2001

Methods development for the identification of unknown proteins

Laura Brumfield Neal

Methods Development for the Identification of Unknown Proteins

A Thesis Presented to the Department of Chemistry
at Marshall University in Partial Fulfillment of the Requirements for the
Degree of Master of Science

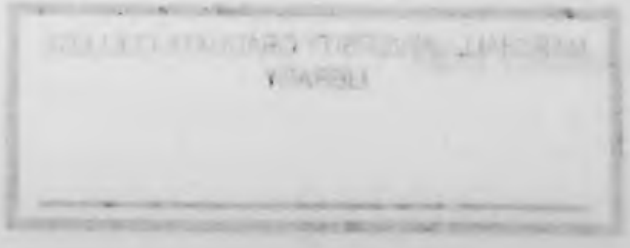
By

Laura Brumfield Neal

Marshall University

Huntington, West Virginia

July 24, 2001



MARSHALL UNIVERSITY

DEPARTMENT OF CHEMISTRY

THIS THESIS WAS ACCEPTED ON August 8, 2001

AS MEETING THE RESEARCH REQUIREMENTS FOR THE DEGREE

OF MASTER OF SCIENCE.

Dr. William Price

Advisor, Department of Chemistry

Dr. Leslie Frost

Member, Department of Chemistry

Dr. Michael L. Norton

Member, Department of Chemistry

Dr. Leonard J. Deutsch

Dean, Graduate School

To My Husband and Parents

ACKNOWLEDGMENTS

I would like to thank Dr. William Price, my research advisor, and Dr. Leslie Frost for their guidance and extreme patience through the course of my thesis research. I would also like to thank Dr. Gary Wright and Dr. William McCumbee at the Marshall University School of Medicine for their collaboration on this project. I thank my parents for their encouragement and support for which I will always be grateful and indebted. Also, very special thanks goes out to my husband, who has been my life support in my pursuit of a Master of Science degree.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS AND ACRONYMS

AcCN	Acetonitrile
AcOH	Acetic Acid
Apo A-I	Apolipoprotein A-I
BME	β-Mercaptoethanol
BSA	Bovine Serum Albumin
CBB	Coomassie Brilliant Blue
dH_2O	Deionized Water
DNF	Dinitrofluorobenzene
DTT	Dithiothreitol
ESI-MS	Electrospray Ionization Mass Spectrometry
HCCA	α-Cyano-4-hydroxycinnamic Acid
HDL	High-Density Lipoprotein
HPLC	High Performance Liquid Chromatography
IEF	Isoelectric Focusing
LDL	Low-Density Lipoprotein
m/z	Mass to Charge Ratio
MALDI-TOF	Matrix-Assisted Laser Desorption Ionization Time of Flight
Mb	Myoglobin
MeOH	Methanol
MS	Mass Spectrometry
MW	Molecular Weight

pI	Isoelectric Point
RP-HPLC	Reverse-Phase High Performance Liquid Chromatography
RSA	Rat Serum Albumin
SA	Sinapinic Acid
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
SEC	Size Exclusion Chromatography
TFA	Trifluoroacetic Acid

pI	Isoelectric Point
RP-HPLC	Reverse-Phase High Performance Liquid Chromatography
RSA	Rat Serum Albumin
SA	Sinapinic Acid
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
SEC	Size Exclusion Chromatography
TFA	Trifluoroacetic Acid

ABSTRACT

The chemistry of transport, storage, energy production, and reproduction in physiological systems is controlled by the actions and interactions of proteins. A species full complement of these amazing machines is often referred to as its proteome. The new science of proteomics utilizes a set of powerful and evolving tools to characterize the nature of the proteome, and further to fully elucidate the functional attributes of the proteins and their associations.

A methodology was developed by combining several proteomic techniques, such as electrophoresis, RP-HPLC, and mass spectrometry, to correctly identify proteins. Three known proteins (myoglobin, albumin, and actin) were employed for development of these techniques to achieve optimal results. The meticulously refined methodology was then applied to an unknown protein in Zucker serum to obtain information about the protein's activity, structure, and possible post-translational modifications.

The protein was identified as apolipoprotein A-I and studied in Zucker serum to analyze its implications for hereditary obesity. Apo A-I was found to be much more abundant in the obese serum than in the lean. Apo A-I is normally found in the phospholipid layer of serum, but this particular protein was found in the aqueous layer of Zucker serum. These results indicate a difference in the protein. Delipidated Zucker serum analyzed by MALDI-TOF revealed several modifications to the protein, each ~200 Da apart. Two-dimensional electrophoresis of the delipidated serum suggested a multitude of differences in protein levels between the lean and obese. This research has only scratched the surface of understanding the role apo A-I plays in hereditary obesity.

CHAPTER 1

DEVELOPMENT OF METHODS FOR IDENTIFICATION OF PROTEINS

INTRODUCTION

Recently, the study of genomics has drastically increased with the introduction of the Human Genome Project (HGP). The era of complete genome sequences has arrived and with it vast amounts of data that must be interpreted, cross referenced, and incorporated into the regulatory networks, which define the physiology of an organism. Gene function and interaction are investigated by simultaneous measurement of the expression levels of all genes in a cell and the mapping and quantitation of all proteins expressed within a cell. Genomic sequences when complemented with proteomic analysis heralds a new era of holistic cellular biology.[1]

Proteomic techniques offer an understanding of the function of genomes, as it directly analyzes the end product of the genome, an organism's complete set of genes and chromosomes. In 1995 Wasinger et al. introduced the term 'proteome' and defined it as the protein complement of the genome. The genome can be thought of as relatively static in that information cannot be directly obtained regarding when the gene is transcribed and translated, at what rate, what its end products are, or what the nature of its function is. The proteome, however, is highly dynamic. The protein expression and modification will vary with respect to environment, physiological state, stress, and other stimuli.[2] The proteome can thus provide information on (i) if and when predicted gene products are translated, (ii) the relative concentrations of gene products, and (iii) the extent of post-translational modification, none of which can be predicted from the nucleic acid sequence

alone.[3] Because proteomics provides this additional information, it may be the preferred tool for providing an understanding of how genomes function.

Proteome analysis is based on four technological principles: (i) a fast and simple procedure for purifying small amounts of proteins from complex mixtures, (ii) a fast and sensitive method to generate limited, but sufficient, structural information from the proteins of interest, (iii) access to extended protein or DNA sequence databases, and (iv) computer algorithms capable of translating and linking the DNA sequence language with various types of protein structural information, such as N-terminal protein or internal peptide sequences, amino acid compositions, pI, peptide mass fingerprints, MS fragmentation patterns, or sequence tags of selected peptides.[4] Two specific methods are used in order to investigate these platforms: the top-down approach and the bottom-up approach. The former technique identifies proteins completely by mass spectrometry. First, an accurate MW of the protein must be obtained using ESI-MS, which initiates the breakdown of the protein into a manifold of highly charged states. These highly charged ions are then subject to fragmentation by electron capture dissociation (ECD). If needed, further fragmentation can be obtained by collision activated dissociation (CAD). These sequence ions can then be "stitched together" for identification of the protein. Alternatively, the bottom-up approach begins by first cleaving the protein using enzymatic digests. The digest delivers peptides, which can be used for identification purposes (peptide map). MALDI-TOF analysis of the digest generates a peak list, which is then subjected to a protein database search. This search results in a list of proposed proteins.[5] Each of these methods provides a detailed analysis of proteins.

Many techniques have been employed to investigate the characteristics of proteins. These techniques can be used separately or in combination with each other to obtain vast amounts of information about the protein's activity, structure, and possible post-translational modifications. A number of these techniques are outlined below.

Size exclusion chromatography (SEC) is a technique based upon the use of a porous gel in the form of insoluble beads placed into a column. As a solution of proteins is passed through the column, small proteins can penetrate into the pores of the beads and, therefore, are impeded in their rate of travel through the column. The larger the protein, the less likely it will enter the pores. Different beads with different pore sizes can be used depending upon the desired protein size separation profile.

Another chromatographic technique is ion exchange chromatography. Each individual protein exhibits a distinct overall net charge at a given pH. Some proteins will be negatively charged and some will be positively charged at the same pH. This property of proteins is the basis for ion exchange chromatography. Fine cellulose resins are used that are either negatively (cation exchanger) or positively (anion exchanger) charged. Proteins of opposite charge to the resin are retained as a solution of proteins is passed through the column. The bound proteins are then eluted by passing a solution of ions bearing a charge opposite to that of the column. By utilizing a gradient of increasing ionic strength, proteins with increasing affinity for the resin are progressively eluted.

Proteins also have high affinities for their substrates, co-factors, prosthetic groups, receptors, or antibodies raised against them. This affinity can be exploited in the purification of proteins, namely affinity chromatography. A column of beads bearing the high affinity compound can be prepared and a solution of protein passed through the

column. The bound proteins are then eluted by passing a solution of soluble high affinity compounds through the column.[6]

In column chromatography, the smaller and more tightly packed a resin is the greater the separation capability of the column. The limitation of gravity flow columns is the time it takes to pass the solution of proteins through the column. HPLC utilizes tightly packed fine diameter resins to impart increased resolution and overcomes the flow limitations by pumping the solution of proteins through the column under high pressure. Like standard column chromatography, HPLC columns can be used for size exclusion or charge separation. An additional separation technique commonly used with HPLC utilizes hydrophobic resins to slow the movement of nonpolar proteins. The proteins are then eluted from the column with a gradient of increasing concentration of an organic solvent. This latter form is termed reverse-phase HPLC.[7]

Proteins can also be characterized according to size and charge by separation in an electric current (electrophoresis) within solid sieving gels made from polymerized and cross-linked acrylamide. The most commonly used technique is termed SDS-PAGE. The gel is a thin slab of acrylamide polymerized between two glass plates. This technique utilizes a negatively charged detergent (SDS) to denature and solubilize proteins. SDS denatured proteins have a uniform negative charge such that all proteins will migrate through the gel in the electric field based solely upon size. The larger the protein the more slowly it will move through the matrix of the polyacrylamide. Following electrophoresis, the migration distance of unknown proteins relative to known standard proteins is assessed by various staining or radiographic detection techniques.

The use of polyacrylamide gel electrophoresis can also be used to determine the isoelectric charge of proteins (pI). This technique is termed isoelectric focusing. Isoelectric focusing utilizes a thin tube of polyacrylamide made in the presence of a mixture of small positively and negatively charged molecules termed ampholytes. The ampholytes have a range of pIs that establish a pH gradient along the gel when current is applied. Proteins will, therefore, cease migration in the gel when they reach the point where the ampholytes have established a pH equal to the protein's pI.

Prior to sequencing peptides, it is necessary to eliminate disulfide bonds within peptides and between peptides. Several different chemical reactions can be used in order to permit separation of peptide strands and prevent protein conformations that are dependent upon disulfide bonds. The most common treatments are to use either 2-BME or DTT. Both of these chemicals reduce disulfide bonds. To prevent reformation of the disulfide bonds, the peptides are treated with iodoacetamide to alkylate the free sulfhydryls.[2]

There are several techniques used for C-terminal sequence determination. The most reliable chemical technique for C-terminal residue identification is hydrazinolysis. A peptide is treated with hydrazine at high temperature for an extended length of time. This treatment cleaves all of the peptide bonds yielding amino-acyl hydrazides of all the amino acids excluding the C-terminal residue, which can be identified chromatographically, compared to amino acid standards[8] (see Appendix III for a list of amino acids).

There are three major chemical techniques for sequencing peptides and proteins from the N-terminus. These are the Sanger, dansyl chloride, and Edman techniques.

Sanger's reagent utilizes the compound, 2,4-DNF, which reacts with the N-terminal residue under alkaline conditions. The derivatized amino acid can be hydrolyzed and will be labeled with a dinitrobenzene group that imparts a yellow color to the amino acid. Separation of the modified amino acids by electrophoresis and comparison with the migration of standards allows for the identification of the N-terminal amino acid. Like DNF, dansyl chloride reacts with the N-terminal residue under alkaline conditions. Analysis of the modified amino acids is carried out similarly to the Sanger method except that the dansylated amino acids are detected by fluorescence.[9] The utility of Edman degradation allows for additional amino acid sequence to be obtained from the N-terminus inward. Using this method, it is possible to acquire the entire sequence of peptides. This method utilizes phenylisothiocyanate to react with the N-terminal residue under alkaline conditions. The resultant derivatized amino acid is hydrolyzed in anhydrous acid. The hydrolysis reaction results in a rearrangement of the released N-terminal residue to a phenylthiohydantoin derivative. As in the previous methods, the N-terminal residue is tagged with an identifiable marker but now the remainder of the peptide is intact. The entire sequence of reactions can be repeated over and over to obtain the sequences of the peptide. Edman sequencing does, however, have several limitations. There is an upper mass limit due to peptide bond cleavage arising chiefly from N to O acyl shifts. Other limitations include contaminants and blocked amino termini.[2]

Due to the limitations of the Edman degradation technique, peptides longer than ~50 residues cannot be sequenced completely. The ability to obtain peptides of this length from proteins of greater length is facilitated by the use of enzymes, such as

endoproteases, which cleave at specific sites within the primary sequence of proteins (Table 1-1). The resultant smaller peptides can be chromatographically separated and subjected to Edman degradation sequencing reactions.[2]

ENZYME	SOURCE	SPECIFICITY	ADDITIONAL POINTS
Trypsin	Bovine pancreas	Peptide bond C-terminal to R,K but not if next to P	Highly specific for positively charged residues
Chymotrypsin	Bovine pancreas	Peptide bond C-terminal to F, Y, W but not if next to P	Prefers bulky hydrophobic residues, cleaves slowly at N, H, M, L
Elastase	Bovine pancreas	Peptide bond N-terminal to A, G, S, V, but not if next to P	
Thermolysin	Bacillus thermoproteolyticus	Peptide bond N-terminal to I, M, F, W, Y, V, but not if next to P	Prefers small neutral residues, can cleave at A, D, H, T
Pepsin	Bovine gastric mucosa	Peptide bond N-terminal to L, F, W, Y, but not when next to P	Exhibits little specificity, requires low pH
Lys-C	*Lysobacter enzymogenes*	Peptide bond C-terminal to K	

Table 1-1. Specificities of certain endoproteases that can be used to cleave proteins/peptides into smaller pieces.

No reliable chemical techniques exist for sequencing the C-terminal amino acid of peptides. However, there are enzymes, exopeptidases, which cleave peptides at the C-terminal residue that can then be analyzed chromatographically and compared to standard amino acids. This class of exopeptidases is called carboxypeptidases[2] (Table 1-2).

Enzyme	Source	Specificity
Carboxypeptidase A	Bovine pancreas	Will not cleave when C-terminal residue is R, K, or P or if P resides next to terminal residue
Carboxypeptidase B	Bovine pancreas	Cleaves when C-terminal residue is R or K, not when P resides next to terminal residue
Carboxypeptidase C	Citrus leaves	All free C-terminal residues, pH optimum 3.5
Carboxypeptidase Y	Yeast	All free C-terminal residues, slowly at G residues

Table 1-2. Specificities of certain exoproteases that can be used to cleave peptides at the C-terminal end.

Mass spectrometry is now an indispensable tool for rapid protein and peptide structural analysis. All mass spectrometers have three essential components: ion source, mass analyzer, and detector. Gas phase ions produced in the ion source are separated in the mass analyzer by their m/z ratio and then detected. MS data are recorded as "spectra" which display ion intensity versus the m/z value. The two techniques that have become preferred methods for ionization of peptides and proteins are ESI and MALDI due to their effective application on a wide range of proteins and peptides. Although proteins can often be directly identified by their unique molecular weight, it is often best to break them down by the use of enzymes prior to mass spectrometric investigations. A mixture of peptides that are characteristic of the precursor protein are obtained. Mass spectrometry also allows the detection of mutations within proteins, whether they are natural or are derived through mutagenesis. The strategy that is used for the natural mutants consists of comparing the molecular weights of a species of peptides obtained through enzymatic cleavage with those obtained for the native protein. A change in a peptide molecular weight indicates the position of the mutation, while the difference between the native peptide molecular weight and that of the mutant peptide allows a possible determination of the nature of the amino acid that has mutated.[10]

For the purpose of this thesis research, only the most useful of the above techniques was employed to identify proteins. To begin, a variety of known proteins were subjected to electrophoresis, enzymatic digests, mass spectrometry, and other techniques to characterize these proteins. A protocol was then designed to achieve identification of these proteins by proteomic techniques. This methodology was then applied to an unknown protein to determine the structure, characterization, and any modifications of this protein.

MATERIALS

Chemicals

Acros Organics, Fairlawn, NJ

iodoacetamide

Aldrich Chemical Co., Inc., Milwaukee, WI

acetonitrile (HPLC grade)

Bio Rad Laboratories, Hercules, CA

2-β-mercaptoethanol

dithiothreitol

Laemmili sample buffer

10x Tris/Glycine/SDS running buffer

Hewlett Packard, Palo Alto, CA

α-cyano-4-hyroxycinnamic acid matrix

peptide standard

protein standard

sinapinic acid matrix

ICN Biomedicals, Inc., Aurora, OH

actin (rabbit smooth muscle)

myoglobin (horse skeletal muscle)

Promega, Madison, WI

trypsin (sequencing grade)

Sigma-Aldrich, Inc., St. Louis, MO

albumin (bovine serum)

Coomassie Brilliant Blue

silver stain kit

Instrumentation/Accessories

Bio Rad Laboratories, Hercules, CA

Power Pac Junior

Ready Gels (10 well, 15 µL)

Ready Gel Cell

Fisher Scientific, Suwanee, GA

dry bath incubator

Lab-Line Instruments, Boston, MA

shaking bath

Millipore Corporation, Bedford, MA

C_{18} ZipTip

Savant Instruments, Inc., Holbrook, NY

SpeedVac SC110

METHODS

Molecular Weight Determination of Proteins in Solution

Five hundred μL of three protein samples (Mb, BSA, and actin) was prepared at concentrations of 10 pmol/μL each. The molecular weights of these proteins are shown in Table 1-3.

Proteins	Molecular Weight (Da)
Horse Heart Myoglobin	16950
Rabbit Actin, Smooth Muscle	42300
Bovine Serum Albumin	66430

Table 1-3. Molecular weights of protein samples used for experimentation.

The solvent used was 70% AcCN. The sample (0.5 μL) was mixed with an equal volume of the appropriate matrix solution (sinapinic acid). One μL of each was spotted onto a stainless steel MALDI target and allowed to air-dry, forming a co-crystalline sample/matrix complex. Proteins were analyzed in the positive mode. This approach favored the production of protonated $(M+H)^+$ ions, although in appropriate buffer conditions sodium $(M+Na)^+$, potassium $(M+K)^+$, or ammonium $(M+NH_4)^+$ adducts were also formed.

Samples were analyzed on a Bruker Bi-Flex III MALDI-TOF instrument (described in Appendix I), using a N_2 laser at 337 nm to desorb the solute molecules from the sample plate. Following insertion of the probe and evacuation of the system to $<1x10^{-7}$ mbar, a voltage of 20 kV was established in the source region. The proteins were analyzed in the linear mode. The coarse laser energy was set to 40% with fine

adjustment being varied for each sample. Once a good quality spectrum was obtained, 100 laser shots were accumulated to produce a m/z spectrum of 5000-150000 Da. The spectrum was first calibrated externally using an appropriate standard (Table 1-4) under the same instrumental conditions.

Proteins	Molecular Weight (Da)
Ubiquitin	8566
Horse Heart Myoglobin	16950
β-lactoglobulin	18400
Bovine Serum Albumin	66430

Table 1-4. Protein standards used to calibrate the Bruker Bi-Flex III MALDI-TOF to ensure optimal performance.

1-D Electrophoresis

The SDS-reducing sample buffer was prepared by adding 100 μL 2-BME to 900 μL of Laemmili sample buffer (Table 1-5).

	Description (1x buffer concentration)
	62.5 mM Tris-HCl
Laemmili	2% SDS
Buffer	25% Glycerol
	0.01% Bromophenol Blue

Table 1-5. Composition of sample buffer used for 1-D electrophoresis.

Solutions containing three different proteins (Mb, BSA, and actin) were used for 1-D electrophoresis. Six concentrations were used: 1 nM, 10 nM, 100 nM, 1 μM, 10 μM, and 100 μM. Ten μL of sample were diluted with an equal volume of SDS-reducing sample buffer. The samples were then heated at 95°C for 3 minutes.[12] The 1-D electrophoresis cell (Figure 1-1) was assembled.

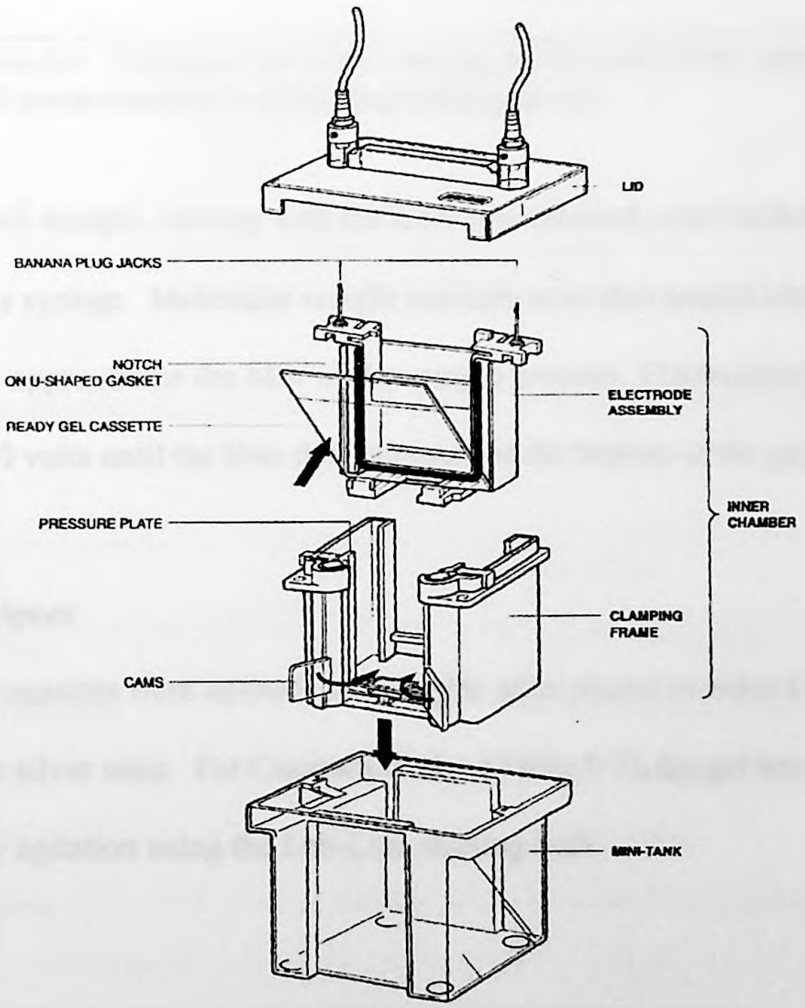


Figure 1-1. Diagram of the Ready Gel Cell and its components.[13]

The running buffer was prepared by diluting 100 µL of Tris/Glycine/SDS buffer with 900 µL dH_2O (Table 1-6). The upper (inner chamber) and lower reservoirs were filled with running buffer.

	Description (1x Buffer Concentration)
Tris/Glycine/SDS	25 mM Tris 192 mM Glycine 0.1% SDS

Table 1-6. Components of the running buffer used in the upper and lower reservoirs of the electrophoresis cell.

Fifteen µL of each sample, starting with the least concentrated, were loaded into the wells of the gel using a syringe. Molecular weight markers were also loaded into adjacent wells in order to approximate the MW of the sample proteins. Electrophoresis was performed at 100 volts until the blue dye had reached the bottom of the gel.[13]

Staining Techniques

The gel cassettes were opened, and the gels were placed in either Coomassie Brilliant Blue or silver stain. For Coomassie Blue (Table 1-7), the gel was stained for 1 hour with gentle agitation using the Lab-Line shaking bath.

Component	Volume
Coomassie Blue R-250	0.1% (w/v)
MeOH	45.5% (v/v)
AcOH	9.0% (v/v)

Table 1-7. Components of Coomassie Brilliant Blue used for staining polyacrylamide gels.

The gels stained with Coomassie Blue were then destained (Table 1-8) for 3 hours changing the solution frequently (~6 changes).

Solution	Volume
MeOH	40%
AcOH	10%

Table 1-8. Components of the destain solution, which removes Coomassie background from gels.

For silver staining, the gels were first placed in fixing solution (Table 1-9).

Solutions	Components
Fixing	30% EtOH, 10% AcOH
Silver Equilibration	1.5 mL Silver Concentrate, 300 mL H_2O
Development	30 mL Developer 1, 0.17 Developer 2, 300 mL H_2O
Stop	1% AcOH
Reducer	2 mL Reducer A, 4 mL Reducer B, 0.7 mL Reducer C, 300 mL H_2O

Table 1-9. Solutions needed for silver staining (Sigma-Aldrich) of polyacrylamide gels.

The gels were fixed 3 times, changing the solution every 20 minutes. The gels were then rinsed 3 times with dH_2O for 10 minutes each rinse. The gels were placed in silver equilibration solution and equilibrated with gentle agitation for 30 minutes. The gels were then rinsed for 10 seconds with dH_2O. Developer solution was placed over the gels and remained until achieving dark bands. The staining was then terminated using stop solution. The gels were again rinsed 3 times with dH_2O. Tap H_2O was run over the gel for one minute followed by 3 rinses with dH_2O.[14]

The gels were stored at room temperature in dH_2O in a container with a lid until needed.

Digestion of Proteins

The protein bands of interest were excised and digested using trypsin. A protocol for in-gel tryptic digestion of proteins that is rapid and generally applicable is described in Table 1-10.

Step	Method
1	Excise protein spot/band and dehydrate in AcCN for 10 minutes. Remove AcCN and SpeedVac until dry.
2	Reswell gel pieces at 37°C overnight in buffer containing 1 μM trypsin and 50 mM NH_4HCO_3 (pH ~8) at a 1:100 ratio. Solutions should just cover gel pieces.
3	Centrifuge gel pieces and collect supernatant. Further extract peptides by one change of 20 mM NH_4HCO_3 (centrifuge then collect), and 3 changes of 5% formic acid in 50% AcCN (20 minutes between changes) at room temperature.
4	Dry sample down in SpeedVac until desired volume has been reached.

Table 1-10. Protocol for an in-gel digestion method that is rapid and widely applicable.[2]

A similar method (Table 1-11), which is longer and involves reduction of disulfide bonds with DTT and alkylation of cysteine residues with iodoacetamide, is used when a higher sequence coverage of the protein is needed.

Step	Method
1	Excise protein spot/band and dehydrate in AcCN for 10 minutes. Remove AcCN and SpeedVac until dry.
2	Cover gel pieces with 10 mM DTT solution in 100mM NH_4HCO_3. Reduce proteins for 1 hour at 56°C.
3	Cool to room temperature. Remove DTT solution and add equal volume of 55 mM iodoacetamide in 100 mM NH_4HCO_3. Incubate for 45 minutes in dark at room temperature.
4	Wash gel pieces with 100 mM NH_4HCO_3 for 10 minutes. Dehydrate with AcCN, reswell in 100 mM NH_4HCO_3, and shrink again with AcCN.
5	Remove liquid phase and SpeedVac.
6	Reswell gel pieces at 37°C overnight in buffer containing 1 μM trypsin and 50 mM NH_4HCO_3 (pH ~8) at a 1:100 ratio. Solutions should just cover gel pieces.
7	Centrifuge gel pieces and collect supernatant. Further extract peptides by one change of 20 mM NH_4HCO_3 (centrifuge then collect) and 3 changes of 5% formic acid in 50% AcCN (20 minutes between changes) at room temperature.
8	Dry sample down in SpeedVac until desired volume had been reached.

Table 1-11. Longer protocol for an in-gel digestion method that results in higher sequence coverage of the protein.[2]

The long digest method was employed for cleavage of the proteins into peptide fragments.[2]

Sample Purification for Mass Spectrometry

Sample purification of the digested solutions was achieved using ZipTips, which are intended for concentrating, desalting, and removing detergents from the samples. The solutions needed for ZipTips are outlined in Table 1-12.

Solution	$ZipTip_{C18}$
Wetting	50% AcCN
Equilibration	0.1% TFA
Wash	0.1% TFA
Elution	50% AcCN in 0.1% TFA

Table 1-12. Aqueous solutions employed when utilizing ZipTips for sample purification.

The tip was prewet by aspirating 10 μL of wetting solution into the tip. The solution was dispensed to waste, and this step was repeated. The tip was equilibrated for binding by washing it twice with the equilibration solution. The peptides were bound to the ZipTip by aspirating and dispensing the sample for 10 cycles. The tip was washed and then dispensed using 3 cycles of wash solution. Five μL of elution solution was aspirated and dispensed into a clean Ependorf tube.[15]

MALDI-TOF Mass Spectrometry

The purified samples (0.5 μL each) were mixed with an equal volume of α-cyano-4-hyroxycinnamic acid (HCCA) matrix. One μL was spotted onto a MALDI target and allowed to air dry. The samples were analyzed using MALDI-TOF. The peptides were investigated in the reflectron mode (Appendix I). The coarse laser energy was set to 20% with fine adjustment varying for each sample.[11] Once a quality spectrum was obtained, 100 laser shots were accumulated to produce a m/z spectrum of 500-4000

Da. The spectrum was first calibrated externally using an appropriate standard analyzed under the same instrumental conditions (Table 1-13).

Peptides	Molecular Weights (Da)
Arginine-8-Vasopressin	1084.25
Angiotensin I	1281.49
Somatostatin	1637.90
Chicken Atrial Natriuretic Peptide	3160.66

Table 1-13. Peptide standards employed for calibration of the Bruker Bi-Flex III MALDI-TOF to ensure optimal results.

The resulting peptide peaks from the tryptic digest were then entered into MASCOT, a protein-searching database. This database compares the experimental tryptic peaks to the tryptic peaks of proteins in the database, resulting in a list of proposed proteins.

RESULTS AND DISCUSSION

Determination of Molecular Weights by MS

The MALDI-TOF mass spectrometer was first calibrated with a protein standard solution (Table 1-4) to ensure optimal quality and accuracy. A two-point calibration was performed using ubiquitin and BSA. Each of these peaks was calibrated to their exact mass. Figure 1-2 depicts the resulting calibration spectrum.

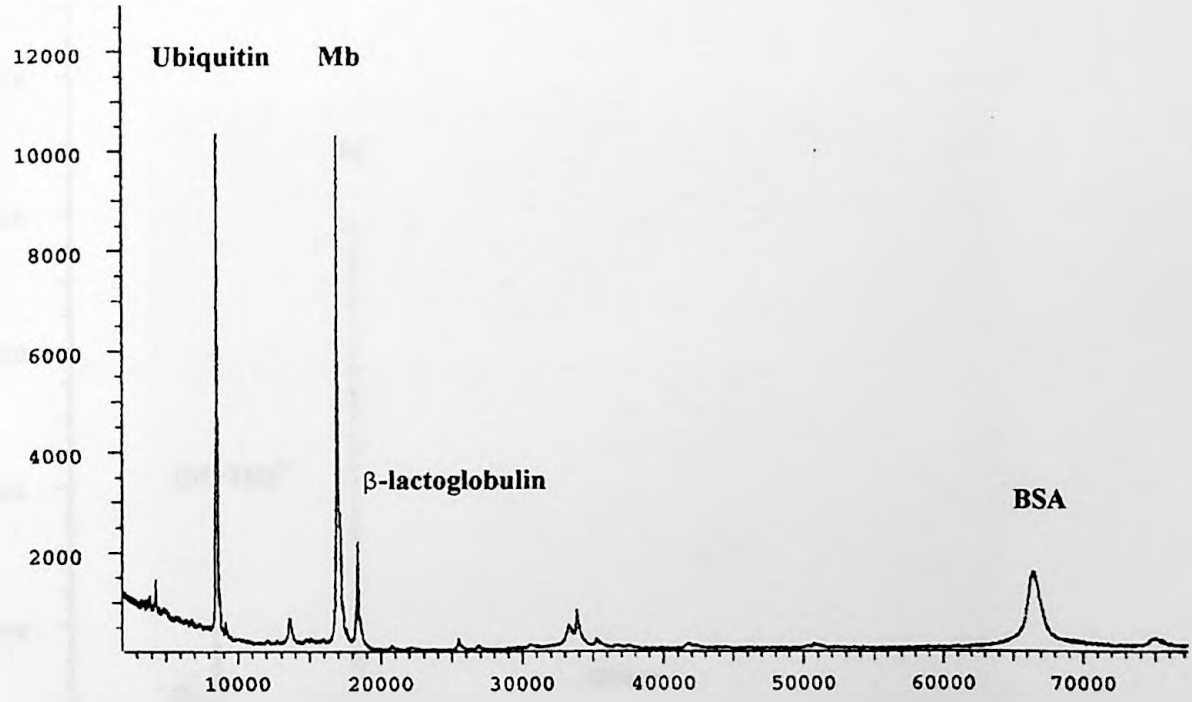


Figure 1-2. MALDI spectum of a protein standard used for calibration of the instrument.

The x-axis represents the m/z ratio of the protein, while the y-axis corresponds to the absolute intensity of that protein. The peak at m/z 66430 corresponded to the monoisotopic form of BSA. The $(M+2H)^{2+}$ ion of BSA was also observed at

~m/z 33220. The molecular ion of Mb appeared at m/z 16945. The peak at ~m/z 33900 was determined to be the dimer of Mb. Each of the other components of the protein standard was identified. The molecular weight of each protein was extremely accurate (0.03% error).

Once the instrument was calibrated, the samples were ready to be analyzed. Figure 1-3 is the resulting spectrum of the Mb sample.

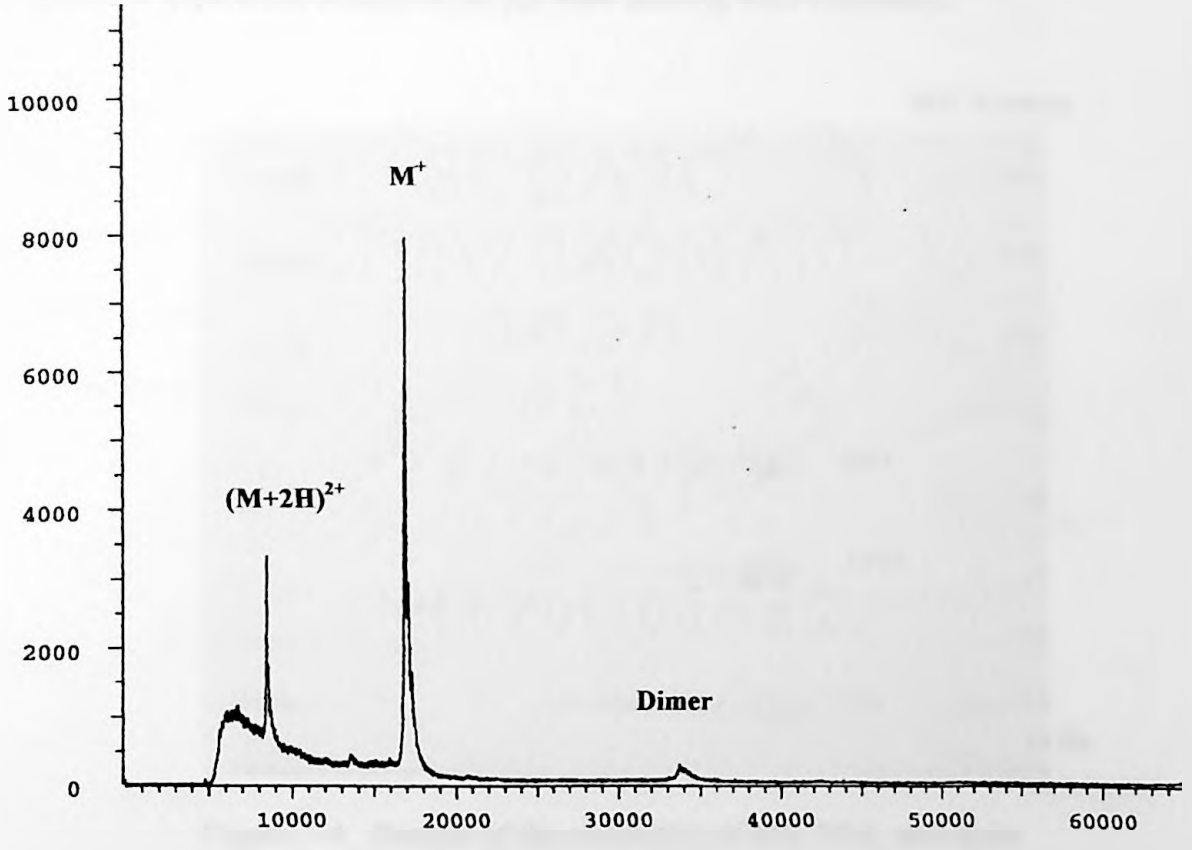


Figure 1-3. MALDI spectrum of Mb in solution.

The molecular ion of Mb was observed at m/z 16970. This was a 0.15% deviation from the theoretical mass of Mb (16950 Da). The $(M+2H)^{2+}$ ion (m/z 8450) and dimer

(m/z 3390) were also present. The spectra of BSA and actin yielded the same precise results, enabling MW determination.

Separation of Proteins by Electrophoresis

Electrophoresis provided a way to separate proteins according to their physical properties of size and charge. Staining methods allowed for the visualization of the proteins by penetrating into the gel and binding to the protein within the gel.

Each of the samples was visually analyzed after electrophoresis and staining. Figure 1-4 depicts the results of the gel after staining with Coomassie.

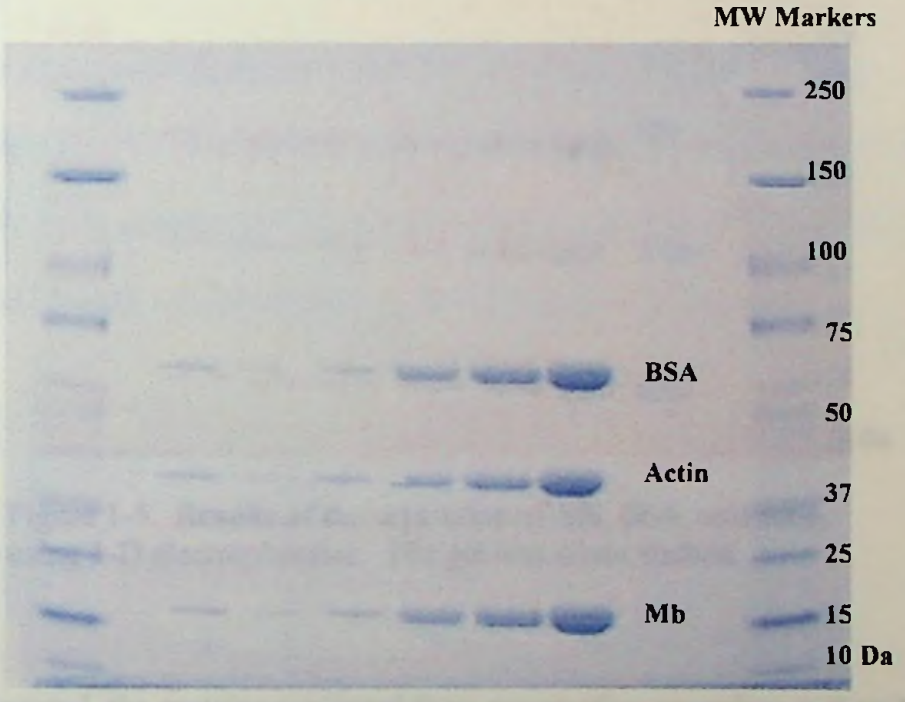


Figure 1-4. Results of the separation of Mb, BSA, and actin using 1-D electrophoresis. Gel was stained with Coomassie.

As the concentration of the samples increased (left to right), the intensity of the band increased. As the gel revealed, the proteins were successfully separated form each other. Because Mb had the lightest mass of the three proteins, it migrated farthest through the

gel. Comparing the Mb band to the MW markers, the mass of the protein was approximated at 17 kDa. BSA, on the other hand, had the heaviest mass, and as a result, did not travel far through the gel. Actin with its MW of ~44 kDa migrated farther than BSA but not as far as Mb. The same results were drawn from the silver stained gels (Figure 1-5).

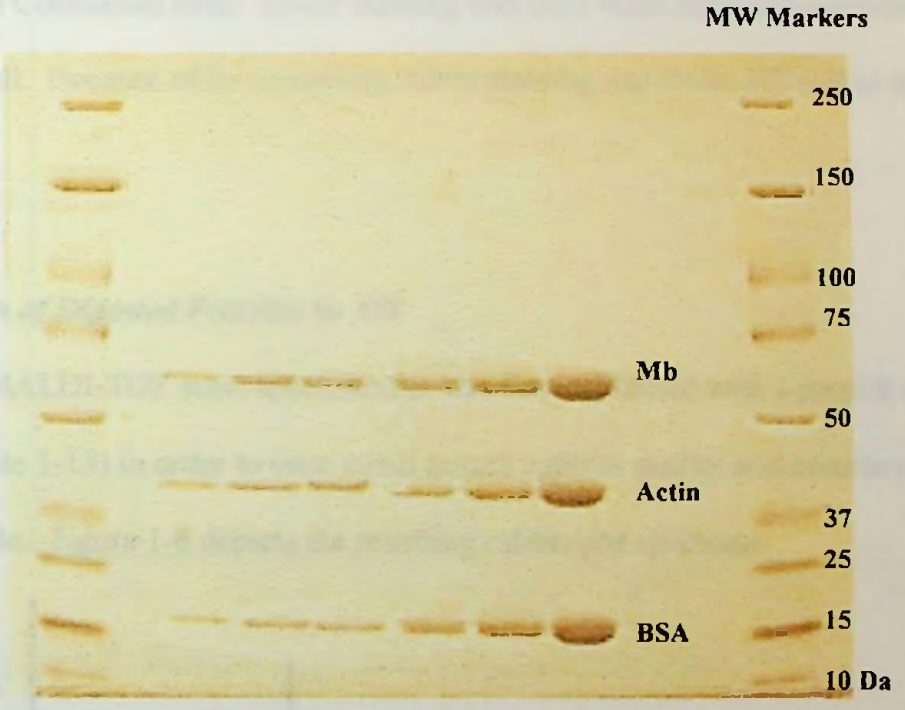


Figure 1-5. Results of the separation of Mb, BSA, and actin using 1-D electrophoresis. The gel was silver stained.

As expected, the proteins separated from one another according to their MW when the electric field was applied to the gel. Separation by mass was possible because the different polypeptides bound approximately the same amount of SDS, i.e., they had the same negative charge, but their radii were different and smaller proteins (Mb) migrated through the polyacrylamide faster than bigger ones (BSA). Troubleshooting 1-D electrophoresis is discussed in Appendix II.

Comparison of Staining Techniques

By comparing Coomassie Blue and silver stain results, the advantages and disadvantages of each become apparent. Coomassie Blue proved to be a short and simple procedure for staining gels. This stain required 1.0 μg/ band in order to visualize. Silver staining will detect as little as 0.01 μg/ band. Silver stains are one hundred times more sensitive than Coomassie Blue. Silver staining was used when the protein concentration was very small. Because of its sensitivity, silver staining was more difficult to use and optimize.

Identification of Digested Proteins by MS

The MALDI-TOF mass spectrometer was first calibrated with a peptide standard solution (Table 1-13) in order to once again ensure optimal quality and accuracy of the desired sample. Figure 1-6 depicts the resulting calibration spectrum.

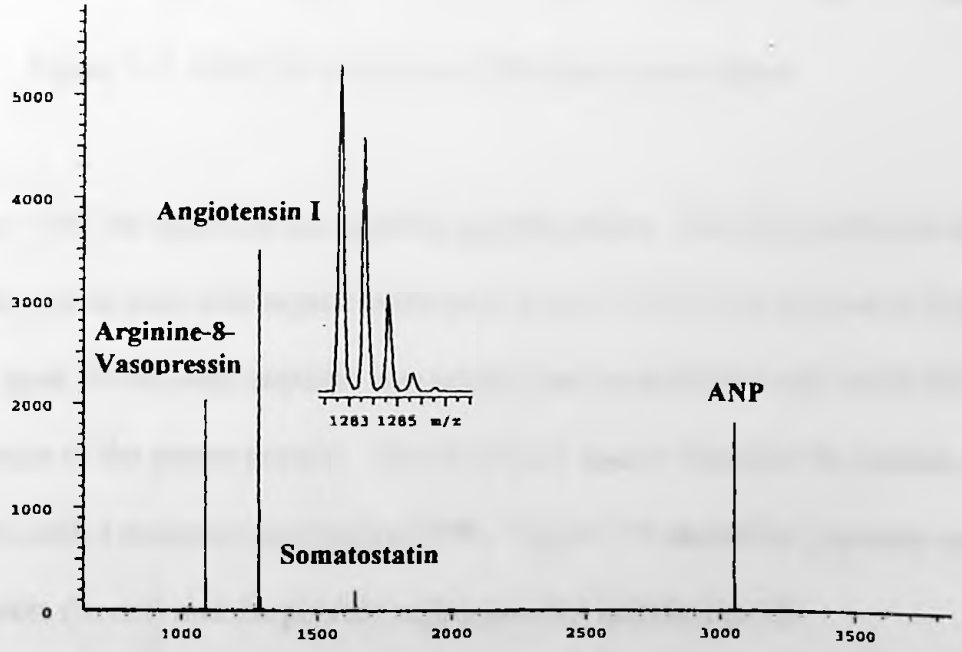


Figure 1-6. MALDI spectrum of the peptide standard used to calibrate the instrument.

Each protein excised from the gel was digested with trypsin, which cleaves the protein specifically after R and K residues (see Appendix III for a list of amino acids). The samples were spotted on the MALDI plate with HCCA matrix. Figure 1-7 is the MALDI spectrum of the Mb digest.

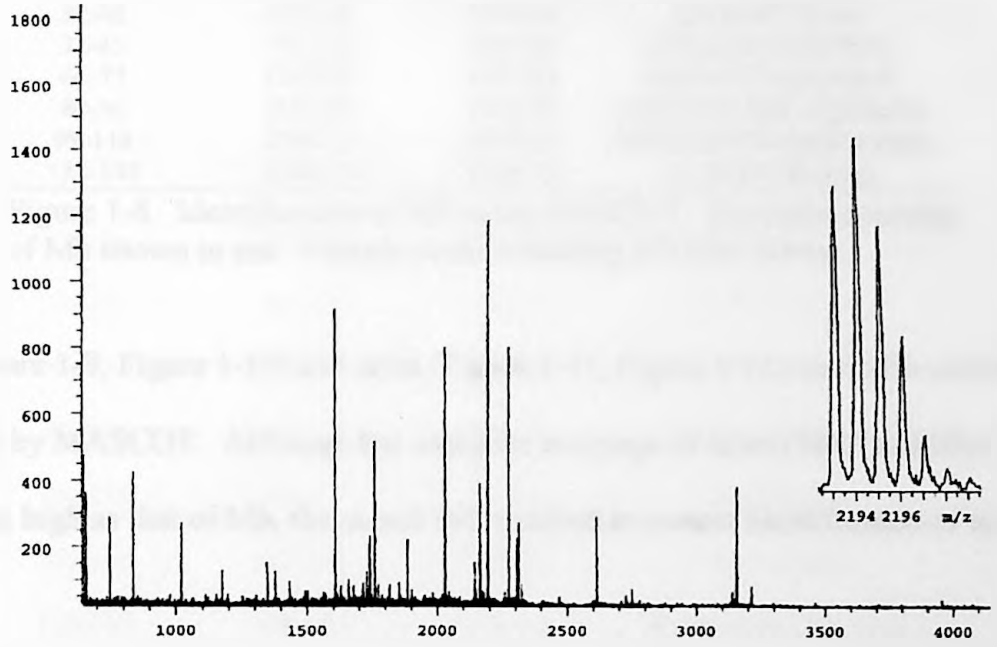


Figure 1-7. MALDI spectrum of Mb after tryptic digest.

The peaks >500 Da represent the resulting peptide pieces. The high resolution of the instrument can be seen with regard to the peak at m/z 2193.5 (see subview of Figure 1-7).

A peak list of these peptides was created and searched through MASCOT for identification of the parent protein. The MASCOT search classified the protein as horse myoglobin with a sequence coverage of 70%. Figure 1-8 shows the sequence coverage of the protein (in red) and the peptide sequences that matched to Mb.

1 **GLSDGEWQQV LNVWGKVEAD IAGHGQEVLI RLFTGHPETL EKFDK**FKHLK
51 TEAEMKASED LKK**HGTVVLT ALGGILKKKG HHEAELKPLA QSHATK**HK**IP**
101 **IKYLEFISDA IIHVLHSK**HP GNFGADAQGA MTK**ALELFRN DIAAK**YKELG
151 FQG

Start-End	Observed (Da)	Mr (Da)	Sequence
1-16	1815.87	1814.90	GLSDGEWQQVLNVWGK
17-31	1607.05	1605.85	VEADIAGHGQEVLIR
32-42	1271.95	1270.66	LFTGHPETLEK
32-45	1661.94	1660.85	LFTGHPETLEKFDK
64-77	1379.06	1377.83	HGTVVLTALGGILK
80-96	1853.92	1852.95	GHHEAELKPLAQSHATK
99-118	2336.33	2335.33	IPIKYLEFISDAIIHVLHSK
134-145	1360.75	1359.75	ALELFRNDIAAK

Figure 1-8. Identification of Mb using MASCOT. Sequence coverage of Mb shown in red. Peptide peaks matching Mb also shown.

BSA (Figure 1-9, Figure 1-10) and actin (Figure 1-11, Figure 1-12) were also correctly identified by MASCOT. Although the sequence coverage of actin (58%) and BSA (33%) was not as high as that of Mb, the search still resulted in correct identification of each protein.

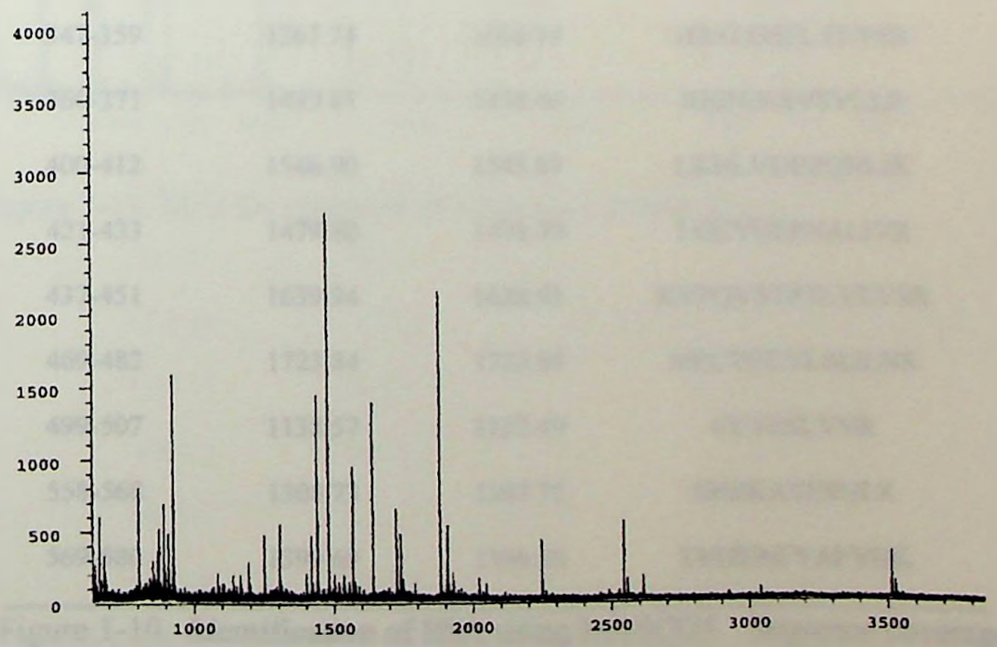


Figure 1-9. MALDI spectrum of BSA after tryptic digest.

1 **MKWVTFISLL LLFSSAYSR**G VFRRDTHKSE IAHR**FKDLGE EHFK**GLVLIA
51 FSQYLQQCPF DEHVK**LVNEL TEFAK**TCVAD ESHAGCEK**SL HTLFGDELCK**
101 VASLRETYGD MADCCEKQEP ERNECFLSHK **DDSPDLPK**LK PDPNTLCDEF
151 KADEKKFWGK **YLYEIARRHP YFYAPELLYY ANK**YNGVFQE CCQAEDKGAC
201 LLPK**IETMRE K**VLASSARQR LRCASIQKFG ERALKAWSVA RLSQKFPKAE
251 FVEVTKLVTD LTKVHKECCH GDLLECADDR ADLAKYICDN QDTISSKLKE
301 CCDKPLLEKS HCIAEVEKDA IPENLPPLTA DFAEDKDVCK NYQEAK**DAFL**
351 **GSFLYEYSRR HPEYAVSVLL R**LAKEYEATL EECCAKDDPH ACYSTVFDK**L**
401 **KHLVDEPQNL IK**QNCDQFEK **LGEYGFQNAL IVR**YTR**KVPQ VSTPTLVEVS**
451 **R**SLGKVGTRC CTKPESER**MP CTEDYLSLIL NR**LCVLHEKT PVSEKVTK**CC**
501 **TESLVNR**RPC FSALTPDETY VPKAFDEKLF TFHADICTLP DTEKQIKKQT
551 ALVELLK**HKP KATEEQLKTV MENFVAFVDK** CCAADDKEAC FAVEGPKLVV
601 STQTALA

Start-End	Observed (Da)	Mr (Da)	Sequence
1-19	2262.24	2261.23	MKWVTFISLLLLFSSAYSR
35-44	1249.62	1248.61	FKDLGEEHFK
66-75	1163.63	1162.62	LVNELTEFAK
89-100	1418.74	1418.69	SLHTLFGDELCK
131-138	886.42	885.41	DDSPDLPK
161-168	1083.60	1082.59	YLYEIARR
169-183	1889.00	1887.92	HYFYAPELLYYANK
205-211	906.47	905.46	IETMREK
347-359	1567.74	1566.74	DAFLGSFLYEYSR
360-371	1439.81	1438.80	RHPEYAVSVLLR
400-412	1546.90	1545.89	LKHLVDEPQNLIK
421-433	1479.80	1478.79	LGEYGFPNALIVR
437-451	1639.94	1638.93	KVPQVSTPTLVEVSR
469-482	1723.84	1723.83	MPCTEDYLSLILNR
499-507	1138.57	1137.49	CCTESLVNR
558-568	1308.73	1307.72	HKPKATEEQLK
569-580	1399.69	1398.69	TVMENFVAFVDK

Figure 1-10. Identification of BSA using MASCOT. Sequence coverage of BSA shown in red. Peptide peaks matching BSA are also shown.

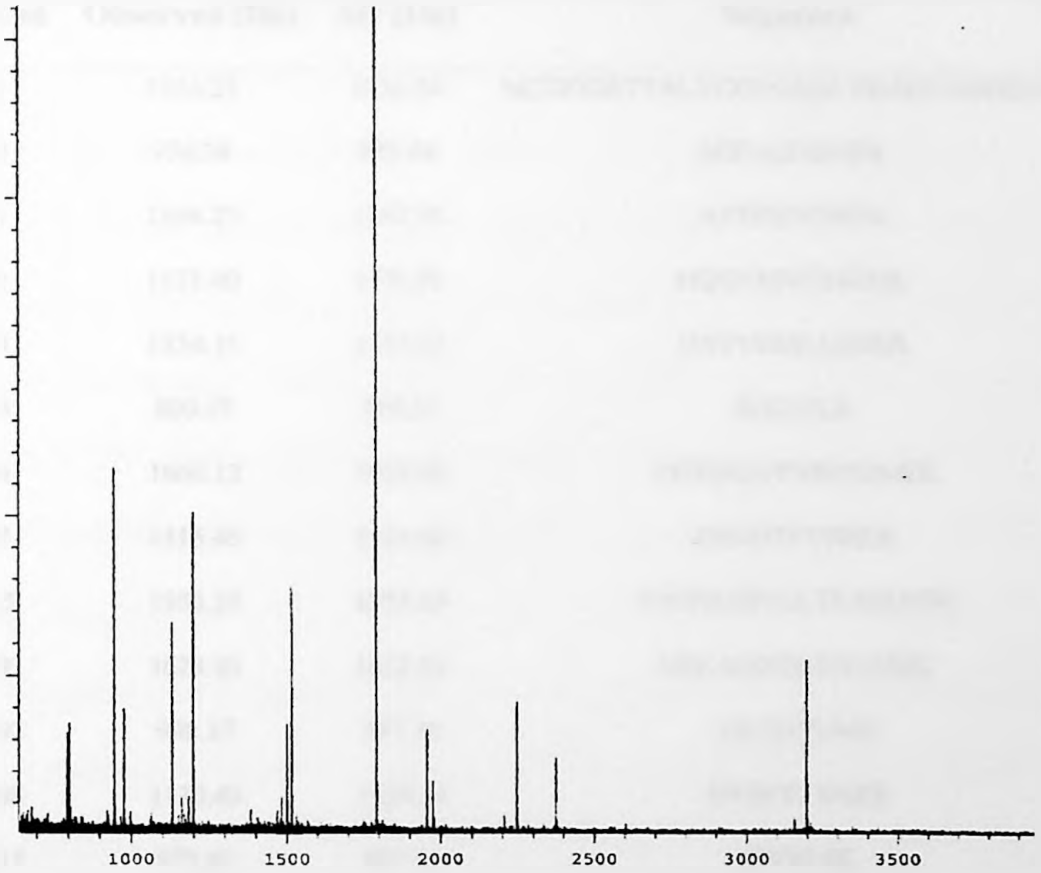


Figure 1-11. MALDI spectrum of actin after tryptic digest.

1 **MCDDDETTAL VCDNGSGLVK AGFAGDDAPR AVFPSIVGRP RHQGVMVGMG**
51 **QKDSYVGDEA QSKRGILTLK YPIEHGIITN WDDMEKIWHH TFYNELRVAP**
101 **EEHPTLLTEA PLNPK**ANREK MTQIMFETFN VPAMYVAIQA VLSLYASGRT
151 TGIVLDAGDG VTHNVPVYEG YALPHAIMRL **DLAGRDLTDY LMK**ILTER**GY**
201 **SFVTTAEREI VRDIK**EKLCY VALDFENEMA TAASSSSLEK **SYELPDGQVI**
251 **TIGNER**FRCP ETLFQPSFIG MESAGIHETA YNSIMK**CDID IRKDLYANNV**
301 **LSGGTTMYPG IADR**MQK**EIT ALAPSTMK**IK **IIAPPERK**YS VWIGGSILAS
351 LSTFQQMWIS K**QEYDEAGPS IVHR**KCF

Start-End	Observed (Da)	Mr (Da)	Sequence
1-30	3156.25	3156.24	MCDDDETTALVCDNGSGLVKAGFAGDDAPR
21-30	976.38	975.44	AGFAGDDAPR
31-41	1198.25	1197.70	AVFPSIVGRPR
42-52	1171.40	1170.56	HQGVMVGMGQK
53-64	1354.15	1353.62	DSYVGDEAQSKR
64-70	800.67	799.53	RGILTLK
71-86	1960.12	1959.90	YPIEHGIITNWDDMEK
87-97	1515.46	1515.46	IWHHTFYNELR
98-115	1956.24	1955.04	VAPEEHPTLLTEAPLNPK
180-193	1623.43	1622.83	LDLAGRDLTDYLMK
186-193	998.37	997.48	DLTDYLMK
199-208	1130.43	1129.54	GYSFVTTAER
209-215	872.81	871.51	EIVRDIK
241-256	1790.47	1789.88	SYELPDGQVITIGNER
287-292	791.25	790.36	CDIDIR
294-314	2228.36	2227.06	DLYANNVLSGGTTMYPGIADR
318-328	1161.61	1160.61	EITALAPSTMK
331-337	795.49	794.47	IIAPPER
331-338	923.50	922.56	IIAPPERK
362-374	1500.43	1499.70	QEYDEAGPSIVHR

Figure 1-12. Identification of actin using MASCOT. Sequence coverage of actin shown in red. Peptide peaks matching actin are also shown.

The entire methodology explained above proved to be a powerful tool for protein identification. Each of the three known proteins were successfully analyzed and characterized by electrophoresis, staining procedures, tryptic digests, and mass spectrometry. These same techniques were then applied to an unknown protein in order to investigate structure and activity.

CHAPTER 2

APPLICATIONS:

IDENTIFICATION OF UNKNOWN PROTEINS

BY METHODS DEVELOPMENT

INTRODUCTION

Obesity is a disease that affects at least 70 million Americans: more than one-third of all adults and one in five children. Each year obesity causes at least 300,000 excess deaths in the United States and costs the country more than $100 billion in health care. In the last 10 years, the proportion of the population that is obese has increased from 25% to 32%.[16] West Virginia, in particular, has a large population of obese people. In fact, it has ranked near the top among states with high obesity levels in 12 of the last 13 years.[17]

The term obesity refers to an excess in body weight relative to height. Obesity is defined as relative weight over 20% above ideal body weight. The body mass index (BMI) is well correlated with measures of body fat and is defined as weight (kg) divided by height^2 (m^2). Obesity is described as a BMI of >30 kg/m^2.[6]

The cause of most cases of obesity is not known. Genetic facts tend to interact with environmental factors. Obesity is the leading cause of many health risks. It increases one's risk of developing conditions such as high blood pressure, diabetes, heart disease, stroke, and cancer of the breast, prostate, and colon.[18]

A large amount of research has been done in the area of obesity regulation, particularly the relationship between obesity and leptin. Leptin is considered to be the

major player in the regulation of body fat. This 16 kDa protein is made by adipocytes (fat cells) and operates by interacting with its receptor in the hypothalamus to regulate energy expenditure and food intake. When the leptin receptor is defective, the signaling function of leptin is lost, which eventually leads to obesity.[19]

Researchers originally thought that injecting leptin would solve the problem of obesity. However, there is no difference in the structure of leptin between lean and obese human subjects. This suggests that the problem in obese individuals might be decreased sensitivity to leptin. Different approaches are now being taken in the search for obesity regulation, namely determining the function of the leptin receptor.[20]

Occasional spontaneous mutations in laboratory animal colonies have permitted the development of a number of valuable models for human diseases. Mutations in rodents leading to a genetically determined obesity have been prominent in this group. Two specific models are the fatty Zucker rat and the LA/N-corpulent (cp) rat. Each of these strains of rats contains a defect in their leptin receptor. The leptin receptor in the LA/N-cp rat contains a nonsense mutation, which causes formation of a stop codon from one that encodes an amino acid. This defect leads to premature termination of translation of the leptin receptor gene, rendering an inactive protein. The leptin receptor in the fatty Zucker rat contains a missense mutation, which arises from a base change that causes incorporation of a different amino acid in the encoded protein. As a consequence of these mutations to the leptin receptor gene, both strains of rats become severely obese, weighing as much as 3x the normal.[21]

These two rats are similar in some respects, yet differ in certain diseases that they develop. The LA/N rats heterozygous for the cp gene or homozygous normal are lean.

Homozygous cp rats end up hyperlipidemic, insulin-resistant, and develop atherosclerosis, which is a type of cardiovascular disease. Zucker rats will be lean if homozygous normal or heterozygous for the fatty gene. If homozygous for the fatty gene, the rat will be obese. The homozygous fatty rats have been reported to be hyperlipidemic and hyperinsulinemic, but do not develop atherosclerosis.[22]

The project began with the anticipation of discovering an early marker for hereditary obesity. A collaboration with a group of professors from Marshall University School of Medicine was undertaken to investigate this possibility. The group at the medical school began by analyzing the results of a 1-D gel on the aqueous layer of Zucker pup serum (both obese and lean). A particular protein was found to be of extreme interest. This protein was highly intense in the fatty serum but could not be seen at all in the lean sample. Further investigations found that if these lean rats mated, some of the pups would in fact be obese. These lean rats were considered to be heterozygous for the fatty gene based on this information.

The same procedure was performed on the serum of the Sprague Dawley rats. The Sprague Dawley was used as the control group as it does not contain any type of obesity mutation. The protein of interest was not found in the aqueous layer of this rat. Thus, investigations of this protein in the Zucker rats became a priority.

The protein was eventually identified as apolipoprotein A-I (Figure 2-1), which will be discussed later.

PRE | PRO | MATURE

MKAAVLAVAL VFLTGCQA WE FWQQ DEPQSQ WDRVKDFATV YVDAVKDSGR
51 DYVSQFESST LGKQLNLNLL DNWDTLGSTV GRLQEQLGPV TQEFWANLEK
101 ETDWLRNEMN KDLENVKQKM QPHLDEFQEK WNEEVEAYRQ KLEPLGTELH
151 KNAKEMQRHL KVVAEEFRDR MRVNADALRA KFGLYSDQMR ENLAQRLTEI
201 RNHPTLIEYH TKAGDHLRTL GEKAKPALDD LGQGLMPVLE AWKAKIMSMI
251 DEAKKKLNA

Figure 2-1. Amino acid sequence of apolipoprotein A-I.

Apo A-I is initially synthesized as a preproprotein. This protein is secreted by the cells of the liver (HDL) and small intestine (chylomicrons) in a proform with a mass of ~28 kDa. This proapo A-I is then converted to the mature protein, ~27 kDa, in the serum.[23]

Apo A-I contains helices which have their hydrophobic and hydrophilic residues on opposite sides of the helical cylinder (Figure 2-2).

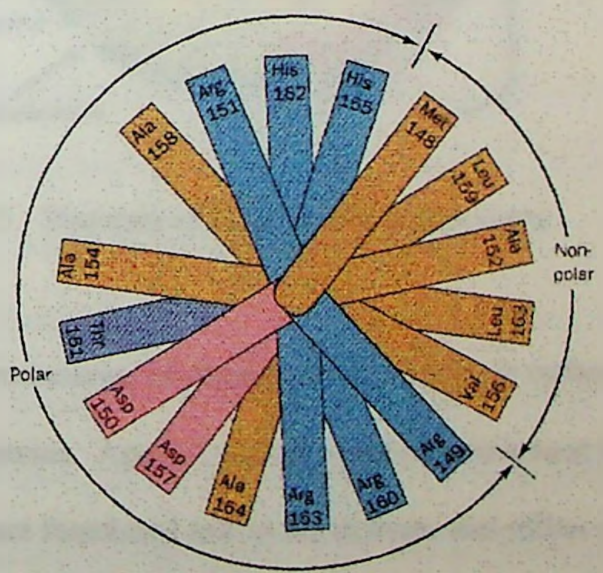


Figure 2-2. Helical wheel of apo A-I showing the hydrophobic and hydrophilic residues.

Furthermore, the polar helix face has a dipolar character because its negatively charged residues project from the center of this face, whereas its positively charged residues are located at its edges. This suggests that lipoprotein α helices float on phospholipid surfaces, much like logs on water.[19]

Serum apo A-I is the most important protein found in HDL. HDL is composed of a neutral lipid core consisting of cholesteryl ester and triglyceride (Figure 2-3).

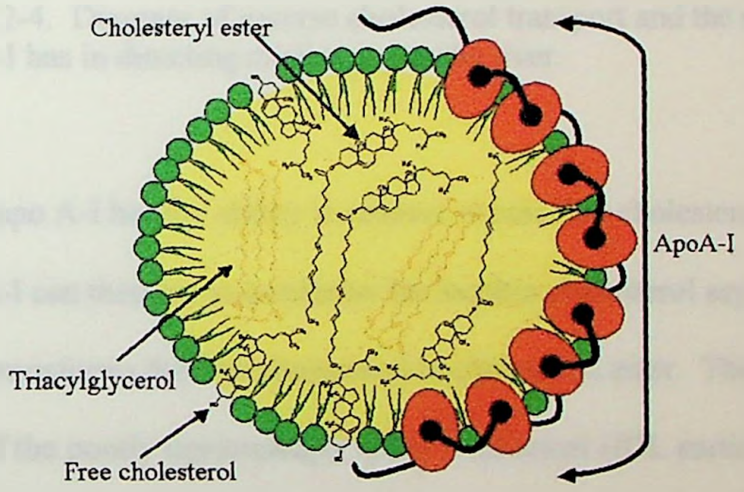


Figure 2-3. Structure of HDL and its components.

Surrounding the core is a monolayer of phospholipid in which is embedded free cholesterol and apolipoproteins. Apo A-I not only provides structural framework for HDL but also serves a direct functional role in the delivery and efflux of lipids to and from cells.[6,8]

Apo A-I has the ability to mediate reverse cholesterol transport (Figure 2-4).

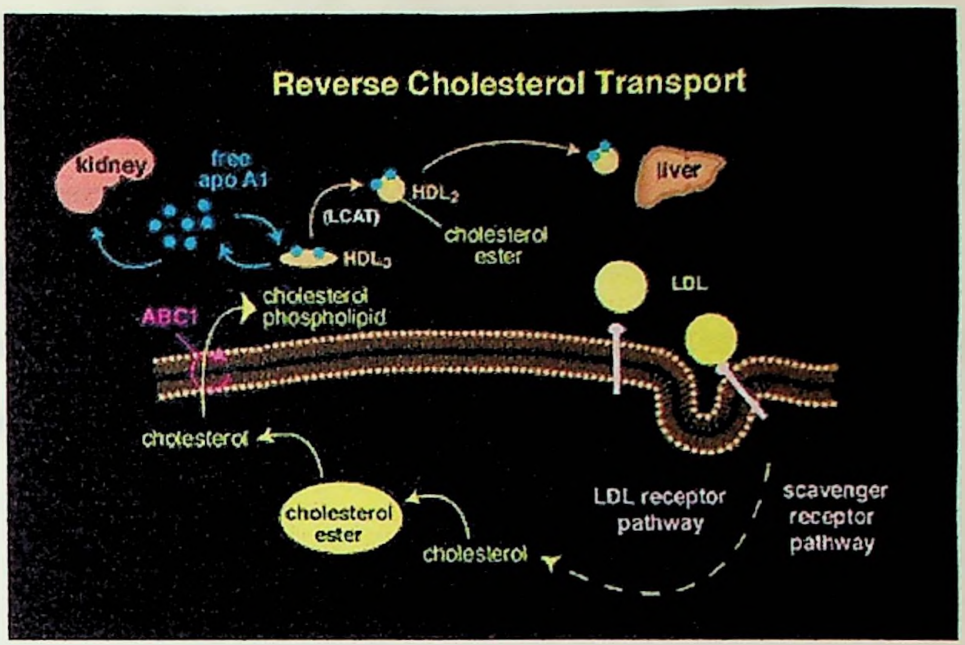


Figure 2-4. Diagram of reverse cholesterol transport and the role Apo A-I has in directing cholesterol to the liver.

Poorly lipidated apo A-I has the ability to remove excess free cholesterol from peripheral cells. The apo A-I can then act as a cofactor for lecithin cholesterol acyltransferase (LCAT), which transforms the free cholesterol to cholesteryl ester. This event initiates the conversion of the poorly lipidated apo A-I to a spherical HDL particle. The final step in reverse cholesterol transport involves the delivery of the HDL cholesterol to the liver where it can be excreted or repackaged into new lipoproteins.[6]

Unlike HDL, LDL transports cholesterol from the liver to the tissues. This causes an increase in cholesterol, which is why LDL is known as bad cholesterol. HDL, however, has the opposite function of LDL since it removes cholesterol from the tissues. This is why it is referred to as good cholesterol. HDL cholesterol levels are highly correlated to the plasma level of apo A-I. Thus, decreased levels of apo A-I are associated with an increased risk for atherosclerosis.[6] This information explains why

fatty Zucker rats, with their increased levels of apo A-I, do not suffer from atherosclerosis.

The purpose of this research was to investigate the possible role of apo A-I in hereditary obesity. Characterization of the protein's structure and function was performed using the proteomic approach. Utilization of electrophoresis, RP-HPLC, MS, and other techniques provided detailed information about apo A-I.

MATERIALS

Chemicals

Acros Organics, Fairlawn, NJ

isopropanol

Aldrich Chemical Co., Inc., Milwaukee, WI

ethanol

Bio Rad Laboratories, Hercules, CA

Tris

Fisher Scientific, Suwanee, GA

diethyl ether

formic acid

hydrocholoric acid

sodium chloride

Genomic Solutions, Inc., Chelmsford, MA

bromophenol blue

phosphoric acid

sample buffer mix

sodium hydroxide

Tris/Acetate equilibration buffer

Tris/Acetate pre-mixed lower buffer concentrate

Tris/Tricine/SDS pre-mixed running buffer powder

Instrumentation/Accessories

Fisher Scientific, Suwanee, GA

sonicator FS30

Vortex Genie 2

Genomic Solutions, Inc., Chelmsford, MA

precast IEF gels (carrier ampholyte, pH 4-8)

precast homogenous slab gels

Perkin Elmer, Norwalk, CT

C_4 Brownlee columns (7 μ, 30x2.1 nm)

METHODS

Digest of Unknown Protein

The unknown protein was excised from the 1-D gel containing the sample from the obese Zucker rat. A tryptic digest (long method, Table 1-11) was performed on the gel.[2] The resulting sample was mixed with an equal volume of HCCA matrix. One uL was spotted on a MALDI target and allowed to air dry. The sample was again analyzed using MALDI-TOF (Appendix I). The instrument was calibrated using the mixture described in Table 1-13. The peptides were investigated in the reflectron mode. The coarse laser energy was set to 30% with fine adjustment varying for each sample.[11] Once a quality spectrum was obtained, 100 laser shots were collected to produce a m/z spectrum of 500-4000 Da. The resulting peptide peaks were entered into MASCOT for identification. The same procedure was used to digest the protein using the enzymes chymotrypsin and Lys-C (Table 1-1).

Separation of Aqueous and Lipid Layers of Serum

Zucker serum, both lean and obese, was provided by Marshall University School of Medicine. The rats were fasted for 36 hours before extracting the serum. In order to separate the hydrophobic (lipid) and hydrophilic (aqueous) layers, each sample was first homogenated. This step was carried out by sonicating the serum for 30 minutes and then vortexing for 30 seconds. One mL aliquots were removed from each sample and placed in 1.5 mL Ependorf tubes. The samples were then centrifuged at 13000 rpm for 1.5 hours. The bottom aqueous layer was removed from the lipid layer and placed in a

separate tube. The same procedure was performed on Sprague Dawley rat serum for use as a control group.

Reverse-Phase HPLC

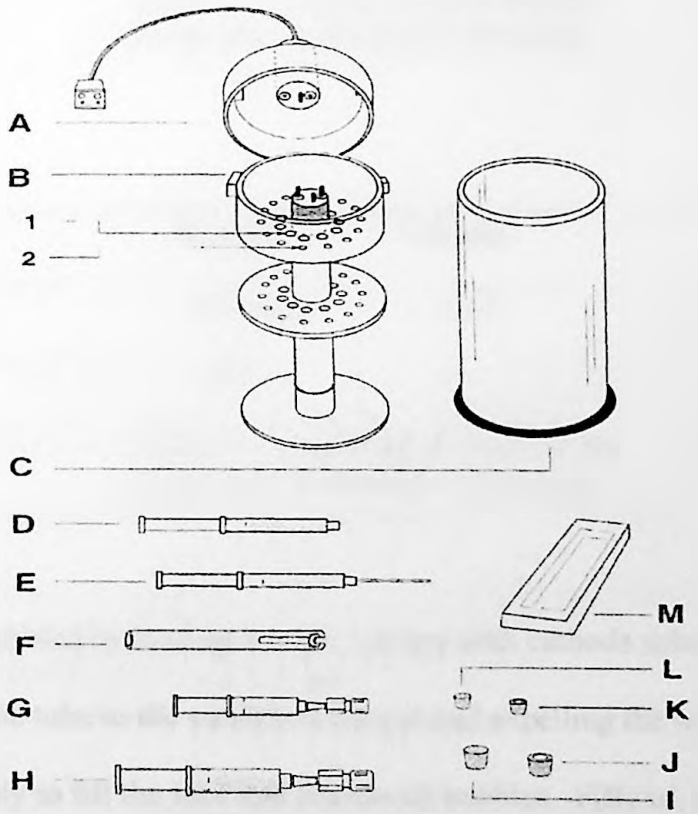
The aqueous layer of both the lean and obese samples was purified on an Applied Biosystems model 130A Separation System. Fifty µL of the sample was injected onto a C_4 column. Solvent A consisted of 0.1% AcOH and solvent B was 0.085% AcOH in 70% AcCN. The proteins were eluted from the column with a gradient of 10% B (0 to 10 min, 2 min to ramp) and 100% B (10 to 24 min, 2 minutes to ramp) at a flow of 200 µL/min. Column effluent was monitored at 214 nm, and proteins were collected simultaneously by peak. The samples were spun to dryness by vacuum centrifugation and then resuspended in 1% AcOH.

The protein extracts were then fractionated on the HPLC instrument. Fifty µL of sample was injected onto the C_4 column. The same solvents were again used. The proteins were eluted from the column with a gradient of 20% B (0 to 3 min) and 100% B (3 to 24 min) at a flow rate of 200 µL/min. Column effluent was monitored at 214 nm, and protein fractions were collected every minute into 1.5 mL Ependorf tubes. The samples were spun to dryness and resuspended in 1% AcOH. The samples were mixed with SA matrix and analyzed using MALDI-TOF in the linear mode.

2-D Electrophoresis

Two-dimensional electrophoresis was performed using the Genomic Solutions

2-D Investigator system. The following samples were used: HPLC fraction 18 for the Zucker obese, HPLC fraction 16 for the Zucker lean, and HPLC fraction 17 for Sprague Dawley. Each of the samples was found to contain high levels of apo A-I when analyzed by MALDI. The samples were prepared by diluting 20 µL of sample with 30 µL of Sample Buffer Mix (Genomic Solutions). The first dimension of the technique was to perform isoelectric focusing. The 1-D running tank was assembled (Figure 2-5).



1-D Running System Components

A	Cover and Cable	H	Preparative Extrusion
	Syringe		
B	Running Rack	I	Preparative Plug
C	Running Tank	J	Preparative Grommet
D	Hamilton Syringe	K	Analytical Grommet
E	Hamilton Syringe with needle	L	Analytical Plug
F	Installer Tool	M	Extrusion Tray
G	Analytical Extrusion	1	Preparative Holes
	Syringe (with adapter)	2	Analytical Holes

Figure 2-5. Components of the Investigator 1-D running system used for isoelectric focusing.[24]

The IEF gels (pH 4-8) were loaded into the tank (refer to the Investigator 2-D Operating and Maintenance Manual). The lower reservoir was filled with anode solution (Table 2-1) and the upper reservoir with cathode solution (Table 2-2).

Reagent	Volume
85% Phosphoric acid	1.4 mL
dH_2O	2 L

Table 2-1. Anode solution for the lower tank in isoelectric focusing.

Reagent	Volume
10N NaOH	10 mL
dH_2O	1 L

Table 2-2. Cathode solution for the upper tank in isoelectric focusing.

The gels were debubbled by loading a 1 mL syringe with cathode solution and lowering the needle within the tube to the surface of the gel and expelling the solution. The needle was drawn up slowly to fill the tube and release all bubbles. Fifty µL of sample was drawn into the syringe. The needle was inserted in the IEF tube, and the sample was expelled on top of the gel. The above step was repeated for the remaining samples. Table 2-3 describes the power supply program used to perform isoelectric focusing.

Display	Parameters
Function Setup	2F
# of Gels	2
Max Voltage	2000 v
Holding Voltage	125 v
Duration	17.5 h
Max Current	110 μA
Volt Hours	180000 v-h

Table 2-3. Parameters entered into the 2-D power supply for isoelectric focusing.

Once the run was complete, the tubes were then removed from the 1-D running system and placed on ice for 10 minutes. A 1 mL IEF gel extrusion syringe fitted with an extrusion adapter was filled with dH_2O. The thread flush end of the IEF tube was inserted into the gel extrusion adapter. The adapter was tightened. Constant pressure was exerted on the syringe until the gel began to move out of the tube. The gel was slowly extruded into the extrusion tray filled with equilibration buffer (Table 2-4) and incubated for 2 minutes at room temperature.

Reagent	Volume
Tris/Acetate equilibration buffer	100 mL
DTT	0.77 g
Bromophenol blue	0.01 g

Table 2-4. Equilibration buffer used for IEF gel.

The gels were ready for the second dimension of electrophoresis. The lower chamber of the 2-D tank (Figure 2-6) was filled with anode buffer (Table 2-5) and pre-chilled on the maximum setting of the Peltier chiller.

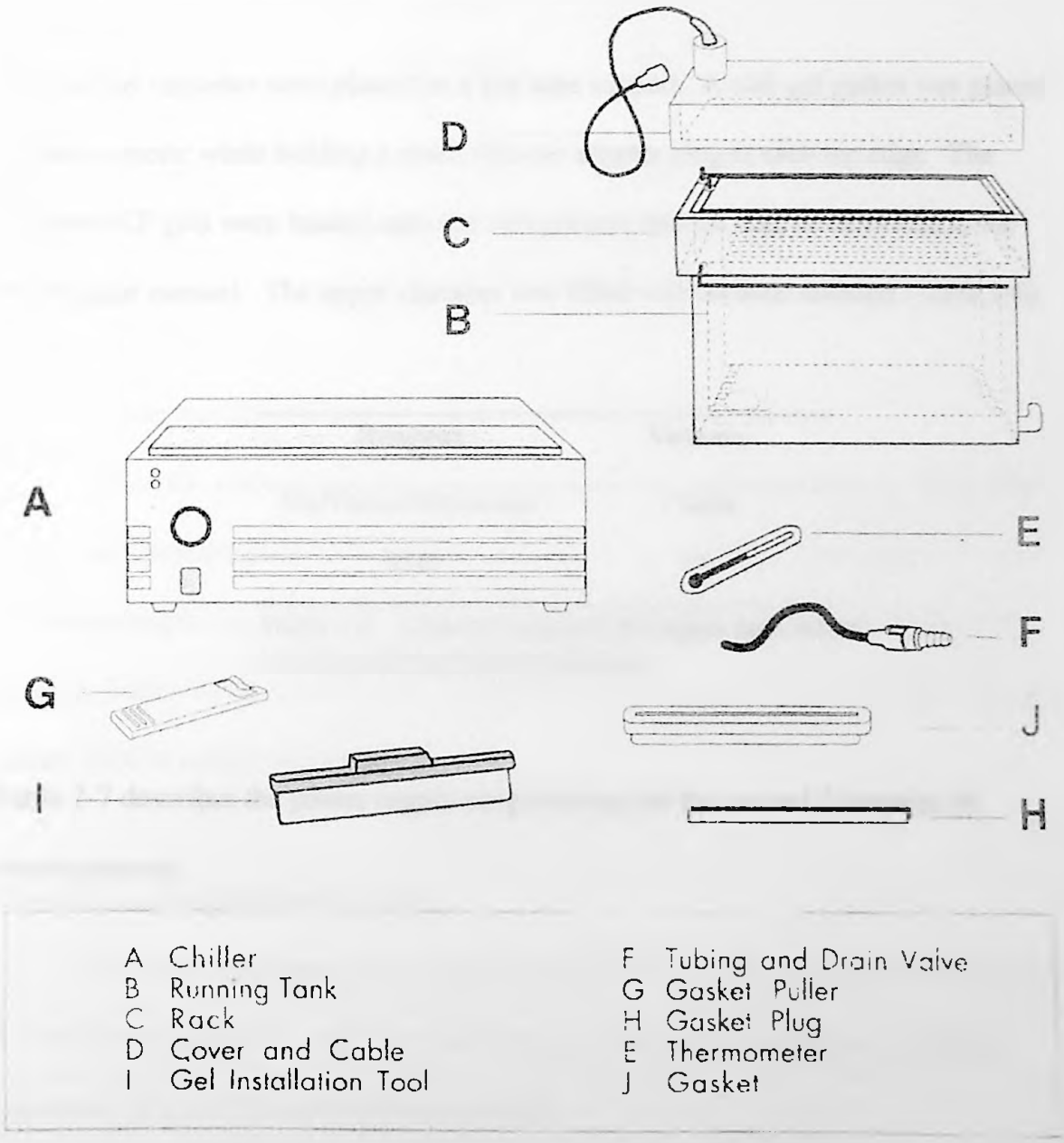


Figure 2-6. Components of the Investigator 2-D running system used for slab gel electrophoresis.[24]

Reagent	Volume
25x Tris/Acetate concentrate	440 mL
dH_2O	11 L

Table 2-5. Anode solution for lower tank when running slab gel electrophoresis.

The slab gel cassettes were placed on a test tube support. A slab gel gasket was placed on each cassette while holding a small silicone adapter plug at each top edge. The extruded IEF gels were loaded onto the slab gel and into the tank as described in the Investigator manual. The upper chamber was filled with cathode solution (Table 2-6).

Reagent	Volume
Tris/Tricine/SDS powder	1 bottle
dH_2O	3 L

Table 2-6. Cathode solution for upper tank when running slab gel electrophoresis.

Table 2-7 describes the power supply programming for the second dimension of electrophoresis.

Display	Parameters
Function Setup	3S
Slab Output Select	1L
# of Gels	2
Max Voltage	500 v
Duration	8 h
Max Power	20000 mW

Table 2-7. Parameters entered into the 2-D power supply for slab gel electrophoresis.

The gels were run until the blue dye front ran off the gel.[24]

When the run was complete, the gels were removed from the tank. They were stained in Coomassie overnight and then destained for ~6 hours. The spots corresponding to the protein of interest were excised and digested with trypsin (long method, Table 1-11). The samples were analyzed by MALDI-TOF MS in the reflectron mode. HCCA matrix was used.

Dialysis and Delipidation of Serum

The obese and lean serums were homogenized and 100 μL of each were placed in 1.5 mL Ependorf tubes. Dialysis was then performed using the Dispo-Biodialyzer apparatus. Figure 2-7 outlines the procedure.

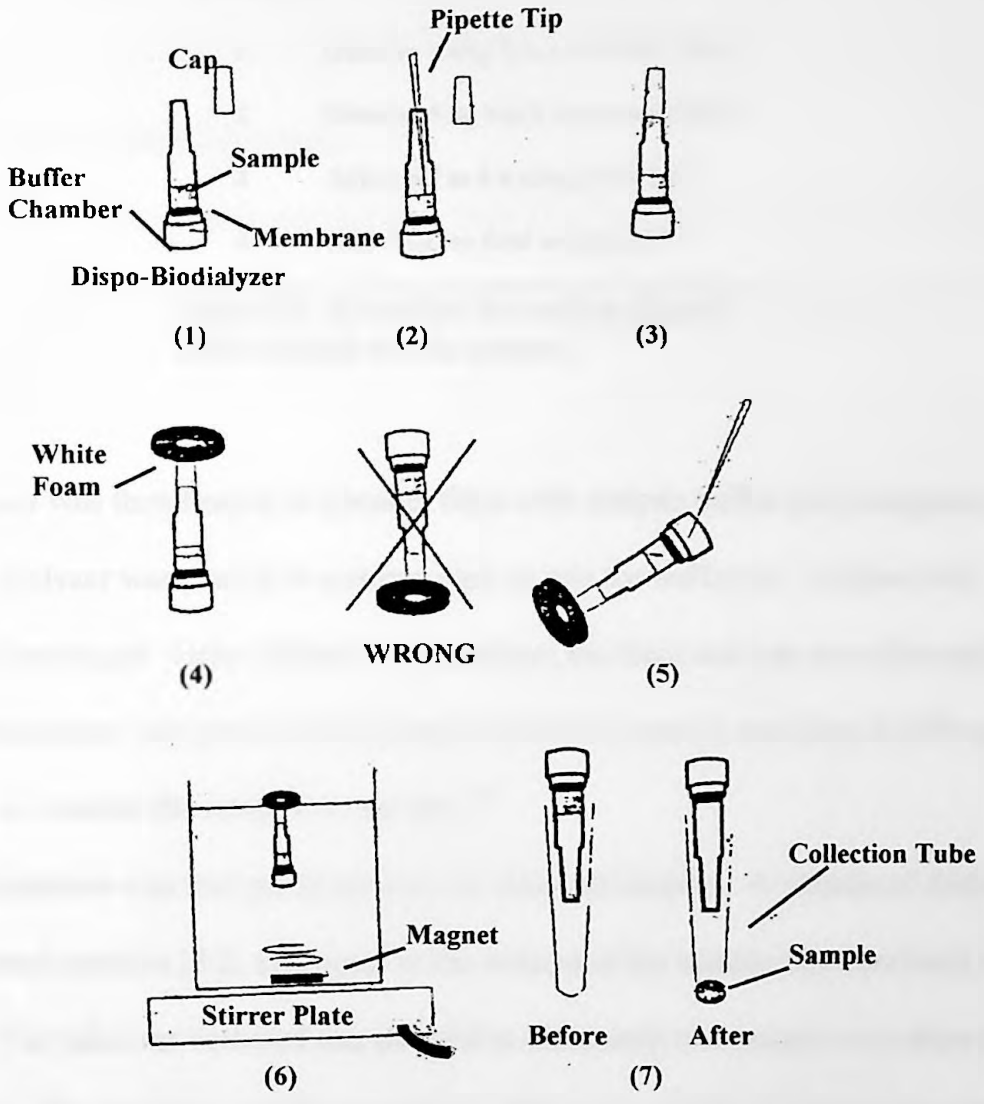


Figure 2-7. Procedure for using the Bio-dialyzer to perform dialysis on samples.[25]

The cap was removed (1) and 100 μL of the serum was placed in the dialyzer (2). The cap was replaced and closed firmly (3). The white foam was then placed over the cap for flotation (4). Five hundred μL of dialysis buffer (Table 2-8) was placed in the lower cavity buffer chamber during dialysis to displace air bubbles (5).

Step	Procedure
1	Dissolve 1.84g Tris in 500 mL dH_2O
2	Dissolve 5.8g NaCl in above solution
3	Adjust pH to 8.4 using 0.5M HCl
4	Add dH_2O to final volume of 1 L

Table 2-8. Procedure for making dialysis buffer needed for the dialyzer.

The dialyzer was then floated in a beaker filled with dialysis buffer and a magnetic stir bar. The dialyzer was placed on a stirrer plate to mix the buffer (6). Dialysis was continued overnight. Once dialysis was complete, the foam and cap were removed. The Dispo-Biodialyzer was placed in the sample collection tube (7) and spun at 2000 rpm for 1 second to transfer the samples to the tube.[25]

Delipidation was then performed on the dialyzed samples. A volume of diethyl ether/ethanol solution (3:2, v/v) equal to the volume of the sample was combined with the sample. The tube was vortexed and allowed to stand until two phases were observed (~20 min). The top layer was then removed. The same volume of diethyl ether/ethanol (3:1, v/v) was added. The sample was vortexed, allowed to stand 5 minutes, and then the top layer removed. This extraction process was repeated 5 times. The ether was then evaporated using the SpeedVac, resulting in a powder. The powder was reconstituted in 2.5% formic acid.[26]

Analysis was then carried out in several ways. The proteins of the delipidated serum (obese and lean) were investigated by MALDI-TOF in the linear mode. SA matrix was used. The samples were also examined by 2-D electrophoresis.

RESULTS AND DISCUSSION

Identification of Unknown Protein

The unknown protein excised from the 1-D gel of the aqueous layer of the Zucker obese rat had been digested by trypsin and the resulting MALDI spectrum is depicted in Figure 2-8.

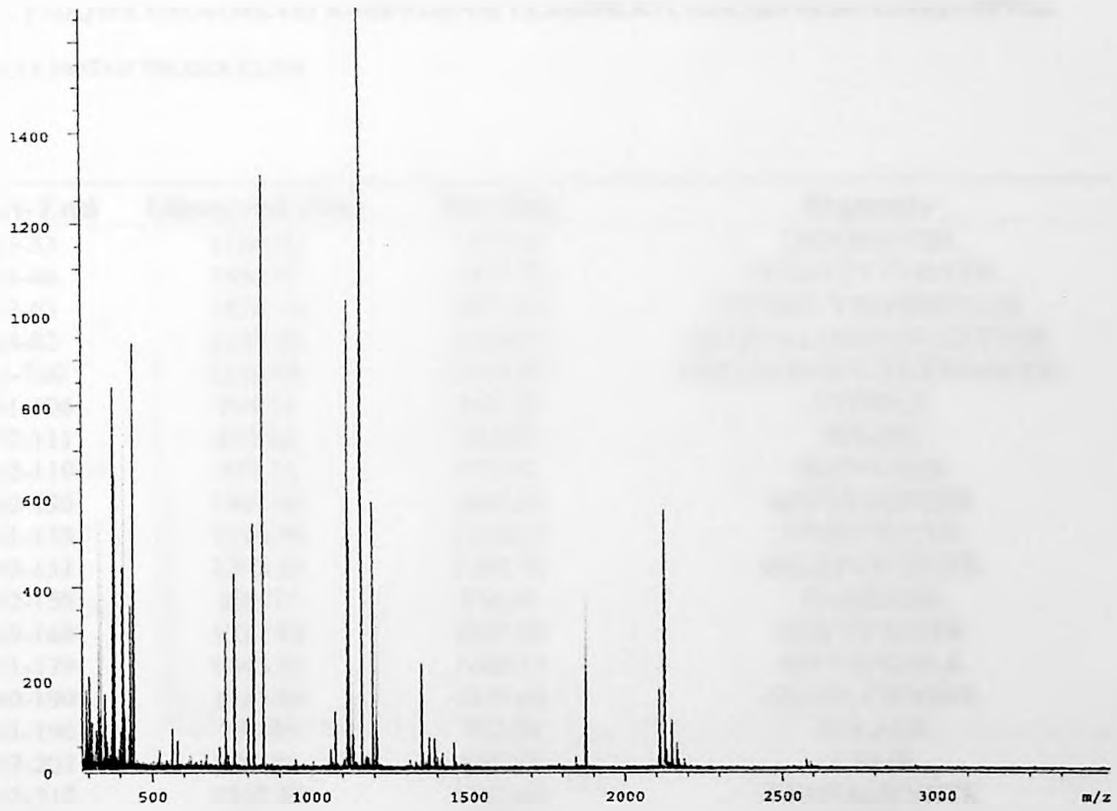


Figure 2-8. MALDI spectrum of unknown protein digested by trypsin.

Each of the peaks >500 Da corresponded to the peptides cleaved from the protein, specifically after R and K. A peak list of these peptides was generated by MALDI and searched by MASCOT. The database identified the protein to be the mature form of

apolipoprotein A-I with a sequence coverage of 93% (Figure 2-9). The pre and pro pieces of the protein were not seen because each is cleaved to activate the mature form of apo A-I once it enters the serum.

1
MKAAVLAVAL VFLTGCQAWE FWQQ**DEPQSQ WDRVKDFATV YVDAVKDSGR DYVSQFESST**
61
LGKQLNLNLL DNWDTLGSTV GRLQEQLGPV TQEFWANLEK ETDWLRNEMN KDLENVKQKM
121
QPHLDEFQEK WNEEVEAYRQ KLEPLGTELH KNAKEMQRHL KVVAEEFRDR **MRVNADALRA**
181
KFGLYSDQMR ENLAQRLTEI KNHPTLIEYH TKASDHLKTL GEK**AKPALDD LGQGLMPVLE**
241
AWKAK**IMSMI DEAK**KKLNA

Start-End	Observed (Da)	Mr (Da)	Sequence
25-33	1160.00	1160.00	DEPQSQWDR
34-46	1454.97	1454.78	VKDFATVYVDAVK
47-63	1876.18	1875.86	DSGRDYVSQFESSTLGK
64-82	2129.46	2129.09	QLNLNLLDNWDTLGSTVGR
83-100	2130.08	2130.08	GRLQEQLGPVTQEFWANLEK
101-106	819.75	819.39	ETDWLR
107-111	635.26	635.27	NEMNK
112-119	973.12	973.52	DLENVKQK
120-130	1402.09	1401.64	MQPHLDEFQEK
131-139	1195.98	1195.53	WNEEVEAYR
140-151	1393.21	1392.78	QKLEPLGTELHK
152-158	876.77	876.43	NAKEMQR
159-168	1227.84	1227.68	HLKVVAEEFR
171-179	1045.62	1045.55	MRVNADALR
180-190	1315.89	1315.64	AKFGLYSDQMR
191-196	730.80	730.38	ENLAQR
197-201	603.21	603.37	LTEIK
202-212	1352.89	1352.69	RNHPTLIEYHTK
213-218	670.46	670.34	ASDHLK
224-243	2152.43	2152.14	AKPALDDLGQGLMPVLEAWK
246-254	1037.68	1037.49	IMSMIDEAK

Figure 2-9. Summarizes the matching peaks (bottom) and sequence coverage (top in red) obtained from the tryptic digest for apo A-I.

The protein was also digested by chymotrypsin (cleaves after F, Y, W) and Lys-C (cleaves after K) for further confirmation (Figure 2-10 and Figure 2-11, respectively).

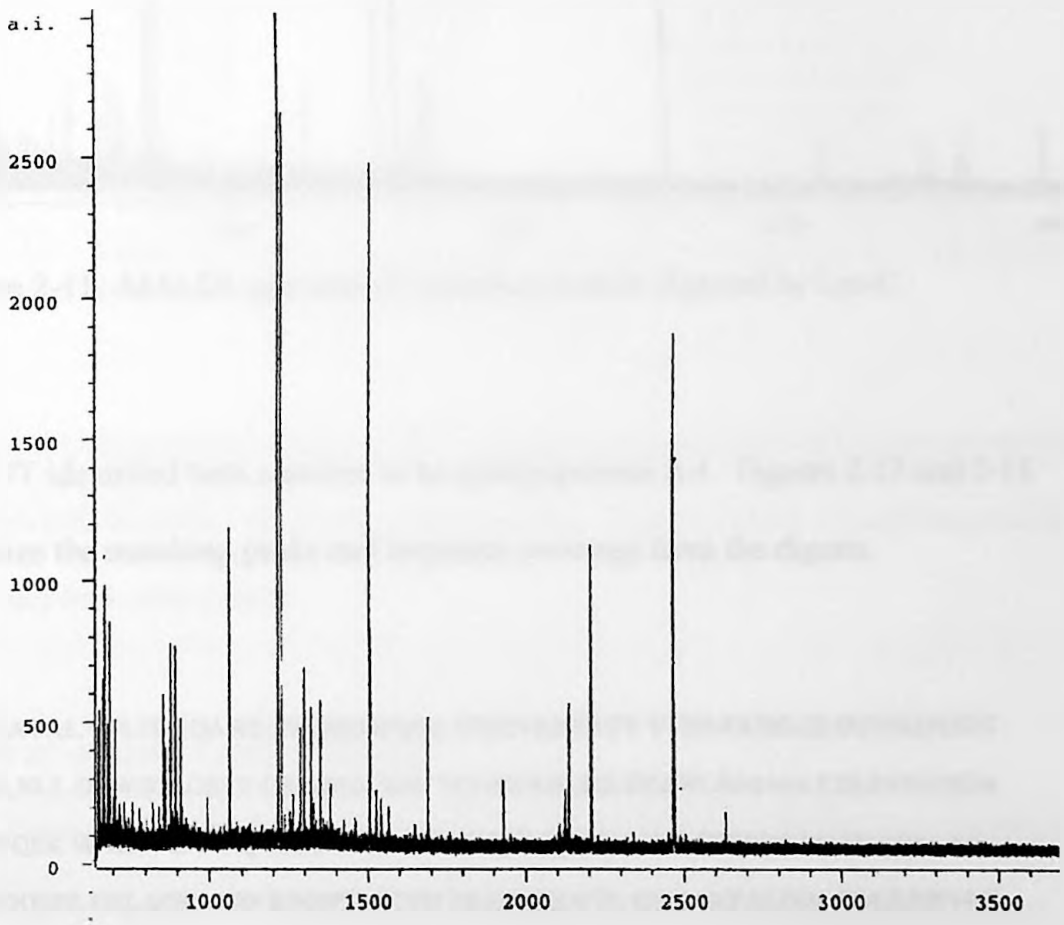


Figure 2-10. MALDI spectrum of unknown protein digested by chymotrypsin.

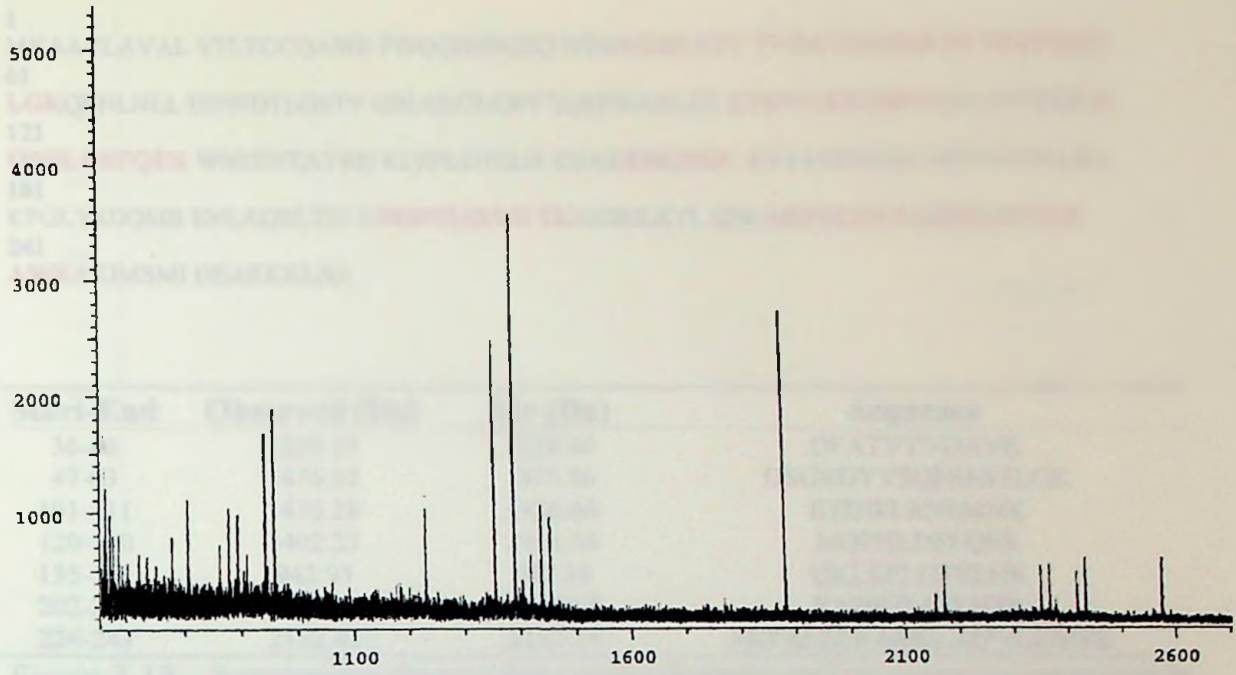


Figure 2-11. MALDI spectrum of unknown protein digested by Lys-C.

MASCOT identified both searches to be apolipoprotein A-I. Figures 2-12 and 2-13 summarize the matching peaks and sequence coverage from the digests.

1
MKAAVLAVAL VFLTGCQAWE FWQQDEPQSQ W**DRVKDFATV YVDAVKDSGR DYVSQF**ESST
61
LGKQLNLNLL DNWDTLGSTV GRLQEQLGPV TQEF**WANLEK ETDW**LRNEMN KDLENVKQKM
121
QPHLDEFQEK W**NEEVEAY**RQ KLEPLGTELH KNAKEMQRHL KVVAEEFRDR MRVNADALRA
181
KF**GLYSDQMR ENLAQRLTEI KNHPTLIEYH TKASDHLKTL GEKAKPALDDLGQGLMPVLE**
241
AWKAKIMSMI DEAKKKLNA

Start-End	Observed (Da)	Mr (Da)	Sequence
32-37	779.54	779.41	DRVKDF
38-56	2121.52	2120.03	ATVYVDAVKDSGRDYVSQF
95-104	1291.83	1291.60	WANLEKETDW
132-138	853.79	853.36	NEEVEAY
183-209	3204.97	3204.64	GLYSDQMRENLAQRLTEIKNHPTLIEY
210-242	3572.05	3571.88	HTKASDHLKTLGEKAKPALDDLGQGLMPVLEAW

Figure 2-12. Summarizes the matching peaks (bottom) and sequence coverage (top in red) obtained from the chymotrypsin digest for apo A-I.

1
MKAAVLAVAL VFLTGCQAWE FWQQDEPQSQ WDR**VKDFATV YVDAVKDSGR DYVSQFESST**
61
LGKQLNLNLL DNWDTLGSTV GRLQEQLGPV TQEFWANLEK **ETDWLRNEMN K**DLENVKQK**M**
121
QPHLDEFQEK WNEEVEAYRQ KLEPLGTELH KNAK**EMQRHL K**VVAEEFRDR MRVNADALRA
181
KFGLYSDQMR ENLAQRLTEI K**NHPTLIEYH TK**ASDHLKTL GEK**AKPALDD LGQGLMPVLE**
241
AWKAKIMSMI DEAKKKLNA

Start-End	Observed (Da)	Mr (Da)	Sequence
36-46	1228.14	1228.40	DFATVYVDAVK
47-63	1876.62	1875.86	DSGRDYVSQFESSTLGK
101-111	1436.28	1436.60	ETDWLRNEMNK
120-130	1402.23	1401.64	MQPHLDEFQEK
155-161	942.93	942.14	QKLEPLGTELHK
202-212	1353.25	1352.69	RNHPTLIEYHTK
224-243	2152.43	2152.14	AKPALDDLGQGLMPVLEAWK

Figure 2-13. Summarizes the matching peaks (bottom) and sequence coverage (top in red) obtained from the Lys-C digest for apo A-I.

Although the chymotrypsin and Lys-C digests did not result in a high sequence coverage, the search still identified the protein to be apo A-I. Thus, a correct identification of the protein had been determined.

Analysis of Aqueous Layer of Serum

The purified samples of the aqueous layers of Sprague Dawley, Zucker obese, and Zucker lean serums were first analyzed using mass spectrometry. Zucker serum was compared to the Sprague Dawley serum (control group) to investigate possible modifications. Figure 2-8 is the MALDI spectrum of the aqueous layer of Sprague Dawley rat serum.

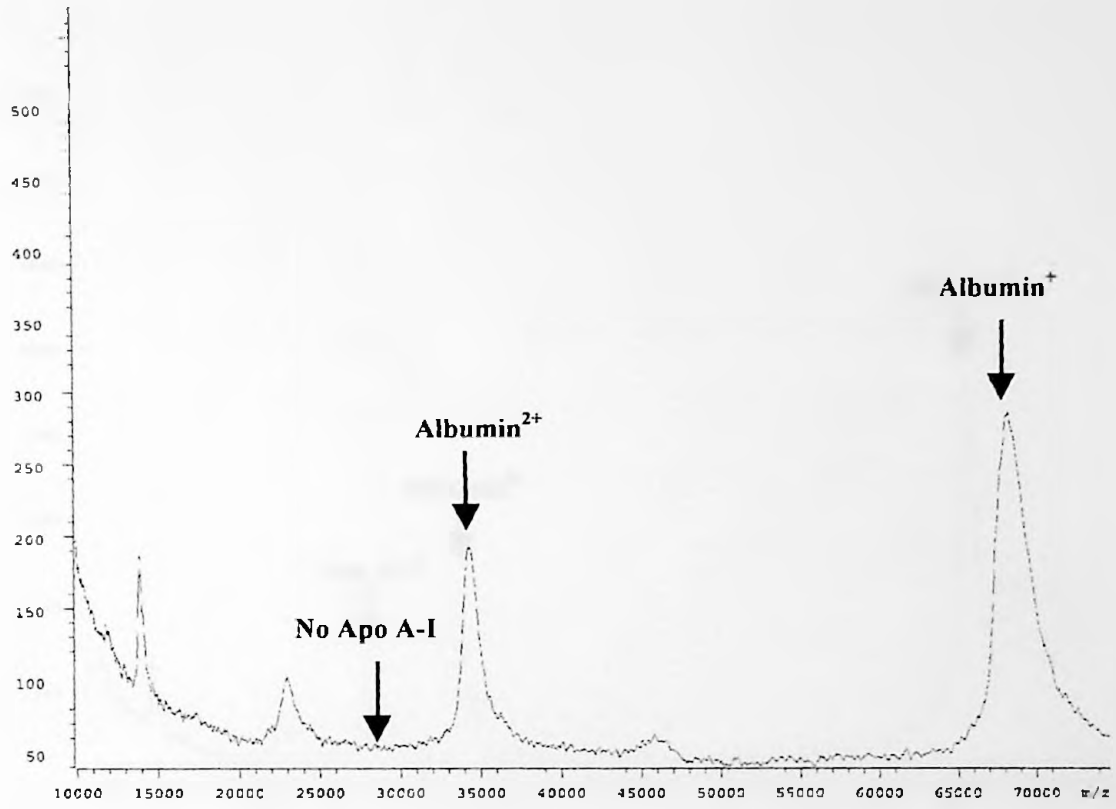


Figure 2-14. MALDI spectrum of the aqueous layer of Sprague Dawley serum. Apo A-I was not present.

The protein found in the greatest abundance was RSA (m/z ~68000). The $(M+2H)^{2+}$ and $(M+3H)^{3+}$ ions of RSA were also observed (m/z ~34000 and m/z ~22350). As expected, apo A-I was not present in the aqueous layer of Sprague Dawley serum. Figure 2-15 represents the proteins in the aqueous layer of Zucker lean serum.

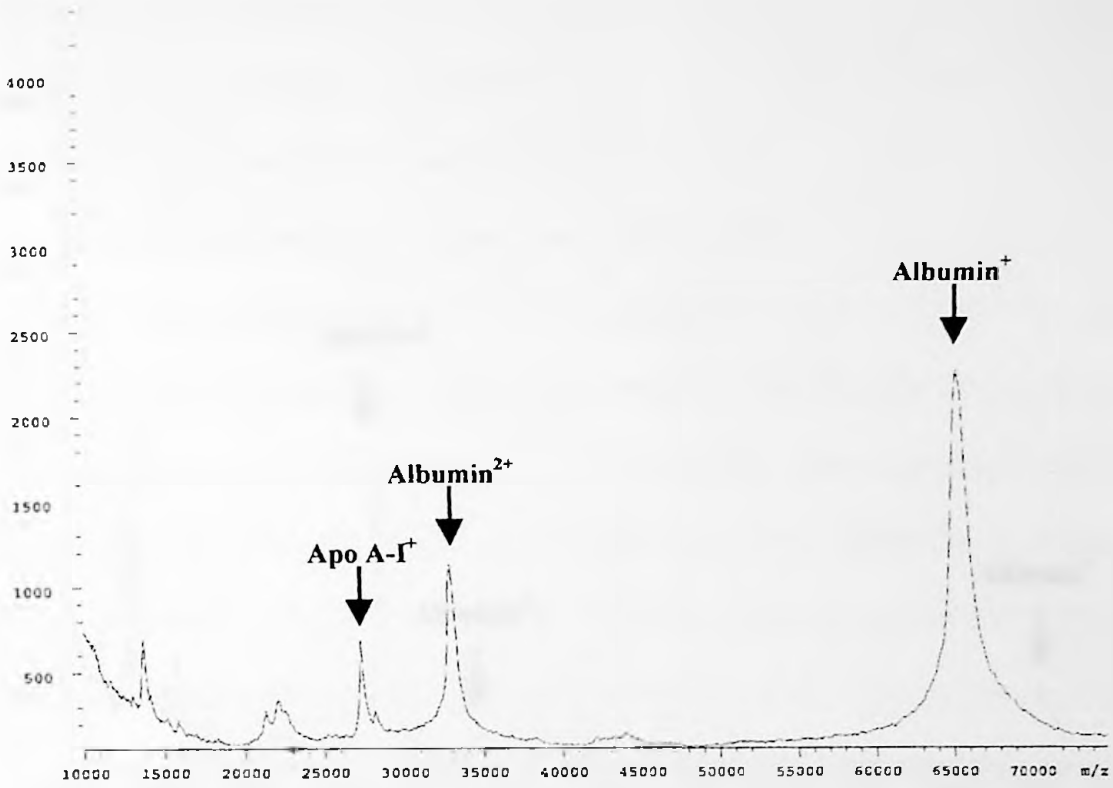


Figure 2-15. MALDI spectrum of the aqueous layer of Zucker lean serum. Apo A-I was present.

RSA was again present in both the $(M+H)^+$ and $(M+2H)^{2+}$ ion forms. Interestingly, apo A-I was also detected. Both the mature (m/z ~27000) and pro (m/z ~28000) forms of apo A-I were observed. The same proteins were also found in the Zucker obese aqueous layer (Figure 2-16).

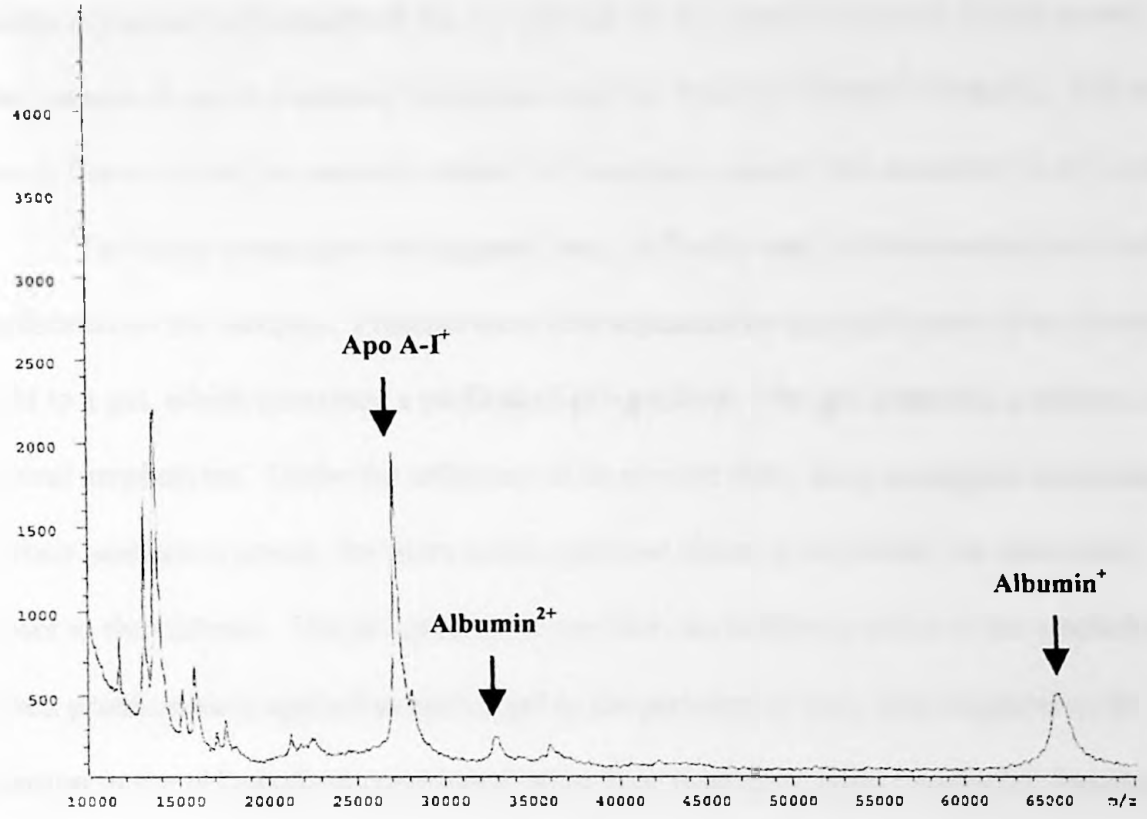


Figure 2-16. MALDI spectrum of the aqueous layer of Zucker obese serum. Apo A-I was present.

Apolipoprotein A-I was present in both the lean and obese aqueous layers of Zucker serum. The separation methodology was confirmed to be correct by the control group in which apo A-I was not seen. As stated previously, apo A-I is normally found in the lipid layer where it is noncovalently bound to such lipids. The fact that apo A-I was found in the aqueous layer of Zucker serum proved a difference in this protein of this particular rat. The differences could be a concentration issue in that apo A-I was so abundant that it was actually being forced into the aqueous layer. A disposition issue could be another possibility in that the rat simply cannot get rid of the protein.

Comparisons of the spectra of the Zucker lean to the Zucker obese aqueous layer indicated apo A-I to be much more abundant in the obese serum than in the lean. These

results coincided with results of the 1-D gel run on the aqueous layer of Zucker serum. The amount of apo A-I seemed to increase with the inherited chance of obesity. Whether apo A-I could in fact be an early marker for hereditary obesity still remained to be seen.

To further investigate the aqueous layer of Zucker rats, 2-D electrophoresis was performed on the samples. Proteins were first separated by the application of an electric field to a gel, which contained a preformed pH-gradient. The gel contained a mixture of several ampholytes. Under the influence of an electric field, these segregated according to their isoelectric points; the more acidic gathered closer to the anode, the more basic closer to the cathode. The pH-gradient arose from the buffering action of the ampholytes. When proteins were applied to such a gel in the presence of urea, they migrated to the position in the pH-gradient corresponding to their isoelectric point (isoelectric focusing). Second-dimension separation occurred when the IEF-gel was soaked in SDS-containing buffer and then positioned onto a slab gel to which another electric field was applied. Focused proteins were thereby resolved by their molecular weight like a conventional SDS-PAGE experiment.[2] For troubleshooting 2-D electrophoresis, see Appendix IV.

Figure 2-17 depicts the results of 2-D electrophoresis on the aqueous layer of Zucker obese serum.

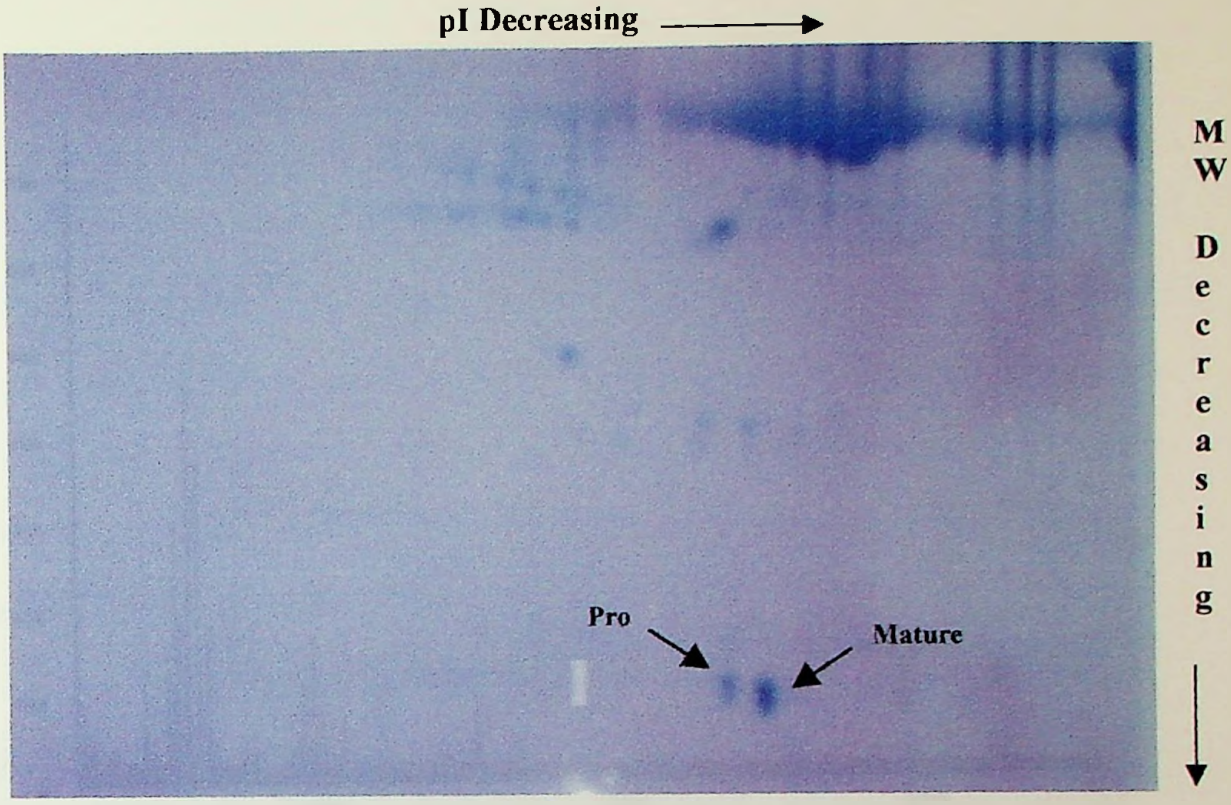


Figure 2-17. Zucker obese aqueous layer after 2-D electrophoresis was performed. Both mature and pro forms of apo A-I were seen.

Based on its molecular weight and pI (Table 2-10), the vicinity of apo A-I was determined.

Protein	MW (Da)	pI
Proapo A-I	28276	5.38
Mature Apo A-I	27372	5.26

Table 2-9. Molecular weight and pI of proapo A-I and mature apo A-I.[28]

Two spots were present in the general area. Both spots were excised and digested with trypsin to correctly determine their identification. The first spot (left to right) was classified as the pro form of apo A-I (Figure 2-18).

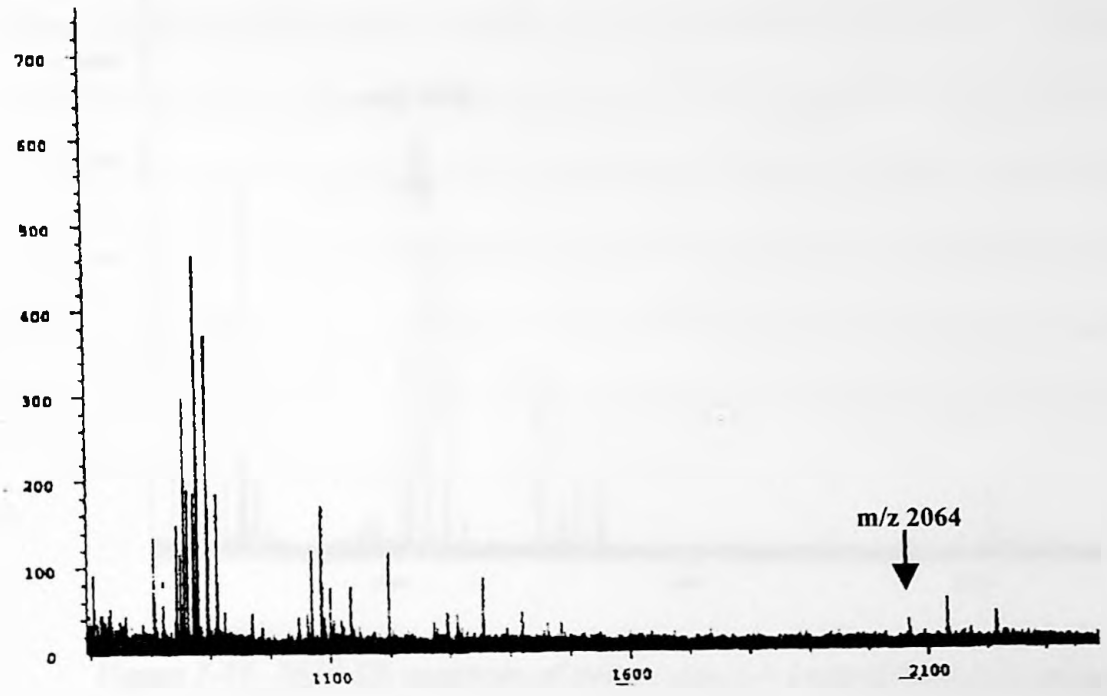


Figure 2-18. MALDI spectrum of proapo A-I excised from 2-D gel and digested using trypsin.

The protein was identified by MASCOT and determined to be proapo A-I by the peak at m/z 2064. This peak corresponded to fragment 19-33 of the amino acid sequence (Figure 2-1). This segment includes the hexapeptide pro piece. The second spot was identified as the mature protein based on the peak at m/z 1160 (Figure 2-19).

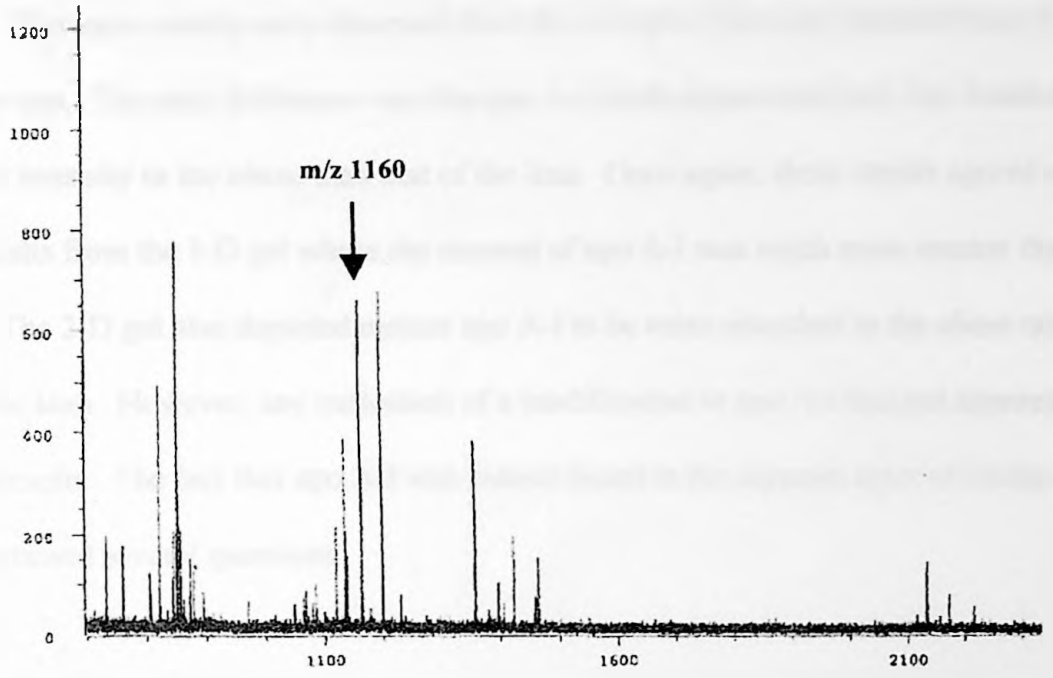


Figure 2-19. MALDI spectrum of mature apo A-I excised from 2-D gel and digested using trypsin.

This peak is fragment 25-33 of the apo A-I amino acid sequence, the segment directly after the pro piece (fragment 19-24). This protein would not normally cleave after the pro piece when using trypsin to fragment because trypsin will only cleave after K or R, not Q. Also, no peak was found at m/z 2064. Thus, no pro piece was observed, indicating the protein to be the mature form.

These results coincided with the results from the 2-D gel in that the mature migrated further through the gel based on MW and pI than the proprotein (Figure 2-17). In addition, the mature protein was found in greater abundance than the proprotein (based on the intensity of the spot). As mentioned previously, the proprotein is converted to the active mature protein when entering the serum. So, a larger amount of mature apo A-I was expected.

The same results were observed from the 2-D gel of the lean aqueous layer of Zucker rats. The only difference was that apo A-I (both mature and pro) was found in greater intensity in the obese than that of the lean. Once again, these results agreed with the results from the 1-D gel where the amount of apo A-I was much more intense than the lean. The 2-D gel also depicted mature apo A-I to be more abundant in the obese rather than the lean. However, any indication of a modification to apo A-I was not apparent in these results. The fact that apo A-I was indeed found in the aqueous layer of Zucker serum raised several questions.

Analysis of Delipidated Serum

The separation of apolipoproteins from Zucker serum was achieved by dialysis followed by delipidation. Figure 2-20 portrays the MALDI results of delipidated serum of the Zucker obese rat.

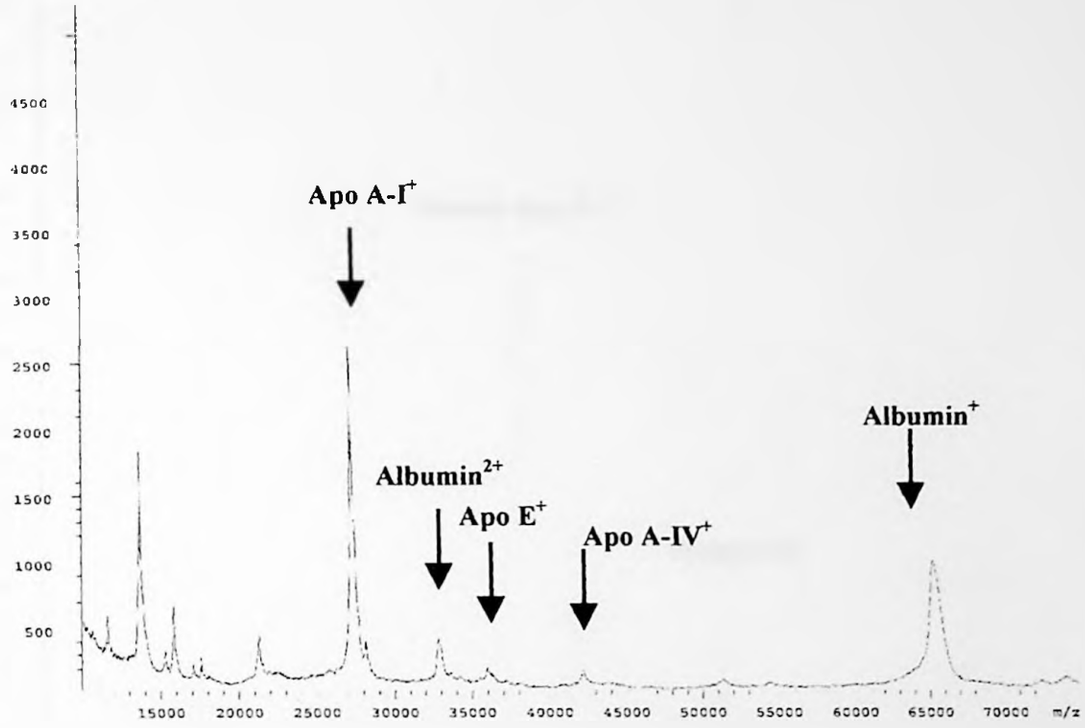


Figure 2-20. MALDI spectrum of delipidated Zucker obese serum.

Most of the proteins that were detected were apolipoproteins. Apo A-I was again the most abundant protein found. Both the mature and the proprotein were seen. Other proteins found include RSA, apo E, and apo A-IV. Upon closer inspection of apo A-I, the protein appeared to contain several modifications to both the mature and proprotein (Figure 2-21).

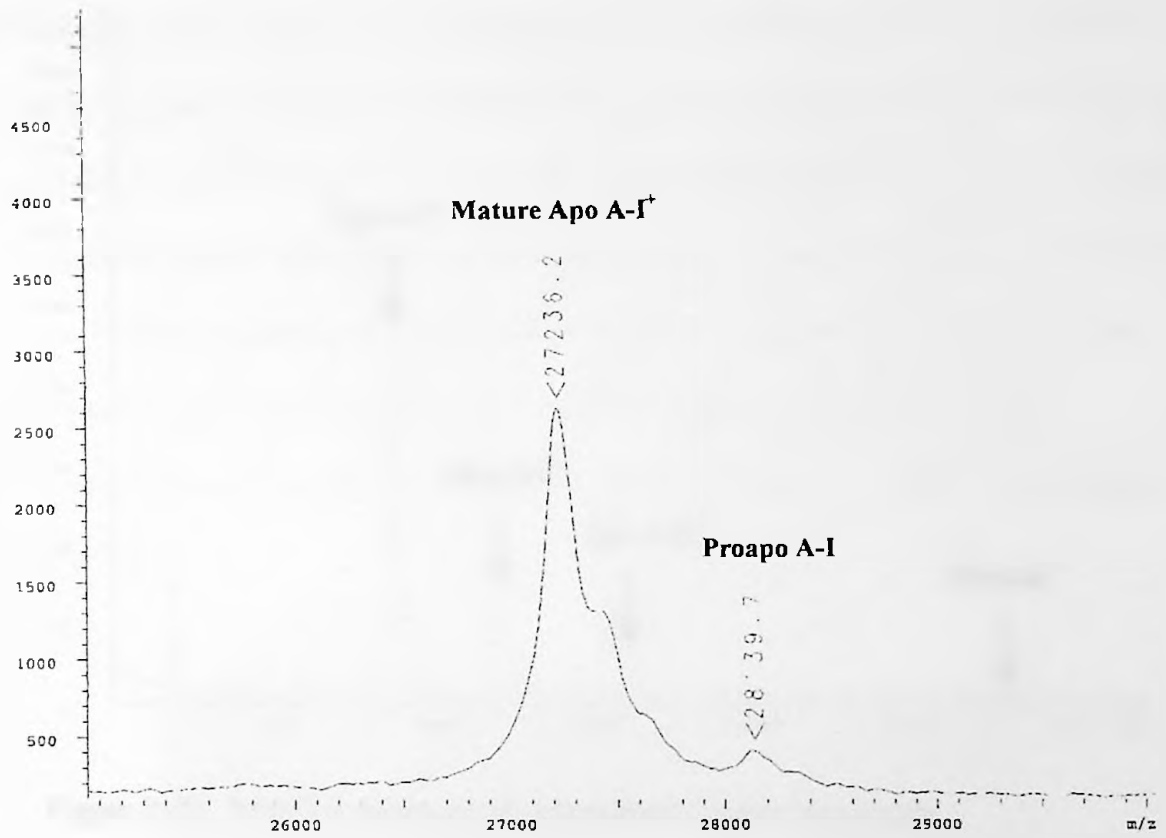


Figure 2-21. Zoomed area of a MALDI spectrum of apo A-I in delipidated Zucker obese serum.

Each of these peaks were ~200 Da apart. The results obtained from the lean serum are shown in Figure 2-22 and Figure 2-23 (subview).

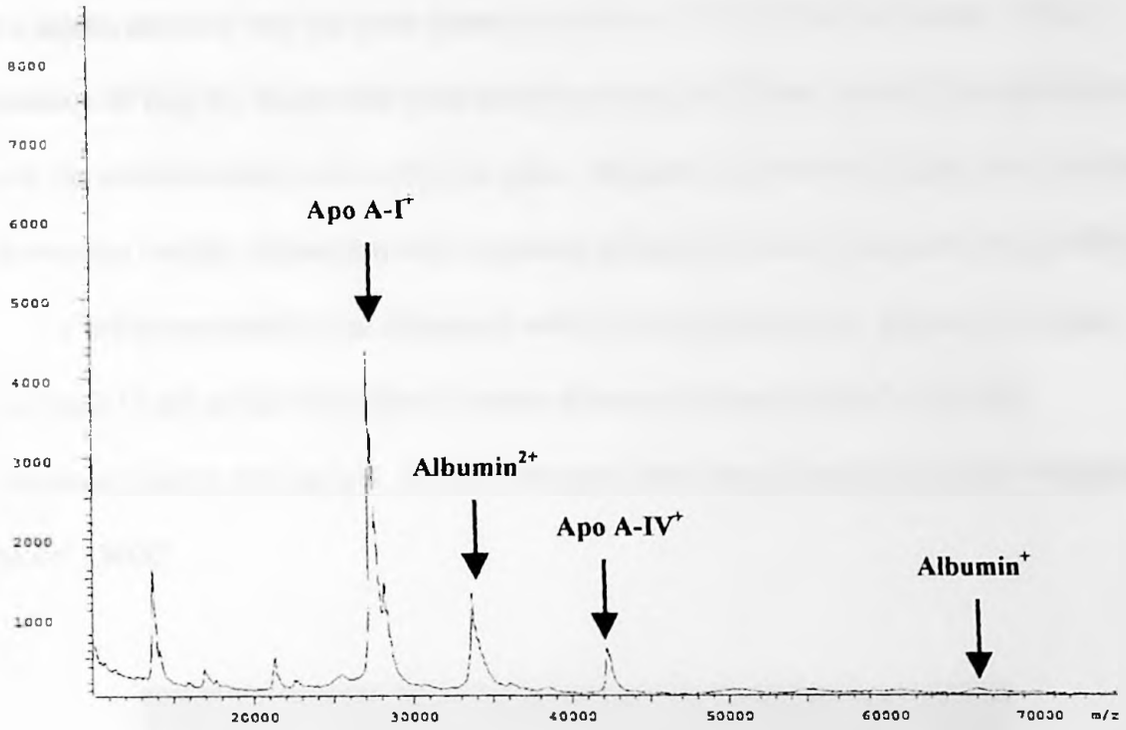


Figure 2-22. MALDI spectrum of delipidated Zucker lean serum.

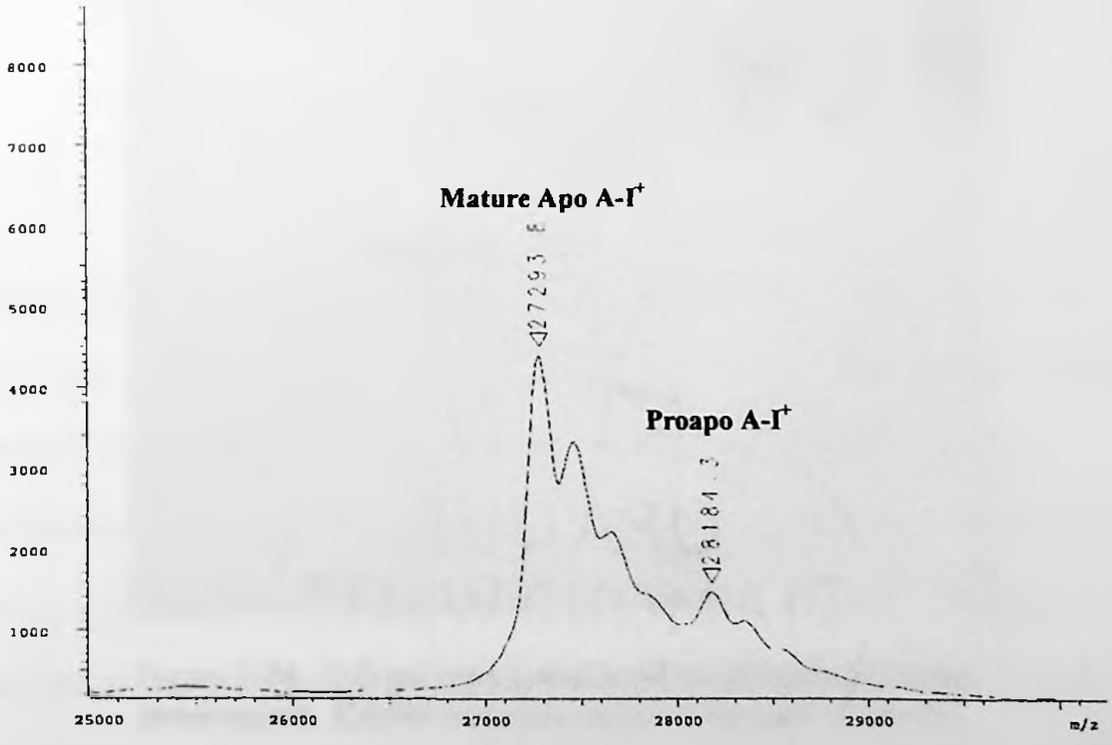


Figure 2-23. Zoomed area of a MALDI spectrum of apo A-I in delipidated Zucker lean serum.

Once again, apo A-I was the most abundant protein in the delipidated serum. Closer inspection of apo A-I found that both the pro and mature forms contained modifications. Again the modifications were ~200 Da apart. Because these modifications were so close in molecular weight, separation was extremely difficult and not achieved with the HPLC.

Further separation was attempted with 2-D electrophoresis. Figure 2-24 is the resulting 2-D gel of the delipidated Zucker obese serum and Figure 2-25 to the delipidated Zucker lean serum. Each of the gels were imaged using a Hewlett Packard ScanJet 3300C.

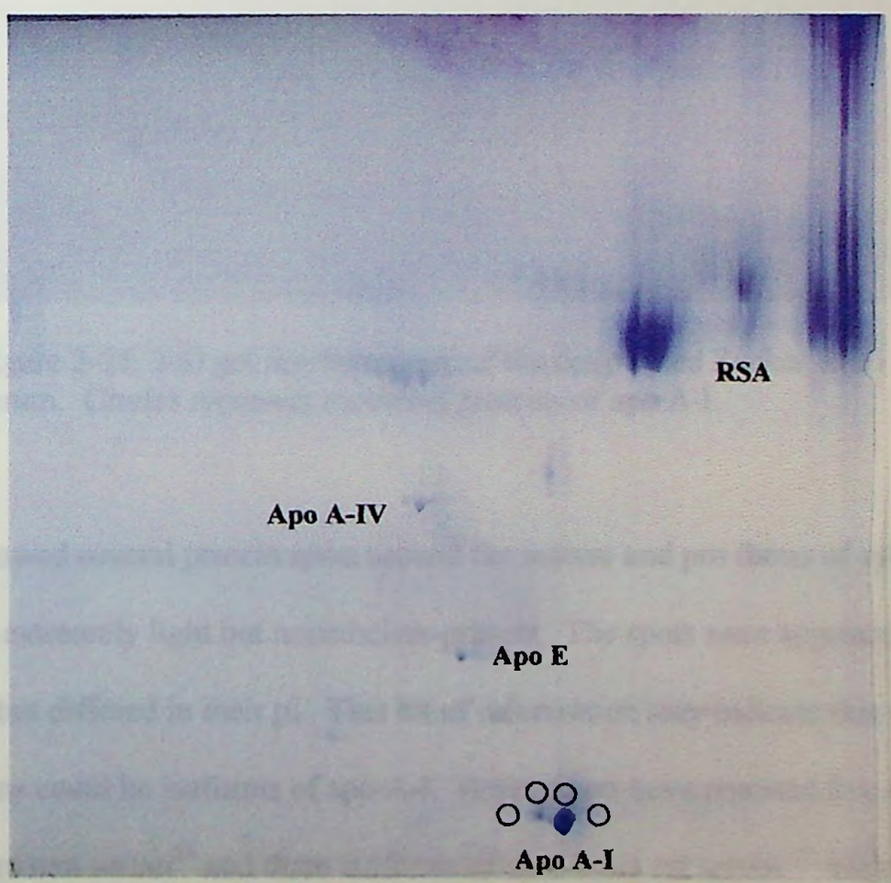


Figure 2-24. 2-D gel representation of the delipidated Zucker obese serum. Circles represent modified proteins of apo A-I.

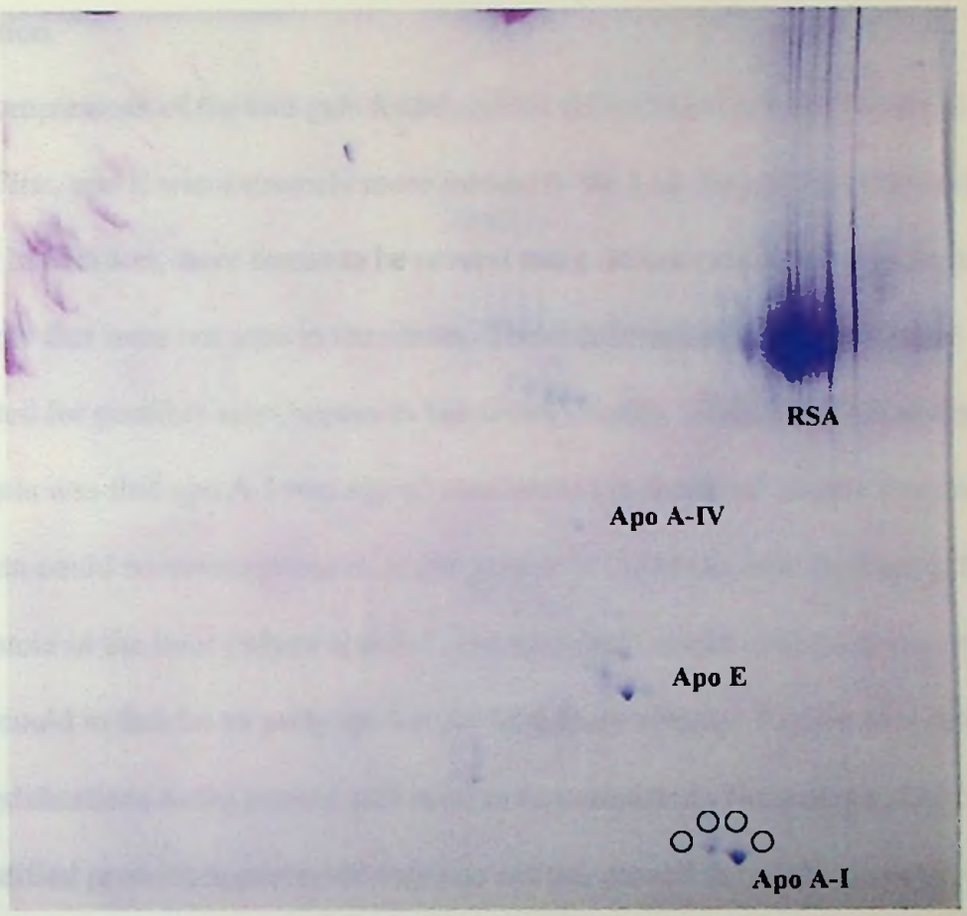


Figure 2-25. 2-D gel representation of the delipidated Zucker lean serum. Circles represent modified proteins of apo A-I.

Each gel showed several protein spots around the mature and pro forms of apo A-I. These were extremely light but nonetheless present. The spots were approximately the same MW, but differed in their pI. This bit of information may indicate that the modifications could be isoforms of apo A-I. Researchers have reported five isoforms of apo A-I in human serum[29] and three isoforms of apo A-I in rat serum.[27] Neither study classified the actual structural differences between the isoforms. So, these different proteins may or may not actually be isoforms at all. The spots on the gel surrounding the

pro and mature proteins were not concentrated enough to extract from the gel for further identification.

Comparisons of the two gels found several differences between the obese and lean serums. First, apo E was extremely more intense in the lean than in the obese delipidated samples. In addition, there seems to be several more differences to this protein in the lean sample that were not seen in the obese. These differences should be further investigated for possible associations to hereditary obesity. Another difference between the two gels was that apo A-I was again more intense in the obese sample than the lean. One reason could be overexpression of the protein in the obese rats. Analyzing the levels of the protein in the liver (where apo A-I is synthesized) would confirm this assumption. Apo A-I could in fact be an early marker for hereditary obesity. Further characterization of the modifications to the protein still need to be completed. Nonetheless, the fact that these modified proteins appeared as separate entities proved they differ in some structural way.

CONCLUSIONS

Proteomic analysis has proved to be a key tool in understanding proteins and their functions. A powerful methodology for the identification of proteins was developed by combining several techniques, such as electrophoresis, HPLC, and mass spectrometry. This methodology was developed using known proteins (Mb, actin, and BSA) in order to achieve optimal results. This approach was successfully applied to an unknown protein to better understand its function and structure.

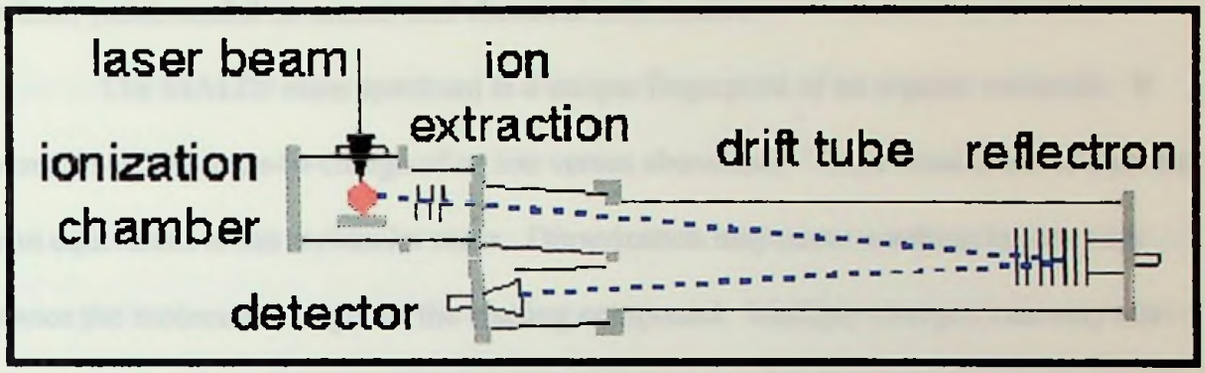
Apolipoprotein A-I was identified and studied using these techniques. This protein was first enzymatically cleaved using trypsin, chymotrypsin, and Lys-C. The peptide map was analyzed by MALDI-TOF and identified using MASCOT, a protein-searching database. The protein (both the pro and mature) was investigated and found in the aqueous layer of Zucker serum. Apo A-I is normally found in the aqueous layer where it is noncovalently bound. A full understanding of apo A-I in the aqueous layer remains elusive. Apo A-I was found to be of greater abundance in the obese rat than in the lean. Further experimentation on the delipidated serum resulted in the appearance of several modification to apo A-I. These modifications were found to be ~200 Da apart. Two-dimensional electrophoresis separated these different forms of the protein from each other, but the modified proteins were not concentrated enough to further identify. Because these different forms were close in MW but differed by pI, these modified proteins may in fact represent isoforms of apo A-I. Again, apo A-I was seen in greater intensity in the obese samples rather than the lean. Apo E was present in greater

abundance in the delipidated lean samples. There were also several modifications of this protein present in the lean serum that were not seen in the obese.

Further experimentation needs to be completed on Zucker serum. A multitude of differences in protein levels has been observed between Zucker rats. This work has only scratched the surface. Apo E should be analyzed further by identifying the modifications seen on the 2-D gel. This particular protein may in fact play some role in hereditary obesity. More investigations need to be done on apo A-I to fully understand why this protein is found in the aqueous layer and what structural modifications are actually present. Because the modifications are only 200 Da apart, separation has been extremely difficult. Further separation techniques could include coupling capillary electrophoresis to the mass spectrometer. Also, accurate molecular weights of the modifications need to be completed. FT/MS instrumentation can be used to gather this information. This high resolution instrument has a much higher number of theoretical plates ($\sim 10^9$) than that of HPLC ($\sim 10^6$).[7] Also identification of the modified proteins represented by MS and 2-D electrophoresis could give further insight into the structural differences which may lead to information on associations of apo A-I to hereditary obesity. Once this is accomplished, experimentation on apo A-I in human serum should be performed.

APPENDIX I

MALDI-TOF Theory



Matrix-Assisted Laser Desorption Ionization Time-of-Flight mass spectrometry (MALDI-TOF) is a soft desorption ionization technique which allows (1) a rapid and accurate mass analysis of biomolecules, (2) the use of low sample amounts (low pmol range), (3) protein identification by peptide mapping, and (4) accurate structural information by post-source decay (PSD). This instrument utilizes a pulsed nitrogen laser beam to ionize a sample. The ion source is maintained at 25 kV. The sample is dissolved in a suitable matrix solvent, forming a sample:matrix crystal. The laser is fired at the sample, resulting in a desorption event, which occurs if the matrix crystals absorb the photon energy of the wavelength of the laser. Energy deposition into the matrix molecules results in their emission of absorbed energy as heat. This process ionizes the sample. These ionized molecules are repelled from the surface (which is at high voltage) and accelerated toward a series of lenses at close to ground, thereby directing them into a field-free drift region. The time-of-flight tube measures the arrival time of ions at the detector. Ions are accelerated from the source with equal kinetic energy. Different

masses travel at different speeds ($E=1/2mv^2$). Ions with lower mass will arrive at the detector ahead of ions of higher mass. The time-of-flight is converted to mass using a calibration function. The resulting mass spectrum represents m/z versus abundance, which yields useful structural and chemical information.

The MALDI mass spectrum is a unique fingerprint of an organic molecule. It consists of the mass-to-charge of an ion versus abundance. There usually exists a parent ion equivalent to the molecular mass. Dimerization may occur resulting in ions with twice the molecular weight of the starting compound. Multiply-charged ions may also exist. For example, doubly-charged ions appear as peaks at half the mass of the parent ion.

Two independent detectors are present in MALDI-TOF: linear and reflectron. The linear detector is located at the end of the flight tube, while the reflectron detector is located at the end of the angled branch near the ion source. In linear mode, the ions travel down a linear flight path. The linear mode of operation is useful for determining molecule weight, especially of very high mass proteins. However, resolution and sensitivity decrease as the size of the protein increases. The mass accuracy in linear mode is expected to be ~0.15% when using an external calibration. In reflectron mode, ions are reflected back at a slight angle towards a detector. This method is useful because it allows MS/MS data to be collected via PSD. The reflectron mode allows detection below m/z ~5000 easily at high resolution for increased mass accuracy, but above m/z 5000 sensitivity decreases significantly. However, this mode is useful for determining the sequence of peptides from a protein digest.[30]

Ions formed by MALDI can undergo metastable decay using PSD, i.e., ions may acquire excess energy from multiple collisions with matrix ions in the source during desorption/acceleration process, resulting in sufficient energy to fragment the ion. Because metastable fragmentation has occurred 'post' source/ionization, the fragment ions will all have the same velocity as their precursor, but only a fraction of its kinetic energy. Therefore, the precursor and fragment ions (both charged and neutral) will arrive at the linear detector at the same time and same m/z. Fragments can be further resolved by exploiting kinetic energy differences through the use of the reflectron. The reflectron focuses ions of limited kinetic energy range. The voltage of the reflectron is decreased over 7-14 steps (bringing fragmentations of lower and lower kinetic energy into focus) until the last spectrum is acquired at 5-10% of the initial voltage. The spectra are then stitched together and calibrated to yield complete fragmentation spectrum. The reflectron has an ion gate just after the focusing lenses that enables low-resolution selection of the precursor ion from the mixture.

APPENDIX II

Troubleshooting Guide for 1-D Electrophoresis

Problem	Cause/Solution
Artifact bands observed at ~65 kDa in reduced samples	Reduce skin keratins – wear gloves to avoid contamination
"Smile" effect: band pattern curves upward at both sides of the gel	Center of the gel running hotter than either side, probably due to excessive power
Vertical streaking	1. Sample overload - dilute sample 2. Sample precipitation – centrifuge sample
Skewed or distorted band; lateral band spreading	1. Excess salts in sample – remove salts by dialysis 2. Excessive pressure applied to the gel plates
Run too fast; poor resolution; gel temperature is high	Running buffers too concentrated
Doublets observed where a single protein species is expected	Protein sample re-oxidized
Observe fewer bands than expected and one heavy band at dye front	1. Gel percentage is too low for MW range of sample 2. Degradation of sample
Observe more bands than expected for purified protein	Sample degradation (proteolysis)
Observe fewer bands than expected for purified protein	Bands are diffusing during staining

APPENDIX III

Molecular Weights and Residue Masses of the Common Amino Acids

Amino Acid	Abbreviations	Molecular Weight	Incremental Mass
Glycogen	Gly, G	75	57
Alanine	Ala, A	89	71
Serine	Ser, S	105	87
Proline	Pro, P	115	97
Valine	Val, V	117	99
Threonine	Thr, T	119	101
Cysteine	Cys, C	121	103
Leucine	Leu, L	131	113
Isoleucine	Ile, I	131	113
Asparagine	Asn, N	132	114
Aspartic acid	Asp, D	133	115
Glutamine	Gln, Q	146	128
Lysine	Lys, K	146	128
Glutamic acid	Glu, E	147	129
Methionine	Met, M	149	131
Histidine	His, H	155	137
Phenylalanine	Phe, F	165	147
Arginine	Arg, R	174	156
Tyrosine	Tyr, Y	181	163
Tryptophan	Trp, W	204	186

APPENDIX IV

Troubleshooting Guide for 2-D Electrophoresis[24]

Problem	Cause/Solution
Non-reproducible gel patterns	Due to poor ampholytes or chemicals
Gaps in pattern associated with IEF dimension	1. Due to poor ampholytes or acrylamide 2. Breaks in the IEF gel when loading
Pattern occupies only a portion of the expected 18 cm length of the IEF gel	Isoelectric focusing was not performed for the necessary length of time
All proteins appear to be multiplied in IEF dimension	Due to carbamylation of sample in urea – never heat sample in urea
Tailing or streaking of all proteins associated with 1-D gels	Samples precipitated during IEF, poor sample solubilization – add 10 mM CHAPS to 1-D gel
Poorly resolved high MW proteins. Most of the protein remains at the basic origin of the gel during IEF	Improper sample solubilization – centrifuge samples prior to application to IEF gel
Proteins are doubled or streaked in the second dimension	Due to ion depletion of the upper running buffer
Two bands appear at 55 kDa and 69 kDa across the entire pattern upon staining	Due to epidermal keratin contamination from hands or hair
Low MW proteins are missing or greatly reduced	IEF gels equilibrated too long – equilibrate 2 min

High MW proteins are missing or greatly reduced	Proteolysis during sample preparation
High silver stain background	Due to low quality water and dirty equipment
Pattern bows out at either vertical edge of the gel	Due to electrical leakage through the spacers
Greater than expected sieving in MW dimension. Gel runs very quickly	The 1.5 M Tris gel buffer is too basic
Less than expected sieving in MW dimension. Gel runs very slowly	The 1.5 M Tris gel buffer is too acidic
Some proteins migrate differently form run to run, while other migrate consistently from run to run	Make sure 2-D runs are always performed at the same temperature

REFERENCES

1. James, P. *Biochem. Biophys. Res. Commun.* **1997**, *231*, 1-6.

2. Rabilloud, T., ed. *Proteome Research: Two-Dimensional Gel Electrophoresis and Identification Methods*; Springer: New York, **2000**.

3. Humphery-Smith, I.; Blackstock, W. *J. Protein Chem.*, **1997**, *16*, 537-544.

4. Grevaert, K.; Vandekerckhove, J. *Electrophoresis* **2000**, *21*, 1145-1154.

5. "The Human Genome Project" *MCFA News* **2000**, 32.

6. Devlin, T., ed. *Textbook of Biochemistry with Clinical Correlations, 4th ed.*; Wiley-Liss: New York, **1997**.

7. Skoog, D.; Holler, F.J.; Nieman, T. *Principles of Instrumental Analysis, 5th ed.*; Harcourt Brace: Philadelphia, **1998**.

8. Nelson, D. L.; Cox, M. M. *Lehninger Principles of Biochemistry, 3rd ed.*; Worth: New York, **2000**.

9. Sorell, T. *Organic Chemistry*; University Science Books: Sausalito, **1999**.

10. Barker, J. *Mass Spectrometry, 2nd ed.*; Wiley: West Sussex, **1999**.

11. "MALDI MS" *ABRFNews* June **1997**.

12. http://www.crchul.ulaval.cal/crchul/en/serv/sspeg_protocol2.asp

13. Bio Rad Laboratories, Inc. *Ready Gel Cell Instruction Manual*; Hercules, **1996**.

14. Sigma-Aldrich, Inc. *Product Information – Silver Stain Kit for Polyacrylamide Gels*; St. Louis, **1999**.

15. Millipore Corp. *User Guide for Reversed-Phase ZipTip*; Bedford, **1999**.

16. http://www.obesity.org

17. Jordan, C.; Rafaela, E. *WV Family Magazine*; **2000**, 65.

18. http://www.coloradohealthnet.org/obesity

19. Voet, D.; Voet, J. G.; Pratt, C.W. *Fundamentals of Biochemistry*; Wiley: New York, **1999**.

20. http://aci.mta.ca/Courses/Biochemistry

21. Schuller, E.; Patel, N.; Item, C. *Int. J. of Obesity* **2000**, *24*, 460-467.

22. Amy, R. M. *Atherosclerosis* **1988**, *69*, 199-209.

23. Gordon, J. I.; Smith, D. P.; Andy, R. *J. Biol. Chem.* **1982**, *257*, 971-978.

24. Genomic Solutions, Inc. *Investigator 2-D Electrophoresis System Manual*; Chelmsford, **1999**.

25. AmiKa, Inc. *Dispo-Biodialyzer Manual*; Elsboro, **2001**.

26. Eberini, I. Personal Communication. **2001**.

27. Dolphin, P. J. *FEBS Letters* **1980**, *117*, 252-258.

28. http://www.embl-heidelberg.de/~toldo/JaMBW/3/1/6/

29. Contiero, E.; Ferrari, R.; Vaselli, G. M. *Electrophoresis* **1997**, *18*, 122-126.

30. Bruker Daltonics, Inc. *Beginners Training Manual for Understanding the Principles of MALDI TOF MS*; Billerica, **1998**.