

**A NEW METHOD FOR IDENTIFYING CORONARY ARTERY DISEASE
VIA ANALYSIS OF HUMAN SERUM**

An Undergraduate Research Scholars Thesis

by

JACK RILEY MARTINEZ
and
JENNIFER BACHAND
and
AMY CHUNG
and
ALEXIS LUEDKE
and
NYLA VASQUEZ

Submitted to Honors and Undergraduate Research
Texas A&M University
in partial fulfillment of the requirements for the designation as an

UNDERGRADUATE RESEARCH SCHOLAR

Approved by
Research Advisor:

Dr. Ronald Macfarlane

May 2015

Majors: Biomedical Sciences
Mathematics
Biology
Microbiology
Biology

TABLE OF CONTENTS

	Page
ABSTRACT.....	1
DEDICATION.....	3
ACKNOWLEDGEMENTS.....	4
NOMENCLATURE.....	5
CHAPTER	
I INTRODUCTION.....	6
II METHODS.....	7
Materials.....	7
Human Serum Collection.....	8
Separation Conditions.....	8
Preparation of Molecular Weight Ladder.....	8
Human Sera Preparation.....	9
Analysis of Serum.....	10
Statistical Analysis of Serum.....	10
III RESULTS.....	12
IV CONCLUSION.....	20
REFERENCES.....	22
APPENDIX.....	23

ABSTRACT

A New Method for Identifying Coronary Artery Disease via Analysis of Human Serum. (May 2015)

Jack Riley Martinez
Department of Biomedical Sciences
Texas A&M University

Jennifer Bachand
Department of Mathematics
Texas A&M University

Amy Chung
Department of Biology
Texas A&M University

Alexis Luedke
Department of Microbiology
Texas A&M University

Nyla Vasquez
Department of Biology
Texas A&M University

Research Advisor: Dr. Ronald Macfarlane
Department of Chemistry

Coronary artery disease (CAD), the number one cause of death in the United States, has traditionally been combated by reactionary diagnostics. However, the current healthcare dynamic is shifting towards a preventative approach, which aims to treat disease prior to the manifestation of symptoms. A minimally invasive diagnostic method which provides insight into an individual's blood chemistry has the potential to facilitate this transition. This study analyzes the protein chemistry of human blood serum in order to differentiate between CAD and Non-CAD serum, with the goal of discovering potential CAD biomarkers. The Beckman/Coulter P/ACE MDQ Capillary Electrophoresis System was coupled with a commercial SDS-Gel based buffer

system in order to achieve highly resolved serum protein spectrums. Electropherograms for nineteen participants were analyzed in 32 Karat™ Software by integrating protein peak areas, which were then normalized using an internal standard. Based on these integrations, a Linear Discriminant Analysis (LDA) algorithm was generated using OriginPro 8.6 Data Analysis and Graphing Software for the purpose of distinguishing between CAD and Non-CAD individuals. Cohorts of twelve CAD and seven Non-CAD participants were used for training purposes. This training data was classified with 100% accuracy, and an obtained p-value of 0.00332 indicated a significant difference between the CAD and Non-CAD participants' peak distributions. This study establishes the utilization of capillary gel electrophoresis techniques for serum classification purposes, and ultimately points toward the development of a new, minimally invasive diagnostic tool for CAD based on blood protein chemistry rather than symptom onset.

DEDICATION

We dedicate this study to persons afflicted with coronary artery disease and their beloved families and friends.

ACKNOWLEDGEMENTS

We would like to thank Texas A&M University for supplying us with the facility and specialized equipment to perform our research. In particular, we would like to thank Katharine Wigginton for her advice and assistance throughout our research. We would also like to thank the members of the Laboratory for Cardiovascular Chemistry for their help at various stages of this project, including Delton Tatum for his support and comments on our original work. Finally, we would like to thank the TAMUS LSAMP NSF grant (NSF HRD-1304975) for providing support to Nyla Vasquez as an undergraduate researcher.

NOMENCLATURE

CAD: Coronary Artery Disease

CGE: Capillary Gel Electrophoresis

HDL: High Density Lipoprotein

IBC: Institutional Biosafety Committee

ID: Internal Diameter

IRB: Institutional Review Board

LDA: Linear Discriminant Analysis

LDL: Low Density Lipoprotein

MW: Molecular Weight

PI: Principal Investigator

CDA: Canonical Discriminant Analysis

CHAPTER I

INTRODUCTION

As the number one cause of death in the United States, CAD takes the lives of approximately 380,000 Americans every year (1). Many of these deaths can be attributed to the progressive and typically irreversible nature of plaque build-up within the arteries (2). Due to current reactionary medical tactics, CAD-afflicted individuals often suffer from advanced stages of the disease that are difficult to reverse. Current heart disease prediction models have extremely low accuracy, debilitating the advancement towards early heart disease detection. The Framingham Study estimates the risk for CAD and myocardial infarction based on total cholesterol levels and other factors such as systolic blood pressure. It both underestimates and overestimates the risk of patients with and without heart disease respectively, due to poor sensitivity and specificity (3,4). It is evident that a new technique to identify heart disease at pre-symptomatic stages is necessary to eliminate the life-threatening manifestations that are common-place in today's society.

In light of this, recent research has attempted to investigate the subtle changes in chemistry that are ultimately responsible for the gross anatomical and physiological changes associated with disease development (5). Analysis of specific chemical components, known as biomarkers, have been shown to play a key role in the identification and assessment of a given disease (6,7). This pilot study capitalizes on the myriad of proteins in human blood serum in the hopes of identifying chemical changes that may be associated with CAD development.

CHAPTER II

METHODS

After obtaining Institutional Review Board (IRB) and Institutional Biosafety Committee (IBC) approval (IRB2014-0375M and IBC2013-161), a library of human serum samples of CAD and Non-CAD participants were run through the P/ACE MDQ Capillary Electrophoresis System in order to obtain electropherograms displaying the protein peaks of each separation.

Analysis of protein chemistry in human blood serum was performed by using the Beckman/Coulter PACE MDQ Capillary Electrophoresis System in capillary gel electrophoresis (CGE) mode. The Beckman/Coulter System incorporates the 32 Karat™ Software, which provides a detailed analysis of each sample in the form of an electropherogram.

Materials

All capillaries, buffers, protein standard and wash solutions were obtained from the Beckman Coulter SDS-MW Kit (# 390953). Contents of the kit that were used in the analyses include capillaries of 50 μm I.D. bare-fused silica, a proprietary formulation of SDS-MW Gel Buffer (pH 8), 0.2% SDS, SDS-MW Sample Buffer (100 mM Tris-HCl pH 9.0, 1%SDS), SDS-MW Size Standard—*referred to as the MW ladder* (10 to 225 kDa, 16 mg/mL), Acidic Wash Solution (0.1 N HCl), and Basic Wash Solution (0.1 N NaOH). β -mercaptoethanol was acquired from Sigma-Aldrich (PN M6250). Twelve CAD and seven Non-CAD clinical serum samples were obtained from the -85°C sample repository of the Laboratory for Cardiovascular Chemistry (LCC).

Human Serum Collection

The serum samples were acquired from study participants at Scott & White Hospital in Temple, Texas with informed consent. All participants were of Caucasian race and their ages ranged from 44-81 years old. These participants were clinically diagnosed with the presence or absence of CAD by their medical histories (e.g, coronary artery bypass graft(s), angioplasty/stent(s)) or unremarkable angiographies, respectively. The participants underwent a 12-hour fast before the blood was collected into 9.5 mL Vacutainer™ tubes which had a silica activator and polymer gel (#366510, Beckton Dickinson Systems). The sera were isolated from the erythrocytes through centrifugation at 3200 rpm at 5°C for 20 minutes. Then, the supernatant was aspirated into 500 µL Eppendorf tubes. These sample tubes were stored at -80°C preceding the sera analysis.

Separation Conditions

The CGE conditions for all separations were fixed with a wavelength of 214 nm and a stable system temperature of 25.0°C. A capillary length of 31.2 cm was used, with a separation length of 21.0 cm, and an internal diameter of 50.0 µm. All samples were eluted through the capillary to the anode.

Preparation of Molecular Weight Ladder

A commercial molecular weight (MW) ladder standard was run prior to the separation of the nineteen serum samples in order to calibrate the instrument and ensure validity. The standard is comprised of seven proteins of known molecular weights: 10, 20, 35, 50, 100, 150, and 225 kDa. The peaks of the standard produced in the electropherogram were verified using scientific publications. The nineteen clinical samples and the MW ladder standard were prepared in the

Biosafety Level-2 Laboratory (BSL-2) and analyzed using the following CGE method to produce electropherograms.

The MW ladder run was prepared by adding 10 μL of the commercial MW ladder standard to 40 μL of Sample Buffer. Following the addition of 2 μL of β -mercaptoethanol, the mixture was heated for five minutes in a 95°C water bath. After heating, the sample was allowed to de-gas in a water bath sonicator for five minutes. The standard was then placed in the sample tray and four rinses of the capillary were performed. The capillary was successively washed with 0.1N NaOH for three minutes at 70.0 psi, 0.1N HCl for one minute at 70.0 psi, distilled water for one minute at 70.0 psi, and lastly sodium dodecyl sulfate (SDS) gel rinse for ten minutes at 70.0 psi.

After the capillary rinse was completed, a voltage of 5.0 kV was applied for a duration of 20.0 seconds during the sample injection in order to reverse the polarity, then a voltage of 15.0 kV was applied for 45.00 minutes separated the proteins of the MW ladder. Finally, an auto-zero function was performed five minutes into the separation. The separation data was recorded throughout the run using the 32 Karat™ Software, generating electropherograms comprised of separated protein peaks.

Human Sera Preparation

For the preparation of the serum samples, 25 μL of the patient serum was added to 25 μL of Sample Buffer. The method used for the MW ladder separation was implemented for the analysis of individual serum samples, excluding the addition of β -mercaptoethanol and exposure to a hot water bath. These exclusions were made to preserve integrity of the serum proteins to prevent the coagulation of the albumin and other proteins when exposed to high temperatures.

Analysis of Serum

After acquiring the serum electropherograms with the 32 Karat™ Software, a double-peak alignment of internal markers was performed for each serum sample using the software's 'Stretch' operation. This operation was employed to negate fluctuation in the actual migration times of each separation which is a caveat to using gel systems. A consistent, characteristic divet with an actual migration time of approximately 8.5 minutes was aligned at the 10-minute mark and the consistently-predominating peak in the serum profiles was aligned at the 30-minute mark in each profile. A set of parameters, called 'integration events' (in the software), was utilized to obtain the integrated area of each individual protein peak. The integration events consist of setting the width to 0.2 and the threshold to 75 for the duration of the separation and turning the integration off for the first eight minutes of the separation. The areas of these peaks were manually inputted based on their relative retention times into an Excel file used as a database. Peaks were recorded in six-second intervals (every tenth of a minute), and were then assigned peak numbers based on their relation to the peaks of other serum profiles.

A calibration curve was generated in the software by using the 'Qualitative Analysis' function based on the profile of the MW ladder. The migration times of each of the MW ladder proteins of known molecular weights were used to generate the curve. Analysis of the serum protein peaks against the curve provided an estimated molecular weight of the unknown serum proteins.

Statistical Analysis of Serum

Seven of the most distinct and prevalent peaks were chosen for statistical analysis based on migration time uniformity and a low signal to noise ratio. The data from these seven peaks was used to compare the CAD cohort profiles against the Non-CAD cohort profiles using OriginPro

8.6 Data Analysis and Graphing Software. A Canonical Discriminant Analysis (CDA) was performed on the seven peaks of the nineteen samples, which treated each peak as an independent variable and generated Canonical Coefficients, or coefficients used in a linear combination that, when applied to each sample's peak data, produced a Canonical Score for that sample. The Canonical Score ultimately classified each sample as CAD or Non-CAD, based on its calculated value. Furthermore, a Linear Discriminant Analysis (LDA) was performed to determine the relative significance of each peak toward classifying samples as CAD or Non-CAD. The nineteen participants were used as training data, and their classification by the CDA algorithm was compared to the actual CAD vs. Non-CAD statuses of the participants. A statistical analysis was performed on the results of the Canonical Scores in order to determine statistical significance. Lastly, a cross-validation analysis was performed to extrapolate our findings to a larger sample size.

CHAPTER III

RESULTS

The MW ladder protein standard is displayed in **Figure 1**. The ladder is composed of seven proteins of known molecular weights revealed by the seven predominate peaks of the electropherogram. The ladder demonstrates differences in elution time of the proteins due to their varying sizes as smaller, lighter proteins travel more quickly than heavier ones. **Figure 2** shows the calibration curve of the MW ladder standard. A cubic fit type was used for the qualitative analysis and goodness of fit was calculated to be 0.999916. Using this calibration curve, the molecular weights of unknown peaks from other samples can be estimated. The migration times of the sera protein peaks were cross-referenced to this curve to approximate the sera protein sizes. **Figure 3** shows a complete human serum electropherogram. The full 45-minute separation of the human proteomic display can be seen. The predominate peak in the serum profile is albumin—identified by both its predominating concentration in serum and based on where it falls on the calibration curve with regard to migration time. Albumin, which has a molecular weight of 66.5 kDa, came in at approximately 29 minutes. (8) **Figure 4** shows the human serum sample separations of the seven Non-CAD participants. As reported in the Methods section, the profiles were aligned. The electropherograms were scaled to show the region between the 10-minute and 36-minute marks as this region holds the majority of the protein peaks. **Figure 5** and **Figure 6** display the human serum sample separations of the twelve CAD participants, which were aligned and scaled in the same way as the electropherograms in **Figure 4**.

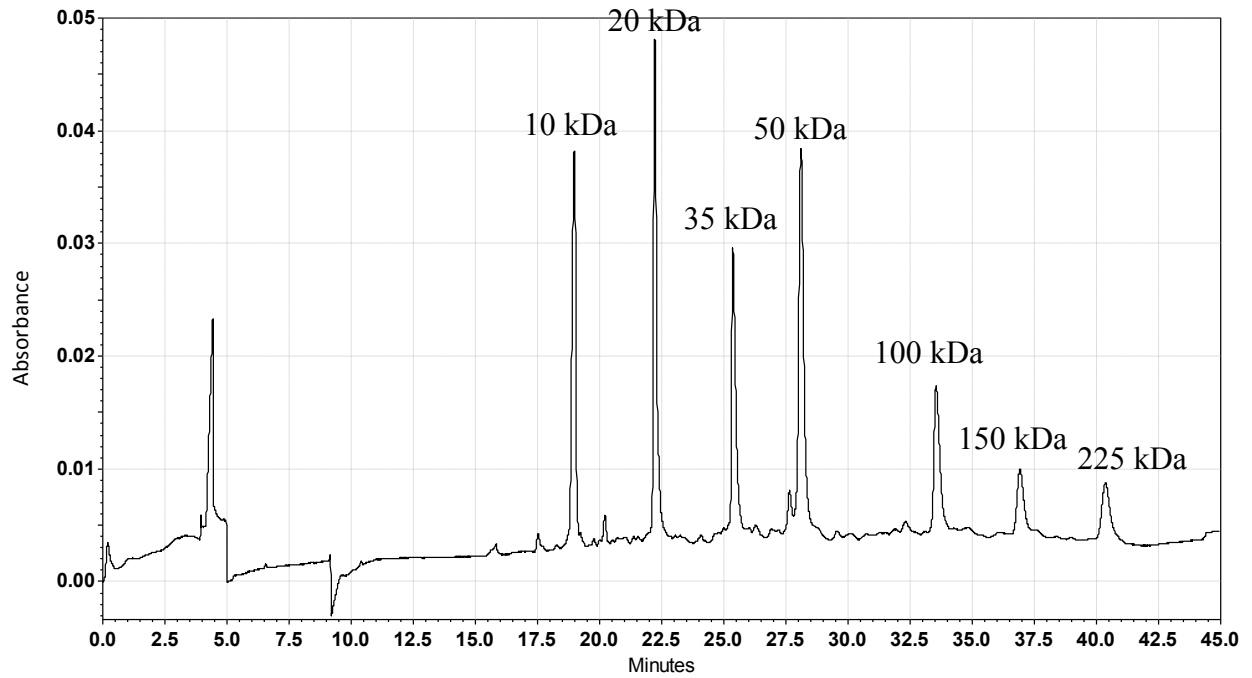


Figure 1 Molecular weight ladder electropherogram displays seven proteins with known molecular weights: 10, 20, 35, 50, 100, 150, and 225 kDa.

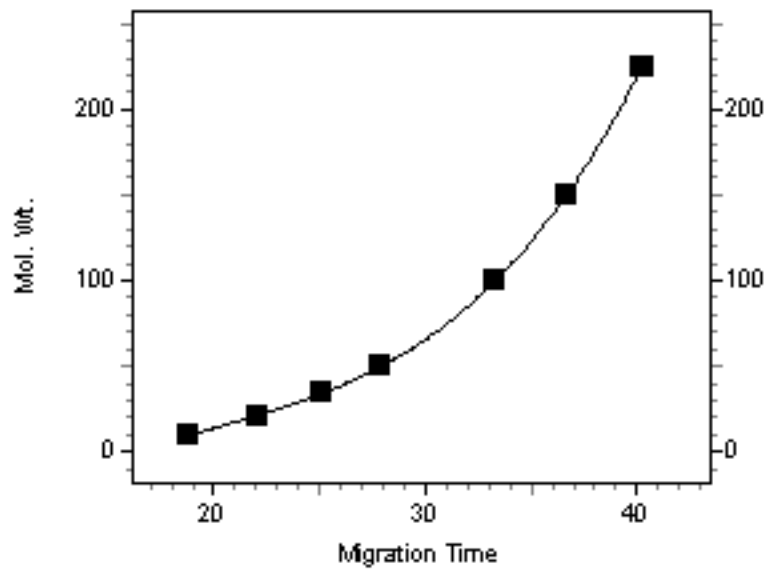


Figure 2 The calibration curve generated from the molecular weight ladder standard.

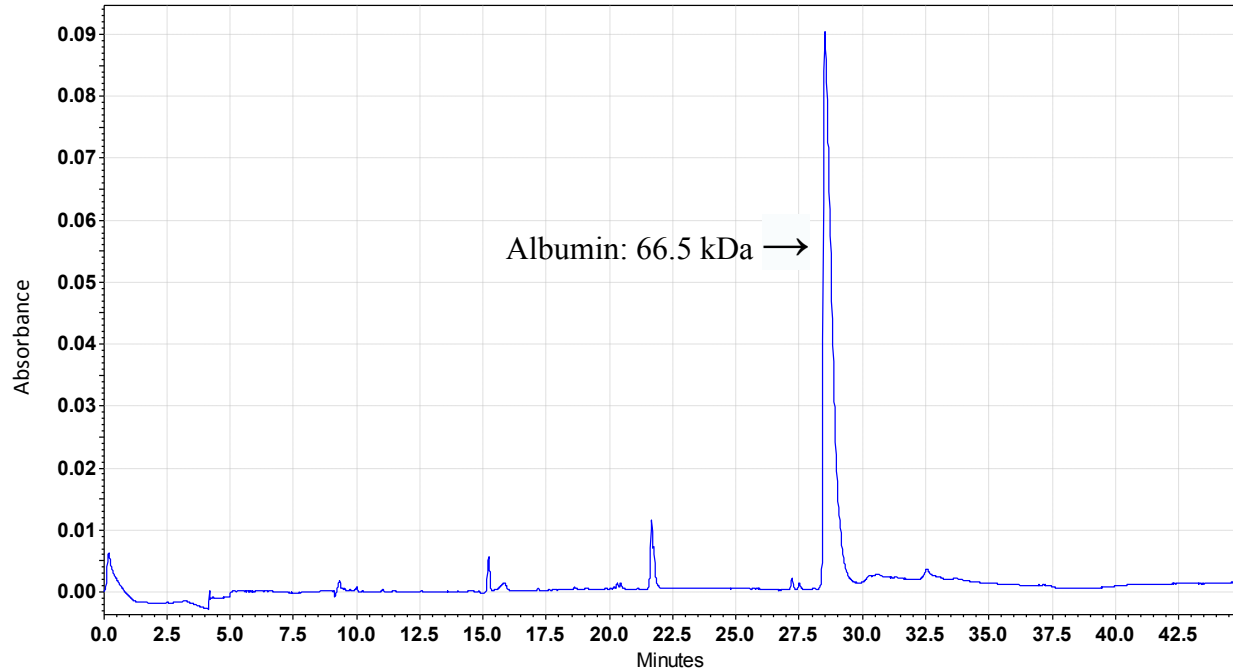


Figure 3 Human serum electropherogram of the full 45-minute separation

Figures 4-6 enumerate the seven peaks of interest that were used to distinguish the CAD from the Non-CAD cohorts. These integrated peak areas were used for our statistical analysis to discriminate between the two cohorts. Peak numbers are not listed in numerical order because more than the seven peaks involved in the analysis were identified and assigned numbers. Two additional electropherograms have been added in the appendix showing the peaks numbered 1-21. The stacking of the electropherograms was performed to depict the differences in the profiles from one participant to another.

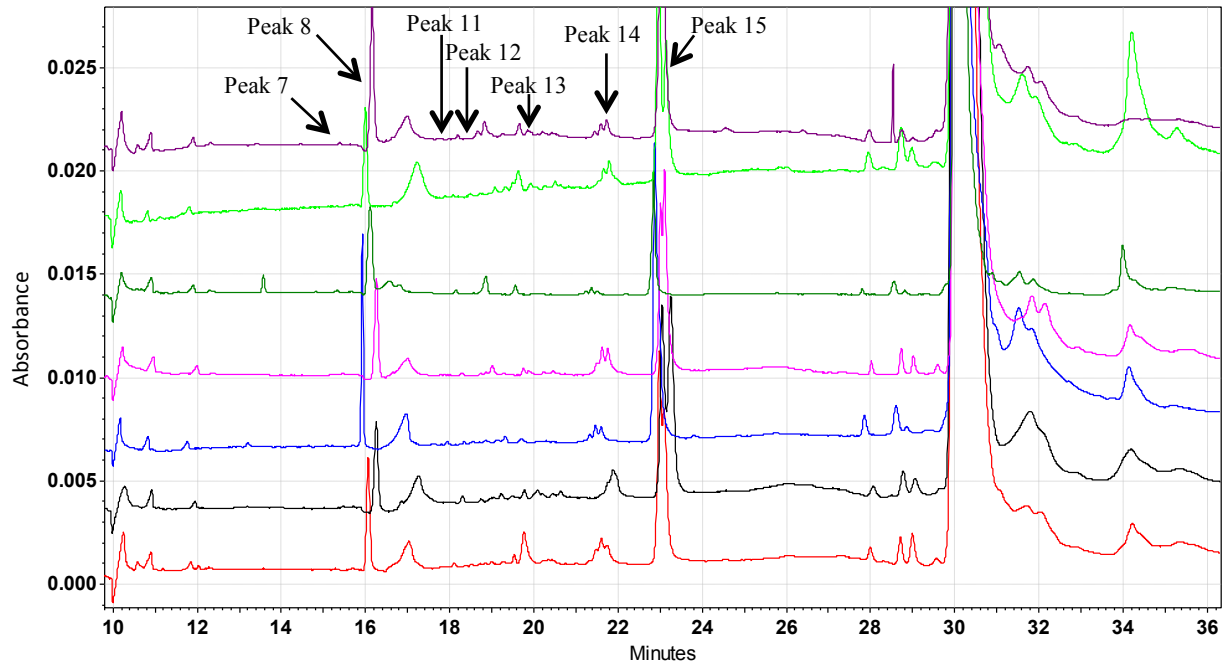


Figure 4 Serum Electropherograms for Non-CAD participants 1-7 with seven highlighted peaks of interest

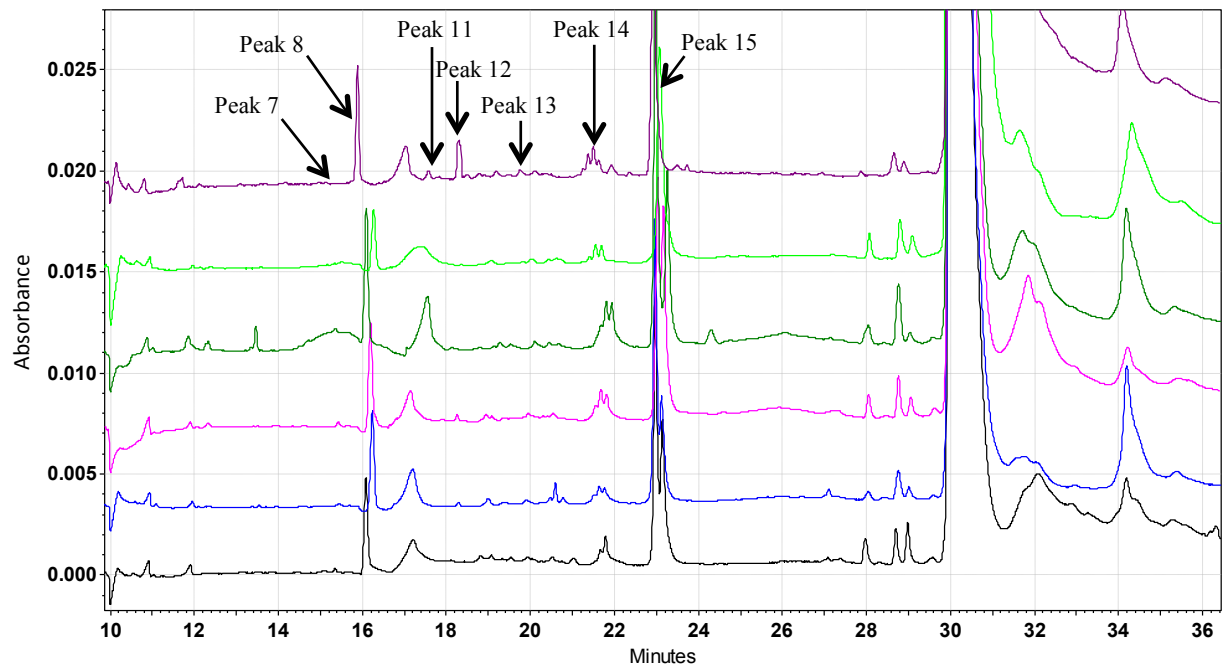


Figure 5 Serum Electropherograms for CAD Participants 1-6 with seven highlighted peaks of interest

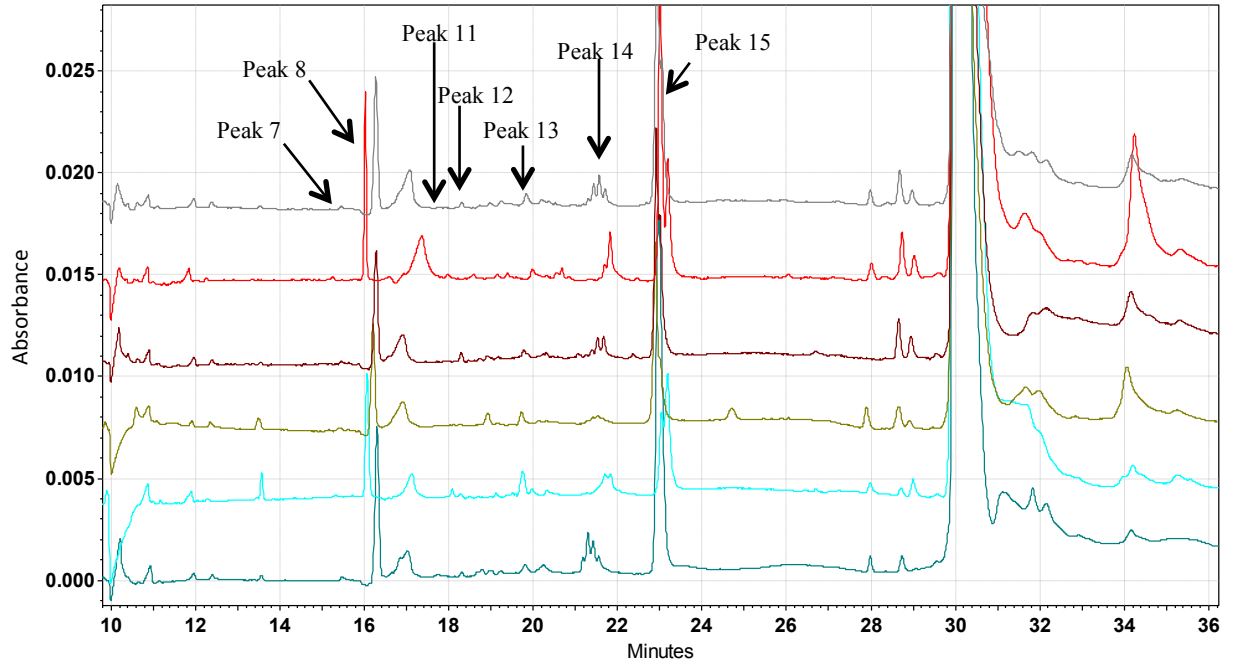


Figure 6 Serum Electropherograms for CAD Participants 7-12 with seven highlighted peaks of interest

Figure 7 displays the Linear Discriminant Analysis coefficients for the CAD and Non-CAD cohorts. The dominance of the coefficient of peak #7 indicates that this peak is relatively influential on the deviation of CAD sample electropherogram peak areas from the Non-CAD sample peak areas. Alternately, peak #8 is less influential, as indicated by its small value for both the CAD and Non-CAD LDA function.

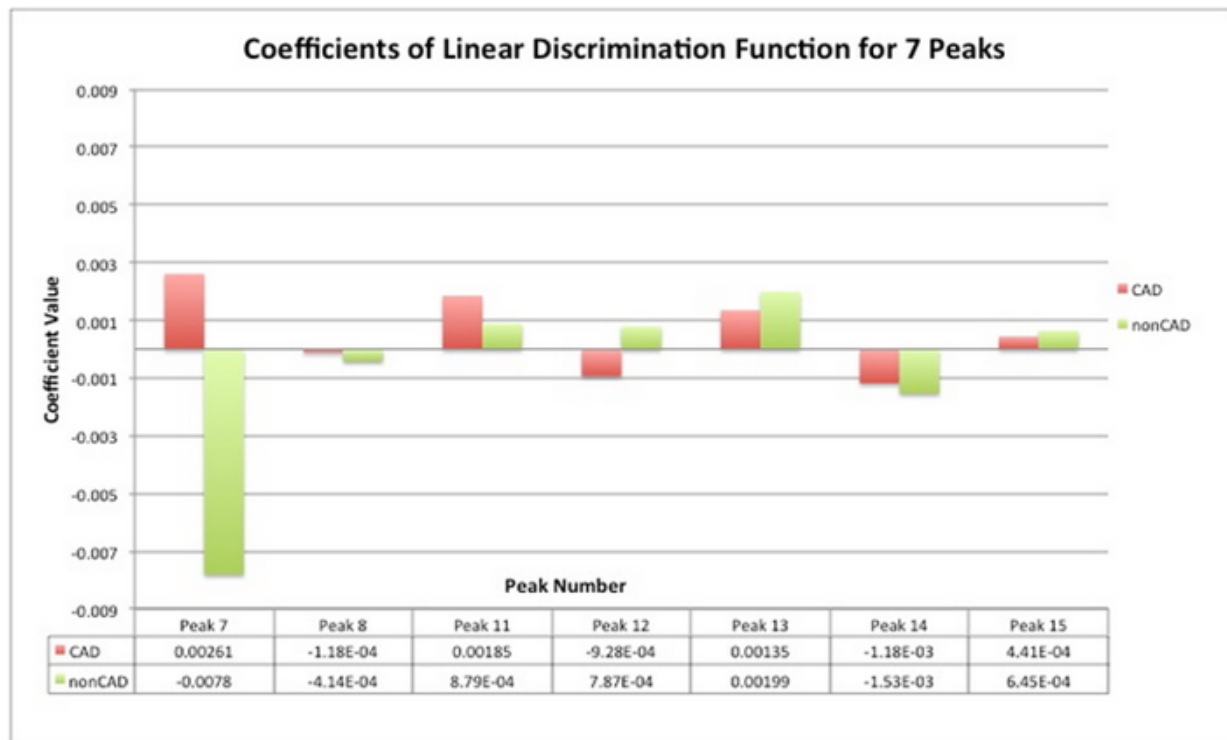


Figure 7 The LDA coefficients show each peak's relative contribution to the discrimination between nineteen CAD and Non-CAD participants

In the following figure, the Canonical Scores for the nineteen samples are displayed in **Figure 7**. Coronary artery disease classification is expressed as a value of -1 on the y -axis, while Non-CAD classification is expressed as a value of +1. The equation generated by the Canonical Discriminant Analysis (CDA) was applied to the peak data for each sample, as shown in **Equation 1**. A negative Canonical Score classified the sample as CAD, while a positive Canonical Score classified the sample as Non-CAD. A Canonical Score deviating farther away from 0, or the y -axis, shows a stronger classification and indicates that the particular sample exhibited typical CAD or Non-CAD indicators, while a Canonical Score close to 0 indicates a weaker classification. As shown in **Figure 8**, samples were classified as CAD or Non-CAD with 100% accuracy by application of the CDA function. Furthermore, a cross-validation was performed using the training data as a substitute for the test data. An error rate of 12.5% resulted

from the cross-validation measure. It was discovered that the cross-validation has high specificity for Non-CAD participants and low sensitivity for CAD participants. This was determined by the 100% accurate classification of the Non-CAD participants and a 75% accurate classification of the CAD participants.

Equation 1 The Canonical Discriminant Analysis function (shown below), was applied to the peak data of the nineteen samples and generated a Canonical Score, 'C', which classified each sample as CAD or Non-CAD.

$$C = -3.52074 + (-0.0027)(Peak7) + (-7.6923 \times 10^{-5})(Peak8) + (-2.58107 \times 10^{-4})(Peak11) + (4.45374 \times 10^{-4})(Peak12) + (1.67349 \times 10^{-4})(Peak13) + (-9.01718 \times 10^{-5})(Peak14) + (5.31353 \times 10^{-5})(Peak15)$$

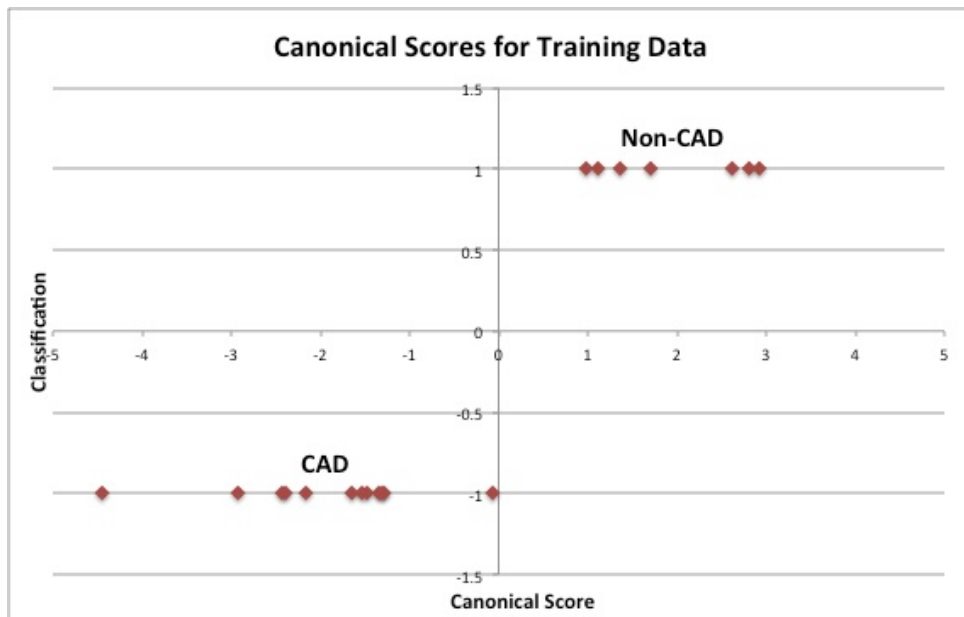


Figure 8. A score was generated for each of the nineteen-study participant by a canonical analysis of training data. A classification accuracy of 100% was obtained.

Table 1. A p-value of 0.00332 was recorded, indicating a significant difference between the CAD and Non-CAD participants' peak distributions.

Classification Summary for Training Data			
	CAD	Non-CAD	TOTAL
CAD	12 100.00%	0 0.00%	12 100.00%
NON-CAD	0 0.00%	7 100.00%	7 100.00%

p-value: 0.00332

Finally, the variation in peak areas was analyzed for statistical significance. Our null hypothesis, H_0 , stated that no difference exists between CAD and Non-CAD serum electropherogram peak area distributions. Our alternate hypothesis, H_A , stated that a difference does exist between CAD and Non-CAD serum electropherogram peak area distributions. A Wilks-Lambda test was performed on the peak means and a p value of 0.00332 was obtained, allowing us to reject the null hypothesis. Thus, our results indicate a difference between the peak area distributions of CAD and Non-CAD serum sample electropherograms.

CHAPTER IV

CONCLUSION

In conclusion, the generated algorithms allow us to quantitatively compare CAD serum samples to Non-CAD serum samples. The resulting data from this comparison shows promise for CGE as a new technique for diagnosing CAD, as results were statistically significant. Thus, analyzing human sera with this method provides a simple, effective approach for distinguishing between CAD and Non-CAD individuals.

Linear discriminant analysis of the electropherogram data allowed for the identification of several discrepancies between the CAD and Non-CAD cohorts. These differences in protein chemistry, upon further analysis, may prove to be associated and indicative of CAD development. Each of these peaks was compared via their assigned algorithm coefficients, as shown in **Figure 7**. The magnitude of each coefficient describes the weight each peak holds in the CAD versus Non-CAD designation relative to the respective peak integration values. Peak #7 in particular is shown to be critical in the Non-CAD designation, with a coefficient four times larger than the other Non-CAD coefficients. Similarly, peak #7 plays the most significant role in the CAD designation, with a coefficient approximately two times larger than the other CAD coefficients. Identification of the protein responsible for peak #7 may reveal a critical biomarker for CAD and provide insight into the physiological changes caused by protein chemistry alterations. Peaks #11, #12, #13 and #14 carry a significant amount of weight in both the CAD and Non-CAD designations, while peaks #8 and #15 appear to be less significant. As each of the peaks is identified, exploration into the effects of other variables on protein chemistry involved

in CAD and Non-CAD classification will be pursued, e.g., diet, physical activity, and therapies. Further examination of the relationship between protein biomarker patterns and the onset of CAD using our novel method may lead to the development of a minimally invasive CAD screening tool requiring only one drop of human serum.

The future goal is to obtain a larger study population to validate the finding of the significant protein peaks in the serum samples. If the future findings are congruent with these results, the identification of each statistically significant peak will be pursued using protein standards. Furthermore, analysis of the protein profiles will be completed to calculate statistical coefficients of the protein peaks that were not examined in this study. This future examination may lead to the discovery of the other biomarkers for CAD.

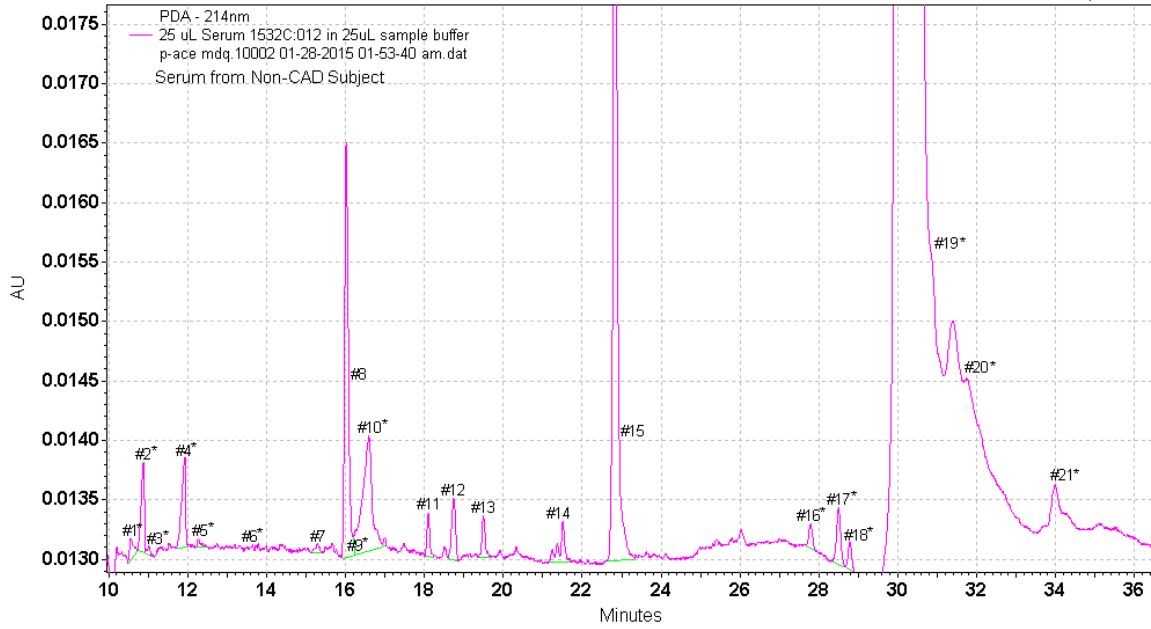
REFERENCES

1. "Heart Disease Facts." *Centers for Disease Control and Prevention*. 29 Oct. 2014. Web. 02 Dec. 2014.
2. "Can Heart Disease Be Reversed?" *Track Your Plaque*. Heart Scan Resource Center, 2006. Web. 23 Feb. 2015.
3. Park, Alice, and Alice Park. "Rethinking the Framingham Score: Is There a Better Way to Predict Heart Disease? | TIME.com." *Time*. 14 Sept. 2011. Web. 03 Dec. 2014.
4. Seiko Otokozawaa, Masumi Aia, Bela F. Asztalosa, Charles C. Whiteb, Serkalem Demissie-Banjawb, L. Adrienne Cupplesb, Katsuyuki Nakajimaa, Peter W.F. Wilsonc, Ernst J. Schaefer. "Direct assessment of plasma low density lipoprotein and high density lipoprotein cholesterol levels and coronary heart disease: Results from the Framingham Offspring Study." *Atherosclerosis*. 251-255. Vol.213.
5. Mozaffaria, Dariush, & Wu, Jason H. Y. "Omega-3 Fatty Acids and Cardiovascular Disease : Effects on Risk Factors, Molecular Pathways, and Clinical Events." *Journal of the American College of Cardiology* 58(20):2047–2067.
6. Thomas M. Stulnig, M.D., Enrico Ammirati, M.D., et al. "C-Reactive Protein, Fibrinogen, and Cardiovascular Risk." *New England Journal of Medicine* 368.1 (2013): 84-86. Print.
7. Atkinson AJ, Colburn WA, DeGruttola VG, DeMets DL, Downing GJ, Hoth DF, et al. "Biomarkers and surrogate endpoints: preferred definitions and conceptual framework." *Clin Pharmacol Ther*. 2001;69:89 –95.
8. Masuelli, Martin A. "Study of Bovine Serum Albumin Solubility in Aqueous Solutions by Intrinsic Viscosity Measurements." *Hindawi*. Hindawi Publishing Corporation, 9 May 2013. Web. 09 Apr. 2015.

APPENDIX

Human Serum Protein Profile Laboratory for Cardiovascular Chemistry Texas A&M University, College Station

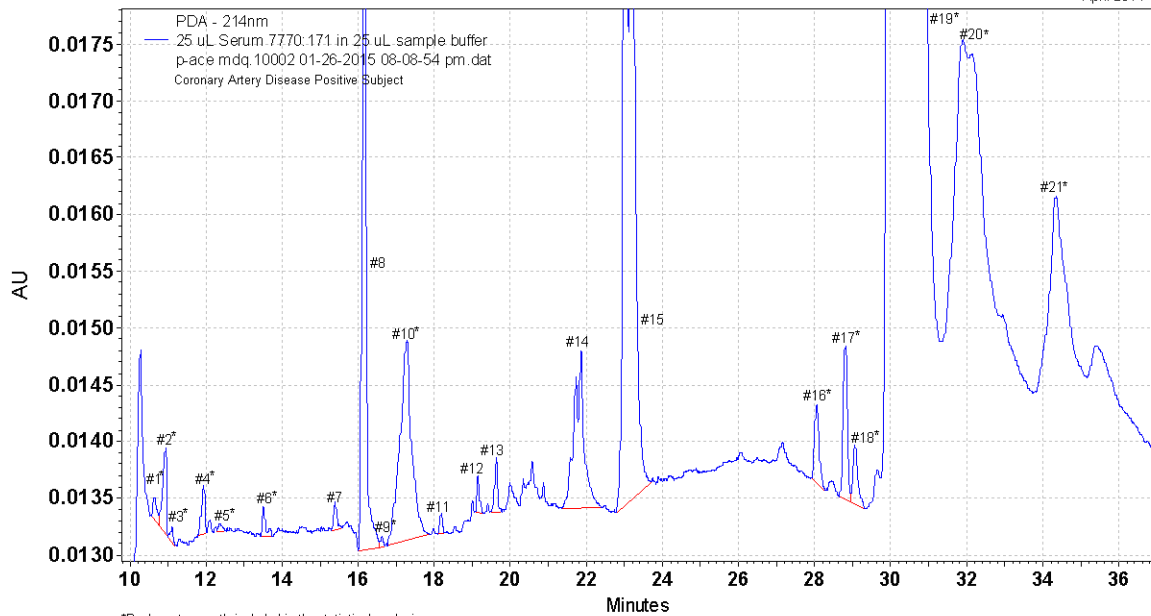
April 2014



*Peaks not currently included in the statistical analysis

Human Serum Protein Profile Laboratory for Cardiovascular Chemistry Texas A&M University, College Station

April 2014



*Peaks not currently included in the statistical analysis