BEHAVIOR OF ALPHA-2-MACROGLOBULIN UNIQUE PEPTIDES IN BIOLOGICAL SAMPLES

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

PELİN YILDIZ

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
CHEMISTRY

AUGUST 2021

Approval of the thesis:

BEHAVIOR OF ALPHA-2-MACROGLOBULIN UNIQUE PEPTIDES IN BIOLOGICAL SAMPLES

submitted by **PELİN YILDIZ** in partial fulfillment of the requirements for the degree of **Master of Science in Chemistry, Middle East Technical University** by,

Prof. Dr. Halil Kalıpçılar
Dean, Graduate School of **Natural and Applied Sciences** ________________

Prof. Dr. Özdemir Doğan
Head of the Department, **Chemistry** ________________

Assist. Prof. Dr. Süreyya Özcan Kabasakal
Supervisor, **Chemistry, METU** ________________

Examining Committee Members:

Prof. Dr. Gülay Ertaş
Chemistry, METU ________________

Assist. Prof. Dr. Süreyya Özcan Kabasakal
Chemistry, METU ________________

Assoc. Prof. Dr. Ezel Boyacı
Chemistry, METU ________________

Assoc. Prof. Dr. Özgül Persil Çetinkol
Chemistry, METU ________________

Assoc. Prof. Dr. Tunca Doğan
Computer and Artificial Intelligence Eng., Hacettepe Üni. ________________
Institute of Informatics, Hacettepe Üni.

Date: 09.08.2021

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name : Pelin Yıldız

Signature :

ABSTRACT

BEHAVIOR OF ALPHA-2-MACROGLOBULIN UNIQUE PEPTIDES IN BIOLOGICAL SAMPLES

Yıldız, Pelin
Master of Science, Chemistry
Supervisor: Assist. Prof. Dr. Süreyya Özcan Kabasakal

August 2021, 132 pages

Proteomics is the comprehensive study of proteins and proteoforms. Proteomics research enables the identification of new protein biomarkers for diagnostic applications and investigates novel targets for drug development. In bottom-up (shotgun) proteomics, proteins are digested using proteases, and corresponding peptides are analyzed by mass spectrometry (MS). The peptide-centric approach focuses on MS-based identification and/or quantification of only unique peptide(s) of the protein. However, proteins often contain multiple unique peptides. Therefore, the selection of the unique peptide representing the protein is crucial for both qualitative and quantitative proteomics. Here, we investigated the relationship between protein concentration and unique peptide responses under conventional proteolytic digestion conditions. Alpha-2-macroglobulin (A2MG), a clinically important protein associated with liver, lung, neurological diseases, and prostate cancer, was selected as a reference protein. Two common proteases, trypsin, and a trypsin/Lys-C mixture, were used for proteolytic digestion. Protein-peptide correlation, digestion efficiency, matrix-effect, and concentration-effect were evaluated for protein standard, human serum, and bovine serum. Twelve unique A2MG peptides were monitored using triple quadrupole mass spectrometry (QQQ-MS) operated in multiple reaction monitoring mode (MRM).

The protein-unique peptide correlations were assessed at eight protein concentration levels. While a linear correlation was observed for all unique peptides at high concentration levels (0.0536 – 0.1071 µg/µl), peptide correlation at low concentration levels (0.0071 – 0.0357 µg/µl) were variable for both enzymes. The same investigation was performed in human serum and bovine serum at three A2MG protein levels. The results showed that the change in protein level was not reflected in peptide levels.

We further investigated the relationship between protein-peptide correlation and certain peptide parameters such as pI value, peptide length, locations in the protein structure, and the presence of reactive amino acids. Outcomes of the research suggested that location of the peptides in the protein structure is the main factor which affects the linear peptide correlation since peptides located inner regions of the structure did not show linear correlation with other target peptides. Also, the peptides with lowest pI values show opposite correlation among all twelve A2MG unique peptides. It was observed that the twelve A2MG unique peptides behave different at different protein concentrations, as well as, various biological matrices.

This is the first study investigating dynamic protein-peptide correlations in biological samples. The behavior of peptides at different concentrations and biological environments is critical for protein-based biomarker studies.

Keywords: targeted proteomics, mass spectrometry, multiple reaction monitoring (MRM), alpha-2-macroglobulin, unique peptide

ÖZ

BİYOLOJİK NUMUNELERDE ALFA-2-MAKROGLOBÜLİN ÖZGÜN PEPTİTLERİNİN DAVRANIŞI

Yıldız, Pelin
Yüksek Lisans, Kimya
Tez Yöneticisi: Dr. Öğr. Üyesi Süreyya Özcan Kabasakal

Ağustos 2021, 132 sayfa

Proteomik, proteinlerin ve proteoformların kapsamlı bir çalışmasıdır. Proteomik araştırma, teşhis uygulamaları için yeni protein biyobelirteçlerinin tanımlanmasını sağlar ve ilaç geliştirme için yeni hedefleri araştırır. 'Bottom-up' (Aşağıdan yukarıya) proteomikte, proteinler proteazlar kullanılarak sindirilir ve karşılık gelen peptitler kütle spektrometrisi (MS) ile analiz edilir. Peptit merkezli yaklaşım, proteinin yalnızca özgün peptit(ler)inin MS tabanlı tanımlanmasına ve/veya nicelenmesine odaklanır. Bununla birlikte, proteinler genellikle birden fazla özgün peptit içerir. Bu nedenle, proteini temsil eden özgün peptidin seçimi, hem kalitatif hem de kantitatif proteomik için çok önemlidir. Burada, geleneksel proteolitik sindirim koşullarında protein derişimini ile özgün peptit tepkileri arasındaki ilişkiyi araştırılmıştır. Karaciğer, akciğer, nörolojik hastalıklar ve prostat kanseri ile ilişkili klinik olarak önemli bir protein olan alfa-2-makroglobulin (A2MG) referans protein olarak seçilmiştir. Proteolitik sindirim için iki yaygın proteaz, tripsin ve bir tripsin/Lys-C karışımı kullanılmıştır. Protein-peptit korelasyonu, sindirim verimliliği, matriks etkisi ve konsantrasyon etkisi, protein standardı, insan serumu ve sığır serumu için değerlendirilmiştir. On iki A2MG özgün peptidi, çoklu reaksiyon izleme modunda (MRM) çalıştırılan üçlü dört kutuplu kütle spektrometrisi (QQQ-MS) kullanılarak izlenmiştir.

Proteine özgü peptit korelasyonları, sekiz protein derişimi seviyesinde değerlendirilmiştir. Tüm özgün peptitler için yüksek derişim seviyelerinde (0.0536 – 0.1071 µg/µl) doğrusal bir korelasyon gözlemlenirken, düşük derişim seviyelerinde (0.0071 – 0.0357 µg/µl) peptit korelasyonu her iki enzim için değişkendir. Aynı araştırma, insan serumu ve sığır serumunda üç A2MG protein seviyesinde gerçekleştirilmiştir. Sonuçlar, protein seviyesindeki değişimin peptit seviyelerini yansıtmadığını göstermiştir.

Protein-peptit korelasyonu ile pI değeri, peptit uzunluğu, protein yapısındaki yerler ve reaktif amino asitlerin varlığı gibi belirli peptit parametreleri arasındaki ilişki de araştırılmıştır. Araştırma sonuçları, yapının iç bölgelerinde yer alan peptitlerin diğer hedef peptitlerle doğrusal korelasyon göstermediğinden, protein yapısındaki peptitlerin lokasyonunun doğrusal peptit korelasyonunu etkileyen ana faktör olduğunu göstermiştir. Ayrıca, en düşük pI değerlerine sahip peptitler, on iki A2MG özgün peptitinin tümü arasında zıt korelasyon gösterir. On iki A2MG özgün peptidinin, çeşitli biyolojik matrikslerin yanı sıra farklı protein derişimlerinde farklı davrandığı gözlemlenmiştir.

Bu, biyolojik örneklerde dinamik protein-peptit korelasyonlarını araştıran ilk çalışmadır. Peptitlerin farklı derişimlerde ve biyolojik ortamlardaki davranışı, protein bazlı biyobelirteç çalışmaları için kritik öneme sahiptir.

Anahtar Kelimeler: hedefli proteomik, kütle spektrometresi, çoklu reaksiyon izleme (MRM), alfa-2-makroglobülin, özgün peptit

I close my eyes, only for a moment and the moment's gone
All my dreams pass before my eyes, a curiosity

Dust in the wind
All we are is dust in the wind
Dust in the wind
Everything is dust in the wind

Same old song, just a drop of water in an endless sea
All we do crumbles to the ground, though we refuse to see

Dust in the wind
All we are is dust in the wind
Dust in the wind
Everything is dust in the wind

Now, don't hang on, nothing lasts forever but the earth and sky
It slips away, and all your money won't another minute buy

Dust in the wind
All we are is dust in the wind
Dust in the wind
Everything is dust in the wind

KANSAS – Dust in the Wind (1977)

I would like to dedicate this thesis to my family, my biggest supporters

my mom Ayça, my dad Arslan, my brother Kaan

and my lovely daughter Mişo ...

ACKNOWLEDGMENTS

I'd like to express my heartfelt and genuine gratitude to my advisor, Assist. Prof. Dr. Süreyya Özcan Kabasakal, for guiding me through this brief but challenging process, broadening my horizons, always remaining positive and optimistic, making a meaningful contribution to overcoming my problems, always motivating me. I am also grateful for your continuous support, encouragement, love, and patience throughout this study, as well as your constant concern about and attention to my occupational safety and health during the pandemic process.

I would like to thank Assoc. Prof. Dr. Çağdaş Son for a serum sample.

I am thankful to my analytical chemistry professors, Prof. Dr. Mürvet Volkan, Prof. Dr. Gülay Ertaş, and Assoc. Prof. Dr. Ezel Boyacı, for always being helpful, sharing, morally supportive, and understanding.

I am appreciative to members of the analytical chemistry community (C Block fellows) for allowing me to use their laboratories and for never refusing to assist me when I needed it, especially, Canan Höçük, Sezin Özdemir, Begüm Avcı, and Cemre Mertoğlu.

I'd like to express my sincere thanks to my labmate Hatice Akkulak and my gorgeous and dear friends, Hande Güler, Özge Börekçi, Sibel Kırılmaz, Necla Güçlü, and Zeynep Suvacı, as well, for being there for me throughout my academic career, always willing to help, thinking of me, listening to my problems, and assisting me in overcoming them. I wouldn't have been able to complete this study without their friendship and support.

Last but not least, I'd like to express my genuine, deepest, and sincere gratitude to my family, my mother, my father, and my brother, to whom I owe everything. I couldn't have accomplished any of this without their help. I'd like to give special thanks to my incomparable and unique family for being helpful, supportive, encouraging, optimistic, thoughtful, and understanding.

TABLE OF CONTENTS

ABSTRACT v

ÖZ vii

ACKNOWLEDGMENTS x

TABLE OF CONTENTS xi

LIST OF TABLES xiv

LIST OF FIGURES xv

LIST OF ABBREVIATIONS AND SYMBOLS xx

1 INTRODUCTION 1

1.1 Introduction to Proteomics 1

1.1.1 Proteins 2

1.1.2 Post-Translational Modifications 4

1.1.3 Targeted and Untargeted Proteomics 6

1.1.4 Enzymatic Digestion in Proteomics 10

1.1.5 Proteomics Based Disease Biomarker Studies 12

1.2 Introduction to Mass Spectrometry 14

1.2.1 A General Overview to Mass Spectrometry 14

1.2.2 Ionization Techniques 15

1.2.3 Mass Analyzers 16

1.2.4 The Vacuum System 20

1.2.5 Mass Resolution and Resolving Power 20

1.2.6 Collision Induced Dissociation 21

1.3 Model Protein: Alpha-2-Macroglobulin 24

1.3.1 The Clinical Significance of A2MG 24

1.3.2 The Structure and Function of A2MG 25

1.3.3 The Modifications of A2MG 27

1.4 Motivation 29

2 MATERIALS AND METHODS 31

2.1 Materials 31

2.1.1 Samples 32

2.1.2 Enzymatic In-Solution Digestion Procedure 35

2.1.3 The Sample Preparation for LC-MS Analysis 36

2.2 LC-MS Analysis 38

2.2.1 Instrumentation 38

2.2.2 LC-MS Conditions 38

2.3 Targeted Proteomics Method Development 40

2.4 Data Analysis 41

2.4.1 Data Pre-Processing 42

2.4.2 Statistical Data Processing 42

3 RESULTS AND DISCUSSION 44

3.1 MRM Method Development 44

3.1.1 The Selection of Unique Peptides of A2MG Protein 44

3.1.2 The Localization of Unique Peptides 47

3.1.3 The Association of Unique Peptides with Literature 50

3.1.4 The Selection of Unique Peptide Standards 52

3.1.5 The Determination of Retention Times 56

3.1.6 The Determination of Predominant Charge States 58

3.1.7 The Selection of MRM Transitions....................60
3.2 Targeted Protein Analysis....................62
3.2.1 Data Pre-Processing....................62
3.2.1 The MRM Method Performance....................65
3.2.1 The Digestion Reproducibility....................69
3.2.2 Statistical Data Processing....................71
3.2.1 The Peptide Normalization....................72
3.3 The Effect of Concentration....................76
3.4 The Effect of Enzymes on Digestion Process....................80
3.5 The Peptide Behavior in Different Matrices....................83
4 CONCLUSION....................93
REFERENCES....................97
A. HPLC and MS Results of Peptide Standards and IS....................127

LIST OF TABLES

Table 1.1. *List of post-translational modifications in UniProt database* 28
Table 2.1. *Peptide and internal standard stock solutions and their concentrations* 33
Table 2.2. *The MS operating parameters*...39
Table 2.3. *The list of A2MG unique peptides*...40
Table 3.1. *List and information of unique tryptic peptides in A2MG protein*........46
Table 3.2. *The color codes of unique peptides in Figure 3.2*...............................47
Table 3.3. *The list of unique peptides located near to glycosylation*....................49
Table 3.4. *The number of publications of A2MG unique peptides in literature*.....50
Table 3.5. *The criteria for identifying unique representative peptides*.................53
Table 3.6. *The retention times of unique peptides*...57
Table 3.7. *The selected predominant charges of unique peptides*.........................59
Table 3.8. *The list of selected MRM transitions of unique peptides of A2MG protein* 60
Table 3.9. *List of A2MG protein's unique peptides shared to both matrices*.........84

LIST OF FIGURES

Figure 1.1. *The branches of proteomics. The figure is generated from "Applications of Proteomics" by Abhilash, M., 2008, The Internet Journal of Genomics and Proteomics, 4(1), 1-7.* 1
Figure 1.2. *The list of the most common PTMs. The figure is regenerated from "Proteome-Wide Post-Translational Modification Statistics: Frequency Analysis and Curation of The Swiss-Prot Database" by Khoury, G. A., Baliban, R. C., Floudas, C. A., 2011, Scientific Reports, 1(5), 1–5.* 5
Figure 1.3. *The following steps for protein identification and protein quantification in proteomics.* 7
Figure 1.4. *The workflow of methods commonly used in proteomics. The figure is regenerated from "The Pros and Cons of Peptide-Centric Proteomics" by Duncan, M. W., Aebersold, R., Caprioli, R. M., 2010, Nature Biotechnology, 28(7), 659–664.* 8
Figure 1.5. *The representation of peptide-centric approach in proteomics. The figure is regenerated from "Proteomics: Technologies and Their Applications" by Aslam, B., Basit, M., Nisar, M. A., Khurshid, M., Rasool, M. H., 2017, Journal of Chromatographic Science, 55(2) 182–196.* 9
Figure 1.6. *The illustration of protein digestion steps with trypsin enzyme* 11
Figure 1.7. *The schematic representation of mass spectrometer instrumentation* . 14
Figure 1.8. *The schematic interpretation of the positive mode of the ESI process*. 16
Figure 1.9. *The schematic representation of MRM mode on a QQQ instrument*... 18
Figure 1.10. *The interpretation of six types of transition on Arg-Cys-Lys residues* 22
Figure 1.11. *The representation of fragmentation of peptide FEVQVTVPK in +2 charge state using ion trap MS result of NIST peptide library* 23
Figure 1.12. *The 3D structure of A2MG from ExPasy Swiss Model Database* 26
Figure 2.1. *The reference systems which were used in the study*.......................... 32
Figure 2.2. *The workflow of sample preparation for in-solution digestion*........... 35

Figure 2.3. *Experimental study design for each sample set* 36
Figure 2.4. *The illustration of the statistical data treatment process* 42
Figure 3.1. *Positions of lysine and arginine in the A2MG protein. The purple color code is used for lysine residues and the yellow color code for the arginine residues* 45
Figure 3.2. *The 3D model of A2MG protein in PyMOL software.* ***A.*** *The locations of unique peptides in the protein structure.* ***B.*** *The localization of unique peptides and glycosylation of the protein structure.* 47
Figure 3.3. *A word cloud plot represents the incidence of A2MG unique peptides in the literature.* 51
Figure 3.4. *The determined retention times of 12 unique peptides in serum digest*56
Figure 3.5. *The representation of determining the predominant charge state on target unique peptide QGIPFFGQVR in A2MG protein standard.* ***A.*** *The chromatogram of the +2 charge state of the peptide.* ***B.*** *The chromatogram of the +3 charge state of the unique peptide* 58
Figure 3.6. *The representation of interference filtration on the unique peptide FEVQVTVPK in Skyline software* ***A.*** *The peak of the unique peptide is suppressed by interference.* ***B.*** *The interference eliminated version of the chromatogram* 63
Figure 3.7. *The interpretation of reintegration on one of the unique peptides AAQVTIQSSGTFSSK in Skyline software.* ***A.*** *The auto-integrated peak of the peptide in the serum sample.* ***B.*** *The reintegrated peak of the peptide in the serum sample.* ***C.*** *The auto-integrated peak of the peptide in the A2MG protein standard sample.* ***D.*** *The reintegrated peak of the peptide in the A2MG protein standard sample.........*64
Figure 3.8. *The run stability of QC serum in the QQQ instrument* ***A.*** *The normalized peak area of each peptide* ***B.*** *CV values of unique peptides in QC serum* 65
Figure 3.9. *The Log2 normalized peak areas versus CV values of QC serum*.......66
Figure 3.10. *The replicate stability of unique peptide FEVQVTVPK of QC serum in Skyline software* 67
Figure 3.11. *The linearity of unique peptide FEVQVTVPK in formic acid, bovine serum and human serum samples* 68

Figure 3.12. *The replicates of standard neat human serum samples for both enzymes.* ***A.*** *In the presence of trypsin, each peptide's normalized response is reported in replicates of neat human serum.* ***B.*** *In the presence of trypsin/Lys-C mixture enzyme, each peptide's normalized response is reported in replicates of neat human serum.* ***C.*** *CV values of peptides digested with trypsin* ***D.*** *CV values of peptides digested with trypsin/Lys-C mixture* 70
Figure 3.13. *The Log2 normalized peak areas versus CV values of neat human serum samples* ***A.*** *in trypsin* ***B.*** *in trypsin/Lys-C mixture* 71
Figure 3.14. *The error column plots of all unique peptides in six serum digest replicates for both enzymes.* ***A.*** *The plot of obtained peak areas of unique peptides for both enzymes.* ***B.*** *The plot of log2 normalized peak areas of unique peptides for both enzymes.* 72
Figure 3.15. *The five selected unique peptides' response* ***A.*** *in equimolar peptide mixture* ***B.*** *in protein standard in the presence of trypsin* ***C.*** *in protein standard in the presence of trypsin/Lys-C mixture* ***D.*** *in human serum in the presence of trypsin* ***E.*** *in human in the presence of trypsin/Lys-C mixture* 73
Figure 3.16. *Coefficient of variation (CV) of commonly used normalization approaches in neat human serum for both enzyme digestions* ***A.*** *CV values of different normalization approaches for trypsin digestion* ***B.*** *CV values of different normalization approaches for trypsin/Lys-C mixture digestion*...........................75
Figure 3.17. *A heat map with a dendrogram plot demonstrating the association of A2MG unique peptides in protein standard samples presents different enzymes* ***A.*** *Correlation in low concentration protein standard samples in the presence of trypsin.* ***B.*** *Correlation in low-concentration protein standard samples with trypsin/Lys-C mixture.* ***C.*** *The correlation in high concentration protein standards presence of trypsin.* ***D.*** *Correlation in high-concentration protein standard samples with trypsin/Lys-C mixture*..77
Figure 3.18. *The oppositely correlated peptides in Figure 3.17C with the highlighted CV values* ***A.*** *The peptide correlation due to trypsin digestion in Figure 3.17C. The*

oppositely correlated peptides are highlighted. ***B.*** *The CV values of these peptides are highlighted.*..79

Figure 3.19. *A heatmap with a dendrogram plot depicting the correlation of A2MG unique peptides in protein standard and serum digest samples due to change in enzyme.* ***A.*** *The correlation in protein standard samples upon trypsin digestion.* ***B.*** *The correlation in protein standard samples upon trypsin/Lys-C mixture digestion.* ***C.*** *The correlation in serum samples upon trypsin digestion.* ***D.*** *The correlation in serum samples upon trypsin/Lys-C mixture digestion.*...81

Figure 3.20. *The close inspection of oppositely correlated peptides in Figure 3.19D. The red-colored amino acids represent chemically active amino acids, and the blue-colored amino acids are the ones that affect activity of trypsin.*...........................82

Figure 3.21. *A heatmap with a dendrogram plot illustrating the association of A2MG unique peptides in human and bovine serum-protein spiked samples in the presence of various enzymes.* ***A*** *Peptide correlation in the presence of trypsin in protein standard.* ***B.*** *The correlation in the presence of trypsin/Lys-C mixture in protein standard.* ***C.*** *The correlation in the presence of trypsin in bovine serum* ***D.*** *The correlation in the existence of trypsin/Lys-C mixture in bovine serum* ***E.*** *The correlation in the presence of trypsin in human serum.* ***F.*** *The correlation in the presence of trypsin/Lys-C mixture in human serum*..85

Figure 3.22. *A heatmap with a dendrogram plot highlighting the association of A2MG unique peptides in standard human serum and ERM certificated human serum samples with the participation of different enzymes* ***A.*** *The correlation of trypsin digested ERM serum.* ***B.*** *The correlation of trypsin/Lys-C mixture digested ERM serum samples.* ***C.*** *The correlation of trypsin digested standard human serum.* ***D.*** *The correlation of trypsin/Lys-C mixture digested human serum samples.*.......87

Figure 3.23. *The shared unique peptides with the strongest opposite correlation in both enzyme digests were marked. The A2MG protein structure's structure is shown on the left panel.*..89

Figure 4.1. *HPLC chromatogram of FEVQVTVPK*..127

Figure 4.2. *Mass spectrum of FEVQVTVPK*...127

Figure 4.3. *HPLC chromatogram of DMYSFLEDMGLK*.................................128
Figure 4.4. *Mass spectrum of DMYSFLEDMGLK*..128
Figure 4.5. *HPLC chromatogram of QGIPFFGQVR*.......................................129
Figure 4.6. *Mass spectrum of QGIPFFGQVR*..129
Figure 4.7. *HPLC chromatogram of VGFYESDVMGR*....................................130
Figure 4.8. *Mass spectrum of VGFYESDVMGR*...130
Figure 4.9. *HPLC chromatogram of LPPNVVEESAR*.......................................131
Figure 4.10. *Mass spectrum of LPPNVVEESAR*...131
Figure 4.11. *HPLC chromatogram of TFLLR (Internal Standard)* 132
Figure 4.12. *Mass spectrum of TFLLR (Internal Standard*................................132

LIST OF ABBREVIATIONS AND SYMBOLS

ABBREVIATIONS

A2MG	Alpha-2-macroglobulin
ABC	Ammonium bicarbonate
ACN	Acetonitrile
CE	Collision energy
CID	Collision-induced dissociation
DTT	Dithiothreitol
ELISA	Enzyme-linked immunosorbent assay
ERM	European reference materials
ESI	Electrospray ionization
FTICR	Fourier-transform ion cyclotron resonance
HPLC	High-pressure liquid chromatography
IAA	Iodoacetamide
IS	Internal Standard
LC	Liquid chromatography
LC-MS	Liquid chromatography-mass spectrometer
LOD	Limit of detection
LOQ	Limit of quantification
LTQ	Linear ion trap
m/z	Mass-to-charge ratio

MALDI	Matrix-assisted laser desorption/ionization
MRM	Multiple reaction monitoring
MS	Mass spectrometry
MS/MS	Tandem mass spectrometer
PTM	Post-translational modification
Q	Quadrupole
QIT	Quadrupole ion trap
QQQ	Triple quadrupole
Q-TOF	Quadrupole time-of-flight
SRM	Selected reaction monitoring
TOF	Time-of-flight

CHAPTER 1

INTRODUCTION

1.1 Introduction to Proteomics

Proteomics is the comprehensive investigation of protein structure and function. The term was used first time by Marc Wilkins in 1994[1,2]. It is derived by combining the words proteome and the suffix omics.

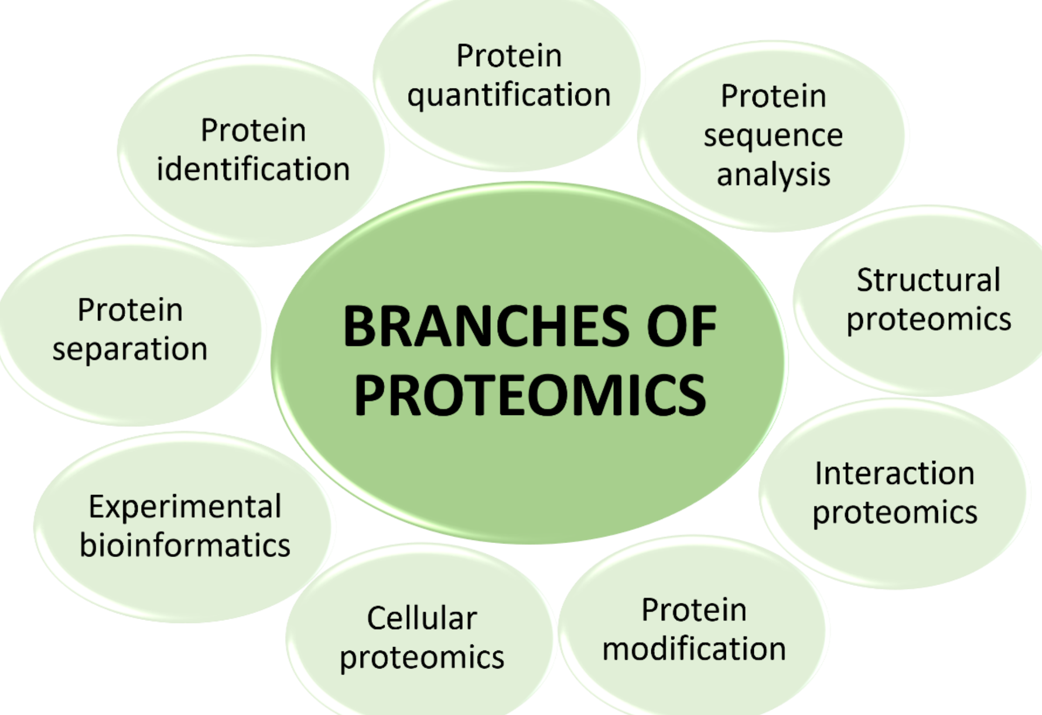

Figure 1.1. *The branches of proteomics. The figure is generated from "Applications of Proteomics" by Abhilash, M., 2008, The Internet Journal of Genomics and Proteomics, 4(1), 1-7.*

The proteome expresses an entire set of proteins in a living organism[3], and adding the suffix 'omics' to a molecular definition implies a thorough examination of molecules.

By merging clinical data, the study of omics has been widely employed to generate great insight into biological processes. Also, proteomics investigates and clarifies the cause-and-effect relationship between proteins and their biological activities, as well as how proteins influence biological processes. As a result, a wide range of omics approaches has been emerged. Omics studies also include genomics, transcriptomics, metabolomics, glycomics but are not limited to proteomics[4].

Proteomics is such a wide area of study. Because of that, it is subdivided into sub-branches, as shown in **Figure 1.1**[5].

Each branch concentrates on a different aspect of the protein's research. The studies in proteomics are intended to better understand the structure and function of proteins, how they perform their roles in living organisms, the relationship among proteins and the impact of this relationship, and how they affect living systems. In other words, proteomics is at the core of studies into the whole complement of proteins, including concentrations, functions, isoforms, structures, interactions, modifications, regulation, and localization in living organisms[6–8].

1.1.1 Proteins

Proteins are macromolecules composed of amino acids, which are known as building blocks of proteins. The synthesis of a protein starts with the peptide bond formation between two amino acids. As a consequence of the peptide bond formation between several amino acids, polypeptides, polymers, are produced. Proteins are polypeptide chains that contain hundreds of amino acids[9].

Proteins are at the center of body systems as they participate in numerous essential biological reactions[10,11]. Proteins play a variety of roles in the body, including acting

as a catalyst. These catalyst proteins, well-known as enzymes, are present in almost all biological processes and help to provide a chemical platform for biochemical reactions[12]. Proteins also play a critical part in transportation within the body. Transport proteins help small molecules or ions to be transferred from one location in the body to the other[13]. One of the most prominent examples is hemoglobin, which transports oxygen from the lungs to and tissues within the body[14]. In addition, proteins participate in the blood coagulation system[11]. The coagulation happens due to the formation of insoluble fibrous proteins. Moreover, proteins are responsible for gene regulation. These regulatory proteins control protein synthesis by participating in replication, transcription, and translation[15]. Furthermore, proteins function as antibodies in the body. Antibodies are also well-known as immunoglobulins which detect unknown organisms such as viruses or bacteria to the human system, and neutralize them[16,17]. In addition to the functions described above, proteins play crucial roles in various biological processes such as producing cell movement, transferring signals, including nerve impulse transmission, providing mechanical support to cells and tissues, between or within cells, and etc[18]. Protein structural changes serve such a wide range of biological functions because the four protein structures mentioned below enable proteins to differentiate from each other.

Proteins have complex structures, and the chemical bonding between amino acids assists protein's stability and shape. Furthermore, proteins can be folded, looped, and curled differently from a one-dimensional structure to form a three-dimensional molecule capable of performing various biological activities[18]. Protein structures are classified into four main categories. These are primary, secondary, tertiary, and quaternary structures[9]. *The primary structure* is defined as the linear amino acid sequence that forms the proteins. *The secondary structure* is the 3D form of protein coiling or folding and is classified mainly into two types: alpha-helix and beta-sheet[9]. *The tertiary structure* is the three-dimensional shape of a protein provided by various bonds and forces such as hydrogen bonding, ionic bonding, disulfide bonding, and hydrophobic interactions[19]. *The quaternary structure* is defined as the spatial

arrangement of subunits formed by interactions between multiple polypeptide chains. Each polypeptide chain is referred as a subunit. The quaternary structures of proteins may either contain many same or different subunits[20]. The quaternary structure is classified based on the number of subunits as monomeric, dimeric, tetrameric, etc.

Protein complexity is increased not only from four different protein structures but also from proteoforms. The term proteoform is used to refer to all the various molecular forms of the protein. They are generated by a single gene, but the variation in the structure is caused by genetic variants, alternatively spliced RNA transcripts, and post-translational modifications (PTMs)[21]. The differences in post-translational modifications cause alterations in the protein structure and function. The following section will go over post-translational modifications in detail.

1.1.2 Post-Translational Modifications

Post-translational modification (PTM) is a vital mechanism used by living organisms to significantly regulate and alter their biological or chemical function or activity[22,23]. These modifications are generally enzymatic modifications and enhance the protein's nature by diversifying its functions and structures.

Glycosylation, phosphorylation, and acetylation are just a few of the PTMs that result in many additional protein variants. **Figure 1.2**[24] represents the most common post-translational modifications in human proteins from the Swiss-Prot database. The most common three modifications are N-linked glycosylation, phosphorylation, and acetylation, respectively.

Glycosylation is the term used for the biological process of binding of glycans (complex sugar proteins) covalently getting attached to proteins and is named according to the atom to which it is bound[25,26]. It is the process of attaching sugars to proteins, and it produces more proteome variability than other PTMs since

different sugars combine with various combinations, and the diversity rises when the linkage isomers is included. N-linked glycosylation and O-linked glycosylation are the two common kinds of glycosylation. The attachment of sugar through the nitrogen atom (N) of asparagine amino acid of the protein is known as N-linked glycosylation, which is the most frequently observed modification[26] as seen in **Figure 1.2**, whereas the attachment of sugar through the oxygen atom (O) of the amino acid serine or threonine is known as O-linked glycosylation.

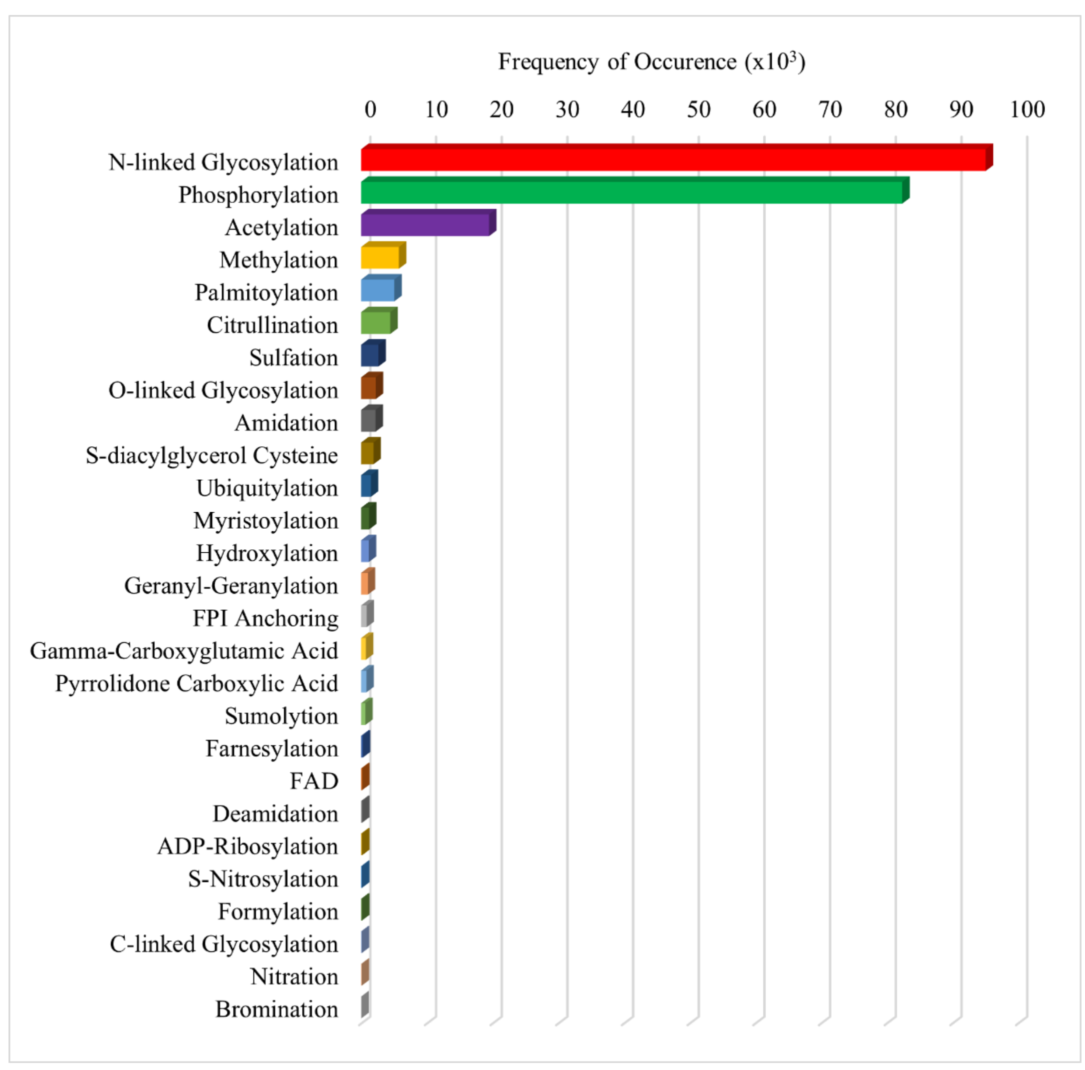

Figure 1.2. *The list of the most common PTMs. The figure is regenerated from "Proteome-Wide Post-Translational Modification Statistics: Frequency Analysis and Curation of The Swiss-Prot Database" by Khoury, G. A., Baliban, R. C., Floudas, C. A., 2011, Scientific Reports, 1(5), 1–5.*

Phosphorylation is a biological process in which a phosphoryl group is attached to a polypeptide chain[27]. One of the most common PTMs is protein acetylation, which involves attaching an acetyl group to a specific location on a protein[28].

The protein isoforms are one of the factors that contribute to increased protein complexity. It is a protein variant that is a part of a group of nearly identical proteins which all derive from the same gene or gene family but diverge genetically. Protein variants are caused by PTMs, which generate proteoforms and isoforms of the protein. Furthermore, these modifications affect biological processes reversibly or irreversibly[22]. As a result, they alter the activity, function, structure, interactions, and location of the protein, as well as increasing its diversity.

Protein identification and quantification are becoming more challenging due to the presence of many protein forms described above. To this end, robust and sensitive targeted and untargeted proteomics approaches should be utilized for clinical applications.

1.1.3 Targeted and Untargeted Proteomics

Targeted proteomics has risen in importance as a method for detecting proteins of interest with high sensitivity, quantitative precision, and repeatability in mass spectrometry-based protein quantification[29]. On the other hand, untargeted proteomics is a common technique for identifying and/or quantifying as many proteins as feasible, but only in a relative manner[30].

The studies of proteomics can be categorized as qualitative and quantitative proteomics. The main purpose of qualitative analysis is to identify as many proteins as possible in biological mixtures. Because of this manner, qualitative analysis is based on an untargeted approach. Protein identification is the process following steps: i) cleaving proteins with a protease, ii) analyzing protein products by mass spectrometry (MS) iii) using reference peptide libraries to identify the proteins. In

addition, protein identification means matching experimental mass spectral data of peptides with MS peptide libraries to determine the proteins. The procedure described in **Figure 1.3** is followed for protein identification. Nevertheless, increased biological sample complexity, different proteoforms, and a wide dynamic range of proteins result in limiting peptide identification reproducibility and quantification consistency in untargeted approaches31. On the other hand, the primary purpose of quantitative proteomics in clinical research is to identify and quantify many proteins in various biological samples.

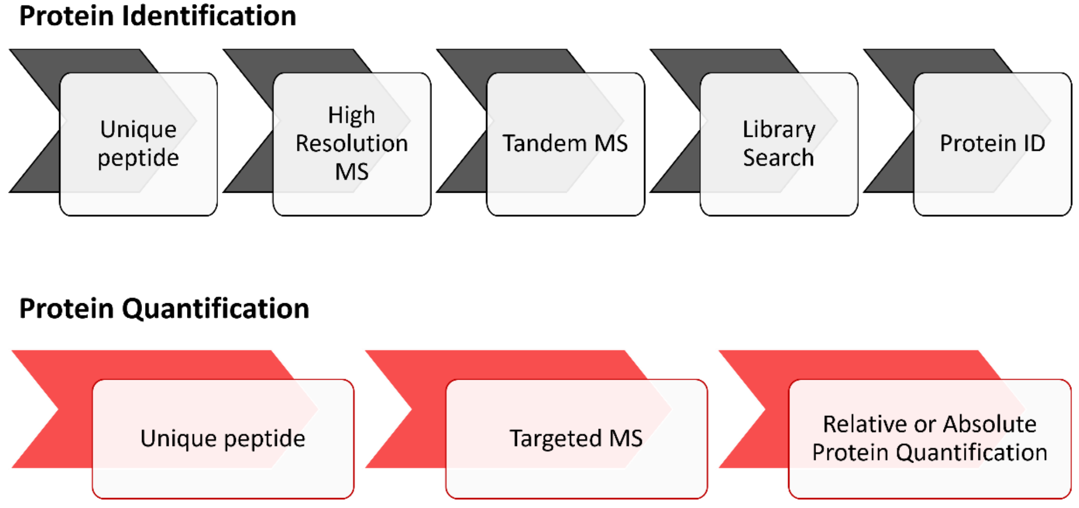

Figure 1.3. *The following steps for protein identification and protein quantification in proteomics*

In quantitative proteomics, enzyme-linked immunosorbent assay (ELISA) is one of the widely used chemiluminescence techniques to discover biomarkers for protein quantification in clinical research[31]. Although ELISA is frequently used for protein quantification, it is expensive[32], has batch-to-batch variations[31], and challenges with multiplexed analysis due to cross-reactivity[32]. Therefore, MS-based techniques are better alternative tools because it is more reproducible and has multiplex characteristics, meaning that several analytes can be analyzed simultaneously. In

addition, quantitative analysis is based on a targeted approach, and analysis steps are shown in **Figure 1.3**. The targeted strategy produces widespread, accurate, and reproducible data. This analysis can be used with either top-down or bottom-up proteomics[31].

There are two main approaches used in qualitative and/or quantitative proteomics: bottom-up and top-down proteomics in **Figure 1.4**[33].

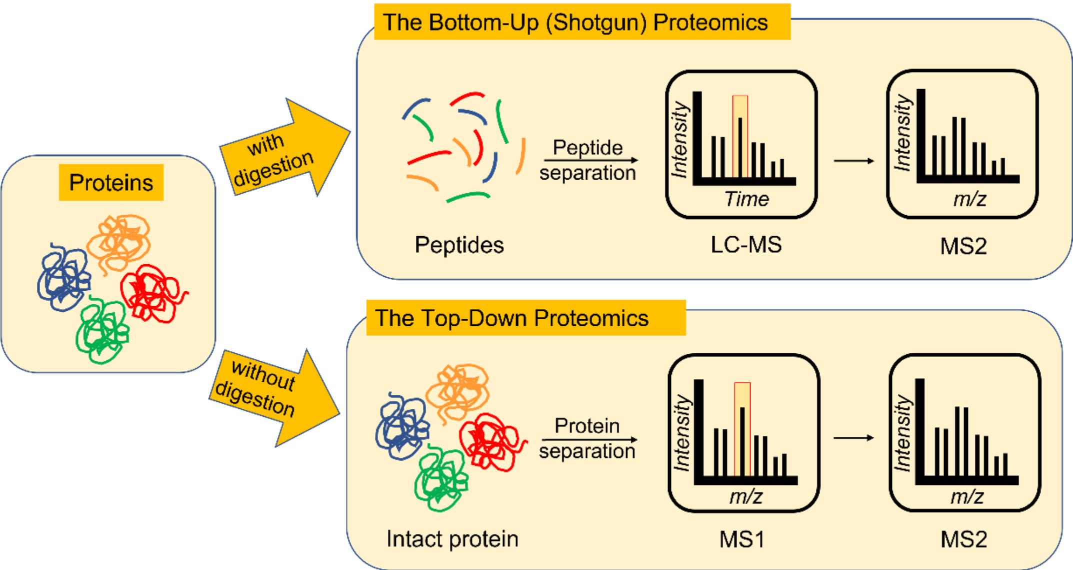

Figure 1.4. *The workflow of methods commonly used in proteomics. The figure is regenerated from "The Pros and Cons of Peptide-Centric Proteomics" by Duncan, M. W., Aebersold, R., Caprioli, R. M., 2010, Nature Biotechnology, 28(7), 659–664.*

In the top-down approach, intact proteins are examined by mass spectrometry without any enzymatic digestion. Top-down proteomics' ability to achieve intact protein characterization has made it particularly valuable for analyzing single proteins or simple mixtures of biological interest. Furthermore, the top-down proteomics is generally used for the characterization of proteins. Besides, the bottom-up proteomics is used to characterize and quantify the proteins. However,

top-down proteomics has fallen behind bottom-up proteomics in terms of proteome coverage, sensitivity, and throughput due to the technological challenges of proteome-wide analysis at the intact protein level[34].

Bottom-up proteomics is also referred to as shotgun proteomics. In bottom-up proteomics, the peptide-centric approach is used for identifying and/or quantifying proteins in the biological environment. The primary assumption of this strategy is illustrated in **Figure 1.5**[35].

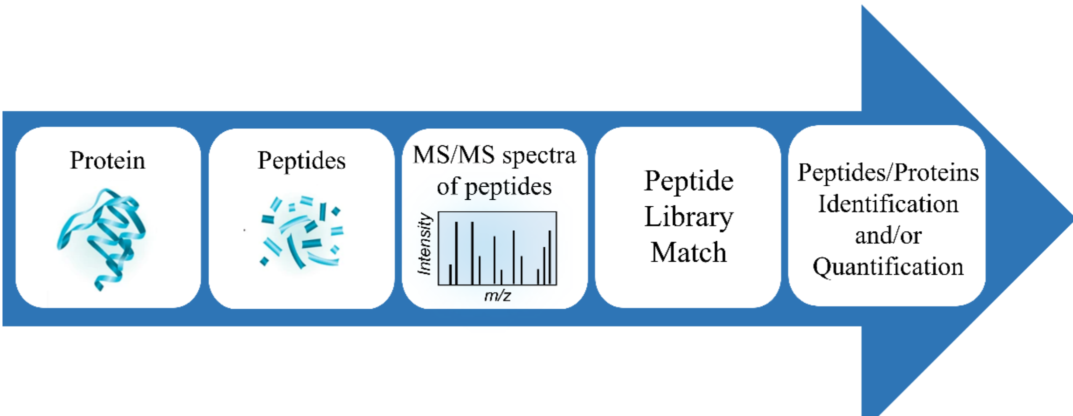

Figure 1.5. *The representation of peptide-centric approach in proteomics. The figure is regenerated from "Proteomics: Technologies and Their Applications" by Aslam, B., Basit, M., Nisar, M. A., Khurshid, M., Rasool, M. H., 2017, Journal of Chromatographic Science, 55(2) 182–196.*

To briefly summarize the diagram, proteins are digested into peptides. The unique peptides of the proteins are selected among all obtained peptides. Then, high-resolution mass spectrometry determines the proteins in the original biological sample by using their unique peptides[33]. The term "unique peptide" is used to describe peptides that are distinct to a protein. The digested protein is identified by these specific peptides thanks to this characteristic. Protein identification is performed by matching the obtained mass results with the peptide and/or protein libraries. In addition, the peptide-centric approach can be used for protein quantification.

In brief, heavy isotope-labeled synthetic peptides[36] are utilized to peptides of targeted proteins in the biological environment. As a result, specified proteins can be quantified with high precision, either relative or absolute.

Proteins are identified by their unique peptides in the peptide-centric approach, as discussed previously. This approach assumes a linear correlation between the selected unique peptide and identified protein. However, proteins frequently contain many unique peptides. According to this strategy, all unique peptides in the protein should behave the same in different circumstances. The selection of unique peptides for representing protein becomes critical. There are criteria to be considered in the selection of peptides in the literature. Despite there are numerous criteria for selecting unique peptides, such as peptide length, lack of post-translational modifications, and avoidance of chemically active amino acids residues that can cause oxidation, acetylation, and so on, there is no evidence in the literature on determining the unique peptide that represents the protein the best[37].

1.1.4 Enzymatic Digestion in Proteomics

One of the essential processes in proteomics study is the sample preparation, which directly impacts the experiment performance[35]. In proteomics, the traditional enzymatic digestion method is to obtain a mixture of peptides by breaking down the intact protein with a protease. This process occurs in the presence of external stress, including heat, radiation, and urea[38,39]. Following that, the mixture of peptides is separated using reverse-phase liquid chromatography. Then, the peptides are identified and/or quantified using mass spectrometry (MS).

Protease selection is a critical element in protein digestion because the number of proteins is mainly affected by the protease's specificity. Higher specificity proteases can identify more proteins than lower specificity proteases caused by incomplete digestion or missed cleavage. Since the peptide databases are made up of particular peptides generated by enzymes, it is simpler to identify peptides acquired from

enzymes with high specificity[40]. In proteomics, trypsin is accepted as 'the gold standard enzyme' for protein digestion. In other words, trypsin is the most common enzyme to digest the proteins[8] into their peptides in mass spectrometry (MS)-based proteomics since it has high cleavage efficiency and specificity[41]. The following are the optimized parameters mostly used in the literature for complete tryptic protein digestion[42–69]: incubation temperature is 37°C, incubation time is 16 hours, and the protein:enzyme ratio is 50:1 (w/w).

Trypsin cleaves proteins at the carboxyl side of the lysine and arginine amino acid residues, as shown in **Figure 1.6**. Other proteases are also mentioned in the literature: chymotrypsin[46–52], Lys-C[42,49–51], Asp-N[52], Arg-C[42], Glu-C[52], and so on.

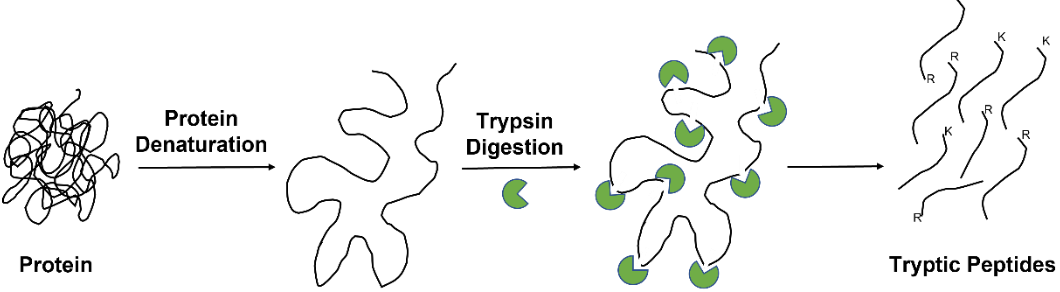

Figure 1.6. *The illustration of protein digestion steps with trypsin enzyme*

The chymotrypsin cleaves at the carboxyl side of tryptophan, tyrosine, and phenylalanine, amino acid residues of proteins. Furthermore, Lys-C cleaves the C-terminus of lysine amino acids, Asp-C cleaves the C-terminus of asparagine amino acids, Arg-C cleaves the C-terminus of arginine amino acids, and Glu-C cleaves the C-terminus of glutamine amino acid residues of proteins.

In order to increase specificity protein digestion, enzyme combinations are recently employed in clinical research[8,70]. The purpose of combining multiple enzymes is to improve the enzyme's digestion efficiency and prevent missed cleavage during protein digestion. In other words, when compared to trypsin digestion alone, this method yields a better number of protein identifications[71]. For example, trypsin

digestion is affected by amino acid residues close to the cleavage site and other factors such as PTMs, local conformation, tertiary structure, and experimental conditions[72]. Consequently, trypsin cleaves less efficiently the protein from all of the lysine (K) and arginine (R) amino acids, resulting in incomplete digestion. Three major factors influence trypsin's efficiency[73]:

1. The presence of a proline residue close to the cleavage site.

2. The basic residues such as lysine (K) and arginine (R) are present in the sequence.

3. The surrounding of the cleavage site by the negatively charged residues glutamate (E) and aspartate (D).

Furthermore, highly folded proteins are resistant to proteolysis, and many protein preparation chemicals limit trypsin activities[74]. Trypsin digestion, in particular, cleavage of lysine (K) amino acid residues becomes more difficult. Combining trypsin with Lys-C improves these deficiencies[74].

1.1.5 Proteomics Based Disease Biomarker Studies

The development of new disease-associated biomarkers by proteomic analysis of commonly available body fluids such as plasma and serum utilizing mass spectrometry (MS)-based technologies provides exciting potential for better patient care[75]. The main purpose of clinical and translational proteomics is to improve existing clinical practice by early accurate diagnosis, new biomarker identification, and personalized medicine monitoring of disease progression and potentially harmful effects[76].

A biomarker is an expression for a disease indicator. The proteins are good sources for biomarkers as they are easily found in biological fluids in living organism[77]. Protein biomarkers play an essential part in diagnosing diseases, disease progress predictions, and tracking disease treatment response in clinical studies[77–79]. Before the final clinical evaluation, the biomarker pipeline is generally considered as a series

of preclinical phases, such as biomarker identification and verification. The process of biomarker verification is a test or procedure to see how effectively biomarker measures, shows, and/or predictions of diseases. Furthermore, the capacity of a test to properly predict a clinically significant response is described as clinical verification.

Furthermore, large-scale verification studies of protein biomarkers are now possible thanks to the robustness and high throughput of MS-based proteomics. The proteomics-based tools for biomarker discovery have shown potential[80,81] since changes in protein expression, protein abundance, structure, or function can be applied as indicators of pathological anomalies before the development of clinical symptoms. In addition, biomarkers for the diseases can be identified by MS-based clinic proteomics focused on identifying proteins in easily accessible body fluids, such as cerebrospinal fluid, serum, or blood[82]. An ideal biomarker would be present in the blood before clinical verification of the disease, have high sensitivity and specificity, and be reproducible[83].

Because MS-based clinic proteomics is a promising area for disease diagnosis and treatment, accurate and reliable analytical methods must be developed and implemented. Despite significant advances in MS-based proteomics technologies, the limits and challenges of using proteomics techniques as a routine diagnostic tool in clinical practice should be validated[83,84].

1.2 Introduction to Mass Spectrometry

1.2.1 A General Overview to Mass Spectrometry

Mass spectrometry (MS) is a powerful, sensitive, and selective analytical tool for analyzing biological samples. It is widely applied in a variety of omics fields, including proteomics, lipidomics, and metabolomics[85]. MS is an irreplaceable instrument for proteomics research, as it can be used in both qualitative and quantitative protein and peptide analysis. The working principle of MS is illustrated in **Figure 1.7**. According to this principle, the molecules are first introduced to an ion source compartment with or without inlet systems such as liquid chromatography (LC). The ion source has two modes: one is a positive mode, where the analyte is protonated, while another is a negative mode, in which the analyte is deprotonated. The introduced analyte is ionized, and then these ionized molecules are filtered in a mass analyzer according to their mass-to-charge (m/z) ratio. The ions are sent to a detector and measured to obtain the mass spectrum. The remaining ions of undesired molecules are sent to the waste compartment.

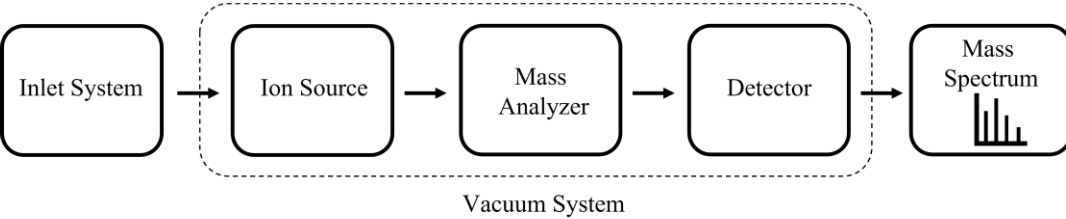

Figure 1.7. *The schematic representation of mass spectrometer instrumentation*

In general instrumentation, the MS components are in a vacuum environment. There are various types of instruments available, depending on various combinations of their compartments. The compartments of MS and instrument types are described in the following sections in detail.

1.2.2 Ionization Techniques

In the ionization process in MS, molecules move into the mass spectrometer through the inlet system and first arrive at the ionization source section. The negatively or positively charged gas-phase ions are then directed to the mass analyzer section.

The ionization techniques, which are used in the analysis, are determined by the properties of the analyte. Since proteins are non-volatile, polar, thermally labile with high molecular weights, soft ionization techniques are favorable in protein and peptide studies[86]. The term 'soft ionization' refers to the fact that it produces little fragmentation. Two soft ionization techniques, listed below, are extensively used in proteomics research.

Matrix-assisted laser desorption/ionization[87] *(MALDI)* is a soft ionization technique that requires laser energy to generate ions. The biomolecules are mixed with a solution of an organic-based absorbent, well-known as the 'matrix'. This part of the process is only peculiar to this ionization technique. A pulsed laser beam irritates the sample on the plate after it crystallizes within the matrix; generally, forming the single-charged ions (protonated or deprotonated)[88,89]. The biggest limitation of this application is the ion suppression issue often observed during the analysis of complex mixtures.

Electrospray ionization[90] *(ESI)* is another commonly used soft ionization technique. The technique was first developed by John B. Fenn, Koichi Tanaka, and Kurt Wüthrich, who were awarded the Nobel Prize in Chemistry in 2002. The invention of this ionization technique has opened up new opportunities for research in the field of proteomics. The basic working principle is that ESI utilizes electrical power to move gas-phase ions. With the help of electric potential, the analyte solution that arrives at the nozzle with the mobile phase is dispersed as charged droplets, which then condense into smaller droplets. The analyte ions in the gas phase are analyzed by the mass analyzer, shown in **Figure 1.8**. In addition, this technique allows the

analysis of large biomolecules by converting them to ionic form[86,91–93]. Hence, it has become an important technique in proteomics studies[92].

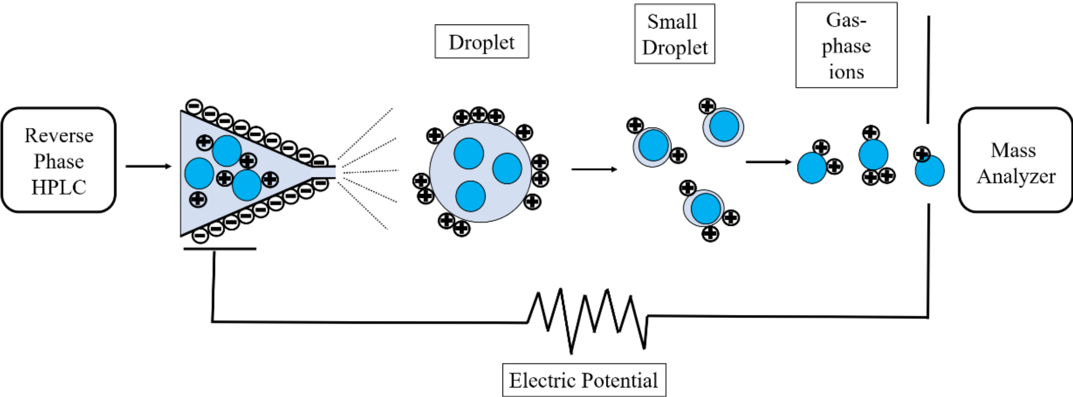

Figure 1.8. *The schematic interpretation of the positive mode of the ESI process*

Because both techniques have pros and cons, combining them, if possible, provides higher performance and more information about the analyte[94]. ESI is often used in combination with time-of-flight (TOF), triple quadrupole (QQQ), and ion trap MS. MALDI is frequently utilized in combination with TOF analyzers to determine the mass of intact peptides[95,96]. Both techniques for measuring protein and peptides have advantages and disadvantages. When they are compared, although MALDI is a faster technique, ESI has higher reproducibility[94].

1.2.3 Mass Analyzers

The mass analyzer is the most essential component of the MS instruments. Moreover, mass spectrometers are named corresponding to their mass analyzers since this compartment is where mass separation methods are used. Generally, four kinds of analyzers are generally utilized in proteomics area[97].

Quadrupole (Q) mass analyzers are made up of four rods that are parallel to each other. It allows ions to travel by following the certain mass-to-charge ratio using the

electric potential provided by the radio frequency (RF) and the direct current (DC) voltage applied by two opposite parallel rods[98,99]. It essentially functions as a mass filter within the MS instrument[2].

Time-of-flight (TOF) mass analyzers use time-of-flight measurements to calculate an ion's mass-to-charge (m/z) ratio. The ions are accelerated by a known electric field, allowing them to obtain kinetic energy. Since the energy provided to ions varies based on their mass-charge ratio, it allows them to be differentiated from one another[100,101]. The ions with a particular m/z ratio enter the detector after being separated from the others.

Ion trap mass analyzers are a method for entrapping ions by combining electric and magnetic fields. While ion trap mass analyzers work similarly to quadrupole mass analyzers, the main difference is that they capture ions[102]. The captured masses are stored because the hyperbolic metal electrodes located between the parallel rods ensure that the trapped ions follow a circular flight direction[98]. As a result, they do not work as a mass filter[102].

Fourier-transform ion cyclotron resonance (FTICR) mass analyzers enable ions to gain cyclotron frequency while being accelerated by a cyclotron. The mass analyzers use this frequency to calculate the m/z value of the ions[103,104].

These mass spectrometers combined with multiple mass analyzers are referred to as 'hybrid instruments'. Furthermore, tandem mass spectrometry (MS/MS) is another term for hybrid instruments. There are numerous MS/MS available, including triple quadrupole (QQQ), triple quadrupole ion trap (QQ-LIT), quadrupole time-of-flight (Q-TOF), tandem time-of-flight (TOF-TOF), and linear ion trap Fourier transform ion cyclotron resonance (LTQ-FTICR).

The triple quadrupole mass spectrometer (QQQ-MS) is one of the tandem mass spectrometers. QQQ-MS is a good sensitive and specific analytical tool for quantifying proteins and peptides in a biological environment[105,106]. In addition, it has multiple operational modes that can be used to perform different kinds of

discovery and quantification[107]. The theory of operation of triple quadrupole mass spectrometers is similar to that of a single quadrupole mass spectrometer. However, using consecutive quadrupoles in QQQ instruments, it is possible to perform multiple mass filtrations simultaneously with this more sophisticated approach. The working principle of the MRM mode of the triple quadrupole instrument is demonstrated in **Figure 1.9**.

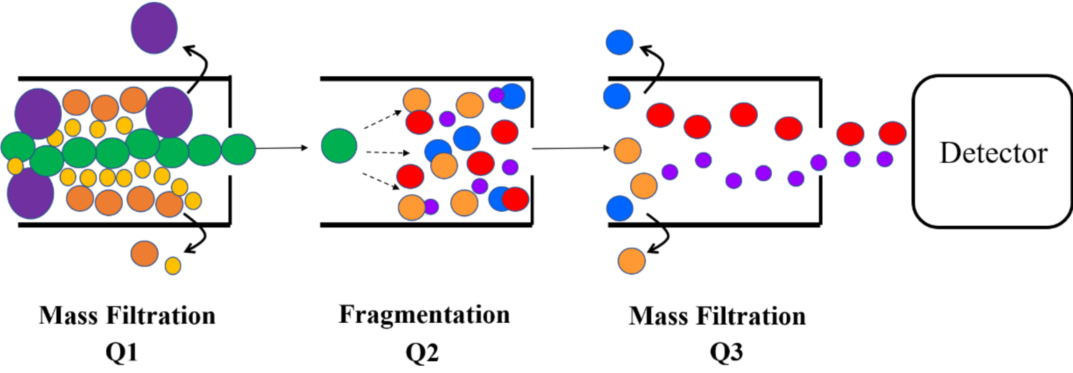

Figure 1.9. *The schematic representation of MRM mode on a QQQ instrument*

To explain the diagram, gas-phase ions are produced in the ion source, and they enter the first quadrupole (Q1 or MS1). The specific m/z values of intact peptide (parent ion or precursor ion) are picked and directed to the second quadrupole (Q2). Fragmentations of the precursor ions form when gas atoms collide with precursor ions in the second quadrupole. The second quadrupole is also named the 'collision cell,' and the process is well-known as 'collision-induced dissociation' (CID). These terms will be explained in detail in the following section. Finally, specific fragment ions are selected among all those produced fragment ions in the third quadrupole (Q3 or MS2)[108]. These chosen fragment ions are also described as 'daughter ion or product ions'. Then, product ions are sent to the instrument detector. In contrast to typical MS techniques, MS/MS systems enable a mass analysis to happen sequentially in various locations of the instruments[109].

Multiple reaction monitoring (MRM) or selected reaction monitoring (SRM) is a scanning mode of the QQQ-MS. This mode is used to quantify target proteins in biological specimens. The first and third quadrupoles serve as mass filters, while the second quadrupole produces fragment ions, as mentioned above. These targeted precursor and product ion pairs are also known as 'MRM transition'[110]. The precursor ions of the peptides that identify the protein are specified as targets in MRM-based protein quantification. These peptides' fragment ions are generated, and the earlier selected product ions are monitored. Since only MRM transitions of the target peptide are scanned in this method, the instrument's selectivity is significantly enhanced[111]. In addition, in the MRM process, only the MRM transition of one peptide is seen in the separated time window defined for each scan. As a result, its selectivity increases considerably as other transitions do not interfere. The increase in selectivity enables quantifying low abundant proteins in highly complex biological environments[112]. Furthermore, this approach offers absolute or relative protein quantification. Relative protein quantification measures variations in protein concentration in biological samples, while absolute protein quantification defines changes in protein concentration in biological samples based on standards[110,111].

To summarize this, QQQ-MS is a sensitive and precise analytical method for quantifying proteins and peptides in biological environments. Although QQQ instruments have low resolution, they offer great specificity for protein analysis[113,114]. The MRM/ SRM, which is a targeted MS-based method, improves the selectivity and sensitivity of the QQQ instruments. Because of its high specificity, precision, and accuracy, the MRM/ SRM mode of QQQ-MS is considered a good fit in protein and/or peptide quantification in complex biological mixtures[115], and more than one peptide can be quantified at the same time in a single analysis[110].

1.2.4 The Vacuum System

A high vacuum system is required for MS instruments. A high vacuum is generally generated by two pumps which are mechanical and turbomolecular pumps. These pumps provide very low pressure between 10^{-3} and 10^{-6} torr[86,116] to separate analyte ions by eliminating the collision efficiently.

The ions in the gas phase are accelerated due to transferred from the ion source to the detector, which is the main reason to use a high vacuum. Because the produced ions are highly reactive and have a short lifetime[86], they should arrive at the detector as fast as possible without colliding with other molecules such as air molecules[116]. Vacuum system forms mean free paths to prevent undesired collisions because the analyte ion can collide with another molecule, may be neutralized, scattered, reacted, or fragmented. This procedure causes the loss of ions used in detection. This collision changes ions' velocity or their forms. Therefore, interference caused by collisions is observed in the mass spectrum due to the formation of molecules or any change in the ion because of the collision. In other words, a high vacuum is used to improve the mean free path of ions significantly. The mean free path is described as the average distance a molecule travels without interacting or colliding with another molecule.

1.2.5 Mass Resolution and Resolving Power

Mass resolution is used to define the ability to separate two mass spectral peaks that have the closest m/z ratio depending on their masses and widths[117]. Mass resolution is among the most crucial factors in determining the efficiency of a mass analyzer since it signifies the quality of the instrument's performance. The following equation is used to calculate the term:

$$R = \frac{m}{\Delta m}$$

where ‘R’ denotes mass resolution, ‘m’ denotes a specific mass-to-charge ratio, and ‘Δm’ denotes the smallest difference in a specific m/z. The formula clearly shows that mass resolution does not have a unit.

The ‘mass resolving power’ is often confused with the term mass resolution. Mass resolving power, also known as resolving power, is described as the ability to separate two closest mass spectral peaks with the same intensity or height.

1.2.6 Collision Induced Dissociation

The gas molecules collide with one another, and collisions between molecules result in any apparent or important chemical change. In MS instruments, the collision takes place in the collision cell. This compartment contains inert gas molecules such as helium (He), nitrogen (N_2), and argon (Ar) at low pressure. Furthermore, nitrogen gas is the most common and easily accessible inert gas.

The ions become energized and have translational energy when an electric potential is applied to them[118]. As the energized precursor ions collide with the inert gas molecules because of the applied electric voltage, the energy is transformed into molecular vibration. As a result of these vibrations, ion fragments are generated. This phenomenon is known as collision-induced dissociation (CID)[119] or collision-activated dissociation (CAD).

When peptide fragment ions with low energy are produced in the collision cell, many ions form during the CID process, as represented in **Figure 1.10**. The most common transition ions are y-type and b-type ions, which are observed upon cleavage of the amide backbone due to apparent collision-induced dissociation. That is, they are ions produced upon the peptide bond cleavage. As a result, a significant number of overlapping b- (N-terminal ions) and y-type ions are formed (C-terminal ions)[120,121]. The energy needed to form fragment ions is referred to as collision energy (CE).

Figure 1.10. *The interpretation of six types of transition on Arg-Cys-Lys residues*

CE must be optimized for each peptide because it can differ from peptide to peptide. Therefore, it arranges optimal values for each target peptide to obtain proper peptide fragmentation. Nevertheless, if the collision energy is below the optimum value, precursor ions cannot be fragmented into product ions; however, if the collision energy is above the optimum value, fragment ions with high energy are generated instead of the low-energy fragment ions[122]. Therefore, it is one of the most significant factors influencing the quality of mass spectrometer data. Because the collision energies of and target peptide and protein vary, it is not easy to optimize this parameter[123].

A representative fragmentation of peptide FEVQVTVPK in a +2 charge state resulting from the CID process is depicted in **Figure 1.11**. The cleavage intervals in the spectrum are equal to the masses of amino acids because the peptide is divided from amino acid sites. **Figure 1.11** illustrates a method for fragmenting peptides. The intensity of the fragments created by the breakdown of each amino acid is depicted in this diagram.

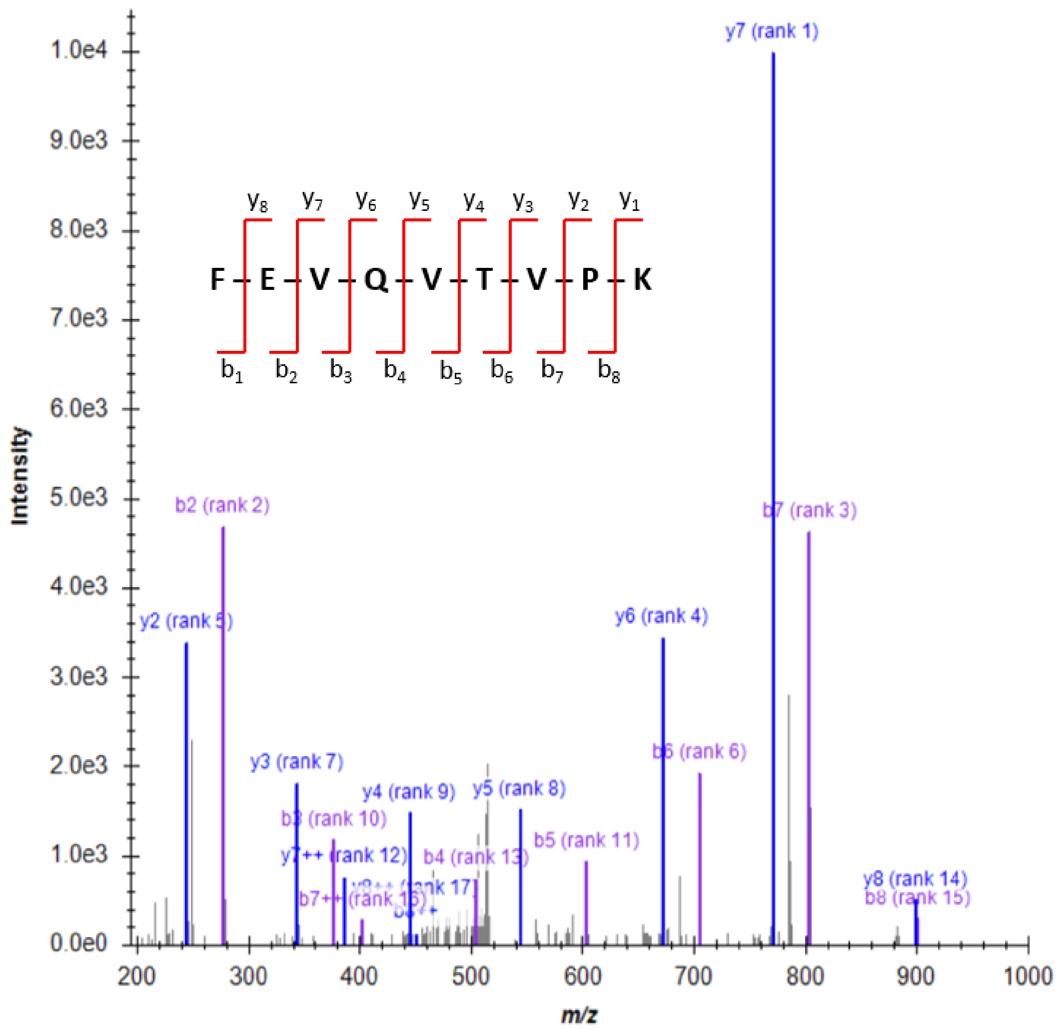

Figure 1.11. *The representation of fragmentation of peptide FEVQVTVPK in +2 charge state using ion trap MS result of NIST peptide library*

Peptide identification can be performed in-house or using publicly available peptide and/or protein databases such as the National Institute of Standards and Technology (NIST) Peptide Library. This technique can be used to identify peptides or to verify peptide identification. **Figure 1.11** was created using data from the NIST peptide library. It is assessed by considering ionization and collision conditions.

1.3 Model Protein: Alpha-2-Macroglobulin

Alpha-2-macroglobulin (A2MG) is one of the largest and the most abundant proteins in human serum. Alpha-2-macroglobulin is also known as C3 and PZP-like alpha-2-macroglobulin domain-containing protein 5. It primarily acts as a broad-spectrum proteinase inhibitor[124]. A2MG is a multifunctional protein that plays a crucial role in blood homeostasis, and its concentration in human blood decreases with age, from approximately 4.0 to 1.5 mg/ml[125]. Hence, it takes part in blood homeostasis and aging-related diseases, it is of great clinical importance[126]. A2MG is selected as a reference protein because of its clinical significance and its well-defined structure.

1.3.1 The Clinical Significance of A2MG

A2MG is a secreted plasma protein, which plays an essential role in biological processes. It is predominantly synthesized by the liver in the human body. On average, it is found as 1.5-2.0 mg/ml in human blood plasma[127] and 1.0-3.6 mg/ml in cerebral spinal fluid[128]. This protein is directly and indirectly associated with the serious diseases described below.

One role of A2MG in the body is in the immune system. It plays a role in inflammation. When the protein forms an A2MG-proteinase complex at the inflammation sites, it inhibits the proteinases[129,130]; meaning that, it is has a role in the dysregulation of the immune system in inflammatory diseases. For instance, it and inhibits aspartic proteinases such as, aspartic proteinase of HIV[131]. Also, A2MG has a role in the regulation of cytokines in inflammatory processes. When A2MG is oxidized, its regulation ability changes, and it starts regulating cytokines differently[131,132]. And the change in the activity of cytokinesis is connected to the tumor development[133,134].

It also plays an active role in the coagulation system. A2MG repairs and destroys tissues. It forms a complex with the activated protein C[135]. The activated protein C

is an essential regulator of thrombosis. This complex decreases the anticoagulant activity of the activated protein C, and thrombin is produced[135]. When the activated protein C deficiency occurs, it causes fatal thrombotic diseases such as venous thromboembolism[136] and uremia[137–139].

Another role of A2MG in the body is due to its effects on growth factors. It potentiates growth factor signaling, as pro-nerve growth factor. Hence, it can inhibit this growth factor's activity[140–142]. It can cause the accumulation of misfolded proteins in the neuro-system. In this way, A2MG is related to neurological diseases such as Alzheimer's disease[143,144], Parkinson's disease[128,140], and motor neuron diseases[145]. Also, A2MG is linked to many other diseases such as cardiac diseases[146], diabetes[147], prostate cancer[148], chronic liver disease[149], obesity[150], and so much more.

To sum up, A2MG is a plasma glycoprotein that can be discovered in various biological fluids, including blood, serum, and saliva. The liver produces most of A2MG, in addition to the locally producing in macrophages, fibroblasts, and adrenocortical cells. It is biologically active because it inhibits a wide range of proteases and works as a disease protection barrier due to its capacity to bind to foreign peptides and particles. A2MG can assist the reversible or irreversible capture of proteins with various biological activities with its different reactive sites. Understanding the regulation of proteostasis by A2MG and homolog proteins is now possible. A2MG has become a biomarker for various disorders over time. Consequently, the search for the A2MG protein is critical in clinical applications.

1.3.2 The Structure and Function of A2MG

A2MG is a secreted homotetrameric protein composed of four identical monomeric subunits. The A2M gene produces and codes for the protein alpha-2-macroglobulin.

The sequence of A2MG consists of 1474 amino acids, and it has a mass of 720 kDa[151,152]. In its sequence, the signal peptide consists of the first 23 amino acids,

whereas the chain includes the remaining amino acids (between amino acids 24 and 1474). The amino acid sequence of the A2MG protein, which has the protein ID P01023 in the UniProt[153,154] database, was retrieved from this database.

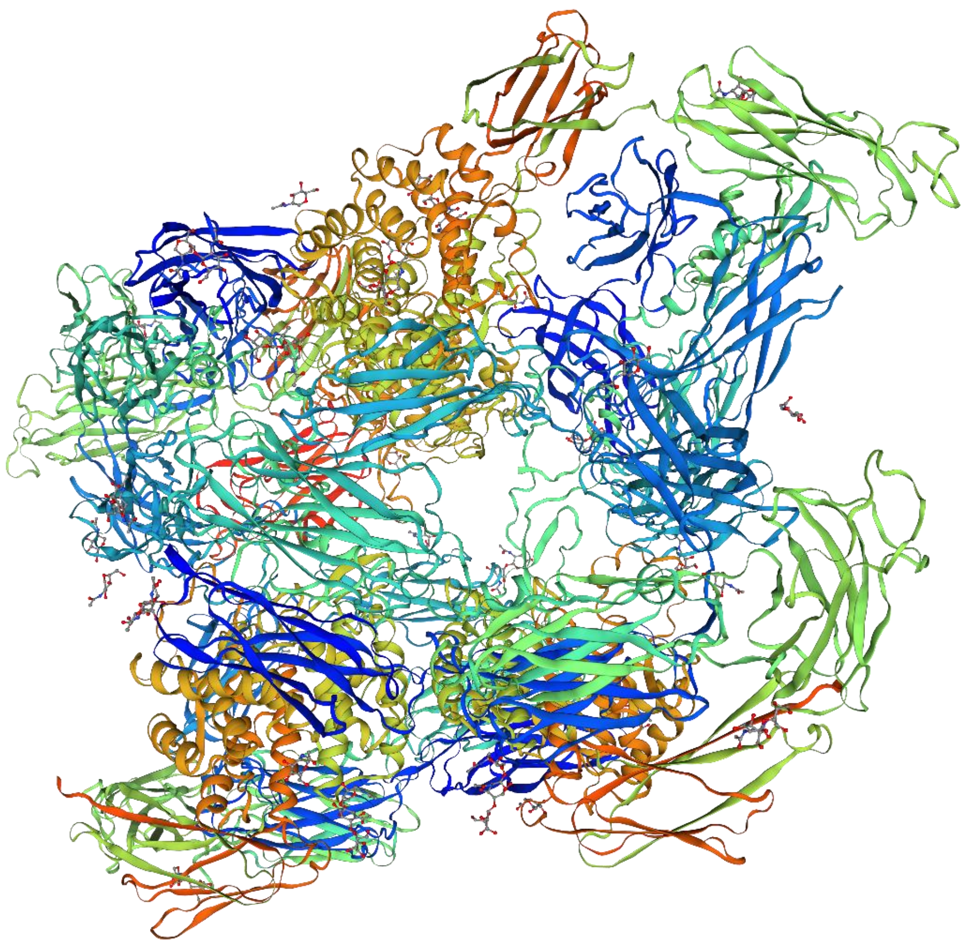

Figure 1.12. *The 3D structure of A2MG from ExPasy Swiss Model Database*

As illustrated in **Figure 1.12**[155], it is a tetrameric protein with four identical subunits. Furthermore, thanks to a unique 'trapping' mechanism, it can inhibit all four types of proteinases. It is also known as the 'Venus flytrap' mechanism[156,157]. The cage-like quaternary structure of this protein is a peptide region with specific cleavage sites for different proteinases. When a proteinase cleaves the bait region, the protein undergoes a conformational change that captures the proteinase. When entrapped,

the enzyme remains active against low molecular weight substrates, but its activity against high molecular weight substrates decreases dramatically. Following the cleavage at the bait region, a thioester bond is hydrolyzed, allowing the protein to form a covalent bond with the proteinase. This unique trapping mechanism has yet to be explained entirely in the literature[131]. The bait region in the 3D structure of A2MG protein is located between the amino acids 690 and 728 in the polypeptide chain sequence[131]. However, the exact location of the bait region in 3D structure of the protein is not known. The one of the most recent research, AlphaFold[158] interface, predicts the A2MG protein's bait region per-residue confidence score (pLDDT) that is between 22.31 and 59.57 and the mean of the score is 27.91. It is clear that the confidence score in this region is very low (pLDDT<50). This indicates that the precise location of this region has not been identified due to a lack of information.

A2MG is engaged in various biological processes due to its several activities according to UniProt and Reactome[159] databases. The biological mechanisms involved are as follows[160] negative regulation of complement activation, the lectin pathway, negative regulation of endopeptidase activity, stem cell differentiation, etc. A2MG has some of the PTMs mentioned in the previous protein modifications section. As with other proteins, this modification impacts the structure, function, and biological mechanisms of the A2MG protein. The PTMs of this protein will be reviewed in the following section.

1.3.3 The Modifications of A2MG

The A2MG protein structure and its modifications were investigated through the UniProt database. There are three major PTMs of this protein: N-linked glycosylation, cross-linking, and disulfide bonding, as indicated in **Table 1.1**[153].

The A2MG protein sequence contains chemically active amino acid residues. Because these amino acids are reactive, they can cause various undesirable reactions such as oxidation, acetylation, phosphorylation, and methylation. However,

chemically combining compounds with two or more reactive ends capable of chemically binding to particular functional groups on proteins by a covalent bond is known as cross-linking. The crosslinking is enabled by a free thiol group in each subunit of A2MG protein[152,161].

A disulfide bond is a type of PTM in proteins that occurs when the sulfur atoms of two cysteine residues come together during the cell's production. Disulfide bonds are critical in protein folding because they affect both the structure and function of the protein. In the literature, A2MG has two intrachain disulfide bonds. In addition, the protein structure has eleven intrachain bridges[162]. Disulfide bonds bind four subunits together to form covalently coupled dimers, which non-covalently associate to complete A2MG's cage-like quaternary structure[152,157]. The bait region, which is the A2MG protein's unique trapping mechanism, is constructed in this manner.

The attachment of sugar to the asparagine (N) amino acid residue is known as N-linked glycosylation which is described previous section (Part 1.1.2). A2MG has eight N-linked glycosylation sites. Also, these PTMs with locations at the protein were summarized in **Table 1.1**.

Table 1.1. *List of post-translational modifications in UniProt database*

Post-translational Modifications	**Position in the sequence**
Disulfide bond	278, 431, 48 ↔ 86, 251 ↔ 299, 269 ↔ 287, 470 ↔ 563, 595 ↔ 771, 642 ↔ 689, 821 ↔ 849, 847 ↔ 883, 921 ↔ 1321, 1079 ↔ 1127, 1352 ↔ 1467
N-Glycosylation	55, 70, 247, 396, 410, 869, 991, 1424
Cross-link	693, 694, 972 ↔ 975

1.4 Motivation

In peptide-centered bottom-up proteomics, unique peptides are used for qualitative and quantitative purposes. However, a single protein often has more than one unique tryptic peptide sequence. Although there are non-consensus criteria for selecting unique peptides, no rules and regulations have been issued by authorities for clinical applications. Common practice is to use the most abundant peptide(s) for biomarker research and report the outcome in protein level. Thus, the peptide representing the same target protein may differ from one study to another. While general assumption is that all unique peptides representing the same protein behave the same, the dynamic protein-unique peptide relationship is yet to be discovered.

This study is attempted to understand quantitative protein-unique peptide relationship. To this end, conventional proteomics sample processing procedure is applied to A2MG protein standard, human and bovine serums at different concentration levels and proteins were digested using two most common proteases. The outcomes of this research have potential to be utilized for clinical applications towards effective and reproducible biomarker research.

CHAPTER 2

MATERIALS AND METHODS

2.1 Materials

A2MG protein standard, fetal bovine serum, and human serum were obtained from Sigma-Aldrich (St. Louis, MO USA). Reference human serum was purchased from European Reference Materials (ERM®- DA470k/IFCC, Geel, Belgium). LC-MS grade water, acetonitrile, and formic acid were bought from Merck (Darmstadt, Germany). Dithiothreitol (DTT), iodoacetamide (IAA), and ammonium bicarbonate (ABC) were purchased from Sigma-Aldrich (St. Louis, MO USA). Sequencing grade trypsin and trypsin/Lys-C enzyme mixture were bought from Promega (Madison, WI USA), and a different brand of sequencing grade trypsin was also purchased from Roche (Mannheim, Germany). The peptides and internal standard were synthesized by PeptiTeam (Ankara, Turkey). An InfinityLab Poroshell 120 EC C18 (3.0 x 150 mm, 2.7 microns) reverse-phase column was used for peptide separation in LC-MS systems from Agilent (Santa Clara, CA). 2-20 µL, 20-200 µL, and 100-1000 µL range Eppendorf micropipettes were used for preparing stock solutions, working solutions, and standards. Polypropylene (PP) microtubes and falcon tubes were used during all experimental processes instead of glass tubes since glass surfaces cause loss of proteins and contaminations.

2.1.1 Samples

The similarities and differences in the behavior of unique peptides in biological systems were investigated using reference systems such as the peptide standard, the protein standard, bovine and human serum, as shown in **Figure 2.1**.

The peptide and protein standards were used for MS performance evaluation and analytical method optimization. The behavior of unique peptides in protein standards and serum samples was studied under different biological conditions.

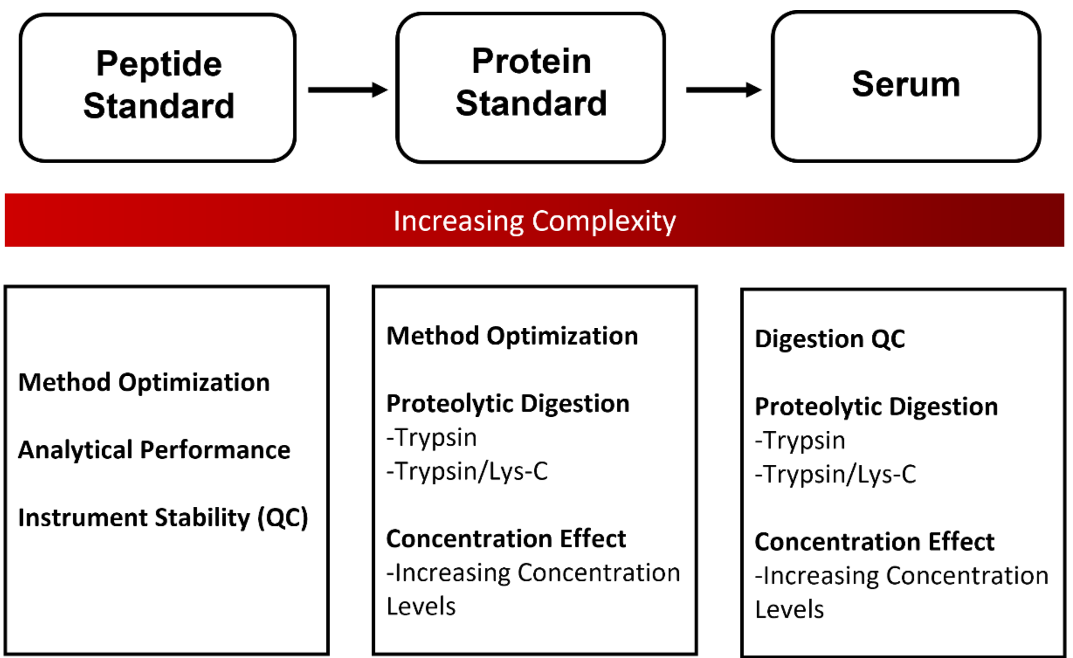

Figure 2.1. *The reference systems which were used in the study*

- ***Preparation of Protein and Serum Standard Solutions***

The protein and serum standards were prepared according to the manufacturer's instructions. After preparing a 50 mM ABC buffer solution with a pH of 7.8, a 495 µl of 50 mM ABC and 5 µl of pure formic acid were taken to prevent mold formation in the solution for a long time and added to 1 mg A2MG protein standard vial to

prepare 2.0 µg/µl stock solution. Then, stock solution was diluted with ABC buffer solution to obtain a final concentration of 0.2 µg/µl.

ERM serum samples were prepared by pipetting 990 µl of 50 mM ABC and 10 µl of formic acid to the ERM standard vial, followed by gentle shaking in particular periods, and the final concentration is calculated as 1.59 µg/µl.

- ***Peptide Standard Solutions***

Lyophilized five peptide standards were dissolved in 1% (v/v) formic acid and vortexed to yield 10 µg/µl stock solutions. The prepared stock solutions were listed in **Table 2.1**. Five peptide standard solutions were mixed so that the final concentration of the stock peptide mixture was 1.8 µg/µl and diluted with ABC buffer solution to working solutions with concentrations of 0.5 µg/µl and 0.05 µg/µl, respectively.

Table 2.1. *Peptide and internal standard stock solutions and their concentrations*

Peptide Sequence	Solution Type	Amount	Purity	Concentration
FEVQVTVPK	Peptide standard	1.03 mg	99.60%	9.96 µg/µl
QGIPFFGQVR	Peptide standard	0.86 mg	90.0%	9.00 µg/µl
VGFYESDVMGR	Peptide standard	1.08 mg	98.8%	9.88 µg/µl
DMYSFLEDMGLK	Peptide standard	1.03 mg	98.0%	9.80 µg/µl
LPPNVVEESAR	Peptide standard	1.07 mg	98.4%	9.84 µg/µl
TFLLR	Internal standard	1.1 mg	>99%	10.0 µg/µl

Furthermore, lyophilized the internal standard was dissolved in 1% (v/v) formic acid and vortexed to obtain 10 µg/µl stock solution. Then, it was added to all samples.

- ***Dilution Solvent***

Dilution solvent was prepared by adding 25 µl of 10 µg/µl the internal standard solution and 50 µl pure formic acid to LC-MS grade water and bringing final volume up to 50 ml. As a result, 50 ml of dilution solvent was prepared with a final concentration of 0.05% (v/v) internal standard and 0.1% (v/v) formic acid.

- ***Quality Controls***

Two levels of quality controls (QCs) were used to check the instrument's stability and sample processing variation. Method performance was controlled to prevent systematic error. QCs were made for pooled human serum digest and a mixture of peptide standard samples. Serum QC and peptide QC were prepared as their concentration were 0.5 and 0.0065 µg/µl, respectively.

- ***Calibrants***

Calibrants were prepared in three different background matrices. Calibrants were prepared by adding the peptide standard mixture and internal standard with a final concentration of 0.0004 µg/µl.

- ***External Calibration***

The peptide standard mixture solution was added to formic acid to obtain nine levels with concentrations of 0.0005, 0.0035, 0.0065, 0.0095, 0.0125, 0.0155, 0.0185, 0.0215, and 0.0245 µg/µl.

- ***Matrix-Matched Calibration***

The peptide standard mixture solution was added to neat bovine serum to obtain five levels with concentrations of 0.0035, 0.0065, 0.0125, 0.0155, and 0.0215 µg/µl.

- ***Standard Addition Calibration***

The peptide standard mixture solution was added to neat human serum to obtain five levels with concentrations of 0.0005, 0.0035, 0.0065, 0.0095, and 0.0125 µg/µl.

2.1.2 Enzymatic In-Solution Digestion Procedure

The 5 µl thawed aliquots of serum and protein standard samples were transferred in 0.5 ml Eppendorf tubes, vortexed and digested according the digestion procedure is visualized in **Figure 2.2**.

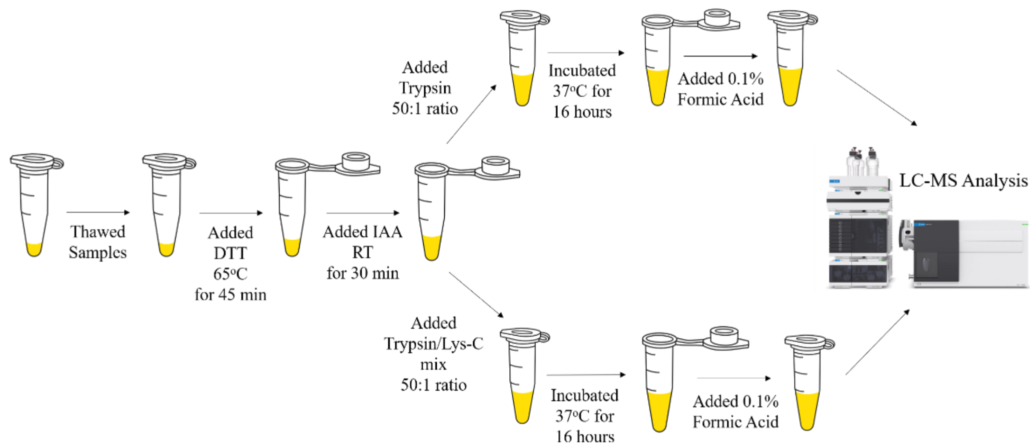

Figure 2.2. *The workflow of sample preparation for in-solution digestion*

- ***Reduction Process***

10 µl of 45 mM DTT solution was added to the thawed samples and mixed. The samples were left at 65°C for 45 minutes. The reduction process is necessary to reduce the disulfide bonds and protein denaturation. In addition, it is also used to prevent inter and intra-molecular disulfide formation between cysteine (C) amino acids in the protein.

- ***Alkylation Process***

10 µl of 100 mM IAA solution was added to the samples and vortexed. The samples were left in the dark for 30 minutes at room temperature before being alkylated. The alkylation process is necessary to stabilize free sulfhydryl groups.

And then, 50 mM ABC buffer solution was added to each sample so that the total volume was 100 µl. Half of the samples were treated with trypsin at a protein:enzyme

ratio of 50:1, while the other half were treated with a trypsin/Lys-C enzyme mixture at a protein:enzyme ratio of 50:1. All prepared samples were vortexed before being incubated at 37°C for 16 hours. After the incubation, a 1% (v/v) stock formic acid solution was added to the samples at a final concentration of 0.1% (v/v) to stop the digestion. During the experiment, fresh stock solutions were prepared. The samples were prepared for mass spectrometric analysis after the digestion.

2.1.3 The Sample Preparation for LC-MS Analysis

In total, 213 runs with QC and calibration samples were performed, and 108 samples were analyzed with LC-MS. A 150 µl of dilution solvent containing 0.05% internal standard was added into 40 µl of prepared serum and A2MG protein standard samples. The volume was finalized at 200 µl by adding 10 µl of 0.1% (v/v) formic acid to each sample set.

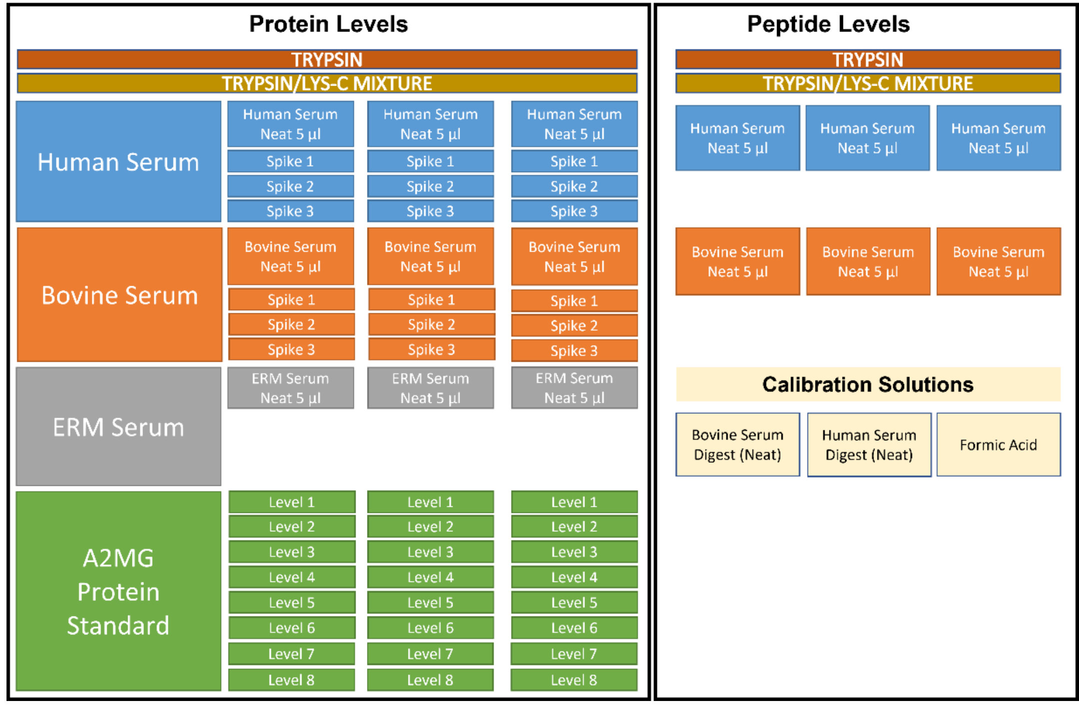

Figure 2.3. *Experimental study design for each sample set*

Consequently, the samples were diluted five times and vortexed immediately after each addition to ensure the solution homogeneity, which is important in protein measurements.

The samples were prepared and evaluated in triplicate and included human serum, bovine serum, ERM serum, and A2MG protein standard samples**. Figure 2.3** depicts the prepared work plan for the experiment. The concentration levels of samples indicated in this diagram were explained in detail below.

In order to investigate the analytical variability, two variables are controlled: repeatability and reproducibility[163]. The serum samples were analyzed in triplicate, and A2MG protein samples were measured in duplicate to control repeatability. The cycle of blanks and quality control (QC) samples was repeated after each sample set. The coefficient of variation (CV) was calculated to assess reproducibility.

The total protein content of human serum has been reported to be between 6 and 8 g/dl[164]. The total A2MG protein concentration in human serum has been found as 1.5 mg/ml to 2.4 mg/ml[165–167]. Also, total protein concentration in fetal bovine serum has been reported as 3.0-4.5 g/dl by manufacturers. Consequently, the values used in the calculations are based on approximate values (total protein in human serum 7.0 g/dl (70 µg/µl), total A2MG protein in human serum 2.0 mg/ml (2.0 µg/µl), and total protein in bovine serum 4.0 g/dl (40 µg/µl)).

5 µl of human serum samples were digested, and the total protein and A2MG protein concentrations in samples were theoretically calculated as 1.5909 and 0.0455 µg/µl, respectively. A2MG protein standard samples were spiked into human serum samples at three concentration levels (0.0591, 0.0727, and 0.1000 µg/µl), where humans serum samples were also prepared at three concentration levels (20%, 30%, 50% of 5 µl of serum was added to the human serum samples).

ERM human serum has been certificated to have A2MG concentration of 1.43 g/l (1.43 µg/µl) with a 0.06 g/l uncertainty. After sample preparation, the final total

protein and A2MG protein concentrations in 5 µl of ERM human serum samples were theoretically calculated as 1.5909 and 0.0325 µg/µl, respectively.

5 µl of bovine serum samples were prepared, and the total protein concentration was calculated as 1.2121 µg/µl. A2MG protein standard was spiked into bovine serum at three concentration levels (0.0030, 0.0061, and 0.0121 µg/µl).

A2MG protein standard samples were digested at eight different concentrations (0.0071, 0.0143, 0.0214, 0.0357, 0.0536, 0.0714, 0.0893, and 0.1071 µg/µl).

2.2 LC-MS Analysis

2.2.1 Instrumentation

Quantitation of all samples was performed by an Agilent 1260 Infinity II HPLC system coupled to an Agilent 6470A triple-quadrupole (QQQ) system (Santa Clara, CA). Agilent MassHunter Quantitative Analysis software was used for data acquisition and processing.

2.2.2 LC-MS Conditions

- ***LC Conditions***

Mobile phase A was LC-MS grade water containing 0.1% formic acid, and mobile phase B was pure acetonitrile containing 0.1% formic acid. Gradient elution was used for LC separation, which involved altering the compositions of the mobile phases with time. In addition, the flow rate was adjusted to 0.3 ml/min. The needle wash was performed in the autosampler at a draw speed of 100 l/min to avoid unnecessary carry-over. The separation gradient was 15% of mobile phase B during 0 and 4 minutes and increased to 65% at 16.1 minutes. Then it was held at 65% for 2.4 minutes before dropping back to 5%. It remained at for the last 3 minutes. The

Agilent InfinityLab Poroshell 120 C18 (3.0 x 150 mm, 2.7 µm) column was used as an HPLC column. The reversed-phase C18 column was incubated during analysis in an oven at 50°C for better separation.

- ***MS Conditions***

The unique peptides were introduced to the QQQ by ESI with Jet Stream of Agilent Technologies after being separated in the LC system. In the Q1 and Q3 compartments, each scan window was operated at a unit resolution. In the analysis, the dynamic MRM mode was selected with retention time windows of 2 minutes. The highest sensitivity for all target peptides and transitions was achieved by optimizing MS instrument conditions described in **Table 2.2**:

Table 2.2. *The MS operating parameters*

Gas temperature, °C	300
Gas flow, L/min	11
Nebulizer pressure, psi	40
Sheath gas temperature, °C	400
Sheath gas flow, L/min	11
Capillary voltage, V	3500
Collision energy, V	6.9 – 22.4

The mass spectrometer was performed under a positive polarity with MRM acquisition parameters and was optimized so that the electron multiplier voltage (EMV, +) value was 500. The cycle time was 500 ms, while the dwell time was between 125 and 167 ms. The voltage of the fragmentor was set to 135 V. Collision energy varies for each peptide and is in the range displayed in **Table 2.2**. The following sections provide detailed information on the collision energies of the unique peptides. The identification of each targeted peptide of A2MG protein was based on three or more transitions; it is clarified following sections in this chapter.

2.3 Targeted Proteomics Method Development

After optimizing LC-MS settings, the Multiple Reaction Monitoring (MRM) method was developed following the processes outlined below. The information presented in this section was expanded in Chapter 3.

- ***The Selection of Unique Peptides***

The unique peptides of the A2MG protein were assessed utilizing protein and/or peptide databases widely used in the proteomics field after the tryptic peptides were determined. It was revealed that fifteen of the 124 tryptic peptides were unique peptides, listed in **Table 2.3**. The two unique peptides, QTVSWAVTPK and GGVEDEVTLSAYITIALLEIPLTVTHPVVR, are excluded from the study.

Table 2.3. *The list of A2MG unique peptides*

Peptide No	Unique Peptide Sequence
1	IAQWQSFQLEGGLK
2	FEVQVTVPK
3	QGIPFFGQVR
4	LHTEAQIQEEGTVVELTGR
5	HNVYINGITYTPVSSTNEK
6	DMYSFLEDMGLK
7	VGFYESDVMGR
8	LVHVEEPHTETVR
9	QTVSWAVTPK
10	DTVIKPLLVEPEGLEK
11	LPPNVVEESAR
12	GGVEDEVTLSAYITIALLEIPLTVTHPVVR
13	ALLAYAFALAGNQDK
14	AAQVTIQSSGTFSSK
15	FQVDNNNR

- ***Localization of Unique Peptides in A2MG Protein***

Using the PyMOL[168] software (1.7.4.5 Edu, Schrödinger LLC), the positions of thirteen unique peptides on the three-dimensional A2MG protein structure were determined. The positions of peptides in the structure's inner, outer, and bait regions, together with their closeness to PTM sites like glycosylation, were assessed.

- ***Determination of Retention Times***

The unscheduled run was used to determine the retention times of thirteen unique peptides. For each charge state of each peptide, six transitions were determined, and the retention time was determined from the protein standard digest. The retention time of the unique peptide ALLAYAFALAGNQDK could not be established due to an ionization problem, which was excluded from further research. The detailed information was provided in Chapter 3.

- ***Determination of Predominant Charge states***

Twelve unique peptides were screened in the +1, +2, and +3 charge states. The predominant charge states for each unique peptide were identified from these three charge states.

- ***The Selection of MRM Transitions***

Three or four MRM transitions were detected between six transitions for twelve unique peptides after determining the predominant charge states of each unique peptide. One transition was chosen as a quantifier, while the others were utilized as qualifiers.

2.4 Data Analysis

Raw MS data were taken from the Agilent QQQ instrument in .d files format. The results were loaded into the Skyline[169] software package (20.2.0.343, MacCoss Lab, UW). Data pre-processing was done using Skyline software. It is software to develop

methods for mass spectrometry-based protein quantification and analyze quantitative data. It is free to use in both academic and commercial[169].

2.4.1 Data Pre-Processing

The Skyline software was used to remove interferences and reintegrate obtained peaks of samples. When the results were reviewed via Skyline, interferences were detected and manually removed. In addition, peak reintegration was completed manually after interference filtration. Peak reintegration was used to confirm that the peak areas of the unique peptides are consistent in all samples.

2.4.2 Statistical Data Processing

The pre-processed MS data was subjected to statistical data analysis. The following sections provide a detailed description of the data treatment using various techniques.

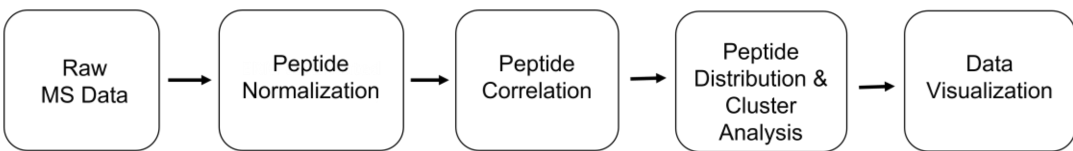

Figure 2.4. *The illustration of the statistical data treatment process*

The MS data were normalized, correlated, and distribution and clustering analyses were done, respectively. The general data treatment workflow is shown in **Figure 2.4**. Statistical data were treated and visualized by using OriginPro 2018 SR1 (b9.5.1.195, OriginLab Corporation) software.

- ***Peptide Normalization***

Normalization is a statistical approach used in proteomics studies to make data more comparable and representative by minimizing non-biological systematic variations[45,170,171]. Four kinds of normalization transformations were performed, which are peak area, Log2, median, and peak area of the sample to the internal standard ratio.

- ***Peptide Correlation***

The correlation is a statistical technique to evaluate the relationship between multiple continuous variables[172]. Two different correlation coefficients were calculated on A2MG unique peptides: Pearson and Spearman. In brief, the Pearson correlation coefficient represents a linear relationship between multiple continuous variables. In other words, as the relationship between variables changes, it gradually decreases or increases. The Spearman correlation coefficient, on the other hand, demonstrates a monotonic relationship between them. It means that one variable rises while the other falls, or vice versa, indicating a non-linear relationship between them. By using the calculated correlation coefficients of unique peptides of A2MG, their relationship was investigated with multivariate analysis.

- ***Peptide Distribution and Cluster Analysis***

The distribution and hierarchical cluster analysis were examined. These analyses visualized complex data sets so that the relationship between them can be observed appropriately. Therefore, charts as a heatmap, a dendrogram, and a combination of them were plotted. Color mapping with hierarchical cluster analysis was used to investigate twelve unique peptides of A2MG. Heat map analysis changes the order of the peptides in rows and columns to show the similarities and/or differences of the peptides[173]. The colors provide an efficient way to visualize the peptides' familiarity in biological samples.

CHAPTER 3

RESULTS AND DISCUSSION

3.1 MRM Method Development

3.1.1 The Selection of Unique Peptides of A2MG Protein

It is essential to use unique tryptic peptides which are specific to protein for qualitative and quantitative proteomics. The standard approach in clinical studies is to use a limited number of unique peptides for quantitative analysis[128,157]. This study assesses behavior of A2MG unique peptides in biological environments. For this purpose, the selection of unique peptides for the study is of crucial importance.

In order to investigate the dynamic unique peptide-protein relationship, A2MG was used as a reference protein. First, the A2MG protein sequence was obtained from the UniProt database, and then tryptic peptides were calculated using the ExPasy PeptideCutter[174] interface. As a result of this analysis, 124 tryptic peptides were obtained. **Figure 3.1** shows the enzyme cleavage sites. The lysine (K) and arginine (R) sites in the A2MG protein's 3D structure are highlighted differently. Lysine (K) residues are color-coded purple, while arginine (R) residues are color-coded yellow. 80 of the 124 tryptic peptides were cleaved from the lysine (K) residues, while 44 of them were cleaved from the arginine (R) amino acid residues. As stated in an earlier section (Part 1.1.4), since the trypsin enzyme cuts lysine (K) amino acid residues with poor digestion efficiency, Lys-C protease that cleaves from the C-terminal side of the lysine (K) amino acid residues might be added to digestion to increase the efficiency.

The uniqueness of the peptides was determined using three independent databases: UniProt, NextProt[175,176], and BLAST[177]. The outcomes of the search were compared

to ensure protein specificity. At the end of this analysis, 15 unique peptides were determined among 124 tryptic peptides. These unique peptide compositions, their positions, and sizes are shown in **Table 3.1** below. The number of amino acids they contain varies between 8-30.

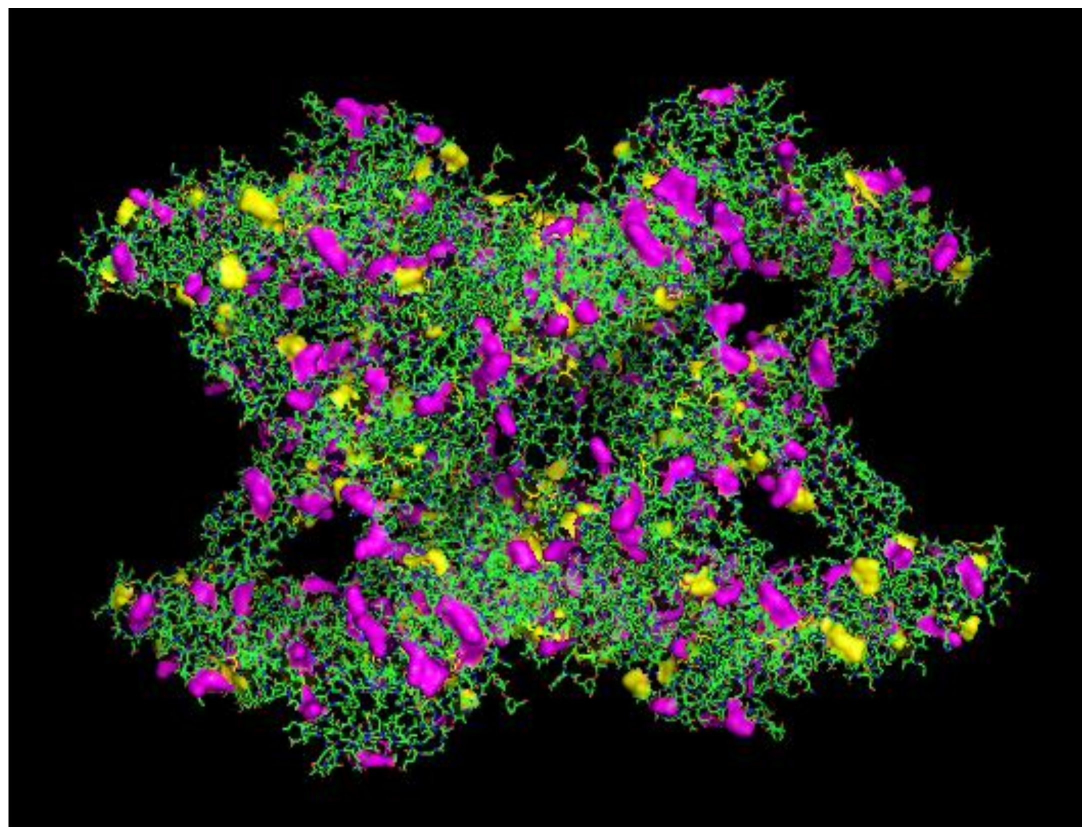

Figure 3.1. *Positions of lysine and arginine in the A2MG protein. The purple color code is used for lysine residues and the yellow color code for the arginine residues*

We further investigated the behavior of these unique peptides and their correlations under different biological conditions. Fifteen unique peptides were further filtered, and two were excluded from the study. This research did not include the unique peptides QTVSWAVTPK and GGVEDEVTLSAYITIALLEIPLTVTHPVVR.

Peptide length influences mass spectrometry-based sequence identification. The optimum peptide length range to be studied for unique peptide is given in the

literature as 8-25 amino acids[178]. Peptides, including six amino acids or less [41], are small peptides, as they are not specific peptides.

Table 3.1. *List and information of unique tryptic peptides in A2MG protein*

Position of the cleavage	Position of Peptide Sequence	Peptide Sequence	Peptide Length (Number of amino acids)
188	175-188	IAQWQSFQLEGGLK	14
237	229-237	FEVQVTVPK	9
370	361-370	QGIPFFGQVR	10
338	320-338	LHTEAQIQEEGTVVELTGR	19
664	646-664	HNVYINGITYTPVSSTNEK	19
676	665-676	DMYSFLEDMGLK	12
715	705-715	VGFYESDVMGR	11
732	720-732	LVHVEEPHTETVR	13
863	854-863	QTVSWAVTPK	10
912	897-912	DTVIKPLLVEPEGLEK	16
945	935-945	LPPNVVEESAR	11
1122	1093-1122	GGVEDEVTLSAYITIALLEI PLTVTHPVVR	30
1162	1148-1162	ALLAYAFALAGNQDK	15
1289	1275-1289	AAQVTIQSSGTFSSK	15
1297	1290-1297	FQVDNNNR	8

The peptide GGVEDEVTLSAYITIALLEIPLTVTHPVVR is also not in the optimal peptide length range. It was excluded from the study because too-long peptides interfere with the enzyme’s digestion ability, resulting in incomplete enzymatic digestion. On the other hand, because of a surrogate matrix, the unique peptide QTVSWAVTPK was also not considered in the study as a unique representative

peptide. It is not just a unique peptide for humans. It is also found in other animals such as Cattle, Sumatran orangutans, brown rats, and house mice.

3.1.2 The Localization of Unique Peptides

The locations unique peptides in 3D structure of A2MG protein was investigated. This research was made using the RCSB[179] protein database and the PyMOL[168] software, which is a tool to visualize proteins and/or peptides.

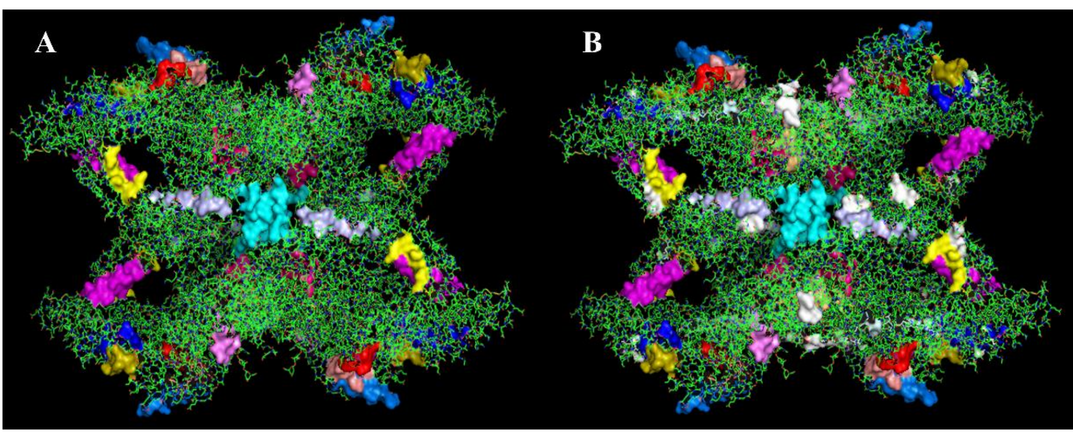

Figure 3.2. *The 3D model of A2MG protein in PyMOL software.* ***A.*** *The locations of unique peptides in the protein structure.* ***B.*** *The localization of unique peptides and glycosylation of the protein structure.*

Table 3.2. *The color codes of unique peptides in Figure 3.2*

Sequence Position	Peptide Sequence	Color Name	Color Code
175-188	IAQWQSFQLEGGLK	Red	
229-237	FEVQVTVPK	Yellow	
320-338	LHTEAQIQEEGTVVELTGR	Lilac	
361-370	QGIPFFGQVR	Pale Blue	
646-664	HNVYINGITYTPVSSTNEK	Cyan	

Table 3.2. Continued

665-676	DMYSFLEDMGLK	Pink	
705-715	VGFYESDVMGR	-	-
720-732	LVHVEEPHTETVR	-	-
897-912	DTVIKPLLVEPEGLEK	Blue	
935-945	LPPNVVEESAR	Olive	
1148-1162	ALLAYAFALAGNQDK	Violet	
1275-1289	AAQVTIQSSGTFSSK	Sea Blue	
1290-1297	FQVDNNNR	Salmon	

The locations of the peptides in the A2MG protein structure were established as a result of this research. **Figure 3.2A** shows the distribution of unique peptides on the protein's structure; **Figure 3.2B** includes glycosylation sites to the unique peptide distribution. Each unique peptide is represented in **Figure 3.2** by a different color, and the white areas represent the glycosylation sites in **Figure 3.2B**. Furthermore, **Table 3.2** indicates the color codes for unique peptides. The PyMOL protein simulation of A2MG protein does not contain the bait region, between 690 and 728 amino acid sequences. The unique peptides VGFYESDVMGR and LVHVEEPHTETVR were found in the protein structure’s bait region, so their structural position could not be determined in the simulation[71]. By analyzing the positions of the unique peptides in the simulation, the following observations can be made:

The unique peptide FQVDNNNR is on the outside of the protein, exhibited by the salmon color code. The unique peptides AAQVTIQSSGTFSSK and IAQWQSFQLEGGLK are very close to it. The unique peptide FEVQVTVPK is also found on the outside of the protein, as shown by the yellow color code. The unique peptide LHTEAQIQEEGTVVELTGR and glycosylation site are very near to the unique peptide FEVQVTVPK. The light blue unique peptide QGIPFFGQVR is near

the glycosylation site and is found in the protein's outer region. The unique peptide LPPNVVEESAR encoded in olive green is very close to the unique peptide DTVIKPLLVEPEGLEK and is found in the protein's outer region. The protein's interior contains the unique peptide DMYSFLEDMGLK, shown by the vibrant pink color code. The unique peptide IAQWQSFQLEGGLK, marked by the red color code, is found outside the protein, close to the unique peptide FQVDNNNR. The ALLAYAFALAGNQDK unique peptide, shown in violet, is observed on the protein's surface. The unique peptide AAQVTIQSSGTFSSK, denoted by the color code sea blue, is located on the outside of the protein, close to the unique peptide FQVDNNNR. The unique peptide DTVIKPLLVEPEGLEK, shown in blue, is located on the protein's outer side, very close to the glycosylation site, and is adjacent to the peptide LPPNVVEESAR. The unique peptide LHTEAQIQEEGTVVELTGR, shown in magenta, coexists with the unique peptide FEVQVTVPK and the unique peptide remains on the protein's surface. The unique peptide HNVYINGITYTPVSSTNEK, encoded in cyan, is located at the center of the protein's monomers.

Table 3.3. *The list of unique peptides located near to glycosylation*

Peptide Sequence	**Peptide Length**	**Peptide Position**	**PTM**
FEVQVTVPK	9	229-237	Next to Glycosylation at 247
QGIPFFGQVR	10	361-370	Next to Glycosylation at 410
DTVIKPLLVEPEGLEK	16	897-912	Next to Glycosylation at 869

As mentioned before, PTM is one of the important parameters when examining a protein's structure. When the unique peptides on the protein and the glycosylation

sites in **Figure 3.2B** are investigated, the unique peptides are not found at these locations. However, it appears to be located with the unique peptide FEVQVTVPK, QGIPFFGQVR, and DTVIKPLLVEPEGLEK near to glycosylation, shown in **Table 3.3**. The digestion efficiency of these unique peptides may be altered due to their close locations to glycosylation sites.

3.1.3 The Association of Unique Peptides with Literature

After determining the location of unique peptides on A2MG protein, the incidence of unique peptides used in studies published in the literature was analyzed. The frequency of unique peptides utilized in the literature for identification and/or quantification of A2MG protein is listed in **Table 3.4**.

Table 3.4. *The number of publications of A2MG unique peptides in literature*

The Number of Publications	Peptide Sequence
13[48,50,52,61,63,180–187]	IAQWQSFQLEGGLK
18[46,48,50,52,58,60–63,180–185,188–190]	FEVQVTVPK
13[48,52,58,61,63,180,183–187,191,192]	LHTEAQIQEEGTVVELTGR
18[46,50,52,58,59,61–63,180,181,183–188,193,194]	QGIPFFGQVR
10[48,50–52,61,180,185–187,195]	HNVYINGITYTPVSSTNEK
11[48,50,52,59,61,180,184–188]	DMYSFLEDMGLK
13[46,48,50,52,61,63,180,183–188]	VGFYESDVMGR
14[46,48,50,52,58,61,63,180,185–188,193,194]	LVHVEEPHTETVR
15[46,48,50,52,58,59,61,63,180,181,183–185,187,188]	DTVIKPLLVEPEGLEK
18[46,48,50,52,57,59,61–63,180–183,185–189]	LPPNVVEESAR
11[48,52,58,62,63,180,182,183,185,187,189]	AAQVTIQSSGTFSSK
9[46,52,63,180–183,185,186]	FQVDNNNR

The relationship between the number of publications in this table and the unique peptides has been visualized using a word cloud graph to make it easier to interpret in **Figure 3.3**. The peptides with larger font are frequently studied, while the peptides with a smaller font are investigated rarely in the literature based on data on **Table 3.4**.

The prevalence of unique peptides in the literature is portrayed in **Figure 3.3**, revealing the four unique peptides, FEVQVTVPK, LPPNVVEESAR, QGIPFFGQVR, and DTVIKPLLVEPEGLEK, are frequently encountered. The peptide FQVDNNNR, on the other hand, has received the least amount of consideration in the literature.

Figure 3.3. *A word cloud plot represents the incidence of A2MG unique peptides in the literature.*

When the relationship between peptide variation and the incidence of unique peptides in the literature is investigated, it has been observed that the four most studied unique peptides were found on the protein surface. Furthermore, the lengths of the frequently used unique peptides range from 9 to 16 amino acids.

Researchers usually choose a unique peptide to represent a protein at random or one which appears in high abundance on the screen. Although the selection of unique peptides to represent the protein is not regulated, the literature has established certain criteria for selecting unique peptides. Peptide length, consecutive amino acids that affect trypsin efficiency and/or ragged ends (KK, RK, RR, and KR), chemically active amino acids in the peptide sequence, and peptide locations on the protein tertiary structure are all factors to consider as criteria for selection of unique peptides.

3.1.4 The Selection of Unique Peptide Standards

When the localization of unique peptides and their literature investigation were examined, the unique peptide DMYSFLEDMGLK was selected as the best unique representative peptide because it is the only unique peptide located in the interior. Other unique peptides can be found in the protein's outer regions. The glycosylation sites are surrounded by the unique peptides FEVQVTVPK, QGIPFFGQVR, and DTVIKPLLVEPEGLEK, which may impact digestion products. One of these unique peptides can be chosen as the representative specific peptide when studying the effect of post-translational modifications quantitatively. Chemical reactions can occur when reactive groups in free amino acids and proteins, such as amino, carboxyl, sulfhydryl, phenolic, hydroxyl, thioether, imidazole, and guanyl, bind to other small organic molecules. These reactions can alter the physicochemical and functional properties of proteins and peptides, resulting in post-translational modifications. The order of reactivity of amino acids is as follows[196]:

Cysteine (C)> Methionine (M)> Tryptophan (W)> Tyrosine (Y)> Histidine (H)> Leucine (L), Isoleucine (I)> Arginine (R), Lysine (K), Valine (V)> Serine (S),

Threonine (T), Proline (P)> Glutamine (Q), Glutamate (E)> Aspartate (D), Asparagine (N)> Alanine (A)> Glycine (G)

Significant post-translational modifications are caused by the deamidation of glutamine (Q) and asparagine (N) active amino acid residues[197]. In addition to this information, the criteria used to select unique peptides are summarized in **Table 3.5** as peptide length, peptide active amino acid residues, and enzyme efficiency. Each unique peptide is colored red in the table to indicate the number of chemically active amino acid residues.

Table 3.5. *The criteria for identifying unique representative peptides*

Peptide Sequence	Peptide Length	Peptide Position	Sequential AA affecting trypsin efficiency	Chemically Active Amino Acid Residues					
				Methionine (M)	Cysteine (C)	Glutamine (Q)	Tryptophan (W)	Histidine (H)	Asparagine (N)
IAQWQSFQLE GGLK	14	175-188		0	0	3	1	0	0
FEVQVTVPK	9	229-237	K/E	0	0	1	0	0	0
LHTEAQIQEE GTVVELTGR	19	320-338	EK, R/K	0	0	2	0	1	0
QGIPFFGQVR	10	361-370		0	0	2	0	0	0
HNVYINGITY TPVSSTNEK	19	646-664		0	0	0	0	1	3
DMYSFLEDM GLK	12	665-676		2	0	0	0	0	0
VGFYESDVM GR	11	705-715		1	0	0	0	0	0

Table 3.5. Continued

LVHVEEPHTE TVR	13	720-732	R/K	0	0	0	0	2	0
DTVIKPLLVE PEGLEK	16	897-912	RK	0	0	0	0	0	0
LPPNVVEESA R	11	935-945		0	0	0	0	0	1
ALLAYAFAL AGNQDK	15	1148-1162	K/RK	0	0	1	0	0	1
AAQVTIQSSG TFSSK	15	1275-1289		0	0	2	0	0	0
FQVDNNNR	8	1290-1297		0	0	1	0	0	3

This section includes information on all A2MG unique peptides. In addition, the FEVQVTVPK peptide is unique in that it only has one active residue, glutamine (Q) amino acid. The glutamic acid (E) comes after lysine (K) amino acid residue, making trypsin difficult to cleavage. A glycosylation site surrounds this unique peptide at position 247 in the sequence. The unique peptide QGIPFFGQVR contains two of the amino acids glutamine (Q) with the active residue and is close to the glycosylation site located at position 410 on the sequence. One of the active residues methionine (M), is found in the VGFYESDVMGR unique peptide. The PyMOL protein simulation does not include location of this peptide in A2MG protein structure, known as the bait region. As a practical matter, its location within the protein structure is unknown, as mentioned previous section (Part 1.3.2). Two of the methionine (M) active amino acid residues are present in the unique peptide DMYSFLEDMGLK. Unlike other peptides found on the protein's surface (outer) region, this peptide is found in the inner region of the protein.

These unique peptides, located at various locations on the protein, were preferred to represent the protein best. The unique peptide LPPNVVEESAR is located on the outside of the protein. Because it has two proline (P) residues too close to the cleavage site, which may affect enzymatic digestion because 'Keil Rule'[73] says that trypsin cleaves sequences from arginine (R) or lysine (K) residues if they are not placed next to a proline (P).

When the chemically active amino acid residues of the unique peptides that are frequently used in the literature are examined, the unique peptides FEVQVTVPK and QGIPFFGQVR contain glutamine (Q), whereas the peptide LPPNVVEESAR contains asparagine (N). However, the unique peptide LPPNVVEESAR has two prolines (P) near its cleavage site. Proline (P), as previously stated, has a negative impact on activity of trypsin. The last unique peptide DTVIKPLLVEPEGLEK does not contain an active amino acid residue, yet it is a peptide that is made as a consequence of the impact of proline (P) on trypsin activity. The trypsin enzyme does not cut it because proline (P) is located after lysine (K) amino acid.

In this study, we aimed to cover all A2MG unique peptides. However, three of the fifteen unique peptides were eliminated, and twelve unique peptides were analyzed in serum and protein standard samples. In addition, five unique synthetic representative peptides, were purchased depending on cost viability, were also investigated. The selection was based on peptide intensity, frequency of observation in the literature, and its positions in the protein. These are:

1. FEVQVTVPK

2.QGIPFFGQVR

3.VGFYESDVMGR

4. DMYSFLEDMGLK

5. LPPNVVEESAR

The synthetic peptide TFFLR was used as an internal standard (IS). The most important goal of using IS peptide is to improve the accuracy and precision of the analysis[198]. Although the unique peptides were chosen for the best candidates to present A2MG protein, the expected results may be influenced by the active residues and the factors affecting the enzymatic digestion performance. Outputs may be obtained differently from the expected results.

3.1.5 The Determination of Retention Times

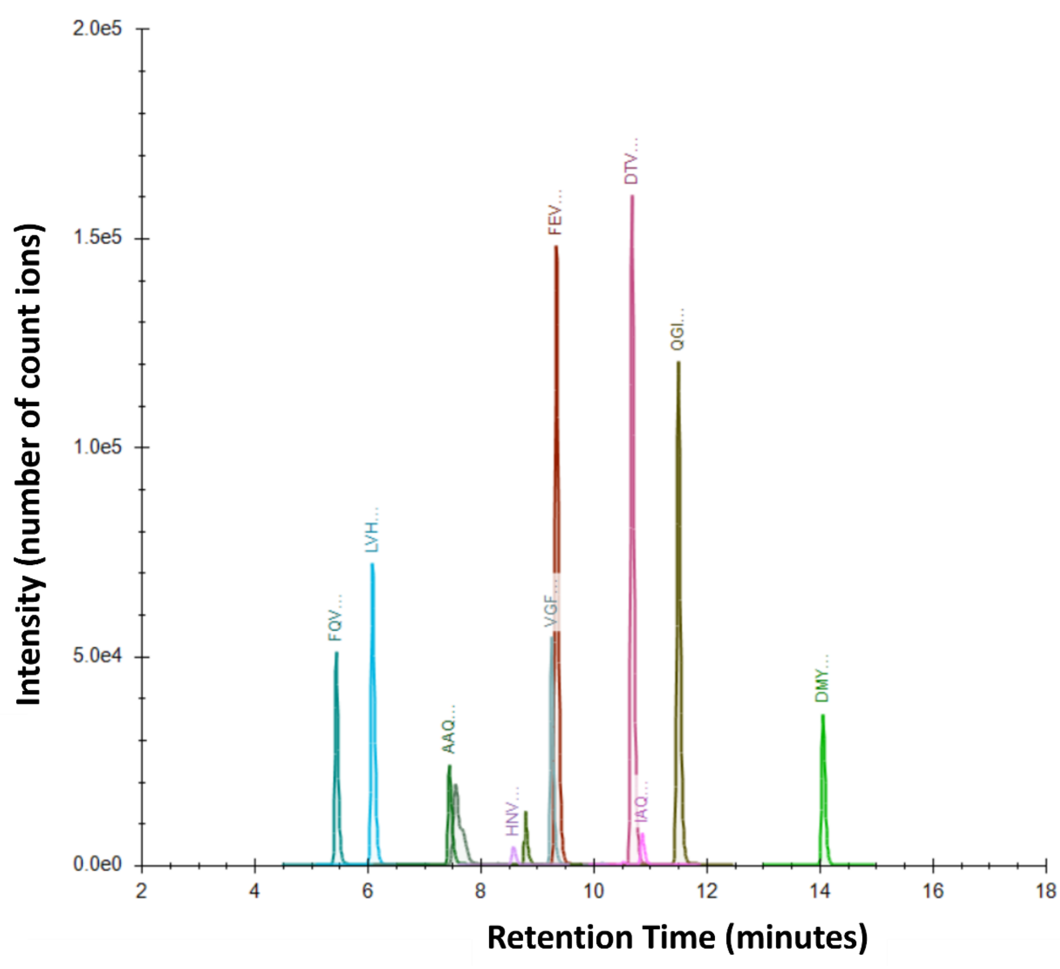

Figure 3.4. *The determined retention times of 12 unique peptides in serum digest*

An unscheduled run was used for determining the retention time of unique peptides in this study, which are seen in **Figure 3.4** and listed in **Table 3.6**. Retention times have been determined over the gradient run. The MRM approach was used to detect each peptide one by one with retention time windows.

The retention time of the unique peptide ALLAYAFALAGNQDK could not be determined since the peptide was not ionized sufficiently. Because the unique peptide's intensity was low to analyze, it was removed in further investigation. After that, the peptides' predominant charge states were selected in the following step.

Table 3.6. *The retention times of unique peptides*

Peptide Sequence	Retention Time (min)
IAQWQSFQLEGGLK	10.9
FEVQVTVPK	9.4
LHTEAQIQEEGTVVELTGR	8.8
QGIPFFGQVR	11.5
HNVYINGITYTPVSSTNEK	8.6
VGFYESDVMGR	14.0
LVHVEEPHTETVR	9.3
DMYSFLEDMGLK	6.1
DTVIKPLLVEPEGLEK	10.7
LPPNVVEESAR	7.6
ALLAYAFALAGNQDK	-
AAQVTIQSSGTFSSK	7.5
FQVDNNNR	5.5

The retention times of some unique peptides in **Figure 3.4** are very close to each other. They separated from each other by using narrow retention time windows specific to each peptide's transitions.

3.1.6 The Determination of Predominant Charge States

As mentioned earlier (Part 2.3), the charge states of proteins and peptides are mainly influenced by their molecular weight and the number of accessible basic sites such as arginine (R), histidine (H), and lysine (K)[199]. All of the target unique peptides were screened with commonly used charge states of +1, +2, and +3. Because it is unstable, charge states +4 and higher are not employed[200]. Some target peptides have an abundant and interference-free peak in the +2 charge state, while others have a more intense peak in the +3 charge state.

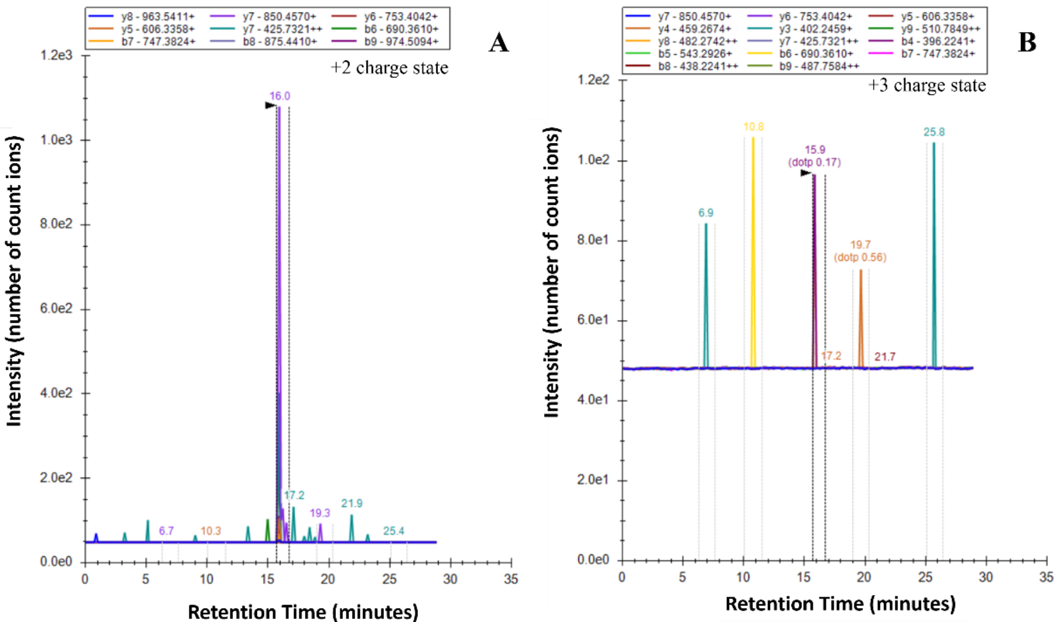

Figure 3.5. *The representation of determining the predominant charge state on target unique peptide QGIPFFGQVR in A2MG protein standard.* ***A.*** *The chromatogram of the +2 charge state of the peptide.* ***B.*** *The chromatogram of the +3 charge state of the unique peptide*

The heavier target unique peptides have higher charge states due to more ionization sites, increased number of basic sites, as seen in **Figure 3.5**. As an example, **Figure 3.5** shows the charge states of the unique peptide QGIPFFGQVR. **Figure 3.5A**

depicts the peptide in a +2 charge state, while **Figure 3.5B** represents the peptide in a +3 charge state. The peptide QGIPFFGQVR in the +2 charge state has a higher abundance than the +3 charge state. In addition, quite intense interferences were seen in the +3 charge state chromatogram. Hence, the peptide in a +2 charge state was selected as the predominant charge state.

Briefly, targeted unique peptides' charge states were selected depending on their abundance and interfering conditions. The most intense and interference-free charge state was carefully chosen for each target peptide, which was shown one by one in **Table 3.7**.

Table 3.7. *The selected predominant charges of unique peptides*

Peptide Sequence	Number of Basic Sites (R, H, K)	Molecular Weight (g/mol)	Predominant Charge States
IAQWQSFQLEGGLK	1	1604.80	+3
FEVQVTVPK	1	1046.22	+2
LHTEAQIQEEGTVVELTGR	2	2110.28	+3
QGIPFFGQVR	1	1148.31	+2
HNVYINGITYTPVSSTNEK	2	2137.31	+3
VGFYESDVMGR	1	1259.39	+2
LVHVEEPHTETVR	3	1545.70	+2
DMYSFLEDMGLK	1	1448.66	+3
DTVIKPLLVEPEGLEK	2	1780.07	+3
LPPNVVEESAR	1	1210.34	+2
AAQVTIQSSGTFSSK	1	1511.63	+2
FQVDNNNR	1	1006.03	+2

3.1.7 The Selection of MRM Transitions

The transition ions were picked from among the most common y- and b-ions in the proteomics field. The MRM transitions were listed in **Table 3.8**. In the analysis, the most intense MRM transitions in the consensus spectral libraries were chosen. The multiple transitions were selected to ensure specificity.

Table 3.8. *The list of selected MRM transitions of unique peptides of A2MG protein*

Peptide Sequence	Precursor ion (m/z)	Transition ion (y,b)	Transitions (m/z)	Collision Energy (V)
IAQWQSFQLEGGLK	535.62	y7 y6 b10	744.42+ **616.37**+ 616.31+	14.5
FEVQVTVPK	523.71	y7 y6 y5 b6	**770.48**+ 671.49+ 543.35+ 704.36+	14.1
LHTEAQIQEEGTVVELTGR	704.03	y9 y7 b7 b15	**931.52**+ 773.45+ 793.42+ 832.91++	20.5
QGIPFFGQVR	574.81	y7 y6 y5 y7	**850.46**+ 753.40+ 606.34+ 425.73++	15.9
HNVYINGITYTPVSSTNEK	713.02	y9 y8 y8 b11	962.48+ **861.43**+ 431.22++ 638.82++	20.9

Table 3.8. *Continued*

DMYSFLEDMGLK	724.83	y9 y8 y7	**1039.52+** 952.48+ 805.41+	21.3
VGFYESDVMGR	630.29	y10 y8 y7 y6	1160.50+ 956.41+ 793.35+ **664.31+**	17.9
LVHVEEPHTETVR	515.94	y11 y10 y7 b6	667.33++ 598.80++ **420.22++** 707.37+	13.8
DTVIKPLLVEPEGLEK	594.01	y6 y6 y14 y13	**672.36+** 336.68++ 782.47++ 732.94++	16.6
LPPNVVEESAR	605.83	y6 y10 y9 y9	690.34+ **549.28++** 1000.51+ 500.76++	17.0
AAQVTIQSSGTFSSK	756.39	y11 y10 y9 y8	**1142.57+** 1041.52+ 928.44+ 800.38+	22.4
FQVDNNNR	503.74	y7 y6 y5 y4	859.40+ **731.34+** 632.28+ 517.25+	13.3

Table 3.8. Continued

TFFLR (internal standard)	325.21	y4 y3 b3 b4	548.36+ **401.29+** 362.21+ 475.29+	6.9

When transitions with high noise are eliminated, desired protein can be monitored with high signal-to-noise. The selection was finalized into three or four transitions for each targeted peptide. One of the multiple transitions is a quantifier, and the others are qualifiers. All chosen transitions and collision energies of each target peptide were shown in **Table 3.8**, and quantifiers are represented in bold.

3.2 Targeted Protein Analysis

3.2.1 Data Pre-Processing

Interference-causing transitions were detected in the samples, as they allow interference to be scanned in Skyline. Furthermore, the program enables peak areas to be re-integrated.

- ***Interference Filtration***

The interference filtration process was executed using the Skyline software. Transitions that interfere with the transitions of target unique peptides were removed.

Figure 3.6 shows the chromatogram of the unique peptide FEVQVTVPK before the interference filtration process (**Figure 3.6A**) and the chromatogram of the same peptide after the process (**Figure 3.6B**). By removing the interference with a retention time of around 10.5 minutes, potential ion suppression of quantifier transition was eliminated, as shown in **Figure 3.6B**. And, when the interference-

causing transition was eliminated, the suppressed peak could be observed more clearly.

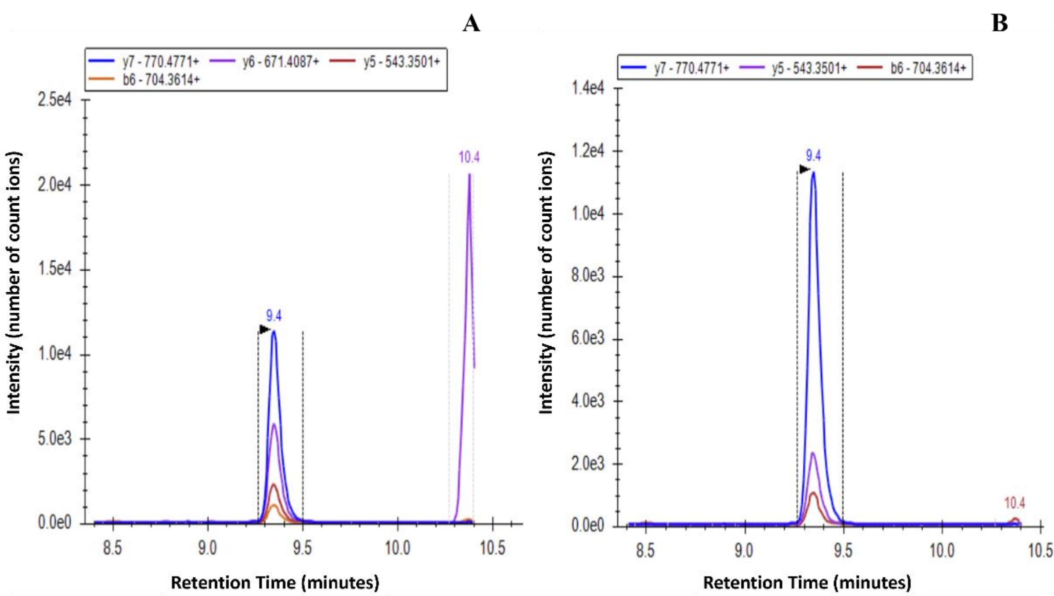

Figure 3.6. *The representation of interference filtration on the unique peptide FEVQVTVPK in Skyline software* ***A.*** *The peak of the unique peptide is suppressed by interference.* ***B.*** *The interference eliminated version of the chromatogram*

- ***Peak Reintegration***

The following step after the interference filtration is peak area reintegration. Peak reintegration was performed to ensure that the peak area of the unique peptides was consistent across all samples in the Skyline software.

Consistency is an essential parameter for data comparison. The integrated peak areas may differ from one another during automated peak area integration. Therefore, it is necessary to check the consistency between peak areas in all samples for each target unique peptide. Accordingly, the dissociation or association between the samples can be revealed.

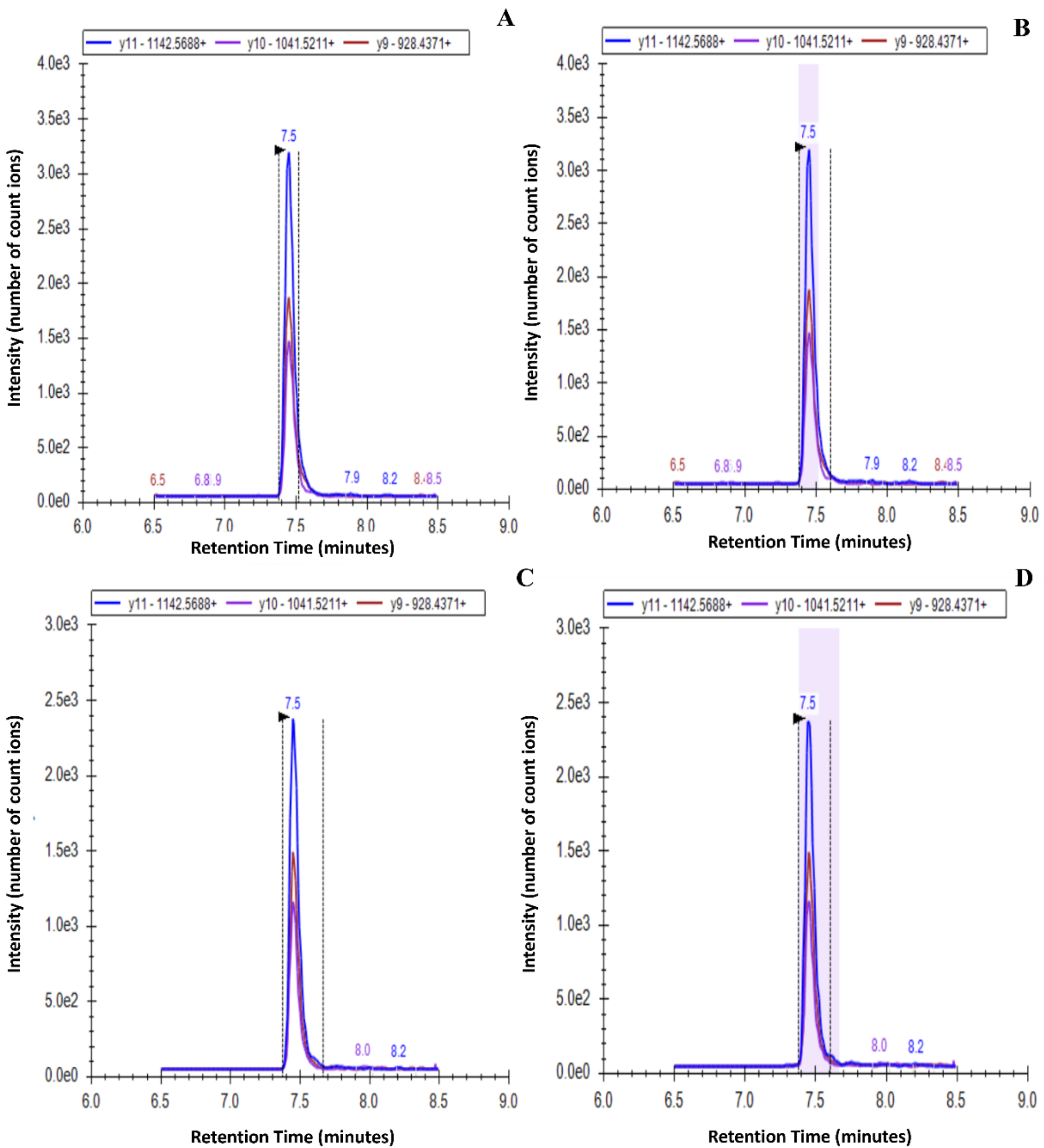

Figure 3.7. *The interpretation of reintegration on one of the unique peptides AAQVTIQSSGTFSSK in Skyline software.* ***A.*** *The auto-integrated peak of the peptide in the serum sample.* ***B.*** *The reintegrated peak of the peptide in the serum sample.* ***C.*** *The auto-integrated peak of the peptide in the A2MG protein standard sample.* ***D.*** *The reintegrated peak of the peptide in the A2MG protein standard sample.*

In **Figure 3.7**, the chromatograms of two samples are shown before and after the peak reintegration process. While the auto-integration peak area in the serum sample was narrow (**Figure 3.7A**), the peak area in the protein standard sample was broad

(**Figure 3.7C**). The peak area of the two samples after peak reintegration, on the other hand, were identical (**Figure 3.7B** and **Figure 3.7D**). The peak areas of the samples became consistent and comparable as a result of this process.

3.2.1 The MRM Method Performance

During the total four-day study period, the same sample was injected into the same volume at different time periods. The method of generating proteomics data on biological materials involves several phases. There are currently practical MS quantitation approaches that allow for highly precise quantification on a small or large number of specimens. One of the MS-based proteomics methods, the MRM method, was used in this investigation.

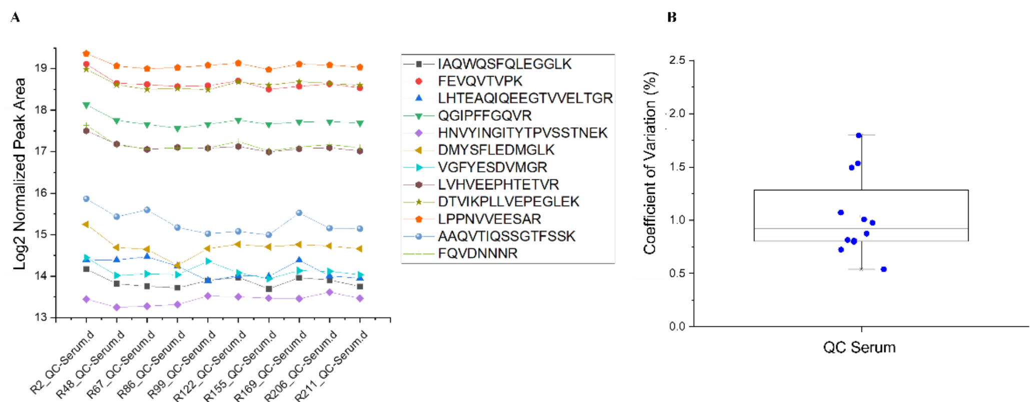

Figure 3.8. *The run stability of QC serum in the QQQ instrument* ***A.*** *The normalized peak area of each peptide* ***B.*** *CV values of unique peptides in QC serum*

The performance of the MRM technique is evaluated before studying the behaviors of peptides under various conditions to verify that the method is reproducible. The serum replicates were used to evaluate the run stability of the instrument using Log2 normalized QC serum replicates for each unique peptide in **Figure 3.8**. The run has decreased and then stabilized over time. **Figure 3.8A** shows log2 normalized peak

areas of peptides in the QC serum sample injected at various times, while **Figure 3.8B** shows CV values for each target peptide. It is seen that the CV values of each unique peptide do not exceed 2%, and the CV values are below the 20% value, which is accepted in the literature[201–203].

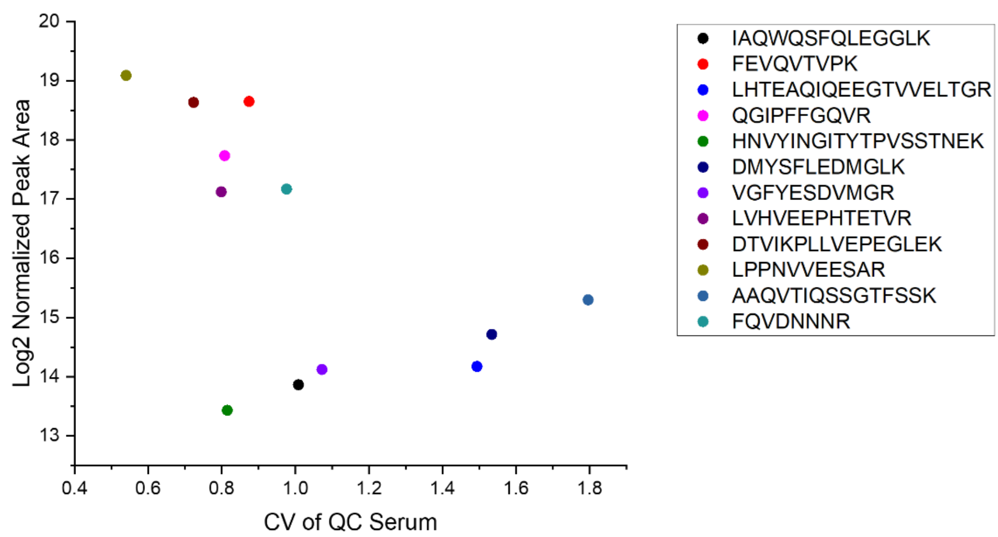

Figure 3.9. *The Log2 normalized peak areas versus CV values of QC serum*

The coefficient of variation of each unique peptide is shown in **Figure 3.9**. Each target peptide is represented with a different color. When the CV values of the peptides are compared, the unique peptides showing the highest variation are LHTEAQIQEEGTVVELTGR, AAQVTIQSSGTFSSK, and DMYSFLEDMGLK. On the other hand, the target unique peptides with the lowest variation are LPPNVVEESAR and DTVIKPLLVEPEGLEK. Furthermore, the unique peptide LPPNVVEESAR has the highest peak intensity while presenting the least variation. However, peak intensity does not decrease or show same behavior as variation increases.

The serum replicates were prepared using the synthetic unique peptide standards, which are listed in the previous section. The stability of representative peptide of serum replicates and transitions of its fragments were used to evaluate the method's performance (**Figure 3.10**).

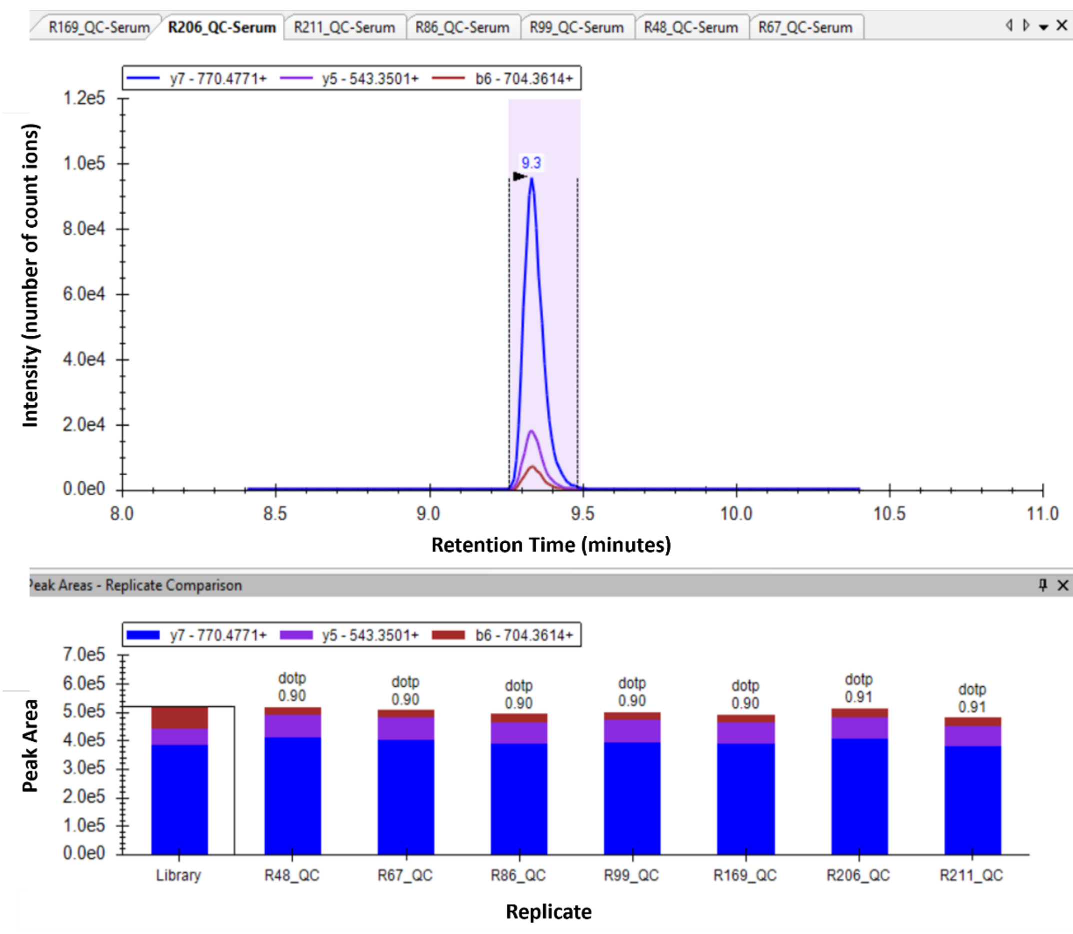

Figure 3.10. *The replicate stability of unique peptide FEVQVTVPK of QC serum in Skyline software*

This process was performed for all peptide standards. **Figure 3.10** shows the peptide FEVQVTVPK results in Skyline software. The upper window in the figure shows transitions of the unique peptide, and the lower window demonstrates the changes in transition weights of serum replicates and libraries. The stability of serum replicates

for a given unique peptide, as well as the consistency of serum replicates, demonstrates that the MRM method is highly reproducible.

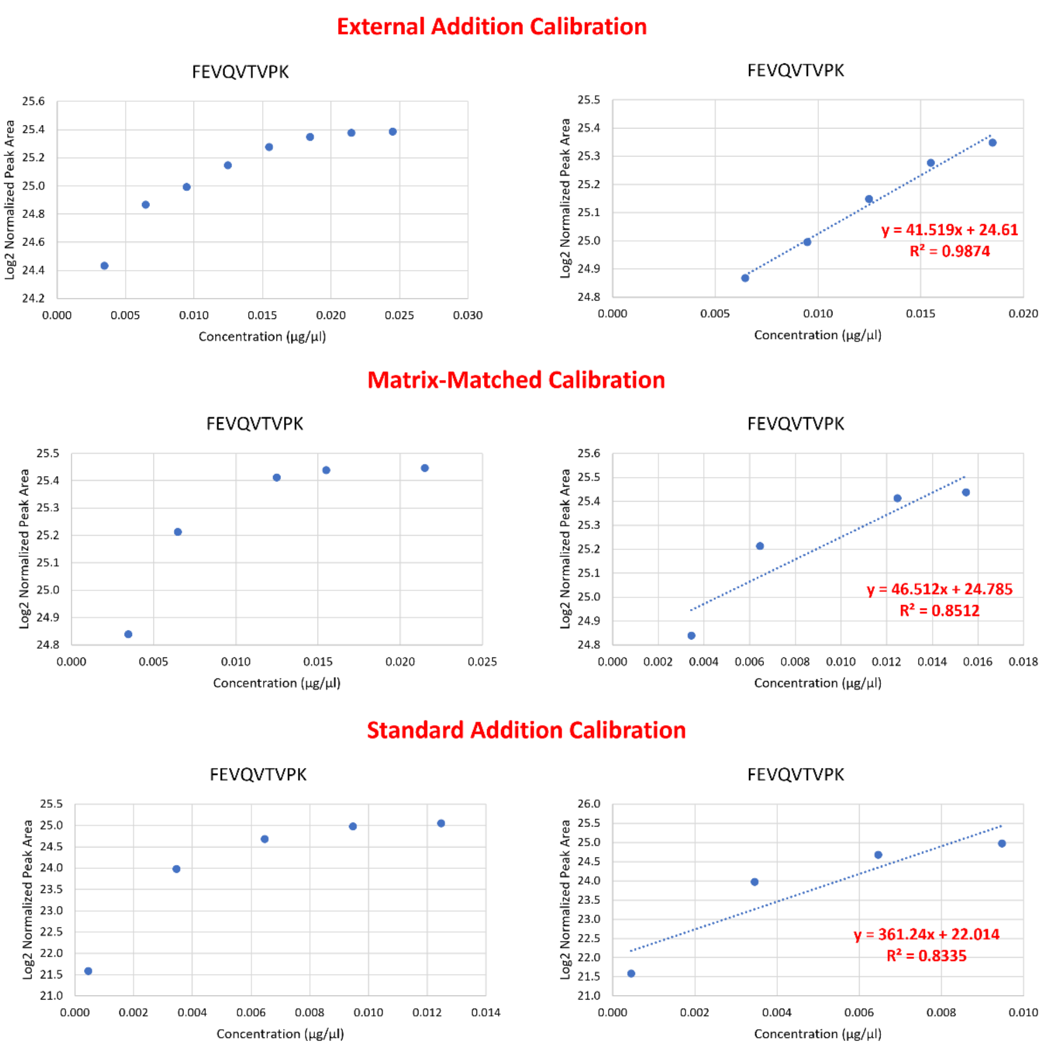

Figure 3.11. *The linearity of unique peptide FEVQVTVPK in formic acid, bovine serum and human serum samples*

The samples obtained by adding the peptide mixture were assessed for linearity. The linearity of a representative unique peptide FEVQVTVPK in different matrices was shown in **Figure 3.11**. Formic acid is used to generate the external calibration curve, bovine serum is used to prepare the matrix-matched calibration, and human serum is

used to prepare the standard addition calibration. External calibration curves are the most linear, whereas linearity decreases in matrix-matched and standard addition calibration curves, respectively. The results revealed that as sample complexity rises (matrix-effect), linearity diminishes.

3.2.1 The Digestion Reproducibility

Most recent bottom-up proteomics applications involve quantitative assessment of the absolute and relative protein quantities. And, protein quantification approaches are impacted from several factors such as sample preparation, MS recovery, and data analysis. Protein quantification depends on the efficiency and reproducibility of protein digestion[204]. Thus, the ability to create consistent digestions lack of missing or non-specific cleavages is critical for protein quantification research.

Until recently, investigations have been performed digestion products at the protein level; however, the digestion products of protein at peptide level were investigated in this study. For this reason, the reproducibility of digestion is examined in this section.

When investigating the digestion reproducibility, log2 normalized serum digest replicates' results were obtained for both enzymes as shown in **Figure 3.12**. Trypsin enzyme used replicates were displayed in **Figure 3.12A**, and enzyme mixture results were exhibited in **Figure 3.12B**. The following comments can be made by comparing the graphs: digestion is reproducible in highly abundant unique peptides for both enzymes, whereas the reproducibility is lower in less abundant unique peptides. At the bottom of the figures, related graphs of coefficient of variation of log2 normalized peak areas of neat human serum samples are displayed (**Figure 3.12C** and **Figure 3.12D**). In both experiments, the CV values are lower than 10%. It is within the acceptable CV range in the literature[201–203] (below 20%).

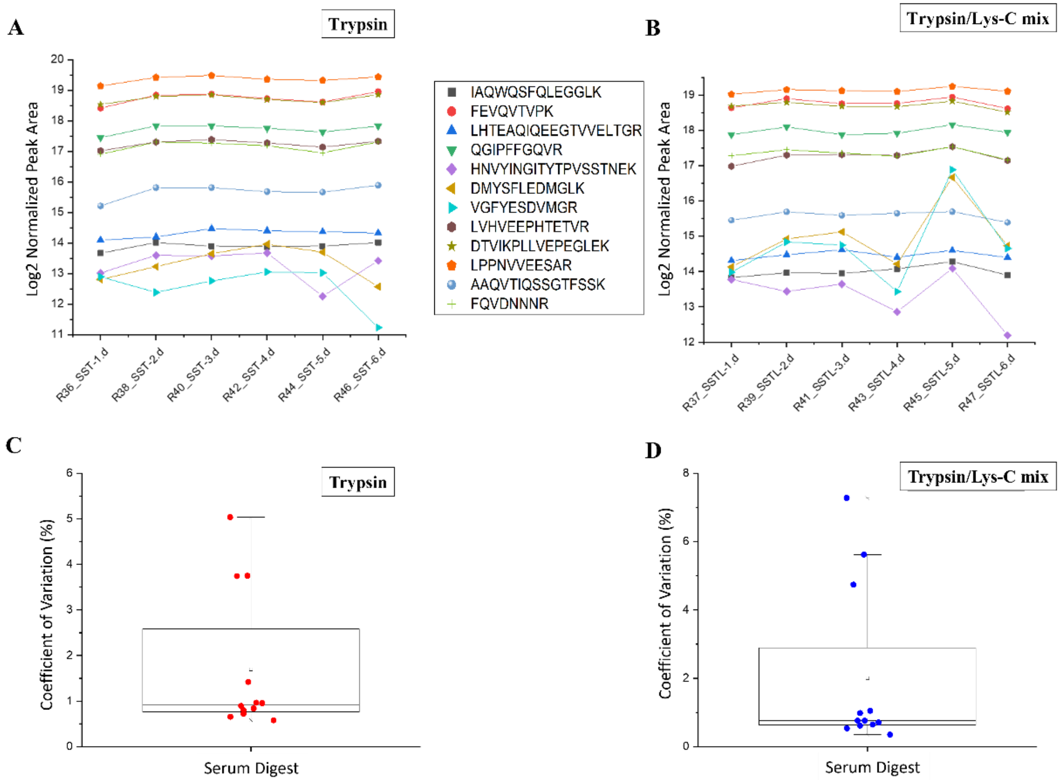

Figure 3.12. *The replicates of standard neat human serum samples for both enzymes.* ***A.*** *In the presence of trypsin, each peptide's normalized response is reported in replicates of neat human serum.* ***B.*** *In the presence of trypsin/Lys-C mixture enzyme, each peptide's normalized response is reported in replicates of neat human serum.* ***C.*** *CV values of peptides digested with trypsin* ***D.*** *CV values of peptides digested with trypsin/Lys-C mixture*

To investigate the coefficient of variation results in detail, CV values of each unique peptide and log2 normalized peak area results were combined. **Figure 3.13A** illustrates the results of trypsin digestion of neat human serum, while **Figure 3.13B** shows the results for the use of combination of enzymes. VGFYESDVMGR, DMYSFLEDMGLK, and FQVDNNNR are the unique peptides with the most variation in the presence of trypsin, as seen in **Figure 3.13A**. On the other hand, the unique peptides with the largest variation in the enzyme combination results, as seen in **Figure 3.13B**, are VGFYESDVMGR, DMYSFLEDMGLK, and HNVYINGITYTPVSSTNEK.

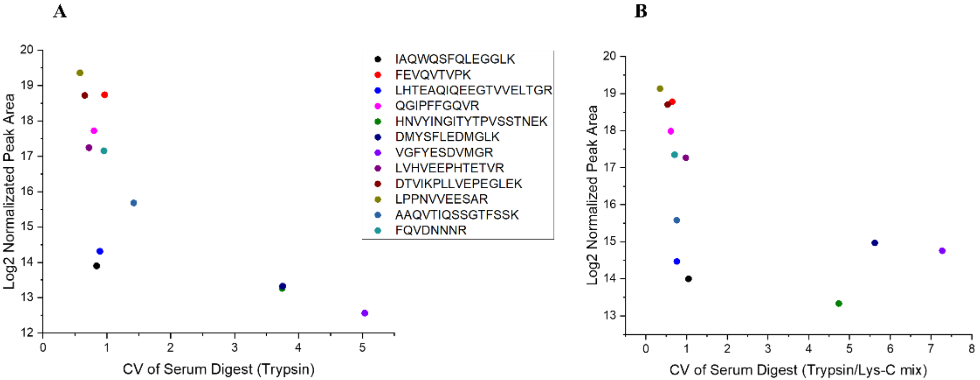

***Figure 3.13.** The Log2 normalized peak areas versus CV values of neat human serum samples **A.** in trypsin **B.** in trypsin/Lys-C mixture*

Figure 3.13A shows that when the log2 normalized peak area of unique peptides decreases, the variation increased, and vice versa. The highly abundant peptides have less variation in this case, whereas the less abundant peptides have more significant variation. However, the enzyme mixture result, as shown in **Figure 3.13B**, does not demonstrate the same behavior observed in **Figure 3.13A**. It was difficult to specify a relationship between the variation and log2 normalized peak area for unique peptides.

3.2.2 Statistical Data Processing

Statistical approaches were used to evaluate the behavior of unique peptides under different conditions. The MS data of the unique peptides acquired were normalized and correlated using different approaches. In addition, distribution and clustering analyses of these unique peptides were performed, as detailed below.

3.2.1 The Peptide Normalization

Several peptide normalization approaches that are frequently used in the literature are evaluated in this section. **Figure 3.14** shows an error bar plot of the twelve unique peptides in six replicates serum digests obtained by using trypsin and trypsin/Lys-C.

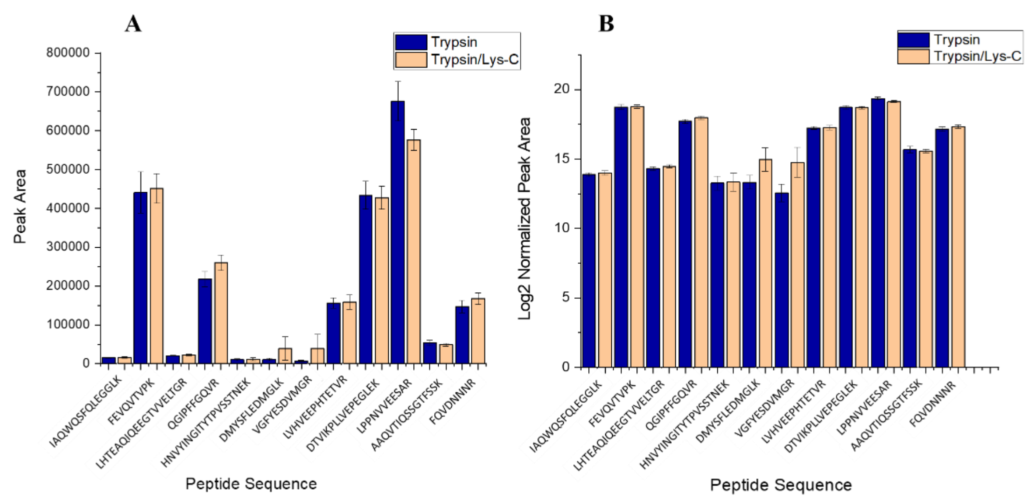

Figure 3.14. *The error column plots of all unique peptides in six serum digest replicates for both enzymes.* ***A.*** *The plot of obtained peak areas of unique peptides for both enzymes.* ***B.*** *The plot of log2 normalized peak areas of unique peptides for both enzymes.*

The peptides obtained from trypsin digestion are represented in the dark blue, while the peptides obtained from trypsin/Lys-C digestion are represented in dark yellow. The error bar plot in **Figure 3.14A** shows the obtained peak areas of each unique peptide for both enzymes, whereas the second plot shown in **Figure 3.14B** illustrates the log2 normalization. By the way, the most frequently used normalization in clinical proteomics is log2 normalization[205].

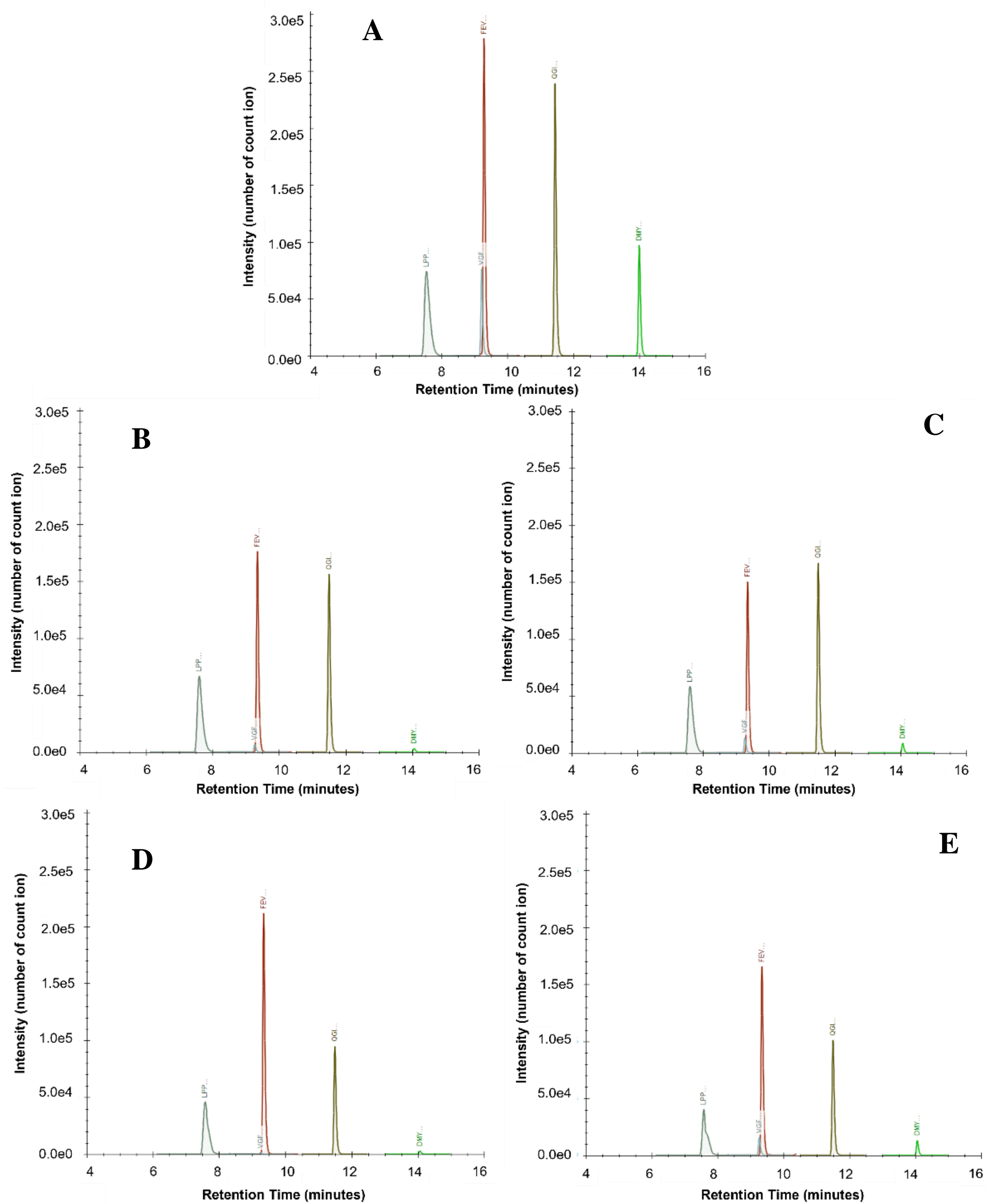

Figure 3.15. *The five selected unique peptides' response* ***A.*** *in equimolar peptide mixture* ***B.*** *in protein standard in the presence of trypsin* ***C.*** *in protein standard in the presence of trypsin/Lys-C mixture* ***D.*** *in human serum in the presence of trypsin* ***E.*** *in human in the presence of trypsin/Lys-C mixture*

The chromatograms of the five unique peptide standards purchased are shown in **Figure 3.15**. The equimolar mixture of five synthetic peptides is given in **Figure 3.15A**. These unique peptides are FEVQVTVPK, QGIPFFGQVR, VGFYESDVMGR, DMYSFLEDMGLK, and LPPNVVEESAR as mentioned earlier in the section (Part 3.1.4). The chromatograms of the five unique peptides in the A2MG protein standard are given in **Figure 3.15B** and **Figure 3.15C** for trypsin and trypsin/Lys-C, respectively. Furthermore, **Figure 3.15D** and **Figure 3.15E** show chromatograms of five unique peptides in neat human serum digested with trypsin and trypsin/Lys-C, respectively.

A response could not be produced for all peptides in the protein standard and serum. If there was no digestion effect on protein standard samples, the response of unique peptides would be the same as the equimolar peptide mixture response. This difference in peak intensities of unique peptides demonstrates that the variation is due to protein digestion rather than ionization. The fact that the response of the peptides alters when the enzyme is changed supports the statement. However, both digestion and the matrix effect affected the response change in serum samples.

Normalization is a technique for analyzing data comparisons more conveniently[170]. It was observed that the selected unique peptides in the literature were randomly selected either from the highly or the scarcely abundant ones.

When assessing digestion reproducibility, the coefficients of variations (CV) of alternative normalizing approaches were also evaluated for both enzyme digestions. Four typical normalization techniques were considered to determine peptide performance, as shown in **Figure 3.16**. Each data point corresponds to a single peptide. The area, log2, and median normalizations each included twelve distinct peptides. Furthermore, the ideal peak area ratio between the analyte and IS peptide is 1:1.

However, the complexities of optimization rise with the number of protein peptides. As a result, the recommended range of ratio between sample peak area and IS peak

area for clinical research is between a 1:10 to 10:1 ratio[45,206,207] in the literature. However, the area internal standard area ratio only had three peptides. These unique peptides are FEVQVTVPK, DTVIKPLLVEPEGLEK, and LPPNVVEESAR, respectively. The reason is that the ratio of these peptides is within the acceptable range in the proteomics field.

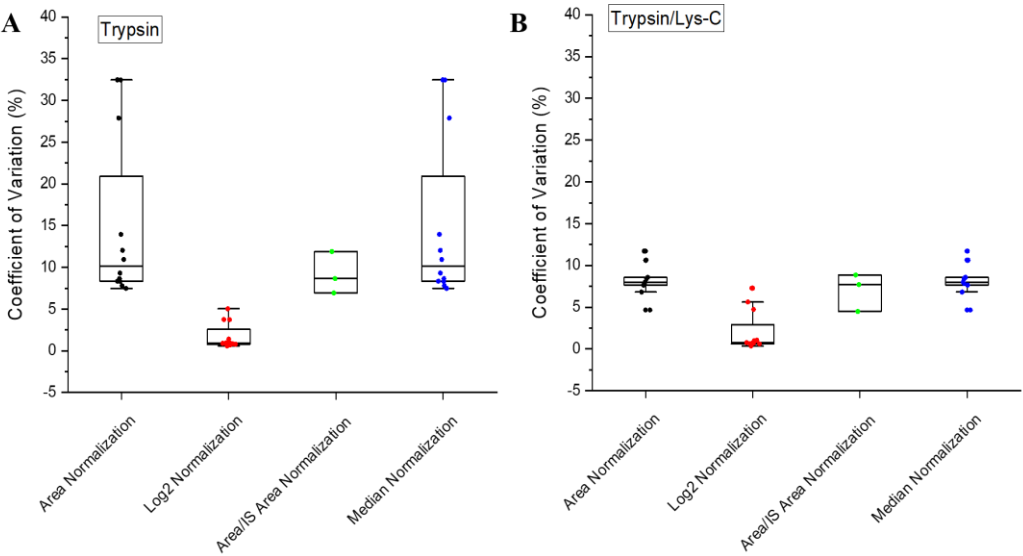

Figure 3.16. *Coefficient of variation (CV) of commonly used normalization approaches in neat human serum for both enzyme digestions* ***A.*** *CV values of different normalization approaches for trypsin digestion* ***B.*** *CV values of different normalization approaches for trypsin/Lys-C mixture digestion*

In the error bar plots, the whisker range was drawn using a coefficient of 1.5 in respect to the outlier. **Figure 3.16A** shows that most of the CV values of the peptides in the trypsin digestion graph are under 10%, and two peptides have more than 20% variance. **Figure 3.16B** shows the results of the trypsin/Lys-C digestion, which show that all the CV values are less than 20% variation. These CV values indicate that each target peptide has less variance.

The coefficient of variation demonstrates analytical variation. Thus, the lower the CV value is considered as better. When the analytical variation is less, the biological variation is observed more clearly. Although some sources suggest that CV values should be less than 10% [61,208,209], the literature indicates that the threshold is 20% [201–203].

3.3 The Effect of Concentration

The concentration effect of A2MG protein standard on unique peptides was visualized using heatmaps with a dendrogram graph. This graph was chosen because it clearly shows the variations in peptide distribution and clusters showing similar behaviors for the peptides. The peptides are highly correlated when the correlation is represented in yellow color, and the opposite correlation is stronger whenever the correlation is represented in red color.

To investigate the impact of concentration, A2MG protein standard samples were compared to one another. For the comparison to be valid, Pearson correlated plots were drawn in the same correlation scales (between -1 to +1), shown in **Figure 3.17**.

When evaluating the concentration effect, A2MG protein samples prepared in eight concentration levels were used. The graphs in **Figure 3.17** are divided into two categories based on the concentration of protein standard samples: high concentration and low concentration. Because in biomarker studies, the protein concentration in the samples may increase or decrease. The samples were investigated at both concentrations so that the effect of concentration could be observed. The low concentration A2MG protein standard samples shown in **Figure 3.17A** and **Figure 3.17B** have a concentration range of 0.0071 – 0.0357 µg/µl, whereas the high concentration protein standards demonstrated in **Figure 3.17C** and **Figure 3.17D** have a concentration range of 0.0536 – 0.1071 µg/µl. When the cluster analysis is evaluated, the dendrograms of the four plots in **Figure 3.17** are divided into two major subgroups. Any association between the location and abundance of

the peptides included in these clusters was not found when the subgroups were examined.

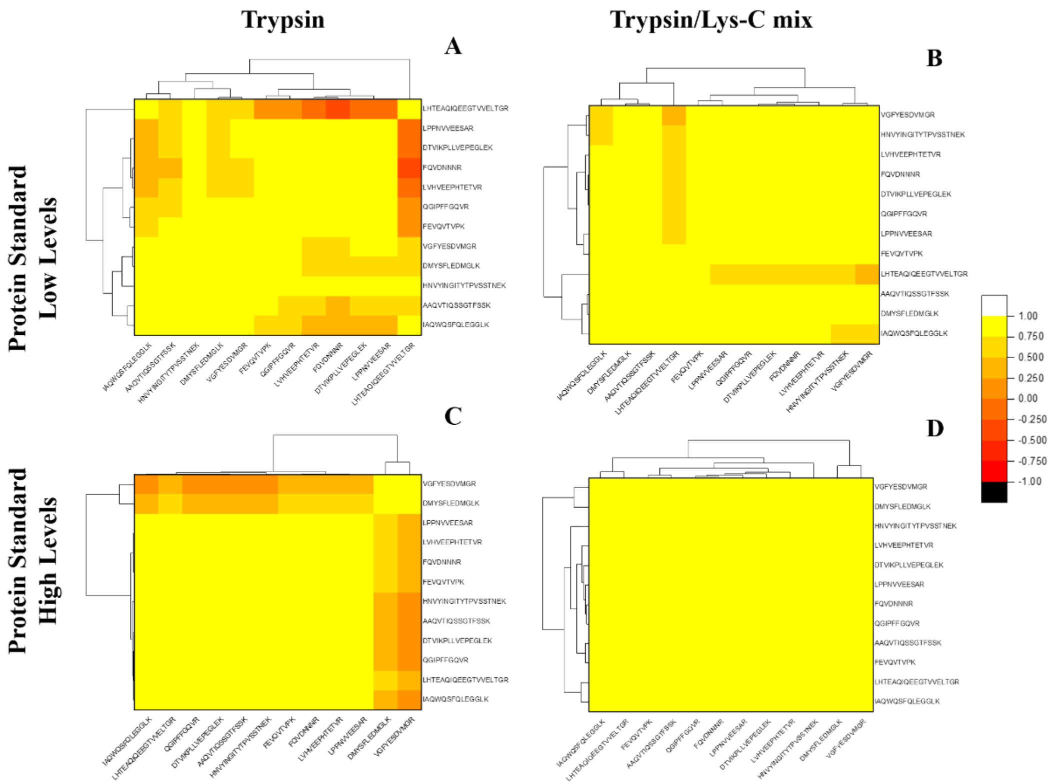

Figure 3.17. *A heat map with a dendrogram plot demonstrating the association of A2MG unique peptides in protein standard samples presents different enzymes* ***A.*** *Correlation in low concentration protein standard samples in the presence of trypsin.* ***B.*** *Correlation in low-concentration protein standard samples with trypsin/Lys-C mixture.* ***C.*** *The correlation in high concentration protein standards presence of trypsin.* ***D.*** *Correlation in high-concentration protein standard samples with trypsin/Lys-C mixture*

When we compare high (**Figure 3.17B**) and low concentration (**Figure 3.17D**) of A2MG protein standard samples digested by trypsin and trypsin/Lys-C mixture, the linear correlation is found to be more significant at high concentration. The opposite linear correlation appears to be stronger in protein standard samples with low concentration (**Figure 3.17A** and **Figure 3.17B**) than in protein standard samples

with high concentration (**Figure 3.17C** and **Figure 3.17D**). This weaker signal may have resulted in a different correlation. Apart from the unique peptide HNVYINGITYTPVSSTNEK, it is observed that eleven unique peptides behave differently in the presence of trypsin at low protein concentrations. Also, the unique peptide LHTEAQIQEEGTVVELTGR in the presence of a trypsin/Lys-C mixture at low protein concentration levels is not correlated with other unique peptides. These differences in correlation may be due to insufficient information gathered due to poor MS signal response of protein standard samples at low concentrations. When the correlation of protein standard samples digested with trypsin is compared in **Figure 3.17A** and **Figure 3.17C**, it is seen that samples with higher concentrations have a higher correlation. However, the unique peptide DMYSFLEDMGLK and VGFYESDVMGR show opposite linear correlations.

Similarly, in **Figure 3.17B** and **Figure 3.17D**, the unique peptides in the presence of trypsin/Lys-C mixture are correlated better in high concentration protein standard samples. Once again, trypsin/Lys-C mixture provides successful cleavage in the inner regions of the protein and better digestion than trypsin. On the other hand, it appears to have a weaker opposite linear correlation with the unique peptides, DMYSFLEDMGLK, AAQVTIQSSGTFSSK, and VGFYESDVMGR. The strong correlation could be attributed to the fact that these peptides are located very close to glycosylation sites in the protein structure. On the other hand, the unique peptides' active amino acid residues such as methionine (M), glutamine (Q), and histidine (H) could account for the weak correlation.

At this time, the unique peptide LHTEAQIQEEGTVVELTGR was to have an opposite correlation with the unique peptide VGFYESDVMGR but a weaker opposite correlation with the unique peptides LPPNVVEESAR, LVHVEEPHTETVR, QGIPFFGQVR, DTVIKPLLVEPEGLEK, HNVYINGITYTPVSSTNEK, and FQVDNNNR in **Figure 3.17B**. This relationship could have occurred between peptides whose digestion was incomplete. Since the peptide LHTEAQIQEEGTVVELTGR is one of the longest peptides, an increment

in peptide length might be a negative factor incomplete digestion. Furthermore, this unique peptide has consecutive amino acid residues such as EK or RK, that might be reducing the complete digestion. Based on these results, it can be deduced that digestion of this unique peptide is incomplete and that it has a relation with the peptides with which it is oppositely correlated through this factor. In addition, the highest correlation was observed in high concentration levels of protein standard samples digested with trypsin/Lys-C mixture (**Figure 3.17D**). While all peptides correlate with each other in **Figure 3.17D**, the unique peptides in high concentration levels of protein standard digested with trypsin shows an opposite correlation in **Figure 3.17C**.

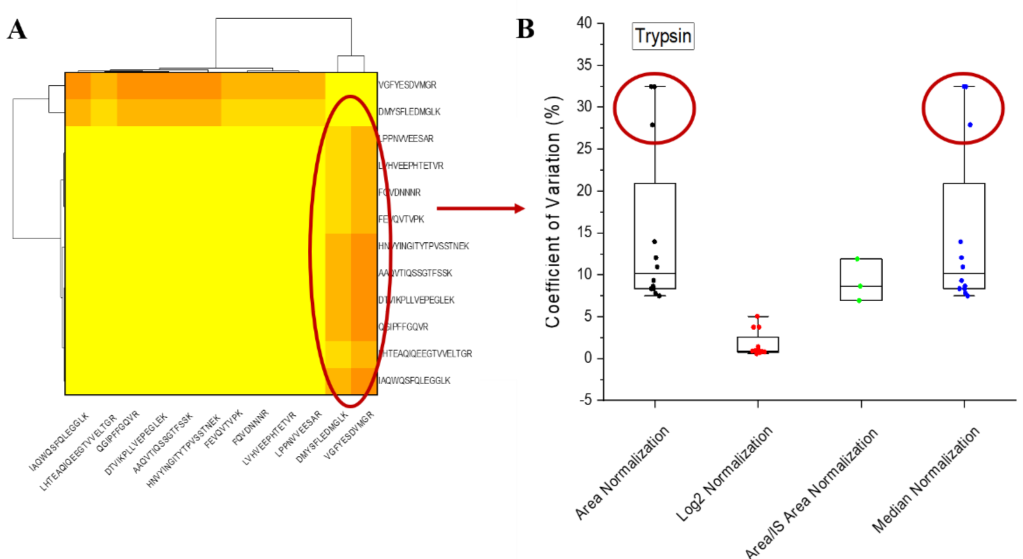

Figure 3.18. *The oppositely correlated peptides in Figure 3.17C with the highlighted CV values* ***A.*** *The peptide correlation due to trypsin digestion in Figure 3.17C. The oppositely correlated peptides are highlighted.* ***B.*** *The CV values of these peptides are highlighted.*

These unique peptides showing opposite correlation are DMYSFLEDMGLK and VGFYESDVMGR. **Figure 3.17C**, and **Figure 3.18A** shows a zoomed-in version of

the heatmap with a dendrogram graph for these peptides. In the coefficient of variation figure from the previous section (Part 3.2.1), these unique peptides likewise had the highest CV values, which are marked in **Figure 3.18B**. These are the peptides with the lowest pI values (3.54 and 3.93, respectively) and are found within the protein core region. Also, pI is the pH at which the net charge in the peptides is zero. Furthermore, they have nearly the same peptide length (12 and 11, respectively). The result leads to the following conclusion: a low pI value may be a factor that negatively affects the functioning of trypsin, resulting in poor digestion in the inner regions of the protein.

3.4 The Effect of Enzymes on Digestion Process

The effect of different enzymes on the digestion process was investigated by comparing digestion of A2MG protein standard and human serum samples by trypsin and trypsin/Lys-C mixture. The Pearson correlation was graded on a scale of -1 to 1 in **Figure 3.19**. High concentration levels of the A2MG protein standard were utilized to create this heatmap graph, as well as three concentration levels generated by spiking A2MG protein into the human serum and neat human serum. The dendrograms reveal two major clusters of twelve unique peptides.

Figure 3.19A and **Figure 3.19B** represent the correlation of peptides in protein standard samples, while **Figure 3.19C** and **Figure 3.19D** represent the correlation in standard serum samples. Thus, the comparison in the pure protein standard and serum sample with complex matrix has been investigated. Trypsin and trypsin/Lys-C mixture has been used to compare digestion efficiency with missed cleavage.

The highest correlation was observed in protein standard samples. The peptides VGFYESDVMGR and DMYSFLEDMGLK have a complete opposite correlation with other peptides as shown in **Figure 3.19A**. Their lengths are 11 and 12, and their pI values are 3.93 and 3.54, respectively. They are located at similar sites in the 3D structure. The peptide VGFYESDVMGR is located in the bait region (core region),

whereas the peptide DMYSFLEDMGLK is located at the inner region of the protein. This result exhibits that these two peptides have very similar properties and are different from other unique peptides.

When **Figure 3.19A** and **Figure 3.19B** are compared, the dissimilarity coefficient decreases for trypsin/Lys-C mixture digestion compared to trypsin digestion. The main reason for this differentiation could be the location of these longer peptides. This behavior is parallel to what is reported in the literature. Trypsin/Lys-C mixture performs a better cleavage because enzymes can reach the inner regions considerably better than only trypsin and prevent incomplete digestion with missed cleavage.

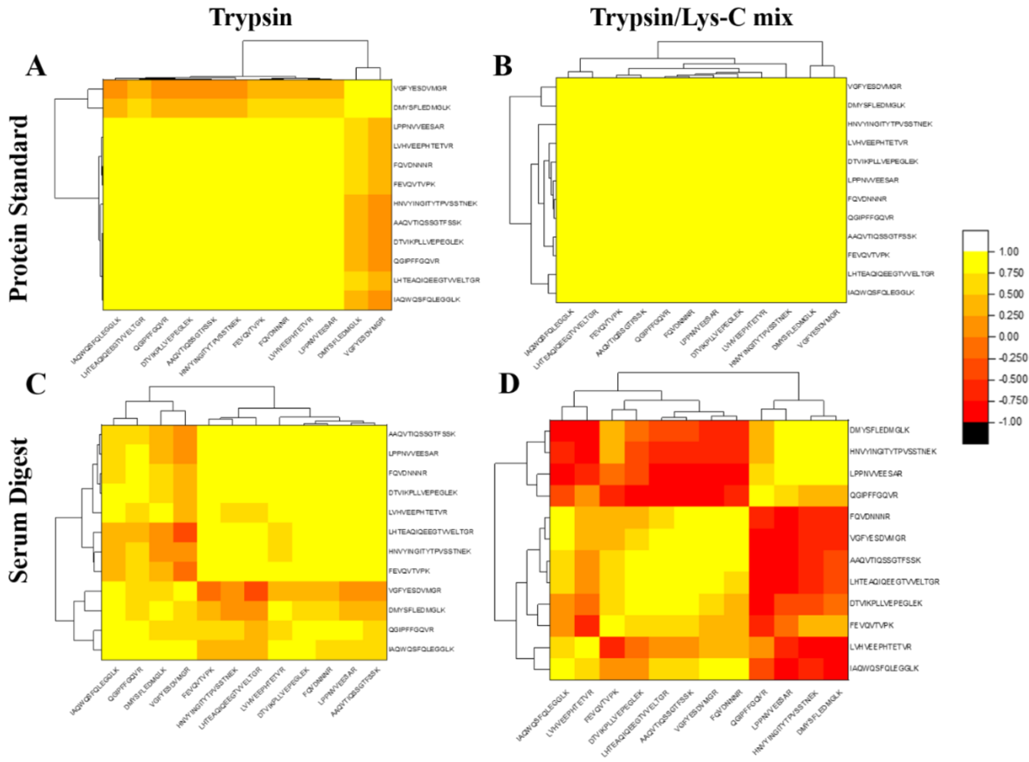

Figure 3.19. *A heatmap with a dendrogram plot depicting the correlation of A2MG unique peptides in protein standard and serum digest samples due to change in enzyme.* ***A.*** *The correlation in protein standard samples upon trypsin digestion.* ***B.*** *The correlation in protein standard samples upon trypsin/Lys-C mixture digestion.* ***C.*** *The correlation in serum samples upon trypsin digestion.* ***D.*** *The correlation in serum samples upon trypsin/Lys-C mixture digestion.*

Besides, when all unique peptides in **Figure 3.19A** and **Figure 3.19B** are compared, the decrease in linear correlation validates our thought. On the other hand, serum digest samples' peptide correlations are not as simple as protein standard samples because of higher complexity, including other proteins and different forms of proteins. When the increase in environment's heterogeneity, VGFYESDVMGR and DMYSFLEDMGLK peptides are still highly opposite correlated as displayed in **Figure 3.19C**, though correlation was not as substantial as the correlation displayed in **Figure 3.19A**.

The unique peptides HNVYINGITYTPVSSTNEK, LHTEAQIQEEGTVVELTGR, and LVHVEEPHTETVR, are correlated with each other as shown in **Figure 3.19C**. The reason for the linear correlation may be presence of histidine (H) residues in these sequences. The chemical modifications can cause a change in peptide forms [111]. A modified form of these peptides may have occurred, which may have also affected the peptides' correlation.

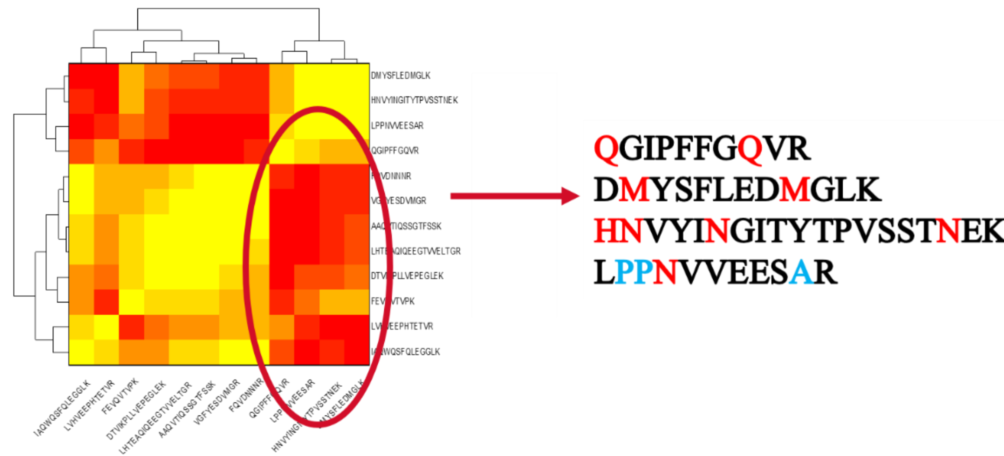

Figure 3.20. *The close inspection of oppositely correlated peptides in Figure 3.19D. The red-colored amino acids represent chemically active amino acids, and the blue-colored amino acids are the ones that affect activity of trypsin.*

The four unique peptides QGIPFFGQVR, LPPNVVEESAR, DMYSFLEDMGLK, and HNVYINGITYTPVSSTNEK, are highly correlated **Figure 3.19D**, and **Figure 3.20** shows a zoomed-in version of this heatmap with a dendrogram graph. They have chemically active residues, and it is possible that they may be modified, as shown in **Figure 3.20**, or that they have been digested improperly. Because the first three unique peptides QGIPFFGQVR, HNVYINGITYTPVSSTNEK, and DMYSFLEDMGLK in **Figure 3.20**, are located in the protein's inner region. Unlike the other peptides, the reason for the peptide LPPNVVEESAR's missed cleavage is that it has two proline (P) amino acid residues near the peptide ends. Trypsin digests results in limited proteolysis[210] as a result of these amino acid residues. In other words, proline (P)[73] amino acids have been shown to have a negative effect on trypsin activity in the literature.

3.5 The Peptide Behavior in Different Matrices

To investigate the behavior of the A2MG protein's unique peptides in various matrices, the bovine serum matrix was used as an alternative matrix to the human serum matrix in this study. The reason is that the bovine serum is easily available, has a complex environment as human serum, and does not contain all of the unique peptides of A2MG protein found in human sera. **Table 3.9** below lists the unique peptides in the human A2MG protein as well as the same peptides in the bovine A2MG protein. In this part of the analysis, bovine serum and human serum were spiked with A2MG protein standard and digested with trypsin enzyme and a trypsin/Lys-C mixture. The serum samples are compared with protein standards.

A2MG concentrations in spiked A2MG protein into human serum range from 0.0455 to 0.1000 µg/µl, in the A2MG protein standard from 0.0536 to 0.1071 µg/µl, and spiked A2MG protein into bovine serum from 0 to 0.0121 µg/µl. In addition, A2MG concentration was considered nonexistent in bovine serum since it does not contain all of the human specific A2MG unique peptides.

Table 3.9. *List of A2MG protein's unique peptides shared to both matrices*

	Peptide Sequence
1	HNVYINGITYTPVSSTNEK
2	DMYSFLEDMGLK**
3	DTVIKPLLVEPEGLEK
4	LPPNVVEESAR**
5	ALLAYAFALAGNQDK*
6	AAQVTIQSSGTFSSK

**The peptide has not been included due to low ionization during MS analysis*

***The synthetic unique peptides were purchased.*

Figure 3.21 depicts all heatmaps with dendrogram using Pearson's correlation on the same scale, which is between -1 to +1. According to the hierarchical cluster analysis, the included protein standard and serum samples were separated into two classes. In bovine serum samples, there is an opposite correlation. Furthermore, the opposite correlation in unique peptides in trypsin digested samples was lower than in the trypsin/Lys-C digested samples (**Figure 3.21C** and **Figure 3.21D**, respectively). Likewise, the opposite correlation was stronger in unique peptides in presence of the trypsin/Lys-C digested samples than in the trypsin digested human serum samples in **Figure 3.21E** and **Figure 3.21F**, respectively.

When all of the plots in **Figure 3.21** are examined, the protein standard plots exhibit the strongest linear peptide correlation compared to the serum plots. This finding is to be expected, considering it is proportionate to the biological environment's complexity. When the graphs are evaluated over the digestion process, it can be seen that the enzyme mixture results are the strongest positive or negative peptide correlation than the trypsin enzyme.

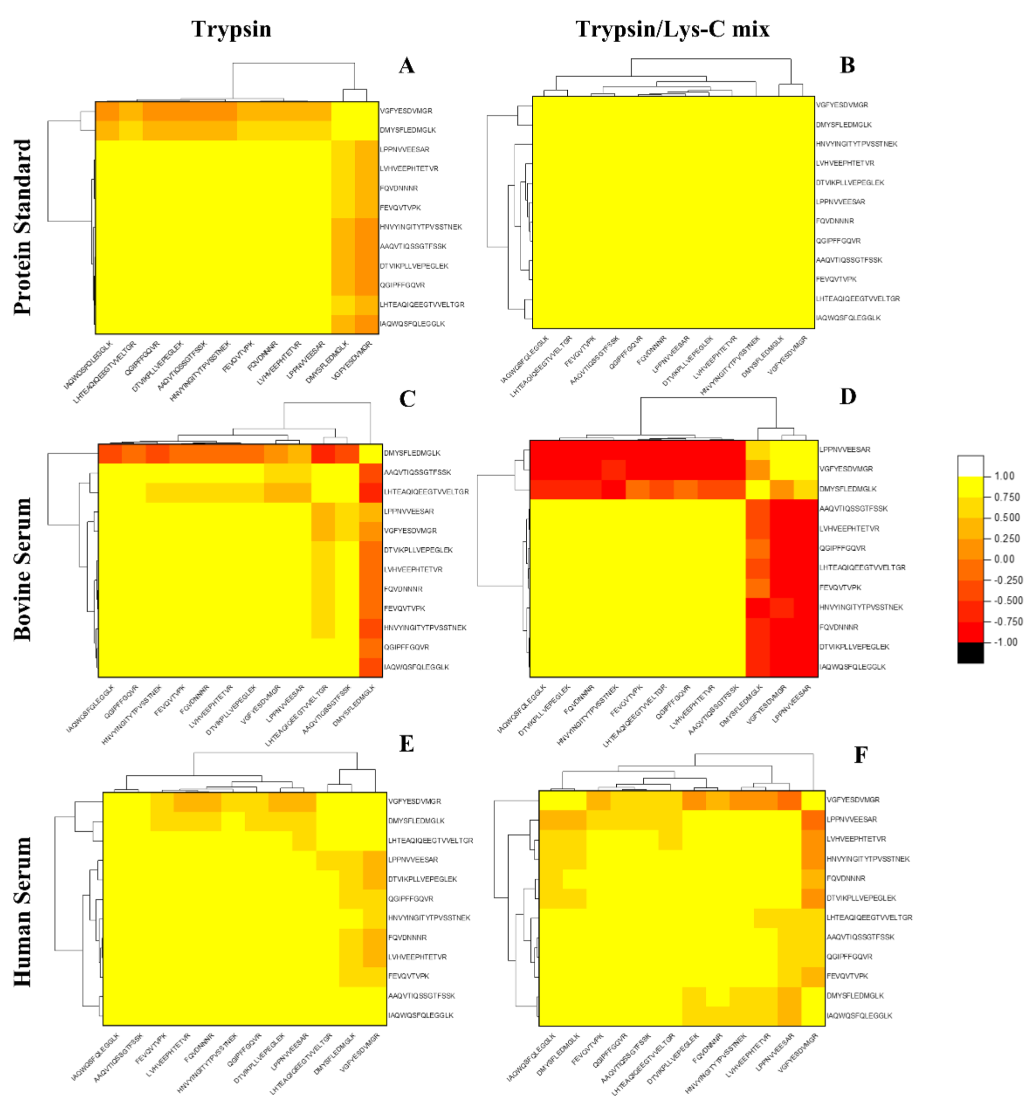

Figure 3.21. *A heatmap with a dendrogram plot illustrating the association of A2MG unique peptides in human and bovine serum-protein spiked samples in the presence of various enzymes.* ***A*** *Peptide correlation in the presence of trypsin in protein standard.* ***B.*** *The correlation in the presence of trypsin/Lys-C mixture in protein standard.* ***C.*** *The correlation in the presence of trypsin in bovine serum* ***D.*** *The correlation in the existence of trypsin/Lys-C mixture in bovine serum* ***E.*** *The correlation in the presence of trypsin in human serum.* ***F.*** *The correlation in the presence of trypsin/Lys-C mixture in human serum*

When human and bovine serum samples are compared, human serum samples are correlated better than bovine serum samples because they already contain these twelve unique peptides of the A2MG protein. Thus, the peptides' concentration in human serum is higher than in bovine serum, revealing a better linear correlation. **Figure 3.21C** shows that the unique peptide DMYSFLEDMGLK has a high opposite correlation with other A2MG unique peptides. The unique peptide DMYSFLEDMGLK with one of the lowest pI values is found inside the A2MG protein. Although the linear correlation in human serum samples is higher than in bovine serum samples in the presence of trypsin enzyme, the unique peptides DMYSFLEDMGLK and VGFYESDVMGR differ from other unique peptides found in trypsin digested human serum samples, as shown in **Figure 3.21E**. However, trypsin digested human serum samples have the highest linear correlation of these four serum plots in **Figure 3.21**.

The highly opposite correlations of three unique peptides, DMYSFLEDMGLK, LPPNVVEESAR, and VGFYESDVMGR, are demonstrated in **Figure 3.21D**. These three unique peptides' lengths are 12, 11, and 11, respectively. The active methionine (M) amino acid residue is found in the peptides DMYSFLEDMGLK and VGFYESDVMGR. This residue may be a distinguishing factor since only these two unique peptides have this active amino acid residue among the twelve unique peptides. In addition, these three unique peptides have extremely low pI values among the twelve unique peptides of A2MG, which are 3.54, 3.93, and 4.15, correspondingly. As a result, it may be one of the major factors driving this correlation. Furthermore, since the unique peptides DMYSFLEDMGLK and VGFYESDVMGR are found in the inner region of the protein, even the unique peptide VGFYESDVMGR is located in the bait region. Thus, incomplete digestion may be one of the factors affecting the correlation.

In addition to those, as an alternative matrix, ERM-approved human serum samples were compared to standard human serum samples. The A2MG protein concentration in an ERM-certified human serum is theoretically calculated as 1.43 µg/µl (with

±0.06 uncertainty). Graphics were analyzed using the same correlation scale (between -1 to +1). A heatmap with a dendrogram graph was applied to compare human serum and ERM-certified serum, as shown in **Figure 3.22**. When comparing all graphs in **Figure 3.22**, the peptide correlation in the trypsin digested samples is more oppositely correlated than in the enzyme mixture digested samples in both human sera. In all heatmap plots, two primary clusters were founded for twelve unique peptides.

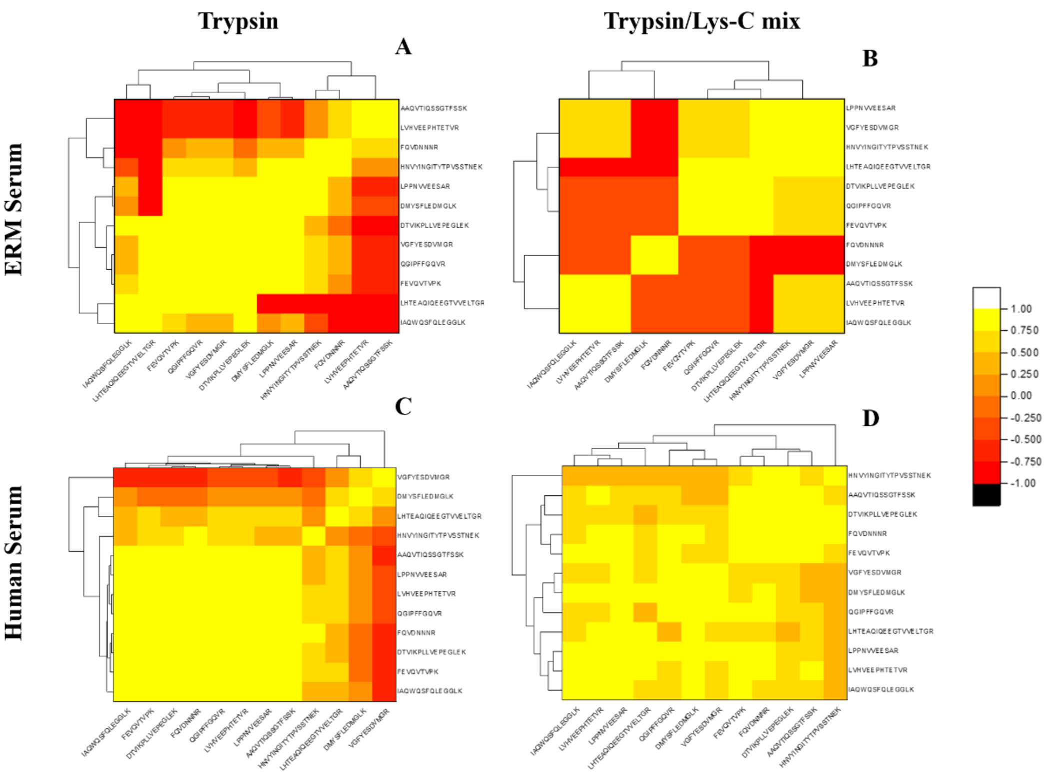

Figure 3.22. *A heatmap with a dendrogram plot highlighting the association of A2MG unique peptides in standard human serum and ERM certificated human serum samples with the participation of different enzymes* ***A.*** *The correlation of trypsin digested ERM serum.* ***B.*** *The correlation of trypsin/Lys-C mixture digested ERM serum samples.* ***C.*** *The correlation of trypsin digested standard human serum.* ***D.*** *The correlation of trypsin/Lys-C mixture digested human serum samples.*

Figure 3.22 shows unique peptides in ERM-certified serum that have a high opposite correlation for both enzymes. The four unique peptides with opposite correlations are FQVDNNNR, LVHVEEPHTETVR, AAQVTIQSSGTFSSK, and HNVYINGITYTPVSSTNEK which are listed in **Figure 3.22A**. The lengths of the peptides are 8, 13, 15, and 19, respectively. This grouping is a little odd because the longest and shortest peptides are clustered together. As a result, peptide length may not be the factor causing the variation.

When considering their positions in the protein structure, the peptide LVHVEEPHTETVR is found in the bait regions, while the peptides FQVDNNNR and AAQVTIQSSGTFSSK are located next to each other on the protein's surface. The peptide HNVYINGITYTPVSTNEK, on the other hand, is found in the protein's inner region. Another factor could be that they have chemically active amino acids. There are two histidine (H) amino acids in the unique peptide LVHVEEPHTETVR. The FQVDNNNR peptide has one glutamine (Q) and three asparagines (N), whereas the AAQVTIQSSGTFSSK peptide has two glutamines (Q). Finally, there is one histidine (H) and three asparagines (N) in the peptide HNVYINGITYTPVSTNEK.

The following unique peptides in **Figure 3.22B** with the opposite correlation are DMYSFLEDMGLK, FQVDNNNR, LVHVEEPHTETVR, AAQVTIQSSGTFSSK, and IAQWQSFQLEGGLK. These peptides have lengths of 12, 8, 13, 15, and 14, respectively. Investigations based on the presence of chemically active amino acid residues revealed that the unique peptide DMYSFLEDMGLK has two methionine (M), whereas unique peptide FQVDNNNR has one glutamine (Q) and three asparagines (N). The unique peptide IAQWQSFQLEGGLK, on the other hand, has three glutamine (Q) and one tryptophan (W) residue. Two histidines (H) are found in the unique peptide LVHVEEPHTETVR, whereas two cysteines (C) are present in the unique peptide AAQVTIQSSGTFSSK. When the locations of these five unique peptides in the protein structure were examined, it was discovered that three of them, FQVDNNNR, AAQVTIQSSGTFSSK, and IAQWQSFQLEGGLK, were found next to each other, while the unique peptide LVHVEEPHTETVR was found in the

bait region, and the unique peptide DMYSFLEDMGLK was found in the inner region. Furthermore, when the pI values of the five peptides were compared, it was observed that they differed significantly. These five unique peptides, IAQWQSFQLEGGLK, LVHVEEPHTETVR, AAQVTIQSSGTFSSK, DMYSFLEDMGLK, and FQVDNNNR, have pI values of 6.68, 5.19, 10.19, 6.4, and 3.54, respectively. As a result, it can be concluded that the pI value is not the most critical element determining the linear correlation.

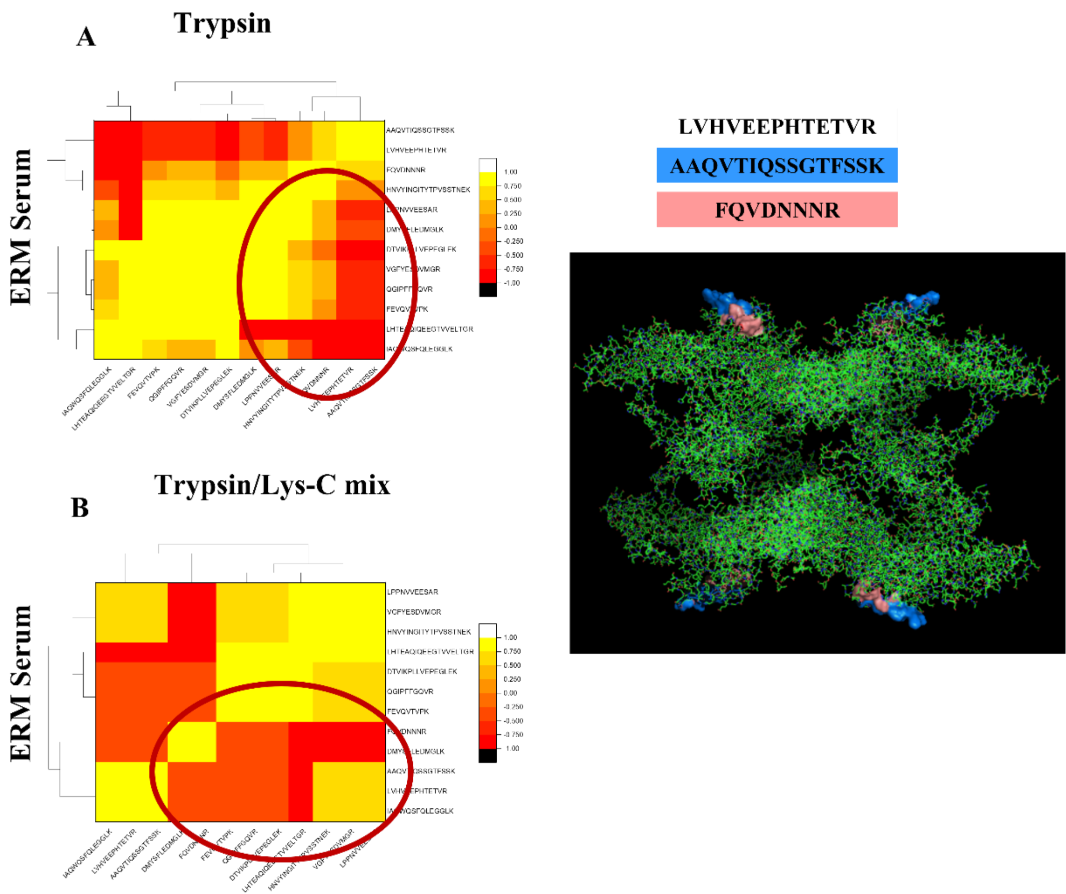

Figure 3.23. *The shared unique peptides with the strongest opposite correlation in both enzyme digests were marked. The A2MG protein structure's structure is shown on the left panel.*

The three peptides FQVDNNNR, LVHVEEPHTETVR, and AAQVTIQSSGTFSSK behave differently in both enzyme digests, as shown above in **Figure 3.23**, resulting in a more significant opposite correlation. As mentioned earlier, these three unique peptides have different lengths, chemically active amino acid residues, and pI values. Two unique peptides are positioned next to each other, and one of them is located in the bait region. However, because both peptides are present on the protein's surface, as illustrated in **Figure 3.23**, it is difficult to determine why they exhibit the same behavior based on their locations in the protein structure. The unique peptide's exact position in the bait region is uncertain, but it is assumed to be in the inner region. Unfortunately, it has not been determined why these three unique peptides show the same behavior. However, based on the information, the following conclusion may be drawn: the factors we assessed are not the primary cause for the grouping.

Comparing **Figure 3.22C** and **Figure 3.22D** revealed a decrease in the opposite peptide correlation between trypsin digest and trypsin/Lys-C mixture digest human serum. The distribution of peptides is highly correlated in trypsin/Lys-C digest. The oppositely correlated peptides in **Figure 3.22C** are DMYSFLEDMGLK, HNVYINGITYTPVSTNEK, LHTEAQIQEEGTVVELTGR, and VGFYESDVMGR. These unique peptides have lengths of 12, 11, 19, and 19, respectively. Two glutamines (Q) and one histidine (H) are the chemically active residues in LHTEAQIQEEGTVVELTGR.

Furthermore, the HNVYINGITYTPVSTNEK peptide has one histidine (H) and three asparagines (N). The unique peptide DMYSFLEDMGLK has two methionines (M), while the unique peptide VGFYESDVMGR only has one methionine (M). The VGFYESDVMGR peptide is found to be in the bait region when the positions of these unique peptides in the 3D protein structure are investigated. As previously stated in **Figure 3.18**, the unique peptides DMYSFLEDMGLK and VGFYESDVMGR have a coefficient of variations higher than the others. The three unique peptides DMYSFLEDMGLK, LHTEAQIQEEGTVVELTGR, and HNVYINGITYTPVSTNEK, are all found in the protein's inner regions. Because

trypsin activity reduces in the inner locations of the protein, this might be taken to mean that peptides low ability to be digested (**Figure 3.22C**). The fact that these unique peptides exhibit a weaker linear peptide correlation in the trypsin/Lys-C digests (**Figure 3.22D**), supports the accuracy of this explanation.

CHAPTER 4

CONCLUSION

In this study, protein-peptide relationship was investigated under conventional proteolytic digestion conditions in order to assess the clinical utility of peptide centric biomarker studies. To this end, A2MG selected as a reference protein and the MRM method was developed to monitor twelve A2MG unique peptides. The replicate QC injections were performed to assure reproducibility of the measurements. Proteolytic digestion reproducibility was tested through parallel sample processing.

We investigated the quantitative protein-peptide relationship. Thus, the enzyme on the digestion process, concentration, and the matrix effects were investigated to find factors affecting the peptide behavior. A2MG protein standard and protein spiked human serum and bovine serum samples at varying concentration levels were digested using two common proteases (trypsin and trypsin/Lys-C mixture).

The relative abundances of unique peptides were monitored and compared with corresponding protein concentration levels. The highest protein-peptide correlation was observed in high concentration levels of the protein standard for both enzymes. The correlation of eleven A2MG unique peptides were variable at low concentration levels of the protein standard. Apart from the unique peptide HNVYINGITYTPVSSTNEK, eleven A2MG unique peptides have been discovered to behave differently in the presence of trypsin at low protein concentrations. In the presence of a trypsin/Lys-C mixture at low protein concentration levels, the unique peptide LHTEAQIQEEGTVVELTGR is not correlated with any other unique peptides.

For the enzyme effect on protein digestion, the A2MG protein standard and human serum were compared. Enzyme mixture (trypsin/Lys-C) has better digestion efficiency at all different protein concentrations and protein spiked into biological matrix than trypsin. The protein standard had a higher peptide correlation in presence of trypsin/Lys-C mixture and all A2MG unique peptides were correlated linearly. In A2MG spiked into serum samples, the peptide correlation was not linear since serum has complex protein/peptide environment, the results may be affected by matrix-effect.

In addition, the matrix-effect was investigated using A2MG protein standard, human serum, bovine serum, and ERM-certified serum. It was observed that the peptide correlation was weaker and varied in serum samples compared to the protein standard.

In this study, we used a systematic approach to show dynamic protein-peptide correlation. Results suggests that the change in A2MG protein concentration at protein level is not reflected in peptide level. The variation in peptide level might be affected by various parameters such as the presence of chemically active amino acids in the sequence, the ragged peptide ends that impact activity of trypsin, peptide length, pI values, the locations in the protein structure and potential modifications on peptide sequence. The location of the peptides in the protein structure is the main factor which affects the linear peptide correlation since peptides located inner regions of the structure did not show linear correlation with other target peptides. Also, the peptides with lowest pI values show opposite correlation among all twelve A2MG unique peptides.

As a first step, linearity is shown in the calibration curves in different biological conditions. However, figure of merits such as limit of detection (LOD), limit of quantification (LOQ), dynamic range and recovery should be performed for A2MG absolute protein quantification.

Even though parallel measurements of individual samples are reproducible, all unique peptides in the protein behave differently in serum regardless of the enzyme used. This is clearly showing that selection of a unique peptide for any clinical application is critical as one could get different results depending on the choice.

The common practice in proteomics-based clinical studies is to report the biomarkers in protein level even though the measurements are performed in peptide level. This may mislead the researchers in the field since their behavior changes from one peptide to another.

This is first study showing a comprehensive quantitative protein peptide relationship. Future work needs to focus on the root cause of the variation observed in peptide level to facilitate more precise and sensitive biomarker studies in clinical proteomics.

REFERENCES

(1) Wasinger, V. C., Cordwell, S. J., Poljak, A., Yan, J. X., Gooley, A. A., Wilkins, M. R., Duncan, M. W., Harris, R., Williams, K. L., and Humphery-Smith, I. (1995) Progress with Gene-Product Mapping of The Mollicutes: Mycoplasma Genitalium. *Electrophoresis 16(7)*, 1090–1094.

(2) Parker, C. E., Warren, M. R., and Mocanu, V. (2010) Mass Spectrometry for Proteomics, in *Neuroproteomics* (Alzate O, Ed.), 71–91. CRC Press/Taylor & Francis, Boca Raton (FL).

(3) Molnar, C., and Gair, J. (2015) Introduction to Biotechnology, in *Concepts of Biology* 1st Canadian edition. BCcampus.

(4) Hasin, Y., Seldin, M., and Lusis, A. (2017) Multi-Omics Approaches to Disease. *Genome Biology 18(1)*, 83–98.

(5) Abhilash, M. (2008) Applications of Proteomics. *The Internet Journal of Genomics and Proteomics 4(1)*, 1–7.

(6) Choudhary, C., and Mann, M. (2010) Decoding Signalling Networks by Mass Spectrometry-Based Proteomics. *Nature Reviews Molecular Cell Biology 11(6)*, 427–439.

(7) Bensimon, A., Heck, A. J. R., and Aebersold, R. (2012) Mass Spectrometry-Based Proteomics and Network Biology. *Annual Review of Biochemistry 81(1)*, 379–405.

(8) Dau, T., Bartolomucci, G., and Rappsilber, J. (2020) Proteomics Using Protease Alternatives to Trypsin Benefits from Sequential Digestion with Trypsin. *Analytical Chemistry 92(14)*, 9523–9527.

(9) Stollar, E. J., and Smith, D. P. (2020) Uncovering Protein Structure. *Essays in Biochemistry 64(4)*, 649–680.

(10) Yoithapprabhunath, T. R., Nirmal, R. M., Santhadevy, A., Anusushanth, A., Charanya, D., Rojiluke, Sri Chinthu, K. K., and Yamunadevi, A. (2015) Role of Proteomics in Physiologic and Pathologic Conditions of Dentistry: Overview. *The Journal of Pharmacy and Bioallied Sciences 7(6)*, 344–349.

(11) Anson, M. L., and Mirsky, A. E. (1925) On Some General Properties of Proteins. *The Journal of General Physiology 9(2)*, 169–179.

(12) Cooper, G. M. (2000) The Central Role of Enzymes as Biological Catalysts, in *The Cell: A Molecular Approach* 2nd edition. Sinauer Associates, Sunderland (MA).

(13) Quick, M., and Javitch, J. A. (2007) Monitoring The Function of Membrane Transport Proteins in Detergent-Solubilized Form. *Proceedings of the National Academy of Sciences of the United States of America 104(9)*, 3603–3608.

(14) Coates, C. J., and Decker, H. (2017) Immunological Properties of Oxygen-Transport Proteins: Hemoglobin, Hemocyanin and Hemerythrin. *Cellular and Molecular Life Sciences 74(2)*, 293–317.

(15) Cooper, G. M. (2000) Regulation of Transcription in Eukaryotes, in *The Cell: A Molecular Approach* 2nd edition. Sinauer Associates, Sunderland (MA).

(16) Ali, M. G., Zhang, Z., Gao, Q., Pan, M., Rowan, E. G., and Zhang, J. (2020) Recent Advances in Therapeutic Applications of Neutralizing Antibodies for Virus Infections: An Overview. *Immunologic Research 68(6)*, 325–339.

(17) Forthal, D. N. (2014) Functions of Antibodies. *Microbiology Spectrum 2(4)*, 1–17.

(18) Berg, J. M., Tymoczko, J. L., and Stryer, L. (2002) Protein Structure and Function, in *Biochemistry* 5th edition. W H Freeman, New York.

(19) Rehman, I., Farooq, M., and Botelho, S. (2018) Biochemistry, Secondary Protein Structure. *StatPearls*. StatPearls Publishing, Treasure Island (FL).

(20) Berg, J. M., Tymoczko, J. L., and Stryer, L. (2002) Quaternary Structure: Polypeptide Chains Can Assemble into Multisubunit Structures, in *Biochemistry* 5th edition. W H Freeman, New York.

(21) Smith, L. M., Kelleher, N. L., and The Consortium for Top Down Proteomics. (2013) Proteoform: A Single Term Describing Protein Complexity. *Nature Methods 10(3)*, 186–187.

(22) Wang, Y.-C., Peterson, S. E., and Loring, J. F. (2013) Protein Post-Translational Modifications and Regulation of Pluripotency in Human Stem Cells. *Cell Research 24(2)*, 143–160.

(23) Aebersold, R., Agar, J. N., Amster, I. J., Baker, M. S., Bertozzi, C. R., Boja, E. S., Costello, C. E., Cravatt, B. F., Fenselau, C., Garcia, B. A., Ge, Y., Gunawardena, J., Hendrickson, R. C., Hergenrother, P. J., Huber, C. G., Ivanov, A. R., Jensen, O. N., Jewett, M. C., Kelleher, N. L., Kiessling, L. L., Krogan, N. J., Larsen, M. R., Loo, J. A., Loo, R. R. O., Lundberg, E., MacCoss, M. J., Mallick, P., Mootha, V. K., Mrksich, M., Muir, T. W., Patrie, S. M., Pesavento, J. J., Pitteri, S. J., Rodriguez, H., Saghatelian, A., Sandoval, W., Schlüter, H., Sechi, S., Slavoff, S. A., Smith, L. M., Snyder, M. P., Thomas, P. M., Uhlén, M., Eyk, J. E. van, Vidal, M., Walt, D. R., White, F. M., Williams, E. R., Wohlschlager, T., Wysocki, V. H., Yates, N. A., Young, N. L., and Zhang, B. (2018) How Many Human Proteoforms Are There? *Nature Chemical Biology 14(3)*, 206–214.

(24) Khoury, G. A., Baliban, R. C., and Floudas, C. A. (2011) Proteome-Wide Post-Translational Modification Statistics: Frequency Analysis and Curation of The Swiss-Prot Database. *Scientific Reports 1(1)*, 1–5.

(25) Apweiler R, Hermjakob H, and Sharon N. (1999) On The Frequency of Protein Glycosylation, As Deduced from Analysis of the Swiss-Prot Database. *Biochimica et Biophysica Acta 1473(1)*, 4–8.

(26) Tsai, P.-L., and Chen, S.-F. (2017) A Brief Review of Bioinformatics Tools for Glycosylation Analysis by Mass Spectrometry. *Mass Spectrometry 6(2017)*, S0064–S0069.

(27) Ubersax, J. A., and Ferrell, J. E. J. (2007) Mechanisms of Specificity in Protein Phosphorylation. *Nature Reviews Molecular Cell Biology 8(7)*, 530–541.

(28) Menzies, K. J., Zhang, H., Katsyuba, E., and Auwerx, J. (2015) Protein Acetylation in Metabolism — Metabolites and Cofactors. *Nature Reviews Endocrinology 12(1)*, 43–60.

(29) Borràs, E., and Sabidó, E. (2017) What Is Targeted Proteomics? A Concise Revision of Targeted Acquisition and Targeted Data Analysis in Mass Spectrometry. *Proteomics 17(17–18)* , 1700180–1700188.

(30) Sobsey CA, Ibrahim S, Richard VR, Gaspar V, Mitsa G, Lacasse V, Zahedi RP, Batist G, and Borchers CH. (2020) Targeted and Untargeted Proteomics Approaches in Biomarker Development. *Proteomics 20(9)*, e1900029–e1900045.

(31) Schubert, O. T., Röst, H. L., Collins, B. C., Rosenberger, G., and Aebersold, R. (2017) Quantitative Proteomics: Challenges and Opportunities in Basic and Applied Research. *Nature Protocols 12(7)*, 1289–1294.

(32) Feng, F., Thompson, M. P., Thomas, B. E., Duffy, E. R., Kim, J., Kurosawa, S., Tashjian, J. Y., Wei, Y., Andry, C., and Stearns-Kurosawa, D. J. (2019) A Computational Solution to Improve Biomarker Reproducibility During Long-Term Projects. *PLoS One 14(4)*, e0209060.

(33) Duncan, M. W., Aebersold, R., and Caprioli, R. M. (2010) The Pros and Cons of Peptide-Centric Proteomics. *Nature Biotechnology 28(7)*, 659–664.

(34) Catherman, A. D., Skinner, O. S., and Kelleher, N. L. (2014) Top Down Proteomics: Facts and Perspectives. *Biochemical and Biophysical Research Communications 445(4)*, 683–693.

(35) Aslam, B., Basit, M., Nisar, M. A., Khurshid, M., and Rasool, M. H. (2017) Proteomics: Technologies and Their Applications. *Journal of Chromatographic Science 55(2)*, 182–196.

(36) Kuhn, E., Wu, J., Karl, J., Liao, H., Zolg, W., and Guild, B. (2004) Quantification of C-Reactive Protein in The Serum of Patients with Rheumatoid Arthritis Using Multiple Reaction Monitoring Mass Spectrometry and 13C-Labeled Peptide Standards. *Proteomics 4(4)*, 1175–1186.

(37) Plubell, D. L., Käll, L., Webb-Robertson, B.-J., Bramer, L., Ives, A., Kelleher, N. L., Smith, L. M., Montine, T. J., Wu, C. C., and Maccoss, M. J. (2021) Can We Put Humpty Dumpty Back Together Again? What Does Protein Quantification Mean in Bottom-Up Proteomics? *bioRxiv preprint*.

(38) Norrgran, J., Williams, T. L., Woolfitt, A. R., Solano, M. I., Pirkle, J. L., and Barr, J. R. (2009) Optimization of Digestion Parameters for Protein Quantification. *Analytical Biochemistry 393(1)*, 48–55.

(39) Switzar, L., Giera, M., Lingeman, H., Irth, H., and Niessen, W. M. A. (2011) Protein Digestion Optimization for Characterization of Drug-Protein Adducts Using Response Surface Modeling. *The Journal of Chromatography A 1218(13)*, 1715–1723.

(40) Thiede, B., Höhenwarter, W., Krah, A., Mattow, J., Schmid, M., Schmidt, F., and Jungblut, P. R. (2005) Peptide Mass Fingerprinting. *Methods 35(3)*, 237–247.

(41) Tsiatsiani, L., and Heck, A. J. R. (2015) Proteomics Beyond Trypsin. *The FEBS Journal 282(14)*, 2612–2626.

(42) Wu, Z., Huang, J., Huang, J., Li, Q., and Zhang, X. (2018) Lys-C/Arg-C, a More Specific and Efficient Digestion Approach for Proteomics Studies. *Analytical Chemistry 90(16)*, 9700–9707.

(43) Chambers, A. G., Percy, A. J., Yang, J., Camenzind, A. G., and Borchers, C. H. (2013) Multiplexed Quantitation of Endogenous Proteins in Dried Blood Spots by

Multiple Reaction Monitoring - Mass Spectrometry. *Molecular and Cellular Proteomics 12(3)*, 781–791.

(44) Glatter, T., Ludwig, C., Ahrné, E., Aebersold, R., Heck, A. J. R., and Schmidt, A. (2012) Large-Scale Quantitative Assessment of Different In-Solution Protein Digestion Protocols Reveals Superior Cleavage Efficiency of Tandem Lys-C/trypsin Proteolysis Over Trypsin Digestion. *Journal of Proteome Research 11(11)*, 5145–5156.

(45) Ozcan, S., Cooper, J. D., Lago, S. G., Kenny, D., Rustogi, N., Stocki, P., and Bahn, S. (2017) Towards Reproducible MRM Based Biomarker Discovery Using Dried Blood Spots. *Scientific Reports 7(1)*, 1–10.

(46) Chantaraamporn, J., Champattanachai, V., Khongmanee, A., Verathamjamras, C., Prasongsook, N., Mingkwan, K., Luevisadpibul, V., Chutipongtanate, S., and Svasti, J. (2020) Glycoproteomic Analysis Reveals Aberrant Expression of Complement C9 and Fibronectin in The Plasma of Patients with Colorectal Cancer. *Proteomes 8(3)*, 26–45.

(47) Gaither, C., Popp, R., Mohammed, Y., and Borchers, C. H. (2020) Determination of The Concentration Range for 267 Proteins from 21 Lots of Commercial Human Plasma Using Highly Multiplexed Multiple Reaction Monitoring Mass Spectrometry. *Analyst 145(10)*, 3634–3644.

(48) Yang, Y., Liu, X., Shen, C., Lin, Y., Yang, P., and Qiao, L. (2020) In Silico Spectral Libraries by Deep Learning Facilitate Data-Independent Acquisition Proteomics. *Nature Communications 11(1)*, 1–11.

(49) Hu, A., Wang, Z., Xie, J., Lin, L., Yang, P., Zhang, L., Shen, H., and Cao, Y. (2020) Fast and Precise Quantification of Serum Biomarkers and Simultaneous Recognition of Multiple Diseases Enabled by a Stable Isotope-Labelled Peptides Assisted High-Throughput MRM Strategy. *Analyst 145(15)*, 5299–5306.

(50) Sepiashvili, L., Waggott, D., Hui, A., Shi, W., Su, S., Ignatchenko, A., Ignatchenko, V., Laureano, M., Huang, S. H., Xu, W., Weinreb, I., Waldron, J., O'Sullivan, B., Irish, J. C., Boutros, P. C., Liu, F. F., and Kislinger, T. (2014) Integrated Omic Analysis of Oropharyngeal Carcinomas Reveals Human Papillomavirus (HPV) - Dependent Regulation of The Activator Protein 1 (AP-1) Pathway. *Molecular and Cellular Proteomics 13(12)*, 3572–3584.

(51) Kim, H., Park, J., Kim, Y., Sohn, A., Yeo, I., Jong Yu, S., Yoon, J. H., Park, T., and Kim, Y. (2017) Serum Fibronectin Distinguishes The Early Stages of Hepatocellular Carcinoma. *Scientific Reports 7(1)*, 1–9.

(52) Khoonsari, E., Häggmark, A., Lonnberg, M., Mikus, M., Kilander, L., Lannfelt, L., Bergquist, J., Ingelsson, M., Nilsson, P., Kultima, K., and Shevchenko, G. (2016) Analysis of The Cerebrospinal Fluid Proteome in Alzheimer's Disease Payam. *PLoS One 11(3)*, e0150672.

(53) Breen, M. S., Uhlmann, A., Ozcan, S., Chan, M., Pinto, D., Bahn, S., and Stein, D. J. (2017) Parallel Changes in Serum Proteins and Diffusion Tensor Imaging in Methamphetamine-Associated Psychosis. *Scientific Reports 7(1)*, 1–10.

(54) van Rees, G. F., Lago, S. G., Cox, D. A., Tomasik, J., Rustogi, N., Weigelt, K., Ozcan, S., Cooper, J., Drexhage, H., Leweke, F. M., and Bahn, S. (2018) Evidence of Microglial Activation Following Exposure to Serum from First-Onset Drug-Naive Schizophrenia Patients. *Brain, Behavior, and Immunity 67(1)*, 364–373.

(55) Han, S. Y. S., Cooper, J. D., Ozcan, S., Rustogi, N., Penninx, B. W. J. H., and Bahn, S. (2019) Integrating Proteomic, Sociodemographic and Clinical Data to Predict Future Depression Diagnosis in Subthreshold Symptomatic Individuals. *Translational Psychiatry 9(1)*, 1–11.

(56) Cooper, J. D., Han, S. Y. S., Tomasik, J., Ozcan, S., Rustogi, N., van Beveren, N. J. M., Leweke, F. M., and Bahn, S. (2019) Multimodel Inference for Biomarker

Development: An Application to Schizophrenia. *Translational Psychiatry 9(1)*, 1–10.

(57) Bislev, S. L., Kusebauch, U., Codrea, M. C., Beynon, R. J., Harman, V. M., Røntved, C. M., Aebersold, R., Moritz, R. L., and Bendixen, E. (2012) Quantotypic Properties of QconCAT Peptides Targeting Bovine Host Response to Streptococcus Uberis. *Journal of Proteome Research 11(3)*, 1832–1843.

(58) Miyamoto, T., Yi Lo, P. H., Saichi, N., Ueda, K., Hirata, M., Tanikawa, C., and Matsuda, K. (2017) Argininosuccinate Synthase 1 Is An Intrinsic Akt Repressor Transactivated by p53. *Science Advances 3(5)*, e1603204.

(59) Varnum, S. M., Webb-Robertson, B. J. M., Hessol, N. A., Smith, R. D., and Zangar, R. C. (2011) Plasma Biomarkers for Detecting Hodgkin's Lymphoma in HIV Patients. *PLoS One 6(12)*, e29263.

(60) Akuffo, E. L., Davis, J. B., Fox, S. M., Gloger, I. S., Hosford, D., Kinsey, E. E., Jones, N. A., Nock, C. M., Roses, A. D., Saunders, A. M., Skehel, J. M., Smith, M. A., and Cutler, P. (2008) The Discovery and Early Validation of Novel Plasma Biomarkers in Mild-to-Moderate Alzheimer's Disease Patients Responding to Treatment with Rosiglitazone. *Biomarkers 13(6)*, 618–636.

(61) Perrin, R. J., Payton, J. E., Malone, J. P., Gilmore, P., Davis, A. E., Xiong, C., Fagan, A. M., Townsend, R. R., and Holtzman, D. M. (2013) Quantitative Label-Free Proteomics for Discovery of Biomarkers in Cerebrospinal Fluid: Assessment of Technical and Inter-Individual Variation. *PLoS One 8(5)*, e64314.

(62) Cutler, P., Akuffo, E. L., Bodnar, W. M., Briggs, D. M., Davis, J. B., Debouck, C. M., Fox, S. M., Gibson, R. A., Gormley, D. A., Holbrook, J. D., Hunter, A. J., Kinsey, E. E., Prinjha, R., Richardson, J. C., Roses, A. D., Smith, M. A., Tsokanas, N., Willé, D. R., Wu, W., Yates, J. W., and Gloger, I. S. (2008) Proteomic Identification and Early Validation of Complement I Inhibitor and Pigment

Epithelium-Derived Factor: Two Novel Biomarkers of Alzheimer's Disease in Human Plasma. *Proteomics - Clinical Applications 2(4)*, 467–477.

(63) Sleddering, M. A., Markvoort, A. J., Dharuri, H. K., Jeyakar, S., Snel, M., Juhasz, P., Lynch, M., Hines, W., Li, X., Jazet, I. M., Adourian, A., Hilbers, P. A. J., Smit, J. W. A., and van Dijk, K. W. (2014) Proteomic Analysis in Type 2 Diabetes Patients Before and After A Very Low Calorie Diet Reveals Potential Disease State and Intervention Specific Biomarkers. *PLoS One 9(11)*, e112835.

(64) Zhong, X., Chen, H., and Zare, R. N. (2020) Ultrafast Enzymatic Digestion of Proteins by Microdroplet Mass Spectrometry. *Nature Communications 11(1)*, 1–9.

(65) Wiśniewski, J. R. (2016) Quantitative Evaluation of Filter Aided Sample Preparation (FASP) and Multienzyme Digestion FASP Protocols. *Analytical Chemistry 88(10)*, 5438–5443.

(66) Martin, N. J., Bunch, J., and Cooper, H. J. (2013) Dried Blood Spot Proteomics: Surface Extraction of Endogenous Proteins Coupled with Automated Sample Preparation and Mass Spectrometry Analysis. *The Journal of the American Society for Mass Spectrometry 24(8)*, 1242–1249.

(67) Stachowicz, A., Siudut, J., Suski, M., Olszanecki, R., Korbut, R., Undas, A., and Wiśniewski, J. R. (2017) Optimization of Quantitative Proteomic Analysis of Clots Generated from Plasma of Patients with Venous Thromboembolism. *Clinical Proteomics 14(1)*, 38–46.

(68) Liu, Y., Zhang, K., Zheng, H., Shang, Y., Guo, J., Tian, H., Lu, G., Jin, Y., He, J., Cai, X., and Liu, X. (2011) Proteomics Analysis of Porcine Serum Proteins by LC-MS/MS after Foot-and-Mouth Disease Virus (FMDV) Infection. *The Journal of Veterinary Medical Science 73(12)*, 1569–1572.

(69) Boerema, D. J., An, B., Gandhi, R. P., Papineau, R., Regnier, E., Wilder, A., Molitor, A., Tang, A. P., and Kee, S. M. (2017) Biochemical Comparison of Four

Commercially Available Human α1-Proteinase Inhibitors for Treatment of α1-Antitrypsin Deficiency. *Biologicals 50(1)*, 63–72.

(70) Wiśniewski, J. R., Wegler, C., and Artursson, P. (2018) Multiple-Enzyme-Digestion Strategy Improves Accuracy and Sensitivity of Label- and Standard-Free Absolute Quantification to a Level That Is Achievable by Analysis with Stable Isotope-Labeled Standard Spiking. *Journal of Proteome Research 18(1)*, 217–227.

(71) Swaney, D. L., Wenger, C. D., and Coon, J. J. (2010) The Value of Using Multiple Proteases for Large-Scale Mass Spectrometry-Based Proteomics. *Journal of Proteome Research 9(3)*, 1323–1329.

(72) Siepen, J. A., Keevil, E. J., Knight, D., and Hubbard, S. J. (2007) Prediction of Missed Cleavage Sites in Tryptic Peptides Aids Protein Identification in Proteomics. *Journal of Proteome Research 6(1)*, 399–408.

(73) Rodriguez, J., Gupta, N., Smith, R. D., and Pevzner, P. A. (2008) Does Trypsin Cut Before Proline? *Journal of Proteome Research 7(1)*, 300–305.

(74) Saveliev, S., Bratz, M., Zubarev, R., Szapacs, M., Budamgunta, H., and Urh, M. (2013) Trypsin/Lys-C Protease Mix for Enhanced Protein Mass Spectrometry Analysis. *Nature Methods 10(11)*, 1–2.

(75) Simpson, K. L., Whetton, A. D., and Dive, C. (2009) Quantitative Mass Spectrometry-Based Techniques for Clinical Use: Biomarker Identification and Quantification. *The Journal of Chromatography. B, Analytical Technologies in the Biomedical and Life Sciences 877(13)*, 1240–1249.

(76) Beretta L. (2007) Proteomics from The Clinical Perspective: Many Hopes and Much Debate. *Nature Methods 4(10)*, 785–786.

(77) Amiri-Dashatan, N., Koushki, M., Abbaszadeh, H.-A., Rostami-Nejad, M., and Rezaei-Tavirani, M. (2018) Proteomics Applications in Health: Biomarker and Drug Discovery and Food Industry. *Iranian Journal of Pharmaceutical Research 17(4)*, 1523–1536.

(78) Anderson, D. C., and Kodukula, K. (2014) Biomarkers in Pharmacology and Drug Discovery. *Biochemical Pharmacology 87(1)*, 172–188.

(79) Poste, G. (2011) Bring on The Biomarkers. *Nature 469(1)*, 156–157.

(80) Patel, S. (2014) Role of Proteomics in Biomarker Discovery: Prognosis and Diagnosis of Neuropsychiatric Disorders. *Advances in Protein Chemistry and Structural Biology 94(1)*, 39–75.

(81) Serna, G., Ruiz-Pace, F., Cecchi, F., Fasani, R., Jimenez, J., Thyparambil, S., Landolfi, S., Elez, E., Vivancos, A., Hembrough, T., Tabernero, J., Dienstmann, R., and Nuciforo, P. (2019) Targeted Multiplex Proteomics for Molecular Prescreening and Biomarker Discovery in Metastatic Colorectal Cancer. *Scientific Reports 9(1)*, 1–10.

(82) Füzéry, A. K., Levin, J., Chan, M. M., and Chan, D. W. (2013) Translation of Proteomic Biomarkers into FDA Approved Cancer Diagnostics: Issues and Challenges. *Clinical Proteomics 10(1)*, 13–27.

(83) Crutchfield, C. A., Thomas, S. N., Sokoll, L. J., and Chan, D. W. (2016) Advances in Mass Spectrometry-Based Clinical Biomarker Discovery. *Clinical Proteomics 13(1)*, 1–12.

(84) Li, D., and Chan, D. W. (2014) Proteomic Cancer Biomarkers from Discovery to Approval: It's Worth the Effort. *Expert Review of Proteomics 11(2)*, 135–136.

(85) Altuntaş, E., Weber, C., Kempe, K., and Schubert, U. S. (2013) Comparison of ESI, APCI and MALDI for The (Tandem) Mass Analysis of Poly(2-Ethyl-2-Oxazoline)s with Various End-Groups. *European Polymer Journal 49(8)*, 2172–2185.

(86) Banerjee, S., and Mazumdar, S. (2012) Electrospray Ionization Mass Spectrometry: A Technique to Access the Information beyond the Molecular Weight of the Analyte. *International Journal of Analytical Chemistry 2012(1)*, 1–40.

(87) Karas, M., and Hillenkamp, F. (1988, October 1) Laser Desorption Ionization of Proteins with Molecular Masses Exceeding 10,000 Daltons. *Analytical Chemistry*. Anal Chem.

(88) Singhal, N., Kumar, M., Kanaujia, P. K., and Virdi, J. S. (2015) MALDI-TOF Mass Spectrometry: An Emerging Technology for Microbial Identification and Diagnosis. *Frontiers in Microbiology 6*.

(89) Aichler, M., and Walch, A. (2015) MALDI Imaging Mass spectrometry: Current Frontiers and Perspectives in Pathology Research and Practice. *Laboratory Investigation 95(4)*, 422–431.

(90) Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., and Whitehouse, C. M. (1989) Electrospray Ionization for Mass Spectrometry of Large Biomolecules. *Science 246(4926)*, 64–71.

(91) Ho, C. S., Lam, C. W. K., Chan, M. H. M., Cheung, R. C. K., Law, L. K., Lit, L. C. W., Ng, K. F., Suen, M. W. M., and Tai, H. L. (2003) Electrospray Ionisation Mass Spectrometry: Principles and Clinical Applications. *The Clinical Biochemist Reviews 24(1)*, 3–12.

(92) Mann, M. (2019) The Ever Expanding Scope of Electrospray Mass Spectrometry—A 30 Year Journey. *Nature Communications 10(1)*, 1–3.

(93) Wilm, M. (2011) Principles of Electrospray Ionization. *Molecular and Cellular Proteomics 10(7)*, 1–8.

(94) Zhang, Y., Wang, J., Liu, J., Han, J., Xiong, S., Yong, W., and Zhao, Z. (2016) Combination of ESI and MALDI Mass Spectrometry for Qualitative, Semi-Quantitative and In-Situ Analysis of Gangliosides in Brain. *Scientific Reports 6(1)*, 1–11.

(95) Aebersold, R., and Mann, M. (2003) Mass Spectrometry-Based Proteomics. *Nature 422(1)*, 198–207.

(96) Aebersold, R., and Goodlett, D. R. (2001) Mass Spectrometry in Proteomics. *Chemical Reviews 101(2)*, 269–296.

(97) Han, X., Aslanian, A., and Yates, J. R. (2008) Mass Spectrometry for Proteomics. *Current Opinion in Chemical Biology 12(5)*, 483–490.

(98) Clarke, W. (2017) Mass Spectrometry in The Clinical Laboratory: Determining The Need and Avoiding Pitfalls, in *Mass Spectrometry for the Clinical Laboratory*, 1–15. Academic Press, Baltimore.

(99) Smith, R. W. (2013) Mass Spectrometry. *Encyclopedia of Forensic Sciences* (Siegel, J., and Saukko, P., Eds.) 2nd Edition. Academic Press.

(100) Cho, Y. T., Su, H., Wu, W. J., Wu, D. C., Hou, M. F., Kuo, C. H., and Shiea, J. (2015) Biomarker Characterization by MALDI-TOF/MS. *Advances in Clinical Chemistry 69(1)*, 209–254.

(101) Guilhaus, M., Selby, D., and Mlynski, V. (2000) Orthogonal Acceleration Time-of-Flight Mass Spectrometry. *Mass Spectrometry Reviews 19(2)*, 65–107.

(102) Sleeman, R., and Carter, J. F. (2004) Mass Spectrometry - Overview. *Encyclopedia of Analytical Science: Second Edition* (Worsfold, P., Townshend, A., and Poole, C., Eds.) 2nd Edition. Elsevier Inc.

(103) Cheng, L., Fenton, R. A., and Pisitkun, T. (2014) Mass spectrometry in Systems Biology. *Recent Advances in Systems Biology Research 500(1)*, 89–107.

(104) Li, D., Tang, Y., and Xu, W. (2016) Ion Collision Cross Section Measurements in Fourier Transform-Based Mass Analyzers. *Analyst 141(12)*, 3554–3561.

(105) Liebler Daniel C., and Zimmerman Lisa J. (2013) Targeted Quantitation of Proteins by Mass Spectrometry. *Biochemistry 52(22)*, 3797–3806.

(106) Yang L, Amad M, Winnik WM, Schoen AE, Schweingruber H, Mylchreest I, and Rudewicz PJ. (2002) Investigation of An Enhanced Resolution Triple Quadrupole Mass Spectrometer for High-Throughput Liquid

Chromatography/Tandem Mass Spectrometry Assays. *Rapid Communications in Mass Spectrometry 16(21)*, 2060–2066.

(107) Wang, E. H., Combe, P. C., and Schug, K. A. (2016) Multiple Reaction Monitoring for Direct Quantitation of Intact Proteins Using a Triple Quadrupole Mass Spectrometer. *The Journal of the American Society for Mass Spectrometry 27(5)*, 886–896.

(108) Anderson, L., and Hunter, C. L. (2006) Quantitative Mass Spectrometric Multiple Reaction Monitoring Assays for Major Plasma Proteins. *Molecular and Cellular Proteomics 5(4)*, 573–588.

(109) Johnson, J. v., Yost, R. A., Kelley, P. E., and Bradford, D. C. (1990) Tandem-in-Space and Tandem-in-Time Mass Spectrometry: Triple Quadrupoles and Quadrupole Ion Traps. *Analytical Chemistry 62(20)*, 2162–2172.

(110) Kennedy, J. J., Abbatiello, S. E., Kim, K., Yan, P., Whiteaker, J. R., Lin, C., Kim, J. S., Zhang, Y., Wang, X., Ivey, R. G., Zhao, L., Min, H., Lee, Y., Yu, M. H., Yang, E. G., Lee, C., Wang, P., Rodriguez, H., Kim, Y., Carr, S. A., and Paulovich, A. G. (2014) Demonstrating The Feasibility of Large-Scale Development of Standardized Assays to Quantify Human Proteins. *Nature Methods 11(2)*, 149–155.

(111) Lange, V., Picotti, P., Domon, B., and Aebersold, R. (2008) Selected Reaction Monitoring for Quantitative Proteomics: A Tutorial. *Molecular Systems Biology 4(222),* 1-14.

(112) Lange, V., Malmström, J. A., Didion, J., King, N. L., Johansson, B. P., Schäfer, J., Rameseder, J., Wong, C. H., Deutsch, E. W., Brusniak, M. Y., Bühlmann, P., Björck, L., Domon, B., and Aebersold, R. (2008) Targeted Quantitative Analysis of Streptococcus Pyogenes Virulence Factors by Multiple Reaction Monitoring. *Molecular and Cellular Proteomics 7(8)*, 1489–1500.

(113) Wolf-Yadlin, A., Hautaniemi, S., Lauffenburger, D. A., and White, F. M. (2007) Multiple reaction Monitoring for Robust Quantitative Proteomic Analysis of

Cellular Signaling Networks. *Proceedings of the National Academy of Sciences of the United States of America 104(14)*, 5860–5865.

(114) Mann, M., and Kelleher, N. L. (2008) Precision Proteomics: The Case for High resolution and High Mass Accuracy. *Proceedings of the National Academy of Sciences of the United States of America 105(47)*, 18132–18138.

(115) Thomas, S. N., and Zhang, H. (2016) Targeted Proteomic Assays for The Verification of Global Proteomics Insights. *Expert Review of Proteomics 13(10)*, 897–899.

(116) Cañas Montalvo, B., López-Ferrer, D., Ramos-Fernández, A., Camafeita, E., and Calvo, E. (2006) Mass Spectrometry Technologies for Proteomics. *Briefings Functinal Genomic and Proteomics 4(4)*, 295–320.

(117) G. Marshall, A., T. Blakney, G., Chen, T., K. Kaiser, N., M. McKenna, A., P. Rodgers, R., M. Ruddy, B., and Xian, F. (2013) Mass Resolution and Mass Accuracy: How Much Is Enough? *Mass Spectrometry 2(2013)*, S0009-S0014.

(118) Mitchell Wells, J., and McLuckey, S. A. (2005) Collision-Induced Dissociation (CID) of Peptides and Proteins. *Methods in Enzymology 402(1)*, 148–185.

(119) Sleno, L., and Volmer, D. A. (2004) Ion Activation Methods for Tandem Mass Spectrometry. *Journal of Mass Spectrometry 39(10)*, 1091–1112.

(120) Keshishian, H., Addona, T., Burgess, M., Kuhn, E., and Carr, S. A. (2007) Quantitative, Multiplexed Assays for Low Abundance Proteins in Plasma by Targeted Mass Spectrometry and Stable Isotope Dilution. *Molecular and Cellular Proteomics 6(12)*, 2212–2229.

(121) Kay, R. G., Gregory, B., Grace, P. B., and Pleasance, S. (2007) The Application of Ultra-Performance Liquid Chromatography/Tandem Mass Spectrometry to The Detection and Quantitation of Apolipoproteins in Human Serum. *Rapid Communications in Mass Spectrometry 21(16)*, 2585–2593.

(122) Szabó, D., Schlosser, G., Vékey, K., Drahos, L., and Révész, Á. (2021) Collision Energies on QTof and Orbitrap Instruments: How to Make Proteomics Measurements Comparable? *The Journal of Mass Spectrometry 56(1)*, e4693.

(123) Révész, Á., Rokob, T. A., Jeanne Dit Fouque, D., Turiák, L., Memboeuf, A., Vékey, K., and Drahos, L. (2018) Selection of Collision Energies in Proteomics Mass Spectrometry Experiments for Best Peptide Identification: Study of Mascot Score Energy Dependence Reveals Double Optimum. *Journal of Proteome Research 17(5)*, 1898–1906.

(124) Zhu, M., Zhao, B., Wei, L., and Wang, S. (2021) alpha-2-Macroglobulin, a Native and Powerful Proteinase Inhibitor, Prevents Cartilage Degeneration Disease by Inhibiting Majority of Catabolic Enzymes and Cytokines. *Current Molecular Biology Reports 7(1)*, 1–7.

(125) Birkenmeier, G., Müller, R., Huse, K., Forberg, J., Gläser, C., Hedrich, H., Nicklisch, S., and Reichenbach, A. (2003) Human α2-Macroglobulin: Genotype-Phenotype Relation. *Experimental Neurology 184(1)*, 153–161.

(126) Thieme, R., Kurz, S., Kolb, M., Debebe, T., Holtze, S., Morhart, M., Huse, K., Szafranski, K., Platzer, M., Hildebrandt, T. B., and Birkenmeier, G. (2015) Analysis of Alpha-2 Macroglobulin from the Long-Lived and Cancer-Resistant Naked Mole-Rat and Human Plasma. *PLoS One* (Roemer, K., Ed.) *10(6)*, e0130470.

(127) Destefano, A., and Hoffman, M. (1991) The Effect of Alpha-2-Macroglobulin-Proteinase Complexes on Macrophage IA Expression in Vivo. *Immunological Investigations 20(1)*, 33–43.

(128) Cater, J. H., Wilson, M. R., and Wyatt, A. R. (2019) Alpha-2-Macroglobulin, a Hypochlorite-Regulated Chaperone and Immune System Modulator. *Oxidative Medicine and Cellular Longevity 2019(1)*, 1–9.

(129) Athauda, S. B. P., Nishigai, M., Arakawa, H., Ikai, A., Ukai, M., and Takahashi, K. (2003) Inhibition of Human Pepsin and Gastricsin by α2-

Macroglobulin. *The Journal of Enzyme Inhibition and Medicinal Chemistry 18(3)*, 219–224.

(130) Heumann, D., and Vischer, T. L. (1988) Immunomodulation by α2-Macroglobulin and α2-Macroglobulin-Proteinase Complexes: The Effect on The Human T Lymphocyte Response. *European Journal of Immunology 18(5)*, 755–760.

(131) Wu, S. M., Patel, D. D., and Pizzo, S. v. (1998) Oxidized α2-Macroglobulin (α2M) Differentially Regulates Receptor Binding by Cytokines/Growth Factors: Implications for Tissue Injury and Repair Mechanisms in Inflammation. *The Journal of Immunology 161(8)*, 4356–4365.

(132) Businaro, R., Fabrizi, C., Fumagalli, L., and Lauro, G. M. (1992) Synthesis and secretion of α2-macroglobulin by human glioma established cell lines. *Experimental Brain Research 88(1)*, 213–218.

(133) Cvirn, G., Gallistl, S., and Muntean, W. (2001) Alpha-2-Macroglobulin Inhibits The Anticoagulant Action of Activated Protein C in Cord and Adult Plasma. *Haemostasis 31(1)*, 1–11.

(134) Hoogendoorn, H., Toh, C. H., Nesheim, M. E., and Giles, A. R. (1991) α2-Macroglobulin Binds and Inhibits Activated Protein C. *Blood 78(9)*, 2283–2290.

(135) Martos, L., Ramón, L. A., Oto, J., Fernández-Pardo, Á., Bonanad, S., Cid, A. R., Gruber, A., Griffin, J. H., España, F., Navarro, S., and Medina, P. (2018) α2 -Macroglobulin Is a Significant in Vivo Inhibitor of Activated Protein C and Low APC:α2M Levels Are Associated with Venous Thromboembolism. *Thrombosis and Haemostasis 118(4)*, 630–638.

(136) Mezzano, D., España, F., Panes, O., Medina, P., Pais, E., Marshall, G., Tagle, R., Downey, P., Caceres, S., González, F., Quiroga, T., and Pereira, J. (2001) Increased Activation of Protein C, But Lower Plasma Levels of Free, Activated Protein C in Uraemic patients: Relationship with Systemic Inflammation and Haemostatic Activation. *British Journal of Haematology 113(4)*, 905–910.

(137) Mathew, S., Arandjelovic, S., Beyer, W. F., Gonias, S. L., and Pizzo, S. v. (2003) Characterization of The Interaction Between α2-Macroglobulin and Fibroblast Growth Factor-2: The Role of Hydrophobic Interactions. *Biochemical Journal 374(1)*, 123–129.

(138) Šunderić, M., Mihailović, N., and Nedić, O. (2014) Protein Molecular Forms of Insulin-Like Growth Factor Binding Protein-2 Change with Aging. *Experimental Gerontology 58(1)*, 154–158.

(139) Feige, J. J., Negoescu, A., Keramidas, M., Souchelnitskiy, S., and Chambaz, E. M. (1996) α2-Macroglobulin: A Binding Protein for Transforming Growth Factor-β and Various Cytokines. *Hormone Research 45(3–5)* , 227–232.

(140) Kovacs, D. M. (2000) α2-Macroglobulin in Late-Onset Alzheimer's Disease. *Experimental Gerontology 35(4)*, 473–479.

(141) Veerhuis, R., Boshuizen, R. S., and Familian, A. (2005) Amyloid Associated Proteins in Alzheimer's and Prion Disease. *Current Drug Targets: CNS and Neurological Disorders 4(3)*, 235–248.

(142) Huang, X., Cuajungco, M. P., Atwood, C. S., Moir, R. D., Tanzi, R. E., and Bush, A. I. (2000) Alzheimer's Disease, β-Amyloid Protein and Zinc. *The Journal of Nutrition 130(5)*, 1488S-1492S.

(143) Gupta, A. K., Pokhriyal, R., Khan, M. I., Kumar, D. R., Gupta, R., Chadda, R. K., Ramachandran, R., Goyal, V., Tripathi, M., and Hariprasad, G. (2019) Cerebrospinal Fluid Proteomics for Identification of α2-Macroglobulin as a Potential biomarker to Monitor Pharmacological Therapeutic Efficacy in Dopamine Dictated Disease States of Parkinson's Disease and Schizophrenia. *Neuropsychiatric Disease and Treatment 15(1)*, 2853–2867.

(144) Varma, V. R., Varma, S., An, Y., Hohman, T. J., Seddighi, S., Casanova, R., Beri, A., Dammer, E. B., Seyfried, N. T., Pletnikova, O., Moghekar, A., Wilson, M. R., Lah, J. J., O'Brien, R. J., Levey, A. I., Troncoso, J. C., Albert, M. S., and

Thambisetty, M. (2017) Alpha-2 Macroglobulin in Alzheimer's Disease: A Marker of Neuronal Injury Through The RCAN1 Pathway. *Molecular Psychiatry 22(1)*, 13–23.

(145) Meyer, C., Hinrichs, W., and Hahn, U. (2012) Human α2-Macroglobulin-Another Variation on the Venus Flytrap. *Angewandte Chemie International Edition 51(21)*, 5045–5047.

(146) Burgess, E. F., Ham, A. J. L., Tabb, D. L., Billheimer, D., Roth, B. J., Chang, S. S., Cookson, M. S., Hinton, T. J., Cheek, K. L., Hill, S., and Pietenpol, J. A. (2008) Prostate Cancer Serum Biomarker Discovery Through Proteomic Analysis of Alpha-2 Macroglobulin Protein Complexes. *Proteomics - Clinical Applications 2(9)*, 1223–1233.

(147) Yoshino, S., Fujimoto, K., Takada, T., Kawamura, S., Ogawa, J., Kamata, Y., Kodera, Y., and Shichiri, M. (2019) Molecular Form and Concentration of Serum α2-Macroglobulin in Diabetes. *Scientific Reports 9(1)*, 1–10.

(148) Housley, J. (1968) Alpha-2-Macroglobulin Levels in Disease in Man. *The Journal of Clinical Pathology 21(1)*, 27–31.

(149) Eubanks, L. M., Stowe, G. N., de Lamo Marin, S., Mayorov, A. v., Hixon, M. S., and Janda, K. D. (2011) Identification of α2-Macroglobulin As a Major Serum Ghrelin Esterase. *Angewandte Chemie International Edition 50(45)*, 10699–10702.

(150) Tabb, D. L., Vega-Montoto, L., Rudnick, P. A., Variyath, A. M., Ham, A.-J. L., Bunk, D. M., Kilpatrick, L. E., Billheimer, D. D., Blackman, R. K., Cardasis, H. L., Carr, S. A., Clauser, K. R., Jaffe, J. D., Kowalski, K. A., Neubert, T. A., Regnier, F. E., Schilling, B., Tegeler, T. J., Wang, M., Wang, P., Whiteaker, J. R., Zimmerman, L. J., Fisher, S. J., Gibson, B. W., Kinsinger, C. R., Mesri, M., Rodriguez, H., Stein, S. E., Tempst, P., Paulovich, A. G., Liebler, D. C., and Spiegelman, C. (2010) Repeatability and Reproducibility in Proteomic

Identifications by Liquid Chromatography-Tandem Mass Spectrometry. *Journal of Proteome Research 9(2)*, 761–776.

(151) Rehman, A. A., Ahsan, H., and Khan, F. H. (2013) Alpha-2-Macroglobulin: A Physiological Guardian. *The Journal of Cellular Physiology 228(8)*, 1665–1675.

(152) Wyatt, A. R., Kumita, J. R., Farrawell, N. E., Dobson, C. M., and Wilson, M. R. (2015) Alpha-2-Macroglobulin Is Acutely Sensitive to Freezing and Lyophilization: Implications for Structural and Functional Studies. *PLoS One 10(6)*, e0130036.

(153) Bateman, A., Martin, M. J., O'Donovan, C., Magrane, M., Alpi, E., Antunes, R., Bely, B., Bingley, M., Bonilla, C., Britto, R., Bursteinas, B., Bye-AJee, H., Cowley, A., da Silva, A., de Giorgi, M., Dogan, T., Fazzini, F., Castro, L. G., Figueira, L., Garmiri, P., Georghiou, G., Gonzalez, D., Hatton-Ellis, E., Li, W., Liu, W., Lopez, R., Luo, J., Lussi, Y., MacDougall, A., Nightingale, A., Palka, B., Pichler, K., Poggioli, D., Pundir, S., Pureza, L., Qi, G., Rosanoff, S., Saidi, R., Sawford, T., Shypitsyna, A., Speretta, E., Turner, E., Tyagi, N., Volynkin, V., Wardell, T., Warner, K., Watkins, X., Zaru, R., Zellner, H., Xenarios, I., Bougueleret, L., Bridge, A., Poux, S., Redaschi, N., Aimo, L., ArgoudPuy, G., Auchincloss, A., Axelsen, K., Bansal, P., Baratin, D., Blatter, M. C., Boeckmann, B., Bolleman, J., Boutet, E., Breuza, L., Casal-Casas, C., de Castro, E., Coudert, E., Cuche, B., Doche, M., Dornevil, D., Duvaud, S., Estreicher, A., Famiglietti, L., Feuermann, M., Gasteiger, E., Gehant, S., Gerritsen, V., Gos, A., Gruaz-Gumowski, N., Hinz, U., Hulo, C., Jungo, F., Keller, G., Lara, V., Lemercier, P., Lieberherr, D., Lombardot, T., Martin, X., Masson, P., Morgat, A., Neto, T., Nouspikel, N., Paesano, S., Pedruzzi, I., Pilbout, S., Pozzato, M., Pruess, M., Rivoire, C., Roechert, B., Schneider, M., Sigrist, C., Sonesson, K., Staehli, S., Stutz, A., Sundaram, S., Tognolli, M., Verbregue, L., Veuthey, A. L., Wu, C. H., Arighi, C. N., Arminski, L., Chen, C., Chen, Y., Garavelli, J. S., Huang, H., Laiho, K., McGarvey, P., Natale, D. A., Ross, K., Vinayaka, C. R., Wang, Q., Wang, Y., Yeh, L. S., and Zhang, J. (2004)

UniProt: The Universal Protein Knowledgebase. *Nucleic Acids Research 32(D1)*, D115–D119.

(154) Pundir, S., Martin, M. J., and O'Donovan, C. (2017) UniProt Protein Knowledgebase. *Methods in Molecular Biology 1558(1)*, 41–55.

(155) Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D., and Bairoch, A. (2003) ExPASy: The Proteomics Server for In-Depth Protein Knowledge and Analysis. *Nucleic Acids Research 31(13)*, 3784–3788.

(156) Garcia-Ferrer, I., Marrero, A., Gomis-Rüth, F. X., and Goulas, T. (2017) α2-Macroglobulins: Structure and Function. *Sub-Cellular Biochemistry 83(1)*, 149–183.

(157) Marrero, A., Duquerroy, S., Trapani, S., Goulas, T., Guevara, T., Andersen, G. R., Navaza, J., Sottrup-Jensen, L., and Gomis-Rüth, F. X. (2012) The Crystal Structure of Human α2-Macroglobulin Reveals a Unique Molecular Cage. *Angewandte Chemie International Edition 51(14)*, 3340–3344.

(158) Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A. W., Kavukcuoglu, K., Kohli, P., and Hassabis, D. (2021) Highly Accurate Protein Structure Prediction with AlphaFold. *Nature 596(1)*, 583–589.

(159) Croft, D., O'Kelly, G., Wu, G., Haw, R., Gillespie, M., Matthews, L., Caudy, M., Garapati, P., Gopinath, G., Jassal, B., Jupe, S., Kalatskaya, I., Mahajan, S., May, B., Ndegwa, N., Schmidt, E., Shamovsky, V., Yung, C., Birney, E., Hermjakob, H., D'Eustachio, P., and Stein, L. (2011) Reactome: A Database of Reactions, Pathways and Biological Processes. *Nucleic Acids Research 39(D1)*, D691–D697.

(160) Ambrus G, Gál P, Kojima M, Szilágyi K, Balczer J, Antal J, Gráf L, Laich A, Moffatt BE, Schwaeble W, Sim RB, and Závodszky P. (2003) Natural Substrates

and Inhibitors of Mannan-Binding Lectin-Associated Serine Protease-1 and -2: A Study on Recombinant Catalytic Fragments. *The Journal of immunology 170(3)*, 1374–1382.

(161) Salvesen GS, Sayers CA, and Barrett AJ. (1981) Further Characterization of The Covalent Linking Reaction of Alpha 2-Macroglobulin. *The Biochemical journal 195(2)*, 453–461.

(162) Hyldgaard Jensen, P. E., and Sottrup-Jensen, L. (1986) Primary Structure of Human α2-Macroglobulin. Complete Disulfide Bridge Assignment and Localization of Two Interchain Bridges in the Dimeric Proteinase Binding Unit. *The Journal of Biological Chemistry 261(34)*, 15863–15869.

(163) European Medicines Agency. (2011) Guideline on Bioanalytical Method Validation. United Kingdom.

(164) Busher, J. T. (1990) Serum Albumin and Globulin, in *Clinical Methods: The History, Physical, and Laboratory Examinations* (Walker, K. H., Hall, D. W., and Hurst Willis J, Eds.) 3rd Edition. Butterworths, Boston.

(165) Wang, S., Wei, X., Zhou, J., Zhang, J., Li, K., Chen, Q., Terek, R., Fleming, B. C., Goldring, M. B., Ehrlich, M. G., Zhang, G., and Wei, L. (2014) Identification of α2-Macroglobulin As A Master Inhibitor of Cartilage-Degrading Factors That Attenuates The Progression of Posttraumatic Osteoarthritis. *Arthritis and Rheumatology 66(7)*, 1843–1853.

(166) Sottrup-Jensen, L. (1989) α-Macroglobulins: Structure, Shape, and Mechanism of Proteinase Complex Formation. *The Journal of Biological Chemistry 264(20)*, 11539–15542.

(167) Biltoft, D., Gram, J. B., Larsen, A., Münster, A. M. B., Sidelmann, J. J., Skjoedt, K., and Palarasah, Y. (2017) Fast Form Alpha-2-Macroglobulin - A Marker for Protease Activation in Plasma Exposed to Artificial Surfaces. *Clinical Biochemistry 50(18)*, 1203–1208.

(168) Rigsby, R. E., and Parker, A. B. (2016) Using The PyMOL Application to Reinforce Visual Understanding of Protein Structure. *Biochemistry and Molecular Biology Education 44(5)*, 433–437.

(169) Maclean, B., Tomazela, D. M., Shulman, N., Chambers, M., Finney, G. L., Frewen, B., Kern, R., Tabb, D. L., Liebler, D. C., and Maccoss, M. J. (2010) Gene Expression Skyline: An Open Source Document Editor for Creating and Analyzing Targeted Proteomics Experiments. *Bioinformatics 26(7)*, 966–968.

(170) Välikangas, T., Suomi, T., and Elo, L. L. (2016) A Systematic Evaluation of Normalization Methods in Quantitative Label-Free Proteomics. *Briefings in Bioinformatics 19(1)*, 1–11.

(171) Chawade, A., Alexandersson, E., and Levander, F. (2014) Normalyzer: A Tool for Rapid Evaluation of Normalization Methods for Omics Data Sets. *Journal of Proteome Research 13(6)*, 3114–3120.

(172) Mukaka, M. M. (2012) Statistics Corner: A Guide to Appropriate Use of Correlation Coefficient in Medical Research. *Malawi Medical Journal 24(3)*, 69–71.

(173) Key, M. (2012) A Tutorial in Displaying Mass Spectrometry-Based Proteomic Data Using Heat Maps. *BMC Bioinformatics 13(16)*, S10–S23.

(174) Wilkins, M. R., Gasteiger, E., Bairoch, A., Sanchez, J. C., Williams, K. L., Appel, R. D., and Hochstrasser, D. F. (1999) Protein Identification and Analysis Tools in The ExPASy Server. *Methods in Molecular Biology 112(1)*, 531–552.

(175) Gaudet, P., Michel, P. A., Zahn-Zabal, M., Britan, A., Cusin, I., Domagalski, M., Duek, P. D., Gateau, A., Gleizes, A., Hinard, V., de Laval, V. R., Lin, J. J., Nikitin, F., Schaeffer, M., Teixeira, D., Lane, L., and Bairoch, A. (2017) The neXtProt Knowledgebase on Human Proteins: 2017 Update. *Nucleic Acids Research 45(D1)*, D177–D182.

(176) Schaeffer, M., Gateau, A., Teixeira, D., Michel, P. A., Zahn-Zabal, M., and Lane, L. (2017) The neXtProt Peptide Uniqueness Checker: A tool for The Proteomics Community. *Bioinformatics 33(21)*, 3471–3472.

(177) Pundir, S., Martin, M. J., and O'Donovan, C. (2016) UniProt Tools. *Current Protocols in Bioinformatics 53(1)*, 1.29.1-1.29.15.

(178) Song, E., Gao, Y., Wu, C., Shi, T., Nie, S., Fillmore, T. L., Schepmoes, A. A., Gritsenko, M. A., Qian, W.-J., Smith, R. D., Rodland, K. D., and Liu, T. (2017) Targeted Proteomic Assays for Quantitation of Proteins Identified by Proteogenomic Analysis of Ovarian Cancer. *Scientific Data 4(1)*, 1–13.

(179) Burley, S. K., Berman, H. M., Christie, C., Duarte, J. M., Feng, Z., Westbrook, J., Young, J., and Zardecki, C. (2018) RCSB Protein Data Bank: Sustaining a Living Digital Data Resource That Enables Breakthroughs in Scientific Research and Biomedical Education. *Protein Science 27(1)*, 316–330.

(180) Perrin, R. J., Craig-Schapiro, R., Malone, J. P., Shah, A. R., Gilmore, P., Davis, A. E., Roe, C. M., Peskind, E. R., Li, G., Galasko, D. R., Clark, C. M., Quinn, J. F., Kaye, J. A., Morris, J. C., Holtzman, D. M., Townsend, R., and Fagan, A. M. (2011) Identification and Validation of Novel Cerebrospinal Fluid Biomarkers for Staging Early Alzheimer's Disease. *PLoS One 6(1)*, e16032.

(181) Pierce, J. M., and Abbott, K. L. (2010, June 11) Glycoprotein Cancer Biomarker.

(182) Murugan, B. D. (2015, June 25) Proteomic Investigation of Blood-based Biomarkers for the Diagnosis of HIV-associated Neurocognitive Disorders. University of Cape Town, Cape Town.

(183) Singh, V., van Pelt, E. D., Stoop, M. P., Stingl, C., Ketelslegers, I. A., Neuteboom, R. F., Berrevoets, C. E. C., Luider, T. M., and Hintzen, R. Q. (2015) Gray Matter-Related Proteins Are Associated with Childhood-Onset Multiple Sclerosis. *Neurology: Neuroimmunology and NeuroInflammation 2(5)*, 1–10.

(184) Mann, B. F., Goetz, J. A., House, M. G., Schmidt, C. M., and Novotny, M. v. (2012) Glycomic and Proteomic Profiling of Pancreatic Cyst fluids Identifies Hyperfucosylated Lactosamines on The N-Linked Glycans of Overexpressed Glycoproteins. *Molecular and Cellular Proteomics 11(7)*, 1–11.

(185) Krishnan, S., Whitwell, H. J., Cuenco, J., Gentry-Maharaj, A., Menon, U., Pereira, S. P., Gaspari, M., and Timms, J. F. (2017) Evidence of Altered Glycosylation of Serum Proteins Prior to Pancreatic Cancer Diagnosis. *International Journal of Molecular Sciences 18(12)*, 2670–2686.

(186) Murgia, M., Toniolo, L., Nagaraj, N., Ciciliot, S., Vindigni, V., Schiaffino, S., Reggiani, C., and Mann, M. (2017) Single Muscle Fiber Proteomics Reveals Fiber-Type-Specific Features of Human Muscle Aging. *Cell Reports 19(11)*, 2396–2409.

(187) Kim, H. J., Ahn, S. J., Woo, S. J., Hong, H. K., Suh, E. J., Ahn, J., Park, J. H., Ryoo, N. K., Lee, J. E., Kim, K. W., Park, K. H., and Lee, C. (2016) Proteomics-Based Identification and Validation of Novel Plasma Biomarkers Phospholipid Transfer Protein and Mannan-Binding Lectin Serine Protease-1 in Age-Related Macular Degeneration. *Scientific Reports 6(1)*, 1–12.

(188) Wu, C. C., Lin, J. der, Chen, J. T., Chang, C. M., Weng, H. F., Hsueh, C., Chien, H. P., and Yu, J. S. (2018) Integrated Analysis of Fine-Needle-Aspiration Cystic Fluid Proteome, Cancer Cell Secretome, and Public Transcriptome Datasets for Papillary Thyroid Cancer Biomarker Discovery. *Oncotarget 9(15)*, 12079–12100.

(189) O'Brien, D. (2011) The Development and Implementation of a TMT-SRM Assay for the Validation of Candidate Biomarkers of Alzheimer's Disease. United Kingdom.

(190) Nielsen, C. T., Østergaard, O., Rasmussen, N. S., Jacobsen, S., and Heegaard, N. H. H. (2017) A Review of Studies of The Proteomes of Circulating Microparticles: Key Roles for Galectin-3-Binding Protein-Expressing

Microparticles in Vascular Diseases and Systemic Lupus Erythematosus. *Clinical Proteomics 14(1)*.

(191) Bradford, C., Severinsen, R., Pugmire, T., Rasmussen, M., Stoddard, K., Uemura, Y., Wheelwright, S., Mentinova, M., Chelsky, D., Hunsucker, S. W., Kearney, P., Hickok, D., Fleischer, T. C., Ichetovkin, I., Boniface, J. J., Critchfield, G. C., and Peltier, J. M. (2017) Analytical Validation of Protein Biomarkers for Risk of Spontaneous Preterm Birth. *Clinical Mass Spectrometry 3(1)*, 25–38.

(192) Bennett, S. J. (2010, August) Oxidative Stress Biomarkers in Dementia. United Kingdom.

(193) Garin, J., Masselon, C., Vlahou, A., Makrydakis, M., Zoidakis, I., Domon, B., Duriez, E., Radvanyi, F., Kamoun, A., Allory, Y., Malats Riera, N., Marquez Cid, M., and Court, M. (2017, June 29) Bladder Cancer Biomarker Proteins.

(194) Singh, S., Akhter, M. S., Dodt, J., Volkers, P., Reuter, A., Reinhart, C., Krettler, C., Oldenburg, J., and Biswas, A. (2019) Identification of Potential Novel Interacting Partners for Coagulation Factor XIII B (FXIII-B) Subunit, A Protein Associated with A Rare Bleeding Disorder. *International Journal of Molecular Sciences 20(11)*, 2682–2694.

(195) Chen, C. L., Lai, Y. F., Tang, P., Chien, K. Y., Yu, J. S., Tsai, C. H., Chen, H. W., Wu, C. C., Chung, T., Hsu, C. W., Chen, C. de, Chang, Y. S., Chang, P. L., and Chen, Y. T. (2012) Comparative and Targeted Proteomic Analyses of Urinary Microparticles from Bladder Cancer and Hernia Patients. *Journal of Proteome Research 11(12)*, 5611–5629.

(196) Xu, G., and Chance, M. R. (2005) Radiolytic Modification and Reactivity of Amino Acid Residues Serving as Structural Probes for Protein Footprinting. *Analytical Chemistry 77(14)*, 4549–4555.

(197) Kempkes, L. J. M., Martens, J., Grzetic, J., Berden, G., and Oomens, J. (2016) Deamidation Reactions of Asparagine- and Glutamine-Containing Dipeptides

Investigated by Ion Spectroscopy. *The Journal of the American Society for Mass Spectrometry 27(11)*, 1855–1869.

(198) Tan, A., and Awaiye, K. (2013) Use of Internal Standards in LC-MS Bioanalysis, in *Handbook of LC-MS Bioanalysis: Best Practices, Experimental Protocols, and Regulations* (Li, W., Zhang, J., and Tse, F. L. S., Eds.), 217–227. John Wiley & Sons, Ltd.

(199) Mead, J. A., Bianco, L., Ottone, V., Barton, C., Kay, R. G., Lilley, K. S., Bond, N. J., and Bessant, C. (2009) MRMaid, The Web-Based Tool for Designing Multiple Reaction Monitoring (MRM) Transitions. *Molecular and Cellular Proteomics 8(4)*, 696–705.

(200) Klammer, A. A., MacCoss, M. J., Wu, C. C., and Noble, W. S. (2005) Peptide Charge State Determination for Low-Resolution Tandem Mass Spectra, in *Proceedings - 2005 IEEE Computational Systems Bioinformatics Conference*, 175–185.

(201) Bian, Y., Zheng, R., Bayer, F. P., Wong, C., Chang, Y.-C., Meng, C., Zolg, D. P., Reinecke, M., Zecha, J., Wiechmann, S., Heinzlmeir, S., Scherr, J., Hemmer, B., Baynham, M., Gingras, A.-C., Boychenko, O., and Kuster, B. (2020) Robust, Reproducible and Quantitative Analysis of Thousands of Proteomes by Micro-Flow LC–MS/MS. *Nature Communications 11(1)*, 1–12.

(202) Qian WJ, Liu T, Petyuk VA, Gritsenko MA, Petritis BO, Polpitiya AD, Kaushal A, Xiao W, Finnerty CC, Jeschke MG, N, J., ME, M., RJ, M., LL, M., RW, D., RG, T., DN, H., DG, C., and RD, S. (2009) Large-Scale Multiplexed Quantitative Discovery Proteomics Enabled by The Use of an (18)O-Labeled “Universal” Reference Sample. *Journal of Proteome Research 8(1)*, 290–299.

(203) Piehowski, P. D., Petyuk, V. A., Orton, D. J., Xie, F., Ramirez-Restrepo, M., Engel, A., Lieberman, A. P., Albin, R. L., Camp, D. G., Smith, R. D., and Myers, A. J. (2013) Sources of Technical Variability in Quantitative LC-MS Proteomics:

Human Brain Tissue Sample Analysis. *Journal of Proteome Research 12(5)*, 2128–2137.

(204) Chiva C, Ortega M, and Sabidó E. (2014) Influence of The Digestion Technique, Protease, and Missed Cleavage Peptides in Protein Quantitation. *Journal of Proteome Research 13(9)*, 3979–3986.

(205) Dubois, E., Galindo, A. N., Dayon, L., and Cominetti, O. (2020) Comparison of Normalization Methods in Clinical Research Applications of Mass Spectrometry-Based Proteomics, in *2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology, CIBCB 2020*, 1–10. Institute of Electrical and Electronics Engineers Inc.

(206) Kuzyk MA, Smith D, Yang J, Cross TJ, Jackson AM, Hardie DB, Anderson NL, and Borchers CH. (2009) Multiple Reaction Monitoring-Based, Multiplexed, Absolute Quantitation of 45 Proteins in Human Plasma. *Molecular and Cellular Proteomics 8(8)*, 1860–1877.

(207) Chace D. H., Barr J. R., Duncan M. W., Matern D., Morris M. R., Palmer-Toy D. E., Rockwood A. L., Siuzdak G., Urbani A., Yergev A. L., and Chan Y. M. (2007) Mass Spectrometry in The Clinical Laboratory: General Principles and Guidance; Approved Guideline. *Clinical and Laboratory Standards Institute, 27(24).*

(208) Maia, T. M., Staes, A., Plasman, K., Pauwels, J., Boucher, K., Argentini, A., Martens, L., Montoye, T., Gevaert, K., and Impens, F. (2020) Simple Peptide Quantification Approach for MS-Based Proteomics Quality Control. *ACS Omega 5(12)*, 6754–6762.

(209) Anjo, S. I., Simões, I., Castanheira, P., Grãos, M., and Manadas, B. (2018) A Generic Normalization Method for Proper Quantification in Untargeted Proteomics Screening. *bioRxiv preprint*.

(210) Pandeswari, P. B., and Sabareesh, V. (2019) Middle-Down Approach: A Choice to Sequence and Characterize Proteins/Proteomes by Mass Spectrometry. *RSC Advances 9(1)*, 313–344.

APPENDICES

A. HPLC and MS Results of Peptide Standards and IS

The HPLC chromatograms and MS spectra of the synthesized peptide standards were represented in this section.

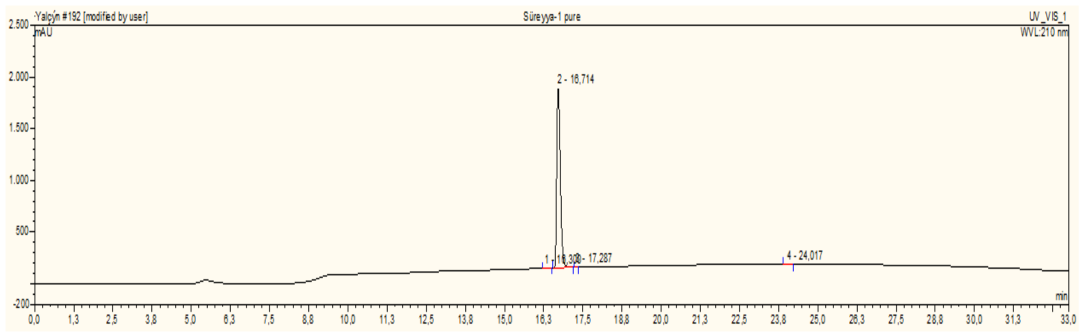

Figure 4.1. *HPLC chromatogram of FEVQVTVPK*

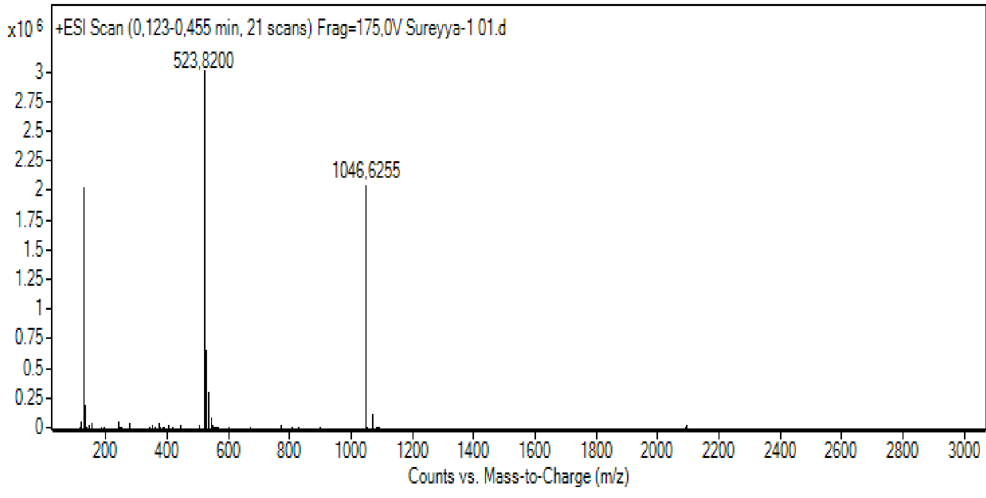

Figure 4.2. *Mass spectrum of FEVQVTVPK*

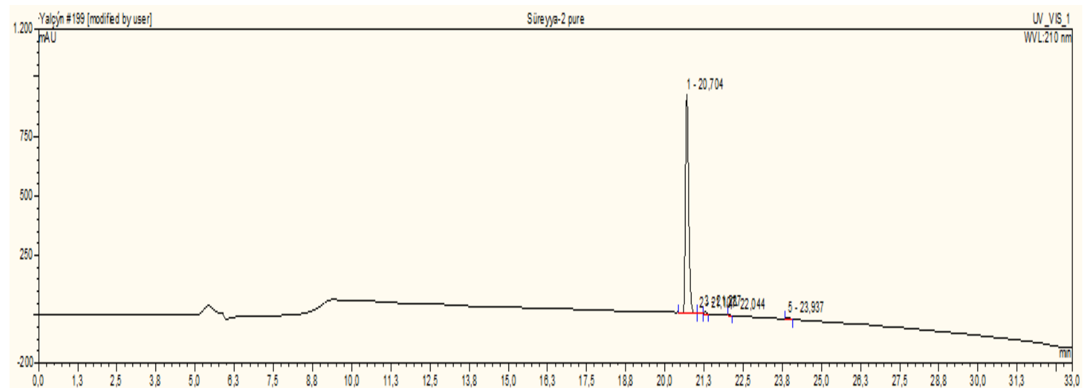

Figure 4.3. *HPLC chromatogram of DMYSFLEDMGLK*

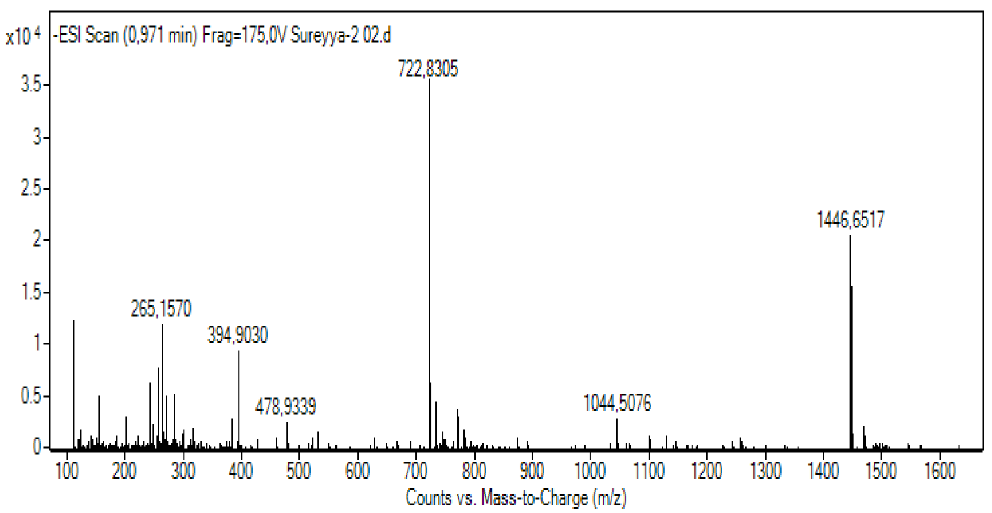

Figure 4.4. *Mass spectrum of DMYSFLEDMGLK*

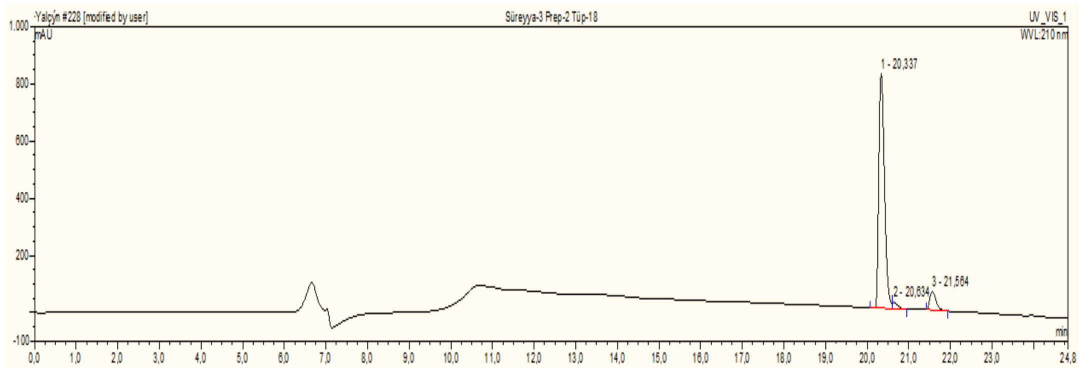

Figure 4.5. *HPLC chromatogram of QGIPFFGQVR*

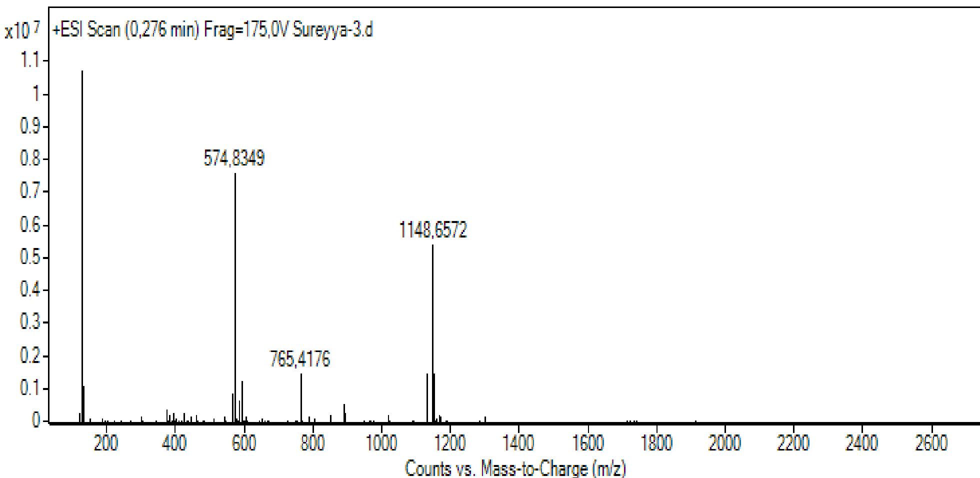

Figure 4.6. *Mass spectrum of QGIPFFGQVR*

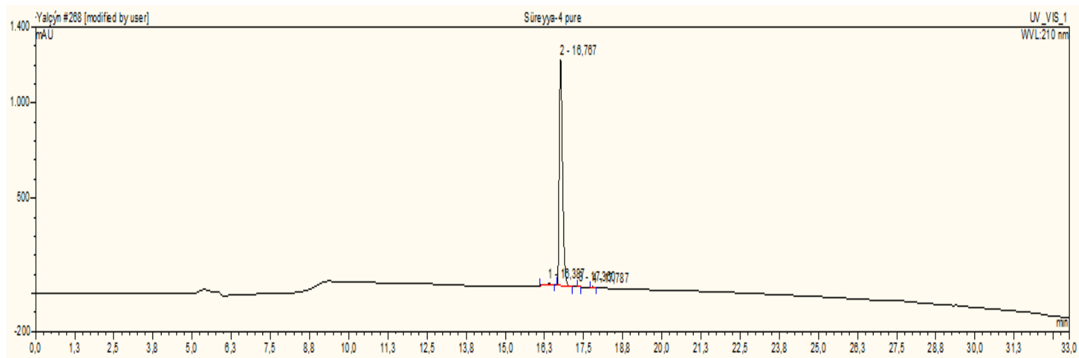

Figure 4.7. *HPLC chromatogram of VGFYESDVMGR*

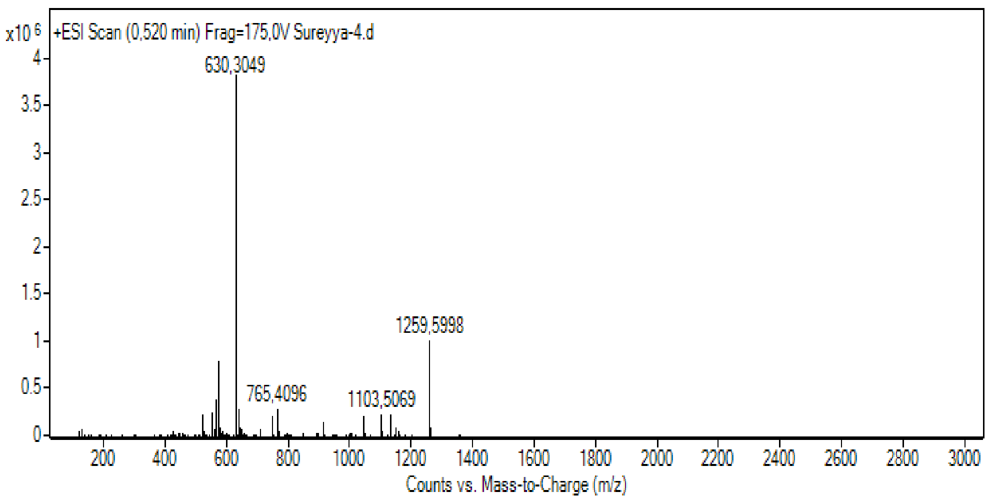

Figure 4.8. *Mass spectrum of VGFYESDVMGR*

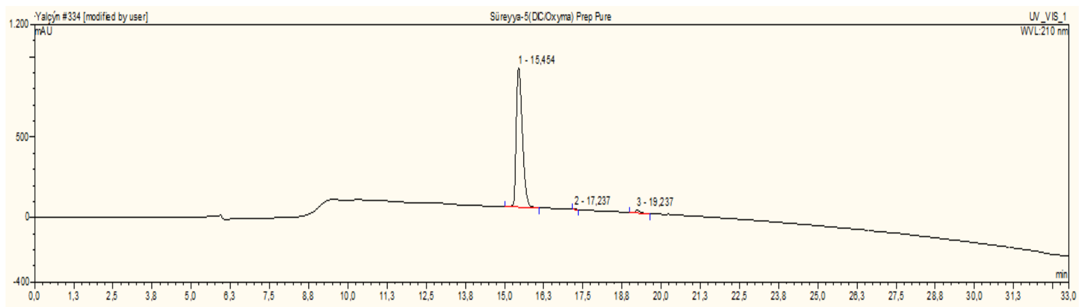

Figure 4.9. *HPLC chromatogram of LPPNVVEESAR*

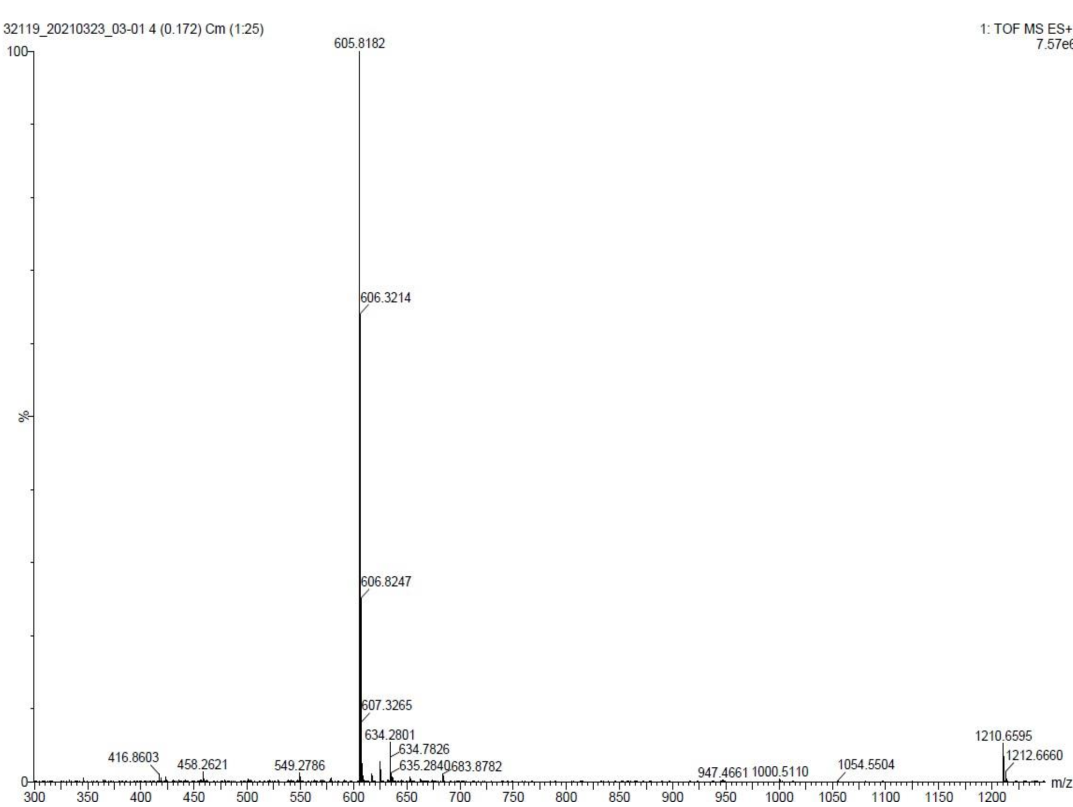

Figure 4.10. *Mass spectrum of LPPNVVEESAR*

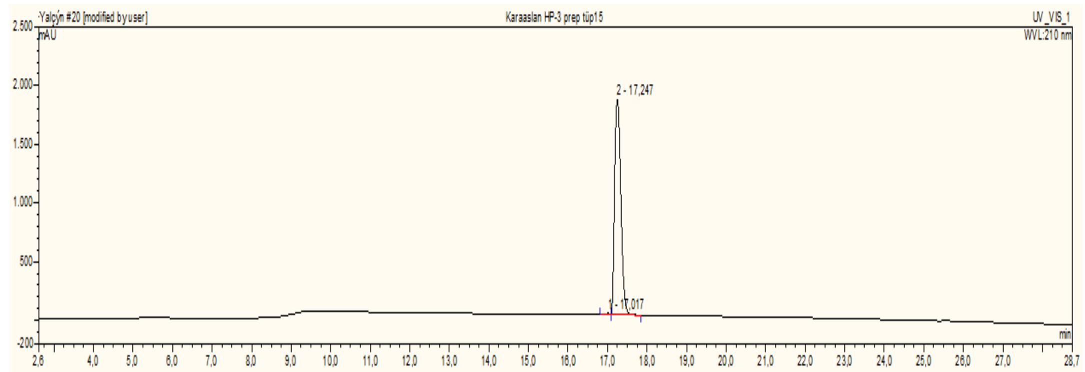

Figure 4.11. *HPLC chromatogram of TFLLR (Internal Standard)*

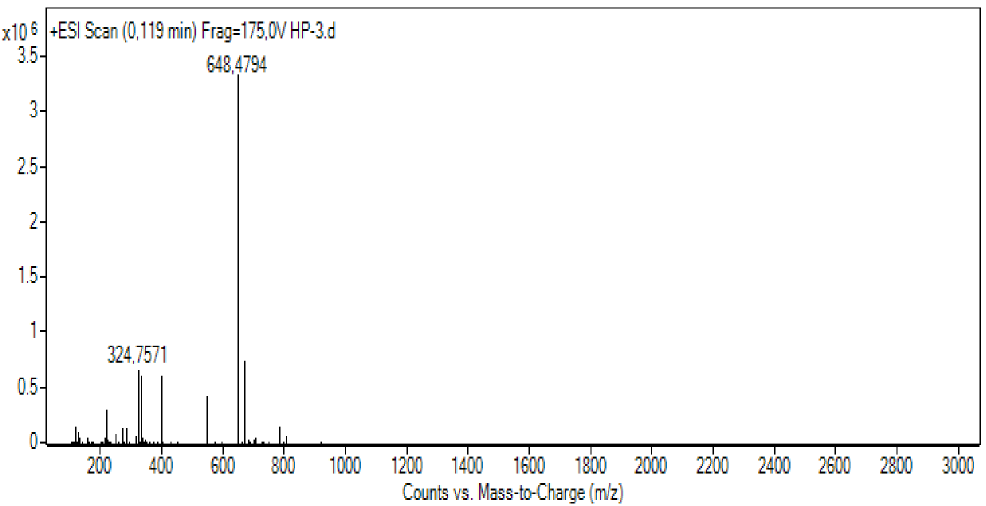

Figure 4.12. *Mass spectrum of TFLLR (Internal Standard*