory effect in the presence of insulin (p-value = 0.0153).

Table 22. Glucose uptake modulation of chemically synthesized peptides representing the Ile-X-Ile and Ile-X-Tyr motifs, at the indicated peptide concentration, in the (-) absence or (+) presence of insulin following the methodology outlined in Section 3.3.

Concentration	Glucose Uptake (fold activity to control)					
	1 $\mu\text{g/mL}$		1 ng/mL		1 pg/mL	
	Insulin (-)	Insulin (+)	Insulin (-)	Insulin (+)	Insulin (-)	Insulin (+)
Control	1.00 \pm 0.00	1.76 \pm 0.03	1.00 \pm 0.00	1.76 \pm 0.03	1.00 \pm 0.00	1.72 \pm 0.02
Ile-Ile-Ile	-	-	1.00 \pm 0.04	2.02 \pm 0.27	-	-
Ile-Ala-Ile	-	-	1.04 \pm 0.04	1.80 \pm 0.09	-	-
Ile-Gly-Ile	-	-	1.04 \pm 0.04	1.75 \pm 0.17	-	-
Combined	0.91 \pm 0.07	1.56 \pm 0.07*	1.08 \pm 0.05	1.80 \pm 0.15	-	-
Ile-Ile-Tyr	-	-	1.06 \pm 0.03	1.88 \pm 0.11	0.97 \pm 0.04	1.61 \pm 0.07
Ile-Ala-Tyr	-	-	1.13 \pm 0.05*	1.97 \pm 0.08*	0.86 \pm 0.04*	1.55 \pm 0.04*
Ile-Gly-Tyr	-	-	1.05 \pm 0.06	2.02 \pm 0.05*	0.98 \pm 0.06	1.49 \pm 0.10*
Combined	0.98 \pm 0.07	1.48 \pm 0.09	0.86 \pm 0.05*	1.70 \pm 0.13	0.91 \pm 0.03*	1.59 \pm 0.06

Values are means \pm SEM, from n = 6 individual experiments, performed in triplicate. * p-value < 0.05 vs. control.

At 1 ng/mL, the stimulation of glucose uptake was only observed for peptides that contained the C-terminal Tyr residue, therefore the activities of each of the three Ile-X-Tyr peptides were further evaluated at 1 pg/mL. At this diluted concentration, Ile-Ala-Tyr actually inhibited glucose uptake in the absence (p-value = 0.0036) and presence (p-value = 0.0070) of insulin, while Ile-Gly-Tyr also inhibited glucose uptake but only in the presence of insulin (p-value = 0.0435). A mixture of the three peptides containing 1 pg/mL each (3 pg/mL total) also maintained the inhibition of glucose uptake (p-value = 0.0075) in the absence of insulin similarly observed at 3 ng/mL.

The Ile-X-Ile motif was identified from a peptide mixture with glucose uptake stimulating activity, but the activities of Ile-Ala-Ile, Ile-Gly-Ile and Ile-Ile-Ile did not yield this effect.

In contrast, the Ile-X-Tyr motif was identified from a peptide mixture with glucose uptake inhibiting activity, but purified peptides Ile-Ala-Tyr and Ile-Gly-Tyr had a stimulating effect when applied at 1 ng/mL and an inhibiting effect at 1 pg/mL. Importantly, regardless of concentration, Ile-Ala-Tyr and Ile-Gly-Tyr both maintained their activity targeting conditions where insulin is absent and present, respectively.

According to Song et al. (2017), the modulation of glucose uptake by bioactive peptides typically occurs as a result of their effect on AMPK activation in the

absence of insulin or by their effect on Akt signaling pathways in response to insulin stimulation. Also, according to Song et al. (2017), these two effects can occur simultaneously, indicating that the bioactivities of Ile-Ala-Tyr and Ile-Gly-Tyr may occur through their effects on different cellular targets.

Interestingly, the mixtures of three peptides representing each motif exhibit bioactivities that were dissimilar to any of their component peptides. The Ile-X-Tyr peptide mixture at 3 ng/mL (1 ng/mL each) inhibited glucose uptake even though each of Ile-Ala-Tyr and Ile-Gly-Tyr at 1 ng/mL previously exhibited a stimulating effect. According to Bhattacharyya et al. (2006), the combined activation of different cellular targets can activate novel signaling cascades, perhaps reflecting a scenario where the co-activation of Ile-Ala-Tyr and Ile-Gly-Tyr results in novel functions. However, further confirmation is required.

5.5.4 Prediction of Peptide Interactions using Structural Homology

Swiss Target Prediction (SIB, Swiss Institute of Bioinformatics) (Table 23) was used to compare the structural similarities of each chemically synthesized peptide to libraries of bioactive compounds so as to identify potential cellular functions/targets based on a structural homology of each peptide to established agonists and antagonists of cellular functions (Gfeller et al., 2014). All chemically synthesized peptides were equivalently predicted as agonists/antagonists to metalloproteases ACE I and ACE II, as well as E3 ubiquitin-protein ligase XIAP (XIAP/BIRC4) and baculoviral IAP repeat-containing protein 3 (BIRC3). BIRC3

and BIRC4 are genes that encode for proteins that act as inhibitors of apoptosis through their downstream interactions with cysteine-aspartic proteases in NF- κ B and Akt signaling pathways, respectively, but have not been directly linked as targets for modulation of glucose uptake in cultured L6 myotubes.

The ACE-inhibiting activities of all chemically synthesized peptides in the present study were previously reported (Wu et al., 2006), but glucose uptake stimulating effects by ACE-inhibiting peptide Ile-Gln-Trp at 100 μ M in L6 myotubes (Son et al., 2018) was only recently described. The Ile-Gln-Trp peptide was demonstrated to down-regulate the expression of the angiotension type 1 receptor and contribute an antioxidative effect that each promoted increased GLUT4 translocation to the cell membrane.

In the current study, the concentration of synthetic peptides tested at 1 ng/mL in cultured L6 myotubes was equivalent to 2.80 nM (Ile-Ile-Ile), 3.17 nM (Ile-Ala-Ile), 3.32 nM (Ile-Gly-Ile), 2.45 nM (Ile-Ile-Tyr), 2.74 nM (Ile-Ala-Tyr), and 2.85 nM (Ile-Gly-Tyr), or equivalent to 2.45 pM (Ile-Ile-Tyr), 2.74 pM (Ile-Ala-Tyr) and 2.85 pM (Ile-Gly-Tyr) when screened at 1 pg/mL. These lower peptide concentrations indicate that favourable glucose uptake modulation by Ile-Ala-Tyr and Ile-Gly-Tyr are likely to occur independent of their modulating effects on ACE, but further confirmation is required. Alternatively, lab-to-lab variations that occur as a result of unique sensitivities to insulin stimulation from distinct L6 myocyte cell lines could also be an important variable to consider these large differences.

Swiss Target Prediction of peptides representing the Ile-X-Ile motif uniquely resembled inhibitors of serine and cysteine proteases including DPP-IV, fibroblast activation protein and tripeptidyl-peptidase 2, while peptides representing the Ile-X-Tyr motif uniquely resembled ligands of mu- (μ), kappa- (κ) and delta- (δ) type opioid receptors, the nociceptin receptor and neprilysin. The regulation of opioid receptors is a common activity described for bioactive peptides (Liu and Udenigwe, 2019), and importantly whose regulation by peptides has been demonstrated to stimulate glucose uptake at 100 nM in L6 myotubes in the absence of insulin (Kairupan et al., 2018).

Although the glucose uptake modulating activity of chemically synthesized dipeptides in cultured L6 myotubes was not evaluated by the current study, the putatively identified dipeptides in fractions 7 and 9 of GF Separation 4 (Figure 16), along with others composed of one BCAA and aromatic residue, were also evaluated using Swiss Target Prediction (SIB, Swiss Institute of Bioinformatics) (Table 24). Consistent with tripeptides, nearly all dipeptides were also predicted as inhibitors of metallo- (including ACE) and cysteine proteases (calpains). These dipeptides containing C-terminal Tyr residue were also similarly predicted as ligands to the opioid receptor. Ile-Trp, Leu-Trp, Trp-Ile, and Trp-Val were uniquely predicted as ligands to Endothelin receptors A and B, whose activation and downstream signaling has also been demonstrated to influence skeletal muscle glucose uptake (Ottoson-Seeberger et al., 1997).

Table 23. Swiss Target Prediction of synthetic peptides. Peptides are indicated using single letter amino acid codes.

Protein Target	Uniprot ID	Probability					
		IAI	IGI	III	IAY	IGY	IYI
ACE I	P12821	0.90	0.88	0.90	0.91	0.86	0.91
ACE II	Q9BYF1	0.90	0.88	0.90	0.91	0.86	0.91
E3 ubiquitin-protein ligase XIAP	P98170	0.87	0.83	0.91	0.90	0.81	0.93
Baculoviral IAP repeat-containing protein 3	Q13489	0.87	0.83	0.91	0.90	0.81	0.93
DPP-IV	P27487	0.90	0.87	0.89	-	-	-
Fibroblast Activation Protein	Q12884	0.90	0.87	0.89	-	-	-
Tripeptidyl-peptidase 2	P29144	0.89	0.82	0.90	-	-	-
Calpain-1 catalytic subunit	P07384	0.81	0.79	0.85	-	-	-
Mu-type opioid receptor	P35372	-	-	-	0.90	0.81	0.92
Delta-type opioid receptor	P41143	-	-	-	0.90	0.81	0.92
Kappa-type opioid receptor	P41145	-	-	-	0.89	0.79	0.91
Nociceptin Receptor	P41146	-	-	-	0.89	0.79	0.91
Neprilysin	P08473	-	-	-	0.83	0.72	0.87

Table 24. Swiss Target Prediction of dipeptides containing one branched chain amino acid and one aromatic amino acid. Peptides are indicated using single letter amino acid codes.

C-Terminal Aromatic Amino Acid		Peptide (Probability)									
Protein Target	Uniprot ID	IY	LY	VY	IF	LF	VF	IW	LW	VW	
ACE I	P12821	0.94	0.95	0.90	0.89	0.90	0.79	0.92	0.93	0.92	
ACE II	Q9BYF1	0.94	0.95	0.90	0.89	0.90	0.79	0.92	0.93	0.92	
Calpain-1	P07384	0.85	0.85	0.87	0.90	0.90	0.93	0.70	-	0.72	
Calpain-2	P17655	0.85	0.85	0.87	0.90	0.90	0.93	0.70	-	0.72	
Calpain-3	P20807	0.85	0.85	0.87	0.90	0.90	0.93	-	-	0.72	
Delta-type opioid receptor	P35372	0.82	0.83	0.76	-	-	0.71	0.81	0.80	0.75	
Mu-type opioid receptor	P41143	0.82	0.83	0.76	-	-	0.71	0.81	0.80	0.75	
kappa-type opioid receptor	P41145	0.78	0.76	-	-	-	-	0.77	0.75	0.75	
Endothelin A	P25101	-	-	-	-	-	-	0.77	0.80	-	
Endothelin B	P24530	-	-	-	-	-	-	0.77	0.80	-	
Solute carrier family 15 member 2	Q16348	-	-	0.75	-	-	0.76	-	-	0.75	
Peptidyl-prolyl cis-trans isomerase A	P62937	-	-	-	-	-	-	-	0.77	-	
Tyrosine-tRNA ligase	P54577	-	-	0.90	-	-	0.76	-	-	-	

N-Terminal Aromatic Amino Acid		Peptide (Probability)									
Protein Target	Uniprot ID	YI	YL	YV	FI	FL	FV	WI	WL	WV	
ACE I	P12821	0.80	0.83	0.80	0.90	0.89	0.81	0.86	0.85	0.87	

ACE II	Q9BYF1	0.80	0.83	0.80	0.90	0.89	0.81	0.86	0.85	0.87
Calpain-1	P07384	0.71	0.73	0.78	0.82	0.83	0.86	-	-	-
Calpain-2	P17655	0.71	0.73	0.78	0.82	0.83	0.86	-	-	-
Calpain-3	P20807	-	-	-	0.82	0.83	0.86	-	-	-
Delta-type opioid receptor	P35372	0.88	0.87	0.83	0.82	0.82	-	0.77	0.77	0.78
Mu-type opioid receptor	P41143	0.88	0.87	0.83	0.82	0.82	-	0.77	0.77	0.78
kappa-type opioid receptor	P41145	0.83	0.83	-	0.82	0.82	-	0.75	0.75	0.76
Endothelin A	P25101	-	-	-	-	-	-	-	-	-
Endothelin B	P24530	-	-	-	-	-	-	0.74	-	0.70
Solute carrier family 15 member 2	Q16348	0.73	0.76	0.86	-	-	0.87	0.77	-	0.85
Peptidyl-prolyl cis-trans isomerase A	P62937		-	-	-	-	-	0.80	0.84	0.75
Tyrosine-tRNA ligase	P54577	0.87	0.84	0.91	-	-	0.74	-	-	-

The sequences Ile-Ala-Tyr and Ile-Gly-Tyr may represent potentially therapeutic bioactive peptides, but further investigations are necessary to identify the concentration required for optimal activity and a possible mechanism of action. Additionally, the primary representation by dipeptides containing BCAAs and aromatic amino acids in other bioactive SPF fractions (Table 19) suggests these peptides could also be mediators of bioactivity in the fractions from which they are identified. Importantly, the dipeptides and tripeptides containing a C-terminal Tyr screened by Swiss Target Prediction were uniquely identified as ligands to opioid membrane receptors – targets regulated endogenously by peptide hormones to stimulate glucose uptake from the blood into peripheral skeletal muscle (Lord et al., 1977) and recently been proposed as a potentially therapeutic pathway to overcome impairments to insulin receptor signaling (Cheng et al., 2013). The endogenous peptide hormones to opioid receptors, Leu-enkephalin (Tyr-Gly-Gly-Phe-Leu) and Met-enkephalin (Tyr-Gly-Gly-Phe-Met), each contain the conserved Tyr-Gly-Gly-Phe tetrapeptide motif. Other opioid peptide agonists have been identified represented by Tyr-X-Phe, Tyr-X-X-Phe, or Tyr-X-X-X motifs, demonstrating some permissible flexibility of the endogenous motifs at internal and C-terminal positions but maintain the requirement for the N-terminal Tyr. Furthermore, opioid peptides have been identified from within numerous food-derived protein hydrolysates; casein (casomorphins) and soy (soymorphins) are most frequently investigated (Liu and Udenigwe, 2019), but other sources are also identified in the BIOPEP bioactive peptide database.

Due to the flexibility of opioid-binding peptide motifs and their frequent identification within food-derived protein hydrolysates (Dziuba and Darewicz, 2007), putatively identified peptides in bioactive chromatographic fractions of SPF were screened to identify sequences that mimic the endogenous and non-endogenous opioid receptor binding motifs (Table 25). Sequences consistent to either of the endogenous opioid agonists were not identified, but others related to the non-endogenous opioid binding peptide motifs were similar to other food-derived opioid peptides with confirmed *in vitro* activity and described in BIOPEP. Amongst the candidate opioid peptides with potential glucose uptake modulating activities in Table 25, the peptide Tyr-Pro-Ile-Glu putatively identified from fraction 2 from GF Separation 4, as well as in fraction 6 SAX Separation 5, closely resembles the first four residues of Rubiscolin-6 (Tyr-Pro-Leu-Asp-Leu-Phe) (Kairupan et al., 2018), a peptide derived from the spinach protein Rubisco and was demonstrated to stimulate glucose uptake by ~ 1.5-fold the controls in the absence of insulin in both L6 and C₂C₁₂ myotubes at 100 nM, and with targeting to both delta- and mu-type opioid receptors.

Table 25. Peptides identified in bioactive SPF fractions resembling opioid receptor binding motifs.

Fraction	Potential Opioid Peptide Sequences
F2_S4	YPIE
F3_S4	ITDY, GGVPLY
F6_S4	GF, PF, VY, YL, LY, IY, MY, GIW, MIY, IFF, NLW, PAY, PVY, IGY, IAY, ITY, IVY, FGY, ILY
F7_S4	GYSF, PVW, SVW, VTW
F9_S4	TWPW, VW, LW, IW, FY, YF, MW, AW, WS, WVF
F6_S5	YVEG, YIEN, GGVDY, ITDY, YNEI, YPIE, YEF, GDYPLE, YELPDG, YELPDGQ, IYEF, TWPW, DY, MY, DSY, AEW, DIY, DLY, ELY, GDLY, YDNSL
F7_S5	YDDSL, DYY, YDNEFG, IDTEY, YDDSL, YDNEF, YPEM, YGESDL, YEF, YELPDG, YELPDGQ, LEFDY, TWPW, FY, GYSF, TF, IF, LF, YI, LY, IY, VW, IW, LW, DW, WD, EW, DSY, IDF, EIF, LEF, DEF, DIY, DLY, ELY, EEY, YDF, FDY

5.5.5 Conclusion

In conclusion, the present study shows that chemically synthesized peptides Ile-Ala-Tyr and Ile-Gly-Tyr, derived from bioactive fractions of the SPF, were validated as glucose uptake modulating peptides in cultured L6 myotubes. The most interesting finding was that Ile-Ala-Tyr and Ile-Gly-Tyr exhibited stimulation of glucose uptake in L6 myotubes at 2.7 and 2.8 nM, respectively, but inhibition at 2.7 and 2.8 pM. Novel mechanisms of glucose uptake modulation were suspected by each peptide due to their specificity to conditions without and with insulin-stimulation, and the modulating effects of peptide mixtures containing these peptides could not be predicted by their activities when evaluated independently.

The importance of Tyr residues within functional peptide sequences to exhibit bioactivity was further supported by predictions that the chemically synthesized peptide sequences resembled opioid receptor-binding peptides and represents a potential mechanism of the regulation of glucose metabolism by SPF. The similarity of other identified sequences within the SPF to these motifs reveals that many peptides are likely to contribute both stimulating or inhibiting activities that cumulatively influence glucose uptake modulation by the SPF.

5.5.6 Future Work

The above findings describe the challenges associated with the selection of peptides from within functional peptide mixtures. Many peptide sequences consistent with both Ile-X-Ile and Ile-X-Tyr motifs were identified in bioactive

fractions but not tested for bioactivity and could perhaps be more or less bioactive than the evaluated sequences.

The screening of individual sequences identified within protein hydrolysates should be evaluated over a range of peptide (1 pM – 1 nM) and insulin (10^{-6} M to 10^{-12} M) concentrations to comprehensively evaluate the influence of bioactive peptide sequences due to the unique activity of purified peptides between these concentrations. Both functional peptides could be further evaluated to reveal the mechanisms of their regulation, using inhibitors for the receptors that peptides are predicted to interact with, or for other possible targets of interest. All other peptides identified as significant components of functional fractions should also be screened *in vitro*. With a further understanding of the modulating effects of peptides, the methodology used to develop the SPF could be enhanced to further concentrate the glucose uptake stimulation.

REFERENCES

- Agilent. (2018). *Agilent G2721AA Spectrum Mill MS Proteomics Workbench Site Preparation Guide*. Retrieved from <https://www.agilent.com/cs/library/usermanuals/public/G2721-90049ApplicationGuide.pdf>
- Anderson, R. A., Broadhurst, C. L., Polansky, M. M., Schmidt, W. F., Khan, A., Flanagan, V. P., Schoene, N. W., and Graves, D. J. (2004). Isolation and characterization of polyphenol type-A polymers from cinnamon with insulin-like biological activity. *Journal of Agricultural and Food Chemistry*, 52(1), 65–70. <https://doi.org/10.1021/jf034916b>
- Arroume, N., Froidevaux, R., Kapel, R., Cudennec, B., Ravallec, R., Flahaut, C., Bazinet, L., Jacques, P., and Dhulster, P. (2016). Food peptides: purification, identification and role in the metabolism. *Current Opinion in Food Science*, 7, 101–107. <https://doi.org/10.1016/j.cofs.2016.02.005>
- Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., De Castro, E., Duvaud, S., Flegel, V., Fortier, A., Gasteiger, E., Grosdidier, A., Hernandez, C., Ioannidis, V., Kuznetsov, D., Liechti, R., Moretti, S., Mostaguir, K., Redaschi, N., Rossier, G., Xenarios, I., and Stockinger, H. (2012). ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Research*, 40(W1), W597–W603. <https://doi.org/10.1093/nar/gks400>
- Atef, M., and Mahdi Ojagh, S. (2017). Health benefits and food applications of bioactive compounds from fish byproducts: a review. *Journal of Functional Foods*, 35, 673–681. <https://doi.org/10.1016/j.jff.2017.06.034>
- Bahar, A., and Ren, D. (2013). Antimicrobial peptides. *Pharmaceuticals*, 6(12), 1543–1575. <https://doi.org/10.3390/ph6121543>
- Bailey, C., and Day, C. (2004). Metformin: its botanical background. *Practical Diabetes International*, 21(3), 115–117. <https://doi.org/10.1002/pdi.606>
- Beermann, C., Euler, M., Herzberg, J., and Stahl, B. (2009). Anti-oxidative capacity of enzymatically released peptides from soybean protein isolate. *European Food Research and Technology*, 229(4), 637–644. <https://doi.org/10.1007/s00217-009-1093-1>
- Capriotti, A. L., Cavaliere, C., Foglia, P., Piovesana, S., Samperi, R., Zenezini Chiozzi, R., and Laganà, A. (2015). Development of an analytical strategy for the identification of potential bioactive peptides generated by *in vitro* tryptic digestion of fish muscle proteins. *Analytical and Bioanalytical Chemistry*, 407(3), 845–854. <https://doi.org/10.1007/s00216-014-8094-z>

- Chaijan, M., Benjakul, S., Visessanguan, W., Lee, S., and Faustman, C. (2007). Effect of ionic strength and temperature on interaction between fish myoglobin and myofibrillar proteins. *Journal of Food Science*, 72(2), C89-95. <https://doi.org/10.1111/j.1750-3841.2006.00236.x>
- Chalamaiah, M., Jyothirmayi, T., Bhaskarachary, K., Vajreswari, A., Hemalatha, R., and Dinesh Kumar, B. (2013). Chemical composition, molecular mass distribution and antioxidant capacity of rohu (*Labeo rohita*) roe (egg) protein hydrolysates prepared by gastrointestinal proteases. *Food Research International*, 52(1), 221–229. <https://doi.org/10.1016/j.foodres.2013.03.020>
- Chen, Y. C., and Jaczynski, J. (2007). Protein recovery from rainbow trout (*Oncorhynchus mykiss*) processing byproducts via isoelectric solubilization/precipitation and its gelation properties as affected by functional additives. *Journal of Agricultural and Food Chemistry*, 55(22), 9079–9088. <https://doi.org/10.1021/jf071992w>
- Chen, Y., Zhang, J., Xing, G., and Zhao, Y. (2009). Mascot-derived false positive peptide identifications revealed by manual analysis of tandem mass spectra. *Journal of Proteome Research*, 8(6), 3141–3147. <https://doi.org/10.1021/pr900172v>
- Chenau, J., Michelland, S., Sidibe, J., and Seve, M. (2008). Peptides OFFGEL electrophoresis: a suitable pre-analytical step for complex eukaryotic samples fractionation compatible with quantitative iTRAQ labeling. *Proteome Science*, 6(1), 9. <https://doi.org/10.1186/1477-5956-6-9>
- Cheng, K.-C., Asakawa, A., Li, Y.-X., Liu, I.-M., Amitani, H., Cheng, J.-T., and Inui, A. (2013). Opioid μ -receptors as new target for insulin resistance. *Pharmacology & Therapeutics*, 139(3), 334–340. <https://doi.org/10.1016/j.pharmthera.2013.05.002>
- Cheng, Y., Chen, J., and Xiong, Y. L. (2010). Chromatographic separation and tandem MS identification of active peptides in potato protein hydrolysate that inhibit autoxidation of soybean oil-in-water emulsions. *Journal of Agricultural and Food Chemistry*, 58(15), 8825–8832. <https://doi.org/10.1021/jf101556n>
- Cheung, H. S., Wang, F. L., Ondetti, M. A., Sabo, E. F., and Cushman, D. W. (1980). Binding of peptide substrates and inhibitors of angiotensin-converting enzyme. Importance of the COOH-terminal dipeptide sequence. *Journal of Biological Chemistry*, 255(2), 401–407. Retrieved from <http://www.jbc.org/cgi/content/short/255/2/401>

- Chevrier, G., Mitchell, P. L., Rioux, L.-E., Hasan, F., Jin, T., Roblet, C. R., Doyen, A., Pilon, G., St-Pierre, P., Lavigne, C., Bazinet, L., Jacques, H., Gill, T., McLeod, R. S., and Marette, A. (2015). Low-molecular-weight peptides from salmon protein prevent obesity-linked glucose intolerance, inflammation, and dyslipidemia in LDLR^{-/-}/ApoB^{100/100} mice. *The Journal of Nutrition*, 145(7), 1415–1422. <https://doi.org/10.3945/jn.114.208215>
- Comerford, K. B., and Pasin, G. (2016). Emerging evidence for the importance of dietary protein source on glucoregulatory markers and type 2 diabetes: different effects of dairy, meat, fish, egg, and plant protein foods. *Nutrients*, 8(8). <https://doi.org/10.3390/nu8080446>
- Conlon, J. M. (2007). Purification of naturally occurring peptides by reversed-phase HPLC. *Nature Protocols*, 2(1), 191–197. <https://doi.org/10.1038/nprot.2006.437>
- Crooks, G. E., Hon, G., Chandonia, J.-M., and Brenner, S. E. (2004). WebLogo: a sequence logo generator. *Genome Research*, 14(6), 1188–1190. <https://doi.org/10.1101/gr.849004>
- Daliri, E. B.-M., Lee, B. H., and Oh, D. H. (2018). Current trends and perspectives of bioactive peptides. *Critical Reviews in Food Science and Nutrition*, 58(13), 2273–2284. <https://doi.org/10.1080/10408398.2017.1319795>
- Darewicz, M., Borawska, J., and Pliszka, M. (2016). Carp proteins as a source of bioactive peptides - an *in silico* approach. *Czech Journal of Food Sciences*, 34(2), 111–117. <https://doi.org/10.17221/300/2015-CJFS>
- de Castro, R. J. S., and Sato, H. H. (2015). Biologically active peptides: processes for their generation, purification and identification and applications as natural additives in the food and pharmaceutical industries. *Food Research International*, 74, 185–198. <https://doi.org/10.1016/j.foodres.2015.05.013>
- Deng, R. (2012). A review of the hypoglycemic effects of five commonly used herbal food supplements. *Recent Patents on Food, Nutrition & Agriculture*, 4(1), 50–60. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/22329631>
- Dongré, A. R., Jones, J. L., Somogyi, Á., and Wysocki, V. H. (1996). Influence of peptide composition, gas-phase basicity, and chemical modification on fragmentation efficiency: evidence for the mobile proton model. *Journal of the American Chemical Society*, 118(35), 8365–8374. <https://doi.org/10.1021/ja9542193>

- Doyen, A., Udenigwe, C. C., Mitchell, P. L., Marette, A., Aluko, R. E., and Bazinet, L. (2014). Anti-diabetic and antihypertensive activities of two flaxseed protein hydrolysate fractions revealed following their simultaneous separation by electro dialysis with ultrafiltration membranes. *Food Chemistry*, 145, 66–76. <https://doi.org/10.1016/j.foodchem.2013.07.108>
- Dziuba, M., and Darewicz, M. (2007). Food proteins as precursors of bioactive peptides - classification into families. *Food Science and Technology International*, 13(6), 393–404. <https://doi.org/10.1177/1082013208085933>
- Erdmann, K., Cheung, B. W. Y., and Schröder, H. (2008). The possible roles of food-derived bioactive peptides in reducing the risk of cardiovascular disease. *Journal of Nutritional Biochemistry*, 19(10), 643–654. <https://doi.org/10.1016/j.jnutbio.2007.11.010>
- Escudero, E., Sentandreu, M. A., Arihara, K., and Toldrá, F. (2010). Angiotensin I-converting enzyme inhibitory peptides generated from *in vitro* gastrointestinal digestion of pork meat. *Journal of Agricultural and Food Chemistry*, 58(5), 2895–2901. <https://doi.org/10.1021/jf904204n>
- FAO. (2018). B-23 Salmons, trouts, smelts. Fishery and Aquaculture Statistics 2016. *FAO Yearbook*, 81–84.
- Fekete, S., Veuthey, J. L., and Guillarme, D. (2012). New trends in reversed-phase liquid chromatographic separations of therapeutic peptides and proteins: theory and applications. *Journal of Pharmaceutical and Biomedical Analysis*, 69, 9–27. <https://doi.org/10.1016/j.jpba.2012.03.024>
- Ferro, E. S., Hyslop, S., and Camargo, A. C. M. (2004). Intracellular peptides as putative natural regulators of protein interactions. *Journal of Neurochemistry*, 91(4), 769–777. <https://doi.org/10.1111/j.1471-4159.2004.02757.x>
- Fisher Scientific. (2017). Common background contamination ions in mass spectroscopy. Retrieved from https://beta-static.fishersci.ca/content/dam/fishersci/en_US/documents/programs/scientific/brochures-and-catalogs/posters/fisher-chemical-poster.pdf
- GE Healthcare. (2007). *Protein purification - handbook* (18th-1132nd–29th ed.).
- Gehrig, P. M., Hunziker, P. E., Zahariev, S., and Pongor, S. (2004). Fragmentation pathways of NG-methylated and unmodified arginine residues in peptides studied by ESI-MS/MS and MALDI-MS. *Journal of the American Society for Mass Spectrometry*, 15(2), 142–149. <https://doi.org/10.1016/j.jasms.2003.10.002>

- Gfeller, D., Grosdidier, A., Wirth, M., Daina, A., Michielin, O., and Zoete, V. (2014). SwissTargetPrediction: a web server for target prediction of bioactive small molecules. *Nucleic Acids Research*, *42*(W1), W32–W38. <https://doi.org/10.1093/nar/gku293>
- Ghassem, M., Arihara, K., Babji, A. S., Said, M., and Ibrahim, S. (2011). Purification and identification of ACE inhibitory peptides from Haruan (*Channa striatus*) myofibrillar protein hydrolysate using HPLC-ESI-TOF MS/MS. *Food Chemistry*, *129*(4), 1770–1777. <https://doi.org/10.1016/j.foodchem.2011.06.051>
- Ghorbani, Z., Hekmatdoost, A., and Mirmiran, P. (2014). Anti-hyperglycemic and insulin sensitizer effects of turmeric and its principle constituent curcumin. *International Journal of Endocrinology and Metabolism*, *12*(4), 18081. <https://doi.org/10.5812/ijem.18081>
- Girgih, A. T., Nwachukwu, I. D., Hasan, F. M., Fagbemi, T. N., Malomo, S. A., Gill, T. A., and Aluko, R. E. (2016). Kinetics of *in vitro* enzyme inhibition and blood pressure-lowering effects of salmon (*Salmo salar*) protein hydrolysates in spontaneously hypertensive rats. *Journal of Functional Foods*, *20*, 43–53. <https://doi.org/10.1016/j.jff.2015.10.018>
- Girgih, A. T., Udenigwe, C. C., Hasan, F. M., Gill, T. A., and Aluko, R. E. (2013). Antioxidant properties of salmon (*Salmo salar*) protein hydrolysate and peptide fractions isolated by reverse-phase HPLC. *Food Research International*, *52*(1), 315–322. <https://doi.org/10.1016/j.foodres.2013.03.034>
- Giri, A., and Ohshima, T. (2012). Bioactive marine peptides: nutraceutical value and novel approaches. *Advances in Food and Nutrition Research*, *65*, 73–105. <https://doi.org/10.1016/B978-0-12-416003-3.00005-6>
- Goodman, M., Zapf, C., and Rew, Y. (2001). New reagents, reactions, and peptidomimetics for drug design. *Biopolymers*, *60*(3), 229–245. [https://doi.org/10.1002/1097-0282\(2001\)60:3<229::AID-BIP10034>3.0.CO;2-P](https://doi.org/10.1002/1097-0282(2001)60:3<229::AID-BIP10034>3.0.CO;2-P)
- Gu, R.-Z., Li, C.-Y., Liu, W.-Y., Yi, W.-X., and Cai, M.-Y. (2011). Angiotensin I-converting enzyme inhibitory activity of low-molecular-weight peptides from Atlantic salmon (*Salmo salar* L.) skin. *Food Research International*, *44*(5), 1536–1540. <https://doi.org/10.1016/j.foodres.2011.04.006>
- Guijas, C., Montenegro-Burke, J. R., Domingo-Almenara, X., Palermo, A., Warth, B., Hermann, G., Koellensperger, G., Huan, T., Uritboonthai, W., Aisporna, A. E., Wolan, D. W., Spilker, M. E., Benton, H. P., and Siuzdak, G. (2018). METLIN: a technology platform for identifying knowns and unknowns. *Analytical Chemistry*, *90*(5), 3156–3164. <https://doi.org/10.1021/acs.analchem.7b04424>

- Halim, N. R. A., Yusof, H. M., and Sarbon, N. M. (2016). Functional and bioactive properties of fish protein hydrolysates and peptides: a comprehensive review. *Trends in Food Science and Technology*, *51*, 24–33. <https://doi.org/10.1016/j.tifs.2016.02.007>
- Han, B. K., Lee, H. J., Lee, H.-S., Suh, H. J., and Park, Y. (2016). Hypoglycaemic effects of functional tri-peptides from silk in differentiated adipocytes and streptozotocin-induced diabetic mice. *Journal of the Science of Food and Agriculture*, *96*(1), 116–121. <https://doi.org/10.1002/jsfa.7067>
- Harrison, A. G. (2001). Sequence-specific fragmentation of deprotonated peptides containing H or alkyl side chains. *Journal of the American Society for Mass Spectrometry*, *12*(1), 1–13. [https://doi.org/10.1016/S1044-0305\(00\)00199-9](https://doi.org/10.1016/S1044-0305(00)00199-9)
- Hazato, T., and Kase, R. (1986). Isolation of angiotensin-converting enzyme inhibitor from porcine plasma. *Biochemical and Biophysical Research Communications*, *139*(1), 52–55. [https://doi.org/10.1016/S0006-291X\(86\)80078-X](https://doi.org/10.1016/S0006-291X(86)80078-X)
- Henaus, L., Thibodeau, J., Pilon, G., Gill, T., Marette, A., and Bazinet, L. (2019). How charge and triple size-selective membrane separation of peptides from salmon protein hydrolysate orientate their biological response on glucose uptake. *International Journal of Molecular Sciences*, *20*(8), 1939. <https://doi.org/10.3390/ijms20081939>
- Howard, P. L., Chia, M. C., Del Rizzo, S., Liu, F.-F., and Pawson, T. (2003). Redirecting tyrosine kinase signaling to an apoptotic caspase pathway through chimeric adaptor proteins. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(20), 11267–11272. <https://doi.org/10.1073/pnas.1934711100>
- Hu, X., Tao, N., Wang, X., Xiao, J., and Wang, M. (2016). Marine-derived bioactive compounds with anti-obesity effect: a review. *Journal of Functional Foods*, *21*, 372–387. <https://doi.org/10.1016/j.jff.2015.12.006>
- Huang, S., and Czech, M. P. (2007). The GLUT4 glucose transporter. *Cell Metabolism*, *5*(4), 237–252. <https://doi.org/10.1016/j.cmet.2007.03.006>
- Hultin, H. O. (1984). Postmortem biochemistry of meat and fish. *Journal of Chemical Education*, *61*(4), 289. <https://doi.org/10.1021/ed061p289>
- Ibáñez, C., Simó, C., García-Cañas, V., Cifuentes, A., and Castro-Puyana, M. (2013). Metabolomics, peptidomics and proteomics applications of capillary electrophoresis-mass spectrometry in foodomics: a review. *Analytica Chimica Acta*. <https://doi.org/10.1016/j.aca.2013.07.042>

- Jesus Perez de Vega, M., Martin-Martinez, M., and Gonzalez-Muniz, R. (2007). Modulation of protein-protein interactions by stabilizing/mimicking protein secondary structure elements. *Current Topics in Medicinal Chemistry*, 7(1), 33–62. <https://doi.org/10.2174/156802607779318325>
- Jimsheena, V. K., and Gowda, L. R. (2011). Angiotensin I-converting enzyme (ACE) inhibitory peptides derived from arachin by simulated gastric digestion. *Food Chemistry*, 125(2), 561–569. <https://doi.org/10.1016/j.foodchem.2010.09.048>
- Jin, T. (2012). *Separation and purification of antidiabetic bioactive peptide from salmon and cod waste*. Dalhousie University. Retrieved from <https://dalspace.library.dal.ca/handle/10222/37804>
- Johnson, R. S., Martin, S. A., Biemann, K., Stults, J. T., and Watson, J. T. (1987). Novel fragmentation process of peptides by collision-induced decomposition in a tandem mass spectrometer: differentiation of leucine and isoleucine. *Analytical Chemistry*, 59(21), 2621–2625. <https://doi.org/10.1021/ac00148a019>
- Kairupan, T. S., Cheng, K.-C., Asakawa, A., Amitani, H., Yagi, T., Ataka, K., Rokot, N. T., Kapantow, N. H., Kato, I., and Inui, A. (2018). Rubiscolin-6 activates opioid receptors to enhance glucose uptake in skeletal muscle. *Journal of Food and Drug Analysis*. <https://doi.org/10.1016/j.jfda.2018.06.012>
- Kanegawa, N., Suzuki, C., and Ohinata, K. (2010). Dipeptide Tyr-Leu (YL) exhibits anxiolytic-like activity after oral administration via activating serotonin 5-HT_{1A}, dopamine D₁ and GABA_A receptors in mice. *FEBS Letters*, 584(3), 599–604. <https://doi.org/10.1016/j.febslet.2009.12.008>
- Keil, B. (1992). *Specificity of Proteolysis*. Berlin, Heidelberg, New York: Springer-Verlag. <https://doi.org/10.1007/978-3-642-48380-6>
- Khedr, S., Martin, M., and Deussen, A. (2015). Inhibitory efficacy and biological variability of tryptophan containing dipeptides on human plasma angiotensin converting enzyme activity. *Journal of Hypertension: Open Access*, 04(02). <https://doi.org/10.4172/2167-1095.1000198>
- Kobayashi, Y., Suzuki, M., Satsu, H., Arai, S., Hara, Y., Suzuki, K., Miyamoto, Y., and Shimizu, M. (2000). Green tea polyphenols inhibit the sodium-dependent glucose transporter of intestinal epithelial cells by a competitive mechanism. *Journal of Agricultural and Food Chemistry*, 48(11), 5618–5623. <https://doi.org/10.1021/jf0006832>
- Korhonen, H., and Pihlanto, A. (2006). Bioactive peptides: production and functionality. *International Dairy Journal*, 16(9), 945–960. <https://doi.org/10.1016/j.idairyj.2005.10.012>

- Kristinsson, H. G., and Rasco, B. A. (2000). Biochemical and functional properties of Atlantic salmon (*Salmo salar*) muscle proteins hydrolyzed with various alkaline proteases. *Journal of Agricultural and Food Chemistry*, *48*, 657–666. <https://doi.org/10.1021/jf990447v>
- Kurian, L. A., Silva, T. A., and Sabatino, D. (2014). Submonomer synthesis of azapeptide ligands of the insulin receptor tyrosine kinase domain. *Bioorganic and Medicinal Chemistry Letters*, *24*(17), 4176–4180. <https://doi.org/10.1016/j.bmcl.2014.07.046>
- Lacroix, I. M. E., and Li-Chan, E. C. Y. (2014). Overview of food products and dietary constituents with antidiabetic properties and their putative mechanisms of action: a natural approach to complement pharmacotherapy in the management of diabetes. *Molecular Nutrition and Food Research*, *58*(1), 61–78. <https://doi.org/10.1002/mnfr.201300223>
- Lahrichi, S. L., Affolter, M., Zolezzi, I. S., and Panchaud, A. (2013). Food peptidomics: large scale analysis of small bioactive peptides - a pilot study. *Journal of Proteomics*, *88*, 83–91. <https://doi.org/10.1016/j.jprot.2013.02.018>
- Lammi, C., Zanoni, C., Arnoldi, A., Paliyath, G., and Segura-Carretero, A. (2015). Three peptides from soy glycinin modulate glucose metabolism in human hepatic HepG2 cells. *International Journal of Molecular Sciences*, *16*(11), 27362–27370. <https://doi.org/10.3390/ijms161126029>
- Lan, V. T. T., Ito, K., Ohno, M., Motoyama, T., Ito, S., and Kawarasaki, Y. (2015). Analyzing a dipeptide library to identify human dipeptidyl peptidase IV inhibitor. *Food Chemistry*, *175*, 66–73. <https://doi.org/10.1016/j.foodchem.2014.11.131>
- Le Maux, S., Nongonierma, A. B., and FitzGerald, R. J. (2015). Improved short peptide identification using HILIC–MS/MS: retention time prediction model based on the impact of amino acid position in the peptide sequence. *Food Chemistry*, *173*, 847–854. <https://doi.org/10.1016/J.FOODCHEM.2014.10.104>
- Lee, H.-J., Wang, N. X., Shao, Y., and Zheng, J. J. (2009). Identification of tripeptides recognized by the PDZ domain of Dishevelled. *Bioorganic & Medicinal Chemistry*, *17*(4), 1701–1708. <https://doi.org/10.1016/j.bmc.2008.12.060>
- Liu, H., Lin, D., and Yates, J. R. (2002). Multidimensional separations for protein/peptide analysis in the post-genomic era. *BioTechniques*, *32*(4), 898–911. <https://doi.org/10.2144/02324pt01>

- Liu, Q., Chen, L., Hu, L., Guo, Y., and Shen, X. (2010). Small molecules from natural sources, targeting signaling pathways in diabetes. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, 1799(10–12), 854–865. <https://doi.org/10.1016/j.bbagr.2010.06.004>
- Liu, R., Zheng, W., Li, J., Wang, L., Wu, H., Wang, X., and Shi, L. (2015). Rapid identification of bioactive peptides with antioxidant activity from the enzymatic hydrolysate of *Mactra veneriformis* by UHPLC–Q-TOF mass spectrometry. *Food Chemistry*, 167, 484–489. <https://doi.org/10.1016/j.foodchem.2014.06.113>
- Liu, Z., and Udenigwe, C. C. (2019). Role of food-derived opioid peptides in the central nervous and gastrointestinal systems. *Journal of Food Biochemistry*, 43(1), e12629. <https://doi.org/10.1111/jfbc.12629>
- Lord, J. A. H., Waterfield, A. A., Hughes, J., and Kosterlitz, H. W. (1977). Endogenous opioid peptides: multiple agonists and receptors. *Nature*, 267(5611), 495–499. <https://doi.org/10.1038/267495a0>
- Ma, M. S., In, Y. B., Hyeon, G. L., and Yang, C. B. (2006). Purification and identification of angiotensin I-converting enzyme inhibitory peptide from buckwheat (*Fagopyrum esculentum Moench*). *Food Chemistry*, 96(1), 36–42. <https://doi.org/10.1016/j.foodchem.2005.01.052>
- Marles, R. J., and Farnsworth, N. R. (1995). Antidiabetic plants and their active constituents. *Phytomedicine*, 2(2), 137–189. [https://doi.org/10.1016/S0944-7113\(11\)80059-0](https://doi.org/10.1016/S0944-7113(11)80059-0)
- Maruyama, S., Mitachi, H., Tanaka, H., Tomizuka, N., and Suzuki, H. (1987). Studies on the active site and antihypertensive activity of angiotensin I-converting enzyme inhibitors derived from casein. *Agricultural and Biological Chemistry*, 51(6), 1581–1586. <https://doi.org/10.1080/00021369.1987.10868244>
- Matak, K. E., Tahergorabi, R., and Jaczynski, J. (2015). A review: protein isolates recovered by isoelectric solubilization/precipitation processing from muscle food by-products as a component of nutraceutical foods. *Food Research International*, 77, 697–703. <https://doi.org/10.1016/j.foodres.2015.05.048>
- Mathers, C. D., and Loncar, D. (2006). Projections of global mortality and burden of disease from 2002 to 2030. *PLoS Medicine*, 3(11), 2011–2030. <https://doi.org/10.1371/journal.pmed.0030442>
- Matsufuji, H., Matsui, T., Seki, E., Osajima, K., Nakashima, M., and Osajima, Y. (1994). Angiotensin I-converting enzyme inhibitory peptides in an alkaline protease hydrolyzate derived from sardine muscle. *Bioscience, Biotechnology, and Biochemistry*, 58(12), 2244–2245. <https://doi.org/10.1271/bbb.58.2244>

- Matsui, T., Li, C.-H., Tanaka, T., Maki, T., Osajima, Y., and Matsumoto, K. (2000). Depressor effect of wheat germ hydrolysate and its novel angiotensin I-converting enzyme inhibitory peptide, Ile-Val-Tyr, and the metabolism in rat and human plasma. *Biological & Pharmaceutical Bulletin*, 23(4), 427–431. <https://doi.org/10.1248/bpb.23.427>
- Matsui, T., Oki, T., and Osajima, Y. (1999). Isolation and identification of peptidic α -glucosidase inhibitors derived from sardine muscle hydrolyzate. *Zeitschrift Fur Naturforschung - Section C Journal of Biosciences*, 54(3–4), 259–263. <https://doi.org/10.1515/znc-1999-3-417>
- Mekata, T., Kono, T., Satoh, J., Yoshida, M., Mori, K., Sato, T., Miyazato, M., and Ida, T. (2017). Purification and characterization of bioactive peptides RYamide and CCHamide in the kuruma shrimp *Marsupenaeus japonicus*. *General and Comparative Endocrinology*, 246, 321–330. <https://doi.org/10.1016/j.ygcen.2017.01.008>
- Menting, J. G., Gajewiak, J., MacRaild, C. A., Chou, D. H. C., Disotuar, M. M., Smith, N. A., Miller, C., Erchegeyi, J., Rivier, J. E., Olivera, B. M., Forbes, B. E., Smith, B. J., Norton, R. S., Safavi-Hemami, H., and Lawrence, M. C. (2016). A minimized human insulin-receptor-binding motif revealed in a *Conus geographus* venom insulin. *Nature Structural and Molecular Biology*, 23(10), 916–920. <https://doi.org/10.1038/nsmb.3292>
- Minkiewicz, P., Dziuba, J., Iwaniak, A., Dziuba, M., and Darewicz, M. (2008). BIOPEP database and other programs for processing bioactive peptide sequences. *Journal of AOAC International*, 91(4), 965–980.
- Mojica, L., and de Mejía, E. G. (2015). Characterization and comparison of protein and peptide profiles and their biological activities of improved common bean cultivars (*Phaseolus vulgaris* L.) from Mexico and Brazil. *Plant Foods for Human Nutrition*, 70(2), 105–112. <https://doi.org/10.1007/s11130-015-0477-6>
- Morifuji, M., Koga, J., Kawanaka, K., and Higuchi, M. (2009). Branched-chain amino acid-containing dipeptides, identified from whey protein hydrolysates, stimulate glucose uptake rate in L6 myotubes and isolated skeletal muscles. *Journal of Nutritional Science and Vitaminology*, 55(1), 81–86. <https://doi.org/10.3177/jnsv.55.81>
- Mullally, M. M., Meisel, H., and Fitzgerald, R. J. (1997). Angiotensin-I-converting enzyme inhibitory activities of gastric and pancreatic proteinase digests of whey proteins. *International Dairy Journal*, 7(5), 299–303. [https://doi.org/10.1016/S0958-6946\(97\)00018-6](https://doi.org/10.1016/S0958-6946(97)00018-6)

- Murray, K. K., Boyd, R. K., Eberlin, M. N., Langley, G. J., Li, L., and Naito, Y. (2013). Definitions of terms relating to mass spectrometry (IUPAC Recommendations 2013)*. *Pure Appl. Chem*, *85*(7), 1515–1609. <https://doi.org/10.1351/PAC-REC-06-04-06>
- Najafian, L., and Babji, A. S. (2014). Production of bioactive peptides using enzymatic hydrolysis and identification antioxidative peptides from patin (*Pangasius sutchi*) sarcoplasmic protein hydrolysate. *Journal of Functional Foods*, *9*(1), 280–289. <https://doi.org/10.1016/j.jff.2014.05.003>
- Neduva, V., and Russell, R. B. (2006). Peptides mediating interaction networks: new leads at last. *Current Opinion in Biotechnology*, *17*(5), 465–471. <https://doi.org/10.1016/j.copbio.2006.08.002>
- Nesvizhskii, A. I., Vitek, O., and Aebersold, R. (2007). Analysis and validation of proteomic data generated by tandem mass spectrometry. *Nature Methods*, *4*(10), 787–797. <https://doi.org/10.1038/nmeth1088>
- Neves, A. C., Harnedy, P. A., O’Keeffe, M. B., Alashi, M. A., Aluko, R. E., and FitzGerald, R. J. (2017). Peptide identification in a salmon gelatin hydrolysate with antihypertensive, dipeptidyl peptidase IV inhibitory and antioxidant activities. *Food Research International*, *100*, 112–120. <https://doi.org/10.1016/j.foodres.2017.06.065>
- Neves, A. C., Harnedy, P. A., O’Keeffe, M. B., and FitzGerald, R. J. (2017). Bioactive peptides from Atlantic salmon (*Salmo salar*) with angiotensin converting enzyme and dipeptidyl peptidase IV inhibitory, and antioxidant activities. *Food Chemistry*, *218*, 396–405. <https://doi.org/10.1016/j.foodchem.2016.09.053>
- Newgard, C. B., An, J., Bain, J. R., Muehlbauer, M. J., Stevens, R. D., Lien, L. F., Haqq, A. M., Shah, S. H., Arlotto, M., Slentz, C. A., Rochon, J., Gallup, D., Ilkayeva, O., Wenner, B. R., Yancy, W. S., Eisenson, H., Musante, G., Surwit, R. S., Millington, D. S., Butler, M. D., and Svetkey, L. P. (2009). A Branched-Chain Amino Acid-Related Metabolic Signature that Differentiates Obese and Lean Humans and Contributes to Insulin Resistance. *Cell Metabolism*, *9*(4), 311–326. <https://doi.org/10.1016/j.cmet.2009.02.002>
- Nongonierma, A. B., and FitzGerald, R. J. (2013). Inhibition of dipeptidyl peptidase IV (DPP-IV) by tryptophan containing dipeptides. *Food & Function*, *4*(12), 1843. <https://doi.org/10.1039/c3fo60262a>
- Nongonierma, A. B., and FitzGerald, R. J. (2017). Strategies for the discovery and identification of food protein-derived biologically active peptides. *Trends in Food Science and Technology*, *69*, 289–305. <https://doi.org/10.1016/j.tifs.2017.03.003>

- Ono, S., Hosokawa, M., Miyashita, K., and Takahashi, K. (2003). Isolation of peptides with angiotensin I-converting enzyme inhibitory effect derived from hydrolysate of upstream chum salmon muscle. *Journal of Food Science*, *68*(5), 1611–1614. <https://doi.org/10.1111/j.1365-2621.2003.tb12300.x>
- Ottoson-Seeberger, A., Lundberg, J. M., Alvestrand, A., and Ahlborg, G. (1997). Exogenous endothelin-1 causes peripheral insulin resistance in healthy humans. *Acta Physiologica Scandinavica*, *161*(2), 211–220. <https://doi.org/10.1046/j.1365-201X.1997.00212.x>
- Palzs, B., and Suhal, S. (2005). Fragmentation pathways of protonated peptides. *Mass Spectrometry Reviews*, *24*(4), 508–548. <https://doi.org/10.1002/mas.20024>
- Panchaud, A., Affolter, M., and Kussmann, M. (2012). Mass spectrometry for nutritional peptidomics: how to analyze food bioactives and their health effects. *Journal of Proteomics*, *75*(12), 3546–3559. <https://doi.org/10.1016/j.jprot.2011.12.022>
- Pangestuti, R., and Kim, S. K. (2017). Bioactive peptide of marine origin for the prevention and treatment of non-communicable diseases. *Marine Drugs*, *15*(3). <https://doi.org/10.3390/md15030067>
- Pergande, M., and Cologna, S. (2017). Isoelectric point separations of peptides and proteins. *Proteomes*, *5*(1), 4. <https://doi.org/10.3390/proteomes5010004>
- Picot, L., Ravallex, R., Fouchereau-Péron, M., Vandanion, L., Jaouen, P., Chaplain-Derouiniot, M., Guérard, F., Chabeaud, A., LeGal, Y., Alvarez, O. M., Bergé, J.-P., Piot, J.-M., Batista, I., Pires, C., Thorkelsson, G., Delannoy, C., Jakobsen, G., Johansson, I., and Bourseau, P. (2010). Impact of ultrafiltration and nanofiltration of an industrial fish protein hydrolysate on its bioactive properties. *Journal of the Science of Food and Agriculture*, *90*(11), 1819–1826. <https://doi.org/10.1002/jsfa.4020>
- Pihlanto, A., and Korhonen, H. (2003). Bioactive peptides and proteins. *Advances in Food and Nutrition Research*, *47*(03), 175–276. [https://doi.org/10.1016/S1043-4526\(03\)47004-6](https://doi.org/10.1016/S1043-4526(03)47004-6)
- Pilon, G., Ruzzin, J., Rioux, L. E., Lavigne, C., White, P. J., Frøyland, L., Jacques, H., Bryl, P., Beaulieu, L., and Marette, A. (2011). Differential effects of various fish proteins in altering body weight, adiposity, inflammatory status, and insulin sensitivity in high-fat-fed rats. *Metabolism: Clinical and Experimental*, *60*(8), 1122–1130. <https://doi.org/10.1016/j.metabol.2010.12.005>
- Punthakee, Z., Goldenberg, R., and Katz, P. (2018). Definition, classification and diagnosis of diabetes, prediabetes and metabolic syndrome. *Canadian Journal of Diabetes*, *42*, S10–S15. <https://doi.org/10.1016/j.cjcd.2017.10.003>

- Qian, Z. J., Jung, W. K., Lee, S. H., Byun, H. G., and Kim, S. K. (2007). Antihypertensive effect of an angiotensin I-converting enzyme inhibitory peptide from bullfrog (*Rana catesbeiana* Shaw) muscle protein in spontaneously hypertensive rats. *Process Biochemistry*, 42(10), 1443–1448. <https://doi.org/10.1016/j.procbio.2007.05.013>
- Ramos, Y., Gutierrez, E., Machado, Y., Sánchez, A., Castellanos-Serra, L., González, L. J., Fernández-de-Cossio, J., Pérez-Riverol, Y., Betancourt, L., Gil, J., Padrón, G., and Besada, V. (2008). Proteomics based on peptide fractionation by SDS-free PAGE. *Journal of Proteome Research*, 7(6), 2427–2434. <https://doi.org/10.1021/pr700840y>
- Righetti, P. G., and Boschetti, E. (2013). Chromatographic and Electrophoretic Prefractionation Tools in Proteome Analysis. In *Low-Abundance Proteome Discovery* (pp. 13–40). Elsevier. <https://doi.org/10.1016/B978-0-12-401734-4.00002-6>
- Righetti, P. G., Boschetti, E., and Boschetti, P. G. R. (2013). Current Low-Abundance Protein Access. In *Low-abundance Proteome Discovery* (pp. 41–77). <https://doi.org/http://dx.doi.org/10.1016/B978-0-12-401734-4.00003-8>
- Roblet, C., Akhtar, M. J., Mikhaylin, S., Pilon, G., Gill, T., Marette, A., and Bazinet, L. (2016). Enhancement of glucose uptake in muscular cell by peptide fractions separated by electro dialysis with filtration membrane from salmon frame protein hydrolysate. *Journal of Functional Foods*, 22, 337–346. <https://doi.org/10.1016/j.jff.2016.01.003>
- Roblet, C., Doyen, A., Amiot, J., Pilon, G., Marette, A., and Bazinet, L. (2014). Enhancement of glucose uptake in muscular cell by soybean charged peptides isolated by electro dialysis with ultrafiltration membranes (EDUF): activation of the AMPK pathway. *Food Chemistry*, 147, 124–130. <https://doi.org/10.1016/j.foodchem.2013.09.108>
- Rodríguez, A. A., Ständker, L., Zaharenko, A. J., Garateix, A. G., Forssmann, W. G., Béress, L., Valdés, O., Hernández, Y., and Laguna, A. (2012). Combining multidimensional liquid chromatography and MALDI-TOF-MS for the fingerprint analysis of secreted peptides from the unexplored sea anemone species *Phymanthus crucifer*. *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences*, 903, 30–39. <https://doi.org/10.1016/j.jchromb.2012.06.034>
- Ryder, K., Bekhit, A. E. D., McConnell, M., and Carne, A. (2016). Towards generation of bioactive peptides from meat industry waste proteins: generation of peptides using commercial microbial proteases. *Food Chemistry*, 208, 42–50. <https://doi.org/10.1016/j.foodchem.2016.03.121>

- Saavedra, L., Hebert, E. M., Minahk, C., and Ferranti, P. (2013). An overview of 'omic' analytical methods applied in bioactive peptide studies. *Food Research International*, 54(1), 925–934. <https://doi.org/10.1016/j.foodres.2013.02.034>
- Saiga, A., Tanabe, S., and Nishimura, T. (2003). Antioxidant activity of peptides obtained from porcine myofibrillar proteins by protease treatment. *Journal of Agricultural and Food Chemistry*, 51(12), 3661–3667. <https://doi.org/10.1021/jf021156g>
- Saito, K., Jin, D.-H., Ogawa, T., Muramoto, K., Hatakeyama, E., Yasuhara, T., and Nokihara, K. (2003). Antioxidative properties of tripeptide libraries prepared by the combinatorial chemistry. *Journal of Agricultural and Food Chemistry*, 51(12), 3668–3674. <https://doi.org/10.1021/jf021191n>
- Saito, Y., Wanezaki (Nakamura), K., Kawato, A., and Imayasu, S. (1994). Structure and activity of angiotensin I-converting enzyme inhibitory peptides from sake and sake lees. *Bioscience, Biotechnology, and Biochemistry*, 58(10), 1767–1771. <https://doi.org/10.1271/bbb.58.1767>
- Saltiel, A. R., and Kahn, C. R. (2001). Insulin signalling and the regulation of glucose and lipid metabolism. *Nature*, 414(6865), 799–806. <https://doi.org/10.1038/414799a>
- Satake, M., Enjoh, M., Nakamura, Y., Takano, T., Kawamura, Y., Arai, S., and Shimizu, M. (2002). Transepithelial transport of the bioactive tripeptide, Val-Pro-Pro, in human intestinal Caco-2 cell monolayers. *Bioscience, Biotechnology, and Biochemistry*, 66(2), 378–384. <https://doi.org/10.1271/bbb.66.378>
- Schanbacher, F. L., Talhouk, R. S., and Murray, F. a. (1997). Biology and origin of bioactive peptides in milk. *Livestock Production Science*, 50(1–2), 105–123. [https://doi.org/10.1016/S0301-6226\(97\)00082-1](https://doi.org/10.1016/S0301-6226(97)00082-1)
- Schlichtherle-Cerny, H., Affolter, M., and Cerny, C. (2003). Hydrophilic interaction liquid chromatography coupled to electrospray mass spectrometry of small polar compounds in food analysis. *Analytical Chemistry*, 75(10), 2349–2354. <https://doi.org/10.1021/ac026313p>
- Segura-Campos, M., Chel-Guerrero, L., Betancur-Ancona, D., and Hernandez-Escalante, V. M. (2011). Bioavailability of bioactive peptides. *Food Reviews International*, 27(3), 213–226. <https://doi.org/10.1080/87559129.2011.563395>
- Sellers, J. R., Goodson, H. V, and Wang, F. (1996). A myosin family reunion. *Journal of Muscle Research and Cell Motility*, 17(1), 7–22. <https://doi.org/0142-4319>

- Shao, W., and Lam, H. (2017). Tandem mass spectral libraries of peptides and their roles in proteomics research. *Mass Spectrometry Reviews*, 36(5), 634–648. <https://doi.org/10.1002/mas.21512>
- Shevchenko, A., Tomas, H., Havliš, J., Olsen, J. V., and Mann, M. (2007). In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nature Protocols*, 1(6), 2856–2860. <https://doi.org/10.1038/nprot.2006.468>
- Shtatland, T., Guettler, D., Kossodo, M., Pivovarov, M., and Weissleder, R. (2007). PepBank - a database of peptides based on sequence text mining and public peptide data sources. *BMC Bioinformatics*, 8(1), 280. <https://doi.org/10.1186/1471-2105-8-280>
- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., Thompson, J. D., and Higgins, D. G. (2014). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, 7(1), 539–539. <https://doi.org/10.1038/msb.2011.75>
- Silvestre, M. P. C. (1997). Review of methods for the analysis of protein hydrolysates. *Food Chemistry*, 60(2), 263–271. [https://doi.org/10.1016/S0308-8146\(96\)00347-0](https://doi.org/10.1016/S0308-8146(96)00347-0)
- Soga, M., Ohashi, A., Taniguchi, M., Matsui, T., and Tsuda, T. (2014). The dipeptide Trp-His activates AMP-activated protein kinase and enhances glucose uptake independently of insulin in L6 myotubes. *FEBS Open Bio*, 4(1), 898–904. <https://doi.org/10.1016/j.fob.2014.10.008>
- Son, M., Chan, C. B., and Wu, J. (2018). Egg white ovotransferrin-derived ACE inhibitory peptide ameliorates angiotensin II-stimulated insulin resistance in skeletal muscle cells. *Molecular Nutrition & Food Research*, 62(4), 1700602. <https://doi.org/10.1002/mnfr.201700602>
- Stefansson, G., and Hultin, H. O. (1994). On the solubility of cod muscle proteins in water. *Journal of Agricultural and Food Chemistry*, 42(12), 2656–2664. <https://doi.org/10.1021/jf00048a002>
- Stevenson, C. L., Augustijns, P. F., and Hendren, R. W. (1999). Use of Caco-2 cells and LC/MS/MS to screen a peptide combinatorial library for permeable structures. *International Journal of Pharmaceutics*, 177(1), 103–115. [https://doi.org/10.1016/S0378-5173\(98\)00331-7](https://doi.org/10.1016/S0378-5173(98)00331-7)
- Strohalm, M., Hassman, M., Košata, B., and Kodíček, M. (2008). mMass data miner: an open source alternative for mass spectrometric data analysis. *Rapid Communications in Mass Spectrometry*, 22(6), 905–908. <https://doi.org/10.1002/rcm.3444>

- Tahergorabi, R., Beamer, S. K., Matak, K. E., and Jaczynski, J. (2012). Isoelectric solubilization/precipitation as a means to recover protein isolate from striped bass (*Morone saxatilis*) and its physicochemical properties in a nutraceutical seafood product. *Journal of Agricultural and Food Chemistry*, *60*(23), 5979–5987. <https://doi.org/10.1021/jf3001197>
- Tahrani, A. A., Bailey, C. J., Del Prato, S., and Barnett, A. H. (2011). Management of type 2 diabetes: new and future developments in treatment. *The Lancet*, *378*(9786), 182–197. [https://doi.org/10.1016/S0140-6736\(11\)60207-9](https://doi.org/10.1016/S0140-6736(11)60207-9)
- Tan, M.-J., Ye, J.-M., Turner, N., Hohnen-Behrens, C., Ke, C.-Q., Tang, C.-P., Chen, T., Weiss, H.-C., Gesing, E.-R., Rowland, A., James, D. E., and Ye, Y. (2008). Antidiabetic activities of triterpenoids isolated from bitter melon associated with activation of the AMPK pathway. *Chemistry & Biology*, *15*, 263–273. <https://doi.org/10.1016/j.chembiol.2008.01.013>
- Tang, Y., Li, R., Lin, G., and Li, L. (2014). PEP search in MyCompoundID: detection and identification of dipeptides and tripeptides using dimethyl labeling and hydrophilic interaction liquid chromatography tandem mass spectrometry. *Analytical Chemistry*, *86*(7), 3568–3574. <https://doi.org/10.1021/ac500109y>
- Tompa, P., Davey, N. E., Gibson, T. J., and Babu, M. M. (2014). A million peptide motifs for the molecular biologist. *Molecular Cell*, *55*(2), 161–169. <https://doi.org/10.1016/j.molcel.2014.05.032>
- Udenigwe, C. C., Li, H., and Aluko, R. E. (2012). Quantitative structure–activity relationship modeling of renin-inhibiting dipeptides. *Amino Acids*, *42*(4), 1379–1386. <https://doi.org/10.1007/s00726-011-0833-2>
- Van Der Ven, C., Gruppen, H., De Bont, D. B. A., and Voragen, A. G. J. (2001). Reversed phase and size exclusion chromatography of milk protein hydrolysates: relation between elution from reversed phase column and apparent molecular weight distribution. *International Dairy Journal*, *11*(1–2), 83–92. [https://doi.org/10.1016/S0958-6946\(01\)00032-2](https://doi.org/10.1016/S0958-6946(01)00032-2)
- Vik, R., Tillander, V., Skorve, J., Vihervaara, T., Ekroos, K., Alexson, S. E. H., Berge, R. K., and Bjørndal, B. (2015). Three differently generated salmon protein hydrolysates reveal opposite effects on hepatic lipid metabolism in mice fed a high-fat diet. *Food Chemistry*, *183*, 101–110. <https://doi.org/10.1016/j.foodchem.2015.03.011>
- Wang, P. A., Vang, B., Pedersen, A. M., Martinez, I., and Olsen, R. L. (2011). Post-mortem degradation of myosin heavy chain in intact fish muscle: effects of pH and enzyme inhibitors. *Food Chemistry*, *124*, 1090–1095. <https://doi.org/10.1016/j.foodchem.2010.07.093>

- Wang, P., and Wilson, S. R. (2013). Mass spectrometry-based protein identification by integrating *de novo* sequencing with database searching. *BMC Bioinformatics*, 14(S2), S24. <https://doi.org/10.1186/1471-2105-14-S2-S24>
- Wasswa, J., Tang, J., and Gu, X. (2007). Utilization of fish processing by-products in the gelatin industry. *Food Reviews International*, 23(2), 159–174. <https://doi.org/10.1080/87559120701225029>
- Watt, P. M. (2006). Screening for peptide drugs from the natural repertoire of biodiverse protein folds. *Nature Biotechnology*, 24(2), 177–183. <https://doi.org/10.1038/nbt1190>
- Wee, S., O'Hair, R. A. J., and McFadyen, W. D. (2006). The role of the position of the basic residue in the generation and fragmentation of peptide radical cations. *International Journal of Mass Spectrometry*, 249–250, 171–183. <https://doi.org/10.1016/j.ijms.2006.01.007>
- Whitaker, J. R., Feeney, R. E., and Sternberg, M. M. (1983). Chemical and physical modification of proteins by the hydroxide ion. *CRC Critical Reviews in Food Science and Nutrition*, 19(3), 173–212. <https://doi.org/10.1080/10408398309527375>
- Wiernsperger, N. F., and Bailey, C. J. (1999). The antihyperglycaemic effect of metformin. *Drugs*, 58(S1), 31–39. <https://doi.org/10.2165/00003495-199958001-00009>
- Wilcox, G. (2005). Insulin and insulin resistance. *The Clinical Biochemist Reviews*, 26(2), 19–39. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/16278749>
- Wu, J., Aluko, R. E., and Nakai, S. (2006). Structural requirements of angiotensin I-converting enzyme inhibitory peptides: quantitative structure–activity relationship study of di- and tripeptides. *Journal of Agricultural and Food Chemistry*, 54(3), 732–738. <https://doi.org/10.1021/jf051263l>
- Wysocki, V. H., Tsaprailis, G., Smith, L. L., and Brechi, L. A. (2000). Mobile and localized protons: a framework for understanding peptide dissociation. *Journal of Mass Spectrometry*, 35(12), 1399–1406. [https://doi.org/10.1002/1096-9888\(200012\)35:12<1399::AID-JMS86>3.0.CO;2-R](https://doi.org/10.1002/1096-9888(200012)35:12<1399::AID-JMS86>3.0.CO;2-R)
- Xiao, Y., Vecchi, M. M., and Wen, D. (2016). Distinguishing between leucine and isoleucine by integrated LC–MS analysis using an orbitrap fusion mass spectrometer. *Analytical Chemistry*, 88(21), 10757–10766. <https://doi.org/10.1021/acs.analchem.6b03409>

- Xiao, Z., Conrads, T. P., Lucas, D. A., Janini, G. M., Schaefer, C. F., Buetow, K. H., Issaq, H. J., and Veenstra, T. D. (2004). Direct ampholyte-free liquid-phase isoelectric peptide focusing: application to the human serum proteome. *Electrophoresis*, 25(1), 128–133. <https://doi.org/10.1002/elps.200305700>
- Yano, S., Suzuki, K., and Funatsu, G. (1996). Isolation from α -zein of thermolysin peptides with angiotensin I-converting enzyme inhibitory activity. *Bioscience, Biotechnology, and Biochemistry*, 60(4), 661–663. <https://doi.org/10.1271/bbb.60.661>
- Yoon, M.-S. (2016). The emerging role of branched-chain amino acids in insulin resistance and metabolism. *Nutrients*, 8(7), 405. <https://doi.org/10.3390/nu8070405>
- Zamyatnin, A. A. (2006). The EROP-Moscow oligopeptide database. *Nucleic Acids Research*, 34(Database issue), D261–D266. <https://doi.org/10.1093/nar/gkj008>
- Zhang, J., Xue, C., Zhu, T., Vivekanandan, A., Pennathur, S., Ma, Z. A., and Chen, Y. E. (2013). A tripeptide diapin effectively lowers blood glucose levels in male type 2 diabetes mice by increasing blood levels of insulin and GLP-1. *PLoS ONE*, 8(12), e83509. <https://doi.org/10.1371/journal.pone.0083509>
- Zhou, R., Grant, J., Goldberg, E. M., Ryland, D., and Aliani, M. (2019). Investigation of low molecular weight peptides (<1 kDa) in chicken meat and their contribution to meat flavor formation. *Journal of the Science of Food and Agriculture*, 99(4), 1728–1739. <https://doi.org/10.1002/jsfa.9362>
- Zielińska, E., Baraniak, B., and Karaś, M. (2018). Identification of antioxidant and anti-inflammatory peptides obtained by simulated gastrointestinal digestion of three edible insects species (*Gryllodes sigillatus*, *Tenebrio molitor*, *Schistocerca gregaria*). *International Journal of Food Science & Technology*, 53(11), 2542–2551. <https://doi.org/10.1111/ijfs.13848>

APPENDIX I

Table A1. Peptides identified by LC-MS in fraction 1 from SAX Separation 1. The Sequest HT program was used to generate XCorr scores for each database search hit.

Sequence	Retention Time (Min)	XCorr Score	Precursor Ion Mass (Da)
AAADGPMK	3.04	2.03	760.36504
IQAQ GK	3.17	1.30	615.38146
LVQEAR	3.68	1.64	715.40862
KYDSTHGR	4.51	2.31	963.46268
PTNKEVR	4.52	1.91	843.46854
ATQKTVDGPS GK	4.70	2.87	1188.62409
TVDGPS GK	5.11	2.32	760.38268
KVINAQTK	5.14	1.80	901.54729
VGDEAQS KR	5.17	3.19	989.50066
AVDHV NK	5.31	2.03	782.41417
GKPSADDMANKR	5.71	2.71	1289.62847
TLPPHNSR	5.74	2.20	921.49126
TGESGAGK	5.77	3.04	706.33636
KPNMVTAGH	6.06	2.88	954.48381
HERDPANIK	6.62	2.79	1079.56107
ATAASSSSLEK	6.80	2.33	1051.52849
ITSTTPSR	7.22	2.56	862.46361
AASSSSLEK	7.65	2.27	879.44328
KITSTTPSR	8.26	2.25	990.55754
AGPPGGDGQPGAK	8.35	2.76	1108.53777
ITGESGAGK	8.55	2.87	819.41919
GGAGVAGAK	8.97	2.74	687.37909
VMGVNHEK	9.60	2.80	913.45770
AVDHV NKDIAAK	9.83	4.28	1280.69597
AVTATQKTVDGPS GK	10.19	2.83	1459.77500
TSVAPGSDPVKK	10.46	2.67	1185.64543
TTIEKASTHLK	10.51	2.22	1228.68967
TFPEHV KSR	10.82	2.55	1100.58414
AILGKPSADDmANKR	11.14	3.06	1602.82395
LSGTGSAGATVR	11.20	4.13	1076.57048
LQASALK	11.41	1.99	730.44567
IIPASTGAAK	11.59	2.83	928.54547
PGIADRMQK	11.63	2.73	1015.53504
VVYPQTK	11.68	1.42	834.47051
TEAPLNPK	12.06	2.26	869.47118
LTVTNPK	12.22	1.53	772.45531
AGPSIVH	12.77	1.67	680.37395

Sequence	Retention Time (Min)	XCorr Score	Precursor Ion Mass (Da)
MATAASSSSLEK	12.87	3.60	1182.56682
ALGQNPTNKEVR	12.93	3.60	1326.71308
VVIGHVDSGK	13.04	1.75	1010.56163
IAPPERK	13.43	1.48	810.48387
IIAPPERK	13.63	2.24	923.56665
KIIAPPER	13.90	2.92	923.56664
LKPNMVTAGH	13.93	2.45	1067.56670
AAVPSGASTGVH	14.26	3.30	1053.53276
ALAPSTMK	14.51	1.94	818.44243
RVEELSKFSKK	14.64	2.53	1350.77334
VGNEYVTK	15.40	1.83	909.46581
ATQKTVDGPSGKL	15.42	3.09	1301.70636
SLSETHANKL	15.58	1.88	1099.57475
SKSTPSGF	15.58	1.94	810.39813
ISTAGTEASGTGNMK	15.71	3.66	1424.66765
TQKTVDGPSGKL	15.75	2.69	1230.66648
TDAGKTFPEHVK	15.85	3.02	1329.67885
VEELSKFSKK	16.10	2.59	1194.67060
IMGNPAVAK	16.19	2.30	900.49828
FHGEVKAEGGKL	16.32	2.39	1271.67231
IVGDKDYSTTAHSKV	16.50	2.88	1620.82010
PSIVGRP	16.50	2.10	725.42900
KAPGIIPR	16.57	2.88	851.54686
VDGPSGKL	16.64	2.15	772.41895
AYPDQKPGTSGLR	16.66	2.99	1389.71088
KGLKPGEL	16.75	2.83	841.51512
ILGKPSADDMANKR	16.80	3.19	1515.79397
DQGSSEKIQKL	16.93	2.65	1319.68064
IISKLENHEGVR	17.14	4.27	1394.77457
DVVPTNIHQVR	17.43	2.55	1178.62859
VERPESTTVIK	17.54	3.22	1258.70176
GFPDSDGAAGPK	17.62	2.76	1060.50757
IAIKPDGVQR	17.88	2.75	1096.64573
IGSVTESIK	17.93	3.15	933.52587
FHLLPR	17.93	2.82	919.52488
LEKPASYDAIKK	17.95	3.15	1362.76471
FHETLYQKTDAGK	17.97	2.81	1537.76419
ASMGQLNVK	17.97	3.65	947.49847
PIISDRHSGY	18.00	2.53	1144.57248
ADLSPGSAPVK	18.13	2.91	1041.55559
VIGSGTNLDVGR	18.19	2.83	1175.60088
LIPAQTQHPIR	18.23	2.71	1273.73955
GNEQFISASK	18.38	2.47	1080.53264
LGKPSADDMANKR	18.44	4.08	1402.70917

Sequence	Retention Time (Min)	XCorr Score	Precursor Ion Mass (Da)
LTEAPLNPK	18.49	3.39	982.55615
VGMGQKDSYVGDEAQSQR	18.50	4.24	1954.92514
PSADDMANKR	18.70	1.77	1104.50945
VSTLGEGAFGK	18.76	2.45	1065.55927
PSNLGTGLR	18.85	3.24	914.50609
AILGKPSADDMANKR	18.92	3.80	1586.82932
TTVPAGLPDHK	18.93	2.14	1135.60930
IIDQNRDGIISK	18.96	2.78	1371.75766
VITIGNER	19.00	1.90	901.50957
LEKPASYDAIK	19.01	2.96	1234.66669
VEELSKFSK	19.06	2.09	1066.57587
ITSTTPSRL	19.08	2.43	975.54570
LTmFGEK	19.10	1.63	841.41380
AADESTGSVAKRF	19.28	2.94	1338.66428
LEVPKDIPR	19.30	2.11	1066.62448
ANVSEAVANSVR	19.44	3.47	1216.62737
VTIIDAPGHR	19.45	3.19	1078.60039
AVTATQKTVDGPSGKL	19.50	3.72	1572.85855
KVIEVQGIK	19.51	3.10	1013.63603
KVNQIGSVTESIK	19.55	4.67	1402.79094
KIKIIAPPER	19.55	2.68	1164.74585
VLVGEPVGSR	19.67	2.99	1012.57713
SNRPAFMPSEGK	19.69	3.04	1320.63828
DGRGASQNIIPASTGAAK	19.81	4.64	1713.88725
QIGSVTESIK	19.86	2.81	1061.58196
KPASYDAIK	19.89	2.01	992.54198
SQNIIPASTGAAK	19.90	2.98	1257.67864
GASQNIIPASTGAAK	19.95	3.86	1385.73613
GTTmYPGIADR	19.98	2.60	1197.55620
EKPASYDAIK	20.01	1.64	1121.58196
YPGIADRMQK	20.02	2.86	1178.59879
ASQNIIPASTGAAK	20.08	3.63	1328.71489
PASYDAIK	20.08	3.00	864.44482
SGVNVAGVSLK	20.14	3.02	1030.58769
VVYPQTKTY	20.16	1.84	1098.58196
EITALAPSTmK	20.24	2.16	1177.61211
AILGKPSADDMANK	20.28	3.30	1430.72980
NIIPASTGAAK	20.33	2.34	1042.58965
TAYNSIMK	20.53	1.99	927.45952
ALEKPARPF	20.58	1.86	1028.58725
TPTPAQTPTPSLK	20.60	3.77	1338.72844
GPEIRTGL	20.71	2.19	842.47252
PVPALPEK	20.72	1.70	850.50383
PFLTPDQK	20.82	2.92	945.50340

Sequence	Retention Time (Min)	XCorr Score	Precursor Ion Mass (Da)
VIPELNGK	20.92	1.66	869.50749
VESILKNLHK	20.92	2.54	1180.70295
AAQEEFIKR	21.18	3.00	1091.58332
SGGTTmYPGIADRMQK	21.22	3.81	1728.80216
IKIIAPPER	21.29	2.84	1036.64975
LSNRPAFMPSEGK	21.60	2.96	1433.71790
GLMSTVHAVTATQK	21.61	3.12	1443.76215
IASRPGVF	21.72	2.27	846.48271
VIVSAFK	21.72	1.53	763.47191
IPTMPEGPK	21.79	2.17	969.50633
TIIDQNRDGIISK	21.91	4.09	1472.80730
VIEVQGIK	21.95	2.97	885.53876
IALNDHFVK	21.97	2.22	1056.58298
VNQIGSVTESIK	22.00	3.56	1274.69487
SMPTSGALDNVAK	22.01	2.03	1290.63481
ITIGNERFR	22.03	1.69	1105.61185
AADESTGSVAKRFQSIN	22.12	3.46	1780.88028
IALDTKGPEIR	22.23	3.19	1212.69294
VVEIQGIK	22.29	2.10	885.53880
ALDTKGPEIR	22.32	3.08	1099.60881
LDTKGPEIR	22.37	2.09	1028.57182
VVDGVKL	22.41	1.49	729.44957
GTTMYPGIADR	22.56	2.81	1181.55974
TLLKPNMVTAGH	22.71	1.96	1281.69641
ILPIGASN	22.89	1.51	784.45508
ALTFSYGR	22.91	2.06	914.47423
AITFSYGR	22.91	2.06	914.47423
SGGTTMYPGIADR	22.92	2.98	1325.61333
IELPPTHPIR	22.96	3.33	1172.67890
GPGTIEYRPVA	23.02	3.25	1159.60845
PLVVTPPQ	23.13	1.71	850.50224
GTTMYPGIADRMQK	23.14	2.57	1568.75370
TTMYPGIADRMQK	23.24	3.16	1511.73484
VFPSIVGRPR	23.28	3.17	1127.66709
GLMSTVHAVT	23.31	2.15	1015.52318
VPIVPSGIK	23.35	2.40	909.57640
LLTEAPLNPK	23.49	2.57	1095.64055
SSGTSYPDVLK	23.49	2.67	1153.57231
IGGIGTVPVGR	23.55	3.15	1025.60991
IVLGTLGEK	23.73	2.60	929.56651
SGGTTMYPGIADRMQK	23.75	3.49	1712.80661
VLGTLGEK	23.86	2.51	816.48193
PIVIPGKP	23.86	2.00	820.52764
KLEKGEAIDGMIPAQK	23.92	2.91	1727.93593

Sequence	Retention Time (Min)	XCorr Score	Precursor Ion Mass (Da)
TVRNDITLLK	24.00	3.44	1172.69858
AADTFNFK	24.10	2.09	913.44286
QVSLQDKTGF	24.28	1.90	1122.57739
VNQIGSVTESIKA	24.35	2.94	1345.73198
APPPLDLSTVK	24.36	2.22	1137.65251
EITALAPSTMK	24.55	2.48	1161.61675
AVFPSIVGRPR	24.64	3.13	1198.70597
VLASMGQLNVK	24.78	2.82	1159.65080
GEAIDGMIPAQK	24.79	2.30	1229.61809
AGDDAPRAVFPSIVGRPR	24.83	4.98	1881.00600
FLTEIQSPR	24.90	1.48	1090.58745
VDSLVPIGR	24.98	1.88	955.55638
HPTLLTEAPLNPK	25.21	4.41	1430.79814
VGGAALKPEF	25.25	2.74	988.54448
VLSIGDGIAR	25.30	2.90	1000.57726
FPSIVGRPR	25.31	2.94	1028.60117
ILPIGASNFHEAmR	25.32	2.21	1571.79947
VIGSGTNLDSGRF	25.34	3.23	1322.66875
VTVDDKSDTSVTITWRPPK	25.56	2.72	2145.11464
NDIMLIK	25.65	1.98	846.47521
MLESMIK	25.70	2.04	851.43657
VFPSIVGR	25.78	1.67	874.51598
NNDIMLIK	25.84	2.73	960.51860
RPVAIALDTKGPEIR	25.85	5.05	1635.95688
AAPPPLDLSTVK	25.92	2.66	1208.69084
KEAFTIIDQNRDGIISK	26.02	4.44	1948.04789
VAYNQVADIMR	26.03	3.18	1279.64433
LKVNQIGSVTESIKA	26.08	2.93	1586.91373
VVIPAGVPR	26.14	2.36	1006.64214
DAVGMSLIK	26.15	2.42	933.50786
LDVNLKPVKPM	26.22	2.45	1253.72815
INNDITLLK	26.39	3.02	1043.60857
LEGTLLKPNMVTAGH	26.41	3.38	1580.84465
IALDTKGPEIRTGLIK	26.46	3.03	1725.03029
RVPTPNVSVVDLTVR	26.52	4.53	1651.94693
ASLGELIK	26.58	2.12	830.49755
AGDDAPRAVFPSIVGR	26.83	4.58	1627.85373
FPSIVGR	26.97	2.13	775.44554
IALDTKGPEIRTGL	27.02	2.94	1483.84494
AVFPSIVGRPRHQGVM	27.03	2.21	1750.95266
VDVSKGFL	27.08	2.45	864.48149
ATQMLESNIK	27.08	2.48	1151.57793
MVVPVESPIR	27.13	2.75	1126.63066
GTTMYPGIADRM	27.14	1.85	1312.60051

Sequence	Retention Time (Min)	XCorr Score	Precursor Ion Mass (Da)
FQTNLVPYPR	27.16	2.58	1234.65972
NMGNILATYK	27.36	2.67	1124.57585
VAAPPPLDLSTVK	27.38	3.18	1307.75749
PTLLTEAPLNPK	27.41	3.38	1293.73894
TVPPAVPGVT	27.42	1.51	937.53485
VIDYKPTPF	27.44	1.74	1079.57927
TLNNDImLIK	27.70	2.04	1190.64336
ANNVLSGGTTMYPGIADR	27.81	4.21	1836.89018
PPPLDLSTVK	27.86	2.78	1066.61345
AFTIIDQNRDGIISK	27.88	3.08	1690.91269
ILPIGASNFHEAMR	27.91	2.78	1555.80429
TINNDITLLK	27.93	3.54	1144.65630
LVDPLGPGLK	27.96	2.90	1008.61077
VPIVPSGIKY	28.04	2.16	1072.63970
AVFPSIVGR	28.06	2.93	945.54974
LPVYRTW	28.76	1.43	934.51445
NDIMLIK	28.87	2.40	1087.65483
FVPIVPSGIK	29.05	2.24	1056.64531
AVFPSIVGRP	29.18	2.42	1042.60625
VVAATLQDIVR	29.32	3.52	1184.69744
ALQASALKAW	29.34	2.52	1058.59783
ILPIPDNNAVIGR	29.38	2.91	1391.79778
AGFAGDDAPRAVFPSIVGRPR	29.43	4.28	2156.13247
IIEPVGPPASIK	29.55	2.76	1234.73906
FDKPVSPLL	29.76	2.50	1015.58299
VVYPQTKTYFSHW	29.80	2.30	1655.82035
VPTPNVSVVDLTVR	29.87	3.96	1495.84831
TLNNDIMLIK	30.06	2.42	1174.64873
KPEFVDIINAK	30.36	4.12	1273.71306
AYEPVWAIGTGK	30.49	1.92	1291.66863
NTLNNDIMLIK	30.68	2.06	1288.69353
SNTLNNDIMLIK	30.84	3.00	1375.72270
TIIP LISQATPK	30.84	3.08	1281.77703
LLIPQIVK	30.92	2.27	923.62969
NSNTLNNDIMLIK	30.97	3.58	1489.76616
AGFAGDDAPRAVFPSIVGR	31.09	3.39	1902.98398
NSKYNSLTINNDITLLK	31.24	3.88	1951.04898
LVVDGVKL	31.35	1.88	842.53484
LTINNDITLLK	31.39	3.26	1257.73882
PSFNRTPIGW	31.41	3.46	1174.59856
AAPPPLDLSTVKV	31.61	1.94	1307.75701
NSLTINNDITLLK	31.90	4.14	1458.81804
GLVDKFPL	31.94	1.84	888.51726
SLTINNDITLLK	32.04	3.06	1344.77239

Sequence	Retention Time (Min)	XCorr Score	Precursor Ion Mass (Da)
GAALKPEFVDIINAK	32.22	3.07	1585.89255
LFDKPVSPLL	32.36	2.22	1128.66496
VAAPPPLDLSTVKV	32.44	2.59	1406.82329
FIIPQIVK	32.53	1.81	957.61193
GTVSGAELRIVL	32.55	2.30	1214.70891
VGGAALKPEFVDIINAK	32.79	3.09	1741.98502
TVPPAVPGVTFLSGGQ	32.87	2.18	1526.81828
INPQSMFDFQVK	32.95	2.73	1453.71257
ILPIGASNF	33.00	1.15	931.52313
LMPIVEIVR	33.04	2.19	1069.64189
YNSLTINNDITLLK	33.15	3.17	1621.87676
SIVHPSYNSNTLNNDIMLIKLK	33.64	3.86	2514.33901
PPAVPGITF	33.90	2.18	898.50220
TVPPAVPGVTF	34.06	2.56	1084.60437
PSIVGRPR	34.14	2.23	881.53099
TVPPAVPGITFLSGGQ	34.44	2.05	1540.83611
AVPGSWPWQVSLQDK	34.92	2.78	1697.86260
FDKPVSPLLLSAGM	36.60	1.96	1474.79485
VGGLFWMFR	36.83	1.77	1112.56890
TVPPAVPGVTFL	37.30	1.65	1197.68787
GLVDKFPLF	37.43	2.49	1035.58611
TVDGPSGKL	38.11	1.66	873.46776
AVPGSWPWQVSLQDKTGF	38.37	2.20	2002.99870
AAIIIQIIELPGPPASIK	39.43	2.06	1844.12908
ERGGAGVAGAKGNTGEPG	40.24	1.32	1584.77263
VVDVGK	40.30	1.68	616.36583
FGNAEGEFCKFPFMfmgKEYNSC			
TNQGRDDGFLWC	41.21	0.23	4091.69643
PASTGAAK	42.04	2.64	702.37787
VVDGVK	43.83	1.55	616.36662
VDIINAK	43.92	1.50	772.45629
PIISDR	45.57	1.44	700.39995
VGMGQK	46.76	1.31	619.32335
NEKLGGMHSLDMPTAVNYLSMSD			
TAmQVLGAAYIQHQC	47.19	0.47	4140.90346
ALAPSTM	49.62	1.60	690.34796
IIAPPER	58.49	2.05	795.47203
PSIVGR	58.93	1.43	628.37737

Table A2. Peptides identified by LC-MS in fraction 2 from SAX Separation 1. The Sequest HT program was used to generate XCorr scores for each database search hit.

Sequence	Retention Time (Min)	XCorr Score	Precursor Ion Mass (Da)
VITHGEEK	4.91	1.69	912.47678
AGDDAPR	4.93	2.24	701.32250
ADESTGSVAK	4.96	1.90	964.45885
VIENRASKDEEK	5.24	2.04	1417.73003
HVAEEGK	5.25	2.61	769.38542
SPDDPAR	5.47	2.56	757.34642
LENHEGVR	5.47	2.25	953.47913
KLEEA EK	5.49	1.53	846.45629
DSGDGVTH	5.64	2.60	787.32225
DKENALDR	5.87	2.09	960.47307
QREEQAEPDGTE	6.15	2.96	1388.59160
DISNADR	6.16	2.48	790.36761
HERDPAN	6.31	2.28	838.38109
DIGGER	6.52	1.97	646.31529
TEHTPEGVK	6.81	1.51	997.49639
DESTGSVAK	7.03	2.72	893.42273
TLKPEEEK	7.09	1.44	973.52110
VITHGEEKEE	7.19	1.98	1170.56559
AELSEGK	7.84	1.80	733.37163
HIAEEADR	8.00	2.76	940.44945
DTDGGDISTK	8.34	2.22	1065.46965
AADESTGSVAK	8.46	3.99	1035.49487
TKLEEA EK	8.98	2.94	947.50469
TASPAQAQDVHDK	9.01	3.12	1367.65764
LGDAETVK	10.21	2.14	832.44243
APAEPAPEQPK	10.33	2.88	1134.57891
DSYVGDEAQSK	10.41	3.36	1198.52039
DSYVGDEAQSKR	10.61	3.63	1354.62073
DEAGPSIVHR	10.70	3.17	1080.54399
MESAGIHE	11.00	2.80	873.37628
YVGDEAQSKR	11.15	1.76	1152.56340
VGmGQKDSYVGDEAQSK	11.42	2.12	1814.82401
SQKEDKYEEEIK	12.58	4.80	1525.73638
VEEELDRAQER	12.69	2.74	1373.66728
FVTTAERE	12.88	2.19	952.47246
IEEELGAK	13.10	2.97	888.46771
GPAPEGITDK	13.23	2.92	984.50109
HFADNIKD	13.42	2.84	959.45978
DEAGPSIVH	13.58	2.66	924.44194

Sequence	Retention Time (Min)	XCorr Score	Precursor Ion Mass (Da)
GFAGDDAPR	13.76	2.18	905.41014
EAGPSIVH	14.15	1.85	809.41545
IDSEPAVAR	14.48	2.71	957.50170
IIHGSDSLDSANKE	14.79	3.48	1485.71941
FSVVDQEK	14.93	1.84	951.47734
DKENALDRAEGAEGDKK	15.49	2.52	1845.89646
IGmESAGIHET	15.54	2.24	1160.52312
LELDGTENK	15.67	2.99	1018.50397
RFQSINTENTEENR	16.13	2.11	1737.81828
AAVPSGASTGVHE	16.39	3.02	1182.57439
IEEHDRPTL	16.50	2.36	1109.55925
LEKPASYDA	16.54	1.85	993.48808
IGmESAGIHE	16.54	2.48	1059.47551
YDEAGPSIVHR	16.61	2.72	1243.60430
AGFAGDDAPR	17.07	3.71	976.44971
SEQISTAGTEASGTGNMK	17.14	3.70	1768.80449
KVEEEYPDLTKHNNHMAK	17.28	3.89	2183.05434
KEEAPAPAPAEAAPAE	17.41	4.60	1548.75566
LFTADDR	17.54	2.00	837.40886
LSQTIDKVDEER	17.61	2.61	1432.73032
FQSINTENTEENR	17.65	3.77	1581.71489
FQSINTENTEENRR	17.74	3.14	1737.81845
VEEVDAMDAGK	17.93	2.21	1163.52434
GEAIDGmIPAQK	18.00	2.17	1245.61626
VGMGQKDSYVGDEAQSK	18.06	5.60	1798.82878
SFVTTAERE	18.14	1.82	1039.50359
EYDEAGPSIVHR	18.19	2.67	1372.65104
FDADLSEK	18.48	1.84	924.43224
TATDAEADVASLNRR	18.59	1.95	1589.78996
LGQNPTREELDE	19.06	2.50	1400.66789
KVEEEYPDLTK	19.27	3.43	1350.67949
YDEAGPSIVH	19.28	2.31	1087.50664
EKIPAPDEQLK	19.41	2.35	1267.68694
LKGTEDELDKY	19.51	2.20	1310.64910
VDWTDAEK	19.64	1.53	963.44335
IGMESAGIH	19.77	2.76	914.43865
GKDATNVGDEGGFAPN	19.87	1.65	1548.69451
TLESEKF	19.95	1.21	853.42919
IVLDSGDGVTH	19.98	3.72	1112.55920
NIASGGPAPEGITDK	20.03	2.79	1426.71807
SLEAQAEKY	20.11	3.37	1038.51050
AGDDAPRAVF	20.13	2.92	1018.49474
IGMESAGIHET	20.18	2.39	1144.53227
PEILPDGDHDLKR	20.24	4.39	1504.77307

Sequence	Retention Time (Min)	XCorr Score	Precursor Ion Mass (Da)
ITKQEYDEAGPSIVH	20.27	2.45	1686.83562
EYDEAGPSIVH	20.29	1.89	1216.54827
HIIEGLmSTVH	20.30	2.23	1252.63628
TATDAEADVASLNR	20.31	4.62	1433.68437
TLESEKFDNMER	20.31	3.33	1498.68181
IPAPDEQLK	20.36	2.21	1010.54991
VVGDDLTVTNPK	20.36	2.33	1257.66826
PEILPDGDHD	20.38	1.84	1107.49626
VVAGEFDQGSSEK	20.56	3.60	1538.73516
QEYDEAGPSIVH	20.58	3.30	1344.60625
VDWTD AER	20.86	1.84	991.44805
DLDFKSPDDPAR	20.91	2.67	1375.65129
AAAPAPAPEPEV	21.06	2.62	1119.56677
ATALTKLEEA EK	21.10	3.14	1303.71269
VVAGEFDQGSSEKI QK	21.18	4.28	1907.97368
DISNADRLGSSEVDQ	21.19	3.34	1605.73674
DFENEmATAASSSLEK	21.24	3.86	1832.78642
IGEVVNHD PVIGDR	21.25	3.07	1519.78754
LELDGTENKSKF	21.74	2.77	1380.69806
PDGDHDLK	21.84	2.26	896.41119
AAVPSGASTGVHEAL	22.00	2.64	1366.69756
GEFDQGSSEKI QKL	22.31	3.01	1652.81211
WVNEEDHL	22.40	1.53	1041.46416
VIIESDLER	22.41	3.04	1073.58435
VAPEEHPTLL	22.56	2.35	1105.58826
SDVVAGEFDQGSSEKI QK	22.81	3.85	2110.03301
AAAPAPAPEPDVV	22.87	2.95	1204.61890
AAQE EFK	22.87	2.34	935.48247
DVVAGEFDQGSSEK	22.95	3.26	1653.76018
DVVAGEFDQGSSEKI QK	22.96	4.81	2022.99956
IGMESAGIHETAYN	23.01	2.42	1492.67583
VIIESDLERTEER	23.02	2.47	1588.81748
PEILPDGDHDLKRTQY	23.09	3.24	1896.94609
VIIESDLERTEE	23.18	2.48	1432.71697
KAGDADGDGMIGIDE	23.50	3.06	1463.63176
GEAIDGMIPAQK	23.60	2.97	1229.62187
VIQTGV DNP GHPF	23.83	2.15	1380.69267
DAGAGIALNDHFVK	23.83	2.47	1427.72989
IPDGEKVDFDDIQK	24.00	1.99	1618.79825
AASSSLEKSYELPDGQ	24.07	3.02	1768.82012
GAGIALNDHFVK	24.29	2.38	1241.66191
IPDGDKVDFDDIQK	24.30	3.20	1604.77823

Sequence	Retention Time (Min)	XCorr Score	Precursor Ion Mass (Da)
TDLNFENLKG	24.41	2.30	1150.57158
LDLAGRDLTD	24.80	2.23	1088.55974
LEIPEYDGK	24.93	1.58	1063.53252
VWVNEEDHL	24.98	2.11	1140.53191
VAIQLNDTHPAM	25.01	1.95	1309.65935
WITKQEYDEAGPSIVH	25.20	3.97	1872.91155
IQVVGDDLTVTNPK	25.39	2.90	1498.81206
VAPEEHPTLLTEAPLNPK	25.47	4.42	1956.04228
IDDGLMSLK	25.50	2.44	991.51293
RIQLVEEELDRAQER	25.77	4.24	1883.99631
GGDDLDPNYVLSSR	25.90	3.56	1507.69951
HIIEGLMSTVH	25.90	3.45	1236.63872
LVIIESDLERTEER	25.90	2.47	1701.89836
TDLNFENLK	25.91	1.70	1093.55425
GIITNWDDmEK	25.93	1.99	1337.60674
GDDLDPNYVLSSR	25.96	2.70	1450.67839
IGEVVNHDPVIGDRL	26.09	2.90	1632.87248
PEILPDGDHDL	26.09	2.67	1220.58142
PTVEVDLYTAK	26.31	2.93	1235.65361
PGLAEVIAER	26.44	3.26	1054.59062
DAGAGIALNDHF	26.55	3.33	1200.56470
IPVGPETL	26.59	2.18	825.47302
LVIIESDLER	26.75	3.47	1186.66687
FQPSFIGmESAGIHE	26.96	2.15	1665.76091
VVESTGVFTTIEK	27.01	3.12	1409.75493
ILPIGASNFHE	27.26	2.10	1197.62444
EAFITDQNRDGIISK	27.27	3.42	1819.95452
PIVEPEILPDGDHDLK	27.28	2.76	1786.92496
SDVVVAGEFDQGSSEKI QKL	27.32	4.56	2223.11699
DVVVAGEFDQGSSEKI KL	27.55	5.88	2136.07793
TITGIPNSEATHF	27.82	1.93	1387.68779
HIIEGLMSTVHAVT	27.89	2.80	1507.79448
EATESFGPGTIEYRPVA	28.09	3.23	1823.88371
VAPEEHPTLLTEAPLN	28.26	2.85	1730.89592
DFENEMATAASSSSLEK	28.29	3.14	1816.78923
AGFAGDDAPRAVF	28.30	1.68	1293.62236
AAFPPDVAGNV DYK	28.43	2.92	1463.71709
DSGDGVTHNVPIYEGY	28.49	3.01	1722.76152
GIITNWDDMEK	28.57	2.02	1321.60881
ASGRITGIVLDSGDGVTH NVPIYEG	28.63	3.83	2515.23868
GTLDDYVEGLR	29.43	2.12	1237.60771

Sequence	Retention Time (Min)	XCorr Score	Precursor Ion Mass (Da)
AASSSSLEKSYELPDGQVI	29.51	2.42	2082.01958
T			
ADLFDPIK	29.85	2.13	1017.55986
SGLTDVIHM	30.19	1.58	972.48082
AIHFPADFTPEVH	30.27	1.98	1480.71941
HGIVPIVEPEILPDGDHDLK	30.27	4.34	2193.15378
VDIQVVGDDLTVTNP	30.38	3.21	1712.90605
DEHAQTGTATVAAGIPAG	30.49	4.75	1852.88518
W			
GDLGIEIPTEKV	30.49	2.24	1270.68962
YELPDGQVITIGNER	30.67	3.54	1703.86296
LQGLIDAHF	30.70	2.06	1013.54009
LITAIGEVDNHPVIGDR	30.93	4.06	1918.03562
TAAVDIQVVGDDLTVTNP	30.98	5.29	1956.02825
VIDQDASGFIEVEELK	31.30	4.17	1791.89946
VASGDSAAAGDSL FVANH	31.31	3.68	1922.88970
AY			
IVLDSGDGVTHNVPIYEGY	31.39	1.68	2047.99468
SYELPDGQVITIGNER	32.01	2.65	1790.88994
DVIEDPVEIIDNER	32.89	2.82	1655.81011
VVDTGDPIRIPVGPETL	33.60	4.32	1777.97038
FQPSFIGMESAGIHE	33.90	2.98	1649.76396
SSGEQISEEEIDELLK	34.19	4.03	1805.86614
KFSDEEEFPDLSL	35.27	2.89	1555.71453
TAPTPSLEPGNGTQAAPL	36.42	1.44	1721.87163
PEALERWPIDL	36.49	3.89	1338.70745
AISEELDNLNDMTSI	36.54	2.15	1735.80522
VAAPPPLDLSTVKVEF	36.89	2.69	1682.93596
IVNGEEAVPGSWPWQVSL	38.37	2.03	2644.29961
QDKTGF			
KGVEIVAINDPFIDL DY	38.55	3.44	1920.99578
AGYDPKIIIIGMDVAASEFY	39.61	2.68	2060.00786
TGDWFNILEHY	39.64	2.43	1394.63811
VIDQDASGFIEVEELKLF	40.46	1.71	2052.05766
SVADLVESILK	41.02	2.69	1173.67034
IVAINDPFIDL DYmVY	41.19	1.95	1916.93572
PEILPDGDHDLK	42.76	3.31	1348.67173
IGMESAGIHE	43.64	2.11	1043.48393
LVLCDNRIQ	44.92	2.01	1073.57720
VVISAPSADAPMF	45.43	1.04	1304.65557
DSGDGVTHNVPIYEG	45.53	2.38	1559.69829
IQLVEEELDR	46.17	2.01	1243.65337
VAPEEHPTL	48.03	2.44	992.50548
IALNDHF	49.25	1.39	829.42174

Sequence	Retention Time (Min)	XCorr Score	Precursor Ion Mass (Da)
PVLELDGK	56.01	1.00	870.49291
LVPELQAT	58.23	1.31	870.49285
QYLVTGSVDGF	58.39	1.28	1185.57866

Table A3. Peptides validated using the modified criteria in fraction 2 GF Separation 4. Scores, SPI (%), and spectrum intensity were calculated by the Spectrum Mill MS Workbench Program using the Agilent MassHunter software. Peptide sequences are reported using the single letter amino acid codes.

Sequence	RT (Min)	MH+ (Da)	pI [†]	Score	SPI (%)	Spectrum Intensity
TIE	2.72	362.192	3.351	4.66	58.4	1.89E+06
MVE	3.07	378.169	3.351	5.56	64.5	1.40E+06
AITD	3.10	419.214	3.135	3.62	62.0	1.02E+06
IVE	3.47	360.213	3.351	5.66	65.9	1.11E+07
EGSI	3.58	405.198	3.351	3.27	55.5	2.26E+06
HIIE	3.60	511.288	5.171	4.41	60.8	3.69E+06
SEI	3.60	348.176	3.351	5.28	62.1	3.09E+06
IVD	4.02	346.197	3.135	3.89	65.1	2.77E+06
IVE	4.47	360.213	3.351	5.08	71.9	4.83E+06
TTIE	4.47	463.240	3.351	3.64	57.3	1.69E+06
ETI	5.02	362.192	3.351	3.13	50.6	6.99E+06
IID	6.12	360.213	3.135	3.46	68.5	1.04E+06
VEM	6.15	378.169	3.351	5.10	70.5	1.95E+06
IAASP	6.48	458.261	5.974	5.37	72.6	6.40E+05
VAID	9.70	417.234	3.135	3.22	52.7	1.39E+06
PIVE	9.93	457.266	3.351	6.12	73.4	8.00E+06
IEM	10.80	392.185	3.351	4.07	62.0	4.21E+06
IDV	11.13	346.197	3.135	4.68	71.0	1.32E+06
VDI	11.70	346.197	3.135	4.40	58.2	1.44E+07
ISDI	11.75	447.245	3.135	6.32	67.5	1.26E+06
VEI	11.82	360.213	3.351	4.92	66.0	1.50E+07
TIIDQ	12.70	589.319	3.135	8.54	81.5	1.17E+06
IDSI	12.93	447.245	3.135	6.38	71.9	9.46E+05
NVPAM	13.43	531.260	5.974	6.70	80.5	4.77E+06
EVF	13.50	394.197	3.351	3.79	54.5	2.03E+06
MENF	13.87	540.212	3.351	3.39	61.7	1.46E+06
VEF	14.40	394.197	3.351	3.54	58.2	2.60E+06
YPIE	15.08	521.261	3.351	7.51	71.6	1.74E+06
VVDI	16.28	445.266	3.135	3.71	52.0	1.74E+06
IDI	17.40	360.213	3.135	3.52	72.0	2.07E+07
IVI	18.38	344.254	5.974	4.24	74.8	3.37E+06

Sequence	RT (Min)	MH+ (Da)	pI [†]	Score	SPI (%)	Spectrum Intensity
IEI	18.77	374.229	3.351	4.36	67.8	7.95E+06
IEF	19.68	408.213	3.351	4.73	63.1	1.90E+06
IGDGI	20.18	474.256	3.135	7.91	69.0	1.23E+06
IDF	20.22	394.197	3.135	6.68	72.7	2.10E+06
WDAPD	21.67	603.241	2.957	5.42	82.1	2.23E+06
IVEV	22.07	459.281	3.351	3.81	68.7	3.63E+06
IPEI	23.42	471.281	3.351	5.79	74.2	2.31E+06
VGVDGF	24.05	593.293	3.135	16.94	92.1	4.52E+06
IVW	24.58	417.250	5.974	3.76	61.3	2.26E+06
VGVP	24.80	484.313	5.974	6.59	77.0	5.60E+06
VGVDGF	25.13	593.293	3.135	14.83	84.0	2.15E+06
VDII	25.67	459.281	3.135	3.32	65.9	2.01E+06
INDPF	25.68	605.293	3.135	6.09	79.4	4.03E+06
IPVQI	27.05	569.366	5.974	6.44	70.6	2.04E+06
IDAGF	27.35	522.256	3.135	6.68	67.7	2.90E+06
VIPEL	28.25	570.350	3.351	7.69	86.5	4.59E+06

[†] Calculated using tool available at www.isoelectric.org

Table A4. Peptides validated using the modified criteria in fraction 3 from GF Separation 4. Scores, SPI (%), and spectrum intensity were calculated by the Spectrum Mill MS Workbench Program using the Agilent MassHunter software. Peptide sequences are reported using the single letter amino acid codes.

Sequence	RT (Min)	MH+ (Da)	pI [†]	Score	SPI (%)	Spectrum Intensity
VIT	3.27	332.218	5.974	6.25	75.2	5.40E+06
AAI	3.48	274.176	5.974	3.15	70.3	5.19E+05
IVA	3.63	302.207	5.974	3.15	72.7	1.04E+06
TGI	3.90	290.171	5.974	4.38	71.8	1.34E+05
VIT	4.43	332.218	5.974	5.58	55.8	7.89E+06
IVSA	4.62	389.240	5.974	4.70	58.3	2.71E+06
AVTV	4.67	389.240	5.974	4.45	58.6	2.42E+06
PGIA	5.00	357.213	5.974	6.16	76.1	7.21E+06
ATAI	5.13	375.224	5.974	4.03	56.8	1.32E+06
VSV	5.47	304.187	5.974	3.10	58.0	1.80E+06
IYE	6.12	424.208	3.351	4.18	67.2	3.22E+06
AIAPS	6.48	458.261	5.974	5.45	70.1	1.24E+06
IHALT	6.73	554.330	7.792	3.07	62.7	2.19E+06
TPI	6.77	330.202	5.974	3.62	55.2	2.31E+06

Sequence	RT (Min)	MH+ (Da)	pI†	Score	SPI (%)	Spectrum Intensity
IGIGA	7.23	430.266	5.974	3.55	55.9	1.76E+07
VVVA	7.27	387.260	5.974	4.32	72.1	1.14E+06
VVV	7.52	316.223	5.974	3.83	70.7	5.71E+06
VSTI	7.85	419.250	5.974	4.15	51.6	2.24E+06
IGGI	8.27	359.229	5.974	4.47	68.8	1.89E+06
VGGI	8.28	345.213	5.974	5.79	67.0	1.01E+06
SVI	8.30	318.202	5.974	4.75	72.1	2.89E+06
AVI	8.37	302.207	5.974	3.70	58.1	1.05E+06
TVI	8.40	332.218	5.974	4.63	69.0	2.90E+06
VIVS	8.52	417.271	5.974	6.23	70.2	2.30E+06
IANM	8.53	448.222	5.974	6.16	64.5	1.06E+06
SIM	8.78	350.174	5.974	4.35	66.8	4.05E+06
ITGMA	9.22	492.249	5.974	7.26	67.0	2.52E+06
VVIS	9.32	417.271	5.974	3.09	56.7	1.27E+06
IAM	9.40	334.180	5.974	3.22	62.0	2.65E+06
VTGI	9.72	389.240	5.974	3.85	52.5	1.70E+06
ITDY	9.82	511.240	3.135	7.77	73.9	1.08E+06
IGV	10.02	288.192	5.974	6.06	77.7	4.16E+05
IIPQ	10.13	470.297	5.974	4.24	59.6	1.20E+06
VAI	10.13	302.207	5.974	3.06	62.3	1.53E+06
VAGI	10.65	359.229	5.974	3.16	52.0	2.85E+06
ISV	10.77	318.202	5.974	4.62	75.6	1.28E+07
IGAGM	11.10	448.222	5.974	6.07	67.1	1.60E+06
VIM	11.33	362.211	5.974	3.68	60.9	2.27E+06
TGAIM	11.38	492.249	5.974	5.37	59.1	2.07E+06
IVV	12.27	330.239	5.974	3.61	63.2	2.30E+06
INSI	12.45	446.261	5.974	3.86	58.8	2.14E+06
TVF	12.47	366.202	5.974	3.92	72.5	3.49E+06
VIVSA	12.97	488.308	5.974	7.00	72.9	5.87E+06
ISGI	13.15	389.240	5.974	5.04	63.7	2.58E+06
AII	13.95	316.223	5.974	3.69	69.4	5.40E+06
INI	14.08	359.229	5.974	3.87	60.5	1.43E+06
SII	14.13	332.218	5.974	3.84	76.7	1.32E+07
VISM	14.45	449.243	5.974	5.56	58.6	3.64E+06
VGPI	14.58	385.245	5.974	5.91	74.7	2.07E+06
VVISA	14.60	488.308	5.974	7.07	77.3	5.91E+06
IGVV	15.17	387.260	5.974	5.54	65.0	1.02E+06
TII	15.27	346.234	5.974	4.46	71.0	4.76E+06
INI	15.28	359.229	5.974	4.37	66.1	4.46E+06
VIIA	15.28	415.291	5.974	4.89	67.9	1.53E+06
STGVF	15.38	510.256	5.974	7.69	74.8	1.76E+06
VGIP	15.40	385.245	5.974	3.82	64.8	3.38E+06
SII	15.62	332.218	5.974	3.84	75.7	5.57E+06
IAI	15.93	316.223	5.974	4.04	83.8	2.71E+07

Sequence	RT (Min)	MH+ (Da)	pI†	Score	SPI (%)	Spectrum Intensity
VVGI	16.08	387.260	5.974	5.81	55.7	4.09E+06
IIVA	16.43	415.291	5.974	3.46	71.5	2.89E+06
ITI	16.72	346.234	5.974	3.22	69.3	5.07E+06
MVI	16.85	362.211	5.974	4.73	62.9	3.73E+06
ITI	17.72	346.234	5.974	3.02	63.9	1.49E+07
IIM	18.00	376.226	5.974	3.75	75.8	2.19E+06
IDI	18.35	360.213	3.135	3.26	68.3	5.45E+06
APII	18.37	413.276	5.974	4.11	64.5	1.31E+06
IVI	18.42	344.254	5.974	4.34	76.0	9.00E+07
EIF	19.63	408.213	3.351	4.65	55.6	3.35E+06
IEF	19.67	408.213	3.351	4.93	63.5	2.46E+06
IDF	20.22	394.197	3.135	4.00	61.8	9.65E+06
GGVPLY	21.67	605.329	5.915	6.60	83.7	7.34E+06
VIGI	21.68	401.276	5.974	4.93	51.4	1.63E+07
III	21.97	358.270	5.974	4.01	79.5	6.28E+06
IGVL	22.53	401.276	5.974	7.01	74.5	6.24E+06
IVPI	23.40	441.307	5.974	4.38	58.1	1.19E+07
III	23.45	358.270	5.974	3.83	75.7	1.67E+07
VGII	23.83	401.276	5.974	3.06	51.9	4.30E+06
III	24.32	358.270	5.974	3.02	81.5	1.11E+07
VGVP	24.82	484.313	5.974	5.74	75.2	6.61E+06
IGGISI	25.15	559.345	5.974	7.97	77.4	9.14E+06
ITIIG	26.02	516.339	5.974	10.21	86.5	2.05E+06
FQPSFIG	29.23	795.404	5.974	12.69	87.9	2.28E+06

† Calculated using tool available at www.isoelectric.org

Table A5. Peptides validated using the modified criteria in fraction 6 from GF Separation 4. Scores, SPI (%), and spectrum intensity were calculated by the Spectrum Mill MS Workbench Program using the Agilent MassHunter software. Peptide sequences are reported using the single letter amino acid codes.

Sequence	RT (Min)	MH+ (Da)	pI†	Score	SPI (%)	Spectrum Intensity
PAY	3.70	350.171	5.915	4.68	71.2	2.53E+06
IYT	6.28	396.213	5.915	3.28	66.1	2.40E+06
PVY	7.37	378.202	5.915	4.11	73.3	7.79E+06
IGY	9.83	352.187	5.915	4.93	76.6	8.25E+06
IAY	10.23	366.202	5.915	4.76	76.5	1.10E+07
VYM	10.55	412.190	5.915	6.46	61.9	3.53E+06
ITY	10.58	396.213	5.915	3.78	63.7	2.66E+06
IVY	12.37	394.234	5.915	4.13	71.5	4.39E+06
FGY	14.57	386.171	5.915	4.13	64.2	6.87E+06
VSW	15.53	391.198	5.974	4.17	50.2	3.89E+07

IYY	15.93	408.249	5.915	4.68	74.0	3.02E+07
IYY	17.18	408.249	5.915	3.09	67.5	3.36E+06
VVW	20.18	403.234	5.974	3.13	57.5	5.27E+06
VFF	28.08	412.223	5.974	4.70	68.3	1.17E+07
WVNI	28.10	531.293	5.974	5.31	54.3	2.96E+06
VAVSW	30.13	561.303	5.974	8.50	77.3	4.96E+06
IFF	32.63	426.239	5.974	4.31	69.8	5.47E+06

† Calculated using tool available at www.isoelectric.org

Table A6. Peptides validated using the modified criteria in fraction 7 from GF Separation 4. Scores, SPI (%), and spectrum intensity were calculated by the Spectrum Mill MS Workbench Program using the Agilent MassHunter software. Peptide sequences are reported using the single letter amino acid codes.

Sequence	RT (Min)	MH+ (Da)	pI†	Score	SPI (%)	Spectrum Intensity
GYSF	16.42	473.203	5.915	5.00	67.3	5.30E+06
PVW	18.17	401.218	5.974	3.80	71.0	7.99E+06

† Calculated using tool available at www.isoelectric.org

Table A7. Peptides validated using the modified criteria in fraction 9 from GF Separation 4. Scores, SPI (%), and spectrum intensity were calculated by the Spectrum Mill MS Workbench Program using the Agilent MassHunter software. Peptide sequences are reported using the single letter amino acid codes.

Sequence	RT (Min)	MH+ (Da)	pI†	Score	SPI (%)	Spectrum Intensity
TWPW	37.48	589.277	5.974	7.76	86.2	6.05E+05

† Calculated using tool available at www.isoelectric.org

Table A8. Peptides validated using the modified criteria in fraction 6 from SAX Separation 5. Scores, SPI (%), and spectrum intensity were calculated by the Spectrum Mill MS Workbench Program using the Agilent MassHunter software. Peptide sequences are reported using the single letter amino acid codes.

Sequence	RT (Min)	MH+ (Da)	pI†	Score	SPI (%)	Spectrum Intensity
IEE	2.73	390.187	3.183	5.57	75.4	4.88E+06
SIED	3.12	463.204	3.048	3.34	53.3	2.02E+06
YVEG	3.40	467.214	3.351	5.60	58.1	1.40E+06
DID	3.53	362.156	2.957	3.66	67.6	2.32E+06
IAEE	3.83	461.224	3.183	6.00	69.7	1.09E+06
IGEE	4.07	447.209	3.183	8.85	77.0	2.06E+06
DESV	4.17	449.188	3.048	6.13	64.3	1.27E+06
VAPEE	4.17	544.261	3.183	7.97	70.5	1.00E+06
IGEE	4.40	447.209	3.183	3.88	65.7	2.34E+06
GEAID	5.13	504.230	3.048	6.63	62.6	1.54E+06
E EI	5.77	390.187	3.183	3.26	64.5	8.16E+06
IYE	6.18	424.208	3.351	6.10	73.0	3.99E+06
DEI	6.63	376.171	3.048	4.71	61.9	2.16E+06
ESGEI	6.90	534.241	3.183	6.35	71.3	9.91E+05
YIEN	7.12	538.251	3.351	7.56	56.8	4.66E+05
ADIET	7.52	548.256	3.048	9.54	78.7	2.40E+06
DADI	7.98	433.193	2.957	3.08	53.7	1.76E+06
GGVDY	8.43	510.219	3.135	5.23	65.7	1.21E+06
EEF	8.45	424.171	3.183	5.62	64.7	2.27E+06
EAGITE	8.93	619.293	3.183	5.95	73.1	7.43E+05
IIEE	8.98	503.271	3.183	5.82	64.1	1.27E+06
ITDY	9.83	511.240	3.135	7.77	73.3	3.07E+06
YDNSI	10.12	611.267	3.135	9.05	68.8	7.91E+05
EAEF	10.33	495.209	3.183	5.71	67.9	1.50E+06
EIPDGQ	10.45	658.304	3.048	10.97	87.0	8.18E+05
QIPDE	10.50	601.283	3.048	7.44	77.8	2.12E+06
DGEF	10.65	467.177	3.048	3.83	58.6	1.78E+06
WVNE	10.73	547.251	3.351	7.17	59.8	1.07E+06
IEY	11.17	424.208	3.351	5.94	72.5	1.07E+06
VEIE	11.40	489.255	3.183	6.95	60.5	2.54E+06
YNEI	11.50	538.251	3.351	3.14	50.3	2.16E+06
WGDAGAT	11.97	677.289	3.135	8.08	76.1	9.13E+05
MEDGI	12.80	564.233	3.048	7.01	64.2	2.33E+06
IDDI	14.30	475.240	2.957	6.06	74.6	3.11E+06
VIDPEA	14.73	643.330	3.048	12.38	87.6	1.85E+06
VIVDE	14.73	574.308	3.048	9.96	76.9	2.73E+05
YPIE	15.08	521.261	3.351	7.54	75.7	4.59E+06
IEGDI	15.13	546.277	3.048	9.46	76.0	1.10E+06

Sequence	RT (Min)	MH+ (Da)	pI†	Score	SPI (%)	Spectrum Intensity
IDVE	15.37	475.240	3.048	3.72	61.2	1.19E+06
IEEI	15.63	503.271	3.183	8.23	77.0	2.04E+06
YEI	15.67	424.208	3.351	4.96	57.1	1.19E+06
IGIDE	16.32	546.277	3.048	7.73	64.2	4.80E+05
IDDGI	16.38	532.261	2.957	7.79	67.1	2.67E+06
IQVVGDD	16.93	745.373	2.957	13.10	88.9	1.25E+06
TDYI	16.95	511.240	3.135	4.37	55.2	4.70E+06
VVDGDI	17.20	617.314	2.957	7.03	70.1	5.02E+05
YEF	18.15	458.192	3.351	4.48	55.8	1.83E+06
IETDI	18.80	590.303	3.048	7.02	71.6	7.50E+05
ITDDI	18.87	576.288	2.957	8.12	79.2	8.57E+05
VPIYE	18.87	620.329	3.351	5.36	70.0	3.78E+05
EEII	18.88	503.271	3.183	4.32	62.7	5.59E+06
GDYPLE	21.08	693.309	3.048	9.45	81.1	8.12E+04
EYIPDG	21.18	693.309	3.048	7.83	74.4	6.10E+05
EYLPDGQ	21.27	821.368	3.048	11.99	89.8	7.39E+05
AAAPLYE	21.60	734.372	3.351	11.70	89.8	6.21E+05
IEIPTE	24.95	701.372	3.183	10.51	90.3	8.14E+05
EYVF	25.50	557.261	3.351	6.01	70.7	1.24E+06
ITDYL	25.67	624.324	3.135	10.10	80.7	1.62E+06
LVPELDG	26.03	742.398	3.048	11.21	89.6	9.80E+05
IYEF	31.17	571.276	3.351	4.75	58.8	1.48E+06
LGEVWA	31.33	674.351	3.351	10.65	76.7	2.19E+06
EGIVW	31.90	603.314	3.351	7.21	75.9	7.27E+06
VLDPEATGF	32.10	948.467	3.048	10.14	85.3	6.07E+05
TWPW	37.33	589.277	5.974	7.78	86.9	6.23E+06

† Calculated using tool available at www.isoelectric.org

Table A9. Peptides validated using the modified criteria in fraction 7 from SAX Separation 5. Scores, SPI (%), and spectrum intensity were calculated by the Spectrum Mill MS Workbench Program using the Agilent MassHunter software. Peptide sequences are reported using the single letter amino acid codes.

Sequence	RT (Min)	MH+ (Da)	pI†	Score	SPI (%)	Spectrum Intensity
IEE	2.60	390.187	3.183	4.76	65.8	3.01E+06
IEEE	2.97	519.230	3.084	6.17	73.2	2.57E+05
SIED	3.02	463.204	3.048	3.78	61.9	1.35E+06
IEEE	3.48	519.230	3.084	9.56	81.4	6.30E+05
FENE	3.73	538.214	3.183	5.03	60.7	1.49E+06
QYTEE	3.92	669.273	3.183	7.79	73.5	3.11E+05
IDDDA	4.35	548.220	2.852	7.91	75.1	6.70E+05

Sequence	RT (Min)	MH+ (Da)	pI†	Score	SPI (%)	Spectrum Intensity
DIDD	5.10	477.183	2.852	4.63	66.6	2.78E+06
FEE	5.58	424.171	3.183	3.19	52.8	1.65E+06
EIEE	5.67	519.230	3.084	5.58	62.2	1.15E+06
E EI	5.70	390.187	3.183	4.89	60.3	2.13E+06
AGDEI	6.73	504.230	3.048	4.86	63.9	2.21E+06
MPSDED	6.82	693.240	2.911	7.30	71.4	2.88E+05
VEEE	7.72	505.214	3.084	6.36	73.6	2.88E+06
FTADE	8.30	582.241	3.048	8.85	74.1	4.20E+05
EEF	8.40	424.171	3.183	3.41	52.7	2.90E+06
EEEI	8.47	519.230	3.084	5.12	65.6	2.06E+06
DEF	9.95	410.156	3.048	3.55	56.7	1.54E+06
IDVE	10.35	475.240	3.048	6.78	77.5	1.47E+06
AGLPDE	10.45	601.283	3.048	6.82	71.8	7.82E+05
IEYE	10.95	553.250	3.183	7.65	66.4	3.48E+05
DSYVGD	12.07	655.257	2.957	6.57	71.4	3.72E+05
ETEDGF	12.33	697.268	2.984	7.60	68.5	6.19E+05
YDDSL	12.58	612.251	2.957	8.70	67.4	3.15E+05
DYY	13.17	460.171	3.135	4.28	63.9	2.11E+06
MEIDD	13.72	622.239	2.911	7.77	79.3	5.72E+05
YDDSL	13.73	612.251	2.957	10.55	75.5	1.98E+06
VEEEI	14.05	618.298	3.084	11.70	83.9	3.97E+05
IDDI	14.27	475.240	2.957	4.70	70.9	2.26E+06
YDNEFG	14.42	744.284	3.048	8.61	70.7	3.21E+05
QQVDDL	14.55	717.341	2.957	4.56	71.8	3.14E+05
LVDPEA	14.70	643.330	3.048	11.67	86.6	5.92E+05
EVEDI	14.82	604.282	2.984	9.49	76.4	2.03E+06
IDTEY	14.87	640.282	3.048	7.61	59.2	2.09E+06
YDDSL	15.23	612.251	2.957	6.68	70.8	3.93E+05
YDNEF	15.33	687.262	3.048	9.35	70.5	5.37E+05
YEP EM	16.27	668.260	3.183	8.18	85.3	8.61E+05
IDDVA	16.35	532.261	2.957	7.62	66.0	3.54E+05
YGESDL	16.40	683.288	3.048	13.23	80.8	5.16E+05
IEEEI	17.92	632.314	3.084	12.71	88.6	3.32E+05
IEDDI	17.95	604.282	2.911	7.77	72.6	7.20E+05
YEF	18.10	458.192	3.351	3.86	62.8	1.32E+06
TIDDL	18.78	576.288	2.957	8.12	82.2	6.22E+05
FDDI	19.08	509.224	2.957	5.67	54.4	1.60E+06
EFDI	19.68	523.240	3.048	6.09	70.1	6.19E+05
IEDDI	19.77	604.282	2.911	7.92	71.4	1.48E+06
EYIPDG	21.17	693.309	3.048	9.81	83.8	9.16E+05
EYLPDGQ	21.23	821.368	3.048	13.52	90.5	4.07E+05
PTCPDA	21.65	603.244	3.135	5.16	81.7	1.00E+07
IQEDDL	21.70	732.341	2.911	8.22	75.2	5.95E+05
DVWDTA	22.83	706.304	2.957	10.37	70.2	9.32E+05

Sequence	RT (Min)	MH+ (Da)	pI†	Score	SPI (%)	Spectrum Intensity
DIDDI	23.93	590.267	2.852	8.72	84.0	5.81E+06
SSGELW	25.45	678.309	3.351	8.39	75.5	6.51E+05
LDAVWE	30.13	732.356	3.048	10.33	79.2	2.16E+06
LEFDY	30.40	686.303	3.048	10.06	82.4	4.63E+05
IDTEYF	30.87	787.351	3.048	8.36	70.0	1.46E+06
ADLVW	31.88	603.314	3.135	7.08	72.7	6.20E+05
TWPW	37.30	589.277	5.974	6.40	80.3	5.92E+05

† Calculated using tool available at www.isoelectric.org

Table A10. Primary sequences of identified progenitor proteins from the low-alkali solubilized salmon muscle, selected from the UniProtKB protein database of the *Salmo salar* species (Taxon ID: 8030) representing the curated protein library for *in silico* digestions.

Accession No.: A0A1S3QJ13 – Myosin heavy chain, fast skeletal muscle-like
Gene: LOC106593382

MSTDAEMQIYGKAAIYLRKPERERIEAQTAPFDSKNSCYVTDKEELYLKGLVTARAD
GKCTVTVTKPDGKKEEGKEFKDADIYEMNPPKYDKIEDMAMMTYLNEASVLYNLKE
RYAAWMIYTYSGLFCATVNPYKWLVPYDMEVNNAYRGKKRMEAPPHIFSVSDNAFQ
FMLIDKENQSVLITGESGAGKTVNTRVIQYFATIAVSGGEKKKEADPGKMQGSLED
QIIAANPLLEAYGNAKTVRNDNSSRFGKFIRIHFQGGKLAADIETYLLKSRVTFQLP
DERGYHIFFQMMTNHKPEIVEMALITNPNYDFPMCSQGQITVASINDNEELDATDDAI
TILGFSNEEKQAIYKLTGAVLHHGNLKFQKQREEQAEPDGTVEVADKIGYLLGLNSAE
MLKALCYPRVKVGNVYVTKGQTVPQVNNNSVSALAKSIYERMFLWMVIRINEMLDTKQ
PRQFYIGVLDIAGFEIFDYSMEQLCINFTNEKLQQFFNHTMFVLEQEEYKKEGIVWA
FIDFGMDLAACIELIEKPLGIFSILEEECMFPKSSDTTFKDKLYSQHLGKTKAFEKPKPA
KGKAEAHFSLVHYAGTVDYNITGWLEKNKDPLNDSVCQLYGKSSVKLLAALYPAAPP
EDTTKKGKGGKSMQTVSSQFRENHLKLMTNLRSTHPPHFVRCLIPNESKTPGLME
NFLVIHQLRNGVLEGIKIRKGFPSRIYADFKQRYKVLNASVIPEGQFMDNKKASEK
LLGSIDVNHEDYKFGHTKVFFKAGLLGVLEEMRDEKLATLVGMVQALSARGFLMRREF
SKMMERRESIYSIQYNIRSFMNVKTWPWMKLYFKIKPLLQSAETEKELANMKENYEK
MTADLAKALSTKKQMEEKLVALTQEKNLALQVASEGESLNDAAERCEGLIKSKIQL
EAKLKETTERLEDEEEINAELTAKKRKLEDECSELKDDIDLELTLAKVEKEKHATENK
VKNLTEEMASMDSEVAKLTKEKKALQEAHQQTLDLQAEEDKVNTLTKAKTKLEQQ
VDDLEGSLEQEKKLRMDLERSKRKLEGLDKLAQESIMDLENDKQQADEKIKKKEFET
TQLLSKIEDEQSLGAQLQKKIKELQARIEELEEIEAERAARAKVEKQRADLSRELEEI
SERLEEAGGATAAQIEMNKKREAEFQKLRRDLEESTLQHEATAAALRKKQADSVLAE
GEQIDNLQRVKQKLEKEKSEYKMEIDDLSSNMEAVAKAKGNLEKMCRTLEDQLSEL
KTKNDENVRQVNDISGQRARLLTENGEFGRQLEEKEALVSQLTRGKQAFQTQQVEEL
KRATEEEVKAKNALAHSVQSARHDCDLLREQFEEEQEAKAELQRGMSKANSEVAQ
WRTKYETDAIQRTEELEEAKKKLAQRLQEAETIEATNSKSCASLEKTKQRLQGEVED
LMVDVERANALANLDKKQRNFDKVLAEWKQKYEEGQAELEGAQKEARSMSTELF
KMKNSEYEEALDHELETLKRENKNLQQEISDLTEQIGETGKSIHELEKAKKTVETEKSEI
QTALEEAEGTLEHEESKILRVQLELNQIKGEVDRKIAEKDEEMEQIKRNSQRVVDMS
QSTLDSEVRSRNDALRVKKKMEGDLNEMEIQLSHSNRQAAEAQKQLRNQVGGQLKD
AQLHLDDAVRAAEDMKEQAAMVERRNGLMVAEIEELRVALEQTERGRKVAETELVD
ASERVGLLHSQNTSLLNTKKKLETDLVQVQGEVDDIVQEARNAEEKAKKAITDAAMM
SEELKKEQDTSSHLERMKKNLEVTVKDLQHRLEAENLAMKGGKKQLQKLESRVRE
LETEVEAEQRRGVDVAVKGVKRYERRVKELTYQTEEDKKNVNRLQDLVDKLMKVK
AYKRQAEAAEAANQHMSKFRKVQHELEEAERADIAETQVNKLRAKTRDSGKGKE
AAE

Accession No.: Q78BU2 – Actin alpha 1-1

Gene: LOC106564730

MCDDDETTALVCDNGSGLVKAGFAGDDAPRAVFPSIVGRPRHQGVMVGMGQKD
SYVGDQAQSKRGILTLKYPIEHGIITNWDDMEKIWHHTFYNELRVAPEEHPTLLTEA
PLNPKANREKMTQIMFETFNVPAMYVAIQAVLSLYASGRTTGIVLDSGDGVTHNVP
IYEGYALPHAIMRLDLGRDLTDYLMKILTERGYSFVTTAEREIVRDIKEKLCYVALD
FENEMATAASSSSLEKSYELPDGQVITIGNERFRCPETLFQPSFIGMESAGIHETAY
NSIMKCDIDIRKDLYANNVLSGGTTMYPGIADRMQKEITALAPSTMKIKIIPPERKY
SVWIGGSILASLSTFQAMWITKQEYDEAGPSIVHRKCF

Accession No.: Q91472 – Fast myotomal muscle tropomyosin (TM)

Gene: tpm1

MDAIKKKMQMLKLDKENALDRAEGAEGDKKAAEDKSKQLEDDLVALQKCLKGTE
DELDKYSESLKDAQEKLEVAEKTATDAEADVASLNRRRIQLVEEELDRAQERLATAL
TKLEEAEKAADESERGMKVIENRASKDEEKMELQDIQLKEAKHIAEEADRKYEEVA
RKLVIIESDLERTEERAELSEGKCSELEEEELKTVTNLKSLEAQAEEKYSQKEDKYEE
EIKVLTDKLKEAETRAEFAERSVAKLEKTIDDELELYAQKLKYKAISEELDNALND
MTSI

Accession No.: B5DGG5 – Creatine Kinase-4 (CK)

Gene: ckm3

MTKNCHNDYKMKFSDEEEFPDLSLHNNHMAKVLTKDMYKLRKSKSTPSGFTLDD
CTQTGVDNPGHPFIMTVGCVAGDEECYEVFKDMFDPIISDRHGGYKPTDKHKTDL
NFENLKGDDLDPAYVLSSRVRTGRSIKGYTLPPHNSRGERRMVEKLSIEALATLD
GEFKGKYYPNGMTDAEQDQLIADHFLFDKPVSPLLL SAGMARDWPDARGIWHN
DAKSFLVWVNEEDHLRVISMEKGGNMKEVFRFCVGLQKIEAVFKKHNHGFMWN
EHLGYVLTCPSNLGTGLRGGVHVKLPKLSHAKFEEILTRLRLQKRGTTGGVDTASV
GGIFDISNADRLGSSEVQQVQMVVDGVKLMVEMEKKLEKGEAIDGMIPAQK

Accession No.: B5DGR3 – Glyceraldehyde-3-phosphate dehydrogenase (G3P Dh)

Gene: G3P

MVKVGVNGFGRIGRLVTRAAFHSHKKGVEIVAINDPFIDLDMVYMFKYDSTHGRFH
GEVKAEGGKLVIDGHKITVFHERDPANIKWGDAGATYVVESTGVFTTIEKASTHLK
GGAKRVVISAPSADAPMFVMGVNHEKYDNSLKVSNASCTTNCLAPLAKVIHDNY
HIIELMSTVHAVTATQKTVDGPSGLWRDGRGASQNIIPASTGAAKAVGKVIPEL
NGKITGMAFRVPTPNVSVDLTVRLEKPASYDAIKKVVKAAADGPMKILGYTEQQ
VVSSDFNGDTHSSIFDAGAGIALNDHFVVKLVTWYDNEFGYSNRVIDLMAHMATKE

Accession No.: B5DGM7 – Fructose-bisphosphate aldolase A (ALDOA)

Gene: N/A

MPHAFPFLLTPDQKKELSDIALKIVAKGKGILAADESTGSVAKRFQSINTENTEENRR
LYRQLLFTADDRAGPCIGGVIFFHETLYQKTDAGKTFPEHVKSRGWVVGKVDKGV
VPLAGTNGETTTQGLDGLYERCAQYKKDGCDFAKWRCVLKITSTTPSRLAIMENC
NVLARYASICQMHGIVPIVEPEILPDGDHDLKRTQYVTEKVLAAAMYKALSDDHVVYLE
GTLLKPNMVTAGHSCSHKYTHQEIAMATVTALRRTVPPAVPGVTFLSGGQSEEEA

SINLNMNQCPLHRPWALTFSYGRALQASALKAWGGKPGNGKAAQEEFIKRALA
NSLACQGKYVASGDSAAAGDSLFFVANHAY

Accession No.: B5DGM7 – Fructose-bisphosphate aldolase A (ALDOA)

Gene: N/A

MPHAFPFLTPDQKKELSDIALKIVAKGKGILAADESTGSAKRFQSINTENTEENRR
LYRQLLFTADDRAGPCIGGVIFFHETLYQKTDAGKTFPEHVKSARGWVVGIVDKGV
VPLAGTNGETTTQGLDGLYERCAQYKKGDCDFAKWRCVLKITSTTPSRILMENC
NVLARYASICQMHGIVPIVEPEILPDGDHDLKRTQYVTEKVLAAAMYKALSDHHVYLE
GTLLKPNMVTAGHSCSHKYTHQEIAMATVTALRRTVPPAVPGVTFLSGGQSEEEA
SINLNMNQCPLHRPWALTFSYGRALQASALKAWGGKPGNGKAAQEEFIKRALA
NSLACQGKYVASGDSAAAGDSLFFVANHAY

Accession No.: B5XH68 – Triosephosphate Isomerase (TPIS)

Gene: TPIS

MNGDKASLGELIKTLNSAKLDPNTEVVCGAPSIYLEFARAKLDPKIGVAAQNCYKV
KGAFTGEISPAMIKDVGVHVVILGHSERRWVFGTDELIGQKCAHALENGLGVIA
CIGEKLDEREAGITEKVINAQTKHFADNIKDWSKVVLAYEPVWAIGTGKTASPAQA
QDVHDKLRQWVKANVSEAVANSVRIIYGGSVTGGTCKELGGMKDVDGFLVGGAA
LKPEFVDIINAKQ

Accession No.: B5DGT9 – Phosphoglycerate Mutase (PGM)

Gene: LOC100194644

MTTAHKLIVIRHGESEWNQYNKFCGWFDADLSEKGLEEAKRGAKAIKDAGMKFDI
CHTSVLKRAVKTLWTIMEGTDQMWLVPVYRTWRLNERHYGGLTGLNKAETAEKHG
EEQVKIWRRSFDTPPPMEHDHAFHKIISESRRYKGLKPGELPTCESLKDTIARALP
YWNDVIAPEIKAGKNVIAAHGNSLRGIVKHLEGMSDAAIMELNLPTGIPIVYELDAN
LKPVKPMAFLGDAETVKKAMEAVAAQGKAKK

Accession No.: Q7ZZN0 – Myosin Regulatory Light Chain 2-2 (MLC-2)

Gene: mlc-2

MAPKKAKRRGAAAEGGSSNVFSMFEQSQIQEYKEAFTIIDQNRDGIISKDDLRLDVL
ASMGQLNVKNEELEAMVKEASGPINFTVFLTMFGEKLGADPEDVIVSAFKVLDPE
ATGFIKKEFLQELLTTQCDRFSAEEMKNLWAAFPPDVAGNVDYKQICYVITHGEEK
EE

Accession No.: B5DGQ7– Beta Enolase (ENO)

Gene: ENO3

MSITKIHAREILDSRGNPTVEVDLYTAKGRFRAAVPSGASTGVHEALELRDGDKSR
YLGKGTAVKAVDHVNKDIAAKLIEKFSVVDQEKIDHFMLELDGTENKSKFGANAILG
VSLAVCKAGAAEKGVPLYRHIADLAGHKDVILPCPAFNVINGGSHAGNKLAMQEF
MILPIGASNFHEAMRIGAEVYHNLKNVIKAKYKDATNVGDEGGFAPNILENNEALE
LLKTAIEKAGYPDKIIGMDVAASEFYKAGKYDLDFKSPDDPARYITGDQLGDLYKS
FIKGYPVQSIEDPFDQDDWAAWTKFTAANDIQVVGDDLTVTNPKRIQQAVEKKAC
NCLLLKVNQIGSVTESIKACKLAQSNGWGMVSHRSGETEDTFIADLVVGLCTGQI
KTGAPCRSERLAKYNQLMRIIEEELGAKAKFAGKDYRHPKIN

Table A11. Compound list generated using the Find by algorithm with Molecular Feature Extraction using Agilent MassHunter Software on hydrolysates generated from the extracted proteins from band 2. Compound identities were predicted using *in silico* digestions of the myosin heavy chain (Accession #: A0A1S3QJ13) and by comparison to common contaminants during mass spectrometry analysis.

RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	<i>In silico</i> Digestion		Contaminants	
			Non-Specific	Specific	PPG Adduct	PEG Adduct
2.55	142.1588	0.27	No Match	No Match	No Match	No Match
2.85	173.0786	2.99	No Match	No Match	No Match	Na+
2.86	302.1966	1.41	No Match	No Match	No Match	No Match
2.94	114.0550	1.91	No Match	No Match	No Match	No Match
2.99	166.0863	1.57	F	F	No Match	No Match
2.99	103.0543	1.01	No Match	No Match	No Match	No Match
2.99	120.0811	7.59	No Match	No Match	No Match	No Match
3.28	207.8932	1.87	No Match	No Match	No Match	No Match
3.62	252.0362	0.34	No Match	No Match	No Match	No Match
4.10	130.1589	0.68	No Match	No Match	No Match	No Match
4.24	245.0996	0.44	No Match	No Match	No Match	No Match
5.05	316.2118	1.67	No Match	No Match	No Match	No Match
5.42	246.1700	0.47	No Match	No Match	No Match	No Match
5.45	188.0709	2.37	No Match	No Match	No Match	No Match
5.80	333.0551	0.51	No Match	No Match	No Match	No Match
5.88	233.0789	0.43	No Match	No Match	No Match	K+
5.88	217.1052	8.09	No Match	No Match	No Match	Na+
5.94	157.0837	0.43	No Match	No Match	Na+	No Match
6.01	212.0530	0.45	No Match	No Match	No Match	No Match
6.52	231.0841	0.33	No Match	No Match	No Match	No Match
6.82	218.0235	0.34	No Match	No Match	No Match	No Match

248

RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	<i>In silico</i> Digestion		Contaminants	
			Non-Specific	Specific	PPG Adduct	PEG Adduct
6.82	262.1346	1.73	No Match	No Match	No Match	No Match
6.83	277.1045	1.52	No Match	No Match	No Match	K+
7.31	316.2123	11.23	No Match	No Match	No Match	No Match
8.27	272.1861	0.61	No Match	No Match	No Match	No Match
9.74	277.1045	0.70	No Match	No Match	No Match	K+
9.74	261.1314	6.14	No Match	No Match	No Match	Na+
9.74	239.1491	1.01	No Match	No Match	No Match	H+
10.42	335.1674	1.03	No Match	No Match	No Match	No Match
10.50	275.1104	0.50	No Match	No Match	No Match	No Match
13.50	195.0883	0.62	No Match	No Match	No Match	No Match
13.64	305.1575	3.84	No Match	No Match	No Match	No Match
13.64	283.1755	1.58	No Match	No Match	No Match	No Match
13.65	321.1305	0.62	No Match	No Match	No Match	K+
13.65	300.2012	0.43	No Match	No Match	No Match	No Match
13.75	316.1753	0.48	No Match	No Match	No Match	No Match
14.56	319.1361	0.36	No Match	No Match	No Match	No Match
16.14	303.1417	0.53	No Match	No Match	No Match	No Match
17.30	327.2014	1.16	No Match	No Match	No Match	H+
17.30	344.2277	1.46	No Match	No Match	No Match	No Match
17.30	349.1834	2.49	SSR	No Match	No Match	Na+
20.54	430.2437	8.87	No Match	No Match	No Match	No Match
20.64	371.2274	0.59	No Match	No Match	No Match	No Match
20.65	393.2094	1.89	No Match	No Match	No Match	No Match
20.65	388.2539	1.81	No Match	No Match	No Match	No Match
20.65	409.1829	0.49	No Match	No Match	No Match	K+
23.17	330.1910	0.65	No Match	No Match	No Match	No Match

RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	<i>In silico</i> Digestion		Contaminants	
			Non-Specific	Specific	PPG Adduct	PEG Adduct
23.51	481.2615	0.72	No Match	No Match	No Match	Na+
23.51	459.2794	0.57	QRR	No Match	No Match	No Match
23.73	437.2356	1.57	No Match	No Match	No Match	Na+
23.73	453.2092	0.35	No Match	No Match	No Match	No Match
23.73	432.2798	1.70	No Match	No Match	No Match	No Match
26.53	476.3061	1.63	No Match	No Match	No Match	No Match
26.53	497.2352	0.24	No Match	No Match	No Match	No Match
26.53	481.2616	1.23	No Match	No Match	No Match	Na+
29.07	525.2876	0.85	No Match	No Match	No Match	Na+
29.07	520.3320	1.22	No Match	No Match	No Match	No Match
31.40	564.3578	0.82	No Match	No Match	No Match	No Match
31.41	569.3136	0.68	No Match	No Match	No Match	Na+
33.57	608.3839	0.59	No Match	No Match	No Match	No Match
35.57	652.4103	0.33	No Match	No Match	No Match	No Match

Table A12. Compound list generated using the Find by algorithm with Molecular Feature Extraction using Agilent MassHunter Software on hydrolysates generated from the extracted proteins from band 4. Compound identities were predicted using *in silico* digestions of actin (Accession #: Q78BU2) and by comparison to common contaminants during mass spectrometry analysis.

RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	<i>In silico</i> Digestion		Contaminants	
			Non-Specific	Specific	PPG Adduct	PEG Adduct
2.55	142.1588	0.47	No Match	No Match	No Match	No Match
2.65	250.1440	0.27	No Match	No Match	No Match	No Match
2.81	222.1493	0.35	No Match	No Match	No Match	No Match

250

RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	<i>In silico</i> Digestion		Contaminants	
			Non-Specific	Specific	PPG Adduct	PEG Adduct
2.84	302.1965	1.40	No Match	No Match	No Match	No Match
2.84	173.0784	2.17	No Match	No Match	No Match	Na+
2.94	114.0550	0.76	No Match	No Match	No Match	No Match
2.99	166.0863	1.56	F	F	No Match	No Match
2.99	103.0544	0.96	No Match	No Match	No Match	No Match
2.99	120.0811	8.23	No Match	No Match	No Match	No Match
3.27	207.8932	1.40	No Match	No Match	No Match	No Match
3.38	187.0940	0.25	No Match	No Match	No Match	No Match
4.10	130.1588	0.66	No Match	No Match	No Match	No Match
4.19	177.0633	0.84	No Match	No Match	No Match	No Match
5.03	316.2117	1.67	No Match	No Match	No Match	No Match
5.42	246.1699	0.50	No Match	No Match	No Match	No Match
5.44	188.0709	2.57	No Match	No Match	No Match	No Match
5.80	333.0549	0.47	No Match	No Match	No Match	No Match
5.89	217.1050	5.58	No Match	No Match	No Match	Na+
6.83	218.0236	0.26	No Match	No Match	No Match	No Match
6.83	261.1314	12.58	No Match	No Match	No Match	Na+
6.83	292.0606	0.74	No Match	No Match	No Match	No Match
6.83	277.1044	1.26	No Match	No Match	No Match	No Match
6.84	263.1074	0.32	No Match	No Match	No Match	No Match
7.30	316.2123	11.19	No Match	No Match	No Match	No Match
7.52	227.1756	0.40	No Match	No Match	No Match	No Match
8.26	272.1860	0.59	No Match	No Match	No Match	No Match
9.73	239.1490	0.74	No Match	No Match	No Match	H+
9.73	277.1048	0.55	No Match	No Match	No Match	K+
9.73	261.1312	4.37	No Match	No Match	No Match	Na+

251

RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	<i>In silico</i> Digestion		Contaminants	
			Non-Specific	Specific	PPG Adduct	PEG Adduct
10.42	335.1675	0.82	No Match	No Match	No Match	No Match
10.52	275.1102	0.33	No Match	No Match	No Match	No Match
12.58	336.1915	0.52	AVF	No Match	No Match	No Match
13.39	239.0895	0.35	No Match	No Match	No Match	No Match
13.49	340.9522	0.73	No Match	No Match	No Match	No Match
13.65	300.2015	0.40	No Match	No Match	No Match	No Match
13.65	283.1755	1.31	No Match	No Match	No Match	No Match
13.65	305.1574	3.14	No Match	No Match	No Match	Na+
13.95	491.7816	0.75	No Match	No Match	No Match	No Match
14.24	416.2275	1.66	No Match	No Match	No Match	No Match
16.13	303.1415	0.40	No Match	No Match	No Match	No Match
17.30	349.1835	2.01	No Match	No Match	No Match	No Match
17.30	344.2277	1.17	No Match	No Match	No Match	No Match
17.30	327.2013	0.95	No Match	No Match	No Match	H+
20.55	430.2436	8.47	No Match	No Match	No Match	No Match
20.67	393.2093	1.53	No Match	No Match	No Match	No Match
20.67	388.2539	1.44	No Match	No Match	No Match	No Match
20.67	371.2279	0.47	No Match	No Match	No Match	H+
23.18	330.1910	0.62	No Match	No Match	No Match	No Match
23.52	481.2616	0.66	No Match	No Match	No Match	Na+
23.52	459.2793	0.53	No Match	No Match	No Match	No Match
23.74	437.2356	1.19	No Match	No Match	No Match	Na+
23.74	432.2797	1.58	No Match	No Match	No Match	No Match
26.53	476.3061	1.38	No Match	No Match	No Match	No Match
26.53	481.2613	1.05	No Match	No Match	No Match	No Match
27.63	502.3227	0.52	VITIG	No Match	No Match	No Match

RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	<i>In silico</i> Digestion		Contaminants	
			Non-Specific	Specific	PPG Adduct	PEG Adduct
29.07	525.2876	0.68	No Match	No Match	No Match	Na+
29.08	520.3321	1.00	No Match	No Match	No Match	No Match
31.41	569.3137	0.53	No Match	No Match	No Match	Na+
32.29	233.0787	0.66	No Match	No Match	No Match	K+

Table A13. Compound list generated using the Find by algorithm with Molecular Feature Extraction using Agilent MassHunter Software on hydrolysates generated from the extracted proteins from band 6. Compound identities were predicted using *in silico* digestions of glyceraldehyde-3-phosphate dehydrogenase (Accession #: B5DGR3) and fructose bisphosphate aldolase (Accession #: B5DGM7) and by comparison to common contaminants during mass spectrometry analysis.

RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	<i>In silico</i> Digestion		Contaminants	
			Non-Specific	Specific	PPG Adduct	PEG Adduct
2.85	173.0785	1.94	No Match	No Match	No Match	Na+
2.85	302.1963	1.09	No Match	No Match	No Match	No Match
2.94	114.0550	1.46	No Match	No Match	No Match	No Match
2.98	120.0809	2.26	No Match	No Match	No Match	No Match
3.40	187.0939	0.78	No Match	No Match	No Match	No Match
5.04	316.2119	1.22	No Match	No Match	No Match	No Match
5.87	217.1052	5.22	No Match	No Match	No Match	Na+
6.03	212.0531	1.09	No Match	No Match	No Match	No Match
6.84	218.0236	0.94	No Match	No Match	No Match	No Match
6.85	263.1072	1.15	No Match	No Match	No Match	No Match

RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	<i>In silico</i> Digestion		Contaminants	
			Non-Specific	Specific	PPG Adduct	PEG Adduct
6.85	277.1046	3.99	No Match	No Match	No Match	K+
6.85	261.1313	41.50	No Match	No Match	No Match	Na+
7.29	316.2123	7.94	No Match	No Match	No Match	No Match
9.74	261.1311	3.39	No Match	No Match	No Match	Na+
10.42	335.1678	4.48	No Match	No Match	No Match	No Match
10.43	176.0705	1.50	No Match	No Match	No Match	No Match
13.50	340.9524	2.02	No Match	No Match	No Match	No Match
13.65	283.1755	0.83	No Match	No Match	No Match	No Match
13.65	305.1573	1.92	No Match	No Match	No Match	Na+
16.15	303.1415	1.95	No Match	No Match	No Match	No Match
17.30	349.1834	1.23	No Match	No Match	No Match	Na+
20.56	430.2437	6.67	No Match	No Match	No Match	No Match
20.68	388.2538	0.90	No Match	No Match	No Match	No Match
21.84	363.1989	1.69	No Match	No Match	No Match	No Match
23.52	481.2615	1.70	No Match	No Match	No Match	Na+
23.52	459.2796	1.14	No Match	No Match	No Match	H+

Table A14. Compound list generated using the Find by algorithm with Molecular Feature Extraction using Agilent MassHunter Software for fraction 2 from GF Separation 4. The sequences of the detected ions were compared using the peptides identified by database searching and of predictions from *in silico* digestions with non-specific and specific hydrolysis of the curated protein library, outlined in Section 3.15. No Match indicates the ion was not identified by the indicated peptide identification methodology. Database search hits are reported to include the detected product ions, where / indicates the identification of the b-series ion, \ indicates the identification of the y-series ion, and | indicates the identification of both b- and y-series ions.

254

RT (Min)	Precursor		Database Searching	Peptide Identification Methodology	
	Ion Mass (Da)	Peak Vol. (%)		<i>In silico</i> Digestions	
				Specific	Non-Specific
2.18	362.1921	2.77	No match	TIE, TEL, LTE, ETL	IET, ELT, LTE, TLE, ETI, TIE, TEL, LET, ETL, EIT, ITE
2.32	229.1551	1.34	No match	PL	PL, LP, IP, PI
2.58	332.1819	2.81	No match	LEA, IEA, EIA, EAL, ALE, AEL	IEA, LEA, ELA, AEL, EAL, VDV, LAE, ALE, IAE, VVD, AEI, EAI, EIA, AIE
2.70	229.1549	1.84	No match	PL	PL, LP, IP, PI
2.72	358.1973	1.73	P E	PIE	PEI, IPE, PIE, ELP, PEL
2.82	362.1922	6.66	T E	TIE, TEL, LTE, ETL	IET, ELT, LTE, TLE, ETI, TIE, TEL, LET, ETL, EIT, ITE
2.84	389.2032	3.17	No match	LEGA, ELGA	AELG, LEGA, DLAA, ADLA, EAVA, ADIA, EQL, LEQ, QLE, LQE, ELQ, QIE, EQI, QEI, EIQ, VVDG, VDG, GEAI, LAAD,

RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	Peptide Identification Methodology		
			Database Searching	<i>In silico</i> Digestions	
				Specific	Non-Specific
2.93	332.1824	7.83	No match	LEA, IEA, EIA, EAL, ALE, AEL	DVGV, EAGI, DAAI, IQE, QEL, ELGA, DIAA, IGAE, VVGD IEA, LEA, ELA, AEL, EAL, VDV, LAE, ALE, IAE, VVD, AEI, EAI, EIA, AIE
3.00	334.1613	2.39	No match	SDL	VTD, TVD, SID, ESV, DLS, DIS, SEV, ISD, SDL, LDS, VDT, VES, LSD, SDI, DSL, VSE
3.01	346.1974	15.89	No match	VVE, VDL	EVV, VLD, IDV, LVD, DLV, DIV, VID, VVE, VDL, VDI, DVI, DVL, VEV
3.01	213.1602	2.67	No match	No Match	No Match
3.05	120.0804	14.73	No match	No Match	No Match
3.05	166.0865	2.98	No match	No Match	No Match
3.06	403.2187	3.14	V/I D\G	No Match	GVLD, VDAV, EALA, VIDG
3.07	378.1697	6.14	M V E	No Match	No Match
3.11	419.2136	4.45	A I T\D	SIEA, ITGE, ITDA, GTLE, GITE, EGTL	TEVA, VVDS, ITGE, LDAT, DAIT, TADL, TDAI, ASLE, IGET, EGTL, GTLE, VAET, AITD, ITDA, AELS, SLEA, AISE, ATLD, SIEA, SVVD, LEGT, EASI, TGEI, GITE, DTIA, AETV
3.24	332.2171	2.36	No match	No Match	LVT, SIL, TLV, LLS, SLL, ITV, VIT, LSL, VLT, IIS, LSI, LTV

255

RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	Peptide Identification Methodology		
			Database Searching	<i>In silico</i> Digestions	
				Specific	Non-Specific
3.28	332.1821	4.01	No match	LEA, IEA, EIA, EAL, ALE, AEL	IEA, LEA, ELA, AEL, EAL, VDV, LAE, ALE, IAE, VVD, AEI, EAI, EIA, AIE
3.35	318.1665	3.00	No match	No Match	No Match
3.35	231.1708	1.92	No match	No Match	VL, VI, LV, IV
3.41	362.1560	2.18	No match	DDL	DID, IDD, DDL, LDD, VED, DVE, EVD, DDI, DLD, EDV
3.43	375.1880	7.59	No match	No Match	No Match
3.45	360.2129	37.12	I/V E	VLE, LVE, IVE, IDL	EIV, VLE, ELV, LEV, LID, IVE, LDI, DLL, LDL, IDI, VIE, LVE, VEI, IDL, DII, IID, ILD
3.48	213.1601	3.11	No match	No Match	No Match
3.56	511.2875	10.59	H/I I E	No Match	IHEL, APLNP, HIIE
3.58	405.1984	6.67	E G S I	VASE, SLGE, SDIA, KEE, GSLE, GESL, EKE, EGSL	KEE, EASV, GSLE, EEK, GTEV, EKE, VASE, GESL, ESVA, LEGS, EGSL, SVAE, SEVA, ADLS, LSEG, TLDG, VDTA, SDIA, ALSD, TTPS, STTP, SLGE, VSEA, SEAV, GEIS, ITGD, LDGT
3.60	348.1768	8.58	S/E I	VTE, TDL, SLE, SIE, SEL, LSE, ISE	No Match
3.61	247.1296	6.45	No match	No Match	No Match
3.62	302.2075	1.94	I V A	No Match	No Match
3.64	211.0868	3.15	No match	No Match	No Match
3.72	375.1878	5.55	No match	No Match	No Match
4.01	132.1017	1.83	No match	L	I, L

256

RT (Min)	Precursor Ion Mass (Da) Peak Vol. (%)		Peptide Identification Methodology			
			Database Searching	<i>In silico</i> Digestions		
				Specific	Non-Specific	
4.01	318.1664	14.25	No match	AVE, DIA, DLA, EVA, GEL, IGE, LDA, LGE, VAE, VEA	LEG, ELG, VEA, ADI, LDA, DAI, EVA, DIA, EGI, DLA, ADL, EGL, VAE, LGE, EAV, ALD, IGE, DAL, IAD, AID, IEG, TTP, GEL, GEI, GLE, AEV, AVE	
4.06	346.1978	6.27	No match	No Match	No Match	
4.14	332.2179	2.38	No match	SIL	No Match	
4.35	389.2035	9.18	No match	No Match	No Match	
4.37	231.1709	1.99	No match	No Match	VL, VI, LV, IV	
4.45	463.2401	4.94	T T I/E	TTIE	LHHG, CRKG, TTIE	
4.47	360.2130	14.35	I/V E	VLE, LVE, IVE, IDL	EIV, VLE, ELV, LEV, LID, IVE, LDI, DLL, LDL, IDI, VIE, LVE, VEI, IDL, DII, IID, ILD	
4.51	201.0875	4.38	No match	No Match	No Match	
4.51	332.1818	33.61	No match	LEA, IEA, EIA, EAL, ALE, AEL	IEA, LEA, ELA, AEL, EAL, VDV, LAE, ALE, IAE, VVD, AEI, EAI, EIA, AIE	
4.56	314.1714	2.10	No match	No Match	PTP, TPP	
4.60	446.2244	1.46	No match	No Match	QIGE, ELAN, EINA, EQIG, LNEA, INAE, NAEL, DLQA, LGEQ, GEQI, EVAQ, DAIQ, LQGE, DAQL, VQEA, AENL, ENLA, ENAL, ALEN, TASP, ADLAG, NEAL, KSPD, QAVE	
4.60	419.2138	5.73	No match	SIEA, ITGE, ITDA, GTLE, GITE, EGTL	TEVA, VVDS, ITGE, LDAT, DAIT, TADL, TDAI, ASLE, IGET, EGTL, GTLE, VAET, AITD, ITDA, AELS, SLEA,	

RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	Peptide Identification Methodology		
			Database Searching	<i>In silico</i> Digestions	
				Specific	Non-Specific
					AISE, ATLD, SIEA, SVVD, LEGT, EASI, TGEI, GITE, DTIA, AETV
4.69	348.1768	8.57	No match	VTE, TDL, SLE, SIE, SEL, LSE, ISE	No Match
4.72	375.1878	5.04	No match	No Match	No Match
4.95	357.2135	4.13	No match	No Match	QLP, PGIA, PLAG, PIGA, AAVP
5.00	362.1920	24.33	E/T I	TIE, TEL, LTE, ETL	IET, ELT, LTE, TLE, ETI, TIE, TEL, LET, ETL, EIT, ITE
5.16	304.1508	4.93	No match	No Match	VAD, GDL, GEV, DAV, VDA, PST, ADV, DVA, LDG, IDG, TPS, STP, GVE, GLD, DGL, LGD, DGI, AVD
5.17	433.2294	5.58	No match	TIEA, TALE, LTEA, ELTA	SIDV, AELT, ELTA, TIEA, IEAT, TALE, IAET, LTEA, EITA, VLDS, VVES, TAIE
5.27	460.2401	2.68	No match	LQEA, AQIE	GVDAV, IEAQ, TKPD, QLEA, ALQE, LQEA, LQAE, LAQE, ELQA, AQIE, VNDI, AELQ, QAEL, ALEQ, EVVN, IDVN, LEAQ, KPTD, PTDK, QEIA, EAVAA, NDVI, VVDQ
5.35	362.1564	3.49	No match	DDL	DID, IDD, DDL, LDD, VED, DVE, EVD, DDI, DLD, EDV
5.50	392.1849	3.63	I/M E	MLE, IME, EML	EFP, FPE, PEF, IEM, EML, LME, MEI, MEL, IME, MLE

258

	Peptide Identification Methodology					
	RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	Database Searching	<i>In silico</i> Digestions	
					Specific	Non-Specific
	5.69	378.1691	1.72	No match	MVE	PFD, DFP, FPD, FDP, DPF, MEV, VEM, MLD, MDL, IMD, DLM, MVE
	6.10	360.2130	3.46	I/I D	VLE, LVE, IVE, IDL	EIV, VLE, ELV, LEV, LID, IVE, LDI, DLL, LDL, IDI, VIE, LVE, VEI, IDL, DII, IID, ILD
	6.12	372.7165	3.97	No match	No Match	No Match
	6.15	378.1697	5.44	No match	No Match	No Match
	6.39	316.2234	3.71	No match	LIA, IIA	ALI, LAL, IIA, LLA, ILA, LIA, IAL, LAI, AIL
259	6.45	458.2606	2.66	I A A S P	No Match	ALAPS, DKPV, PLAGT
	6.52	346.2337	4.34	No match	LTL	No Match
	6.58	249.1402	3.62	No match	No Match	No Match
	6.62	374.2286	24.53	No match	LLE, LIE, LEL, ELL, EIL	No Match
	6.67	352.1501	2.18	No match	GFE, FDA, ADF	GFE, ADF, GEF, EFG, FDA, DFA, FGE, FAD
	6.68	417.2346	2.31	No match	No Match	No Match
	6.85	341.6983	5.11	No match	No Match	No Match
	7.27	344.1820	5.39	No match	PDL	LPD, DPL, PDL, DPI, LDP, VEP, EPV
	7.41	316.6851	2.80	No match	No Match	No Match
	7.44	396.1769	8.60	No match	ETF	No Match
	7.64	358.1977	7.16	No match	No Match	No Match
	7.75	360.2127	3.14	No match	VLE, LVE, IVE, IDL	EIV, VLE, ELV, LEV, LID, IVE, LDI, DLL, LDL, IDI, VIE, LVE, VEI, IDL, DII, IID, ILD
	8.01	350.7037	3.02	No match	No Match	No Match

	Peptide Identification Methodology					
	Precursor			<i>In silico</i> Digestions		
	RT (Min)	Ion Mass (Da)	Peak Vol. (%)	Database Searching	Specific	Non-Specific
260	8.37	424.1711	2.84	E E F	FEE, EEF	EFE, FEE, EEF
	8.38	365.1823	2.68	No match	QAF	SKM, MSK, AFQ, QAF, AGFA, FQA
	8.65	336.6719	4.16	No match	No Match	No Match
	8.67	350.7036	2.99	No match	No Match	No Match
	8.74	328.7030	3.58	No match	No Match	No Match
	9.19	300.1928	7.58	No match	No Match	No Match
	9.44	360.2132	8.15	No match	VLE, LVE, IVE, IDL	EIV, VLE, ELV, LEV, LID, IVE, LDI, DLL, LDL, IDI, VIE, LVE, VEI, IDL, DII, IID, ILD
	9.48	374.2288	2.84	I I E	LLE, LIE, LEL, ELL, EIL	No Match
	9.56	212.1188	3.90	No match	No Match	No Match
	9.56	431.2498	3.14	No match	IADL	LEGI, EALV, VLAE, LDIA, DLAL, EGLI, VAEL, VAEI, VALE, LDLA, LEVA, IEAV, LIAD, EIVA, IEGL, IIEG, DIAL, ELIG, LGEL, GELI, IADL
	9.70	417.2344	2.65	V A I D	LVDA, LGDL, GVLE	VLEG, EGIV, GVLE, LVDA, VALD, DLVA, GVEI, LDGL, DVIA, DGII, DVLA, LGDL, AVDI, ADLV
	9.92	457.2653	19.02	P/I V E	No Match	PEIV, VIPE, DP II, PIVE, IVEP, ILPD
	10.39	387.2238	2.25	T/P G I	ISPA	TPGL, LAPS, ISAP, IPAS, APSI, ISPA, LPTG, PTGI, TGIP

	Peptide Identification Methodology					
	RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	Database Searching	<i>In silico</i> Digestions	
					Specific	Non-Specific
261	10.69	378.1694	2.95	No match	MVE	PFD, DFP, FPD, FDP, DPF, MEV, VEM, MLD, MDL, IMD, DLM, MVE
	10.74	475.2764	7.14	No match	ELTL	LELT, ELTL, LLTE, LETL, ILTE, EILT, ELLT
	10.78	392.1852	9.03	I/E M	MLE, IME, EML	EFP, FPE, PEF, IEM, EML, LME, MEI, MEL, IME, MLE
	11.09	301.6509	4.11	No match	No Match	No Match
	11.11	346.1975	2.96	I/D V	No Match	No Match
	11.30	371.2295	7.79	No match	No Match	No Match
	11.51	323.6639	6.06	No match	No Match	No Match
	11.68	346.1975	37.64	V/D I	No Match	No Match
	11.73	447.2449	2.85	No match	ISDL	ISDL, TDLV, EVTV, VLTD, DLSD, IISD, VDLT, DLTV, LSDI, TEVV, ILDS, TVEV
	11.80	229.1187	1.98	No match	No Match	No Match
	11.80	360.2128	29.91	V/E I	VLE, LVE, IVE, IDL	EIV, VLE, ELV, LEV, LID, IVE, LDI, DLL, LDL, IDI, VIE, LVE, VEI, IDL, DII, IID, ILD
	11.87	474.2558	4.89	No match	VDIQ, GLDGL	IDNL, GVLEG, IDNL, DLVQ, DIVQ, NLEV, QDLV, VIEN, ISAPS, GLDGL, DIIN, VDIQ, DIQV
	12.38	398.2397	100.00	No match	No Match	No Match
	12.38	569.3038	4.17	No match	HGEVK	FFKAG, APPER, VTHNV, HGEVK, IDGHK, DGHKI, KAVDH
	12.38	227.1756	2.80	No match	No Match	No Match
	12.59	437.2031	2.33	No match	No Match	MSTV, AEFA, PASY, VDFG

RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	Peptide Identification Methodology		
			Database Searching	<i>In silico</i> Digestions	
				Specific	Non-Specific
12.69	589.3189	2.42	T/I D/Q	TIIDQ, EKEAL	EKELA, EKEAL, LEEAK, ELEKA, DLVVK, QLSEL, LLTEN, EIQLS, KLEEA, LKEAE, LATLDG, ASLGEL, TIIDQ, TAAVDI
12.92	447.2448	2.30	I/D S I	ISDL	ISDL, TDLV, EVTV, VLTD, DLVL, IISD, VDLT, DLTV, LSDI, TEVV, ILDS, TVEV
12.92	488.3081	7.02	No match	ITALA, ALATL	LIDK, EKLK, VKEL, ITALA, LKLD, KLEV, LATAL, KVIE, EIKV, VEKL, ALATL, VVISA, EKVL, EKVI, VIVSA, LGVSL, VSLAV, ILGVS, DKII
12.93	346.1978	6.67	No match	No Match	No Match
12.93	215.1029	2.42	No match	No Match	No Match
13.39	198.1278	2.96	No match	No Match	No Match
13.43	531.2602	14.51	N V/P/A/M	No Match	NVPAM, PLNGM, DHFL, VFHE, IDHF
13.46	394.1974	5.40	E V F	LDF, IDF, EVF	IFD, FID, IDF, LDF, EVF, LFD, FDI, EFV
13.85	540.2118	2.57	M E N F	MENF	No Match
14.36	394.1973	4.67	V/E F	LDF, IDF, EVF	IFD, FID, IDF, LDF, EVF, LFD, FDI, EFV
14.38	378.1693	2.64	No match	MVE	PFD, DFP, FPD, FDP, DPF, MEV, VEM, MLD, MDL, IMD, DLM, MVE
14.58	488.3077	3.99	V V S\A	ITALA, ALATL	LIDK, EKLK, VKEL, ITALA, LKLD, KLEV, LATAL,

RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	Peptide Identification Methodology		
			Database Searching	<i>In silico</i> Digestions	
				Specific	Non-Specific
14.59	392.1849	3.11	No match	MLE, IME, EML	KVIE, EIKV, VEKL, ALATL, VVISA, EKVL, EKVI, VIVSA, LGVSL, VSLAV, ILGVS, DKII EFP, FPE, PEF, IEM, EML, LME, MEI, MEL, IME, MLE
14.95	374.2291	14.92	No match	LLE, LIE, LEL, ELL, EIL	No Match
15.01	417.2344	2.14	No match	LVDA, LGDL, GVLE	VLEG, EGIV, GVLE, LVDA, VALD, DLVA, GVEI, LDGL, DVIA, DGII, DVLA, LGDL, AVDI, ADLV
263 15.08	521.2604	4.52	Y/P I E	No Match	PIYE, YPIE, YELP, ETGKS, QTVSS, TVSSQ, STGSVA, SVTGGT
16.27	445.2655	3.12	V/V D I	ALEL	No Match
16.82	374.2290	7.42	No match	LLE, LIE, LEL, ELL, EIL	No Match
17.39	360.2129	36.88	I/D I	VLE, LVE, IVE, IDL	EIV, VLE, ELV, LEV, LID, IVE, LDI, DLL, LDL, IDI, VIE, LVE, VEI, IDL, DII, IID, ILD
17.39	229.1186	2.08	No match	No Match	No Match
17.59	374.2285	28.70	No match	LLE, LIE, LEL, ELL, EIL	No Match
18.31	360.2131	13.46	No match	VLE, LVE, IVE, IDL	EIV, VLE, ELV, LEV, LID, IVE, LDI, DLL, LDL, IDI, VIE, LVE, VEI, IDL, DII, IID, ILD
18.37	344.2549	6.03	No match	No Match	VLI, LVI, IVL, VII, VIL

	Precursor			Peptide Identification Methodology		
	RT (Min)	Ion Mass (Da)	Peak Vol. (%)	Database Searching	<i>In silico</i> Digestions	
					Specific	Non-Specific
	18.72	374.2289	17.39	No match	LLE, LIE, LEL, ELL, EIL	No Match
	19.60	408.2133	6.17	E/I F	LEF, EIF	No Match
	20.17	474.2556	2.44	I/G D G I	VDIQ, GLDGL	IDNL, GVLEG, IDNL, DLVQ, DIVQ, NLEV, QDLV, VIEN, ISAPS, GLDGL, DIIN, VDIQ, DIQV
	20.18	394.1973	4.11	I/D F	LDF, IDF, EVF	IFD, FID, IDF, LDF, EVF, LFD, FDI, EFV
	21.09	415.2147	9.90	No match	No Match	IPEG, VAPE, LDPA, PGEL, GELP
264	21.18	408.2135	8.76	No match	No Match	No Match
	21.65	603.2406	4.04	W D/A/P D	No Match	No Match
	22.06	459.2814	5.22	I V E V	No Match	QRR, RRGV, VLDI, IVLD, VEIV, LVID, VIDL, VDII, DVIL
	23.39	471.2813	5.25	I/P\E I	PEIL	PLLE, IPEL, PEIL, EILP
	23.82	488.2867	3.33	No match	No Match	IVWA, VWAI, RVVD, RIEA, ARIE, NKNL, QVNK, ERLA, RAEL, NNLK, NLKGG, AGKNV, AREI, NLKN, KVNQ
	24.03	593.2935	10.75	No match	No Match	No Match
	24.54	417.2497	3.70	No match	LVW	No Match
	24.79	484.3135	9.46	V G V P I	No Match	No Match
	25.11	593.2928	3.58	V/G V D G F	TVNPY	AMATVT, MATVTA, TCKEL, PFDSK, TVNPY, IFDAGA, AFLGDA, DFKSP, FKSPD, YPVQS
	25.64	459.2811	3.20	No match	No Match	QRR, RRGV, VLDI, IVLD, VEIV, LVID, VIDL, VDII, DVIL

RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	Database Searching	Peptide Identification Methodology	
				<i>In silico</i> Digestions	
				Specific	Non-Specific
25.66	605.2929	6.87	I N/D P F	INDPF	TGWLE, INDPF, NDPFI, DLAACI, EQLCI, IACIGE, IGMDVA, GESGAGK, RATEE, DSEVR, KENQS, TAERE, TEERA, EAETR, ETRAE, GKDATN
27.04	569.3656	4.16	No match	No Match	No Match
27.33	522.2557	4.41	No match	FLGDA, DIAGF	CTVTV, MATVT, DIAGF, PKYD, ENFL, LNFE, FENL, IFDAG, AFLGD, FLGDA, YPDK
28.24	570.3498	7.41	No match	No Match	VIPEL

Table A15. Compound list generated using the Find by algorithm with Molecular Feature Extraction using Agilent MassHunter Software for fraction 3 from GF Separation 4. The sequences of the detected ions were compared using the peptides identified by database searching and of predictions from *in silico* digestions with non-specific and specific hydrolysis of the curated protein library, outlined in Section 3.15. No Match indicates the ion was not identified by the indicated peptide identification methodology. Database search hits are reported to include the detected product ions, where / indicates the identification of the b-series ion, \ indicates the identification of the y-series ion, and | indicates the identification of both b- and y-series ions.

266

RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	Peptide Identification Methodology		
			Database Searching	<i>In silico</i> Digestion	
				Specific	Non-Specific
2.65	304.1871	1.68	No match	TLA, TIA, TAL, LTA, ITA, ATL	ATI, ATL, TAL, TIA, AIT, LAT, ALT, LTA, TLA, ITA, VVS, VSV, SVV, TAI
2.73	331.1980	4.92	No match	LQA, IQA, AVAA, ALQ	QAI, QAL, VVN, PKS, ALQ, LQA, LAQ, AQL, AQI, AIQ, IQA, AGIA, GAAL, AVAA, KSP
2.83	317.1827	2.89	No match	No Match	QIG, GQI, LNA, VQA, LAN, QVA, INA, NAL, VAQ, LQG, ANL, GQL, NLA, QAV, ALN, GLQ, GAGI, AIN, ANI, QGL, GVAA, IGQ, QLG, NAI
2.87	320.1817	2.24	No match	STL	TVT, LST, STL, TSL, VTT, TSI, ITS, SIT

RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	Peptide Identification Methodology		
			Database Searching	<i>In silico</i> Digestion	
				Specific	Non-Specific
2.88	439.1823	1.91	No match	QEY, EGYA	EAYG, EGYA, QEY
2.89	290.1716	7.73	No match	No Match	No Match
2.89	159.0765	1.33	No match	No Match	No Match
2.90	288.1926	2.50	No match	No Match	No Match
3.06	120.0809	1.96	No match	No Match	No Match
3.19	462.2558	1.96	T A/V TVA	VTDK, TTQL, LTTQ	MRR, RRM, VTDK, TVKD, TTQL, SLKD, AVTAT, KTVD, ATVTA, ISKD, LTTQ
3.21	347.1928	4.45	No match	SLGA, SIQ, QSL, KEA, KAE, GTGL, EKA, EAK	KEA, TGAV, KAE, AGTV, EAK, EKA, AEK, SLGA, NIT, NLT, QLS, QTV, TNL, SIQ, LQS, NTL, QSL, SQL, LNT, TQV, SAGI, ITN, GTGL, LSAG, LGTG, VTAG, QSI, INT, ASLG, IGTG, TLN, GGLT, GLTG, SQI, IGAS, VAAS
3.23	290.1715	11.69	A/S L U	No Match	No Match
3.24	159.0765	1.47	No match	No Match	No Match
3.25	332.2177	15.66	V/I T	SIL	No Match
3.28	213.1606	3.89	No match	No Match	No Match
3.33	302.2078	3.31	No match	No Match	No Match
3.39	260.1611	1.82	No match	VAA, QL, LQ, LGA, IQ, GIA, AVA	IAG, AGL, LGA, AVA, QI, QL, IQ, LQ, LAG, AGI, GIA, VAA, AIG, AAV, IGA
3.49	132.1018	2.28	No match	L	I, L
3.51	274.1763	11.17	A/A I	LAA, ALA, AAL	IAA, ALA, LAA, AAI, AAL, VGV, VVG, GVV

	Peptide Identification Methodology					
	Precursor Ion			<i>In silico</i> Digestion		
	RT (Min)	Mass (Da)	Peak Vol. (%)	Database Searching	Specific	Non-Specific
268	3.51	143.0816	2.60	No match	No Match	No Match
	3.62	247.1297	3.22	No match	No Match	No Match
	3.65	302.2074	27.04	I VA	VAL, LGL, IVA, GLL, AVL	VLA, IAV, ILG, AVL, LLG, LGL, LGI, GLL, LVA, VAL, GLI, ALV, GIL, GII, VAI, IVA, VIA, LIG, LAV, IIG
	3.65	213.1601	6.70	No match	No Match	No Match
	3.66	320.1819	1.74	No match	STL	TVT, LST, STL, TSL, VTT, TSI, ITS, SIT
	3.74	375.1879	3.46	No match	No Match	No Match
	3.80	361.2083	1.71	No match	VDK, QTL, LTQ, ASAL	LTGA, SALA, DKV, VKD, VDK, QIT, QTL, TQL, TLQ, QLT, LTQ, IQT, TQI, AVTA, KVD, LAGT, ASAL, KDV, AGIT, AIGT, TAAV
	3.92	290.1721	3.28	T/G I	No Match	No Match
	3.93	304.1868	21.03	No match	TLA, TIA, TAL, LTA, ITA, ATL	ATI, ATL, TAL, TIA, AIT, LAT, ALT, LTA, TLA, ITA, VVS, VSV, SVV, TAI
	3.93	173.0927	3.00	No match	No Match	PG, GP
	4.06	402.2352	1.88	No match	VVNA	VVNA, KTPG, NGVL, AAQI, AIQA, KPAS, ALQA, VING, QVVG
	4.14	332.2184	4.26	No match	SIL	No Match
	4.26	334.1977	2.61	No match	No Match	TLT, ITT, TTI, LTT
	4.38	231.1711	2.18	No match	No Match	VL, VI, LV, IV
	4.42	213.1605	1.92	No match	No Match	No Match
	4.42	332.2180	13.25	V I T	SIL	No Match

	Peptide Identification Methodology					
	RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	Database Searching	<i>In silico</i> Digestion	
					Specific	Non-Specific
269	4.51	389.2396	2.50	No match	No Match	GLVT, KIE, ASVL, LKE, LEK, VASI, VSAL, EKL, IEK, ASVI, LGSI, TLVG, KEL, KLE, ELK, EKI, IKE, IAVS, LLGS, ALVS, TGIV, KEI, GSIL, AVLS, VASL, EIK, VISA, VLAS, IVSA, GIIS, VSLA, SLAV
	4.62	302.2076	5.62	No match	No Match	No Match
	4.66	242.1510	2.31	No match	No Match	No Match
	4.84	389.2399	2.19	No match	No Match	GLVT, KIE, ASVL, LKE, LEK, VASI, VSAL, EKL, IEK, ASVI, LGSI, TLVG, KEL, KLE, ELK, EKI, IKE, IAVS, LLGS, ALVS, TGIV, KEI, GSIL, AVLS, VASL, EIK, VISA, VLAS, IVSA, GIIS, VSLA, SLAV
	4.88	324.1670	7.08	No match	PHA	PHA
	4.98	357.2136	12.32	P/G I\A	No Match	QLP, PGIA, PLAG, PIGA, AAVP
	5.09	306.1489	1.95	No match	No Match	ALC, ACI, VGM, GMV, GVM, MVG, VMG, CLA, MGV, LAC, IAC
	5.12	375.2241	1.70	A T A/I	VEK, LDK, KVE, GVSL, DIK, ATIA, ATAL	ATIA, VEK, LKD, VKE, IDK, KVE, DLK, EVK, LDK, DKI, DKL, KDI, KDL, SGLV, TALA, SIVG, DIK, VLGS,

RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	Peptide Identification Methodology			
			Database Searching	<i>In silico</i> Digestion		
				Specific	Non-Specific	
270	5.46	304.1872	3.36	V/S V	TLA, TIA, TAL, LTA, ITA, ATL	KLD, ATAL, LATA, KEV, ALAT, EKV, IKD, KID, LGVS, GVSL, IGSV ATI, ATL, TAL, TIA, AIT, LAT, ALT, LTA, TLA, ITA, VVS, VSV, SVV, TAI
	5.94	302.2079	3.99	No match	No Match	No Match
	5.96	380.6854	1.86	No match	No Match	No Match
	6.09	424.2078	5.13	I/Y E	ELY	No Match
	6.22	300.1925	3.49	No match	No Match	No Match
	6.41	316.2235	5.25	No match	No Match	No Match
	6.48	458.2611	2.30	I A A S P	No Match	ALAPS, DKPV, PLAGT
	6.53	346.2340	2.32	No match	LTL	No Match
	6.57	446.2613	2.69	No match	AKVE, DAIK, EVKA, GEKL, GVSLA, IKDA, IKGE, KGLE, LKDA, LVSQ, SINL, TAAAL, VKAE, VSQL	KEGI, TAAAL, IKGE, LKDA, KLEG, EVKA, IAVSG, KADI, ADKI, DLAK, AKVE, KVAE, QSVL, VNTL, LVSQ, VSQL, SLLN, IGGSI, GGSIL, DAIK, VAEK, LEKG, VKAE, SINL, AKLD, KELG, IGEK, GEKL, EKGL, AIKD, IKDA, KGLE, VKEA, KDIA, GVSLA, AVEK, VQSI
	6.71	554.3296	2.39	I H/A L T	No Match	No Match
	6.77	235.6715	6.75	No match	No Match	No Match
	6.94	330.2029	1.71	No match	No Match	PTL, TLP, LTP, LPT
	6.94	352.1400	1.37	No match	No Match	No Match

RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	Peptide Identification Methodology		
			Database Searching	<i>In silico</i> Digestion	
				Specific	Non-Specific
7.22	430.2657	32.88	I G G A	VALQ, VAIQ, NAIL, LQVA, LIGQ	AGLLG, VQAL, ALQV, LQVA, VAIQ, IQAV, QAVL, VALQ, KPVS, IALN, IGVAA, IINA, LIGQ, NAIL
7.23	371.2295	4.09	No match	No Match	No Match
7.25	387.2604	2.01	V/V VA	VVGL, LAAL, IIAA	LAAL, IIAA, LLAA, VVGI, ILAA, VVGL, LVVG
7.44	396.1767	5.57	No match	ETF	No Match
7.51	446.2611	2.31	No match	AKVE, DAIK, EVKA, GEKL, GVSLA, IKDA, IKGE, KGLE, LKDA, LVSQ, SINL, TAAAL, VKAE, VSQL	KEGI, TAAAL, IKGE, LKDA, KLEG, EVKA, IAVSG, KADI, ADKI, DLAK, AKVE, KVAE, QSVL, VNTL, LVSQ, VSQL, SLLN, IGGSI, GGSIL, DAIK, VAEK, LEKG, VKAE, SINL, AKLD, KELG, IGEK, GEKL, EKGL, AIKD, IKDA, KGLE, VKEA, KDIA, GVSLA, AVEK, VQSI
7.54	316.2235	7.20	V/V V	No Match	No Match
7.80	346.2335	10.63	No match	LTL	No Match
7.81	419.2501	2.44	No match	No Match	TVTV, VTVT, TSVL
8.01	316.2237	2.35	No match	No Match	No Match
8.02	350.7036	7.16	No match	No Match	No Match
8.25	359.2292	4.74	No match	LNL, GVIA	GAVL, GLLG, VIQ, LLN, LQV, VQL, LVQ, IVQ, QVI, QLV, GILG, NII, INL, IGVA, GVIA, IIN, LNL, NIL, IQV
8.25	345.2138	2.58	N V i	No Match	No Match

RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	Peptide Identification Methodology		
			Database Searching	<i>In silico</i> Digestion	
				Specific	Non-Specific
8.27	318.2027	6.89	S/V I	No Match	No Match
8.28	187.1084	2.22	No match	PA	AP, PA
8.37	171.1131	10.35	No match	No Match	No Match
8.38	302.2076	31.42	A/V I	No Match	No Match
8.38	143.1181	2.56	No match	No Match	No Match
8.38	332.2183	5.56	T/VI	SIL	No Match
8.52	417.2713	4.39	V/I V S	LTLA, LATL, ITAL	ALIT, AITI, LATL, LTLA, ITAL, VVIS, VIVS
8.52	288.1927	2.95	No match	No Match	No Match
8.53	448.2227	2.02	I/A N M	No Match	KDW, GGFAP, AAWT, PAFN, FAPN, MVQA, LANM, NLAM, VGCVA, CIGGV, CNVL, VVCGA, MGQL
8.71	350.7040	5.18	No match	No Match	No Match
8.73	334.1799	5.22	No match	MAL	MAL, LAM, LVC, AIM, LMA, CVL, IAM, AMI
8.78	350.1750	6.36	No match	No Match	No Match
9.20	288.1925	1.90	No match	No Match	No Match
9.20	492.2491	4.58	I/T G M A	No Match	EKLC, TQIM, MTQI, ITGMA, AMATV, CKEL, MKDV, SLAVC, VGLCT, GVMVS
9.31	417.2710	2.52	V/V I S	LTLA, LATL, ITAL	ALIT, AITI, LATL, LTLA, ITAL, VVIS, VIVS
9.38	334.1800	4.62	I/A M	MAL	MAL, LAM, LVC, AIM, LMA, CVL, IAM, AMI
9.56	212.1190	2.89	No match	No Match	No Match
9.60	302.2079	3.19	I/A V	No Match	No Match

	Peptide Identification Methodology					
	RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	Database Searching	In silico Digestion	
					Specific	Non-Specific
273	9.61	399.2606	3.36	No match	No Match	PLGI, VPLA, VIAP, GIPI, LPIG
	9.65	389.2399	1.86	V/T G/I	No Match	GLVT, KIE, ASVL, LKE, LEK, VASI, VSAL, EKL, IEK, ASVI, LGSI, TLVG, KEL, KLE, ELK, EKI, IKE, IAVS, LLGS, ALVS, TGIV, KEI, GSIL, AVLS, VASL, EIK, VISA, VLAS, IVSA, GIIS, VSLA, SLAV
	9.73	345.2134	6.36	No match	No Match	ALAA, AAAL, TKP, KTP, VLN, VQV, NVL, VGGI, KPT, QVV, IGGV, GGVI, LNV, GLGV, LVGG, VIN, NVI
	9.80	511.2397	2.23	I/T D Y	YVTE, TDYL, ESIY	No Match
	9.85	350.1750	3.17	No match	No Match	No Match
	9.92	359.2291	1.70	No match	LNL, GVIA	GAVL, GLLG, VIQ, LLN, LQV, VQL, LVQ, IVQ, QVI, QLV, GILG, NII, INL, IGVA, GVIA, IIN, LNL, NIL, IQV
	10.04	288.1923	10.57	I G V	GVL	GLV, IGV, GVL, GIV, LGV, VGL, LVG, IVG, GVI, VGI
	10.10	470.2974	2.32	I I P/Q	No Match	PLLQ, LPIGA
	10.18	171.1130	6.79	No match	No Match	No Match
	10.18	302.2075	47.83	No match	No Match	No Match
	10.27	391.2011	2.13	No match	CVGL	No Match

RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	Peptide Identification Methodology		
			Database Searching	<i>In silico</i> Digestion	
				Specific	Non-Specific
10.41	387.2238	7.35	T/P G I	ISPA	TPGL, LAPS, ISAP, IPAS, APSI, ISPA, LPTG, PTGI, TGIP
10.45	332.2178	10.68	No match	SIL	No Match
10.54	359.2289	13.99	No match	LNL, GVIA	GAVL, GLLG, VIQ, LLN, LQV, VQL, LVQ, IVQ, QVI, QLV, GILG, NII, INL, IGVA, GVIA, IIN, LNL, NIL, IQV
10.63	364.1908	2.93	No match	No Match	LMT, IMT, TIM, LTM
10.65	328.2235	15.76	No match	No Match	VIP, LPV, VPI, VPL, PIV, IVP
10.76	318.2025	17.51	I/S V	No Match	No Match
11.05	330.2390	1.95	No match	No Match	VVI, VVL, VIV, LVV
11.06	332.2181	12.34	No match	SIL	No Match
11.08	448.2225	1.76	I/G A G/M	No Match	KDW, GGFAP, AAWT, PAFN, FAPN, MVQA, LANM, NLAM, VGCVA, CIGGV, CNVL, VVCGA, MGQL
11.19	328.2240	7.90	No match	No Match	VIP, LPV, VPI, VPL, PIV, IVP
11.32	362.2113	2.75	V/I M	No Match	VFP, MVI, LMV
11.33	492.2488	2.31	No match	No Match	EKLC, TQIM, MTQI, ITGMA, AMATV, CKEL, MKDV, SLAVC, VGLCT, GVMVS
12.22	330.2393	3.22	No match	No Match	VVI, VVL, VIV, LVV
12.41	446.2611	2.78	No match	AKVE DAIK EVKA GEKL GVSLA IKDA	KEGI, TAAAL, IKGE, LKDA, KLEG, EVKA, IAVSG, KADI,

RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	Peptide Identification Methodology			
			Database Searching	<i>In silico</i> Digestion		
				Specific	Non-Specific	
275	12.44	366.2031	4.07	No match	No Match	ADKI, DLAK, AKVE, KVAE, QSVL, VNTL, LVSQ, VSQL, SLLN, IGGSI, GGSIL, DAIK, VAEK, LEKG, VKAE, SINL, AKLD, KELG, IGEK, GEKL, EKGL, AIKD, IKDA, KGLE, VKEA, KDIA, GVSLA, AVEK, VQSI
	12.66	362.2112	1.94	No match	No Match	No Match
	12.90	360.1957	2.56	No match	No Match	VFP, MVI, LMV
	12.92	330.2389	26.96	V/V I	No Match	MIP, QRG, RAN, ARN, RNA, GQR, GRQ, ANR, NRA, GRGA
	12.92	199.1445	4.73	No match	No Match	VVI, VVL, VIV, LVV
	12.93	334.1800	2.37	No match	MAL	No Match
	12.94	488.3076	10.57	No match	ITALA, ALATL	MAL, LAM, LVC, AIM, LMA, CVL, IAM, AMI
	12.95	346.1979	2.80	No match	No Match	LIDK, EKL, DLKL, VKEL, ITALA, LKLD, KLEV, LATAL, KVIE, EIKV, VEKL, ALATL, VVISA, EKVL, EKVI, VIVSA, LGVSL, VSLAV, ILGVS, DKII
	13.13	350.1749	1.79	No match	No Match	No Match
	13.14	389.2400	2.80	V/A S L U	No Match	GLVT, KIE, ASVL, LKE, LEK, VASI, VSAL, EKL, IEK, ASVI, LGSIL, TLVG, KEL,

RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	Peptide Identification Methodology		
			Database Searching	<i>In silico</i> Digestion	
				Specific	Non-Specific
					KLE, ELK, EKI, IKE, IAVS, LLGS, ALVS, TGIV, KEI, GSIL, AVLS, VASL, EIK, VISA, VLAS, IVSA, GIIS, VSLA, SLAV
13.56	364.1909	5.46	No match	No Match	LMT, IMT, TIM, LTM
13.59	362.2114	2.78	No match	No Match	VFP, MVI, LMV
13.60	346.2341	4.50	No match	LTL	No Match
13.79	387.2606	2.02	No match	No Match	No Match
13.87	330.2390	2.31	No match	No Match	VVI, VVL, VIV, LVV
13.89	403.2553	7.50	No match	LLSA, ITVA, GTLL	ATLV, LVTA, LITG, TILG, TIAV, ITVA, VALT, TALV, ITIG, SILA, GILT, GIIT, ILAS, LASL, LLSA, VSVV, GTLL, VTAL, LTGL
13.93	316.2234	9.36	No match	LIA, IIA	ALI, LAL, IIA, LLA, ILA, LIA, IAL, LAI, AIL
13.93	185.1290	2.58	No match	No Match	No Match
14.01	413.2761	3.15	No match	No Match	IIAP, LAPL, IIPA
14.06	359.2291	2.31	I/V/A G	LNL, GVIA	GAVL, GLLG, VIQ, LLN, LQV, VQL, LVQ, IVQ, QVI, QLV, GILG, NII, INL, IGVA, GVIA, IIN, LNL, NIL, IQV
14.11	332.2181	21.12	S/I/I	SIL	No Match
14.11	201.1238	5.11	No match	No Match	No Match
14.38	394.1971	2.27	V/E F	LDF, IDF, EVF	IFD, FID, IDF, LDF, EVF, LFD, FDI, EFV
14.43	449.2433	5.63	No match	No Match	VISM, YPGI, VFPS, PAYV

276

RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	Peptide Identification Methodology		
			Database Searching	<i>In silico</i> Digestion	
				Specific	Non-Specific
14.51	302.2077	7.63	G/I/I	No Match	No Match
14.51	171.1132	2.49	No match	No Match	No Match
14.57	385.2451	2.57	V/G P I	No Match	GIVP, GVPL
14.58	488.3076	7.34	V V I S A	ITALA, ALATL	LIDK, EKLV, DLKL, VKEL, ITALA, LKLD, KLEV, LATAL, KVIE, EIKV, VEKL, ALATL, VVISA, EKVL, EKVI, VIVSA, LGVSL, VSLAV, ILGVS, DKII
14.81	373.2446	15.19	No match	IQL, GIAL	AGLL, QII, LLQ, IQL, QLL, QLI, GIAL, VLAA, GILA, AILG
14.90	302.2079	5.58	No match	No Match	No Match
15.15	387.2605	1.77	No match	No Match	No Match
15.23	359.2292	6.97	No match	LNL, GVIA	GAVL, GLLG, VIQ, LLN, LQV, VQL, LVQ, IVQ, QVI, QLV, GILG, NII, INL, IGVA, GVIA, IIN, LNL, NIL, IQV
15.24	346.2338	8.17	No match	LTL	No Match
15.25	415.2918	2.52	V/I I A	LLGL	No Match
15.37	510.2557	1.95	S T G V F	YKEA, STGVF, SIQY, EFSK, AEKY	TFKD, YAGTV, EFSK, YNIT, YSIQ, TYLN, LYSQ, SIQY, LYASG, AEKY, GATYV, STGVF, TQYV, QYVT, YGGLT, YKEA
15.39	385.2452	3.57	V G I P	No Match	GIVP, GVPL
15.48	362.2114	2.84	No match	No Match	VFP, MVI, LMV
15.58	332.2182	7.19	S/I/I	SIL	No Match

	Peptide Identification Methodology					
	RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	Database Searching	<i>In silico</i> Digestion	
					Specific	Non-Specific
278	15.61	316.2233	9.33	No match	LIA, IIA	ALI, LAL, IIA, LLA, ILA, LIA, IAL, LAI, AIL
	15.61	185.1289	2.00	No match	No Match	No Match
	15.74	302.2079	3.56	No match	No Match	No Match
	15.78	344.2547	1.92	No match	No Match	VLI, LVI, IVL, VII, VIL
	15.81	342.2394	3.49	No match	No Match	PLL, LIP, PII, IIP, ILP, IPI, LPI
	15.92	316.2232	37.96	I/A I	LIA, IIA	ALI, LAL, IIA, LLA, ILA, LIA, IAL, LAI, AIL
	15.92	185.1290	4.71	No match	No Match	No Match
	16.06	387.2608	4.88	V/V G I	No Match	No Match
	16.18	373.2446	24.16	No match	IQL, GIAL	AGLL, QII, LLQ, IQL, QLL, QLI, GIAL, VLAA, GILA, AILG
	16.42	415.2920	3.01	I V IA	LLGL	No Match
	16.43	302.2076	7.19	I/G L U	No Match	No Match
	16.48	332.2183	5.56	No match	SIL	No Match
	16.70	346.2341	7.88	I/T I	LTL	No Match
	16.78	344.2546	14.73	No match	No Match	VLI, LVI, IVL, VII, VIL
	16.78	213.1605	1.96	No match	No Match	No Match
	16.84	362.2115	5.41	No match	No Match	No Match
	17.08	316.2236	7.64	No match	No Match	No Match
	17.17	373.2447	1.53	No match	IQL, GIAL	AGLL, QII, LLQ, IQL, QLL, QLI, GIAL, VLAA, GILA, AILG
	17.35	342.2390	13.40	No match	No Match	PLL, LIP, PII, IIP, ILP, IPI, LPI

	Peptide Identification Methodology					
	RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	Database Searching	<i>In silico</i> Digestion	
					Specific	Non-Specific
	17.45	417.2710	2.17	No match	LTLA, LATL, ITAL	ALIT, AITI, LATL, LTLA, ITAL, VVIS, VIVS
	17.47	332.2183	2.08	No match	SIL	No Match
	17.69	346.2337	19.65	I/T I	LTL	No Match
	17.98	376.2267	2.18	I/I M	No Match	No Match
	18.18	292.2134	2.62	No match	No Match	No Match
	18.32	360.2131	7.71	No match	VLE, LVE, IVE, IDL	EIV, VLE, ELV, LEV, LID, IVE, LDI, DLL, LDL, IDI, VIE, LVE, VEI, IDL, DII, IID, ILD
279	18.36	687.5014	1.35	No match	No Match	No Match
	18.38	213.1601	13.63	No match	No Match	No Match
	18.40	344.2546	100.00	No match	No Match	VLI, LVI, IVL, VII, VIL
	18.65	415.2914	12.64	No match	No Match	LLGL, LVAL, IVAI, VIIA, IIIG
	18.83	413.2760	1.99	No match	No Match	IIAP, LAPL, IIPA
	19.61	408.2135	3.52	E/I F	No Match	No Match
	20.18	344.2546	21.25	No match	No Match	VLI, LVI, IVL, VII, VIL
	20.19	213.1604	3.09	No match	No Match	No Match
	20.20	394.1974	8.84	I/D F	LDF, IDF, EVF	IFD, FID, IDF, LDF, EVF, LFD, FDI, EFV
	21.03	376.2265	7.91	No match	No Match	No Match
	21.22	408.2133	7.07	No match	LEF, EIF	No Match
	21.51	376.2268	4.38	No match	No Match	No Match
21.66	605.3293	8.11	G G V/P/L Y	No Match	IAMATV, MIKDV, AALYPA, ALYPAA, TFQLP, TLFQP, NVPIY, FDKPV, SGKGKE, SGGEKK, KTKND, DKSKQ, ASTGAAK, STGAAKA, ENKSK	

RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	Peptide Identification Methodology		
			Database Searching	<i>In silico</i> Digestion	
				Specific	Non-Specific
21.68	401.2760	17.88	V/I G I	LGVL, IGVL	IGVL, LLGV, LGVL, VGLL, GIVL, VILG, LGVI, VVLA, ILGV
21.91	358.2705	6.46	No match	No Match	No Match
22.53	401.2763	4.87	I/G V L	LGVL, IGVL	IGVL, LLGV, LGVL, VGLL, GIVL, VILG, LGVI, VVLA, ILGV
23.36	441.3072	11.44	I/V P I	No Match	IVPI, IPIV, VILP
23.42	358.2702	14.34	No match	No Match	LLL, III
23.81	401.2763	3.23	V G I I	LGVL, IGVL	IGVL, LLGV, LGVL, VGLL, GIVL, VILG, LGVI, VVLA, ILGV
24.29	358.2703	7.93	I I I	No Match	LLL, III
24.80	484.3133	4.75	V G V P I	No Match	LAPLA, GVVPL
25.13	559.3450	6.67	No match	LVSQI, KLEVA, IGGSIL, EKLVA	EKLVA, LAKVE, DLKLA, KEALV, KVLAE, EKLLG, EGLIK, QSVLI, LVSQI, IGGSIL, KLEVA, VAKLE, KIEAV, EKVLA, DIALK, GELIK, IGEKL, AILGVS, LGVSLA
25.78	358.2707	4.32	No match	No Match	No Match
25.99	516.3393	2.09	I/T I I G	No Match	RGRK, KRRG, LATLV, ITILG, LVALT, GILTL, ILASL, LLLSA
26.03	559.3447	2.52	I G G S I L	LVSQI, KLEVA, IGGSIL, EKLVA	EKLVA, LAKVE, DLKLA, KEALV, KVLAE, EKLLG, EGLIK, QSVLI, LVSQI,

RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	Peptide Identification Methodology		
			Database Searching	<i>In silico</i> Digestion	
				Specific	Non-Specific
					IGGSIL, KLEVA, VAKLE, KIEAV, EKVLA, DIALK, GELIK, IGEKL, AILGVS, LGVSLA
26.29	479.2867	5.55	No match	No Match	AIYL, IFSI, FSIL, IIYA, ITVF, VTFL, TVFL, VFLT
27.25	479.2866	6.26	No match	No Match	AIYL, IFSI, FSIL, IIYA, ITVF, VTFL, TVFL, VFLT
28.79	459.3176	11.00	No match	ITIL	KRR, ITIL, ILTL
29.21	795.4036	2.76	F/Q P S/F IG	No Match	NSSRFGK, FQPSFIG, AMYVAIQ, MYVAIQA, PDLSLHN, ARYITGD

Table A16. Compound list generated using the Find by algorithm with Molecular Feature Extraction using Agilent MassHunter Software for fraction 6 from GF Separation 4. The sequences of the detected ions were compared using the peptides identified by database searching and of predictions from *in silico* digestions with non-specific and specific hydrolysis of the curated protein library, outlined in Section 3.15. No Match indicates the ion was not identified by the indicated peptide identification methodology. Database search hits are reported to include the detected product ions, where / indicates the identification of the b-series ion, \ indicates the identification of the y-series ion, and | indicates the identification of both b- and y-series ions.

RT (Min)	Precursor Ion		Peptide Identification Methodology		
			Peak Vol. (%)	<i>In silico</i> Digestion	
Mass (Da)	Database Searching	Specific		Non-Specific	
2.74	281.1499	43.05	No match	VY	YV, VY
2.95	288.2036	3.11	No match	No Match	LR, RI, IR, RL
3.08	103.0535	6.66	No match	No Match	No Match
3.08	166.0855	18.59	No match	F	No Match
3.09	120.0804	100.00	No match	No Match	No Match
3.34	166.0867	2.74	No match	No Match	No Match
3.36	223.1078	13.73	No match	GF	FG, GF
3.45	120.0808	5.50	No match	No Match	No Match
3.47	166.0865	2.42	No match	No Match	No Match
3.50	237.1237	7.04	No match	No Match	AF, FA
3.69	350.1712	3.11	P/A Y	YPA, PSF	YPA, FPS, PSF, PAY, SIM, MTV, ISM, LMS, MVT, MSI
4.39	313.1217	13.16	No match	MY	MY, YM
4.92	263.1394	43.84	No match	PF	PF, FP

			Peptide Identification Methodology		
RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	Database Searching	<i>In silico</i> Digestion	
				Specific	Non-Specific
6.21	295.1654	64.52	No match	YL, LY, IY	IY, YL, LY, YI
6.25	396.2136	5.08	I/Y\T	No Match	No Match
6.63	288.2037	4.72	No match	No Match	LR, RI, IR, RL
7.33	378.2026	9.67	No match	No Match	PVY, YPV
7.79	265.1555	5.53	No match	No Match	FV, VF
7.93	295.1654	82.74	No match	YL, LY, IY	IY, YL, LY, YI
7.93	136.0756	2.59	No match	No Match	No Match
8.40	257.6904	4.23	No match	No Match	No Match
9.68	347.1718	6.78	No match	No Match	AAW, WAA
9.93	423.2243	3.71	No match	SGLF, KFE, KEF, GYAL, EFK	KEF, SGLF, LGFS, GIFS, EFK, FEK, QIY, QLY, IQY, SFIG, GYAL, KFE, TGVF, GVFT, GVTF, FLSG, LYQ, VSAF
9.96	352.1875	4.53	No match	No Match	IYG, IGY, GYL, YIG, LYG, FSV, YVA, SFV, AYV, LGY, GLY, VFS, YLG
10.26	366.2024	11.54	No match	VTF, TVF, LYA, ALY, AIY	AIY, VTF, FSI, FSL, ALY, IYA, IFS, LYA, YAL, FVT, SFI, SFL, TVF, VFT, SIF, SLF, FLS, LAY, FTV
10.52	412.1903	4.40	No match	MVY	HQQ, ANHA, MYV, YMV, MVY, VYM
10.55	396.2132	2.82	I/T Y	TYL, TLY, LTY	TYL, IYT, LTY, YTL, TLY, LYT, YIT
12.36	394.2339	5.02	I/V Y	YVL	VLY, YVL, VYL, IVY, YVI
13.41	279.1707	10.53	No match	No Match	LF, IF, FI, FL
14.55	386.1712	8.28	No match	FGY	FGY

	Peptide Identification Methodology					
	Precursor Ion			<i>In silico</i> Digestion		
	RT (Min)	Mass (Da)	Peak Vol. (%)	Database Searching	Specific	Non-Specific
284	14.93	426.2060	3.01	No match	MIY	GVNH, MIY, YLM
	15.12	416.1818	2.95	No match	No Match	YSF, FSY
	15.50	205.0979	48.53	No match	No Match	No Match
	15.50	391.1977	7.02	V/S W	SVW	AAMV, MQI, LQM, QIM, QML, CVGL, VGLC, QLM
	15.72	279.1710	4.80	No match	No Match	LF, IF, FI, FL
	15.91	408.2494	38.77	I/I Y	LYL, IYL, ILY	IYL, LYL, YLL, ILY
	15.92	375.2031	4.93	No match	GIW	PMK, KPM, VWA, GWL, WIG, GIW
	16.10	375.2029	3.94	No match	GIW	PMK, KPM, VWA, GWL, WIG, GIW
	16.26	405.2133	25.93	No match	VTW	VTW, AAFP, TER, RTE, ERT, ETR, VGMV, VMVG, ALVC, GVMV, VMGV, LAAM, IAMA, VIAC, AAIM, LAVC
	16.56	299.1611	18.53	No match	No Match	No Match
	17.15	408.2499	4.02	I/I Y	No Match	No Match
	18.12	462.2351	6.13	No match	No Match	EWK, EARS, TERG, ASER, AERS, LCIN, VCQL, GVMVG, GVIAC, LAMQ, NCLL
	18.84	313.1550	38.72	No match	FF	FF
	20.01	290.1509	4.83	No match	No Match	No Match
	20.15	403.2345	4.43	No match	No Match	VWV, WVV
	21.26	460.2555	11.60	No match	No Match	No Match
	23.60	432.2244	4.20	No match	NLW	PGKM, PMKG, GPMK, YHI, WIGG, QWV, NLW, ARAD,

RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	Peptide Identification Methodology		
			Database Searching	<i>In silico</i> Digestion	
				Specific	Non-Specific
28.04	412.2233	12.94	No match	VFF	QRE, REQ, EQR, RAEG, QER, REAG
28.07	531.2926	3.12	W/V/N/I	No Match	VFF LVHY, KTWP, VWIGG, HVYL, DAVRA, RINE, RENL, RLQD, GIADR, DLAGR, LAGRD, AGRDL, NELR, IENR, DARGI, ADRLG, RLNE, LNER
29.59	424.2580	8.45	No match	No Match	No Match
30.10	561.3035	4.76	V A V S W	No Match	No Match
32.62	426.2391	4.98	I F F	No Match	IFF, FLF

Table A17. Compound list generated using the Find by algorithm with Molecular Feature Extraction using Agilent MassHunter Software for fraction 7 from GF Separation 4. The sequences of the detected ions were compared using the peptides identified by database searching and of predictions from *in silico* digestions with non-specific and specific hydrolysis of the curated protein library, outlined in Section 3.15. No Match indicates the ion was not identified by the indicated peptide identification methodology. Database search hits are reported to include the detected product ions, where / indicates the identification of the b-series ion, \ indicates the identification of the y-series ion, and | indicates the identification of both b- and y-series ions.

RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	Peptide Identification Methodology		
			Database Searching	<i>In silico</i> Digestion	
				Specific	Non-Specific
3.11	103.0540	8.11	No match	No Match	No Match
3.12	120.0809	100.00	No match	No Match	No Match
3.20	166.0860	4.87	No match	F	No Match
5.32	295.1658	2.38	No match	No Match	No Match
6.23	295.1658	2.60	No match	No Match	No Match
9.70	347.1718	4.04	No match	No Match	AAW, WAA
9.76	329.1501	4.74	No match	YF, FY	No Match
15.25	318.1819	2.59	No match	No Match	No Match
15.59	391.1980	2.84	No match	SVW	AAMV, MQI, LQM, QIM, QML, CVGL, VGLC, QLM
16.28	405.2134	7.52	No match	VTW	VTW, AAFP, TER, RTE, ERT, ETR, VGMV, VMVG, ALVC, GVMV, VMGV, LAAM, IAMA, VIAC, AAIM, LAVC

RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	Peptide Identification Methodology		
			Database Searching	<i>In silico</i> Digestion	
				Specific	Non-Specific
16.38	473.2034	3.81	G Y S F	GYSF	PPME, GYSF, FGYS, FSYG
18.14	401.2190	3.95	No match	No Match	PVW
21.39	375.2031	2.15	No match	GIW	PMK, KPM, VWA, GWL, WIG, GIW

Table A18. Compound list generated using the Find by algorithm with Molecular Feature Extraction using Agilent MassHunter Software for fraction 9 from GF Separation 4. The sequences of the detected ions were compared using the peptides identified by database searching and of predictions from *in silico* digestions with non-specific and specific hydrolysis of the curated protein library, outlined in Section 3.15. No Match indicates the ion was not identified by the indicated peptide identification methodology. Database search hits are reported to include the detected product ions, where / indicates the identification of the b-series ion, \ indicates the identification of the y-series ion, and | indicates the identification of both b- and y-series ions.

RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	Peptide Identification Methodology		
			Database Searching	<i>In silico</i> Digestion	
				Specific	Non-Specific
0.40	121.0509	100.00	No match	No Match	No Match
1.06	141.1130	5.76	No match	No Match	No Match
1.78	121.0509	21.13	No match	No Match	No Match
3.10	173.0785	1.91	No match	No Match	No Match

RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	Peptide Identification Methodology		
			Database Searching	<i>In silico</i> Digestion	
				Specific	Non-Specific
3.20	103.0539	1.12	No match	No Match	No Match
3.20	120.0807	23.19	No match	No Match	No Match
3.21	166.0864	4.79	No match	F	No Match
4.38	215.1289	1.58	No match	No Match	No Match
5.40	345.1448	2.43	No match	No Match	No Match
5.51	292.1293	3.64	No match	No Match	WS
5.78	188.0709	6.63	No match	No Match	No Match
5.97	276.1352	24.19	No match	No Match	AW, WA
6.22	217.1049	3.17	No match	No Match	No Match
9.80	329.1498	2.47	No match	YF, FY	No Match
10.11	304.1661	11.88	No match	VW	VW, WV
11.94	329.1502	16.59	No match	YF, FY	No Match
12.38	444.2128	2.54	No match	No Match	QAEP, VDNP
13.85	336.1378	9.43	No match	MW	WM, MW, DTT, EST, STE, TES
15.28	318.1814	88.62	No match	LW, IW	WL, LW, IW, WI
17.30	318.1817	19.27	No match	LW, IW	WL, LW, IW, WI
19.72	318.1820	26.67	No match	LW, IW	WL, LW, IW, WI
19.72	301.1550	2.72	No match	No Match	No Match
24.12	330.1812	4.80	No match	No Match	No Match
28.91	439.1974	1.95	No match	No Match	GLFC
34.00	451.2338	2.58	No match	WVF	WVF, FRE, REF, ERF
37.43	589.2763	2.01	No match	No Match	TWPW
50.08	128.9504	5.27	No match	No Match	No Match
50.10	186.0581	1.87	No match	No Match	No Match
50.62	141.1131	0.71	No match	No Match	No Match

Table A19. Compound list generated using the Find by algorithm with Molecular Feature Extraction using Agilent MassHunter Software for fraction 6 from SAX Separation 5. The sequences of the detected ions were compared using the peptides identified by database searching and of predictions from *in silico* digestions with non-specific and specific hydrolysis of the curated protein library, outlined in Section 3.15. No Match indicates the ion was not identified by the indicated peptide identification methodology. Database search hits are reported to include the detected product ions, where / indicates the identification of the b-series ion, \ indicates the identification of the y-series ion, and | indicates the identification of both b- and y-series ions.

289

RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	Peptide Identification Methodology		
			Database Searching	<i>In silico</i> Digestion	
				Specific	Non-Specific
2.44	390.1870	30.93	No match	LEE, IEE, ELE, EIE, EEL	HPH, EEL, LEE, EEI, IEE, ELE, EIE
2.45	376.1714	8.01	No match	VEE, DIE, DEL	IED, LED, ELD, EID, VEE, EEV, DIE, DLE, EVE, LDE, EDL, DEL
2.61	297.1082	4.54	No match	DY	YD, DY
2.71	362.1917	4.09	T E	TIE, TEL, LTE, ETL	IET, ELT, LTE, TLE, ETI, TIE, TEL, LET, ETL, EIT, ITE
2.71	243.1341	3.93	No match	No Match	No Match
2.72	412.1689	5.91	No match	No Match	ETY, YET, YTE
2.73	390.1869	100.00	I/E E	LEE, IEE, ELE, EIE, EEL	HPH, EEL, LEE, EEI, IEE, ELE, EIE
2.77	398.1553	4.83	No match	No Match	SEY, SYE, TDY, YSE

RT (Min)	Precursor Ion		Peptide Identification Methodology		
			Peak Vol. (%)	<i>In silico</i> Digestion	
				Database Searching	Specific
2.80	295.1291	9.79	No match	FE, EF	EF, FE
2.95	439.1826	36.71	No match	No Match	No Match
3.01	361.1716	3.30	No match	AAEA	AAEA, GVDA, IND, LND, ND, QVD, IDN, DNL, ENV, NDI, NLD, DLN, LDN, VAGD, TPSG, VNE, SAPS, GLDG, DNI, QDV, AAAE, DVAG, VDQ
3.03	173.0786	2.73	No match	No Match	No Match
3.04	491.2339	4.69	No match	No Match	THPH, LTEE, TEEL, EETI, ETIE, ETEL, ELET, LETE
3.06	384.1405	23.44	No match	No Match	No Match
3.09	382.1607	4.76	No match	SEF, EAY, DTF, AYE	EAY, EFS, AYE, FDT, SEF, DTF
3.09	334.1609	2.03	No match	SDL	VTD, TVD, SID, ESV, DLS, DIS, SEV, ISD, SDL, LDS, VDT, VES, LSD, SDI, DSL, VSE
3.10	463.2034	51.25	S I E D	SDLE, DLSE	SLED, TLDD, EISD, LDSE, DLTD, IESD, ESDL, SDLE, TIDD, ELSD, DLSE, SIED, DDLT
3.16	120.0809	24.09	No match	No Match	No Match
3.16	166.0864	5.40	No match	F	No Match
3.16	103.0544	2.99	No match	No Match	No Match

	Peptide Identification Methodology					
	RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	Database Searching	In silico Digestion	
					Specific	Non-Specific
291	3.19	378.1688	2.86	M/V E	MVE	PFD, DFP, FPD, FDP, DPF, MEV, VEM, MLD, MDL, IMD, DLM, MVE
	3.20	412.1712	5.15	No match	No Match	No Match
	3.20	371.6723	16.56	No match	No Match	No Match
	3.23	398.1555	11.78	No match	TDY, SEY	No Match
	3.24	410.1919	4.59	No match	EVY, DLY, DIY	DIY, EYV, DYL, DLY, YEV, LDY, VYE, EVY, YDL
	3.25	390.1870	10.94	No match	LEE, IEE, ELE, EIE, EEL	HPH, EEL, LEE, EEI, IEE, ELE, EIE
	3.33	447.2083	16.35	No match	No Match	ELDA, ELGE, EALD, VEAE, ELEG, LDEA, EVEA, ADIE, DIAE, EEVA, EVAE, EAID, GLEE, EELG
	3.41	467.2140	27.03	Y V E G	GDLY, DGLY	DGLY, EAFT, GDLY
	3.51	377.6903	6.81	No match	No Match	No Match
	3.53	362.1561	25.43	D/I D	DDL	DID, IDD, DDL, LDD, VED, DVE, EVD, DDI, DLD, EDV
	3.53	229.1185	8.09	No match	No Match	No Match
	3.55	410.1916	3.49	No match	EVY, DLY, DIY	DIY, EYV, DYL, DLY, YEV, LDY, VYE, EVY, YDL
	3.55	375.1872	6.08	No match	QVE, QDL, NEL, LNE, LGDA, ENL, DQL, DIQ, ADVA	DIAG, QID, EIN, LEN, NLE, QVE, ELN, LNE, DQI, INE, ENL, DLQ,

RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	Peptide Identification Methodology		
			Database Searching	<i>In silico</i> Digestion	
				Specific	Non-Specific
					DQL, VQE, LQD, QDL, GIAD, DLAG, NEL, APST, ADVA, IEN, QDI, DIQ, AIDG, EVQ, PAST, ELGG, TASP, LGDA, EQV, IDQ, GAEV, AAVD, DVAA
3.56	473.2232	2.69	P E V E	No Match	LPDE, ELPD, VEPE, LDPE, IEDP
3.68	518.2448	3.53	No match	No Match	AELGE, LEEAG, AELEG, ELEGA, ADIAE, LEQE, QLEE, LAADE, GLEEA, VGDDL, EELGA
3.70	348.1765	3.80	S E I	VTE, TDL, SLE, SIE, SEL, LSE, ISE	No Match
3.70	247.1290	2.66	No match	VE, DL	VE, DI, EV, ID, LD, DL
3.72	447.2078	3.40	No match	No Match	ELDA, ELGE, EALD, VEAE, ELEG, LDEA, EVEA, ADIE, DIAE, EEVA, EVAE, EAID, GLEE, EELG
3.82	384.1398	7.03	No match	DSY	DSY, YDS, SYD
3.84	461.2240	17.49	I A E E	VDDL, IEAE, EIEA, AELE	LEEA, ELEA, EEAL, ALEE, AEIE, VDDI, DDIV, DLVD, DDLV, IAEE, ELEA, VEVD, EALE

	Peptide Identification Methodology					
	RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	Database Searching	<i>In silico</i> Digestion	
					Specific	Non-Specific
293	3.89	538.2137	5.98	F/E N E	No Match	FENE, KACNC
	3.96	504.2291	3.81	I G D A E	No Match	LEDQ, EQID, QVEE, EEIN, EDQI, NEEL, EDQL, AEADV, EADVA, QLED, ELQD, EAIDG, GEAID, LGDAE, EEQV, VVGDD
	4.07	447.2087	23.84	I/E/G E	VEAE, EVAE, ELGE, ELDA, EEVA, DIAE	No Match
	4.15	362.1561	19.66	No match	DDL	DID, IDD, DDL, LDD, VED, DVE, EVD, DDI, DLD, EDV
	4.17	449.1878	13.74	D E S V	ETAE	No Match
	4.18	544.2605	10.79	V/A/P/E E	No Match	No Match
	4.21	395.1557	3.97	No match	No Match	NFD, DFN
	4.23	458.7035	7.96	No match	No Match	No Match
	4.34	518.2449	7.38	E I/Q E	No Match	AELGE, LEEAG, AELEG, ELEGA, ADIAE, LEQE, QLEE, LAADE, GLEEA, VGDDL, EELGA
	4.35	376.1712	8.56	No match	VEE, DIE, DEL	IED, LED, ELD, EID, VEE, EEV, DIE, DLE, EVE, LDE, EDL, DEL
	4.41	447.2086	32.25	I G E E	VEAE, EVAE, ELGE, ELDA, EEVA, DIAE	No Match

	Peptide Identification Methodology					
	RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	Database Searching	<i>In silico</i> Digestion	
					Specific	Non-Specific
294	4.47	490.2134	7.89	I D D Q	No Match	AEEAA, DDLQ, DLEN, LEND, DLNE, VGDEA, ELDN, VAGDE, VNEE, VDQE
	4.49	313.1218	3.15	No match	MY	MY, YM
	4.63	267.1036	5.14	No match	No Match	No Match
	4.72	520.2240	7.71	No match	No Match	GSLED, DGTEV, EEDK, DLEGS, EKDE, VDASE, DKEE, DESVA, DSVAE, KDEE, DEEK, DADLS, AETAE, GDDLT
	4.76	534.2392	3.63	No match	No Match	PCPAF, LDATD, TDDAI, DDAIT, TEVAD, EGESL, EGSLE, EEKE, ELSEG, TLDGE, ADLSE, DAETV, E, ELDGT, LDGTE
	4.79	277.1189	19.43	No match	No Match	No Match
	4.79	295.1292	8.41	No match	FE, EF	EF, FE
	4.82	348.1766	4.52	No match	VTE, TDL, SLE, SIE, SEL, LSE, ISE	No Match
	4.85	447.2080	11.69	No match	No Match	ELDA, ELGE, EALD, VEAE, ELEG, LDEA, EVEA, ADIE, DIAE, EEVA, EVAE, EAID, GLEE, EELG
	4.99	376.1717	20.25	D I E	No Match	No Match

				Peptide Identification Methodology		
	RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	Database Searching	<i>In silico</i> Digestion	
					Specific	Non-Specific
295	5.13	504.2296	17.73	G/E A I\ D	QVEE, LGDAE, EDQL, EADVA	No Match
	5.15	362.1918	3.87	No match	TIE, TEL, LTE, ETL	IET, ELT, LTE, TLE, ETI, TIE, TEL, LET, ETL, EIT, ITE
	5.26	497.2234	4.34	V/Y E S	No Match	YVTD, TVDY, FETT
	5.31	345.1445	72.70	No match	No Match	No Match
	5.39	316.2119	2.69	No match	No Match	No Match
	5.40	376.1715	17.77	No match	No Match	No Match
	5.40	358.1606	3.28	No match	No Match	No Match
	5.40	295.1656	4.47	No match	No Match	No Match
	5.47	132.1022	4.55	No match	L	I, L
	5.47	231.0613	4.96	No match	No Match	No Match
	5.48	362.1557	69.23	No match	DDL	DID, IDD, DDL, LDD, VED, DVE, EVD, DDI, DLD, EDV
	5.49	489.2548	3.31	V I E/E	No Match	QWR, RQW, EIVE, VLEE, DLEL, VEEL, LVEE, DELI, EILD, LELED
	5.61	453.1974	3.31	F/T A/D	No Match	YVGD, FTAD, FTGE, FGET, FSAE, ASEF
	5.69	363.6750	8.17	No match	No Match	No Match
	5.76	390.1868	72.98	E E I	LEE, IEE, ELE, EIE, EEL	HPH, EEL, LEE, EEI, IEE, ELE, EIE
	5.76	372.1765	11.12	No match	No Match	No Match
	5.79	188.0712	17.20	No match	No Match	No Match
	5.84	447.2079	3.85	No match	No Match	ELDA, ELGE, EALD, VEAE, ELEG, LDEA,

	Peptide Identification Methodology					
	RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	Database Searching	<i>In silico</i> Digestion	
					Specific	Non-Specific
296	6.03	380.6852	45.47	No match	No Match	EVEA, ADIE, DIAE, EEVA, EVAE, EAID, GLEE, EELG
	6.13	410.1916	5.03	No match	EVY, DLY, DIY	No Match DIY, EYV, DYL, DLY, YEV, LDY, VYE, EVY, YDL
	6.14	267.1037	5.15	No match	No Match	No Match
	6.18	424.2073	84.81	I/Y E	No Match	ELY, IYE, YEL, LYE, YLE
	6.23	217.1050	3.96	No match	No Match	No Match
	6.25	281.1135	11.45	No match	No Match	No Match
	6.26	605.2763	3.81	No match	No Match	RCPET, NVMNQ, VAAQNC, ATDDAI, KDIDD, EEKEA, NLTEE, SLEQE, TLEDQ, DLTEQ, VLDSGD, EEAEK, AELSEG, ATLDGE, LAADES, LGDAET
	6.30	295.1655	6.15	No match	No Match	No Match
	6.39	409.1712	4.29	No match	No Match	MCR, GEFQ, ENF, FAGD, FEN, NFE, FDAG, NEF, EGGF, FDQ
	6.47	382.1605	4.25	No match	SEF, EAY, DTF, AYE	EAY, EFS, AYE, FDT, SEF, DTF
	6.47	518.2453	13.75	A D A E	No Match	AELGE, LEEAG, AELEG, ELEGA, ADIAE, LEQE, QLEE, LAADE, GLEEA, VGDDL, EELGA

RT (Min)	Precursor Ion		Peptide Identification Methodology		
			Peak Vol. (%)	Database Searching	<i>In silico</i> Digestion
					Specific
6.57	605.2765	11.75	I D/G A E T	No Match	No Match
6.64	245.0769	5.78	No match	No Match	No Match
6.64	376.1713	45.37	D\E I	VEE, DIE, DEL	IED, LED, ELD, EID, VEE, EEV, DIE, DLE, EVE, LDE, EDL, DEL
6.75	410.1916	8.44	I Y D	EVY, DLY, DIY	DIY, EYV, DYL, DLY, YEV, LDY, VYE, EVY, YDL
6.78	504.2299	28.32	No match	QVEE, LGDAE, EDQL, EADVA	No Match
6.79	477.2184	9.34	D I E T	No Match	EISE, TLED, DLTE, VETE, LETD, SEEL, ETEV, TEVE, DIET, EEIS, LEES, ETDL, ELSE, SELE, ISEE, TDEL
6.86	424.2072	3.45	No match	No Match	ELY, IYE, YEL, LYE, YLE
6.87	463.2028	4.33	D I T\D	SDLE, DLSE	SLED, TLDD, EISD, LDSE, DLTD, IESD, ESDL, SDLE, TIDD, ELSD, DLSE, SIED, DDLT
6.90	534.2399	6.62	D T G E I	No Match	No Match
7.14	461.2238	7.84	A/E E I	VDDL, LEEA, IEAE, EIEA, AELE	VEEA, EIEA, EEAL, ALEE, AEIE, VDDI, DDIV, DLVD, DDLV, IAEE, EIEA, VEVD, EALE
7.14	538.2496	5.09	I Y E N	No Match	EFKD, YLNE, YNEL, GDSLFL, GAEVY

	Peptide Identification Methodology					
	RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	Database Searching	<i>In silico</i> Digestion	
					Specific	Non-Specific
298	7.16	410.1918	9.09	No match	EVY, DLY, DIY	DIY, EYV, DYI, DLY, YEV, LDY, VYE, EVY, YDL
	7.19	477.2184	8.04	No match	No Match	EISE, TLED, DLTE, VETE, LETD, SEEL, ETEV, TEVE, DIET, EEIS, LEES, ETDL, ELSE, SELE, ISEE, TDEL
	7.19	424.2073	3.43	No match	No Match	ELY, IYE, YEL, LYE, YLE
	7.38	382.1606	11.07	No match	SEF, EAY, DTF, AYE	EAY, EFS, AYE, FDT, SEF, DTF
	7.44	534.2396	5.25	No match	No Match	No Match
	7.49	396.1769	39.04	No match	ETF	No Match
	7.51	366.1656	5.24	E/A F	FAE, EFA, EAF, AEF	AFE, AEF, EFA, FAE, EAF
	7.53	548.2559	25.32	A D E T	TEVEA, EGTLE	RCNGV, ELDAT, TEVEA, ADIET, ETDAI, EGTLE, VAETE, DIAET, AELSE, AISEE, EEASI
	7.82	349.1796	5.31	No match	No Match	SSR
	7.90	518.2452	8.55	A E D A	No Match	AELGE, LEEAG, AELEG, ELEGA, ADIAE, LEQE, QLEE, LADE, GLEEA, VGDDL, EELGA
	8.00	295.1655	5.61	No match	No Match	No Match
	8.02	473.2233	6.02	I/P\D E	No Match	LPDE, ELPD, VEPE, LDPE, IEDP

RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	Peptide Identification Methodology			
			Database Searching	<i>In silico</i> Digestion		
				Specific	EGDL,	Non-Specific
8.06	433.1925	9.30	No match	GEVE, DADL	EGDL,	No Match
8.09	424.2077	7.21	No match	ELY		No Match
8.21	309.6484	5.18	No match	No Match		No Match
8.43	510.2191	13.87	N V D Y	No Match		MCRT, FTNE, VDYN, ETFN, NVDY, FEQS
8.45	424.1716	24.71	E/E F	No Match		No Match
8.70	366.6698	20.37	No match	No Match		No Match
8.85	437.1928	41.40	No match	No Match		No Match
8.94	619.2919	7.85	A/E G T\E	No Match		EEDKV, EAEGTL, DKIED, AEGTLE, ADIAET, LEQTE, DELDK, EAGITE, VGDDL, GDDLTV
8.98	503.2707	13.40	I E E	IEEL, EEIL		No Match
9.08	530.2447	6.76	E P/D G	No Match		ELPDG, DLDPA
9.09	481.2287	6.62	Y D A	No Match		IYAD, ADIY, DLYA, IYEG, FDIS, YDAI, SIFD, GLYE, DSL, YLEG, AEVY
9.43	554.2446	4.47	S V\E G Y	No Match		GTVDY, EDYK, FSVSD, DKYE, EDKY, EKDY, VSSDF
9.84	511.2396	32.92	I/T D Y	YVTE, ESY	TDYL,	No Match
9.85	461.2236	12.27	No match	VDDL, IEAE, EIEA, AELE	LEEA, ELEA, EEAL,	VDDL, EIEA, IEAE, LEEA, VDVE, AELE, EEAL, ALEE, AEIE, VDDI, DDIV, DLVD, DDLV, IAEE, ELEA, VEVD, EALE

	Peptide Identification Methodology					
	RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	Database Searching	<i>In silico</i> Digestion	
					Specific	Non-Specific
300	9.96	364.1849	10.59	No match	No Match	TFP
	10.00	410.1554	8.70	D E F	No Match	DFE
	10.12	611.2660	8.15	Y D\N S\I	No Match	HNHGF, TVDYN, YDNSL, YVASGD
	10.13	261.1311	2.85	No match	No Match	No Match
	10.32	495.2083	15.63	E\A E F	No Match	EAEF, EFAE
	10.39	475.2397	33.81	I D V E	VEDL, IDDL	No Match
	10.46	658.3033	6.00	E I P/D/G/Q	No Match	ELPDGQ, KSPDDP
	10.49	601.2821	15.55	Q/I P/D E	No Match	QLPDE
	10.66	467.1772	16.13	D G E F	DGEF	MSTE, DGEF, FDAD
	10.73	405.1767	11.63	No match	AEW	AEW, SPAM
	10.73	547.2503	12.49	W/V N/E	No Match	WVNE, EVYH, DPGKM, DRAEG, DGPMK, ADDRA, EENR, CGAPSI
	10.80	475.2393	53.90	No match	No Match	RHY, YRH, ELVD, IDDL, LDDL, VEDL, DIDI, IDLD, EDVI, EVDL
	10.94	424.2075	4.25	No match	ELY	No Match
	11.17	424.2077	7.88	I/E Y	ELY	No Match
	11.34	589.2815	6.69	No match	No Match	RHGGY, RHYGG
	11.39	489.2556	33.11	V/E I/E	VEEL, LVEE	No Match
	11.49	538.2502	19.23	Y N E I	No Match	EFKD, YLNE, YNEL, GDSLF, GAEVY
	11.54	410.1923	24.16	I/D Y	EVY, DLY, DIY	DIY, EYV, DYI, DLY, YEV, LDY, VYE, EVY, YDL
	11.66	428.1483	6.78	No match	No Match	YDM, DMY, DYM

	Peptide Identification Methodology					
	RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	Database Searching	<i>In silico</i> Digestion	
					Specific	Non-Specific
301	11.98	677.2879	7.05	W G DVA\GVA\T	No Match	NALNDM, HGFMW, WGDAGAT, YTHQE
	12.11	408.2126	5.61	I/F E	LEF, EIF	No Match
	12.19	424.2073	5.59	No match	No Match	ELY, IYE, YEL, LYE, YLE
	12.37	410.1919	5.71	No match	EVY, DLY, DIY	DIY, EYV, DYL, DLY, YEV, LDY, VYE, EVY, YDL
	12.38	394.1971	5.47	No match	LDF, IDF, EVF	IFD, FID, IDF, LDF, EVF, LFD, FDI, EFV
	12.77	404.6958	8.63	No match	No Match	No Match
	12.80	564.2330	18.19	M/E D G I	No Match	No Match
	12.80	387.1774	10.74	No match	No Match	No Match
	13.20	511.2392	10.41	No match	No Match	HAFH, EYVT, SIYE, ESIY, TDYL, SYEL, LTDY, YVTE, DLYT
	13.75	424.2074	6.07	No match	No Match	ELY, IYE, YEL, LYE, YLE
	13.91	540.2114	6.70	M E N\F	No Match	QFMD, MENF
	14.30	475.2397	22.53	I/D D I	VEDL, IDDL	No Match
	14.46	356.6747	4.08	No match	No Match	No Match
	14.66	349.6668	9.34	No match	No Match	No Match
	14.73	574.3072	4.65	I V V D\E	No Match	ERRN, ENRR, EDVIV, VEVDL, EALEL
	14.74	643.3296	34.09	V I/D P E\A	No Match	No Match
	14.86	489.2549	5.33	No match	No Match	QWR, RQW, EIVE, VLEE, DLEL, VEEL, LVEE, DELI, EILD, LELED

	Peptide Identification Methodology					
	RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	Database Searching	<i>In silico</i> Digestion	
					Specific	Non-Specific
302	15.09	521.2603	52.85	Y/P E	No Match	PIYE, YPIE, YELP, ETGKS, QTVSS, TVSSQ, STGSVA, SVTGGT
	15.14	546.2763	12.09	I E D G	No Match	MWLP, ELVDA, GVLEE, LEGDL, DELIG, LELDG
	15.31	523.2749	5.63	No match	No Match	VYLE, IVYE, VYEL
	15.34	410.1920	70.70	No match	EVY, DLY, DIY	DIY, EYV, DYL, DLY, YEV, LDY, VYE, EVY, YDL
	15.35	229.1186	8.08	No match	No Match	No Match
	15.43	489.2550	9.10	I E D	No Match	QWR, RQW, EIVE, VLEE, DLEL, VEEL, LVEE, DELI, EILD, LELD
	15.63	503.2710	22.87	I/E E	IEEL, EEIL	No Match
	15.66	475.2390	5.27	No match	No Match	RHY, YRH, ELVD, IDDL, LDDL, VEDL, DIDI, IDLD, EDVI, EVDL
	15.67	424.2076	11.10	Y/E	ELY	No Match
	16.08	492.2430	7.30	No match	No Match	VNPY
	16.17	323.6458	7.89	No match	No Match	No Match
	16.30	546.2761	4.08	I G I/D E	No Match	MWLP, ELVDA, GVLEE, LEGDL, DELIG, LELDG
	16.39	532.2610	16.86	I/D D G I	DDLVA	No Match
	16.94	745.3716	8.07	I Q V V G D D	No Match	No Match
	16.97	511.2398	31.81	T D Y	YVTE, ESY	TDYL, No Match

	Peptide Identification Methodology					
	RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	Database Searching	<i>In silico</i> Digestion	
					Specific	Non-Specific
303	17.17	410.1918	7.14	No match	EVY, DLY, DIY	DIY, EYV, DYI, DLY, YEV, LDY, VYE, EVY, YDL
	17.20	617.3130	5.62	V/V D G D\I	No Match	PGKMQG, THGRF, VAELGE, LEDQI, VLEQE, LEDQL, QVEEL, EDQII, QLVEE, ELQDI, VVGDDL
	17.44	419.1922	7.07	No match	No Match	RDE, DER, SQGQ, ERD, ASQN, AGTNG, AQSN
	17.66	536.2510	7.91	No match	No Match	No Match
	18.05	589.2818	5.42	G D E/L G V	No Match	RHGGY, RHYGG
	18.17	458.1919	13.20	Y E F	No Match	EFY
	18.70	523.2391	7.51	E F D	No Match	TAASSS, EIFD, LDFE
	18.79	590.3022	10.32	I E T/D\I	No Match	QWRT, RNGLM, QLCIN, QGVMVG, NLNVM, LNVMN, SILEE, DLELT, LEEIS, ETELV, LETDL, LETEV, ISEEL, TDELI
	18.86	576.2864	10.73	I T D D I	No Match	KTGAPC, TELVD, TLDDL, EISDL, ETDLV, IESD, IESDL, TIDDL, ELSDI, EILDS
	18.87	620.3279	4.40	V/P\ Y\E	No Match	VPIYE, PIVYE, EKEKS, KEKSE, AEKTAT, TTTQGL
	18.89	503.2710	61.58	E E I	IEEL, EEIL	No Match
	19.11	522.2450	21.23	No match	No Match	FRCP, STLDS, SSSLE, MQML, SVTES

	Peptide Identification Methodology					
	RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	Database Searching	<i>In silico</i> Digestion	
					Specific	Non-Specific
	19.68	408.2127	4.49	I/E F	LEF, EIF	No Match
	19.86	704.3448	3.56	L/G I D S A\E	No Match	STHGRF, PGKMQGS, NLNVMN, EEEVKA, IAEKDE, ADSVAEL, SVAELGE, QTLDDL, LEDQLS, SLEDQI, TLDDLQ, DQLSEL, QEISDL, ETDLVQ, EADVASL, EGGLEE, LGDAETV
304	20.26	394.1971	5.36	No match	LDF, IDF, EVF	IFD, FID, IDF, LDF, EVF, LFD, FDI, EFV
	20.53	450.2306	31.57	No match	No Match	FVAN, AFNV
	20.86	489.2554	20.10	No match	No Match	QWR, RQW, EIVE, VLEE, DLEL, VEEL, LVEE, DELI, EILD, LELED
	21.02	408.2060	6.42	No match	No Match	No Match
	21.04	647.3236	8.27	No match	No Match	RSRND, LPCPAF, TELVDA, DDAITI, LEGSLE, LEEKE, EEELK, EEEIK, ELSDIA, TDELIG, LELDGT
	21.14	618.2868	5.31	W D Q G I	No Match	EHLGY, AEVYH, RNAEE, ADPGKM, DRAEGA, DRAQE, ADGPMK, TNCLAP
	21.18	693.3080	10.02	E Y I P D G	No Match	LTEEMA, GETTTQG, YELPDG, DLDPAY

305

RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	Peptide Identification Methodology		
			Database Searching	<i>In silico</i> Digestion	
				Specific	Non-Specific
21.22	476.2135	7.91	No match	No Match	DERG, GASQN, ADDR, DDRA, AQSNG
21.27	821.3664	17.29	E\Y L P D G\Q	No Match	YELPDGQ, LNVMNQC, DKIEDMA, SIMDLEN
21.27	408.2125	7.46	No match	LEF, EIF	No Match
21.35	352.1658	11.75	No match	No Match	WF
21.61	734.3707	5.17	A A A P/L/Y/E	No Match	RSTHPH, EIAMATV, KTKNDE, SKANSEV, EDKSKQ, SQKEDK, PETLFQ, ETLFQP, NVPIYE, KPEFVD, PEATGFI
22.67	447.2233	4.58	No match	No Match	WLE, AGGATA, GGATAA, NQSV, NGSGL, GQKD, VTNN, NDAK, INGGS
23.54	241.0688	10.93	No match	No Match	No Match
23.67	447.2233	5.11	No match	No Match	WLE, AGGATA, GGATAA, NQSV, NGSGL, GQKD, VTNN, NDAK, INGGS
23.83	603.2973	4.52	I D D\Q I	No Match	RMEAP, LDDLQ, DIVQE, TGEISP
24.20	352.1657	12.64	No match	No Match	WF
24.38	415.7057	8.08	No match	No Match	No Match
24.58	417.2491	6.29	I/V W	No Match	IVW, VWI, LVW, WVI, LER, ELR, ERI, RIE, KVG N, ERL, RLE, REL,

RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	Peptide Identification Methodology			
			Database Searching	<i>In silico</i> Digestion		
				Specific	Non-Specific	
306	24.96	701.3704	8.00	I E I/P T/E	No Match	LRE, REI, GAAKA, KAAQ, GKNV, KAGAA NCLAPLA, RNVQQQ, MKLYF
	25.51	557.2601	9.76	E Y V F	No Match	PSADAP, IFDY, YEVF
	25.68	624.3233	11.18	I/T D Y/L	No Match	LTDYL, SIYLE
	26.05	742.3969	7.81	L V/P E L D G	No Match	No Match
	26.29	428.1813	11.49	No match	No Match	No Match
	26.44	442.1969	6.84	No match	No Match	YYP
	26.54	415.7061	49.54	No match	No Match	No Match
	26.99	530.2602	11.94	No match	No Match	PERE, GKYY, EPVW
	28.23	687.3195	8.82	No match	No Match	No Match
	28.86	439.1971	4.48	No match	No Match	GLFC
	31.17	571.2756	10.65	I Y/E F	No Match	YLEF
	31.34	674.3500	13.57	L G E V W/A	No Match	PAMIKD, EGIVWA, YHIIE, ERADIA, RADIAE, RLEEAG, AEAQKQ, RQLEE, ANNVLSG, ALDRAE, EREAGI
	31.92	603.3136	47.84	E G I V W	EGIVW	No Match
	32.10	948.4657	8.45	V L/D/P E A T G F	No Match	EQIGETGKS, KTGAPCRSE, VTFQLPDE, VLDPEATGF, EELEAMVK, ELEAMVKE
	37.28	488.2286	7.04	No match	No Match	HFQG, WPW

RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	Peptide Identification Methodology		
			Database Searching	<i>In silico</i> Digestion	
				Specific	Non-Specific
37.34	589.2769	38.65	T W P/W	TWPW	AELEGA, LQAEE, QVDDL, LQEAE, QAELE, DDIVQ, QDLVD, VVDQE

Table A20. Compound list generated using the Find by algorithm with Molecular Feature Extraction using Agilent MassHunter Software for fraction 7 from SAX Separation 5. The sequences of the detected ions were compared using the peptides identified by database searching and of predictions from *in silico* digestions with non-specific and specific hydrolysis of the curated protein library, outlined in Section 3.15. No Match indicates the ion was not identified by the indicated peptide identification methodology. Database search hits are reported to include the detected product ions, where / indicates the identification of the b-series ion, \ indicates the identification of the y-series ion, and | indicates the identification of both b- and y-series ions.

RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	Peptide Identification Methodology		
			Database Searching	<i>In silico</i> Digestion	
				Specific	Non-Specific
2.61	390.1871	19.89	I/E E	LEE, IEE, ELE, EIE, EEL	HPH, EEL, LEE, EEI, IEE, ELE, EIE
2.62	412.1686	1.67	No match	No Match	ETY, YET, YTE
2.84	439.1820	7.26	No match	QEY, EGYA	EAYG, EGYA, QEY
2.87	491.2337	2.79	No match	No Match	THPH, LTEE, TEEL, EETI, ETIE, ETEL, ELET, LETE

RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	Peptide Identification Methodology			
			Database Searching	<i>In silico</i> Digestion		
				Specific	Non-Specific	
2.88	505.2132	7.03	I/E D E	No Match	EELD, LEDE, IEDE, DLEE, EEEV, VEEE, EDEL	
2.95	173.0786	4.61	No match	No Match	No Match	
2.97	519.2289	5.24	I/E E E	No Match	LEEE, EEEI, EELE, ELEE, EEIE, EIEE, EEEL, IEE	
2.97	384.1397	5.04	No match	DSY	DSY, YDS, SYD	
3.00	463.2030	25.19	E/I S/D	SDLE, DLSE	SLED, TLDD, EISD, LDSE, DLTD, IESD, ESDL, SDLE, TIDD, ELSD, DLSE, SIED, DDLT	
3.01	554.2082	2.27	No match	No Match	SMSTE, ENYE, YDEAG, QEYD	
3.02	302.1963	4.95	No match	No Match	No Match	
3.08	412.1709	3.16	E/T Y	No Match	No Match	
3.10	166.0862	3.23	No match	F	No Match	
3.11	120.0808	13.47	No match	No Match	No Match	
3.13	197.1285	2.28	No match	No Match	No Match	
3.13	398.1554	1.37	No match	No Match	SEY, SYE, TDY, YSE	
3.15	390.1867	5.04	No match	LEE, IEE, ELE, EIE, EEL	HPH, EEL, LEE, EEI, IEE, ELE, EIE	
3.17	447.2078	4.36	I/G E E	No Match	ELDA, ELGE, EALD, VEA, ELEG, LDEA, EVEA, ADIE, DIAE, EEVA, EVAE, EAID, GLEE, EELG	
3.20	377.6900	2.59	No match	No Match	No Match	
3.24	621.2346	3.65	No match	No Match	QTEED, TDAEAD, AADESE, ENTEE, QSEEE	

RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	Peptide Identification Methodology			
			Database Searching	<i>In silico</i> Digestion		
				Specific	Non-Specific	
3.29	491.1982	9.79	I E/D/D	EDDL	EIDD, DDLE, EVED, LEDD, EDDL, DELD, DLED	
3.43	362.1558	11.17	No match	DDL	DID, IDD, DDL, LDD, VED, DVE, EVD, DDI, DLD, EDV	
3.43	229.1180	3.71	No match	No Match	No Match	
3.50	519.2295	8.55	I/E E E	LEEE, EIEE, EEEL	ELEE, EEIE,	No Match
3.60	497.1868	2.97	No match	No Match	FPMC, CMFP, YEEG, YDEA, FSDE	
3.61	505.2130	7.56	I E/E/D	No Match	EELD, LEDE, IEDE, DLEE, EEEV, VEEE, EDEL	
3.64	247.1289	2.57	No match	VE, DL	VE, DI, EV, ID, LD, DL	
3.66	440.1659	9.32	No match	EEY	EEY, YEE	
3.66	422.1551	1.33	No match	No Match	DNSS	
3.71	461.2235	4.94	I A E E	VDDL, IEAE, EIEA, AELE	LEEA, ELEA, EEAL,	VDDL, EIEA, IEAE, LEEA, VDVE, AELE, EEAL, ALEE, AEIE, VDDI, DDIV, DLVD, DDLV, IAEE, ELEA, VEVD, EALE
3.73	426.1501	6.24	No match	No Match	EDY, EYD, YDE	
3.74	288.5847	2.30	No match	No Match	No Match	
3.74	538.2141	27.77	F/E N E	No Match	FENE, KACNC	
3.84	509.1539	5.03	No match	No Match	EEEC, DDME	
3.86	267.1343	1.18	No match	TF	TF, FT	
3.92	669.2710	4.30	Q Y T E E	No Match	AAQNCY, YQTEE, SYVGDE	

RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	Peptide Identification Methodology		
			Database Searching	<i>In silico</i> Digestion	
				Specific	Non-Specific
3.96	447.2077	1.94	No match	No Match	ELDA, ELGE, EALD, VEA, ELEG, LDEA, EVEA, ADIE, DIAE, EEVA, EVAE, EAID, GLEE, EELG
3.98	544.2606	2.52	No match	No Match	No Match
4.05	318.1659	1.19	No match	AVE, DIA, DLA, EVA, GEL, IGE, LDA, LGE, VAE, VEA	LEG, ELG, VEA, ADI, LDA, DAI, EVA, DIA, EGI, DLA, ADL, EGL, VAE, LGE, EAV, ALD, IGE, DAL, IAD, AID, IEG, TTP, GEL, GEI, GLE, AEV, AVE
4.06	449.1873	3.65	No match	No Match	AETE, ATEE, AEET, DSEV, DDLS, DESV, EAET, ETAE
4.07	362.1557	9.13	No match	DDL	DID, IDD, DDL, LDD, VED, DVE, EVD, DDI, DLD, EDV
4.19	497.1870	2.09	No match	No Match	FPMC, CMFP, YEEG, YDEA, FSDE
4.20	505.2130	3.43	No match	No Match	EELD, LEDE, IEDE, DLEE, EEEV, VEEE, EDEL
4.21	396.1397	1.52	No match	No Match	No Match
4.21	518.2447	5.13	No match	No Match	AELGE, LEEAG, AELEG, ELEGA, ADIAE, LEQE, QLEE, LADE, GLEEA, VGDDL, EELGA
4.22	513.1816	1.87	No match	No Match	No Match
4.23	412.1349	7.35	No match	No Match	No Match

	Peptide Identification Methodology					
	RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	Database Searching	<i>In silico</i> Digestion	
					Specific	Non-Specific
311	4.26	376.1712	3.70	I D E	VEE, DIE, DEL	IED, LED, ELD, EID, VEE, EEV, DIE, DLE, EVE, LDE, EDL, DEL
	4.27	604.2814	2.26	No match	No Match	ERHY, CATVNP, DGPMKG, EIDDL, IDDLE, DDLEL, EVEDL, LEDDL
	4.33	447.2078	3.50	I G E\ E	No Match	ELDA, ELGE, EALD, VEAE, ELEG, LDEA, EVEA, ADIE, DIAE, EEVA, EVAE, EAID, GLEE, EELG
	4.33	523.1694	2.30	No match	No Match	DEEM
	4.35	548.2189	10.58	I/D D DVA	No Match	GEVED, AEEAE, EAEEA, DDLEG
	4.37	410.1550	2.26	No match	No Match	DFE
	4.37	490.2134	2.12	No match	No Match	AEEAA, DDLQ, DLEN, LEND, DLNE, VGDEA, ELDN, VAGDE, VNEE, VDQE
	4.42	320.1237	2.39	No match	DW	WD, DW
	4.55	598.2342	6.33	No match	No Match	AQNCY, YETDA, SYEEA
	4.55	267.1036	5.21	No match	No Match	No Match
	4.60	332.1816	2.11	No match	LEA, IEA, EIA, EAL, ALE, AEL	IEA, LEA, ELA, AEL, EAL, VDV, LAE, ALE, IAE, VVD, AEI, EAI, EIA, AIE
	4.60	426.1501	4.96	No match	No Match	EDY, EYD, YDE
	4.63	520.2253	3.31	No match	VDASE, KDEE, EKDE, DSVAE, DEEK	No Match

RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	Peptide Identification Methodology		
			Database Searching	<i>In silico</i> Digestion	
				Specific	Non-Specific
4.72	277.1182	1.24	No match	No Match	No Match
4.73	348.1766	1.02	No match	VTE, TDL, SLE, SIE, SEL, LSE, ISE	No Match
4.73	592.2446	2.66	No match	No Match	DLEES, SEELD, ESDLE
4.76	447.2079	3.37	V A E E	No Match	ELDA, ELGE, EALD, VEAЕ, ELEG, LDEA, EVEA, ADIE, DIAE, EEVA, EVAE, EAID, GLEE, EELG
4.92	376.1714	10.92	D I E	VEE, DIE, DEL	IED, LED, ELD, EID, VEE, EEV, DIE, DLE, EVE, LDE, EDL, DEL
4.96	263.1404	1.18	No match	No Match	MI, ML, LM, IM
5.00	505.2136	17.96	No match	VEEE, EDEL	No Match
5.02	511.2027	1.51	No match	No Match	YEEA, EDTF
5.06	530.0930	2.84	No match	No Match	No Match
5.07	477.1824	31.71	D I D D	No Match	DIDD, EAEE, EEAE, EVDD, DDLD, EEEA
5.07	362.1916	2.21	No match	TIE, TEL, LTE, ETL	IET, ELT, LTE, TLE, ETI, TIE, TEL, LET, ETL, EIT, ITE
5.07	258.0690	1.93	No match	No Match	No Match
5.13	663.2816	5.61	L/T/E/D G E	No Match	ELDATD, GTEDEL, GETDEL, EE, ELDGTE
5.16	647.2867	1.43	No match	No Match	GEVDDI, VDDLEG, LEEAEG, LEQEE, IAEEAD, EEELGA
5.17	334.1397	3.15	No match	EW	EW

	Peptide Identification Methodology					
	RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	<i>In silico</i> Digestion		
				Database Searching	Specific	Non-Specific
313	5.18	434.1666	1.76	No match	No Match	CGAPS, NWD, WND
	5.24	345.1450	31.35	No match	No Match	No Match
	5.30	316.2118	3.96	No match	No Match	No Match
	5.32	376.1712	7.15	No match	VEE, DIE, DEL	IED, LED, ELD, EID, VEE, EEV, DIE, DLE, EVE, LDE, EDL, DEL
	5.34	295.1654	2.04	No match	YL, LY, IY	IY, YL, LY, YI
	5.40	132.1020	2.07	No match	L	I, L
	5.40	362.1562	29.01	No match	DDL	DID, IDD, DDL, LDD, VED, DVE, EVD, DDI, DLD, EDV
	5.40	200.5555	1.48	No match	No Match	No Match
	5.45	292.1289	2.62	No match	No Match	WS
	5.46	590.2658	5.64	A I E E/E	No Match	CPSNLG, DYRH, GAPCRS, DDDL, VDDLE, EEIEA, EIEAE, ELEE, LEEAE, AEIEE, EVDDI, EDDLV, EELEA
	5.46	410.1556	13.99	F/D E	DFE	No Match
	5.57	424.1715	29.32	F E E	No Match	No Match
	5.69	591.2245	8.94	G G/D\D\L D	No Match	QAEED, GGDDL
	5.70	519.2290	14.94	E/I E E	No Match	LEEE, EEEI, EELE, ELEE, EEIE, EIEE, EEEL, IEE
	5.70	390.1873	24.32	E E I	LEE, IEE, ELE, EIE, EEL	HPH, EEL, LEE, EEI, IEE, ELE, EIE
	5.70	372.1762	3.70	No match	No Match	No Match
	5.74	188.0713	22.88	No match	No Match	No Match
	5.75	205.0973	2.79	No match	W	W
	5.75	146.0599	1.38	No match	No Match	No Match

RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	Peptide Identification Methodology		
			Database Searching	<i>In silico</i> Digestion	
				Specific	Non-Specific
5.89	424.1710	6.33	No match	FEE, EEF	EFE, FEE, EEF
5.90	663.2817	1.82	No match	No Match	ELDATD, GTEDEL, GETDEL, EE, ELDGTE
6.10	424.2075	15.72	I/Y E	ELY	No Match
6.20	481.1537	2.86	No match	No Match	No Match
6.21	512.0827	4.10	No match	No Match	No Match
6.21	249.0638	4.64	No match	No Match	No Match
6.26	295.1655	1.64	No match	No Match	No Match
6.28	442.2113	21.34	No match	No Match	No Match
6.29	448.1822	3.14	No match	No Match	TGAPC, WNE, WGDA, EWN
6.48	693.2378	2.62	No match	No Match	No Match
6.52	605.2763	10.45	Q V T E E	No Match	RCPET, NVMNQ, VAAQNC, ATDDAI, KDIDD, EEKEA, NLTEE, SLEQE, TLEDQ, DLTEQ, VLDSGD, EEAEK, AELSEG, ATLDGE, LAADES, LGDAET
6.54	332.1242	5.02	No match	No Match	No Match
6.57	712.2769	2.81	No match	No Match	NSYEEA, FPMCSQ
6.60	376.1714	15.77	D E I	VEE, DIE, DEL	IED, LED, ELD, EID, VEE, EEV, DIE, DLE, EVE, LDE, EDL, DEL
6.60	245.0768	1.90	No match	No Match	No Match
6.65	527.1976	4.17	No match	No Match	YETD, SYEE
6.67	491.1977	6.24	No match	EDDL	EIDD, DDLE, EVED, LEDD, EDDL, DELD, DLED

	Peptide Identification Methodology					
	RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	<i>In silico</i> Digestion		
				Database Searching	Specific	Non-Specific
315	6.69	374.2281	1.56	No match	No Match	LLE, IEL, ELI, LIE, ILE, LEL, IIE, EIL, ELL
	6.74	504.2295	23.24	A G D E I	QVEE, LGDAE, EDQL, EADVA	No Match
	6.74	477.2182	4.26	No match	No Match	EISE, TLED, DLTE, VETE, LETD, SEEL, ETEV, TEVE, DIET, EEIS, LEES, ETDL, ELSE, SELE, ISEE, TDEL
	6.83	693.2379	2.98	M/P S D E D	No Match	No Match
	6.85	534.2396	2.97	D T G E I	No Match	No Match
	6.92	582.2389	1.57	No match	No Match	YVGDE, FSAEE
	6.97	491.1976	2.71	No match	EDDL	EIDD, DDLE, EVED, LEDD, EDDL, DELD, DLED
	7.02	382.1665	4.06	No match	No Match	No Match
	7.05	434.1668	3.83	No match	No Match	CGAPS, NWD, WND
	7.06	568.2237	4.69	No match	No Match	QEEY, FTADD, FGETD
	7.16	477.2181	1.78	No match	No Match	EISE, TLED, DLTE, VETE, LETD, SEEL, ETEV, TEVE, DIET, EEIS, LEES, ETDL, ELSE, SELE, ISEE, TDEL
	7.25	396.1399	1.53	No match	No Match	No Match
	7.31	366.1656	1.96	No match	FAE, EFA, EAF,	AFE, AEF, EFA, FAE, EAF
	7.34	446.1663	2.80	No match	No Match	No Match
	7.40	534.2393	1.94	No match	No Match	PCPAF, LDATD, TDDAI, DDAIT, TEVAD, EGESL, EGSLE, EEKE, ELSEG,

	Peptide Identification Methodology					
	RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	Database Searching	<i>In silico</i> Digestion	
					Specific	Non-Specific
316	7.44	396.1762	6.67	No match	No Match	TLDGE, ADLSE, DAETV, E, ELDGT, LDGTE
	7.52	590.2658	1.86	No match	No Match	VDY, VYD, FET, ETF CPSNLG, DYRH, GAPCRS, DDDL, VDDLE, EEIEA, EIEAE, ELEE, LEEAE, AEIEE, EVDDI, EDDLV, EELEA
	7.56	211.1441	1.46	No match	No Match	No Match
	7.56	334.1399	5.98	No match	EW	EW
	7.56	316.1295	4.09	No match	No Match	PCP
	7.58	435.1869	2.12	No match	No Match	SGGQS, GMDL, ACIE, PSTM, IDGM, DGMI, IGMD, MDVA
	7.61	364.1665	2.68	No match	No Match	No Match
	7.65	316.2120	3.54	No match	No Match	No Match
	7.68	695.2867	2.27	No match	No Match	No Match
	7.68	381.1771	15.62	No match	YAQ	NFT, SQF, FTN, TFN, YAQ, AQY, FQS
	7.70	505.2137	37.96	V E E E	VEEE, EDEL	No Match
	7.70	228.0655	3.55	No match	No Match	No Match
	7.70	272.0847	2.44	No match	No Match	No Match
	7.77	346.1396	5.21	No match	No Match	No Match
	7.79	349.1794	2.54	No match	No Match	No Match
	7.96	295.1656	2.51	No match	No Match	No Match
	8.04	410.1555	13.40	No match	DFE	No Match
	8.05	424.2075	2.37	No match	ELY	No Match

RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	Peptide Identification Methodology		
			Database Searching	<i>In silico</i> Digestion	
				Specific	Non-Specific
8.17	505.2133	22.65	No match	No Match	EELD, LEDE, IEDE, DLEE, EEEV, VEEE, EDEL
8.25	582.2396	5.23	F/T A D E	No Match	No Match
8.31	381.1771	18.83	No match	YAQ	NFT, SQF, FTN, TFN, YAQ, AQY, FQS
8.32	250.1803	4.94	No match	No Match	No Match
8.38	510.2188	6.48	N V D Y	No Match	MCRT, FTNE, VDYN, ETFN, NVDY, FEQS
8.40	406.1603	4.51	No match	No Match	SDNA, SNAD, NGDT
8.41	424.1715	34.40	E E F	No Match	No Match
8.47	364.1846	4.63	No match	No Match	TFP
8.48	519.2293	20.32	E E E I	No Match	LEEE, EEEI, EELE, ELEE, EEIE, EIEE, EEEL, IEEL
8.66	366.6695	11.70	No match	No Match	No Match
8.72	460.1820	4.89	No match	No Match	No Match
8.80	437.1922	3.64	No match	No Match	MVGM
8.94	503.2704	5.19	I I E E	No Match	RDGR, ELIE, ILEE, IEEL, LEEI, EEIL
8.99	437.1916	10.71	No match	No Match	MVGM
8.99	381.1769	1.26	No match	YAQ	NFT, SQF, FTN, TFN, YAQ, AQY, FQS
9.02	530.2445	4.37	L E P D G	No Match	ELPDG, DLDPA
9.09	619.2558	1.54	No match	No Match	EAEAAA, DNEEL, NEELD, DAEADV, QLEDD, EELDN, VAGDEE
9.23	525.2179	2.76	No match	No Match	EFET, DLDY, YDLD
9.56	217.0976	3.68	No match	No Match	No Match
9.56	144.0807	4.15	No match	No Match	No Match

RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	Peptide Identification Methodology		
			Database Searching	<i>In silico</i> Digestion	
				Specific	Non-Specific
9.56	421.1718	1.97	No match	SEW	SEW, CEGE, CIGE, AMEA, GMDV
9.64	748.3340	1.56	No match	No Match	CPSNLGTG, EAEGTLE, DDLEGLS, LEEAEGT, EEAEGTL, EDELDK
9.75	329.1495	1.81	No match	YF, FY	No Match
9.76	419.1919	2.71	No match	No Match	RDE, DER, SQGQ, ERD, ASQN, AGTNG, AQSN
9.80	511.2390	7.00	I/T D Y	No Match	HAFH, EYVT, SIYE, ESIY, TDYL, SYEL, LTDY, YVTE, DLYT
9.81	461.2237	7.54	V/D D I	VDDL, IEAE, EIEA, AELE	LEEAE, ELEE, EEAL, AEIE, VDDI, DDIV, DLVD, DDLV, IAEE, ELEE, VEVD, EALE
9.84	478.1924	3.91	No match	No Match	AEMQ, AMQE
9.93	736.2767	4.24	P/F/D Q D D	No Match	MDLEND, DPFDQD, PFDQDD
9.96	410.1555	17.65	D E F	DFE	No Match
9.97	245.0770	2.26	No match	No Match	No Match
9.99	320.1241	5.66	No match	DW	WD, DW
10.06	304.1656	2.72	No match	VW	VW, WV
10.13	433.2075	3.45	No match	No Match	LTCP, LPTC
10.14	448.1820	1.98	No match	No Match	TGAPC, WNE, WGDA, EWN
10.28	495.2080	11.47	EVAE F	No Match	EAEF, EFAE
10.34	525.2181	4.11	No match	No Match	EFET, DLDY, YDLL

RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	Peptide Identification Methodology		
			Database Searching	<i>In silico</i> Digestion	
				Specific	Non-Specific
10.34	475.2394	20.21	No match	No Match	RHY, YRH, ELVD, IDDL, LDDL, VEDL, DIDI, IDLD, EDVI, EVDL
10.35	536.2345	13.68	No match	No Match	SQNTS, APFDS, MEEK, IGMES, EEKM, EKME, EMEK, EGLMS, LEGMS, EEMK
10.42	539.2339	1.83	No match	No Match	DIYE, YEEV, DELY, YELD
10.45	601.2817	6.76	No match	No Match	QLPDE
10.60	467.1769	10.94	D/G E F	DGEF	MSTE, DGEF, FDAD
10.62	505.2131	5.79	No match	No Match	EELD, LEDE, IEDE, DLEE, EEEV, VEEE, EDEL
10.68	405.1766	6.94	No match	AEW	AEW, SPAM
10.69	547.2503	4.15	W/V N/E	No Match	WVNE, EVYH, DPGKM, DRAEG, DGPMK, ADDRA, EENR, CGAPSI
10.74	539.2335	2.89	No match	No Match	DIYE, YEEV, DELY, YELD
10.75	475.2392	13.83	No match	No Match	RHY, YRH, ELVD, IDDL, LDDL, VEDL, DIDI, IDLD, EDVI, EVDL
10.91	435.1871	15.00	No match	No Match	SGGQS, GMDL, ACIE, PSTM, IDGM, DGMI, IGMD, MDVA
10.96	553.2493	2.90	I E Y E	No Match	EELY
11.19	448.1821	2.12	No match	No Match	TGAPC, WNE, WGDA, EWN
11.32	464.1767	3.49	No match	No Match	NMEA, NEMA, CVAGD

	Peptide Identification Methodology					
	RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	Database Searching	<i>In silico</i> Digestion	
					Specific	Non-Specific
320	11.34	489.2546	5.72	V E EI	No Match	QWR, RQW, EIVE, VLEE, DLEL, VEEL, LVEE, DELI, EILD, LEED
	11.43	553.2493	1.69	No match	No Match	EELY
	11.53	663.2816	2.64	No match	No Match	ELDATD, GTEDEL, GETDEL, EE, ELDGTE
	11.59	719.3445	1.26	No match	No Match	EQLCIN, TTNCLAP, MEAVAAQ, SILEEE, ELEEIS, LEEISE, ELETEV, LETEVE, ETDELI
	11.74	602.2655	2.07	No match	No Match	No Match
	11.75	346.1971	2.47	No match	VVE, VDL	EVV, VLD, IDV, LVD, DLV, DIV, VID, VVE, VDL, VDI, DVI, DVL, VEV
	11.76	683.2359	2.00	No match	No Match	EMASMD, MASMDE
	11.79	748.3338	1.82	No match	No Match	CPSNLGTG, EAEGTLE, DDLEGSL, LEEAEGT, EEAEGTL, EDELDK
	11.87	360.2129	1.74	No match	VLE, LVE, IVE, IDL	EIV, VLE, ELV, LEV, LID, IVE, LDI, DLL, LDL, IDI, VIE, LVE, VEI, IDL, DII, IID, ILD
	11.90	329.1493	3.96	No match	No Match	YF, FY
	11.97	724.2771	2.99	No match	No Match	No Match
	12.06	655.2555	3.95	D S Y V G D	No Match	No Match
	12.10	663.2820	7.65	No match	No Match	ELDATD, GTEDEL, GETDEL, EE, ELDGTE
	12.11	511.2028	5.18	D V D Y	No Match	YEEA, EDTF

	Peptide Identification Methodology					
	RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	Database Searching	<i>In silico</i> Digestion	
					Specific	Non-Specific
321	12.31	697.2659	6.22	E T E D G F	No Match	EQEEY, FGETDE
	12.34	410.1917	5.45	I/D Y	EVY, DLY, DIY	DIY, EYV, DYL, DLY, YEY, LDY, VYE, EVY, YDL
	12.59	612.2497	2.78	Y/D D SL	No Match	TEDTF
	12.67	589.2819	3.15	No match	No Match	RHGGY, RHYGG
	12.72	404.6954	5.55	No match	No Match	No Match
	12.75	564.2325	7.00	M/E D G I	No Match	No Match
	12.76	387.1771	5.81	No match	No Match	No Match
	12.93	807.3171	2.73	No match	No Match	EGDLNEM, GDLNEME, MEGDLNE
	13.17	460.1710	13.47	D/Y Y	No Match	No Match
	13.45	279.1706	2.96	No match	No Match	LF, IF, FI, FL
	13.65	612.2501	8.96	No match	No Match	TEDTF
	13.71	622.2377	4.67	M E I D D	No Match	PMCSQG, EDPFD, LEEEC, MEIDD
	13.83	561.2653	1.48	No match	No Match	PNMVT, RAAED, QREE, REEQ, QVNNS, EREAG
	13.85	612.2501	9.33	No match	No Match	TEDTF
	13.99	305.1572	2.81	No match	No Match	No Match
	14.05	618.2968	8.90	V E E E I	No Match	LVEEE, VEEEL, LEDEL
	14.13	433.2072	3.01	No match	No Match	EER, ERE, REE, VNNS, NNSV, NGDK, NAQT
	14.15	539.2338	9.98	No match	No Match	DIYE, YEEV, DELY, YELD
	14.18	525.2183	3.43	No match	No Match	EFET, DLDY, YDL
	14.19	368.1609	54.08	No match	No Match	FCV, WY, MAF, YW
	14.20	422.1554	4.16	No match	No Match	DNSS
	14.25	553.2493	4.16	No match	No Match	EELY
	14.27	475.2395	22.76	I/D D I	VEDL, IDDL	No Match

RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	Peptide Identification Methodology		
			Database Searching	<i>In silico</i> Digestion	
				Specific	Non-Specific
14.31	447.2233	9.31	No match	No Match	WLE, AGGATA, GGATAA, NQSV, NGSSL, GQKD, VTNN, NDAK, INGGS
14.34	539.2339	10.06	No match	No Match	DIYE, YEEV, DELY, YELD
14.42	689.3338	5.87	No match	No Match	GRGASQN, AADGPMK, PNMVTAG, EVDDIV, VAEIEE, IEEELG
14.43	740.2906	2.44	No match	No Match	FENEMA
14.44	596.2552	6.79	Y D G E I	No Match	MTKNC, DADIY, DGLYE
14.44	744.2820	5.12	Y D N E F G	No Match	YDNEFG, DNEFGY
14.56	717.3397	5.82	Q Q V D D L	No Match	YSFVTT, QAELEGA, AELEGAQ, QQVDDL, GDQLGDL
14.59	559.2499	5.33	No match	No Match	No Match
14.63	349.6667	4.96	No match	No Match	No Match
14.65	640.2815	1.57	No match	DIETY	DIETY, GCDFAK, RCAQY
14.70	574.3069	1.98	I V V D E	No Match	ERRN, ENRR, EDVIV, VEVDL, EALEL
14.71	643.3285	6.77	V I D P E A	No Match	No Match
14.83	604.2818	17.09	E V E D I	LEDDL	TGAPCR
14.88	640.2817	16.92	I D T E Y	DIETY	DIETY, GCDFAK, RCAQY
14.89	279.1706	2.01	No match	No Match	LF, IF, FI, FL
15.24	612.2499	4.65	Y D D S L	No Match	TEDTF
15.27	318.1810	3.56	No match	LW, IW	WL, LW, IW, WI
15.30	410.1917	7.40	No match	EVY, DLY, DIY	DIY, EYV, DYL, DLY, YEV, LDY, VYE, EVY, YDL
15.33	687.2606	4.27	Y D N E F	No Match	SCYVTD, YDNEF

RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	Peptide Identification Methodology		
			Database Searching	<i>In silico</i> Digestion	
				Specific	Non-Specific
15.36	520.2391	1.75	No match	No Match	NNSVS, ESER, GGSSNV, MDLAA, MEAVA, AAMVE, QIEM, EMQI, MEQL, MEQI, MEIQ, ALVCD, MELQ, MVVDG, LTCPS, DAAIM, AMEAV
15.39	590.3019	1.74	I T I D E	No Match	QWRT, RNGLM, QLCIN, QGVMVG, NLNVM, LNVMN, SILEE, DLELT, LEEIS, ETELV, LETDL, LETEV, ISEEL, TDELI
15.51	391.1973	2.24	V/S W	No Match	SVW, FQP, GFAP, SRE, SER, TRD, RES, ERS, ESR, RSE
15.55	539.2339	15.76	No match	No Match	DIYE, YEEV, DELY, YELD
15.59	503.2699	3.83	I/E E I	No Match	RDGR, ELIE, ILEE, IEEL, LEEI, EEIL
15.68	502.7190	3.98	No match	No Match	No Match
16.26	546.2760	1.62	I G I D\E	No Match	MWLP, ELVDA, GVLEE, LEGDL, DELIG, LEIDG
16.26	668.2584	9.81	Y E P E\M	No Match	DSMQST, DFNGDT
16.35	532.2602	2.98	No match	No Match	RGYH, RHYG, ALEEA, LDDAV, DDLVA, IAEEA
16.40	683.2869	4.17	Y/G E S D L	No Match	DKYEE, EDKYE
16.76	846.3825	13.76	No match	No Match	DDRAGPCI, LEQQVDD, INDNEEL, EQQVDDL, QQVDDLE
16.76	435.6745	5.79	No match	No Match	No Match

	Peptide Identification Methodology					
	RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	<i>In silico</i> Digestion		
				Database Searching	Specific	Non-Specific
324	16.76	450.1511	3.68	No match	No Match	DSGDG
	16.85	433.2076	2.49	No match	No Match	LTCP, LPTC
	16.91	745.3715	2.71	I Q V V G D D	No Match	No Match
	16.94	511.2390	3.43	No match	No Match	HAFH, EYVT, SIYE, ESIY, TDYL, SYEL, LTDY, YVTE, DLYT
	17.13	488.2132	9.63	No match	No Match	WPDA
	17.22	433.2076	3.37	No match	No Match	LTCP, LPTC
	17.30	318.1811	2.87	No match	LW, IW	WL, LW, IW, WI
	17.38	419.1923	17.88	No match	No Match	RDE, DER, SQGQ, ERD, ASQN, AGTNG, AQSN
	17.46	675.3083	2.03	No match	No Match	GAEVYH, DPGKMQ, ADGPMKG
	17.46	360.2132	2.71	No match	VLE, LVE, IVE, IDL	EIV, VLE, ELV, LEV, LID, IVE, LDI, DLL, LDL, IDI, VIE, LVE, VEI, IDL, DII, IID, ILD
	17.63	536.2509	3.89	No match	No Match	No Match
	17.64	349.1835	1.71	No match	No Match	No Match
	17.66	374.2283	1.79	No match	No Match	LLE, IEL, ELI, LIE, ILE, LEL, IIE, EIL, ELL
	17.71	691.3130	2.67	No match	No Match	TGAPCRS, EIDDLs, EAEETI, AEETIE, EETIEA, CNGVLEG, ALVCDNG, TCPSNLG, CPSNLGT, ENCNVL
	17.73	486.2270	2.78	No match	No Match	No Match
	17.76	355.6669	6.28	No match	No Match	No Match

325

RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	Peptide Identification Methodology		
			Database Searching	<i>In silico</i> Digestion	
				Specific	Non-Specific
17.92	632.3124	4.33	I/E E E I	No Match	RGASQN, PLNGMT, PNMVTA, VVCGAPS, ILEEE, IEELE, LEEEI, ELEEI, EIEEL, LEEEL, IEEEL
17.94	604.2815	7.18	No match	LEDDL	TGAPCR
17.98	380.1607	33.45	No match	No Match	No Match
18.03	492.7348	3.85	No match	No Match	No Match
18.08	444.1763	2.04	No match	No Match	YDF, FDY
18.12	458.1918	9.97	Y E F	No Match	EFY
18.28	635.2658	6.65	V D/W T D	No Match	DESER, SGDSAAAG, LNEME, NEMEI, MELQD, EAIDGM
18.36	360.2128	1.62	I/D I	VLE, LVE, IVE, IDL	EIV, VLE, ELV, LEV, LID, IVE, LDI, DLL, LDL, IDI, VIE, LVE, VEI, IDL, DII, IID, ILD
18.36	574.2347	2.43	No match	No Match	DDLDP, DPEDV
18.52	532.2032	5.35	No match	No Match	DWPD
18.67	523.2390	4.61	E F D	No Match	TAASSS, EIFD, LDFE
18.75	590.3021	1.48	I E T/DI	No Match	QWRT, RNGLM, QLCIN, QGVMVG, NLNVM, LNVMN, SILEE, DLELT, LEEIS, ETELV, LETDL, LETEV, ISEEL, TDELI
18.77	374.2284	1.98	I/E I	No Match	LLE, IEL, ELI, LIE, ILE, LEL, IIE, EIL, ELL
18.80	576.2866	5.83	T I D D L	No Match	No Match

	Peptide Identification Methodology					
	RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	<i>In silico</i> Digestion		
				Database Searching	Specific	Non-Specific
326	18.83	503.2704	2.34	No match	No Match	RDGR, ELIE, ILEE, IEEL, LEEI, EEIL
	19.09	509.2239	15.08	F/D D I	DLDF	AASSSS, EFVD, DLDF
	19.26	554.2511	27.03	No match	No Match	No Match
	19.62	648.2972	2.40	No match	No Match	ITNWD, TADDRA, TEENR, EQAAMV, QAAMVE, AENLAM, QMVVDG
	19.63	408.2126	2.63	E I F	LEF, EIF	No Match
	19.68	523.2389	6.58	E F D I	No Match	TAASSS, EIFD, LDFE
	19.75	604.2817	13.45	I E D D I	LEDDL	TGAPCR
	19.78	455.1921	12.19	No match	No Match	CYVA, AAMY, TMFG
	19.85	501.2184	3.77	No match	No Match	No Match
	19.89	559.2386	1.58	No match	No Match	QAEPD
	20.21	394.1969	2.42	I/D F	LDF, IDF, EVF	IFD, FID, IDF, LDF, EVF, LFD, FDI, EFV
	20.26	465.1798	2.24	No match	No Match	EEST, EDTT, DETT, TEDT
	20.30	810.3504	11.09	No match	No Match	NSIMKCD, LEQEEY, DAPMFVM, FGETDEL, EDTFIAD
	20.40	419.1920	4.72	No match	No Match	RDE, DER, SQGQ, ERD, ASQN, AGTNG, AQSN
	20.42	817.3926	6.37	E E I N A E L	No Match	AAADGPMKG, YKQICY, EVDDIVQ, VDDIVQE, EEINAEL
	20.67	500.2133	8.70	No match	No Match	No Match
	20.75	810.3501	3.67	No match	No Match	NSIMKCD, LEQEEY, DAPMFVM, FGETDEL, EDTFIAD

	Peptide Identification Methodology					
	RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	<i>In silico</i> Digestion		
				Database Searching	Specific	Non-Specific
327	20.91	430.2428	1.38	No match	No Match	No Match
	20.95	626.2652	2.13	No match	No Match	EFETT, DLTDY, HDHAF, DHAFH
	21.00	393.2086	1.21	No match	No Match	No Match
	21.00	388.2535	1.17	No match	No Match	No Match
	21.11	370.1641	2.50	No match	No Match	TYS
	21.16	459.7091	17.76	No match	No Match	No Match
	21.16	373.6143	3.29	No match	No Match	No Match
	21.17	693.3081	14.47	E Y P\D G	No Match	LTEEMA, GETTTQG, YELPDG, DLDPAY
	21.17	359.1377	2.33	No match	No Match	No Match
	21.18	476.2136	6.04	No match	No Match	DERG, GASQN, ADDR, DDRA, AQSNG
	21.22	437.6433	2.28	No match	No Match	No Match
	21.23	423.1667	1.92	No match	No Match	No Match
	21.23	821.3662	8.66	E Y L P/D/G Q	No Match	YELPDGQ, LNVMNQC, DKIEDMA, SIMDLEN
	21.23	408.2128	3.35	No match	LEF, EIF	No Match
	21.24	818.4127	2.38	No match	No Match	TWRLNE, NLWAAFP, QVQMVVD, MVKEASGP, IDDLELT, DDLELTL, IIESDLE
	21.32	352.1655	3.22	No match	No Match	WF
	21.49	636.2861	3.32	E Y P D	No Match	YELPD, ACNCLL, GETTTQ, ETTTQG
	21.52	561.2710	6.76	No match	No Match	No Match
	21.64	603.2404	100.00	No match	No Match	DWPDA, GDTHSS
	21.64	314.1037	5.76	No match	No Match	No Match

	Peptide Identification Methodology					
	RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	<i>In silico</i> Digestion		
				Database Searching	Specific	Non-Specific
328	21.64	656.1509	3.05	No match	No Match	No Match
	21.64	130.0650	8.91	No match	No Match	No Match
	21.64	321.0982	4.03	No match	No Match	No Match
	21.71	732.3398	4.71	I Q E D D\L	No Match	DPGKMQG, QLEDDL
	22.00	578.1903	2.30	No match	No Match	No Match
	22.42	855.3175	3.49	No match	No Match	No Match
	22.64	447.2233	10.95	No match	No Match	WLE, AGGATA, GGATAA, NQSV, NGSSL, GQKD, VTNN, NDAK, INGGS
	22.76	433.2077	6.63	No match	No Match	LTCP, LPTC
	22.84	706.3031	8.21	D V W D TV	No Match	DMEVVN, DAEMQI, ADESER, ASGDSAAAG, GAAAEGGSS
	22.88	822.3501	3.13	No match	No Match	SIEDPFD, SILEEEC, MEIDDL, PPPMEHD, FDICHTS, CTTNCLAP, IMENCNV, MENCNVL
	23.00	415.6786	4.46	No match	No Match	No Match
	23.19	681.3071	1.88	No match	No Match	MEAPPH, ATAASSSS, PMFVMG, TLDGEF, LFTADD, DTFIAD
	23.22	546.2913	7.19	I V G G N S	No Match	QEAQA, AQKEA, AEAQK, KEQAA, ERIE, KVGNE, ERLE, RIEE, RELE, RLEE, EELR, QGQIT, NGSLV, EREI, NVLSSG, NNVLS, AQAEK, KAAQE, INAQT,

	Peptide Identification Methodology					
	RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	Database Searching	<i>In silico</i> Digestion	
					Specific	Non-Specific
329						KAGAAE, AGAAEK, VINGGS
	23.30	631.2875	3.32	No match	No Match	No Match
	23.50	241.0688	14.19	No match	No Match	No Match
	23.62	631.2873	4.35	No match	No Match	NDHFV
	23.65	447.2233	5.29	No match	No Match	WLE, AGGATA, GGATAA, NQSV, NGSGL, GQKD, VTNN, NDAK, INGGS
	23.72	695.3231	1.90	No match	No Match	NSIMKC, APMFVM, EDTFIA
	23.92	643.1769	2.86	No match	No Match	No Match
	23.92	590.2666	42.11	D I D D I	VDDLE, EIEAE, EEIEA	No Match
	24.08	570.7689	2.58	No match	No Match	No Match
	24.16	352.1656	1.85	No match	No Match	WF
	24.26	465.7090	15.65	No match	No Match	No Match
	24.35	415.7058	2.51	No match	No Match	No Match
	24.46	684.2864	2.72	No match	No Match	ETAYNS
	24.74	674.2653	3.00	No match	No Match	No Match
	24.74	852.3504	2.19	No match	No Match	No Match
	25.04	447.2231	2.10	No match	No Match	WLE, AGGATA, GGATAA, NQSV, NGSGL, GQKD, VTNN, NDAK, INGGS
	25.27	557.2596	1.38	No match	No Match	PSADAP, IFDY, YEVF
	25.43	447.2232	2.73	No match	No Match	WLE, AGGATA, GGATAA, NQSV, NGSGL, GQKD, VTNN, NDAK, INGGS

330

RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	Peptide Identification Methodology		
			Database Searching	<i>In silico</i> Digestion	
				Specific	Non-Specific
25.46	678.3082	5.92	S S G E L W	No Match	QTAPFD, MARDW, RSGETE, AQESIM, VVDSMQ, TEVVCGA
25.65	591.2761	6.45	E G S I W	No Match	NNSVSA, NGDKAS, MEVFN, AQIEM, AEMQI, QMVVD, QEIAM, AMEAVA, MEAVAA, LAMQE
25.84	504.2444	2.95	No match	No Match	GWLE, AEER, EAER, EERA, REAE, NDKQ, ENKN, AERE, ERAE, EKGGN, NGDKA, EREA
25.90	481.2441	2.12	No match	No Match	PEHV, AFPF, MAFL
26.50	415.7060	18.05	No match	No Match	No Match
26.63	584.2553	6.21	No match	No Match	MFVMG
27.12	525.2286	5.28	No match	No Match	No Match
27.37	603.3127	7.84	V A V W E	No Match	No Match
27.64	542.7563	10.12	No match	No Match	No Match
27.71	638.3018	1.69	No match	No Match	KSSDTT, THPHF, IDLDY, VYELD, EVDLY
27.94	551.6982	2.58	No match	No Match	No Match
27.94	525.2423	9.17	No match	No Match	No Match
28.06	616.2734	10.34	No match	No Match	No Match
28.56	610.7929	12.14	No match	No Match	No Match
28.81	591.2761	1.85	No match	No Match	NNSVSA, NGDKAS, MEVFN, AQIEM, AEMQI, QMVVD, QEIAM, AMEAVA, MEAVAA, LAMQE
28.83	439.1974	19.88	No match	No Match	GLFC

331

RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	Peptide Identification Methodology		
			Database Searching	<i>In silico</i> Digestion	
				Specific	Non-Specific
28.84	274.1186	1.65	No match	No Match	No Match
29.04	433.2077	6.06	No match	No Match	LTCP, LPTC
29.62	693.3230	2.55	No match	No Match	ADGKCTV, YEPWW
29.82	433.2075	1.88	No match	No Match	LTCP, LPTC
30.08	481.2075	4.12	No match	No Match	No Match
30.13	732.3554	19.82	L D A V W E	No Match	EGAQKEA, RLDEAE, EERADI, VASINDN, QVNDISG, VNDISGQ, NESKTPG, EQSLGAQ, EELDRA, ENLKGGD, GKAAQEE, LDEREA, DDLRDV, NVKNEE, EQSQIQ, QSQIQE
30.24	364.1654	3.80	No match	No Match	KNC
30.41	686.3018	4.15	L E F D Y	No Match	EIFDY
30.87	787.3499	10.90	I/D T E Y F	No Match	No Match
30.91	735.3217	2.45	No match	No Match	GMARDW, IEDPFD, DGTKEEG, SEGESLN, ILEEEC, MEIDDL, EVEDLM
31.89	603.3125	4.72	E G I V W	No Match	No Match
34.84	608.2759	15.60	No match	No Match	LSDHH, SMEKGG, CTQTGV
34.96	493.2074	4.14	No match	No Match	ADGKC, ASMGQ
35.74	967.4502	5.34	No match	No Match	VDYNITGW, EQLCINFT, GGTTMYPGIA
36.18	739.3492	2.06	No match	No Match	DLTDYL, EVDLYT
36.53	602.2602	2.53	No match	No Match	No Match

RT (Min)	Precursor Ion Mass (Da)	Peak Vol. (%)	Peptide Identification Methodology		
			Database Searching	<i>In silico</i> Digestion	
				Specific	Non-Specific
36.87	1007.4486	4.25	No match	No Match	IITNWDDM
37.32	589.2759	4.66	T W P/W	No Match	TWPW
39.01	670.3068	2.40	No match	No Match	FEIFD

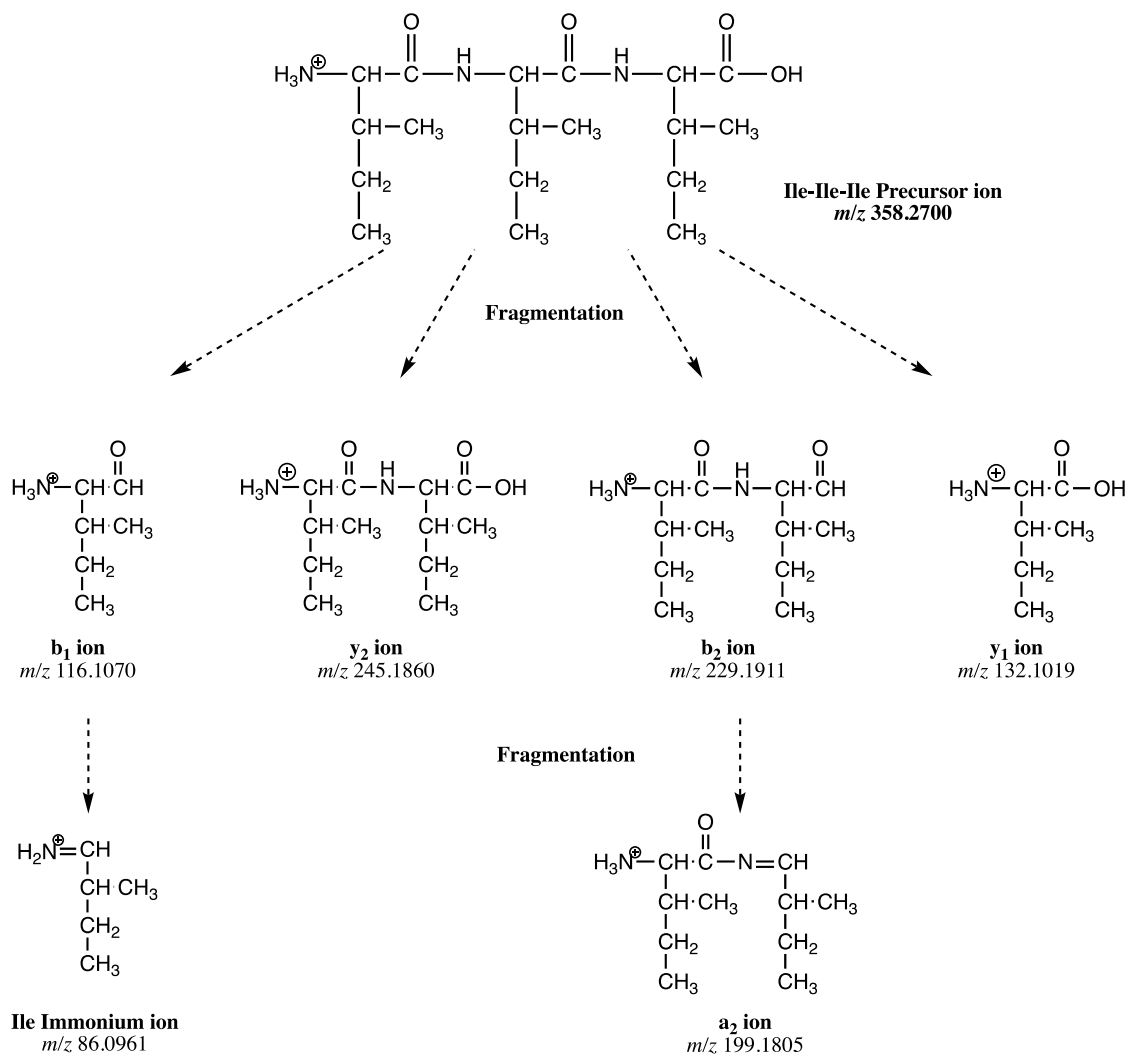


Figure A1. Simulated fragmentation of the Ile-Ile-Ile peptide during LC-MS analysis to facilitate its *de novo* sequencing through their comparison to the detected product ions in its MS/MS spectrum.

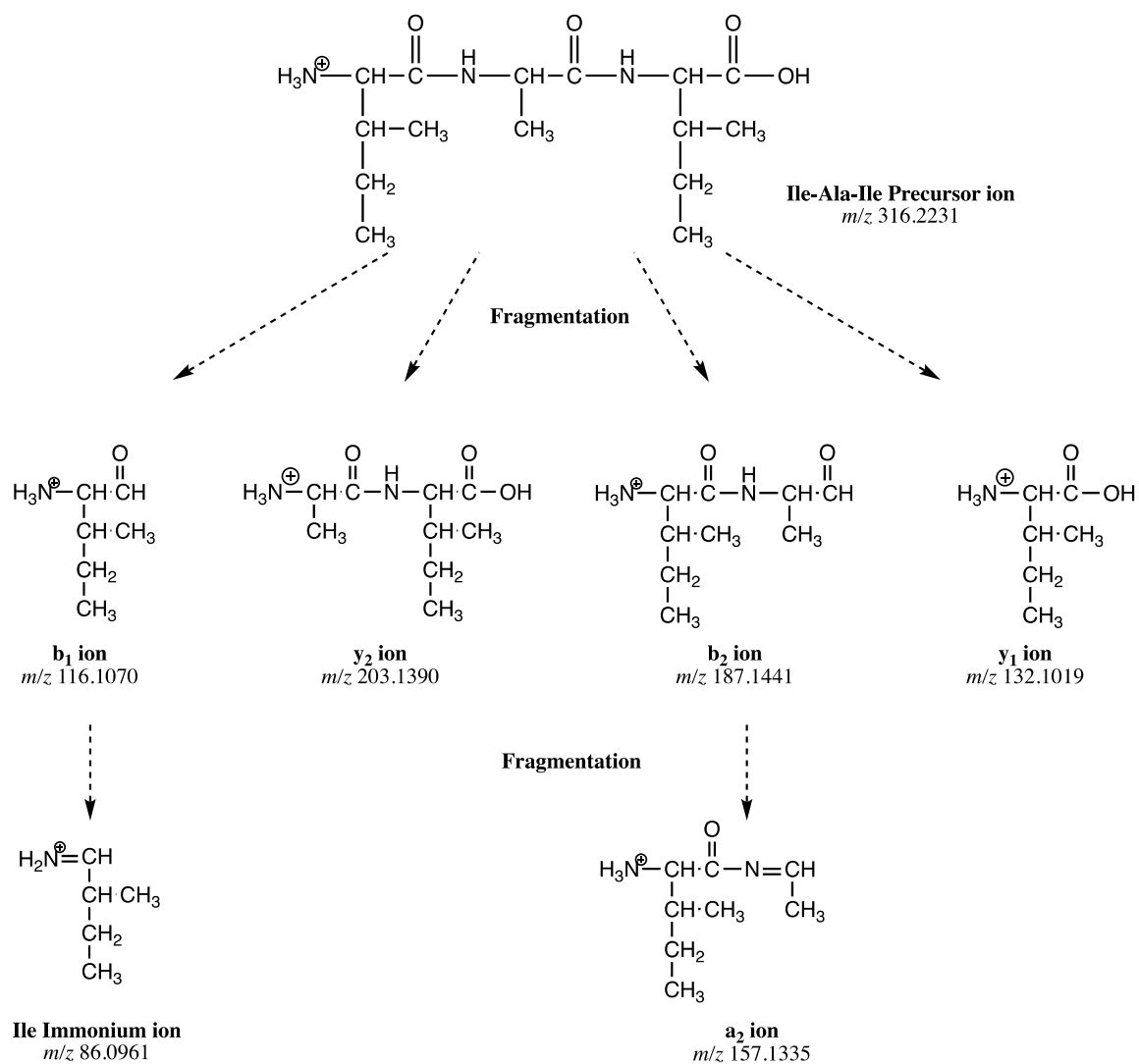


Figure A2. Simulated fragmentation of the Ile-Ala-Ile peptide during LC-MS analysis to facilitate its *de novo* sequencing through their comparison to the detected product ions in its MS/MS spectrum.

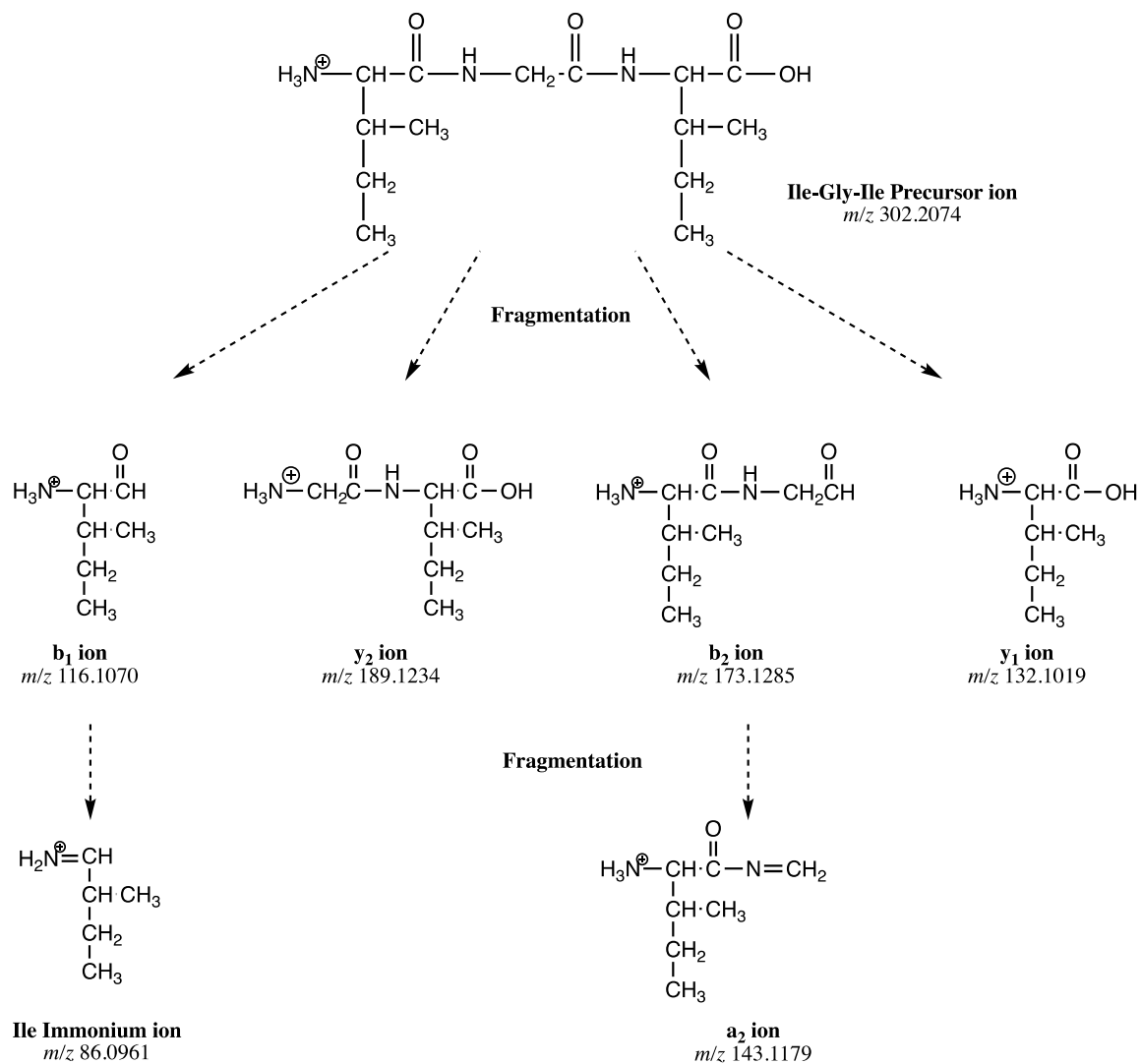


Figure A3. Simulated fragmentation of the Ile-Gly-Ile peptide during LC-MS analysis to facilitate its *de novo* sequencing through their comparison to the detected product ions in its MS/MS spectrum.

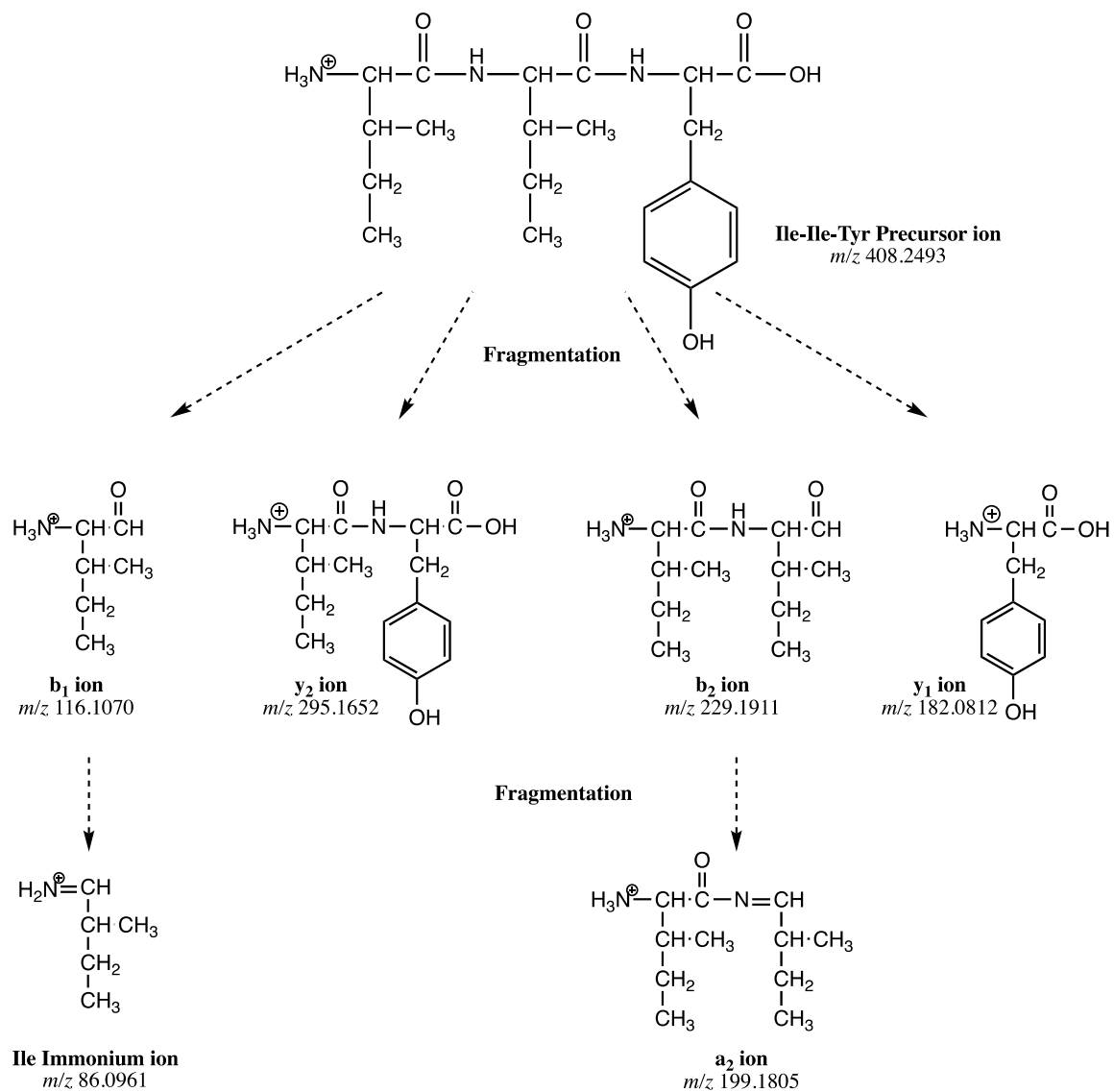


Figure A4. Simulated fragmentation of the Ile-Ile-Tyr peptide during LC-MS analysis to facilitate its *de novo* sequencing through their comparison to the detected product ions in its MS/MS spectrum.

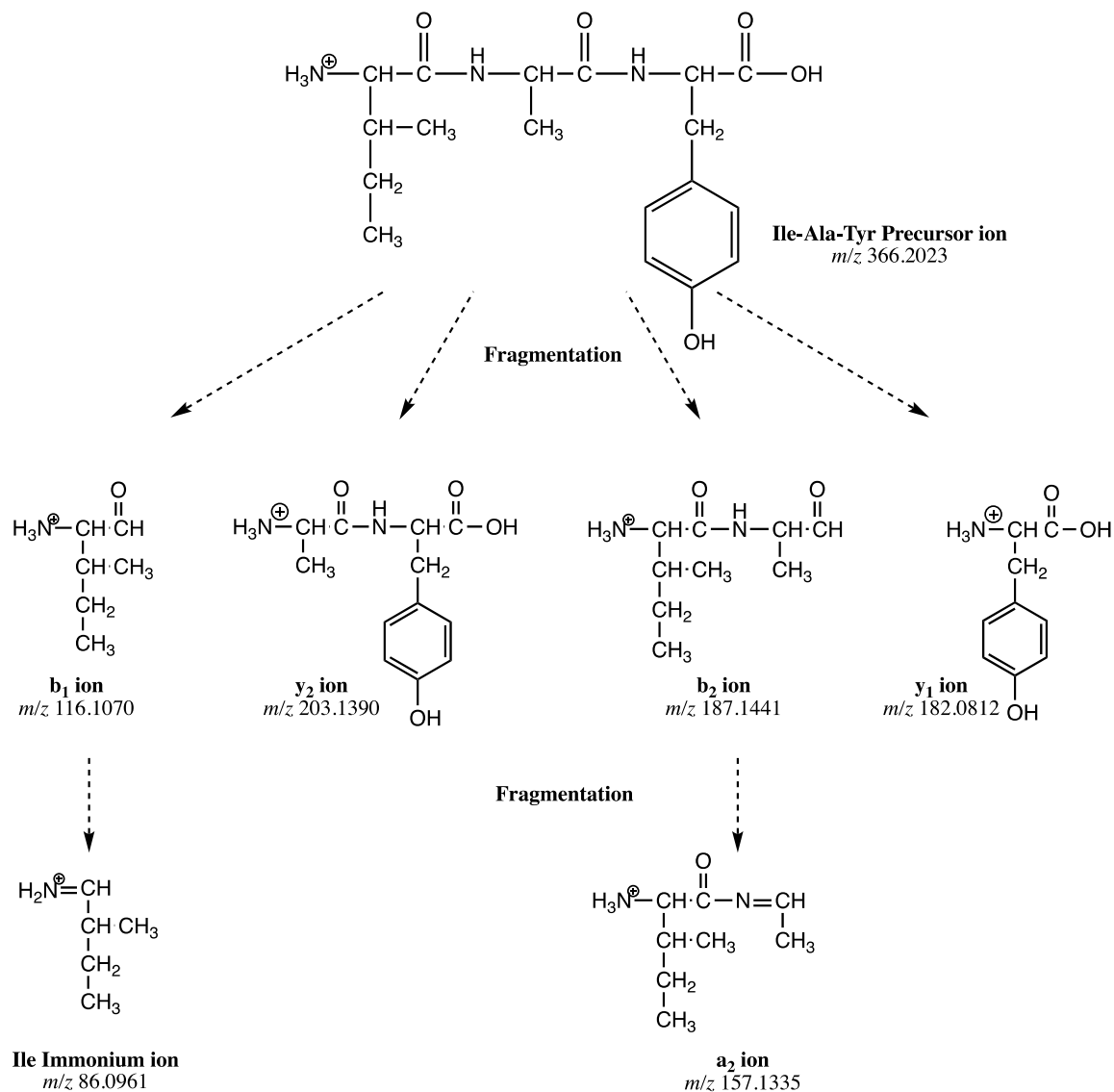


Figure A5. Simulated fragmentation of the Ile-Ala-Tyr peptide during LC-MS analysis to facilitate its *de novo* sequencing through their comparison to the detected product ions in its MS/MS spectrum.

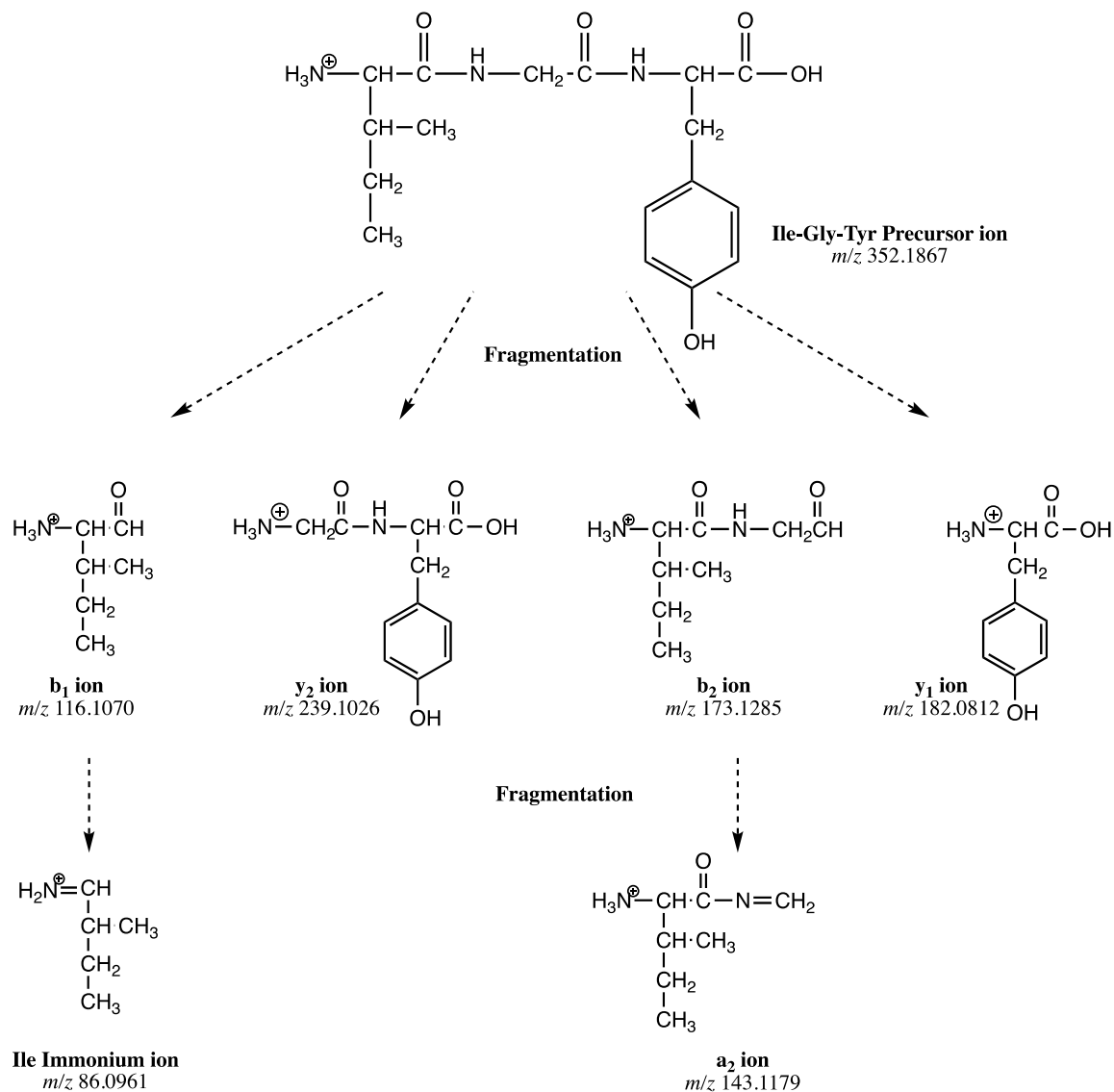


Figure A6. Simulated fragmentation of the Ile-Gly-Tyr peptide during LC-MS analysis to facilitate its *de novo* sequencing through their comparison to the detected product ions in its MS/MS spectrum.