Applications of peptide-based systems in diagnosis of cancer and COVID-19

by

Sumia Ehsan

B.S., University of Sargodha, 2016
M.S., Western Illinois University, 2018

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Chemistry
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2022

Abstract

Early detection is critical in diagnosing any disease as it increases the chances of survival and saves a patient from experiencing severe symptoms and related treatments, especially in diseases like cancer and coronavirus. At the molecular level, these conditions are found to be associated with an altered expression of specific biomarkers (biological molecules indicative of normal behavior or abnormality in a physiological process or condition), such as proteases and arginases. As these biomarkers are present not only in the tissues but also in the blood and other body fluids, liquid biopsy can be conveniently performed to detect these disease-specific enzymes, reducing patient discomfort, and allowing for repeated tests over time to track the development of the disease and monitoring the progress of treatment. To target such biomarkers, the use of peptides as a targeting moiety is getting substantial attention from scientists because of the ease of synthesis, flexibility in structure and function, cost-effectiveness, high specificity and selectivity towards a receptor, and high biocompatibility.

This work focuses on developing peptide-based detection platforms for the clinical diagnosis of cancer (specifically lung cancer) at its early stages and coronavirus disease via liquid biopsy. The design of biosensors consists of an enzyme-specific peptide sequence labeled with tetrakis-carboxyphenyl-porphyrin (TCPP) and explosion graphene core-shell attached to polyethylenimine (PEI, to link peptide via amide bond). These consensus peptide sequences act as substrates for the enzymes/proteases that are overexpressed in lung cancer and COVID-19 diseases, hence, measuring the activity of lung cancer and coronavirus protease and arginase biomarkers. These detection systems work on the principle of Förster resonance energy transfer, where the cleaving of consensus peptide sequence by relevant proteases results in the termination of the fluorescence quenching of TCPP by graphene. The increase in the fluorescence intensity

of TCPP upon the termination of quenching is measured using a plate reader. Furthermore, we optimized the solid-phase peptide synthesis protocol to find a suitable set of conditions to get the maximum purity of synthetic peptides, as confirmed by ultra-performance liquid chromatography-mass spectrometry (UPLC-MS). Hence, the newly developed biosensors with ultrapure peptides are highly sensitive and can detect the overly expressed disease biomarkers using a conventional plate reader down to the sub-femtomolar level and can potentially be used in laboratories with limited resources without any compromise on efficiency and sensitivity.

Applications of peptide-based systems in diagnosis of cancer and COVID-19

by

Sumia Ehsan

B.S., University of Sargodha, 2016
M.S., Western Illinois University, 2018

A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Chemistry
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2022

Approved by:

Co-Major Professor
Dr. Christine Aikens

Approved by:

Co-Major Professor
Dr. Stefan H. Bossmann

Copyright

Abstract

Early detection is critical in diagnosing any disease as it increases the chances of survival and saves a patient from experiencing severe symptoms and related treatments, especially in diseases like cancer and coronavirus. At the molecular level, these conditions are found to be associated with an altered expression of specific biomarkers (biological molecules indicative of normal behavior or abnormality in a physiological process or condition), such as proteases and arginases. As these biomarkers are present not only in the tissues but also in the blood and other body fluids, liquid biopsy can be conveniently performed to detect these disease-specific enzymes, reducing patient discomfort, and allowing for repeated tests over time to track the development of the disease and monitoring the progress of treatment. To target such biomarkers, the use of peptides as a targeting moiety is getting substantial attention from scientists because of the ease of synthesis, flexibility in structure and function, cost-effectiveness, high specificity and selectivity towards a receptor, and high biocompatibility.

This work focuses on developing peptide-based detection platforms for the clinical diagnosis of cancer (specifically lung cancer) at its early stages and coronavirus disease via liquid biopsy. The design of biosensors consists of an enzyme-specific peptide sequence labeled with tetrakis-carboxyphenyl-porphyrin (TCPP) and explosion graphene core-shell attached to polyethylenimine (PEI, to link peptide via amide bond). These consensus peptide sequences act as substrates for the enzymes/proteases that are overexpressed in lung cancer and COVID-19 diseases, hence, measuring the activity of lung cancer and coronavirus protease and arginase biomarkers. These detection systems work on the principle of Förster resonance energy transfer, where the cleaving of consensus peptide sequence by relevant proteases results in the termination of the fluorescence quenching of TCPP by graphene. The increase in the fluorescence intensity

of TCPP upon the termination of quenching is measured using a plate reader. Furthermore, we optimized the solid-phase peptide synthesis protocol to find a suitable set of conditions to get the maximum purity of synthetic peptides, as confirmed by ultra-performance liquid chromatography-mass spectrometry (UPLC-MS). Hence, the newly developed biosensors with ultrapure peptides are highly sensitive and can detect the overly expressed disease biomarkers using a conventional plate reader down to the sub-femtomolar level and can potentially be used in laboratories with limited resources without any compromise on efficiency and sensitivity.

Table of Contents

Table of Contents viii
List of Figures xi
List of Tables xvii
List of Schemes xviii
Acknowledgements xix
Dedication xxii
Chapter 1 - Potential of Synthetic Peptides in Diagnosis and Therapy of Cancer and COVID-19 1
1.1. Cancer: a continuous battle 1
1.2. History and importance of peptides in cancer research 4
1.3. Biomedical applications of peptides in cancer diagnostics and therapy 5
1.3.1. Peptides in the detection of cancer 5
1.3.2. Peptides in the subcellular-targeted drug delivery[1] 9
1.3.3. Peptides in imaging applications 11
1.4. Biomedical applications of peptides in COVID-19 diagnostics and therapy 13
1.5. Conclusion 15
1.6. References 17
Chapter 2 - Peptide synthesis and optimization 21
Abstract 21
2.1. Introduction 21
2.2. Materials and Methods 24
2.2.1. Selection of Resin 25
2.2.2. Removal of the Fmoc-group 26
2.2.2.1. Fmoc-deprotection of short fragments 26
2.2.2.2. The Fmoc-deprotection of GKK fragments 27
2.2.2.3. The Fmoc-deprotection of MMP2 peptide sequences 28
2.2.3. Coupling of Amino Acids 28
2.2.4. Labeling of MMP2 peptides with dye 29
2.2.5. Acetic Anhydride 29
2.2.6. Cleaving of resin 29
2.2.7. Analysis of the synthesized peptides 30
2.3. Results and Discussion 30
2.3.1. Optimization of Fmoc-deprotection conditions 37
2.3.1.1. Diethylamine for the Fmoc-deprotection 37
2.3.1.1.1. Optimization of the concentration of diethylamine at room temperature 37
2.3.1.1.2. Optimization of the concentration of diethylamine at higher temperature 44
2.3.1.1.3. Diethylamine for the synthesis of longer peptide chains 49
2.3.1.2. Piperidine for the Fmoc-deprotection 53
2.3.1.2.1. Optimization of the concentration of piperidine at room temperature 54
2.3.1.2.2. Optimization of the concentration of piperidine at higher temperature 62
2.3.1.3. Multiple incubations with piperidine and diethylamine 65
2.3.1.4. Piperidine for the synthesis of longer peptide chains 70
2.4. Conclusion 73
2.5. References 75
Chapter 3 - Developing peptide-based nanoplatform for lung cancer diagnosis 78

Abstract 78
3.1. Introduction 79
3.1.1. Biomarkers for cancer detection 80
3.1.2. Platform for lung cancer detection 82
3.1.2.1. Graphene 84
3.2. Materials and Methods 86
3.2.1. Gene expression analysis 86
3.2.2. Synthesis of peptides 86
3.2.2.1. Labeling of consensus peptide sequences with the dye TCPP 87
3.2.2.2. Cleaving of the resin 87
3.2.2.3. Analysis of the synthesized peptides 88
3.2.3. Synthesis of carboxygraphene 88
3.2.4. Loading of peptide 89
3.2.5. Enzymatic digestion experiments 90
3.2.6. Fluorescence measurements with detonated graphene-biosensors 91
3.3. Results and Discussion 91
3.3.1. Peptide syntheses and characterizations 94
3.3.2. Reaction kinetics analysis between enzymes and peptide substrates 96
3.3.3. Carboxygraphene and carboxygraphene-PEI characterizations 98
3.3.4. Dynamic Light Scattering and Zeta-Potential measurements of carboxygraphene, carboxygraphene-PEI, and nanobiosensors 101
3.3.5. Ultraviolet/Visible (UV/Vis) spectroscopy analysis 104
3.3.6. Transmission electron microscopy (TEM) analysis 106
3.3.7. Optimizations of designed nanobiosensors 107
3.3.7.1. Optimization of the peptide concentration 107
3.3.7.2. Optimization of the nanobiosensor concentration 109
3.3.7.3. Activity in buffer 110
3.3.7.4. Adjusting the pH 110
3.3.7.5. Adjusting the incident wavelength 111
3.3.7.6. Determination of limit of detection (LOD) 112
3.4. Conclusion 114
3.5. References 115
Chapter 4 - Developing peptide-based nanoplatform for COVID-19 diagnosis 118
Abstract 118
4.1. Introduction 119
4.1.1. Biomarkers for COVID-19 detection 121
4.2. Materials and Methods 122
4.2.1. Synthesis of peptides 122
4.2.1.1. Labeling of consensus peptide sequences with dye 123
4.2.1.2. Cleaving of the resin 123
4.2.1.3. Analysis of the synthesized peptides 123
4.2.2. Enzymatic digestion experiments 124
4.2.3. Synthesis of nanoparticles 124
4.2.4. Assembly of nanobiosensors 124
4.2.5. Fluorescence measurements with Fe/Fe_3O_4-biosensors 125
4.2.6. Fluorescence measurements with detonated graphene-biosensors 126

4.3. Results and Discussion 126
4.3.1. Iron/Iron oxide-based nanobiosensors for Cysteine Protease (PLpro) and 3C-like Serine Protease (3CLpro) 126
4.3.2. Designing consensus sequences for Cysteine Protease (PLpro) and 3C-like Serine Protease (3CLpro) 128
4.3.3. UPLC-MS characterization of consensus peptide sequences 129
4.3.4. Bioinformatics: CABS-dock server for flexible protein-peptide docking 131
4.3.5. Reaction kinetics analysis between enzymes and peptide substrates 132
4.3.6. Activity of iron/iron oxide-based nanobiosensors for Cysteine Protease (PLpro) and 3C-like Serine Protease (3CLpro) detection 135
4.3.6.1. Detection of Cysteine Protease (PLpro) activity 135
4.3.6.2. Detection of 3C-like Serine Protease (3CLpro) activity 137
4.3.7. Graphene-based biosensors for measuring the activity of Cysteine Protease (PLpro) and 3C-like Serine Protease (3CLpro) 138
4.3.7.1. Serum Sample Assays – Graphene nanobiosensor protease activity measurements procedure 139
4.4. Conclusion 142
4.5. Outlook: Verification the concept of SARS-CoV-2 protease activity monitoring in a COVID-19 mouse model 144
4.5.1. Calibration and validation of the optical nanobiosensors for Papain-like Cysteine Protease (PLpro) and 3C-like Serine Protease (3CLpro) 144
4.5.1.1. Calibration of optical nanobiosensors 144
4.5.1.2. In-vitro approach 145
4.5.2. Assessment of sensor performance in pre-clinical studies 145
4.5.2.1. Approach 146
4.5.2.2. Expected results and potential alternatives 147
4.5.3. Broader impacts 148
4.6. References 149
Chapter 5 - Summary and Future Outlook 153
5.1. References 156
Appendix_ Chapter 2 (S) 157
Appendix_ Chapter 3 (A) 175
UPLC-MS Characterization of consensus peptide sequences 175
Enzymatic digestion results 183
HEPES vs MES buffer– 19 nanobiosensors and control serum 188
TEM images of all nanobiosensors 196

List of Figures

Figure 1.1. Estimated number and percentage of new cases in both genders and all ages in 2022.[10] 2
Figure 1.2. Schematic of microcantilever biosensor. (a) Microcantilever with non-specific reference peptide (Ref-1or Ref-2) showing no deflection to the presence of either normal or cancer cells. (b) Microcantilever functionalized with cancer-targeting peptides showing strong deflection to cancerous cells.[26] (Reprinted with permission from ref 26)..... 6
Figure 1.3. Real-time detection of cancer cells (a) Detection of MCF7 cells mimicking the circulating tumor cells at 25±5 cells mL–1 using cantilever array functionalized with four different peptides. (b) Capture yield of cancer cells corresponding to all four peptide-based sensors.[26] (Reprinted with permission from ref 26)..... 8
Figure 1.4. Mitochondria; structure and compartments.[1] (Reprinted with permission from ref 1). 9
Figure 1.5. T1-weighted axial MRI images obtained before (pre) and at 10, 20, and 30 min after intravenous injection of each contrast agent. Arrows indicate tumors.[41] (Reprinted with permission from ref 41). 13
Figure 1.6. Illustration of ACE2, TMPRSS2, and CD147 targets of SARS-CoV-2.[43] (Reprinted with permission from ref 43). 15
Figure 2.1. (A) UPLC chromatogram of unlabeled MMP2 peptide. The peak at 2.49 min corresponds to the MMP2 peptide product while the peak at 2.91 min corresponds to the peptide with a missing lysine at position 3 in the sequence, (B) Mass spectrum of the peak eluted at 2.49 min corresponding to the mass (M+1 mass peak) of MMP2 peptide, (C) Mass spectrum of the peak eluted at 2.91 min corresponding to the major peptide impurity with a missing lysine at position 3 in the sequence. 33
Figure 2.2. (A) The peak at 2.91 min corresponds to the RhB-labeled peptide while the major peak at 3.86 min corresponds to the peptide with Fmoc-protected MMP2, (B) Mass spectrum of the peak eluted at 2.91 min corresponding to the mass (M+1 mass peak, 1678 m/z) of the RhB-labeled MMP2 peptide, (C) Mass spectrum of the peak eluted at 3.86 min corresponding to the Fmoc-protected MMP2 peptide (m/z=1438). 34
Figure 2.3. (A) UPLC chromatogram of GKK peptide. The peak at 0.78 min corresponds to GKK peptide product and the peak at 2.70 min corresponds to the GHH peptide attached to the Fmoc group, (B) Mass spectrum of the peak eluted at 0.78 min corresponding to the mass (M+1 mass peak, 332 m/z) of GKK peptide, (C) Mass spectrum of the peak eluted at 2.70 min corresponding to the mass (M+1 mass peak, 554 m/z) of GKK peptide attached to the Fmoc group. 36
Figure 2.4. (A) UPLC chromatogram of GKK-2. The peak at 0.78 min corresponding to GKK peptide product with minor impurities at 2.73 min and 4.02 min corresponding to the Fmoc-protected peptide and KK peptide fragment, respectively. (B) Mass spectrum of GKK-2 corresponding to the GKK at 0.78 min (M+1 mass peak, 332 m/z), (C) Mass spectrum of GKK-2 corresponding to Fmoc-protected GKK at 2.73 min. 41
Figure 2.5. (A) UPLC chromatogram of GKK-3. The peak at 0.77 min corresponds to GKK peptide product with the minor impurity peak at 2.70 min corresponding to the Fmoc-protected peptide, (B) Mass spectrum of GKK-3 corresponding to the major peak at 0.78 min, (C) Mass spectrum of GKK-3 corresponding to the Fmoc-protected GKK-peptide.... 42
Figure 2.6. (A) UPLC chromatogram of GKK-6. The peak at 0.75 min corresponds to GKK peptide product with the impurity peaks at 2.63 min and 3.91 min corresponding to the

Fmoc-protected GKK peptide impurity and KK peptide fragment peak, respectively (B) Mass spectrum of GKK-6 corresponding to the major GKK peptide peak at 0.76 min (C) Mass spectrum of GKK-6 corresponding to the Fmoc-protected peptide impurity peak at 2.62 min. 43
Figure 2.7. (A) UPLC chromatogram of GKK-7. The peak at 0.77 min corresponds to GKK peptide product with the impurity peaks at 2.68 min and 3.99 min corresponding to the Fmoc-protected GKK peptide impurity and KK peptide fragment peak, respectively (B) Mass spectrum of GKK-7 corresponding to the major GKK peptide peak at 0.77 min, (C) Mass spectrum of GKK-7 corresponding to the minor Fmoc-protected peptide peak at 2.71 min. 47
Figure 2.8. (A) UPLC chromatogram of GKK-10. The peak at 0.75 min corresponds to GKK peptide product with the impurity peaks at 2.66 min and 3.91 min corresponding to the Fmoc-protected GKK peptide impurity and KK peptide fragment peak, respectively (B) Mass spectrum of GKK-10 corresponding to the minor GKK peptide peak at 0.75 min, (C) Mass spectrum of GKK-10 corresponding to the Fmoc-protected peptide peak at 2.66 min. 48
Figure 2.9. (A) UPLC chromatogram of unlabeled MMP2 (Trial#1). The peak at 2.58 min corresponds to the unlabeled MMP2 peptide product and the peak at 3.15 min corresponds to the impurity (peptide fragment without a cysteine and a water molecule), (B) Mass spectrum of the peak at 2.57 min corresponds to the mass (1158) of the MMP2 peptide, (C) Mass spectrum of the peak at 3.15 min corresponds to the mass of the impurity (peptide without a cysteine and one less water molecule). 51
Figure 2.10. (A) UPLC chromatogram of MMP2 (Trial# 2). The peak at 2.52 min corresponds to MMP2 peptide product and the peak at 3.02 min corresponds to the impurity (peptide fragment with a missing cysteine and a water molecule), (B) Mass spectrum of the peak at 2.53 min corresponding to the mass of MMP2 peptide product, (C) Mass spectrum of the peak at 3.02 min corresponding to the mass of the impurity (peptide fragment with a missing cysteine and one less water molecule)..... 52
Figure 2.11. (A) UPLC chromatogram of MMP2 (Trial#3). The peak at 2.52 min corresponds to the MMP2 peptide product and the peak at 3.97 min corresponding to the number of peptide fragments, (B) Mass spectrum of the peak at 2.52 min corresponding to the MMP2 peptide product, (C) Mass spectrum of the peak at 3.97 min corresponding to multiple peptide fragments..... 53
Figure 2.12. (A) UPLC chromatogram of GKK-11. The peak at 0.78 min corresponds to GKK peptide product with the impurity peaks at 2.72 min and 4.00 min corresponding to the Fmoc-protected GKK peptide impurity and KK peptide fragment peak, respectively (B) Mass spectrum of GKK-11 corresponding to the major GKK peptide product eluted at 0.78 min, (C) Mass spectrum of GKK-11 eluted at 3.99 min corresponding to the mass of KK peptide fragment. 58
Figure 2.13. (A) UPLC chromatogram of GKK-12. The peak at 0.76 min corresponds to GKK peptide product with the impurity peak at 3.92 min corresponding to the KK peptide fragment, (B) Mass spectrum of GKK-12 eluted at 0.76 min, corresponding to the GKK peptide, (C) Mass spectrum of GKK-12 corresponding to the KK peptide fragment eluted at 3.92 min. 59
Figure 2.14. (A) UPLC chromatogram of GKK-16. The peak at 0.76 min corresponds to the GKK peptide product and the peak at 3.93 min corresponds to the KK peptide fragment

impurity, (B) Mass spectrum of GKK-16 corresponding to the GKK peptide eluted at 0.75 min, (C) Mass spectrum of GKK-16 corresponding to the KK peptide fragment eluted at 3.92 min. 60
Figure 2.15. (A) UPLC chromatogram of GKK-19. The peak at 0.75 min corresponds to GKK peptide product and the peak at 4.01 min corresponds to the KK peptide fragment impurity, (B) Mass spectrum of GKK-19 corresponding to the GKK peptide product eluted at 0.76 min, (C) Mass spectrum of GKK-19 corresponding to the KK peptide fragment eluted at 4.02 min. 61
Figure 2.16. (A) UPLC chromatogram of GKK-20. The peak at 0.75 min corresponds to the GKK peptide product with the impurity peaks at 2.68 min, 3.91 min and 6.73 min corresponding to the Fmoc-protected GKK peptide, KK peptide and a free K amino acid, respectively (B) Mass spectrum of GKK-20 corresponding to the GKK peptide eluted at 0.75 min, (C) Mass spectrum of GKK-20 corresponding to the minor peak at 6.73 min (M+1 mass peak for free lysine). 63
Figure 2.17. (A) UPLC chromatogram of GKK-21. The peak at 0.75 min corresponds to the GKK peptide product with multiple impurity peaks. The peaks eluted at 2.50, 3.89 min and 6.64 min correspond to the Fmoc-protected GKK peptide, KK peptide and a free K amino acid, respectively (B) Mass spectrum of GKK-21 corresponding to the GKK peptide eluted at 0.75 min, (C) Mass spectrum of GKK-21 corresponding to the Fmoc-protected GKK peptide peak eluted at 2.49 min. 64
Figure 2.18. (A) UPLC chromatogram of GKK-23. The peak at 0.76 min corresponds to the GKK peptide product with multiple impurity peaks. The peaks eluted at 3.93 min and 6.53 correspond to the KK peptide and a free K amino acid, respectively (B) Mass spectrum of GKK-23 corresponding to the GKK peptide peak at 0.76 min, (C) Mass spectrum of GKK-23 corresponding to the impurity peak at 6.53 min (M+1 mass peak for free lysine). 68
Figure 2.19. (A) UPLC chromatogram of GKK-26. The peak at 0.77 min corresponds to the GKK peptide product with multiple impurities. The peak eluted at 3.88 min corresponds to the KK peptide, (B) Mass spectrum of GKK-26 corresponding to the GKK peptide eluted at 0.77 min, (C) Mass spectrum of GKK-26 corresponding to the KK fragment peak at 3.89 min. 69
Figure 2.20. (A) UPLC chromatogram of MMP2 (Trial#1). The peak at 2.68 min corresponds to MMP2 peptide product and the peak at 3.19 min corresponds to the impurity (peptide sequence without a cysteine and a water molecule), (B) Mass spectrum of MMP2 corresponding to the MMP2 peptide peak eluted at 2.68 min, (C) Mass spectrum of MMP2 corresponding to the impurity peak (peptide fragment without a cysteine and a water molecule, 1040 m/z) eluted at 3.19 min. 71
Figure 2.21. (A) UPLC chromatogram of MMP2 (Trial#2). The major peak at 2.62 min corresponds to the MMP2 peptide product and the peak at 2.72 min corresponds to the peptide sequence without a cysteine, (B) Mass spectrum of MMP2 corresponding to the M+1 mass peak of MMP2 peptide (1159 m/z) eluted at 2.72 min, (C) Mass spectrum of MMP2 corresponding to the M+1 mass of the impurity peak (peptide without a cysteine, 1056) at 2.62 min. 72
Figure 2.22. (A) UPLC chromatogram of MMP2 (Trial#3). The major peak at 2.38 min corresponds to the MMP2 peptide product with minor impurities, (B) Mass spectrum of MMP2 corresponding to the M+1 mass peak of MMP2 peptide (1170 m/z) eluted at 2.38 min. 73

Figure 3.1. Design principles of a nanobiosensor for protease detection (A) Iron/iron oxide-peptide detection system[17] (Reprinted with permission from ref 17) (B) Graphene-peptide detection system. 83
Figure 3.2. Raman spectra of graphite and graphene nanoparticles.[19] (Reprinted with permission from ref 19). 84
Figure 3.3. UV/Vis spectroscopic analysis of graphene and Fenton-oxidized graphene nanoparticles.[20] (Reprinted with permission from ref 20) 85
Figure 3.4. Representation of carboxygraphene-polyethylenimine-based platform. 90
Figure 3.5. Gene expression differences (displayed as natural logarithm on y-axis) between control tissue (non-cancerous lung tissue cells) and cancerous tissue (NSCLC, on x-axis).[23] The y-axis is normalized to natural logarithm of mRNA expression for the respective enzyme found in tumor tissues before surgery. The error bars indicate the deviation of mRNA's for the proteases found in tissue. 92
Figure 3.6. Gene expression differences (displayed as natural logarithm on y-axis) between control tissue (non-cancerous lung tissue cells) and cancerous tissue (SCLC, on x-axis).[23] The y-axis is normalized to natural logarithm of comparative mRNA expression for the respective enzyme found in tumor tissues before surgery. CTSL1 and CTSL2 are isoforms of cathepsin L. Gene splicing is very common in the tumor environment.[26] The error bars indicate the deviation of mRNA's for the proteases found in tissue. 93
Figure 3.7. Gene expression differences (displayed as natural logarithm on y-axis) between Small Cell Lung Cancer (SCLC) and Non-small Cell Lung Cancer (NSCLC) tissue (on x-axis).[23] The error bars indicate the deviation of mRNA's for the proteases found in tissue. 94
Figure 3.8. Neutrophil elastase (unlabeled) characterization via UPLC-MS. Chromatogram (green) shows the relevant peaks when mass filter was applied. 96
Figure 3.9. Neutrophil elastase (TCPP-labeled) characterization via UPLC-MS. Chromatogram (green) shows the relevant peaks when mass filter was applied. 96
Figure 3.10. Neutrophil elastase consensus sequence digestion with commercial Neutrophil elastase protease, created using UPLC-MS results. (A) Trend-line for peptide fragment product (B) Trend-line for digested peptide substrate. 97
Figure 3.11. Thermogravimetric Analysis of Carboxygraphene and Carboxygraphene-PEI.[20] (Reprinted with permission from ref 20). 100
Figure 3.12. Fourier-transform infrared spectroscopy (FTIR) analysis of carboxygraphene.[20] (Reprinted with permission from ref 20). 101
Figure 3.13 Ultraviolet/visible optical extinction spectra of carboxygraphene (Black), carboxygraphene-PEI (Grey), and the MMP1 nanobiosensor (Red, for comparison) in water. The absorption peak of TCPP is marked with a red arrow.[20] (Reprinted with permission from ref 20). 104
Figure 3.14. Transmission electron microscopy images of A) Graphene, B) Carboxygraphene, C) Carboxygraphene-PEI, and D) MMP1 nanobiosensor.[20] (Reprinted with permission from ref 20). 107
Figure 3.15. Measurement of fluorescence intensity for the optimization of peptide-TCPP (MMP1-TCPP) per 100 mg of nanobiosensor.[20] (Reprinted with permission from ref 20). 108
Figure 3.16. Measurement of fluorescence intensity for the optimization of (MMP1) nanobiosensor concentration.[20] (Reprinted with permission from ref 20). 109

Figure 3.17. Measurement of fluorescence intensity in HEPES vs. MES buffer using CTSE nanobiosensor. 110
Figure 3.18. Measurement of fluorescence intensity for the optimization of pH using MMP1 nanobiosensor.[20] (Reprinted with permission from ref 20)..... 111
Figure 3.19. Measurement of fluorescence intensity for the wavelength optimization using MMP1 nanobiosensor (before and after incubation). All sera used for the principal testing of the MMP1 biosensor were provided by Hawkeye Bio LLC.[20] (Reprinted with permission from ref 20). 112
Figure 3.20. Limit of detection estimation indicated by plotting signal vs - log molar enzyme concentration. (Blue: fluorescence before adding MMP1 in various molar concentrations to 0.01 mg/ml optimized MMP1 nanobiosensor in HEPES; Orange: fluorescence after 60 min. incubation at 37 °C).[20] (Reprinted with permission from ref 20). 113
Figure 4.1. The determination of the prevalence of clinical or scientific application of a developed method.[6] (Reprinted with permission from ref 6). 120
Figure 4.2. Genomes and Structures of Coronviruses.[10] (Reprinted with permission from ref 10). 121
Figure 4.3. Fe/Fe_3O_4 core/shell nanoparticle-based nanobiosensor for protease detection. 127
Figure 4.4. UPLC-MS analysis of PLpro1 peptide sequence. 129
Figure 4.5. UPLC-MS analysis of PLpro2 peptide sequence. 130
Figure 4.6. UPLC-MS analysis of 3CLpro peptide sequence. 131
Figure 4.7. Docking Simulation (CABS-dock (20-25)) of GAGVTLQSAESAG to the active center of human coronavirus 229E main peptidase (monomer). The peptide is shown in red. 132
Figure 4.8. UPLC-MS (total mass range detection), signal intensity in percent vs. retention time in minutes. A) Digestion of GAGKAAG-GKVSAG peptide (25 nM) with PLpro enzyme (100 nM) in modified Tris buffer. B) Digestion of GAGKQGA-GDAGAG peptide (25 nM) with PLpro enzyme (100 nM) in modified Tris buffer. C) Digestion of GAGVTLQ-SAESAG peptide (25 nM) with PLpro enzyme (100 nM) in modified Tris buffer. UPLC chromatograms of digestion fractions collected at 0 and 4 hours are shown. 133
Figure 4.9. Intensities of the peptides PLpro-1: GAGKAAGGKVSAG, m/z: 1030.13;..... 134
Figure 4.10. Calibration curve of the nanobiosensor 1 for Cysteine Protease (PLpro1) in HEPES buffer (pH=7.2). Plot of normalized fluorescence intensity (Synergy H1, BioTek) vs. decadic logarithm of PLpro concentration. I90min: fluorescence signal at 650nm (excitation: 420nm) after 90 min of incubation. I0min: fluorescence signal at 650nm immediately after mixing. AC: assay control: fluorescence at 650nm after 90 min. of incubation in the absence of PLpro. 135
Figure 4.11. (Failed) Calibration curve of the nanobiosensor 2 for Cysteine Protease (PLpro2) in HEPES buffer (pH=7.2). Plot of normalized fluorescence intensity (Synergy H1, BioTek) vs. decadic logarithm of PLpro concentration. I90min: fluorescence signal at 650nm (excitation: 420nm) after 90 min. of incubation. I0min: fluorescence signal at 650nm immediately after mixing. AC: assay control: fluorescence at 650nm after 90 min. of incubation in the absence of PLpro. 136
Figure 4.12. Calibration curve of the nanobiosensor for 3C-like Serine Protease (3CLpro) in HEPES buffer (pH=7.2). Plot of normalized fluorescence intensity (Synergy H1, BioTek) vs. decadic logarithm of PLpro concentration. I90min: fluorescence signal at 650nm (excitation: 420nm) after 90 min. of incubation. I0min: fluorescence signal at 650nm

immediately after mixing. AC: assay control: fluorescence at 650nm after 90 min. of incubation in the absence of PLpro. 137
Figure 4.13. Protease activity in serum samples. 140
Figure 5.1. Mitochondria[5] (regenerated with permission from ref 5). 155

List of Tables

Table 2.1. List of resins used for the peptides. 25
Table 2.2. Summary of the synthesis of short (3 amino acids) peptides. 26
Table 2.3. Peak purities of the 3-amino acid peptide sequences. 36
Table 2.4. Summary of GKK synthesis to optimize the concentrations of diethylamine at room temperature. 39
Table 2.5. Peak purities of GKK-2, GKK-3, GKK-4, GKK-5 and GKK-6..... 44
Table 2.6. Summary of GKK synthesis to optimize the concentrations of diethylamine and time of incubation at higher temperature. 45
Table 2.7. Peak purities of GKK-7, GKK-8, GKK-9 and GKK-10..... 48
Table 2.8. Summary of MMP2 synthesis (with diethylamine)..... 49
Table 2.9. Summary of GKK synthesis to optimize the concentration of piperidine and time of incubation at room temperature. 57
Table 2.10. The peak purities of the GKK peptides (GKK-11 to GKK-19) synthesis to optimize the concentration of piperidine and time of incubation at room temperature. 61
Table 2.11. Summary of GKK synthesis to optimize the concentration of piperidine and time of incubation at higher temperature. 63
Table 2.12. Peak purities of GKK synthesized (GKK-20 and GKK-21) using piperidine at higher temperature. 64
Table 2.13. Summary of GKK synthesis using piperidine with multiple incubations (more than two). 66
Table 2.14. Summary of GKK synthesis using diethylamine with multiple incubations (more than two). 66
Table 2.15. Peak purities of GKK synthesized (GKK-22 to GKK-27) using piperidine and diethylamine with multiple incubations (more than two). 69
Table 2.16. Summary of MMP2 synthesis (with piperidine)..... 71
Table 3.1. Consensus peptide sequences for lung cancer proteases. 94
Table 3.2. CHNO elemental analysis results.[20]..... 99
Table 3.3. Dynamic light scattering and Zeta Potential measurements of carboxygraphene and carboxygraphene-PEI..... 102
Table 3.4. Dynamic light scattering and Zeta Potential measurements of all 19 nanobiosensors. 102
Table 3.5. Concentration of graphene-attached TCPP for each nanobiosensor. 105
Table 4.1. Consensus sequences for PLpro and 3CLpro detection..... 128
Table 4.2. Significance Table – Protease Activity differences..... 141

List of Schemes

Scheme 2.1. Flow-chart diagram showing Merrifield-SPPS synthesis steps (regenerated with permission from ref 16) with an additional step of using 4% acetic anhydride. 24
Scheme 2.2. Scheme for the 9-fluorenylmethoxycarbonyl deprotection in solid phase peptide synthesis.[18] (Regenerated with permission from ref 18). 30
Scheme 2.3. Side-chain acetylation of lysine with acetic anhydride.[21] (Regenerated with permission from ref 21). 32
Scheme 2.4. The Fmoc-group removal with piperidine (in DMF).[18] (Regenerated with permission from ref 18). 54
Scheme 3.1. Activation and addition of dye to the peptide and resin cleaving. 88
Scheme 3.2. Synthesis of carboxygraphene. 99

Acknowledgements

Foremost, I thank Allah for blessing me so generously! I would like to express my gratitude to my advisor, Dr. Stefan H. Bossmann, for all his advice, inspiration, support, and assistance over the past four years. Thank you for allowing me to see what the research world was like as your graduate student. There are not enough words to adequately convey my gratitude for having Dr. Bossmann as the best mentor and role model. Thank you so much for always being there whenever I needed help with anything, whether academic or non-academic. I would not have accomplished so much if I did not have his support and encouraging words at different steps during my Ph.D.

Additionally, I want to thank Katrin Bossmann for all her support throughout the years. Truly the best person to talk to whenever I had anything bothering me. Thank you for always helping with all the challenges I faced and for your emotional support.

I want to thank my co-major professor, Dr. Cristine Aikens, for all the support and advice she gave at different steps during this process. She is one of the nicest people, and an inspiring female scientist role model for me.

I would like to extend my gratitude to my committee members, Dr. Daniel A. Higgins, and Dr. Mathew Basel, for their time and valued suggestions.

I also want to thank Dr. Chris Culbertson and Dr. Shu Jia for allowing me to work in their research lab. It was a great and fun learning experience which has taught me a lot about analytical chemistry instrumentation and techniques.

I would like to thank Dr. Anne Culbertson for giving me the opportunity to be her GTA. I enjoyed every single moment working with her and learning teaching skills. She has been a great support and a counselor for me. One of the best and kindest people I ever got to work with! I am

very grateful for all the suggestions, support, and our time-to-time conversations about how things are going.

I would like to thank Dr. Tej Shrestha and Marla Pyle for the trainings to be able to work in Veterinary Medical at Kansas State University.

Additionally, I would like to thank the Chemistry department at Kansas State University and all the faculty and staff of the Chemistry department for their help at various steps during this journey.

I want to thank my family, especially my parents, Mr. Ehsan Ahmed, and Mrs. Riffat Parveen, who allowed me to come across the continents to pursue higher education. Thank you for ignoring financial and social pressure, and instead let all your children attain higher education which set a role model for the whole community in town. I would like to thank my siblings, Mr. Muhammad Shoaib, Mr. Muhammad Zohaib, Mr. Muhammad Tayyab, Mr. Muhammad Abu Hurairah, Mr. Muhammad Abu Bakar, and my sisters Dr. Iqra Parveen and Maryam Rani. I miss you all! It was a hard decision for me to pursue my education and achieve my goals while living so far away from you all. Nothing in this world can banish my regret that I missed your childhood days and could not watch you grow and progress in your life. A huge thanks to my big brother Mr. Muhammad Zohaib who has fulfilled his responsibilities not only as a son and a brother, but also as a best friend. He has done so much to support the whole family, and I will always be grateful for that.

I would also like to thank my fiancé, Mr. Muhammad Ali Raza, for being such a fantastic friend and emotionally supporting to me. I am fortunate to have him as a friend and a partner. Thank you for your unconditional love and support in completing this journey. There were so

many days when I wanted to quit, but he always encouraged and convinced me, saying that he knew I could do it, and I did!

I want to thank my advisors at Western Illinois University especially Dr. Jin Jin and Dr. TK Vinod who always helped me whenever needed and encouraged my decision of experiencing research deeply by choosing to pursue a PhD.

I am also thankful to all my teachers from my middle and high schools and the University of Sargodha, especially Mr. Aslam Nasir and Mr. Muhammad Riaz, who saw my potential and always encouraged me.

I am particularly thankful to my uncle Mr. Masood Ahmed (late), and my grandfather Mr. Muhammad Amin (late), who were the source of inspiration for me, and they always encouraged me to do what I wanted to do with my life. I am also very thankful to all the other members of my family whom I lost during my time in the US, including my beloved aunt Mrs. Humaira Shiraz, my grandfather Mr. Muhamamd Allah Ditta and my grandmother Amina Bibi. I miss you all!

Finally, I would like to thank all the members of the Bossmann group who made this an unforgettable experience for me, including Dr. Obdulia Covarrubias, Dr. Jose Covarrubias, Dr. Kayla Eschliman, Dr. Macy Payne, Ivina Mali, and Raul Neri and all other members of the group whom I have interacted with. A special thanks to Obdulia Covarrubias for always being there whenever I needed anything and teaching me so much.

I am also thankful to all the friends I made during my time at Kansa State University, especially Obdulia Covarrubias, Farhat Khurshid and Ivina Mali with whom I always had the best time.

Dedication

I am very happy to dedicate this effort to my beloved family, especially my parents Mr. Ehsan Ahmed and Mrs. Riffat Parveen. I was never going to be able to accomplish this without their support. They never cared for their own needs and wishes but they worked hard in life to give a better future to their children. I would also like to dedicate this work to my uncle Masood Ahmed (late).

Chapter 1 - Potential of Synthetic Peptides in Diagnosis and Therapy of Cancer and COVID-19

1.1. Cancer: a continuous battle[1]

Cancer is a broad term referring to more than 100 distinctive diseases associated with cells' abnormal and uncontrolled growth.[1,2] This uncontrolled growth leads to the formation of masses of cells called tumors. Tumors are classified as benign tumors, premalignant tumors, and malignant tumors. Benign tumors are noncancerous as they stay at the primary site of the origin and do not invade other tissues or organs. Premalignant tumors have the potential to be malignant tumors while staying noncancerous. In contrast to benign and premalignant tumors, malignant tumors are cancerous and spread to distant locations via metastasis.[2,3] Benign tumors can be removed entirely by surgical procedures and do not grow back, while malignant tumors may grow back. Premalignant tumors may require careful monitoring and need to be surgically removed or treated if they lead to the formation of "dysplasia", masses in which cells reproduce faster than normal, or "carcinoma in situ", in which cells are highly abnormal to a point where they can potentially start invading other cells and is also referred as "0-stage cancer".[4]

Cancer can arise anywhere in the body and is mostly named after the region of origin. Most cancer types show symptoms at earlier stages, but not all the cancer types show obvious symptoms until stage 3 or 4. The lack of symptoms at earlier stages makes it harder to properly treat cancer which, ultimately, reduces the chances of survival. Such cancers are named "silent

[1] Ehsan, S.; Covarrubias-Zambrano, O.; Bossmann, S. H. Mitochondrial Targeting Peptide-Based Nanodelivery for Cancer Treatment. *Curr. Protein Pept. Sci.* **2022**. https://doi.org/10.2174/1389203723666220520160435.

killers". Examples include ovarian cancer, colorectal cancer, cervical cancer, lung cancer, and breast cancer (the most common cancer).[5,6]

This destructive disease has the second-highest mortality rate worldwide, after heart diseases being the primary cause of deaths.[7] According to statistics from the American Cancer Society, there are expected 1,918,030 new cancer cases and almost 609,360 cancer deaths in the year 2022 in the US alone (approximately four new cases and one death per minute).[8] This indicates the severity of this evil disease which is not only impacting human lives but also increasing the socioeconomic pressure worldwide. The cost of managing the cancer treatment or the individual's inability to work are significant outcomes that lead to poverty. It is particularly true for third-world countries that lack proper health care systems and cancer care centers.[9]

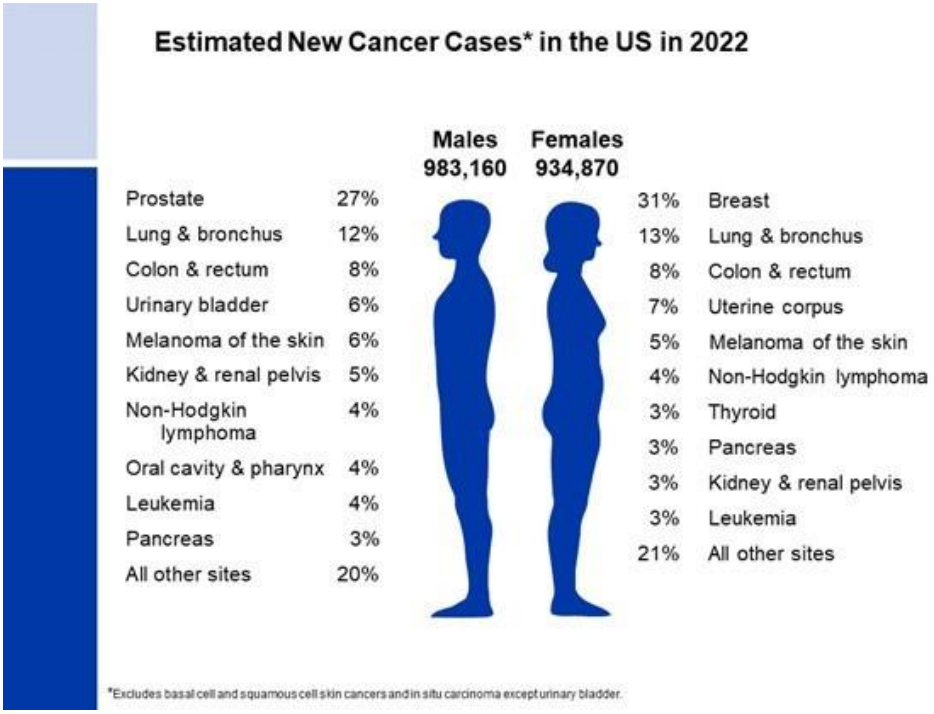


Figure 1.1. Estimated number and percentage of new cases in both genders and all ages in 2022.[10]

Sadly, anyone can be a victim of this deadly disease without any discrimination! Thanks to cancer research, the number of survivors has increased over the last decade, giving hope to the people suffering from this disease.[11]

Several factors are known to trigger cancer, and the outcomes may vary from person to person. A well-known cause of cancer is gene mutations. There are two types of these mutations: acquired mutations and germline mutations. Acquired mutations occur due to damage to genes of a particular cell at any point during a person's life. Factors that can trigger an acquired gene mutation include ultraviolet radiation, tobacco use, viral/bacterial infections, and age. Germline mutations come from a mutation in the egg cell or the sperm cell and can pass from one generation to the next generation. These types of mutations lead to cancer that is passed from generation to generation and is named "inherited cancer".[12] Genetic mutations occur in the genes that control the cell proliferation mechanisms by either over-activating the growth-promoting genes or suppressing the tumor-suppressor genes.[13] Errors in DNA repair genes can be acquired or inherited as the genetic mistakes may remain uncorrected and transferred to the next generation.[12]

Depending on the type and stage of cancer, a patient may undergo different treatments. The most common cancer treatments include surgical removal of tumors and cancers, chemotherapy, and radiation therapy. With recent advances in cancer research, some new technologies have emerged as means of more effective cancer treatment. These include hormone therapy which uses hormones to control the cellular behaviors, immunotherapy that uses a person's self-defense system to target and kill cancerous cells, bone marrow or stem-cell transplant in which defective cells are replaced by bone marrow/stem cells that eventually grow into normal cells, and targeted therapy in which tumor or cancer cells are specifically targeted

with minimal or no impact on the normal cells. Drug-based treatments (such as chemotherapy, immunotherapy, and hormone therapy) are often named systemic treatments as they usually affect the whole body.[14] Radiation-based therapies are useful for small and localized tumors but may also affect the growth of normal cells or sometimes induce them to generate secondary tumors. Surgical procedures for cancer removal and bone marrow transplant are invasive and expensive procedures with common side effects of developing infections and loss of blood, resulting in weakened immune systems.[14,15] Thanks to the scientific research going on in cancer diagnostics and treatment, the number of cancer survivors has increased over the past few decades. According to statistics, there were 16.9 (5% population) million cancer survivors in the US, which are expected to rise by 22.2 million (31.4% population) survivors by 2030.[16] Early detection of tumors and better ways of cancer therapies are two significant areas of concern to increase the cancer survival rates.

1.2. History and importance of peptides in cancer research

Most conventional cancer therapies suffer from adverse side effects, off-targeting, drug resistance, and high costs. An example is the most employed chemotherapeutic drug for cancer treatment, "doxorubicin", but its use is limited by potential oxidative stress-mediated injury to the heart, kidneys, and brain.[17] In comparison to the side effects associated with the use of toxic chemotherapeutics, the use of peptides is a very safe, efficient, and biocompatible alternative.[18] Peptides are linear chains of amino acids having 500-5000 Da molecular weight and lie in between small molecules and big-sized proteins but differ therapeutically and biochemically.[19] Because of the ability of peptides to participate in intrinsic signaling pathways for many physiological activities, they provide an opportunity for therapeutic interventions that mimic the natural pathways.[17,18] These peptide drugs are also called "replacement therapies" because they

supplement the peptide hormones when endogenous levels are low or absent. The use of insulin in diabetics, which started in the 1920s, is an example of the first therapeutic use of peptides as a drug when the patient could not produce enough insulin. Not only insulin, but adrenocorticotropic hormone (ACTH) was also among the life-saving peptide-therapeutics in the early 1900s. Later after the discovery of methods to synthesize the peptides artificially, synthetic vasopressin and oxytocin were also used clinically. Afterward, new peptides with therapeutic potential were identified and isolated using natural sources such as venoms of cephalopods. With the development in science and technology, genomic sequencing made it possible to identify and characterize the receptors for many endogenous peptide hormones to evaluate the possible use of peptide-based ligand for these peptide/protein receptors.[20] Hence, the potential of peptides to be used in therapeutics encouraged academia and industry to begin to pursue novel research in this field. To date, there are many peptide-based drugs approved by Food and Drug Administration (FDA) for different medical conditions, and many are under clinical trials. There are 23 peptides and oligonucleotide therapeutics among a total of 228 drugs that the FDA approved from 2016 to 2020. In just the year 2020, the FDA approved 53 drugs out of which six were peptides and oligonucleotide therapeutics.[21]

1.3. Biomedical applications of peptides in cancer diagnostics and therapy

1.3.1. Peptides in the detection of cancer

Early detection of cancer is crucial because it increases the likelihood of successful treatment and lowers the death rates without the need for harsh treatments.[22] The traditional diagnostic methods for disease detection include endoscopy, tissue biopsies, X-ray scans, magnetic resonance imaging, computed tomography, and positron emission tomography.[23,24] However, these diagnostics are not practical for repeated screenings to evaluate the treatment

progress and are not accessible to a large population.[23] There is a critical need for improved clinical diagnostic tools and procedures that are sensitive, reliable, rapid, and cost-effective.[25] Recently, Etayash and coworkers developed a real-time detection system to detect circulating tumor cells in the event of breast cancer from human blood samples.[26]

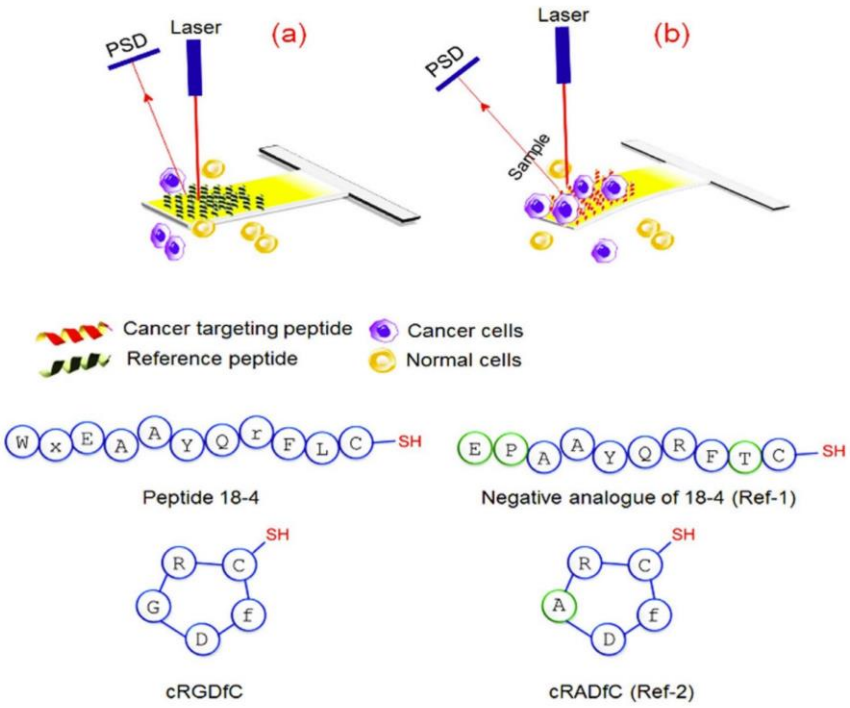


Figure 1.2. Schematic of microcantilever biosensor. (a) Microcantilever with non-specific reference peptide (Ref-1or Ref-2) showing no deflection to the presence of either normal or cancer cells. (b) Microcantilever functionalized with cancer-targeting peptides showing strong deflection to cancerous cells.[26] (Reprinted with permission from ref 26).

The group used microcantilever arrays functionalized with breast cancer-specific peptide sequence (WxEAAYQrFL) and tumor homing sequence RGD peptide (cRGDfC) to detect the circulating tumor cells in the bloodstream at the early stages of cancer, as clinical and molecular findings suggest the blood invasion of cancer cells may occur during the early stages. Circulating tumor cells can potentially give early prognostic characteristics and indications of a specific tumor.[27] Hence, the authors developed a cantilever decorated with the peptides specific for breast

cancer cells (MDA-MB-231 and MCF7) using an additional cysteine in the peptide sequence to make a thiol-gold linkage between the peptides and gold interfaces.

The interaction between the immobilized peptides and the cancer cells (adsorption through analyte-ligand interaction) on the cantilever surface caused a reduction in the surface free energy, which in turn caused a differential surface stress between the functionalized and non-functionalized sides of the lever. As a result of the differential surface stress, the cantilever beam showed deflection or bending to a certain degree which is indicative of cancer cell capture by the designed system. The surface stress generated was expressed using Stoney's formula[28], as shown below:

$$\Delta s = \left[\frac{Et^2\delta}{3L^2(1-v)}\right]$$

Where Δs *is* differential surface stress, δ is the deflection of the cantilever beam, E is the substrate's Young's modulus, t is the cantilever's thickness, L is cantilever's length, and v is Poisson's ratio (deformation in directions perpendicular to the direction of loading).[28]

The system was able to give real-time detection of cancer cells with a detection limit of 50–100 cancer cells mL^{-1}. The cancer cells capture yield was calculated to be 80% from the spiked whole blood samples with a flow rate of 1 mL per hour.[26]

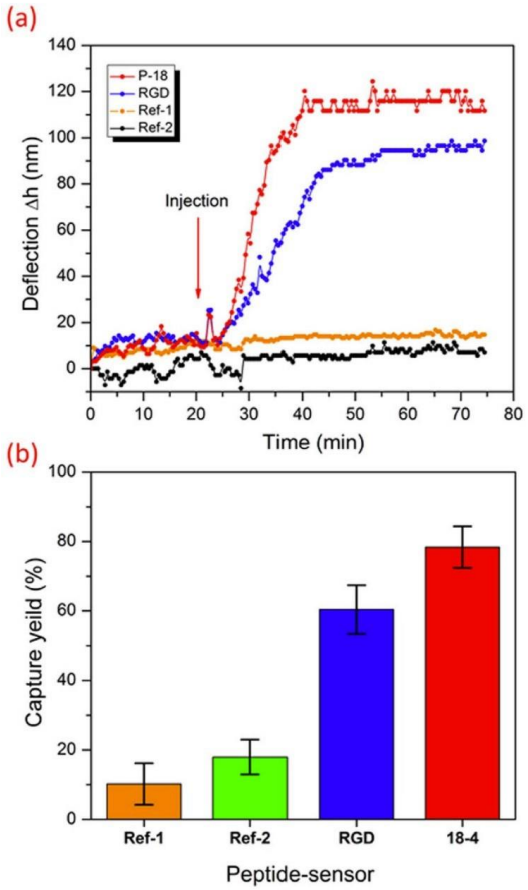


Figure 1.3. Real-time detection of cancer cells (a) Detection of MCF7 cells mimicking the circulating tumor cells at 25±5 cells mL–1 using cantilever array functionalized with four different peptides. (b) Capture yield of cancer cells corresponding to all four peptide-based sensors.[26] (Reprinted with permission from ref 26).

This efficient method did not require any labeling or sample preparation and showed a potential to not only offer the easy detection of cancer but can also be used to find recognition elements/biomarkers to develop peptide-based nanobiosensors for detection and targeted drug delivery platforms.[26] Scientists have used several other approaches to develop peptide-based detection systems to detect different biomarkers and proteases such as using nanoparticle conjugation, fluorescent tags, microfluidic devices, fluorescence resonance energy transfer (FRET) technique, protein chips or microarrays using immobilized antibodies, etc. However,

there is still a need for an efficient system with the potential to detect cancer at early stages with ease and accuracy in clinical settings.

1.3.2. Peptides in the subcellular-targeted drug delivery[1]

Targeted drug delivery to the organelles in the tumor cells eliminates the side effects and risks associated with nuclear targeting, such as cardiotoxicity induced by nuclear targeting of doxorubicin.[17,29] In this regard, the importance of mitochondria in cancer physiology makes them a universal pharmacological target for almost all cancer types. Mitochondria consist of dual membranes with a negative potential and are involved in cellular respiration and programmed cell death processes.[30]

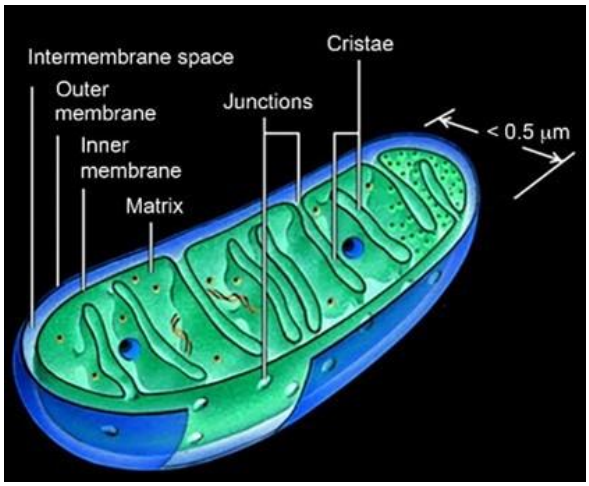


Figure 1.4. Mitochondria; structure and compartments.[1] (Reprinted with permission from ref 1).

Scientists have developed certain chemical moieties to target mitochondria, for example, fluorescent dyes (cyanine, rhodamine, pyridinium, boron-dipyrromethane, etc.) and lipophilic cationic species (triethylammonium, triphenylphosphine or TPP, phenylsulfonylfuroxan, dequalinium, guanidine, and indolinium). However, such chemical species are known to induce fetal toxicity under physiological conditions, and synthetic protocols are very complex. Additionally, linkers are often needed for the conjugation of other components of the designed

formulations. The cellular uptake of such chemical species is essentially charge-dependent resulting in the accumulation of these compounds by the normal cells as well.[31,32] Contrary to that, targeting the intracellular organelle (mitochondria) using synthetic peptides is a promising approach because of high biocompatibility and specificity. The use of peptides for mitochondrial targeting started with the use of modified cell-penetrating TAT peptides. The original TAT peptide was modified by adding arginine and cyclohexyl groups to increase the lipophilic character for mitochondrial accumulation.[33] Scientists have divided mitochondrial targeting peptides into three subdivisions: mitochondria penetrating peptides, Szeto-Schiller peptides (SS, a class of cell-permeable peptides with antioxidative activity), and the mitochondrial targeting sequence peptide (MTS).[34]

Mitochondria penetrating peptides are synthetic peptides (consisting of 4-16 amino acids) with the ability to penetrate the cell membrane and accumulate specifically in the mitochondrial matrix. The electrochemical gradient is the driving force of the translocation of these peptides through the cell and mitochondrial membranes. Mitochondria penetrating peptides are formulated to have an overall positive character and hydrophobic amino acid residues to interact with negatively charged lipid membranes.[35]

SS peptides are a class of short amino acid sequences having aromatic and basic amino acid residues (tyrosine, phenylalanine (Phe), dimethyltyrosine (Dmt), arginine, and lysine). The amino acid residues that are basic in nature give an overall positive character to these peptides to permeate the cell membrane and selectively accumulate in the inner mitochondrial membrane. The presence of aromatic amino acids attributes to the antioxidant activity against ROS to stabilize free radicals. SS peptides are known to follow energy-independent uptake mechanism for membrane translocation.[35,36] Among other reported SS peptides, only SS-31 peptide (*D*-

Arginine-Dmt-Lysine-Phe-NH_2) has been evaluated for effectiveness in ischemia-reperfusion and microvascular injuries in patients during clinical trials.[36]

MTS peptides are also short amino acid chains that consist of 20-40 amino acid residues. These peptides act as ligands to the receptors present on the mitochondrial surface.[36,37] The use of such peptides to target the intracellular organelle offers better protein-protein and protein-lipid interactions that are not as effective with chemical species (small drug molecules or nanoparticles alone).[30] To develop mitochondria-targeting moieties, the lipophilicity (log P value, defined as the logarithm of the octanol−water partition) of peptide and its overall charge (Z) are two crucial factors to be considered because of the highly non-polar nature and highly negative membrane potential of the mitochondrial membranes.[30,35] It has been found that the peptides with an overall charge greater than 0 and a log P value greater than −1.7 result in a high mitochondrial accumulation.[29] In a study conducted by Horton and coworkers, a series of peptides was developed for mitochondrial accumulation by combining hydrophobic amino acid phenylalanine with positively charged amino acids (lysine and arginine) and a chemically modified non-natural amino acid cyclohexylalanine. This study revealed that peptide sequences with an overall charge of +3 or above with log P values of −2 accumulate well inside the mitochondria.[38] Hence, any tumor-targeting moieties can be combined with such mitochondrial targeting peptide sequences to selectively deliver cytotoxic payloads and imaging agents to the mitochondria of the tumor cells, hence, limiting the possibility of off-targeting and unwanted toxicity.

1.3.3. Peptides in imaging applications

Selective receptor-targeting using peptide moieties is gaining considerable importance in molecular imaging of tumor cells because tumors overexpress certain corresponding peptide

receptors that can be targeted to deliver the imaging probes at tumor sites. Peptides are ideal probes for imaging a tumor site because they provide the benefit of rapid clearance from blood circulation as well as high affinities for their targets, hence removing the chances of induced toxicities and off-targeting.[39] For imaging applications, a targeting peptide is labeled with an imaging moiety such as magnetic nanoparticle or contrast agents for magnetic resonance imaging (MRI), radioisotope for single photon emission computed tomography (SPECT) or positron emission tomography (PET), and fluorescent dyes for optical imaging.[40] Several peptides have been successfully explored for receptor-targeted imaging applications, such as vasoactive intestinal peptide, somatostatin, and Arg-Gly-Asp (RGD) peptide, etc.[39] Recently, Li and coworkers developed a peptide-based imaging probe for the detection of prostate cancer.[41] The group identified four short peptide sequences (GVK, IGK, SGV, and ZD2) via phage display that are specific to overly expressed extradomain B fibronectin (EDB-FN) in certain cancers. High expression of oncofetal fibronectin (including EDB-FN) is directly related to the high rate of metastasis and low chances of survival in patients with prostate, ovarian, breast, neck, and head cancers. The imaging probe was synthesized by conjugating cyanin (Cy) 5.5-labeled peptides with tetraazacyclododecane-1,4,7,10-tetraacetic acid (DOTA) in the solid phase, followed by complexation with $GdCl_3$. This newly synthesized Gd–DOTA showed high in vivo chelation stability when used for MR molecular imaging in BALB/c mice with aggressive PC3 prostate tumor xenografts. High-resolution 2D axial MR images of tumors at T1-weighted spin-echo were taken before (pre) and after intravenous injection at 10, 20, and 30 min at a dose of 0.1 mmol/kg body weight. The results showed very strong signal enhancement at 10 min to up to 30 min post-injection for all the designed contrast agents showing the effectiveness of the contrast agents specific to EDB-FN. Similar results were obtained with T1-weighted axial MRI images

for whole-body analysis. The results were also consistent with fluorescence microscopy, where the fluorescence from Cy 5.5 showed a 1/3 decrease in the intensity at 3 hours compared to the intensity at 15 minutes with successful clearance of contrast agents from kidneys after 24 hours post-injection. Overall, the analysis revealed better tumor-signal enhancement when Gd-DOTA decorated with peptide was used compared to clinical contrast agent Gd(HP-DO3A) alone.[41]

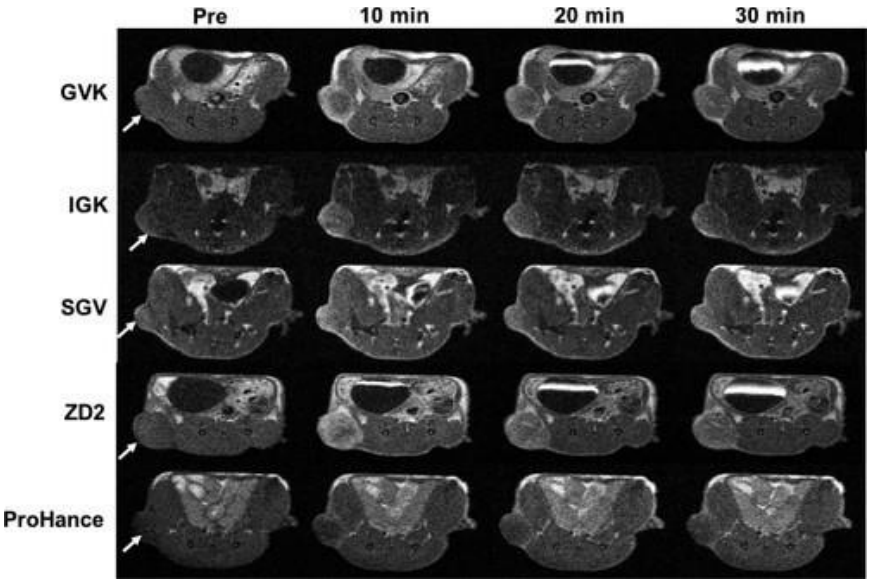


Figure 1.5. T1-weighted axial MRI images obtained before (pre) and at 10, 20, and 30 min after intravenous injection of each contrast agent. Arrows indicate tumors.[41] (Reprinted with permission from ref 41).

1.4. Biomedical applications of peptides in COVID-19 diagnostics and therapy

The COVID-19 pandemic caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has changed the world entirely since it started. Every individual has been affected by this deadly virus one way or the other. We have lost so many precious lives to COVID-19, and the number is still increasing to date, as there are no effective measures to completely avoid and treat the virus. There is a great need for developing detection platforms

that may give accurate results with no false positives, as well as, effective and specific therapy to treat the patients.[42] In this regard, peptides are promising candidates as they can be used to identify new biomarkers to develop efficient, easy-to-use, and cost-effective detection platforms, as well as, they have the potential to selectively deliver the drugs at the targeted site using the ligand-receptor approach. Peptides also can be used as drugs by inducing a response resulting from a receptor-specific interaction with peptide substrates. Studies have revealed that angiotensin-converting enzyme 2 (ACE2) is used as an input receptor by the viral spike SARS-CoV. [43] Hence ACE2-expressing organs, mainly the lungs (83%), are the target of the virus. The heart and the kidneys are known to be affected in case of secondary infections because of the 4-7.5% expression of ACE2 by these organs.[42,43] As the entry point of the virus, ACE2 interaction with the virus is the most critical part of the onset of the infection, hence, targeting this point of interaction could be a possible way to discover a COVID-19-specific drug. Other possible targets are cell serine proteases TMPRSS2 which primes the viral protein spike entailing the cleavage of the spike at two sites: Arg685/Ser686 and Arg815/Ser816, and cathepsin L (CTSL) that activates the virus. CD147 is another attractive target for COVID-19 detection and treatment, as it is an activator of matrix metalloproteinases (MMPs). CD147 increases the expression of MMPs under asthmatic complications, and studies have found enhanced levels of CD147 and MMPs in inflammatory processes and cancer progression.[43]

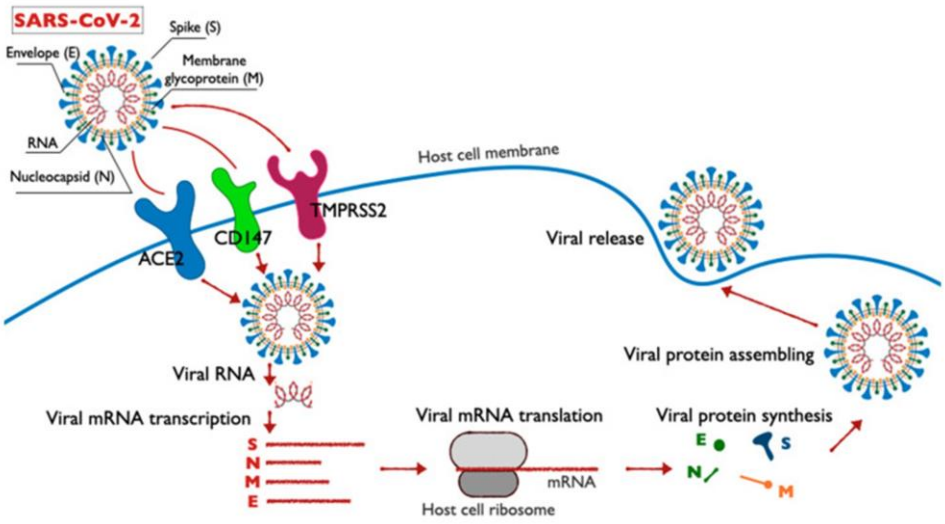


Figure 1.6. Illustration of ACE2, TMPRSS2, and CD147 targets of SARS-CoV-2.[43] (Reprinted with permission from ref 43).

Hence inhibiting CD147 or other similar proteins/enzymes using receptor-binding peptides for COVID-19 therapy or measuring the increased expression levels of CD147, MMPs, or other related proteins/enzymes using peptides to develop a highly efficient and cost-effective COVID-19 detection platform are two very up-and-coming applications of peptides in COVID-19 diagnostics and therapy.

Among other published peptides, SBP1 (targets SARS-CoV-2-RBD), Spikeplug (targets SARS-CoV-2-RBD, EK1 (targets S protein-HR1), EK1C4 (targets S protein-HR1), 2019-nCoV-HR1P & HR2P (target S protein-HR1), IBP01 (targets S protein-HR1) and P9R (targets viral binding and endosomal acidification) are in preclinical trails while CoVac-1 which is a peptide-vaccine composed of a combination of viral SARS-CoV-2 epitopes is in phase II clinical trials.[42,44]

1.5. Conclusion

Cancer and COVID-19 are two ongoing battles that have led to the loss of many precious lives. There is still a lack of efficient early detection platforms and completely safe treatments.

There is a particular need for developing detection systems that have the potential to detect cancer and COVID-19-like diseases at the early onset so that the treatment could be possible to save the life of at-risk patients. The use of peptides is getting a lot of attention from scientists because of the ease of synthesis, flexibility in structure and function, cost-effectiveness, high specificity and selectivity towards a receptor, and high biocompatibility. The use of peptides in developing the detection systems and imaging applications is of great importance because of rapid clearance from the body and high specificity if used as an immobilized receptor for detecting the biomarkers from blood or serum samples. There are certain aspects that limit the use of peptides for such biological applications, such as peptides are prone to proteolytic digestion and offer low bioavailability, have a short half-life, and suffer from production and manufacturing challenges. There is a need for developing methods to synthesize ultra-pure peptides with increased stability to further extend the scope of their biomedical applications, especially in disease diagnosis. Developing efficient detection systems with the potential to detect deadly diseases at early stages should have some characteristics which are superior to the traditional diagnostic tools, such as non-invasiveness, low cost, no need for expensive instruments and extensive sample preparation protocols, and shorter processing time. Hence, developing such technologies is crucial and needs to be the priority of the scientific community.

1.6. References

(1) Ehsan, S.; Covarrubias-Zambrano, O.; Bossmann, S. H. Mitochondrial Targeting Peptide-Based Nanodelivery for Cancer Treatment. *Curr Protein Pept Sci* **2022**. https://doi.org/10.2174/1389203723666220520160435.

(2) *Medical Definition of Cancer*. MedicineNet. https://www.medicinenet.com/cancer/definition.htm (accessed 2021-06-24).

(3) *Tumors: Benign, premalignant, and malignant*. https://www.medicalnewstoday.com/articles/249141 (accessed 2021-06-24).

(4) *The Differences Between Benign and Malignant Tumors*. Healthline. https://www.healthline.com/health/cancer/difference-between-benign-and-malignant-tumors (accessed 2021-06-25).

(5) Silent Killers - Why Cancers Are the Most Deadly. *ISET® - Isolation by Size of Tumor cells*, 2014.

(6) *Cancer today*. http://gco.iarc.fr/today/home (accessed 2021-06-27).

(7) CDCBreastCancer. *An Update on Cancer Deaths in the United States*. Centers for Disease Control and Prevention. https://www.cdc.gov/cancer/dcpc/research/update-on-cancer-deaths/index.htm (accessed 2022-05-23).

(8) *American Cancer Society | Cancer Facts & Statistics*. American Cancer Society | Cancer Facts & Statistics. http://cancerstatisticscenter.cancer.org/ (accessed 2022-05-23).

(9) Kimman, M.; Jan, S.; Kingston, D.; Monaghan, H.; Sokha, E.; Thabrany, H.; Bounxouei, B.; Bhoo-Pathy, N.; Khin, M.; Cristal-Luna, G.; Khuhaprema, T.; Hung, N. C.; Woodward, M. Socioeconomic Impact of Cancer in Member Countries of the Association of Southeast Asian Nations (ASEAN): The ACTION Study Protocol. *Asian Pac J Cancer Prev* **2012**, *13* (2), 421–425. https://doi.org/10.7314/apjcp.2012.13.2.421.

(10) *Cancer Facts & Figures 2022| American Cancer Society*. https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures-2022.html (accessed 2022-05-23).

(11) *Life after cancer: More survivors live longer, face new health challenges*. https://www.usatoday.com/in-depth/news/50-states/2019/02/13/life-after-cancer-survivors-oncology-survivorship-plans-long-term-health/2794121002/ (accessed 2021-06-28).

(12) *The Genetics of Cancer*. Cancer.Net. https://www.cancer.net/navigating-cancer-care/cancer-basics/genetics/genetics-cancer (accessed 2021-06-28).

(13) *Cell Division, Cancer | Learn Science at Scitable*. https://www.nature.com/scitable/topicpage/cell-division-and-cancer-14046590/ (accessed 2021-06-24).

(14) *Types of Cancer Treatment | American Cancer Society*. https://www.cancer.org/treatment/treatments-and-side-effects/treatment-types.html (accessed 2021-06-28).

(15) *Side effects of cancer drugs | Treatment for cancer | Cancer Research UK*. https://www.cancerresearchuk.org/about-cancer/cancer-in-general/treatment/cancer-drugs/side-effects (accessed 2021-06-28).

(16) *Statistics and Graphs | Division of Cancer Control and Population Sciences (DCCPS)*. https://cancercontrol.cancer.gov/ocs/statistics#footnote1 (accessed 2022-05-23).

(17) Marqus, S.; Pirogova, E.; Piva, T. J. Evaluation of the Use of Therapeutic Peptides for Cancer Treatment. *J Biomed Sci* **2017**, *24* (1), 21. https://doi.org/10.1186/s12929-017-0328-x.

(18) Recio, C.; Maione, F.; Iqbal, A. J.; Mascolo, N.; De Feo, V. The Potential Therapeutic Application of Peptides and Peptidomimetics in Cardiovascular Disease. *Front Pharmacol* **2017**, *7*, 526. https://doi.org/10.3389/fphar.2016.00526.

(19) Wang, L.; Wang, N.; Zhang, W.; Cheng, X.; Yan, Z.; Shao, G.; Wang, X.; Wang, R.; Fu, C. Therapeutic Peptides: Current Applications and Future Directions. *Sig Transduct Target Ther* **2022**, *7* (1), 48. https://doi.org/10.1038/s41392-022-00904-4.

(20) Lau, J. L.; Dunn, M. K. Therapeutic Peptides: Historical Perspectives, Current Development Trends, and Future Directions. *Bioorganic & Medicinal Chemistry* **2018**, *26* (10), 2700–2707. https://doi.org/10.1016/j.bmc.2017.06.052.

(21) Al Musaimi, O.; Al Shaer, D.; Albericio, F.; de la Torre, B. G. 2020 FDA TIDES (Peptides and Oligonucleotides) Harvest. *Pharmaceuticals (Basel)* **2021**, *14* (2), 145. https://doi.org/10.3390/ph14020145.

(22) *Exosomes and microvesicles in normal physiology, pathophysiology, and renal diseases*. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6244861/#CR1 (accessed 2021-03-20).

(23) Cheng, N.; Du, D.; Wang, X.; Liu, D.; Xu, W.; Luo, Y.; Lin, Y. Recent Advances in Biosensors for Detecting Cancer-Derived Exosomes. *Trends Biotechnol* **2019**, *37* (11), 1236–1254. https://doi.org/10.1016/j.tibtech.2019.04.008.

(24) Park, S.-M.; Aalipour, A.; Vermesh, O.; Yu, J. H.; Gambhir, S. S. Towards Clinically Translatable in Vivo Nanodiagnostics. *Nat Rev Mater* **2017**, *2* (5). https://doi.org/10.1038/natrevmats.2017.14.

(25) Drain, P. K.; Hyle, E. P.; Noubary, F.; Freedberg, K. A.; Wilson, D.; Bishai, W.; Rodriguez, W.; Bassett, I. V. Evaluating Diagnostic Point-of-Care Tests in Resource-

Limited Settings. *Lancet Infect Dis* **2014**, *14* (3), 239–249. https://doi.org/10.1016/S1473-3099(13)70250-0.

(26) Etayash, H.; Jiang, K.; Azmi, S.; Thundat, T.; Kaur, K. Real-Time Detection of Breast Cancer Cells Using Peptide-Functionalized Microcantilever Arrays. *Sci Rep* **2015**, *5* (1), 13967. https://doi.org/10.1038/srep13967.

(27) Gupta, G. P.; Massagué, J. Cancer Metastasis: Building a Framework. *Cell* **2006**, *127* (4), 679–695. https://doi.org/10.1016/j.cell.2006.11.001.

(28) Azmi, S.; Jiang, K.; Stiles, M.; Thundat, T.; Kaur, K. Detection of Listeria Monocytogenes with Short Peptide Fragments from Class IIa Bacteriocins as Recognition Elements. *ACS Comb. Sci.* **2015**, *17* (3), 156–163. https://doi.org/10.1021/co500079k.

(29) Jean, S. R.; Tulumello, D. V.; Riganti, C.; Liyanage, S. U.; Schimmer, A. D.; Kelley, S. O. Mitochondrial Targeting of Doxorubicin Eliminates Nuclear Effects Associated with Cardiotoxicity. *ACS Chem. Biol.* **2015**, *10* (9), 2007–2015. https://doi.org/10.1021/acschembio.5b00268.

(30) Constance, J. E.; Lim, C. S. Targeting Malignant Mitochondria with Therapeutic Peptides. *Ther Deliv* **2012**, *3* (8), 961–979.

(31) Dong, L.; Neuzil, J. Targeting Mitochondria as an Anticancer Strategy. *Cancer Communications* **2019**, *39* (1), 63. https://doi.org/10.1186/s40880-019-0412-6.

(32) Kang, H. C. Mitochondria-Targeting Theranostics. *Biomaterials Research* **2018**, *22* (1), 34. https://doi.org/10.1186/s40824-018-0145-7.

(33) Horton, K. L.; Stewart, K. M.; Fonseca, S. B.; Guo, Q.; Kelley, S. O. Mitochondria-Penetrating Peptides. *Chemistry & Biology* **2008**, *15* (4), 375–382. https://doi.org/10.1016/j.chembiol.2008.03.015.

(34) Battogtokh, G.; Cho, Y.-Y.; Lee, J. Y.; Lee, H. S.; Kang, H. C. Mitochondrial-Targeting Anticancer Agent Conjugates and Nanocarrier Systems for Cancer Treatment. *Front. Pharmacol.* **2018**, *9*. https://doi.org/10.3389/fphar.2018.00922.

(35) Liew, S. S.; Qin, X.; Zhou, J.; Li, L.; Huang, W.; Yao, S. Q. Smart Design of Nanomaterials for Mitochondria-Targeted Nanotherapeutics. *Angewandte Chemie International Edition* **2021**, *60* (5), 2232–2256. https://doi.org/10.1002/anie.201915826.

(36) Lu, P.; Bruno, B. J.; Rabenau, M.; Lim, C. S. Delivery of Drugs and Macromolecules to the Mitochondria for Cancer Therapy. *Journal of Controlled Release* **2016**, *240*, 38–51. https://doi.org/10.1016/j.jconrel.2015.10.023.

(37) Wang, Z.; Guo, W.; Kuang, X.; Hou, S.; Liu, H. Nanopreparations for Mitochondria Targeting Drug Delivery System: Current Strategies and Future Prospective. *Asian Journal of Pharmaceutical Sciences* **2017**, *12* (6), 498–508. https://doi.org/10.1016/j.ajps.2017.05.006.

(38) Jeena, M. T.; Kim, S.; Jin, S.; Ryu, J.-H. Recent Progress in Mitochondria-Targeted Drug and Drug-Free Agents for Cancer Therapy. *Cancers (Basel)* **2019**, *12* (1), 4. https://doi.org/10.3390/cancers12010004.

(39) Sun, X.; Li, Y.; Liu, T.; Li, Z.; Zhang, X.; Chen, X. Peptide-Based Imaging Agents for Cancer Detection. *Adv Drug Deliv Rev* **2017**, *110–111*, 38–51. https://doi.org/10.1016/j.addr.2016.06.007.

(40) James, M. L.; Gambhir, S. S. A Molecular Imaging Primer: Modalities, Imaging Agents, and Applications. *Physiological Reviews* **2012**, *92* (2), 897–965. https://doi.org/10.1152/physrev.00049.2010.

(41) Li, Y.; Han, Z.; Roelle, S.; DeSanto, A.; Sabatelle, R.; Schur, R.; Lu, Z.-R. Synthesis and Assessment of Peptide Gd–DOTA Conjugates Targeting Extradomain B Fibronectin for Magnetic Resonance Molecular Imaging of Prostate Cancer. *Mol. Pharmaceutics* **2017**, *14* (11), 3906–3915. https://doi.org/10.1021/acs.molpharmaceut.7b00619.

(42) Shah, J. N.; Guo, G.-Q.; Krishnan, A.; Ramesh, M.; Katari, N. K.; Shahbaaz, M.; Abdellattif, M. H.; Singh, S. K.; Dua, K. Peptides-Based Therapeutics: Emerging Potential Therapeutic Agents for COVID-19. *Therapie* **2021**. https://doi.org/10.1016/j.therap.2021.09.007.

(43) Khavinson, V.; Linkova, N.; Dyatlova, A.; Kuznik, B.; Umnov, R. Peptides: Prospects for Use in the Treatment of COVID-19. *Molecules* **2020**, *25* (19), E4389. https://doi.org/10.3390/molecules25194389.

(44) Moroy, G.; Tuffery, P. Peptide-Based Strategies Against SARS-CoV-2 Attack: An Updated In Silico Perspective. *Frontiers in Drug Discovery* **2022**, *2*.

Chapter 2 - Peptide synthesis and optimization[2]

Abstract

Therapeutic peptides are known to have significant pharmaceutical and commercial importance due to their high selectivity, versatility, and safe mode of action. Solid-phase peptide synthesis is the most used and relatively efficient protocol for peptide synthesis. However, it is a multi-step process in which there is a possibility that many undesired reactions will occur. fluorenylmethoxycarbonyl (Fmoc)-based solid-phase peptide chemistry provides a good solution to avoid side reactions but frequently suffers from deprotection issues. This study focuses on the optimization of the Fmoc-based solid-phase peptide synthesis protocol, employing different reagents and reaction conditions. Short peptide fragments were used as an example to determine the effectiveness of each parameter during the optimization. The results were analyzed and confirmed by means of ultra-high performance liquid chromatography-mass spectrometry (UPLC-MS).

2.1. Introduction

Proteins and peptides are essential constituents of living systems and are also known to have significant pharmaceutical applications as drug targets[1] and targeting moieties[2]. Peptide-based therapeutics offer lower production complexity and higher targeting potency than biopharmaceuticals and small molecule drugs.[2] Therapeutic activity of the peptides is attributed to high selectivity, versatility, and relatively safe mode of action with fewer side effects. These peptides have the ability to bind with cell surface receptors and ion channels as ligands to trigger

[2] Reprinted with permission from Optimization of the Fmoc-deprotection in Solid-phase Peptide Synthesis, Chemistry. Sumia Ehsan, Obdulia Covarrubias-Zambrano, and Stefan H. Bossmann, *submitted.*

intracellular effects.[3] There are over 60 peptide-based medicines approved by the US Food and Drug Administration (FDA) to date. The number is projected to increase significantly in the future as almost 150 peptide drugs are currently in clinical trials.[4]

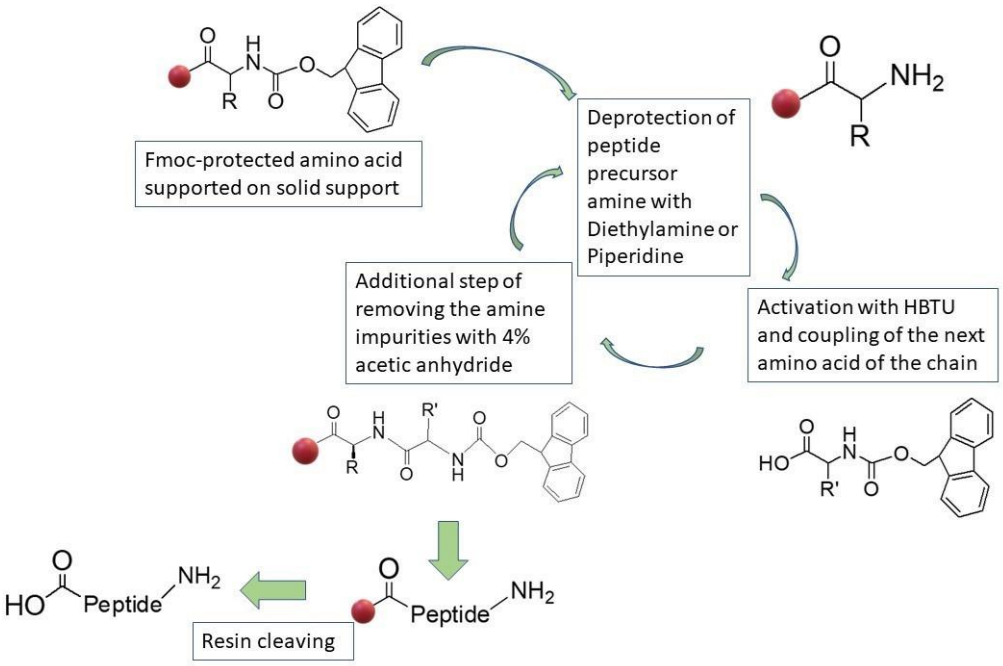
Recent advances in proteomics and genomics have opened new doors to identifying a large number of target peptides and proteins relating to a number of diseases.[1] This evolution resulted in higher demands of the potential therapeutic peptides; hence peptide synthesis has become an essential part of peptide-based therapeutic procedures. Solid-phase peptide synthesis (SPPS) is a suitable synthesis protocol that offers an easy and reliable synthesis of the peptides. Chemically, SPPS follows a condensation reaction between a carboxylic acid functional group and an amine functional group of two α-amino acids that leads to the formation of an amide bond (also called a peptide bond).[5] In the SPPS process, amino acids are sequentially conjugated ($C \rightarrow N$ direction) to a growing chain attached to a solid support or resin. There is usually a cleavable handle between the *C*-terminal of the first amino and the solid support that can be cleaved easily after completing the desired chain.[6]

There are 22 natural amino acids found in polypeptides.[7] 20 are genetically encoded, whereas selenocysteine and pyrrolysine are synthesized via enzymatic modifications of serine[8] and lysine[9]. Excluding glycine, all naturally occurring amino acids are chiral.[7] Consequently, the nanoenvironment differs around each free - N^{α}-amine group once the protection group has been removed. Furthermore, depending on the amino acid sequence, the nascent oligopeptide will develop distinct peptide dynamics while being tethered to the solid support. The underlying paradigm of this study is that the conditions for each SPPS cycle (reagent concentration, reaction temperature, and incubation time) must be optimized to evaluate the effectiveness of conditions in each cycle to get highly pure peptide products.

From the plethora of available peptide synthesis methodologies, fluorenylmethoxycarbonyl (Fmoc) is the most efficient and commonly used α-amine protecting group[10], which requires basic conditions for the removal[11]. In SPPS, the N^{α}-protecting group (Fmoc) can be removed without affecting the sidechain protecting groups to prepare the polypeptide chain for the next coupling cycle. All the chemical reactions are driven to completion by using soluble reagents in excess that are removed by filtration and washings.[6] The deprotection of the amine group has always been challenging as the effectiveness of this deprotection step may vary depending on reagents used, experimental conditions (temperature), length, and combination of amino acids in the chain.[1,6,10] In order to facilitate effective peptide synthesis, previous studies suggested a convergent approach, in which protected peptide fragments are synthesized first and then assembled to construct the protein.[12,13] Although this approach minimizes the cumulative effects of the stepwise synthetic errors, it requires partially protected peptide fragments and hence needs selectively cleavable protecting groups. Ultimately, this is more complex than achieving a > 99.9% yield in each Fmoc deprotection and consequent addition step.[12–14]

The study reported here has been triggered by the following observation: the ultra-performance liquid chromatography-mass spectrometry (UPLC-MS) analysis of some peptides synthesized in the Bossmann labs has revealed Fmoc-protected peptides and peptide fragments in the product mixture giving highly impure products. This phenomenon is linked to premature peptide termination, the omission of amino acids within the peptide chain, and – ultimately – low peptide yields and significant purification challenges. The consensus sequence for matrix metalloproteinase-2 (MMP-2) GCKLVAN-LVAGDG has been chosen as a representative

synthetic challenge. This study addresses the problems of Fmoc-deprotection during SPPS, as well as the virtually complete reaction of the deprotected -N^{α}-amine group peptide chain.



Scheme 2.1. Flow-chart diagram showing Merrifield-SPPS synthesis steps (regenerated with permission from ref 16) with an additional step of using 4% acetic anhydride.

2.2. Materials and Methods

All commercial grade reagents and solvents were purchased from Fisher. Amines were removed from *N*, *N*-dimethyl-formamide (DMF) by means of azeotropic distillation with water and toluene under N_2 gas. Dichloromethane (DCM) was refluxed over calcium hydride (CaH_2) for 1h and then distilled under argon (Ar) gas. Both solvents were stored and reacted under Ar atmosphere. All other reagents including diethylamine, piperidine, *N*, *N*-diisopropylethylamine (DIEA), and acetic anhydride were degassed with Ar before use. All resins, Fmoc-protected amino acids, and hexafluorophosphate benzotriazole tetramethyl uronium (HBTU) were

purchased from Hanhong Tech, Shanghai, China. Trifluoroacetic acid (TFA) and triisopropyl silane (TRIS) were purchased from Aldrich.

2.2.1. Selection of Resin

SPPS is a multi-step process; the first step of the synthesis is selecting suitable solid support (resin) to grow the peptide chain. The chlorotrityl resin and benzyloxybenzyl alcohol resin (Wang resin) are widely used in SPPS and can be easily removed using an acid, after the completion of the desired peptide chain.[15] Resins used for the peptide sequences described throughout this manuscript are listed in **Table 2.1**. The milli-equivalent ratio of resin to HBTU was 1:2.9, and resin to amino acids was 1:3.

Table 2.1. List of resins used for the peptides.

Peptide sequence	Resin
MMP2	H-Gly-2-ClTrt resin
GDD	H-Asp(OtBu)-2-ClTrt resin
GRR	H-Arg(Pbf)-2-ClTrt resin
GEE	Fmoc-Glu(OtBu)-Wang resin
GHH	Fmoc-His(Trt)-Wang resin
GKK	H-Lys(Boc)-2-ClTrt

2.2.2. Removal of the Fmoc-group

2.2.2.1. Fmoc-deprotection of short fragments

The MMP2 peptide chain and the short peptide fragments (GDD, GRR, GEE, GHH, and GHH) were synthesized following standard SPPS protocol.[5,16] H-Gly-2-ClTrt resin, H-Asp(OtBu)-2-ClTrt resin, H-Arg(Pbf)-2-ClTrt resin, and H-Lys(Boc)-2-ClTrt had the first amino acid of the chain that was not Fmoc-protected (additional mass of 222.247 g/mol due to Fmoc group) so the deprotection step was not performed before adding the next amino acid of the short peptide chain. Experimental conditions for the deprotection of the amine group during the synthesis of short peptide fragments are given in **Table 2.2**.

Table 2.2. Summary of the synthesis of short (3 amino acids) peptides.

Peptide sequence	**Resin**	**Reagent (for the Fmoc-deprotection)**	**Incubation Temperature**	**Incubation time**	**Reagent for removing amine impurities**
GDD	H-Asp(OtBu)-2-ClTrt resin	20% Diethylamine in DMF	RT	1 min, 10 min	4% Acetic anhydride in acetonitrile
GRR	H-Arg(Pbf)-2-ClTrt resin	20% Diethylamine in DMF	RT	1 min, 10 min	4% Acetic anhydride in acetonitrile
GEE	Fmoc-Glu(OtBu)-Wang resin	20% Diethylamine in DMF	RT	1 min, 10 min	4% Acetic anhydride in acetonitrile

GHH	Fmoc-His(Trt)-Wang resin	20% Diethylamine in DMF	RT	1 min, 10 min	4% Acetic anhydride in acetonitrile
GKK	H-Lys(Boc)-2-ClTrt	20% Diethylamine in DMF	RT	1 min, 10 min	4% Acetic anhydride in acetonitrile

2.2.2.2. The Fmoc-deprotection of GKK fragments

The synthesis of GKK-2 to GKK-10 peptides was performed by using the diethylamine base for the Fmoc-deprotection step. Multiple diethylamine concentrations (20-40% in DMF) and incubation times were tested at room temperature for GKK-2 to GKK-6 (**Table 2.4**). Similarly, the synthesis of GKK-7 to GKK-10 was carried out by removing the Fmoc group at a higher temperature (36°C), employing different concentrations of diethylamine incubated for different time intervals (**Table 2.6**).

The synthesis of GKK-11 to GKK-21 peptides was performed by using the piperidine base for the Fmoc-deprotection step. Multiple piperidine concentrations (20-40% in DMF) and incubation times were tested at room temperature (**Table 2.9**). The synthesis of GKK-20 and GKK-21 was carried out by removing the Fmoc group at a higher temperature (36°C), employing 20% piperidine incubated for different time intervals (**Table 2.11**).

To deprotect the amine group of GKK-22, GKK-23, and GKK-24, 20% piperidine was used at room temperature (**Table 2.13**). For GKK-22, 20% piperidine was incubated three times (1 min, 10 min, 10 min) at room temperature. The same conditions were used for the GKK-23

and GKK-24 with four incubations (1 min, 10 min, 10 min, 10 min) and five incubations (1 min, 10 min, 10 min, 10 min, 10 min), respectively.

GKK-25, GKK-26, and GKK-27 were synthesized by using 30% diethylamine at room temperature (**Table 2.14**). For GKK-25, 30% diethylamine was incubated three times (1 min, 10 min, 10 min), while for GKK-26 and GKK-27, it was incubated four times (1 min, 10 min, 10 min, 10 min) and five times (1 min, 10 min, 10 min, 10 min, 10 min), respectively.

2.2.2.3. The Fmoc-deprotection of MMP2 peptide sequences

The MMP2 peptide sequence GCKLVAN-LVAGDG was synthesized by removing the Fmoc group with 20% diethylamine base incubated for 1 minute (at room temperature) followed by 10 minutes incubation (at room temperature).

Another MMP2 peptide sequence, GCGLVAN-LVAGKG, was synthesized using 20% diethylamine for different incubation times at room temperature and 36°C (**Table 2.8**).

The same peptide sequence was also synthesized (3 trials) by deprotecting the Fmoc group with 20% piperidine incubated for 1 minute (at room temperature) followed by 10 minutes (at room temperature) (**Table 2.16**).

2.2.3. Coupling of Amino Acids

To facilitate the formation of an amide bond between the Fmoc-deprotected amine of the growing chain and the carboxyl group of the new amino acid, the carboxylic acid group of the new amino acid was activated before adding to the growing chain by using the coupling agent HBTU in DIEA:DMF (1:23). The amino acid mixture was coupled two times to the growing chain and incubated for 30 minutes each to ensure the coupling of the new amino acid. After

completing the amino acid coupling step, four DMF washings were performed and the cycle of deprotection/coupling of amino acids was repeated to grow the peptide chains further.

2.2.4. Labeling of MMP2 peptides with dye

The MMP2 peptide sequence was labeled with rhodamine B dye (RhB) at the N-terminus of the peptide. For the coupling of RhB dye to peptide chain, the milli-equivalent ratio of the resin to HBTU was 1:2.9, and resin to the dye was 1:3. The carboxylic group of the RhB dye was activated using HBTU in DIEA:DMF (1:23). The dye mixture was incubated for 24 hours with the resin-peptide chain, followed by multiple DMF washings to remove any free dye after the coupling step.

2.2.5. Acetic Anhydride

An additional step of adding 4% acetic anhydride (in acetonitrile) was used to remove potential impurities resulting from amine reactions. After each amino acid was added to the chain, acetic anhydride was added to the resin-peptide mixture for 30 minutes, followed by 5 DMF washings. Acetic anhydride was added to GKK-1, GKK-2, GKK-3, GKK-4, GKK-5, GKK-7, GKK-8, GKK-9, GKK-10, GKK-11, GKK-18, GKK-25, GKK-26, GKK-27 and some of MMP2 trials.

2.2.6. Cleaving of resin

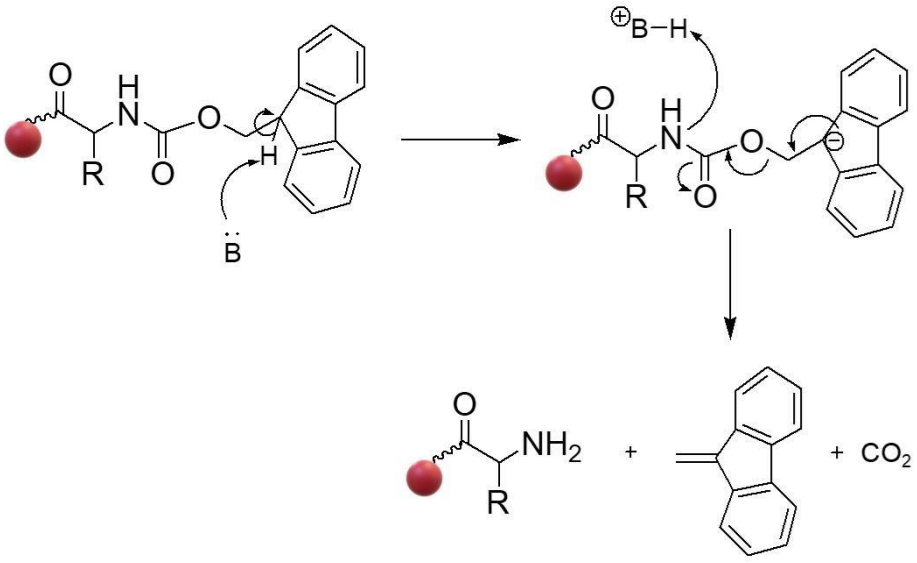
After the desired chain was completed, an acid cocktail solution was used to cleave the peptide sequence from the resin. This cocktail solution consisted of 95% TFA, 2.5% TIPS and 2.5% distilled water. The resin along with the peptide chain was incubated in this cocktail for 3 hours. After the completion of the resin cleaving step, the peptide was precipitated and washed with cold ether, followed by drying with Ar gas.

2.2.7. Analysis of the synthesized peptides

All the samples were prepared in 80% acetonitrile and 20% water v/v to perform UPLC-MS analysis (sample concentration: 0.1 mg/mL) using Waters ACQUITY TQD Tandem Quadrupole UPLC/MS/MS system in ES+ ionization mode using acetonitrile/water gradient. The instrument was equipped with an ACQUITY UPLC peptide CSH C18 column (Waters), 130 Angstrom, 1.7 micrometer 2.1 mm x 150 mm.

2.3. Results and Discussion

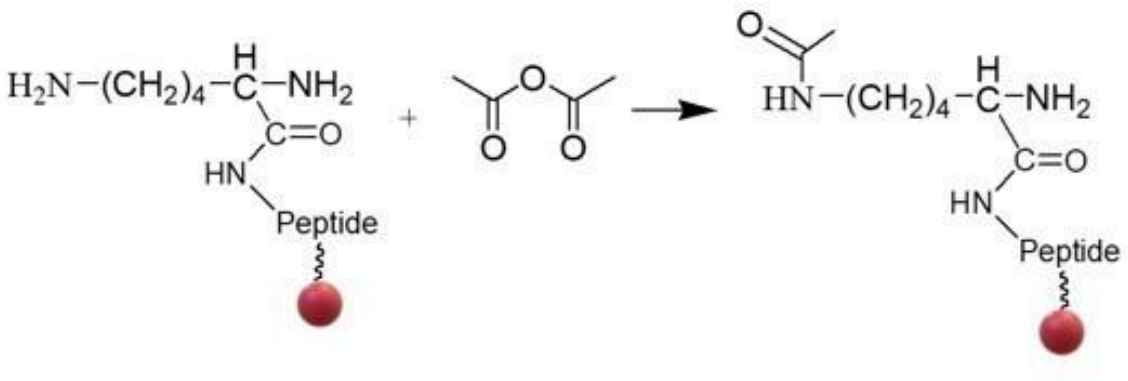
The Fmoc group features an electron-withdrawing fluorenyl group that makes the hydrogen atom of the β-carbon (at position 9) relatively acidic so that a β-elimination reaction using a strong base is able to produce a dibenzofulvene intermediate.[17]



Scheme 2.2. Scheme for the 9-fluorenylmethoxycarbonyl deprotection in solid phase peptide synthesis.[18] (Regenerated with permission from ref 18).

To make a stable adduct, the intermediate is trapped by an excess of amine cleavage agents. This process is faster when a polar solvent medium (such as DMF) is used.[17] Generally, diethylamine is considered an inexpensive base and different concentrations of diethylamine in DMF (up to 20%) have shown high efficiency towards deprotecting the Fmoc group for most of the peptides.[19] Though many studies have suggested variable sets of optimized conditions to standard SPPS protocol[20], further experimental evidence is still needed to find accurate optimized set of conditions[15] to get the high purity of the peptide products. Hence, a few longer and shorter peptides were synthesized using different experimental conditions to test for the maximal removal of the Fmoc group during the deprotection step.

The synthesis of MMP2 peptide was carried out by cleaving the Fmoc group with 20% diethylamine solution incubated for 1 minute followed by 10 minutes (both at room temperature). The additional step of adding the 4% acetic anhydride was performed after the addition of each amino acid of the chain to eliminate the impurities that originate due to the reactions with the amine-groups present on the sidechains of the amino acids or the N-terminal α-NH_2. These NH_2 groups can be modified by acylation with organic acid anhydrides (**Scheme 2.3**).[12,21]



Scheme 2.3. Side-chain acetylation of lysine with acetic anhydride.[21] (Regenerated with permission from ref 21).

The UPLC-MS analysis of MMP2 peptide indicated multiple peaks of high intensities (Figure 2.1A). The theoretical mass for the unlabeled MMP2 peptide sequence (GCKLVAN-LVAGDG) was calculated to be 1216.4 Da which appeared at 2.49 min. The peak purity of the peptide peak calculated using the relative heights of the peaks in the chromatogram was found to be 12.07%. The mass spectrum of the peak at 2.91 min matched the M+1 mass of the peptide with a missing lysine at position 3 in the sequence (1088+1 Da). Interestingly, the UPLC-MS analysis of RhB-labeled MMP2 peptide gave a major peak at 3.86 min corresponding to the MMP2 peptide having Fmoc-protected amine with the mass peak at 1438 m/e (**Figure 2.2A and 2.2C**) while the mass spectrum of the peak at 2.91 min showed the relevant M+1 peak of the RhB-labeled peptide (M+1 mass peak at 1678 m/z) with 17.16% peak purity. These results suggested that the Fmoc group was not completely removed under the conditions used for the Fmoc-deprotection step.

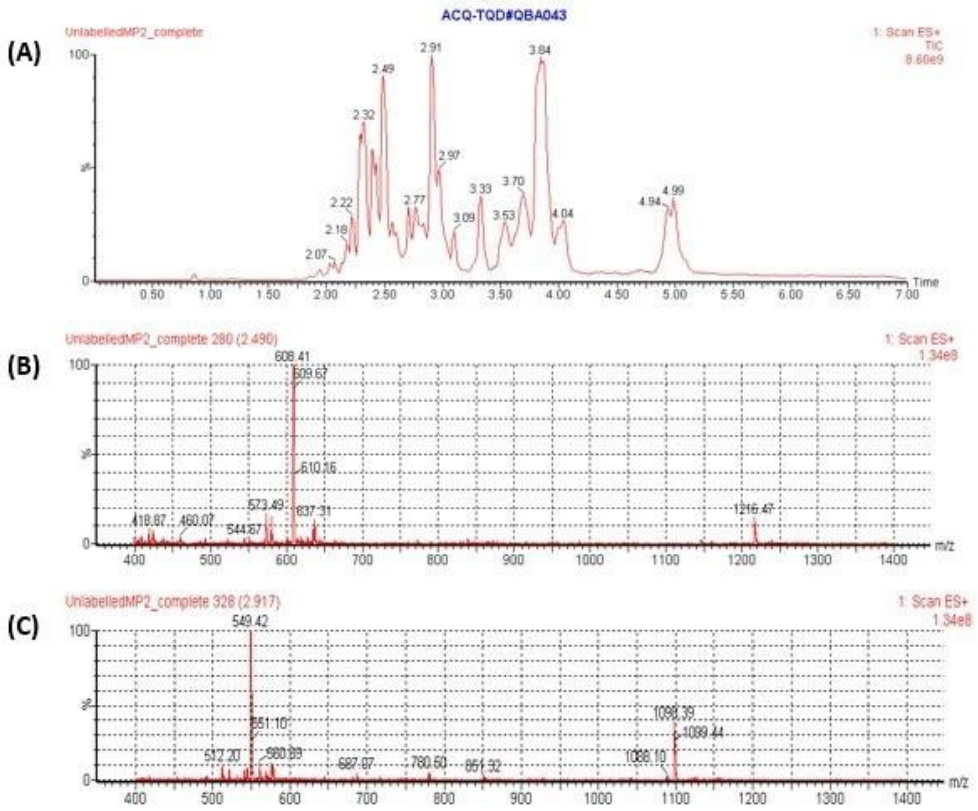


Figure 2.1. (A) UPLC chromatogram of unlabeled MMP2 peptide. The peak at 2.49 min corresponds to the MMP2 peptide product while the peak at 2.91 min corresponds to the peptide with a missing lysine at position 3 in the sequence, **(B)** Mass spectrum of the peak eluted at 2.49 min corresponding to the mass (M+1 mass peak) of MMP2 peptide, **(C)** Mass spectrum of the peak eluted at 2.91 min corresponding to the major peptide impurity with a missing lysine at position 3 in the sequence.

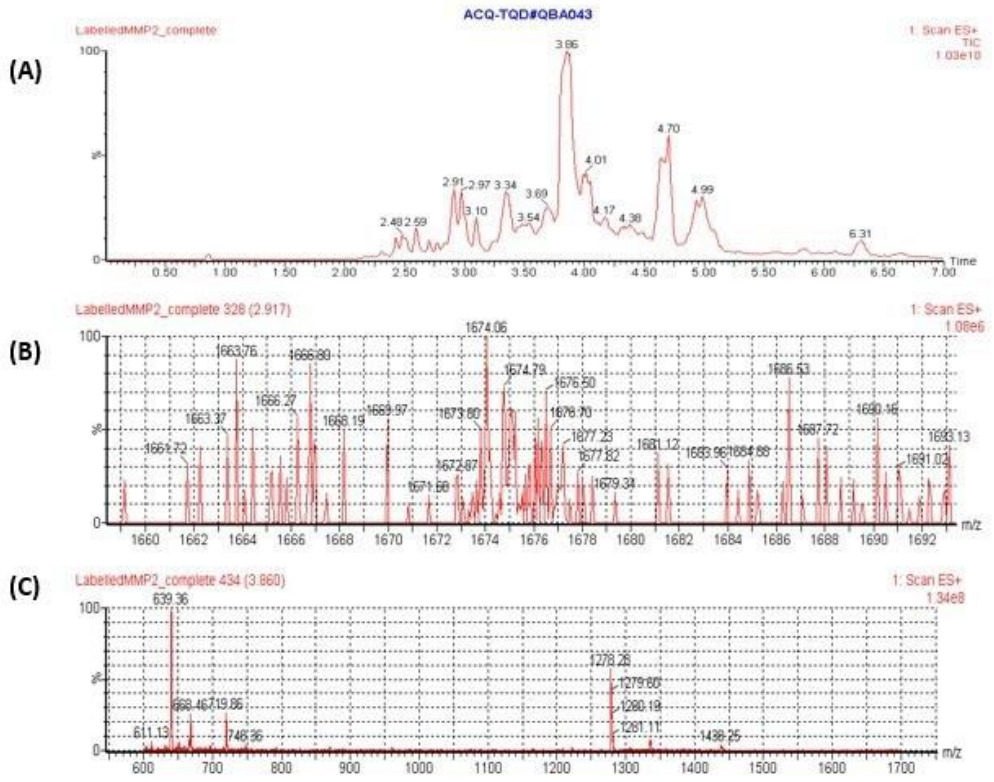


Figure 2.2. (A) The peak at 2.91 min corresponds to the RhB-labeled peptide while the major peak at 3.86 min corresponds to the peptide with Fmoc-protected MMP2, **(B)** Mass spectrum of the peak eluted at 2.91 min corresponding to the mass (M+1 mass peak, 1678 m/z) of the RhB-labeled MMP2 peptide, **(C)** Mass spectrum of the peak eluted at 3.86 min corresponding to the Fmoc-protected MMP2 peptide (m/z=1438).

The use of the small peptide fragments approach was employed to further confirm the identification of the origin of impurities and to optimize the conditions to get rid of those impurities in the longer MMP2 fragment. The summary of the experimental conditions used for the synthesis of short fragments is shown in the **Table 2.2**. The relevant UPLC-MS characterization results are shown throughout the chapter while some are shown in the **Chapter 2 Appendix (S)**.

The GDD peptide showed a broad peak, one major peptide peak that was split in to two peaks, possibly due to dimer formation (**Figure S1-A, Appendix chapter 2**). The peak purity was

found to be 40%. The mass spectrum indicated that both (split) peaks in the chromatogram had the same mass peaks.

The GRR peptide gave a major peptide peak with a peak purity of 51.03% at 0.76 min retention time corresponding to the desired molecular mass peak without the Fmoc-group (387.44+1) and a corresponding M+2 peak (194.06). This fragment indicated that the deprotection of the amine group was efficient, but some impurities of unknown origin were observed (**Figure S2**).

The GEE peptide showed one major peak at 1.06 min retention time corresponding to the mass of the peptide while the second largest peak at 4.306 min retention time corresponds to the peptide with the Fmoc-group (**Figure S3**). The peak purity was calculated to be 57.82%.

The GHH peptide showed two peaks in UPLC chromatogram; the peak at 0.79 min retention time corresponds to the peptide mass peak (peak purity = 32.21%) while the major peak at 2.93 min retention time corresponds to the Fmoc-protected peptide peak (**Figure S4**).

Likewise, the GKK peptide showed two peaks in UPLC chromatogram: the peak at 0.78 min retention time corresponds to the peptide mass peak (second highest peak with a 34.66% peak purity) while the major peak at 2.70 min retention time corresponds to the Fmoc-protected peptide peak fragment (**Figure 2.3**).

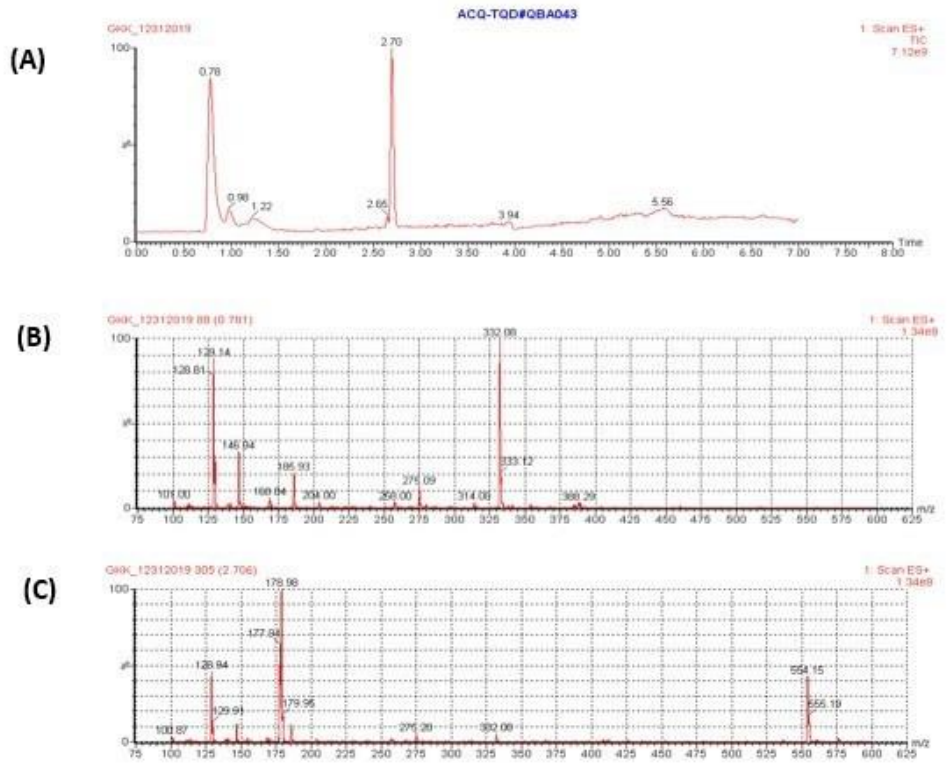


Figure 2.3. (A) UPLC chromatogram of GKK peptide. The peak at 0.78 min corresponds to GKK peptide product and the peak at 2.70 min corresponds to the GHH peptide attached to the Fmoc group, **(B)** Mass spectrum of the peak eluted at 0.78 min corresponding to the mass (M+1 mass peak, 332 m/z) of GKK peptide, **(C)** Mass spectrum of the peak eluted at 2.70 min corresponding to the mass (M+1 mass peak, 554 m/z) of GKK peptide attached to the Fmoc group.

Table 2.3. Peak purities of the 3-amino acid peptide sequences.

Peptide Sequence	Peak purity
GDD	30.49%
GRR	51.03%
GEE	57.82%
GHH	32.21%
GKK	34.66%

2.3.1. Optimization of Fmoc-deprotection conditions

After the synthesis of the MMP2 consensus peptide and all the short peptide fragments, multiple GKK peptides were synthesized to optimize the Fmoc-deprotection step using different concentrations of bases (diethylamine in DMF and piperidine in DMF), time of incubation and temperature.

2.3.1.1. Diethylamine for the Fmoc-deprotection

2.3.1.1.1. Optimization of the concentration of diethylamine at room temperature

The synthesis from GKK to GKK-6 represents different conditions used to optimize the concentration of the diethylamine for the Fmoc group deprotection (**Table 2.4**). The UPLC-MS results from the synthesis of GKK-2 showed a major peptide peak at 0.784 min, in the chromatogram, with some impurities including a peak for the Fmoc-protected GKK peptide and KK fragment at 2.733 and 4.02 min respectively (**Figure 2.4**).

For the GKK-3 peptide, the higher concentration (30% diethylamine) resulted in the major peak corresponding to the peptide at 0.77 min retention time and a minor peak at 2.68 min retention time (**Figure 2.5**), in the chromatogram, corresponding to m/e peak at 332 and 554 in the Mass spectrum (**Figure 2.5**). Overall, 30% diethylamine did not completely deprotect the Fmoc group but gave a very pure peptide without any major impurities including the KK peptide fragment.

The UPLC-MS analysis for GKK-4 indicated a major peak at 2.91 min in the chromatogram corresponding to the mass of the Fmoc-protected peptide (554 g/mol). The second highest peak at 0.80 min in the chromatogram corresponds to the GKK peptide (**Figure S5**).

Similarly, the GKK-5 resulted in a major peak corresponding to the Fmoc-protected peptide at 2.76 min retention time, the second highest peak at 0.76 min retention time corresponds to the GKK peptide while the peak at 4.02 min corresponds to KK-peptide fragment. The presence of the Fmoc-protected peptide peak and KK-peptide fragment indicated that the Fmoc-group was not deprotected completely during the synthesis procedure (**Figure S6**).

To evaluate the effectiveness of 4% acetic anhydride on the Fmoc-deprotection, 4% acetic anhydride was not used for the synthesis of GKK-6 peptide. The UPLC-MS results of GKK-6 showed a major peak corresponding to the peptide at 0.763 min retention time, the second highest peak at 2.626 min retention time corresponds to the Fmoc-protected GKK peptide with a mass peak at 554 in the mass spectrum. The peak at 3.913 min corresponds to KK-peptide fragment with a mass peak at 274 in the mass spectrum. The presence of Fmoc-protected peptide peak indicated that the Fmoc-group was not deprotected completely during the synthesis procedure (**Figure 2.6**). The comparison of GKK-6 with GKK suggested that though the use of 4% acetic anhydride gives more purity of the peptide by removing the unwanted impurities (**Figure 2.3** and **Figure 2.6**) but it resulted in the major Fmoc-protected peptide peak (**Figure 2.3**) as compared to GKK-6 where the major peak corresponds to GKK peptide when 4% acetic anhydride was not used (**Figure 2.6**).

Table 2.4. Summary of GKK synthesis to optimize the concentrations of diethylamine at room temperature.

Peptide sequence	Reagent for the Fmoc-deprotection	Incubation Temperature	First incubation time (min)	Second incubation time (min)	Reagent for removing amine impurities
GKK	20% Diethylamine in DMF	RT	1	10	4% Acetic anhydride in acetonitrile
GKK-2	20% Diethylamine in DMF	RT	1	30	4% Acetic anhydride in acetonitrile
GKK-3	30% Diethylamine in DMF	RT	1	10	4% Acetic anhydride in acetonitrile
GKK-4	30% Diethylamine in DMF	RT	3	14	4% Acetic anhydride in acetonitrile

GKK-5	40% Diethylamine in DMF followed by 20% Diethyamine in DMF	RT	3	10	4% Acetic anhydride in acetonitrile
GKK-6	20% Diethylamine in DMF	RT	1	10	None

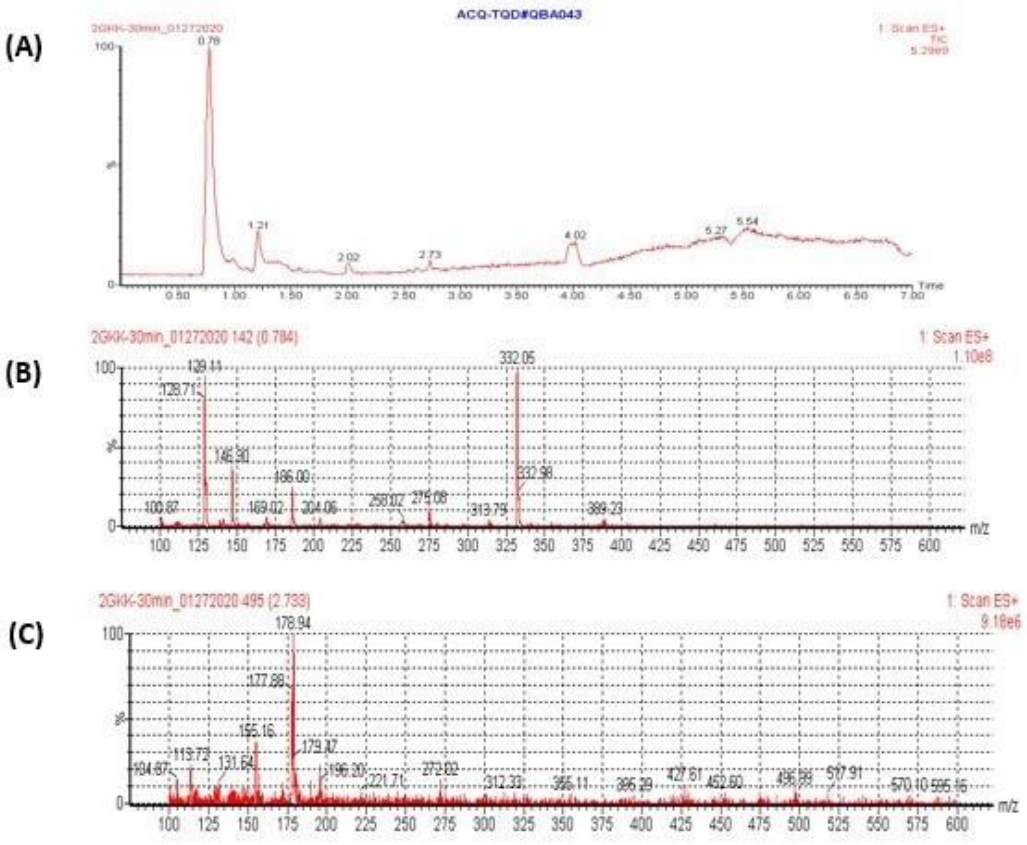


Figure 2.4. (A) UPLC chromatogram of GKK-2. The peak at 0.78 min corresponding to GKK peptide product with minor impurities at 2.73 min and 4.02 min corresponding to the Fmoc-protected peptide and KK peptide fragment, respectively. **(B)** Mass spectrum of GKK-2 corresponding to the GKK at 0.78 min (M+1 mass peak, 332 m/z), **(C)** Mass spectrum of GKK-2 corresponding to Fmoc-protected GKK at 2.73 min.

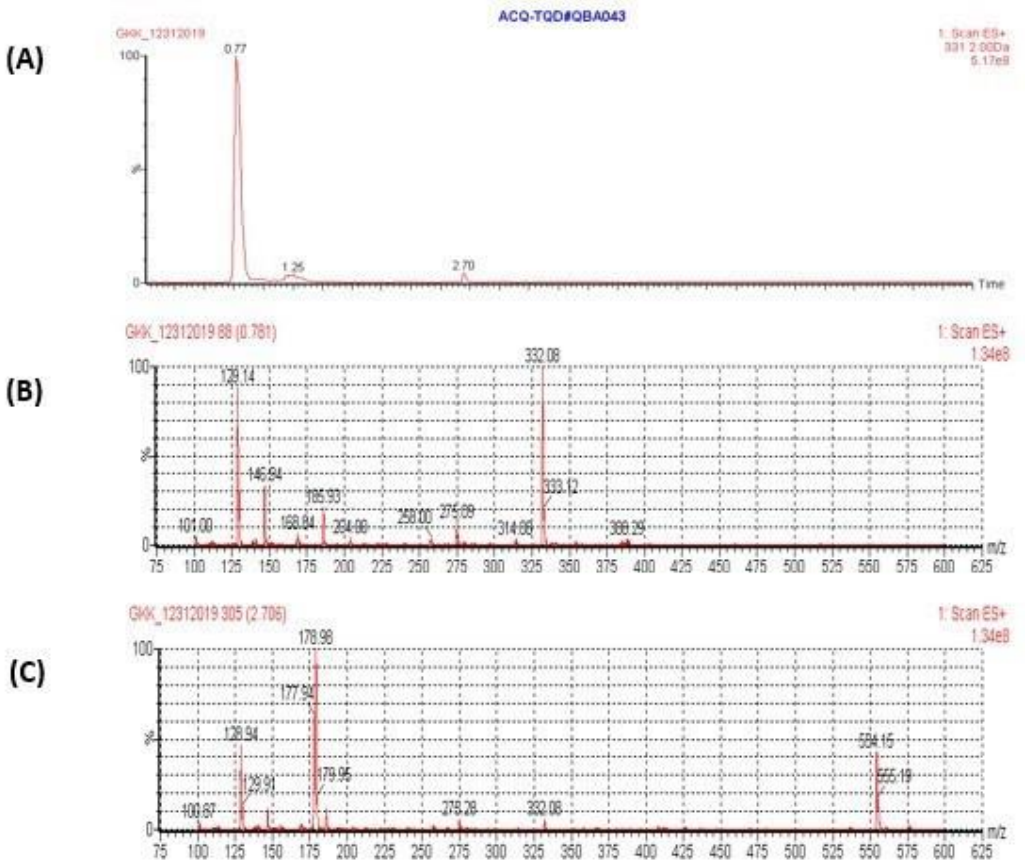


Figure 2.5. (A) UPLC chromatogram of GKK-3. The peak at 0.77 min corresponds to GKK peptide product with the minor impurity peak at 2.70 min corresponding to the Fmoc-protected peptide, **(B)** Mass spectrum of GKK-3 corresponding to the major peak at 0.78 min, **(C)** Mass spectrum of GKK-3 corresponding to the Fmoc-protected GKK-peptide.

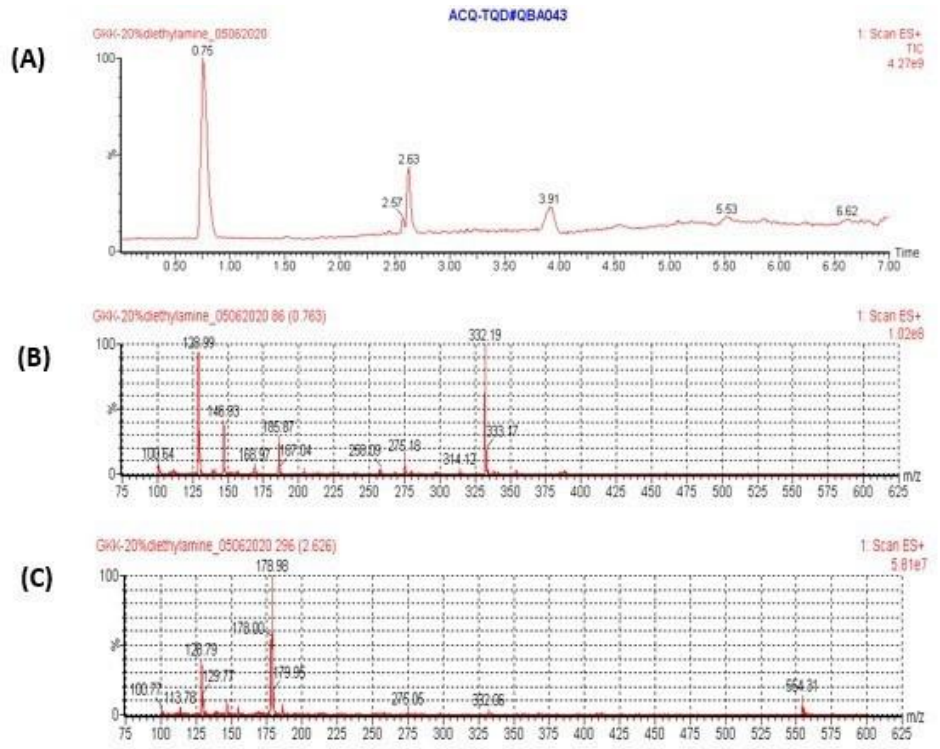


Figure 2.6. (A) UPLC chromatogram of GKK-6. The peak at 0.75 min corresponds to GKK peptide product with the impurity peaks at 2.63 min and 3.91 min corresponding to the Fmoc-protected GKK peptide impurity and KK peptide fragment peak, respectively **(B)** Mass spectrum of GKK-6 corresponding to the major GKK peptide peak at 0.76 min **(C)** Mass spectrum of GKK-6 corresponding to the Fmoc-protected peptide impurity peak at 2.62 min.

Overall, the peptides GKK-2, GKK-3 and GKK-6 showed the GKK peptide peak as the base peak. Though the peak of the Fmoc-protected peptide fragment was present, but it showed very low intensity when compared to the original GKK peptide (**Figure 2.3**). Among the peptides with the major peptide peak, the synthesis conditions used for GKK-3 showed best Fmoc-group deprotection results as confirmed by UPLC-MS (**Figure 2.5**). GKK-3 did not show the peak for KK fragment (3.91-4.02 min retention time, **Figure 2.5**) and gave a very low intensity peak for the Fmoc-protected peptide hence GKK-3 showed the best optimized concentration for deprotecting the Fmoc-group (with 92.97% peak purity) employing

diethylamine base, at room temperature. The peak purities were also high for GKK-2 and GKK-6 peptides (**Table 2.5**).

Table 2.5. Peak purities of GKK-2, GKK-3, GKK-4, GKK-5 and GKK-6.

Peptide sequence	Peak purity
GKK-2	62.16%
GKK-3	92.97%
GKK-4	18.83%
GKK-5	18.51%
GKK-6	60.79%

2.3.1.1.2. Optimization of the concentration of diethylamine at higher temperature

The removal of the Fmoc-protecting group was also carried out at a slightly higher temperature (36°C) to evaluate the effect of the temperature on the deprotection step and ultimately on to the purity of the peptide product (GKK-7 to GKK-10). In the case of GKK-7, the UPLC-MS results indicated the major peptide peak at 0.77 min retention time corresponding to the mass of GKK peptide 332 g/mol (**Figure 2.7**). The presence of minor peak in the chromatogram at 2.7 min retention time corresponds to the m/e peak at 554 (**Figure 2.7C**), indicating that the higher temperature for cleaving the Fmoc-group did not efficiently remove the Fmoc-group from the peptide chain.

The characterization of the results of GKK-8 showed a major peak at 0.77 min in chromatogram corresponding to the GKK peptide (332 g/mol). The presence of Fmoc-protected

m/e peak at 554 suggested that using higher temperature (36°C) and slightly longer incubation time (15 min) did not prove helpful to get rid of the Fmoc-group completely (**Figure S7**).

The Fmoc-group deprotection of GKK-9 employing 20% diethylamine with 4 repeats of 1 minute incubation at 36°C resulted in a major peak at 2.78 min retention time corresponding to the Fmoc-protected peptide peak and the minor peak at 0.77 min corresponding to the peptide peak (**Figure S8**). The higher temperature along with the multiple incubations did not improve the purity of the peptide product and gave a major peak for the Fmoc-protected peptide.

The GKK-10 peptide was synthesized by increasing the temperature of the best optimized condition with diethylamine for the deprotection of the NH_2-protecting group. The UPLC-MS results for this synthesis showed a major peak at 2.66 min in the chromatogram corresponding to the mass of the Fmoc-protected GKK peptide (554 g/mol). The desired peptide was among the one of minor peaks in the chromatogram (at 0.75 min). The use of higher temperature significantly increased the impurity of the product, as confirmed by UPLC-MS (**Figure 2.8**).

Table 2.6. Summary of GKK synthesis to optimize the concentrations of diethylamine and time of incubation at higher temperature.

Peptide sequence	**Reagent for the Fmoc-deprotection**	**Incubation Temperature**	**First incubation time (min)**	**Second incubation time (min)**	**Reagent for removing amine impurities**
GKK-7	20% Diethylamine in DMF	36°C	1	10	4% Acetic anhydride

					in acetonitrile
GKK-8	20% Diethylamine in DMF	36°C	1	15	4% Acetic anhydride in acetonitrile
GKK-9	20% Diethylamine in DMF	36°C	1 (4 repeats)		4% Acetic anhydride in acetonitrile
GKK-10	30% Diethylamine in DMF	36°C	1	10	4% Acetic anhydride in acetonitrile

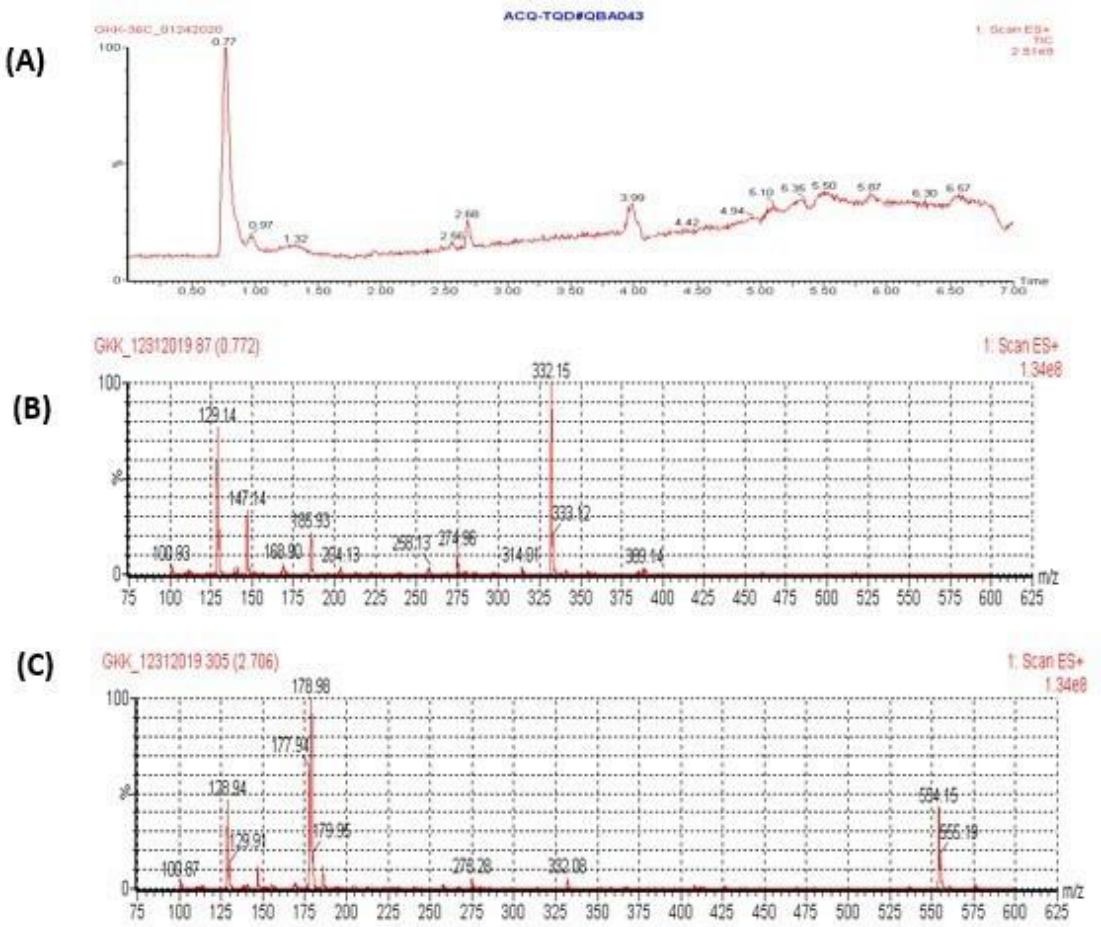


Figure 2.7. (A) UPLC chromatogram of GKK-7. The peak at 0.77 min corresponds to GKK peptide product with the impurity peaks at 2.68 min and 3.99 min corresponding to the Fmoc-protected GKK peptide impurity and KK peptide fragment peak, respectively **(B)** Mass spectrum of GKK-7 corresponding to the major GKK peptide peak at 0.77 min, **(C)** Mass spectrum of GKK-7 corresponding to the minor Fmoc-protected peptide peak at 2.71 min.

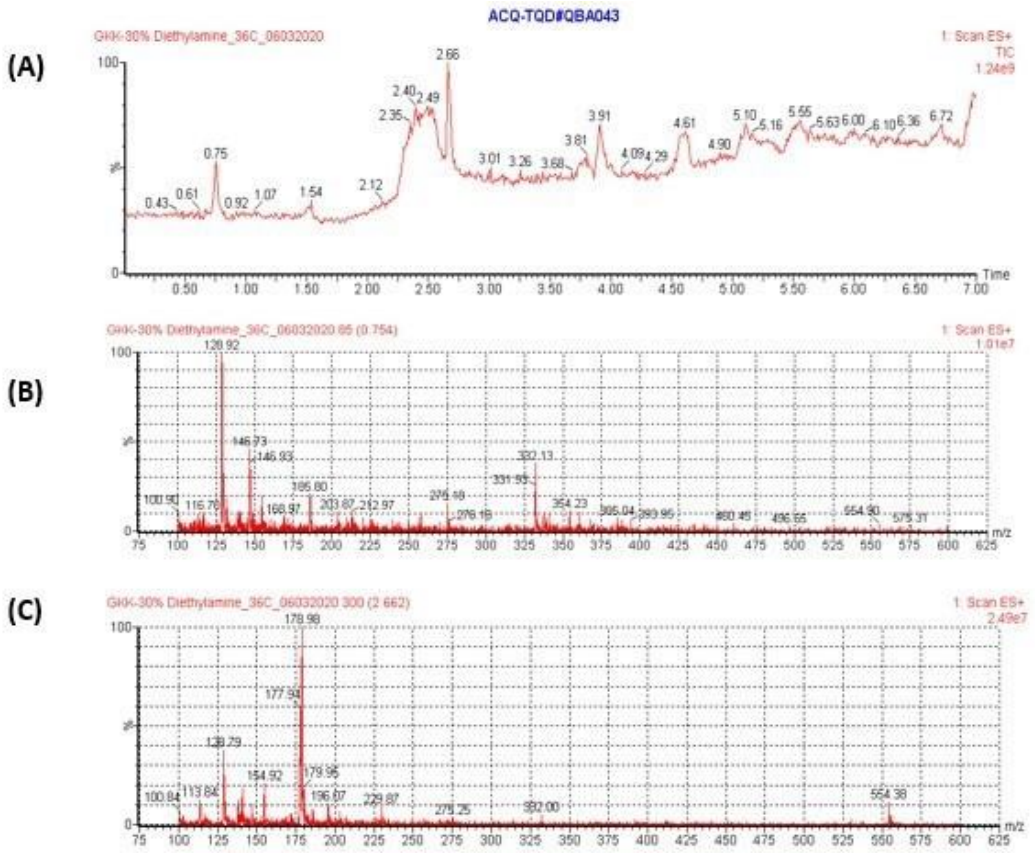


Figure 2.8. (A) UPLC chromatogram of GKK-10. The peak at 0.75 min corresponds to GKK peptide product with the impurity peaks at 2.66 min and 3.91 min corresponding to the Fmoc-protected GKK peptide impurity and KK peptide fragment peak, respectively **(B)** Mass spectrum of GKK-10 corresponding to the minor GKK peptide peak at 0.75 min, **(C)** Mass spectrum of GKK-10 corresponding to the Fmoc-protected peptide peak at 2.66 min.

Table 2.7. Peak purities of GKK-7, GKK-8, GKK-9 and GKK-10.

Peptide sequence	Peak purity
GKK-7	56.66%
GKK-8	54.84%
GKK-9	15.52%
GKK-10	7.43%

2.3.1.1.3. Diethylamine for the synthesis of longer peptide chains

Finally, a new consensus sequence for the MMP2 was developed to evaluate the purity of oligopeptide synthesis under optimized Fmoc-cleaving conditions. The principal knowledge regarding the statistical presence of each amino acid at each position of the four amino acids before and after the cleavage site was obtained from MEROPS.[22] The amino acid sequence developed was GCGLVAN-LVAGKG. The difference between the previously synthesized amino acid sequence and new amino acid sequence is that a lysine at position 3 of the chain is replaced with glycine to eliminate the sidechain having NH_2-group that may lead to unwanted impurities. The aspartic acid at position number 12 is replaced with lysine to avoid the possible unwanted fragments due to the side-chain reactions. The new MMP2 consensus sequence was synthesized employing three different Fmoc-cleaving conditions in three different trials, as shown in **Table 2.8**.

Table 2.8. Summary of MMP2 synthesis (with diethylamine).

Trial#	**Reagent for the Fmoc-deprotection**	**Incubation temperature**	**Incubation time**	**Reagent for removing amine impurities**
Trial#1	20% Diethylamine	RT	1 min, 10 min	4% Acetic anhydride in Acetonitrile
Trial#2	20% Diethylamine	RT	1 min, 30 min	4% Acetic anhydride in Acetonitrile
Trial#3	20% Diethylamine	36°C	1 min, 10 min	4% Acetic anhydride in Acetonitrile

In the first trial, the MMP2 consensus peptide sequence was synthesized by following the standard SPPS protocol; the Fmoc group was cleaved with 20% diethylamine for 1 minute incubation followed by 10 minutes incubation (both at room temperature). The purpose of employing same conditions (for trial#1) as previously synthesized MMP2 peptide was to compare the results from both peptide sequences under the same Fmoc-cleaving conditions. The UPLC-MS results of the complete MMP2 consensus peptide sequence showed fewer impurities (with 21.13% peak purity, **Figure 2.9**) than previously synthesized MMP2 sequence (**Figure 2.1** and **Figure 2.2**) but the two major peaks in the chromatogram did not appear to be the desired mass peak of the MMP2 consensus peptide sequence (**Figure 2.9A**) and gave a small m/e peak at 1158 in the mass spectrum. The MMP2 peptide was labeled with RhB following the same Fmoc-cleaving conditions and UPLC-MS results again showed multiple peaks in the chromatogram with missing RhB-labeled peptide peak (**Figure S9**).

In the trial 2, the MMP2 peptide sequence was synthesized by cleaving the Fmoc-group with 20% diethylamine for 1 minute incubation followed by 30 minutes incubation, at room temperature. The results of the complete MMP2 peptide sequence showed fewer impurities (peak purity = 25.56%) with a major peak at 2.52 min in chromatogram corresponding to the mass of the MMP2 peptide (1158 g/mol) in mass spectrum (**Figure 2.10**).

In the trial 3, the MMP2 peptide sequence was synthesized by cleaving the Fmoc-group with 20% diethylamine for 1 minute incubation followed by 10 minutes incubation, at 36°C in the incubator. The results of the complete MMP2 peptide sequence again showed relatively fewer impurities (peak purity = 23.66%) with a major peak at 2.52 min in chromatogram corresponding to the mass of the MMP2 peptide (1158 g/mol) in mass spectrum (**Figure 2.11**).

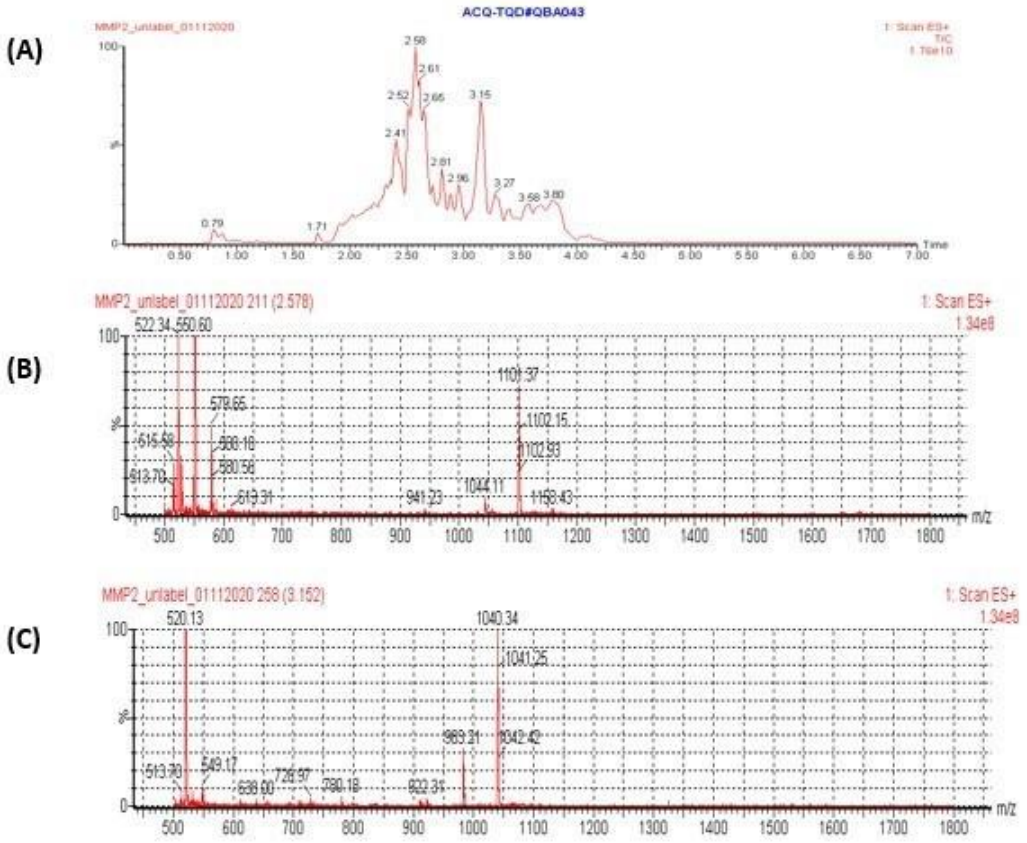


Figure 2.9. (A) UPLC chromatogram of unlabeled MMP2 (Trial#1). The peak at 2.58 min corresponds to the unlabeled MMP2 peptide product and the peak at 3.15 min corresponds to the impurity (peptide fragment without a cysteine and a water molecule), **(B)** Mass spectrum of the peak at 2.57 min corresponds to the mass (1158) of the MMP2 peptide, **(C)** Mass spectrum of the peak at 3.15 min corresponds to the mass of the impurity (peptide without a cysteine and one less water molecule).

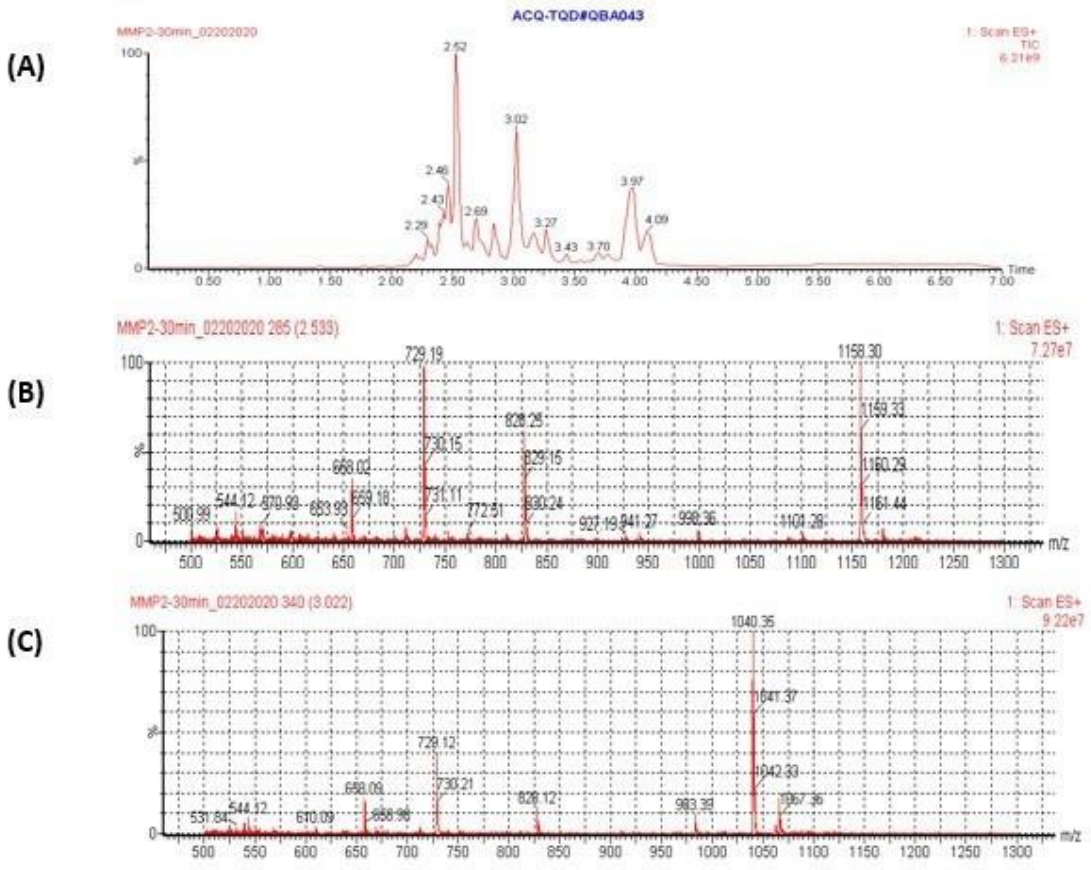


Figure 2.10. (A) UPLC chromatogram of MMP2 (Trial# 2). The peak at 2.52 min corresponds to MMP2 peptide product and the peak at 3.02 min corresponds to the impurity (peptide fragment with a missing cysteine and a water molecule), **(B)** Mass spectrum of the peak at 2.53 min corresponding to the mass of MMP2 peptide product, **(C)** Mass spectrum of the peak at 3.02 min corresponding to the mass of the impurity (peptide fragment with a missing cysteine and one less water molecule).

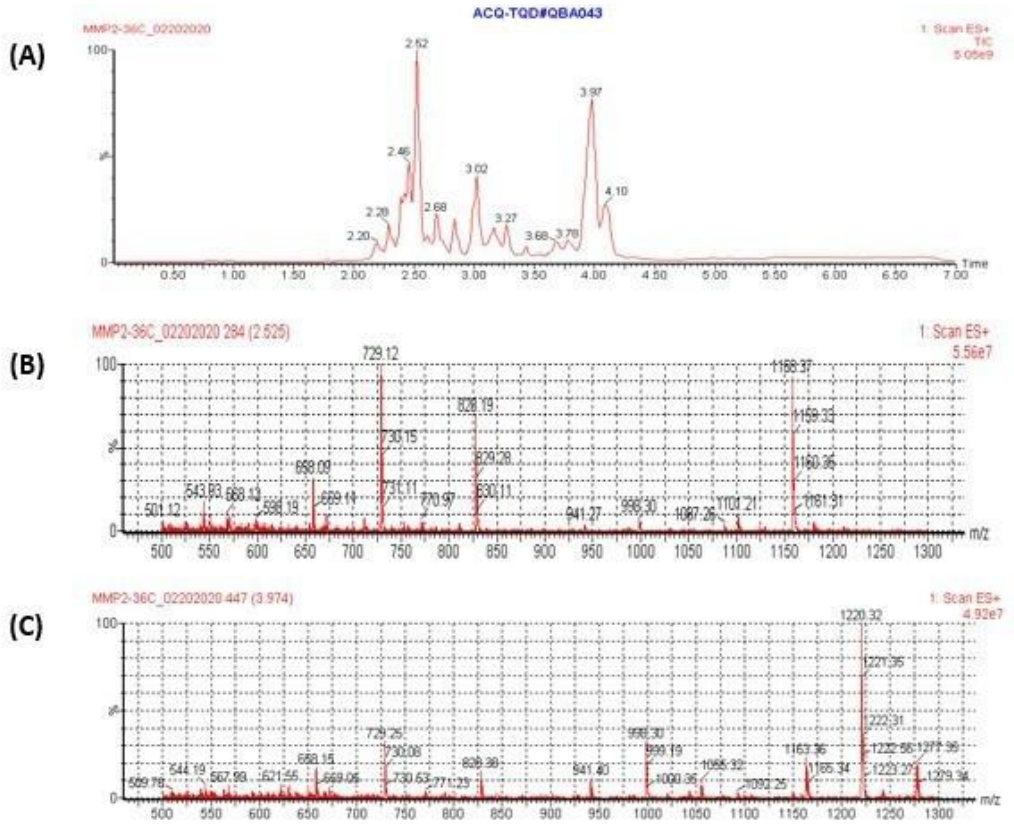
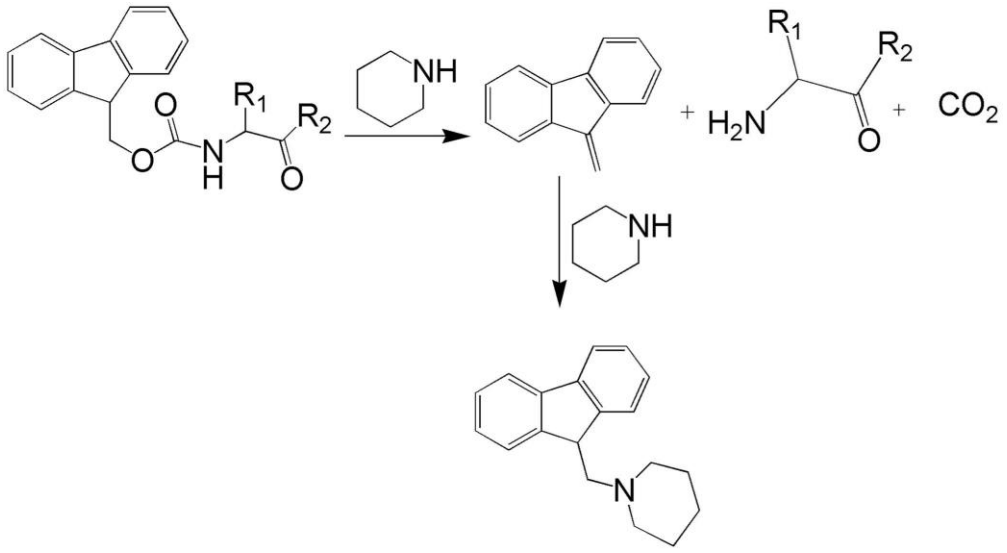


Figure 2.11. (A) UPLC chromatogram of MMP2 (Trial#3). The peak at 2.52 min corresponds to the MMP2 peptide product and the peak at 3.97 min corresponding to the number of peptide fragments, **(B)** Mass spectrum of the peak at 2.52 min corresponding to the MMP2 peptide product, **(C)** Mass spectrum of the peak at 3.97 min corresponding to multiple peptide fragments.

2.3.1.2. Piperidine for the Fmoc-deprotection

Many studies have shown piperidine as the major candidate for removing the Fmoc-group during the SPPS.[23] The concentration and conditions for the reaction have been optimized by many research groups,[23,24] Multiple studies have suggested different set of conditions and concentrations employing piperidine such as 10-20% piperidine [15] and 10-55% piperidine[17] in a variety of solvents for a range of incubation times and temperatures. But a definite concentration along with other conditions (incubation time and temperature) remain a question.



Scheme 2.4. The Fmoc-group removal with piperidine (in DMF).[18] (Regenerated with permission from ref 18).

The use of the piperidine comes with the possible aminolysis of some amino acids, mostly aspartic acid and glutamic acid, leading to the formation of corresponding α and β piperidides that renders the formation of α and β peptides.[15,25]

2.3.1.2.1. Optimization of the concentration of piperidine at room temperature

To further optimize the Fmoc-deprotection conditions another base, piperidine, was used for the Fmoc-group deprotection for GKK-11 to GKK-19 (**Table 2.9**). The UPLC-MS results for GKK-11 showed a major peak at 0.78 min in chromatogram corresponding to the mass of GKK peptide (332 g/mol). The second highest peak in the chromatogram at 4.00 min had a corresponding mass of the KK fragment 274 g/mol. The intensity of the Fmoc-protected peptide peak significantly decreased when piperidine was used for the removal of the Fmoc group (**Figure 2.12**).

The UPLC-MS results for GKK-12 indicated a major peak at 0.76 min in the chromatogram corresponding to the mass of GKK peptide (332 g/mol). Surprisingly, there was

no peak for the Fmoc-protected peptide and the only impurity present was KK peptide fragment (**Figure 2.13**). The results from GKK-12 suggested that avoiding reaction with acetic anhydride offered fewer impurities, especially, the impurities related to the Fmoc-deprotection.

The best optimized concentration (20% piperidine) was also tested for longer incubation time (GKK-13 and GKK-14). The GKK-13 peptide that gave a major peak at 0.75 min in chromatogram corresponding to the mass of GKK peptide (332 g/mol). The second highest peak in the chromatogram at 3.94 min showed a corresponding mass of the KK fragment 274 g/mol (**Figure S10**). The UPLC-MS results for the synthesis of GKK-14 showed the peptide peak (0.75 min in chromatogram) to be a minor product peak among other peaks. The major peak at 2.66 min in the chromatogram corresponds to the mass of the Fmoc-protected GKK peptide (**Figure S11**). Overall, the peptides GKK-13 and GKK-14 showed less purity of the peptide than GKK-11 and GKK-12 upon increased incubation times of piperidine solution.

GKK-15 to GKK-19 were synthesized by using higher concentration of piperidine (more than 20%) for the Fmoc-deprotection step. The synthesis of GKK-15 resulted in a major peak at 0.76 min in chromatogram corresponding to the mass of GKK peptide (332 g/mol). The UPLC chromatogram also indicated a peak at 2.60 min corresponding to the Fmoc-protected GKK peptide peak (**Figure S12**). Another peak in the chromatogram at 3.93 min had a corresponding mass of the KK fragment 274 g/mol. The analysis of GKK-16 indicated a major peak at 0.76 min in chromatogram corresponding to the mass of GKK peptide (332 g/mol). The UPLC chromatogram also indicated a peak at 2.63 min corresponding to the Fmoc-protected GKK peptide peak (**Figure 2.14**). Another peak in the chromatogram at 3.93 min had a corresponding mass of the KK fragment 274 g/mol.

To further eliminate the Fmoc-protected peptide fragments, GKK-17 was synthesized by deprotecting the Fmoc-group with 30% piperidine with slightly longer time of incubation. The UPLC-MS results showed a major peak at 0.76 min corresponding to the mass of GKK peptide (332 g/mol) and a minor peak at 2.70 min corresponding to the Fmoc-protected GKK peptide peak (**Figure S13**) appeared in the UPLC chromatogram. Another peak in the chromatogram at 3.85 min retention time gave a corresponding mass of the KK fragment 274 g/mol.

To compare the purity of the peptide with and without the 4% acetic anhydride step, GKK-18 was synthesized by using the additional step of adding 4% acetic anhydride at the end of each cycle while keeping all other conditions constant as GKK-16. The UPLC-MS results for this synthesis gave a major peak at 0.78 min in the chromatogram corresponding to the M+1 mass (331+1 Da) of the GKK peptide (**Figure S14**). The major impurities (the Fmoc-protected peptide and KK fragment) were also present.

For GKK-19, the UPLC and MS results showed a major peak at 0.75 min corresponding to the mass of GKK peptide (332 g/mol) and a minor peak at 2.64 min corresponding to the Fmoc-protected GKK peptide peak (**Figure 2.15**). The peak in the chromatogram at 4.01 min showed a corresponding mass of the KK fragment 274 g/mol. The peptides GKK-16, GKK-18 and GKK-19 showed fewer impurities than GKK-13, GKK-14, GKK-15, and GKK-17.

Table 2.9. Summary of GKK synthesis to optimize the concentration of piperidine and time of incubation at room temperature.

Peptide sequence	Reagent for the Fmoc-deprotection	Incubation temperature	Incubation time	Reagent for removing amine impurities
GKK-11	20% Piperidine in DMF	RT	1 min, 10 min	4% Acetic anhydride in Acetonitrile
GKK-12	20% Piperidine in DMF	RT	1 min, 10 min	None
GKK-13	20% Piperidine in DMF	RT	1 min, 15 min	None
GKK-14	20% Piperidine in DMF	RT	1 min, 30 min	None
GKK-15	25% Piperidine in DMF	RT	1 min, 10 min	None
GKK-16	30% Piperidine in DMF	RT	1 min, 10 min	None
GKK-17	30% Piperidine in DMF	RT	1 min, 15 min	None
GKK-18	30% Piperidine in DMF	RT	1 min, 10 min	4% Acetic anhydride in acetonitrile
GKK-19	40% Piperidine in DMF	RT	1 min, 10 min	None

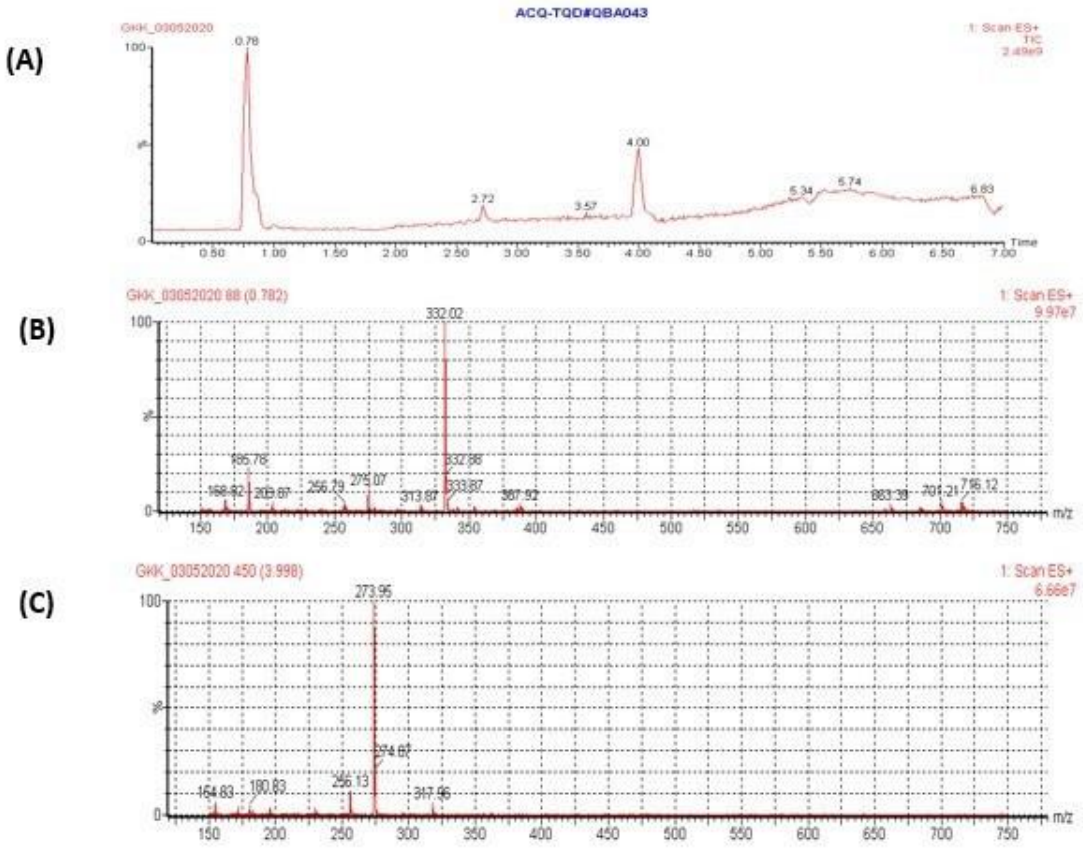


Figure 2.12. (A) UPLC chromatogram of GKK-11. The peak at 0.78 min corresponds to GKK peptide product with the impurity peaks at 2.72 min and 4.00 min corresponding to the Fmoc-protected GKK peptide impurity and KK peptide fragment peak, respectively **(B)** Mass spectrum of GKK-11 corresponding to the major GKK peptide product eluted at 0.78 min, **(C)** Mass spectrum of GKK-11 eluted at 3.99 min corresponding to the mass of KK peptide fragment.

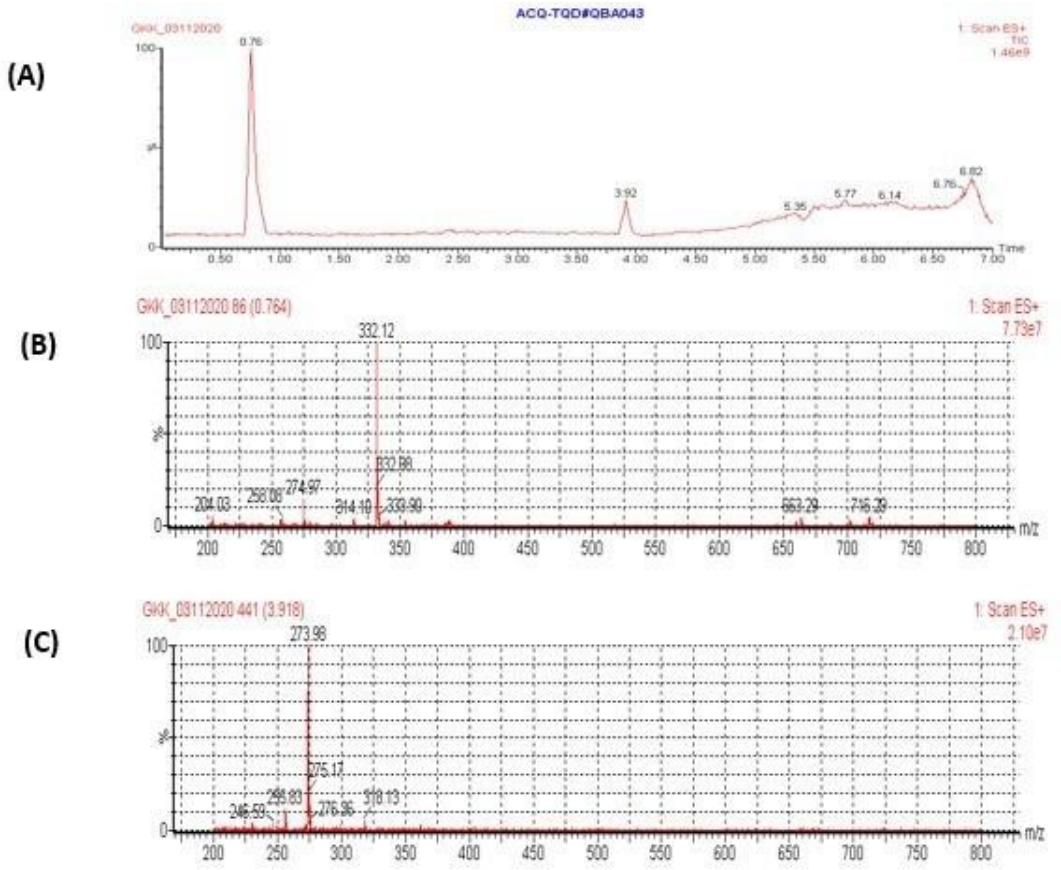


Figure 2.13. (A) UPLC chromatogram of GKK-12. The peak at 0.76 min corresponds to GKK peptide product with the impurity peak at 3.92 min corresponding to the KK peptide fragment, **(B)** Mass spectrum of GKK-12 eluted at 0.76 min, corresponding to the GKK peptide, **(C)** Mass spectrum of GKK-12 corresponding to the KK peptide fragment eluted at 3.92 min.

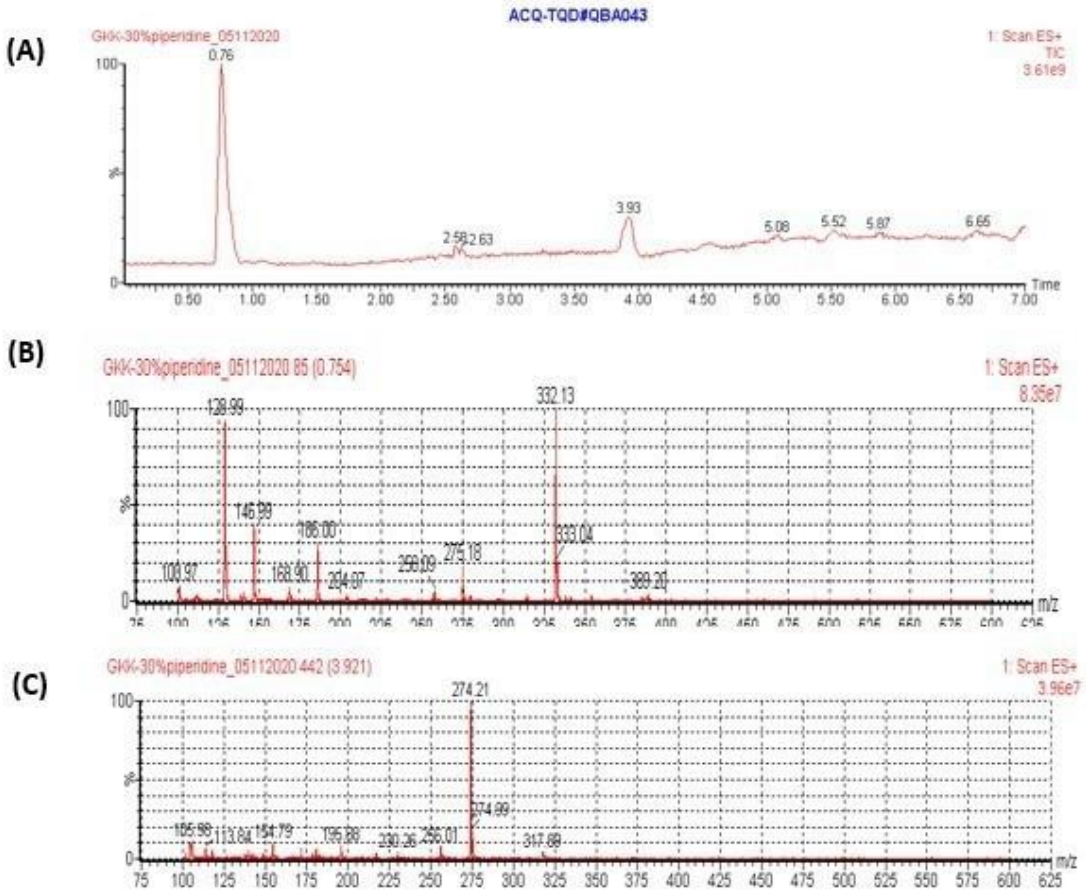


Figure 2.14. (A) UPLC chromatogram of GKK-16. The peak at 0.76 min corresponds to the GKK peptide product and the peak at 3.93 min corresponds to the KK peptide fragment impurity, **(B)** Mass spectrum of GKK-16 corresponding to the GKK peptide eluted at 0.75 min, **(C)** Mass spectrum of GKK-16 corresponding to the KK peptide fragment eluted at 3.92 min.

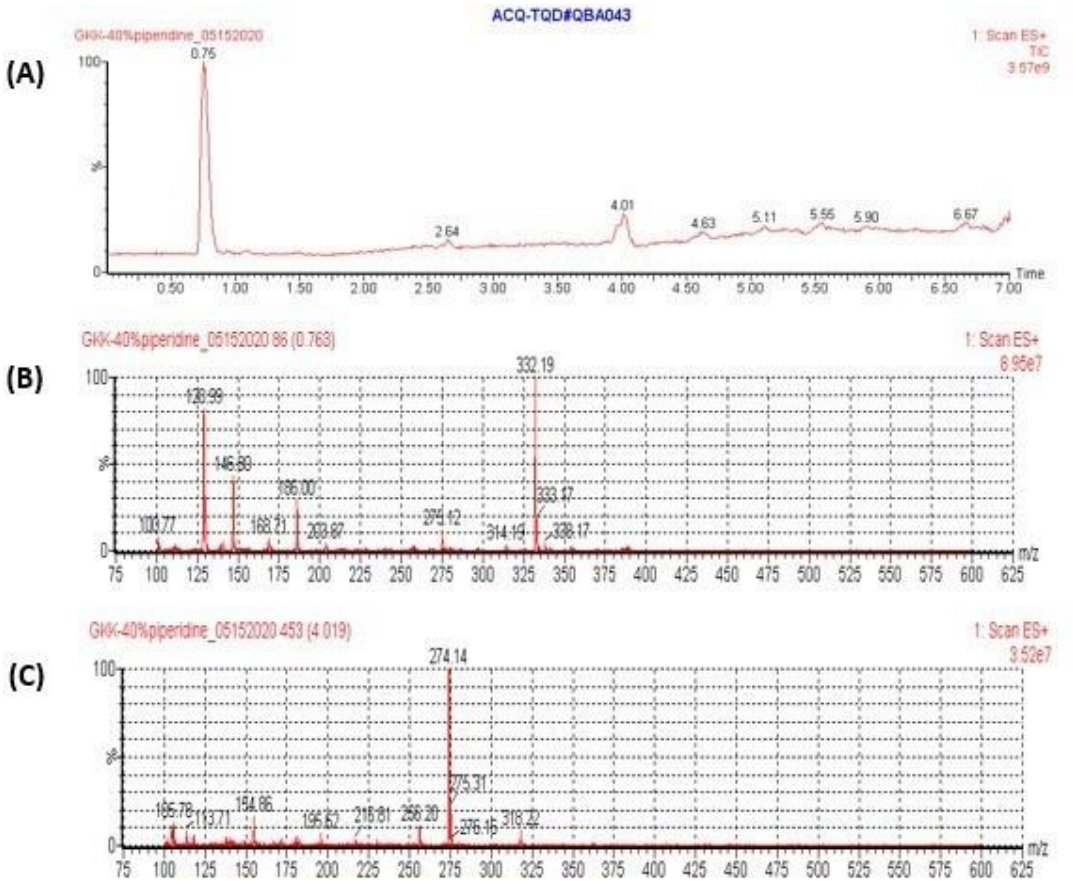


Figure 2.15. (A) UPLC chromatogram of GKK-19. The peak at 0.75 min corresponds to GKK peptide product and the peak at 4.01 min corresponds to the KK peptide fragment impurity, **(B)** Mass spectrum of GKK-19 corresponding to the GKK peptide product eluted at 0.76 min, **(C)** Mass spectrum of GKK-19 corresponding to the KK peptide fragment eluted at 4.02 min.

Table 2.10. The peak purities of the GKK peptides (GKK-11 to GKK-19) synthesis to optimize the concentration of piperidine and time of incubation at room temperature.

Peptide sequence	Peak purity
GKK-11	55.36%
GKK-12	63.99%
GKK-13	25.58%
GKK-14	8.08%
GKK-15	32.71%

GKK-16	49.63%
GKK-17	28.38%
GKK-18	38.78%
GKK-19	38.67%

2.3.1.2.2. Optimization of the concentration of piperidine at higher temperature

To investigate the effect of higher temperature on the efficacy of the Fmoc-deprotection of amino acids, the best optimized condition with piperidine was tested at a higher temperature. For the first trial (GKK-20), 20% piperidine was used for the removal of the Fmoc group for 1 minute and 10 minutes at 36°C while the second trial was performed with the same conditions as trial 1 but with the shorter second incubation (5 minutes). The UPLC-MS results showed that the higher temperature for the Fmoc-deprotection step gave very impure product mixture for both trials. Many unwanted impurities were observed along with GKK peptide and the Fmoc-protected GKK peptide. This approach led to an important finding that the increase in the temperature does not always enhances the efficiency of a reaction (as the trial 1 was performed with the best optimized conditions with piperidine except the temperature) and meanwhile the second trial proved that increasing the temperature and shortening the time of incubation does not always work as well (**Figure 2.16** and **Figure 2.17**).

Table 2.11. Summary of GKK synthesis to optimize the concentration of piperidine and time of incubation at higher temperature.

Peptide sequence	Reagent for the Fmoc-deprotection	Incubation temperature	Incubation time	Reagent for removing amine impurities
GKK-20	20% Piperidine in DMF	36°C	1 min, 10 min	None
GKK-21	20% Piperidine in DMF	36°C	1 min, 5 min	None

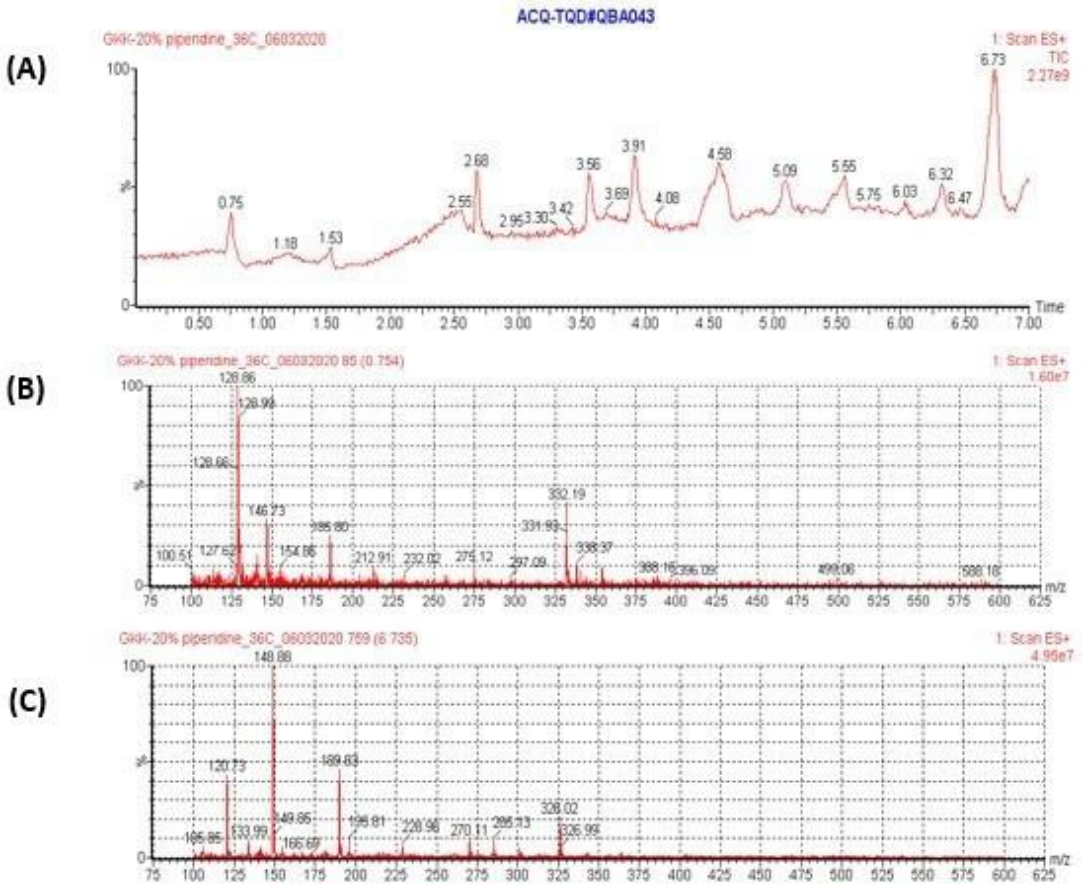


Figure 2.16. (A) UPLC chromatogram of GKK-20. The peak at 0.75 min corresponds to the GKK peptide product with the impurity peaks at 2.68 min, 3.91 min and 6.73 min corresponding to the Fmoc-protected GKK peptide, KK peptide and a free K amino acid, respectively **(B)** Mass spectrum of GKK-20 corresponding to the GKK peptide eluted at 0.75 min, **(C)** Mass spectrum of GKK-20 corresponding to the minor peak at 6.73 min (M+1 mass peak for free lysine).

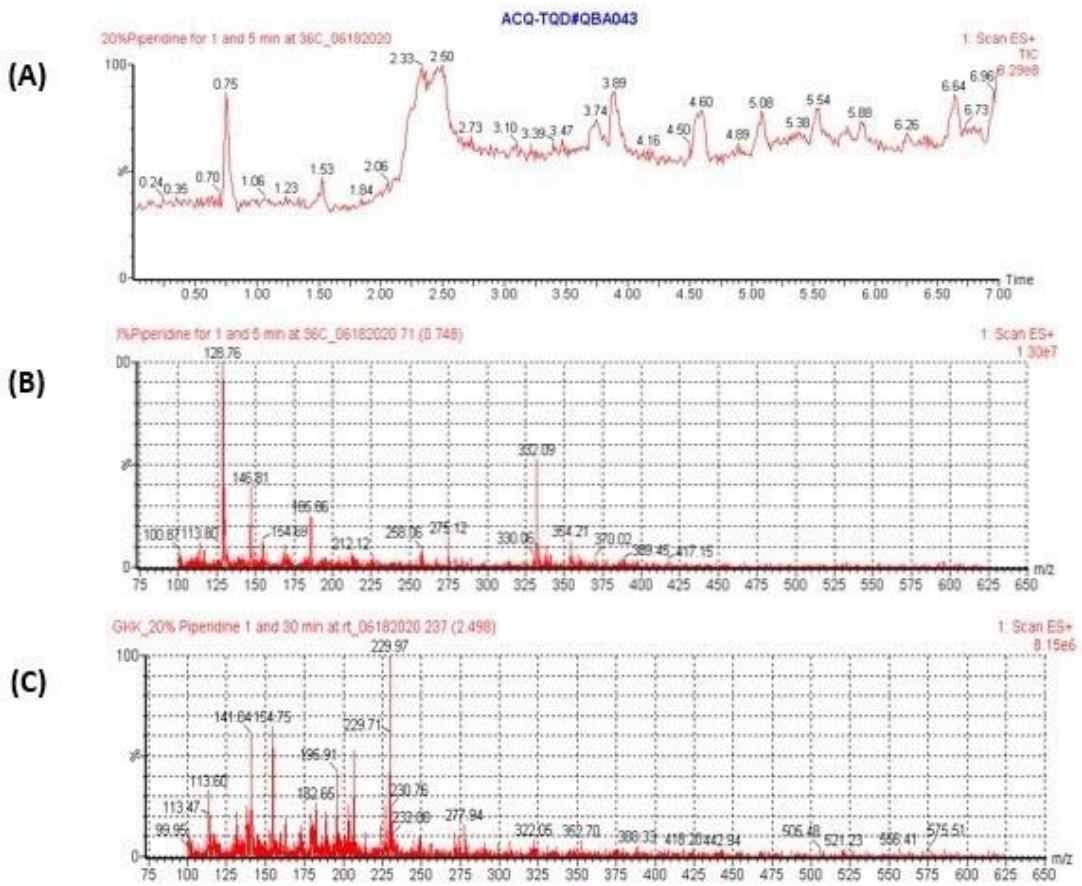


Figure 2.17. (A) UPLC chromatogram of GKK-21. The peak at 0.75 min corresponds to the GKK peptide product with multiple impurity peaks. The peaks eluted at 2.50, 3.89 min and 6.64 min correspond to the Fmoc-protected GKK peptide, KK peptide and a free K amino acid, respectively **(B)** Mass spectrum of GKK-21 corresponding to the GKK peptide eluted at 0.75 min, **(C)** Mass spectrum of GKK-21 corresponding to the Fmoc-protected GKK peptide peak eluted at 2.49 min.

Table 2.12. Peak purities of GKK synthesized (GKK-20 and GKK-21) using piperidine at higher temperature.

Peptide sequence	Peak purity
GKK-20	5.63%
GKK-21	9.77%

2.3.1.3. Multiple incubations with piperidine and diethylamine

The best optimized conditions for the Fmoc-deprotection step employing piperidine and diethylamine were also tested for multiple incubation times to evaluate the best number of incubations for the complete removal of the Fmoc-group. Three trials were performed for piperidine and diethylamine, individually (**Table 2.13** and **Table 2.14**). All three trials with piperidine used 20% piperidine at room temperature without the acetic anhydride step. For the Fmoc-deprotection step, 20% piperidine was incubated 3 times (1 min, 10 min, 10 min) in the first trial, 4 times (1 min, 10 min, 10 min, 10 min) in the second trial while 5 times (1 min, 10 min, 10 min, 10 min, 10 min) in the third trial while all other conditions were kept same as the optimized condition (GKK-12). Similarly, the best optimized condition with diethylamine was also tested in three trials (GKK-25, GKK-26, and GKK-27). The UPLC-MS results for the three trials with piperidine and diethylamine showed that all six peptides did not have Fmoc-protected GKK peptide peak. All six peptides showed that the intensity of the KK fragment peptide peak significantly reduced going from 3 incubations to 5 incubations (**Figure S15**, **Figure 2.18**, **Figure S16**, **Figure S17**, **Figure 2.19,** and **Figure S18**) which indicates that each newly added portion of the base was participating in efficiently removing Fmoc-group, so the chances of having shorter fragments (KK) were eliminated. The GKK peptide peak was the base peak for the peptides with 4 and 5 incubations of the piperidine solution (**Figure 2.18** and **Figure S16**). In case of diethylamine, the deprotection with 4 incubations gave higher intensity peak of the GKK peptide (**Figure 2.19**). The synthesis of all six peptides confirmed that the best optimized conditions both with diethylamine and piperidine gave Fmoc-deprotected peptide as the major product. Some unwanted impurities were also observed along with the desired product in all (six) cases.

Table 2.13. Summary of GKK synthesis using piperidine with multiple incubations (more than two).

Peptide sequence	Reagent for the Fmoc-deprotection	Incubation temperature	Incubation time	Reagent for removing amine impurities
GKK-22	20% Piperidine in DMF	RT	1 min, 10 min, 10 min (3 incubations)	None
GKK-23	20% Piperidine in DMF	RT	1 min, 10 min, 10 min, 10 min (4 incubations)	None
GKK-24	20% Piperidine in DMF	RT	1 min, 10 min, 10 min, 10 min, 10 min (5 incubations)	None

Table 2.14. Summary of GKK synthesis using diethylamine with multiple incubations (more than two).

Peptide sequence	Reagent for the Fmoc-deprotection	Incubation temperature	Incubation time	Reagent for removing amine impurities
GKK-25	30% Diethylamine in DMF	RT	1 min, 10 min, 10 min	4% Acetic anhydride

			(3 incubations)	
GKK-26	30% Diethylamine in DMF	RT	1 min, 10 min, 10 min, 10 min (4 incubations)	4% Acetic anhydride
GKK-27	30% Diethylamine in DMF	RT	1 min, 10 min, 10 min, 10 min, 10 min (5 incubations)	4% Acetic anhydride

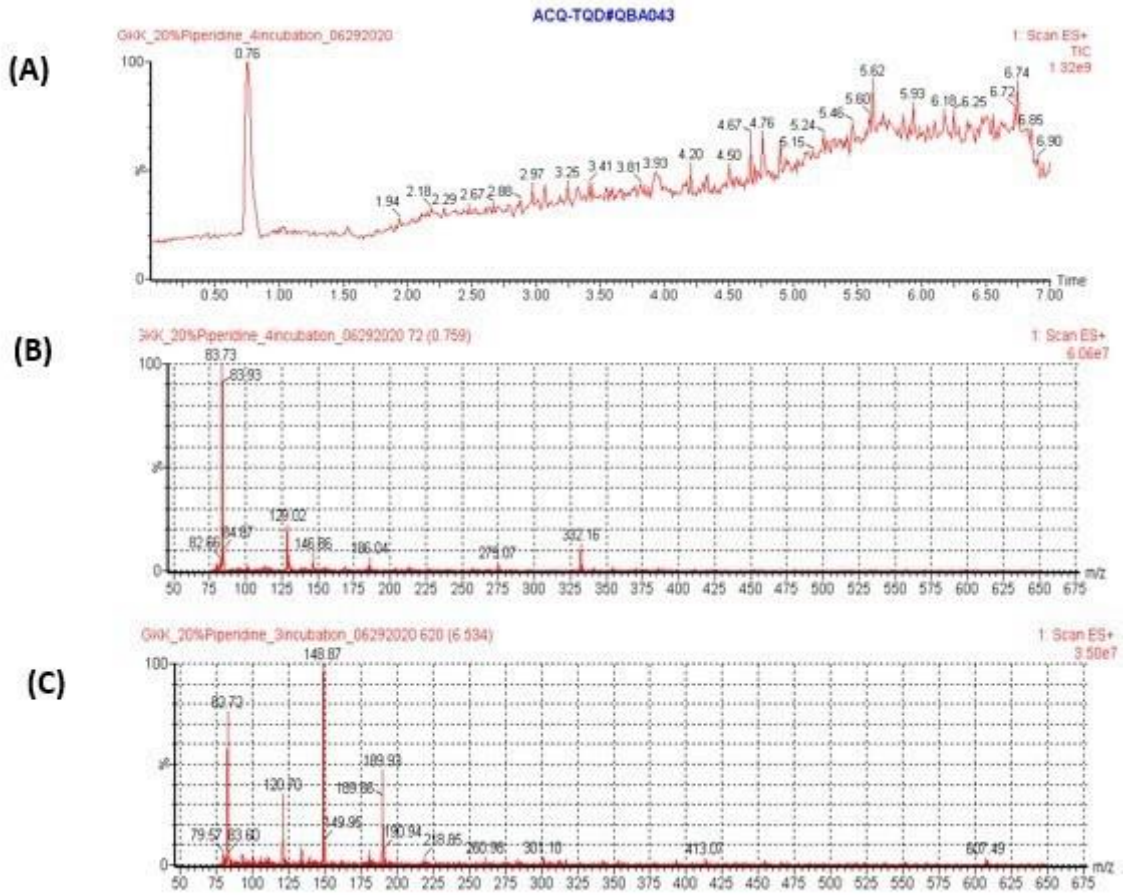


Figure 2.18. (A) UPLC chromatogram of GKK-23. The peak at 0.76 min corresponds to the GKK peptide product with multiple impurity peaks. The peaks eluted at 3.93 min and 6.53 correspond to the KK peptide and a free K amino acid, respectively **(B)** Mass spectrum of GKK-23 corresponding to the GKK peptide peak at 0.76 min, **(C)** Mass spectrum of GKK-23 corresponding to the impurity peak at 6.53 min (M+1 mass peak for free lysine).

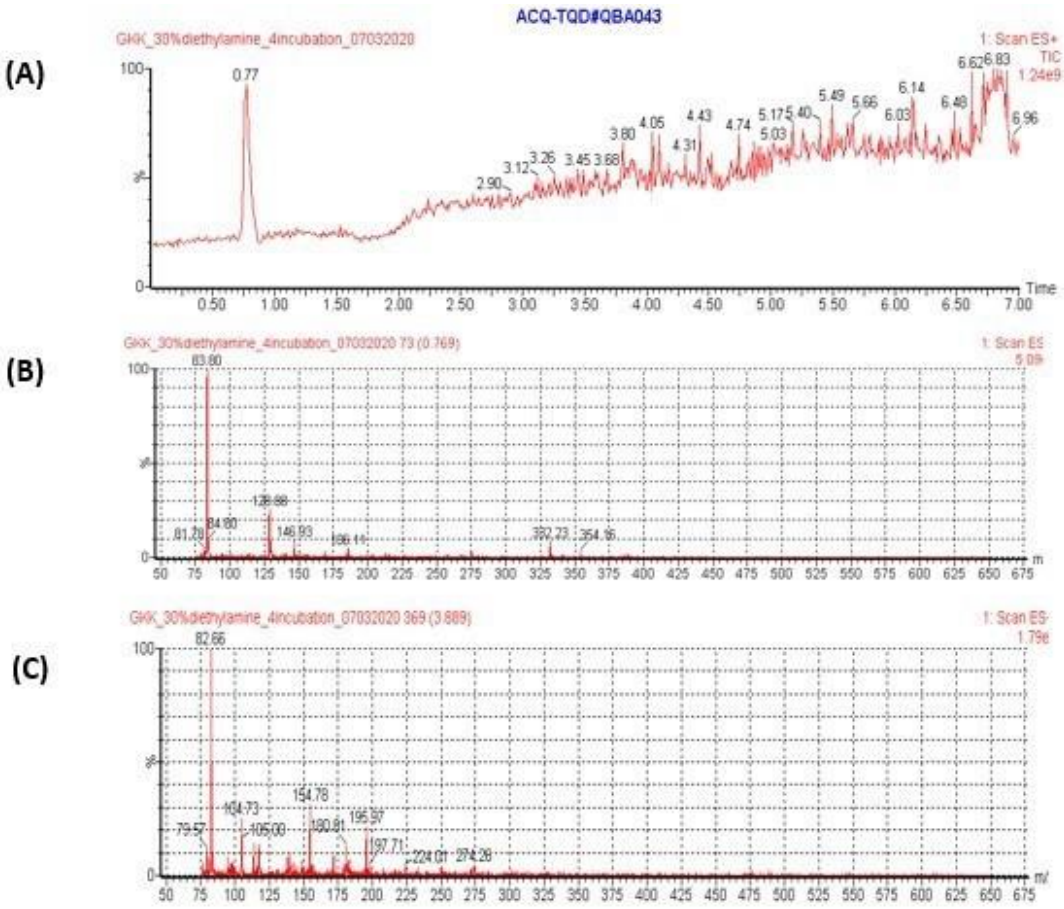


Figure 2.19. (A) UPLC chromatogram of GKK-26. The peak at 0.77 min corresponds to the GKK peptide product with multiple impurities. The peak eluted at 3.88 min corresponds to the KK peptide, **(B)** Mass spectrum of GKK-26 corresponding to the GKK peptide eluted at 0.77 min, **(C)** Mass spectrum of GKK-26 corresponding to the KK fragment peak at 3.89 min.

Table 2.15. Peak purities of GKK synthesized (GKK-22 to GKK-27) using piperidine and diethylamine with multiple incubations (more than two).

Peptide sequence	Peak purity
GKK-22	16.10%
GKK-23	38.79%
GKK-24	37.78%
GKK-25	14.96%
GKK-26	26.77%

GKK-27	27.97%

2.3.1.4. Piperidine for the synthesis of longer peptide chains

The MMP2 consensus sequence was also synthesized by the removal of the Fmoc-group with the best optimal conditions employing piperidine. In the first trial, the peptide sequence was synthesized by cleaving the Fmoc-group with 20% piperidine for 1 minute incubation followed by 10 minutes incubation, at room temperature. The acetic anhydride step was performed for the removal of NH_2-impurities. The UPLC-MS results of the complete MMP2 consensus peptide sequence showed only two peaks in the chromatogram (**Figure 2.20**). The major peak corresponds to the impurity fragment while the second highest peak corresponds to the mass of the desired peptide (1158 g/mol) in the mass spectrum. There were no peaks for the Fmoc-protected peptide in the product mixture. The peak purity of the peptide peak was found to be 32.75%.

The second trial was performed by employing the same conditions as trail # 1 but without the acetic anhydride step. The UPLC-MS results showed a peak of the desired peptide along with very few impurities (**Figure 2.21**). The second trial showed improved intensity of the desired MMP2 peptide peak relative to the major impurity or the base peak with 54.26% peak purity.

To eliminate the impurity caused by cysteine, the third trial was performed with the same peptide sequence but with an alanine amino acid instead of a cysteine amino acid at the second position in the MMP2 peptide sequence. The conditions employed were same as the trial # 2. The UPLC-MS results showed a peak of the desired peptide mass (1069 g/mol) along with very few impurities (**Figure 2.22**). The third trial showed improved purity of 82.86% for the desired MMP2 peptide peak.

Table 2.16. Summary of MMP2 synthesis (with piperidine).

Trial#	Reagent for the Fmoc-deprotection	Incubation temperature	Incubation time	Reagent for removing amine impurities
Trial 1	20% Piperidine	RT	1 min, 10 min	4% Acetic anhydride in Acetonitrile
Trial 2	20% Piperidine	RT	1 min, 10 min	None
Trial 3	20% Piperidine	RT	1 min, 10 min	None

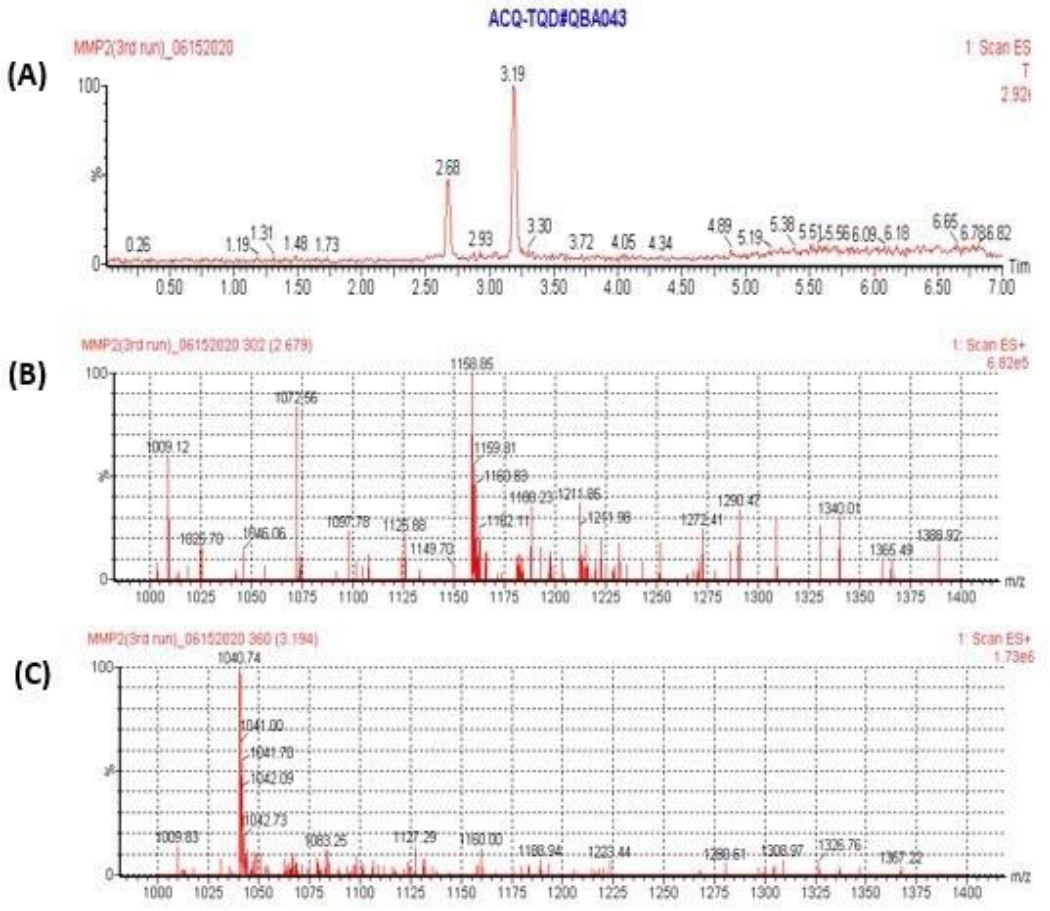


Figure 2.20. (A) UPLC chromatogram of MMP2 (Trial#1). The peak at 2.68 min corresponds to MMP2 peptide product and the peak at 3.19 min corresponds to the impurity (peptide sequence without a cysteine and a water molecule), **(B)** Mass spectrum of MMP2 corresponding to the MMP2 peptide peak eluted at 2.68 min, **(C)** Mass spectrum of MMP2 corresponding to the

impurity peak (peptide fragment without a cysteine and a water molecule, 1040 m/z) eluted at 3.19 min.

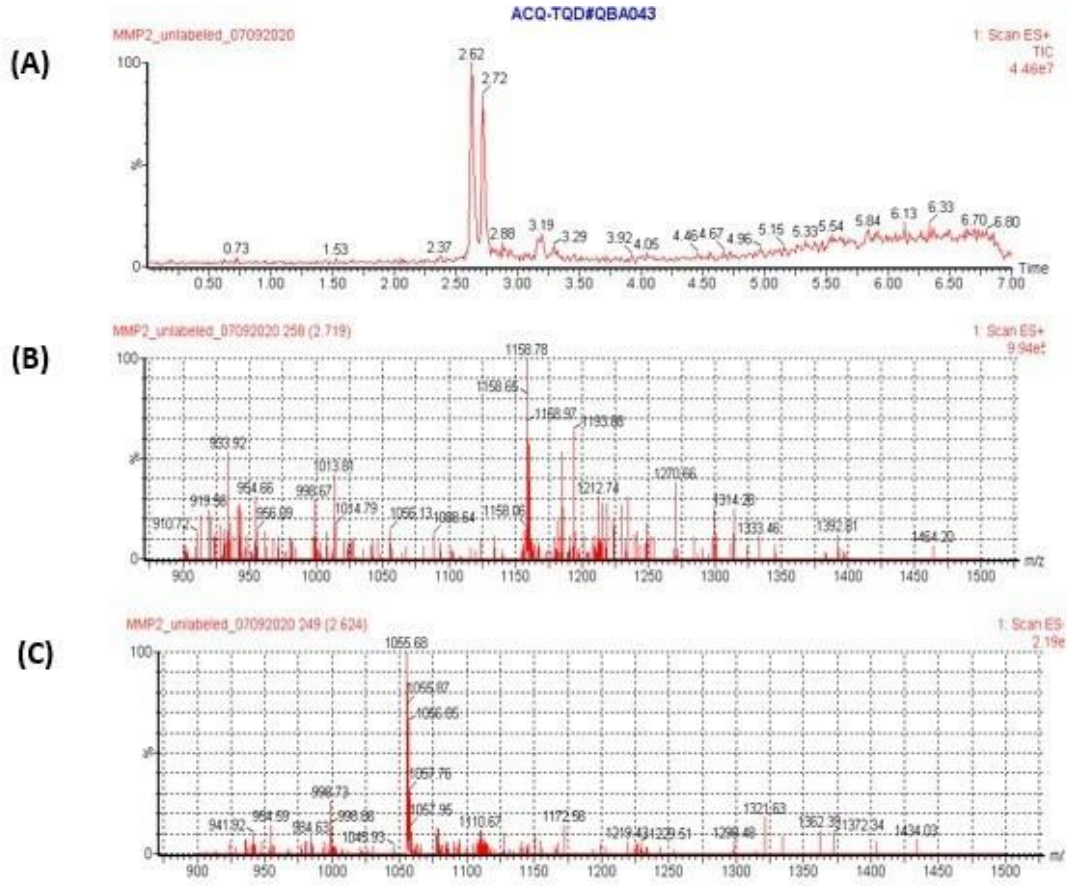


Figure 2.21. (A) UPLC chromatogram of MMP2 (Trial#2). The major peak at 2.62 min corresponds to the MMP2 peptide product and the peak at 2.72 min corresponds to the peptide sequence without a cysteine, **(B)** Mass spectrum of MMP2 corresponding to the M+1 mass peak of MMP2 peptide (1159 m/z) eluted at 2.72 min, **(C)** Mass spectrum of MMP2 corresponding to the M+1 mass of the impurity peak (peptide without a cysteine, 1056) at 2.62 min.

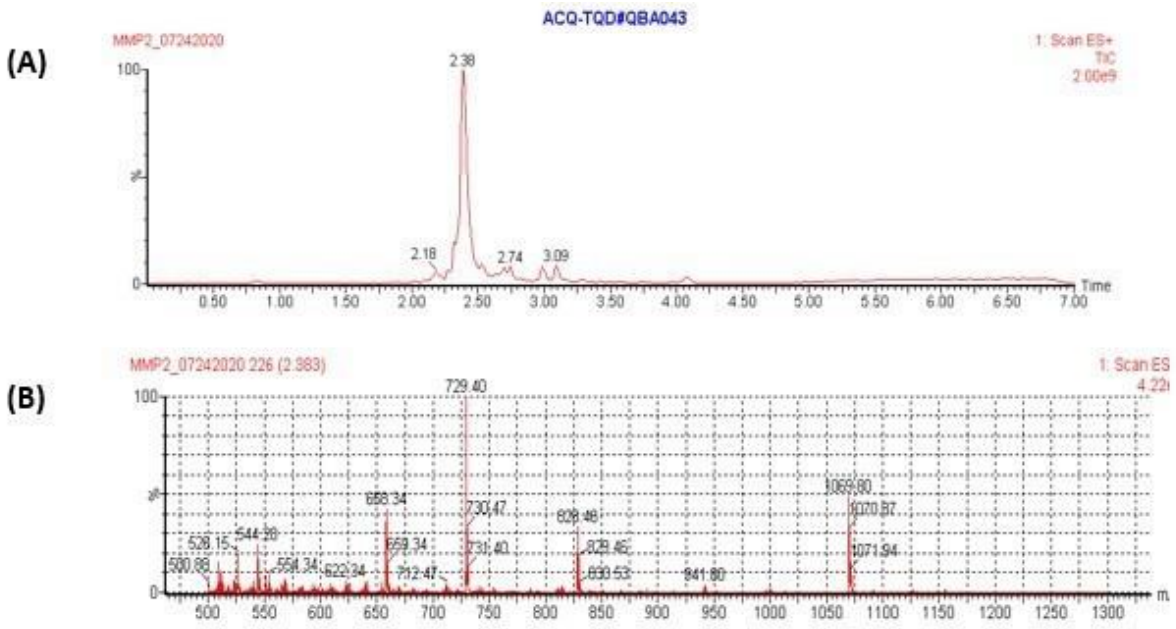


Figure 2.22. (A) UPLC chromatogram of MMP2 (Trial#3). The major peak at 2.38 min corresponds to the MMP2 peptide product with minor impurities, **(B)** Mass spectrum of MMP2 corresponding to the M+1 mass peak of MMP2 peptide (1170 m/z) eluted at 2.38 min.

2.4. Conclusion

The original GKK peptide that was synthesized by using standard SPPS protocol with an additional step of acetic anhydride addition gave a major peak corresponding to the Fmoc-protected peptide and a minor peak corresponding to the desired peptide. The results from the optimization of experimental conditions suggested that diethylamine (in DMF) gave better removal of the Fmoc-group for GKK-2, GKK-3, GKK-6, and GKK-7 as these peptides showed major peptide peaks with fewer impurities. Though the peak for the Fmoc-protected peptide fragment was present, it showed very low intensity compared to the original GKK peptide. Hence, the use of high temperature (36°C) with 20% diethylamine, higher concentration (30% diethylamine in DMF), longer incubation time (30 min) of 20% diethylamine (in DMF) along with the additional step of using 4% acetic anhydride (in acetonitrile) and 20% piperidine without 4% acetic anhydride step turned out to be the most suitable conditions for the Fmoc-

deprotection giving a major peptide product. The absence of the peak for the KK-peptide fragment (GKK-3) suggested that the 30% diethylamine is the best choice for the maximum removal of the Fmoc protecting group along with the least probable impurities. The optimization with 20% piperidine (in DMF), 1 min incubation followed by 10 minutes incubation at room temperature, showed best results as the peak for the Fmoc-protected peptide completely disappeared when the additional acetic anhydride step was not performed (GKK-12). Optimizing the deprotection conditions led to the synthesis of very pure MMP2 consensus sequence (**Figure 2.22**). The synthesis of the relatively long and complex MMP2 consensus peptide sequence employing the optimal conditions with piperidine proved that there was no Fmoc-protected peptide peak. The optimized conditions (both with diethylamine and piperidine) are not only suitable for the maximum removal of the Fmoc group for the peptide chains growing on the polystyrene-based lysine beads (resin) but also for many other resin types and amino acids with the potential Fmoc-deprotection issues. Overall, this study reported optimization of the experimental conditions, both with diethylamine and piperidine, to maximize the Fmoc-deprotection to get the higher purity of the shorter as well as the longer and more complex peptide sequences.

2.5. References

(1) Mäde, V.; Els-Heindl, S.; Beck-Sickinger, A. G. Automated Solid-Phase Peptide Synthesis to Obtain Therapeutic Peptides. *Beilstein J Org Chem* **2014**, *10*, 1197–1212. https://doi.org/10.3762/bjoc.10.118.

(2) Covarrubias-Zambrano, O.; Shrestha, T. B.; Pyle, M.; Montes-Gonzalez, M.; Troyer, D. L.; Bossmann, S. H. Development of a Gene Delivery System Composed of a Cell-Penetrating Peptide and a Nontoxic Polymer. *ACS Appl. Bio Mater.* **2020**, *3* (11), 7418–7427. https://doi.org/10.1021/acsabm.0c00561.

(3) Fosgerau, K.; Hoffmann, T. Peptide Therapeutics: Current Status and Future Directions. *Drug Discovery Today* **2015**, *20* (1), 122–128. https://doi.org/10.1016/j.drudis.2014.10.003.

(4) Lau, J. L.; Dunn, M. K. Therapeutic Peptides: Historical Perspectives, Current Development Trends, and Future Directions. *Bioorganic & Medicinal Chemistry* **2018**, *26* (10), 2700–2707. https://doi.org/10.1016/j.bmc.2017.06.052.

(5) Merrifield, R. B. **Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide**. *J. Am. Chem. Soc.* **1963**, *85* (14), 2149–2154. https://doi.org/10.1021/ja00897a025.

(6) Pedersen, S. L.; Tofteng, A. P.; Malik, L.; Jensen, K. J. Microwave Heating in Solid-Phase Peptide Synthesis. *Chem. Soc. Rev.* **2012**, *41* (5), 1826–1844. https://doi.org/10.1039/C1CS15214A.

(7) Lopez, M. J.; Mohiuddin, S. S. Biochemistry, Essential Amino Acids. In *StatPearls*; StatPearls Publishing: Treasure Island (FL), 2021.

(8) Yuan, J.; O'Donoghue, P.; Ambrogelly, A.; Gundllapalli, S.; Sherrer, R. L.; Palioura, S.; Simonović, M.; Söll, D. Distinct Genetic Code Expansion Strategies for Selenocysteine and Pyrrolysine Are Reflected in Different Aminoacyl-TRNA Formation Systems. *FEBS Lett* **2010**, *584* (2), 342–349. https://doi.org/10.1016/j.febslet.2009.11.005.

(9) Hao, B.; Gong, W.; Ferguson, T. K.; James, C. M.; Krzycki, J. A.; Chan, M. K. A New UAG-Encoded Residue in the Structure of a Methanogen Methyltransferase. *Science* **2002**, *296* (5572), 1462–1466. https://doi.org/10.1126/science.1069556.

(10) *Advances in Fmoc solid-phase peptide synthesis*. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4745034/ (accessed 2021-08-10).

(11) Meienhofer, J.; Waki, M.; Heimer, E. P.; Lambros, T. J.; Makofske, R. C.; Chang, C. D. Solid Phase Synthesis without Repetitive Acidolysis. Preparation of Leucyl-Alanyl-Glycyl-Valine Using 9-Fluorenylmethyloxycarbonylamino Acids. *Int J Pept Protein Res* **1979**, *13* (1), 35–42.

(12) Mendoza, V. L.; Vachet, R. W. Probing Protein Structure by Amino Acid-Specific Covalent Labeling and Mass Spectrometry. *Mass Spectrometry Reviews* **2009**, *28* (5), 785–815. https://doi.org/10.1002/mas.20203.

(13) Albericio, F.; Lloyd-Williams, P.; Giralt, E. [15] Convergent Solid-Phase Peptide Synthesis. In *Methods in Enzymology*; Solid-Phase Peptide Synthesis; Academic Press, 1997; Vol. 289, pp 313–336. https://doi.org/10.1016/S0076-6879(97)89054-4.

(14) Muir, T. W.; Dawson, P. E.; Kent, S. B. H. [13] Protein Synthesis by Chemical Ligation of Unprotected Peptides in Aqueous Solution. In *Methods in Enzymology*; Solid-Phase Peptide Synthesis; Academic Press, 1997; Vol. 289, pp 266–298. https://doi.org/10.1016/S0076-6879(97)89052-0.

(15) *Amino Acid-Protecting Groups | Chemical Reviews*. https://pubs.acs.org/doi/abs/10.1021/cr800323s (accessed 2021-08-10).

(16) Duro-Castano, A.; Conejos-Sánchez, I.; Vicent, M. J. Peptide-Based Polymer Therapeutics. *Polymers* **2014**, *6* (2), 515–551. https://doi.org/10.3390/polym6020515.

(17) Fields, G. B. Methods for Removing the Fmoc Group. In *Peptide Synthesis Protocols*; Pennington, M. W., Dunn, B. M., Eds.; Methods in Molecular Biology; Humana Press: Totowa, NJ, 1995; pp 17–27. https://doi.org/10.1385/0-89603-273-6:17.

(18) Wampenseppl. *English: Fmoc Protecting Group Cleavage with Piperidine Showing E1cB Mechanism.*; 2011.

(19) Bodanszky, M. Solid Phase Peptide Synthesis. In *Peptide Chemistry: A Practical Textbook*; Bodanszky, M., Ed.; Springer: Berlin, Heidelberg, 1988; pp 147–168. https://doi.org/10.1007/978-3-642-97886-9_10.

(20) Adhikary, R.; Dawson, P. E. In Situ Neutralization Protocols for Boc-SPPS. In *Peptide Synthesis: Methods and Protocols*; Hussein, W. M., Skwarczynski, M., Toth, I., Eds.; Methods in Molecular Biology; Springer US: New York, NY, 2020; pp 29–40. https://doi.org/10.1007/978-1-0716-0227-0_3.

(21) Wiefel, L.; Bachmann, F.; Terwort, J.; Steinbüchel, A. In Vitro Modification of Bacterial Cyanophycin and Cyanophycin Dipeptides Using Chemical Agents Towards Novel Variants of the Biopolymer. *Earth Syst Environ* **2019**, *3* (3), 637–650. https://doi.org/10.1007/s41748-019-00107-y.

(22) Rawlings, N. D.; Barrett, A. J.; Thomas, P. D.; Huang, X.; Bateman, A.; Finn, R. D. The MEROPS Database of Proteolytic Enzymes, Their Substrates and Inhibitors in 2017 and a Comparison with Peptidases in the PANTHER Database. *Nucleic Acids Research* **2018**, *46* (D1), D624–D632. https://doi.org/10.1093/nar/gkx1134.

(23) Gioia, M. L. D.; Costanzo, P.; Nino, A. D.; Maiuolo, L.; Nardi, M.; Olivito, F.; Procopio, A. Simple and Efficient Fmoc Removal in Ionic Liquid. *RSC Adv.* **2017**, *7* (58), 36482–36491. https://doi.org/10.1039/C7RA04425A.

(24) Orain, D.; Ellard, J.; Bradley, M. Protecting Groups in Solid-Phase Organic Synthesis. *J. Comb. Chem.* **2002**, *4* (1), 1–16. https://doi.org/10.1021/cc0001093.

(25) Feinberg, R. S.; Merrifield, R. B. Modification of Peptides Containing Glutamic Acid by Hydrogen Fluoride-Anisole Mixtures. .Gamma.-Acylation of Anisole or the Glutamyl Nitrogen. *J. Am. Chem. Soc.* **1975**, *97* (12), 3485–3496. https://doi.org/10.1021/ja00845a034.

Chapter 3 - Developing peptide-based nanoplatform for lung cancer diagnosis

Abstract

By accounting for about 25% of all cancer deaths in the US, lung cancer results in the most cancer-related fatalities. Because the survival rate increases when a tumor is discovered early, a significant effort has been made to identify strategies for the early detection of cancers. To measure the lung cancer-related biomarkers, such as proteases, this research aims to develop a peptide-based detection platform for the clinical diagnosis of lung cancer at its early stages. Hence, we report a rapid and reliable biosensor design that consists of a peptide sequence labeled with tetrakis-carboxyphenyl-porphyrin (TCPP) and explosion-synthesized graphene core-shell attached to polyethylenimine (PEI, to link the peptide via an amide bond). These consensus peptide sequences act as substrates for the enzymes/proteases that are overexpressed in lung cancer, hence, measuring the activity of lung cancer protease biomarkers. This detection system works on the principle of Förster resonance energy transfer, where the cleaving of consensus peptide sequence by relevant proteases results in the termination of the quenching effect between graphene and TCPP. The increase in the fluorescence intensity of TCPP upon the termination of quenching is measured using a plate reader. A total of 19 nanobiosensors were developed for lung cancer detection. These biosensors can detect the overly expressed biomarkers using a conventional plate reader after 1 hour of incubation down to the sub-femtomolar level. Furthermore, this biosensor can be easily used in laboratories with limited resources.

3.1. Introduction

Lung cancer is the deadliest cancer as it is responsible for the most cancer-related deaths in women and men.[1] It is also the most common cancer in men and women, making up to 12.4 % of all cancer types after prostate and breast cancer which are 13.1 % and 14.8 %, respectively.[2] The American Cancer Society estimated 236,740 new lung cancer cases (117,910 in men and 118,830 in women) with approximately 130,180 deaths from lung cancer (68,820 in men and 61,360 in women), in the year 2022. Hence, there is a likelihood of developing lung cancer in every 1 in 15 males and 1 in 17 females during the lifetime.[3] Primary lung cancer starts in the lungs, whereas secondary lung cancer spreads to the lungs from somewhere else in the body. There are two types of primary lung cancers: small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC).[4] Among these two types, SCLC is a less common type of lung cancer causing approximately 13 % of all lung cancer cases. SCLC are classed as neuroendocrine tumors (NETs, tumors that develop in the neuroendocrine system). SCLC are usually fast-spreading cancers and are directly associated with smoking. Around 84 % of lung cancer cases are NSCLC and can be caused by not only smoking but exposure to secondary smoking, asbestos, radon, air pollution and family history of lung cancer.[3,5] These cancers progress slowly and are subdivided into three types: adenocarcinoma, squamous cell carcinoma, and large cell carcinoma.[3] Adenocarcinoma is the most common type of NSCLC and develops in the mucus-making gland cells lining the airways. Squamous cell cancer is originated in the flat cells covering the surface of the airways. They tend to grow near the center of the lung. In large cell carcinoma, cancer cells appear large and round when analyzed under the microscope.[4] Both SCLC and NSCLC differ in prognosis and treatment, but all three types of NSCLC are grouped together because of similar behavior and response to treatment.[3,4] The five-year relative survival

rates of patients diagnosed with NSCLC between 2011 and 2017 suggested 64% survival for localized NSCLC, 37% for regional NSCLC, and 8% for distant NSCLC. The numbers estimated are much lower in the case of SCLC patients as the five-year relative survival rates are found to be 29% for localized, 18% for regional, and 3% for distant SCLC.[6]

Early detection is very crucial as it increases the chances of survival. Detection of lung cancer at stage I increases the likelihood of successful and less painful treatment for the patient. The available diagnostic tests for lung cancer diagnosis include x-rays, low-dose computed tomography (LDCT) scan, magnetic resonance imaging (MRI) and positron emission tomography (PET), and tissue biopsy.[7] These tools require a well-equipped laboratory setting and trained personnel to run the diagnostic tests. Detecting lung cancer at early stages is not always possible due to lack of clear symptoms, and detectable lung cancer cases comprise only small percentage of all lung cancers.[3,7] Hence, there is an urgent need for less invasive and efficient detection method with the potential to enable the detection of lung cancer at early stage with high sensitivity using liquid biological samples (liquid biopsy).

3.1.1. Biomarkers for cancer detection

Biomarkers (also called molecular markers and signature molecules) are biological molecules that indicate normal behavior or abnormality in a physiological process or condition. These biomarkers are found in blood, other body fluids, or tissues. Biomarkers not only indicate a change or abnormality in a biological system but may also indicate the body's response to a specific treatment for a disease.[8] Owing to the severity of lung cancer disease condition, scientists have been searching for suitable biomarkers that may be used to indicate the onset as well as different stages of lung cancer. In this regard, liquid biopsy is a convenient way to extract several biomarkers at once. The key benefit of liquid biopsies is that no tissue samples are

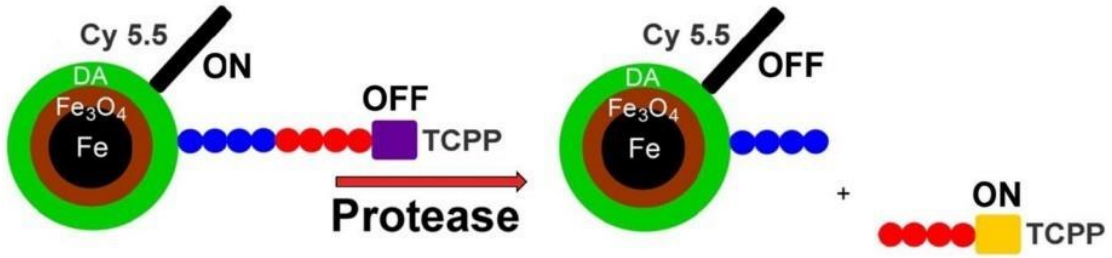
needed to be taken, reducing patient discomfort and allowing for repeated tests over time to track the development of the disease.[9] There is a wide range of biomarkers, including proteins (such as an enzyme or receptor), nucleic acids (such as a microRNA or other non-coding RNA), antibodies, and peptides. A collection of alterations can also be considered a biomarker, such as gene expression, proteomic, and metabolomic signatures.[10] Proteases (class of enzymes that hydrolyze protein) are the largest enzyme gene family in vertebrates.[11] Healthy and diseased cells contain and secrete a vast number of proteases that could be used as a biological signature for healthy or diseased conditions, such as cathepsins (CTSs), urokinase plasminogen activator (uPA), and matrix metalloproteinases (MMPs).[12] An estimated 3% of the human genome (641 genes) comprise the proteases genes that encode different proteases responsible for various physiological functions.[11] In particular, recognition of the connection between proteases and cancer may be traced back to the 1940s, when Fisher postulated that tumor-associated proteolytic activity might be to blame for the breakdown of the cellular matrix and the ensuing invasion of the tumor into the surrounding normal tissue. Then in the 1970s, individual proteases with secreted members of cysteine and serine proteases and metalloproteinase families were identified.[13] Owing to the ability to disintegrate extracellular membrane, these proteases participate in tumor growth, invasion, angiogenesis and cancer metastasis. An example is matrix metalloproteinases (MMPs), which disintegrate the extracellular matrix. The extracellular matrix of cells acts as a physical barrier for cancer cells, preventing them from entering blood or lymphatic vessels to facilitate metastasis. As a result, the effective invasion of tumor cells is facilitated by uncontrolled extracellular matrix breakdown caused by enhanced MMP activity. Additionally, uncontrolled modification of the extracellular matrix makes it easier for angiogenesis to occur, which promotes tumor development and spread.[14] Similarly, the

importance of proteases in lung health and illness cannot be overstated. In the lungs, proteases perform homeostatic roles, which control procedures like regeneration and repair in healthy lungs. The emergence of several forms of lung disorders depends on the dysregulation of the proteases-antiproteases balance. A considerable increase in protease activities is linked to chronic inflammatory lung diseases. As a result, successful infection and inflammation in the lungs depend on the under-or- over regulation of related proteases.[15] Hence, proteases could potentially be used as a tool to indicate the presence of lung inflammations and tumors, and the extent of levels could be indicative of different stages of cancer.

3.1.2. Platform for lung cancer detection

The Bossmann group has been working hard since 2007 to create peptide-based ultra-sensitive protease detection platforms that can detect protease activities over a broad range down to sub-femtomolar limits of detection (LODs).[12,16,17] Consensus peptide sequences are oligopeptides designed to be quickly cleaved by their specific proteases but not by every other protease. As a result, they exhibit extremely high selectivity towards the protease that they are intended to detect.[12] These nanoplatforms are composed of dopamine-coated iron/iron oxide core-shell nanoparticles conjugated to a consensus peptide sequence (that is protease-cleavable) labeled with a fluorescence dye tetrakis (4- carboxyphenyl) porphyrin (TCPP) via an amide linkage, as well as cyanine 5.5, a second dye that is permanently bonded to the dopamine layer and functions as a TCPP quencher. The working principle is based on the Förster resonance energy transfer (FRET) and plasmon resonance quenching. When a suitable protease is present and the consensus sequence is cleaved, TCPP is released and the fluorescence signal increases, providing a "light switch effect" that allows for the highly sensitive detection of protease activity, which is then measured using a plate reader **(Figure 3.1)**.

(A)



(B)

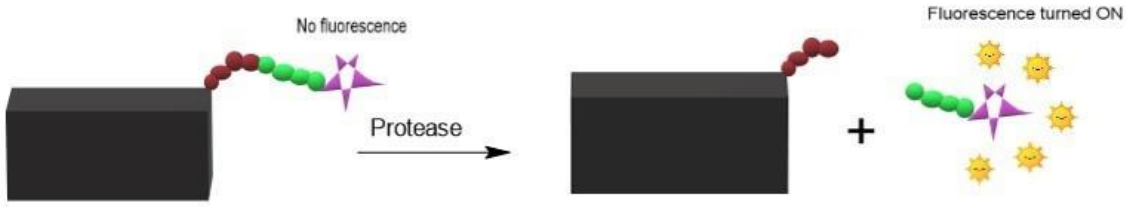


Figure 3.1. Design principles of a nanobiosensor for protease detection **(A)** Iron/iron oxide-peptide detection system[17] (Reprinted with permission from ref 17) **(B)** Graphene-peptide detection system.

Despite the excellent success rates attained, the iron/iron oxide nanobiosensor technique still has certain limitations, such as the cost and short-term stability of the iron/iron oxide nanoparticles, as well as the necessity of using two fluorescent dyes, one for FRET quenching and the other for sensing. To address these issues, we chose long-lasting, highly water/buffer dispersible graphene as the core material that overcomes the shortcomings of the iron/iron oxide nanoparticles. Additionally, by utilizing a graphene-based nanoplatform and taking advantage of graphene's optical absorption property, which enables graphene to behave as a very effective quencher for fluorescent molecules, we have overcome the problem of having to use two dyes.

The study uses detonated carboxygraphene that is linked to polyethylamine (PEI) to provide an anchor for the peptide (labeled with TCPP) conjugation via an amide bond.

3.1.2.1. Graphene

Graphene has exceptional chemical, mechanical, and optical properties that make it a great choice among other drug delivery nanocarrier system candidates. Most important of all the properties is the high surface area that leads to ultra-high drug-loading efficiency.[18] The Sorensen group (Physics at KSU) has developed a one-step, simple, easy, and nontoxic method for the synthesis of graphene from O_2/C_2H_2 detonation at 4000 K without the use of any catalyst. The product forms 1-10mm aerosol gel globules and is very rich in carbon (99.2% carbon, 0. 7% oxygen) with a zeta potential of +60 mV in H_2O. Raman spectra of graphene synthesized with different molar ratios of O_2 and C_2H_2 show two new bands at 1328 and 1610 cm^{-1} along with G and 2D bands at 1580 and 2650 cm^{-1} corresponding to graphene nanoparticles.[19]

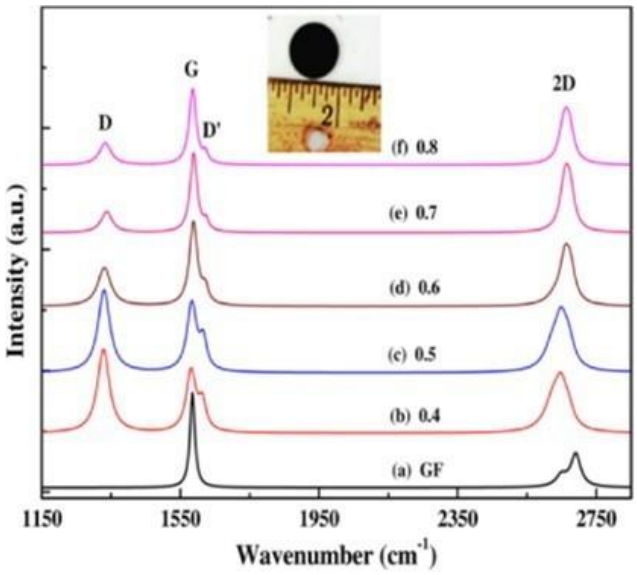


Figure 3.2. Raman spectra of graphite and graphene nanoparticles.[19] (Reprinted with permission from ref 19).

Transmission electron microscopy (TEM) and RAMAN spectroscopic analysis also confirm both the presence of stacked graphene layers (stacking number 7-20) and the fractal

nature of explosion-graphene. UV/Vis (ultraviolet/visible) spectroscopy analysis showed that graphene nanoparticles, as well as derived graphene oxide, can absorb light in the UV region with a high optical extinction value that makes it a great choice for use in photothermal therapy (**Figure 3.3**).

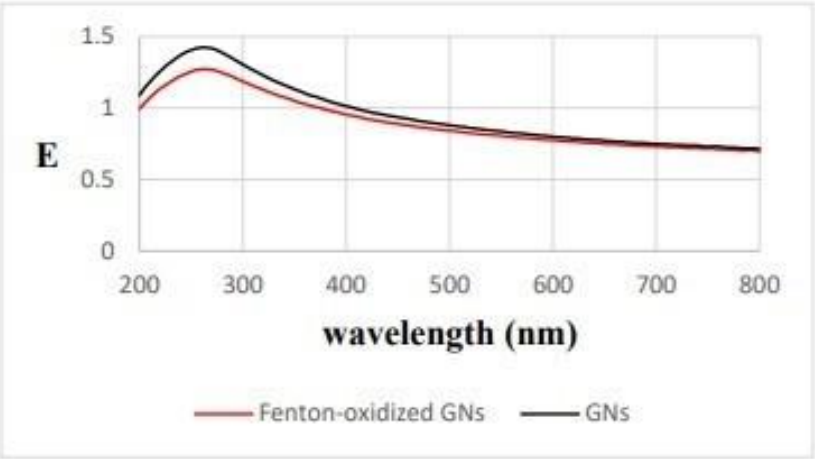


Figure 3.3. UV/Vis spectroscopic analysis of graphene and Fenton-oxidized graphene nanoparticles.[20] (Reprinted with permission from ref 20)

Surface modification of graphene is often needed to enable its usage in different applications and to increase water dispersibility. Hence, the Bossmann group has developed novel methods for the easy surface modifications of explosion-synthesized graphene such as Fenton oxidation and carboxylation. Fenton oxidation involves the use of iron sulfate[20] that may lead to iron toxicities when inside the biological system[21]. To avoid the possible iron toxicity, instead of oxidized graphene (GO), this study will use carboxygraphene. In this regard, method developed for the modification of graphene surface with carboxylic acid functionality in Bossmann lab is very easy to carry out and yields a high surface density of carboxylic acid groups on the surface (4 -COOH groups per nm^2), corresponding to an average space per -COOH group of 0.27 nm^2. Such high labeling density with surface -COOH groups, contributes to much higher water dispersibility for carboxygraphene nanosheets.[20]

3.2. Materials and Methods

3.2.1. Gene expression analysis

Finding the proteases that are overexpressed in solid tumors, such as pancreatic cancer, is simple with the help of gene expression analysis. Databases like NCBI GEO, Entrez Gene ID, Unigene ID, and Gene Symbol provide access to various information.[22]

The NCBI GEO database was used to gather the relevant datasets for the selection of proteases for the detection of SCLC and NSCLC vs. non-cancerous health situations, as well as distinguishing between SCLC and NSCLC.[23] Criteria for datasets included in the analysis were that the investigated species was *Homo sapiens* and the dataset consisted of the samples from both primary tumor samples and healthy human tissue.

3.2.2. Synthesis of peptides

All consensus peptide sequences for developing the lung cancer biosensors were synthesized via our optimized solid phase peptide synthesis (SPPS) conditions **(Scheme 2.1)**. Briefly, a resin containing the first amino acid of the peptide was swollen in DCM for 20 minutes followed by a DMF wash for 1 minute. Meanwhile, the next amino acid (Fmoc-protected) of the peptide chain was activated with HBTU in DIEA:DMF (1:23). After that, the activated amino acid was added to the resin-amino acid. The mixture was swirled for 30 minutes at room temperature followed by another addition of the amino acid mixture. After the addition of new amino acid was completed, the resin-peptide was washed five times with DMF. The next step was to deprotect the amine group that was protected by the Fmoc. To do so, 20% piperidine in DMF was used for 1 minute followed by 10 minutes incubation (both at room temperature). Again, DMF washings were performed to get rid of piperidine. At this point, the next (Fmoc-protected) amino acid was activated and added to the resin having the growing chain of peptide.

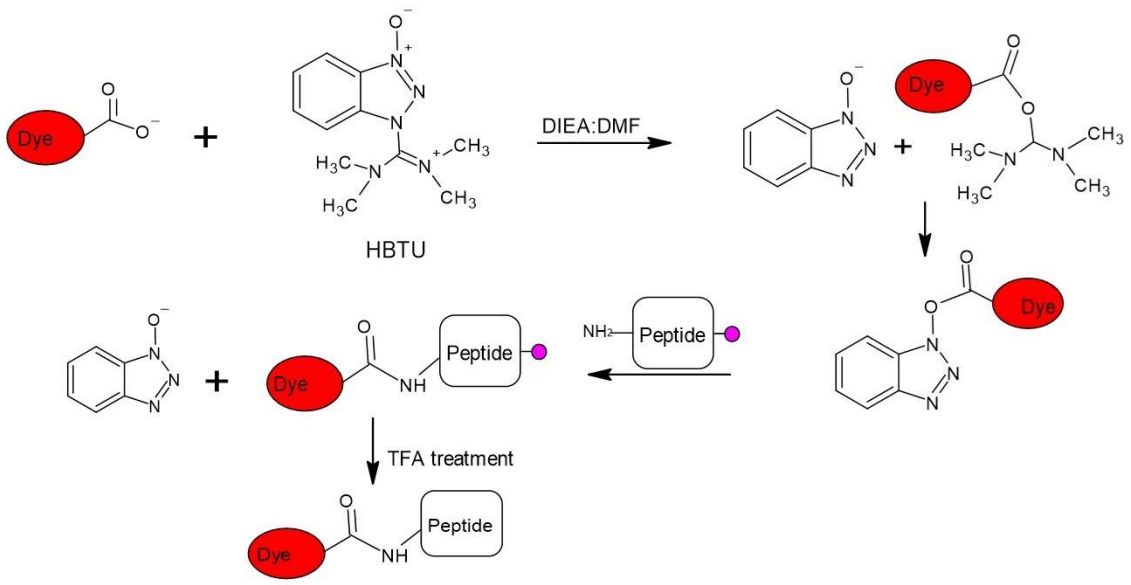
The cycle of Fmoc-deprotection and addition of new amino acid was repeated until the desired amino acid sequence was obtained for all three peptide sequences.

3.2.2.1. Labeling of consensus peptide sequences with the dye TCPP

After the synthesis, peptide sequences were labeled with TCPP dye at the N-terminus of the peptide. For the coupling of dye to peptide chain, the milli-equivalent ratio of the resin to HBTU was 1:2.9, and resin to the dye was 1:3. The carboxylic group of the dye was activated using HBTU in DIEA:DMF (1:23). The dye was incubated for 24 hours with the resin-peptide chain, followed by multiple DMF washings to remove any free dye after the coupling step.

3.2.2.2. Cleaving of the resin

After the successful attachment of the TCPP dye to the peptides, an acid cocktail solution was used to cleave the peptide sequence from the resin. This cocktail solution consisted of 95% TFA, 2.5% TIPS and 2.5% distilled water. The resin along with the peptide chain was incubated in this cocktail for 3 hours. After the completion of the resin cleaving step, the peptide was precipitated and washed with cold ether, followed by drying with Ar gas.



Scheme 3.1. Activation and addition of dye to the peptide and resin cleaving.

3.2.2.3. Analysis of the synthesized peptides

All the samples were prepared in 80% acetonitrile and 20% water v/v to perform UPLC-MS analysis (sample concentration: 0.1 mg/mL) using Waters ACQUITY TQD Tandem Quadrupole UPLC/MS/MS system in ES+ ionization mode using acetonitrile/water gradient. The instrument was equipped with an ACQUITY UPLC peptide CSH C18 column (Waters), 130 Angstrom, 1.7 micrometer 2.1 mm x 150 mm.

3.2.3. Synthesis of carboxygraphene

Detonation-synthesized graphene nanosheets (0.3 O/C, 2g) was suspended in 40 mL DMF at room temperature in a 250 mL three-necked round bottom flask equipped with a magnetic stir bar. First, the temperature of the reaction mixture was increased to 40 °C and 1.0 g of 5-bromo-valeric acid (0.0055 mol) was added to the graphene suspension, followed by the addition of 0.36 g sodium azide crystals (NaN_3) (0.0055 mol). After NaN_3 was fully dissolved, the suspension was slowly heated up to 80 °C (about 1^oC/min.) and stirred for 1 hr. When the

temperature of the mixture reached between 40 °C and 50 °C, a nucleophilic substitution reaction (SN) occurred, in which organic azide and DMF-soluble sodium halogenide was formed. When the temperature reached above 50 °C, the organic azide released nitrogen gas and a nitrene intermediate was formed. The nitrogen of the nitrene intermediate contained an incomplete valance shell so this intermediate underwent a cycloaddition with a pi-pi double bond at the surface of graphene to form a stable azirane anchor (three-membered C-N-C ring). After 1 hour, the suspension was cooled down to room temperature, and carboxygraphene was collected via centrifugation (5 min @ 7,000 rpm) and washed five times with DMF followed by three anhydrous diethyl ether washings. Finally, the resulting product was dried for characterization and then stored under argon.

3.2.4. Loading of peptide

After characterizing the carboxygraphene, the next step was to load the peptide onto the graphene. The first step was to add polyethylenimine groups (PEI) to use later for making an amide bond with the C-terminal of the peptide. To do so, a one-pot reaction was performed by suspending 1.0 g of synthesized carboxygraphene in 30 mL anhydrous DMF (at room temperature) in a 100 mL round bottom flask equipped with a magnetic stir bar. Next, 500 mg of polyethylenimine (PEI), branched, molecular weight 10,000 (PEI), 250 mg of 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC), and 250 mg of 4-dimethylaminopyridine (DMAP) were added to the solution. The reaction mixture was stirred at room temperature overnight. The final product carboxygraphene-polyethylenimine (CG-PEI) was collected via repeated centrifugations (5 min each at 7000 rpm) followed by washings with DMF and diethyl ether. Finally, resulting product was dried under argon.

To attach peptide, 250 mg of carboxygraphene-PEI was suspended in ~8.1 mL DMF in a small glass vial, and the solution was sonicated for one min. Next, 5 mg DMAP and 5 mg EDC were added to the vial followed by 3.5 mg of TCPP-conjugated peptide. The solution was sonicated for about five minutes to ensure the full suspension of all components in DMF. The reaction mixture was then stirred overnight at room temperature. The next day, the product was collected via centrifugation (10 min @ 7000 rpm) followed by three DMF washings and three washings with diethyl ether. After the last washing with diethyl ether, the final product was dried with argon.

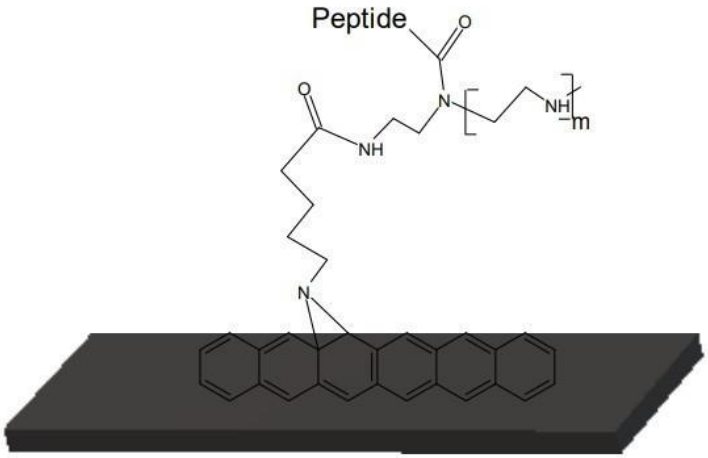


Figure 3.4. Representation of carboxygraphene-polyethylenimine-based platform.

3.2.5. Enzymatic digestion experiments

To run the digestion experiments, commercially available enzymes were used. 25 nM of each peptide (unlabeled) was dissolved in 1.0 mL of a buffer comprised of Tris (50 mM), $CaCl_2$ (10 mM), and NaCl (150 mM), pH=7.5 (adjusted with HCl). The enzyme (specific for the relevant peptide substrate) (100 nM) was dissolved in 1.0 mL of the same buffer. Both solutions were heated to 37 ^{o}C. The reaction started by adding the enzyme solution to the peptide solution. The reaction mixture was kept at 37 ^{o}C for four hours. Samples (25 μl) were taken at 0, 60, 120,

and 240 min. and analyzed using a Waters ACQUITY TQD Tandem Quadrupole UPLC/MS/MS system that was equipped with an ACQUITY UPLC peptide CSH C18 column (Waters), 130 Angstrom, 1.7 micrometer 2.1 mm x 150 mm. An acetonitrile/water gradient was used for eluting the peptides/fragments.

3.2.6. Fluorescence measurements with detonated graphene-biosensors

In accordance with previously published procedures[17,24,25] A BioTek Synergy H1 plate reader with tungsten halogen lamp, analysis bandpass filter: 650 ± 25 nm, excitation bandpass filter: 421 ± 10 nm and 96-well plates was used. Two types of solutions were prepared; solution 1 contained HEPES buffer (25 μmol) (2-[4-(2-hydroxyethyl) piperazin-1-yl]ethanesulfonic acid) that was prepared enriched with Ca(II), Mg(II), and Zn(II) (10 μmol each). The temperature was kept 298 K (pH = 7.2) to ensure the full enzymatic activities. Solution 2 contained the nanobiosensor (0.03 mg of graphene-nanobiosensor) in 1.0 mL of HEPES buffer and was sonicated for 10 min at 298 K. Sample control, assay, assay control and blank were prepared using solution 1 and/or solution 2 to 5 μL of serum sample as follows; Sample Control contained 125 μL of solution 1 + 5 μL serum sample. Assay contained 125 μL of solution 2 + 5 μL of serum sample. Assay Control contained 125 μL of solution 2 + 5 μL of solution 1. Blank contained 130 μL of solution 1. Each sample was loaded (130 μL) into one well of a black 96-well plate and incubated for 90 min. at 37ºC.

3.3. Results and Discussion

For Fe/Fe_3O_4 core/shell-based nanobiosensors the early detection of pancreatic cancer at stages 0 and 1[16], and of NSCLC at stage 1[24] was possible by means of Liquid Biopsies (serum), because of distinctly changes of their protease signatures during the transition from non-

cancerous to cancerous. The data collected for some proteases (out of 19 selected) has been shown in **Figure 3.5**, **3.6** and **3.7**.

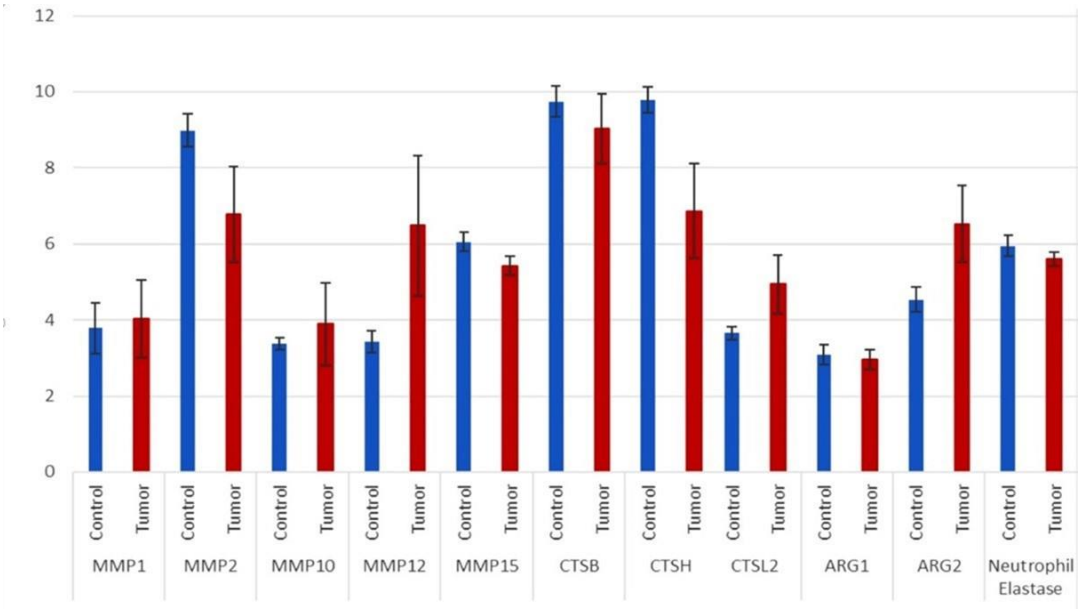


Figure 3.5. Gene expression differences (displayed as natural logarithm on y-axis) between control tissue (non-cancerous lung tissue cells) and cancerous tissue (NSCLC, on x-axis).[23] The y-axis is normalized to natural logarithm of mRNA expression for the respective enzyme found in tumor tissues before surgery. The error bars indicate the deviation of mRNA's for the proteases found in tissue.

Principally, the same selection of proteases/arginase could also differentiate between SCLC and healthy tissue, as shown in **Figure 3.6**.

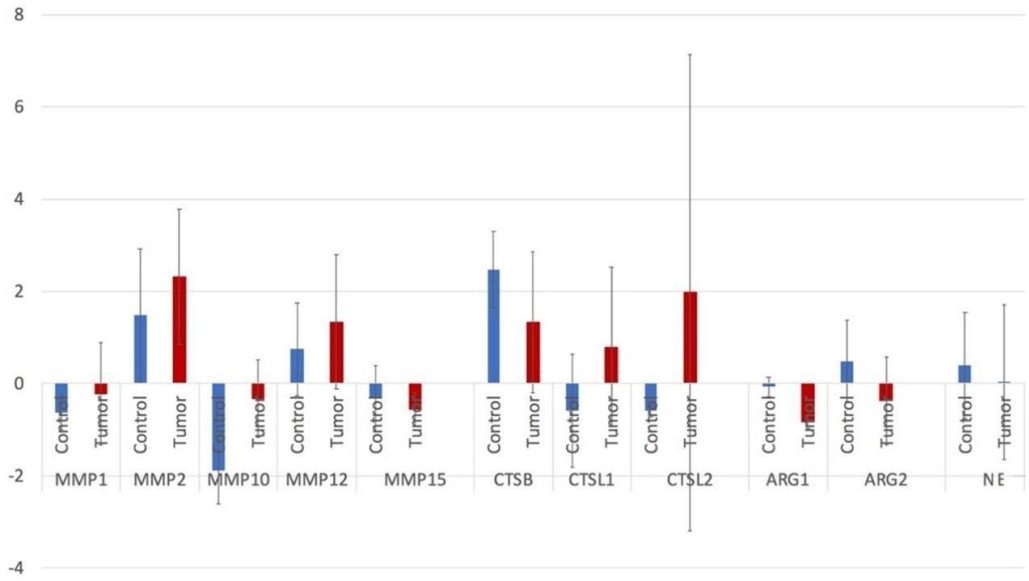


Figure 3.6. Gene expression differences (displayed as natural logarithm on y-axis) between control tissue (non-cancerous lung tissue cells) and cancerous tissue (SCLC, on x-axis).[23] The y-axis is normalized to natural logarithm of comparative mRNA expression for the respective enzyme found in tumor tissues before surgery. CTSL1 and CTSL2 are isoforms of cathepsin L. Gene splicing is very common in the tumor environment.[26] The error bars indicate the deviation of mRNA's for the proteases found in tissue.

The computed differences in gene expression between SCLC and NSCLC are shown in **Figure 3.7**. Due to significant differences in the gene expression levels of SCLC versus NSCLC, MMP10 and CTSH were found suitable to discriminate between SCLC and NSCLC. Differences in the expression levels of NSCLC versus control make them promising biomarkers for lung cancer detection (especially NSCLC, **Figure 3.5**) as well. Among other analyzed proteases, the candidates with significant differences in the gene expression levels of cancerous versus non-cancerous were selected (**Figure 3.5** and **Figure 3.6**) to differentiate between the cancerous lung condition (NSCLC and SCLC) and non-cancerous cells. An example is neutrophil elastase which shows altered expression in both types of lung cancers compared to the control cells (NSCLC and SCLC, **Figure 3.5** and **Figure 3.6**); hence it will be used as an example in the later sections (characterization and enzymatic digestion) of this chapter.

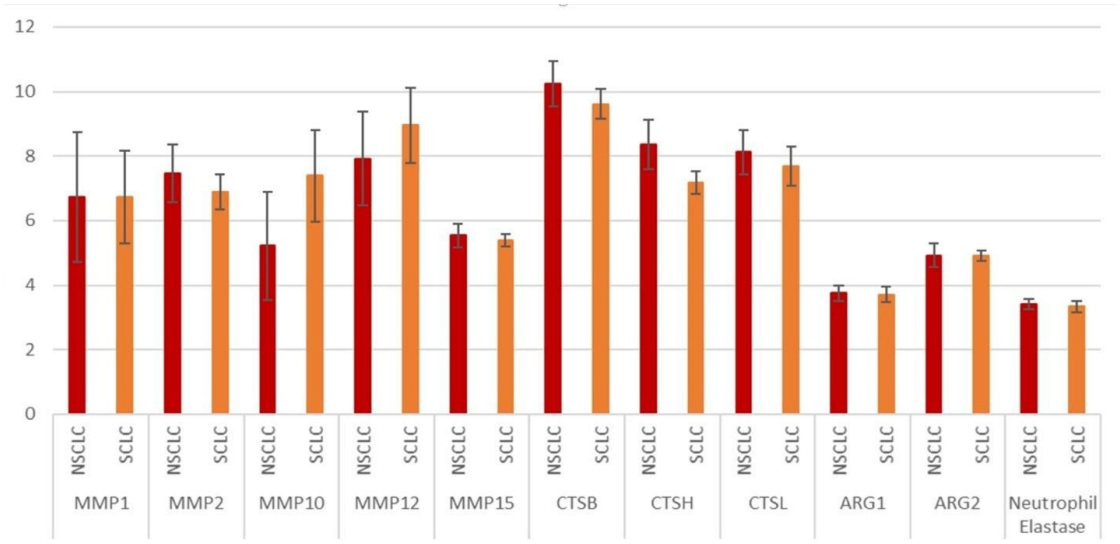


Figure 3.7. Gene expression differences (displayed as natural logarithm on y-axis) between Small Cell Lung Cancer (SCLC) and Non-small Cell Lung Cancer (NSCLC) tissue (on x-axis).[23] The error bars indicate the deviation of mRNA's for the proteases found in tissue.

3.3.1. Peptide syntheses and characterizations

Table 3.1 summarizes all the consensus sequences that were synthesized to develop the 19 nanobiosensors for the detection of lung cancer. These peptides are cleavable in the presence of relevant proteases that are over- or under-expressed in NSCLC and/or SCLC lung cancer.

Table 3.1. Consensus peptide sequences for lung cancer proteases.

Host Proteases	**Consensus peptide sequences**
MMP1	GAGVPMS-MRGGAG
MMP2	GAGIPVS-LRSGAG
MMP3	GAGRPFS-MIMGAG
MMP7	GAGVPLS-LTMGAG
MMP9	GAGVPLS-LYSGAG
MMP10	GAGPSG-LQTGAG
MMP11	GAGGAAN-LVRGAG

MMP12	GAGPKA-LGLAAG
MMP13	GAGPQGLA-GQRGIVAG
MMP15	GAESDA-SQTGAG
Arginase (ARG)	GAGRRRRRRRAG
uPA	GAGSGR-SAG
CTSB	GAGSLLKSR-MVPNFNAG
CTSD	GAGDSG-LGRAG
CTSE	GAGEVAL-VALKAG
CTSH	GAQFVR-SPSGAG
CTSK	GAGAKLK-AENNAG
CTSL	GAGSGVVIA-TVIVITAG
Neutrophil elastase	GAGGEPV-SGLPAG

The characterization results gave high purity of the desired peptide sequences synthesized via our optimized SPPS conditions. An example is the consensus sequence for neutrophil elastase, which secreted by neutrophils, monocytes, macrophages in the tumor environment of lung cancers.[27] The unlabeled peptide showed a highly pure peak at 2.66 min in the chromatogram corresponding to the desired mass peak in the mass spectrum (MW = 1068.134 Da) with closer to 100% peak purity (**Figure 3.8**). The TCPP-labeled neutrophil elastase consensus sequence also showed high purity (**Figure 3.9**) with the mass confirmation via mass spectrum (MW = 1840.924 Da).

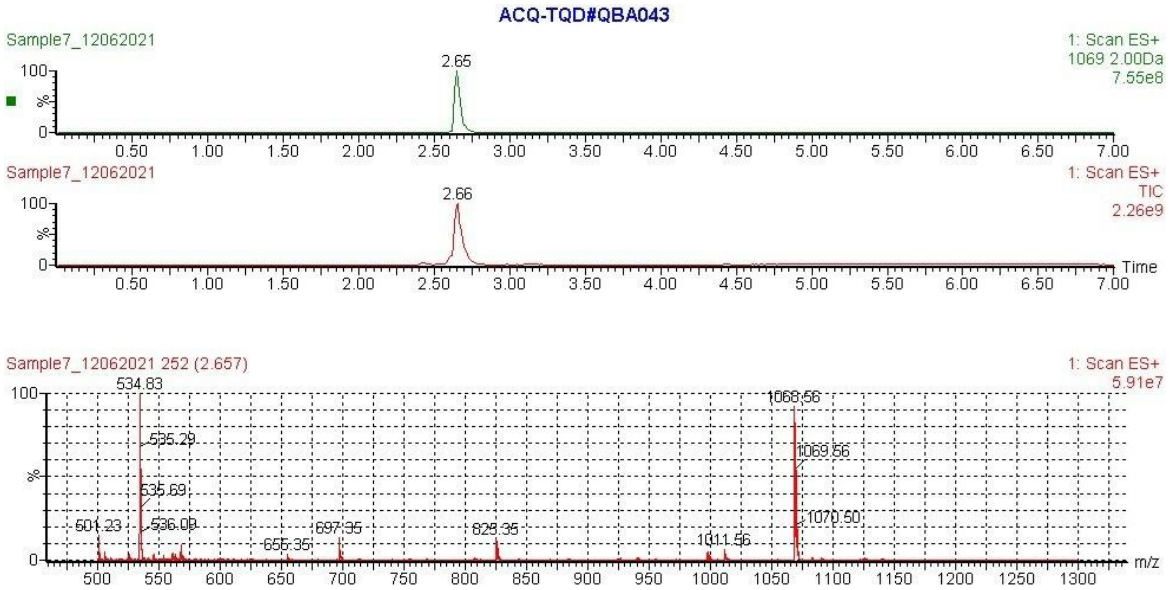


Figure 3.8. Neutrophil elastase (unlabeled) characterization via UPLC-MS. Chromatogram (green) shows the relevant peaks when mass filter was applied.

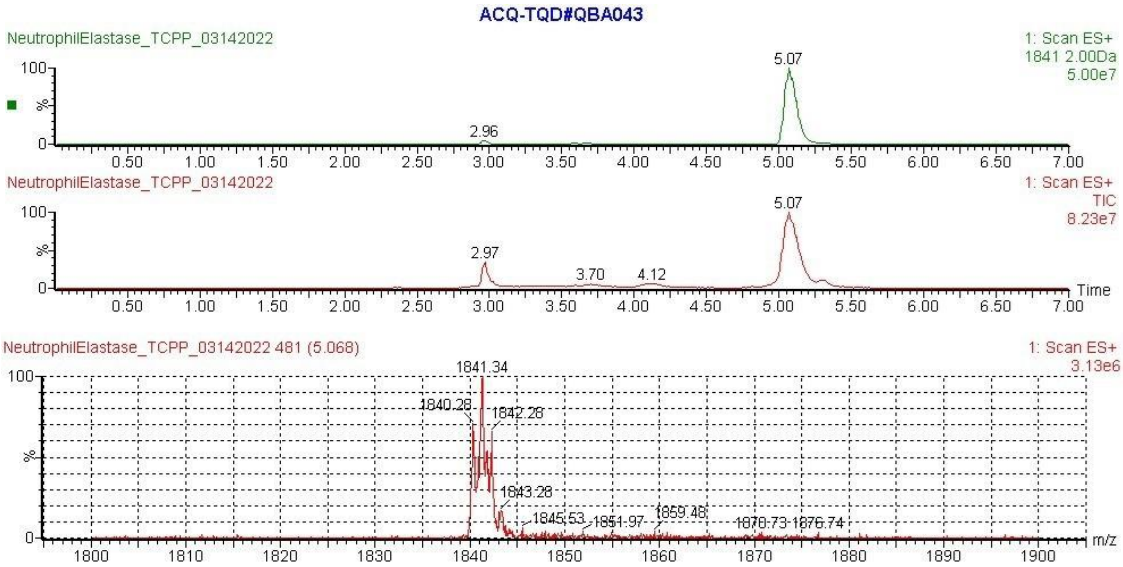


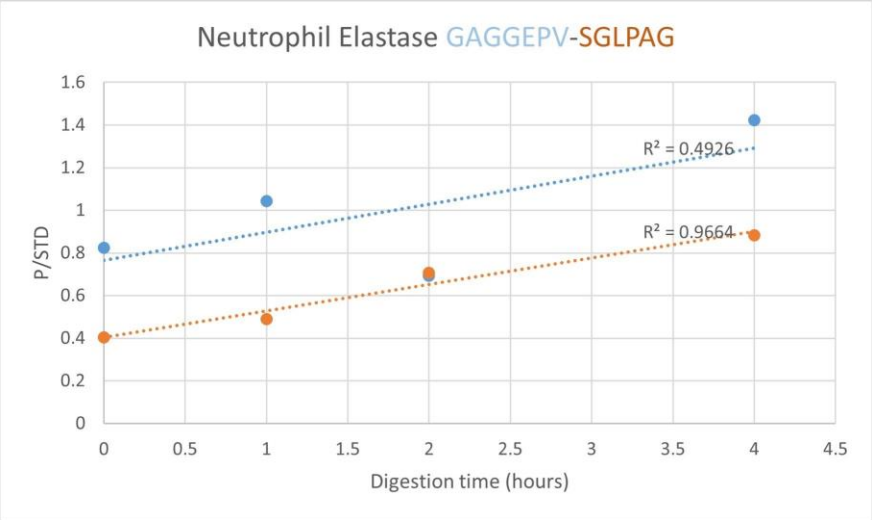
Figure 3.9. Neutrophil elastase (TCPP-labeled) characterization via UPLC-MS. Chromatogram (green) shows the relevant peaks when mass filter was applied.

3.3.2. Reaction kinetics analysis between enzymes and peptide substrates

Enzymatic digestion experiments were performed for unlabeled consensus sequences synthesized for MMP2, MMP9, MMP13, CTSB, CTSD and Neutrophil elastase proteases. These proteases were selected based on the availability of active recombinant proteases. The digestion

fractions were analyzed at 0h, 1h, 2h and 4h intervals and UPLC-MS analysis was performed. It was found that almost all the peptides showed an increasing trend for the concentration of the digested fragments while an expected decreasing trend for the concentration of the peptide, with respect to time. These results confirmed that commercially available proteases were active and specific towards our consensus peptide sequences.

(A)



(B)

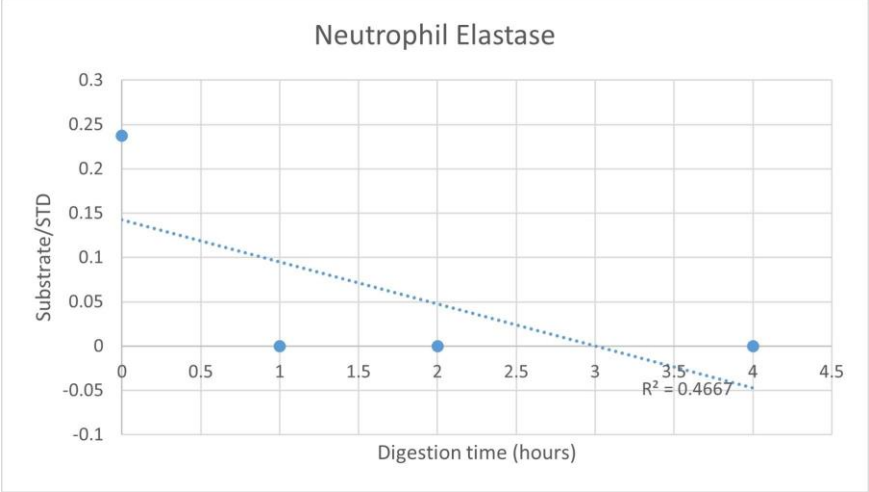
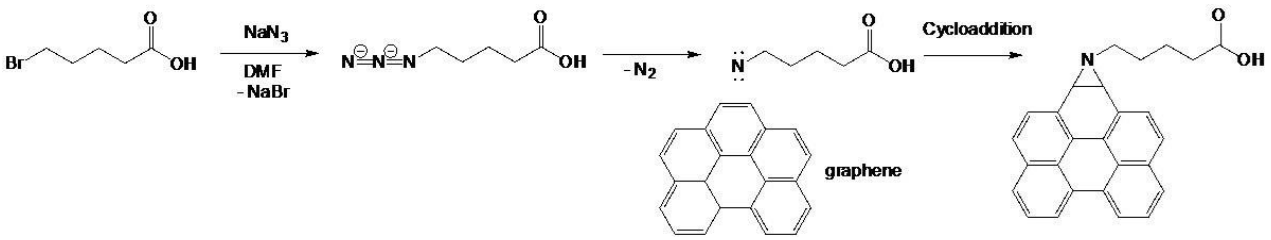


Figure 3.10. Neutrophil elastase consensus sequence digestion with commercial Neutrophil elastase protease, created using UPLC-MS results. **(A)** Trend-line for peptide fragment product **(B)** Trend-line for digested peptide substrate.

The digestion experiments indicate that the commercially available proteases were able to cleave the consensus sequences that were used for the design of nanobiosensors. This is an important result, because it demonstrates that our nanobiosensors principally work as discussed in their general paradigm: upon cleavage of the consensus sequence the fluorescent dye can escape the quenching sphere of the formerly attached nanoparticle or graphene surface. As shown in **Figure 3.10**, principally the same kinetics is found for both fragments of GAGGEPV-SGLPAG. Surprisingly, this was not corroborated by the kinetics of substrate disappearance. The most likely reason for this observation is that the substrate is adsorbed at/in the enzyme and is removed via filtration prior to injection. Another possible explanation is that all MMPs follow Michaelis-Menton kinetics. However, at the chosen concentrations in the digestion experiments described here, both the release of the peptide fragments and the consumption of the consensus sequence should be linearized. It should also be noted that all enzymes used are recombinant. A wide variation of enzyme activity was observed for different enzyme batches throughout my thesis.[28] The results from enzymatic digestion experiments for MMP2, MMP9, MMP13, CTSB, and CTSD are shown in **Appendix A**.

3.3.3. Carboxygraphene and carboxygraphene-PEI characterizations

The reaction of bromovaleric acid with sodium azide leads to the formation of organic azide via nucleophilic substitution reaction that, upon increasing the temperature, releases nitrogen gas to form nitrene intermediate (**Scheme 3.2**). At this point, the nitrogen of the nitrene intermediate contains an incomplete valance shell, so this intermediate undergoes a cycloaddition with a pi-pi double bond at the surface of graphene to form a stable azirane anchor (three-membered C-N-C ring) with a carboxylic acid side chain.[20]



Scheme 3.2. Synthesis of carboxygraphene.

The CHNO elemental analysis of the starting material graphene, carboxygraphene, and carboxygraphene-PEI was performed to find the relative percentage of carbon, hydrogen, nitrogen, and oxygen. Results showed that the inclusion of N in carboxygraphene signifies that a three-membered C-N-C ring called an azirane anchor was formed, indicating that the cycloaddition reaction to conjugate carboxylic acids to graphene's surface was successful. The effective addition process for anchoring polyethylenimine to carboxygraphene is also indicated by the considerable rise in N content for carboxygraphene - PEI. Additionally, we calculated the PEI amount to be 27 ± 2 percent based on the N-content for carboxygraphene-PEI.

Table 3.2. CHNO elemental analysis results.[20]

CHNO Elementary Analysis				
Graphene (0.3)	99.80% C	0.15% H	0.05% O	
Carboxygraphene	94.22% C	1.79% H	2.82% O	1.17% N
Carboxygraphene-PEI	82.58% C	4.65% H	2.12% O	10.65% N

Thermogravimetric analysis (TGA) was used to examine the thermal stability of carboxygraphene and carboxygraphene-PEI. The samples were heated to 600 °C (under nitrogen) at a rate of 10 °C/minute. It was found that before reaching 100 °C, carboxygraphene

experienced an initial mass loss of 28 ± 2 percent by weight (**Figure 3.11**). Graphene was highly stable at temperatures above 100 °C as this behavior is typical of graphene. On the other hand, carboxygraphene-PEI resulted in an initial mass loss of 9 ± 1% by weight below 100 °C and an additional mass loss of 11.5 ± 1% by weight above 150 °C (**Figure 3.11**). The weight loss from carboxygraphene - PEI in the heating range of 150 °C to 600 °C resulted from PEI decomposition. In contrast, the initial weight loss observed was caused by the partial loss of attached valeric acid units in the outer shell of carboxygraphene and the removal of adsorbed water.[29,30]

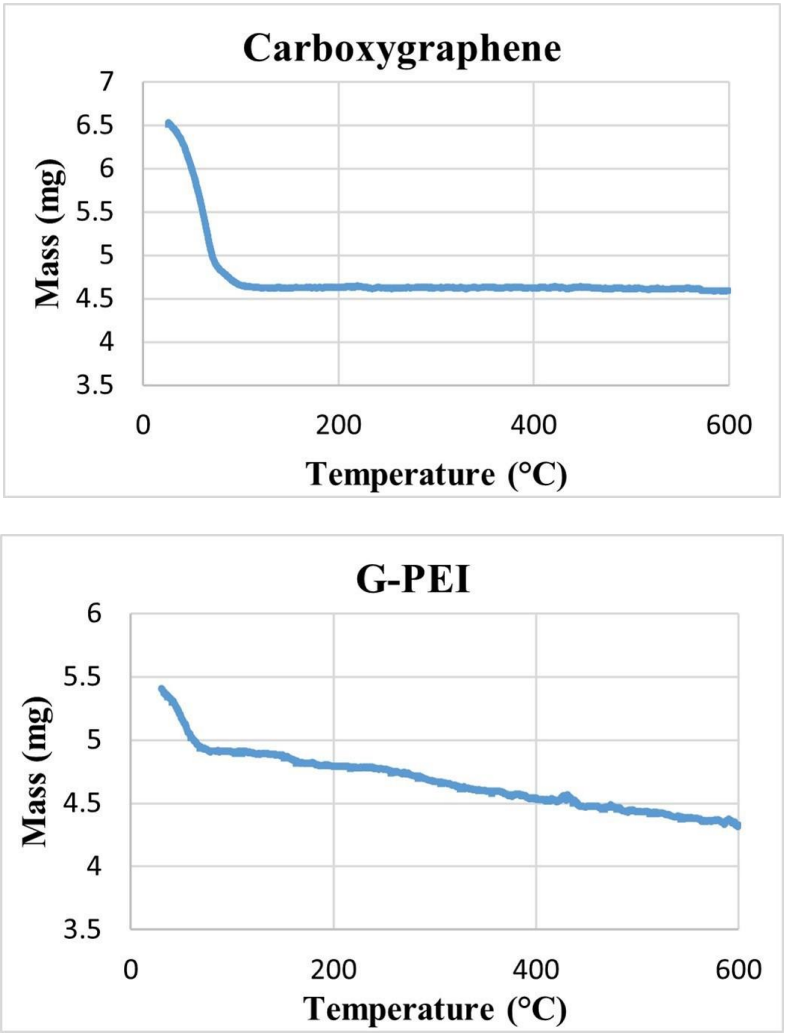


Figure 3.11. Thermogravimetric Analysis of Carboxygraphene and Carboxygraphene-PEI.[20] (Reprinted with permission from ref 20).

The presence of -COOH functional groups on graphene surface was confirmed via Fourier-transform infrared spectroscopy (FTIR analysis) that was recorded as powder ATR (Attenuated Total Reflection. The spectrum in **Figure 3.12** shows data for wavenumber (cm^{-1}) on x-axis and % transmittance on the y-axis. The results showed a distinct signal for the presence of -COOH groups (3400 - 3000 cm^{-1}). The stretching vibrations of C-H bonds were responsible for the absorption band at 2875 cm^{-1}, those of C-O bonds were responsible for the bands at 1250 and 1070 cm^{-1}, and those of C=O bonds were responsible for the band at 1630 cm^{-1}.

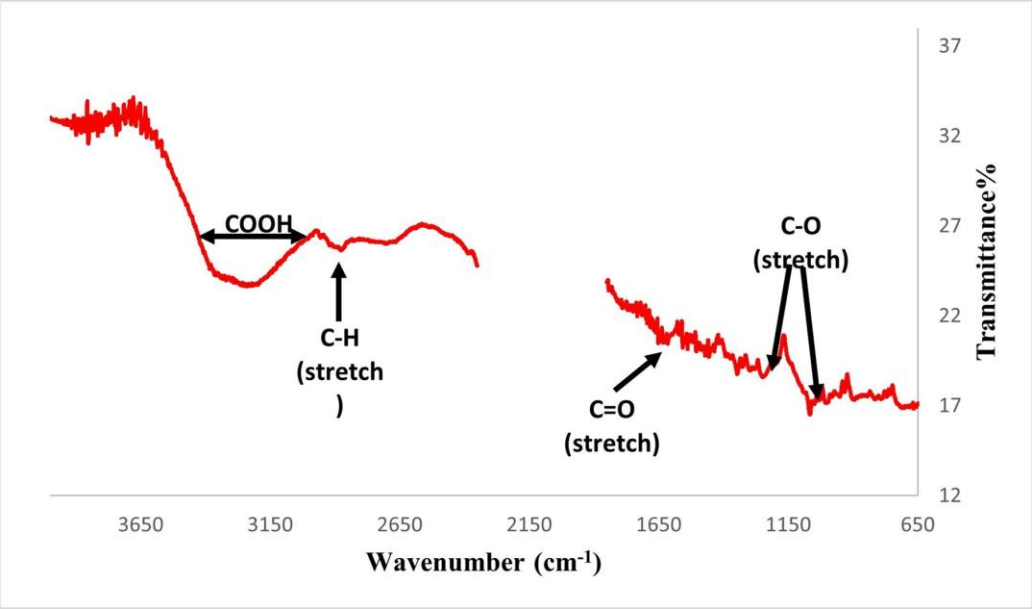


Figure 3.12. Fourier-transform infrared spectroscopy (FTIR) analysis of carboxygraphene.[20] (Reprinted with permission from ref 20).

3.3.4. Dynamic Light Scattering and Zeta-Potential measurements of carboxygraphene, carboxygraphene-PEI, and nanobiosensors

The hydrodynamic size and zeta potential of carboxygraphene, carboxygraphene-PEI, and all 19 nanobiosensors were measured. It was found that agglomeration happens in all samples. Here, two sizes are mentioned for each sample, one from the individual population and the other from the average size (considering possible clusters formed). Based on the findings,

carboxygraphene showed a hydrodynamic diameter of 147.98 nm. The size increased after attaching PEI to the carboxygraphene, which was confirmed by a diameter of 241.53 nm. Similarly, a further increase was observed upon conjugation of TCPP-labeled peptide to carboxygraphene-PEI in the case of all biosensors (**Table 3.4**). Likewise, the zeta potential of carboxygraphene was found to be +22.1 mV, and after covering it with PEI, it rises to +31.9 mV. The rise in charge from carboxygraphene to carboxygraphene-PEI again confirmed that PEI was effectively conjugated to carboxygraphene. After the peptide loading, the zeta potential reduces to almost neutral value that confirms the successful loading of peptide onto the carboxygraphene.

Table 3.3: Dynamic light scattering and Zeta Potential measurements of carboxygraphene and carboxygraphene-PEI

Table 3.3. Dynamic light scattering and Zeta Potential measurements of carboxygraphene and carboxygraphene-PEI.

Sample	**DLS (nm) Individual Population**	**DLS (nm) Average**	**Zeta Potential (mV)**
Carboxygraphene	147.98	839.54	+ 22.1
Carboxygraphene-PEI	241.53	362.32	+ 31.9

Table 3.4. Dynamic light scattering and Zeta Potential measurements of all 19 nanobiosensors.

	DLS (nm)		**Zeta potential (mV)**	
	Batch 1	**Batch 2**	**Batch 1**	**Batch 2**
MMP1	611.07	611.51	1.19	1.33
MMP2	551.19	688.17	0.37	-0.87
MMP3	551.19	603.09	3.66	1.33

MMP7	563.71	636.74	3.13	1.25
MMP9	569.27	670.57	1.68	1.52
MMP10	368.11	419.43	1.96	1.75
MMP11	401.15	404.29	3.33	1.97
MMP12	324.48	363.58	4.10	2.93
MMP13	601.67	676.28	1.85	1.26
MMP15	607.38	684.30	0.55	0.91
CTS B	563.96	657.33	1.42	1.38
CTS D	467.87	533.72	3.88	2.18
CTS E	692.28	794.73	-0.33	-1.81
CTS H	514.39	504.57	2.23	2.35
CTS K	433.65	452.68	3.88	1.42
CTS L	686.77	701.84	0.56	1.19
Neutrophil elastase	504.69	498.07	1.68	0.25
uPA	467.57	509.31	1.92	0.68
ARG	545.25	485.01	3.47	2.06

The data summarized in Table indicates that the sizes of batch 1 and batch 2, as determined by means of light scattering, and their zeta potentials, are comparable. Based on the data available to date, we are not able to formulate a paradigm how to minimize the observed deviations.

3.3.5. Ultraviolet/Visible (UV/Vis) spectroscopy analysis

Very interesting results were obtained from the UV/Vis spectroscopy analysis as the Soret band was clearly visible in the UV/Vis spectra of nanobiosensors at 420 nm which is a characteristic of TCPP confirming the successful loading of peptides labeled with TCPP in all nanobiosensors. It was concluded from the results that TCPP was successfully attached to the polyethylenimine layer through the peptide encoding the consensus MMP-1 sequence. The concentration of TCPP was 1.07 x 10^{-5} moles per gram of carboxygraphene-PEI, based on an estimated absorption coefficient of 135,000 M^{-1} cm^{-1}. Additionally, the findings demonstrated that carboxygraphene-derivatives, such as covalently attached polyethyleneimine could function throughout a wide range of optical wavelengths because the absorption spectrum changes were minimal going from UV to near-infrared (NIR) regions, increasing the biosensor scope with NIR lasers and related biomedical applications. The parameter E (optical extinction) is the sum of A (optical absorption) and Sct (scattering).

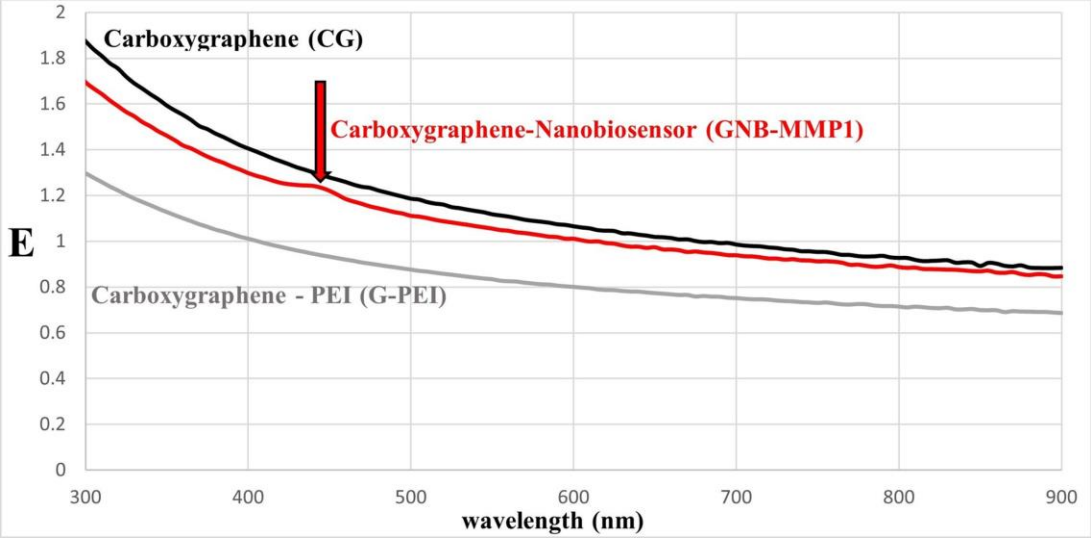


Figure 3.13 Ultraviolet/visible optical extinction spectra of carboxygraphene (Black), carboxygraphene-PEI (Grey), and the MMP1 nanobiosensor (Red, for comparison) in water. The absorption peak of TCPP is marked with a red arrow.[20] (Reprinted with permission from ref 20).

Table 3.5. Concentration of graphene-attached TCPP for each nanobiosensor.

Enzyme	**Consensus Sequence**	**Molecular weight**	**ΔAbs**	**Moles of TCPP per gram of nanobiosensor x 10^6**
MMP1	GAGVPMS-MRGGAG	1920.118	0.014482	1.070
MMP2	GAGIPVS-LRSGAG	1914.068	0.00975	0.722
MMP3	GAGRPFS-MIMGAG	2024.268	0.02255	1.670
MMP7	GAGVPLS-LTMGAG	1903.108	0.02523	1.868
MMP9	GAGVPLS-LYSGAG	1921.058	0.00479	0.355
MMP10	GAGPSG-LQTGAG	1744.798	0.01161	0.860
MMP11	GAGGAAN-LVRGAG	1842.948	0.02065	1.529
MMP12	GAGPKA-LGLAAG	1754.928	0.01552	1.150
MMP13	GAGPQGLA-GQRGIVAG	2181.348	0.010065	0.745

MMP15	GAESDA-SQTGAG	1822.778		
ARG	GAGRRRRRRRAG	2197.418	0.004599	0.341
uPA	GAGSGR-SAG	1491.508	0.02645	1.959
CTSB	GAGSLLKSR-MVPNFNAG	2491.768		
CTSD	GAGDSG-LGRAG	1689.728	0.0078	0.577
CTSE	GAGEVAL-VALKAG	1928.128	0.017625	1.305
CTSH	GAQFVR-SPSGAG	1905.998	0.01583	1.172
CTSK	GAGAKLK-AENNAG	1973.088	0.01393	1.032
CTSL	GAGSGVVIA-TVIVITAG	2257.528		
Neutrophil elastase	GAGGEPV-SGLPAG	1840.928	0.01760	1.304

3.3.6. Transmission electron microscopy (TEM) analysis

To examine the morphology of the samples, TEM images of graphene, carboxygraphene, carboxygraphene-PEI, and carboxygraphene nanobiosensors were collected by Dr. Prem Thapa using a Hitachi H-8100 Transmission Electron Microscope (200 kEV) at the University of Kansas' MAI Research Resource Core Laboratory. The samples dispersed in methanol and were

put onto copper grids, followed by evaporation under a high vacuum. The TEM data demonstrated that the samples' morphology remained unaltered after all derivatizations.

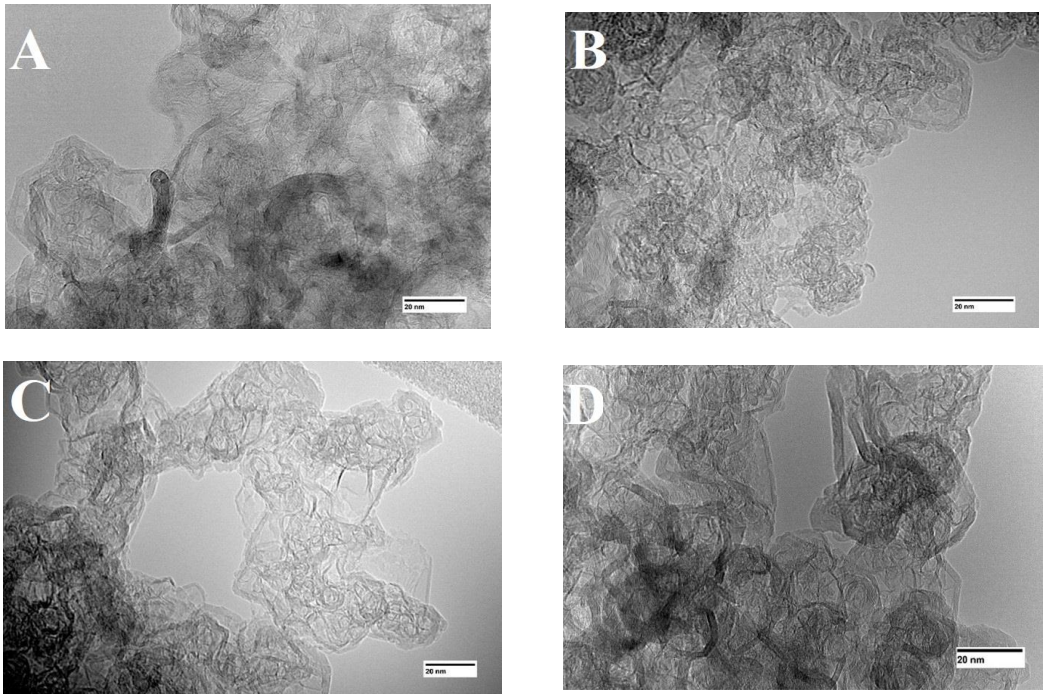


Figure 3.14. Transmission electron microscopy images of A) Graphene, B) Carboxygraphene, C) Carboxygraphene-PEI, and D) MMP1 nanobiosensor.[20] (Reprinted with permission from ref 20).

The TEM Images of all other nanobiosensors synthesized in this thesis are provided in the **Appendix A**. They demonstrate that the fractal structure of explosion graphene-derived nanobiosensors is not changed by the synthetic procedures described here.

3.3.7. Optimizations of designed nanobiosensors

3.3.7.1. Optimization of the peptide concentration

To get the best results from the designed biosensors, the next step was to optimize the working conditions for the nanobiosensor using a serum sample and MMP1 nanobiosensor (as an example). First, the concentration of TCPP-labeled peptide (MMP1-TCPP) loaded onto the nanobiosensor was optimized to get maximum fluorescence intensity after enzymatic cleavage

with recombinant MMP1 (Enzo Lifesciences). For this purpose, 0.30 mg of various graphene nanobiosensors for MMP1 detection possessing multiple TCPP-peptide concentrations (7.5, 14.5, 18, and 24 mg per 100 mg of nanobiosensor) were incubated with 1.0 x 10^{-10} M MMP1 in HEPES buffer at 37° C for 60 min. The outcomes demonstrated that 7.5 mg of TCPP-peptide produced the maximum fluorescence. Based on these findings, the concentration of 7.5 mg TCPP-peptide was selected and used in the subsequent studies. Note that 7.5 mg per 100 mg of nanobiosensor was the lowest concentration that was measurable by means of UV/Vis absorption spectroscopy (see Figure 3.13). At lower TCPP concentrations, we were unable to ascertain the concentration of TCPP-peptide that was bound to the surface of carboxygraphene-PEI.

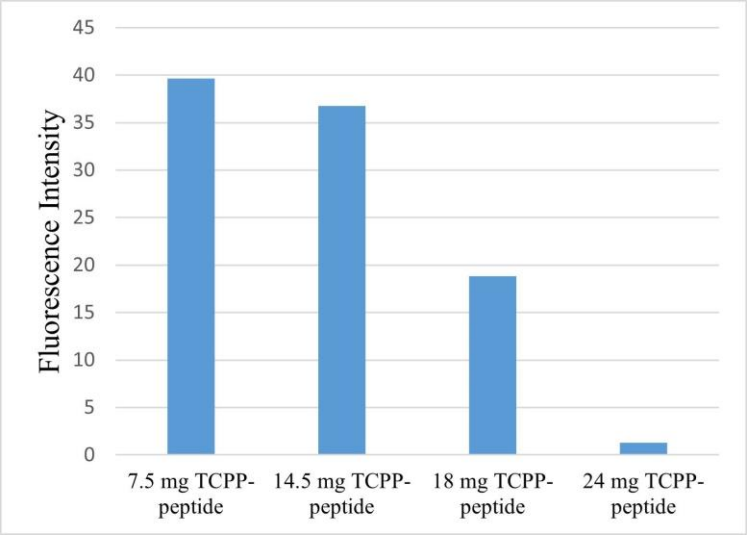


Figure 3.15. Measurement of fluorescence intensity for the optimization of peptide-TCPP (MMP1-TCPP) per 100 mg of nanobiosensor.[20] (Reprinted with permission from ref 20).

Our mechanistic paradigm for the observed release kinetics of peptide-TCPP from the surface of the graphene nanobiosensor is that the attack of the protease is sterically hindered at higher concentrations of consensus-sequence bound TCPP on the surface of the coated few-graphene layer stacks, which acts like a “dense forest”. Because graphene is an excellent

quencher, virtually no fluorescence is observed from bound TCPP, even at higher concentrations on the surface.

3.3.7.2. Optimization of the nanobiosensor concentration

The nanobiosensor concentration was optimized to determine the lowest concentration necessary to produce the highest fluorescence intensity using the same serum sample as the previous experiment. The results demonstrated that the ideal concentration of MMP1 nanobiosensor was between 0.1 mg/mL of nanobiosensor to produce the highest fluorescence intensity. The analysis of samples was repeated three times to get the average values. The average error was ± 2 relative percent. The observed curve shape is typical for the release of fluorescently labeled protease substrates from the surface or few-layer graphene. It should be noted that at 0.5 mg/ml there was no discernible fluorescence. This is an indication for the excellent Fluorescence Resonance Energy Transfer properties of graphene.

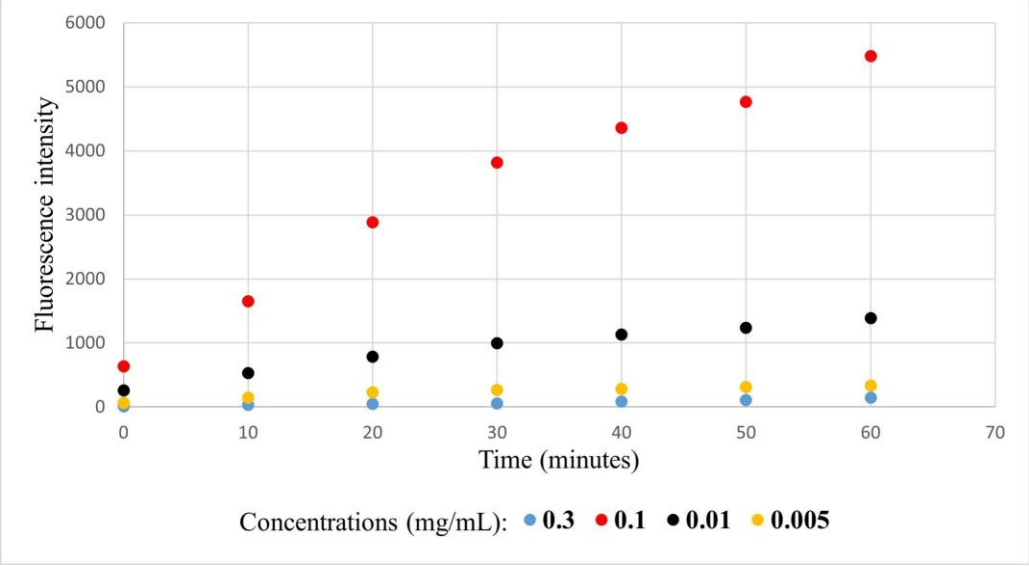


Figure 3.16. Measurement of fluorescence intensity for the optimization of (MMP1) nanobiosensor concentration.[20] (Reprinted with permission from ref 20).

It should be noted that the final optimization of this heterogeneous reaction system is currently being performed by Hawkeye Bio LLC. This has been decided to mitigate any potential conflict of interest with this technology of great identified commercial value.

3.3.7.3. Activity in buffer

The fluorescence intensity was measured in HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid) buffer as well as in MES (2-(N-morpholino)ethanesulfonic acid) buffer to finalize a buffer with most suitable to maintain the physiological conditions during the analysis. The concentration of each buffer used for the analysis was 25 μmol at 298 K temperature and 7.2 pH to ensure the full enzymatic activities. It was observed that most of the nanobiosensors showed higher fluorescence intensities relative to intensities of control samples in HEPES buffer, hence, it was chosen to proceed with. Data for all nanobiosensors is shown in **Appendix A**.

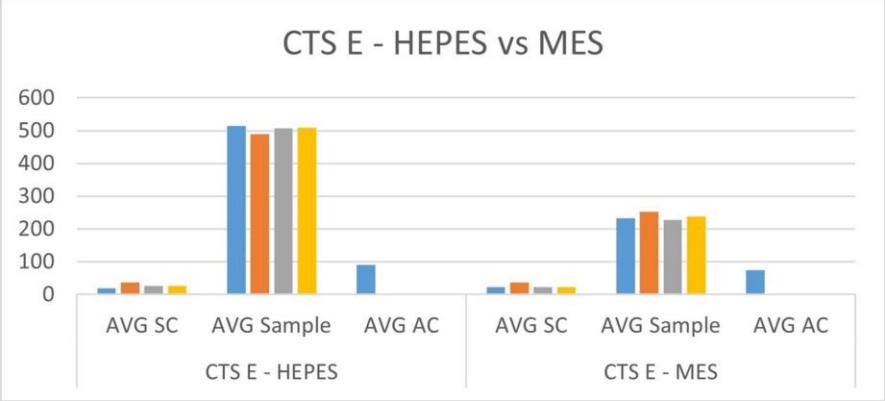


Figure 3.17. Measurement of fluorescence intensity in HEPES vs. MES buffer using CTSE nanobiosensor.

3.3.7.4. Adjusting the pH

The following experiment was carried out to look at how pH affected the fluorescence intensity. When optimized concentration (0.1 mg/mL) was used, it was found that the pH of the sample declined to 6 on its own. It was noticed that the fluorescence intensity significantly

increased again when the pH was adjusted to 7.2 versus when it was not adjusted. There are two main reasons for the observed differences; the enzymatic activity is maximum at physiological pH, which results in maximal cleaving of the peptide substrate leading to higher fluorescence intensity; and acidic pH disturbs the protonation of TCPP dye as TCPP dye displays maximum fluorescence when it is monoprotonated at slightly higher pH.

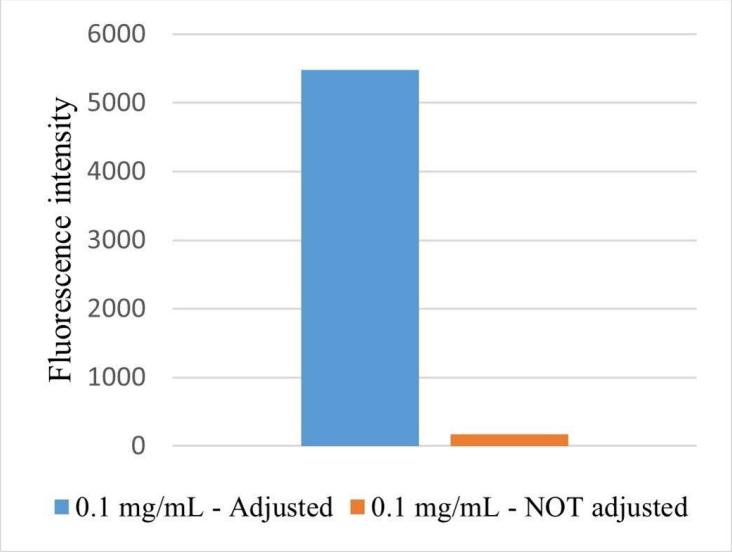


Figure 3.18**.** Measurement of fluorescence intensity for the optimization of pH using MMP1 nanobiosensor.[20] (Reprinted with permission from ref 20).

3.3.7.5. Adjusting the incident wavelength

Since the wavelength is based on the fluorescent moiety employed in the nanobiosensor design, the final optimization experiment was conducted to identify the best wavelength to produce the highest fluorescence intensity. It was discovered that the ideal wavelength for the fluorescence intensity produced by TCPP in the MMP1 nanobiosensor was λ =425 nm.

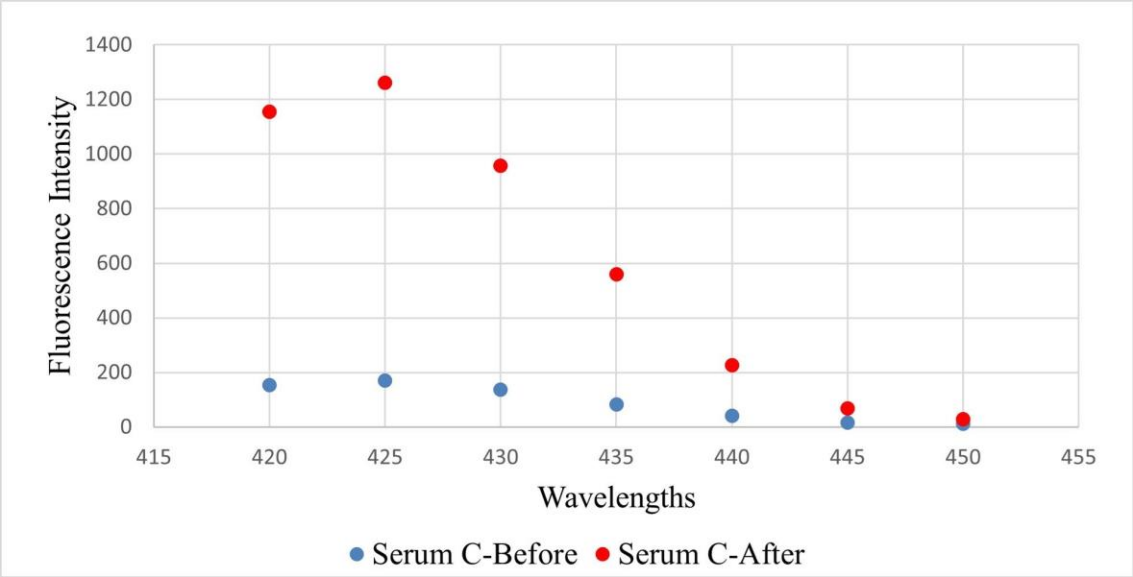


Figure 3.19. Measurement of fluorescence intensity for the wavelength optimization using MMP1 nanobiosensor (before and after incubation). All sera used for the principal testing of the MMP1 biosensor were provided by Hawkeye Bio LLC.[20] (Reprinted with permission from ref 20).

3.3.7.6. Determination of limit of detection (LOD)

To estimate the limit of detection (LOD) of the designed detection platform, the MMP1 nanobiosensor was tested under optimized conditions ((0.1 mg/mL, decreasing enzyme concentration. It was discovered that the detected signal's intensity was correlated with the enzyme concentration. From the data shown here, we estimate a limit of detection (LOD) in the femtomolar range. This is based on the experimental result that with 3% relative error from five independent measurements of the fluorescence intensity after 24h reaction of the optimized nanobiosensor for MMP1 detection (0.1 mg/ml) in HEPES buffer (pH = 7.2) at 37° C for 60 min. The range of MMP1 concentration was varied from 1.0×10^{-11} M to 1.0×10^{-16} M.

The observed fluorescence at λ = 425 nm after 60 min. of incubation was a function of MMP1 concentration, whereas virtually no light was emitted by the intact graphene nanobiosensors for MMP1 detection in the absence of cleaving enzyme (assay control). As it

becomes clear from Figure 3.20, the assay (341 ± 8 counts) at an MMP1 concentration of 1.0 x 10^{-14} M is clearly different from the assay control (12 ± 2 counts). At an MMP1 concentration of 1.0 x 10^{-15} M the assay reads 34 ± 2 counts, whereas the assay control reads 11 ± 1 counts. Based on these findings, we have estimated that the LOD of this assay is at or near 1.0 x 10^{-15} M of MMP1 in HEPES. The assay is about three times larger than the assay control, meaning that at this concentration MMP1 activity is still detectible.

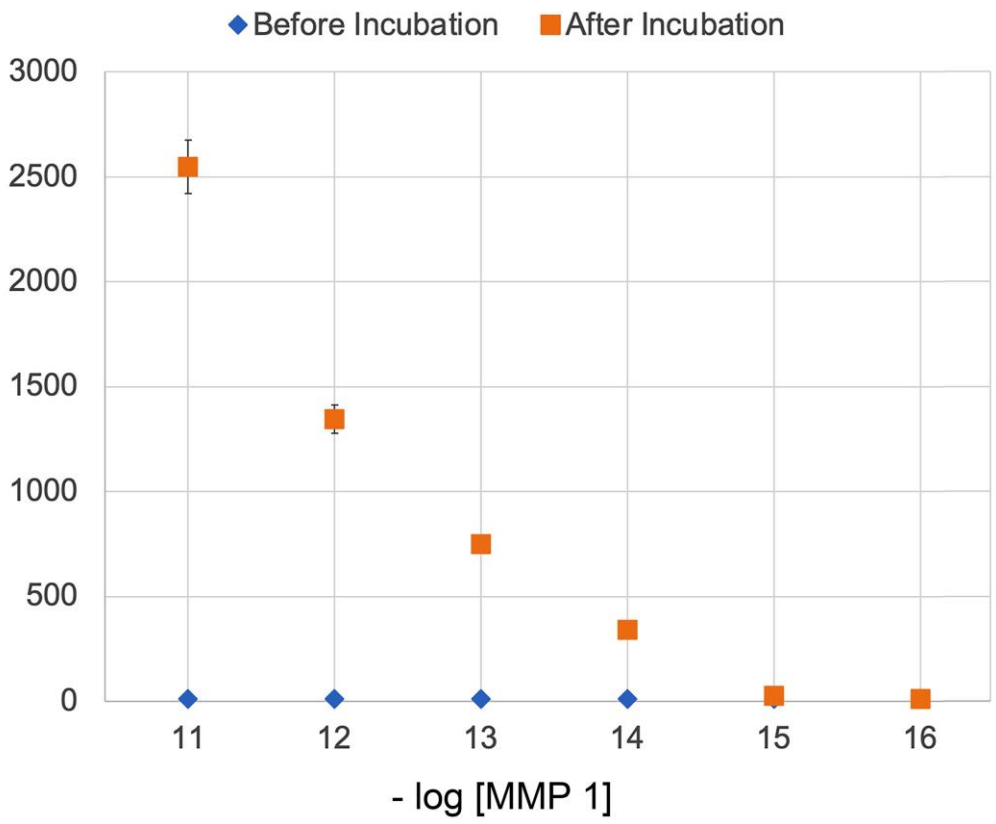


Figure 3.20. Limit of detection estimation indicated by plotting signal vs - log molar enzyme concentration. (Blue: fluorescence before adding MMP1 in various molar concentrations to 0.01 mg/ml optimized MMP1 nanobiosensor in HEPES; Orange: fluorescence after 60 min. incubation at 37 °C).[20] (Reprinted with permission from ref 20).

3.4. Conclusion

In conclusion, this chapter describes a highly efficient detection system for the early detection of lung cancer. The design of the detection system consists of a TCPP-labeled peptide attached to the carboxygraphene core via PEI as a linker and dispersant. The chapter fully describes the synthesis, assembly, characterization, and validation experiments. The designed biosensors were fully optimized to get the best outcomes to detect the lung cancer pathologies. The experimental results showed that after only one hour of incubation, the nanobiosensor was capable of detecting lung cancer biomarkers at as low as sub-femtomolar levels. It has been shown that these nanobiosensors can detect and monitor protease activity in a quantitative and very sensitive manner. This project is in collaboration with Hawkeye Bio LLC, a biotech company located in Torrance/CA. 19 nanobiosensors have been developed with appropriate biomarkers for the detection of lung cancer. They are currently being tested in a Clinical Trial in Shanghai/China for the early detection of lung cancers.

3.5. References

(1) *Lung Cancer Is the Biggest Cancer Killer in Both Men and Women | CDC*. https://www.cdc.gov/cancer/lung/basic_info/mortality-infographic.htm (accessed 2022-06-22).

(2) *Infographic: The Most Common Types of Cancer in the U.S.* Statista Infographics. https://www.statista.com/chart/20692/most-common-types-of-cancer-us/ (accessed 2022-06-22).

(3) *Lung Cancer Statistics | How Common is Lung Cancer?*. https://www.cancer.org/cancer/lung-cancer/about/key-statistics.html (accessed 2022-06-22).

(4) *Types of lung cancer | Cancer Research UK*. https://www.cancerresearchuk.org/about-cancer/lung-cancer/stages-types-grades/types (accessed 2022-06-22).

(5) *Lung Cancer - Non-Small Cell - Risk Factors and Prevention*. Cancer.Net. https://www.cancer.net/cancer-types/lung-cancer-non-small-cell/risk-factors-and-prevention (accessed 2022-06-22).

(6) *Lung Cancer Survival Rates | 5-Year Survival Rates for Lung Cancer*. https://www.cancer.org/cancer/lung-cancer/detection-diagnosis-staging/survival-rates.html (accessed 2022-06-22).

(7) *Stage I Lung Cancer*. WebMD. https://www.webmd.com/lung-cancer/guide/lung-cancer-stage-i-overview (accessed 2022-06-22).

(8) *Definition of biomarker - NCI Dictionary of Cancer Terms - NCI*. https://www.cancer.gov/publications/dictionaries/cancer-terms/def/biomarker (accessed 2022-06-22).

(9) Revelo, A. E.; Martin, A.; Velasquez, R.; Kulandaisamy, P. C.; Bustamante, J.; Keshishyan, S.; Otterson, G. Liquid Biopsy for Lung Cancers: An Update on Recent Developments. *Ann Transl Med* **2019**, *7* (15), 349. https://doi.org/10.21037/atm.2019.03.28.

(10) Henry, N. L.; Hayes, D. F. Cancer Biomarkers. *Molecular Oncology* **2012**, *6* (2), 140–146. https://doi.org/10.1016/j.molonc.2012.01.010.

(11) Bond, J. S. Proteases: History, Discovery, and Roles in Health and Disease. *Journal of Biological Chemistry* **2019**, *294* (5), 1643–1651. https://doi.org/10.1074/jbc.TM118.004156.

(12) Udukala, D. N.; Wang, H.; Wendel, S. O.; Malalasekera, A. P.; Samarakoon, T. N.; Yapa, A. S.; Abayaweera, G.; Basel, M. T.; Maynez, P.; Ortega, R.; Toledo, Y.; Bossmann, L.; Robinson, C.; Janik, K. E.; Koper, O. B.; Li, P.; Motamedi, M.; Higgins, D. A.; Gadbury, G.; Zhu, G.; Troyer, D. L.; Bossmann, S. H. Early Breast Cancer

Screening Using Iron/Iron Oxide-Based Nanoplatforms with Sub-Femtomolar Limits of Detection. *Beilstein Journal of Nanotechnology* **2016**, *7* (1), 364–373. https://doi.org/10.3762/bjnano.7.33.

(13) López-Otín, C.; Matrisian, L. M. Emerging Roles of Proteases in Tumour Suppression. *Nat Rev Cancer* **2007**, *7* (10), 800–808. https://doi.org/10.1038/nrc2228.

(14) Park, K. C.; Dharmasivam, M.; Richardson, D. R. The Role of Extracellular Proteases in Tumor Progression and the Development of Innovative Metal Ion Chelators That Inhibit Their Activity. *Int J Mol Sci* **2020**, *21* (18), 6805. https://doi.org/10.3390/ijms21186805.

(15) Chakraborti, S.; Sarkar, J.; Pramanik, P. K.; Chakraborti, T. Role of Proteases in Lung Disease: A Brief Overview. *Proteases in Human Diseases* **2017**, 333–374. https://doi.org/10.1007/978-981-10-3162-5_16.

(16) Kalubowilage, M.; Covarrubias-Zambrano, O.; Malalasekera, A. P.; Wendel, S. O.; Wang, H.; Yapa, A. S.; Chlebanowski, L.; Toledo, Y.; Ortega, R.; Janik, K. E.; Shrestha, T. B.; Culbertson, C. T.; Kasi, A.; Williamson, S.; Troyer, D. L.; Bossmann, S. H. Early Detection of Pancreatic Cancers in Liquid Biopsies by Ultrasensitive Fluorescence Nanobiosensors. *Nanomedicine: Nanotechnology, Biology and Medicine* **2018**, *14* (6), 1823–1832. https://doi.org/10.1016/j.nano.2018.04.020.

(17) Covarrubias-Zambrano, O.; Motamedi, M.; Ameredes, B. T.; Tian, B.; Calhoun, W. J.; Zhao, Y.; Brasier, A. R.; Kalubowilage, M.; Malalasekera, A. P.; Yapa, A. S.; Wang, H.; Culbertson, C. T.; Troyer, D. L.; Bossmann, S. H. Optical Biosensing of Markers of Mucosal Inflammation. *Nanomedicine: Nanotechnology, Biology and Medicine* **2022**, *40*, 102476. https://doi.org/10.1016/j.nano.2021.102476.

(18) Daniyal, M.; Liu, B.; Wang, W. Comprehensive Review on Graphene Oxide for Use in Drug Delivery System. *Curr Med Chem* **2020**, *27* (22), 3665–3685. https://doi.org/10.2174/1381612825666190201296290.

(19) Nepal, A.; Singh, G. P.; Flanders, B. N.; Sorensen, C. M. One-Step Synthesis of Graphene via Catalyst-Free Gas-Phase Hydrocarbon Detonation. *Nanotechnology* **2013**, *24* (24), 245602. https://doi.org/10.1088/0957-4484/24/24/245602.

(20) Covarrubias, J. J. Rational Chemical Applications of Explosion-Graphene. **2021**.

(21) Eid, R.; Arab, N. T. T.; Greenwood, M. T. Iron Mediated Toxicity and Programmed Cell Death: A Review and a Re-Examination of Existing Paradigms. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **2017**, *1864* (2), 399–430. https://doi.org/10.1016/j.bbamcr.2016.12.002.

(22) Choi, Y.; Kendziorski, C. Statistical Methods for Gene Set Co-Expression Analysis. *Bioinformatics* **2009**, *25* (21), 2780–2786. https://doi.org/10.1093/bioinformatics/btp502.

(23) *Home - GEO - NCBI*. https://www.ncbi.nlm.nih.gov/geo/ (accessed 2022-06-23).

(24) Udukala, D. N.; Wendel, S. O.; Wang, H.; Yapa, A. S.; Covarrubias-Zambrano, O.; Janik, K.; Gadbury, G.; Troyer, D. L.; Bossmann, S. H. Early Detection of Non-Small Cell Lung Cancer in Liquid Biopsies by Ultrasensitive Protease Activity Analysis. *Journal of Cancer Metastasis and Treatment* **2020**, *6*, 25. https://doi.org/10.20517/2394-4722.2020.45.

(25) *Early detection of pancreatic cancers by novel nanobiosensor-based protease biomarkers using hierarchical decision structure. | Journal of Clinical Oncology*. https://ascopubs.org/doi/abs/10.1200/JCO.2021.39.15_suppl.e16273 (accessed 2022-01-18).

(26) Zhang, Y.; Qian, J.; Gu, C.; Yang, Y. Alternative Splicing and Cancer: A Systematic Review. *Sig Transduct Target Ther* **2021**, *6* (1), 78. https://doi.org/10.1038/s41392-021-00486-7.

(27) Lerman, I.; Hammes, S. R. Neutrophil Elastase in the Tumor Microenvironment. *Steroids* **2018**, *133*, 96–101. https://doi.org/10.1016/j.steroids.2017.11.006.

(28) Palmier, M. O.; Van Doren, S. R. Rapid Determination of Enzyme Kinetics from Fluorescence: Overcoming the Inner Filter Effect. *Analytical Biochemistry* **2007**, *371* (1), 43–51. https://doi.org/10.1016/j.ab.2007.07.008.

(29) Chen, B.; Liu, M.; Zhang, L.; Huang, J.; Yao, J.; Zhang, Z. Polyethylenimine-Functionalized Graphene Oxide as an Efficient Gene Delivery Vector. *J. Mater. Chem.* **2011**, *21* (21), 7736–7741. https://doi.org/10.1039/C1JM10341E.

(30) Song, M.; Xu, J. Preparation of Polyethylenimine-Functionalized Graphene Oxide Composite and Its Application in Electrochemical Ammonia Sensors. *Electroanalysis* **2013**, *25* (2), 523–530. https://doi.org/10.1002/elan.201200376.

Chapter 4 - Developing peptide-based nanoplatform for COVID-19 diagnosis

Abstract

The coronavirus pandemic, which started in December 2019, spread around the globe at a rapid pace and proved dangerous, if not deadly, to so many. According to the John Hopkins University and Medicine Coronavirus Resource Center, approximately 535 million people have been affected by the coronavirus, and 6.31 million people have lost their lives to date, and the numbers are still increasing. There is a critical need for a detection system that can diagnose the infected patients with the least resources (especially true for third-world countries and large populations). In this study, we report a rapid and reliable biosensor design that consists of peptide sequence labeled with tetrakis-carboxyphenyl-porphyrin (TCPP) and explosion graphene core-shell. The consensus peptide sequences specifically target and measure the activity of SARS-CoV-2 protease biomarkers, including papain-like cysteine protease (PLpro) and 3C-like serine protease (3CLpro). The detection system works on the principle of Förster resonance energy transfer, where the cleaving of consensus peptide sequence by the PLpro or 3CLpro proteases results in the termination of the quenching effect between graphene and TCPP. The increase in the fluorescence intensity upon the termination of quenching is measured using a plate reader. This platform not only gives a faster detection and quantitative measure of the viral infection, but also offers the opportunity for an easy redesign by modifying the consensus peptide part depending on the genome of the new strain of coronavirus and related proteases. Furthermore, this biosensor can be easily used in laboratories with limited resources.

4.1. Introduction

The coronavirus is now the major pathogen of respiratory disease outbreaks. Recently, at least two major epidemics caused by a severe acute respiratory syndrome coronavirus were reported.[1] In 2002/03, a strain was first detected in Foshan, in Guangdong province, and in 2019/20, another was detected in Wuhan, Hubei province, both in mainland China. Among the severe clinical manifestations of COVID-19 disease, caused by the SARS-CoV-2 that originated in Wuhan, are mainly acute respiratory distress syndrome (ARDS), severe pneumonia, sepsis, and septic shock. Studies have found that a median duration between onset of symptoms and ICU admission is 9 to 10 days, which suggests a gradual deterioration in the majority of cases.[1] The most documented reason for requiring intensive care has been respiratory support, of which two-thirds of the patients have met criteria for ARDS.[2] Currently, the diagnosis of ARDS requires clinical and ventilatory criteria. These are supported by chest imaging, including chest radiograph, CT scan, or lung ultrasound demonstrating bilateral opacities (lung infiltrates > 50%). ARDS is the most common cause of death related to COVID-19.[2] Current diagnostic techniques are expensive and labor-intensive. Therefore, fast, and cost-effective methods for monitoring the development of respiratory failure in SARS-CoV-2 infected patients are urgently needed.

Two major types of SARS-CoV-2 (COVID-19) tests exist to date: A) The majority of SARS-CoV-2 (COVID-19) tests used in clinical diagnostics are based on real-time reverse transcriptase (RT)-PCR (polymerase chain reaction) diagnostic panels[3], including the ID NOW platform from Abbott Diagnostics[4], which obtained emergency use authorization (EUA) from the FDA.[5] Real-time (quantitative) PCR offers rapid high-throughput detection and quantification of virtually any conceivable DNA or RNA target. With respect to detecting SARS-CoV-2, real

time-PCR methods are reliable for detecting more than 100 to 1000 viruses per milliliter.[5] B) The second group consists of serological methods (e.g., radioimmunoassays, ELISA (enzyme-linked immunosorbent assays), Western blots, and recombinant immunoblot assays that are used to determine antibodies titres resulting from (viral) infections.[6] ELISA is widely popular due to low limits of detection (pico- to attomolar) for viral antigens, resulting in high sensitivity, relatively low costs, >90% specificity, and straightforward interpretation of results.[6] However, antibodies that are truly specific for SARS-CoV-2 are not available to date. Therefore, patients who had a coronavirus infection other than SARS-CoV-2 in the past may be incorrectly diagnosed as SARS-CoV-2 positive. However, these patients would not have developed immunity against SARS-CoV-2. A comprehensive list of all SARS-CoV-2 tests that have been approved to date can be found here[7]. Overall, classic approaches (like PCR and serology) are fast but limited to the known pathogens only. Some advanced methods give more data, but are not established in clinic yet, such as HTS (high throughput screening).[6] What is missing is a method to ascertain the activity of a virus (e.g., SARS-CoV-2) close to real-time.

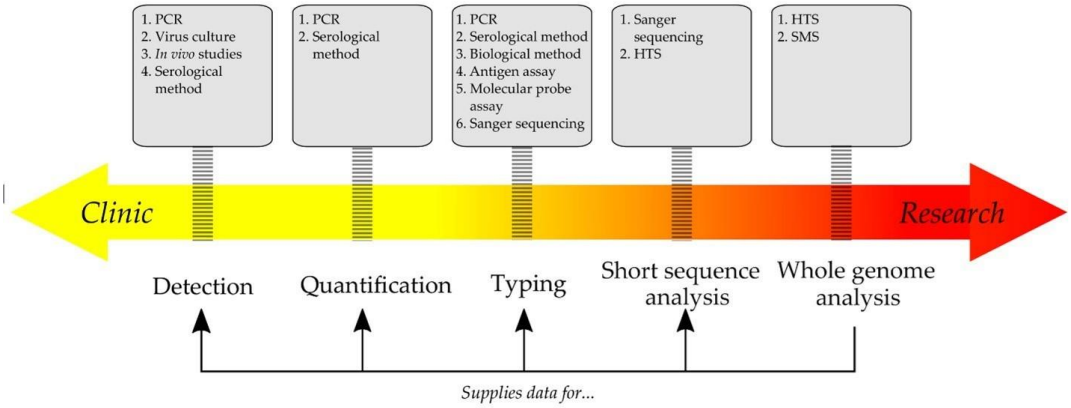


Figure 4.1. The determination of the prevalence of clinical or scientific application of a developed method.[6] (Reprinted with permission from ref 6).

4.1.1. Biomarkers for COVID-19 detection

The typical PCR-based and serological methods have one aspect in common-that they cannot distinguish between patients with active coronavirus infection and patients that have survived the disease; PCR can still give a positive test to viral pieces after a patient survives the infection and serological methods give a positive test to antibodies produced as a result of viral infection.[6] Therefore, a critical need for a technology exists that is capable of detecting the activity of a (corona)virus infection by means of a liquid biopsy using plasma, serum, sputum, or exhaled breath condensate. To fill this need, peptide-based biosensors were developed for detecting the proteolytic activities of SARS-CoV-2 protease biomarkers, including papain-like cysteine protease (PLpro) and 3C-like serine protease (3CLpro). These two signature proteases of coronaviruses are evolutionary conserved[8,9], because they are required to cleave their open-reading frames (ORF, **Figure 4.2**).

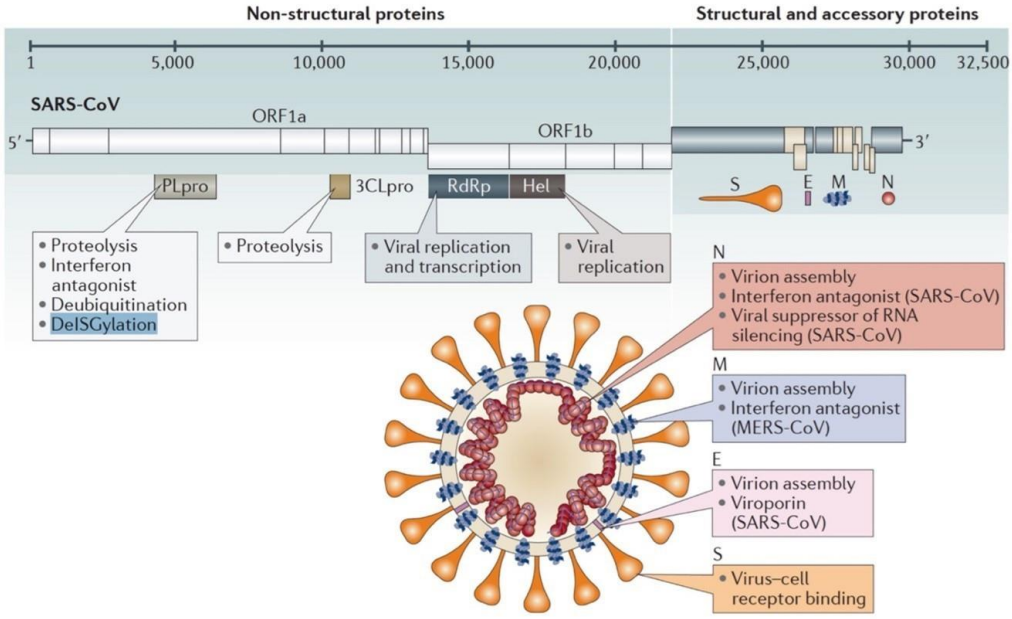


Figure 4.2. Genomes and Structures of Coronviruses.[10] (Reprinted with permission from ref 10).

The typical coronavirus contains a non-segmented and single-stranded ribonucleic acid (RNA) genome (approximately 30 kb). It is arranged in the order of 5′-3′ including replicase

genes, genes that encode for structural proteins (spike glycoprotein (S), envelope protein (E), membrane protein (M) and nucleocapsid protein (N)), and polyadenylated tail. The partial 5′-terminal ORF 1a/b is within the 5′ two-thirds of the covid genome and encodes for replicase polyprotein 1a (pp1a) and pp1ab. These polyproteins are cleaved by specific proteases including papain-like cysteine protease (PLpro) and 3C-like serine protease (3CLpro) to produce the non-structural proteins, including RNA-dependent RNA polymerase and helicase, which are important enzymes for CoVs transcription and replication.[10] It is of importance that, according to the recently published genome of COVID-19, both papain-like cysteine protease (PLpro) and 3C-like serine protease (3CLpro) are genetically conserved.[8,9] Therefore, they are suitable targets for detecting active coronavirus infections, as well as the recurrence of corona viruses after initial immunity.

4.2. Materials and Methods

4.2.1. Synthesis of peptides

All consensus peptide sequences for developing the COVID-19 biosensors were synthesized via our optimized solid phase peptide synthesis (SPPS) conditions **(Scheme 2.1)**. Briefly, a resin containing the first amino acid of the peptide was swollen in DCM for 20 minutes followed by a DMF wash for 1 minute. Meanwhile, the next amino acid (Fmoc-protected) of the peptide chain was activated with HBTU in DIEA:DMF (1:23). After that, the activated amino acid was added to the resin-amino acid. The mixture was swirled for 30 minutes at room temperature followed by another addition of the amino acid mixture. After the addition of new amino acid was completed, the resin-peptide was washed five times with DMF. The next step was to deprotect the amine group that was protected by the Fmoc. To do so, 20% piperidine in DMF was used for 1 minute followed by 10 minutes incubation (both at room temperature).

Again, DMF washings were performed to get rid of piperidine. At this point, the next (Fmoc-protected) amino acid was activated and added to the resin having the growing chain of peptide. The cycle of Fmoc-deprotection and addition of new amino acid was repeated until the desired amino acid sequence was obtained for all three peptide sequences.

4.2.1.1. Labeling of consensus peptide sequences with dye

After the synthesis, peptide sequences were labeled with tetrakis-carboxyphenyl-porphyrin (TCPP) dye at the N-terminus of the peptide. For the coupling of dye to peptide chain, the milli-equivalent ratio of the resin to HBTU was 1:2.9, and resin to the dye was 1:3. The carboxylic group of the dye was activated using HBTU in DIEA:DMF (1:23). The dye mixture was incubated for 24 hours with the resin-peptide chain, followed by multiple DMF washings to remove any free dye after the coupling step.

4.2.1.2. Cleaving of the resin

After the successful attachment of the TCPP dye to the peptides, an acid cocktail solution was used to cleave the peptide sequence from the resin. This cocktail solution consisted of 95% TFA, 2.5% TIPS and 2.5% distilled water. The resin along with the peptide chain was incubated in this cocktail for 3 hours. After the completion of the resin cleaving step, the peptide was precipitated and washed with cold ether, followed by drying with Ar gas.

4.2.1.3. Analysis of the synthesized peptides

All the samples were prepared in 80% acetonitrile and 20% water v/v to perform UPLC-MS analysis (sample concentration: 0.1 mg/mL) using Waters ACQUITY TQD Tandem Quadrupole UPLC/MS/MS system in ES+ ionization mode using acetonitrile/water gradient. The instrument was equipped with an ACQUITY UPLC peptide CSH C18 column (Waters), 130 Angstrom, 1.7 micrometer 2.1 mm x 150 mm.

4.2.2. Enzymatic digestion experiments

To run the digestion experiments, recombinant papain-like cysteine protease (PLpro) and 3C like serine protease (3CLpro) enzymes were used. 25 nM of each peptide (unlabeled) was dissolved in 1.0 mL of a buffer comprised of Tris (50 mM), $CaCl_2$ (10 mM), and NaCl (150 mM), pH=7.5 (adjusted with HCl). 100 nM of either PLpro or 3CLpro were dissolved in 1.0 mL of the same buffer. Both solutions were heated to 37 ^{o}C. The reaction started by adding the enzyme solution to the peptide solution. The reaction mixture was kept at 37 ^{o}C for four hours. Samples (25 μl) were taken at 0, 60, 120, 180, and 240 min. and analyzed using a Waters ACQUITY TQD Tandem Quadrupole UPLC/MS/MS system that was equipped with an ACQUITY UPLC peptide CSH C18 column (Waters), 130 Angstrom, 1.7 micrometer 2.1 mm x 150 mm. An acetonitrile/water gradient was used for eluting the peptides.

4.2.3. Synthesis of nanoparticles

Detailed procedures for Fe/Fe_3O_4-nanobiosensor synthesis have been published.[11] In short, Fe/Fe_3O_4-nanoparticles were synthesized from $Fe(CO)_5$ by means of a seeded growth reaction in the presence of oleylamine and hexadecylammonium chloride (HADxHCl) using 1-octadecene (ODE) as solvent, and then coated with organic dopamine layer.

4.2.4. Assembly of nanobiosensors

In the last step, peptide-bound TCPP and cyanine (Cy) 5.5 were bound to the primary amine groups of the dopamine on the Fe/Fe_3O_4 surface via stable amide bonds. To do so, three reaction mixtures were prepared by taking 100 mg of Fe/Fe_3O_4 nanoparticles coated with dopamine and were suspended in ~15.5 mL DMF in a glass vial individually. All reaction mixtures were sonicated for one min. Next, ~10 mg 4-dimethylaminopyridine (DMAP) and ~10 mg 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) were added to each vial followed by

~14.13 mg of TCPP pre-conjugated with each consensus peptide sequence individually. After doing that, all reaction mixtures were sonicated for five minutes to ensure full suspension of all components in DMF. The reaction mixture was then stirred (vigorously) overnight, at room temperature. After the completion of 24 hours, the nanobiosensors were collected via centrifugation (5 min @ 10000 rpm) followed by multiple DMF washings to get rid of unbound components. After that, washings with cold diethyl ether (263 K) were performed and resulting nanobiosensors were collected via centrifugation. After the washings with diethyl ether, nanobiosensors were dried with argon.

Two batches of nanobiosensors for COVID-19 detection were synthesized, both featuring 35 ± 5 TCPP fluorophors and 50 ± 8 Cyanine 5.5 molecules per Fe/Fe_3O_4 nanoparticle.

4.2.5. Fluorescence measurements with Fe/Fe_3O_4-biosensors

In accordance with previously published procedures[12–14] A BioTek Synergy H1 plate reader with tungsten halogen lamp, analysis bandpass filter: 650 ± 25 nm, excitation bandpass filter: 421 ± 10 nm and 96-well plates was used. Two types of solutions were prepared; solution 1 contained HEPES buffer (25 μmol) (2-[4-(2-hydroxyethyl) piperazin-1-yl]ethanesulfonic acid) that was prepared enriched with Ca(II), Mg(II), and Zn(II) (10 μmol each). The temperature was kept 298 K (pH = 7.2) to ensure the full enzymatic activities. Solution 2 contained the nanobiosensor (0.3 mg Fe/Fe_3O_4) in 1.0 mL of HEPES buffer and was sonication for 10 min at 298 K. Sample control, assay, assay control and blank were prepared using solution 1 and/or solution 2 to 5 μL of serum sample as follows; Sample Control contained 125 μL of solution 1 + 5 μL recombinant enzyme. Assay contained 125 μL of solution 2 + 5 μL of calibration solution containing known concentrations of purchased enzymes. Assay Control contained 125 μL of

solution 2 + 5 μL of solution 1. Blank contained 130 μL of solution 1. Each sample was loaded (130 μL) into one well of a black 96-well plate and incubated for 90 min. at 37ºC.

4.2.6. Fluorescence measurements with detonated graphene-biosensors

Similarly, two types of solutions were prepared; solution 1 contained HEPES buffer (25 μmol) prepared enriched with Ca(II), Mg(II), and Zn(II) (10 μmol each). The temperature was kept 298 K (pH = 7.2) to ensure the full enzymatic activities. Solution 2 contained the nanobiosensor (0.03 mg of graphene-nanobiosensor) in 1.0 mL of HEPES buffer and was sonication for 10 min at 298 K. Sample control, assay, assay control and blank were prepared using solution 1 and/or solution 2 to 5 μL of serum sample as follows; Sample Control contained 125 μL of solution 1 + 5 μL serum sample. Assay contained 125 μL of solution 2 + 5 μL of serum sample. Assay Control contained 125 μL of solution 2 + 5 μL of solution 1. Blank contained 130 μL of solution 1. Each sample was loaded (130 μL) into one well of a black 96-well plate and incubated for 90 min. at 37ºC.

4.3. Results and Discussion

4.3.1. Iron/Iron oxide-based nanobiosensors for Cysteine Protease (PLpro) and 3C-like Serine Protease (3CLpro)

The nanoparticles have a well-defined core-shell structure, with the average Fe(0) core diameter of 13±0.5 nm and the Fe_3O_4 shell thickness of 2.0 ± 0.5 nm, respectively.[11] The high-resolution transmission electron microscopy (HRTEM) image reveals the polycrystalline nature of the nanoparticles. Dopamine forms robust organic coatings with binding constants of the order of 10^{15} l mol^{-1}.[15]

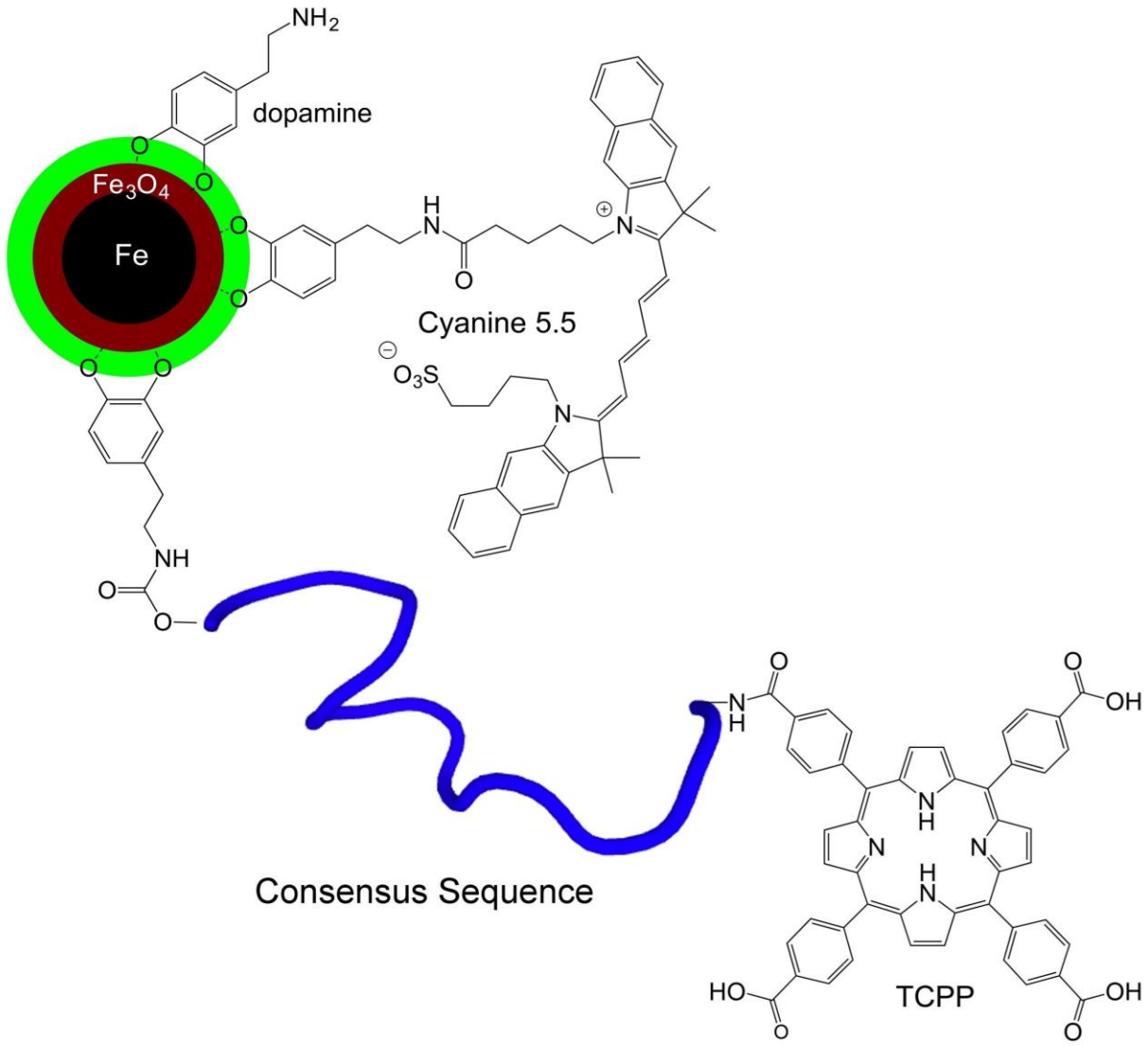


Figure 4.3. Fe/Fe_3O_4 core/shell nanoparticle-based nanobiosensor for protease detection.

It also increases the water-solubility of the resulting nanosensors to > 5g L^{-1}.[11] Porphyrins have been used as cleavable fluorescent dyes because their photophysical properties are excellently characterized.[16] Cyanine 5.5 has been co-attached as FRET quencher due to its large molar extinction coefficient and suitable Förster radius.[17] **Figure 4.3** shows the structure of the nanosensor comprised of dopamine-coated Fe/Fe_3O_4, consensus sequence, TCPP, and Cyanine 5.5. Cy 5.5 is permanently linked to dopamine. Due to plasmonic quenching by the Fe/Fe_3O_4

nanoparticles, Cyanine 5.5 shows only minimal fluorescence. Once the oligopeptide tether is cleaved by a protease with high selectivity for the selected consensus sequence, the fluorescence of TCPP increases significantly. The measure of the fluorescence intensity of the nanosensor after the peptide cleaving gives the extent of the protease activity.

4.3.2. Designing consensus sequences for Cysteine Protease (PLpro) and 3C-like Serine Protease (3CLpro)

The consensus sequences for both viral proteases (Table 1) were taken out of the viral genome. Like many other viruses, SARS-CoV-2 expresses both non-structural and structural/accessory proteins in open reading frames. Papain-like cysteine protease (PLpro) and 3C-like serine protease (3CLpro) cut both distinct cleavage sites of these open reading frames to release proteins. These two signature proteases of coronaviruses are evolutionary conserved and an established measure of virus activity.[1,2] Following consensus sequences were designed to develop fluorescence-based nanobiosensors for measuring the proteolytic activity of corresponding proteases.

Table 4.1. Consensus sequences for PLpro and 3CLpro detection.

Human coronavirus protease	**Common name**	**Consensus sequence(s)**
C16.002: human coronavirus 229E papain-like peptidase 1 (PLpro)	papain-like cysteine protease (PLpro)	GAGKAAG-GKVSAG
C16.002: human coronavirus 229E papain-like peptidase 2 (PLpro)	papain-like cysteine protease (PLpro)	GAGKQGA-GDAGAG

C30.003: human coronavirus 229E main peptidase (3CLpro)	3C-like serine protease (3CLpro)	GAGVTLQ-SAESAG

4.3.3. UPLC-MS characterization of consensus peptide sequences

The UPLC-MS analysis of peptide sequences synthesized via optimized SPPS showed good high purity of peptides with minimum impurities. The PLpro1 peptide sequence showed a major peptide product at 1.77 min. The mass of the highest intensity peak was confirmed by mass spectrum taken for the relevant peak in the chromatogram (1030 Da).

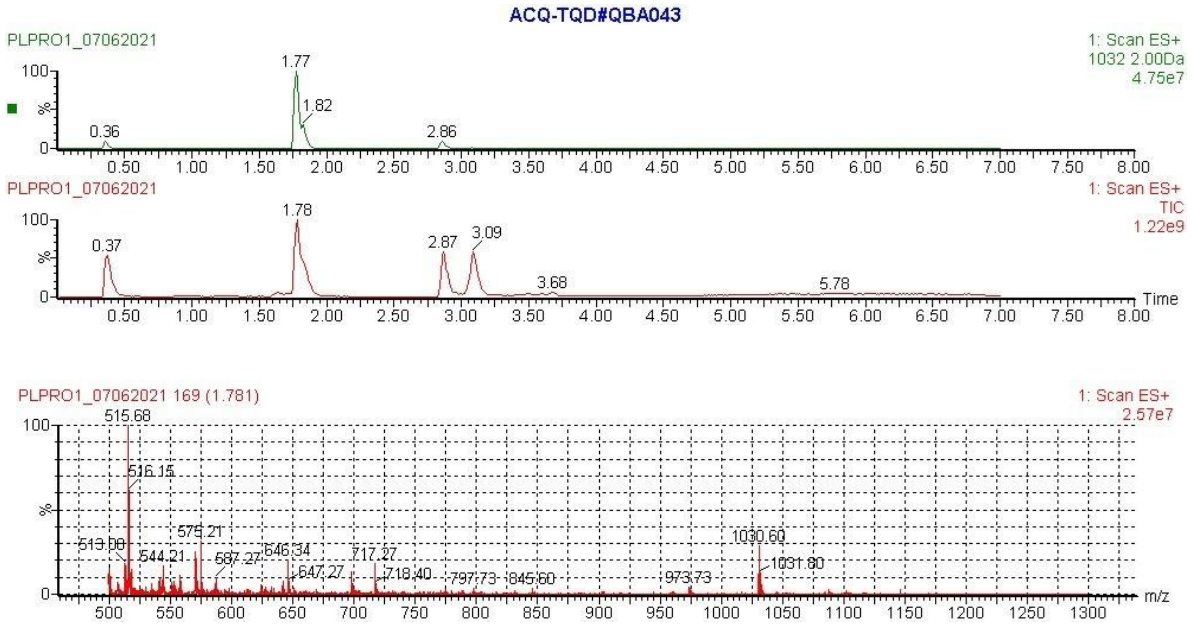


Figure 4.4. UPLC-MS analysis of PLpro1 peptide sequence.

The PLpro2 peptide sequence showed a high purity as the two peaks present in the chromatogram were the peaks corresponding the desired mass (1016 Da) of the peptide, when the mass filter was applied (**Figure 4.5**). The peaks were observed at 1.72 min and 2.17 min. The mass of both high intensity peaks was confirmed by the matching mass spectrum taken for the relevant peaks in the chromatogram (1016 Da).

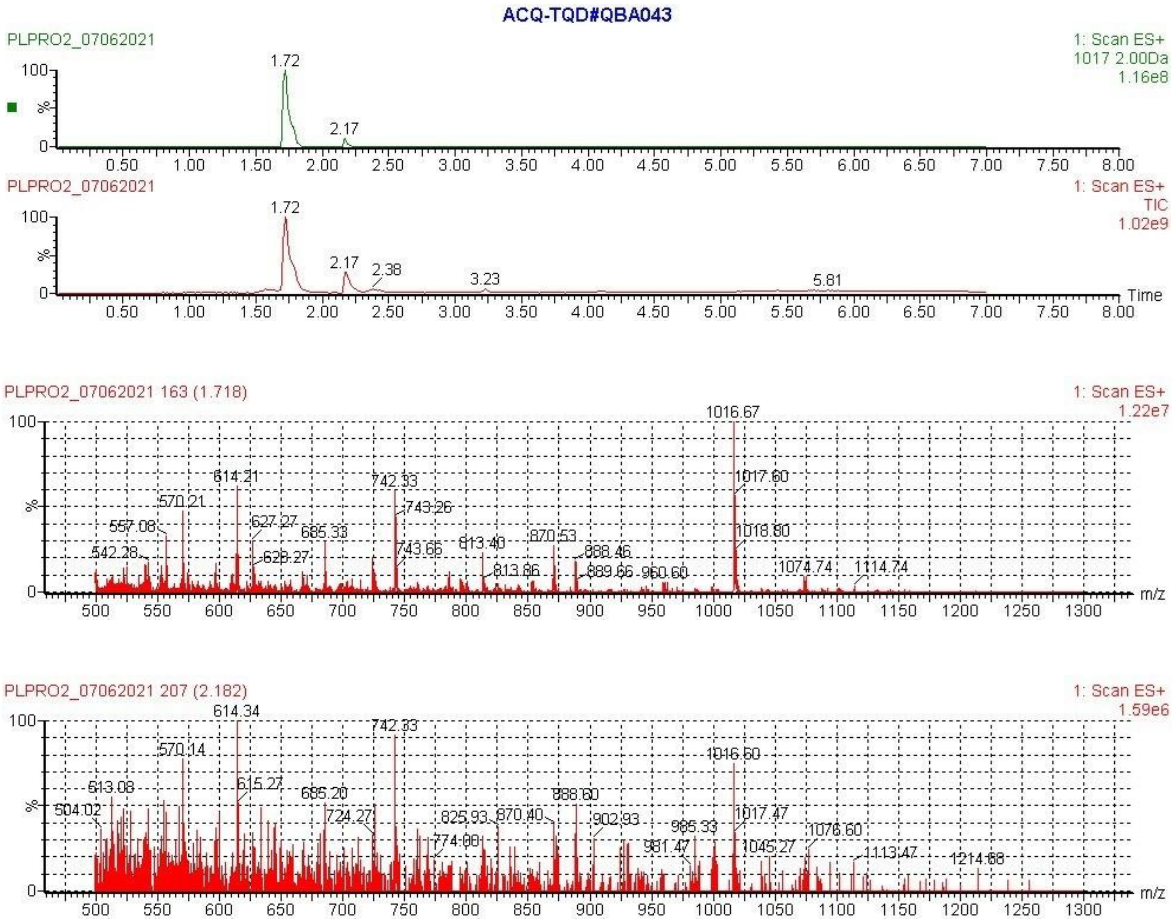
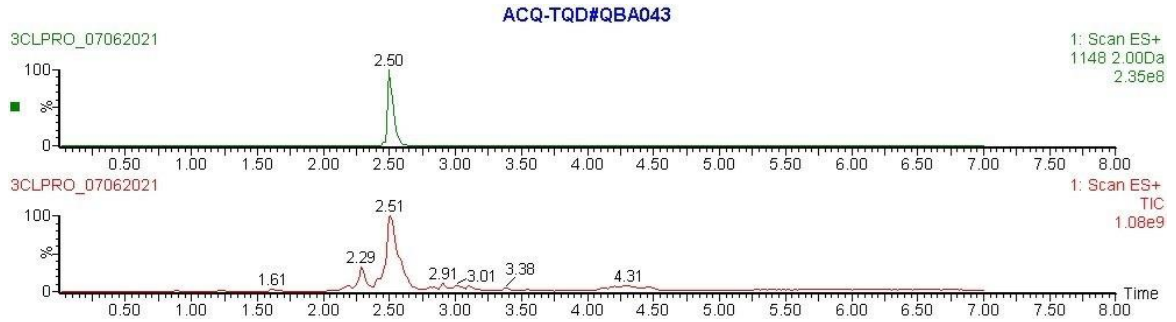


Figure 4.5. UPLC-MS analysis of PLpro2 peptide sequence.

The UPLC-MS analysis of 3CLpro peptide also showed high purity as the highest intensity peak in the chromatogram corresponds molecular mass of the peptide (1147 Da) calculated theoretically at 2.50 min. The mass of the highest intensity peak was confirmed by mass spectrum taken for the relevant peak in the chromatogram (1147 Da, **Figure 4.6**).



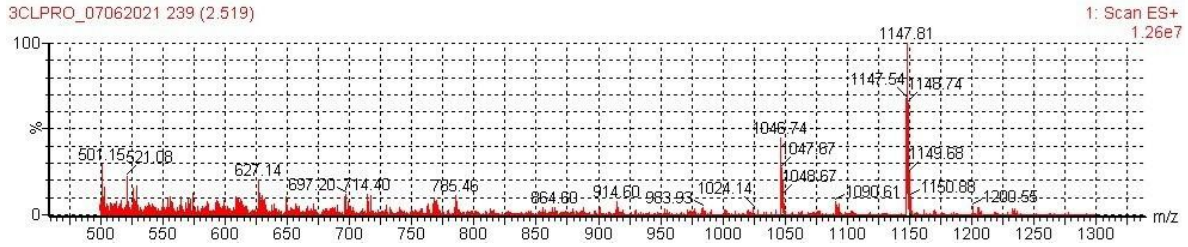


Figure 4.6. UPLC-MS analysis of 3CLpro peptide sequence.

4.3.4. Bioinformatics: CABS-dock server for flexible protein-peptide docking

By using CABS-dock, a computational service provided by the Kolinski research group at the University of Warsau, Poland[18], it was verified by Dr. Bossmann that the proposed consensus sequences are excellent fits for the active center of human coronavirus 229E main peptidase (3CLpro, monomer shown in **Figure 4.7**). CABS-dock performs docking searches and determines preferred binding sites allowing for full flexibility of the peptide and small fluctuations of the receptor backbone. CABS-dock simulations are carried out using a coarse-grained protein model, which is reverted to Protein Data Bank (PDB)-compatible coordinates once the optimization is completed.[19–23]

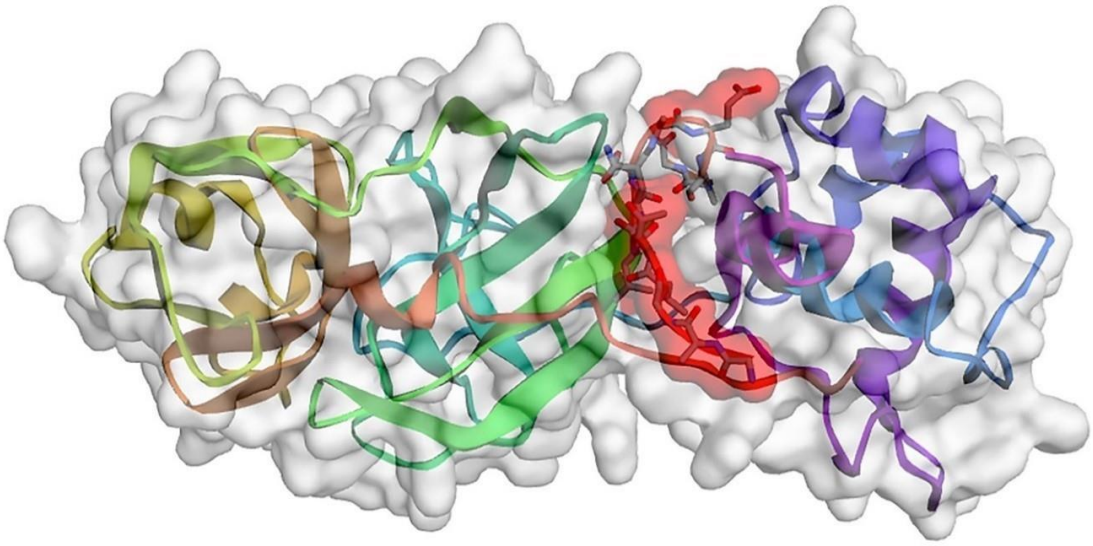
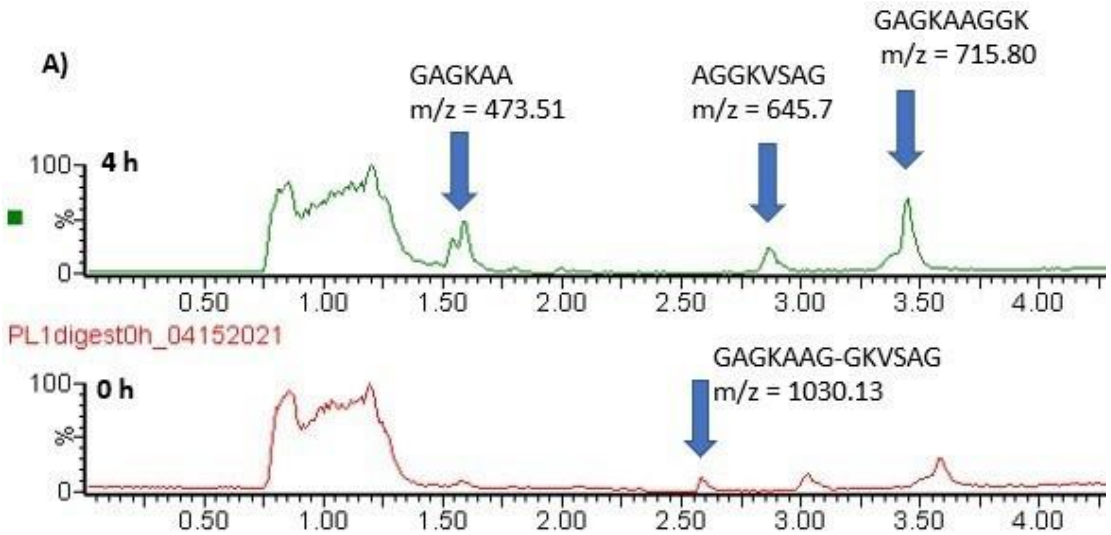


Figure 4.7. Docking Simulation (CABS-dock (20-25)) of GAGVTLQSAESAG to the active center of human coronavirus 229E main peptidase (monomer). The peptide is shown in red.

4.3.5. Reaction kinetics analysis between enzymes and peptide substrates

The reaction kinetics of the consensus sequences (shown in Table 1) was studied by performing enzymatic digestion reactions using recombinant PLpro and 3CLpro enzymes. The UPLC-MS results for digestion fractions collected at different time intervals are shown in **Figure 4.8 (A, B, C)**.



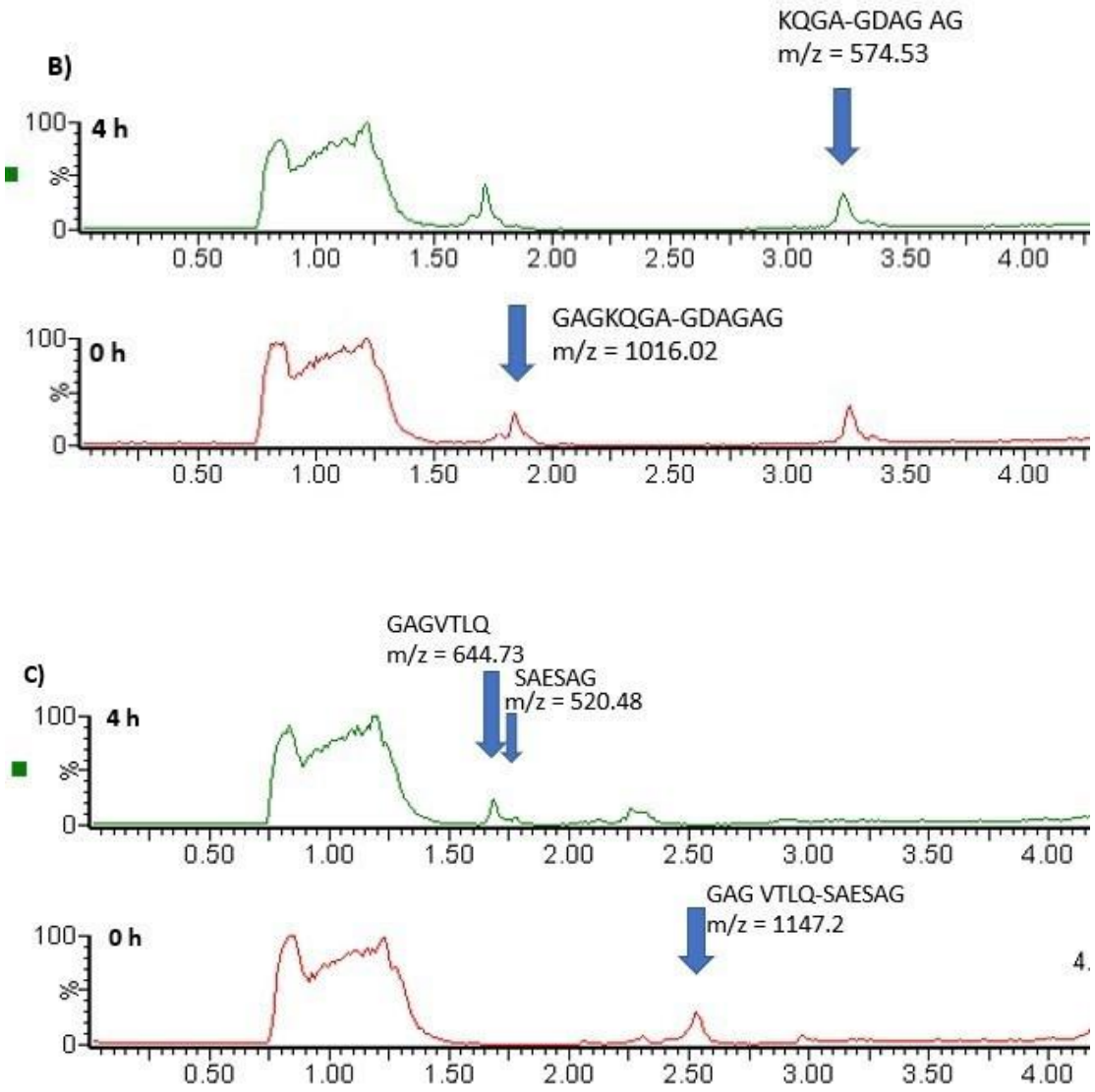


Figure 4.8. UPLC-MS (total mass range detection), signal intensity in percent vs. retention time in minutes. **A)** Digestion of GAGKAAG-GKVSAG peptide (25 nM) with PLpro enzyme (100 nM) in modified Tris buffer. **B)** Digestion of GAGKQGA-GDAGAG peptide (25 nM) with PLpro enzyme (100 nM) in modified Tris buffer. **C)** Digestion of GAGVTLQ-SAESAG peptide (25 nM) with PLpro enzyme (100 nM) in modified Tris buffer. UPLC chromatograms of digestion fractions collected at 0 and 4 hours are shown.

The UPLC-MS results from enzymatic digestion experiment on the peptide substrates proved that PLpro1 substrate was significantly cleaved by enzyme at multiple locations, hence, giving multiple digested fragments. PLpro2 substrate showed very little digestion and only one mass fragment (574.5 mass) was identified. The total ion current for different fractions was

almost similar which indicates minor chemical changes during the digestion process. The data in **Figure 4.8 (C)** proved that recombinant 3CLpro is capable of cleaving GAGVTLQ-SAESAG at the designed position. The major fragment is GAGVTLQ (m/z = 644.73). It is noteworthy that the presence of multiple digested fragments in PLpro1 and 3CLpro peptide substrates indicated the presence of a multiple cleavage site. It should be mentioned that according to our experience, the observed activity of recombinant proteases is generally several magnitudes lower than of correctly folded proteases. Therefore, relatively high concentrations of bothPLpro and 3CLpro were required to verify the cleavage sites.

Figure 4.9 shows the observed reaction kinetics of PLpro with GAGKAAGGKVSAG and GAGKQGAGDAGAG; and of 3CLpro with GAGVTLQSAESAG. This data was obtained by UPLC-MS as described above. GAGKAAG, m/z: 530.56 was detected as a major cleavage product of GAGKAAGGKVSAG by PLpro-1, as predicted. GAGKQGA, m/z: 587.62 was found when GAGKQGAGDAGAG was digested with PLpro-1. As already shown in Figure 1, GAGVTLQ, m/z: 644.73 is the major product that is cleaved by 3Clpro. In all three enzymatic reactions, the C-terminal cleavage products underwent further enzymatic degradation.

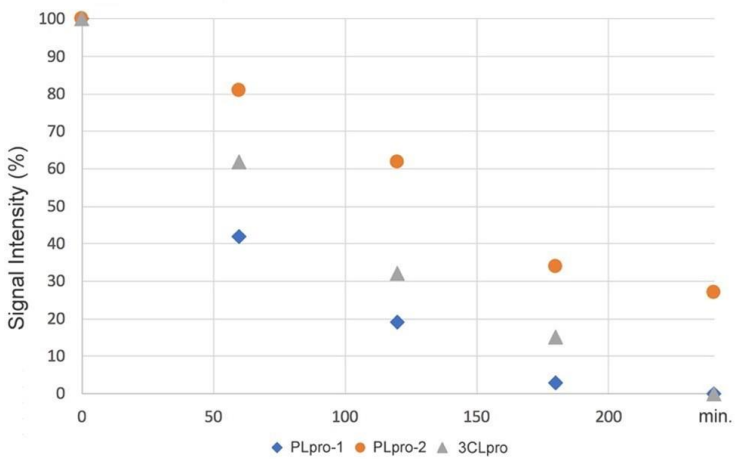


Figure 4.9. Intensities of the peptides PLpro-1: GAGKAAGGKVSAG, m/z: 1030.13;

PLpro-2: GAGKQGAGDAGAG, m/z: 1016.08; and 3CLpro: GAGVTLQSAESAG, m/z: 1147.21 as a function of reaction time in minutes. The peptides (concentration: 25 nM) were digested in the presence of their respective enzymes (100 nm) in modified Tris buffer at 37 °C.

4.3.6. Activity of iron/iron oxide-based nanobiosensors for Cysteine Protease (PLpro) and 3C-like Serine Protease (3CLpro) detection

4.3.6.1. Detection of Cysteine Protease (PLpro) activity

The data summarized in **Figure 4.10** clearly indicates that the range of use for Fe/Fe_3O_4-nanobiosensor 1 for PLpro detection is limited; though the normalized fluorescence intensity trend was random for very low concentration values, the linear trend was observed for higher concentration range only (log PLpro concentration of 10^{-11} to 10^{-6} moles L^{-1}) compared to nanobiosensors for protease[11,13,24,25] and arginase detection[26], which show sub-femtomolar limits of detection. Furthermore, the observed linear correlation ($y = 0.0235x + 0.8249$; $R^2 = 0.844$) is relatively weak, indicating that only semi-quantitative measurements of PLpro activity are possible with this nanobiosensor in the specified activity range.

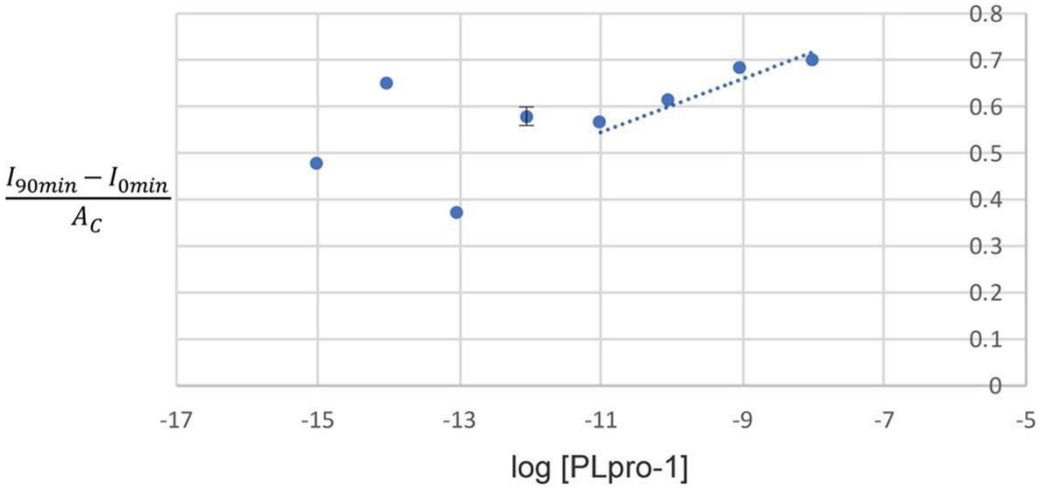


Figure 4.10. Calibration curve of the nanobiosensor 1 for Cysteine Protease (PLpro1) in HEPES buffer (pH=7.2). Plot of normalized fluorescence intensity (Synergy H1, BioTek) vs. decadic logarithm of PLpro concentration. I90min: fluorescence signal at 650nm (excitation: 420nm) after 90 min of incubation. I0min: fluorescence signal at 650nm immediately after mixing. AC: assay control: fluorescence at 650nm after 90 min. of incubation in the absence of PLpro.

Both synthesized batches performed equally well, considering the higher than usual margin of error ($\pm$ 7 %), compared to $\pm$ 3 % that was observed for nanobiosensors for the detection of human proteases.[11,24]

It should be noted thatdespite its suboptimal performance, nanobiosensor 1 for Cysteine Protease (PLpro) activity detection is the first working prototype that demonstrated that the proposed detection strategy of COVID-19 is feasible.

We were unable to calibrate the second nanobiosensor for Cysteine Protease (PLpro) detection, as indicated in **Figure 4.11**. A meaningful calibration curve could not be obtained.

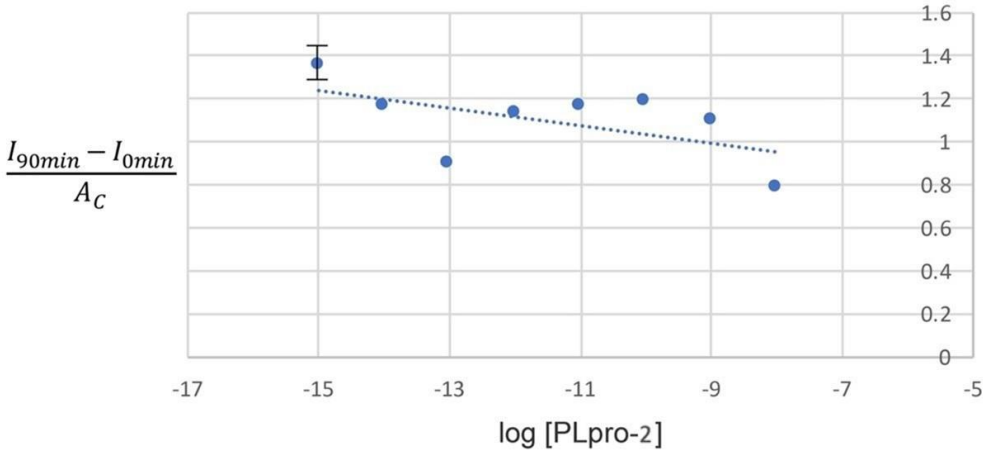


Figure 4.11. (Failed) Calibration curve of the nanobiosensor 2 for Cysteine Protease (PLpro2) in HEPES buffer (pH=7.2). Plot of normalized fluorescence intensity (Synergy H1, BioTek) vs. decadic logarithm of PLpro concentration. I90min: fluorescence signal at 650nm (excitation: 420nm) after 90 min. of incubation. I0min: fluorescence signal at 650nm immediately after mixing. AC: assay control: fluorescence at 650nm after 90 min. of incubation in the absence of PLpro.

It should be noted that in previous nanobiosensor research in the Bossmann group it was discovered that the length and the shape of the consensus sequence (amorphous vs. helical) is critical for the performance of protease nanobiosensors.[11] The mechanistic reason for this behavior is that nanoparticles exhibit both enhanced light scattering and fluorescence

quenching.[27] Therefore, the exact positions of the TCPP-fluorophore before and after proteolytic cleavage are very important for the observed increase vs. decrease of fluorescence intensity. This effect is, quite obviously, also highly concentration dependent.

4.3.6.2. Detection of 3C-like Serine Protease (3CLpro) activity

The calibration date for 3C-like Serine Protease (3CLpro) are summarized in **Figure 4.12**. This Fe/Fe_3O_4-based nanobiosensor yielded a satisfactory linear performance in the concentration range of 10^{-12} to 10^{-6} moles L^{-1} of 3CLpro in HEPES buffer (y = 0.0285x + 1.1165; R^2 = 0.9406).

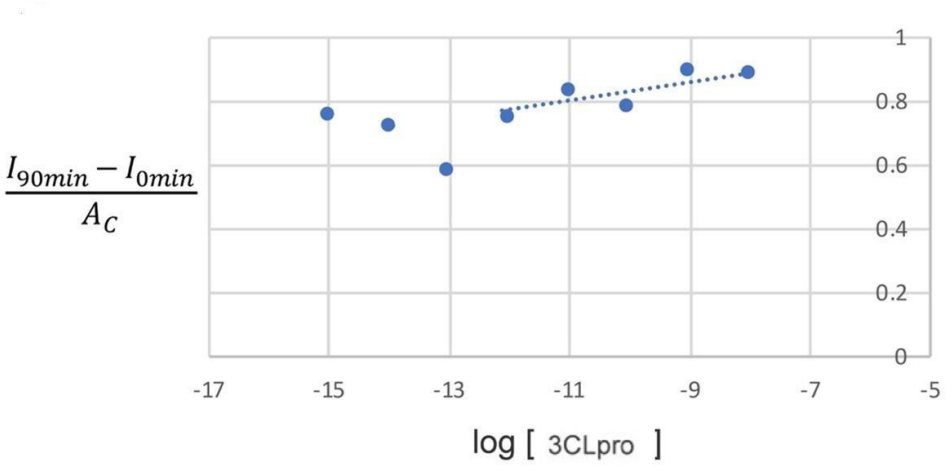


Figure 4.12. Calibration curve of the nanobiosensor for 3C-like Serine Protease (3CLpro) in HEPES buffer (pH=7.2). Plot of normalized fluorescence intensity (Synergy H1, BioTek) vs. decadic logarithm of PLpro concentration. I90min: fluorescence signal at 650nm (excitation: 420nm) after 90 min. of incubation. I0min: fluorescence signal at 650nm immediately after mixing. AC: assay control: fluorescence at 650nm after 90 min. of incubation in the absence of PLpro.

Based on our findings, both COVID-19 signature proteases can be detected, albeit with a limited concentration range. However, it is anticipated that the PLpro and 3CLpro activities will be significantly higher than of human proteases, which range usually from 10^{-14} to 10^{-10} moles L^{-1}, according to our observations.[11,13,24]

4.3.7. Graphene-based biosensors for measuring the activity of Cysteine Protease (PLpro) and 3C-like Serine Protease (3CLpro)

The Bossmann group is known for developing protease-specific nanobiosensors decorated with peptides. In addition, the group has developed nanobiosensors based on Fe/Fe_3O_4 as well as graphene core-shell for various applications.[11] Since the activity of designed dopamine-coated Fe/Fe_3O_4 did not show significant differences in the fluorescence intensity when tested with recombinant enzymes, the nanobiosensors were redesigned by using detonated graphene core-shell because graphene has superior quenching abilities compared to Fe/Fe_3O_4 core/shell nanoparticles. The use of graphene eliminates the need for the co-attached quencher (Cy 5.5). The concentration of the graphene-based nanobiosensor used to analyze the serum samples (0.03 mg/mL) was less than the concentration required by Fe/Fe_3O_4 nanobiosensors owing to higher sensitivity of graphene-nanobiosensors. The procedures for the synthesis of carboxygraphene-polyethylenimine core-shell and assembly for nanobiosensors are same as explained in **Chapter 3**.

A total of ten samples were analyzed for each group of serum samples given as 1) healthy versus (vs) drug-treated 2) healthy vs placebo 3) drug-treated vs placebo 4) healthy vs COVID-infected. The drug utilized was the experimental drug ATI-450 from Aclaris Therapeutics, Inc., which inhibits MK2 kinase. MK2 enables the production of pro- inflammatory cytokines. MK2 is also important in a process known as epithelial-to-mesenchymal transition (EMT), in which epithelial cells transform into a type of stem-like cell that helps organs develop and repairs damaged tissue.

The samples were collected on day zero (D0), day 3 (D3), end of treatment (EOT) and on day 28 (D28). The serum samples were collected in collaboration with Dr. Gregory Gan and Dr. Deepika Polineni at the University of Kansas Medical Center.

4.3.7.1. Serum Sample Assays – Graphene nanobiosensor protease activity measurements procedure

Protease activity for 3 assembled COVID related protease nanobiosensors was quantified using serum samples collected from 10 healthy volunteers and 20 COVID infected patients (group divided into two, 10 treated with placebo and 10 treated with a drug. Serum sample for COVID infected volunteers were collected over time, at days 0, 3, end of treatment (EOT), and 28. To quantify protease activity, a 0.03 mg/mL solution of graphene-based nanobiosensor was prepared in 25 µmol/L HEPES buffer enriched with 10 µmol/L of Ca^{2+}, Mg^{2+}, and Zn^{2+}, pH of 7.20. If pH for HEPES buffer or nanobiosensor solution needed to be adjusted to 7.20, a 0.1 M HCl or 0.1 M NaOH solution was used. A black 96-well plate (Greiner Bio-one, non-binding) was used to plate 3 solutions prepared to measure protease activity, which are as follows:

1. Sample control: 125 µL of HEPES buffer + 5 µL of serum sample
2. Assay control: 125 µL of 0.03 mg/mL graphene solution + 5 µl of HEPES buffer
3. Assay: 125 µL of 0.03 mg/mL graphene solution + 5 µl of serum sample (plated in triplicates)

After solutions were plated, plate was incubated at 37°C for 1 hour and then fluorescence intensity was measured using a BioTek Synergy H1 plate reader using excitation of 425 nm and emission of 650 nm.

The fluorescence measurements taken for all the groups are shown in **Figure 4.13**. Significant differences in the fluorescence intensity were observed in healthy serum samples vs drug treated samples, healthy vs placebo and healthy serum samples vs COVID-infected

samples. The differences were more prominent in the case of PLpro1 and PLpro2 as compared to 3CLpro which was prominent only in the case of healthy vs drug-treated (D28) and healthy vs placebo (D3). The significant differences among different samples within a group were statistically analyzed and those p-values are shown in **Table 4.2**. The yellow boxes indicate the significant differences of fluorescence, gray boxes indicate the borderline values while white boxes indicate non-significant differences. Potentially, the borderline values might be significant with a bigger sample size. The results concluded that the nanobiosensors were able to detect the differences in the healthy vs COVID-infected patients as well as the healthy vs drug-treated, but non-significant differences in placebo vs drug-treated samples showed that though nanobiosensors can differentiate between COVID vs healthy patients, they cannot be used to monitor the progress of an ongoing treatment of the infected patients.

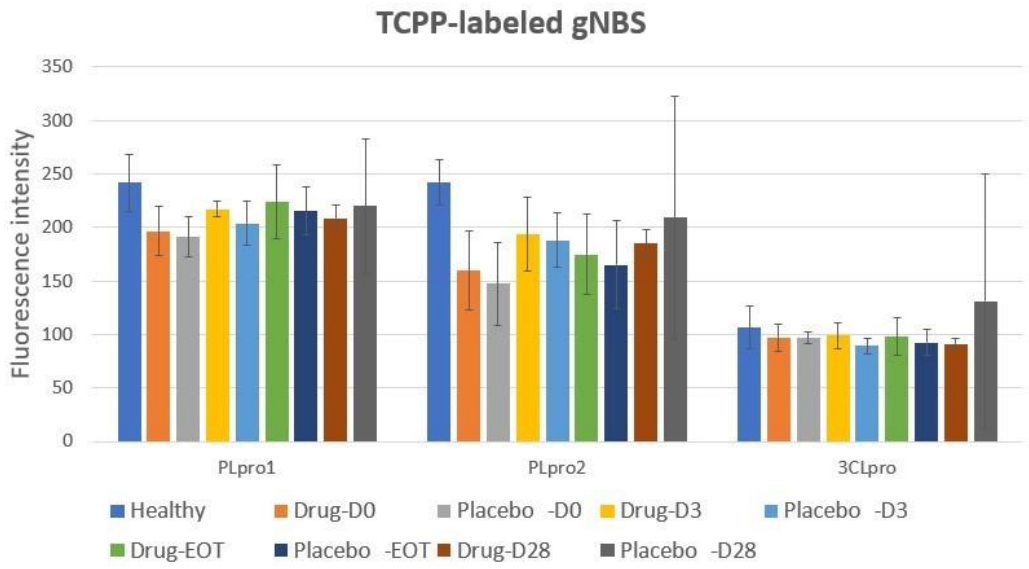


Figure 4.13. Protease activity in serum samples.

The statistical results are summarized in **Table 4.2**.

Table 4.2. Significance Table – Protease Activity differences.

	PLpro1	PLpro2	3CLpro
Healthy vs Drug-D0	7.360E-04	2.221E-05	0.23882
Healthy vs Drug-D3	0.01191	9.743E-04	0.30961
Healthy vs Drug-EOT	0.22176	2.152E-04	0.32935
Healthy vs Drug-D28	0.00392	5.547E-06	0.04248
Healthy vs Placebo-D0	1.436E-04	8.334E-06	0.17625
Healthy vs Placebo-D3	0.00338	0.00026	0.03066
Healthy vs Placebo-EOT	0.02695	1.288E-04	0.07884
Healthy vs Placebo-D28	0.34832	0.40985	0.55828
Drug-D0 vs Placebo-D0	0.54078	0.47576	0.93888
Drug-D3 vs Placebo-D3	0.11521	0.70192	0.05852
Drug-EOT vs Placebo-EOT	0.48972	0.58411	0.42278
Drug-D28 vs Placebo-D28	0.60589	0.54301	0.34670
Healthy vs COVID (D0)	1.738E-04	5.918E-09	0.18842
Healthy vs COVID (D3)	0.00546	2.350E-05	0.11745
Healthy vs COVID (EOT)	0.05112	3.888E-07	0.13981
Healthy vs COVID (D28)	0.07246	0.06257	0.76265

Yellow: $p < 0.05$; Green: $p < 0.1$ (borderline). The R software environment for statistical computing was used to extract the relevant raw data, and calculate p values.[28] The p-value is a

number, calculated from a statistical test (here R software environment), that describes how likely if the null hypothesis is true for the comparison of two data sets. P-values are used in hypothesis testing to help decide whether to reject the null hypothesis

These results suggest that:

1) The nanobiosensors for PLpro1 and PLpro2 are more suitable to detect acute COVID infections at days 0 and 3.
2) There is no clear effect of treatment with ATI-450 on the activities of PLpro1 and PLpro2, because the drug and placebos have shown comparable statistics!

In conclusion, ATI-450 is not active in suppressing the viral load during a COVID infection, as indicated by comparable PLpro1 and PLpro2 activities when ATI-450 and a placebo were given.

These first clinical results indicate that our technology works, and that COVID patients can be clearly distinguished from apparently healthy human subjects. This was the principle aim of this EAGER grant. We will offer this technology to Hawkeye Bio LLC, who are already using our explosion graphene-based nanobiosensors for the early detection of lung cancers in serum. This technology will be easily adaptable to other pandemic viruses, based on their genetic code, which will permit the elucidation of consensus sequences for their viral proteases, which are required to dissect the open reading frames that will be expressed in human cells.

4.4. Conclusion

In conclusion, this chapter describes the development of a highly efficient optical nanobiosensors for the detection of SARS-Covid 19 virus. The chapter fully describes the synthesis, assembly, characterization, and validation experiments. The nanobiosensors designed to measure the proteolytic activities of signature proteases of SARS-CoV-2 (PLpro1 and PLpro2)

at the early days of infection (day 0 and 3). The designed systems have the potential to enable the determination of its functionality and viability, yielding critical diagnostic information beyond how the SARS-CoV-2 is recognized by PCR and QPCR, techniques that are currently used in clinical setting, but are lacking the diagnostic power that is needed to determine the functionality of the virus in the host. The available PCR-based and serological methods cannot distinguish between patients with active coronavirus infection and patients that have survived the disease. Therefore, we have developed a detection system that can detect the activity of a (corona) virus infection in a host by means of a liquid biopsy using plasma, serum, sputum, or exhaled breath condensate. This technology has the potential to detect re-infected patients with high antibody titres from previous coronavirus infection(s) due to coronavirus mutation. Not only detection but the developed system can be easily employed to monitor the progress of a drug in treating lung infection. Our results indicated that the drug used to treat the viral infection (ATI-450) did not prove helpful in reducing the infection. Optical nanobiosensing of viral protease activities would permit the precise diagnosis of coronavirus activity (increasing, maximal, decreasing), which can be used to dose medications against coronavirus infections and their physiological symptoms based on actual measurements. This technology also has the advantage of detecting any new strain of the SARS-CoV virus as the consensus peptide sequence can be easily redesigned/modified as soon as the genome and genome-encoded proteases of the new viral strain are known.

4.5. Outlook: Verification the concept of SARS-CoV-2 protease activity monitoring in a COVID-19 mouse model

Based on our Human study described above, we have developed a validation and further development strategy for protease-based SARS-CoV-2 sensors in rodents, which is summarized below:

4.5.1. Calibration and validation of the optical nanobiosensors for Papain-like Cysteine Protease (PLpro) and 3C-like Serine Protease (3CLpro)

The activity of newly designed optical nanobiosensors for papain-like cysteine protease (PLpro) and 3C-like serine protease (3CLpro) will be validated using the supernatant of SARS-CoV-2 infected cell cultures and serum obtained from a SARS-CoV-2 infected mouse model, both under BSL-3 conditions. The strategy for nanobiosensor validation includes testing and optimization of the sensors in SARS-CoV-2 with infected primary cells, and a unique mouse model of SARS-CoV-2 infection.

4.5.1.1. Calibration of optical nanobiosensors

Plasmids for the expression of C16.002: human coronavirus 229E papain-like peptidase 1 (PLpro, 898-2484 of open reading frame 1 (ORF 1)) and C30.003: human coronavirus 229E main peptidase (3CLpro, 2966-3267 of ORF 1) in *E. coli* will be obtained from collaborative and commercial resources. The plasmids will be expressed in *E. coli*, as described by Zhang et al.[29] After heat disruption of the microorganisms at 70 ^{o}C, the inclusion bodies and other cell debris will be collected via centrifugation 8000 × g.[30] After dissolution in 100 mM oleyl ether (Brij-92), polyoxyethylene, PLpro or 3CLpro will be isolated, making use of their his-tags,[31] and reconstituted in HEPES-buffer containing 0.1mM Brij-92. Both enzymes will be quantified by

means of Bradford assays.[32] After quantification of both enzymes, the calibration of the optical nanobiosensors will be performed as described in the literature.[11,24]

4.5.1.2. In-vitro approach

Two highly relevant to COVID-19 cell types will be used: primary umbilical vein endothelial cells (HUVEC; PSC-100-01) and human airway epithelial cells (EpiAirway HAE cells); both commercially available from ATCC or MatTek. Cells will be grown to confluency in 6 well plates or 6 well inserts as with the HAE cells followed by infection with 10^2, 10^4, or 10^6 plaque forming units (PFU) per mL of a neon GFP-labeled SARS-CoV-2. Conditional media of control and infected cells will be used for validation. Cell culture supernatant on the apical side (HUVEC) or basolateral side (HAE cells) of the cells will be harvested on day 1, 3, 5, and 7 post infection, and cell infectivity will be checked by immunofluorescence microscopy. The viral protease activities in the supernatant will be quantitatively analyzed using the nanobiosensors for PLpro and 3CLpro, based on the calibration procedures described above.

4.5.2. Assessment of sensor performance in pre-clinical studies

Small animal studies will be conducted to assess the ability of the optical biosensor to detect the SARS-CoV-2 Protease activity monitored in infected animals. Currently, the models that are used for understating host response to SARS-CoV-2 as well as studying the modes of transmission of this virus include Collaborative Cross Mice, Syrian Hamster and Ferret as small animal models and NHP as large animal model. To conduct this study, Collaborative Cross Mouse Model of Viral Infection will be used.[33,34] Thirty-six Collaborative Cross (CC) mice will be purchased and used to conduct this study. These mice come from 8 founder strains, purposely outbred to produce over 60 genetically distinct and documented strains, in order to mimic a genetically diverse population, such as the human population, making their use in disease effects

modeling strongly translational. Work done by Galveston National Laboratory (GNL) investigators has shown that the CC001 and CC072 strain of these CC mice are ideal models for coronavirus infections, with demonstration of nearly 20% progressive weight loss by 4 days post-coronavirus infection in young CC001 mice, as compared to 5 other test strains, and nearly the same weight loss by day 3 post-infection in CC072 aged mice, (when subsequent death occurred), while other strains continued living (unpublished results). Prior published work by Dr. Ameredes has also shown a differential susceptibility of CC mouse strains to metabolic derangements of cytochrome P450 function with inhalation of toxicants[35].

4.5.2.1. Approach

CC mice of the CC001 (coronavirus susceptible) and CC003 (coronavirus resistant) strains will be infected with either SARS-CoV-2 (CC001, n=6; CC033, n=6), MERS virus (CC001, n=6; CC033, n=6), or given vehicle (CC001, n=4; CC033, n=4) as sham/infection control. Mice will be weighed and monitored daily, and will be sacrificed on day 6 post infection, or sooner, if near death, or death occurs. Bronchoalveolar lavage (BAL), blood, and lung samples will be obtained, using well-established techniques we have previously published. Briefly, BAL cells will be separated and fixed for histological characterization and differential assessment, and BAL fluid will be quantitatively analyzed by means of nanobiosenor analyses and compared with Nanostring for pro-inflammatory cytokine messages, consistent with the "cytokine storm," e.g., IL-6, that has been observed as characteristic of COVID-19. Blood samples will be obtained tail vein blood sampling that is suitable for collection of small volumes of blood, without premature sacrificing/decapitation of the animal. This will be possible because the optical nanobiosensor technology is ultrasensitive and can work with less than 3 µl of blood. Blood (50µl) will be collected daily (1-6 days post infection) and allow the blood to clot; cell-

free serum will be obtained by filtration (passing through 20µm or 45µm cut-off filters). As with the BAL fluid, Nanostring analyses will be performed for comparison with quantitative nanobiosensor results. Whole lung tissue specimens will be collected at the end of each experiment, with portions being formalin fixed, sectioned, and analyzed histologically, to determine potential airway pathology (e.g., derangement or necrotic destruction of airway epithelium). Data variables will be tested for normality and equal variance, and compared with ANOVA, with discreet differences between treatment groups compared using post-hoc testing with either Tukey's or Student-Newman-Keuls method. If group data fails the normality or equal variance test, comparisons will be made with non-parametric test analogs. All tests will be considered significant at the 0.05 level of significance.

4.5.2.2. Expected results and potential alternatives

As suggested by data with coronavirus infection in CC mice, it is expected that the CC001 mouse strain will be particularly susceptible to both SARS-CoV-2 and MERS, as compared to sham/vehicle controls, but that the nanobiosensor analysis will be able to differentiate between infection with SARS-CoV-2 and MERS. If results are negative in discerning susceptibility to SARS-CoV-2 in the CC001 strain, the CC072 strain will be utilized to test SARS-CoV-2 susceptibility, based on our original preliminary data findings with coronavirus infection. If MERS fails to induce testable infection-associated pathology as a nanobiosensor differentiating comparison with SARS-CoV-2, SARS-CoV-1 may also be considered for comparison. SARS-CoV-1 is a select agent and will require obtaining of specific permissions for its use, which will be pursued, if necessary.

Based on the paradigm that the viral protease activities of SARS-CoV-2 first increase, then reach a plateau, and then eventually decrease, we anticipate that we'll be able to follow

virus maturation in host animals. This will be useful to guide clinical interventions in ICU patients, as well as to understand the maturation of the virus in patients with little or no discernible symptoms.

4.5.3. Broader impacts

The optical nanobiosensor technology will provide virologists and clinicians with an important tool that possess unprecedented capabilities. For the first time, the (reproductive) activity of SARS-CoV-2 will be measurable in mammalian and human host organisms. This will enable *in-vivo* studies, in which the dependence of viral activity on clinical measures can be precisely monitored. This will aid clinicians in precisely dosing viral medication. This technology will also permit the detection of SARS-CoV-2 recurrence in patients with partial immunity, due to previous corona virus infections other than SARS-CoV-2, or the loss of immunity after SARS-CoV-2 infection. Furthermore, this platform technology can be adapted to virtually any virus of clinical or academic importance. Principally, this technology can be extended to studying the activity of self-replicating RNA vaccines[36], which rely – in analogy to RNA viruses – on viral proteases to achieve self-replication. This approach will allow the determination of candidates for a self-replicating RNA vaccine, which will permit the selection of vaccine candidates based on biophysical measurements. This approach will help reduce the required time for vaccine development approx. in half.

4.6. References

(1) Yang, X.; Yu, Y.; Xu, J.; Shu, H.; Xia, J.; Liu, H.; Wu, Y.; Zhang, L.; Yu, Z.; Fang, M.; Yu, T.; Wang, Y.; Pan, S.; Zou, X.; Yuan, S.; Shang, Y. Clinical Course and Outcomes of Critically Ill Patients with SARS-CoV-2 Pneumonia in Wuhan, China: A Single-Centered, Retrospective, Observational Study. *Lancet Respir Med* **2020**, *8* (5), 475–481. https://doi.org/10.1016/S2213-2600(20)30079-5.

(2) Bai, Y.; Yao, L.; Wei, T.; Tian, F.; Jin, D.-Y.; Chen, L.; Wang, M. Presumed Asymptomatic Carrier Transmission of COVID-19. *JAMA* **2020**, *323* (14), 1406–1407. https://doi.org/10.1001/jama.2020.2565.

(3) Arya, M.; Shergill, I. S.; Williamson, M.; Gommersall, L.; Arya, N.; Patel, H. R. H. Basic Principles of Real-Time Quantitative PCR. *Expert Rev Mol Diagn* **2005**, *5* (2), 209–219. https://doi.org/10.1586/14737159.5.2.209.

(4) *Detect COVID-19 in as Little as 5 Minutes*. Abbott. https://www.abbott.com/corpnewsroom/diagnostics-testing/detect-covid-19-in-as-little-as-5-minutes.html (accessed 2022-06-08).

(5) *coronavirus-disease-2019-covid-19. - UpToDate*. https://www.uptodate.com/contents/search?search=coronavirus-disease-2019-covid-19.&sp=0&searchType=PLAIN_TEXT&source=USER_INPUT&searchControl=TOP_PULLDOWN&searchOffset=1&autoComplete=false&language=&max=0&index=&autoCompleteTerm=&rawSentence= (accessed 2022-06-08).

(6) Kiselev, D.; Matsvay, A.; Abramov, I.; Dedkov, V.; Shipulin, G.; Khafizov, K. Current Trends in Diagnostics of Viral Infections of Unknown Etiology. *Viruses* **2020**, *12* (2), 211. https://doi.org/10.3390/v12020211.

(7) SARS-CoV-2 Diagnostic Pipeline. *FIND*.

(8) Itokawa, K.; Sekizuka, T.; Hashino, M.; Tanaka, R.; Kuroda, M. A Proposal of an Alternative Primer for the ARTIC Network's Multiplex PCR to Improve Coverage of SARS-CoV-2 Genome Sequencing. 8.

(9) Sah, R.; Rodriguez-Morales, A. J.; Jha, R.; Chu, D. K. W.; Gu, H.; Peiris, M.; Bastola, A.; Lal, B. K.; Ojha, H. C.; Rabaan, A. A.; Zambrano, L. I.; Costello, A.; Morita, K.; Pandey, B. D.; Poon, L. L. M. Complete Genome Sequence of a 2019 Novel Coronavirus (SARS-CoV-2) Strain Isolated in Nepal. *Microbiol Resour Announc* **2020**, *9* (11), e00169-20. https://doi.org/10.1128/MRA.00169-20.

(10) Zumla, A.; Chan, J. F. W.; Azhar, E. I.; Hui, D. S. C.; Yuen, K.-Y. Coronaviruses — Drug Discovery and Therapeutic Options. *Nat Rev Drug Discov* **2016**, *15* (5), 327–347. https://doi.org/10.1038/nrd.2015.37.

(11) Wang, H.; Udukala, D. N.; Samarakoon, T. N.; Basel, M. T.; Kalita, M.; Abayaweera, G.; Manawadu, H.; Malalasekera, A.; Robinson, C.; Villanueva, D.; Maynez, P.; Bossmann, L.; Riedy, E.; Barriga, J.; Wang, N.; Li, P.; Higgins, D. A.; Zhu, G.; Troyer, D. L.; Bossmann, S. H. Nanoplatforms for Highly Sensitive Fluorescence Detection of Cancer-Related Proteases. *Photochem. Photobiol. Sci.* **2014**, *13* (2), 231. https://doi.org/10.1039/c3pp50260k.

(12) Covarrubias-Zambrano, O.; Motamedi, M.; Ameredes, B. T.; Tian, B.; Calhoun, W. J.; Zhao, Y.; Brasier, A. R.; Kalubowilage, M.; Malalasekera, A. P.; Yapa, A. S.; Wang, H.; Culbertson, C. T.; Troyer, D. L.; Bossmann, S. H. Optical Biosensing of Markers of Mucosal Inflammation. *Nanomedicine: Nanotechnology, Biology and Medicine* **2022**, *40*, 102476. https://doi.org/10.1016/j.nano.2021.102476.

(13) Udukala, D. N.; Wendel, S. O.; Wang, H.; Yapa, A. S.; Covarrubias-Zambrano, O.; Janik, K.; Gadbury, G.; Troyer, D. L.; Bossmann, S. H. Early Detection of Non-Small Cell Lung Cancer in Liquid Biopsies by Ultrasensitive Protease Activity Analysis. *Journal of Cancer Metastasis and Treatment* **2020**, *6*, 25. https://doi.org/10.20517/2394-4722.2020.45.

(14) *Early detection of pancreatic cancers by novel nanobiosensor-based protease biomarkers using hierarchical decision structure. | Journal of Clinical Oncology*. https://ascopubs.org/doi/abs/10.1200/JCO.2021.39.15_suppl.e16273 (accessed 2022-01-18).

(15) Xu, C.; Xu, K.; Gu, H.; Zheng, R.; Liu, H.; Zhang, X.; Guo, Z.; Xu, B. Dopamine as A Robust Anchor to Immobilize Functional Molecules on the Iron Oxide Shell of Magnetic Nanoparticles. *J. Am. Chem. Soc.* **2004**, *126* (32), 9938–9939. https://doi.org/10.1021/ja0464802.

(16) Ishihara, S.; Labuta, J.; Rossom, W. V.; Ishikawa, D.; Minami, K.; Hill, J. P.; Ariga, K. Porphyrin-Based Sensor Nanoarchitectonics in Diverse Physical Detection Modes. *Phys. Chem. Chem. Phys.* **2014**, *16* (21), 9713–9746. https://doi.org/10.1039/C3CP55431G.

(17) Sabanayagam, C. R.; Eid, J. S.; Meller, A. Using Fluorescence Resonance Energy Transfer to Measure Distances along Individual DNA Molecules: Corrections Due to Nonideal Transfer. *J. Chem. Phys.* **2005**, *122* (6), 061103. https://doi.org/10.1063/1.1854120.

(18) *Laboratory of Theory of Biopolymers*. http://biocomp.chem.uw.edu.pl/ (accessed 2022-06-11).

(19) *Flexible docking of peptides to proteins using CABS-dock - PMC*. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6933849/ (accessed 2022-06-11).

(20) Ciemny, M. P.; Kurcinski, M.; Kozak, K. J.; Kolinski, A.; Kmiecik, S. Highly Flexible Protein-Peptide Docking Using CABS-Dock. *Methods Mol Biol* **2017**, *1561*, 69–94. https://doi.org/10.1007/978-1-4939-6798-8_6.

(21) Blaszczyk, M.; Kurcinski, M.; Kouza, M.; Wieteska, L.; Debinski, A.; Kolinski, A.; Kmiecik, S. Modeling of Protein-Peptide Interactions Using the CABS-Dock Web Server for Binding Site Search and Flexible Docking. *Methods* **2016**, *93*, 72–83. https://doi.org/10.1016/j.ymeth.2015.07.004.

(22) Ciemny, M. P.; Debinski, A.; Paczkowska, M.; Kolinski, A.; Kurcinski, M.; Kmiecik, S. Protein-Peptide Molecular Docking with Large-Scale Conformational Changes: The P53-MDM2 Interaction. *Sci Rep* **2016**, *6* (1), 37532. https://doi.org/10.1038/srep37532.

(23) Kurcinski, M.; Jamroz, M.; Blaszczyk, M.; Kolinski, A.; Kmiecik, S. CABS-Dock Web Server for the Flexible Docking of Peptides to Proteins without Prior Knowledge of the Binding Site. *Nucleic Acids Res* **2015**, *43* (W1), W419-424. https://doi.org/10.1093/nar/gkv456.

(24) Udukala, D. N.; Wang, H.; Wendel, S. O.; Malalasekera, A. P.; Samarakoon, T. N.; Yapa, A. S.; Abayaweera, G.; Basel, M. T.; Maynez, P.; Ortega, R.; Toledo, Y.; Bossmann, L.; Robinson, C.; Janik, K. E.; Koper, O. B.; Li, P.; Motamedi, M.; Higgins, D. A.; Gadbury, G.; Zhu, G.; Troyer, D. L.; Bossmann, S. H. Early Breast Cancer Screening Using Iron/Iron Oxide-Based Nanoplatforms with Sub-Femtomolar Limits of Detection. *Beilstein Journal of Nanotechnology* **2016**, *7* (1), 364–373. https://doi.org/10.3762/bjnano.7.33.

(25) Kalubowilage, M.; Covarrubias-Zambrano, O.; Malalasekera, A. P.; Wendel, S. O.; Wang, H.; Yapa, A. S.; Chlebanowski, L.; Toledo, Y.; Ortega, R.; Janik, K. E.; Shrestha, T. B.; Culbertson, C. T.; Kasi, A.; Williamson, S.; Troyer, D. L.; Bossmann, S. H. Early Detection of Pancreatic Cancers in Liquid Biopsies by Ultrasensitive Fluorescence Nanobiosensors. *Nanomedicine: Nanotechnology, Biology and Medicine* **2018**, *14* (6), 1823–1832. https://doi.org/10.1016/j.nano.2018.04.020.

(26) Malalasekera, A. P.; Wang, H.; Samarakoon, T. N.; Udukala, D. N.; Yapa, A. S.; Ortega, R.; Shrestha, T. B.; Alshetaiwi, H.; McLaurin, E. J.; Troyer, D. L.; Bossmann, S. H. A Nanobiosensor for the Detection of Arginase Activity. *Nanomedicine* **2017**, *13* (2), 383–390. https://doi.org/10.1016/j.nano.2016.08.014.

(27) Kalita, M.; Basel, M. T.; Janik, K.; Bossmann, S. H. Optical and Electronic Properties of Metal and Semiconductor Nanostructures. In *Nanoscale Materials in Chemistry*; John Wiley & Sons, Ltd, 2009; pp 537–578. https://doi.org/10.1002/9780470523674.ch16.

(28) Welch, B. L. The Generalization of `Student's' Problem When Several Different Population Variances Are Involved. *Biometrika* **1947**, *34* (1/2), 28–35. https://doi.org/10.2307/2332510.

(29) Zhang, L.; Lin, D.; Sun, X.; Curth, U.; Drosten, C.; Sauerhering, L.; Becker, S.; Rox, K.; Hilgenfeld, R. Crystal Structure of SARS-CoV-2 Main Protease Provides a Basis for Design of Improved α-Ketoamide Inhibitors. *Science* **2020**, *368* (6489), 409–412. https://doi.org/10.1126/science.abb3405.

(30) Marqus, S.; Pirogova, E.; Piva, T. J. Evaluation of the Use of Therapeutic Peptides for Cancer Treatment. *J Biomed Sci* **2017**, *24* (1), 21. https://doi.org/10.1186/s12929-017-0328-x.

(31) Kimple, M. E.; Brill, A. L.; Pasker, R. L. Overview of Affinity Tags for Protein Purification. *Current Protocols in Protein Science* **2013**, *73* (1), 9.9.1-9.9.23. https://doi.org/10.1002/0471140864.ps0909s73.

(32) Wendel, S. O.; Perera, A. S.; Pfromm, P. H.; Czermak, P.; Bossmann, S. H. Adaptation of Mycobacterium Smegmatis to an Industrial Scale Medium and Isolation of the Mycobacterial PorinMspA. *Open Microbiol J* **2013**, *7*, 92–98. https://doi.org/10.2174/1874285801307010092.

(33) Roberts, A.; Paddock, C.; Vogel, L.; Butler, E.; Zaki, S.; Subbarao, K. Aged BALB/c Mice as a Model for Increased Severity of Severe Acute Respiratory Syndrome in Elderly Humans. *J Virol* **2005**, *79* (9), 5833–5838. https://doi.org/10.1128/JVI.79.9.5833-5838.2005.

(34) Roberts, A.; Lamirande, E. W.; Vogel, L.; Jackson, J. P.; Paddock, C. D.; Guarner, J.; Zaki, S. R.; Sheahan, T.; Baric, R.; Subbarao, K. Animal Models and Vaccines for SARS-CoV Infection. *Virus Res* **2008**, *133* (1), 20–32. https://doi.org/10.1016/j.virusres.2007.03.025.

(35) Hartman, J. H.; Miller, G. P.; Caro, A. A.; Byrum, S. D.; Orr, L. M.; Mackintosh, S. G.; Tackett, A. J.; MacMillan-Crow, L. A.; Hallberg, L. M.; Ameredes, B. T.; Boysen, G. 1,3-Butadiene-Induced Mitochondrial Dysfunction Is Correlated with Mitochondrial CYP2E1 Activity in Collaborative Cross Mice. *Toxicology* **2017**, *378*, 114–124. https://doi.org/10.1016/j.tox.2017.01.005.

(36) Rauch, S.; Jasny, E.; Schmidt, K. E.; Petsch, B. New Vaccine Technologies to Combat Outbreak Situations. *Front Immunol* **2018**, *9*, 1963. https://doi.org/10.3389/fimmu.2018.01963.

Chapter 5 - Summary and Future Outlook

Cancer and COVID-19 are two ongoing battles that have led to the loss of many precious lives. Cancer alone is one of the main causes of death on the planet, with the highest mortality caused by lung cancer. The traditional diagnostic methods for lung cancer detection include endoscopy, tissue biopsies, X-ray scans, magnetic resonance imaging, computed tomography, and positron emission tomography.[1] However, these diagnostics may not always indicate early-stage tumors, are not practical for repeated screenings to evaluate the treatment progress and are not accessible to a large population.[2] There is a critical need for improved clinical diagnostic tools and procedures that are sensitive, reliable, rapid, non-invasive, and cost-effective.

Similarly, the COVID-19 pandemic caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has changed the world entirely since it started. Every individual has been affected by this deadly virus one way or the other.[3] There is a great need for a detection system that can give accurate results with no false positives and efficiently diagnose the infected patients with the least resources.

Our study describes the successful development of detection systems based on peptides to detect lung cancer and coronavirus diseases. Before developing the systems, we optimized the solid-phase peptide synthesis (SPPS) procedure condition to get highly pure peptides. We identified that the fluorenylmethoxycarbonyl (Fmoc) removal step in SPPS was not efficiently removing the Fmoc group to get the amine group on peptide precursor deprotected. Two secondary amine bases were tested for different concentrations, time of incubation, number of incubations, and processing temperatures to give maximal removal of the Fmoc group. The ultra-performance liquid chromatography-mass spectrometry (UPLC-MS) analysis revealed that 20% piperidine (in dimethylformamide, DMF) incubated for 1 minute followed by 10 minutes (both incubations at

room temperature) was able to remove the Fmoc group completely. Optimizing the conditions for the amine deprotection/Fmoc removal step significantly increased the purity of the desired peptide products as complete removal of the Fmoc group terminated the chances of formation of incomplete (undesired) peptide fragments during the synthesis of longer chains of amino acids. After the optimization of the peptide synthesis protocol, we applied the peptides to develop detection systems for the detection of lung cancer disease and COVID-19 virus.

The design of the biosensors consists of cleavable peptide sequence specific to the protease biomarker of interest. The C-terminal of the peptide is attached to the graphene core shell via a polyethylenimine linker through an amide bond, while the N-terminal is attached to tetrakis-carboxyphenyl-porphyrin (TCPP). The working principle is based on the Förster resonance energy transfer, where the fluorescence of TCPP dye is quenched when it is in closer proximity to graphene due to its high optical absorption properties.[4] This quenching is terminated when the specific protease from the serum sample (liquid biopsy) cleaves the consensus peptide substrate because of the increased distance between the fluorophore and graphene. A total of 19 nanobiosensors are developed to detect 19 lung cancer protease biomarkers related to lung cancer. The detection platforms detect up to femtomolar concentrations of the proteases with high precision in a quantitative manner. To detect coronavirus disease, three biosensors are developed to detect coronavirus-specific proteases. Out of three biosensors, two biosensors (papain-like cysteine protease, PLpro1 and Plpro2) have shown promising results in detecting the coronavirus infection, specifically on day 0 and day 3. It was also found that the drug used to treat the viral infection (ATI-450) did not treat the infection as there were no significant differences between the fluorescence intensities in drug-treated versus placebo-treated patients.

The chapters discussed in this thesis mainly discuss the use of peptides in disease detection applications, but the similar concept can also be used in cancer therapy and the treatment of other deadly diseases like COVID-19. The high selectivity of the peptides towards their specific proteases/receptors could potentially be used to target the receptors at the tumor sites to deliver the desired payloads. Concerning future work, a peptide-based platform will be designed to deliver an anticancer drug, SN-38 (active metabolite of irinotecan), inside the tumor cells to a specific organelle, mitochondria, to reverse multi-drug resistance[5] in cancer therapy.

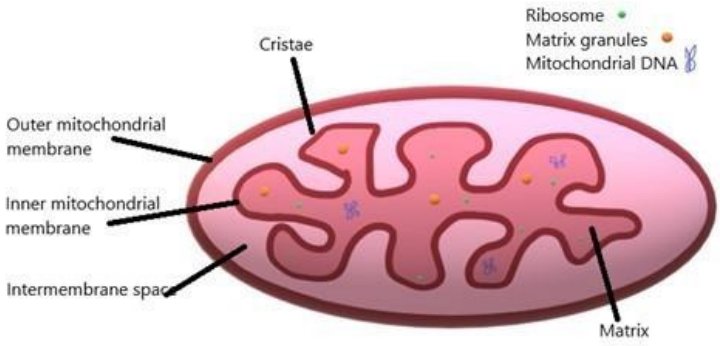


Figure 5.1. Mitochondria[5] (regenerated with permission from ref 5).

5.1. References

(1) *Stage I Lung Cancer*. WebMD. https://www.webmd.com/lung-cancer/guide/lung-cancer-stage-i-overview (accessed 2022-06-22).

(2) Walter, F. M.; Thompson, M. J.; Wellwood, I.; Abel, G. A.; Hamilton, W.; Johnson, M.; Lyratzopoulos, G.; Messenger, M. P.; Neal, R. D.; Rubin, G.; Singh, H.; Spencer, A.; Sutton, S.; Vedsted, P.; Emery, J. D. Evaluating Diagnostic Strategies for Early Detection of Cancer: The CanTest Framework. *BMC Cancer* **2019**, *19* (1), 586. https://doi.org/10.1186/s12885-019-5746-6.

(3) Kesarwani, V.; Gupta, R.; Vetukuri, R. R.; Kushwaha, S. K.; Gandhi, S. Identification of Unique Peptides for SARS-CoV-2 Diagnostics and Vaccine Development by an In Silico Proteomics Approach. *Frontiers in Immunology* **2021**, *12*.

(4) Udukala, D. N.; Wang, H.; Wendel, S. O.; Malalasekera, A. P.; Samarakoon, T. N.; Yapa, A. S.; Abayaweera, G.; Basel, M. T.; Maynez, P.; Ortega, R.; Toledo, Y.; Bossmann, L.; Robinson, C.; Janik, K. E.; Koper, O. B.; Li, P.; Motamedi, M.; Higgins, D. A.; Gadbury, G.; Zhu, G.; Troyer, D. L.; Bossmann, S. H. Early Breast Cancer Screening Using Iron/Iron Oxide-Based Nanoplatforms with Sub-Femtomolar Limits of Detection. *Beilstein Journal of Nanotechnology* **2016**, *7* (1), 364–373. https://doi.org/10.3762/bjnano.7.33.

(5) Ehsan, S.; Covarrubias-Zambrano, O.; Bossmann, S. H. Mitochondrial Targeting Peptide-Based Nanodelivery for Cancer Treatment. *Curr Protein Pept Sci* **2022**. https://doi.org/10.2174/1389203723666220520160435.

Appendix_ Chapter 2 (S)

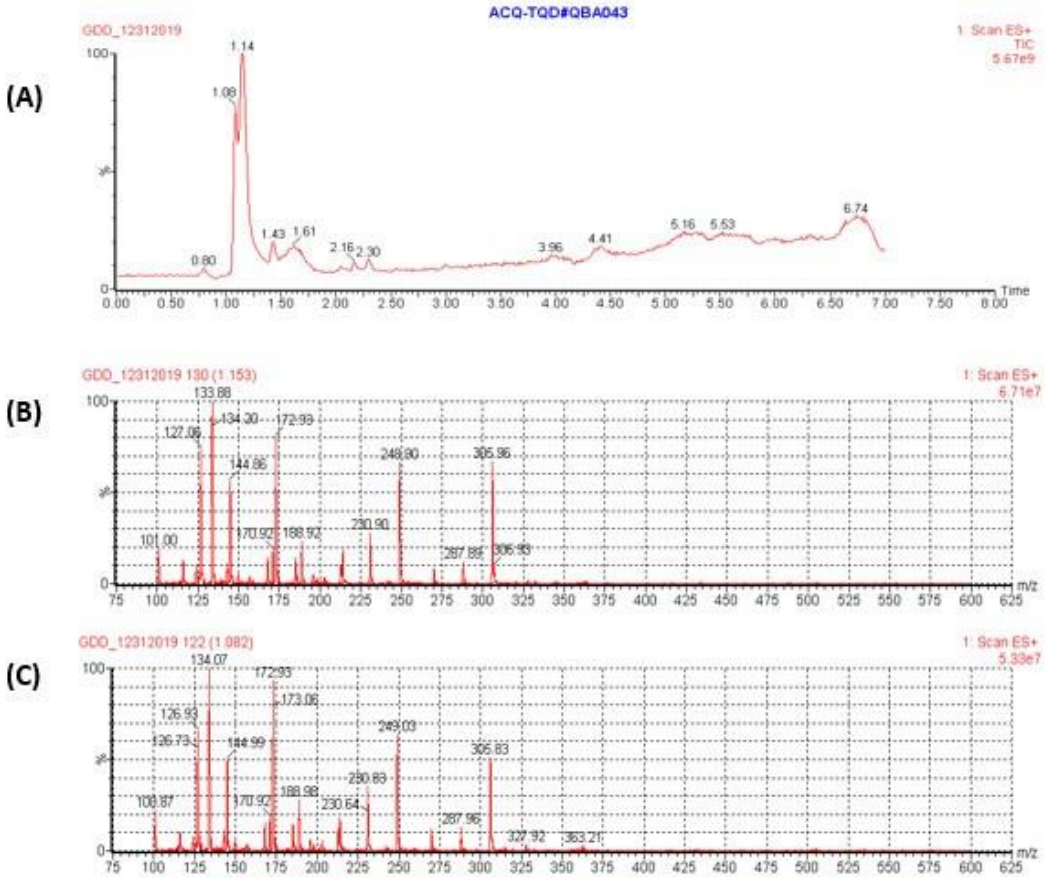


Figure S1. (A) UPLC chromatogram of GDD peptide. The peaks at 1.15 min and 1.08 min correspond to GDD peptide product, **(B)** and **(C)** Mass spectra of the peaks eluted at 1.08 min and 1.15 min, both corresponding to the mass (M+1 mass peak) of GDD peptide.

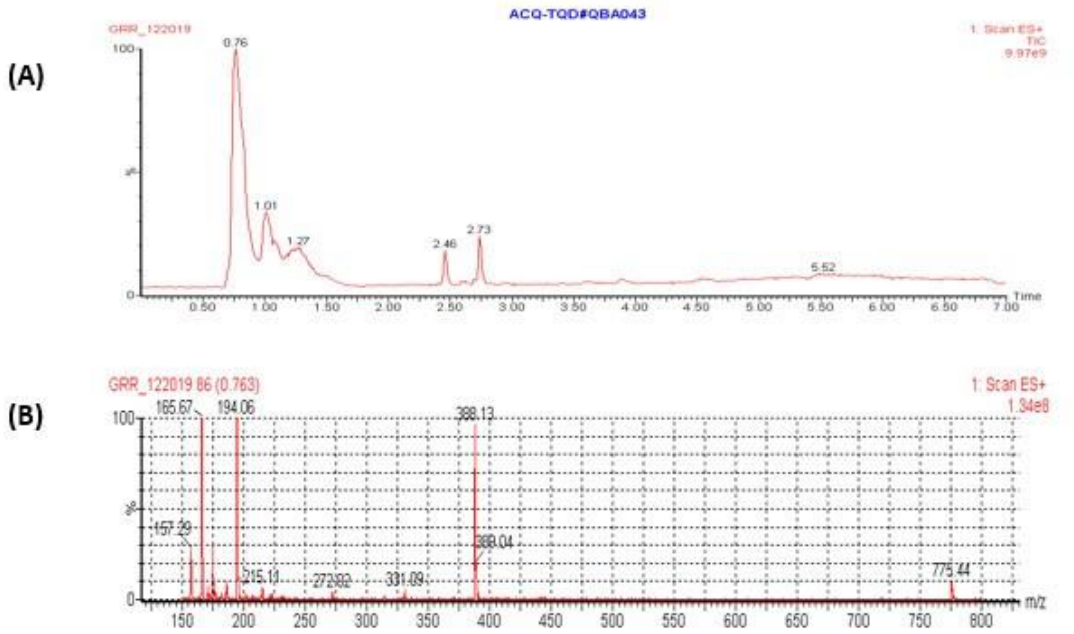


Figure S2. (A) UPLC chromatogram of GRR peptide. The peak at 0.76 min corresponds to GRR peptide product, **(B)** Mass spectrum of the peak eluted at 0.76 min corresponding to the mass (M+1 mass peak, 388 m/z) of GRR peptide.

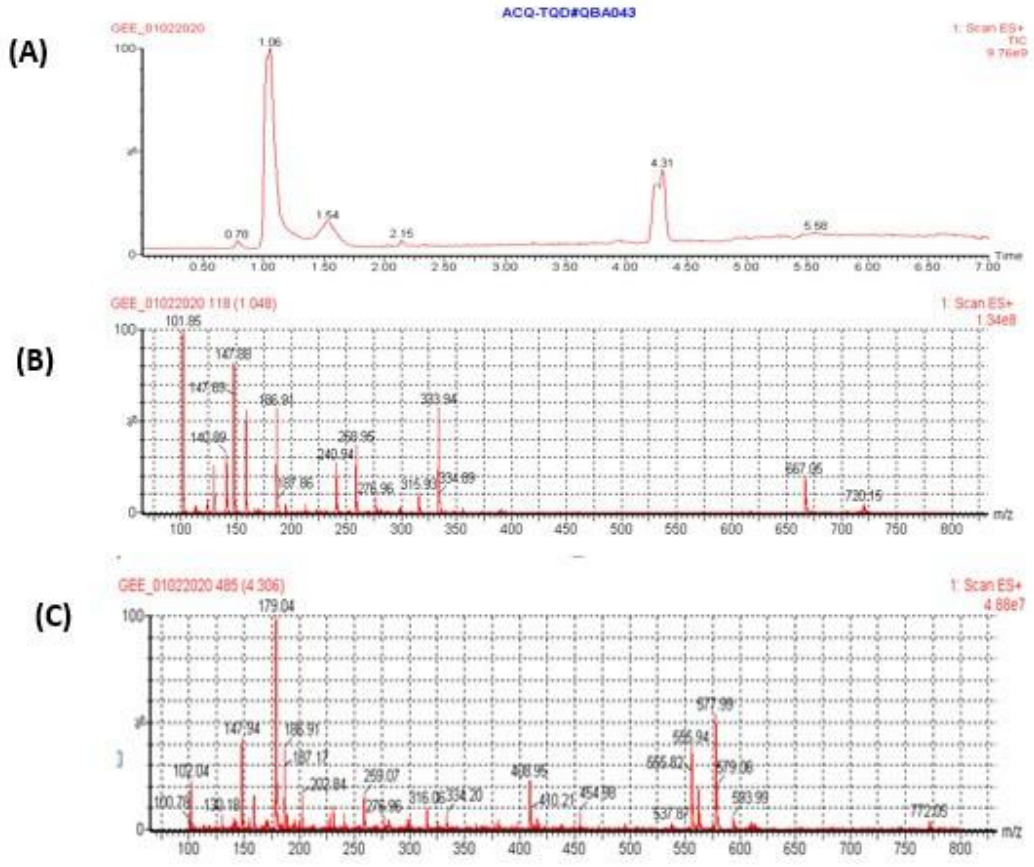


Figure S3. (A) UPLC chromatogram of GEE peptide. The peak at 1.06 min corresponds to GEE peptide product and the peak at 4.31 min corresponds to the peptide with the Fmoc group **(B)** Mass spectrum of the peak eluted at 1.04 min corresponding to the mass (M+1 mass peak, 334 m/z) of GEE peptide, **(C)** Mass spectrum of the peak eluted at 4.31 min corresponding to the peptide with Fmoc-protected peptide (537 m/z).

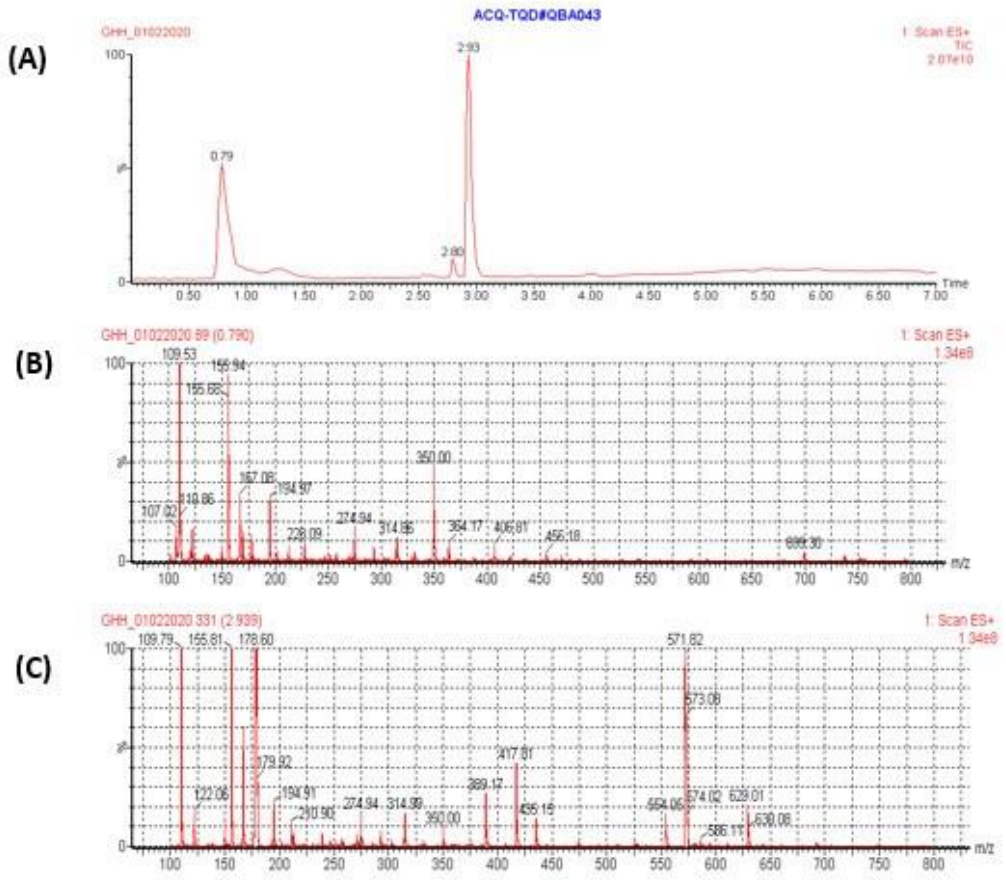


Figure S4. (A) UPLC chromatogram of GHH peptide fragment. The peak at 0.79 min corresponds to GHH peptide product and the peak at 2.93 min corresponding to the Fmoc-protected peptide, **(B)** Mass spectrum of the peak eluted at 0.79 min corresponding to the mass (M+1 mass peak) of GHH peptide, **(C)** Mass spectrum of the peak eluted at 2.93 min corresponding to the mass of GHH peptide attached to the Fmoc-group (571 m/z).

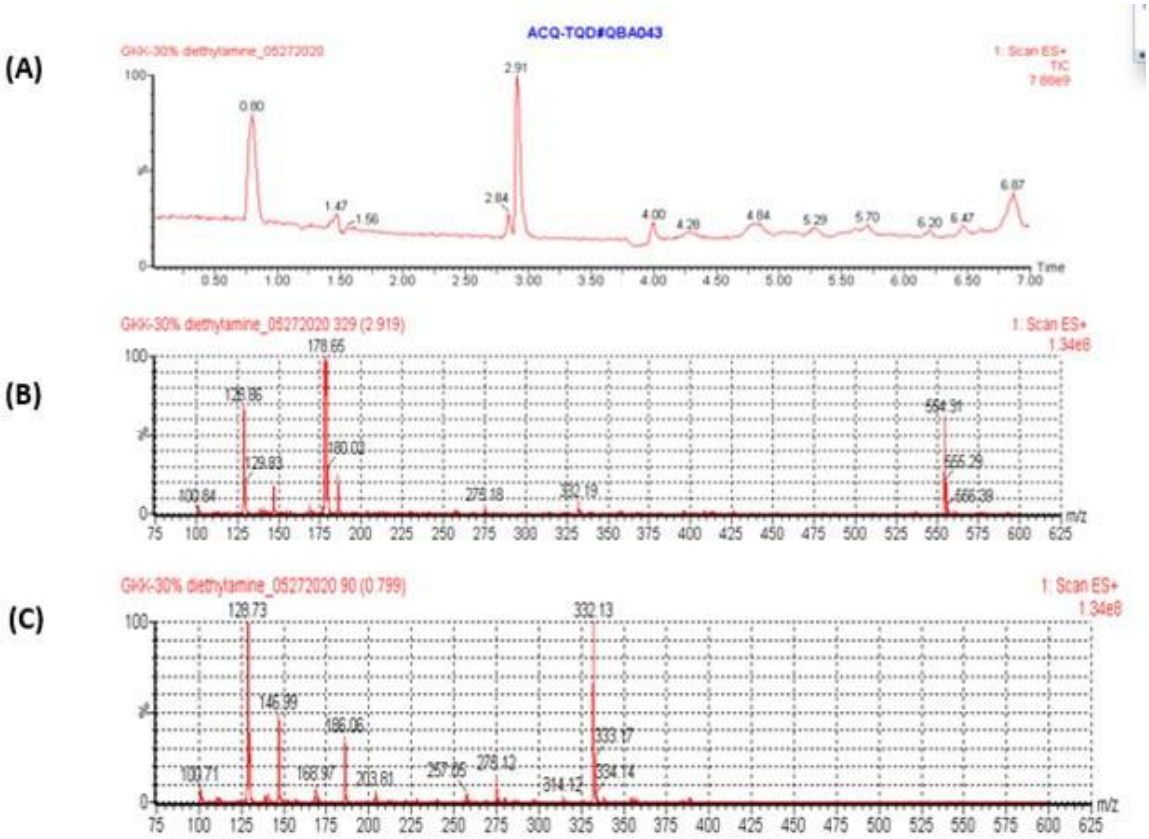


Figure S5. (A) UPLC chromatogram of GKK-4. The peak at 0.80 min corresponds to GKK peptide product with the impurity peaks at 2.91 min and 4.00 min corresponding to the Fmoc-protected GKK peptide impurity and KK peptide fragment peak, respectively **(B)** Mass spectrum of GKK-4 corresponding to the major Fmoc-protected peptide impurity peak at 2.92 min, **(C)** Mass spectrum of GKK-4 corresponding to the minor GKK peak at 0.79 min.

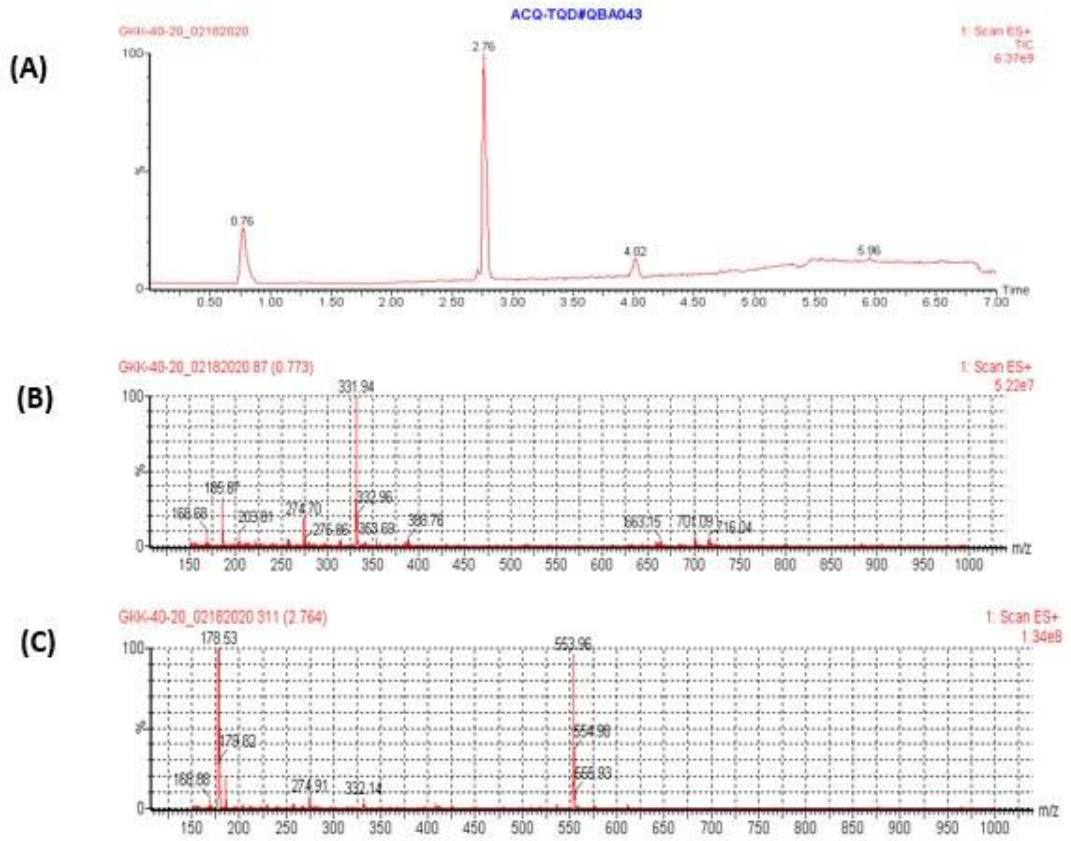


Figure S6. (A) UPLC chromatogram of GKK-5. The peak at 0.76 min corresponds to GKK peptide product with the impurity peaks at 2.76 min and 4.02 min corresponding to the Fmoc-protected GKK peptide impurity and KK peptide fragment peak, respectively **(B)** Mass spectrum of GKK-5 corresponding to the minor GKK peptide peak at 0.77 min **(C)** Mass spectrum of GKK-5 corresponding to the major Fmoc-protected peptide impurity peak at 2.76 min.

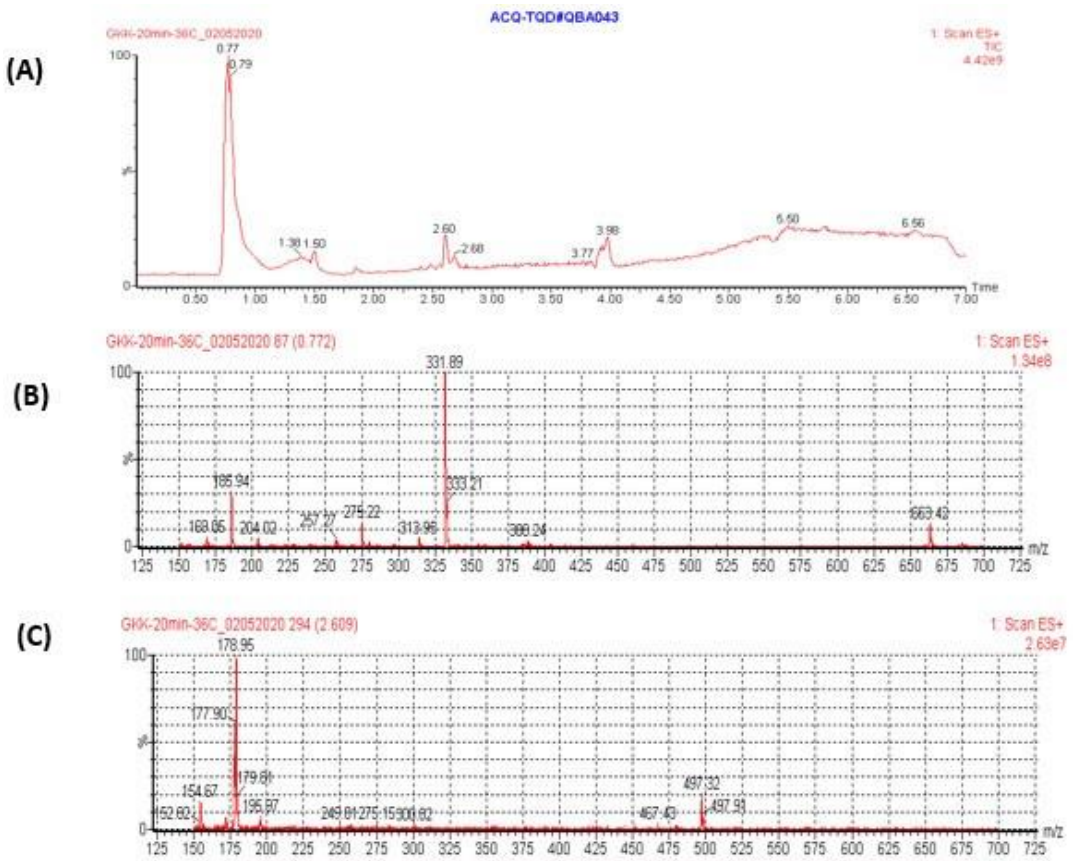


Figure S7. (A) UPLC chromatogram of GKK-8. The peak at 0.77 min corresponds to GKK peptide product with the impurity peaks at 2.60 min and 3.98 min corresponding to the Fmoc-protected KK peptide fragment impurity and KK peptide fragment peak, respectively **(B)** Mass spectrum of GKK-8 corresponding to the major GKK peptide peak at 0.77 min, **(C)** Mass spectrum of GKK-8 corresponding to the Fmoc-protected KK peptide fragment impurity peak at 2.61 min.

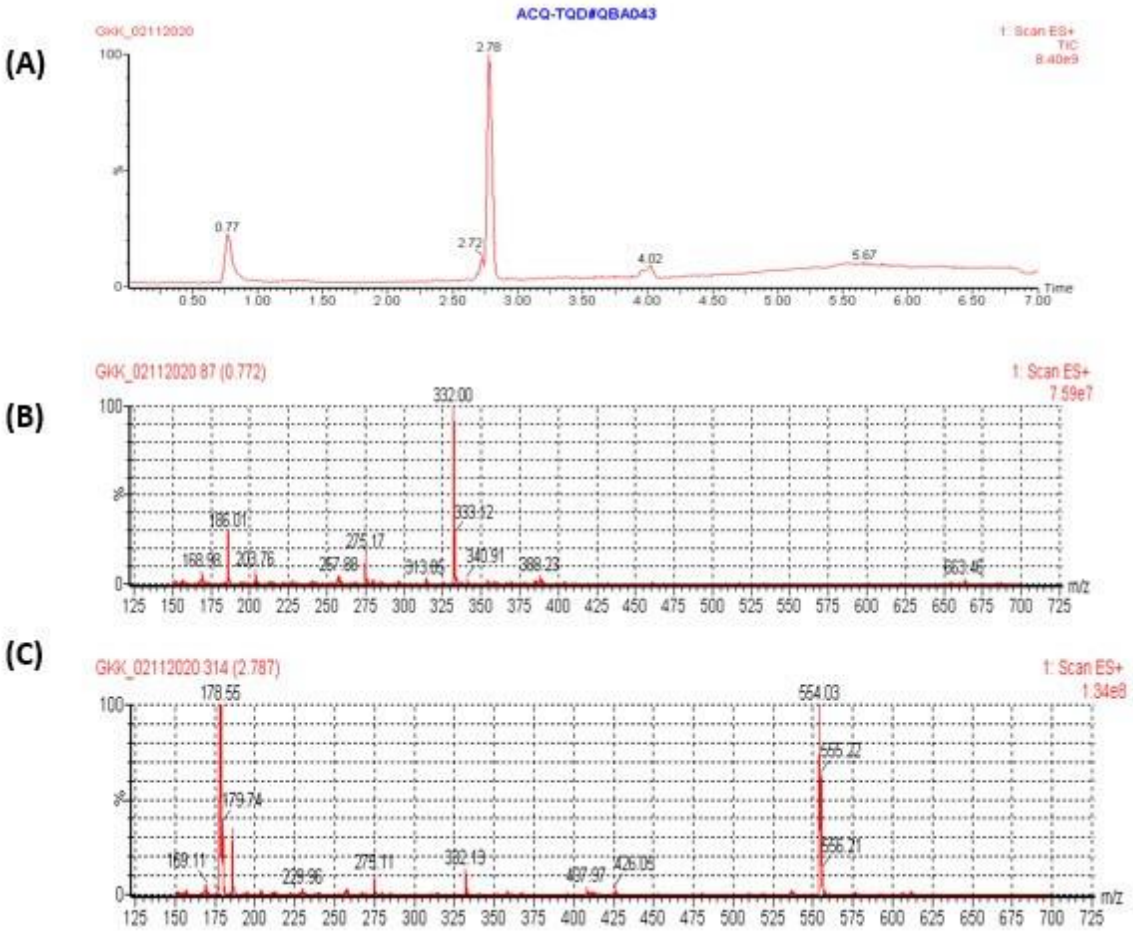


Figure S8. (A) UPLC chromatogram of GKK-9. The peak at 0.77 min corresponds to GKK peptide product with the impurity peaks at 2.78 min and 4.02 min corresponding to the Fmoc-protected GKK peptide impurity and KK peptide fragment peak, respectively **(B)** Mass spectrum of GKK-9 corresponding to the GKK peptide peak at 0.77 min **(C)** Mass spectrum of GKK-9 corresponding to the major Fmoc-protected peptide peak at 2.79 min.

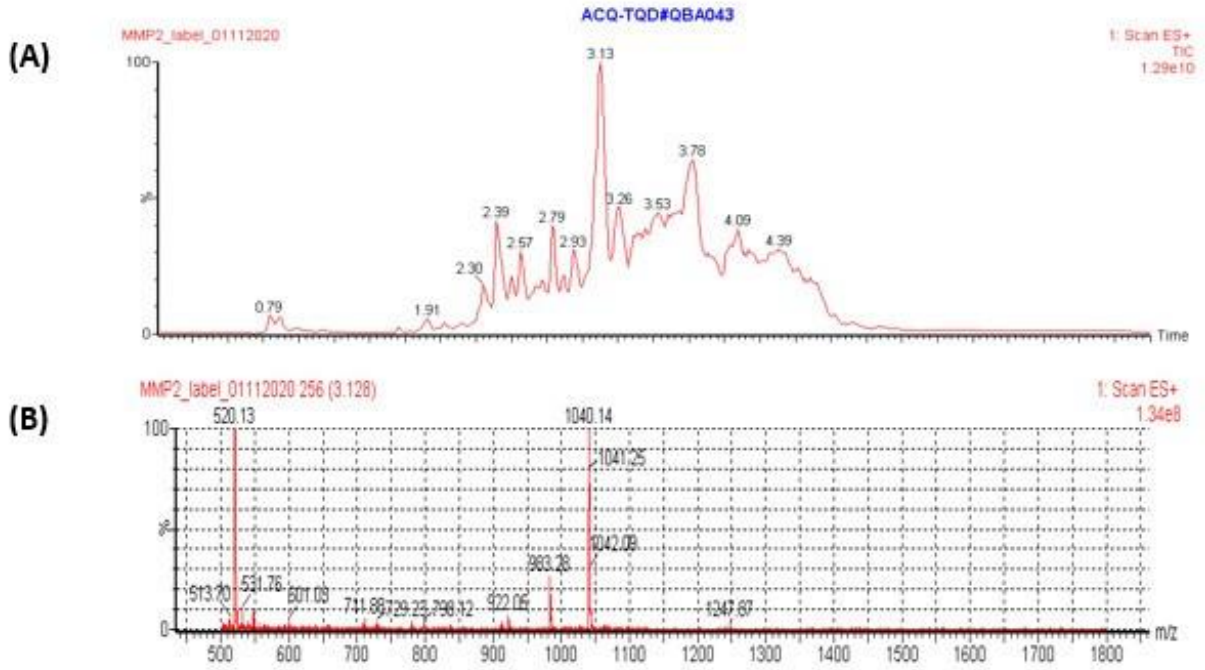


Figure S9: UPLC chromatogram of RhB-labeled MMP2 (Trial 1). The major peak at 3.13 min corresponds to the mass of impurity (peptide fragment with a missing cysteine along-with a water molecule) and multiple peaks of unknown origin, **(B)** Mass spectrum of RhB-labeled MMP2 for the peak eluted at 3.13 min corresponding to the impurity (peptide fragment with a missing cysteine and a water molecule).

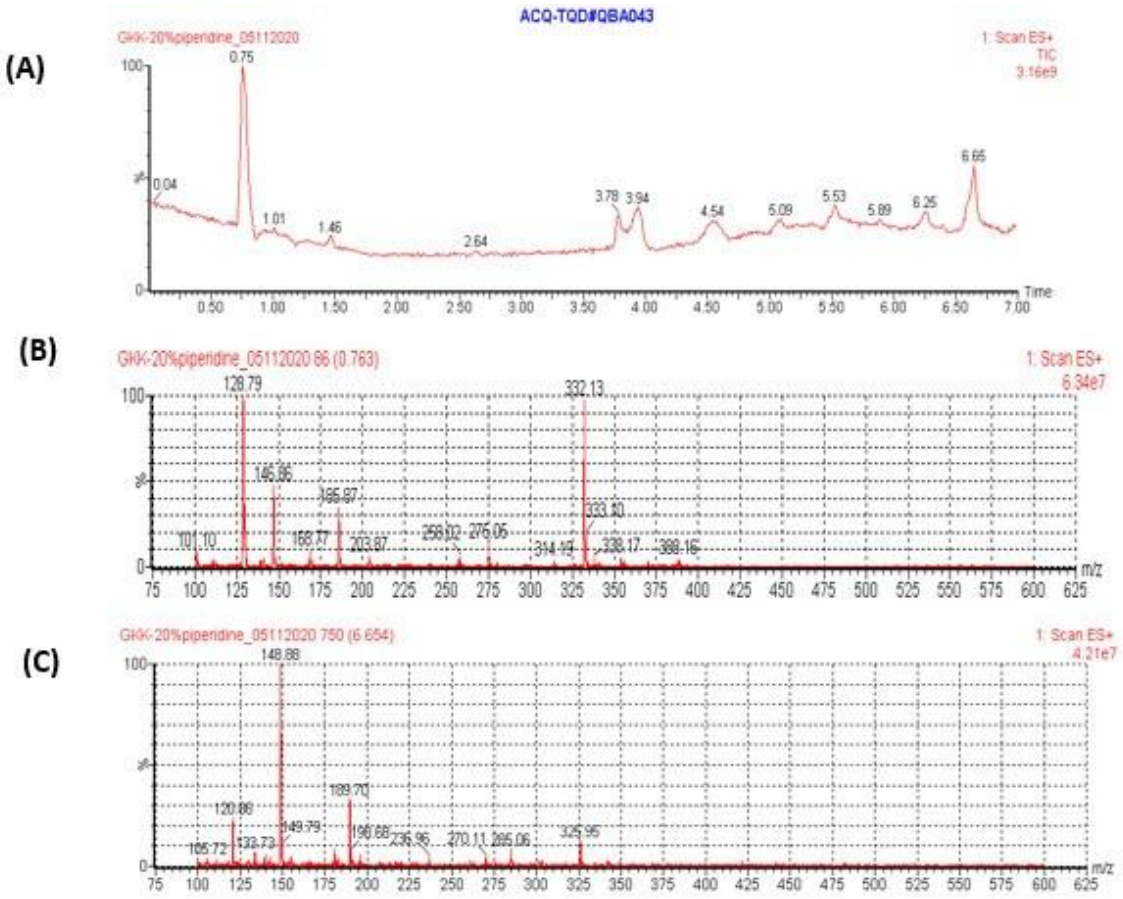


Figure S10. (A) UPLC chromatogram of GKK-13. The peak at 0.75 min corresponds to GKK peptide product with the impurity peaks at 3.94 min and 6.65 min corresponding to the KK peptide impurity and a free K amino acid, respectively **(B)** Mass spectrum of GKK-13 corresponding to the GKK peptide eluted at 0.76 min, **(C)** Mass spectrum of GKK-13 corresponding to the minor peak at 6.65 min (M+1 mass peak for free lysine).

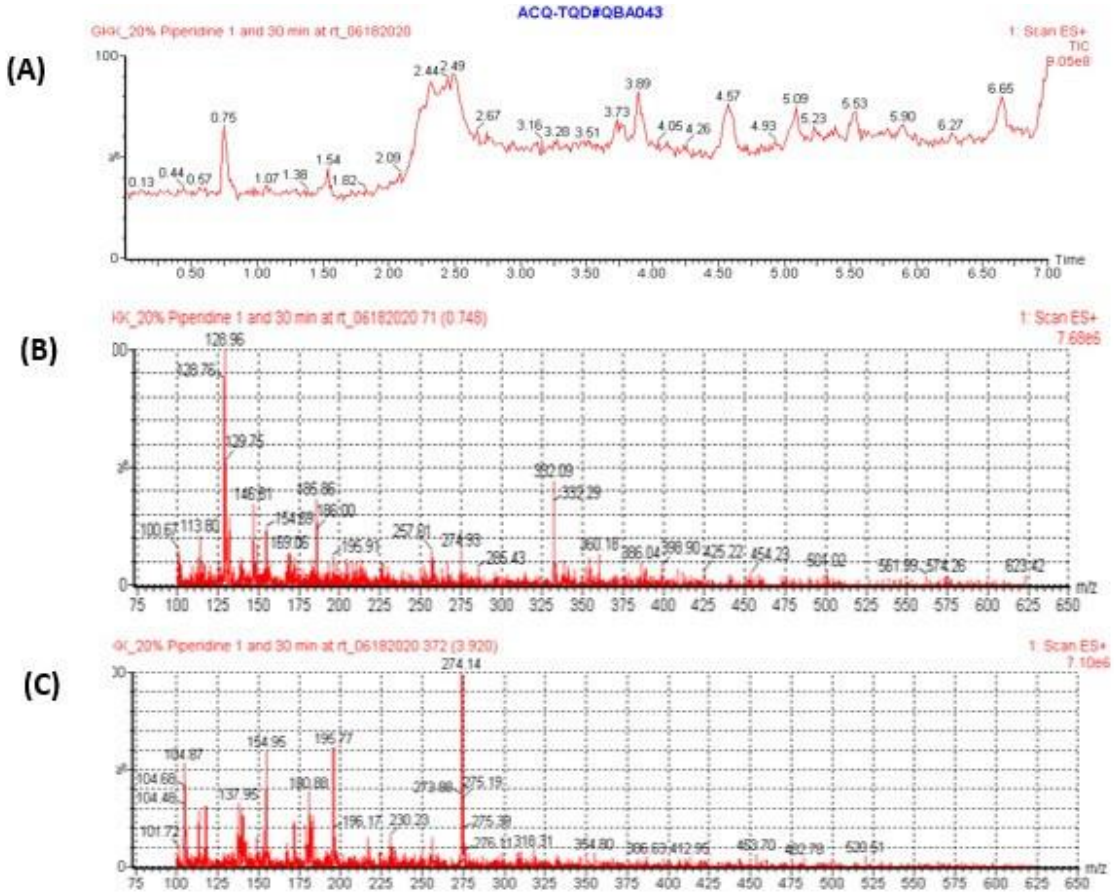


Figure S11. (A) UPLC chromatogram of GKK-14. The peak at 0.75 min corresponds to GKK peptide product with the impurity peaks at 2.44-2.49 min and 3.89 min corresponding to the Fmoc-protected GKK peptide impurity and KK peptide fragment, respectively **(B)** Mass spectrum of GKK-14 corresponding to the GKK peptide eluted at 0.75 min, **(C)** Mass spectrum of GKK-14 corresponding to the KK peptide impurity eluted at 3.92 min.

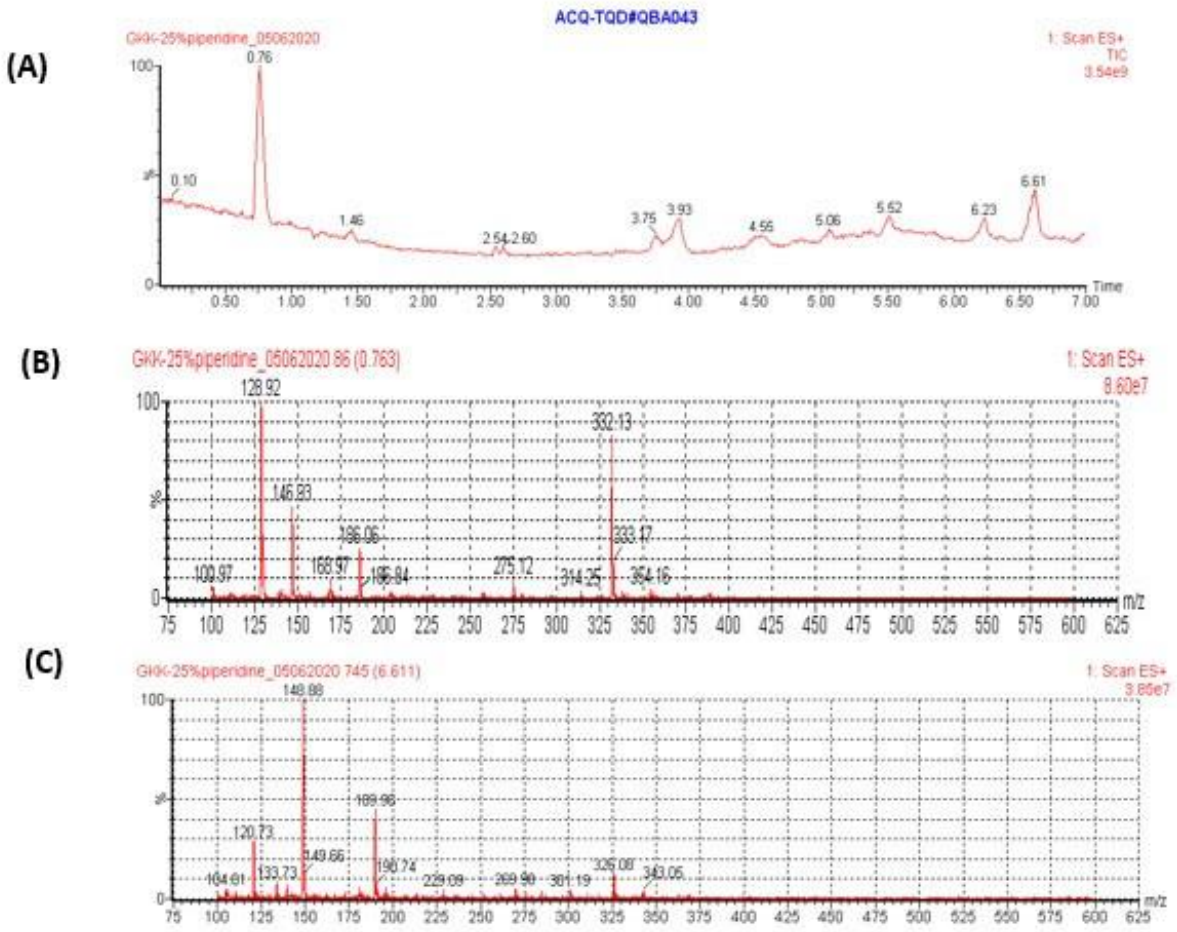


Figure S12. (A) UPLC chromatogram of GKK-15. The peak at 0.76 min corresponds to the GKK peptide product with the impurity peaks at 3.93 min and 6.61 min corresponding to the KK peptide impurity and a free K amino acid, respectively **(B)** Mass spectrum of GKK-15 corresponding to the GKK peptide eluted at 0.76 min, **(C)** Mass spectrum of GKK-15 corresponding to the minor peak at 6.61 min (M+1 mass peak for free lysine).

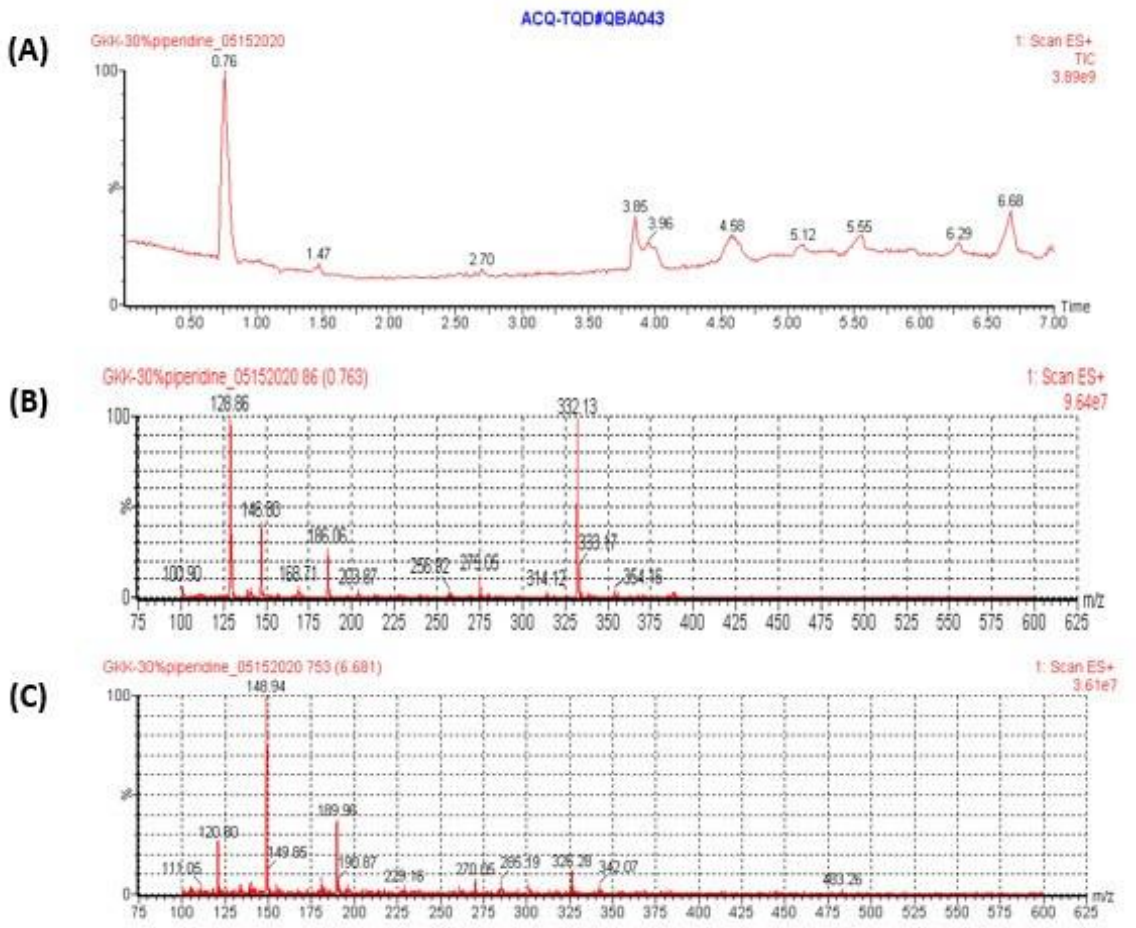


Figure S13. (A) UPLC chromatogram of GKK-17. The peak at 0.76 min corresponds to GKK peptide product with the impurity peaks at 3.85-3.96 min and 6.68 min corresponding to the KK peptide impurity and a free K amino acid, respectively **(B)** Mass spectrum of GKK-17 corresponding to the GKK peptide eluted at 0.76 min, **(C)** Mass spectrum of GKK-17 corresponding to the minor peak at 6.68 min (M+1 mass peak for free lysine).

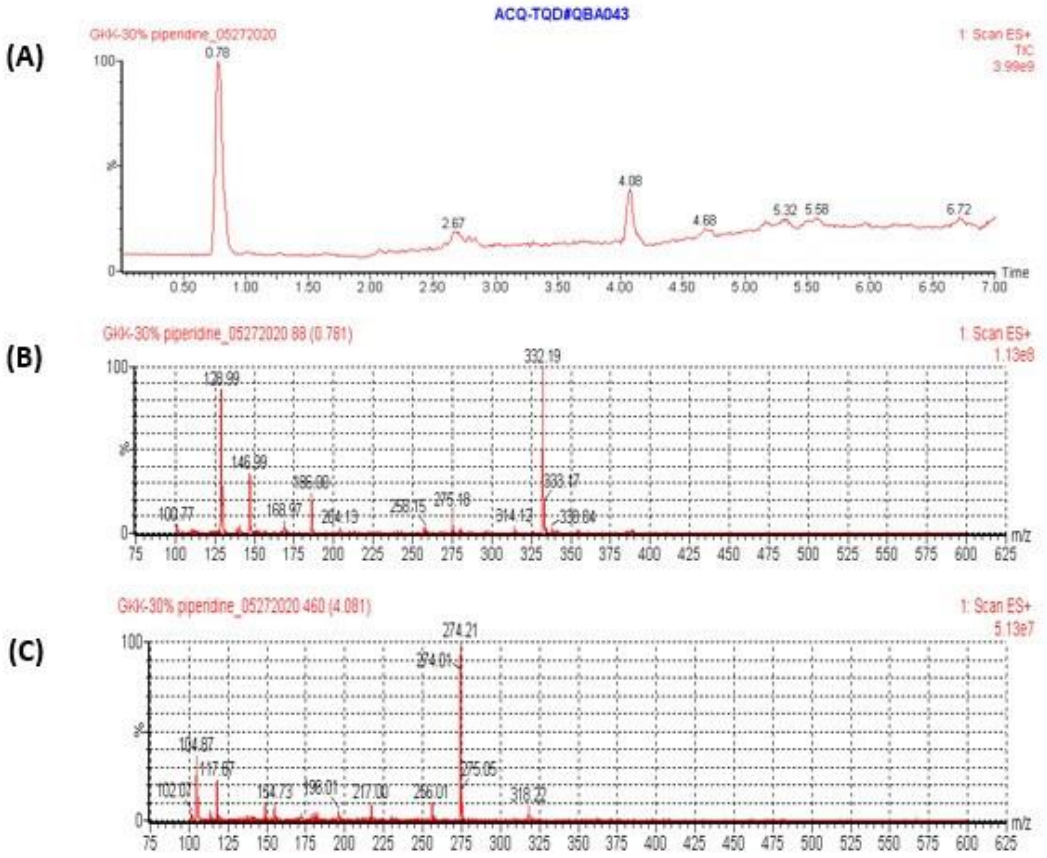


Figure S14. (A) UPLC chromatogram of GKK-18. The peak at 0.78 min corresponds to GKK peptide product and the peak at 4.08 min corresponds to the KK peptide fragment impurity, **(B)** Mass spectrum of GKK-18 corresponding to the GKK peptide eluted at 0.78 min, **(C)** Mass spectrum of GKK-18 corresponding to the KK peptide fragment eluted at 4.08 min.

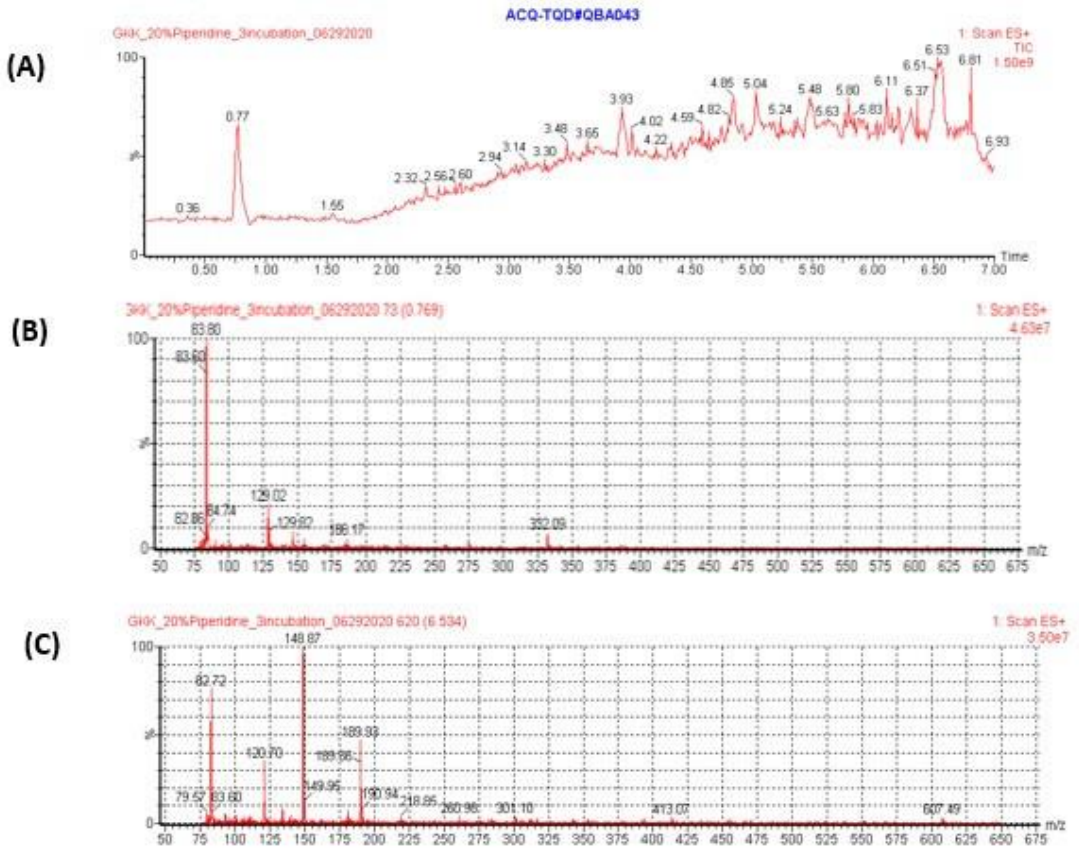


Figure S15. (A) UPLC chromatogram of GKK-22. The peak at 0.77 min corresponds to the GKK peptide product with multiple impurity peaks. The peaks eluted at 3.93 min and 6.53 min correspond to the KK peptide and a free K amino acid, respectively **(B)** Mass spectrum of GKK-22 corresponding to the GKK peptide eluted at 0.77 min, **(C)** Mass spectrum of GKK-22 corresponding to the impurity peak at 6.53 min (M+1 mass peak for free lysine).

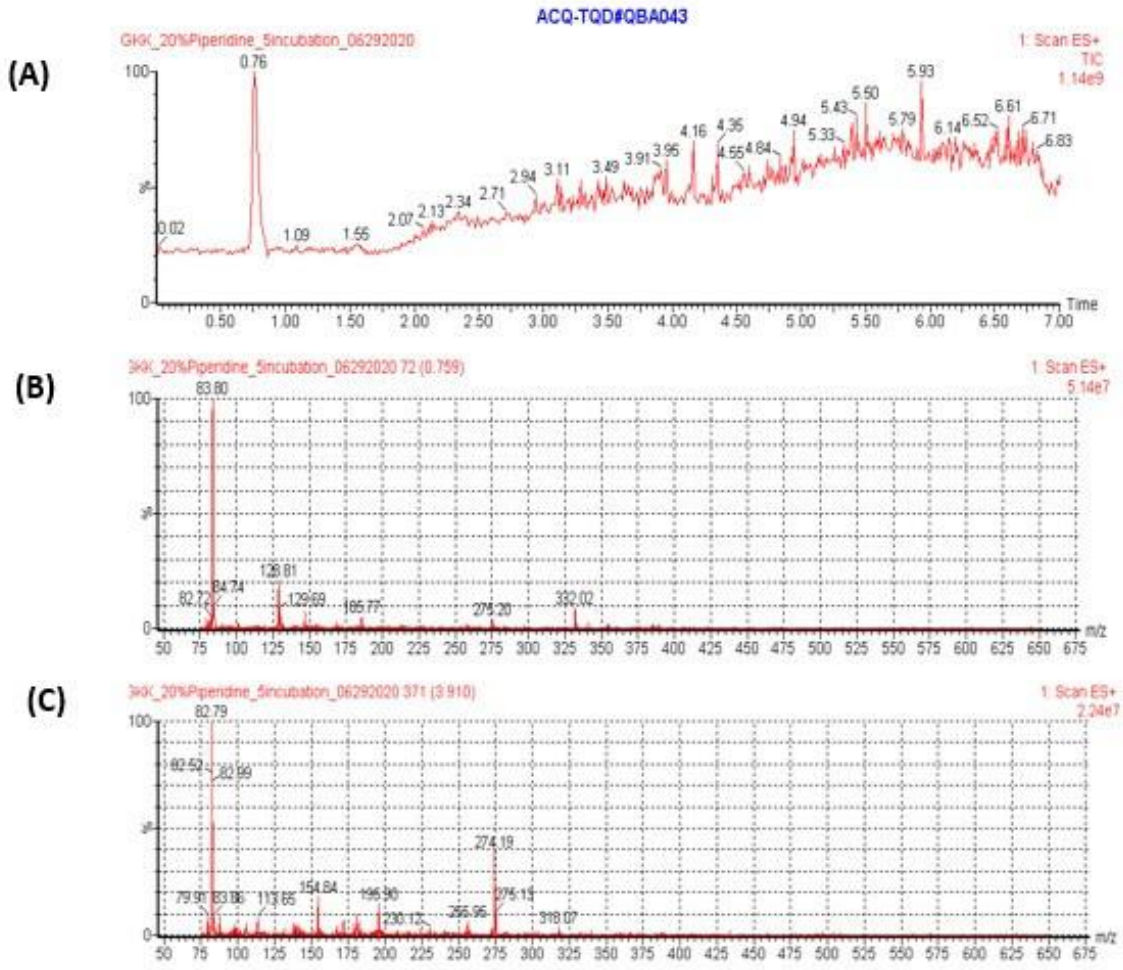


Figure S16. (A) UPLC chromatogram of GKK-24. The peak at 0.76 min corresponds to the GKK peptide product with multiple impurity peaks. The peaks eluted at 3.91 min and 6.71 correspond to the KK peptide and a free K amino acid, respectively **(B)** Mass spectrum of GKK-24 corresponding to the GKK peptide eluted at 0.76 min, **(C)** Mass spectrum of GKK-24 corresponding to the KK fragment peak eluted at 3.91 min.

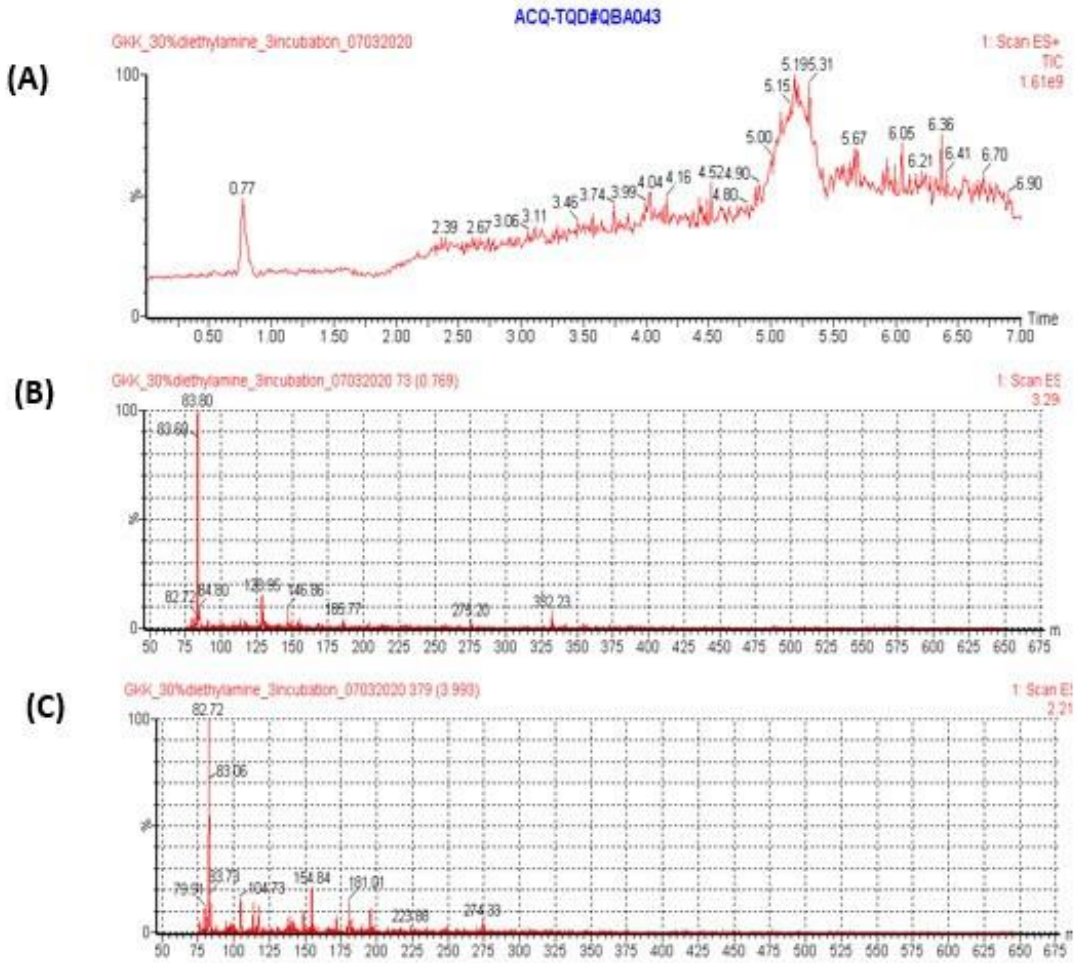


Figure S17. (A) UPLC chromatogram of GKK-25. The peak at 0.77 min corresponds to the GKK peptide product with multiple impurities (peaks with unknown origin in the region of 4.90 to 5.50 min). The peak eluted at 3.93 min corresponds to the KK peptide, **(B)** Mass spectrum of GKK-25 corresponding to the GKK peptide eluted at 0.77 min, **(C)** Mass spectrum of GKK-25 corresponding to the KK fragment peak eluted at 3.99 min.

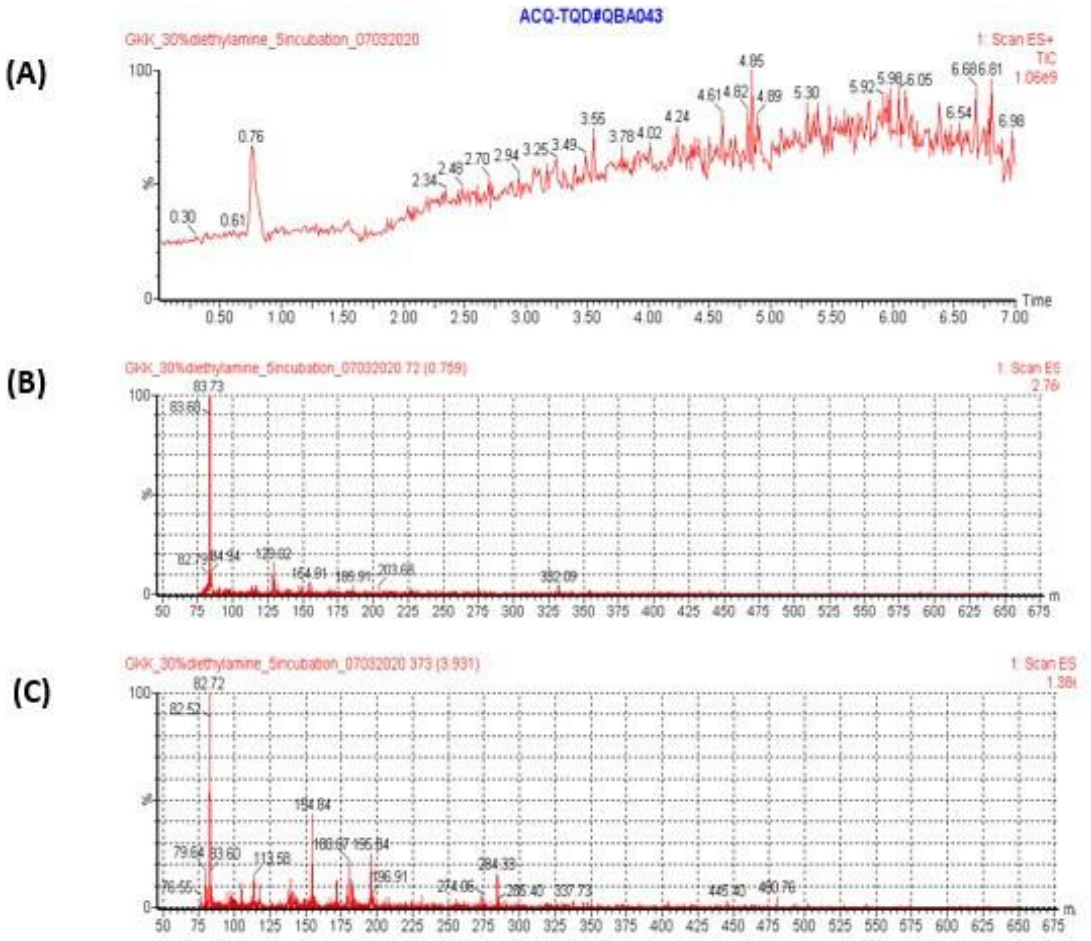


Figure S18. (A) UPLC chromatogram of GKK-27. The peak at 0.76 min corresponds to the GKK peptide product with multiple impurities of unknown origin. The peak eluted at 3.93 min corresponds to the KK peptide, **(B)** Mass spectrum of GKK-27 corresponding to the GKK peptide eluted at 0.76 min, **(C)** Mass spectrum of GKK-27 corresponding to the KK fragment peak eluted at 3.93 min.

Appendix_ Chapter 3 (A)

UPLC-MS Characterization of consensus peptide sequences

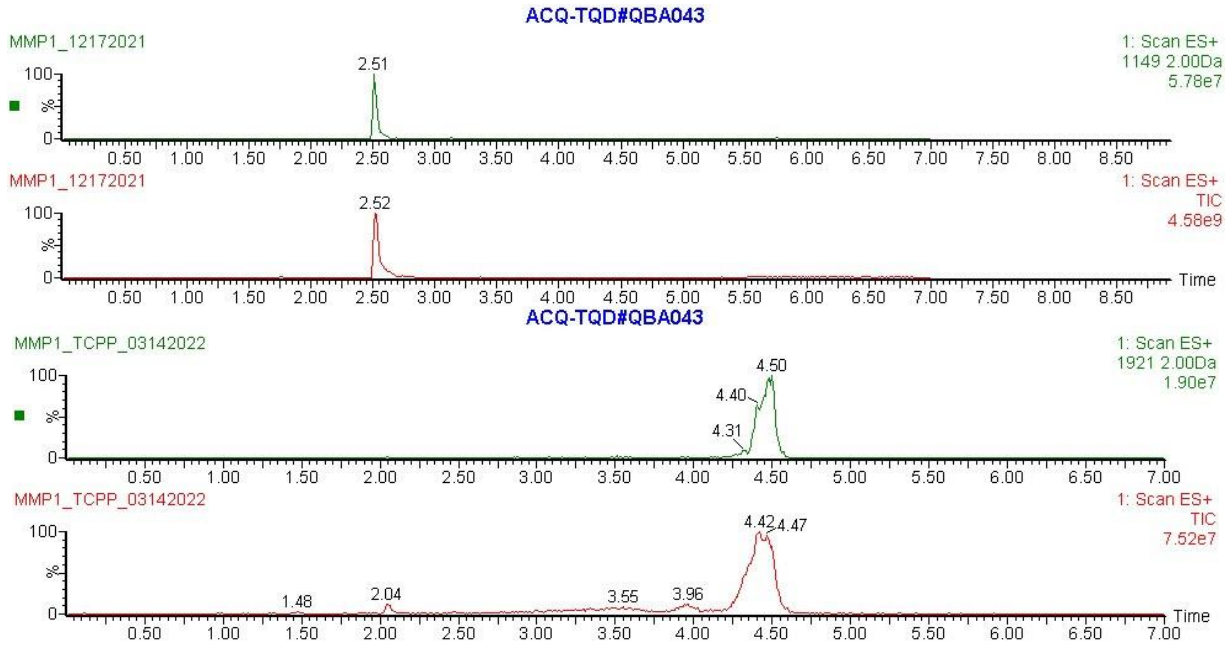


Figure A1. MMP1: Unlabeled and TCPP-labeled (both filtered to the relevant mass, in green).

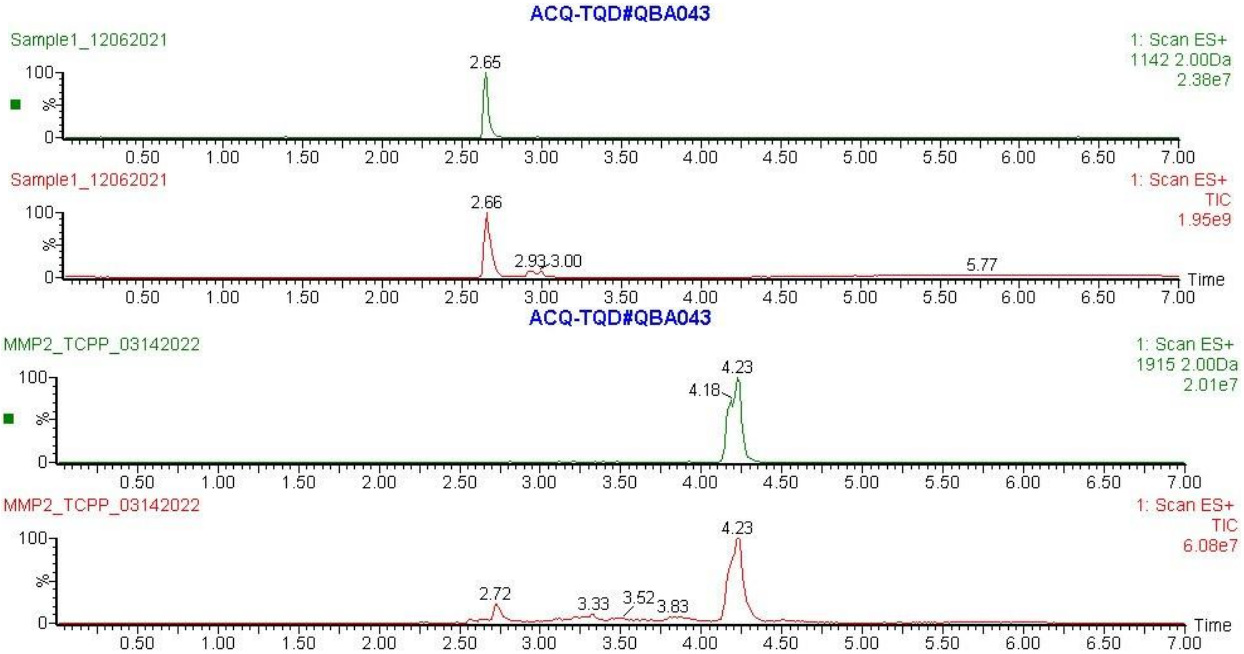


Figure A2. MMP2: Unlabeled and TCPP-labeled (both filtered to the relevant mass, in green).

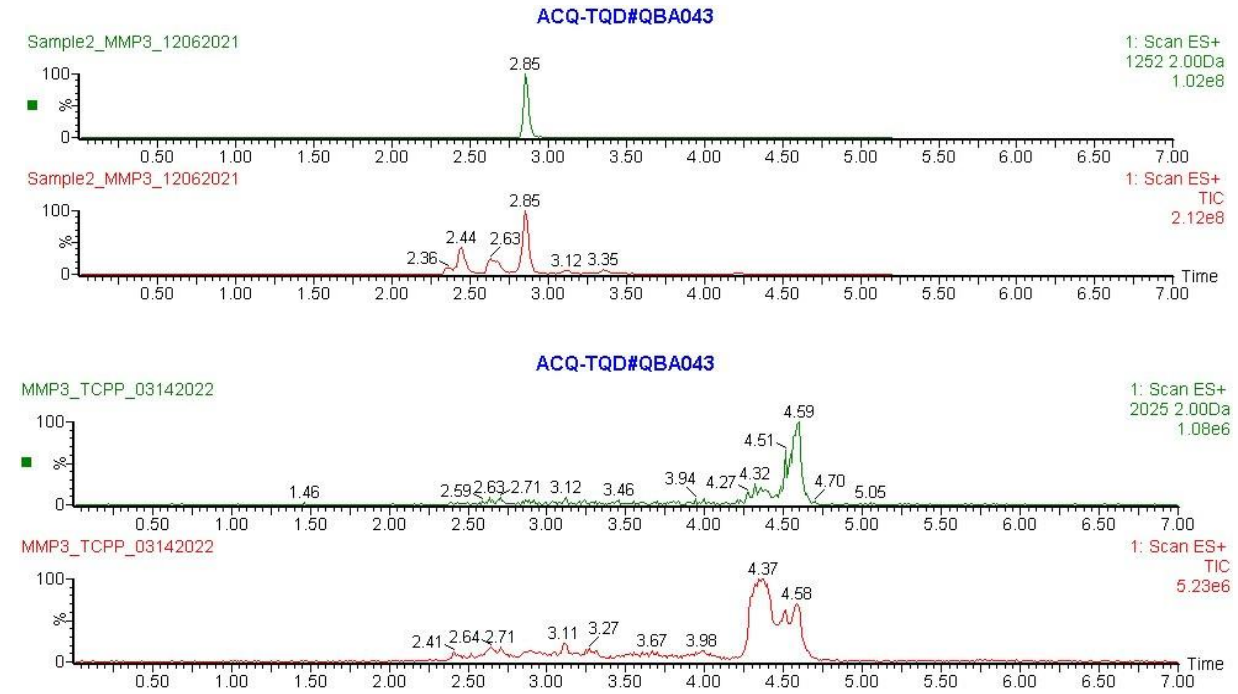


Figure A3. MMP3: Unlabeled and TCPP-labeled (both filtered to the relevant mass, in green).

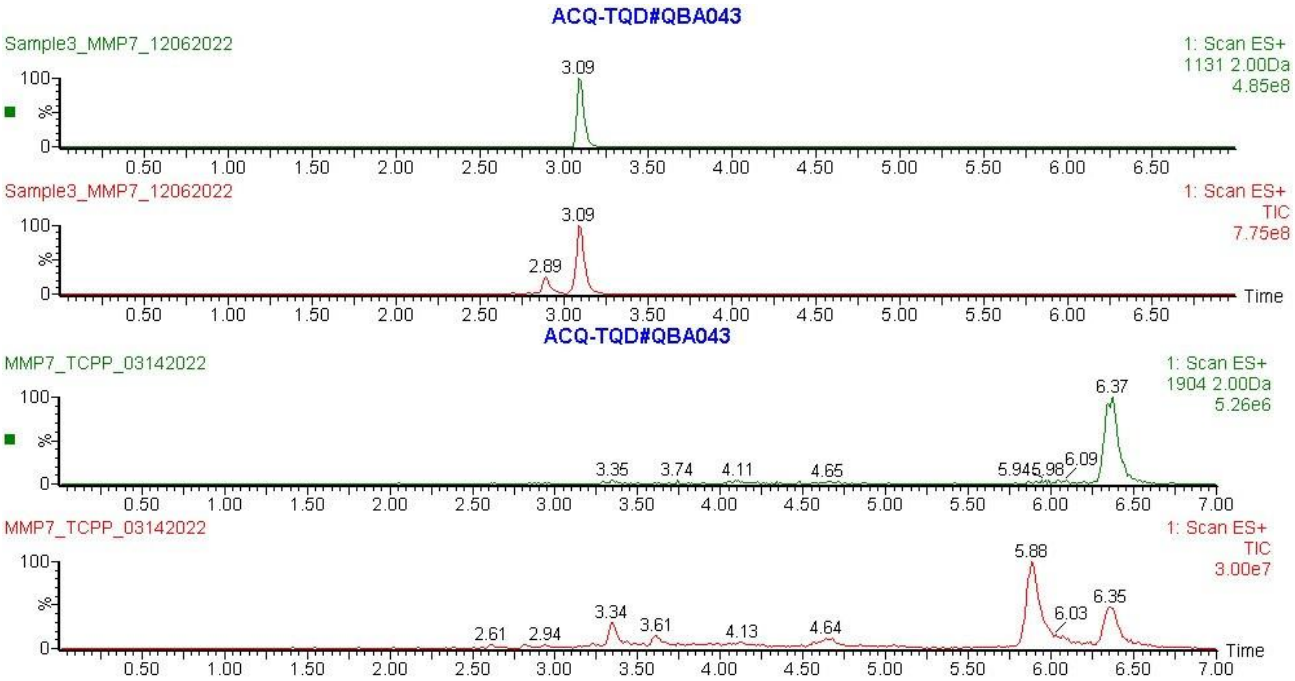


Figure A4. MMP7: Unlabeled and TCPP-labeled (both filtered to the relevant mass, in green).

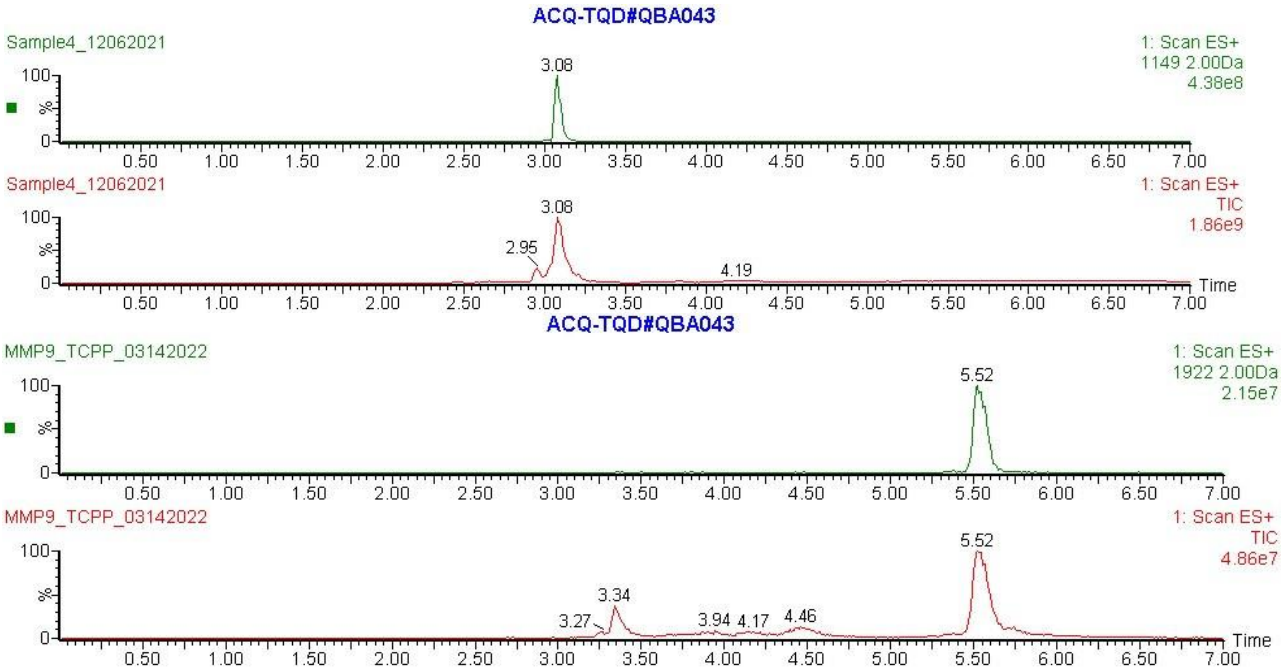


Figure A5. MMP9: Unlabeled and TCPP-labeled (both filtered to the relevant mass, in green).

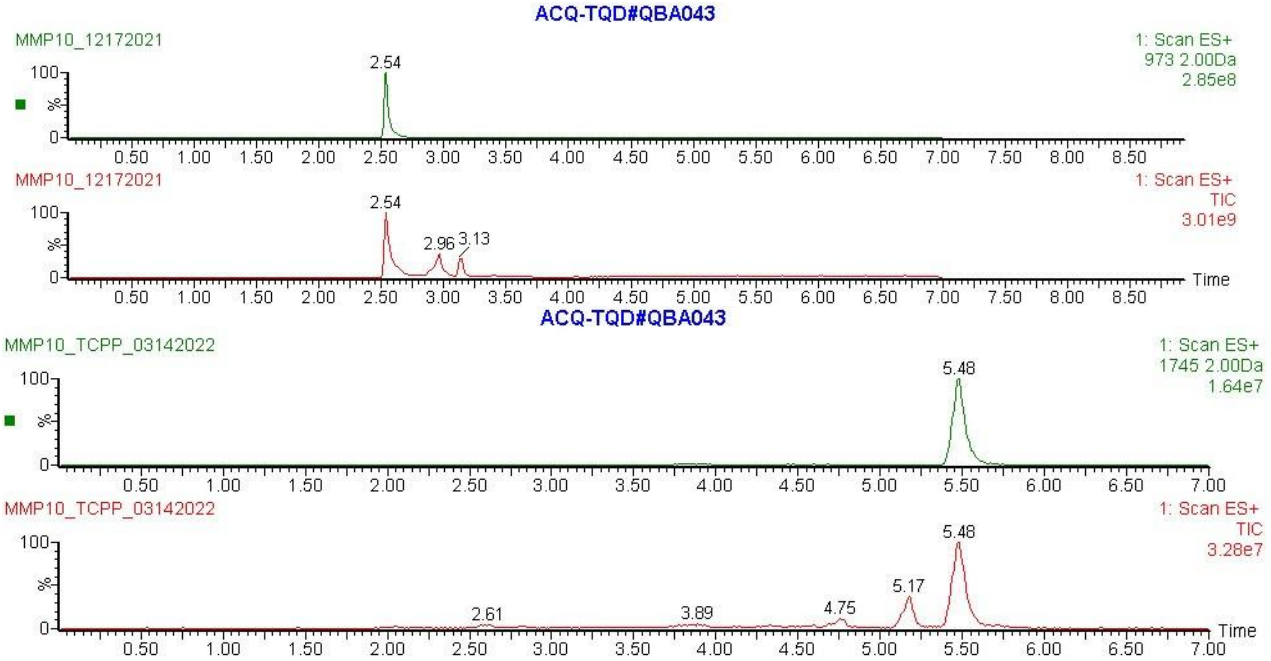


Figure A6. MMP10: Unlabeled and TCPP-labeled (both filtered to the relevant mass, in green).

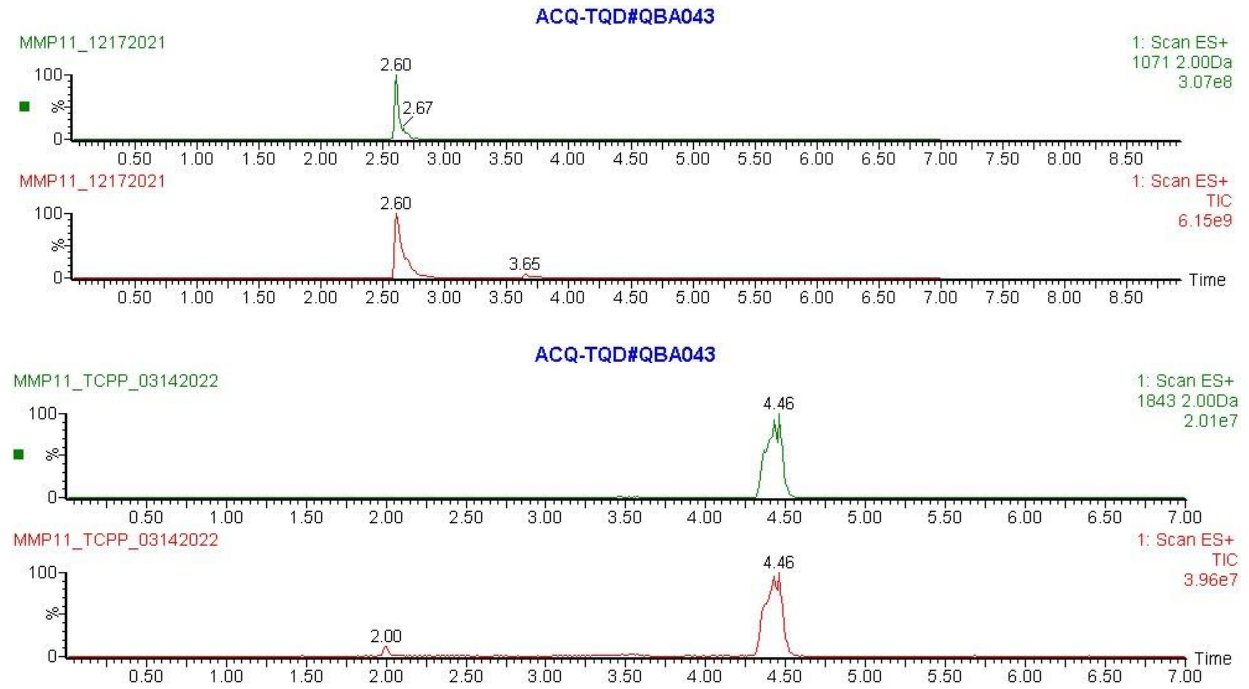


Figure A7. MMP11: Unlabeled and TCPP-labeled (both filtered to the relevant mass, in green).

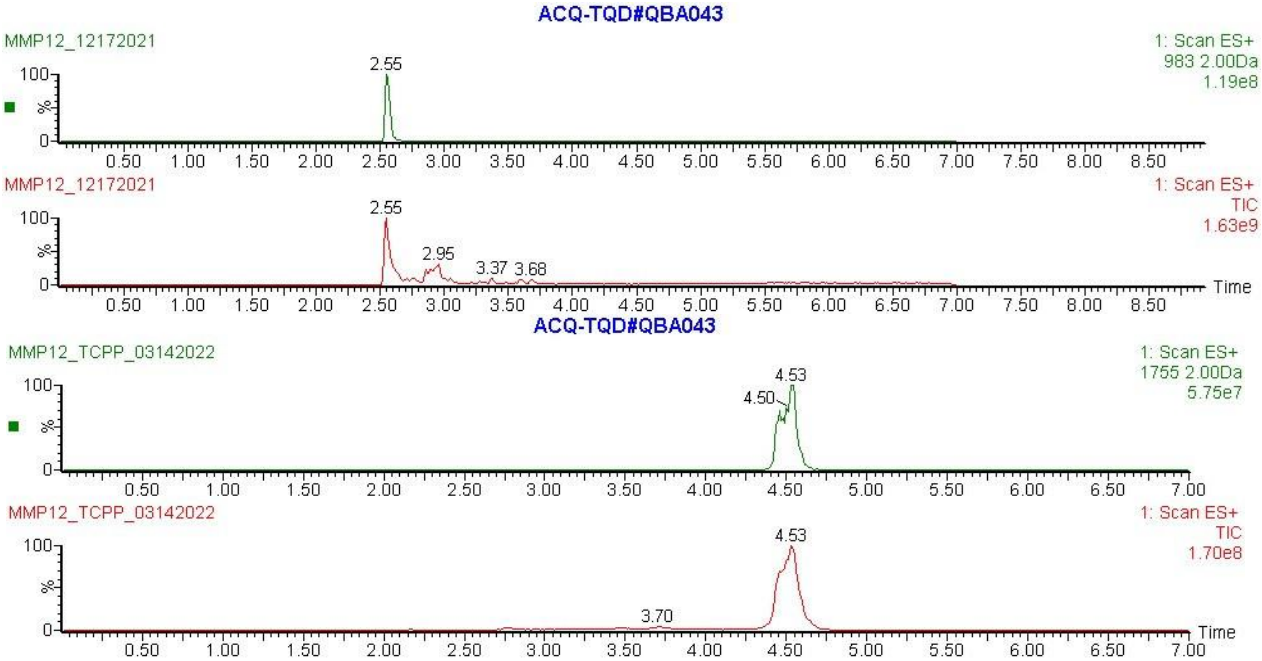


Figure A8. MMP12: Unlabeled and TCPP-labeled (both filtered to the relevant mass, in green).

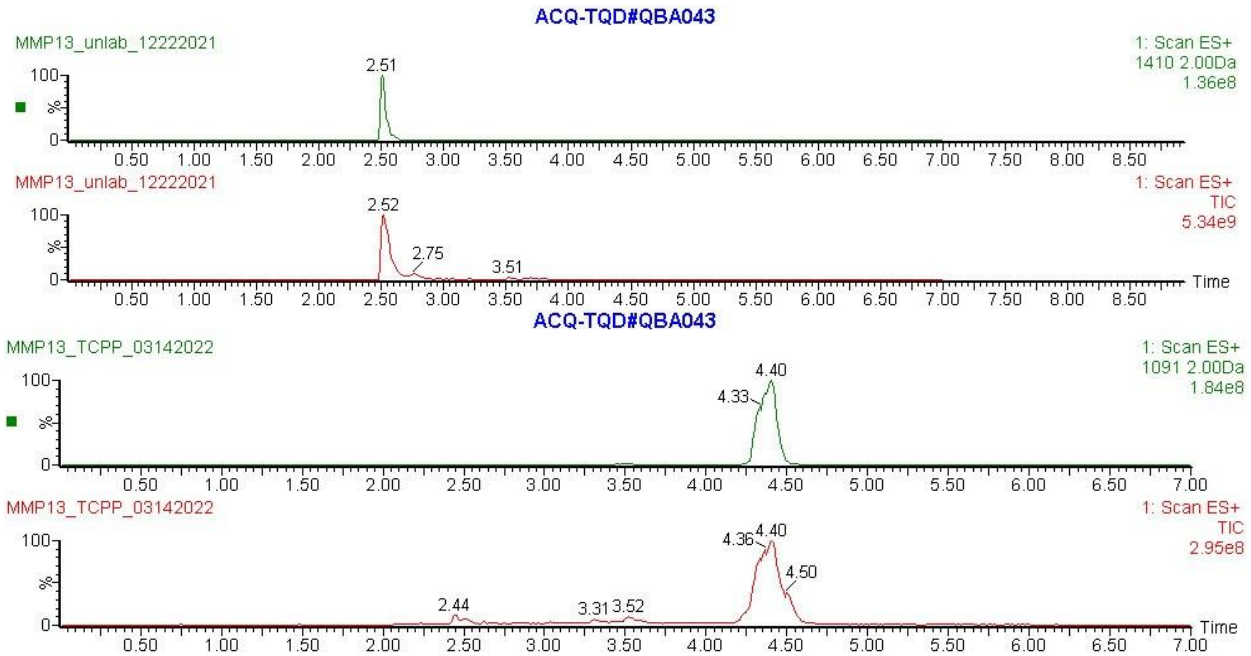


Figure A9. MMP13: Unlabeled and TCPP-labeled (both filtered to the relevant mass, in green).

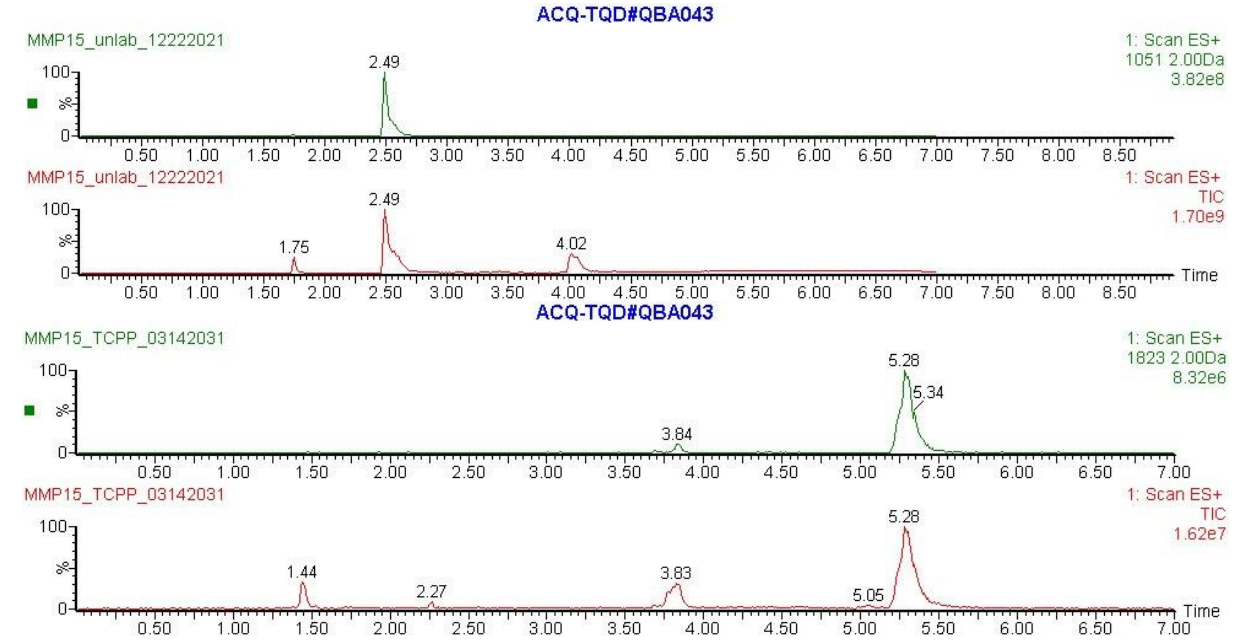
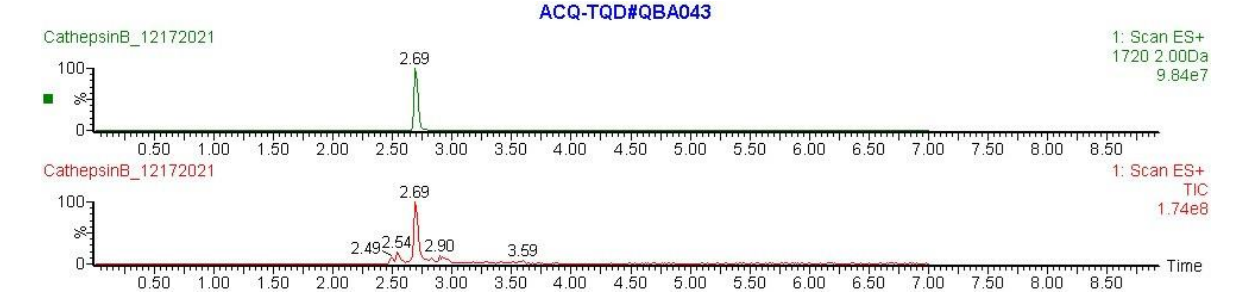


Figure A10. MMP15: Unlabeled and TCPP-labeled (both filtered to the relevant mass, in green).



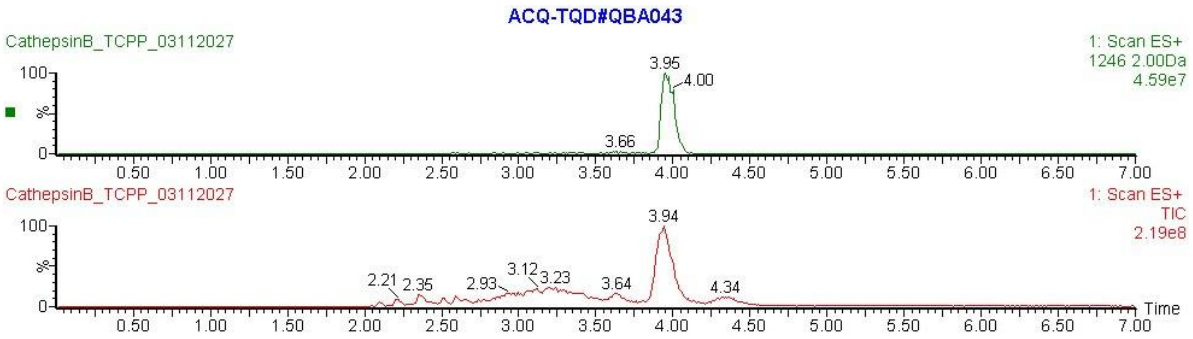


Figure A11. CTSB: Unlabeled and TCPP-labeled (both filtered to the relevant mass, in green).

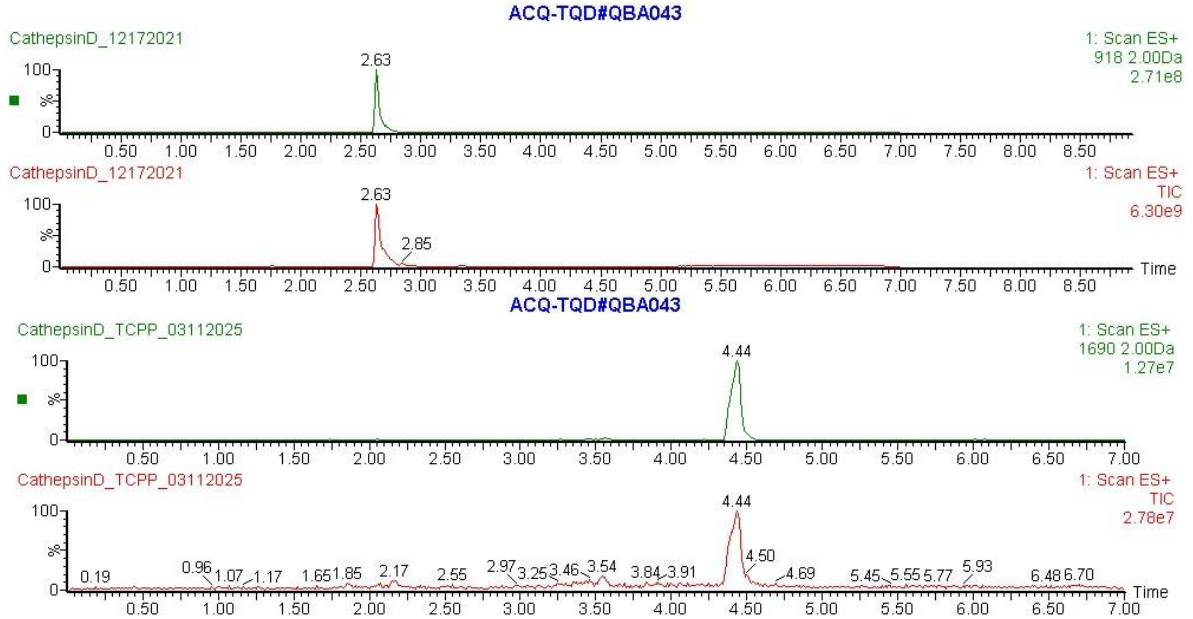


Figure A12. CTSD: Unlabeled and TCPP-labeled (both filtered to the relevant mass, in green).

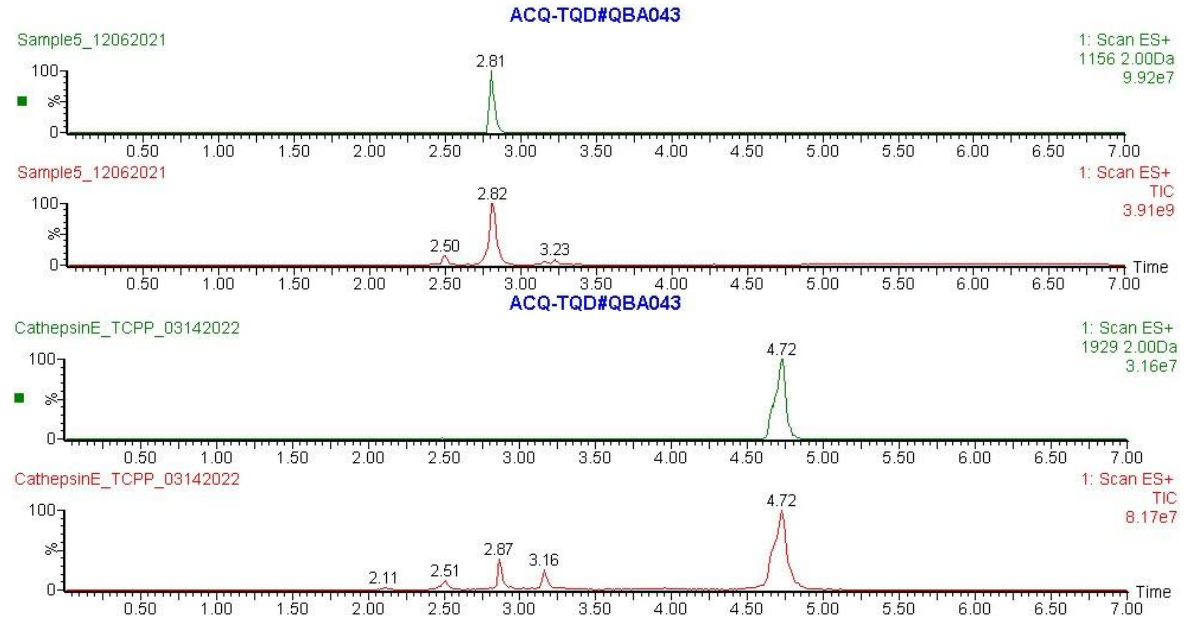


Figure A13. CTSE: Unlabeled and TCPP-labeled (both filtered to the relevant mass, in green).

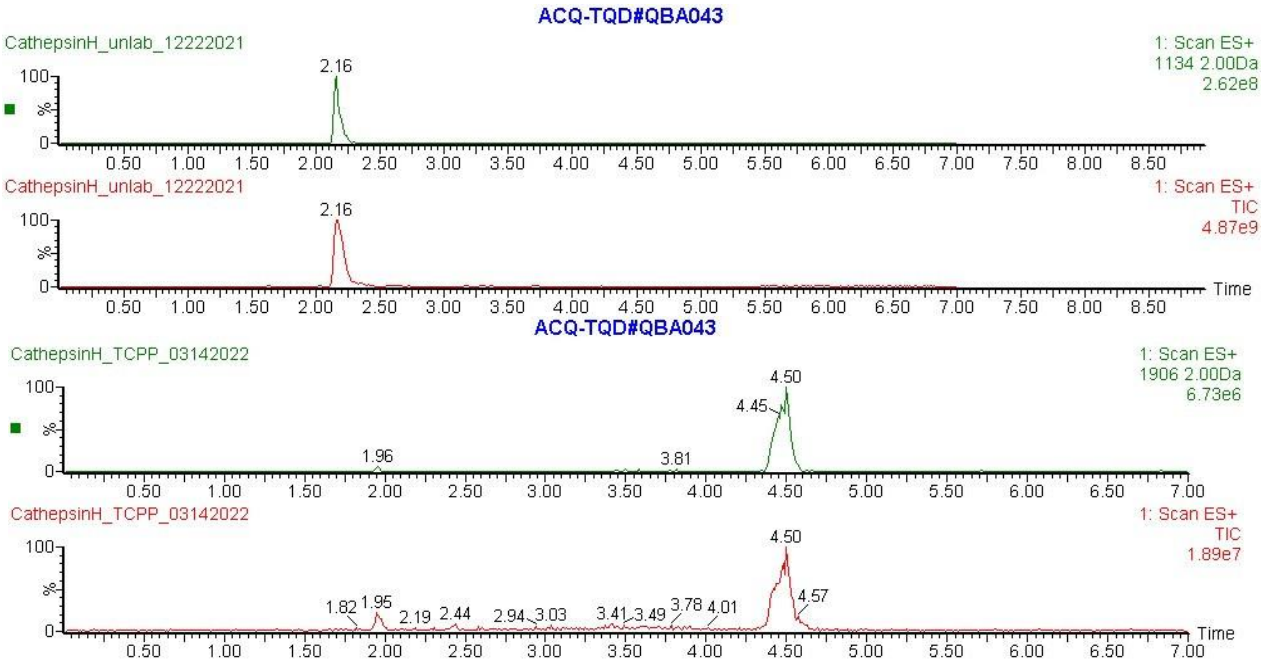


Figure A14. CTSH: Unlabeled and TCPP-labeled (both filtered to the relevant mass, in green).

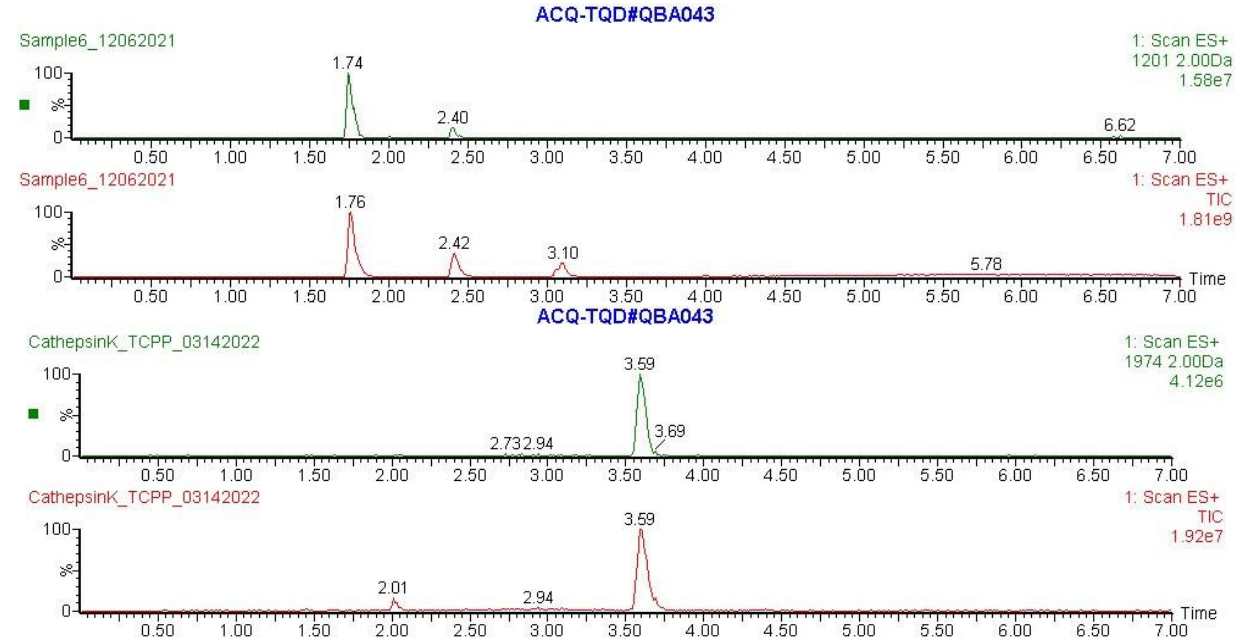


Figure A15. CTSK: Unlabeled and TCPP-labeled (both filtered to the relevant mass, in green).

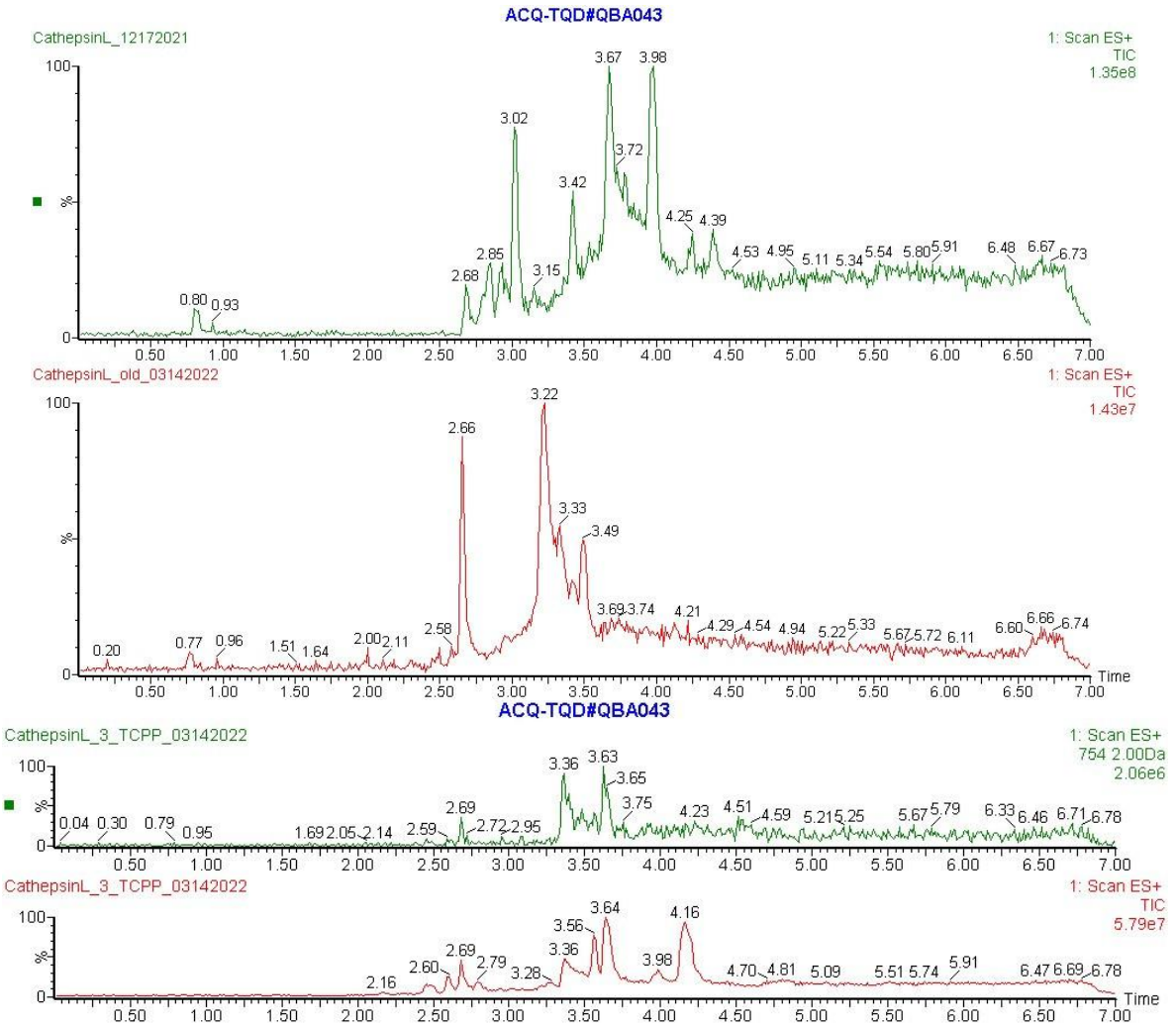


Figure A16. CTSL: Unlabeled and TCPP-labeled (both filtered to the relevant mass, in green).

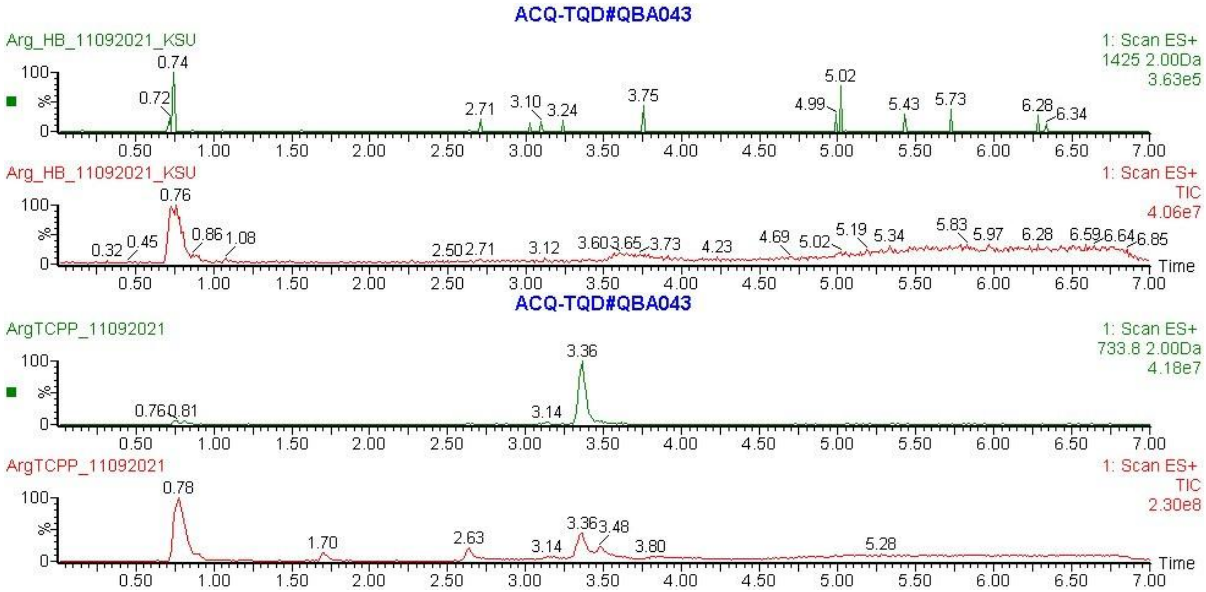


Figure A17. ARG: Unlabeled and TCPP-labeled (both filtered to the relevant mass, in green).

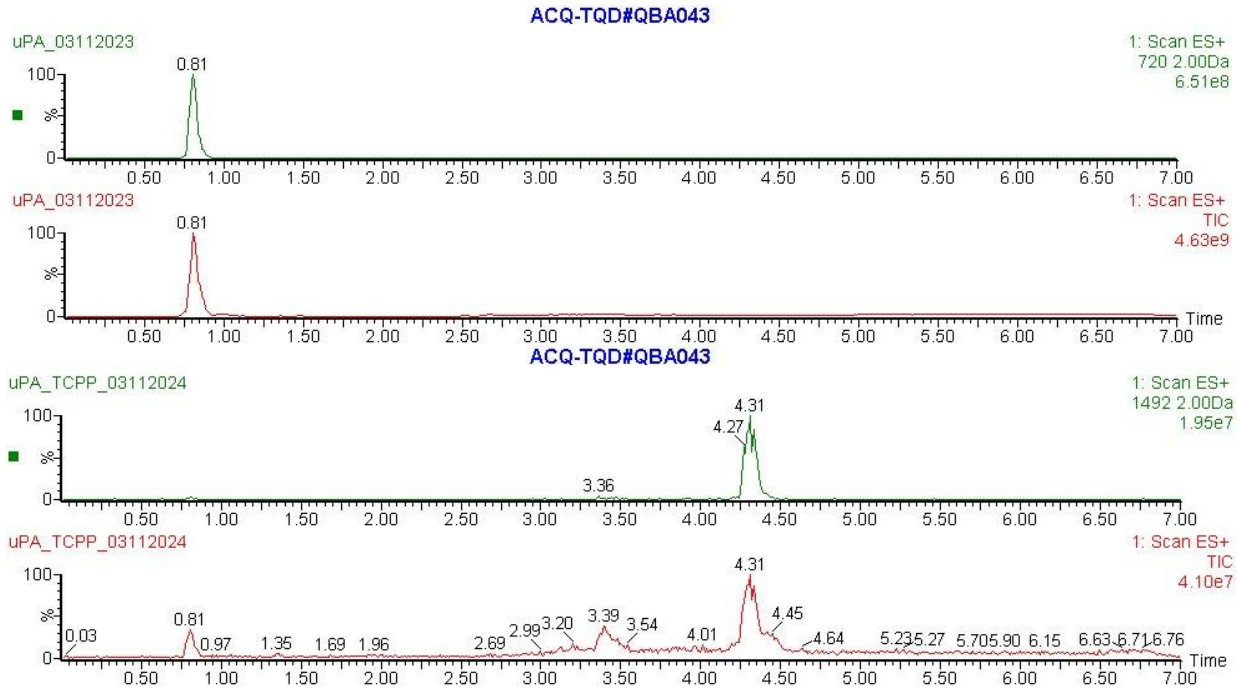
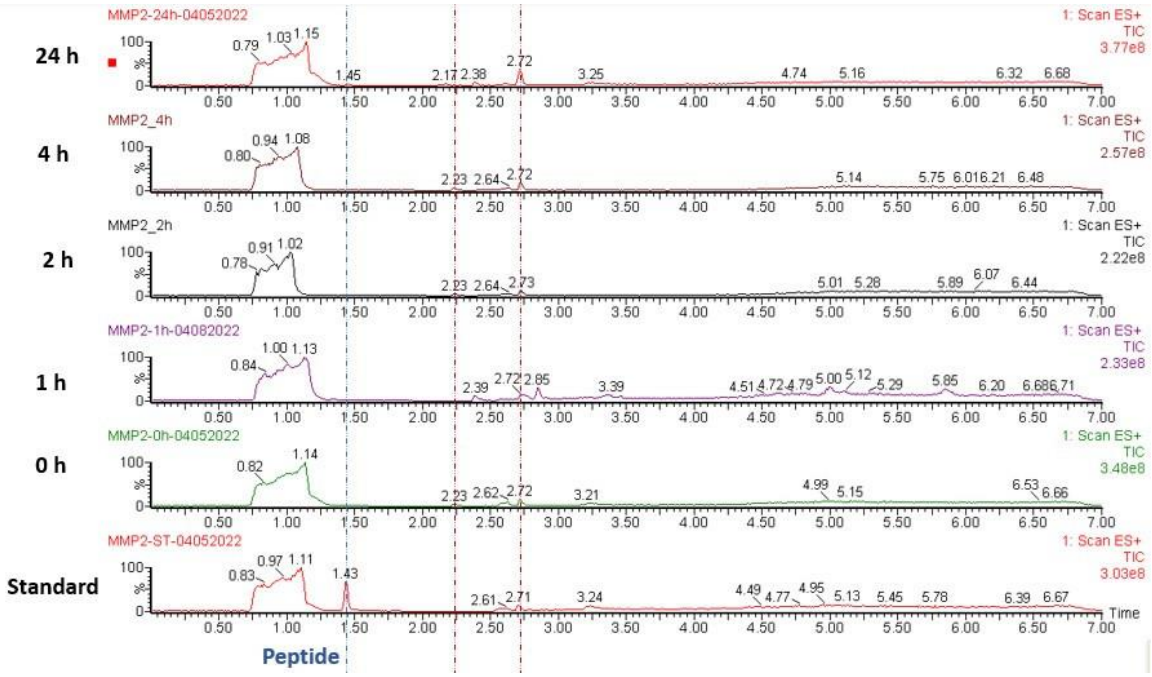


Figure A18. uPA: Unlabeled and TCPP-labeled (both filtered to the relevant mass, in green).

Enzymatic digestion results



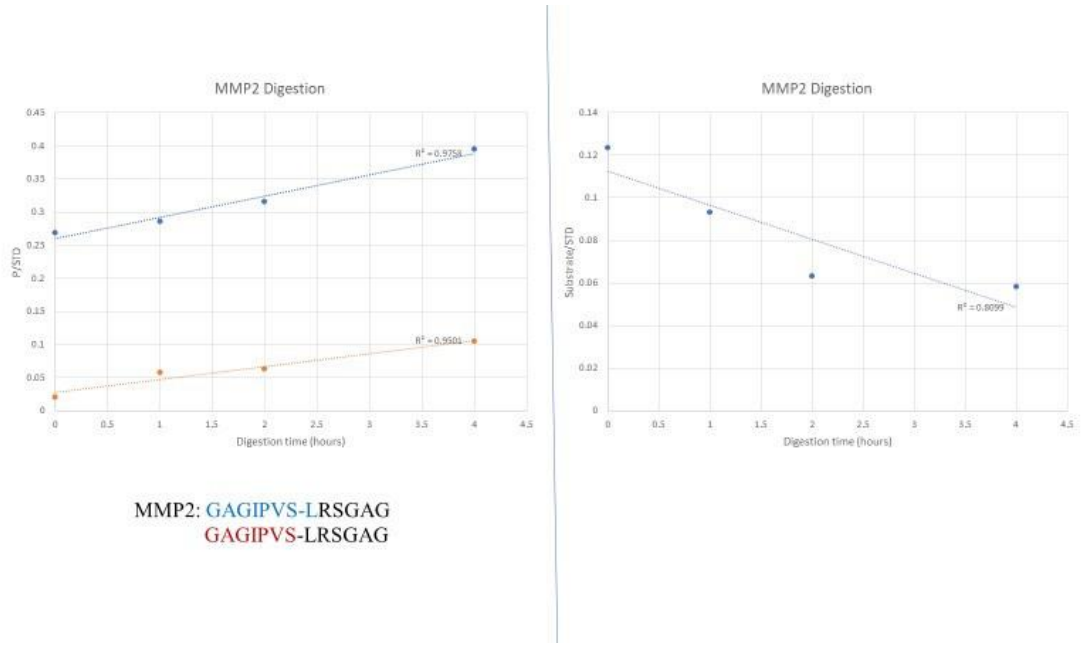
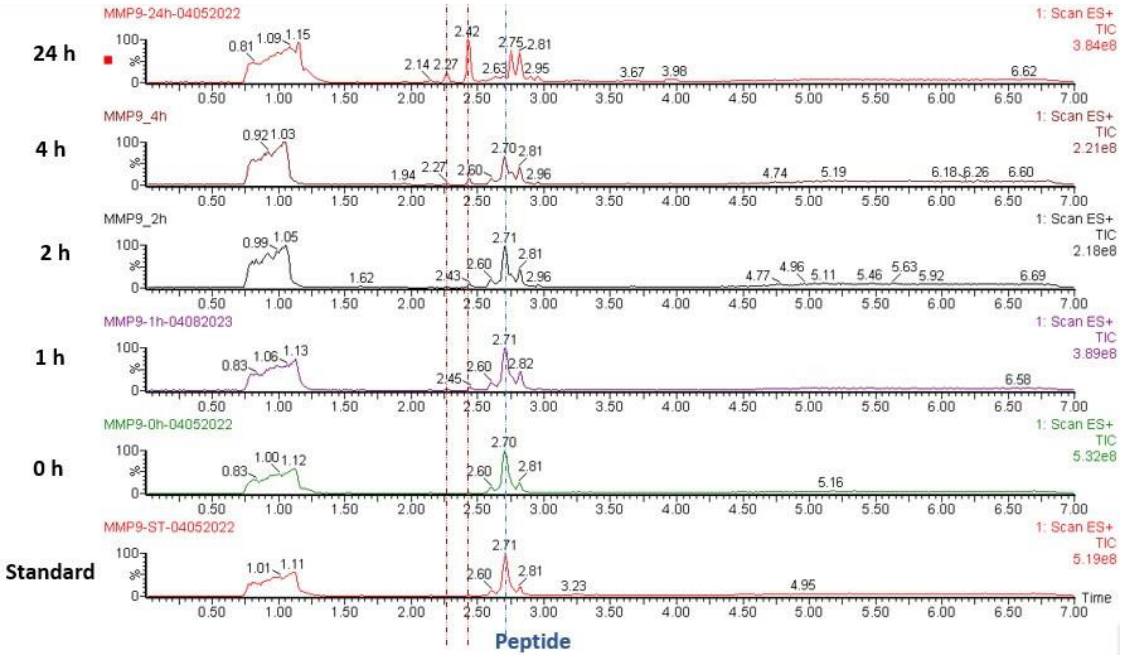


Figure A19. MMP2 consensus sequence digestion with commercial MMP2 protease, created using UPLC-MS results.



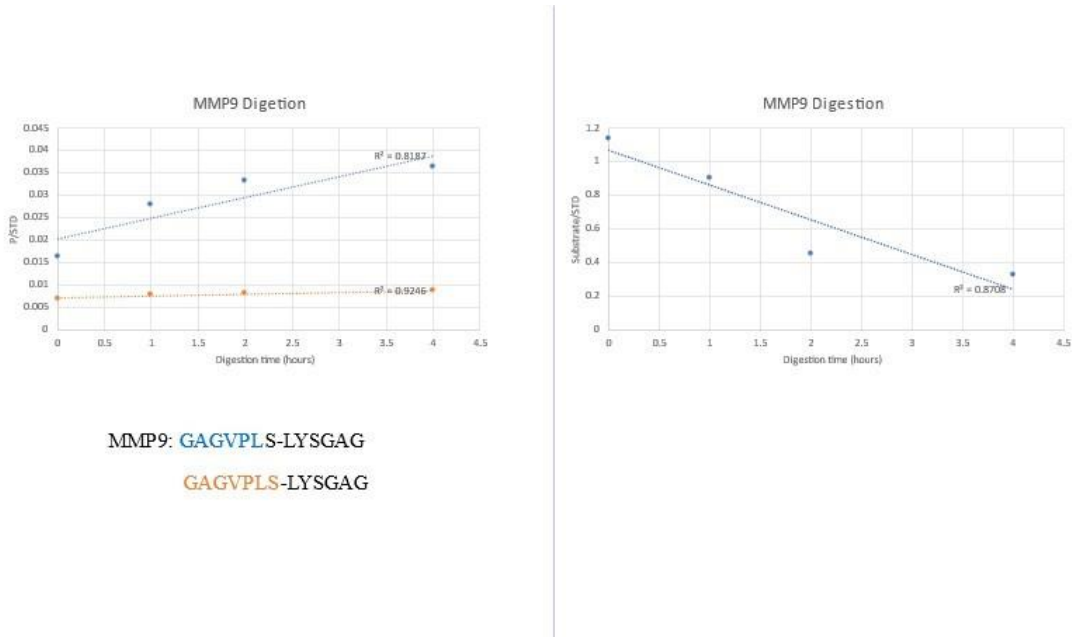
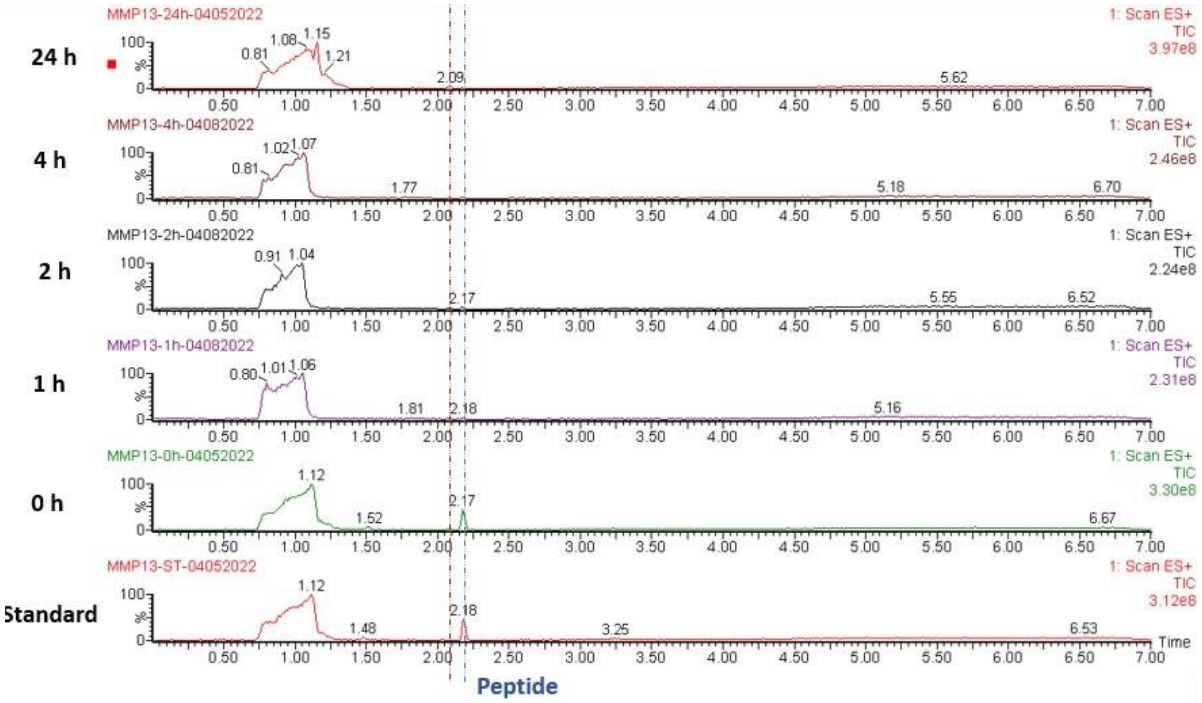
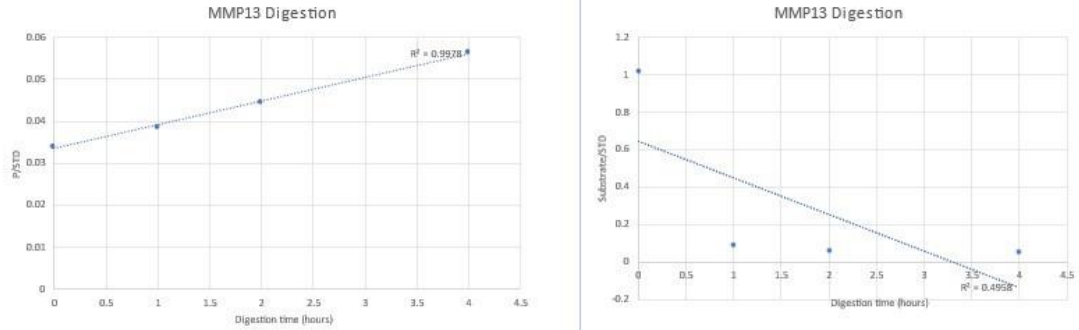


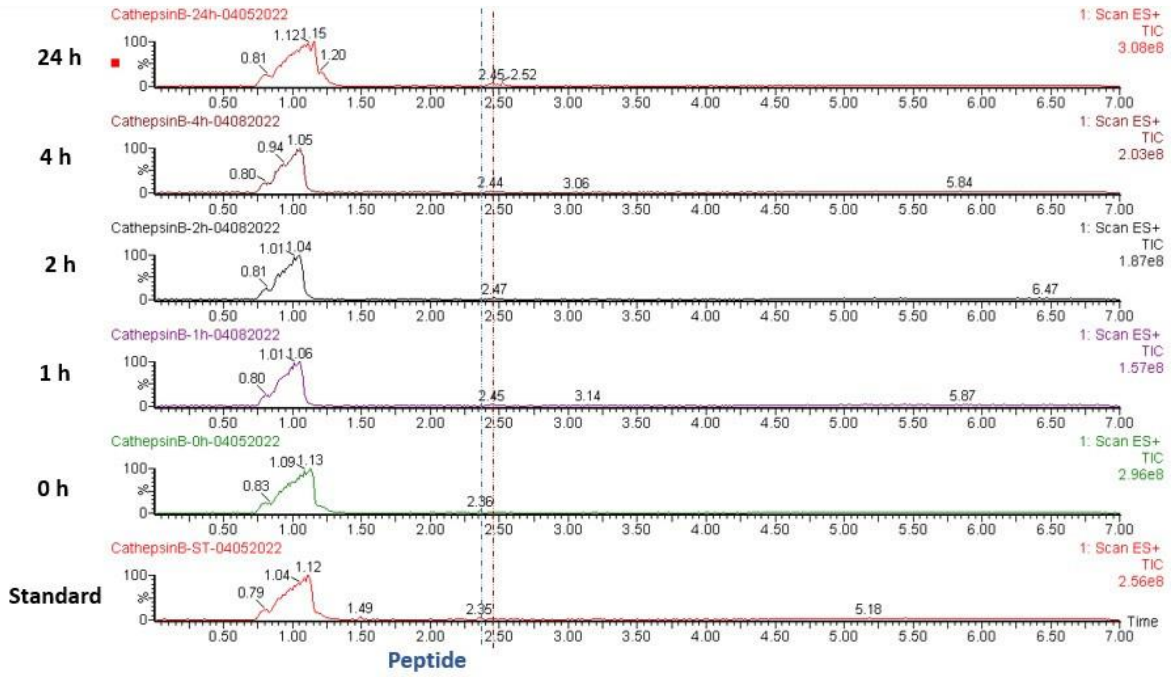
Figure A20. MMP9 consensus sequence digestion with commercial MMP9 protease, created using UPLC-MS results.

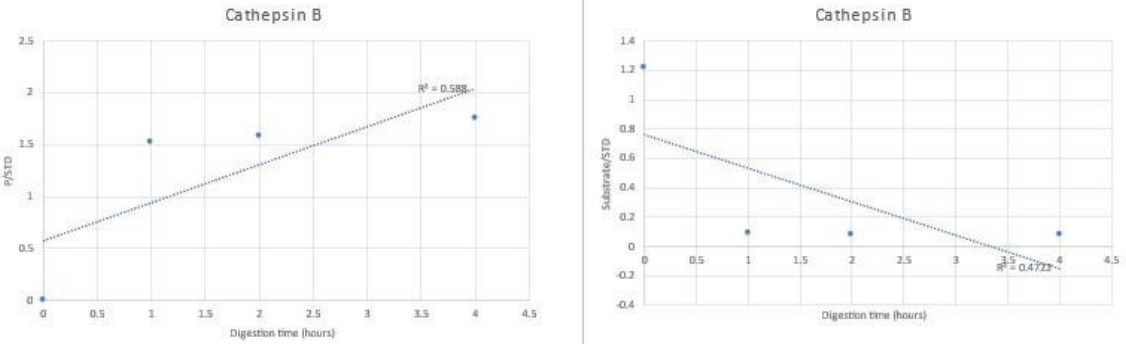




MMP13: GAGPQGLA-GQRGIVAG

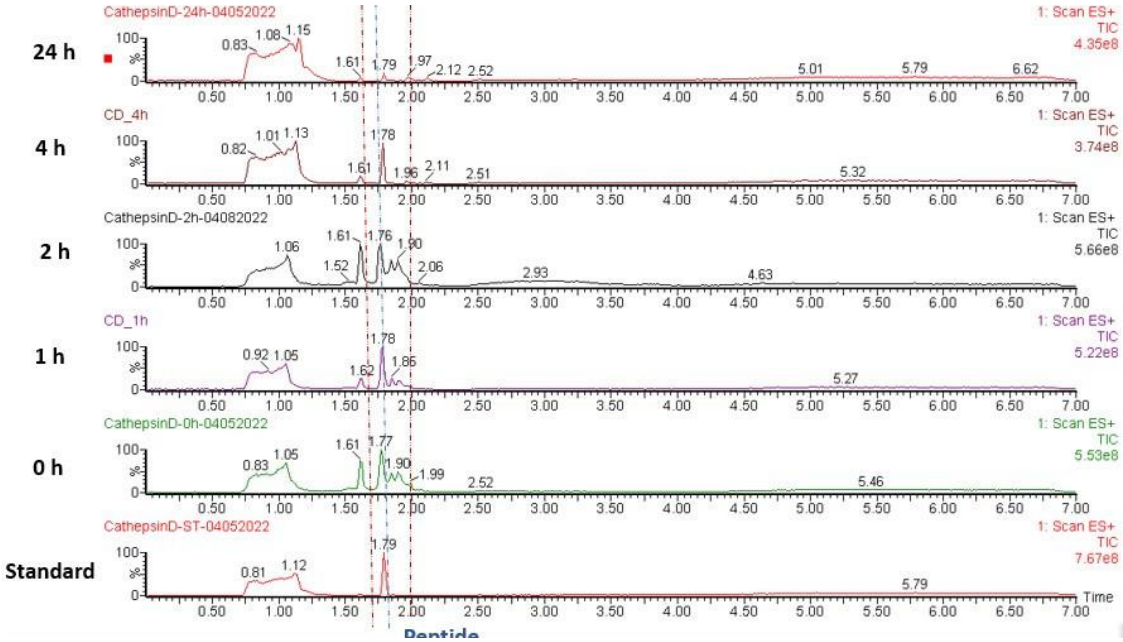
Figure A21. MMP13 consensus sequence digestion with commercial MMP13 protease, created using UPLC-MS results.





Cathepsin B: GAGSLLKSR -MVPNFNAG

Figure A22. CTSB consensus sequence digestion with commercial CTSB protease, created using UPLC-MS results.



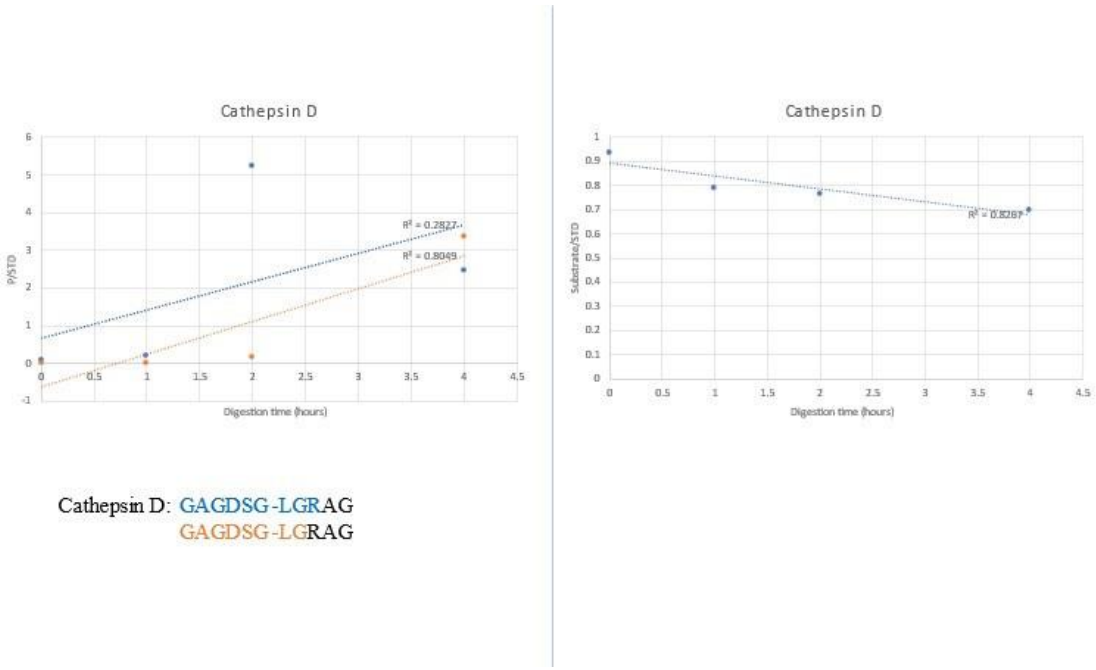
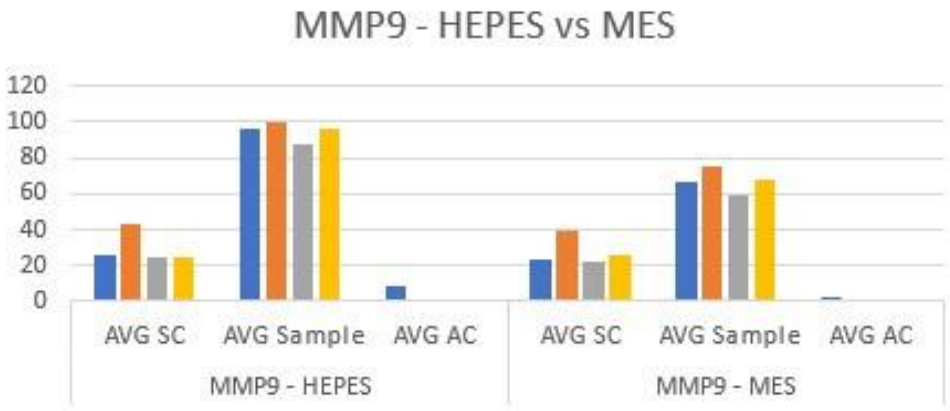
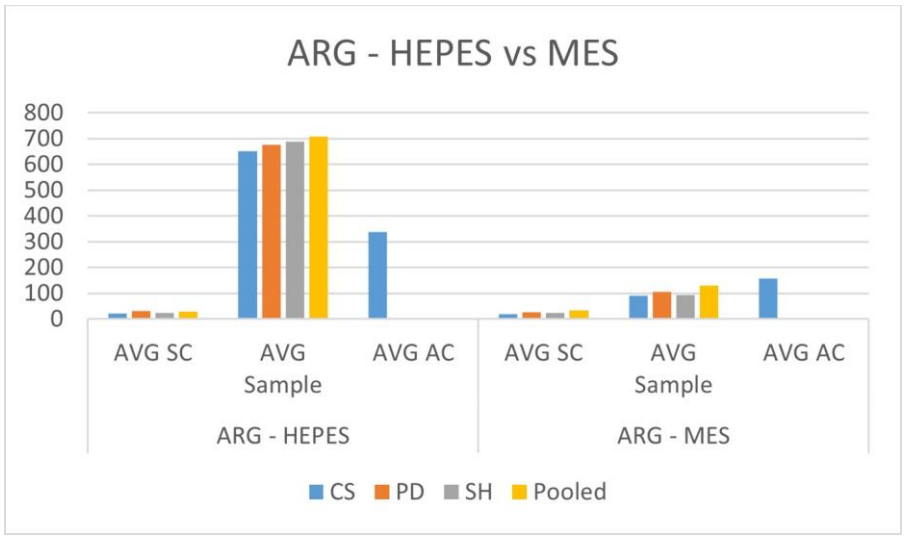
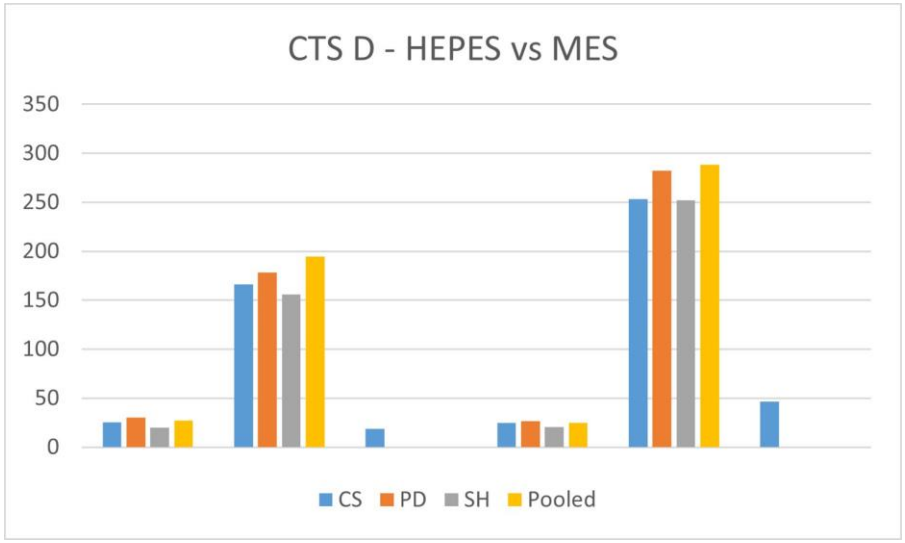


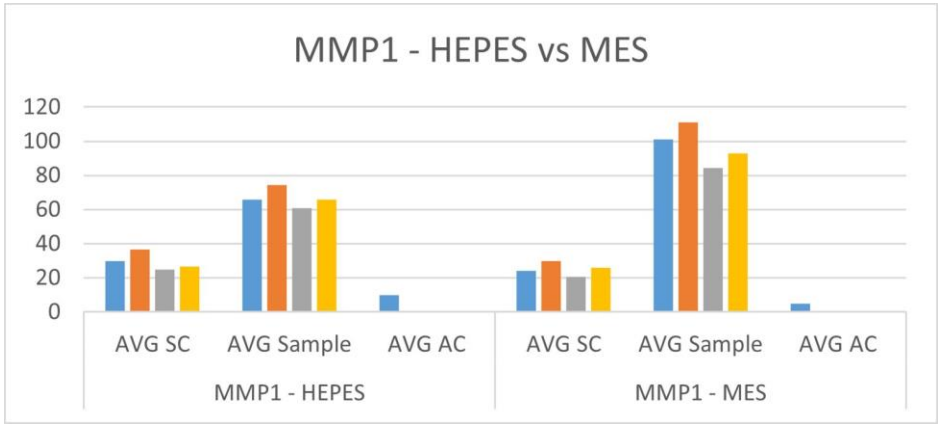
Figure A23. CTSD consensus sequence digestion with commercial CTSD protease, created using UPLC-MS results.

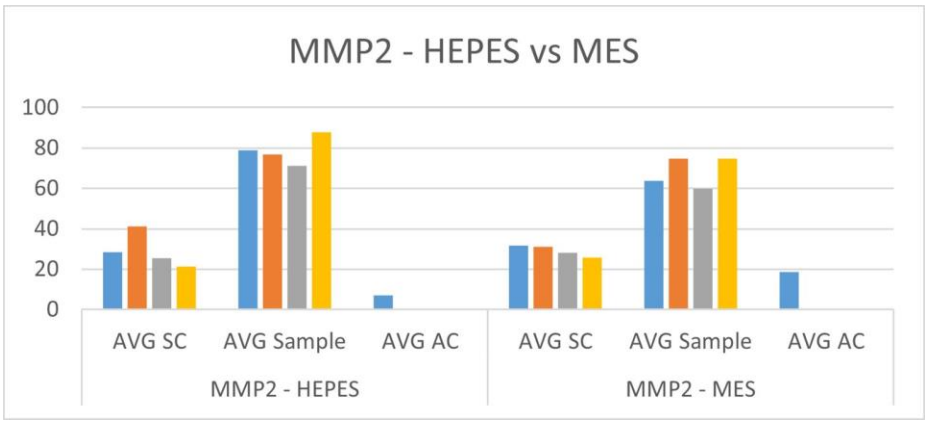
HEPES vs MES buffer– 19 nanobiosensors and control serum

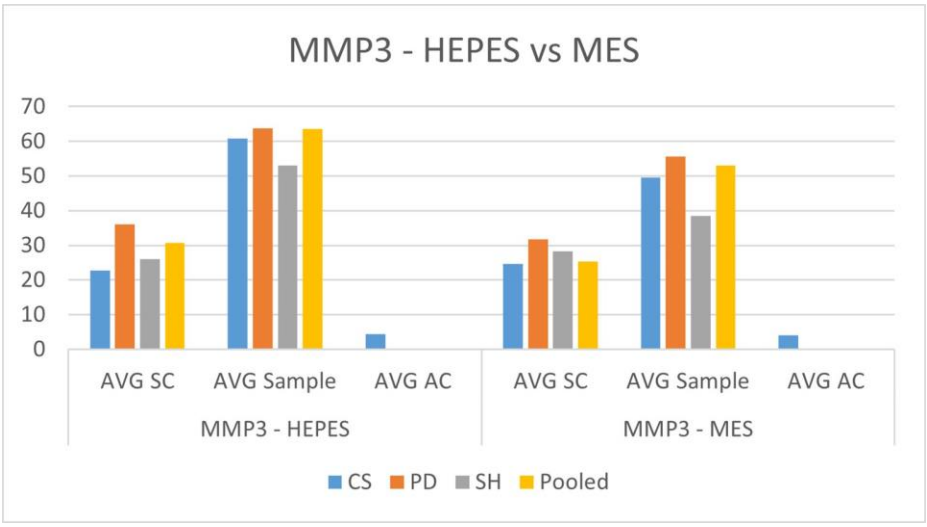


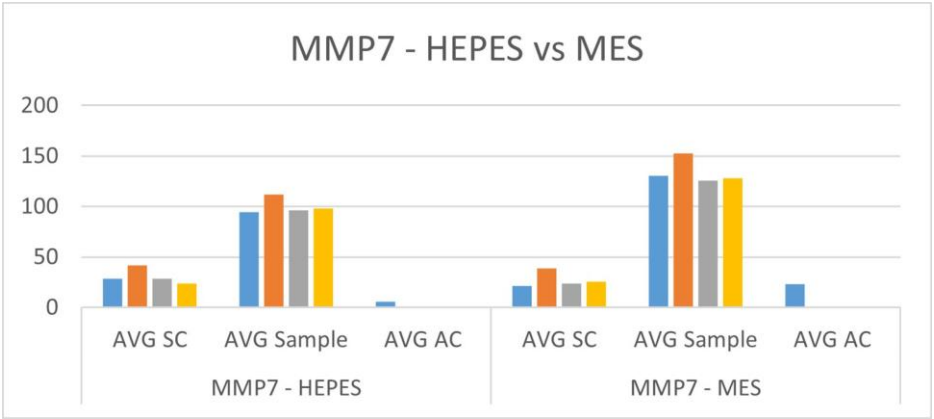

ARG - HEPES vs MES
800
700
600
500
400
300
200
100
0
AVG SC
AVG Sample
AVG AC
AVG SC
AVG Sample
AVG AC
ARG - HEPES
ARG - MES
CS
PD
SH
Pooled

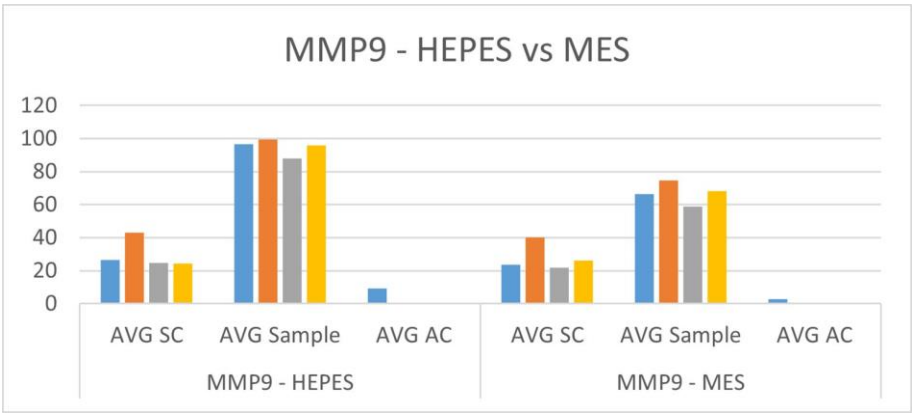

CTS D - HEPES vs MES
350
300
250
200
150
100
50
0
CS
PD
SH
Pooled

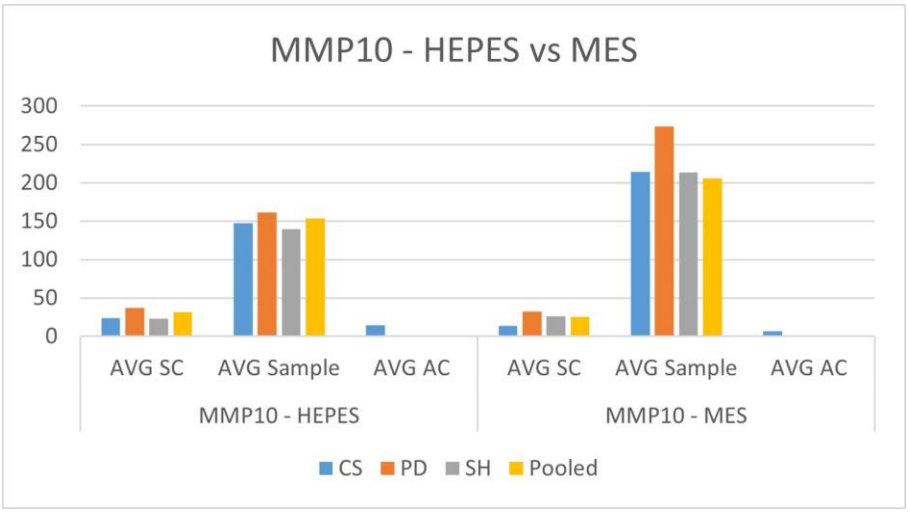

MMP1 - HEPES vs MES
120
100
80
60
40
20
0
AVG SC
AVG Sample
AVG AC
AVG SC
AVG Sample
AVG AC
MMP1 - HEPES
MMP1 - MES

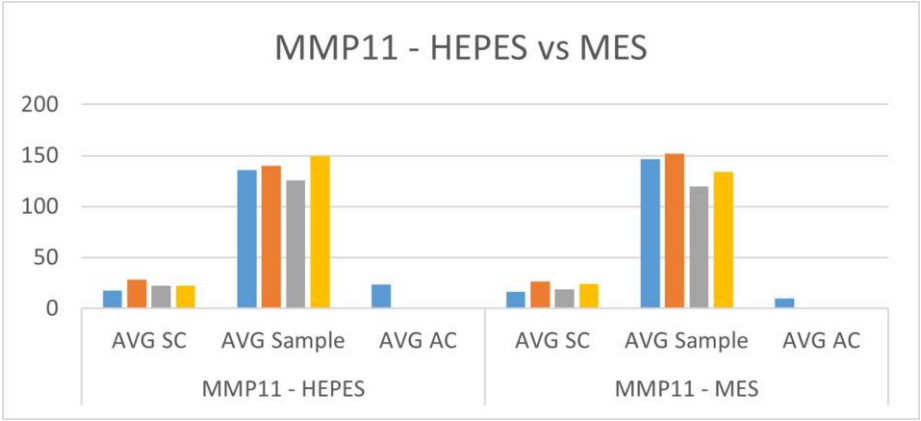

MMP2 - HEPES vs MES
100
80
60
40
20
0
AVG SC
AVG Sample
AVG AC
AVG SC
AVG Sample
AVG AC
MMP2 - HEPES
MMP2 - MES

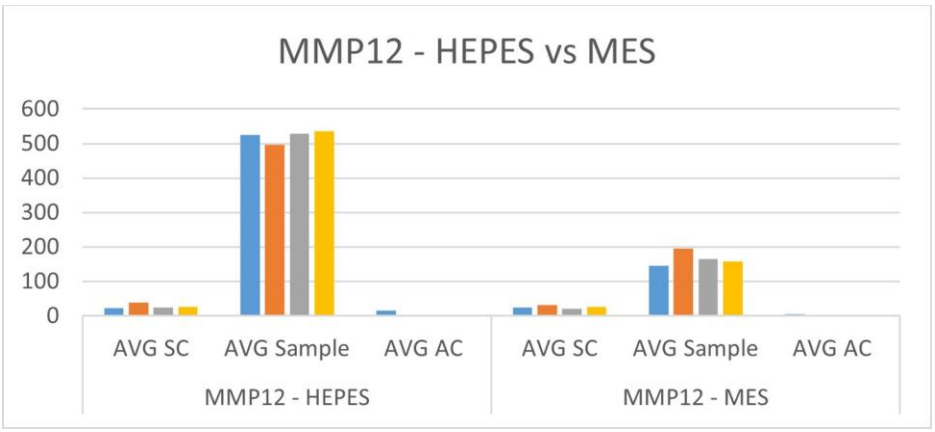

MMP3 - HEPES vs MES
70
60
50
40
30
20
10
0
AVG SC
AVG Sample
AVG AC
AVG SC
AVG Sample
AVG AC
MMP3 - HEPES
MMP3 - MES
CS
PD
SH
Pooled

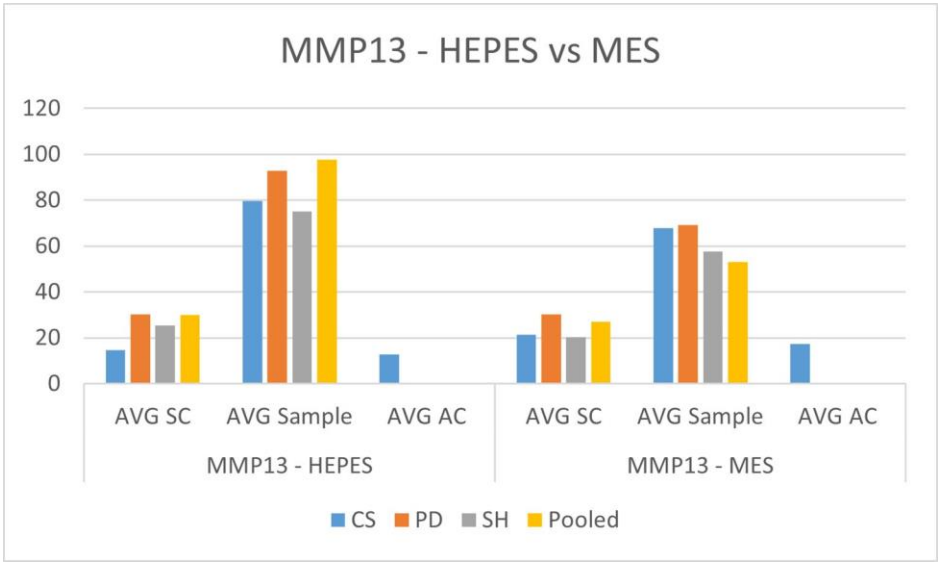

MMP7 - HEPES vs MES
200
150
100
50
0
AVG SC
AVG Sample
AVG AC
AVG SC
AVG Sample
AVG AC
MMP7 - HEPES
MMP7 - MES

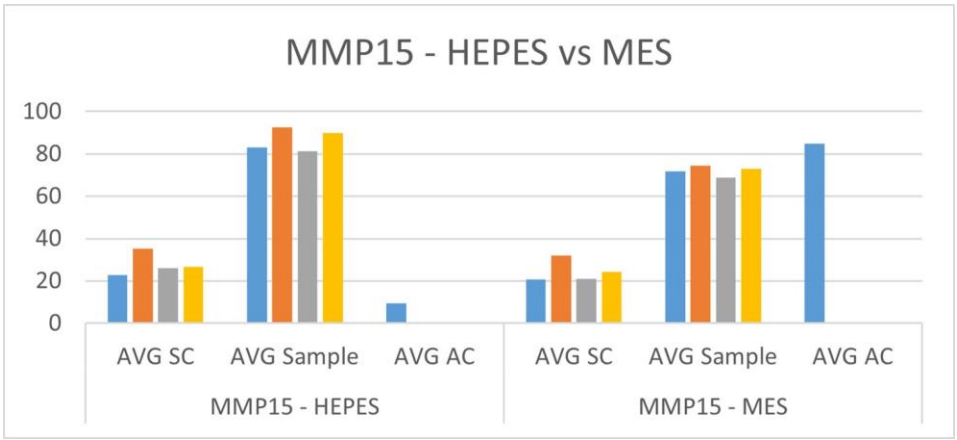

MMP9 - HEPES vs MES
120
100
80
60
40
20
0
AVG SC
AVG Sample
AVG AC
AVG SC
AVG Sample
AVG AC
MMP9 - HEPES
MMP9 - MES

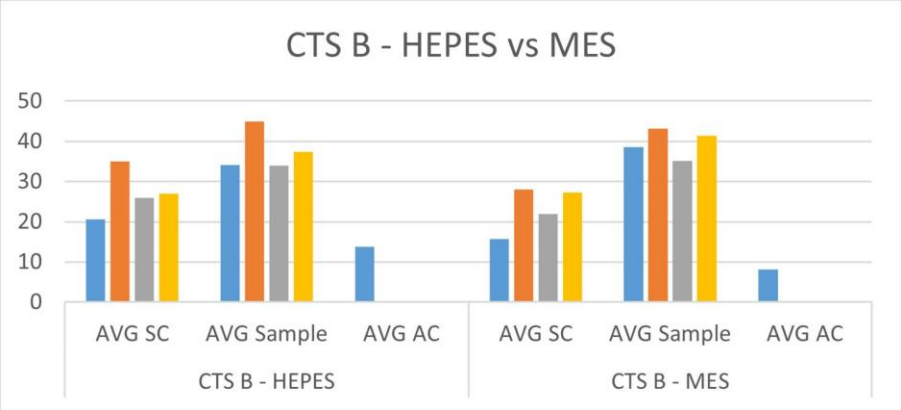

MMP10 - HEPES vs MES
300
250
200
150
100
50
0
AVG SC
AVG Sample
AVG AC
AVG SC
AVG Sample
AVG AC
MMP10 - HEPES
MMP10 - MES
CS
PD
SH
Pooled

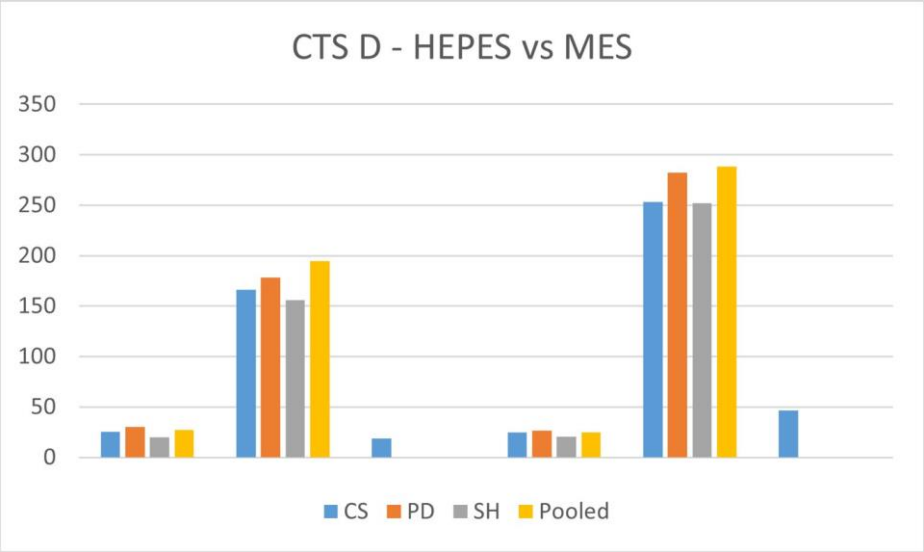

MMP11 - HEPES vs MES
200
150
100
50
0
AVG SC
AVG Sample
AVG AC
AVG SC
AVG Sample
AVG AC
MMP11 - HEPES
MMP11 - MES

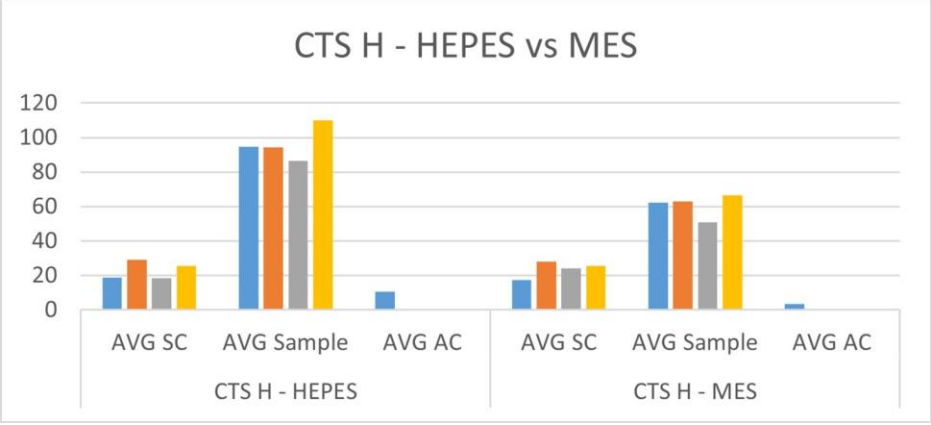

MMP12 - HEPES vs MES
600
500
400
300
200
100
0
AVG SC
AVG Sample
AVG AC
AVG SC
AVG Sample
AVG AC
MMP12 - HEPES
MMP12 - MES

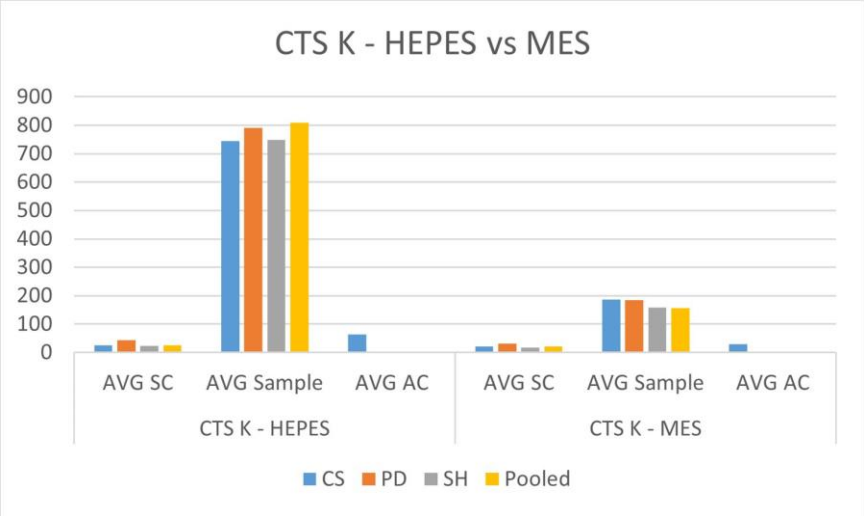

MMP13 - HEPES vs MES
120
100
80
60
40
20
0
AVG SC
AVG Sample
AVG AC
AVG SC
AVG Sample
AVG AC
MMP13 - HEPES
MMP13 - MES
CS
PD
SH
Pooled

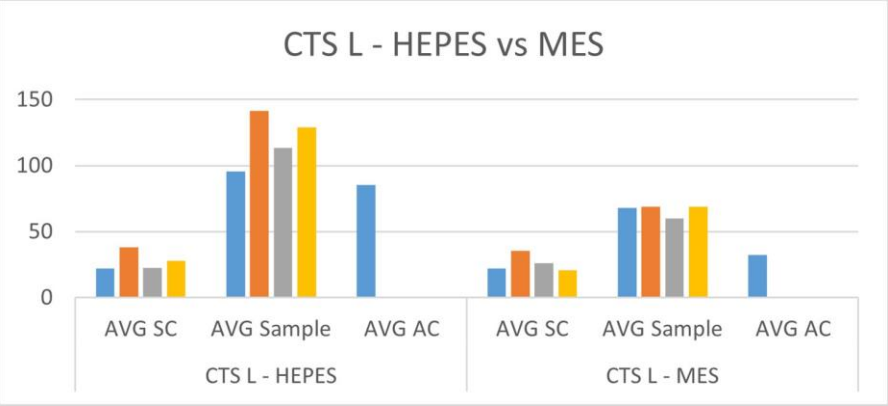

MMP15 - HEPES vs MES
100
80
60
40
20
0
AVG SC
AVG Sample
AVG AC
AVG SC
AVG Sample
AVG AC
MMP15 - HEPES
MMP15 - MES

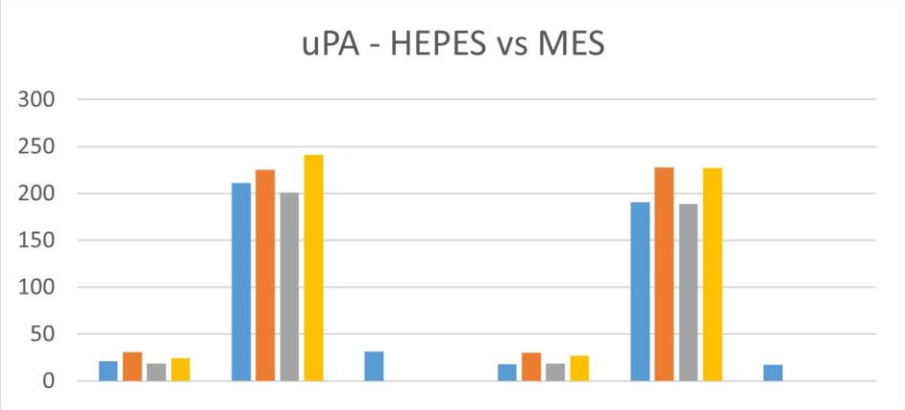

CTS B - HEPES vs MES
50
40
30
20
10
0
AVG SC
AVG Sample
AVG AC
AVG SC
AVG Sample
AVG AC
CTS B - HEPES
CTS B - MES

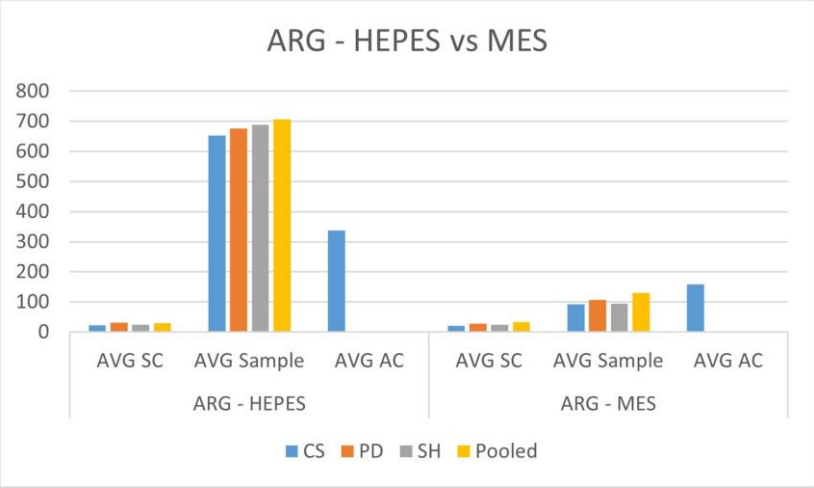

CTS D - HEPES vs MES
350
300
250
200
150
100
50
0
CS
PD
SH
Pooled

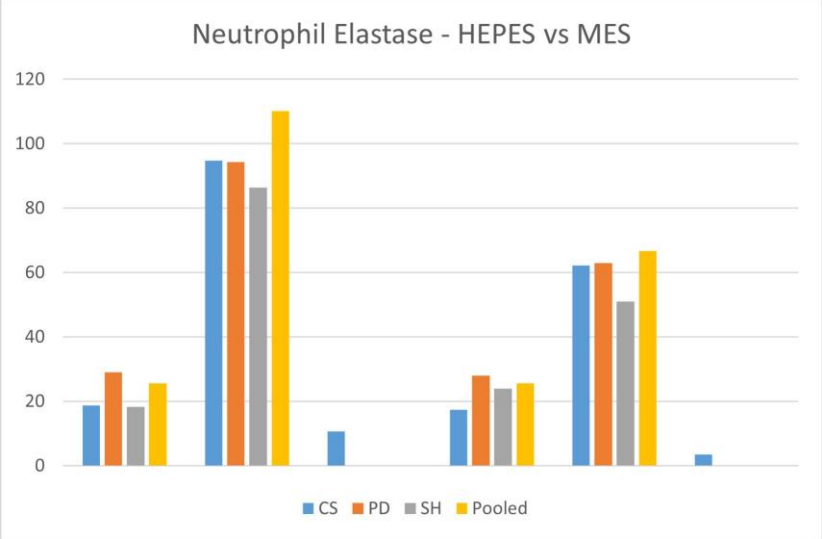

CTS H - HEPES vs MES
120
100
80
60
40
20
0
AVG SC
AVG Sample
AVG AC
AVG SC
AVG Sample
AVG AC
CTS H - HEPES
CTS H - MES


CTS K - HEPES vs MES
900
800
700
600
500
400
300
200
100
0
AVG SC
AVG Sample
AVG AC
AVG SC
AVG Sample
AVG AC
CTS K - HEPES
CTS K - MES
CS
PD
SH
Pooled


CTS L - HEPES vs MES
150
100
50
0
AVG SC
AVG Sample
AVG AC
AVG SC
AVG Sample
AVG AC
CTS L - HEPES
CTS L - MES


uPA - HEPES vs MES
300
250
200
150
100
50
0


ARG - HEPES vs MES
800
700
600
500
400
300
200
100
0
AVG SC
AVG Sample
AVG AC
AVG SC
AVG Sample
AVG AC
ARG - HEPES
ARG - MES
CS
PD
SH
Pooled


Neutrophil Elastase - HEPES vs MES
120
100
80
60
40
20
0
CS
PD
SH
Pooled

TEM images of all nanobiosensors

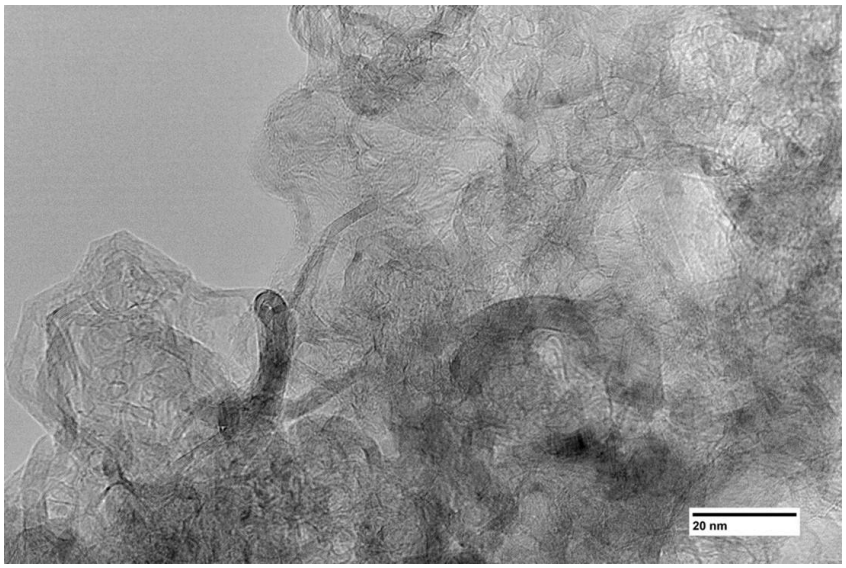


Figure A24. TEM (200 KeV) of Explosion graphene from the Sorensen Group, Physics, Kansas State University. This material is the graphene core of all batch 2 nanobiosensors.

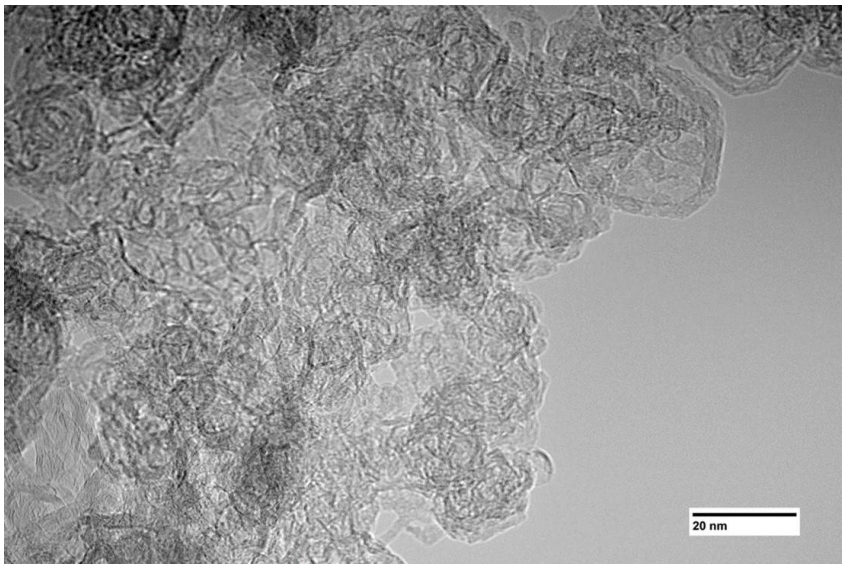


Figure A25. TEM (200 KeV) of Carboxygraphene, batch 2.

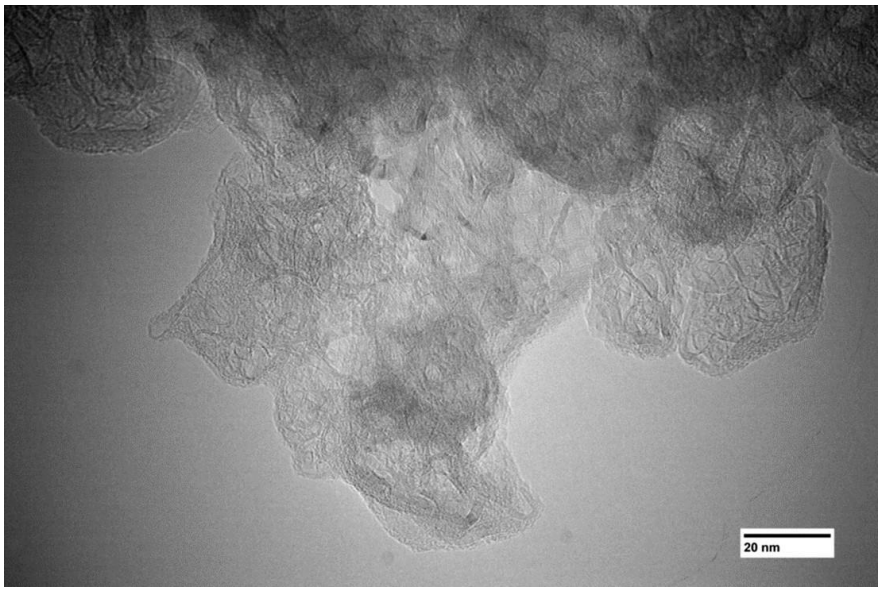


Figure A26. TEM (200 KeV) of explosion graphene-based nanobiosensor for ARG Activity Detection, batch 2.

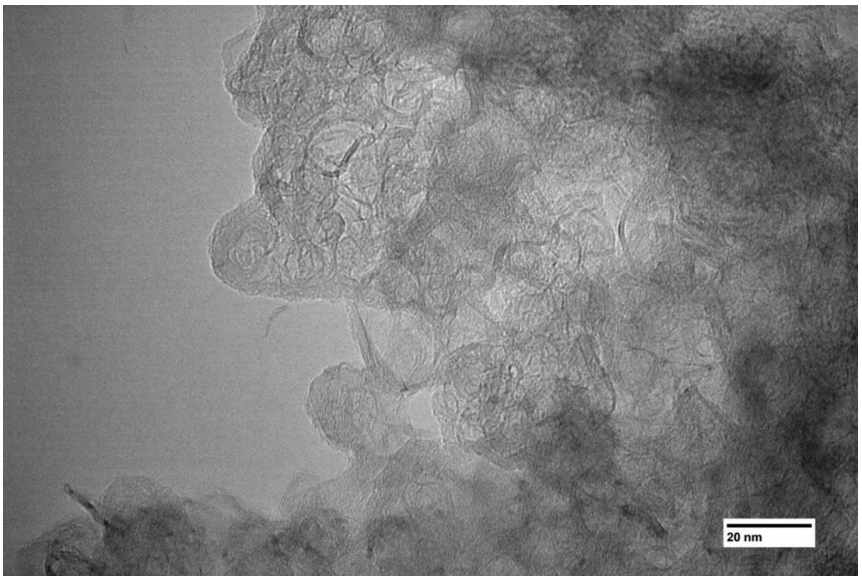


Figure A27. TEM (200 KeV) of explosion graphene-based nanobiosensor for CTSB Activity Detection, batch 2.

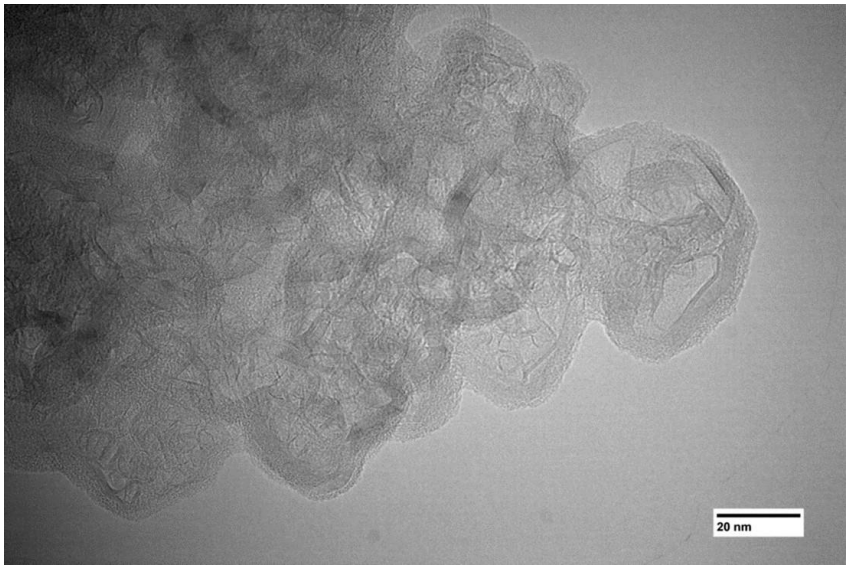


Figure A28. TEM (200 KeV) of explosion graphene-based nanobiosensor for CTSD Activity Detection, batch 2.

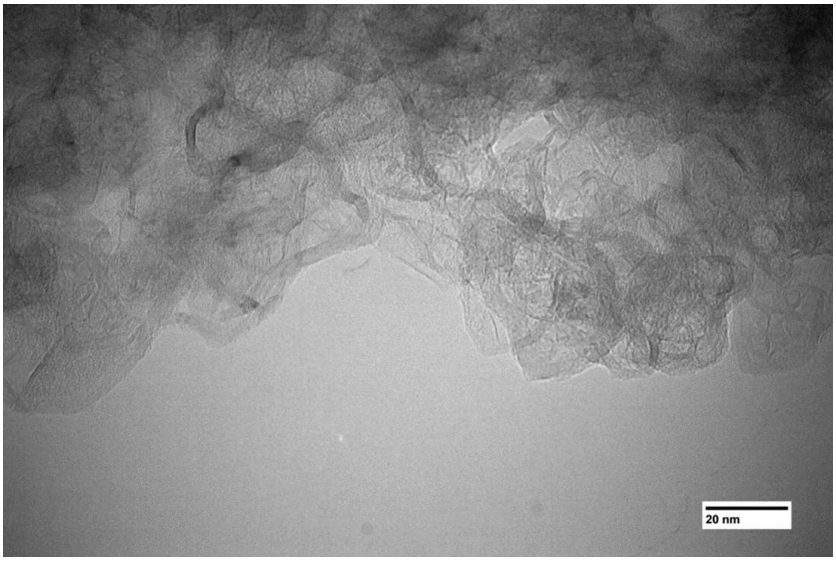


Figure A29. TEM (200 KeV) of explosion graphene-based nanobiosensor for CTSE Activity Detection, batch 2.

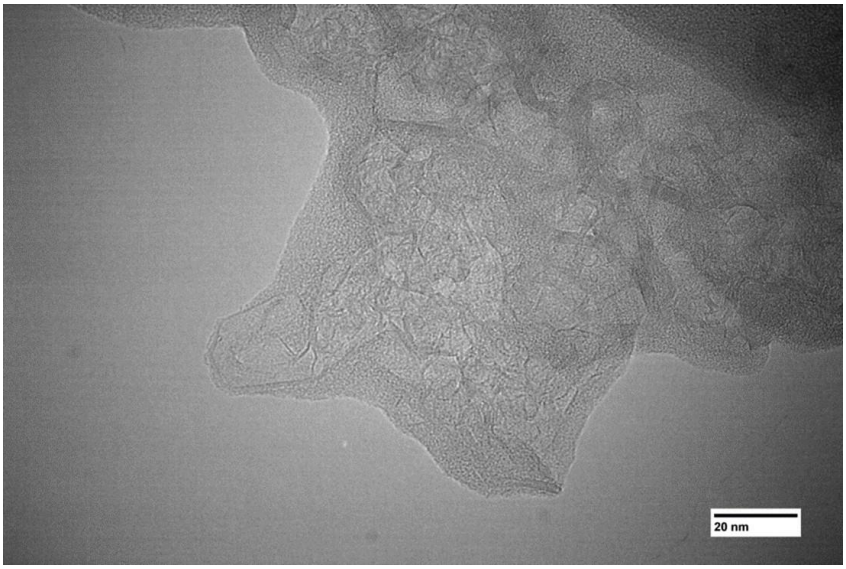


Figure A30. TEM (200 KeV) of explosion graphene-based nanobiosensor for CTSH Activity Detection, batch 2.

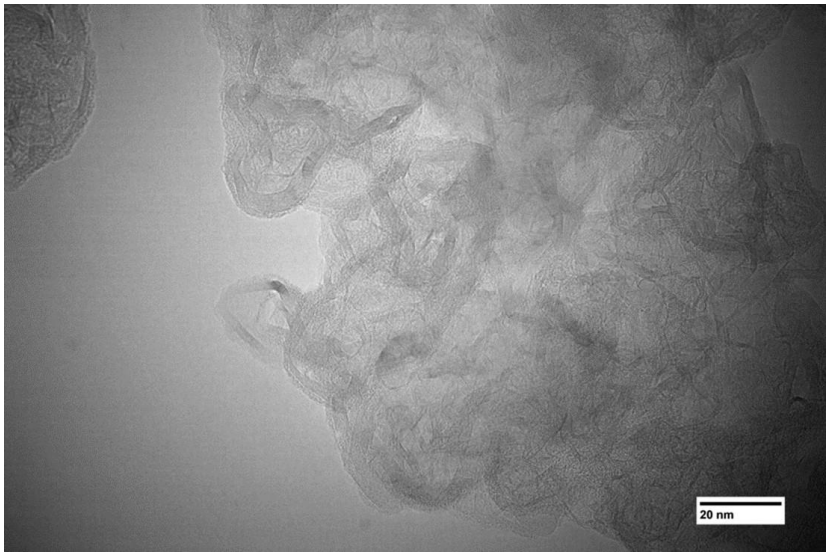


Figure A31. TEM (200 KeV) of explosion graphene-based nanobiosensor for CTSK Activity Detection, batch 2.

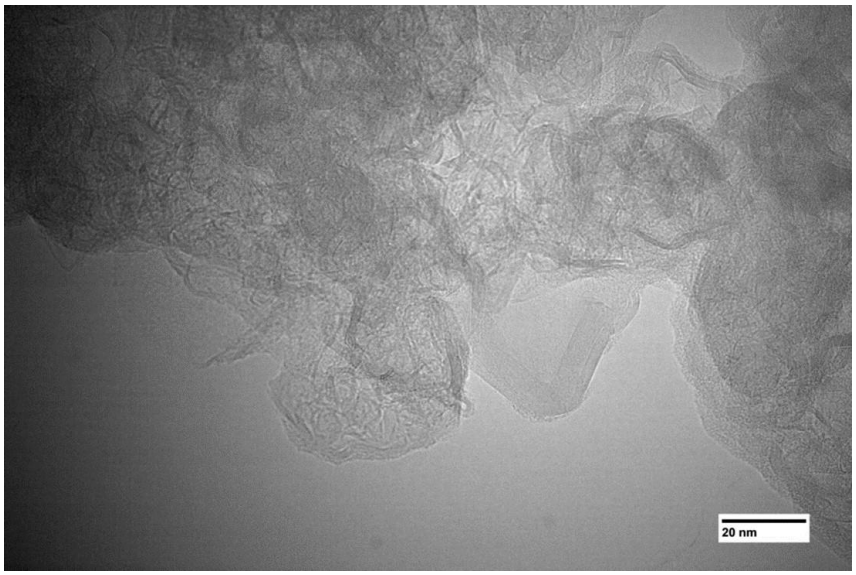


Figure A32. TEM (200 KeV) of explosion graphene-based nanobiosensor for CTSL Activity Detection, batch 2.

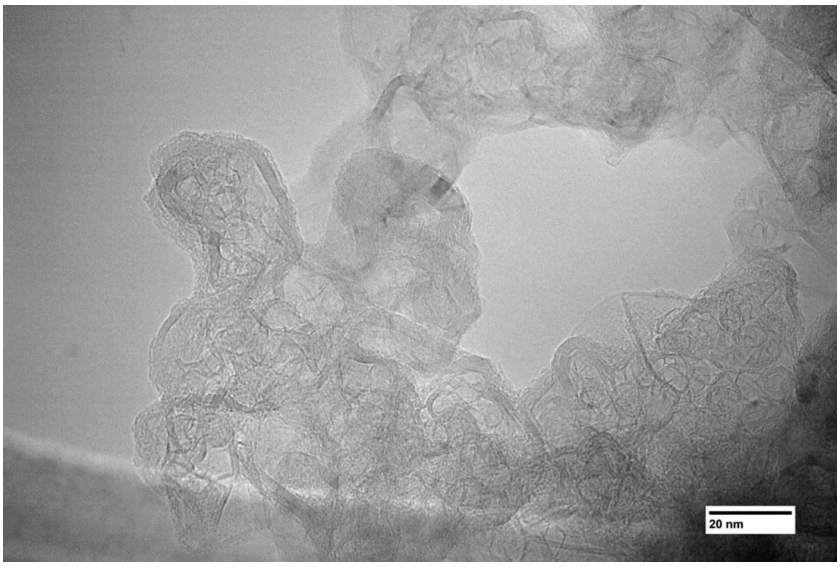


Figure A33. TEM (200 KeV) of explosion graphene-based nanobiosensor for Matrix Metalloproteinase 1 Activity Detection, batch 2.

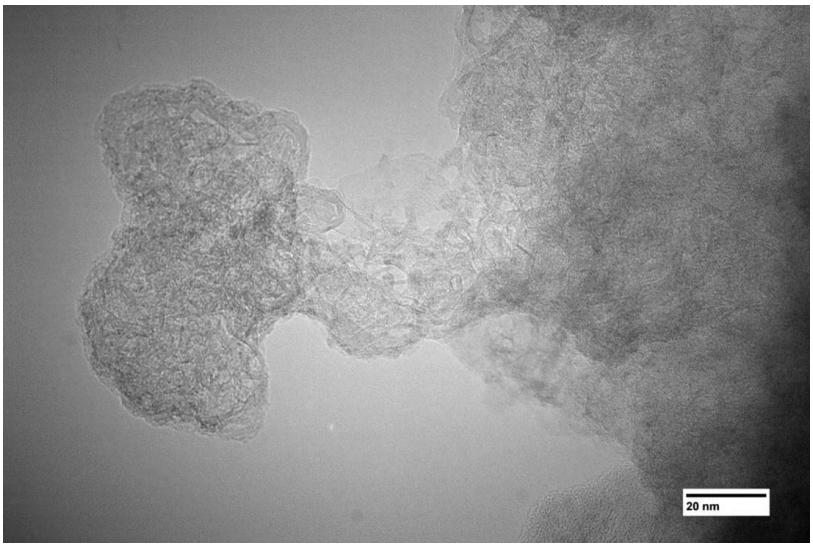


Figure A34. TEM (200 KeV) of explosion graphene-based nanobiosensor for Matrix Metalloproteinase 2 Activity Detection, batch 2.

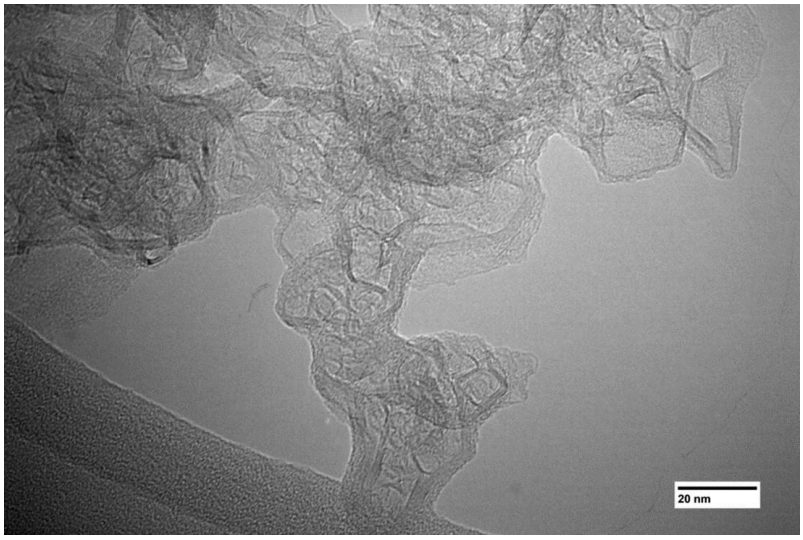


Figure A35. TEM (200 KeV) of explosion graphene-based nanobiosensor for Matrix Metalloproteinase 3 Activity Detection, batch 2.

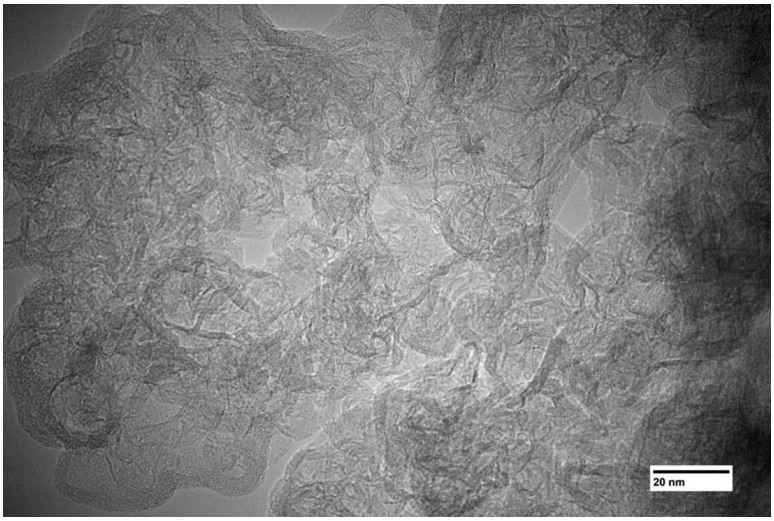


Figure A36. TEM (200 KeV) of explosion graphene-based nanobiosensor for Matrix Metalloproteinase 7 Activity Detection, batch 2.

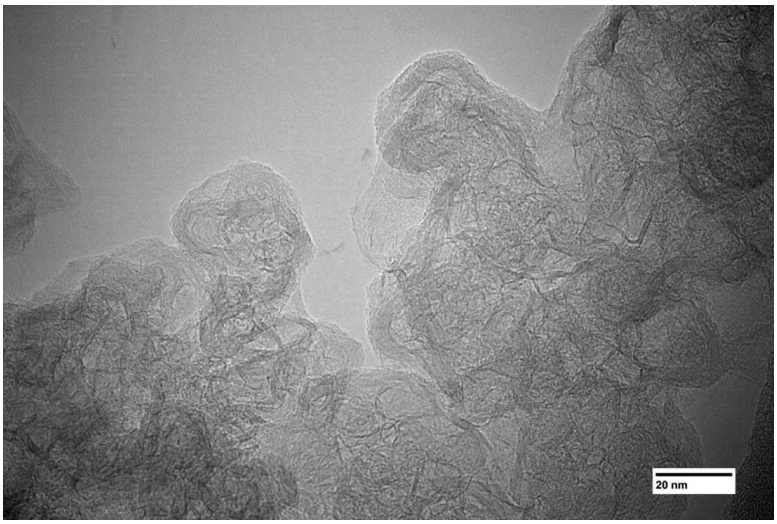


Figure A37. TEM (200 KeV) of explosion graphene-based nanobiosensor for Matrix Metalloproteinase 9 Activity Detection, batch 2.

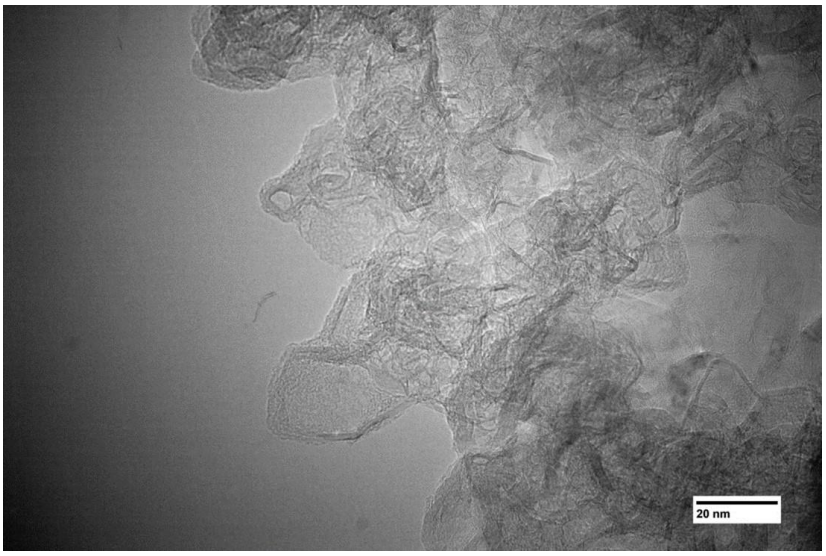


Figure A38. TEM (200 KeV) of explosion graphene-based nanobiosensor for Matrix Metalloproteinase 10 Activity Detection, batch 2.

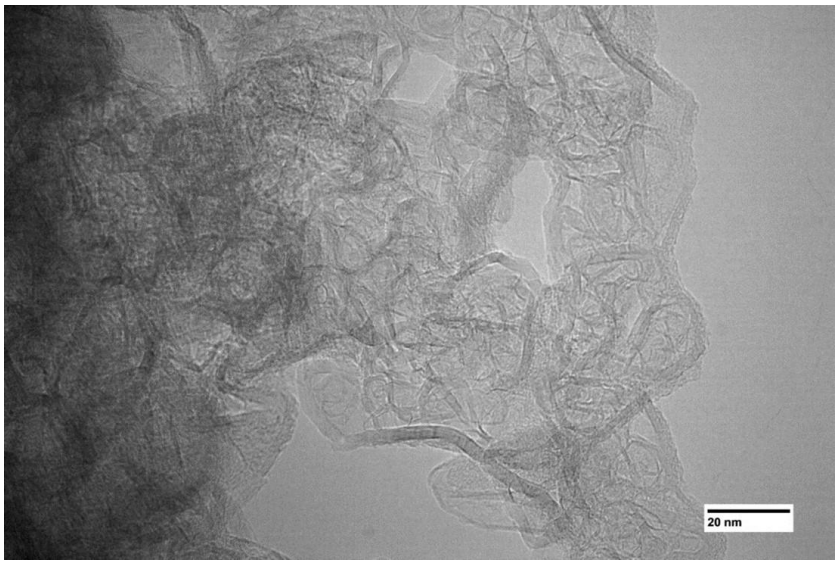


Figure A39. TEM (200 KeV) of explosion graphene-based nanobiosensor for Matrix Metalloproteinase 11 Activity Detection, batch 2.

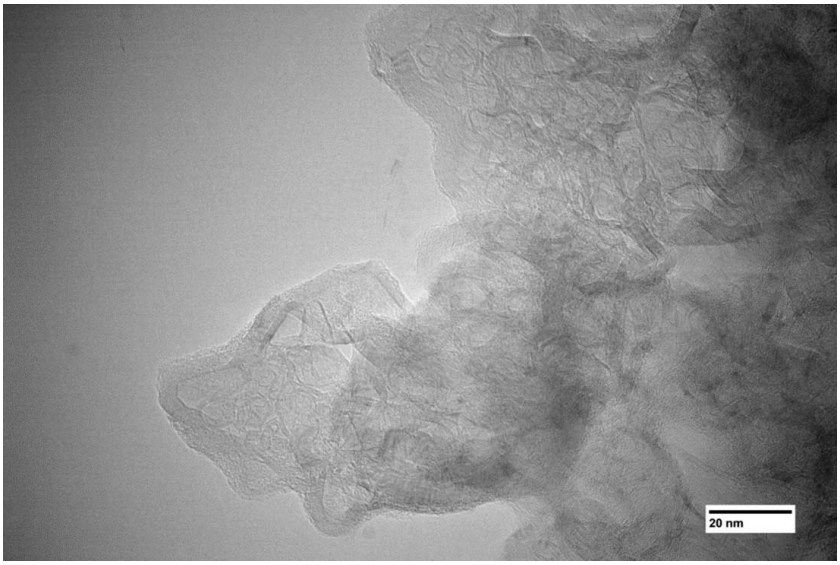


Figure A40. TEM (200 KeV) of explosion graphene-based nanobiosensor for Matrix Metalloproteinase 11 Activity Detection, batch 2.

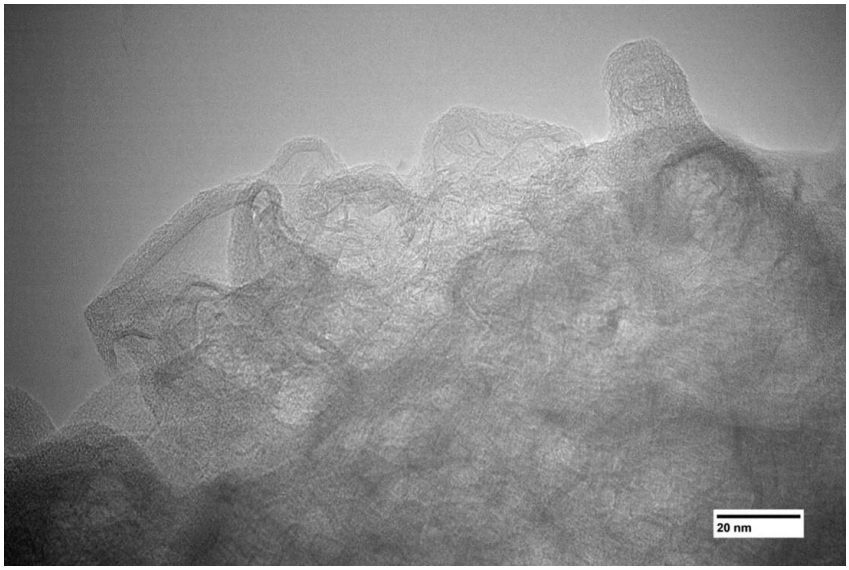


Figure A41. TEM (200 KeV) of explosion graphene-based nanobiosensor for Matrix Metalloproteinase 13 Activity Detection, batch 2.

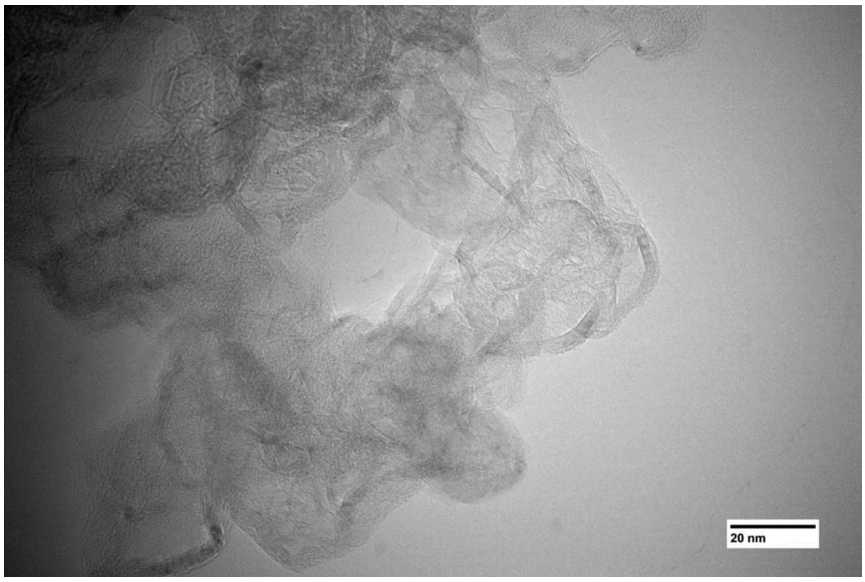


Figure A42. TEM (200 KeV) of explosion graphene-based nanobiosensor for Matrix Metalloproteinase 15 Activity Detection, batch 2.

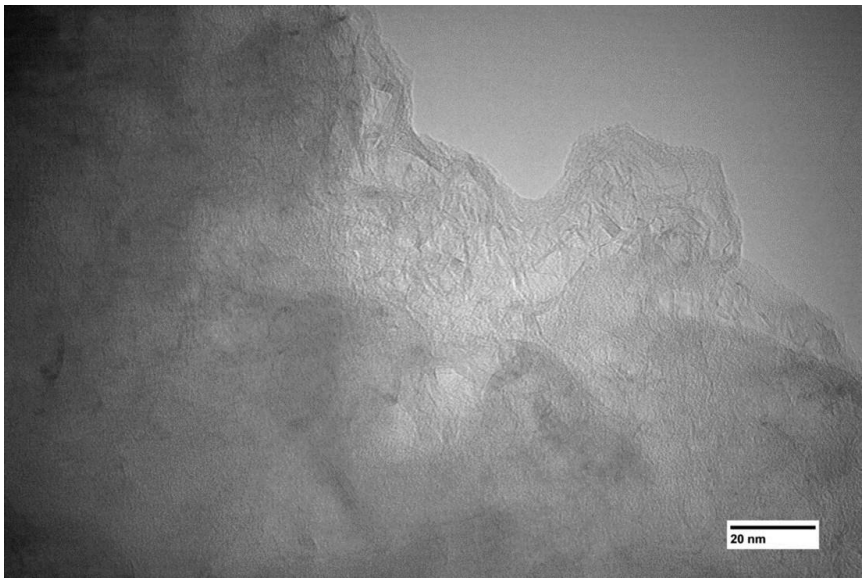


Figure A43. TEM (200 KeV) of explosion graphene-based nanobiosensor for Matrix Neutrophil Elastase Activity Detection, batch 2.

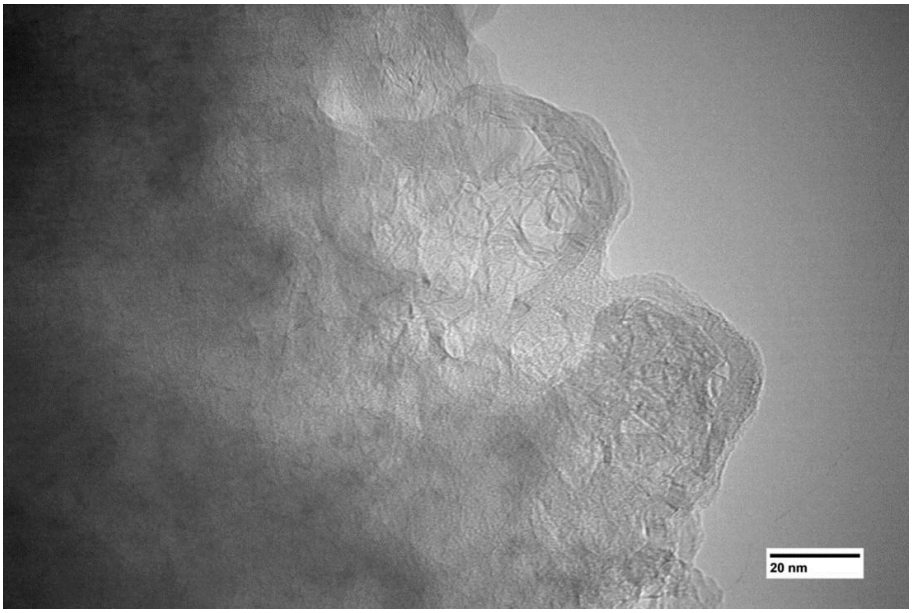


Figure A44. TEM (200 KeV) of explosion graphene-based nanobiosensor for Urokinase Plasminogen Activator 15 Activity Detection, batch 2.