

2017

Protein Found at the Scene of the Crime: The Potential for Using Proteomics for Identification

Gavin R. Tisdale

Follow this and additional works at: https://opencommons.uconn.edu/law_student_papers

 Part of the [Criminal Law Commons](#), and the [Evidence Commons](#)

Recommended Citation

Tisdale, Gavin R., "Protein Found at the Scene of the Crime: The Potential for Using Proteomics for Identification" (2017).
Dissertations and Honors Papers. 2.
https://opencommons.uconn.edu/law_student_papers/2

*Protein Found at the Scene of the Crime: The Potential for Using
Proteomics for Identification*

Gavin R. Tisdale

Hair has long been collected from crime scenes as part of trace evidence. Originally, hair was used for some exclusionary purposes—only general qualities about an unknown source could be determined. Eventually, DNA was used to help identify the source but only if the root was still attached. Within the last two years, however, two major studies have used proteomics—the study of human protein sequences—to extract and identify protein sequences in an unknown source in order to match it to a known source. These two studies support the same hypothesis: proteomics is currently a viable method for narrowing down the source of the hair and will soon be able to identify an individual source. While the science is about a decade away from being comparable to nuclear DNA, the potential of proteomics is undeniable. This paper explores the current status, methods, and future of the science as well as the impact it may have on forensic investigation and criminal prosecution. This paper also explores the potential admissibility of expert testimony on proteomics for identification.

INTRODUCTION

A decade from now, an expert will be able to take the witness stand, and, at the conclusion of her testimony, she will be able to say, “Based on a reasonable degree of scientific certainty, there is less than a one-in-an-eight-billion possibility that the half-inch hair fragment found at the murder scene matches someone other than the defendant.” Today, someone in forensics may scoff at such a conclusion. Usually, hair can only be used for limited purposes as circumstantial evidence: What is the ethnicity of its source? What color hair does its source have? Does its source use chemical hair treatments? Identification from hair can only be done using nuclear DNA analysis, which is impossible with a hair fragment because the hair root must be attached to perform nuclear DNA analysis.¹

Enter proteomics—the study of human protein sequences. Building off the Human Genome Project of the early-2000s, scientists are attempting to sequence the entire human proteome.² Though proteomics has been used for other scientific endeavors, such as cancer research and pharmacology, two recent studies have opened the door for the use of proteomics in forensic science as a method of identifying the source of a hair.

This paper begins with an overview of proteomics, including a brief history. Section II then dives head first into the science, outlining the basic steps of proteomic experiments. Section III discusses how the basic science has evolved into using proteomics for identification. This section also includes a description of the two primary studies that have been performed to date on the use of proteomics for identification purposes. Section IV uses a hypothetical to evaluate

¹ Nancy Mendoza, OPENDEMOCRACY (Nov. 7, 2002), https://www.opendemocracy.net/arts-hair/article_721.jsp. Hair that merely falls out is dead hair, which does not include the root and is not amenable to DNA analysis. *Id.* Mitochondrial DNA can be performed on the shaft, but it can only be used to narrow the source to about one in one hundred. *Id.*

² *History of Proteomics*, PROTEOME SCIENCES, <http://www.proteomics.com/about-us/company/history-of-proteomics> (last visited Oct. 25, 2016).

the admissibility of expert testimony on proteomics in a criminal trial under *Daubert*. Section V discusses how such testimony, if admitted, should be limited and can be attacked. Section VI then gives a few recommendations as to how the forensic community and lawyers can prepare for the eventual use of this type of evidence.

I. WHAT IS PROTEOMICS?

This section will provide some background into the basics of proteins and proteomics before giving a brief history of developments in proteomics and its current uses.

A. *Protein Basics*

Chemists in the nineteenth century understood the importance of proteins—so much so, that the term itself stems from the Greek *proteios*, which means “holding first place.”³ At its most basic, proteins are series of amino acids held together by peptide bonds.⁴ Although amino acids have a specific scientific definition,⁵ they can most easily be understood as a naturally occurring compound that are used as the building blocks of proteins.⁶ There are twenty amino acids that are used to create proteins in the human body.⁷ A peptide is a group of two or more amino acids.⁸ Peptides are bound together by peptide bonds, which are the chemical bonds

³ *Protein*, ENCYCLOPEDIA BRITANNICA, <https://www.britannica.com/science/protein> (last visited Dec. 3, 2016).

⁴ *Protein*, MERRIAM WEBSTER, <http://www.merriam-webster.com/dictionary/protein> (last visited Oct. 28, 2016).

⁵ Specifically, amino acids include “any of a class of organic compounds that contains at least one amino group, -NH₂, and one carboxyl group, -COOH: the alpha-amino acids, RCH(NH₂)COOH, are the building blocks from which proteins are constructed.” *Amino Acid*, DICTIONARY, <http://www.dictionary.com/browse/amino-acid> (last visited Dec. 1, 2016).

⁶ *Id.*

⁷ *What Types of Amino Acids Are There?*, AMINO ACID STUDIES, <http://www.aminoacid-studies.com/amino-acids/what-types-of-amino-acids-are-there.html> (last visited Dec. 2, 2016).

⁸ *Peptide*, MERRIAM WEBSTER, <http://www.merriam-webster.com/dictionary/peptide> (last visited Dec. 1, 2016).

between the carbon terminus (C-terminus) and nitrogen terminus (N-terminus) of an amino acid.⁹ A peptide sequence is typically described, or read, starting with the N-terminus, which is indicated on a peptide chain by an ammonia-based compound (referred to as the free amine group), and ending with the C-terminus, which is indicated on a peptide chain by a carbon-based compound (referred to as the free carboxyl group).¹⁰

From these twenty amino acids, there are hundreds-of-thousands of proteins that can be created as part of the human body.¹¹ Proteins are both specie and organ specific, meaning that the proteins found in one specie differ from those in another and the proteins in one organ in a species differs from proteins in an organ in the same species.¹² The concentration of proteins varies in different parts of the human body; for example, muscle and red blood cells contain about 30% protein, but hair and other tissues with low water content contain higher percentages.¹³

Protein synthesis, or the making of proteins in a cell, happens when the deoxyribonucleic acid (DNA) copies itself into ribonucleic acid (RNA)—a process called transcription.¹⁴ The RNA, which is commonly referred to as a gene, then moves to the ribosome, which is the area in the cell in charge of making proteins.¹⁵ Once in the ribosome, the structure of the RNA decides

⁹ *Peptide Bond*, MERRIAM WEBSTER, <https://www.merriam-webster.com/dictionary/peptide%20bond> (last visited Dec. 5, 2016).

¹⁰ *The Genetic Code*, MICH. ST. UNIV., <https://msu.edu/course/lbs/149h/TRANSLATION.html> (last visited Dec. 3, 2016).

¹¹ PROTEOME SCIENCES, *supra* note 1.

¹² *Protein*, *supra* note 2.

¹³ *Id.*

¹⁴ *Protein Synthesis*, ELMHURST COLL., <http://chemistry.elmhurst.edu/vchembook/584proteinsyn.html> (last visited Dec. 3, 2016).

¹⁵ *Id.*

which amino acid will be added to the protein chain—a process called translation.¹⁶ Amino acids are added, one at a time, to the protein until the RNA chain passes through the ribosome.¹⁷

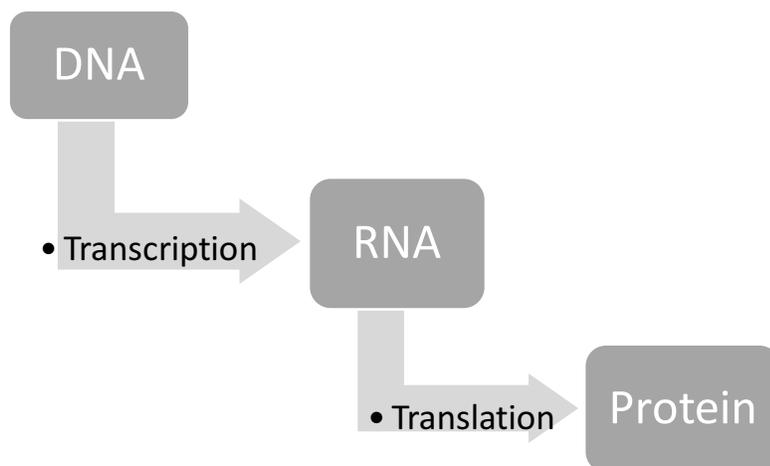


Figure 1: This flowchart shows how DNA is first transcribed into RNA and then translated into protein.¹⁸

Once proteins are developed, they perform varying critical roles in the functioning of an organism.¹⁹ For example, antibodies, such as immunoglobulin, are made up of proteins and help fight viruses and bacteria in the human body.²⁰ Most obviously, proteins form muscles, which allow the body to move, the heart to beat, and lungs to breath.²¹

With the basics of proteins, including what they are, how they are made, and what they are used for, in mind, let us now turn to the basics of proteomics.

¹⁶ *Id.*

¹⁷ *Id.*

¹⁸ Sara J. Tharp-McComas, *The Central Dogma*, <http://kmbiology.weebly.com/index.html> (last visited Dec. 3, 2016).

¹⁹ *What Are Proteins and What Do They Do?*, GENETICS HOME REF., <https://ghr.nlm.nih.gov/primer/howgeneswork/protein> (last visited Dec. 3, 2016).

²⁰ *Id.*

²¹ *Id.*

B. *Proteomics, Generally Speaking*

Academics define proteomics in accordance with their specific purposes. For example, a pharmacologist would define proteomics as the study of “proteins in search[] [of] proteins that are associated with a disease by means of their altered levels of expression.”²² But generally speaking, proteomics is the study of proteomes, which are the “total set of proteins expressed during the lifetime of a cell.”²³ More simply, proteomics is the study of every type of protein that a cell or an organism can produce.

For the purposes of proteomics for identification, proteomics is the study and identification of all of the proteins that *Homo sapiens* can produce.²⁴ The basic concept of proteomics for identification has two primary aspects. The first is that forensic scientists will be able to identify the abundance of certain proteins, called “protein markers,” from a tissue sample, such as a hair.²⁵ Second, scientists will also be able to calculate the probability that these protein markers will be found in an individual by taking random samplings from a given population.²⁶ From there, a statistical analysis can be performed to figure out the likelihood that a hair from a known individual contains the same relative abundance of certain protein markers as a hair from

²² PROTEOME SCIENCES, *supra* note 1.

²³ J. Jason Williams, *Protecting the Frontiers of Biotechnology Beyond the Genome: The Limits of Patent Law in the Face of the Proteomics Revolution*, 58 VAND. L. REV. 955, 962 (2005) (citing P.C. TURNER ET AL., INSTANT NOTES IN MOLECULAR BIOLOGY (2d ed. 2000)).

²⁴ *Infra* Section I.b.

²⁵ The most common protein markers in hair are keratin based. Mendoza, *supra* note 1. Scientists do not need to identify every single protein in the human body to perform proteomics for identification; they just need to identify enough proteins found in human hair to establish statistical uniqueness. Likely, they will continue to identify as many keratin-based proteins as possible, which will allow them to draw conclusions regarding the uniqueness of a piece of hair.

²⁶ *Infra* Section III.

an unknown source.²⁷ From these basic principles, scientists are attempting to develop a science that will parallel, and possibly surpass, DNA for identification.²⁸

C. A Brief History of Protein and Proteomics

In the early nineteenth century, the first proteins were identified by Swedish chemist Jöns Jakob Berzelius, who sought to identify the large groups of macromolecules that were made up of long chains of amino acids and found in living organisms.²⁹ Berzelius is recognized as one of the founders of modern chemistry, and his identification of proteins was a footnote in a long career filled with other notable accomplishments, such as determining atomic weight, developing atomic symbols, and the developing analytic techniques.³⁰

After their initial discovery, several scientists made major headway in the study of proteins. Gerhardus Johannes Mulder studied the elemental composition of proteins and hypothesized that all proteins were comprised of carbon, hydrogen, nitrogen, and oxygen with varying amounts of phosphorus and sulfur.³¹ In 1819, the first amino acid was isolated from a protein.³² It would take until 1936 to identify the twentieth naturally occurring amino acid.³³ In

²⁷ *Id.*

²⁸ *Infra* Section V.

²⁹ Dr. Tomislav Mestrovic, *What is Proteomics?*, NEWS MEDICAL, <http://www.news-medical.net/life-sciences/What-is-Proteomics.aspx> (last visited Sept. 25, 2016). There is, however, some debate as to whether Berzelius was first to discover protein or just the first to name it. Some believe that Mulder was the first scientist to really study protein, though he did not give it a name.

³⁰ *Jöns Jakob Berzelius*, ENCYCLOPEDIA BRITANNICA, <https://www.britannica.com/biography/Jons-Jacob-Berzelius> (last visited Dec. 3, 2016).

³¹ *Gerardus Johannes Mulder*, ENCYCLOPEDIA, <http://www.encyclopedia.com/people/history/historians-miscellaneous-biographies/gerardus-johannes-mulder> (last visited Dec. 10, 2016).

³² CHARLES TANDFORD & JACQUELINE REYNOLDS, *NATURE'S ROBOTS – A HISTORY OF PROTEINS* 30 (OXFORD UNIV. PRESS 2001).

³³ *Id.*

1902, Emil Fisher and Franz Hofmeister discovered the peptide bond.³⁴ Further, in 1925, N. Bjerrum, E. Q. Adams, K. Linderstrom-Lang, and others were able to discover that amino acids have certain isoelectric points,³⁵ a topic discussed in greater detail, *infra*.

In the early 1950s, scientists Herman Branson and Robert Corey published seven research papers in one issue of *Proceedings of the National Academy of Sciences of the United States of America*.³⁶ Their work focused on the structure of protein, and it was groundbreaking—they were the first to successfully identify the helix structure of proteins.³⁷ The story goes that Pauling was bed-ridden with a cold, and to amuse himself, he drew chemical structures on paper.³⁸ While playing with the papers, he twisted one of the papers into spiral form, and the helix structure of proteins was then obvious to him.³⁹

Around the same time, Fred Sanger was the first to successfully sequence the arrangement of amino acids in protein.⁴⁰ This discovery was significant because, up to that time, scientists were unsure of whether protein was a composition of mixed substances or the same substance.⁴¹ Sanger's discovery indicated that each molecule of a protein was made of the same chemical substance.⁴²

³⁴ *Id.* at 31.

³⁵ *Id.* at 65.

³⁶ Christen Brownlee, *The Protein Papers*, PNAS, <http://www.pnas.org/site/classics/classics1.xhtml> (last visited Dec. 3, 2016).

³⁷ *Id.*

³⁸ *Id.*

³⁹ *Id.*

⁴⁰ Francis Sanger, *The Arrangement of Amino Acids in Proteins*, 7 *ADV. PROTEIN CHEM.* 1 (1952).

⁴¹ Antony O. W. Stretton, *The First Sequence: Fred Sanger and Insulin*, 162 *GENETICS* 527 (2002).

⁴² *Id.*

Building upon the work of these scholars, the first attempt to sequence the entire protein composition of an organism occurred in 1975.⁴³ The study mapped the proteins present in small mammals, such as guinea pigs and mice, and in the bacterium, *Escherichia coli* (commonly referred to as *E. Coli*).⁴⁴ The terms “proteome” and “proteomics” were coined in the early 1990s by an Australian student named Marc Wilkins.⁴⁵ The name was designed to mirror the terms “genome” and “genomics,” which is the study, mapping, and identification of genes.⁴⁶

The Human Genome Project was a scientific undertaking to identify and map all of the genes present in our species, *Homo sapiens*.⁴⁷ The Project began on October 1, 1990 and took until April of 2003 to complete.⁴⁸ Upon its completion, scientists were able to, for the first time, identify the entire genetic blueprint of *Homo sapiens*—a total of 19,599 genes.⁴⁹ As explained above, genes dictate which type of protein a cell produces, and the Human Genome Project was a catalyst to reinvigorating interest in proteomics.

Following the completion of the Human Genome Project, an effort for a similarly complete study of proteins was undertaken for a second time.⁵⁰ The first study began shortly after the 1975 study in mammals and bacteria, but lack of funding and limitations in the ability to

⁴³ Mestrovic, *supra* note 29.

⁴⁴ *Id.*

⁴⁵ *Id.*

⁴⁶ *Id.*; *Genomics*, MERRIAM WEBSTER, <http://www.merriam-webster.com/dictionary/genomics> (last visited Dec. 3, 2016).

⁴⁷ *All About the Human Genome Project*, NATIONAL HUMAN GENOME RESEARCH INST., <https://www.genome.gov/10001772/all-about-the--human-genome-project-hgp/> (last visited Oct. 11, 2016).

⁴⁸ *Id.*; PROTEOME SCIENCES, *supra* note 22.

⁴⁹ NATIONAL HUMAN GENOME RESEARCH INST., *supra* note 47; PROTEOME SCIENCES, *supra* note 22.

⁵⁰ PROTEOME SCIENCES, *supra* note 22.

identify individual proteins halted the development.⁵¹ The second study (post-Human Genome Project), even after the developments in genomics, would prove far more complex; for more than *one million* proteins and protein fragments may be identified.⁵² Apart from the plethora of identifiable proteins, proteomics is difficult to study because cells can create different proteins at different times and the concentration of proteins in a given mammalian cell can vary from one to one-hundred thousand.⁵³ Proteomics is also limited by more than its mere complexity: it can take over \$100,000 and several years to identify a single protein.⁵⁴

D. *Proteomic Uses*

These costs might have practically prohibited the development of proteomics, but its application, much like that of genomics, can be quite valuable. Proteomics was used in a major medical malpractice case to identify whether an HPV drug was the cause of heart failure.⁵⁵ It is also being used to identify and study the development of cancer, such as breast cancer.⁵⁶ In its more lucrative form, proteomics is being used for pharmacology.⁵⁷ With applications like these, proteomics has continued to get the requisite funding, and projects like the Swedish Protein Human Protein Atlas have identified over 400,000 proteins.⁵⁸ The vast number of potential

⁵¹ Paul R. Graves et al., *Molecular Biologist's Guide to Proteomics*, 66 MICROBIAL MOLECULAR BIOLOGY REV. 39 (2002), <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC120780/>.

⁵² PROTEOME SCIENCES, *supra* note 22.

⁵³ Kondethimmanahalli Chandramouli & Pei-Yuan Qian, *Proteomics: Challenges, Techniques and Possibilities to Overcome Biological Sample Complexity*, HUMAN GENOMICS & PROTEOMICS (2009), <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2950283/>.

⁵⁴ Keala Chan & Dennis Fernandez, *Patent Prosecutions in Proteomics*, 19 SANTA CLARA COMP. & HIGH TECH. L.J. 457, 461 (2003).

⁵⁵ *Tarsell v. Sec. of Health and Human Serv.*, No. 10-251V, 2016 WL 880223 (Ct. Fed. Claims Feb. 16, 2016).

⁵⁶ JD Wulfkuhle et al., *New Approaches to Proteomics Analysis of Breast Cancer*, 1 PROTEOMICS 10 (2001).

⁵⁷ Graves et al., *supra* note 51.

⁵⁸ *The Human Proteome*, HUMAN PROTEOME ATLAS, humanatlas.org (last visited Oct. 11, 2016).

proteins in the human proteome is due primarily to the large number of basic building blocks, amino acids. Whereas four nucleotides—the building block for DNA—make up about 20,000 identified human genes, the twenty amino acids can create exponentially more variation among proteins. The purpose of this paper, however, is to discuss the emerging use of proteomics for identification.⁵⁹

II. THE SCIENCE

Though the focus of this paper is on the use of proteomics for identifying people, it is important to understand how basic proteomic experiments are performed. These experiments establish the basic principles used in the tests for identification and are part of the reason why some are so confident about the application of proteomics for forensic identification. Note that many steps in a basic experiment are manual and have certain levels of error or limitations. If an expert was explaining these steps in order to give a scientific opinion, opposing counsel may wish to attack some of these limitations or potential errors—a topic more fully discussed *infra*. Further, the current “best practices” for proteomics for identification are based in large part on the basic methods explained below.

A. Major Breakthroughs

Several scientific developments have allowed for massive progress in proteomics in the last two decades. The first major breakthrough was the development of micro-sequencing techniques for electroblotted proteins, which is a process which the proteins could be readily sequenced once they were separated.⁶⁰

⁵⁹ Discussed *infra* Section III.

⁶⁰ Graves et al., *supra* note 51. This process is more fully described *infra*.

Technological developments in Mass Spectrometry (“MS”) have also allowed for rapid progress.⁶¹ MS is an instrument used for identifying the chemical composition of a substance.⁶² The chemical composition is identified by measuring the mass of ionized atoms or molecules.⁶³ This process will be discussed in more detail, *infra*. The sensitivity of the calculations and accuracy of the results from MS have increased by multiple factors in the 1990s, and it continues to develop today.⁶⁴ The more sensitive and accurate the results, the more precise scientists can be with their conclusions.

B. *Proteomic Experiment Process*

A basic proteomic experiment can be broken into three stages: (1) separation of the proteins from the sample (cell, tissue, or organism); (2) acquisition of information for identification and characterization; and (3) database utilization.⁶⁵ Each of these stages will be discussed, in turn.

i. Separation and Isolation

As was described earlier, proteomics can involve complex mixtures of proteins created within a single cell.⁶⁶ In order to successfully identify each type of protein, the mixture must be separated. A process called polyacrylamide gel electrophoresis (“PGE”) was created over fifty years ago, and unlike other areas of proteomics, there have been no major breakthroughs to replace or improve this tedious process.⁶⁷

⁶¹ *Id.*

⁶² *Mass Spectrometry*, MERRIAM WEBSTER, <http://www.merriam-webster.com/dictionary/mass%20spectrometry> (last visited Sept. 30, 2016).

⁶³ *The Mass Spectrometer*, CHEM GUIDE, <http://www.chemguide.co.uk/analysis/masspec/howitworks.html> (last visited Sept. 30, 2016).

⁶⁴ Graves et al., *supra* note 51.

⁶⁵ *Id.*

⁶⁶ *Id.*

⁶⁷ *Id.*

The most common form of PGE for proteomics is called two-dimensional gel electrophoresis (commonly referred to as “2-DE”).⁶⁸ The first step in this process is that the protein mixture⁶⁹ is placed into a gel with sodium dodecyl sulfate (“SDS”) that gives the molecules a net negative charge across a pH gradient, with low pH levels on one side and higher pH levels on the other.⁷⁰ An electric current is placed on one side of the mixture, and the proteins move to where their isoelectric point—where the molecule carries no net electrical charge—matches the pH level.⁷¹ The result is a one-dimensional spectrum of molecules along an axis (the X Axis) according to the relative isoelectric points of proteins.⁷² If the process stopped here, it would be considered one-dimensional gel electrophoresis.

To complete the second stage of 2-DE, the process is repeated with a polyacrylamide gel (“PAGE”), but the charge is placed on a side perpendicular to the first charge (the “Y Axis”).⁷³ This purpose of the second phase is to further separate the proteins by relative weight, giving a more accurate and particular spectrum of proteins.⁷⁴ The weight is measured in kiloDaltons (“kDA”), each of which is defined as the mass of a single nucleon (a proton or neutron).⁷⁵ The

⁶⁸ *Id.*

⁶⁹ At this point, several steps have been applied to a sample in order to extract the proteins from the sample. The process for extracting the proteins from the sample are discussed in more detail in the context of hair, *infra*. Basically, a series of detergents and chemicals are applied to a hair to create a mush of proteins, which allow for the proteins to separate more easily.

⁷⁰ *Introduction to SDS-PAGE*, EXPERIMENTAL BIOSCIENCES, <http://www.ruf.rice.edu/~bioslabs/studies/sds-page/gellab2.html> (last visited Sept. 30, 2106).

⁷¹ *Introduction to Proteomics*, HUMAN PROTEOME ORGANISATION, http://biol.lf1.cuni.cz/ucebnice/pdf/Introduction_to_Proteomics.pdf (last visited Sept. 30, 2016).

⁷² *Id.*

⁷³ *Id.*

⁷⁴ *Id.*

⁷⁵ *MCB 150 Frequently Asked Questions*, UNIV. of Ill., <http://www.life.illinois.edu/mcb/150/private/faq/index.php?action=artikel&cat=13&id=842&artlang=en> (last visited Sept. 30, 2016).

average weight of an amino acid is 135 Daltons.⁷⁶ When these two stages are used together, it is also referred to as “SDS-PAGE.”⁷⁷

The results of protein analysis by SDS-PAGE show groupings of proteins in different locations associated with particular, identifiable weights and isoelectric points.⁷⁸ As the below figure shows, each dark spot on the graph corresponds with an isoelectric point (the x-axis denoted by the pH scale) and weight (the y-axis denoted by kiloDaltons). For example, the grouping marked by the number two on the graph has an approximate weight of seventy kDa and an approximate isoelectric point of pH4.9.⁷⁹

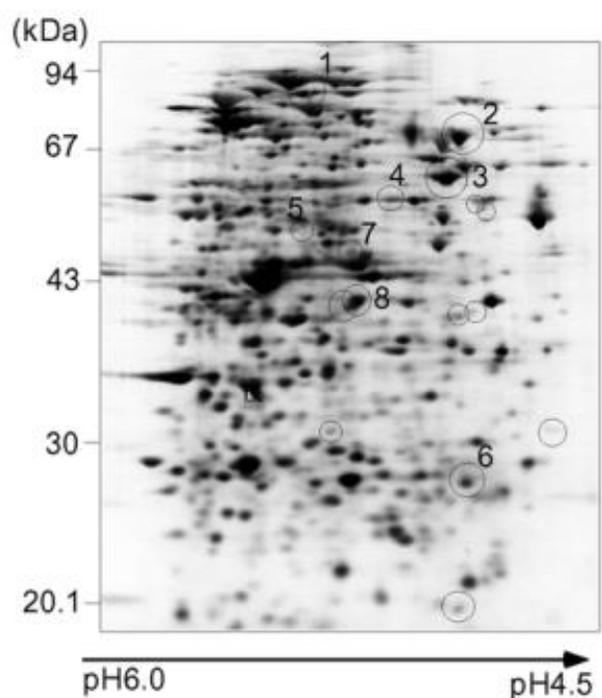


Figure 2: This picture shows the result of a 2-DE test on *E. Coli*.⁸⁰ To some, it may appear as a Rorschach test, but to microbiologists and chemists, the result is an orderly separation of what was once a complex mixture of proteins.

⁷⁶ *Id.*

⁷⁷ HUMAN PROTEOME ORG., *supra* note 71; EXPERIMENTAL BIOSCIENCES, *supra* note 70.

⁷⁸ EXPERIMENTAL BIOSCIENCES, *supra* note 70.

⁷⁹ Fig. 2, *infra*.

⁸⁰ Yuko Yamaguchi et al., *Effects of Disruption of Heat Shock Genes on Susceptibility of Escherichia Coli to Fluoroquinolones*, 3 BMC MICROBIOLOGY 16 (2003),

Using such a graph, scientists can evaluate the relative abundance of a protein and the relative distribution of proteins within a sample.⁸¹ Just by looking at Figure 2, a lay person can see that the grouping marked as number two is more abundant (darker) than the grouping marked as number six. Once the SDS-PAGE test is complete, scientists can then label the different groupings by chemical or radioactive means and compare the results with other control or sample protein mixtures.⁸²

There are, however, several limitations to SDS-PAGE. First, it remains a timely and labor intensive process because there have been no successful attempts at automation.⁸³ A typical test takes two days and is limited to a single sample.⁸⁴ Moreover, in the case of complex protein mixtures, multiple tests have to be run because the spectrum of proteins, in both weight and isoelectric points, is too wide for one test.⁸⁵ Although accuracy has improved, the process still fails to identify proteins in low abundance because more abundant proteins can dominate a sample.⁸⁶

ii. Identification and Characterization

There are two major methods for identification and characterization of proteins. The first is called “Edman sequencing.”⁸⁷ Edman sequencing was one of the earliest methods of protein identification.⁸⁸ Even though use of this process is declining in proteomics, some researchers

https://www.researchgate.net/publication/10619062_Effects_of_disruption_of_heat_shock_genes_on_susceptibility_of_Escherichia_coli_to_fluoroquinolones.

⁸¹ *Id.*

⁸² HUMAN PROTEOME ORG., *supra* note 71.

⁸³ Graves et al., *supra* note 51.

⁸⁴ *Id.*

⁸⁵ *Id.*

⁸⁶ *Id.*

⁸⁷ *Id.*

⁸⁸ *Id.*

still use it because it is a more efficient means of sequencing large numbers of proteins. The second method is the use of MS to identify the individual proteins.⁸⁹ As was discussed *supra*, this method is growing in accuracy and sensitivity, making it the preferred method for contemporary proteomics. But because so much of the original sequencing was done using Edman sequencing, each of the methods will be discussed briefly.

(1) Edman Sequencing

Edman sequencing (also called Edman degradation) is a process by which peptides—groups of two or more amino acids—are sequenced by removing amino acids one-at-a-time from the N-terminus—the start of a protein identified by a free amine group.⁹⁰ Each protein is removed, one-by-one, by adding phenyl isothiocyanate, a commercially-available reagent, which reacts with the N-terminus and “cleaves” the peptide at the end of the chain while leaving the rest of the protein strand intact.⁹¹ When the amino acids are separated, they are identified through chromatography⁹² or electrophoresis.⁹³ One way to think about Edman sequences is to picture a complex protein as a string of colored beads. When the reagent, phenyl isothiocyanate, is added to the string of beads, only one bead is removed at a time. The color of this bead can then be identified. By taking each bead, we can then specify the exact order of the colored beads and compare them to other strings of beads.

⁸⁹ *Id.*

⁹⁰ *Id.*

⁹¹ JM Berg et al., *Amino Acid Sequences Can Be Determined by Automated Edman Degradation in BIOCHEMISTRY* § 4.2 (5th ed. 2002), <https://www.ncbi.nlm.nih.gov/books/NBK22571/>.

⁹² Chromatography is a technique used for separating proteins based on the “relative amounts of each solute distributed between a moving fluid stream and a contiguous stationary phase.” *Chromatography, Encyclopedia Britannica*, <https://www.britannica.com/science/chromatography> (last visited Sept. 30, 2016).

⁹³ Electrophoresis is “the movement of electrically charged particles in a fluid under the influence of an electric field.” *Electrophoresis, Encyclopedia Britannica*, <https://www.britannica.com/science/electrophoresis> (last visited Sept. 30, 2016).

Today, a more advanced method of sequencing requires that the proteins be transferred to a membrane by a process called electroblotting.⁹⁴ Electroblotting essentially transfers and preserves the results of SDS-PAGE, as shown in Figure 2, from the gel to a thin layer of nitrocellulose (the membrane).⁹⁵ Once on the membrane, the protein, especially the less abundant proteins, can be analyzed more accurately and effectively because the gel used to separate the protein no longer inhibits scientists from removing specific samples.⁹⁶ These samples are then tested using the above methods of chromatography or electrophoresis.⁹⁷

(2) Mass Spectrometry

The MS method for identifying and characterizing proteins occurs in four steps. First, an atom or molecule from the protein sample is ionized.⁹⁸ Ionization is the process in which an electronically neutral atom or molecule is given an electric charge—in this case a positive charge.⁹⁹ Second, the charged molecules, called ions, are accelerated in a vacuum so that they all have the same kinetic energy.¹⁰⁰ Kinetic energy is a “form of energy that . . . a particle has by reason of motion.”¹⁰¹ The faster a particle moves, the more kinetic energy it has.¹⁰² Third, a magnetic field is used to deflect the ions, with the lighter ions being deflected more than the

⁹⁴ Graves et al., *supra* note 51.

⁹⁵ *Electrophoresis and Electro Blotting of Proteins*, MARIETTA COLL., <http://w3.marietta.edu/~spilatr/biol309/labexercises/Electrophoresis.pdf> (last visited Dec. 8, 2016).

⁹⁶ Graves et al., *supra* note 51.

⁹⁷ See *supra* notes 92–93 (explaining these processes).

⁹⁸ CHEM GUIDE, *supra* note 63.

⁹⁹ *Ionization*, ENCYCLOPEDIA BRITANNICA, <https://www.britannica.com/science/ionization> (last visited Dec. 4, 2016).

¹⁰⁰ CHEM GUIDE, *supra* note 63.

¹⁰¹ *Kinetic Energy*, ENCYCLOPEDIA BRITANNICA, <https://www.britannica.com/science/kinetic-energy> (last visited Dec. 4, 2016).

¹⁰² *Id.*

heavier ions.¹⁰³ Finally, the ions are detected electronically in the spectrometer, and the degree to which they were deflected is used to identify the contents atoms or molecules.¹⁰⁴

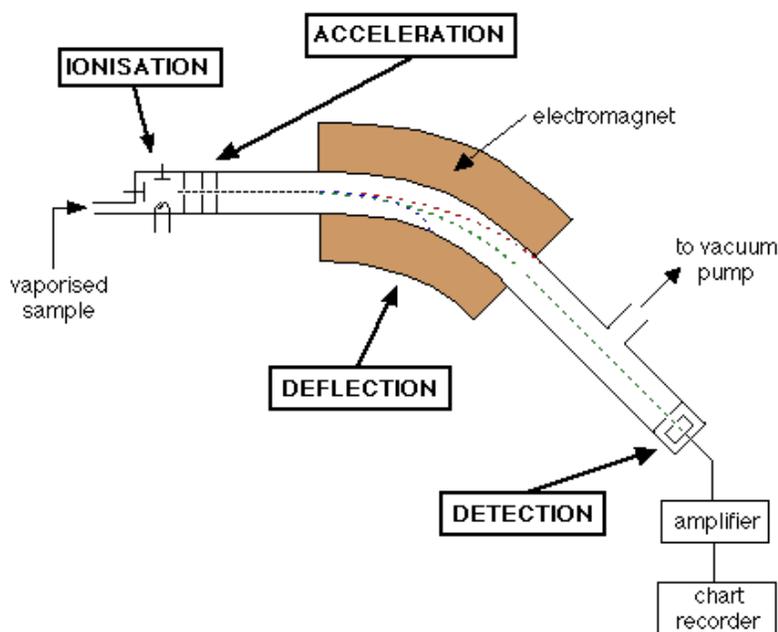


Figure 3: This diagram shows the basic setup of a mass spectrometer.¹⁰⁵

A computer then analyzes the results of the MS test and shows the results on a vertical bar graph.¹⁰⁶ Each bar represents an ion with a specific mass-to-charge ration (“m/z”) and its abundance.¹⁰⁷ When viewing an MS reading, molecules of identical size can readily be distinguished from one another and identified, thus allowing scientists to identify the presence of individual proteins after they are separated through 2-DE.¹⁰⁸

¹⁰³ CHEM GUIDE, *supra* note 63.

¹⁰⁴ *Id.*

¹⁰⁵ *Id.*

¹⁰⁶ *Mass Spectrometry*, MICH. ST. UNIV., <https://www2.chemistry.msu.edu/faculty/reusch/virtxtjml/spectrpy/massspec/masspec1.htm> (last visited Sept. 30, 2016).

¹⁰⁷ *Id.*

¹⁰⁸ *Id.*

iii. Database Utilization

The final stage in a proteome experiment is to use and develop a database.¹⁰⁹ One method for doing so is called “peptide mass fingerprinting.” Peptide mass fingerprinting allows proteins from an unknown source to be compared to a predicted mass from a theoretical calculation in a protein database.¹¹⁰ Because this process can be fully automated, speed is its greatest advantage.¹¹¹ But this method has two major shortcomings. The more complex the protein, the more likely that the predicted mass is wrong or has not been calculated.¹¹² Further, complex mixtures of the same protein cannot be distinguished by this process.¹¹³ So, if each letter in the following sequence, “VAGSE,” represents a different amino acid, that protein will have the same weight, and therefore the same “fingerprint,” as “GSEAV,” which is an entirely different protein.¹¹⁴ This method gives the cumulative weight, not necessarily the order of its parts—a critical difference.

The other primary method is called “amino acid sequence database searching.” This method is more specific, and it allows the full protein to be identified if the amino acid sequence of a peptide is successfully identified.¹¹⁵ Basically, this method allows for a complete protein to be found in a database if the scientist only knows a portion of the amino acid makeup of the protein itself. This method, although more time consuming, allows for more complex proteins to be identified.¹¹⁶

¹⁰⁹ Graves et al., *supra* note 51.

¹¹⁰ *Id.*

¹¹¹ *Id.*

¹¹² *Id.*

¹¹³ *Id.*

¹¹⁴ *Id.*

¹¹⁵ *Id.*

¹¹⁶ *Id.*

With each of these methods, scientists are constantly improving the databases by replacing theoretical masses with known masses, providing the correct order of amino acids in proteins, and adding unknown proteins to the databases.¹¹⁷

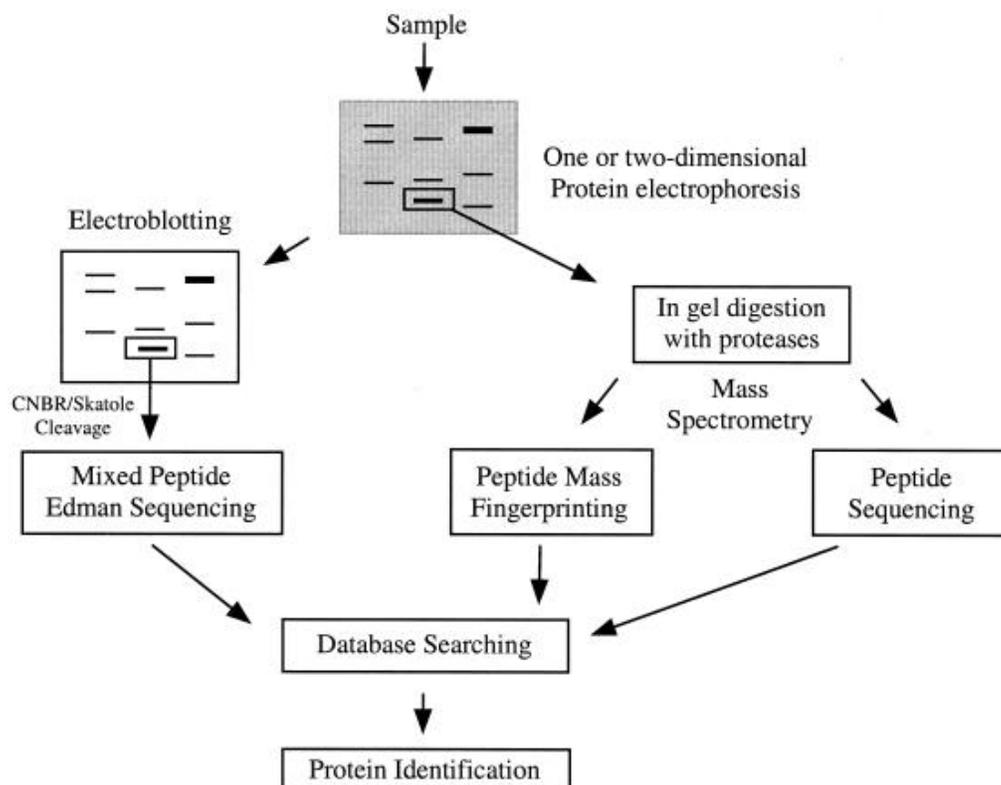


Figure 4: This diagram shows a typical proteome experiment used to identify specific proteins.¹¹⁸

The abovementioned process has allowed for scientists to apply proteomics to cancer research, medical malpractice cases, and even ancestral identification.¹¹⁹ In the last few years, forensic scientists have begun experimenting with proteomics as a means of individual identification from hair.

¹¹⁷ *Id.*

¹¹⁸ *Id.*

¹¹⁹ *Popular Applications of Genomics*, ANNENBERG LEARNER, https://www.learner.org/courses/biology/textbook/genom/genom_10.html (last visited Dec. 4, 2016).

III. PROTEOMICS FOR IDENTIFICATION

Much of the evidence collected at a crime scene is only circumstantial in nature. Trace evidence, for example, can indicate things like the type of clothing and shoes worn, weapons used, or other material present at a crime scene. Though this evidence is used to narrow down suspects or build a case, it does not necessarily indicate that a specific person was there. Some types of forensic evidence, however, can indicate that a specific person was most likely at the crime scene. Latent fingerprints and DNA evidence are types of evidence that can be used for this type of identification.¹²⁰ The basic premise is simple—that variation among latent prints and DNA is so great that scientists can fairly conclude that each set of prints or sample of DNA is from a particular individual (except that identical twins have the same DNA).¹²¹

The future use of proteomics for identification purposes is discussed in very similar terms to the using DNA for identification. For example, the major scientific studies on proteomics compare their level of confidence by comparing it to different types of DNA analysis. Likewise, its evidentiary value is expected to compliment and, perhaps, supersede that of DNA. Finally, as will be discussed *infra*, many of the arguments for and against admitting proteomic evidence are similar to those of DNA. Thus, this paper will give a brief overview of how DNA is used for identification, the limits of DNA, and the future of DNA. This section will then include a

¹²⁰ Note, however, that fingerprints and DNA also have limitations. Even if a O.J.'s bloody fingerprint was found on Nicole Brown, it does not *necessarily* mean that O.J. killed her. But it would be pretty damning evidence.

¹²¹ Brijesh Kumar Sharma, *DNA Fingerprint Introduction and Application*, BIOTECH ARTICLES (May, 20, 2011), <http://www.biotecharticles.com/DNA-Article/DNA-Fingerprinting-Introduction-and-Applications-905.html>. Though the statistical odds of having two people with the same fingerprints or DNA (other than twins) are extraordinarily low, it is theoretically possible that two people could have identical fingerprints or DNA. In other words, nothing technically prohibits two people from having the same fingerprints or DNA, it is just statistically incredibly unlikely.

discussion of the use of hair as forensic evidence. Finally, this section will explain how hair could be used for identification purposes through proteomics.

A. DNA for Identification

The first use of DNA to obtain a conviction in this country was in a rape case from 1987 in Orange County, Florida.¹²² In that case, Tommy Lee Andrews was convicted of rape after blood and semen found in a rape victim matched his DNA.¹²³ The use of DNA evidence was affirmed after an in-depth evaluation of the procedure and application of DNA analysis in a frequently cited case called *New York v. Castro*.¹²⁴ In that case, after an admissibility hearing, the presiding judge held that DNA evidence could be admitted for both inculpatory and exculpatory purposes.¹²⁵ Within fifteen years, forensic DNA evidence became ubiquitous in the criminal justice system.¹²⁶ The use of DNA was enhanced by development in technology, such as the F.B.I.'s Combined DNA Index System, and law, such as the Violent Crime Control and Law Enforcement Act of 1994, which designed uniform standards of DNA testing.¹²⁷

There are two main types of DNA evidence: nuclear DNA and mitochondrial DNA (mtDNA). MtDNA is the DNA present in the mitochondria of a cell, which is responsible for turning energy from food into a form the cell can use.¹²⁸ Compared to nuclear DNA, mtDNA is

¹²² *Andrews v. Florida*, 533 So.2d 841 (Fl. Ct. of App. 1988); Lisa Colandro et al., *Evolution of DNA Evidence for Crime Solving- A Judicial and Legislative History*, FORENSIC MAGAZINE (Jan. 6, 205 3:00 AM), <http://www.forensicmag.com/article/2005/01/evolution-dna-evidence-crime-solving-judicial-and-legislative-history>.

¹²³ *Andrews*, 533 So.2d at 841.

¹²⁴ 545 N.Y.S.2d 985 (Sup. Ct. 1989)

¹²⁵ *Id.* at 995. For a further analysis of this case, see Stephen M. Patton, Note, *DNA Fingerprinting: The Castro Case*, 3 HARV. J.L. & TECH. 223 (1990).

¹²⁶ Colandro, *supra* note 122.

¹²⁷ *Id.*

¹²⁸ *Mitochondrial DNA*, GENETICS HOME REFERENCE, <https://ghr.nlm.nih.gov/mitochondrial-dna> (last visited Dec. 8, 2016).

far more abundant, but it does not allow for individualization.¹²⁹ MtDNA is shared by everyone in the same maternal bloodline.¹³⁰ While this may still be helpful for prosecutors to bolster the circumstantial evidence, it does not allow for specific identification. Nuclear DNA, on the other hand, does allow for specific identification, but it is less abundant.¹³¹

Both types of DNA analysis for forensic purposes follows the basic pattern of the proteomics experiment discussed *supra*. The basic steps of DNA analysis include: (1) isolation of DNA from source (ex. blood, semen, saliva, skin, etc.); (2) processing the DNA to give results; (3) determination of the results; and (4) comparison and interpretation of the results from the unknown source to a known source.¹³²

DNA evidence continues to be an important tool for both the prosecution and the defense of suspected criminals. While many think of DNA as an inculpatory tool, DNA evidence is also frequently used as a post-conviction exculpatory tool.¹³³ This evidence, however, still has some shortcomings. Common critiques of DNA analysis include its potential for cross-contamination, the difficulty of translating statistical certainty to scientific opinion,¹³⁴ and the possibility of human error in drawing conclusions.¹³⁵

But the future of DNA is bright. DNA experts are hopeful that further development will allow them to use DNA from an unknown source that was found at the scene of the crime to

¹²⁹ Hughes, *supra* note 154.

¹³⁰ *Id.*

¹³¹ *Id.*

¹³² *DNA Evidence: Basics of Analyzing*, NAT'L INSTIT. JUST., <http://nij.gov/topics/forensics/evidence/dna/basics/pages/analyzing.aspx> (last visited Oct. 16, 2016).

¹³³ Colandro, *supra* note 122.

¹³⁴ Experts giving statistical conclusions cannot always give a “one in a 7.4 billion” (the rough population of the earth) conclusion. In fact, it is frequently limited to a one in a few million result. While this is not personal identification, it is very significant evidence in a trial when the other evidence is taken into account.

¹³⁵ EDWARD J. IMWINKELRIED ET AL., *SCIENTIFIC EVIDENCE* § 18.04 (5th eds. 2015).

create a range of characteristics for the source, such as skin color, hair color, eye color, and possibly facial characteristics.¹³⁶ At some point in the near future, it may be possible to create a police sketch of what the source looks like without ever seeing the source's face.¹³⁷

B. *Hair Analysis*

The sheer number of individual hairs, about five-million on the human body, and the likelihood of hair exchange on contact between two persons, have prompted forensic scientists to develop a number of techniques to study hair.¹³⁸ Forensic scientists can test hair to determine: whether it is from a human source (instead of an animal or synthetic source), whether the source is male or female, the source's race, and whether the source was poisoned or used drugs, among other conclusions.¹³⁹ These determinations can be made by using basic analytic techniques such as conventional microscopy.¹⁴⁰

Several techniques, however, have been developed to give more specific information about the source of a seized or recovered hair, but only a few have been held to be admissible in court. Ion microprobe is a process that measures the number and mass ions released by a hair sample when struck by an ion beam.¹⁴¹ The theory behind using an ion microprobe for identification is that two pieces of hair from the same source will have sufficiently similar levels of specific ions, while hair from different sources will have dissimilar number of specific ions.¹⁴²

¹³⁶ *Id.* § 18.06.

¹³⁷ *Id.*

¹³⁸ *Id.* § 24.02.

¹³⁹ *Id.*; FED. BUREAU OF INVEST., HAIR EVIDENCE (July 2000), <https://archives.fbi.gov/archives/about-us/lab/forensic-science-communications/fsc/july2000/deedric1.htm>.

¹⁴⁰ IMWINKELRIED ET AL., *supra* note 135, § 24.02.

¹⁴¹ *Id.*

¹⁴² *United States v. Brown*, 557 F.2d 541, 555 (6th Cir. 1977). Similar to other methods, this method relies on statistical significance to put a number to the exact likelihood that a certain hair could come from someone other than the alleged defendant.

However, the Court of Appeals for the Sixth Circuit held that such testimony violated the *Frye v. United States*¹⁴³ admissibility standard because it was not generally accepted.¹⁴⁴

Another type of analysis, called pyrolysis gas chromatography (PGC), separates a sample's chemical components and then measures the time it takes for them to travel through a "PGC column."¹⁴⁵ The time it takes for each chemical to run through the column is unique, and, therefore, a composition of the chemical makeup of the hair can be deduced.¹⁴⁶ The theory behind this method is that the specific chemical composition of a hair is unique to an individual, and the sensitivity of the chromatograph is such that these levels can be distinguished by miniscule measurements.¹⁴⁷ While some scientists believe this method is promising, this technique fails to support a sufficiently specific conclusion as to the source of the hair, particularly because some products, such as hairspray, gels, dyes, and shampoos, can undermine the results.¹⁴⁸

Pyrolysis mass spectrometry (Py-MS) is a similar technique to PGC, but the sample is passed through a mass spectrometer in order to find the chemical composition of the hair.¹⁴⁹ Like PGC, the idea is that an individual's hair will have a unique chemical composition that can be matched to or distinguished from an unknown sample.¹⁵⁰ Although Py-MS is faster and more

¹⁴³ 293 F. 1013 (D.C. Cir. 1923).

¹⁴⁴ *Brown*, 557 F.2d at 555.

¹⁴⁵ IMWINKELRIED ET AL., *supra* note 135, § 24.02.

¹⁴⁶ *Id.*

¹⁴⁷ *Id.*

¹⁴⁸ *Id.* Currently, the sensitivity of chromatographs is not at the requisite levels for individual identification: hairs from the same head can produce varied results that would not allow for a match. *Id.*

¹⁴⁹ *Id.*

¹⁵⁰ *Id.*

standardized than PGC, there has been insufficient research to support claims of individual identification and the sensitivity of the spectrometer is insufficient.¹⁵¹

Finally, scanning electron microscopes (SEM) allow scientists to view incredibly minute details, and possibly individual characteristics, about an individual's hair.¹⁵² But only a few SEM researchers have concluded that their results are reliably individualized, and this type of evidence has not been admitted in court.¹⁵³ While these types of analysis may be helpful in the future, it does not amount to the most important evidence—who was the specific source of the hair.

Of the techniques that are admissible in court, only one can give an individual identification of the source. Nuclear DNA analysis can be performed, as explained *supra*, on hair if the root is attached to the hair sample.¹⁵⁴ But without the root, no nuclear DNA can be extracted.¹⁵⁵ This is a particular problem because hair is unlikely to have the root unless it is pulled out—an unlikely scenario at many crime scenes.¹⁵⁶ Nuclear DNA is the only genetic method from which the source can be individualized.¹⁵⁷ The only DNA testing available for the hair shaft is mtDNA testing,¹⁵⁸ which is shared by everyone in the same maternal bloodline.¹⁵⁹

¹⁵¹ *Id.*

¹⁵² *Id.*

¹⁵³ *Id.*

¹⁵⁴ Robert H. Rice et al., *Human Hair Proteomics- Improved Evidence Discrimination*, NAT'L CRIM. JUST. REFERENCE SERVICE (Aug. 2015), <https://www.ncjrs.gov/pdffiles1/nij/grants/249102.pdf>; Caroline Hughes, *Challenges in DNA Testing and Forensic Analysis of Hair Samples*, FORENSIC MAG. (Apr. 2, 2013, 8:42 AM), <http://www.forensicmag.com/article/2013/04/challenges-dna-testing-and-forensic-analysis-hair-samples>.

¹⁵⁵ Mendoza, *supra* note 1. There are, in very rare circumstances, instances in which nuclear DNA may be found in the shaft. Rice et al., *supra* note 154.

¹⁵⁶ Rice et al., *supra* note 154.

¹⁵⁷ Mendoza, *supra* note 1.

¹⁵⁸ Hughes, *supra* note 154.

¹⁵⁹ *Id.*

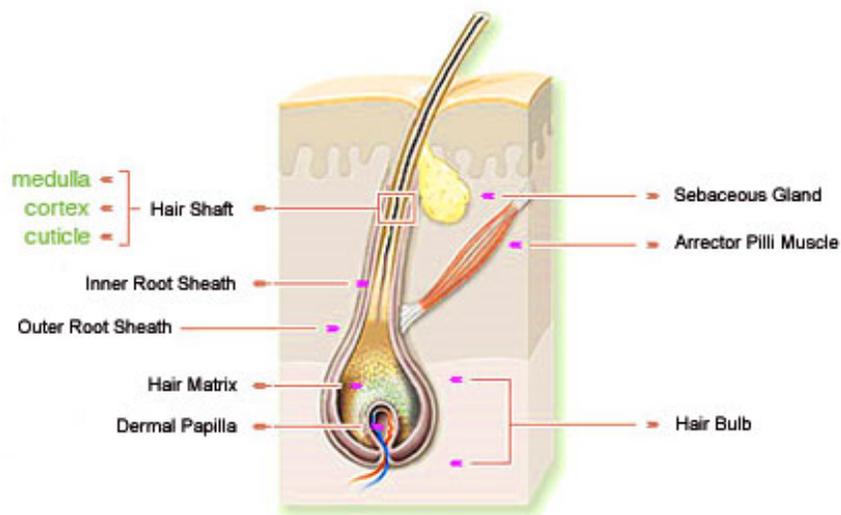


Figure 5: This cross section of a hair shows the area of the hair bulb, which contains the root and is required to perform nuclear DNA analysis. The hair shaft, comprised of the medulla, cortex, and cuticle, can be used to perform mtDNA analysis, but it does not allow for individualization like nuclear DNA.

Regarding the admissibility of testimony about hair, traditionally the majority of the historical methods of analysis (microscopy for information about hair color, for example) have been admitted under *Frye*.¹⁶⁰ The cases that rejected testimony about hair were when new types of procedures were introduced, such as ion microbe.¹⁶¹ There have also been issues of experts giving opinions beyond the limits of the science available at the time.¹⁶²

But skepticism has arisen from specific cases of overstated opinions and botched science. Soon after the 1993 Supreme Court decision, *Daubert v. Merrell Dow Pharmaceuticals, Inc.*,¹⁶³

¹⁶⁰ IMWINKELRIED ET AL., *supra* note 135, § 24.02.

¹⁶¹ *Id.*

¹⁶² *Buie v. McAdory*, 341 F.3d 623 (7th Cir. 2003) (evaluating a violation of due process claim for an expert who overstated his scientific certainty in hair identification); *United States v. Brady*, 595 F.2d 359, 362–63 (6th Cir. 1979) (permitting opinion evidence from conventional microscopy, without requiring an explanation for why other techniques were not used); *United States v. Massey*, 594 F.2d 676, 681 (8th Cir. 1979) (finding that the foundation was inadequate for introducing statistical evidence to quantify the individualization); *United States v. Cyphers*, 553 F.2d 1064, 1072–73 (7th Cir. 1977) (permitting opinion evidence from conventional microscopy, without requiring an explanation for why other techniques were not used).

¹⁶³ 509 U.S. 579 (1993).

some courts began to strike down the more cutting-edge techniques.¹⁶⁴ In 1997, hair testimony was brought into question, in large part, because of the exaggerated testimony of F.B.I. Agent Michael Malone in several high-profile cases.¹⁶⁵ But, it turns out, Mr. Malone was one of thirteen discredited F.B.I. agents who overstated the conclusions of their studies.¹⁶⁶ This required a review of about 6,000 cases between 1996 and 2004.¹⁶⁷ More recently, a 2009 report from the National Academy of Sciences stated that microscopic hair analysis is “highly unreliable.”¹⁶⁸ The F.B.I. is currently conducting a large-scale review of microscopic hair analysis in all cases from the early 1980s to December 31, 1999.¹⁶⁹ As of July 2013, there were 120 convictions, twenty-seven of which were capital cases, that were deemed “problematic.”¹⁷⁰

Even with these concerns, properly performed hair analysis and subsequent testimony continues to be an important component of trace evidence. Scientists focusing in proteomics are optimistic that they can bring proteomics into the category of DNA and use hair for individual source identification in the future.

¹⁶⁴ *Williamson v. Reynolds*, 904 F. Supp. 1529, 1558 (E.D. Okla. 1995), *rev'd on other grounds*, 110 F.3d 1508, 1522-23 (10th Cir. 1997).

¹⁶⁵ IMWINKELRIED ET AL., *supra* note 135, § 24.02.

¹⁶⁶ *Id.*

¹⁶⁷ *Id.*; Spencer S. Hsu, *Convicted Defendants Left Uninformed of Forensic Flaws Found by Justice Department*, Wash. Post (Apr. 16, 2012), https://www.washingtonpost.com/local/crime/convicted-defendants-left-uninformed-of-forensic-flaws-found-by-justice-dept/2012/04/16/gIQAWTcgMT_story.html.

¹⁶⁸ Gianellia, *Microscopic Hair Comparisons: A Cautionary Tale*, 46 *Crim. L. Bull.* 531 (2010).

¹⁶⁹ IMWINKELRIED ET AL., *supra* note 135, § 24.02.

¹⁷⁰ *Id.*

C. Hair for identification

In early September 2016, major news sources, such as the Washington Post,¹⁷¹ picked up on a study from the Forensic Science Center (“FSC Report”).¹⁷² The FSC Report boasted conclusions that demonstrate the potential of proteomics applied to hair shafts for identification of the source of the hair.¹⁷³ While the report’s primary conclusion was that the current use of proteomics for identification was similar to mtDNA (which is not individual but narrows down the identification), some of the scientists involved in the experiment were quick to praise the procedure as the next best breakthrough since DNA identification.¹⁷⁴ Hair proteomics have the potential to corroborate, and even to replace, DNA evidence within the next ten years.¹⁷⁵ Although the FSC Report is one of few studies focused on individual identification of a source through hair proteomics, it builds upon the conclusions of a report from the National Criminal Justice Reference Service in 2015 (“NCJRS Report”).¹⁷⁶ Both of these studies will be explained, in turn.

i. NCJRS Report

The NCJRS Report hypothesized that proteins in human hairs display individual differences that would allow scientists to match unknown samples to a known source.¹⁷⁷ This report focused on two tasks: (1) to determine whether individuals can be distinguished using

¹⁷¹ E.g., Spencer S. Hsu, *Has DNA Met Its Match as a Forensic Tool?*, WASH. POST (Sept. 7, 2016), https://www.washingtonpost.com/local/public-safety/has-dna-met-its-match-as-a-forensic-tool/2016/09/06/247a21d6-6fa8-11e6-8533-6b0b0ded0253_story.html

¹⁷² Glendon J. Parker et al., *Demonstration of Protein-Based Human Identification Using the Hair Shaft Proteome*, FORENSIC SCIENCE CENTER (Sept. 7, 2016) <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0160653>.

¹⁷³ *Id.*

¹⁷⁴ Hsu, *supra* note 171.

¹⁷⁵ *Id.*

¹⁷⁶ Rice et al., *supra* note 154.

¹⁷⁷ *Id.*

shotgun protein profiling; and (2) to develop a method that would allow streamlined testing of multiple unknown hair sources to a known hair source.¹⁷⁸

Regarding the first endeavor, shotgun protein filing (SPF) is a technique that builds off of SDS-PAGE but allows for low abundance proteins to be identified.¹⁷⁹ The separation process in SPF operates similarly to SDS-PAGE, but SPF uses several more steps to separate the proteins.¹⁸⁰ Generally, the same principles are applied: once the protein mixture is separated from the source, its chemical composition allows scientist to separate the different types of proteins in it from one another because of their individual reactions to varying pH levels and electronic charges.¹⁸¹ The NCJRS report was also able to take advantage of the most sensitive mass spectrometers available.¹⁸²

The specific parameters of the first experiment were as follows: scientists removed several hairs from varying areas of the scalp, axillary, (where possible) beard, and pubic region on non-related individuals of Caucasian, African-American, Korean, and Kenyan origin;¹⁸³ there were five subjects from each group of common origin except for Caucasian, where there were six subjects.¹⁸⁴ The hair samples from each body site were then compared using statistical similarities of the proteomic analysis.¹⁸⁵ In other words, the individual makeup of the hair was found, and the levels of certain proteins were used to decide whether there was a sufficient statistical similarity between the hairs to determine if they were from the same source.

¹⁷⁸ *Id.*

¹⁷⁹ *Shotgun Protein Identification*, CREATIVE PROTEOMICS, <http://www.creative-proteomics.com/services/shotgun-protein-identification.htm> (last visited Dec. 9, 2016).

¹⁸⁰ Rice et al., *supra* note 154.

¹⁸¹ *Id.*

¹⁸² *Id.*

¹⁸³ *Id.*

¹⁸⁴ *Id.* It should be noted that the study did not focus on sex-related differences. *Id.*

¹⁸⁵ *Id.*

The primary conclusion of this experiment was that “[i]ndividual differences in hair shaft proteomic profile are clearly evident The results so far permit us to speculate that this method may have discrimination capability similar (but complementary) to that of mitochondrial DNA.”¹⁸⁶ As explained *supra*, mitochondrial DNA does not allow for individualization, but it can exclude anyone who is not from the same maternal lineage is the hair’s source.¹⁸⁷ This Report does not posit that proteomics can only distinguish markers in the maternal lineage. Rather, the conclusion of proteomic analysis on a hair shaft has a similar statistical significance of performing mtDNA on the hair shaft. As is explained *infra*, scientists are optimistic that this statistical significance will soon be on par with nuclear DNA.¹⁸⁸

Further, the differences between ethnic origins were detectable, but the degrees of differences were not as great as among individuals.¹⁸⁹ It is easier to distinguish one person from another than from identifying persons of one ethnic origin from another. Of less note, proteomics can distinguish between what type of hair was found (scalp, axillary, pubic, etc.).¹⁹⁰ For forensic purposes, it is important to note that only hair from the same body site was compared.¹⁹¹ For example, pubic hair was not compared to scalp hair, nor could it be. The Report explained that while hair from each body site could be successfully matched, hairs from

¹⁸⁶ *Id.*

¹⁸⁷ *Supra* notes 154–52.

¹⁸⁸ *Infra* notes 221–11 and accompanying text (explaining that in five to ten years, proteomics will be on par with nuclear DNA).

¹⁸⁹ Rice et al., *supra* note 154.

¹⁹⁰ *Id.*

¹⁹¹ *Id.*

the scalp were the most distinguishable,¹⁹² which means that there are more or more readily identifiable protein markers in scalp hair than elsewhere.¹⁹³

Regarding the second goal of the project—to develop a method of streamlining testing—the experiment was designed as follows: different types of detergents, other than SDS, were used to see if they could remove proteins from the hair with even smaller sample sizes.¹⁹⁴ The Report concluded that new types of detergents, such as sodium dodecanoate and ammonium perfluorooctanoate, are better than SDS for small amounts of hair.¹⁹⁵

In a third experiment that was not part of the original plan, the Report set out to determine if twins could be differentiated from one another via proteomics.¹⁹⁶ They performed this experiment by comparing the hair protein profiles of nine sets of twins from the San Francisco area.¹⁹⁷ The Report concluded that even twins could be distinguished from one another based on proteomics, although their profiles were much more similar to one another than to those of non-related subjects because genetics is a dominant contributor to the formation of hair protein.¹⁹⁸ From the results of this test, the Report also concluded, by comparing a group of twins' habits (sleep, exercise, grooming, eating, etc.), that environmental factors have, at most, a weak impact on hair protein profiles.¹⁹⁹

¹⁹² *Id.*

¹⁹³ While this distinction is important, it does not necessarily change how crime scene investigators would perform their job. For example, if stray pubic hairs were found on a rape victim, pubic hairs from the suspect could be seized for comparison purposes. Likewise, if scalp hair was collected from the crime scene, scalp hair could be seized from a suspect.

¹⁹⁴ *Id.*

¹⁹⁵ *Id.*

¹⁹⁶ *Id.*

¹⁹⁷ *Id.*

¹⁹⁸ *Id.*

¹⁹⁹ *Id.* This conclusion would be particularly important for forensic purposes if, say, a suspect's hair was obtained after the suspect had undergone several hair treatments in order to disguise himself.

While the conclusions of this experiment were promising, the Report was quick to point out that much more research is needed. For example, the report could not conclude whether substantial differences in environment during early development would impact hair protein profiles.²⁰⁰ Further, among the protein markers that appeared to be most readily identifiable and distinguishable was keratin.²⁰¹ But the report suggested that further research was needed as to the validity of using keratin as a distinguishing marker compared to other proteins.²⁰²

ii. FSC Report

The key difference between the NCJRS Report and the FSC Report is that the FSC report set out to find easily identifiable protein markers and to find how easily identifiable a protein sample would be using those protein markers.²⁰³ Additionally, this report set out to see what impact, if any, time has on the scientists' ability to identify an individual using proteomics.²⁰⁴

As far as methods are concerned, the scientists involved in the FSC Report collected hair samples from sixty-six European-American subject, ten subjects with African ancestry, and six archaeological skeletal remains (up to 260 years old).²⁰⁵ Similar to the NCJRS report, the hair was treated with several solutions to separate the proteins from the hair samples and then subjected to liquid chromatography mass spectrometry.²⁰⁶ The results were then filtered, to

²⁰⁰ *Id.*

²⁰¹ *Id.*

²⁰² *Id.*

²⁰³ Parker et al., *supra* note 172.

²⁰⁴ *Id.*

²⁰⁵ *Id.* The primary research was conducted in Europe, but Parker et al. had some help from a scientist in Africa, which explains the mixed sample. *Id.*

²⁰⁶ *Id.*

reduce the possibility of false positives, by removing tests with low result quality (due from a hair sample being too small, machine error, or human error).²⁰⁷

After separating the peptides, cutting-edge mass spectrometers were used to identify the proteins present in the peptides.²⁰⁸ There were 185 protein makers, mostly keratin based, that were used to compare the samples.²⁰⁹ For each protein marker, a statistical probability was calculated that a person would have that markers.²¹⁰ The basis for that probability was from protein data collected as part of the Human Genome Project.²¹¹

In the Report's conclusions, the scientists asserted that there was sufficient variation to permit discrimination similar to that of mtDNA.²¹² Again, this does not allow for individual identification, but rather exclusionary information.²¹³ Specifically, they concluded that there was sufficient variation among the African and European populations to identify ethnicity from the sample.²¹⁴ Just as in the FSC Report, the variation among individuals was greater than that between ethnicities, so individual identification is easier than ethnic distinction.²¹⁵

Regarding the archaeological hair samples, the scientists concluded that proteins are less likely than DNA to be degraded or removed thorough environmental processes.²¹⁶ In other words, while DNA is subject to deterioration after a long period of time, proteomics can be used to correctly identify the proteins found in hair samples from over two-hundred years ago.

²⁰⁷ *Id.* The screening process basically took away results that were far outside the expected range.

²⁰⁸ *Id.*

²⁰⁹ *Id.*

²¹⁰ *Id.*

²¹¹ *Id.* Information on this data set can be found at *The 1000 Genomes Project*, INT'L GENOME SAMPLE RESOURCE, <http://www.internationalgenome.org> (last visited Oct. 28, 2016).

²¹² Parker et al., *supra* note 172.

²¹³ *See supra* notes 154–52.

²¹⁴ Parker et al., *supra* note 172.

²¹⁵ *Id.*

²¹⁶ *Id.*

The FSC Report supports the NCJRS report in two critical ways. First, it supports the proposition that there is sufficient variation among proteins to draw a conclusion of similar magnitude to mtDNA. Second, both reports indicate that proteomics will be just as accurate as DNA testing in the future, and it may sometimes prove to be even more helpful when only small hair samples or old hair samples are available.

Similar to the NCJRS Report, however, the FSC Report is not without its limitations. One critic of the report emphasized that there is still a lot of testing to be done. In the FSC Report there was limited variability among subgroups within the populations that were tested.²¹⁷ Future tests should focus on variability among other populations as well as variation among subgroups, such as regional differences and extended families.²¹⁸ This testing, however, is dependent upon available funding: the FSC Report itself cost well over \$3 million to complete.²¹⁹

D. *Comments*

Both of these reports indicate a step forward in proteomics for identification. Both reports indicated that the significance of their conclusions was similar to that of mtDNA. In many cases, a conclusion about mtDNA would say something like, “there is a 1% chance that the source of the unknown hair came from someone other than the suspect.”²²⁰ This conclusion may seem appealing, but it is a long way off from a conclusion from nuclear DNA testing, which would be around one-in-7.4 billion or roughly 0.000000013514%

²¹⁷ Hsu, *supra* note 171.

²¹⁸ *Id.*

²¹⁹ *Id.*

²²⁰ Terry Melton, *Mitochondrial DNA Examination of Cold Case Crime Scene Hairs*, FORENSIC MAG. (Apr. 1, 2009, 4:00 AM), <http://www.forensicmag.com/article/2009/04/mitochondrial-dna-examination-cold-case-crime-scene-hairs>. Melton posited that the average conclusion for mtDNA analysis is that there is about a one in one-hundred chance that an unknown hair came from someone other than that known source. *Id.*

While this may be the case in 2016, scientists involved in the research believe that it is just a matter of time before proteomics can be used to individualize the source of a hair. As the director of the national laboratory's Forensic Science Center, and co-author of the FSC Report, said, "We are in a very similar place with protein-based identification to where DNA profiling was during the early days of its development."²²¹ Further, Glina S. Cooper, director of science and research at the Innocence Project, said, "The best-case scenario is you will eventually have, in five to ten years, a complementary but separate method that we currently have to correctly identify or to exclude the right person involved in a crime."²²²

As explained below, however, proteomics has a long way to go before it is admitted as evidence in court.

IV. ADMISSIBILITY

In *Daubert v. Merrell Dow Pharmaceuticals, Inc.*,²²³ the United States Supreme Court held that the Federal Rules of Evidence superseded the widely adopted rule from *Frye v. United States*²²⁴ regarding the admissibility of expert testimony based on scientific principle or discovery.²²⁵ Under the old *Frye* rule, "the thing from which a deduction is made must be sufficiently established to have gained general acceptance in the particular field in which it belongs."²²⁶ Applying this test, the D.C. Circuit held that polygraph testing "had not yet gained standing and scientific recognition among physiological and psychological authorities."²²⁷

²²¹ Hsu, *supra* note 171.

²²² *Id.*

²²³ 509 U.S. 579 (1993).

²²⁴ 293 F. 1013 (D.C. Cir. 1923).

²²⁵ *Daubert*, 509 U.S. at 587.

²²⁶ *Frye*, 293 F. at 1014.

²²⁷ *Id.*

After seventy years, however, the Supreme Court in *Daubert* held that the Federal Rules of Evidence superseded the test in *Frye*.²²⁸ Specifically, FRE 702 states that an expert may give an opinion if the opinion: (a) will assist the trier of fact, (b) “is based on sufficient facts or data,” (c) “is the product of reliable principles and methods,” and (d) the expert “reliably applied the principles and methods to the facts of the case.”²²⁹ With FRE 702 as a backdrop, the Court set out to list of non-exhaustive factors to help courts distinguish reliable science from junk science under FRE 702.

Under *Daubert*, a trial judge deciding the admissibility of a proffered technique by evaluating five factors: (1) Is the theory or technique testable and has it been tested?; (2) Has the theory or technique been published or peer reviewed?; (3) Is there a known potential or potential rate of error?; (4) Is the theory or technique subject to operational standards?; and (5) Is the theory generally accepted?²³⁰ In applying these factors, the Court explained that the focus should be on the “principles and methodology, not on the conclusions that they generate,”²³¹ and the Court made it clear that “general acceptance” was no longer a necessary condition for the admissibility of scientific evidence.²³²

A. *Proteomics for Identification under Daubert and FRE 702*

There are only three cases currently available online in which proteomics were discussed.²³³ In two of the cases, proteomics was admitted as evidence in an attempt to link a

²²⁸ *Daubert*, 509 U.S. at 587.

²²⁹ FED. R. EVID. 702.

²³⁰ *Daubert*, 509 U.S. at 593–94.

²³¹ *Id.* at 595.

²³² *Id.* at 597.

²³³ *See, e.g.*, *Tarsell v. Sec. of Health and Human Serv.*, No. 10-251V, 2016 WL 880223 (Ct. Fed. Claims Feb. 16, 2016); *Romero v. Buhimschi*, No. 2:06-cv-10859, 2009 WL 92226 (E.D. Mich. Jan. 14, 2009); *Sullivan v. Sec. of Health & Human Serv.*, No. 10-398V, 2015 WL 1404957 (Ct. Fed. Claims Feb. 13, 2015).

vaccine for HPV, Gardasil, to varying illnesses where the cause of death for the plaintiffs.²³⁴ For example, there was one expert who concluded that “there is likely sufficient homology between the viral components of the Gardasil vaccine and human proteins for the development of autoimmune illness.”²³⁵ In both cases, however, the judges held that plaintiffs had not met their burden of proof in showing that the vaccine was more likely than not the cause of injury.²³⁶ Most importantly, neither of these cases went into detail about the admissibility of the proteomic evidence. All we know is that the judge heard evidence based on proteomic studies. To date, no one has tried to use proteomics using hair samples as a method of identification.

Therefore, in order to supplement the arguments regarding admissibility of proteomics for identification, I will be drawing upon some of the first cases that evaluated the admissibility of both nuclear and mitochondrial DNA evidence. While some of the cases applied *Frye* or its state-equivalent, these cases are still relevant because general acceptance is still evaluated under *Daubert*, albeit as an unnecessary condition. Most importantly, DNA evidence and proteomic evidence are similar in the regard that they were used in contexts other than forensic identification, usually medical research. Therefore, many of the arguments used in these cases will be similar to those raised in a hypothetical *Daubert* hearing for proteomic evidence.

These cases include *New York v. Castro*,²³⁷ which was one of the first in-depth evaluations of DNA, *United States v. Jakobetz*,²³⁸ which was the first federal case admitting

²³⁴ *Tarsell*, 2016 WL 880223 at *10–11; *Sullivan*, 2015 WL 1404957 at *8–9. The third case, *Romero*, mentions and explains proteomics only by reference; it was not evidence in trial. *Romero*, 2009 WL 92226 at *2–3. *Romero* was a dispute arising out of a scientist allegedly failing to contribute authorship to a fellow scientist in a research paper based on proteomics. *Id.* at *1. The judge held that the proper remedy for this disagreement should “be found outside of the courtroom.” *Id.*

²³⁵ *Sullivan*, 2015 WL 1404957 at *8.

²³⁶ *Tarsell*, 2016 WL 880223 at *1; *Sullivan*, 2015 WL 1404957 at *1.

²³⁷ 545 N.Y.S.2d 985 (Sup. Ct. 1989).

DNA under *Frye, Virgin Islands v. Penn*,²³⁹ which was the first federal case admitting DNA post-*Daubert*, *Tennessee v. Ware*,²⁴⁰ which was the first case admitting mitochondrial DNA, and *United States v. Coleman*,²⁴¹ which was the first federal case to evaluate mtDNA post-*Daubert*.

In order to demonstrate whether proteomics would be admissible for identification purposes, let us apply *Daubert* to the following hypothetical scenario:

Before the criminal trial of a man accused of murder, a federal trial court holds a *Daubert* hearing regarding the admissibility of a government expert's testimony. The expert's credentials are not under attack, as the expert has various degrees, including a PhD in biology from a prestigious university, multiple publications about biology generally (and a handful about proteomics), and was part of the research team for the FSC Report. Nor are the methods used to collect the evidence under attack. Three pieces of hair were collected from the clothes of the victim, using the standard procedures for crime scene investigation, immediately after the murder. There are no issues about chain of custody. None of these hairs had the root, which would have allowed for nuclear DNA analysis. While another expert will give testimony regarding the mtDNA analysis, which is consistent with the proteomic evidence, the expert at question in this hypothetical would like to give the following testimony.

After receiving the hair samples from the local forensics unit, she began testing the hair samples using the same methods used in the FSC Report. The expert was able to extract the peptides, use mass spectrometry to identify protein markers, and conduct a statistical analysis of whether any of the hairs could be excluded. The expert concluded that one of the hairs did not

²³⁸ 747 F. Supp. 250 (D. Vt. 1990).

²³⁹ 838 F. Supp. 1054 (D.V.I. 1993).

²⁴⁰ No. 03C01-9705CR00164, 1999 WL 233592 (Crim. App. Tenn. Apr. 20, 1999). This case applied the state-equivalent of *Frye's* general acceptability standard under TENN. R. EVID. 703.

²⁴¹ 202 F. Supp. 2d 962 (E.D. Mo. 2002).

belong to the defendant but two hairs did have a statistical match to the defendant. The expert then calculated the probability that the two hairs that did match could have come from a source other than the defendant. The expert's ultimate conclusion will be that "two of the hairs found on the clothes of the victim were found to have statistically significant similarities the defendant's protein markers. The chances of having a match from a different source is about one in one-hundred."

With the foregoing testimony in mind, the following are the arguments the prosecutor and defense attorney may raise and the conclusion the judge will reach.

i. Is the technique testable? Has it been tested?

The prosecution would argue that the technique is testable and that the expert herself has worked one of the seminal experiments performed on the subject. The FSC Report, as well as the NSCRJ Report, clearly outlined its procedures and methods. It would be fairly straightforward to repeat these procedures to replicate the test and draw conclusions. Moreover, the methods and techniques used in the FSC and NSCRJ reports are not novel, they build off of the same procedure for proteomic studies that have been around since before 1975.²⁴²

The prosecution would also argue that these methods have been tested and documented in various studies,²⁴³ and the particular methods used by the expert were the exact same as the FSC report. Further, in each test, the scientists performed multiple subtests, which tested varying theories on the effects of the environment, aging, and hair sample size. Even though there are

²⁴² See *supra* notes 43 and accompanying text. The argument that the methods and techniques build on established scientific methods was persuasive to the court in *Ware*, which allowed mtDNA into evidence, in part, because it built off the procedures established in nuclear DNA testing. *Ware*, 1999 WL 233592 at *13.

²⁴³ The court in *Castro* made note of the fact that while DNA for identification was novel, the techniques of DNA analysis had long been used in "diagnostics, clinical and experimental settings" *New York v. Castro*, 545 N.Y.S.2d 985, 963, 964 (Sup. Ct. 1989).

only two studies, these studies really have tested more than just the theory that hair proteomics can be used for identification.

The defense might have to concede that method is testable. But the defense could argue that, practically speaking, each step of the complex process would have to be tested and retested, not just the overall theory. In other words, the few studies that have taken place have used varying methods of protein separation: the NCJRS and FSC reports used different gels during the separation stage of the experiments.²⁴⁴ And, in order to test the validity of each of these methods, there would have to be millions of dollars in funding.

Further, the defense will rebut by stating that that two reports are insufficient for purposes of *Daubert*. While both studies reached the same general conclusion that hair proteomics can *likely* be used for identification, the other tests within the reports were not subject to the same retesting. The specific methods used in the two reports were different: the FSC Report did not use the exact same technique as the NCJRS Report. Therefore, the method used by this particular expert has not been tested more than once before. Even if the court found value in the admittance of proteomic evidence in other courts, the purposes and procedures here are different and should be subject to critical examination.²⁴⁵

ii. Has the technique been peer reviewed or published?

Similarly, the prosecution will point to the NCJRS and FSC Reports as two examples of published reports. Additionally, there are many reports that support the basic science behind

²⁴⁴ See *supra* note 69 and accompanying text (explaining that the varying detergents were used).

²⁴⁵ As the court explains in *United States v. Coleman*, 202 F. Supp. 2d 962 (E.D. Mo. 2002), even though judicial notice had been taken of DNA profiling, it is still the court's obligation to inquire about the specific techniques, especially new techniques, being proffered in a given case. *Id.* at 968.

proteomics.²⁴⁶ In fact, there are some medical journals exclusively dedicated to proteomics, including the *Journal of Proteomics & Bioinformatics*²⁴⁷ and *Clinical Proteomics*.²⁴⁸ Further, some of the reports on proteomics have previously been admitted in court.²⁴⁹ Both of those cases used proteomics to find potential side effects for a vaccine, but the basic methods used were similar to those used by the expert. While the NCJRS report did not garner as much attention as the FSC Report, there were many people in the scientific community who had the opportunity to comment, albeit informally, on the report.

The defense will raise a similar argument as it did concerning the first element: while there are two reports that have been published directly on point, they were testing different methods. Further, the scientific community has not had sufficient time to make genuine comments and critiques on the methods used.²⁵⁰ The FSC Report was released in September, and even if other scientific groups were able to get the necessary funding, there has not been enough time to recreate the test in the FSC Report, let alone test the other variables.

iii. Is there a known or potential rate of error?

Here, the prosecution is fighting an uphill battle. To date, there is no definitive information about potential rates of error from the method used in the FSC report. But because

²⁴⁶ The *Coleman* court found the presence of a decade's worth of publications on mtDNA to be a persuasive reason for admitting the evidence, particularly because some reports discussed the use of mtDNA for forensic purposes similar to those at issue in the case. *Id.* at 969.

²⁴⁷ JOURNAL OF PROTEOMICS & BIOINFORMATICS, <https://www.omicsonline.org/proteomics-bioinformatics.php> (last visited Dec. 11, 2016). This journal is on its ninth volume.

²⁴⁸ CLINICAL PROTEOMICS, <https://clinicalproteomicsjournal.biomedcentral.com> (last visited Dec. 11, 2016). This journal is on its thirteenth volume.

²⁴⁹ *Tarsell v. Sec. of Health and Human Serv.*, No. 10-251V, 2016 WL 880223 (Ct. Fed. Claims Feb. 16, 2016); *Sullivan v. Sec. of Health & Human Serv.*, No. 10-398V, 2015 WL 1404957 (Ct. Fed. Claims Feb. 13, 2015).

²⁵⁰ This type of argument was used, though not successfully in *Ware*. The court was concerned about an expert testimony about the limited amount of people in the scientific community who understand mtDNA enough to critique the methods used. *Ware*, 1999 WL 233592 at *15-16.

the conclusions are based in part on statistics, there should be an understanding of the statistical significance of the similarities and, therefore, the potential rate of error. But much more testing needs to be completed before a reliable potential rate of error can be calculated and certainly before a known rate of error can be determined. If the prosecution can show, however, that there are internal procedural measures to limit the theoretical rate of error, much like the FBI's internal procedures limited the chances for erroneous results in the early stages of DNA testing, a court may be more willing allow proteomics.²⁵¹

The defense will drive home the fact that the rate of error goes beyond the basic statistical comparison between the hairs from known and unknown sources. Each step of the procedure—the application of the gels, separation methods, mass spectrometry, etc.—has to be incorporated to get a full picture of the rate of error. This task has yet to be accomplished and would take several studies using the same methods. Most importantly, the lack of testing has not indicated where procedural safeguards should be put into place to diminish false-positives.²⁵² Unlike DNA, which had very specific procedures for the FBI to follow to diminish the rate of error, there has been no standardization of proteomic method for identification.

iv. Is the theory generally accepted?

At best, the prosecution could argue that the basic theory behind proteomics identification is generally accepted and has been studied for many years. Proteomics has been accepted and admitted in multiple medical malpractice cases, so there is some acceptance as to the basic theory of using mass and isoelectric points to separate protein mixtures, methods like MS to identify the contents, and statistical analysis to draw conclusion from the results. Some

²⁵¹ The court in *Penn* relied heavily on the safety mechanisms in place within the FBI's procedures for DNA testing. *Virgin Islands v. Penn*, 838 F.Supp. 1054, 1066–68 (D.V.I. 1993).

²⁵² *Id.*

courts have found general acceptance of the basic theories as a reliable starting block for admitting new applications of these theories as evidence.²⁵³ Usually more indicators of reliability are needed, but the idea that the methods are backed by accepted theories supports admitting the evidence.²⁵⁴

But the defense will argue that this point falls into a similar trap as the others. Namely, this specific method is not generally accepted. In fact, although the initial reaction from the scientific and forensic community was supportive, many were quick to point out the limitations of the results so far.²⁵⁵ While scientists may be in agreement about the potential use of proteomics, only two studies have studied proteomics for identification, and both studies were limited in scope.

B. *The Judge's Ruling*

On the above arguments, a judge would be hard-pressed to find that the expert's testimony passes scrutiny of the *Daubert* test. Though the methods employed by the expert can be tested, all of the other factors fall on the side of the defendant. At this point in time, the specific method has not been used more than twice: it was used in the FSC Report and in the expert's analysis, but it has not been used otherwise. Similarly, the lack of peer review makes it impossible to know just how trustworthy this method is when compared to other methods from unrelated proteomic studies. Regarding the error rate, though there is a theoretical error rate that

²⁵³ For example, the first time an appellate court in New York evaluated DNA evidence, it went on at length about how the theory underlying the DNA identification was generally accepted. *New York v. Castro*, 545 N.Y.S.2d 985, 963 (Sup. Ct. 1989).

²⁵⁴ A military court (one of the first federal courts to evaluate DNA evidence under *Daubert*) explained that even if the specific type of results had not been generally accepted and there was some dispute among experts in the field as to the method's validity, these concerns can be placed in front of the jury to qualify the weight of the evidence. *United States v. Thomas*, 43 M.J. 626 (U.S. Air Force 1995).

²⁵⁵ See *infra* notes 200–98 and 217–15 and accompanying text (explaining the limitations in both reports).

could be measured, there have been no tests to do so. Any attempted calculation would be speculation at best because the data on some of the methods used, especially the protein separation methods, is so limited. Finally, proteomics for identification is far from generally accepted. Again, while many see proteomics as an area of great potential, that day has yet to come.

While there may be some strength in the argument that the basic methods used by the expert are well established and reliable methods, the series of methods used in order to reach this conclusion are less than established. By way of a hypothetical analogy, take a bullet found at a crime scene. For some time, several scientists at the FBI would testify that their methods were able to match a bullet found at the scene to other bullets in the possession of the suspect at that time.²⁵⁶ In other words, they could study a bullet and conclude whether it came from a specific box of ammunition. The basic methods of analysis and procedures used by the FBI were not “junk science,” but their conclusions were. Basically, the FBI measured the level of trace elements found in bullets and presumed that there was enough variation among boxes of ammunition that they could make a statistical match from the bullets to other bullets within a box.²⁵⁷ The actual process used to find the levels of the trace elements had been used in other areas of forensic analysis, such as gunshot residue, and was widely accepted and admitted as evidence.²⁵⁸ While the basic premise was correct, the conclusions that were drawn were insufficiently developed to be admissible on their own.

²⁵⁶ Eric Lichtblau, *F.B.I. Abandons Disputed Test for Bullets from Crime Scenes*, N.Y. TIMES (Sept. 2, 2005), <http://www.nytimes.com/2005/09/02/politics/fbi-abandons-disputed-test-for-bullets-from-crime-scenes.html>.

²⁵⁷ *Id.*

²⁵⁸ Gabriela Vanini, *Forensic Ballistics by Inductively coupled plasma-optical emission spectroscopy: Quantification of gunshot residues and prediction of the number of shots using*

In the case of proteomics, at this point, a judge may be worried of a similar issue. The very basic methods behind proteomics are established, but the combination of methods and the types of conclusions drawn for proteomics for identification purposes are less so. While proteomics is sure to develop into a reliable and admissible science, today it would not pass the court's gatekeeping function under *Daubert*.

V. FUTURE ADMISSIBILITY, POTENTIAL CONCLUSIONS, AND METHODS OF ATTACK

Even if hair proteomics is not yet admissible for circumstantial evidence, and certainly not for individual identification purposes, under *Daubert*, the potential for proteomics is undeniable. As mentioned, scientists are hopeful that proteomics will continue to develop into a technology that can match or surpass that of nuclear DNA, which would permit individualization. So, in ten-or-so years, what will experts be able to conclude? How can opposing counsel limit their conclusions and attack their methods?

A. *Limitations of Expert Testimony*

Expert testimony for proteomics will almost certainly be limited similarly to DNA evidence. DNA evidence, as explained above, is susceptible to exaggeration and mischaracterization.²⁵⁹ First, proteomics evidence should be limited to statistical conclusions. Positive identification—saying that hair from an unknown source *certainly* came from a known source—is overstepping the bounds of proteomics.²⁶⁰ But because the human genome contains

different firearms, 118 MICROCHEMICAL J. 19 (2015), <http://www.sciencedirect.com/science/article/pii/S0026265X14001490>.

²⁵⁹ See also Jonathon J. Koehler, *Error and Exaggeration in the Presentation of DNA Evidence at Trial*, 34 JURIMETRICS J. 21 (1993).

²⁶⁰ See Kimberly Cogdell Boies, *Misuse of DNA Evidence is not Always a "Harmless Error": DNA Evidence, Prosecutorial Misconduct, and Wrongful Conviction*, 17 TEX. WESLEYAN L. REV. 403, 418 (2011) (explaining that DNA evidence can only narrow down a statistical probability that the unknown source matches the known source). Even if the odds of

only about 20,000 genes and the human proteome likely contains over a million proteins, the possibility for better statistical accuracy is greater in proteomics than genomics. With hundreds-of-thousands of proteins to choose from, scientists may be able to identify more protein markers that have a lower statistical probability of occurrence. Thus, it is theoretically possible that statistical accuracy could be many times greater in proteomics than genomics. Nevertheless, experts should be careful to limit their conclusions to the confines of the science and statistics *at the time*, and judges should be careful that proteomics does not garner any more weight than it should.

Further, proteomics evidence, like other trace evidence, must be put into context. From the hypothetical above, it is possible, even if it was the defendant's hair that was found on the clothes of the victim, that the hair arrived at the murder scene through some action other than the suspect murdering the victim—perhaps by an intermediary contact surface. At best, proteomics can help identify hairs from certain individuals that were found in certain places, at certain times. By no means should proteomics become sufficient evidence, by itself, that a certain crime was committed by a certain person. It should remain part of an otherwise strong case that can corroborate other evidence, including DNA evidence.

B. Methods of Attack

Like any piece of evidence, there are multiple lines of attack for an attorney hoping to exclude or limit proteomics evidence. Beyond normal issues of chain of custody, the complex nature of proteomic identification gives it many potential “weak spots.” For example, each step of the process—collection, separation, extraction, mass spectrometry, and statistical analysis—must be completed by scientifically established methods. Thus, if a new chemical was used for

someone else having the same DNA markers is 1 in 4 billion, there is still a possibility that someone else has the same markers.

separation, an attorney may be able to attack the admissibility of that conclusion because that method does not pass muster under *Daubert*. Likewise, if the mass spectrometer had not been calibrated as required by normal standards, an attorney may likewise be able to attack the admissibility or credibility of that test.

A future attorney preparing for a *Daubert* hearing must be familiar with *each* step of the methods used by the expert and how it compares to the scientific standards. With each step, there is an increase in the potential for error and opportunity for attack.

VI. PREPARING FOR PROTEOMICS

While the use of proteomics as evidence of individualization is some years away, there are a few things that scientists, policy makers, and investigators can do in the mean time to help create a smooth transition. As explained, *supra*, there are some major gaps in the research, and further testing for other variables and to recreate previous tests are needed. Ideally, the first area to be tested would be the methods that are used. In other words, variables regarding hair, such as differences in race, ethnic subgroups, time, environmental impact, and the like, are all important, but they should come second to honing in on a preferred process. Once scientists have found an effective and efficient method for testing small amounts of hair, these secondary variables can be tested using the same exact method. This will allow multiple studies to use that method so that it becomes trustworthy and “widely accepted” in the community. In performing the experiments, scientists should also continue to develop the databases so that hypothetical sequences can be confirmed.

Regarding policy makers, standards should be set for the use of proteomics in criminal investigation and prosecution. Just as the American Bar Association created the Criminal Justice Standards on DNA Evidence, the ABA should create standards on proteomics. This will ensure

that there is some consistency among the varying jurisdictions on how proteomics is used and the appropriate limitations. Federal law should also be implemented, much like it was for DNA standards in the Violent Crime Control and Law Enforcement Act of 1994,²⁶¹ in order to set the boundaries and minimal requirements for proteomics. Likewise, a national database should be created that allows state and federal authorities to update and reference the protein sequences of known criminals and unknown hair sources of unsolved crimes. This national database will assist law enforcement in its investigation, similar to the FBI's National DNA Index System.²⁶²

Finally, investigators and forensic scientists should develop standards of collection and preservation of hair samples. Even if the hair sample is small or does not have the bulb, a sample should still be preserved because future proteomics will likely be able to use the small sample. Again, these standards will assist in the transition into proteomics for identification and may allow future investigators to solve cold cases.

CONCLUSION

The future of proteomics is bright. At best, proteomics will be able to use a hair fragment to give more statistically significant conclusions than DNA evidence can today. While proteomics for identification is likely inadmissible today, it will not be long before it will be. When that day comes, an expert's testimony including or excluding a suspect from being the source of a hair found at the crime scene may be enough the tip the scales, one way or another, so that the truth will emerge.

²⁶¹ H.R. 3355, Pub. L. 103-322 (1994).

²⁶² F.B.I., FEDERAL DNA DATABASE, <https://www.fbi.gov/services/laboratory/biometric-analysis/federal-dna-database> (last visited Dec. 13, 2016).