DETECTION AND CLASSIFICATION OF MEASLES VIRUS BY SURFACE-ENHANCED RAMAN SCATTERING (SERS) SPECTROSCOPY AND FOURIER TRANSFORM INFRARED (FTIR) SPECTROSCOPY

by

VINH HOANG

(Under the Direction of Richard A. Dluhy)

ABSTRACT

Measles is a disease most prevalent in young children. The acute infection occurs once in a lifetime, and immunity is attained. Nonetheless, measles can thrive in large susceptible populations, and continue to contribute to the high mortality rate in underdeveloped countries. Currently, diagnosis methods for measles include immunofluorescent antibody assays and enzyme-linked immunoassays (ELISAs). However, these current assays are time-consuming and do not allow for direct analysis of the virus. Unlike the current immunoassays, the high sensitivity of surface-enhanced Raman scattering (SERS) spectroscopy allows for direct and rapid measles detection at low concentration and with minimal sample preparation. Thus, SERS could be the potential diagnostic tool for measles detection. The second method for measles detection was Fourier transform infrared (FTIR) spectroscopy. Rapid detection of pathogens was made possible by FTIR spectroscopy because of the minimal sample preparation and short data collection time.

INDEX WORDS: Measles virus, surface-enhanced Raman scattering spectroscopy, Fourier transform infrared spectroscopy, chemometric analysis

DETECTION AND CLASSIFICATION OF MEASLES VIRUS BY SURFACE-ENHANCED RAMAN SCATTERING (SERS) SPECTROSCOPY AND FOURIER TRANSFORM INFRARED (FTIR) SPECTROSCOPY

by

VINH HOANG

B.S. The University of Georgia, 2005

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2008

DETECTION AND CLASSIFICATION OF MEASLES VIRUS BY SURFACE-ENHANCED RAMAN SCATTERING (SERS) SPECTROSCOPY AND FOURIER TRANSFORM INFRARED (FTIR) SPECTROSCOPY

by

VINH HOANG

Major Professor: Richard Dluhy

Committee: Jonathan Amster
Geoffrey Smith

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
May 2008

DEDICATION

To my parents and grandparents.

ACKNOWLEDGEMENTS

I wish to thank my major Professor, Dr. Richard Dluhy for his guidance throughout the course of the research.

I like to thank my committee, Dr. Jonathan Amster and Dr. Geoffrey Smith for their constructive reviews.

I thank my family and friends for their support and encouragements.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

Chapter I

Introduction

1.1 Summary

Measles is a disease often found in young children. Common symptoms such as rash, fever, and coughing are usually presented during an infection. Measles is an acute infection. Once the host immune system is able to clear the virus from the body, the host acquires lifetime immunity. However, in some cases, measles infection can cause encephalitis or pneumonia. If those complications develop, the disease could be fatal. Currently, because of limited vaccination, measles is endemic in countries in Africa, South America, and Asia. Additionally, recent outbreaks were report in countries where measles vaccines are readily available. The main reasons for these outbreaks are virus importation and low vaccination coverage. For example, new measles outbreaks in the US were linked to imported measles strains; the majority of the cases involved unvaccinated individuals.

Because measles remains a current health concern on the national and global scales, it is important that laboratories have effective methods for detecting the virus. Rapid measles detection will facilitate diagnosis of the disease and contribute to the global measles surveillance. Current methods for measles diagnosis are enzyme-linked immunoassays (ELISAs) and immunofluorescent antibody assays. However, these immunoassays have low sensitivity and poor detection limit. In contrast, the signal enhancement in surface-enhanced Raman spectroscopy (SERS) significantly improves the detection limit of the technique. Thus, in this study, SERS measles detection was achieved with minimal preparation. Additionally, Fourier transform infrared (FTIR) spectroscopy was also applied for detection of measles virus. However, because of the high detection limit, optimal IR measurements required concentrated

measles samples. Finally, the reproducible data obtained from these two spectroscopic techniques were classified using chemometric methods of PCA, HCA, and PLS-DA.

1.2 Backgound on Measles Virus

Measles is an infectious disease that can be traced back to the 7th century. The incubation period and the lifetime immunity were first described by Peter Panum in 1846. The first virus isolation was made by Enders and Peebles in 1954. The isolate was obtained from human and monkey kidney tissue culture [1]. Measles virus (MV) belongs to the family of viruses known as *paramyxoviridae*. The virus is monotypic and is present all over the world. Before vaccination was available, acute measles was prevalent in young children. The acute infection occurs once in a lifetime, and immunity is attained. Most often, the virus is sustained by a large susceptible population. Because children have higher susceptibility to the disease, the infection is more common in the children population. However, with the introduction of measles vaccines, the number of measles cases has decreased. Nonetheless, in underdeveloped countries, a vast population of young children is still exposed to the virus. Children as young as 2 years old can acquire the disease. The mortality rate for the infected children is as high as 10 %. Furthermore, the disease progress is facilitated by malnutrition [2].

In the *paramyxoviridae* family, MV has been placed under the *Morbillivirus* genus. The *Morbillivirus* have been known to infect only humans. The average MV particle is around 120 to 250 nm. These viruses are generally polymorphous. Two essential proteins, locating on the surface of the viral envelope, allowed for initiation of infection. Both proteins are glycosylated integral proteins. The hemaglutinin (H) protein, with a molecular weight of 80 kDa, enables the virus to bind the host cellular receptors, such as receptor CD46. The fusion (F) protein allows for viral fusion with host cell membrane at neutral pH. The F protein is activated by the

cleavage of precursor protein (F0) into a heterodimer (F1/F2) [11]. Currently, the H protein remained the only known antigenic type of MV [1].

The MV genome is composed of a nonsegmented RNA of negative polarity. The MV genome, containing 15,894 nucleotides, codes for six structural proteins and two nonstructural proteins, C and V. The six structural proteins are composed of nucleoprotein, N, phosphoprotein, P, matrix, M, fusion, F, hemagglutinin, H, and large protein, L [3]. The virus RNA genome is stored within the viral ribonucleoprotein (RNP) core structure. The RNP core structure is linked to the MV envelope by the M protein, which interacts with both the H and F proteins. The RNA is protected by the nucleocapsid (N) protein and the phospho-(P)/large (L) protein complex. The L protein participates in viral replication. The P protein in turn, regulates the functions of the L protein in transcription and replication [2].

MV replication occurs within the cytoplasm. The promoter sequence of the virus is located in the noncoding region at the 3' terminal. Transcription is activated at the promoter sequence by the L/P polymerase complex. Following transcription, mRNAs are synthesized, with the most abundant ones being N, P, M, F, H and L-specific mRNAs. The polymerase complex first generates a positive-stranded antigenome. This antigenome serves as the template for the negative polarity RNA genome. The complete viral particle is released from the cell by budding. To attack the host cell, MV binds to the host cell receptor through the H protein. Binding initiates fusion by altering the conformation of the F protein. The conformational change allows the protein fusion domain to fit into the host cell membrane. Finally, the glycoprotein complex enables the infected cell to fuse with neighboring cells. This second fusion provides the means for the virus to spread throughout the host system [2].

The virus maintains its functions while its envelope is intact. However, if the viral structure loses its integrity, the virus is inactivated. Thus, materials such as detergents can be used to inactivate the virus and reduces its infectivity. Additionally, MV is also acid labile and thermolabile. MV is inactivated at a pH below 4.5, and loses its infectivity if stored at 56 ˚C for 30 minutes. However, at 37 ˚C, the MV half-life is 2 hours. For long-term storage, the virus can be kept at – 70 ˚C or freeze dried. In general, according to the Centers for Disease Control (CDC) pink book, heat, acidic pH, and ether can inactivate the virus [1, 2].

In 1963, the Edmonston B strain became the first live attenuated measles vaccine to be distributed in the US. Before the introduction of measles vaccines in 1963, measles epidemics occurred every 2 to 3 years. However, when vaccination was available in 1963, the number of measles cases dropped by more than 98 %. The frequency of measles epidemics was also reduced. However, between 1985 and 1988, 68 % of measles cases were reported for children receiving 1 dose of vaccination. Thus, a second dose was required to provide more effective protection against infections. From 1989 to 1991, a resurgence of measles cases was reported for children under 5 years old. The most likely cause for the measles resurgence was the low vaccination coverage among black and Hispanic children. At the time, measles outbreaks were also detected in other North and Central American countries [1]. After the measles resurgence ended, an increase in vaccination coverage resulted in a significant decrease in measles cases. In the past, the largest measles outbreak reported for preschool-aged children occurred in Los Angeles and lasted from 1987 to 1992. During this outbreak, 12,000 cases were reported; the majority of the cases involved unvaccinated children. In contrast, with increasing vaccination coverage, in 2000, the largest outbreak occurred in New York, involved only 9 cases [1].

In the 60's and 70's, vaccine-induced side-reactions were observed for early vaccine strains. As a result, the inactivated (killed) measles vaccine (KMV) was removed from the market in 1967. KMV stimulated the sensitivity without providing protection in the vaccinated hosts. The side-reactions associated with the vaccines were also known atypical measles. When exposed to wild-types, KMV recipients developed clinical symptoms similar to those observed in typical measles. Symptoms may include rash, fever, pneumonia, and edema. In addition to the KMV strain, the live attenuated Edmonson B vaccine strain was also withdrawn in 1975. The currently acceptable measles vaccine in the US is the attenuated Edmonston-Enders strain, also known as Moraten. This vaccine in found in the MMR combination; mumps and rubella vaccines are included. Measles vaccine causes a mild, but non-communicable infection. 99 % of people who received 2 vaccination doses develop detectable measles immunity. As part of the measles control effort, students in the US are required to provide MMR records before receiving admission. Additionally, because of the higher risk of exposure, MMR immunization records are required for anyone who works in a medical facility. Finally, 2 doses of MMR vaccine are recommended for those traveling outside the US [1].

Measles infections are more common in regions with limited access to vaccines. The number of cases of measles epidemics corresponds to the size of the susceptible population and the frequency of exposure. MV pathogenesis is characterized by tropism; the virus has a tendency of migrating toward the peripheral blood mononuclear cells. Thus, MV is often found in white blood cells, such as lymphocytes and monocytes. Susceptible cells facilitate tropism by displaying the CD46/moesin complex receptor. In healthy individuals, CD46 receptors are expressed to prevent cell lysis [2].

Measles infections begin with respiratory transmission. Then, viral replication occurs in the lymph nodes. In the primary viremia, MV can be detected in the bloodstream within 3 days of exposure. During the second viremia, the infection can spread to other organs. The incubation period takes 10 to 12 days. Clinical symptoms are presented as rash, followed by high fever, coughing or conjunctivitis, and Koplik spots on mucous membranes. The characteristic measles rash developed first on the face or head and then spread to extremities. The rash disappears in the same order. Complications associates with measles cases include diarrhea, pneumonia, and encephalitis. In some cases, pneumonia and encephalitis are the causes of deaths in measles infected patients. In developing countries, measles infections have frequently occurred in children who are 12 months-old or younger. Malnourished children composed the susceptible population for viral circulation. In these children, infection can lead to blindness and death [1].

MV is dispersed by respiratory transmission. The infection begins in the upper respiratory tract and the nose. MV spread through the respiratory tract via the bloodstream. Once the virus population expands, the infection spreads to the skin, kidney, and bladder. Additionally, a rash may appear temporarily. The rash indicates the initial stage of immune response. Antibodies (Ab) arrive at the infection site. In the humoral immune response, three specific Ab are produced against MV, the N-specific Abs, H-protein Abs, and F-specific Abs. Although the N-specific Abs are produced in abundance, these Abs do not provide protection against viral infection. However, the H-protein Abs are protective; they block the virus from binding host cells. H-protein Abs served as virus-neutralizing antibodies and prevent agglutination induced by MV H-protein. Additionally, the F-protein Abs inhibit viral-host fusion. Thus, these antibodies help contained the virus within the infection sites. Furthermore,

the virus is removed from the blood and tissue by T cells. In acute cases, CD4+ and CD8+ T cells are used to target infected cells. However, the drawback to having an induced immune response against measles is the host's increased susceptibility to other antigens. The suppressed immunity associated with measles infections is one of the sources contributing to the high mortality rate in the underdeveloped countries [2].

Further analysis of measles pathogenesis was conducted by McChesney et al.. To analyze the pathogenesis of measles in naïve and vaccinated individuals, the McChesney group used rhesus monkeys as the study models. The rhesus monkeys were chosen because of the shorter incubation period and the similar immune system. The virus isolates were obtained from the 1987 epidemic at the University of California Davis (Davis 87). Sequence analysis of the N and H genes of the Davis 87, Chicago-89, and Moraten vaccine strains showed the Davis 87 strain was closely related to the Chicago-89 strain. 5 monkeys received two doses of Moraten vaccine. About 1 year after vaccination, 5 vaccinated monkeys and 5 naïve monkeys were innoculated with the Davis 87 viruses. Innoculation was carried out conjunctivally and intranasally. Following innoculation, within 7 days, all the naïve monkeys developed symptoms of infections, including skin rash, inflammation, and coughing. However, after the 7-days period, the symptoms disappeared in infected naïve animals that were not euthanized. Secondary bacterial infections, such as bacterial gastroenteritis, were observed. However, these infections were treatable with antimicrobial therapy. Unlike the naïve animals, the vaccinated monkeys did not show signs of infections [4].

In the unvaccinated monkeys, pathologic changes, including inflammation and necrosis were detected in the respiratory tract and gut; the lungs showed signs of bronchitis and pneumonitis. In contrast, the lungs of vaccinated monkeys appeared normal. Remarkably,

hyperplasia was detected in the lymph node of vaccinated monkeys. Hyperplasia was not observed in the lymph node of healthy vaccinated monkeys; these were monkeys that were not innoculated. The cytopathic effect (CPE) of measles infection in the lymphoid tissues was the formation of multinucleated giant cells. Through immunofluorescence, bronchial epithelial cells, endothelial cells, $CD4^+$, $CD8^+$ T cells, and macrophages were identified as target cells in measles infections. Additionally, B lymphocytes in the lymph node follicles appeared to be the most common target cells in the immunofluorescence study. Overall, the animal model showed reproducible symptoms of measles infections. For example, pneumonitis observed in the innoculated naïve monkeys was similar to pneumonia observed in infants and children infected with measles. Additionally, measles infections were known to be systemic. In the animal models, CPE was observed in epithelial cells, endothelial cells, and lymphoid tissues. With the exception of detectable hyperplasia in the lymph node, the immunized monkeys did not present any other signs of infections. Observations from the pathogenic study outlined the infection path in unvaccinated monkeys and identified the target cells. Additionally, the study also showed that the Moraten vaccine was effective, but could not completely prevent infection in immunized individuals who are exposed to the MV wild-types. The observed hyperplasia indicated that infections occurred in vaccinated monkeys, but the infections were limited [4].

The previous study by McChesney et al. showed MV vaccines may produce side-reactions. Consequently, some vaccinated individuals could present symptoms similar to those observed in measles infections. To confirm diagnosis, it is necessary to distinguish symptoms produced by vaccine-induced side-reactions from those resulting from typical measles infections. A study conducted by Kobune et al. demonstrated a promising approach to differentiating side effects from symptoms of typical measles. The differentiation was done by sequence analysis of

the H protein in MV strains. The MV strains were isolated from seven Japanese children who developed fever and rash after vaccination. The differentiation of wild-types from vaccines was based on the nucleotide sequence, the molecular weight, the hemagglutination activity, and the antigenicity of the H protein [5].

Previous research has shown that unlike the vaccine strains, wild-type strains did not coagulate African green monkey red blood cells; the wild-types also had a slightly larger molecular weight. Sequence analysis of the H protein obtained from the vaccinee's virus strains, the wild-types, and the reference vaccines yielded the following results. 6 out of 7 vaccinee's virus strains presented a substitution at nucleotide 1266 in the H gene. The G to A substitution increased the molecular weight of the H protein. All of the wild-type virus strains presented the same G to A substitution at nucleotide 1266. Thus, the H protein of the wild-types also had an increased molecular weight. However, none of the vaccine strains' sequence displayed the substitution. Thus, the size of the H protein remained unchanged. Hemadsorption activity was monitored by combining MV-infected B95a cells with African green monkey red blood cells. The results showed the 6 vaccinee virus strains with the nucleotide substitutions lacked hemadsorption activity. Similarly, the wild-type infected cells also showed no sign of hemadsorption. In contrast, all the vaccine strains displayed hemadsorption activity [5]. Finally, antigenicity was determined by reaction of the virus strains with anti-H MAb. Virus strains from vaccines produced an antibody titre between 96,000 to 120,000. Virus-neutralizing antibody titre of the wild-type strains was less than 20. Based on these four tests, 6 of the 7 vaccinee virus strains were identified as wild-type viruses. Thus, only 1 of the 7 vaccinee presented vaccine-induced side reactions. Overall, the study provided an effective approach for differentiating vaccine-induced side effects and measles infection. Additionally, this method is

also applicable in the evaluation of MV vaccine safety [5]. According to the analysis, the children who presented symptoms were infected with the wild-type strains. Their symptoms did not result from vaccine-induced side reactions.

Even though vaccination programs have effectively reduced the number of measles cases in the US, Japan, and European countries, sporadic outbreaks have been observed in these countries. For example, in 2002, H1 wild-type MV was detected in Osaka, Japan. Genotype analysis was performed using RT-PCR. Through the A BLAST2 search database, the N gene of the Osaka isolate matched the N gene of the H1 reference strain, Hunan.CHN/93/7, with 99.6 % homology. This was the first time H1 viruses were detected in Japan. In the past, D3 and D5 were the predominant genotypes in Japan [6]. Following the H1 outbreak detected in Osaka, Japan, a D9 outbreak was recently reported for Yamagata, Japan in 2004. In the past, D5 was the predominant genotype until 2001. H1 became predominant in 2002 and 2003. Most likely, the D9 strain was imported from regions outside of Japan. The throat swab specimens were obtained from three infected students. MV genome was detected by RT-PCR. Sequence analysis of the N gene identified the isolates as D9 viruses. According to the outbreak report, the majority of the infected students were unvaccinated [7]. These reports showed vaccination is an essential step in measles control.

In addition to vaccination programs, global measles surveillance has also been an important process in measles control. Global measles surveillance provides a useful tool for analyzing the source and routes of MV transmission. Additionally, details concerning the global distribution of measles genotypes can be applied in evaluating the effectiveness of measles vaccines. The following paper reviewed the global distribution of MV and the current state of measles control. According to the review, in 2003, World Health Organization (WHO)

recognized 8 clades of MV. Within these 8 clades, 20 established and 2 proposed genotypes of MV have been reported. Phylogenetic analysis of the genotypes was based on the 450 base pairs (bp) of the C-terminal of the N gene and the entire H gene sequence. Except for the F genotype, the remaining genotypes were each given a reference strain. The reference strain provided information on the first isolation of the virus from a particular genotype. Recently, WHO created a Global Measles Laboratory Network to facilitate measles surveillance on a national, regional, and global scale. Molecular epidemiologic studies of measles showed that sequence variation in the N and H gene of virus strains from the same genotype was quite small. Viruses from the same genotype appeared to be genetically stable. Nonetheless, genetic variations were detected among the wild-types. The variations may represent an antigenic response for survival in an immunized population. However, wild-types with significant variations were still neutralized by antiserum to the vaccine virus. Thus, measles vaccines remained effective, regardless of the differences in the wild-types. Measles vaccines, including the Edmonston B and the Moraten strains, were developed from genotype A strains [8].

The reported cases for the measles surveillance indicated measles are distributed by two routes, endemic transmission and importation. Molecular epidemiologic studies of measles in the US showed that endemic measles was eradicated by 1993. The last reported endemic genotype was D3. The transmission of D3 viruses has been suppressed by vaccination since 1993. The viruses from different genotypes isolated between 1994 and 2001 came from imported sources. Similarly, the diverse genotypes isolated in Australia, Canada, and United Kingdom were also imported. However, measles surveillance is limited in some parts of the world. Endemic measles is still prevalent in these regions. Frequent outbreaks and endemic measles were reported for many countries in Africa and Asia between 1995 and 2001. Within

this time period, Western European countries such as Germany, Spain, and Russia were also identified as the sources for some of the imported cases. In particular, C2, D6, and D7 circulated widely in Western Europe because of the low vaccination rate. Clade B viruses and D2 and D4 viruses have been endemic in Africa. In Asia, India, Nepal, Thailand, Indonesia, China, Korea, and Viet Nam have been identified as the sources for endemic measles for D4, D5, D8, G2, H1, and H2. The continued presence of D3 and D5 in Japan indicated that co-circulation of multiple genotypes could occur within a susceptible population, without genotype displacement [8].

The detection of these different genotypes suggested that even though MV is monotypic, there are variations among the MV strains. For example, MV strains differed in MAb reactivity, temperature sensitivity, and nucleotide sequence. One of the most variable regions in the MV genome is the C-terminal of the N protein. Using the CLUSTAL program, sequence analysis of this region allowed for assignments of 8 different MV genotypes. According Rima et al., the genotype was proposed to have more than 3 nucleotide substitutions, with reference to the central nodes. All the vaccine viruses were placed in Group A genotype. Group B was composed of African isolates obtained from outbreaks in Cameroon and Gabon. Group C viruses had a wider global distribution. The strains were isolated from European and Asian countries, and the US. Group C also contained both wild-types and isolates of SSPE cases. Similarly, Groups D and E strains were also isolated from European countries and the US. Subacute panencephalitis (SSPE) isolates from Spain were identified as Group F strains. Group G consisted of 2 wild-type isolates from the US. Overall, the sequence analysis conducted by Rima et al. revealed that there was no geographic restriction for some of the genotypes, such as Groups C, D, and E. The isolation of MV strains from different regions suggested co-circulation

of different MV genotypes was possible on a global scale. Virus strains from Groups A, E, and F appeared to be inactive; these strains were not isolated within the last 15 years. Sequence analysis showed no sign of recombination. Although the strains were cultured from a variety of cells, the sequence stability was retained [9].

Genetic studies of the N, M, F, H, L, and P proteins found the largest variability was expressed in the 450 nt of the C-terminal coding region of the N-gene. Genetic diversity was also detected in the H protein of the wild-type strains. A comparison of wild-type viruses to vaccine viruses showed that the H protein in wild-types presented new or modified epitopes. The genetic diversity detected in H and N proteins were used to divide the isolated wild-types into different genetic groups. With the exception of Group 6 viruses, which were found to be predominant in Africa, the viruses seemed to circulate without any geographic restriction. Additionally, genetic analysis suggested that MV isolates from the measles resurgence between 1989 and 1992 were different from those isolated during the 1994 and 1995 outbreaks. Sequence data indicated that these isolates were imported from Asia and Europe. As part of the measles surveillance, molecular epidemiology provided a mean for determining the links between different lineages. Although the source of origin is not always certain, these links allow for classification of the viruses as part of the predominant genotypes or the imported strains [3].

Because they contained the most variable regions, the N and H proteins have been the focus of MV sequence analyses. Furthermore, these proteins were known to trigger a strong immune response. Sequence analysis of the N and H proteins were performed to determine if MV groupings based on the N and H sequences were similar. Phylogenetic analysis revealed that MV classification based on either the N or H sequence resulted in the same differentiation of MV genotypes.

While genetic analysis enables MV classification, MV surveillance allows for determination of transmission routes. For example, measles surveillance in Spain indicated there were four genotype switches between 1970 and 1993. In Spain, F was the predominant genotype in 1970. However, by 1979, C1 viruses replaced the F isolates. From April 1992 to July 1993, C2 became the most common lineage. By October 1993, C2 viruses were replaced by D3 viruses. The difference between D3 and C2 viruses was 50 nucleotide substitutions. Because the genotype switch occurred in a short period of time and the difference between the two genotypes was fairly large, it was likely the D3 viruses were imported.

In addition to tracing the transmission route, the mutation rate was also calculated for five strains from the D3 group, isolated between October 1993 and June 1994. The mutation rate was 4×10^{-4} per annum. Within this period, only single and double mutations were observed. Thus, the mutation rate was quite low for MV. Actually, this mutation rate was a magnitude lower than the calculated rate for human influenza A viruses. However, the IFF strain isolated in May 1995 presented 11 mutations. This change in mutations indicated that sequence stability could be maintained over several years, but sudden variations could also occur. Nonetheless, the ratio of expressed to silent mutations was low for both the C2 and D3 groups. Additionally, no fixed mutation was detected in the D3 viruses. With the exception of the IFF strain, the remaining strains presented one expressed mutation. Overall, the study showed that N and H sequence analyses lead to similar classification of the MV genotypes. Additionally, detection of imported virus suggested that measles vaccines could eradicate 'endemic' strains, but measles importation could start new outbreaks. Finally, the resulting mutation rate suggested the gradual variation was due to random drift, and not immune-driven [10].

Previous genetic analyses confirmed that different lineages of MV exist. In 1998, a uniform nomenclature for MV strains must be established to provide effective communication between laboratories. The Expanded Programme on Immunization (EPI) at WHO held a meeting in Geneva to decide on a standard protocol for characterizing MV wild-type strains. According to the recognized convention, the virus strain or sequence will include the following information. The name must distinguish between viruses isolated from cell culture and MV sequence obtained from RNA extracted from clinical samples. The following details are also required. The strain or sequence name must list the city and country of isolation, and the epidemic week and year during which the specimen was collected. Sequence analysis is applied to divide the viruses into clades and genotypes. The clades describe the genetic relationship between the genotypes. In 1998, WHO recognized 8 clades of MV, A, B, C, D, E, F, G, and H. The clades were further divided into genotype subgroups. The genotypes are defined by the sequence diversity in the 450 nucleotides at the C-terminal of the N gene and the genetic variation in the entire H gene sequence. By 1998, 15 genotypes were established, A, B1, B2, C1, C2, D1, D2, D3, D4, D5, D6, E, F, G, and H [11].

In addition to establishing the standard protocol for classifying MV strains, in 2001, WHO also set the criteria for proposing new genotypes. The sequence for the 450 nucleotides at the C-terminal of N gene and the complete H gene must be provided. The minimal variation between the proposed genotype strain and the most closely related strain must be 2.5 % for C-terminal of the N gene and 2.0 % for the H gene. The phylogenetic tree generated based on the diversity of the N or H gene must be similar. The new genotype is analyzed using multiple viral isolates or specimens. Finally, the new genotype must have a designated reference strain. Based on the set criteria for genotype proposal, the following new genotypes were recognized in 2001,

B3, D7, and D8; clades G and H were divided into H1 and H2, and G1 and G2 [12]. Unlike the other genotypes, genotype F was obtained from an SSPE strain. Thus, the F group didn't have an assigned reference strain. However, the remaining genotypes were each given a reference strain. The reference sequence can be accessed via the Genbank database. By 2003, 2 more new genotypes were established, D9 and G3 [13]. The Global Measles Laboratory Network encouraged laboratories to collect viral isolates from all reported measles cases and outbreaks. For virus isolation, the network preferred the use of Vero/hSLAM cells over B95a cells as the cell line. Vero/hSLAM cells are safer for shipment because B95a cells secrete Epstein-Barr virus and are considered to be infectious substance. By 2005, D10 was added to the list of recognized measles genotypes [14].

As part of the global measles surveillance, WHO provided a statistical report on the global reduction of measles. According to the WHO report, from 1995 to 2005, measles-related deaths decreased by 60 %. The number of deaths in 1999 was 873,000. In 2005, the number dropped to 345,000. In 2000, the goal for reducing measles mortality was set for the following 47 priority countries: Afghanistan, Angola, Bangladesh, Benin, Burkina Faso, Burundi, Cambodia, Cameroon, Central African Republic, Chad, Congo, Côte d'Ivoire, Democratic Republic of the Congo, Djibouti, Equatorial Guinea, Eritrea, Ethiopia, Gabon, Ghana, Guinea, Guinea-Bissau, India, Indonesia, Kenya, Lao People's Democratic Republic, Liberia, Madagascar, Mali, Mozambique, Myanmar, Nepal, Niger, Nigeria, Pakistan, Papua New Guinea, Rwanda, Senegal, Sierra Leone, Somalia, Sudan, Timor-Leste, Togo, Uganda, United Republic of Tanzania, Viet Nam, Yemen and Zambia. An important step to achieving this goal is to increase immunization coverage in children 12 months or younger. WHO and the United Nations Children's Fund (UNICEF) recorded 80 % global immunization coverage in 2006. The

improvements in vaccination coverage were most notable in the African, Eastern Mediterranean, and Western Pacific regions. By 2006, the Global Measles Laboratory Network has expanded to 676 labs in 164 countries. From 2000 – 2006, these laboratories conduct enzyme-linked immunosorbent assay (ELISA) tests on serum samples, using measles immunoglobulin M (IgM). Currently, the tests require around 7 days to collect 80 % of the results. From 2000 to 2006, mortality from measles decreased by 68 %. In 2000, approximately 757,000 deaths were recorded. In 2006, 242,000 deaths were recorded. Measles mortality reduction in the African region contributed to 70 % of the global reduction. Reduction in measles mortality was achieved by increasing immunization coverage in the African, Eastern Mediterranean, and Western Pacific regions. Additionally, 478 million children participated in the supplementary immunization activities (SIAs) implemented by WHO and UNICEF. However, the decrease in measles mortality was smaller in the South-East Asia region. This is mostly due to low immunization coverage and the lack of SIAs. Overall, progress in measles mortality reduction is possible with parental and government support of measles immunization programs. Additionally, the Measles Initiative has extended immunization program to the designated priority countries so measles reduction can be achieved on a global scale [15].

For the MV detection, the CDC supplied our lab with genotypes A, D4, D9, and H1. The following are journals written about these genotypes. Genotype A strains were the first MV strains detected in the US in 1954. Genotype A strains later became the prototype for MV vaccines. In 1994, it was uncertain whether the different attenuation methods could affect the genetic diversity of the different commercially-available vaccines. Thus, Rota et al. analyzed the sequences of those vaccine strains to determine the genetic diversity among the strains and also the possible effect that different attenuation conditions may have on the diversity. Analyses of

H, F, and N sequences of the wild-type Edmonston virus (wtEd), vaccines derived from wtEd, and vaccines that were not derived from wtEd were done. The results from the H and F sequence analyses indicated that CAM-70 had the largest sequence diversity. N-gene sequence showed S-191 had the most nucleotide substitutions. The progenitor wild-type virus for CAM-70 was a Japanese wild-type virus and not wtEd. Similarly, S-191 was also a non-Edmonston-derived vaccine, developed in China. Therefore, it was possible the genetic diversity in these two vaccines stemmed from the different MV lineages. However, because the sequences of the CAM-70 and S-191 progenitors were not available at the time, the source of the genetic diversity in MV vaccines could be the passage history or the parental lineage. The high mutation frequency in S-191 N sequence suggested that hypermutation could occur during invitro replication of MV. Although variations existed among the vaccine strains, in general, the vaccines were very similar. For example, a comparison of the complete genome of AIK-C vaccine to that of wtEd showed only 56 nucleotide substitutions. The sequence analysis also showed that the vaccines were most closely-related to the wtEd strain. The sequence similarity between vaccines and wtEd suggested there was homogeneity in MV strains circulating during the pre-vaccine era [16]. In addition to analyzing the genetic diversity in vaccines and the effects of different attenuation methods, the study also provided the reference sequence strain for the H gene of genotype A viruses.

In addition to providing the sequence for the H gene, the Rota's group also presented the sequence for the N gene of the genotype A strains. cDNA clones of N and M genes were sequenced to determine genetic variations in wild-type isolates present in the prevaccine era and in isolates obtained between 1977 and 1989. Earlier comparison of Moraten vaccine strain with MV wild-types indicated a genetic drift in the H protein of the wild-type viruses. These wild-

type isolates presented 19 amino acid substitutions and a new potential glycosylation site. A study was conducted by the Rota's group to compare the wild-type isolates from the 70s and 80s with those isolated during the 50s and 60s. Moraten was used as the reference vaccine strain. Nucleotide (nt) sequence analysis showed that the majority of the substitutions occurred in the last 125 amino acids located at the C-terminal of the N-gene. Sequencing the entire N-protein revealed an additional variable region from residues 130 to 150. Although the C-terminal was variable, the N-terminal of the N protein was highly conserved. The N-terminal served as the binding site for the P protein. Two phenograms were generated to describe the relationship between the wild-types and the vaccine strain. One phenogram was base on the 456 nt at the C-terminal, and the other was created using the entire N protein. The similarity in the two phenograms indicated the source for the genetic diversity in the N protein came from the 456 nt at the C-terminal. Lineage analysis showed the wild-types isolated in the United States between 1983 and 1989 were closely related to wild-types isolated in the United Kingdom between 1983 and 1989. In contrast, wild-types isolated in the US during the 50s and 60s were more closely related to the Moraten vaccine strain. The results suggested vaccination interrupted the transmission of the MV lineage from the prevaccine era. The similarity between wild-types isolated in the US and those isolated in the UK indicated the resurgence of measles cases in the US in the 80s were caused by imported wild-types of different lineages [17].

The reference strain for D9 genotype was isolated in Australia. Between 1999 and 2001, 19 MV strains were isolated from five Australian states. 18 of the strains belonged to 9 different recognized genotypes. The remaining strain was proposed to be a new genotype in clade D (d9). The report on the measles epidemiology in Australia between 1999 and 2001 documented the first isolation of d9 viruses. According to the report, one of the isolates from the 1999 outbreak

in Victoria represented a new genotype in clade D. Subsequent d9 infections were observed in other parts of Australia, including the Northern Territory state. Phylogenetic analysis of the d9 H gene revealed a 424T substitution that was absent in the H gene of other reference strains. Sequence analysis of the 450 nt at the C- terminal of the N gene and the entire H gene suggested isolates from three separate measles outbreaks in Australia belonged to a new genotype in clade D. The new genotype was proposed to be d9, and MVi/Vic.AU/12.99 was designated as the reference strain [6].

H1 strains were first detected in China. Even after the distribution of vaccines in China, 100,000 cases were documented each year between 1987 and 1994. Some of the cases were reported for immunized patients. The following was the first report on genetic and antigenic characterization in China. Sequence analysis of the H and N genes in 14 Chinese wild-types revealed that 13 of these wild-type viruses differed from viruses belonging to previously established genotype groups. Individuals infected with these wild-types presented symptoms commonly observed in measles infections; symptoms included rash, fever, coughing, and conjunctivitis. Compared to wild-types from the known genotypes, the new wild-types had 6.9 % variation in the H gene and 7.0 % variation in the N gene. Further analysis of the N and H proteins indicated that there were distinct amino acid substitutions observed only in the Chinese viruses. Additionally, the N-linked glycosylation site at residue 416 in the H protein, detected in many previously characterized wild-types, was absent in the Chinese viruses. Similarly, the N-linked glycosylation site at residue 187, known to be conserved in most wild-types, was also absent in several of the Chinese viruses. However, antigenic analysis by radioimmunoprecipitation of the 14 wild-types and reaction with seven H MAbs showed no significant difference in binding selection between the Chinese viruses and other known viruses.

Antigenic test of H proteins from the vaccine strains and the Chinese viruses yielded similar antiserum neutralization results. Thus, there was conservation of neutralization epitopes among the wild-types. Because of this conservation, MV vaccine strains, although closely related to genotype A, remained effective on wild-types of different lineages. Before 1997, the virus strains belonging to the new MV group appeared to be more prevalent in China than in any other regions around the world. The possible source of transmission to other regions was importation. At the time that this report was submitted, there was no established system for naming MV genotypes. However, these virus strains were later recognized as belonging to the H1 genotype [18].

In the US, measles vaccination has significantly decreased the number of measles cases. Today, those who developed measles infection and present the common symptoms are the ones who did not receive immunization. However, symptoms such as rash and fever can occur in individuals infected by other pathogens. Thus, to confirm diagnosis, the CDC recommends the use of enzyme-linked immunoassay (ELISA) tests. Currently, ELISA tests are performed on IgM and IgG antibodies to diagnose suspected measles cases. IgM and IgG are detected in individuals exposed to the vaccine or wild-type viruses. IgM is usually present within 1 or 2 months of exposure. However, IgG is still detectable years after exposure. A negative IgM test result meant the person was not infected. Positive IgM capture tests are required by the CDC to confirm the measles diagnosis. Positive results can be obtained when IgM capture tests are performed the day the rash develops. However, if the tests are conducted within 72 hours after the rash is observed, the results may include up to 20 % false-negatives. Alternatively, IgG antibodies test such as ELISA, hemagglutination inhibition (HI), and indirect fluorescent antibody tests are also available. However, the time requirement for IgG tests is much longer

than that for IgM tests. For a complete IgG tests, two serum specimens must be collected. The first serum specimen can be obtained immediately after the rash onset. However, the second specimen is collected 10 to 30 days after the rash onset. The IgG tests for both specimens must be carried out at the same time. Therefore, the diagnosis is delayed until the second specimen is available. Because IgG tests are time-consuming, IgM tests are more preferable for measles diagnosis [1].

1.3 Surface-Enhanced Raman Scattering (SERS) Spectroscopy

Currently, the preliminary diagnosis of MV begins with a review of the patient's exposure history. The MV physical exam often focuses on characteristic symptoms of the disease, including rash and fever. Current diagnosis methods include blood tests such as immunofluorescent antibody assay and ELISA-type assay to confirm the diagnosis [2]. A comparison of mumps-IgM ELISAs conducted by Krause et al.. revealed that the sensitivity of the commercial IgM ELISA assays ranged from 24 to 51% [19]. The low sensitivity was associated with samples from the previously vaccinated patients. This was because IgM response was either delayed or absent in those who received MMR vaccination. However, IgG tests gave positive results for these individuals, confirming the exposure to mumps, possibly through vaccination. Additionally, the specificity was also low for samples from patients with acute infections caused by other agents. The assay low sensitivity resulted in low negative predictive values (NPV). Consequently, a negative IgM test is inconclusive for early phase diagnosis. For diagnosis of mumps in the early phase, within 3 days of onset, IgM assays are efficient for 50 % of the cases [19]. Therefore, because of the low detection efficiency, direct virus detection is the more preferable approach for early diagnosis.

Because of the low sensitivity, a negative IgM-ELISA test was inconclusive for diagnosis of the disease in its early phase. Thus, these commercial IgM ELISA assays were not very effective for viral detection in previously vaccinated patients and those who have the early phase of the disease. Alternative assays such as immunofluorescent antibody assays required sample preparation and a long assay time, approximately 6 – 8 days [20]. One way to improve the detection limit of the immunoassays was to apply reverse transcriptase-polymerase chain reaction (RT-PCR). However, previous study by the Nakayama's group indicated that RT-PCR results are limited by the conditions of the specimens and the primer selection [21]. In patients with natural infections, the detection rate was highest when samples were examined shortly after the onset of rash. For examples, when PCR was performed within 3 days of rash onset, the detection rates were 96.0 % for the NPS sample, 96.4 % for the PBMC sample, and 100 % for the plasma sample. However, detection rate was significantly reduced when PCR was performed 7 days after the onset. H gene nested PCR effectively detected MV genone in all CSF samples of patients with measles encephalitis, acute measles, and SSPE. In addition to the decreased detection rate, there was another limitation to the technique. The sensitivity of RT-PCR changed according to the choice of primers.

Even with RT-PCR amplification, the percentage of positive hits for MV genome detection could fluctuate in the results. A recent study conducted by Riddell et al. showed that RT-PCR yielded a low percentage of positive hits for detection of MV RNA [22]. Specimens were collected from 49 hospitalized children diagnosed with measles. Measles RNA samples were collected from infected Zambian children, within a period of 65 - 118 days after onset. The nucleocapsid (N) and hemagglutinin (H) RNA were detected by RT-PCR. Although IgM and IgG antibodies were detected in all of the children, RT-PCR only detected MV RNA in 18 of the

49 children. The total percentage of positive PCR hits was 37 %. Overall, the low sensitivity of ELISA assays and the limitations of RT-PCR prevented ELISA assays from providing rapid and sensitive viral detection.

Unlike immunoassays, surface enhanced Raman scattering (SERS) spectroscopy does not require complicated sample preparation and allowed for direct analysis of the analytes of interest. Signal enhancement in SERS makes the technique favorable for chemical analysis, environmental monitoring, and biomedical applications. The signal enhancement from the SERS Ag nanorod (AgNR) substrates provides the sensitivity necessary for low level virus detection. Before discussing SERS in details, a review of Raman spectroscopy was provided here to describe the theory behind the technique. Raman spectroscopy is a vibrational spectroscopic technique. The Raman effect is produced by inelastic scattering. Raman scattering was named after Sir C.V. Raman, who was the first to describe the effect in 1928. Inelastic scattering is much weaker than elastic scattering, also called Rayleigh scattering. In Rayleigh scattering, a photon strikes the molecule with an excitation energy at $\tilde{\nu}_0$ frequency. The molecule reemits the light with energy at the same $\tilde{\nu}_0$ frequency. In constrast, in Raman scattering, the energy of the light reemitted is at a lower frequency than $\tilde{\nu}_0$. Stokes Raman scattering occurs when the molecule absorbs some of the energy, and transitions from the ground state to the excited state. Thus, for absorbed energy at $\tilde{\nu}_v$ frequency, light is scattered at a frequency of $\tilde{\nu}_0 - \tilde{\nu}_v$. The opposite is true for anti-Stokes Raman scattering. In anti-Stokes Raman scattering, molecules in the excited state are radiated and release excess energy upon returning to ground state. Thus, the energy of the reemitted light is at a higher frequency, $\tilde{\nu}_0 + \tilde{\nu}_v$. Because of the changes in energy, Raman vibrational bands are defined by the wavenumber shifts from the excitation energy.

Since more molecules populate the ground state at room temperature, Stokes scattering is more common than anti-Stokes scattering for Raman measurements at room temperature. Because the detected Raman vibrational frequencies correspond directly to the surface features of the analytes bound to the substrate, Raman spectroscopy provides a useful method for vibrational analysis of different analytes [23].

In the Raman spectrum, the molecular structure can be determined from the relative intensity of Raman band and frequency. Particularly, for small molecules, the most intense Raman bands arise from totally symmetric vibrations. As for large molecules, the relative band intensities are also dictated by the following general principles, which are based on group intensity approximation. Bond stretching modes produce more intense Raman bands than angular deformation modes. The intensity of a covalent bond stretch is proportional to the bond order. For localized bonds, atoms with larger atomic numbers possess greater vibrational frequencies. For a pair of bond-stretching coordinates, in-phase bond stretch is more intense than out-of-phase stretch. Additionally, cyclic molecules with pseudosymmetric modes usually appeared as strong bands [24].

Since Raman scattering is weak, reduction of stray light is essential. Rayleigh scattering is a major source of stray light in Raman spectroscopy. The Rayleigh line appears on the spectrum as a result of elastic scattering. Rayleigh scattering occurs in greater abundance. Approximately 99.99 % of scattering by molecules in a sample solution is Rayleigh scattering. Because elastic scattering is 10^4 to 10^6 times stronger than Raman scattering, Rayleigh scattering can overwhelm the Raman signal. Thus, the weak Raman signal is improved by rejecting the Rayleigh line [25, 26].

After Rayleigh rejection, the filtered Raman signal is measured by the detector. One type of detectors widely used in Raman spectroscopy is charge-coupled device (CCD) detectors. As described by Ingle and Crouch, CCD is a type of multichannel detector capable of simultaneous detection of dispersed radiation in the focal plane. Current CCD detectors are solid-state sensors in integrated-circuit form; CCD detectors are two-dimensional arrays, containing as many as 4 x 10^6 pixels, on a silicon chip. A thin conducting electrode and p-type silicon substrate covered with insulating oxide make up the components of the pixel. Upon radiation, charges are collected and stored in the metal-oxide semiconductor (MOS) electrodes. Following the generation of electron-hole pairs, about 10^5 to 10^6 electrons are stored in each "potential" well. The charge is moved from pixel to pixel toward the register to generate integrated and amplified signals [27]. The charge accumulation varies linearly with the incident intensity and integration time. The CCD is set up for serial scanning of the stored charge [27].

One disadvantage of CCD detectors is the "blooming" effect that results from strong radiation. "Blooming" occurs when charges from "overilluminated" pixels spread to adjacent pixels. This effect limits the CCD dynamic range at the upper end. Thus, CCD detectors are more useful for weak sources, such as Raman spectroscopy. Furthermore, CCD provides high quantum efficiency (QE) from the UV to near-IR region. The CCD's high sensitivity is useful in Raman spectroscopy; the incorporation of CCD enhances signal-to-noise ratio. Other advantages of CCD include low read noise, low dark current, and readout flexibility [28].

The spectral representation of the signal detected at the CCD detector depends on the type of Raman imaging selected by the operator. In particular, there are three types of Raman imaging. They are single point microscopy, line imaging, and global illumination. In single point imaging, the Raman spectrum is acquired for one point at a time. Moving the sample stage

allows for point-by-point data collection. Thus, single point imaging acquires spatially resolved spectra in a sequential manner. In line imaging, the laser is focused so that spectral data is collected along a line focused on the sample. The scattered light is detected by a CCD. The product is a 2-dimensional representation of the intensity versus the Raman shifts along the imaged line. The simultaneous measurement in line imaging generates a spatially-resolved spectra along the focused line. In global illumination, the laser is defocused to generate an ecclipse-like illumination on the sample. Thus, the illumination covers a larger area on the sample, and a spatially-resolved spectra is obtained for the area. The scattered radiation is detected at the specific Raman shift wavelengths. If the detected shifts are characteristic of the sample components, then the 2D image reflects the chemical composition. The main advantage of these imaging techniques is the reduction in data collection time for a particular pixel dimension. The data collection time is shortest in global imaging. Line imaging has the second shortest collection time. Additionally, the spatial resolution also improves in line imaging because of the simultaneous detection available in line scan. However, the resolution is lower in global illumination. The spatial resolution is dependent on the laser spot size and the collection optics. These two factors, in turn, are limited by diffraction. The Gaussian laser beam diameter, d_l, is directly related to the focal length, f, of the lens and inversely related to the effective diameter of the lens, D. The relationship is shown below [29].

$$d_l = 1.27\lambda\left(\frac{f}{D}\right) \tag{1}$$

Because f/D is also equaled to $f/\#$, d_l is inversely related to the numerical aperture (NA) of the microscope objective. For microscopes, a larger NA value corresponds to a higher resolving power and a higher magnification [29].

$$d_l = 1.27\lambda\left(\frac{1}{2NA}\right) \quad (2)$$

Raman microscopy has been improved through the application of the principles of confocal imaging. Confocal imaging combines the use of spatial filtering with conjugated pinhole diaphrams. The confocal mode focuses the scattered light through the pinhole diaphram. This process constrains the sampling volume. Confocal imaging concentrates the light that comes from the small illuminated region on the sample. Thus, the interferences from the areas that are out-of-focus are removed. Confocal microscopy enhances both the lateral resolution, to a small extent, and the axial resolution, to a large extent. The axial resolution is also known as the depth discrimination. A confocal microscope uses an extra aperture to decrease the depth of focus. A normal objective allows light scattered at different sample depths to enter the spectrometer, given that the depths are within the focused area of the laser beam. However, a confocal objective, through its aperture, has the option of transmitting only light scattered at one of the depths. Consequently, the light scattered at other depths is blocked. If the variations in the composition correspond to different depths, then the confocal aperture enables selective sampling. Therefore, the confocal aperture can confine the sampling depth to a smaller region than that of the conventional optics. Overall, the lateral resolution and the depth discrimination in a confocal microscope are greater than those in a conventional microscope by a power of two. The depth of penetration, Δz, is related to NA as followed. The refractive index is denoted as n [29, 30, 31].

$$\Delta z \geq \pm\frac{4.4n\lambda}{2\pi(NA)^2} \quad (3)$$

Another advantage of confocal imaging is the enhanced lateral discrimination, which decreases the fluorescence background and the interference scattering. However, the lateral discrimination

results in approximately 98 % rejection of the Raman photons. Furthermore, the length of the experiment and the image fidelity are directly proportional to the number of image pixels. Thus, in order to obtain a reasonably high image fidelity, the duration of the experiment must be lengthened [32].

The second type of Raman imaging is line imaging. Line imagining is made possible by line-focused lasers. A laser beam is scanned onto the image plane by a scanning mirror. The microscope objective then focuses the laser beam onto the sample in the shape of a line with an adjustable length. The light is scattered back into the same microscope objective and is focused along the entrance slit of the spectrograph. The CCD detector at the spectrometer simultaneously collects the spatial and spectral data. A process called binning sums the signal of all the pixels in the same column to increase S/N. In conventional Raman spectroscopy, the x-axis of the CCD represents the wavelength, and the y-axis reflects the intensity. In non-imaging Raman spectroscopy, the signal strength is increased for each wavelength detected, when binning is applied [30].

In scanning imaging mode, the x-axis of the CCD follows the Raman shifts in wavenumbers. However, the CCD y-axis column, or the vertical slit axis, is not binned. Instead, the individual pixel along the y-axis of the CCD collects the scattered light from a specific spatial position along the focused laser line. Thus, the y-axis of the CCD is a collection of the spatial positions of the laser line. Thus, the number of spatially collected Raman spectra is directly proportional to the number of vertical pixels. Because this method introduces a spatial dimension to the spectra, the method is often called “hyperspectral” Raman imaging. Hyperspectral Raman imaging provides a combination of spatial and spectral data. Each pixel in the column reflects the intensity of the Raman shift at a spatially resolved position along the laser

line. Meanwhile, the pixels in the CCD rows give the Raman spectrum obtained at each spatially resolved position [30].

In global illumination, the laser is defocused to illuminate a larger area on the sample. This method expands the full view of the microscope. The light from the illuminated area is scattered back into the objective. The scattered light is then filtered by a Rayleigh band rejection (BR) filter and a wavelength selector. The BR filter blocks the light produced from elastic scattering, or the Rayleigh line. The wavelength selector allows only the light within the chosen range of Raman shifts to reach the detector. The use of a BR filter and a wavelength selector in combination with the CCD generates a selective image of the Raman-scattered photons. To obtain Raman spectra and 2D images, the wavelength selector undergoes step scanning for a range of Raman shifts; a CCD image is acquired at each Raman shift. Unlike normal Raman spectroscopy and line-scanning, 2-D Raman global illumination imaging collects the spatial data along the sample at the selected Raman shift for both the x and the y-axes. Thus, the result is a "data cube" of the spectral intensity versus the 3-D position and the Raman shift wavenumbers [32].

Because Raman scattering is weak, without signal enhancement, Raman spectroscopy is not an effective method for sample detection at low concentration. However, the development of surface-enhanced Raman spectroscopy (SERS) allows for Raman measurement with greater signal enhancement. SERS signal enhancement can be described by the classical theory. According to the theory, incident light striking the metal surface induces an oscillation dipole in the metal particles. Light is then scattered from the oscillating particle at the dipole frequency. The dipole moment is expressed as followed.

$$\mu(t) = P \cdot E_{inc}(t) \tag{4}$$

E_{inc} and P are the incident electric field and the molecular polarizability, respectively. Since the Raman intensity is proportional to the square of the dipole, increasing the polarizability or the incident field enhances the Raman signal. Enhancement due to the molecular polarizability is also known as the molecular effect or chemical effect; the electric field enhancement is called the field effect or electromagnetic effect. Under the field effect, signal is enhanced in the presence of additional local fields produced by electromagnetic resonances. The chemical effect results from the molecular polarizability changes that occur when molecules interact with the metal surface [33].

In theory, the two major effects that contribute to SERS are the electromagnetic and the chemical effects. The key factor that contributes to the electromagnetic enhancement in SERS is the surface plasmon resonance. Surface plasmons are described as waves of electrons bound to the metal surface [34, 35]. Radiation of the metal surface causes excitation of surface plasmons. Polarization of the metal nanoparticles occurs at the plasmon frequency. Polarized metal nanoparticles generate local fields that can add to the laser electromagnetic field. Additionally, the analyte molecule within proximity of the enhanced field is polarized, and emits an amplified Raman field. This amplified Raman field polarizes more metal nanoparticles. The induced polarization leads to greater overall Raman enhancement [36, 37].

The second effect that contributes to the signal enhancement in SERS is the chemical effect. In the chemical effect, the molecular polarizability changes when the molecule interacts with the metal surface [38, 39]. One of the models that is used to describe the chemical effect is the charge-transfer model. According to the model, radiation of the metal surface induces an electron transfer from the metal to the charge-transfer excited state of the absorbed molecule. The formerly neutral absorbed molecule becomes negatively charged. Upon nuclear relaxation,

the absorbed molecule returns the electron to the metal. In the process, the molecule becomes vibrationally excited and emits a Raman-shifted photon.

Without signal enhancement, "normal" Raman scattering has a cross-section of approximately 10^{-29} cm^2. However, an enhancement factor of 10^{14} could expand the Raman scattering cross-section to 10^{-16} cm^2. The larger cross-section allows for application of SERS in single-molecule detection. The use of SERS is more preferable than fluorescence because SERS spectra exhibit more detailed chemical structure of the analyte molecules. The ability of SERS to detect single molecule was shown in the detection of single enkephalin protein on fractal silver surface. Exposure of a 1 μm laser spot between 95 – 115 seconds produced SERS spectra displaying a characteristic phenylalanine band at 1000 cm^{-1} for the protein [40].

In addition to detecting single molecules, SERS can also detect microorganisms. In 1998, a study by Efrima et al. showed SERS signal was measured for Escherichia coli bateria coated with Ag colloids. Additionally, because of the large signal enhancement, SERS spectra can be collected at low laser power. Thus, SERS is applicable in studies of live bacteria. Other applications of SERS include extrinsic SERS detection of antigens in immunoasays. As an immunoassay method, extrinsic SERS has advantages over fluorescence techniques. SERS spectra displayed smaller spectral overlaps for reporter molecules. Additionally, a single source can be used to excite different labels in the assay [41].

The potential of extrinsic SERS as an immunoassay method was demonstrated in a study conducted by Driskell and colleagues [42]. In the study, extrinsic SERS was applied in the detection of feline calicivirus (FCV). Experimental results indicated that the salt concentration, the extrinsic Raman label (ERL) binding buffer, and the sample agitation were factors that control the binding efficiency of FCV. In the immunoassay, FCV-specific monoclonal

antibodies (anti-FCV mAb) were immobilized on gold substrates coated with dithiobis (succinimidyl propionate) (DSP). These immobolized antibodies were then exposed to FCVs in the cell culture media. Subsequently, the captured FCVs were allowed to bind ERLs for SERS detection. The ERLs were composed of 60-nm gold particles coupled to anti-FCV mAb and 5,5'-dithiobis(succinimidyl-2-nitrobenzoate) (DSNB). FCV binding was confirmed by measuring the SERS signal of the nitro group from DSNB molecules attached to Au nanoparticles. The effect of ionic strength on viral binding was determined by varying the concentration of NaCl solution. It was found that the number of captured viruses increased at higher NaCl concentration. Driskell and colleagues determined that the optimum conditions for the FCV immunoassay were 2000g centrifugation, 20 μg mAb, 1 % BSA, 8.5 pH, and 150 mM NaCl. Overall, the experiment showed SERS was an effective immunoassay method. However, a disadvantage of the technique is that extrinsic SERS requires a more complicated sample preparation than intrinsic SERS. For example, in this study, the sample solution must be prepared so that the salt concentration favored viral binding without causing ERL aggregation. Additionally, the ERL detection was dependent on the binding efficiency of FCVs to the substrate-bound and ERL-bound anti-FCV mAbs. In contrast, intrinsic SERS enables direct detection of the virus, thus eliminating limitations associated with label binding efficiency.

The potential of intrinsic SERS as a viral detection method was demonstrated in a study conducted by the Dluhy's group [43]. Results from the study showed that Ag nanorod (AgNR) substrates provided the high sensitivity necessary for low level viral detection in SERS spectroscopy. Additionally, the structural information presented in SERS spectra also enabled classification of the viral samples. In the study, reproducible SERS substrates were prepared using the oblique angle deposition (OAD) method.

In oblique evaporation, the evaporated films grow according to the shadowing effect and surface diffusion [44]. For evaporation, the metal atoms must have the necessary escape energy to leave the source. The escape energy decreases when evaporation is carried out at higher temperature. According to Abelmann et al., the adatoms continue to move across the surface by hopping from one nucleation site to the next. The atoms evaporate if they are not buried by newly arriving atoms. Adatom monolayer formation could be interfered by the presence of residual gasses in the evaporation chamber. Contaminant gases may include H_2O, CO, CO_2, N_2, and H_2. Additionally, hydrocarbons from pump fluid could also contribute to the contamination.

To initiate the film growth, the first atoms arriving at the metal surface form the nucleation sites. Because of the shadowing effect, once the atoms adsorbed to the surface, the area behind these adatoms are left vacant because of the atoms' shadow. Thus, secondary nucleation does not occur in the shadowed areas. The next atoms arriving at the surface adsorbed at the nucleation sites. Subsequent adsorption leads to the formation of chains of atoms at the nucleation sites. In surface diffusion, because of conservation of parallel momentum, the adatoms move a short distance parallel to the incident vapour beam upon arrival at the substrate. The adatoms continue to move across the substrate until they reached the trapping point. The trapping point is determined by the oxygen-metal concentration ratio. The concentration ratio is different on the beam-side and shadow side of the nucleus. On the beam-side, the oxygen concentration is less than the metal concentration. However, on the shadow side, the oxygen concentration exceeds the metal concentration. When the relative oxygen concentration reached the critical value, the adatoms are trapped. The position of the trapping point dictates the cross-sectional shape of the nucleus. Elongation of the nuclei may occur either parallel to or perpendicular to the vapour incidence plane [44].

In the film growth analysis, Alberman et al. observed that increasing the oblique incident angle increased surface diffusion [44]. Atoms arriving at the surface lose the perpendicular component of their kinetic energy, but not the parallel component. This is known as conservation of parallel momentum. Additionally, it was suggested that for conservation of total momentum, the lost in the perpendicular component results in a gain in the parallel component. If the incident angle is perpendicular, atoms scatter parallel to the film surface but in random directions. However, if the incident angle is increased, atoms scatter parallel to the direction of the vapour beam projection. Therefore, increasing the oblique incidence increases directional surface diffusion.

These SERS substrates served as biosensors for rapid and cost-efficient low-level viral detection. Previous studies by Shanmukh et al. showed that adenovirus, rhinovirus, and HIV viruses can be detected by SERS spectroscopy [43]. SERS spectra displayed characteristic structural variations for each virus type. Thus, the viruses were differentiable based on visual inspection of the spectra. To confirm that the SERS signal was indicative of the virus and not just the cell lysate media, SERS spectra were collected for uninfected Vero cell lysate (VCL), respiratory syncytial virus (RSV)-infected cell lysate (RSV), and purified RSV. The spectra of infected cell lysate and pure RSV shared some common bands at 1066, 835, 545, and 523 cm^{-1}. However, these bands were absent in the uninfected Vero cell lysate spectra. This shows that SERS signal was measured for both the virus and the cell lysate in the infected cells.

SERS structural selectivity was further tested in detection and classification of different strains of influenza viruses and RSV [43]. Overall, for both types of viruses, the spectra data show variations in frequency shifts that allowed for differentiation of the virus strains [43]. The results suggested SERS was an effective method for viral detection. Current diagnostic

methods for viral detection such as ELISAs and fluorescent antibody assays have low sensitivity. Thus, these methods are often coupled with PCR to improve detection. As a result, the methods are time-consuming and costly. Additionally, the poor sensitivity limits the efficiency of these techniques in low-level viral detection. Compared to the current diagnostic methods, SERS provides a more promising alternative for viral detection because of the higher sensitivity and the structural information presented in SERS spectra.

1.4 Fourier Transform Infrared (FTIR) Spectroscopy

In the past, the common method for pathogen detection has been Fourier transform infrared (FTIR) spectroscopy. Before discussing FTIR detection, the following review provides a background on the technique of infrared spectroscopy. Infrared (IR) radiation was first discovered by Sir William Herchel in the 1800's. After discovering that there was thermal radiation beyond the red spectrum, Herschel used the thermometer as a detector and measured the intensity as a function of the distance beyond the red spectrum. Thus, he recorded the first IR spectrum. In the 1930's, diagnostic IR had its first application in chemical technology and industry. Around the time of WWII, commercial IR spectrometers were introduced to the market by Beckman and Perkin-Elmer. Since then, IR spectroscopy has become a diagnostic and analytical tool in the chemical laboratory [45].

The IR wavelength region is located between 2 – 25 μm; in wavenumbers (cm^{-1}), this region corresponds to 5000 – 400 cm^{-1}. Infrared energy is not strong enough to induce electronic transition or break chemical bonds. However, because the quantized energy for molecular vibrational excitation falls within the IR wavenumber region, IR spectroscopy can be applied in vibrational studies of different molecules [45].

In IR spectroscopy, the vibrational frequency, ν, of a diatomic molecule with masses m_1 and m_2 is expressed as followed.

$$\nu = \frac{1}{2}\pi\sqrt{kG} \qquad \nu = \frac{1}{2}\pi\sqrt{\frac{k}{\mu}} \qquad (1, 2)$$

$$\frac{1}{\mu} = \frac{1}{m_1} + \frac{1}{m_2} = G \qquad (3)$$

The force constant k increases with bond orders. k of single bonds is less than k of double bonds and triple bonds. As shown in equation 2, ν is proportional to the force constant, k, and reduced mass, μ. Since atoms in a molecule possess different atomic masses and form different bonds, distinct vibrational frequencies are detected according to the types of atoms and bonds present in the molecule. Thus, the chemical structure of the molecule can be determined from the characteristic IR wavenumbers. For example, the IR spectrum of CH2=CH-C≡N diplays C-H vibrational stretch in the 2600 – 3600 cm-1 region, ν (C=C) around 2000 – 2400 cm^{-1}, and ν (C≡N) around 1500 – 1800 cm^{-1} [45].

In polyatomic molecules, multiple atoms, bonds, and modes of vibrations are present. Considering a simple molecular system with 2 vibrational motions and different force constants along the x and y axes, the atom that is stretched along the x and y axes return to the center in a straight line. However, if the atom is stretched at an angle between the x and y axes, the restoring force will not pull the atom back to the center. Thus, the atom does not vibrate in a straight line. Similar to describing the off center position of the atom with respect to the x and y axes, vibrational patterns can also be described in terms of modes of straight line vibration. In other word, vibrational patterns of atoms moving back and forth in a straight line can be defined in terms of normal modes. The number of normal modes equal the number of vibrational

degrees of freedom. For n-atomic molecules, there are 3n degrees of freedom. Accounting for 3 translations and 3 rotations, nonlinear molecules have 3n – 6 normal modes. Since linear molecules only have 2 rotations, linear molecules have 3n – 5 normal modes. The magnitude of the displacement during vibration are defined as variable or normal coordinates [45].

A molecule's normal vibrations can be defined with respect to the vibrational symmetry. Symmetry operations are applied to analyze the vibrational data of the geometry of the molecule's equilibrium configuration. The normal modes are given in terms of symmetry species, such as "irreducible representations." The vibrational modes of the species are symmetric. The remaining modes are antisymmetric or asymmetric with reference to the symmetry operations. One type of symmetry operation used to analyze vibrations is inversion. For a molecule with a center of inversion, the normal modes of vibration can fall into one of two categories: symmetric or antisymmetric, or g or u, respectively. The rule of mutual exclusion states that molecules maintaining symmetric (g) vibrational modes, after undergoing an inversion operation, are not infrared-active. In contrast, antisymmetric (u) molecules are not Raman-active [24].

Overall, the normal-mode frequencies are dependent on the atomic mass, molecular geometry, and interatomic forces. The interatomic forces, under the harmonic approximation, are the sum of force constants contributing to the force fields. The atomic masses and molecular geometry are usually known variables. However, it is difficult to determine the force field, even with methods that can calculate the force constants of small molecules. A variable used in force field calculation is group frequencies. Group frequencies are normal mode vibrations characteristic of the functional groups on the molecule. For instance, in infrared spectroscopy, the vibration at 1700 cm^{-1} is assigned to the carbonyl group. However, since the mode is not

completely localized to a functional group, the frequency can vary according to substitutions on neighboring atoms [24].

Because of its distinction and sensitivity, group frequencies are useful in vibrational analysis of large molecules. For example, in the biological field, large molecules of interest such as amides, polypeptides, and proteins were known to display a characteristic band at 1660 cm^{-1} in both the Raman and infrared spectra. In a later study, N-methylacetamide was also found to have a normal vibration at 1660 cm^{-1}. Further analysis indicated 80% of the mode was due to the C=O stretch; C-N and N-H bendings each contributes the remaining 10%. This mode was later defined as the amide I vibration, which is distinctive of peptide linkage. Shifts in the amide I band are useful for reconstructing the geometrical configuration of the molecule. In addition to the amide I band, amide II and amide III bands, characteristic of localized N-H bending and C-N stretching, respectively, are also used to describe peptide linkage [24].

Even though IR spectroscopy provides detailed information on the chemical composition, the high detection limit restricts the technique's application to bulk samples. However, coupling optical microscopy to IR spectroscopy provides high throughput for detection of complex materials. Throughout the past 15 – 20 years, advances in IR were made through the combined use of IR microscope, IR interferometers, and sensitive detectors. These advances contributed to the development of Fourier transform infrared (FTIR) microspectroscopy [46].

In FTIR, Fourier transformation applies a mathematical operation to reduce complex waveforms into simple components. According to Fourier, a mathematical function can be approximated from the superposition of sine and cosine waves. For instance, the wave function F(x) is defined as follows.

$$F(x) = a \cdot \sin(x) + b \cdot \sin(2x) + c \cdot \sin(3x) \quad (4)$$

The process of Fourier transform (FT) is applied to calculate the coefficients a, b, and c. Another example is the Gaussian curve. The Gaussian curve is correctly defined as a mathematical expression or as an approximation of cosine waves. In spectroscopy, the spectral domain is acquired by defining the signal variation as a function of another mathematical expression. The frequency domain is generated from the coefficients of the cosine function. The Fourier transform is used to determine the sine and cosine coefficients, and to describe one domain representation in terms of the other. For a Gaussian function, the cosine coefficients that compose the frequency domain are the real component of the Fourier transform. In contrast, the sine coefficients are the imaginary component in the Fourier transform. The sum of the squares of the real and imaginary components gives a spectrum that is dependent on the frequency. An example is shown in the power spectrum, P(S); $P(s) = S(s)^2 + C(s)^2$. S(s) and C(s) are the sine and cosine components, respectively. The magnitude spectrum can be generated in a similar fashion, and is preferred over the power spectrum because the magnitude spectrum is related to the concentration. In addition, in the time domain, the spectral bandwidth is inversely proportional to the magnitude and the power spectra. Consequently, a larger bandwidth corresponds to a narrower power spectrum and limited high frequency information [47].

In Fourier transform infrared (FT-IR) spectroscopy, the Fourier transform of a theoretical interferogram, I(δ), and spectrum, $E(\tilde{\nu})$, follows these equations.

$$I(\delta) = \int_{-\infty}^{\infty} E(\tilde{\nu}) \cos(2\pi\tilde{\nu}\delta) d\tilde{\nu} \tag{5}$$

$$E(\tilde{\nu}) = \int_{-\infty}^{\infty} I(\delta) \cos(2\pi\tilde{\nu}\delta) d\delta \tag{6}$$

However, the experimental interferogram contains an additional variable, the phase angle, θ. The modified interferogram function is shown below.

$$I(\delta) = \int_{-\infty}^{\infty} E(\tilde{\nu}) \cos(2\pi\tilde{\nu}\delta - \theta) d\tilde{\nu} \tag{7}$$

Most likely, the additional phase components results from non-symmetrical sampling and phase lags in electronic filtering. Another way of expressing cos(α - β) is cos α · cos β + sin α · sin β. Thus, the phase angle introduces the sine components in the cosine wave interferogram. Given that cosθ and sinθ are orthogonal and that the Euler's relationship states cosθ – isinθ = $e^{-i\theta}$, the FT equations above can be rewritten to account for the phase angle in the forward and inverse Fourier transform.

$$E(\tilde{\nu}) = \int_{-\infty}^{\infty} I(\delta) e^{2\pi\tilde{\nu}\delta} d\delta = \mathfrak{F}\{\mathrm{I}(\delta)\} \tag{8}$$

$$I(x) = \int_{-\infty}^{\infty} E(\tilde{\nu}) e^{-2\pi\tilde{\nu}\delta} d\delta = \mathfrak{F}^{-1}\{E(\tilde{\nu})\} \tag{9}$$

$\mathfrak{F}\{\}$ and $\mathfrak{F}^{-1}\{\}$ represent the forward and inverse Fourier transform, respectively. The inclusion of the phase components relate I(δ) to $E(\tilde{\nu})$ in the complex Fourier transform. The sine components correlates to the imaginary part, $i = \sqrt{-1}$, in the complex transform. Thus, the cosine FT is the real component of the transform, and the sine wave is the imaginary component [48, 49].

The addition of phase requires the FT to include both the sine and cosine components. Thus, the complex form of the Fourier transform is used instead of the simpler cosine transform to allow for phase correction. Phase correction combines the sine and cosine transforms. The phase is a measure of the magnitude of sine and cosine components at each frequency. In the

event of a phase shift, the cosine curve does not have a cosine value of 1 at X = 0. Additionally, in the presence of a phase shift, the sine transform is no longer a zero component. Therefore, a phase shift produces a cosine data that is out-of-phase with the sine and cosines transforms. The phase shift is determined below.

$$\Phi(s) = \arctan\left(\frac{-S(s)}{C(s)}\right) \tag{10}$$

The phase spectrum of the real (cos) and imaginary (sin) components are denoted as Φ(s). The phase-corrected spectrum, T(s), is expressed as follows.

$$T(s) = C(s) \cdot \cos[\Phi(s)] + S(s) \cdot \sin[\Phi(s)] \tag{11}$$

The major source for the phase shift is the FT transform sampling because the operator does not have full control over sampling. Fourier transform is applied based on the assumption that the waveform begins at the exact maximum and minimum of the cosine and sine curves, respectively. However, this is not the case for actual sampling because the data is not obtained at these exact values. Thus, the phase correction compensates for the sampling error. The magnitude of the real and imaginary components reflect the deviation of sampling from the expected maximum and minimum [48, 49].

Sampling the waveform is done properly by adhering to the Nyquist sampling criterion. According to the criterion, "the waveforms of sinusoidal function of time or distance can be sampled properly at frequency greater than or equal to twice the bandwidth of the system." This means the sine and cosine curves can be sampled correctly when sampling is done at least twice every oscillation. Thus, the maximum number of unique frequencies for a sine function with N elements is N/2. Digitizing data below the Nyquist frequency results in aliasing; the data are digitized at frequencies other than the true frequencies. The aliased frequency is the difference

between the sampling frequency and the cosine frequency. To avoid aliasing, low-pass filters are used to sample at the selected sampling rate. Thus only N/2 or greater frequencies are digitized [50, 51].

In FTIR, interferometers are an alternative to dispersive monochromators. In dispersive spectrometers, the level of radiation arriving at the detector is controlled by the entrance and exit slits. The resolution of these monochromators depend on the amount of radiation passing through the exit slit. Furthermore, the throughput of the device is the ratio of the incidence and the exiting radiation. Therefore, the size of the exit slit limits the resolution and the throughput of the dispersive spectrometers. Because the dispersive monochromators have low throughput, interferometers are a better choice of spectrometers. One popular type of interferometers is the Michelson interferometer. This interferometer utilizes a beamsplitter to separate the source radiation and direct the separate beams to two different mirrors. Generally, the beamsplitter is composed of material that is half reflective and half transmissive. Different beamsplitter materials are available for the near-IR, mid-IR, and far-IR spectral ranges [48, 51].

Optical retardation is observed in a Michelson interferometer. The optical retardation, δ, is defined as the path difference of the two arms of the interferometer. The retardation is expressed below.

$$\delta = 2(OM - OF) \tag{12}$$

OM defines the distance between the beamsplitter and the moving mirror; OF is the distance between the beamsplitter and the fixed mirror. When there is no retardation and the path difference is zero, the separated beams are in-phase. With zero retardation, the beams recombine at the beamsplitter to produce a constructive interference. When the moving mirror is displaced by ¼ of the wavelength, λ, the radiations are out-of-phase, and their recombination produces

destructive interference. However shifting the mirror to an additional λ/4 gives constructive interference. Thus, the interferometer can be used as a modulator to relate the detector response, I(δ), to the retardation interval, δ.

$$I(\delta) = 0.5 \cdot I(\tilde{\nu})\{1 + \cos(2\pi\tilde{\nu}\delta) \quad (13)$$

$0.5 \cdot I(\tilde{\nu})$ is the constant direct current (DC) component, and $0.5 \cdot I(\tilde{\nu})\{1 + \cos(2\pi\tilde{\nu}\delta)$ is the modulated alternating current (AC) component. The AC component is also known as an interferogram, and is recorded to generate the spectrum [48].

FT spectroscopy is chosen over dispersive spectroscopy because of three major advantages. These advantages are the Fellgett's Advantage (multiplex advantage), Jacquinot's Advantage (throughput advantage), and Conne's Advantage (precision advantage). Because of these advantages, the signal-to-noise ratio (S/N) of FT spectrometers is higher than that of dispersive spectrometers. According to the Fellgett's Advantage, FT spectroscopy allows for shorter measurement time. Because the interferometers do not use slits, the radiations can propagate through the interferometer without attenuation. Additionally, the radiation is measured at all times. Thus, this advantage is also referred to as the multiplex advantage. The S/N of dispersive spectrometers are defined as follows.

$$\frac{S}{N} = \sqrt{\frac{t}{n}} \quad (14)$$

The variables t and n stand for the measurement time and the number of resolution elements, respectively. The variable n is found by dividing the spectral range by the resolution. In contrast, S/N of interferometers are only related to t.

$$\frac{S}{N} = \sqrt{t} \quad (15)$$

S/N of an interferometer is related to S/N of a dispersive spectrometer as follows.

$$\left(\frac{S}{N}\right)_{imp} = \frac{\left(\frac{S}{N}\right)_{\text{int}}}{\left(\frac{S}{N}\right)_{disp}} = \frac{\sqrt{t}}{\sqrt{\frac{t}{n}}} = \sqrt{n} \tag{16}$$

Therefore, the S/N of an interferometer is improved by a factor of $\sqrt{n}$. Thus, the multiplex advantage is more significant at higher resolution. In other word, for a given S/N in a dispersive monochromator, the measurement time required for obtaining the same S/N in an interferometer is much shorter. As seen in equation 16, the measurement time in interferometer is 1/n the time in a dispersive spectrometers [48, 49, 50].

Another advantage of FT spectroscopy is the Jacquinot's Advantage, also known as the throughput advantage. According to the Jacquinot's Advantage, S/N is improved by increasing the strength of the signal observed at the detector. The optical throughput of a grating IR spectrometer and interferometer are shown in equations 17 and 18, respectively. The units for the throughput are expressed in cm^2 sr.

$$\theta_G = \frac{hA^G \Delta\tilde{\nu}}{fa\tilde{\nu}^2} \tag{17}$$

$$\theta_I = 2\pi A^I \frac{\Delta\tilde{\nu}}{\tilde{\nu}_{\max}} \tag{18}$$

The throughput advantage is shown in the ratio of θ_I/θ_G.

$$\frac{\theta_I}{\theta_G} = \frac{2\pi A^I fa\tilde{\nu}^2}{A^G h\tilde{\nu}_{\max}} \tag{19}$$

The irradiated area on the interferometer mirrors is A^I; the illuminated area on the grating is defined as A^G. The variables f, h, and a stand for the focal length of the collimating mirror, the slit height, and the grating constant, respectively. In practice, the detector's optical properties

have the most significant effect on the FTIR throughput. Finally, the Conne's Advantage is the precision advantage in FT spectroscopy. According to the Conne's Advantage, the inherent wavenumber scale in FT spectrometers allows for higher precision. The precision is a result of referencing the moving mirror in the interferometer to the internal laser source. The precise measurement of the mirror position lead to the precise measurements of the wavenumber scale [48, 49, 50].

Advantages of FTIR combined with sensitivity encourage the application of FTIR in pathogen detection. In particular, a novel assay method called infrared fluorescent immunofocus assay (IR-FIFA) was proposed by the Counihan's group to be a promising approach for quantitation of non-cytopathic and mimimally cytopathic viruses in cell culture media [20]. In the assay, cell monolayers were fixed with formaldehyde, stained with Alexa Fluor 680 fluorophores, and exposed to monoclonal antibodies. The foci of infection were treated similar to plaques units in plaque assays. The foci were detected by IR imaging and quantified in units of focus forming units (FFU) per mL. IF-RIFA detection of MV was performed on 10x-dilutions of viral samples. Foci formation was observed 3 days after innoculation. The optimal incubation time for fixed and stained cells was determined to be 5 days. IR-FIFA have several advantages over other viral assays such as radioimmunofocus assay (RIFA) because the detection is not affected by the radioactive decays of the tags and foci can be measured directly from the cell culture plates. However, a disadvantage of the technique is the long assay time. The total assay time is around 6 - 8 days. Thus, because the optimal time required for modifying and staining cells, it would take at least a week before viral quatitation results are available. This suggested that for rapid detection of pathological agents, direct IR detection would be less time-consuming and more efficient than fluorescence infrared spectroscopy.

In the past, FTIR has been used to detect pathological agents in animals and humans. A study conducted by Thomzig et al. showed that FT-IR detection of PrP27-30 prion protein allowed for identification of four different types of transmissible spongiform encephalopathies (TSEs) [52]. These TSEs are scrapie-causing agents in humans and animals. The following scrapie isolates were obtained from Syrian hamsters, 263K, 22A-H, ME7-H, and BSE-H. BSE strain is the agent responsible for Creutzfeldt-Jacob disease. From the second derivative FTIR spectra of the four isolates, the amide I region showed the most spectral variations for the four types of isolates. The secondary structure of the isolates was analyzed through the β-Sheet, α-Helix, and structural turn bands. The bands correlated to the C=O stretching vibrations in the amide I region. In the past, it was difficult to distinguish ME7-H from 22A-H because the two strains have similar electrophoretic mobilities and glycosylation patterns. However, variations presented on the FTIR spectra enable differentiation of these two strains. This signifies the first successful attempt at identifying the two TSE strain via FTIR detection of PrP27-30 prion protein. Not all conformational differences of PrP27-30 are detectable in immunobiochemical typing, which is the current diagnostic method for TSEs detection. Thus, the promising results from this study suggested that FTIR may be the more effective approach for TSE detection. Progress in TSE detection and identification will enhance epidemiological surveillance of TSEs.

Another study of scrapie agents by Lasch and colleagues showed the combined use of FTIR and chemometrics enabled detection and identification of 263K scapie agents from Syrian hamster sera [53]. The infected sera were collected 70, 100, 130, 160 days after infection. After 160 days, the sera were in the terminal stage of the infection. Previous studies have shown that terminal scrapie differed from preclinical scrapies in both structure and conformation. In this study, reproducible variations between terminal and preclinical scrapie spectra confirmed the

existence of structural and conformation differences. Hierarchical classification of the two types of scrapie was developed using the artificial neural networks (ANN). Preclinical sera were distinguished from the negative controls base on the FTIR spectral changes. Furthermore, the spectral changes reflect the molecular differences in infected sera. Thus, FTIR detection of scrapie provides a mean for identifying TSE-infected hosts. The reported sensitivity and specificity for the detection were 92 % and 96 %, respectively.

Before extending its application to TSEs studies, FTIR has been a popular diagnostic tool for bacterial detection and classification. For example, through the use of FTIR spectroscopy, the Helm's group was able to generate a spectral library for 97 strains of Gram-positive and Gram-negative bacteria [54]. For the study, the samples were prepared by suspending bacterial colonies in distilled water. A 25-μL aliquot of the sample was applied to the ZnSe optical plate. The data was collected over a range of 4000 - 500 cm^{-1}. Data analysis was performed on Savitzky-Golay smoothed derivatives. For strain classification, the spectral distance between data was determined by first calculating the Pearson's product correlation coefficients. Then, the correlation coefficients were converted to spectral distance. After determining the spectral distances according to the selected variable regions in the spectra, the spectral library was challenged against a test set composed of 72 of the 97 strains. The classification rates were 94.4 % for separations on the species level and 83.3 % for separation on the strain level. Overall, results from these previous studies indicated that FTIR coupled with chemometric analysis provided an effective approach for sample discrimination.

1.5 Chemometric Analysis

Previous studies have shown that the key advantage of FTIR was rapid bacterial detection without complicated sample preparation [55]. Furthermore, results from the studies indicated the

combined use of FTIR and chemometric analysis allowed for accurate identification of pathological agents. In the studies, to prepare the data for chemometric analysis, preprocessing by normalization with respect to the most intense band was performed. To corrrect for baseline shifts, the data was transformed by taking the first and second derivatives of the Savitzky-Golay smoothed spectra. In chemometric analysis, unsupervised methods such as PCA and HCA were applied to group samples based on statistical similarities and differences. PCA reduced the dimensionality of the data. HCA classified the samples based on interspectral distance. In contrast, supervised methods such as discriminant analysis (DA) separated the classes by decreasing intra-group variability and increasing inter-group variability. Groups were also separated based on the Mahalanobis distance. DA used the training set to create a classification model for the unknowns. The models were then validated by methods such as the "leave-one-out" method. The models yielded a high classification rate for separation of classes. However, the rate decreased for strain separation. Nonetheless, these past studies have shown that chemometric analysis is powerful tool for sample classification.

In chemometric analysis, pattern recognition is divided into two categories, unsupervised and supervised learning [56]. Under unsupervised learning, the samples are not assigned to known classes. Instead, the samples are clustered naturally based on similarities and differences in the measurements. Then, the natural clustering is compared to the grouping of known classes to assess the efficiency of the sample separation. One type of unsupervised method commonly used in chemometric analysis is Principal Component Analysis (PCA). PCA detects and combines variables that contributes significantly to the patterns in the data. PCA scores plots reflect the amount of variance in the data matrix. In multivariate analysis, the variables can be displayed as different dimensions in the same window. However, it is difficult to analyze the

relationship between samples based on more than 3 variables at one time. PCA reduces the dimensionality by assigning a smaller number of factors to represent multiple variables. With the factor assignment, window axes can be defined by factors instead of individual variable. The new axes are also known as principal components (PCs). PC coordinates are also called "scores". The first principal component (PC1) gives the direction of the maximum variation in the data set (Fig. 1.1). PC2 gives the direction of the second greatest variation [56].

PCs are generated from combining the original variables. Thus, the variable contribution to the total variation is reflected in the alignment of the PC axes to the variable axes. For example, for a variable 1 versus variable 2 comparison, if PC1 lies close to variable 1 axis, this means variable 1 makes a greater contribution to PC1 than variable 2. The variable contribution is determined by the $\cos\theta$ value. θ equals the angle formed between the variable and PC axes. The $\cos\theta$ value are termed "loadings". PC loadings range from -1 to 1. If loadings value is 1 for PC1, $\theta = 0$, and all variations of that variable axis can be fully described by PC1. In constrast, if loadings value equals 0 for PC1, none of the variations of that variable axis is described by PC1 [56].

Generally, the optimal number of PCs selected for an analysis depends on the desired percentage of the total variation. Because noise is distributed among all PCs, selecting too many PCs will result in including too much noise contribution in the analysis. In general, the signal-to-noise (S/N) ratio decreases for subsequent PCs. Ideally, excluding these PCs helps filter noise from the data. However, the operator must decide which percentage of variable contribution corresponds to the noise level to avoid removing real data [56].

In addition to PCA, another popular unsupervised method in chemometrics is Hierarchical Cluster Analysis (HCA). HCA groups the samples based on the distance between

the data points [56]. A 2-dimensional plot of the sample grouping is presented in the form of a dendogram to allow for human pattern recognition. Samples with the least row-space (sample-to-sample) distance are clustered together in the dendogram (Fig. 1.2). Clustering is repeated until all the sample groups are linked in the dendogram. There are two routes to linking the clusters in an HCA dendogram. The single-link approach clusters the sample with the nearest neighbor. In the dendogram, the vertical bars indicate which samples are linked in the cluster. The units of the x-axis represent the distance between the samples in the cluster. The position of the vertical bars reflects the distance between linked clusters. In the centroid-link approach, the samples are clustered based on the distance between the centroid of each cluster. The centroid equals the average of all the points in the cluster.

In HCA, the sample distance is determined by calculating either the Euclidian distance, d(x,y) or the Mahalanobis distance between the sample and the reference point. The Euclidian is defined as followed [57].

$$\mathrm{d}(x,y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \ldots + (x_{\mathrm{nvars}} - y_{\mathrm{nvars}})^2} \qquad (1)$$

The coordinates of the variables are represented by x_{nvars} and y_{nvars}. In calculating the Euclidian distance, the region that encompasses the set of variables is a sphere. Thus, the variables are considered to be at an equal distance from the reference point. Calculation of the Mahalanobis distance, however, extends this sphere to accommodate for scaling of different variables and for correlation of variables [57]. As a result, the Mahalanobis distance region resembles an ellipsoid. As mentioned earlier, in the Euclidian distance calculation, all components of x are positioned at an equal Euclidian distance from the center. However, a more statistically appropriate representation of the distance should reflect the variability of the components. Sample weighting should give components with low variability more influence on the data

analysis. Thus, the Euclidian distance equation is modified to give the rescaled components for the Mahalanobis distance equation [57].

$$u = \left(\frac{x_1}{s_1}, \ldots, \frac{x_p}{s_p} \right) \qquad v = \left(\frac{y_1}{s_1}, \ldots, \frac{y_p}{s_p} \right) \tag{2, 3}$$

$$\mathrm{d}(x, y) = d_E(u, v) = \sqrt{\left(\frac{x_1 - y_1}{s_1} \right)^2 + \left(\frac{x_2 - y_2}{s_2} \right)^2 + \ldots + \left(\frac{x_{\mathrm{nvara}} - y_{\mathrm{nvars}}}{s_{\mathrm{nvars}}} \right)^2} \tag{4}$$

A common classifying method in HCA is K-Nearest Neighbor (KNN) analysis. KNN places the unknown in the sample class closest to its multidimensional space [57]. The proximity of the samples is determined either by the Euclidean distance or the Mahalanobis distance. In KNN, a 3-D spherical region encompasses the unknown and the K nearest neighbors. The unknown sample is assigned to the class with the most nearest neighbors (Fig. 1.3). For example, unknown X has 2 nearest neighbors from class A and 1 nearest neighbor from class B. In this case, unknown X is assigned to class A.

In addition to KNN analysis, another supervised method used in classification is partial least squares (PLS) analysis. In the past, partial least squares (PLS) have been used in discrimination of consumable products such as coffee beans, red wheat, and wine [58]. The technique was also applied in pollution studies for waste water and ambient aerosol. According to Barker et al., compared to PCA, PLS is the more preferable approach for discrimination [58]. This is because PCA detects the total variation without distinguishing the variance between the groups from the variance within the group. Thus, PCA discrimination is successful if the inter-group variability exceeds the intra-group variability significantly. The inter-group variability is represented in the total variation, and the groups are correctly separated based on the total variation. However, if the inter-group variability and the intra-group variability are not

differentiable, PCA cannot classify the groups correctly. Unlike PCA, PLS analysis focuses on inter-group variability.

A visual comparison of PLS and PCA discrimination was provided in Figure 1.4. As shown in the figure, when the variance square, σ^2, is small, the PCA model was aligned along the axis of maximum total variability. However, the PLS model was aligned along the axis of inter-group variability in the PLS model. Nonetheless, since the intergroup variability contributes significantly to the total variation, the alignments of the PCA and PLS models were fairly similar. However, when σ^2 was increased, the PCA model no longer aligned with the inter-group variability. Thus, the discrimination efficiency was reduced when PCA was perfomed on systems where the intra-group variability and inter-group variability aren't differentiable. Meanwhile, the PLS discrimination efficiency was not affected because discrimination was still based on inter-group variability [58].

Partial least square discriminant analysis (PLS-DA) is a classification method derived from PLS analysis. PLS-DA focuses on the variation between the classes rather than the variation within the class [57]. PLS-DA searches for factors, also called latent variables, to maximize the variations in the sample data from the x-block that can be used to predict the class from the y-block. In a PLS-DA model, a Y-block is generated to assign the samples to the known classes. The Y-block values alternate between 0 and 1. A value of 0 indicates zero possibility of finding the sample in the assigned class. However, a value of 1 meant that the sample and class match was a perfect fit. PLS-DA analysis is performed on both the x- and y-blocks. The resulting calibrated PLS-DA model is an intermediate matrix of the x and y-blocks. In the calibrated model, a new y-predicted value is calculated for each sample. The calculated y-predicted values do not have the ideal values of 1 and 0. However, the y-predicted values are

close to 1 for samples that belong to a given class in the model. Similarly, samples that do not belong in the class have y-predicted values of approximately 0. In addition to determining the y-predicted value of each sample, a threshold value is also calculated for each class prediction model. The threshold is the optimal y-predicted value that will yield the least false negatives and false positives. For a class prediction, if the sample y-predicted value falls between the calculated threshold for a particular class and a value that is close to 1, the sample is assigned to that class [57].

The goal of PLS-DA is to give a calibration model that maximizes the sensitivity and specificity of the training set. The sensitivity is defined as the number of predicted samples in a class divided by the actual number of samples in the class. The specificity is given as the number of samples predicted not in the class divided by the actual number of samples not in the class. A calibration model with high sensitivity and specificity will most likely predict the class effectively. Furthermore, the calibration model can be cross-validated to test the robustness of the model. One of the cross-validation (CV) methods for testing PLS-DA models is the Venetian Blinds method. In this method, according to the set number of data splits, a certain number of samples are withheld from the model. A new model is generated using the remaining samples. The withheld samples become the unknowns for testing the CV model. The new sensitivity and specificity values for the CV model reflect the classification rate of the validated model [57].

Usually, before performing chemometric analysis, the data is preprocessed to reduce variability among samples of the same class. There are three common methods for preprocessing data, mean centering, variance scaling, and autoscaling [59]. Autoscaling is mean centering followed by variance scaling. In mean centering, the centroid of the data points is moved to the origin of the multivariate coordinate system. Statistically, centering the centroid

prevent any data point from having more influence, or leverage, over other data points. In an analytical sense, the origin of the data set is no longer associated with any external chemical or physical properties. Mean centering is performed on a variable-by-variable basis. Considering a system in which the samples are composed of absorbance spectra and wavelengths are represented as variables, beginning with the first wavelength in the data set, the absorbance of all samples at this wavelength is averaged. The calculation is repeated for the remaining wavelengths in the data set. The absorbance at each wavelength is then subtracted by the average absorbance of the samples. The result is an absorbance matrix with positive and negative values, and the new mean of the matrix is zero. Although the centroid is repositioned, the relative distance between the data points are unchanged [59].

Variance scaling adjusts the data set so that the variance of each variable is equal. Statistically, variance scaling allows for all the variables to have the same influence on the data analysis. In analytical chemistry, variance scaling removes any external chemical or physical significance associated with the coordinate axes. Similar to mean centering, variance scaling is also performed on a variable-by-variable basis. The initial step of scaling is to determine the total variance of the concentration for the first component. The variance is calculated so that all the variables are adjusted to unit variance. The concentrations are scaled according to the variance so that the new variance is at a unity. Adjusting the variance to unity is accomplished by first finding the standard deviation, S_k, of the variables. The matrix component is represented as k.

$$S_k = \frac{\sum_{i=1}^{n} \left(a_{ki} - \bar{a}_k \right)}{\left(n - 1 \right)} \tag{5}$$

Then, the data points are scaled by the standard deviation. To variance scale the data without mean centering, the mean center can be added back.

$$a'_{ki} = \frac{a_{ki} - \bar{a}_k}{S_k} + \bar{a}_k \quad (6)$$

Thus, variance scaling alters the relative positions of the data points without changing the centroid location [59].

In addition to mean centering and variance scaling, normalization is also another method employed in data preprocessing. However, normalization is performed on a sample-by-sample basis. Normalization of the data results in all the normalized samples having the same magnitude. Thus, the significance of the distance the samples and the origin is removed. Instead, the emphasis is placed on the direction of the samples from the origin. Preprocessing data by normalization retains pertinent qualitative information for differentiating the samples while removing the remaining information. This allows for sample separation based on variations other than concentration. For example, spectra of the same sample will not be separated based on slight differences in concentration [59].

One approach for normalizing the data is unit vector normalization. Unit vector normalization divides the spectrum by the norm of the spectrum vector [60]. The vector norm is mathematically expressed as followed.

$$a_i^* = \frac{a_i}{\sqrt{\sum a_i^2}} \quad (7)$$

Vector normalization positions all the data points at an equal distance from the origin. However, the direction of the points from the origin corresponds to the the bandshape, and not the relative intensity.

In addition to eliminating variable contribution from differences in sample concentration, removing noise will also improve the classification accuracy. Data smoothing is a preprocessing method that reduces the noise in the data [56]. Thus, the smoothed data will have a higher signal-to-noise (S/N) ratio. Removal of random variation assumes the noise is present in the low-frequency signals. If this assumption fails, smoothing may remove the noise along with chemical information present in the real data. There are currently five methods of smoothing. One of those methods is polynomial smoothing. In polynomial smoothing, the operator sets the window width so that the points within the window are used to determine the value at the center of the window. Thus, adjusting the window width will alter the smoothing results. Polynomial smoothing with the Savitzky-Golay (S-G) algorithm fits a low-order polynomial function to the points in the window. The jth element equals the polynomial prediction at j. The jth element is used to predict the smoothed value of the data points within the middle region of the window. The effectiveness of S-G smoothing requires that the operator find the optimal window width. The window width may be increased to reduce the noise level. However, after passing a certain width, sharp peaks will be removed from the smoothed spectrum and remaining peaks are distorted. The general rule is to select the window according to the spectral bandwidth. If the window width is larger than the bandwidth, the smoothed peaks will either be removed or distorted. Additionally, S-G smoothing also removes terminal points of the sample vector [56].

For data with complex baseline features, it is difficult to define the points for baseline correction. Thus, errors may be introduced during baseline correction, and these errors contribute to the variation in the data. To minimize variability due to errors associated with baseline correction, one can take the derivatives of the raw data when the points for the baseline cannot be defined with certainty. The first derivative, $\boldsymbol{r'}$, of the sample vector can be taken with

respect to the variable number, **x**. For example, if the following polynomial, **r**, is fitted to the sample vector,

$$\boldsymbol{r} = \tilde{\boldsymbol{r}} + \alpha + \beta \boldsymbol{x} + \gamma \boldsymbol{x}^2 + \delta \boldsymbol{x}^3 \quad (8)$$

then ***r'*** is expressed as follows.

$$\boldsymbol{r}' = \tilde{\boldsymbol{r}}' + 0 + \beta + \gamma 2\boldsymbol{x} + 3\delta \boldsymbol{x}^2 \quad (9)$$

$\tilde{\boldsymbol{r}}'$ is the derivative of the signal of interest. The horizontal offset baseline feature denoted as α becomes zero in the first derivative. If the baseline features are described by more complex features such as β,γ, and δ, then these features can be removed by taking successive derivatives [56].

After the data is preprocessed properly, chemometric analysis is performed to classify the samples. Previous studies have shown that chemometric analysis allowed for accurate identification of a variety of samples, including bacterial strains and candidate genes for cancer cells [61]. According to Musumarra et al., partial least square discriminant analysis (PLS-DA) has shown promising results for identification of candidate genes associated with tumor histotypes. Expression of these genes may be the source for development of colon, leukemia, renal, and central nervous system (CNS) tumor cells. For the experiment, tumor cell lines were divided into 8 classes with the same tissue origin. PLS-DA separated the classes by rotating the projection of latent variables (LVs) so that inter-class variability is maximized. The PLS-DA model calculated the variable influence projection (VIP) values and used these values to rank the gene expression. VIP values were determined from the sum of the variable influence (VIN) contribution of all model dimensions. A high VIP value meant the gene was a likely candidate in the development of a particular type of cancer cells. Because the gene transcription level varied for different types of tumor cells, a different order of VIP rankings was produced for each class

of tumor histotypes. Thus, VIP ranking of gene expression by PLS-DA provided a mean for discriminating different tumor cells. The VIP gene ranking was compared with the current knowledge of candidate genes associated with cancer development. From the comparison, it was determined that the PLS-DA discrimination agrees with the current knowledge of candidate genes.

In another example, chemometric analysis was used to correctly discriminate different strains of cyanobacteria [62]. Over the past 30 years, cyanobacteria have been used as food and energy sources. Additionally, cyanobacteria have also served as models for studies of basic cellular processes. Thus, to understand the differences in the bacterial models, these studies relied on the ability of discriminating cyanobacteria strains. A study conducted by Kansiz et al. confirmed that cyanobacterial strains were differentiable by FTIR microspectroscopy coupled with chemometric analysis. FTIR allowed for rapid bacterial detection at low sample volumes. Characteristic bands corresponding to the bacterial components such as proteins, lipids, polysaccharides, and nucleic acids were displayed on the spectra. For strain discrimination, the particular chemometric approaches applied were Principal Component Analysis (PCA), K-nearest neighbor (KNN), and Soft Independent Modeling of Class Analogy (SIMCA). Mean centered PCA was first carried out on underivatized spectra. Loading plots were also created to show the variance contribution of different variables. To classify the strains, 2/3 of the spectral data were used as the random set for SIMCA and KNN analyses; the remaining 1/3 was assigned to the training set. Optimal classification was achieved when multiple variable regions were selected for the KNN and SIMCA analyses. The main difference between SIMCA and KNN is SIMCA produces a distinct PCA model for each class. In contrast, KNN analysis produces one model for the complete data set.

The success of the statistical separation depended on the reproducibility of the spectral variations for different strains. Although the spectra were reproducible, there were slight variations due to changes in sample concentration or thickness and also due to baseline correction. Therefore, variations between replicate spectra were reduced by taking the first and second derivatives of the Savitzky-Golay smoothed spectra. Derivatizing the data decreases variance within a group. This leads to smaller intra-group spectral distances. Consequently, overlapping between neighboring clusters was also reduced for derivatized spectra. As a result, classification efficiency is improved for the first derivative data set. However, the second derivatives decreased the classification rate because higher order derivatives reduce both intra- and inter-group variance. Overall, these studies showed chemometric analysis allowed for extraction of useful information from the data and classification of samples based on detectable reproducible variability in the data.

1.6 Scanning Electron Microscopy (SEM)

For the SERS viral detection studies, the surface morphology of SERS substrates was characterized using scanning electron microscopy (SEM). SEM is an electron microscopic technique widely used in surface studies [63]. An illustration of the SEM system was provided in Figure 1.5. In a typical SEM, the electron beam is produced by the filament source located in the electron gun. One type of electron guns commonly used in SEM is the thermionic triode electron gun. There are three adjustable parameters in the gun, filament current, grid bias, and accelerating voltage. The filament current controls the filament temperature. Thus, increasing the filament current produces a larger number of emitted electrons. To maximize the number of emitted electrons, one can saturate the filament by increasing the current until the number of emitted electrons plateaus. However, over-saturation of the filament generates more evaporation

of the filament material and reduces the lifetime of the filament. The gun bias controls the total electron current produced by the gun and controls the beam area at the filament. The main purpose of the gun bias is to adjust the electron current to generate the desired image brightness. The accelerating voltage controls the number of electrons being emitted. Decreasing the accelerating voltage reduces the number of emitted electrons, and results in signal loss. Generally, the accelerating voltage ranges from 1 – 30 keV; the optimal voltage is around 20 – 30 keV [63].

In an SEM system, the electron path begins at the filament. The filament is heated to release electrons. The electrons are accelerated through the anode toward the condenser lens. The condenser lens collimates the light to create a fine electron beam. The condenser lens controls the spot size or the beam diameter at the sample. Alternatively, the beam diameter can be adjusted manually by altering the height of the sample platform to optimize the working distance. The best resolution is achieved at the optimal working distance. In general, reducing the working distance decreases the spot size and enhances the resolution. However, a small spot size is produced by decreasing the beam intensity, and less electrons are emitted at lower beam intensities. Consequently, the low signal may cause the image to appear noisy. As an option, a small aperture can be inserted to reduce the current at the sample. With a smaller aperture, more electrons are stopped at the aperture and the convergence angle is decreased. The resulting beam is more parallel, creating a larger depth of field. Increasing the depth of field improves the resolution and produces a sharper image. However, a disadvantage of having a small aperture is that more diffraction occurs at the aperture. Diffraction expands the spot size and decreases the resolution. Therefore, the aperture choice is usually a compromise between the depth of field enhancement and diffraction reduction [63].

After leaving the filament source, the electron beam is passed through microscope lens in the system. The microscope lens establishes an electromagnetic field around the optical axis of the microscope and acts as converging lens. Thus, the lens maintains the electron path along the optical axis. Additionally, scanning coils positioned between the condenser and objective also deflect the electron beam so that the beam travels along the optical axis. When it reaches the objective lens, the beam is focused by the lens onto the sample. Electrons emitted from the sample surface strike the detector. According to the types of electrons emitted, different detectors can be used to measure the signal [63].

In SEM, two types of emitted electrons are detected [64]. Electrons escaping the surface with energy greater than 50 eV are called backscattered electrons. Emitted electrons with energy less than 50 eV are defined as secondary electrons. Secondary electrons are produced by inelastic scattering. Secondary electrons are released by electron-electron interaction between primary electrons and the sample atoms. Inelastic scattering of secondary electrons, σ_{in}, is proportional to $Z^{1/3}$. Z is the atomic number of the sample. Z is raised to the third power because only a small population of secondary electrons is emitted. Emitted secondary electrons have low energy, approximately 2 – 5 eV.

Backscattered electrons are reemitted through multiple scattering of primary electrons emerging from the sample surface. Usually, backscattered electrons are reemitted with a much higher energy than that of secondary electrons. The average energy of backscattered electrons is quite close to the energy of the primary electrons. Backscattered electrons are most likely produced from electron-nuclear interaction between incident electrons and the atomic nucleus of the sample. This type of scattering is also called Rutherford scattering and is proportion to Z^2 [64].

The most common detectors are secondary electron, backscattered electron, and x-ray detectors [63]. Usually, a secondary electron detector is positioned at an angle from the sample platform. The detector is biased positive to attract secondary electron. Electrons travel in curved paths to reach the detector. Because secondary electrons have small energy, the signal detected for secondary electrons is lower than that for backscattered electrons. To measure the signal from backscattered electrons, the detector is biased negative to repel secondary electrons. However, backscattered electrons detector can be positioned directly above the sample to obtain a large collection angle. Higher signal intensity from high energy backscattered electrons produces a stronger contrast in the image. The final image contrast is dominated by either the difference in atomic number or the difference in topography. If the atomic number difference is much larger than the topographical difference, the image contrast reflects the atomic number difference. In contrast, if the atomic number is fairly uniform across the sample surface, then the image contrast is dominated by the topographical differences [64]. For example, if the surface is consists of a uniform Ag film, then the image contrast reflects the different topographical features on the film surface.

The signal measured by the detector is amplified by a photomultiplier. The final image is displayed on the cathode ray tube (CRT) screen. Usually, a 2-D image is displayed if the rectangular raster scan pattern is selected by the scan generator. In other cases, the scan generator can be programmed to produce 3 other types of scan patterns. Single or repeated line scan pattern allows for 1-D tracking of signal variation. Stationary spot scan pattern permits data collection from the smallest region on the surface. Rocked spot pattern outlines the electron channeling patterns and produces crystallographic images [63].

When operating the SEM, the image is aligned at a slow scan rate and small raster size to correct for astigmatism. Astigmatism occurs when the lenses are asymmetrical, and the focus across one diameter does not equal the focus across another diameter. Consequently, the image is enlarged at the best focus point, and elongated on each side of the focus. Additionally, stigmators are also used for adjusting the image. XY stigmators act as the fine focus and controls the sharpness of the image. Amplitude and rotation stigmators remove directional defects in the image. Amplitude and rotation stigmators are adjusted to increase the amplitude to observe the defect. The stigmators then rotate the defect so that the defect is at a right angle to the original direction of the defect. Finally, the stigmators reduce the amplitude to remove the defect.

To image micro-scale and nano-scale features on the sample surface, the operator can adjust the magnification control to produce the necessary image magnification [64]. The magnification control monitors the current flow through the scanning coils. The magnitude of image magnification can be expressed as the relative ratio of the current flow through the detection coil in the column and the current flow in the CRT. Additionally, the magnification is controlled by adjusting the scanned area according to the dimension of the display screen. Thus, the magnification is also expressed as the ratio of scan length in the CRT display and the scan length across the sample surface. For example, for a display screen with 10 x 10 cm dimension, 20x-magnification is produced for a scanned area of 5 x 5 mm. If the scanned area is reduced further to 10 x 10 μm, for example, the magnification increases to 10,000x.

1.7 Summary of Presented Work:

Chapter II describes the detection of different genotypes of measles virus by surface-enhanced Raman scattering (SERS) Spectroscopy. The SERS spectra collected for A, D4, D9,

and H1 measles genotypes and the cell culture media were classified by Principal Component Analysis (PCA), Hierarchical Cluster Analysis (HCA), and Partial Least Square Discriminant Analysis (PLS-DA).

Chapter III describes the detection of different genotypes of measles virus by Fourier transform infrared (FTIR) spectroscopy. Chemometric methods of Principal Component Analysis (PCA), Hierarchical Cluster Analysis (HCA), and Partial Least Square Discriminant Analysis (PLS-DA) were applied to discriminate samples of A, D4, D9, and H1 measles genotypes.

1.8 References

1. Centers for Disease Control and Prevention. Epidemiology and Prevention of Vaccine Preventable Diseases. Atkinson W, Hamborsky J, McIntyre L, Wolfe S, eds. 10th ed. Washington DC: Public Health Foundation, 2007.
2. Schneider-Schaulies, S.; Meulen, V. t.; Allan, G.; Robert, G. W., "Measles Virus (Paramyxoviridae)", in *Encyclopedia of Virology*, Elsevier: Oxford, 1999; pp 952-960.
3. Rota, P. A.; Rota, J. S.; Bellini, W. J., Molecular epidemiology of measles virus. *Seminars in Virology* **1995,** 6, 379-386.
4. McChesney, M. B.; Miller, C. J.; Rota, P. A.; Zhu, Y.-d.; Antipa, L.; Lerche, N. W.; Ahmed, R.; Bellini, W. J., Experimental Measles. I. Pathogenesis in the Normal and the Immunized Host. *Virology* **1997,** 233, 74-84.
5. Kobune, F.; Funatu, M.; Takahashi, H.; Fukushima, M.; Kawamoto, A.; Iizuka, S.; Sakata, H.; Yamazaki, S.; Arita, M.; Wenbo, X.; Zhang, L. B., Characterization of measles viruses isolated after measles vaccination. *Vaccine* **1995,** 13, 370-372.

6. Chibo, D.; Riddell, M.; Catton, M.; Lyon, M.; Lum, G.; Birch, C., Studies of measles viruses circulating in Australia between 1999 and 2001 reveals a new genotype. *Virus Research* **2003,** 91, 213-221.
7. Mizuta, K.; Abiko, C.; Murata, T.; Yamada, K.; Ahiko, T.; Sakamoto, M.; Tsuchida, S.; Matsuzaki, Y.; Hongol, S.; Sunagawa, T.; Kudo, K., An outbreak of measles virus infection due to a genotype D9 at a junior high school in Yamagata, Japan in 2004. *Japanese Journal of Infectious Diseases* **2005,** 58, 98-100.
8. Rota, P. A.; Bellini, W. J., Update on the global distribution of genotypes of wild type measles viruses. *Journal of Infectious Diseases* **2003,** 187, S270-S276.
9. Rima, B. K.; Earle, J. A. P.; Yeo, R. P.; Herlihy, L.; Baczko, K.; Termeulen, V.; Carabana, J.; Caballero, M.; Celma, M. L.; Fernandezmunoz, R., Temporal and geographical distribution of measles-virus genotypes. *Journal of General Virology* **1995,** 76, 1173-1180.
10. Rima, B. K.; Earle, J. A. P.; Baczko, K.; terMeulen, V.; Liebert, U. G.; Carstens, C.; Carabana, J.; Caballero, M.; Celma, M. L.; FernandezMunoz, R., Sequence divergence of measles virus haemagglutinin during natural evolution and adaptation to cell culture. *Journal of General Virology* **1997,** 78, 97-106.
11. World Health Organization, Expanded Programme on Immunization. *Weekly Epidemiological Record* **1998**, 73, 265-272.
12. World Health Organization, Nomenclature for describing the genetic characteristics of wild-type measles viruses (update). *Weekly Epidemiological Record* **2001**, 76, 242-248.
13. World Health Organization, Update of the nomenclature for describing the genetic characteristics of wild-type measles viruses: new genotypes and reference strains. *Weekly Epidemiological Record* **2003**, 78, 229-232.

14. World Health Organization, New genotype of measles virus and update on global distribution of measles genotypes. *Weekly Epidemiological Record* **2005**, 80, 347-351.

15. World Health Organization, Progress in global measles control and mortality reduction, 2000 – 2006. *Weekly Epidemiological Record* **2007**, 82, 418-424.

16. Rota, J. S.; Wang, Z. D.; Rota, P. A.; Bellini, W. J., Comparison of sequences of the H, F, and N coding genes of measles-virus vaccine strains. *Virus Research* **1994,** 31, 317-330.

17. Rota, P. A.; Bloom, A. E.; Vanchiere, J. A.; Bellini, W. J., Evolution of the nucleoprotein and matrix genes of wild-type strains of measles-virus isolated from recent epidemics. *Virology* **1994,** 198, 724-730.

18. Xu, W. B.; Tamin, A.; Rota, J. S.; Zhang, L. B.; Bellini, W. J.; Rota, P. A., New genetic group of measles virus isolated in the People's Republic of China. *Virus Research* **1998,** 54, 147-156.

19. Krause, C. H.; Molyneaux, P. J.; Ho-Yen, D. O.; McIntyre, P.; Carman, W. F.; Templeton, K. E., Comparison of mumps-IgM ELISAs in acute infection. *Journal of Clinical Virology* **2007,** 38, 153-156.

20. Counihan, N. A.; Daniel, L. M.; Chojnacki, J.; Anderson, D. A., Infrared fluorescent immunofocus assay (IR-FIFA) for the quantitation of non-cytopathic and minimally cytopathic viruses. *Journal of Virological Methods* **2006,** 133, 62-69.

21. Nakayama, T.; Mori, T.; Yamaguchi, S.; Sonoda, S.; Asamura, S.; Yamashita, R.; Takeuchi, Y.; Urano, T., Detection of measles-virus genome directly from clinical-samples by reverse-transcriptase polymerase chain reaction and genetic variability. *Virus Research* **1995,** 35, 1-16.

22. Riddell, M. A.; Moss, W. J.; Hauer, D.; Monze, M.; Griffin, D. E., Slow clearance of measles virus RNA after acute infection. *Journal of Clinical Virology* **2007,** 39, 312-317.

23. G. Keresztury, "Raman Spectroscopy: Theory", in *Handbook of Vibrational Spectroscopy*, J.M. Chalmers and P.R. Griffiths, eds. John Wiley and Sons, Chichester, pp. 69-72.

24. T. Turrell and J. Corset, "Molecular vibrations", in *Raman Microscopy Developments and Applications*, Academic Press Inc., San Diego, 1996, pp. 8-10.

25. M.J. Pelletier, "Raman Monochromators and Polychromators", in *Handbook of Vibrational Spectroscopy*, J.M. Chalmers and P.R. Griffiths, eds. John Wiley and Sons, Chichester, 2002, pp. 467-481.

26. R.L. McCreery, "Dispersive Raman Spectrometer", in *Raman Spectroscopy for Chemical Analysis*, John Wiley and Sons, New York, NY, 2000, Chapter 8, pp. 155-171.

27. J.D. Ingle and S.R. Crouch, *Spectrochemical Analysis*, Prentice Hall, Englewood Cliffs, NJ, 1988, Chapter 4, pp. 106-117.

28. CJ. Pommier, L.K. Walton, T.D. Ridder and M.B. Denton, "Array Detectors for Raman Spectroscopy", in *Handbook of Vibrational Spectroscopy*, J.M. Chalmers and P.R. Griffiths, eds. John Wiley and Sons, Chichester, pp. 507-521.

29. K.J. Baldwin, D.N. Batchelder and S. Webster, "Raman Microscopy: Confocal and Scanning Near Field", in *Handbook of Raman Spectroscopy-From the Research Laboratory to the Process Line*, I.R. Lewis and H.G.M. Edwards, eds., Marcel Dekker, Inc., New York, 2001, pp. 145-190.

30. P. Dhamelincourt, "Raman Microscopy", in *Handbook of Vibrational Spectroscopy*, J.M. Chalmers and P.R. Griffiths, eds. John Wiley and Sons, Chichester, 2002, pp. 1419-1428.

31. F. Adar, "Evolution and Revolution of Raman Instrumentation – Application of Available Technologies to Spectroscopy and Microscopy", in *Handbook of Raman Spectroscopy-From the Research Laboratory to the Process Line*, I.R. Lewis and H.G.M. Edwards, eds., Marcel Dekker, Inc., New York, 2001, pp. 11-40.
32. P.J. Treado and M.P. Nelson, "Raman Imaging", in *Handbook of Vibrational Spectroscopy*, J.M. Chalmers and P.R. Griffiths, eds. John Wiley and Sons, Chichester, 2002, pp. 1429-1459.
33. Vo-Dinh, T., Surface-enhanced Raman spectroscopy using metallic nanostructures. *Trac-Trends in Analytical Chemistry* **1998,** 17, 557-582.
34. Willets, K. A.; Van Duyne, R. P., Localized surface plasmon resonance spectroscopy and sensing. *Annu. Rev. Phys. Chem.* **2007,** 58, 267-297.
35. Aroca, R., *Surface-Enhanced Vibrational Spectroscopy: The Interaction of Light With Nanoscopic Metal Particles and Molecules on Smooth Reflecting Surfaces.* John Wiley and Sons: Chichester, 2006; p. 35-72.
36. Schatz, G.C., Van Duyne, R.P., *Handbook of Vibrational Spectroscopy: Electromagnetic Mechanism of Surface-Enhanced Spectroscopy.* John Wiley and Sons: Chichester, 2002; p. 759-774.
37. Moskovits, M., Surface enhanced spectroscopy, *Rev. Mod. Phys*. **1985**, 57, 783-826.
38. Smith, W.E.; Rodger, C., *Handbook of Vibrational Spectroscopy*: *Surface-Enhanced Raman Scattering.* John Wiley and Sons: Chichester, 2002; p. 775-784.
39. Aroca, R., *Surface-Enhanced Vibrational Spectroscopy: Chemical Effects and the SERS Spectrum.* John Wiley and Sons: Chichester, 2006; p. 107-029.

40. Kneipp, K.; Kneipp, H., Single molecule Raman scattering. *Applied Spectroscopy* **2006,** 60, 322A-334A.

41. Kneipp, K.; Kneipp, H.; Itzkan, I.; Dasari, R. R.; Feld, M. S., Surface-enhanced Raman scattering and biophysics. *Journal of Physics-Condensed Matter* **2002,** 14, R597-R624.

42. Driskell, J. D.; Kwarta, K. M.; Lipert, R. J.; Porter, M. D.; Neill, J. D.; Ridpath, J. F., Low-level detection of viral pathogens by a surface-enhanced Raman scattering based immunoassay. *Analytical Chemistry* **2005,** 77, 6147-6154.

43. Shanmukh, S.; Jones, L.; Driskell, J.; Zhao, Y. P.; Dluhy, R.; Tripp, R. A., Rapid and sensitive detection of respiratory virus molecular signatures using a silver nanorod array SERS substrate. *Nano Letters* **2006,** 6, 2630-2636.

44. Abelmann, L.; Lodder, C., Oblique evaporation and surface diffusion. *Thin Solid Films* **1997,** 305, 1-21.

45. E. Brame and J. Grasselli, "An introduction to molecular vibrations", in Infrared and Raman Spectroscopy, Marcel Dekker, Inc., 1976, pp. 2-12.

46. R. Bhargava and I.W. Levin, "Fourier Transform Mid-Infrared Spectroscopic Imaging", in *Spectrochemical Analysis Using Infrared Multichannel Detectors*, R. Bhargava and I.W. Levin, eds., Blackwell Publishing, Oxford, 2005, pp. 1-24.

47. E. Hecht, *Optics*, Addison Wesley, San Francisco, 2002, Chapter 11.

48. P.R. Griffiths and J.A. de Haseth, *Fourier Transform Infrared Spectrometry*, John Wiley & Sons, New York, 1986, Chapter 1.

49. P.R. Griffiths and J.A. de Haseth, *Fourier Transform Infrared Spectrometry*, John Wiley & Sons, New York, 1986, Chapter 3.

50. P.R. Griffiths and J.A. de Haseth, *Fourier Transform Infrared Spectrometry*, John Wiley & Sons, New York, 1986, Chapter 2.

51. A.P. Thorne, *Spectrophysics*, Chapman and Hall, London, 1988, Chapter 7.

52. Tomzig, A.; Spassov, S.; Friedrich, M.; Naumann, D.; Beekes, M., Discriminating scrapie and bovine spongiform encephalopathy isolates by infrared spectroscopy of pathological prion protein. *Journal of Biological Chemistry* **2004,** 279, 33847-33854.

53. Lasch, P.; Beekes, M.; Schmitt, J.; Naumann, D., Detection of preclinical scrapie from serum by infrared spectroscopy and chemometrics. *Analytical and Bioanalytical Chemistry* **2007,** 387, 1791-1800.

54. Helm, D.; Labischinski, H.; Naumann, D., Elaboration of a procedure for identification of bacteria using Fourier transform IR spectral libraries – a stepwise correlation approach. *Journal of Microbiological Methods* **1991,** 14, 127-142.

55. Mariey, L.; Signolle, J. P.; Amiel, C.; Travert, J., Discrimination, classification, identification of microorganisms using FTIR spectroscopy and chemometrics. *Vibrational Spectroscopy* **2001,** 26, 151-159.

56. Beebe, K.R., Pell, R.J., Seasholtz, M.B. *Chemometrics: A Practical Guide*; John Wiley and Sons: New York, 1998.

57. Wise, B.M., Gallagher, *N.B. PLS_Toolbox 4.1 Manual*, Eigenvector Research, Inc., Wenatchee, 2007.

58. Barker, M.; Rayens, W., Partial least squares for discrimination. *Journal of Chemometrics* **2003,** 17, 166-173.

59. Kramer, R., *Chemometric Techniques for Quantitative Analysis: Factor Spaces.* Marcel Dekker: New York, 1998; p. 79-97.

60. Chalmers, J.M., Griffiths, P. R. *Handbook of Vibrational Spectroscopy: Sample Characterization and Spectral Data Processing*; John Wiley and Sons: Chichester, 2002.

61. Musumarra, G.; Barresi, V.; Condorelli, D. F.; Fortuna, C. G.; Scire, S., Potentialities of multivariate approaches in genome-based cancer research: identification of candidate genes for new diagnostics by PLS discriminant analysis. *Journal of Chemometrics* **2004,** 18, 125-132.

62. Kansiz, M.; Heraud, P.; Wood, B.; Burden, F.; Beardall, J.; McNaughton, D., Fourier Transform Infrared microspectroscopy and chemometrics as a tool for the discrimination of cyanobacterial strains. *Phytochemistry* **1999,** 52, 407-417.

63. D. Chescoe and P. Goodhew, "How and why electron microscopes work", in *Operation of Transmission and Scanning Electron Microscopes*, Oxford University Press, Oxford, 1990, pp. 1-24.

64. M. Hayat, "Electron beam and specimen interaction", in *Introduction to biological scanning electron microscopy*, University Park Press, Maryland, 1978, pp. 15-20.

65. S. Chapman, "Imaging" in *Working with a scanning electron microscope*, Lodgemark Press, Chislehurst, 1986, pp. 56-66.

Figure 1.1 Principal Component Analysis (PCA) of variables 1, 2, and 3. Principal component describes the direction of the total maximum variation in the data [57].

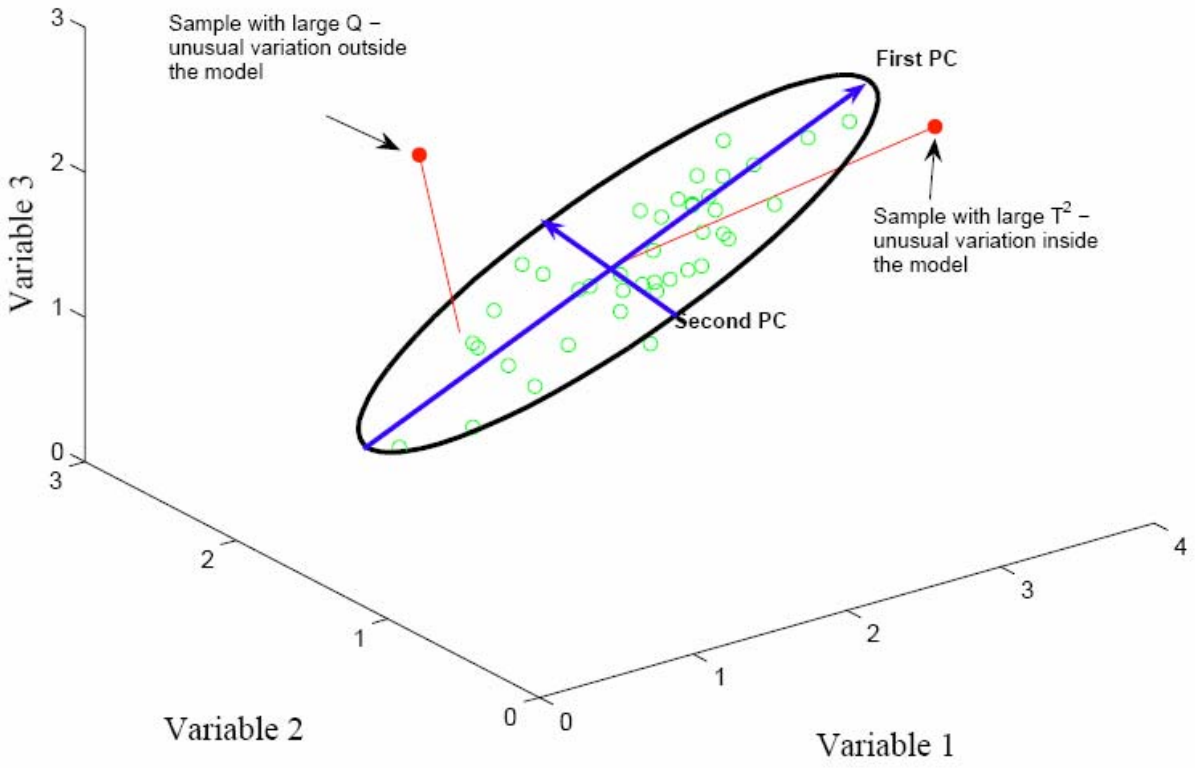
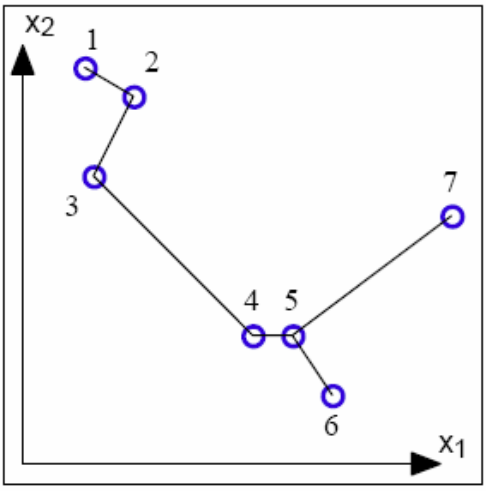
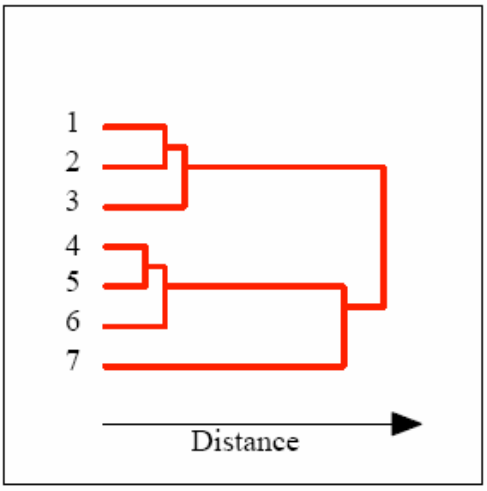

Sample with large Q –
unusual variation outside
the model
First PC
Sample with large T² –
unusual variation inside
the model
Second PC
Variable 3
Variable 2
Variable 1
0
1
2
3
4

Figure 1.2 Hierarchical Cluster Analysis (HCA) dendogram. HCA links the sample with its nearest neighbor(s) to form the clusters on the dendogram [57].



Samples connected to nearest neighbors



Resulting dendrogram

Figure 1.3 K-Nearest Neighbor (KNN) classification of unknown sample. KNN assigns the unknown sample to the class with the most nearest neighbors [57]. In this case, because the unknown is positioned near 2 A samples and 1 B sample, the unknown is identified as belonging to class A.

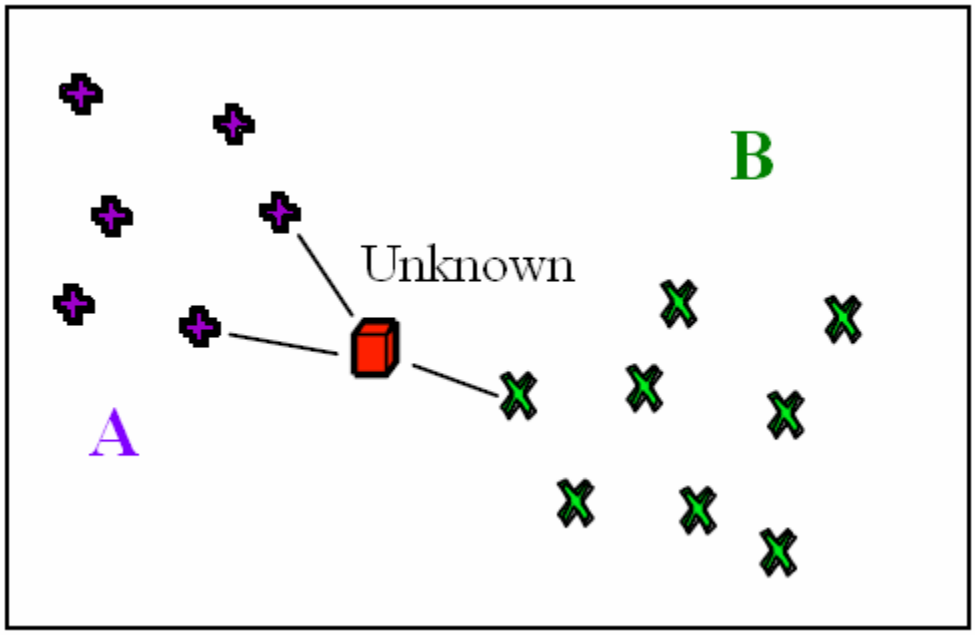
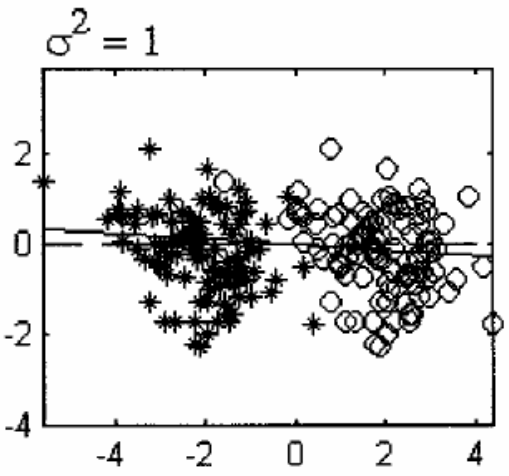
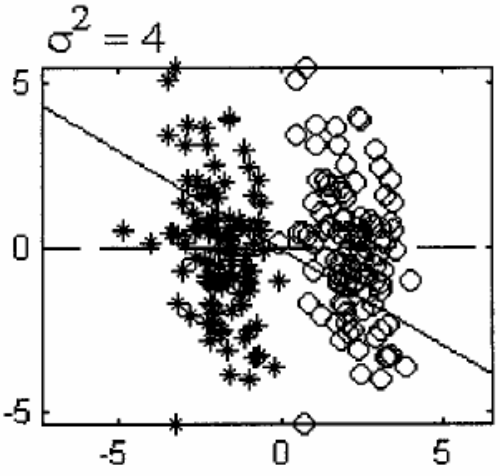
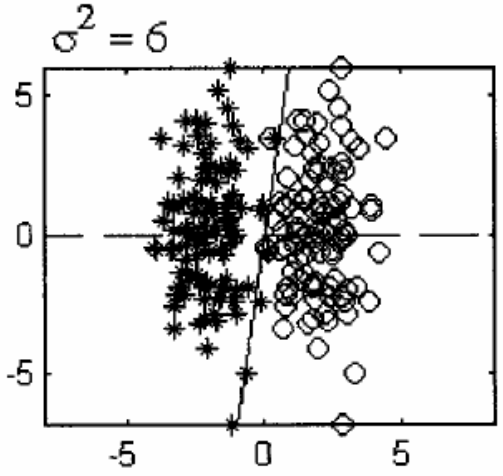

B
Unknown
A

Figure 1.4 Comparison of classification efficiency of Partial Least Square (PLS) and PCA models. The model comparison was performed on data systems with different values of squared variance, σ^2 [58]. For systems with large σ^2 values, the intergroup variability and intragroup variability are not easily distinguishable. For these systems, PCA models cannot discriminate the samples correctly because PCA cannot detect the intergroup variability within the total maximum variation. However, because PLS-DA detects the intergroup variability in the data, PLS-DA models effectively discriminate samples for systems with large σ^2.


$\sigma^2 = 1$
2
0
-2
-4
-4
-2
0
2
4


$\sigma^2 = 4$
5
0
-5
-5
0
5


$\sigma^2 = 6$
5
0
-5
-5
0
5

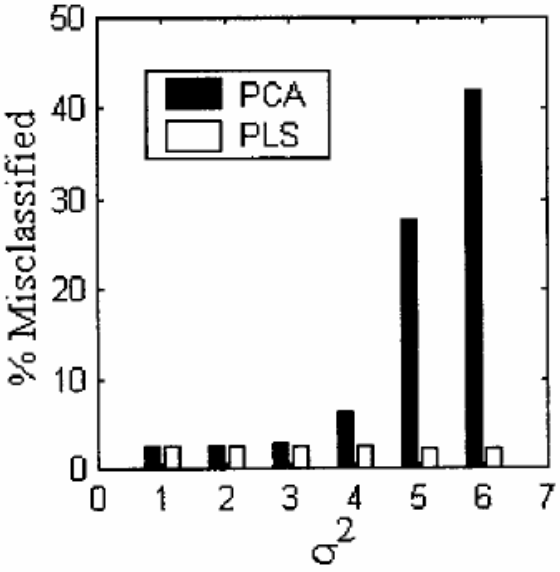
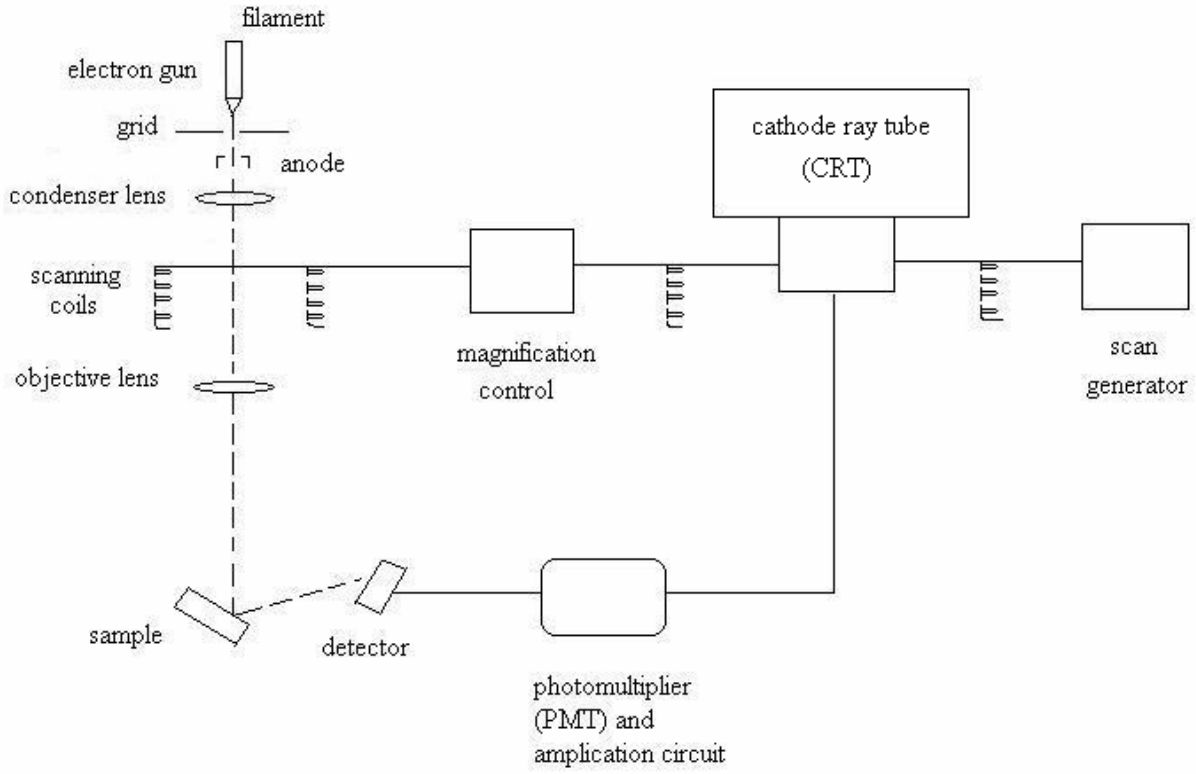

PCA
PLS
% Misclassified
50
40
30
20
10
0
0
1
2
3
4
5
6
7
σ^2

Figure 1.5 Illustration of the scanning electron microscope (SEM) system [63]. The major components in the SEM were labeled in the diagram.


filament
electron gun
grid
anode
condenser lens
cathode ray tube
(CRT)
scanning
coils
magnification
control
scan
generator
objective lens
sample
detector
photomultiplier
(PMT) and
amplication circuit

Chapter II

Surfaced-Enhanced Raman Scattering (SERS) Spectroscopy:

A Potential Diagnostic Tool for Measles Virus Identification[1]

[1] Hoang, V., R.A. Dluhy, Y. Zhao, and R. Tripp. To be submitted to *Analytical and Biological Chemistry*.

Abstract

In the past, measles has been the cause of worldwide epidemics. The disease develops as an acute infection with common symptoms of rash, high fever, and coughing. In some cases, other complications such as pneumonia and encephalitis may develop from the infection. In those cases, the disease could be fatal, especially in young children. Generally, during a measles infection, the host immune system will attempt to repress viral replication and clear the virus from the body. If the attempt is successful, the infection goes away and the host is immune for life.

Currently, IgM and IgG enzyme-linked immunoassays (ELISAs) and immunofluorescent antibody assays are the popular techniques for measles diagnosis. However, these assays are time-consuming and do not allow for direct viral detection. In contrast, because of the signal enhancement in surface-enhanced Raman scattering (SERS) spectroscopy, rapid measles detection was achieved in this SERS study. The SERS spectra were analyzed by chemometric methods of PCA, HCA, and PLS-DA to classify the samples of different measles genotypes.

Introduction

Measles virus thrives successfully in a large susceptible population. Therefore, before the introduction of measles vaccines, the disease was commonly found in young children [1]. However, since measles vaccination was developed in 1963, the number of measles cases has been significantly reduced in the United States [2]. Because of the large vaccination coverage, measles is no longer endemic in the US since 2002. However, in other parts of the world, measles outbreaks occur frequently. A recent report by the CDC stated that measles is the cause of 242,000 deaths worldwide in 2006 [3]. Because of limited access to measles vaccines, measles is still endemic in countries in Africa and Asia [2, 3].

Additionally, even though vaccination interrupted transmission of endemic measles strains, new outbreaks can start from virus importation [4]. For example, recent measles outbreaks in Indiana, Boston, and San Diego were reported between 2005 and 2008. In each outbreak, the cases were traced to a single index patient [5]. The report indicated the patient returned with a measles infection after traveling outside of the country. The index patient then spread the infection to other unvaccinated hosts who came in close contact with the patient. Because measles remains a global disease in many underdeveloped countries and sporadic outbreaks occur from virus importation, there is a continuous need for rapid measles detection to aid in diagnosis of suspected cases and to improve the global measles surveillance.

Currently, the common diagnostic methods available for measles detection are enzyme-linked immunoassay (ELISA) and immunofluorescent antibody assays tests [1, 2]. In particular, the CDC recommends the use of IgM and IgG ELISA tests for diagnosis of measles [2]. Ideally, a negative IgM test result means the individual is not infected. A positive IgM test result is needed to confirm the diagnosis. However, a disadvantage of the IgM test is the detection

efficiency decreases for tests conducted within 72 hours of exposure. In those cases, the IgM tests might produce up to 20 % false negatives [2]. Additionally, a recent study by the Krause's group showed the sensitivity of commercially available ELISA tests for mumps and measles was low [6]. The sensitivity level of these tests was around 24 – 51 %. The low sensitivity is due to the delayed or absence of IgM response during the early phase of infection. Because of the low sensitivity, negative test results could be inconclusive for diagnosis of suspected measles cases. Alternatively, IgG tests can also be used for measles detection. However, IgG tests require a long assay time. The assay time is extended because 2 specimens are required for IgG tests. The first specimen may be collected immediately after the rash develops. However, the second specimen must be acquired 10 – 30 days after the rash onset. After the specimen collection, the test must be conducted on both specimens at the same time. Therefore, IgG test results cannot be obtained until the second specimen is available [2].

In addition to ELISA tests, other current diagnostic methods include immunofluorescent antibody assays. However, these assays have poor detection limit [7]. Thus, sample preparation and reverse transcriptase-polymerase chain reaction (RT-PCR) are often required for signal amplification [8]. Because of the amplification requirements, the total assay time usually take 6 – 8 days to complete [9]. Overall, these limitations reduce the detection efficiency of current diagnostic assays. These tests are time-consuming and do not allow for direct viral detection. In contrast, because of its high sensitivity, surface-enhanced Raman scattering (SERS) spectroscopy provides an effective method for rapid detection of virus at low concentration and with minimal sample preparation [10]. Previous studies have shown that chemometric analysis allows for sample discrimination at a high classification rate [11, 12]. Thus, for this experiment, the

sensitivity of SERS was combined with the specificity from chemometric analysis to identify different genotypes of measles in cell culture media.

Methods

Measles Virus. The CDC supplied 4 genotypes of measles virus (MV). The MV genotypes were MV-A, MV-D4, MV-D9, and MV-H1. Additionally, the CDC also provided two different types of cell culture media to be used as controls. The main difference between the two controls was one control contained the plasmids coding for the SLAM receptor. For simplicity, the control containing the SLAM-coding plasmids was called ctrl 3 (SLAM). The control without the plasmids was named ctrl 8. All the samples were suspended in cell culture media composed of Dulbecco's Modified Eagle's Medium (DMEM) and 10 % fetal bovine serum (FBS). The MV concentration in the sample vials was given as 10^6 plaques forming unit (pfu)/mL.

Substrate preparation. For the SERS viral detection, SERS Ag nanorod (AgNR) substrates were prepared by oblique angle deposition (OAD). In this technique, a stepper motor was used to rotate the sample holder so that the angle between the incoming vapor from the metal source and the substrate normal was 86˚. The nanorods grew on the substrates according to the shadowing effect and surface diffusion. Clean 1 cm x 1 cm glass squares were used as substrates for growing the nanorods. A layer of Ti served as the adhesive layer between the glass substrate and the deposited Ag film. The nanorods were grown on the Ag film at 86˚ from the substrate normal. During the deposition, the length of the AgNRs was monitored by a quartz crystal microbalance (QCM). The morphology and spacing of the nanorods were determined using the LEO 982 Field Emission Scanning Electron Microscope (FE-SEM) system (Way

Peabody, MA). The substrates were stored in a nitrogen-purged glove box to prevent absorption of contaminants in the air.

SERS measurements. Dilution of the samples was carried out in HPLC-grade/ 99.9% assay methanol (MeOH) to determine the optimal concentration for viral detection. 10x-dilution was determined to be the optimal concentration for the diluted sample. 1µL of diluted sample was applied directly on the surface of the AgNR substrates. The substrates exposed to MV samples were air-dried at room temperature. Then, SERS spectra of the exposed substrates were collected using the Renishaw inVia Raman Microscope (Gloucestershire, UK). The SERS signal of the measles samples was measured at five different spots on each substrate. The spectral collection time was 30 seconds.

Data analysis. Using Unscrambler version 9.6 (CAMO Technologies Inc.), the SERS spectral data was preprocessed by taking the Savitzky-Golay first derivative and applying unit vector normalization. To classify the MV samples, the data was further preprocessed by autoscaling and then analyzed by Principal Components Analysis (PCA), using the Partial Least Squares (PLS) Toolbox version 4.1 (Eigenvector Research Inc., Wenatchee, WA) in the MATLAB software version 7.2 (The Mathworks Inc., Natick, MA). Next, through the PLS toolbox, Hierarchical Cluster Analysis (HCA) was carried out to generate dendograms of the classified samples. The clusters in the dendograms were determined based on K-Nearest Neighbor (KNN) grouping and the Mahalanobis distances. Finally, Partial Least Squares Discriminant Analysis (PLS-DA), also provided in the PLS toolbox, was used to classify the samples of the four MV genotypes.

Results

SERS spectra. A comparison between the spectra of measles samples with the background of the substrates suggested that Raman shifts appearing around 1578, 1016, and 920 cm^{-1} on the sample spectra experienced some contribution from the background (Fig. 2.1 – 2.7). However the remaining prominent peaks in the spectra were characteristic of the measles samples. Using previous works on Raman measurements of proteins, peptides, and nucleic acid bases as references, the prominent bands were assigned to components of the virus and the cell culture media [13, 14, 15, 16]. For example, the Raman shifts detected at 1457, 1067, 855, and 501 cm^{-1} were assigned to CH_3 asymmetric deformation, CN stretch and CC stretch and of amino acids, and SS stretch of cysteine, respectively (Table 2.1) [13, 14, 15].

Additionally, there were observable variations in the spectral data of the virus samples, the media controls, and the MeOH solvent (Fig. 2.8). These spectral variations included addition(s) or deletion(s) of peaks and changes in bandshape. For example, the amide I protein bands around 1698 and 1651 cm^{-1} only appeared in the sample spectra, and not in the solvent spectrum (Table 2.2). Additionally, the presence of amino acids and nucleic acid bases in the samples may have contributed to the changes in bandshape, observed in the sample spectra. Comparing the solvent spectrum with the sample spectra, one can see that there was a change in band shape for the overlapping peaks around 1017 and 517 cm^{-1} (Table 2.2). According to previous studies, leucine, alanine, cysteine, and adenine bands usually appeared in these spectral regions. Thus, the presence of these amino acids and nucleic acid base must have contributed to the spectral changes detected at these Raman shifts [13, 14, 15, 16]. There were also notable differences between the spectra of the virus samples and the control media. For example, the peaks appearing around 960 and 1288 cm^{-1} were only seen in the media spectra. These shifts

were probably produced by amino acids and proteins present in the media (Table 2.3) [13, 14]. Finally, comparing the spectra of samples for different measles genotypes, observable spectral variations of H1 and A spectra allowed for separation of H1 and A from D4 and D9 (Table 2.4). However, D4 and D9 spectra were indifferentiable by visual inspection because the two spectra were very similar (Fig. 2.8).

Chemometric analysis. By visual inspection, the spectra of different genotypes and media can be separated according to some of the observable spectral variations. However, these variations did not give a well-defined separation of the samples for different classes. Therefore, chemometric analysis was applied to provide a more standardized and consistent method for separation of samples base on all detectable statistical variations in the data. To interpret the data by chemometrics, the first approach applied was Principal Component Analysis (PCA). In PCA analysis, PC1 versus PC2 scores plot showed that 5 out of 7 sample classes were separable (Fig. 2.9). With respect to PC1 and PC2, PCA analysis of 56 % of the total variation in the data allowed for separation of samples from the A, H1, ctrl 3, ctrl 8, and MeOH classes. However, because of the low variability between D4 and D9 classes, there was an overlap of samples from these 2 classes.

Hierarchical Cluster Analysis (HCA) was the second chemometric method for data interpretation. For the data analysis, HCA was combined with K-Nearest Neighbor classification and Mahalanobis distance calculation to generate a dendogram, showing separate clusters for different classes. HCA analysis of the SERS spectra gave similar results to those produced from PCA analysis (Fig. 2.10). With the exception of D4 and D9 samples, all remaining samples were classified correctly. Additionally, HCA dendogram indicated that samples from MeOH, ctrl 3 and ctrl 8 classes can be separated from the virus samples. According to the dendogram, the

solvent and media samples were determined to be most statistically different from the virus samples.

Similar to the PC scores plots, PLS-DA LV scores plots also displayed well-defined clusters for samples from the A, H1, ctrl 3, ctrl 8, and MeOH classes (Fig. 2.11). However, PLS-DA class prediction models showed that according to the calculated y-predicted values of the samples and the threshold (Table 2.5) defined in the class prediction model, all the samples can be assigned to the correct classes (Fig. 2.12 – 2.18). In particular, PLS-DA model successfully separated and identified samples from D4 and D9. The PLS-DA model was tested by Venetian blinds cross-validation (CV), using 11 data splits. Although the CV model presented 1 or 2 outliers (Fig. 2.19 – 2.25), the model was able to predict the class with 100% sensitivity and 96.4 – 100 % specificity (Table 2.6).

Discussions

SERS spectroscopy has been an important analytical tool in the structural analysis of biological samples. Details on the molecular structures can be extracted from the vibrational analysis of the SERS spectrum. In this study, the SERS spectral results demonstrated the ability of SERS spectroscopy to provide distinct spectral features that correspond to differences in the chemical composition of the virus samples, the media controls, and the MeOH solvent. Variations in the sample composition contributed to the detected Raman shifts and bandshape. For example, the prominent Raman shifts detected at 1067, 855, and 501 cm^{-1} can be assigned to specific vibrational stretches of amino acids in the virus and media samples. Additionally, Raman contribution from the proteins in the samples also reflected in the appearance of amide I and amide III protein bands at 1698 and 1264 cm^{-1}.

In addition to providing the information on the chemical composition of the virus samples, the sensitivity of the SERS technique also allowed for the detection of virus signal in a complex sample system. Even though the virus sample composed of a complex mixture of virus, proteins, and DMEM components, SERS spectra of the virus samples showed distinct spectral variations that could be used to distinguish the spectra of the virus samples from those of the media controls and the MeOH solvent. SERS not only detected the virus signal in a complex mixture; SERS also detected the virus signal at a low viral concentration. Recalling that the original concentration of the virus sample was 10^6 pfu/mL, a 1-μL deposition of the 10x-dilution on a 1 cm x 1 cm substrate gave a viral surface concentration of 100 pfu/cm^2. Thus, the virus signal was detected at a very diluted concentration. Overall, the distinct reproducible spectral variations and the detected virus signal confirmed that AgNR substrates provided high sensitivity necessary for low level virus detection.

The results from the study indicated that the SERS spectra were reproducible for all 7 classes of samples. The classes were composed of 4 MV genotypes, 2 media controls, and MeOH solvent. Because high spectral reproducibility was obtained for samples within the class, the spectral variations between the classes were also reproducible. As a result, the reproducibility of the data enabled chemometrics to successfully differentiate samples of A, D4, D9, and H1 genotypes.

The first chemometric method applied for separation of the measles samples was PCA. PCA separated samples by detecting the maximum total variability in the data [17]. Consequently, if the intergroup variability was much greater than the intragroup variability, the intergroup variability reflected in the total maximum variability, and PCA grouped the samples correctly. As a result, because the significant variations among the A, H1, ctrl 3 (SLAM), ctrl 8,

and MeOH classes were detected by PCA, PC1 versus PC2 scores plot displayed separate clusters for these 5 classes. The PC scores plot indicated 56 % of the total variation was sufficient for discrimation of the A, H1, ctrl 3 (SLAM), ctrl 8, and MeOH classes. However, the spectral variation between the D4 and D9 classes could not be detected by PCA. Therefore, the samples from D4 and D9 classes overlapped in the PC scores plot. The overlapped clusters confirmed that the PCA model could not group samples accurately if the variations between the classes and variations within the class were not distinguishable.

The second chemometric approach for classifying the SERS measles data was HCA. In the analysis, an HCA dendogram was generated according to KNN grouping. Samples were assigned to the class with the most nearest neighbors [17]. Additionally, the Mahalanobis distance was use to determine the proximity between the samples. With the exception of D4 and D9 samples, HCA correctly clustered samples from the A, H1, ctrl 3 (SLAM), ctrl 8, and MeOH classes. In the HCA dendogram, the clustering pattern of the virus, control, and solvent classes correctly reflect the compositional relationship between these classes. For example, the MeOH samples did not contain the measles virus or cell culture components. In the dendogram, the MeOH cluster was considered to be the class with the most statistical differences in the data set. The next most statistically different clusters belonged to ctrl 3 (SLAM) and ctrl 8 classes. Since MeOH was the common solvent for both the virus and the control samples, the diluted control sample contained MeOH and cell culture components. In the HCA dendogram, the two controls appeared as the intermediate clusters between the sample and solvent clusters. For the control sample, the distance along the x axis of the dendogram indicated that the statistical distance between the controls and the virus samples was greater than the virus sample-to-sample distance, and smaller than the distance between the solvent and virus samples. Therefore, the HCA

dendogram correctly presented the controls as the intermediate between the solvent and the virus samples.

The third chemometric method applied in discrimination of the measles samples was PLS-DA. Similar to the PC scores plot, the LV scores plot produced by PLS-DA analysis also displayed separate clusters for the A, H1, ctrl 3 (SLAM), ctrl 8, and MeOH classes. The overlap of D4 and D9 samples could not be resolved in the LV scores plot. However, the calibrated PLS-DA model proved PLS-DA was able to accurately classify samples of D4 and D9 genotypes. According to the calculated y-predicted values for the samples and the defined thresholds for the class prediction model, the samples from the A, H1, D9, ctrl 3 (SLAM), ctrl 8, and MeOH groups were assigned to the correct classes. In the prediction of the D4 class, one D4 sample was misclassified. Thus, the prediction sensitivity for D4 was 90 %. Additionally, 1 MeOH sample was incorrectly assigned to the D4 class. Consequently, the specificity of the model was 98.2 %.

To determine the prediction efficiency, the calibrated model was tested by cross-validation (CV) using the Venetian blinds method. Because 11 data splits were chosen for the cross-validation, the samples were divided into 11 groups. In the first data split, the first group of samples was removed, and a new model was calculated using the remaining samples in the data as the training set. Then, the samples that were withheld from the model calculation were used as the unknowns to test the cross-validated model. After cross-validation, the CV y-predicted plots showed that all the samples were correctly identified, including those from D4 and D9 classes. Therefore, the sensitivity of the cross-validated model was 100 % for the prediction of all 7 classes. However, because 2 samples were misclassified in the D9 prediction, the specificity of the cross-validated model fell between 96.4 and 100 %. Overall, the results from

the PLS-DA analysis indicated that the PLS-DA prediction model effectively resolved the overlap between D4 and D9 classes. Thus, PLS-DA was a powerful chemometric method for discriminating measles samples, especially when both intergroup and intragroup variabilities contributed significantly to the total variation in the data.

In summary, the reproducible variations in the SERS spectra of the virus, control, and MeOH samples showed that SERS could be a promising spectroscopic assay for measles detection. Additionally, SERS detection of virus signal in a complex sample mixture and at low viral concentration indicated that AgNR substrates provided the high sensitivity required for low level viral detection. In the data interpretation, results from the chemometric analysis suggested that PCA and HCA can separate virus samples from H1 and A genotypes. However, PLS-DA was the most effective chemometric method for discriminating measles samples of all 4 genotypes. The successful differentiation of the measles genotypes suggested that SERS AgNRs can be combined with chemometric analysis to create a potential diagnostic tool for rapid and accurate measles identification.

Acknowledgements

The research was funded by the U.S. Army Research Laboratory (W911NF-07-2-0065) grant. The author thanks Ralph Tripp, Paul Rota, and the Centers for Disease Control (CDC) for supplying the measles samples, the Yiping Zhao's group for assistance with substrate preparation, and Jeremy Driskell for assistance with chemometric analysis.

References

1. Schneider-Schaulies, S.; Meulen, V. t.; Allan, G.; Robert, G. W. *Encyclopedia of Virology*; Elsevier: Oxford, 1999.
2. Centers for Disease Control and Prevention. *Epidemiology and Prevention of Vaccine Preventable Diseases*. Atkinson W, Hamborsky J, McIntyre L, Wolfe S, eds. 10th ed. Washington DC: Public Health Foundation, 2007.
3. World Health Organization, Progress in global measles control and mortality reduction, 2000 – 2006. *Weekly Epidemiological Record* **2007**, 82, 418-424.
4. Rima, B. K.; Earle, J. A. P.; Baczko, K.; terMeulen, V.; Liebert, U. G.; Carstens, C.; Carabana, J.; Caballero, M.; Celma, M. L.; FernandezMunoz, R., Sequence divergence of measles virus haemagglutinin during natural evolution and adaptation to cell culture. *Journal of General Virology* **1997,** 78, 97-106.
5. Centers for Disease Control (CDC), Outbreak of Measles --- San Diego, California, January—February 2008. *Morbidity and Mortality Weekly Report* **2008**, 57, 1-4.
6. Krause, C. H.; Molyneaux, P. J.; Ho-Yen, D. O.; McIntyre, P.; Carman, W. F.; Templeton, K. E. Comparison of mumps-IgM ELISAs in acute infection. *Journal of Clinical Virology* **2007,** 38, 153-156.
7. R. Bhargava and I.W. Levin, “Fourier Transform Mid-Infrared Spectroscopic Imaging”, in *Spectrochemical Analysis Using Infrared Multichannel Detectors*, R. Bhargava and I.W. Levin, eds., Blackwell Publishing, Oxford, 2005, pp. 1-24.
8. Riddell, M. A.; Moss, W. J.; Hauer, D.; Monze, M.; Griffin, D. E., Slow clearance of measles virus RNA after acute infection. *Journal of Clinical Virology* **2007,** 39, 312-317.

9. Counihan, N. A.; Daniel, L. M.; Chojnacki, J.; Anderson, D. A., Infrared fluorescent immunofocus assay (IR-FIFA) for the quantitation of non-cytopathic and minimally cytopathic viruses. *Journal of Virological Methods* **2006,** 133, 62-69.
10. Vo-Dinh, T., Surface-enhanced Raman spectroscopy using metallic nanostructures. *Trac-Trends in Analytical Chemistry* **1998,** 17, 557-582.
11. Helm, D.; Labischinski, H.; Naumann, D., Elaboration of a procedure for identification of bacteria using Fourier-transform IR spectral libraries – a stepwise correlation approach. *Journal of Microbiological Methods* **1991,** 14, 127-142.
12. Lasch, P.; Beekes, M.; Schmitt, J.; Naumann, D., Detection of preclinical scrapie from serum by infrared spectroscopy and chemometrics. *Analytical and Bioanalytical Chemistry* **2007,** 387, 1791-1800.
13. Verduin, B. J. M.; Prescott, B.; Thomas, G. J., RNA protein interactions and secondary structures of cowpea chlorotic mottle virus for invitro assembly. *Biochemistry* **1984,** 23, 4301-4308.
14. Stewart, S.; Fredericks, P. M., Surface-enhanced Raman spectroscopy of peptides and proteins adsorbed on an electrochemically prepared silver surface. *Spectrochimica Acta Part a-Molecular and Biomolecular Spectroscopy* **1999,** 55, 1615-1640.
15. Podstawka, E.; Ozaki, Y.; Proniewicz, L. M., Adsorption of S-S containing proteins on a colloidal silver surface studied by surface-enhanced Raman spectroscopy. *Applied Spectroscopy* **2004,** 58, 1147-1156.
16. Koglin, E.; Sequaris, J. M.; Fritz, J. C.; Valenta, P., Surface enhanced Raman-scattering (SERS) of nucleic-acid bases adsorbed on silver colloids. *Journal of Molecular Structure* **1984,** 114, 219-223.

17. Beebe, K.R., Pell, R.J., Seasholtz, M.B., *Chemometrics: A Practical Guide.* John Wiley and Sons: New York, 1998; p. 81-84.

Table 2.1 Band assignments for prominent Raman shifts detected for the virus samples.

Raman shift (cm^{-1})	**Assigments**[a]
1457	CH_2 scissor, Gly, Gln; CH_3 asymmetric deformation; ribose phosphate
1395	CH_3 symmetric deformation; ribose phosphate
1324	CH_2 wag, Gly
1264	amide III; side chain vibration, Tyr
1244	amide III, Cyt
1174	NH^+ deformation, Pro; CH_3 asymmetric rock or out-of-plane, Val, Leu
1134	NH_3^+ deformation, Trp
1067	C-N stretch, Leu, Lys; ribose phosphate
872	Indole NH displacement, Trp; ribose phosphate
855	C-C stretch, Tyr, Leu; ribose phosphate
802	Cyt
768	CH out-of-plane deformation of aromatics, Trp
718	COO^- deformation, amino acid backbone
660	His
604	Gly; ribose phosphate
576	ribose phosphate
539	Ade
519	Ala
501	S-S stretch, Cys
426	ribose phosphate
394	ribose phosphate

[a] The presence of proteins, amino acids, and RNA contributed to the Raman shifts observed in the spectra.

Table 2.2 Comparison of measles sample with the solvent spectra.

Classes[a]	**Raman shift (cm^{-1})**	**Assigments**
A, ctrl 3, ctrl 8	1698	amide I; C=O stretch, Gln
A, ctrl 3, ctrl 8	1651	amide I
MV & ctrl media	1017	Indole asymmetric ring breathing, Trp; C-C str; C-N stretch, Leu
MV & ctrl media	856	Tyr; ribose phosphate
MV & ctrl media	538	Ade
MV & ctrl media	517	Ala
MV & ctrl media	502	S-S stretch, Cys

[a]These Raman shifts were observed only in the sample spectra. For clarity, the classes that presented the detected characteristic Raman shifts were indentified in the table.

Table 2.3 Comparison of measles sample with media control spectra.

Class[a]	**Raman shift (cm^{-1})**	**Assigments**
ctrl 3	1510	Ade
ctrl 3	960	CH_3 asymmetric rock or out-of-plane, Val, Leu
ctrl 3	733	COO^- deformation
ctrl 8	1288	amide III shoulder; H
ctrl 8	1199	Tyr
ctrl 8	601	Gly; ribose phosphate
ctrl 8	577	ribose phosphate
ctrl 8	521	Ala
ctrl 8	501	S-S stretch, Cys

[a] These bands were present only in the spectra of ctrl 3 (SLAM) and ctrl 8.

Table 2.4 Comparison of the sample spectra for different measles genotypes.

Class[a]	**Raman shift (cm^{-1})**	**Assigments**
H1	1394	CH_3 symmetric deformation; ribose phosphate
H1	1325	CH_2 wag, Gly
H1	854	C-C stretch, Leu, Tyr; ribose phosphate
H1	731	Ade
H1	715	COO^- deformation, amino acid backbone
A	801	Cyt
A	771	Trp

[a] These Raman shifts highlighted some of the spectral variations that distinguished the H1 and A genotypes from the D4 and D9 genotypes.

Table 2.5 Threshold values for the PLS-DA classification of SERS data.

Class	**Y predicted**	**Threshold**[a]
MeOH	1	0.1443
A	2	0.3598
D9	3	0.4885
H1	4	0.3180
ctrl 3 (SLAM)	5	0.1800
D4	6	0.4964
ctrl 8	7	0.5132

[a] A distinct threshold was calculated for the y-predicted plots of MeOH, A, D9, H1, ctrl 3 (SLAM), D4, and ctrl 8 classes.

Table 2.6 The sensitivity and specificity values of the PLS-DA calibrated (Cal) and cross-validated (CV) prediction models for SERS data.

Class	**MeOH**	**A**	**D9**	**H1**	**ctrl 3**	**D4**	**ctrl 8**
Sensitivity[a] (Cal)	1.000	1.000	1.000	1.000	1.000	0.900	1.000
Specificity[b] (CV)	1.000	1.000	1.000	1.000	1.000	0.982	1.000
Sensitivity (CV)	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Specificity (CV)	1.000	1.000	0.964	1.000	1.000	0.982	1.000

[a] The sensitivity is defined as the number of predicted samples in a class divided by the actual number of samples in the class. [b] The specificity is given as the number of samples predicted not in the class divided by the actual number of samples not in the class.

Figure 2.1 SERS spectra of samples from the A genotype. The spectral background for the bare AgNR substrate was included at the bottom of the plot. The plot indicated that sample signal was significantly greater than the backround. In the sample spectra, it appeared that the peaks around 1578, 1016, and 920 cm^{-1} experienced some contribution from the background. The sample spectra were collected at 5 different spots on the same substrate. The results showed that the SERS spectra were very reproducible.

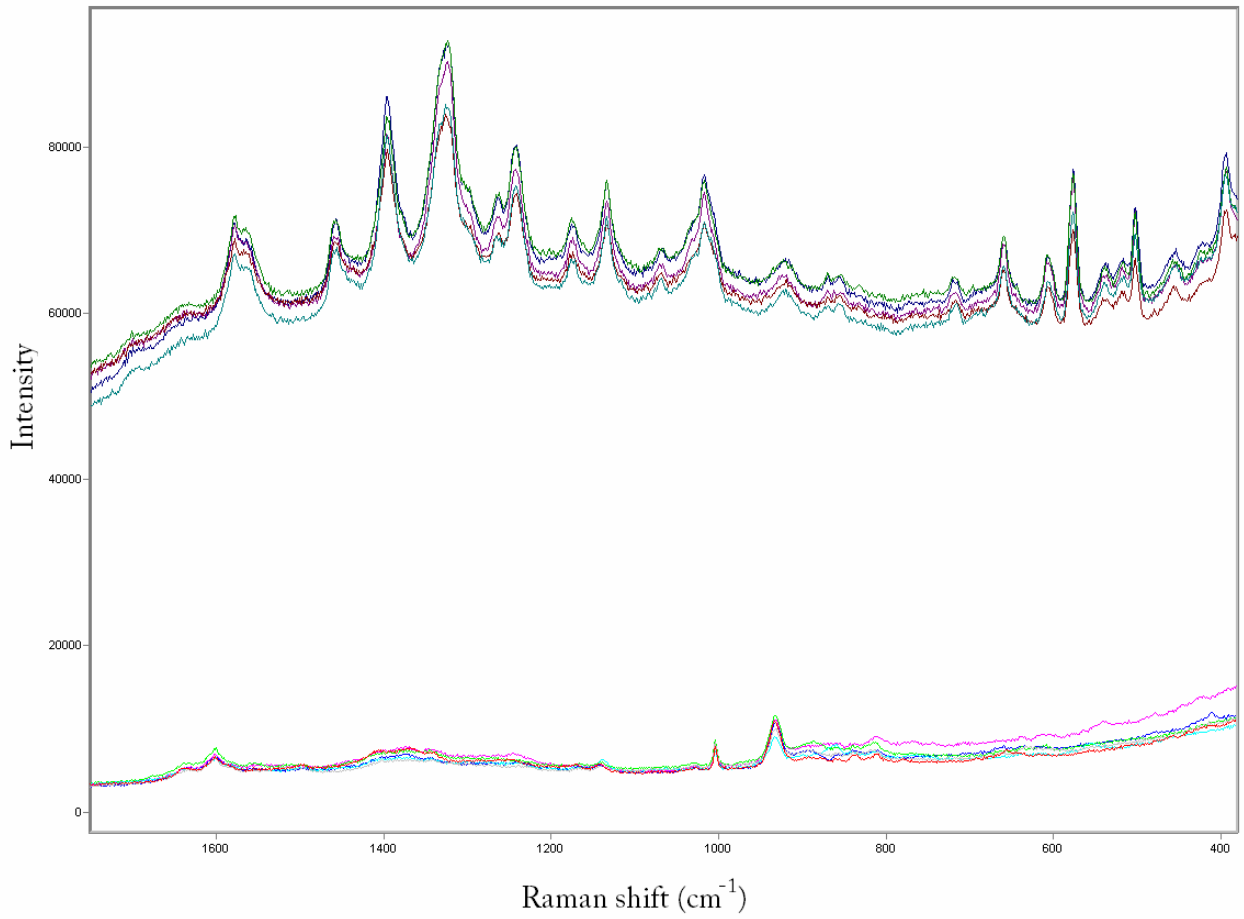

Intensity
80000
60000
40000
20000
0
1600
1400
1200
1000
800
600
400
Raman shift (cm-1)

Figure 2.2 SERS spectra of samples from the D4 genotype. The bottom spectra belonged to the background on the substrate. Reproducible spectra were collected for the D4 samples.

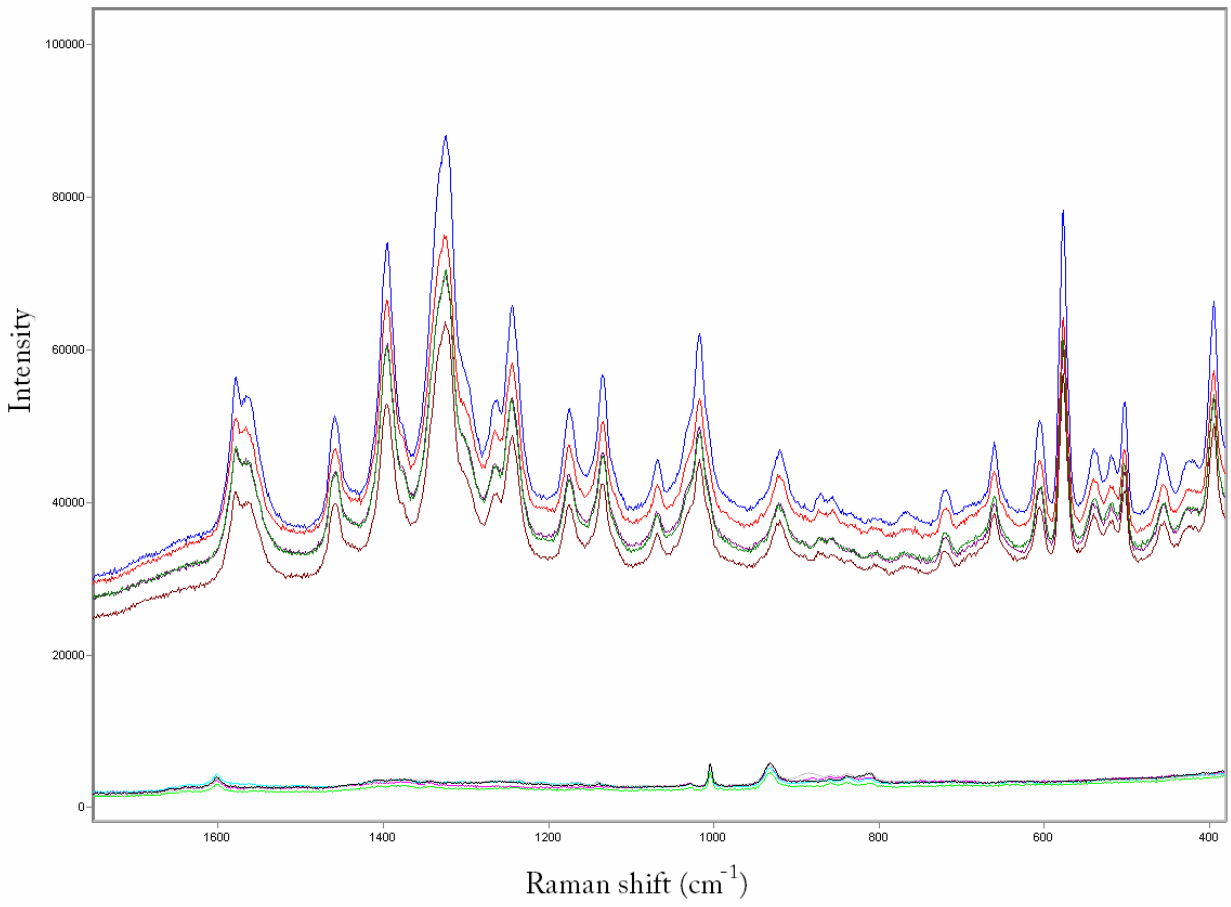

Intensity
100000
80000
60000
40000
20000
0
1600
1400
1200
1000
800
600
400
Raman shift (cm-1)

Figure 2.3 SERS spectra of samples from the D9 genotype. Results indicated reproducible SERS spectra were obtained for D9 samples. The background spectra were presented below the sample spectra.

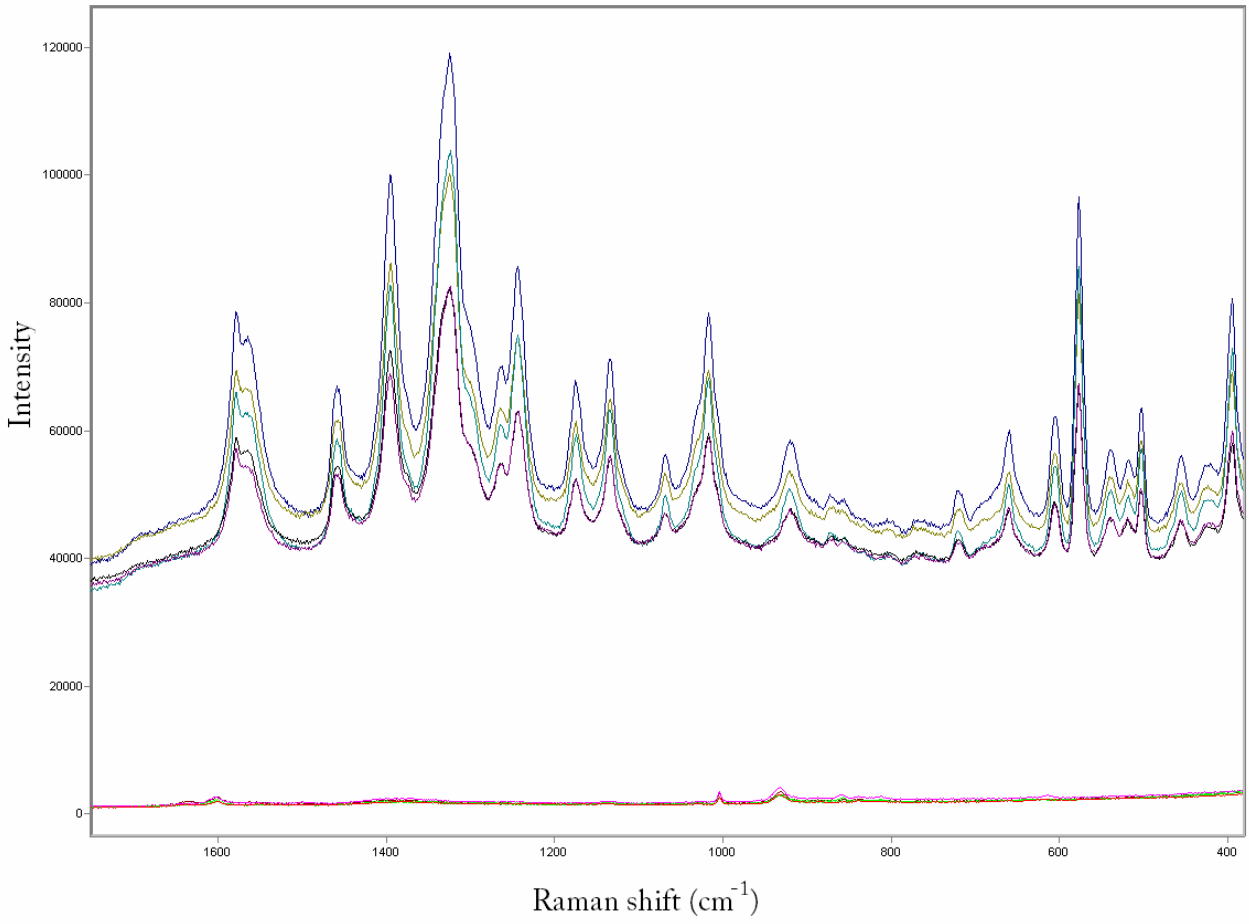

Intensity
120000
100000
80000
60000
40000
20000
0
1600
1400
1200
1000
800
600
400
Raman shift (cm-1)

Figure 2.4 SERS spectra of samples from the H1 genotype. The top spectra belonged to the samples, and the bottom spectra belonged to the background. Reproducible spectra were acquired for both the measles samples and the background.

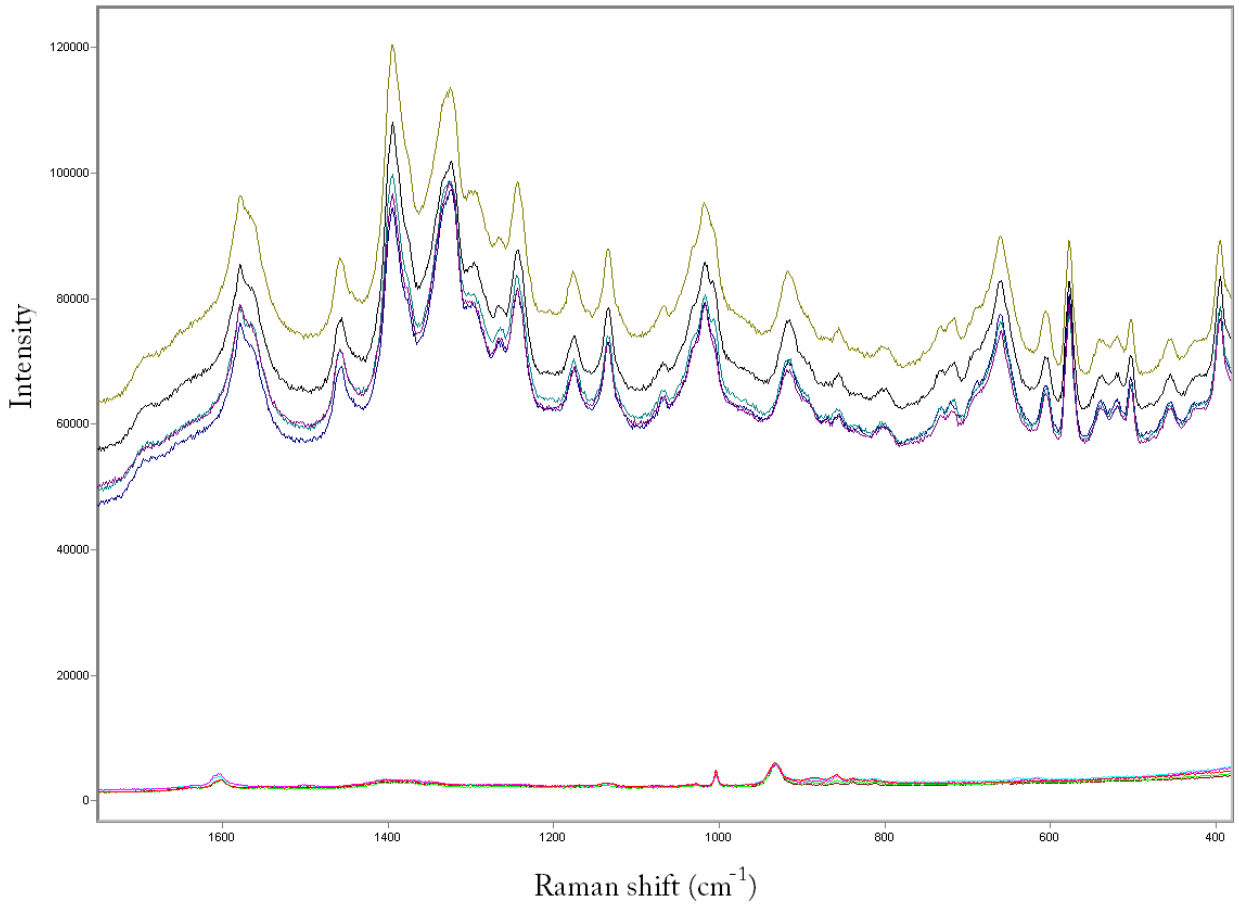

120000
100000
80000
60000
40000
20000
0
Intensity
1600
1400
1200
1000
800
600
400
Raman shift (cm-1)

Figure 2.5 SERS spectra of ctrl 3 (SLAM) media. Ctrl 3 (SLAM) contained the plasmids coding for the SLAM receptor. The background spectra were displayed below ctrl 3 (SLAM) spectra. Similar to the measles sample spectra, the media spectra were also more enhanced than the background.

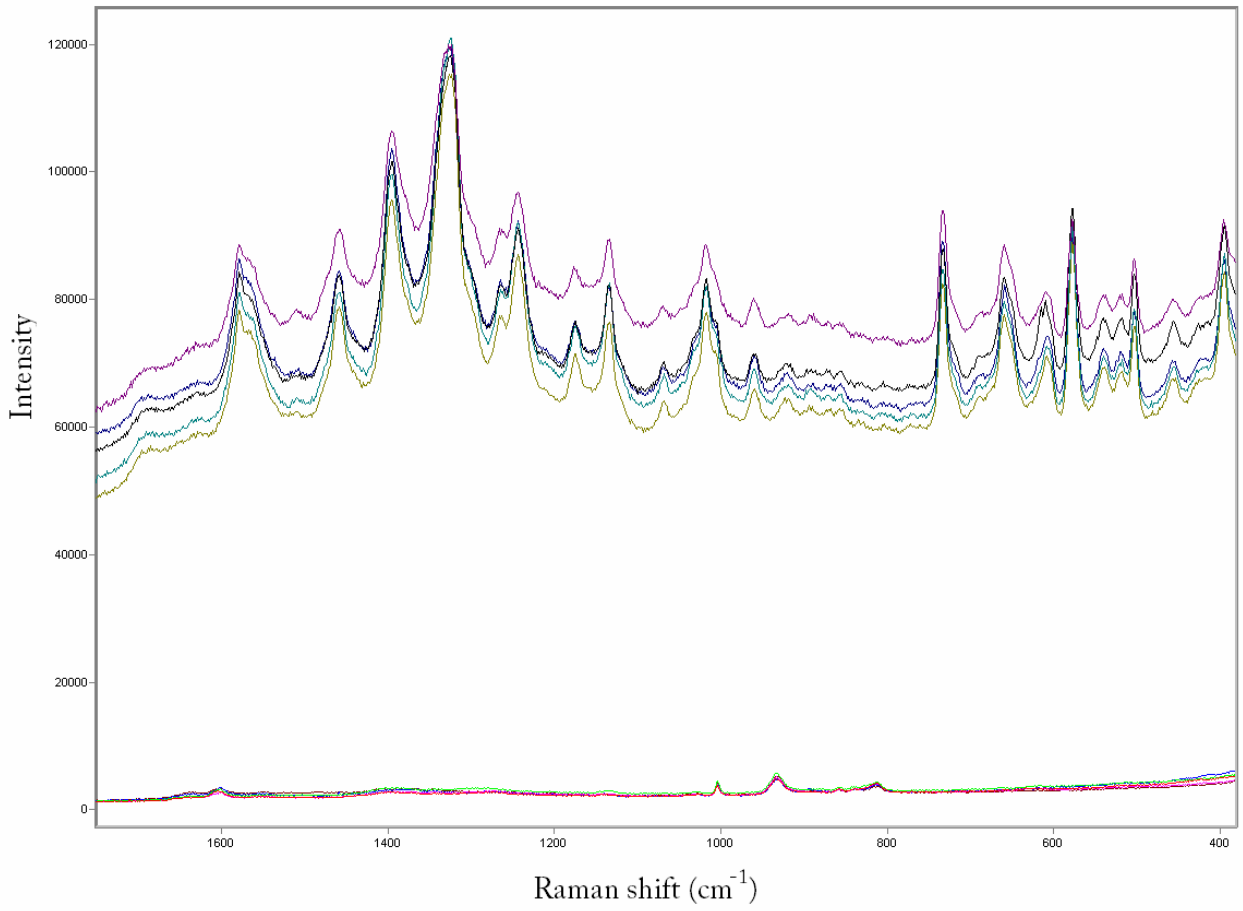

120000
100000
80000
60000
40000
20000
0
Intensity
1600
1400
1200
1000
800
600
400
Raman shift (cm-1)

Figure 2.6 SERS spectra for ctrl 8 media. Ctrl 8 did not contain SLAM receptor-coding plasmids. The measured substrate background was shown below the sample spectra.

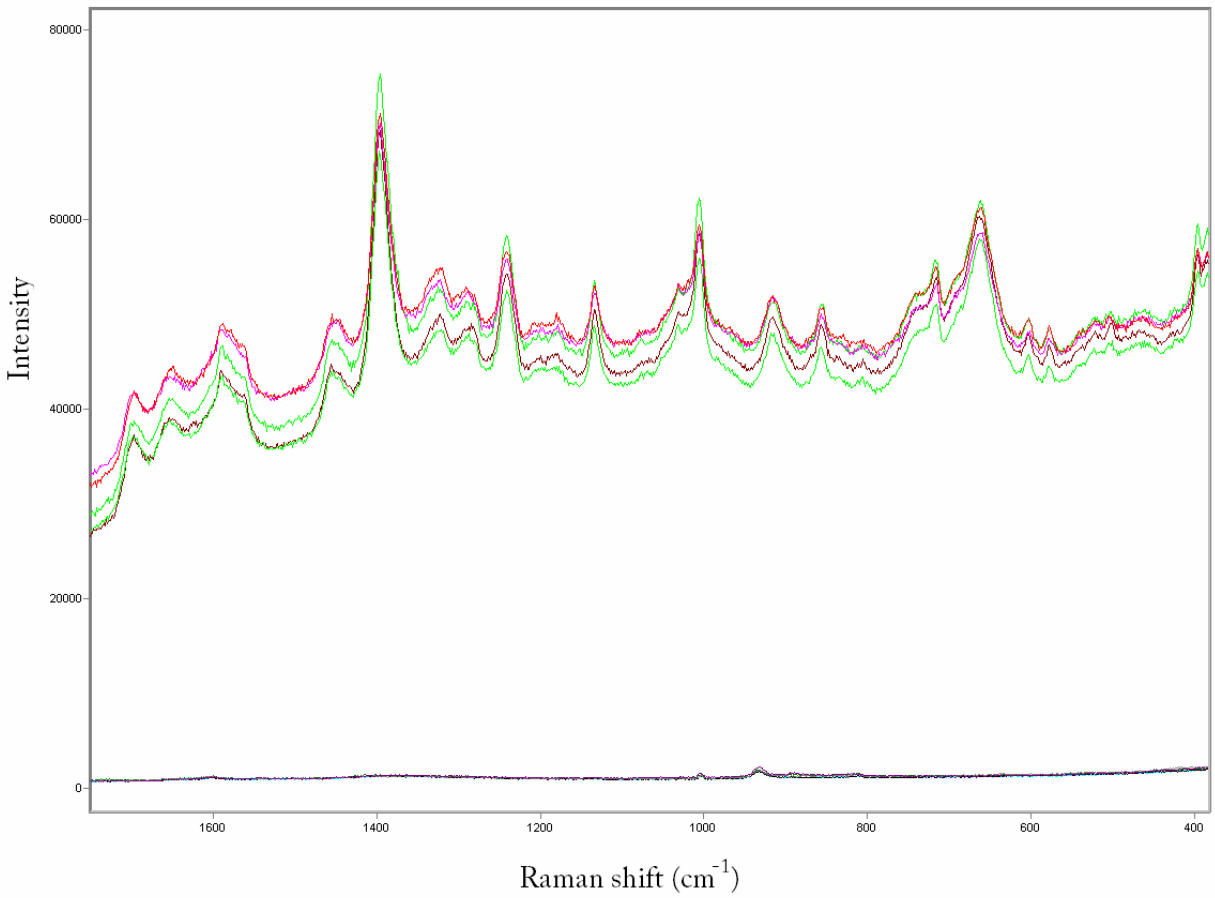

80000
60000
40000
20000
0
Intensity
1600
1400
1200
1000
800
600
400
Raman shift (cm-1)

Figure 2.7 SERS spectra of MeOH solvent on AgNR substrates. The background spectra were shown at the bottom of the plot.

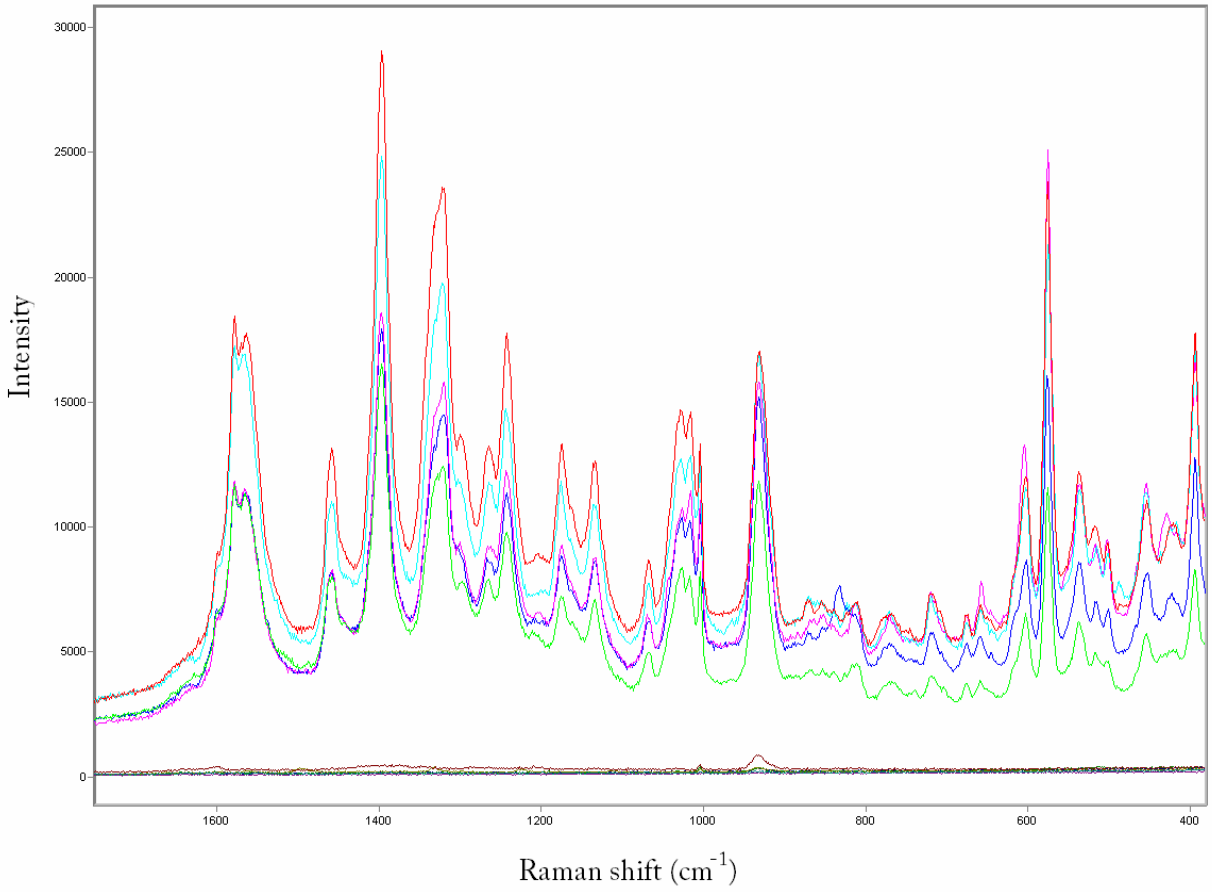

Intensity
30000
25000
20000
15000
10000
5000
0
1600
1400
1200
1000
800
600
400
Raman shift (cm-1)

Figure 2.8 An overlaid of the average SERS spectra collected for the virus samples and media. The average spectra belonged to the 4 measles samples, 2 media controls, and the MeOH solvent. Spectral variations for the samples, media, and solvent were observable. These variations were assigned to different components in the virus and cell culture media.

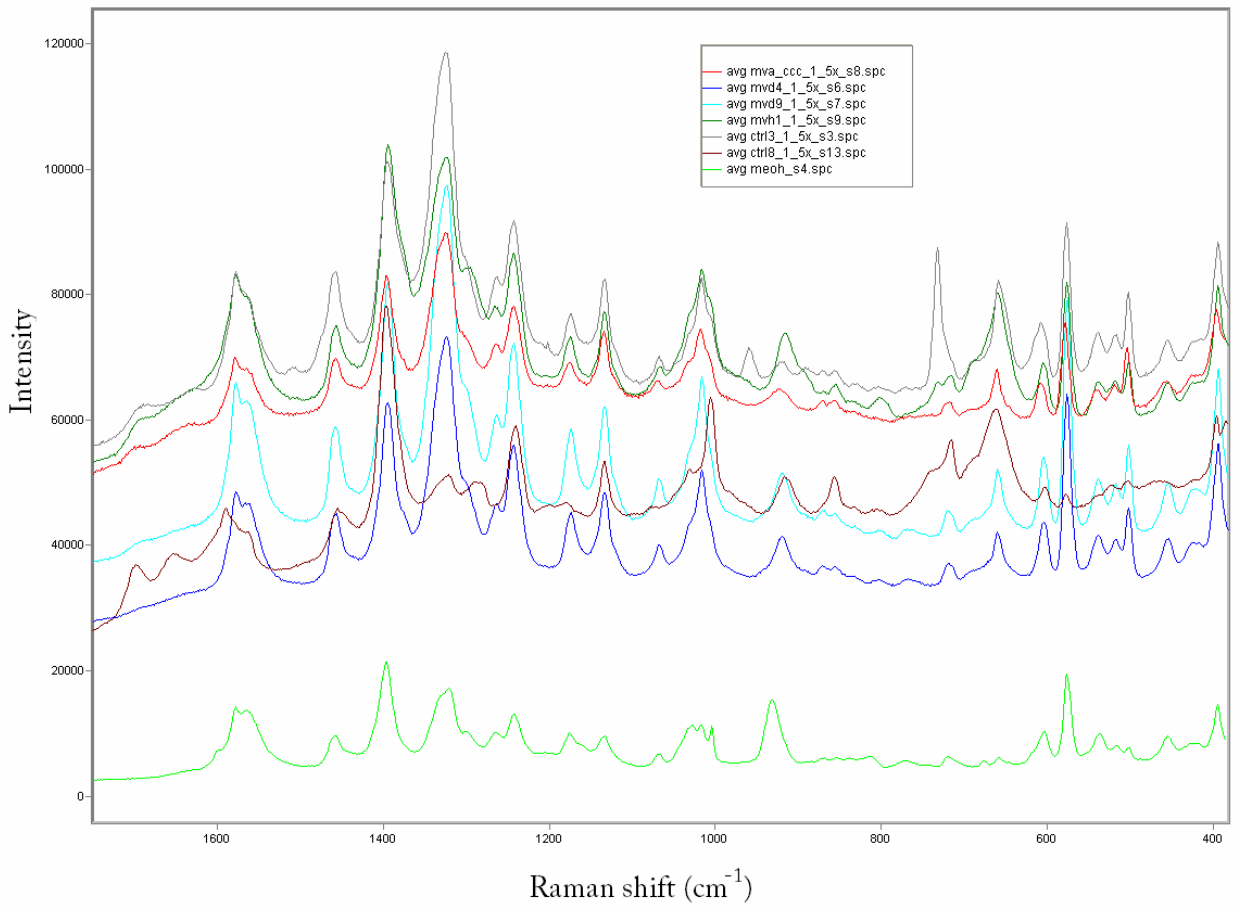

avg mva_ccc_1_5x_s8.spc
avg mvd4_1_5x_s6.spc
avg mvd9_1_5x_s7.spc
avg mvh1_1_5x_s9.spc
avg ctrl3_1_5x_s3.spc
avg ctrl8_1_5x_s13.spc
avg meoh_s4.spc
Intensity
120000
100000
80000
60000
40000
20000
0
1600
1400
1200
1000
800
600
400
Raman shift (cm-1)

Figure 2.9 PC1 versus PC2 scores plot for the SERS data. The scores plot was produced after Principal Component Analysis (PCA) was performed on the SERS data. Separate clusters were displayed for the A, H1, ctrl 3, ctrl 8, and MeOH classes. However, because the small intergroup variability between D4 and D9 classes was not detectable by PCA, D4 and D9 samples overlapped in the scores plot.

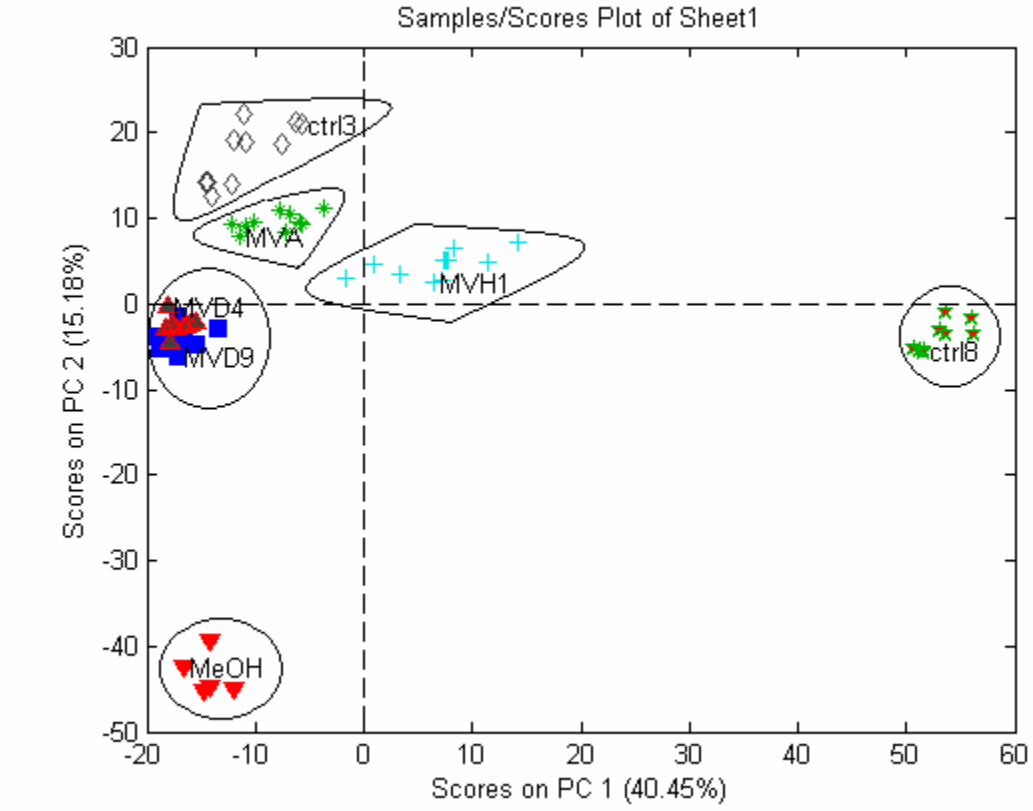

Samples/Scores Plot of Sheet1
Scores on PC 2 (15.18%)
Scores on PC 1 (40.45%)
ctrl3
MVA
MVH1
MVD4
MVD9
ctrl8
MeOH
30
20
10
0
-10
-20
-30
-40
-50
-20
-10
0
10
20
30
40
50
60

Figure 2.10 Dendogram produced by Hierachical Cluster Analysis (HCA) and K-Nearest Neighbor (KNN) classification of the SERS data. The dendogram showed that samples from the A, H1, ctrl 3, ctrl 8, and MeOH classes were correctly identified by HCA. However, similar to PCA, HCA could not distinguish D4 from D9 samples.

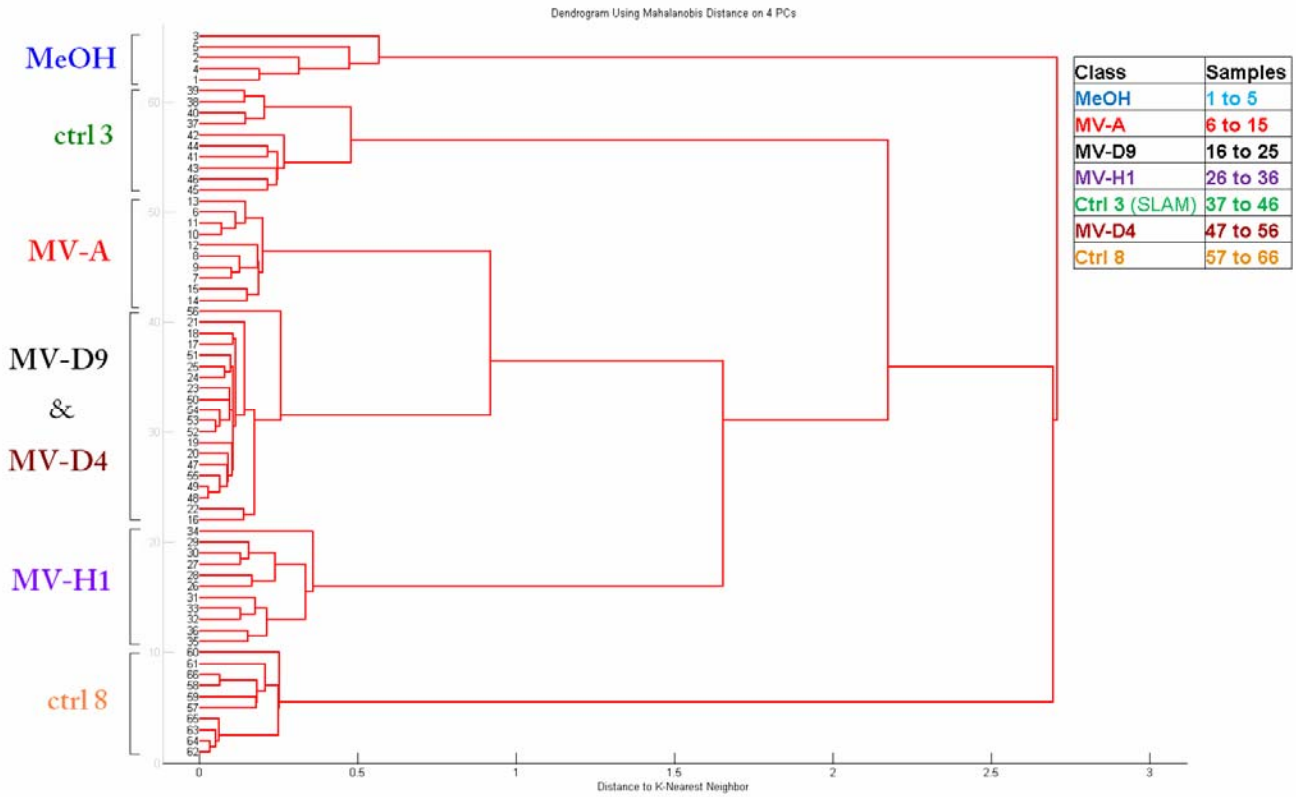

Dendrogram Using Mahalanobis Distance on 4 PCs
MeOH
ctrl 3
MV-A
MV-D9
&
MV-D4
MV-H1
ctrl 8
Class
Samples
MeOH
1 to 5
MV-A
6 to 15
MV-D9
16 to 25
MV-H1
26 to 36
Ctrl 3 (SLAM)
37 to 46
MV-D4
47 to 56
Ctrl 8
57 to 66
0
0.5
1
1.5
2
2.5
3
Distance to K-Nearest Neighbor

Figure 2.11 LV1 versus LV2 scores plot for the SERS data. The LV scores plot indicated that PLS-DA could discriminate samples from the A, H1, ctrl 3, ctrl 8, and MeOH classes. However, PLS-DA LV scores plot could not separate samples from D4 and D9 classes.

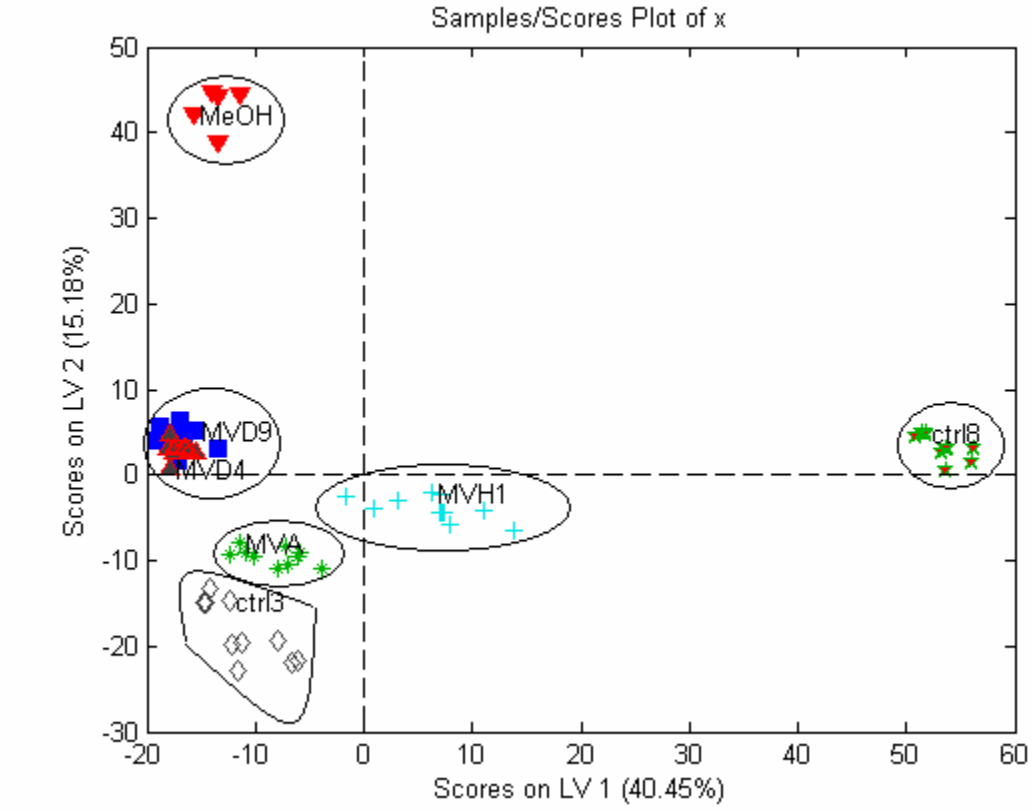

Samples/Scores Plot of x
MeOH
MVD9
MVD4
MVH1
MVA
ctrl3
ctrl8
Scores on LV 2 (15.18%)
Scores on LV 1 (40.45%)
50
40
30
20
10
0
-10
-20
-30
-20
-10
0
10
20
30
40
50
60

Figure 2.12 Y predicted plot for the MeOH class prediction of SERS data. The red dash line represented the threshold for the prediction model. The threshold value defined the cut-off point for the class assignment. Samples that belonged to the MeOH class had a calculated y-predicted value that falls between the threshold (0.1443) and a value that is appromixately close to 1. Samples with y-predicted values below the threshold were identified as not belonging to the MeOH class. According to the y predicted plot, all the samples from the MeOH class were correctly identified by PLS-DA.

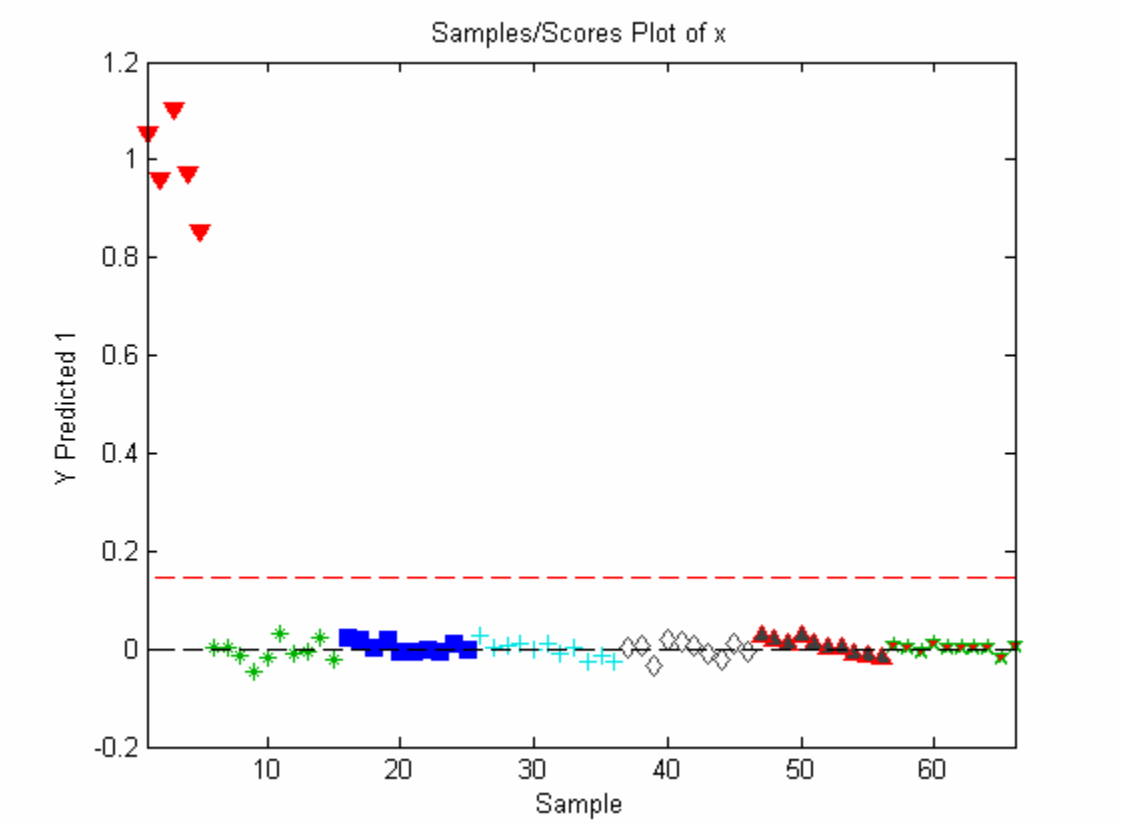

Samples/Scores Plot of x
Y Predicted 1
Sample
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
10
20
30
40
50
60

Figure 2.13 Y predicted plot for the A class prediction of SERS data. Samples with calculated y-predicted values above the 0.3598 threshold were assigned to the A class. In the model, all the A samples were correctly assigned to the A class.

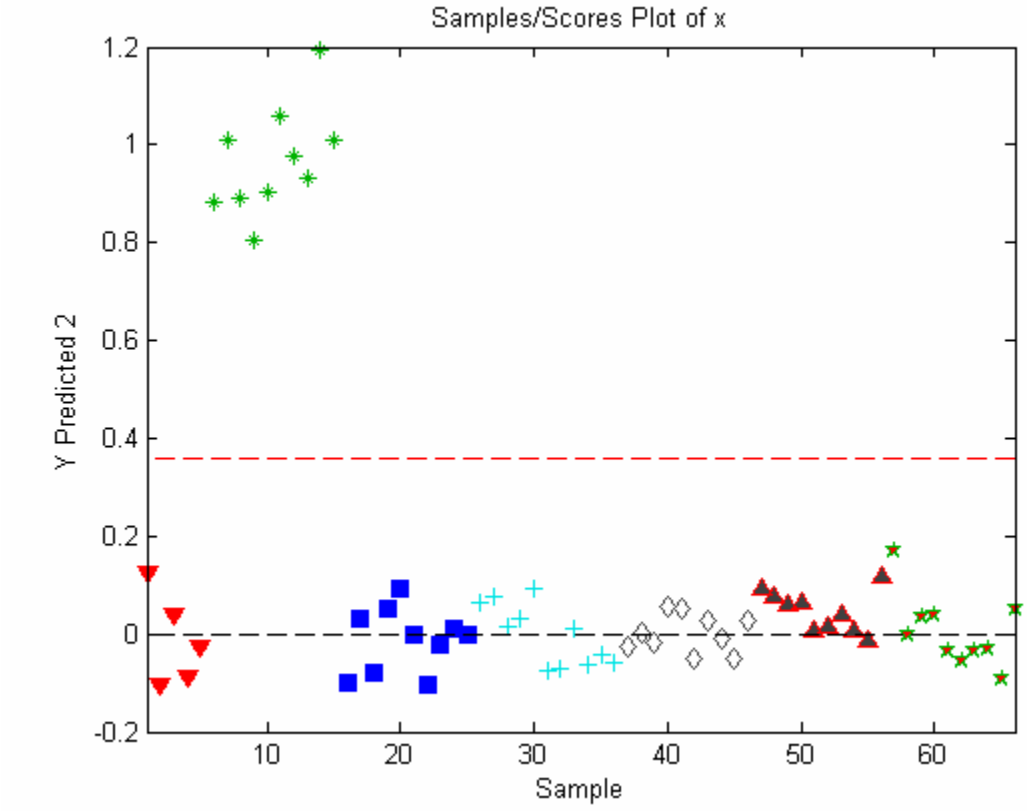

Samples/Scores Plot of x
Y Predicted 2
Sample
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
10
20
30
40
50
60

Figure 2.14 Y predicted plot for the D9 class prediction of SERS data. Samples with calculated y-predicted values above the 0.4885 threshold were assigned to the D9 class. In the model, all the D9 samples were correctly assigned to the D9 class.

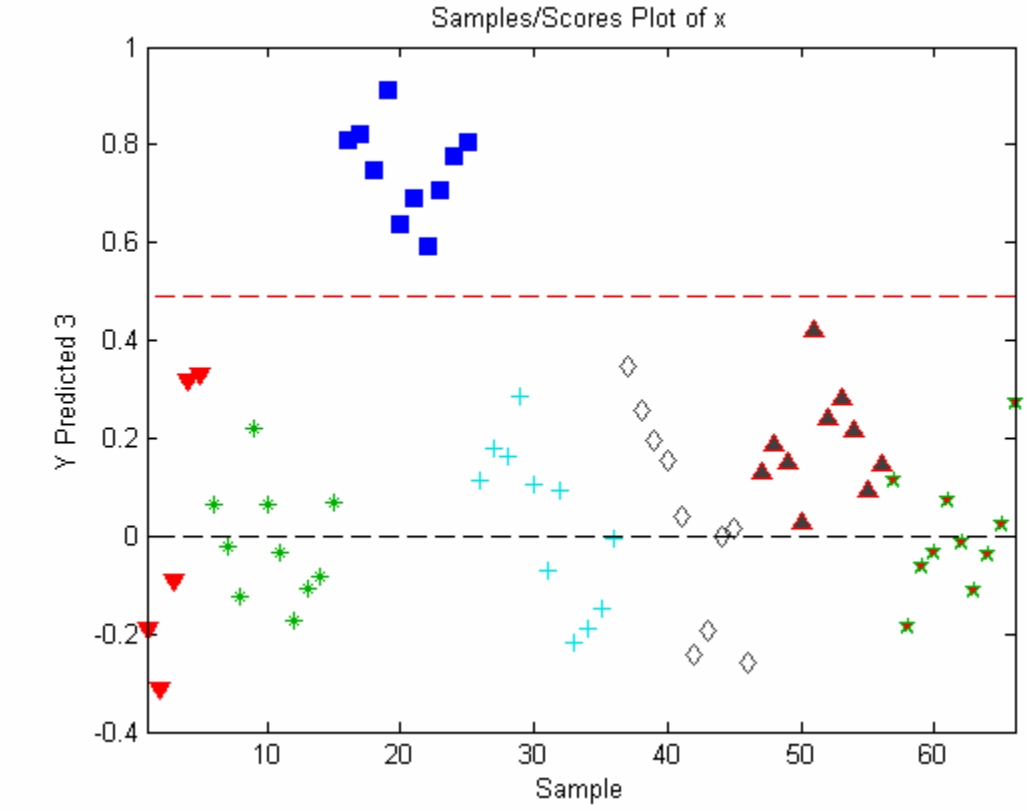

Samples/Scores Plot of x
Y Predicted 3
Sample
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
10
20
30
40
50
60

Figure 2.15 Y predicted plot for the H1 class prediction of SERS data. Samples with calculated y-predicted values above the 0.3180 threshold were assigned to the H1 class. In the model, all the H1 samples were correctly assigned to the H1 class.

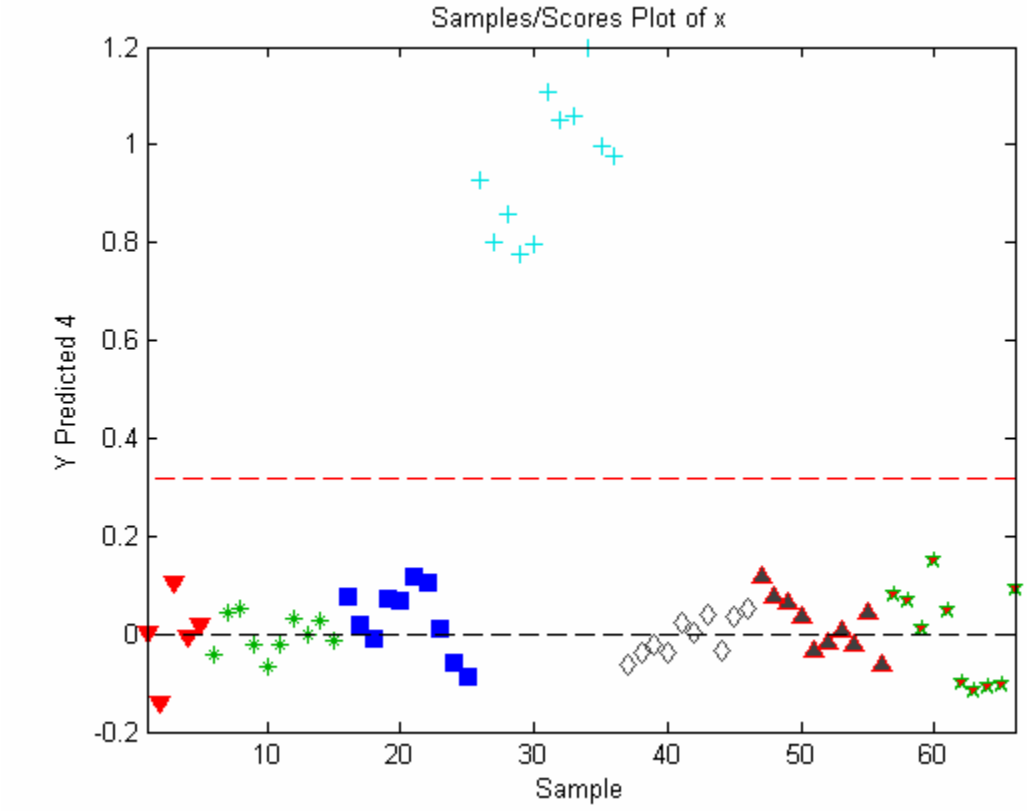

Samples/Scores Plot of x
Y Predicted 4
Sample
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
10
20
30
40
50
60

Figure 2.16 Y predicted plot for the ctrl 3 (SLAM) class prediction of SERS data. Samples with calculated y-predicted values above the 0.1800 threshold were assigned to the ctrl 3 (SLAM) class. In the model, all the ctrl 3 (SLAM) samples were correctly assigned to the ctrl 3 (SLAM) class.

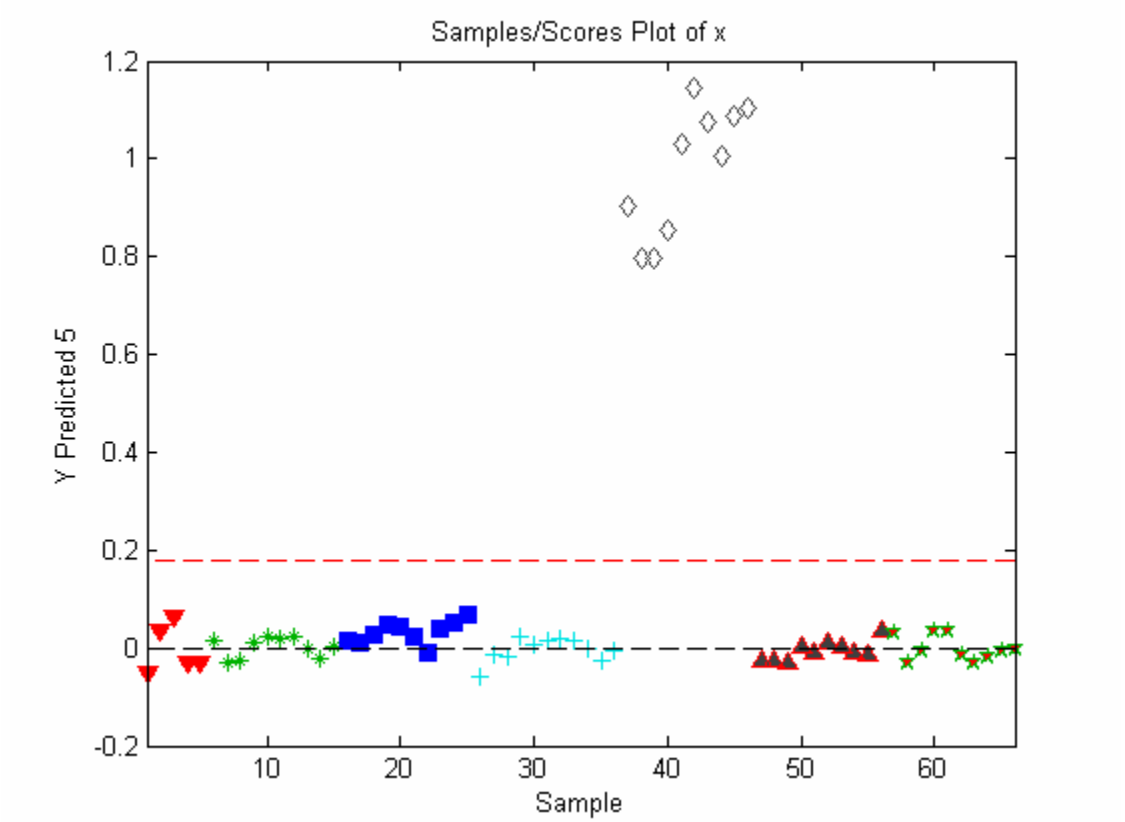

Samples/Scores Plot of x
Y Predicted 5
Sample
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
10
20
30
40
50
60

Figure 2.17 Y predicted plot for the D4 class prediction of SERS data. Samples with calculated y-predicted values above the 0.4964 threshold were assigned to the D4 class. In the model, a D4 sample and a MeOH sample were misclassified. Thus, the sensitivity of calibrate model for the D4 prediction was 90.0 %. The specificity of the calibrated model was 98.2 %.

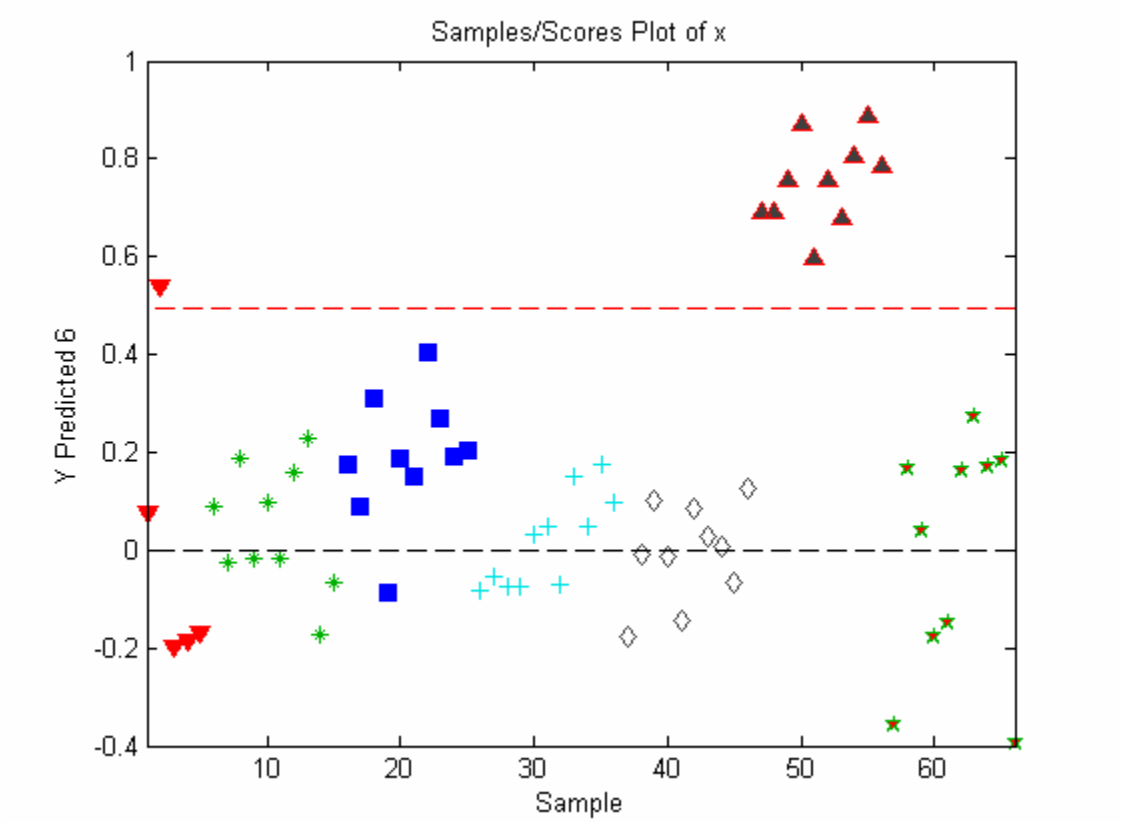

Samples/Scores Plot of x
Y Predicted 6
Sample
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
10
20
30
40
50
60

Figure 2.18 Y predicted plot for the ctrl 8 class prediction of SERS data. Samples with calculated y-predicted values above the 0.5132 threshold were assigned to the ctrl 8 class. In the model, all the ctrl 8 samples were correctly assigned to the ctrl 8 class.

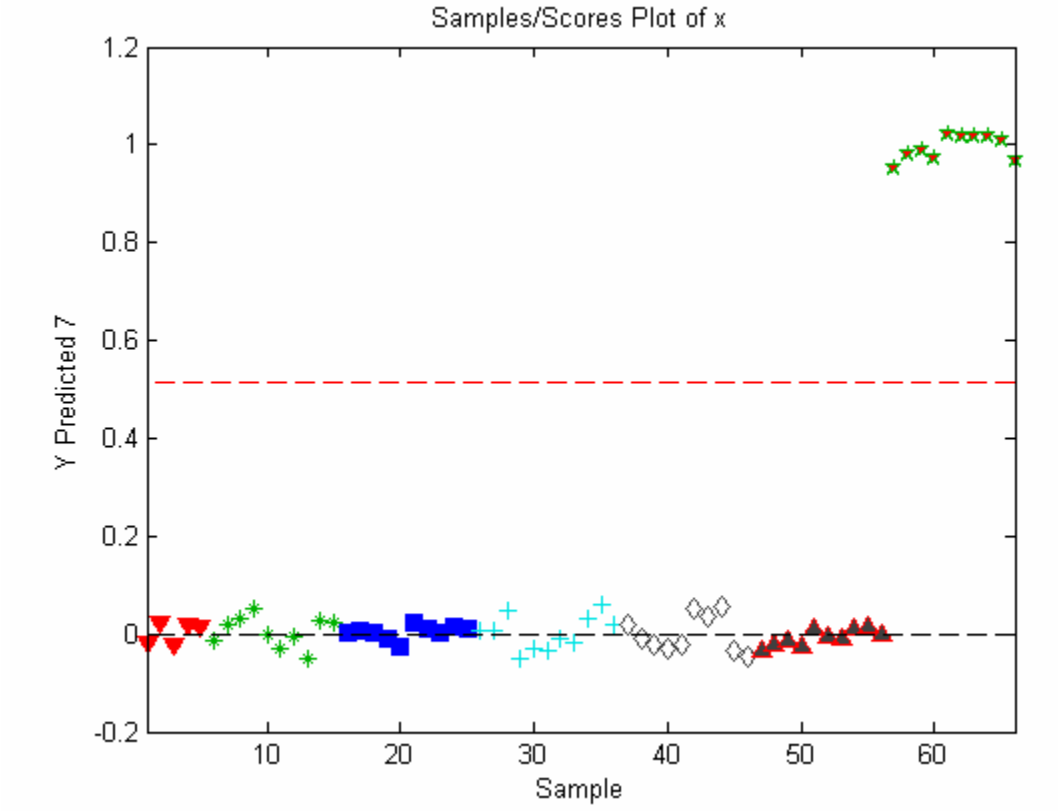

Samples/Scores Plot of x
Y Predicted 7
Sample
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
10
20
30
40
50
60

Figure 2.19 Y CV predicted plot for the MeOH class prediction of SERS data after cross-validation (CV). After cross-validation with the Venetian blinds method, the PLD-DA model correctly predicted the samples from the MeOH class.

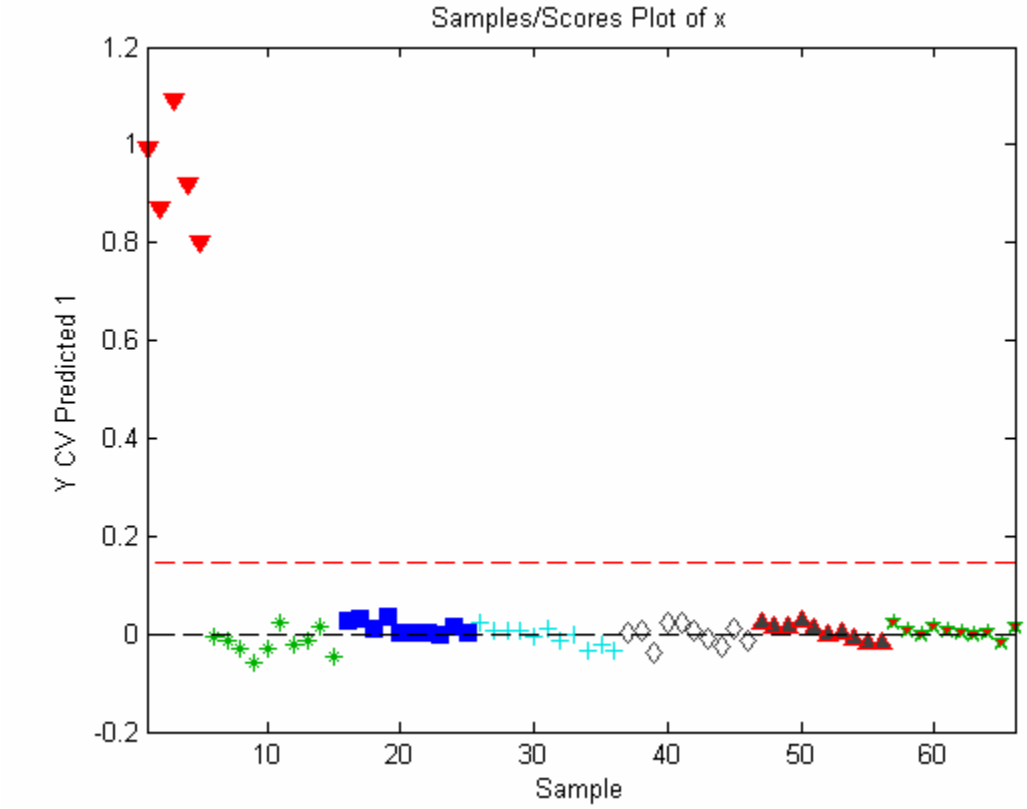

Samples/Scores Plot of x
Y CV Predicted 1
Sample
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
10
20
30
40
50
60

Figure 2.20 Y CV predicted plot for the A class prediction of SERS data after cross-validation (CV). The plot indicated that the A samples remained in the correct class after cross-validation.

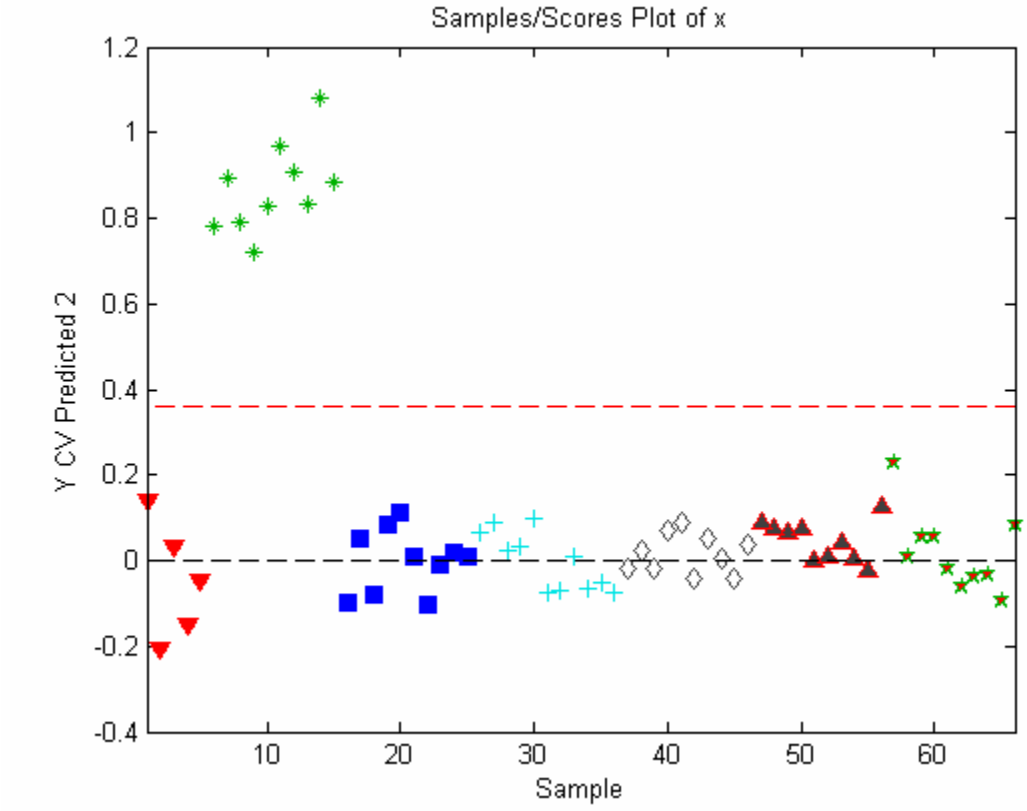

Samples/Scores Plot of x
Y CV Predicted 2
Sample
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
10
20
30
40
50
60

Figure 2.21 Y CV predicted plot for the D9 class prediction of SERS data after cross-validation (CV). The plot indicated that the D9 samples remained in the correct class after cross-validation. However, a D4 sample and a MeOH sample were misclassified. Therefore, the specificity of the cross-validated PLS-DA model was 96.4 % for the prediction of the D9 class.

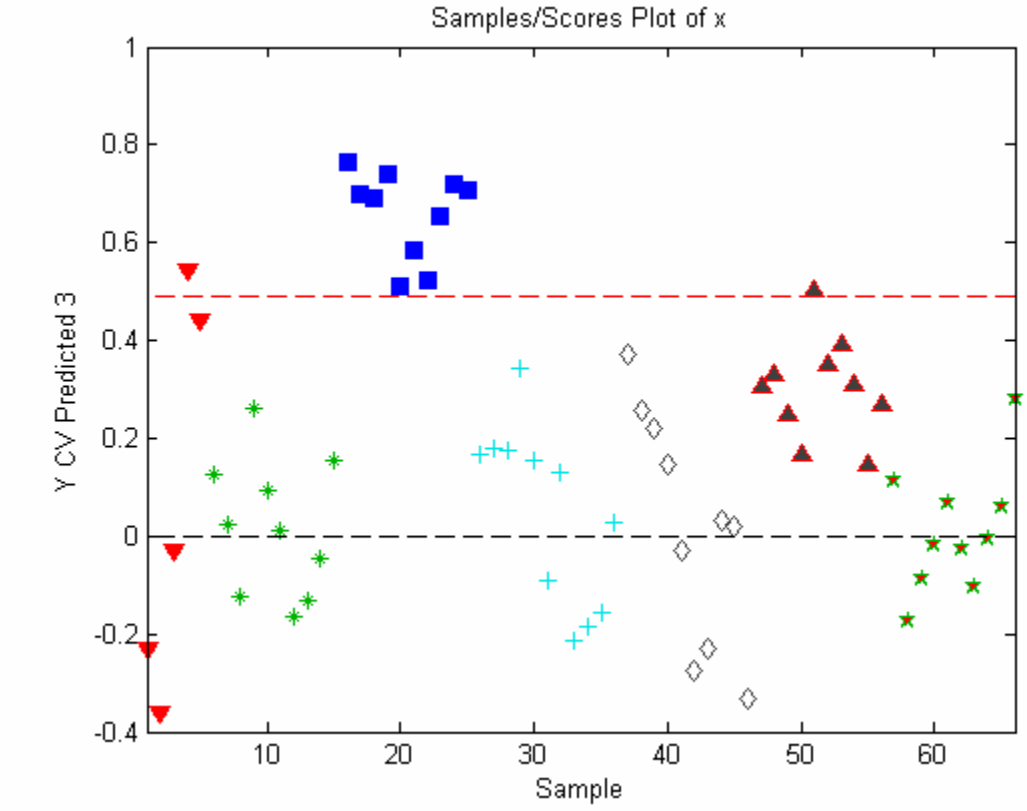

Samples/Scores Plot of x
Y CV Predicted 3
Sample
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
10
20
30
40
50
60

Figure 2.22 Y CV predicted plot for the H1 class prediction of SERS data after cross-validation (CV). The plot indicated that the H1 samples remained in the correct class after cross-validation.

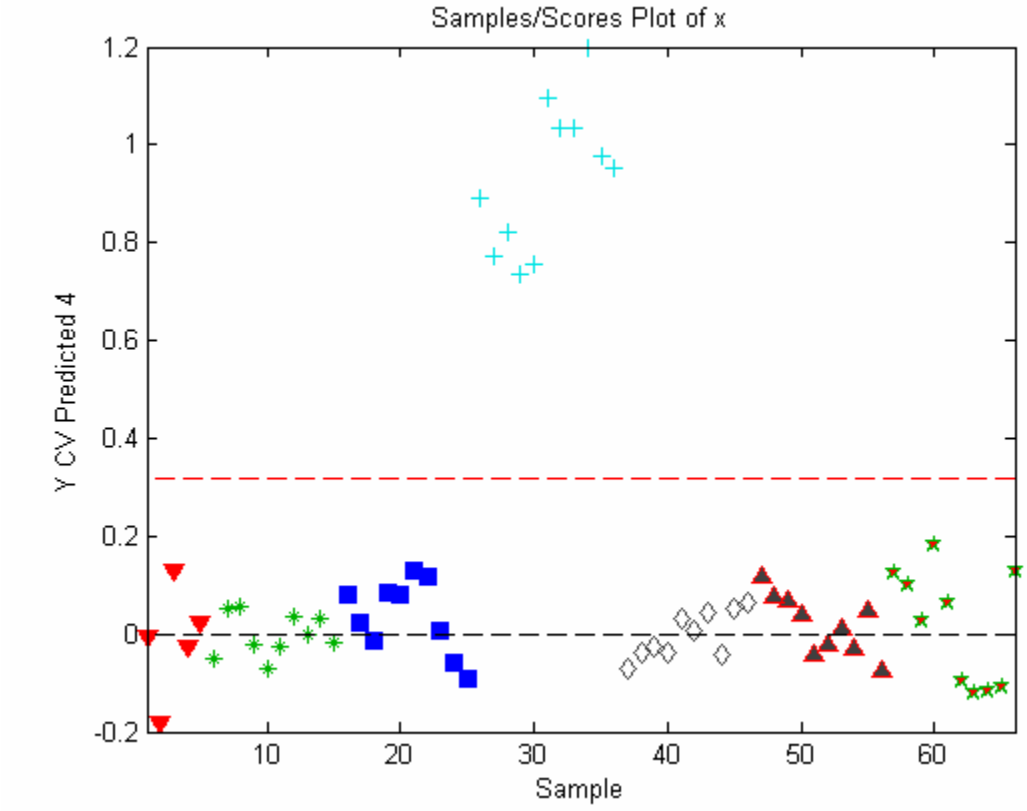

Samples/Scores Plot of x
Y CV Predicted 4
Sample
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
10
20
30
40
50
60

Figure 2.23 Y CV predicted plot for the ctrl 3 (SLAM) class prediction of SERS data after cross-validation (CV). The plot indicated that the ctrl 3 (SLAM) samples remained in the correct class after cross-validation.

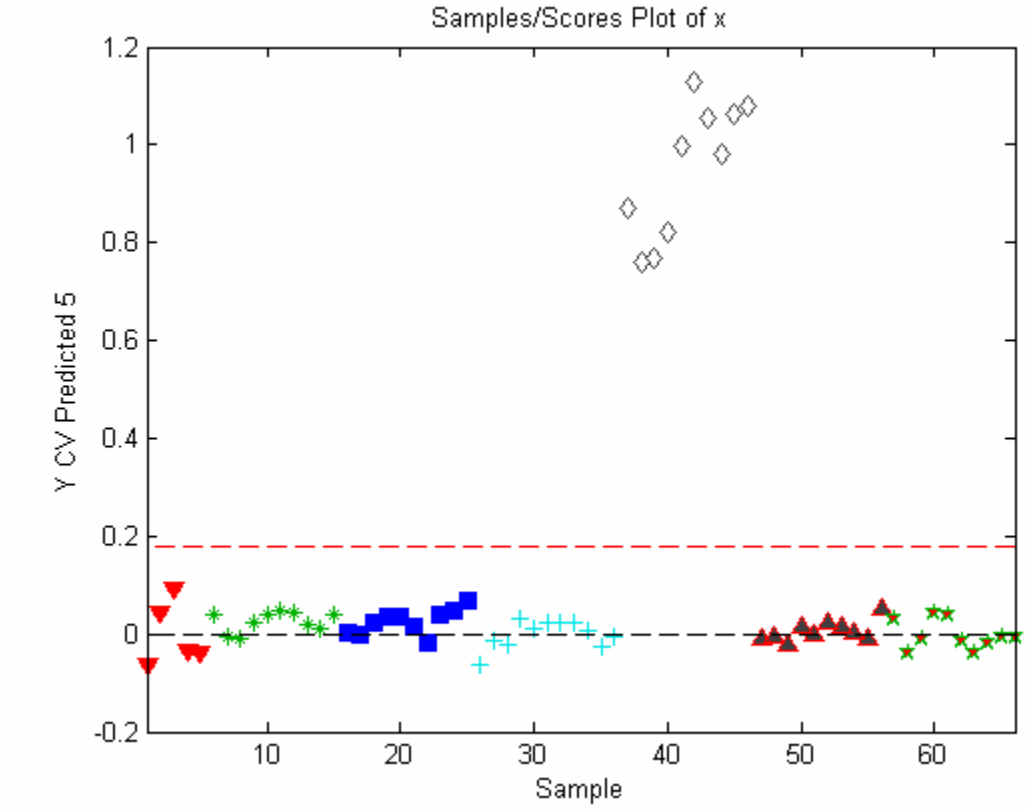

Samples/Scores Plot of x
Y CV Predicted 5
Sample
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
10
20
30
40
50
60

Figure 2.24 Y CV predicted plot for the D4 class prediction of SERS data after cross-validation (CV). The plot indicated that the D4 samples remained in the correct class after cross-validation. However, because one MeOH sample was incorrectly assigned to the D4 class, the specificity of the D4 prediction was 98.2 %.

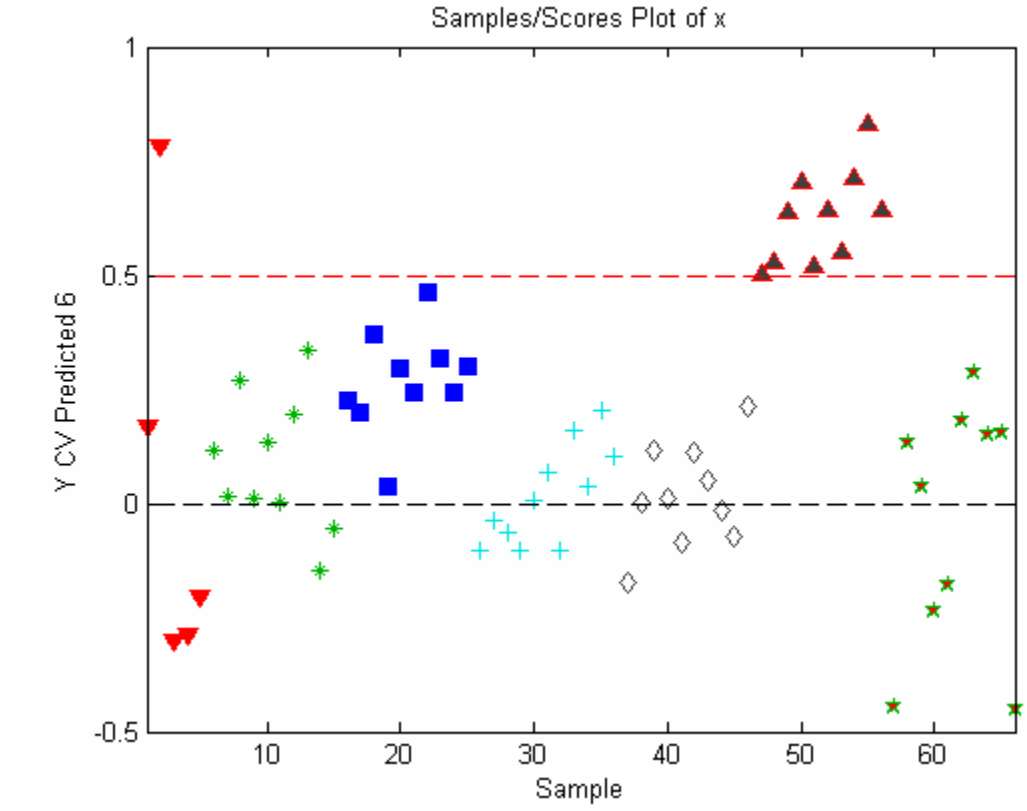
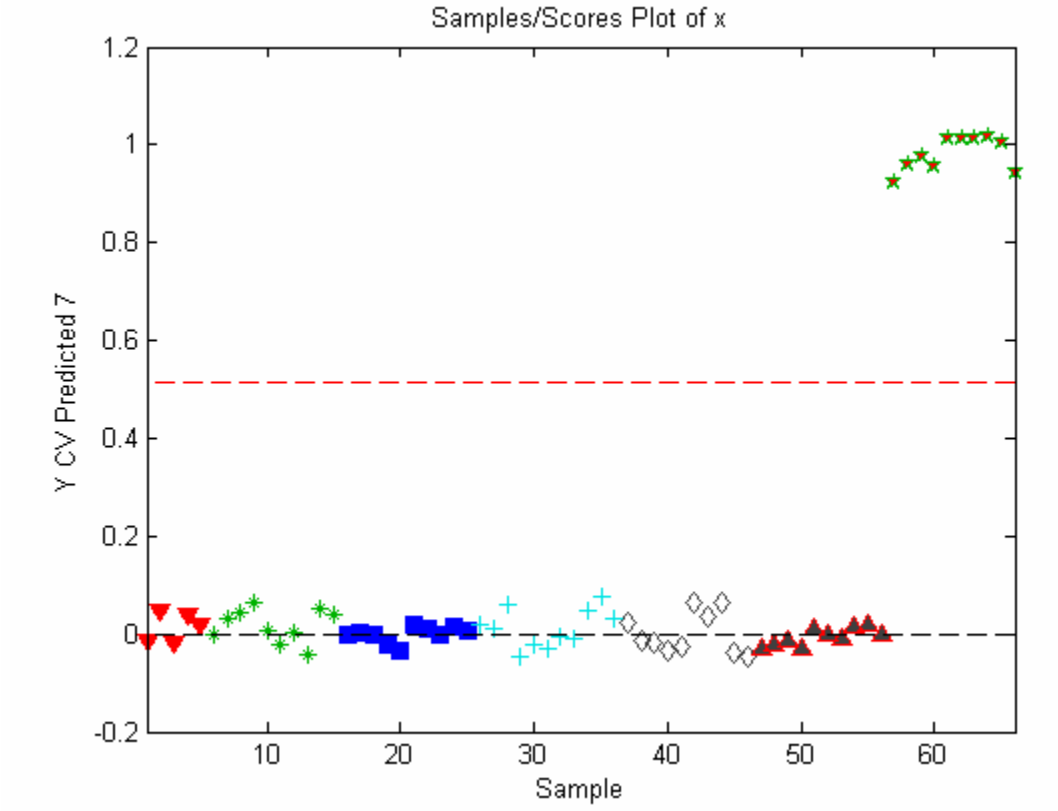

Samples/Scores Plot of x
Y CV Predicted 6
Sample
1
0.5
0
-0.5
10
20
30
40
50
60

Figure 2.25 Y CV predicted plot for the ctrl 8 class prediction of SERS data after cross-validation (CV). The plot indicated that the ctrl 8 samples remained in the correct class after cross-validation.


Samples/Scores Plot of x
Y CV Predicted 7
Sample
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
10
20
30
40
50
60

Chapter III

Detection of Measles Virus by

Fourier Transform Infrared (FTIR) Spectroscopy[2]

[2] Hoang, V., R.A. Dluhy, Y. Zhao, and R. Tripp. To be submitted to *Applied Spectroscopy*.

Abstract

On a global scale, measles is one of the dominant causes of childhood deaths. In 2006, 242,000 measles-related deaths were reported. On a national scale, imported measles strains can lead to measles outbreaks in the US. Often, without effective early phase diagnosis, the infected host could inadvertently spread the infection to other unvaccinated hosts during contact. Thus, rapid and accurate detection of the measles is an essential need in diagnosis and containment of the disease.

In the past, Fourier transform infrared (FTIR) spectroscopy has been the popular method for pathogen detection. For example, previous studies by the Naumann's group demonstrated that FTIR was an effective method for bacterial detection. Additionally, these studies also showed that the coupled application of FTIR with chemometric analysis allowed for accurate discrimination of different bacterial strains. Therefore, in this study, FTIR spectroscopy and chemometric analysis were applied to identify different genotypes of measles.

Introduction

Measles infection is a common disease in infants and young children [1]. In countries where measles vaccines are limited, such as countries in Africa and Asia, measles remains one of the leading causes of childhood deaths [2]. Measles infection is often worse when the disease is combined with secondary factors such as malnutrition. For example, malnutrition is still a major health concern for African children. For these children, the fatality rate from measles infection can be as high as 10 % [2, 3].

Additionally, even with implementation of measles vaccination programs, virus importation can still provide a new route for measles transmission [4]. In the US, the source for recent measles outbreaks reported between 2005 and 2008 has been virus importation. For example, the measles outbreak that occurred in San Diego at the beginning of 2008 was traced to an unvaccinated child who was infected with measles after traveling to Switzerland [5]. The child spread the infection to his classmates at the school. The majority of the infected children were also unvaccinated. This outbreak is an example of how reduced vaccination coverage and virus importation can lead to new measles outbreaks. Because measles is still endemic in Africa and Asia, and virus importation can start new outbreaks, rapid detection of measles remains an essential need in diagnosis of suspected cases.

Current diagnosis methods such as enzyme-linked immunoassay (ELISA) and immunofluorescent antibody assays are time-consuming [1]. Often, these assays require reverse transcriptase-polymerase chain reaction (RT-PCR) coupling to improve the detection limit [6]. However, previous studies with Fourier transform infrared (FTIR) spectroscopy showed that FTIR have two major advantages over immunoassays in pathogen detection [7, 8, 9, 10, 11, 12]. FTIR allows for rapid detection and requires minimal sample preparation.

Because of these advantages, FTIR has been the common spectroscopic technique for detection of pathological agents. Previous studies by the Naumann's group revealed that FTIR is an effective method for detection of different strains of bacteria and scrapie agents, such as transmissible spongiform encephalopathies (TSEs) [7, 8, 12]. In these studies, FTIR measurements yielded reproducible spectral results. The reproducibility of the IR results facilitated sample discrimination. For example, a study conducted by Naumann et al on TSEs demonstrated that FTIR can be coupled with chemometric analysis to discriminate different strains of TSEs [12]. Promising results from these studies suggested that FTIR can be applied in viral detection. Nonetheless, although FTIR has several advantages that favors its potential application in viral detection, FTIR lacks the signal enhancement that is provided by surface-enhanced Raman scattering (SERS) spectroscopy. Consequently, because of the high detection limit, optimal FTIR detection of virus samples requires that the samples are concentrated. Base on the knowledge of the advantages and disadvantages of FTIR, FTIR and chemometric analysis were applied in this study to identify different genotypes of measles virus.

Methods

Measles virus. The same virus and control samples provided by the Centers for Disease Control were used for the FTIR measurements. The media containing the SLAM-coding plasmid, ctrl 3 (SLAM), was used as the control in this experiment.

IR measurements. First, the CaF_2 disk was thoroughly cleaned with HPLC-grade/ 99.9% assay methanol. Next, the IR chamber was purged to remove water vapor. The IR spectra were collected using the Perkin Elmer Spectrum 2000 FT-IR spectrometer system (Beaconsfield Bucks, England). The background of the clean CaF_2 disk was collected over a range of 800 to 4000 cm^{-1} by co-adding 1024 scans of the individual spectrum. After measuring the background,

the cryogenic MV vials were thawed at room temperature prior to sample deposition. Then, 2 μl of the thawed sample was applied directly on the surface of the CaF_2 disk. The deposited sample was left to air-dry for 10 minutes. The IR spectra of the samples were also collected over 800 to 4000 cm^{-1} by co-adding 1024 scans.

Data analysis. Using Unscrambler version 9.6 (CAMO Technologies Inc.), the IR spectral data was preprocessed by taking the Savitzky-Golay first derivative and applying unit vector normalization. To classify the MV samples, the data was further preprocessed by autoscaling and then analyzed by Principal Components Analysis (PCA), Hierarchical Cluster Analysis (HCA) coupled with K-Nearest Neighbor (KNN) classification, and Partial Least Squares Discriminant Analysis (PLS-DA). The chemometric analysis was performed using the Partial Least Squares (PLS) Toolbox version 4.1 (Eigenvector Research Inc., Wenatchee, WA) in the MATLAB software version 7.2 (The Mathworks Inc., Natick, MA).

Results

IR spectra. IR measurements of the virus samples and ctrl 3 (SLAM) produced broad but reproducible IR bands (Fig. 3.1). The detected IR bands can be assigned to the viral membrane, protein, amino acid, and RNA components in the samples (Table 3.1) [7, 9, 11, 13]. For example, the IR band appearing at 2925 cm^{-1} was assigned to the CH stretch in the viral membrane. The spectral bands measured at 1660 and 1629 cm^{-1} were identified as α-helix and amide I protein bands, respectively [7, 9, 11]. The prominent band around 1356 cm^{-1} corresponded to the CO symmetric stretch of the COO^- backbone in amino acids [13]. The presence of RNA in the sample contributed to the P=O symmetric vibration detected at 1076 cm^{-1} [9].

Chemometric analysis. Because the broad IR bands overlap in the spectra, it was difficult to distinguish samples of different classes just by visual inspection of the data (Fig. 3.1). However, the high reproducibility of the IR spectra allowed for sample classification by chemometric analysis. In the first chemometric approach, to classify the measles samples, Principal Component Analysis (PCA) was performed on the data. PCA analysis of the IR data revealed A, D4, and H1 genotypes and ctrl 3 media were differentiable in all PC scores plots (Fig. 3.2). In particular, PC1 vs PC2 scores plot indicated that 99.3 % of the total variation was not useful for identifying samples from the D9 class. However, PC1 versus PC2 and PC2 versus PC4 scores plots confirmed that D9 samples could be classified using 1.9 % of the total variation (Fig. 3.3 – 3.4). Thus, the latter scores plots suggested that the 4 measles genotypes were differentiable by PCA.

The second chemometric method for measles identification was Hierarchical Cluster Analysis (HCA) coupled with K-Nearest Neighbor (KNN) classification. The HCA dendogram correctly clustered all the samples of the virus and media control (Fig. 3.5). Therefore, the total classification rate was 100 %. The high classification rate proved that HCA successfully discriminated samples of different measles genotypes. Additionally, HCA also correctly distinguish the virus samples from the media control.

In the third chemometric approach, the measles samples were differentiated by Partial Least Squares Discriminant Analysis (PLS-DA). The LV1 versus LV2 scores plot showed that PLS-DA correctly grouped samples from the A, D4,D9, H1, and ctrl 3 classes (Fig. 3.6). The well-defined clusters on the scores plot suggested that 70.9 % of the total variation was sufficient for separation of samples by PLS-DA. In addition to finding the intergroup variability, PLS-DA used the detected variability to predict the class for each sample. The resulting y-predicted plots

for the calibrated model showed that PLS-DA assigned all the samples to the correct classes (Fig. 3.7 – 3.11). Therefore, the sensitivity and specificity of the calibrated model were 100 %. The calibrated model was further tested by cross-validation (CV) with the Venetian blinds method. After cross-validation, the CV y-predicted plots indicated that all the samples remained in the right class (Fig. 3.12 – 3.16). Thus, the sensitivity and specificity of the cross-validated model were also 100 % (Table 3.3).

Discussions

Similar to viral detection with SERS, viral detection with FTIR was also achieved with minimal sample preparation. However, because of the high detection limit, to improve the detected signal, a higher sample concentration is required for the IR measurements. For example, 1 μL of a 10x-dilution was sufficient for the SERS measurements. 1 μL of a 10x-dilution was equivalent to 100 pfu/μL. In contrast, for the IR measurements, 2 μL of the undiluted sample was applied to a CaF_2 disk. A 2 μL-aliquot of a sample with 10^6 pfu/ml concentration contained approximately 2000 pfu/μL. Thus, IR viral detection was achieved at 20 times the sample concentration applied in SERS measurement.

IR detection of the concentrated sample produced very reproducible spectra. The prominent bands located within 2925 – 1076 cm^{-1} could be assigned to the membrane, proteins, amino acids, and RNA components in the sample. The appearance of these bands showed that FTIR can detect different chemical composition in the virus samples. However, unlike the Raman shifts observed in the SERS spectra, the IR bands were not sharp or easily distinguishable. Instead, the bands were broad and overlapping. Because the IR bands overlapped, it was difficult to analyze the spectral variations between the virus samples and the

control by visual inspection. Thus, to identify the samples of different measles genotypes, chemometric analysis was performed on the IR data.

The first classification method was PCA. PCA detected the maximum total variation in the data [14]. If the intergroup variability were greater than the intragroup variability, the intergroup variability would have a larger contribution to the total variation in the data [15]. Thus, PCA correctly separated the samples for the classes with large intergroup variability. The resulting PC scores plot indicated that PC1 and PC2 described 99.3 % of the total variation. In the scores plot, separate clusters were displayed for the A, D4, H1, and ctrl 3 classes. This means the intergroup variability of these 4 classes contributed significantly to 99.3% of the total variation. However, unlike the samples from the A, D4, H1, and ctrl 3 classes, the D9 samples did not cluster in the PC1 versus PC2 scores plot. This means the 99.3 % variation captured by PC1 and PC2 did not fully describe the variable contribution from the D9 class. Instead, D9 variable contribution was actually described in the the 1.9 % variation captured by PC2, PC3, and PC4. As a result, PC2 versus PC3 and PC2 versus PC4 scores plot presented a separate cluster for D9 samples. Thus, the latter scores plots indicated that measles samples were differentiable by PCA.

The second classification method was HCA, coupled with KNN grouping. The dendogram created from the HCA analysis showed that HCA correctly identified the samples from the A, D4, D9, H1, and ctrl 3 classes. In addition to identifying the samples, HCA also determined that ctrl 3 was the most statistically different class in the data set. Thus, HCA correctly distinguished the media control from the virus samples. The resulting dendogram confirmed that the measles samples were identified with 100 % accuracy by HCA and KNN classification.

The third classification method was PLS-DA. The LV scores plot produced in PLS-DA provided an excellent example of why PLS-DA is a more effective discrimination method than PCA. Rather than focusing on the total variability in the data, PLS-DA focuses on the variability between the classes [15]. Therefore, even though the intergroup variability of the D9 class was small, compared to that of A, D4, H1, and ctrl 3 classes, PLS-DA was still able to separate D9 from the remaining classes. As a result, LV1 versus LV2 scores plot showed that the 70.9 % variation captured by LV1 and LV2 was sufficient for grouping samples of A, D4, D9, H1, and ctrl 3 classes.

In addition to showing how the samples clustered according to the detected intergroup variability, PLS-DA also provided a calibrated model for predicting the class of each sample. The y-predicted plots generated from the calibrated model indicated that PLS-DA correctly predicted the class for all the samples. Furthermore, the prediction efficiency of the calibrated model was tested with Venetian blinds cross-validation (CV). Because 5 data splits were chosen for the cross-validation, the samples were divided into 5 groups. In the first data split, the first group of samples was removed, and a new model was calculated using the remaining samples in the data as the training set. Then, the samples that were withheld from the model calculation were used to create an unknown set for the cross-validated model. The procedure was repeated for the remaining samples. After cross-validation, the CV y-predicted plots showed that the PLS-DA model effectively classified samples from both the training and the unknown sets. According to the CV y-predicted plots, all the samples remained in the correct classes after cross-validation. Thus, the CV model predicted the class with 100 % sensitivity and specificity.

In summary, the successful differentiation of the measles samples proved that FTIR can be combined with chemometric analysis to identify measles in cell culture media. Like SERS,

FTIR allowed for rapid detection with minimal sample preparation. However, because of the high detection limit, virus detection by FTIR was achieved at a higher sample concentration. Thus, SERS is the more preferable spectroscopic technique for low level viral detection. Additionally, because of the overlapping bands, spectral variations between the virus samples and the control were not very noticeable. Nonetheless, chemometric analysis of the IR data provided an effective method for identifying measles samples of different genotypes. Results from the chemometric analysis showed that measles samples were differentiable by PCA, HCA, and PLS-DA.

Acknowledgements

The research was funded by the U.S. Army Research Laboratory (W911NF-07-2-0065) grant. The author thanks Ralph Tripp, Paul Rota, and the Centers for Disease Control (CDC) for supplying the measles samples and Yu Zhu for assistance with the FTIR measurements.

References

1. Schneider-Schaulies, S.; Meulen, V. t.; Allan, G.; Robert, G. W. *Encyclopedia of Virology*; Elsevier: Oxford, 1999; pp 952-960.
2. Centers for Disease Control and Prevention. Epidemiology and Prevention of Vaccine Preventable Diseases. Atkinson W, Hamborsky J, McIntyre L, Wolfe S, eds. 10th ed. Washington DC: Public Health Foundation, 2007.
3. World Health Organization, Progress in global measles control and mortality reduction, 2000 – 2006. *Weekly Epidemiological Record* **2007**, 82, 418-424.
4. Rima, B. K.; Earle, J. A. P.; Baczko, K.; terMeulen, V.; Liebert, U. G.; Carstens, C.; Carabana, J.; Caballero, M.; Celma, M. L.; FernandezMunoz, R., Sequence divergence of measles virus haemagglutinin during natural evolution and adaptation to cell culture. *Journal of General Virology* **1997,** 78, 97-106.
5. Centers for Disease Control (CDC), Outbreak of Measles --- San Diego, California, January—February 2008. *Morbidity and Mortality Weekly Report* **2008**, 57, 1-4.
6. Krause, C. H.; Molyneaux, P. J.; Ho-Yen, D. O.; McIntyre, P.; Carman, W. F.; Templeton, K. E. Comparison of mumps-IgM ELISAs in acute infection. *Journal of Clinical Virology* **2007,** 38, 153-156.
7. Helm, D.; Labischinski, H.; Naumann, D., Elaboration of a procedure for identification of bacteria using Fourier-transform IR spectral libraries – a stepwise correlation approach. *Journal of Microbiological Methods* **1991,** 14, 127-142.
8. vanderMei, H. C.; Naumann, D.; Busscher, H. J., Grouping of Streptococcus mitis strains grown on different growth media by FT-IR. *Infrared Physics & Technology* **1996,** 37, 561-564.

9. Kansiz, M.; Heraud, P.; Wood, B.; Burden, F.; Beardall, J.; McNaughton, D., Fourier Transform Infrared microspectroscopy and chemometrics as a tool for the discrimination of cyanobacterial strains. *Phytochemistry* **1999,** 52, 407-417.
10. Tomzig, A.; Spassov, S.; Friedrich, M.; Naumann, D.; Beekes, M., Discriminating scrapie and bovine spongiform encephalopathy isolates by infrared spectroscopy of pathological prion protein. *Journal of Biological Chemistry* **2004,** 279, 33847-33854.
11. Spassov, S.; Beekes, M.; Naumann, D., Structural differences between TSEs strains investigated by FT-IR spectroscopy. *Biochimica Et Biophysica Acta-General Subjects* **2006,** 1760, 1138-1149.
12. Lasch, P.; Beekes, M.; Schmitt, J.; Naumann, D., Detection of preclinical scrapie from serum by infrared spectroscopy and chemometrics. *Analytical and Bioanalytical Chemistry* **2007,** 387, 1791-1800.
13. J. Lambert, H. Shurvell, D. Lightner, and R. Cooks, "Group Frequencies: Infrared and Raman", in *Organic Structural Spectroscopy*, Prentice Hall, New Jersey, 1998, pp. 175-199.
14. Beebe, K.R., Pell, R.J., Seasholtz, M.B., *Chemometrics: A Practical Guide.* John Wiley and Sons: New York, 1998; p. 81-84.
15. Barker, M.; Rayens, W., Partial least squares for discrimination. *Journal of Chemometrics* **2003,** 17, 166-173.

Table 3.1 Band assignments for IR wavenumbers.

Wavenumbers (cm^{-1})	**Assignments**[a]
3301	OH stretch
2925	CH stretch in membrane amphiphiles
1660	α-helix
1629	amide I; β-sheet
1356	CH_3, CH_2 asymmetric deformation of proteins; C-O symmetric stretch of COO^-
1076	P=O symmetric stretch of phosphodiester backbone in RNA

[a] The IR bands were assigned to viral membrane, proteins, amino acids, and RNA components in the samples.

Table 3.2 Threshold values for the PLS-DA classification of IR data.

Class	**Y predicted**	**Threshold**[a]
ctrl 3	**1**	0.3801
A	**2**	0.7042
D4	**3**	0.5212
D9	**4**	0.5923
H1	**5**	0.6500

[a] A distinct threshold was determined for the y-predicted plots of ctrl 3 (SLAM), A, D4, D9, and H1 classes.

Table 3.3 The sensitivity and specificity values of the PLS-DA calibrated and cross-validated prediction models for IR data.

Class	**ctrl 3**	**A**	**D4**	**D9**	**H1**
Sensitivity[a] (Cal)	1.000	1.000	1.000	1.000	1.000
Specificity[b] (Cal)	1.000	1.000	1.000	1.000	1.000
Sensitivity (CV)	1.000	1.000	1.000	1.000	1.000
Specificity (CV)	1.000	1.000	1.000	1.000	1.000

[a] The sensitivity is defined as the number of predicted samples in a class divided by the actual number of samples in the class. [b] The specificity is given as the number of samples predicted not in the class divided by the actual number of samples not in the class. Because the all the samples were assigned to the correct class, the sensitivity and specificity were 100 % for both the calibrated (Cal) and cross-validated (CV) models.

Figure 3.1 Composition of average IR spectra for the A, D4, D9, H1, and ctrl 3 (SLAM) classes.

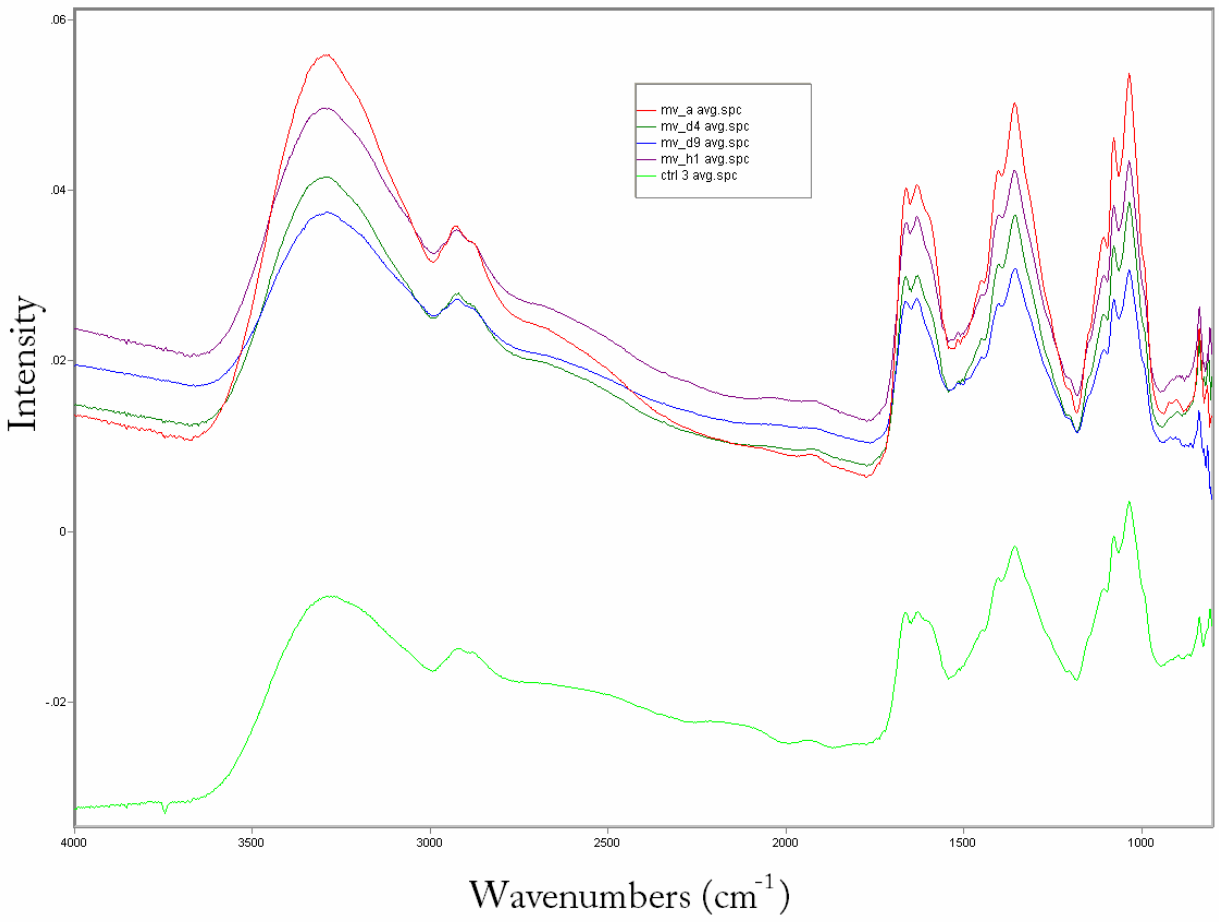

mv_a avg.spc
mv_d4 avg.spc
mv_d9 avg.spc
mv_h1 avg.spc
ctrl 3 avg.spc
.06
.04
.02
0
-.02
Intensity
4000
3500
3000
2500
2000
1500
1000
Wavenumbers (cm-1)

Figure 3.2 PC1 versus PC2 scores plot for the IR data. The scores plot was produced after Principal Component Analysis (PCA) was performed on the IR data. Separate clusters were displayed for the A, H1, D4, and ctrl 3 classes. However, because the small intergroup variability between D9 and the other classes was not detected in the 99.3 % variation captured by PC1 and PC2, D9 samples did not cluster in this scores plot.

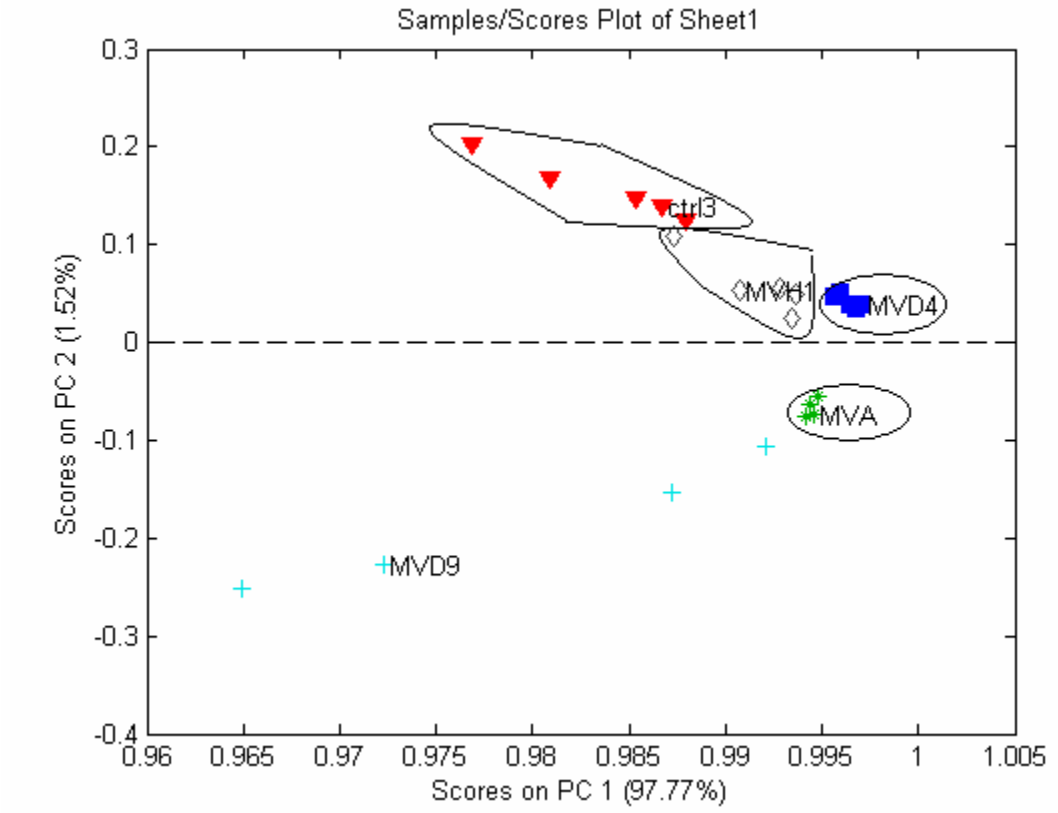

Samples/Scores Plot of Sheet1
Scores on PC 2 (1.52%)
Scores on PC 1 (97.77%)
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
0.96
0.965
0.97
0.975
0.98
0.985
0.99
0.995
1
1.005
ctrl3
MVD4
MVA
MVD9

Figure 3.3 PC2 versus PC3 scores plot for the IR data. The variable contribution from the D9 class was detected in the 1.9 % variation captured by PC2 and PC3. Thus, D9 samples grouped into one cluster in the PC2 versus PC3 scores plot.

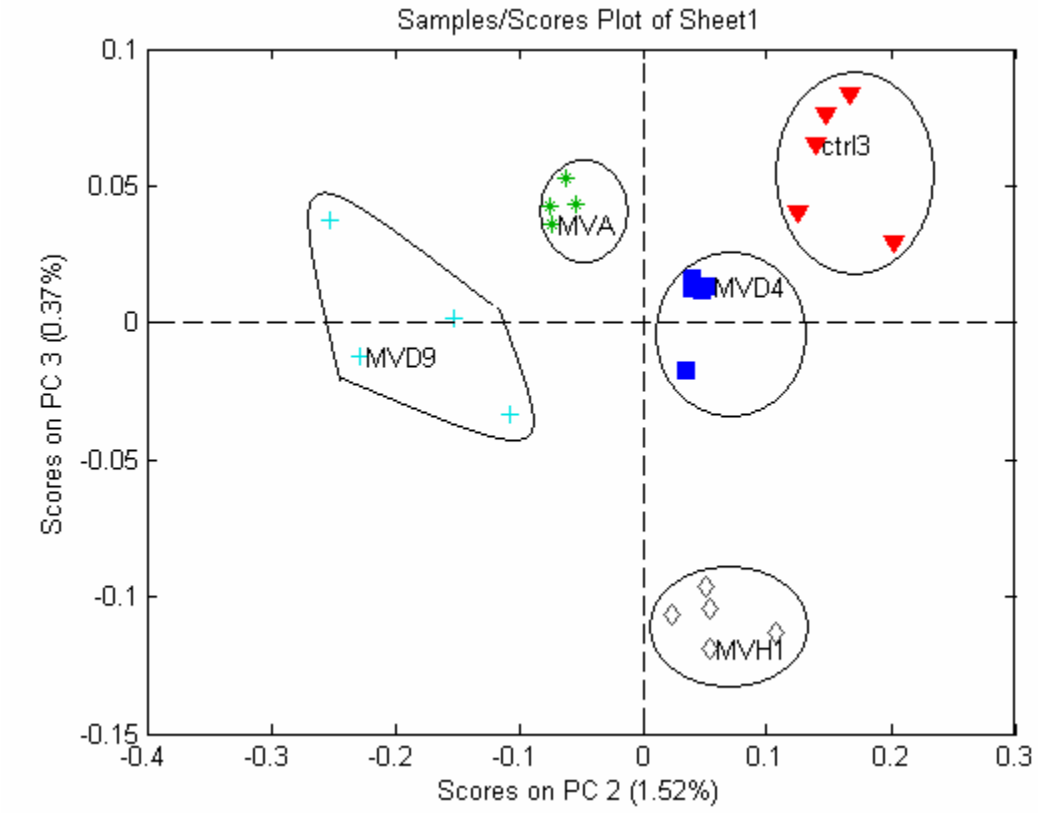

Samples/Scores Plot of Sheet1
Scores on PC 3 (0.37%)
Scores on PC 2 (1.52%)
0.1
0.05
0
-0.05
-0.1
-0.15
-0.4
-0.3
-0.2
-0.1
0
0.1
0.2
0.3
ctrl3
MVA
MVD4
MVD9
MVH1

Figure 3.4 PC2 versus PC4 scores plot for the IR data. The variable contribution from the D9 class was also detected in the 1.7 % variation captured by PC2 and PC4. Therefore, PC2 versus PC4 scores plot also presented a separate cluster for D9 samples.

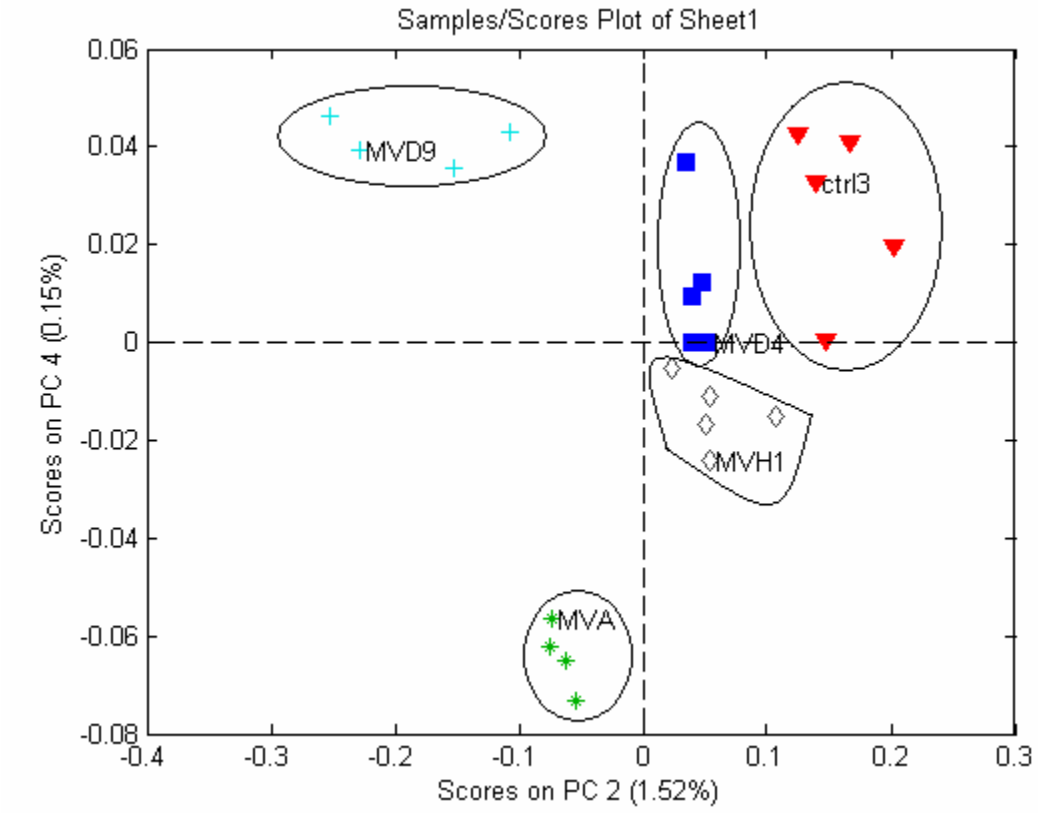

Samples/Scores Plot of Sheet1
Scores on PC 4 (0.15%)
Scores on PC 2 (1.52%)
0.06
0.04
0.02
0
-0.02
-0.04
-0.06
-0.08
-0.4
-0.3
-0.2
-0.1
0
0.1
0.2
0.3
MVD9
ctrl3
MVD4
MVH1
MVA

Figure 3.5 Dendogram produced by Hierachical Cluster Analysis (HCA) and K-Nearest Neighbor (KNN) classification of the IR data. The dendogram showed that HCA classified the samples from the A, D4, D9, H1, and ctrl 3 (SLAM) classes with 100 % accuracy. Additionally, HCA also correctly determined that ctrl 3 (SLAM) media was the most statistically different class in the data set.

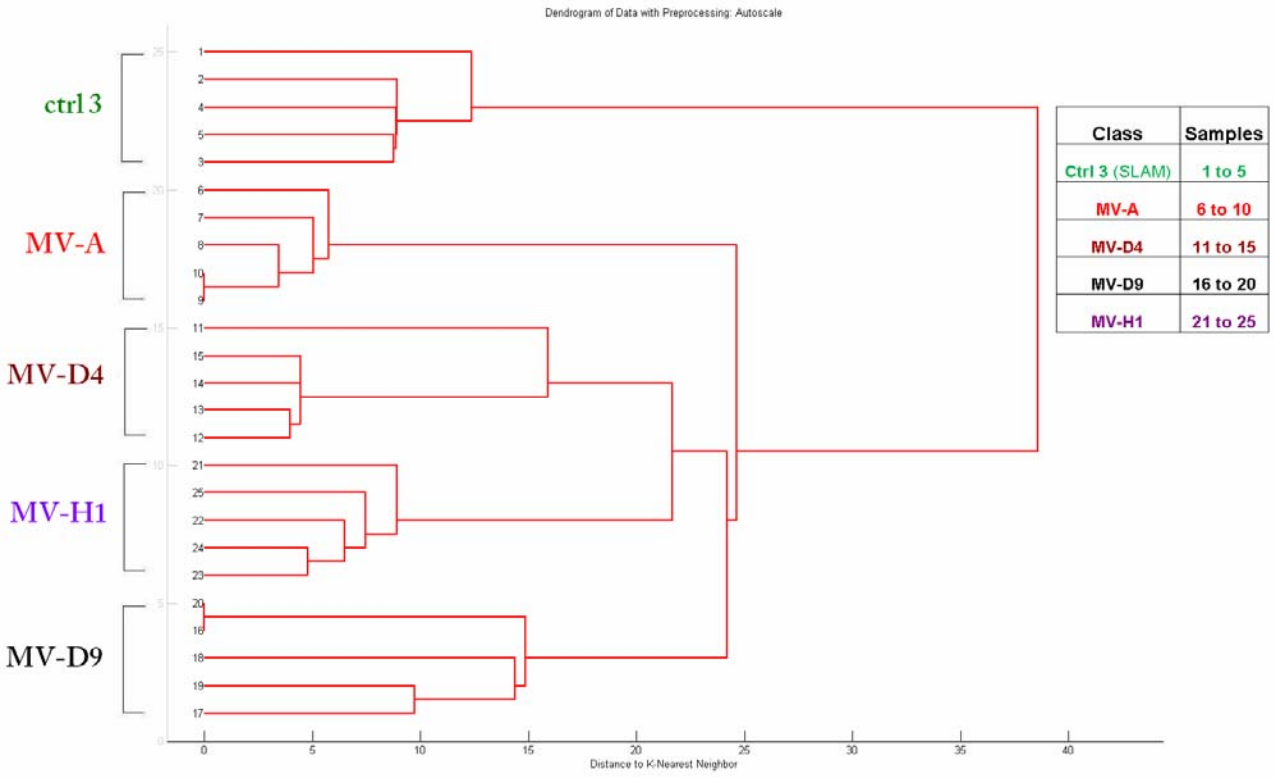

Dendrogram of Data with Preprocessing: Autoscale
ctrl 3
MV-A
MV-D4
MV-H1
MV-D9
1
2
4
5
3
6
7
8
10
9
11
15
14
13
12
21
25
22
24
23
20
16
18
19
17
0
5
10
15
20
25
30
35
40
Distance to K-Nearest Neighbor
Class
Samples
Ctrl 3 (SLAM)
1 to 5
MV-A
6 to 10
MV-D4
11 to 15
MV-D9
16 to 20
MV-H1
21 to 25

Figure 3.6 LV1 versus LV2 scores plot for the IR data. Unlike PCA, PLS-DA separated the samples by detecting the intergroup variability. Thus, LV scores plot showed that PLS-DA grouped all the samples, including those from the D9 class, into well-defined clusters for each sample class.

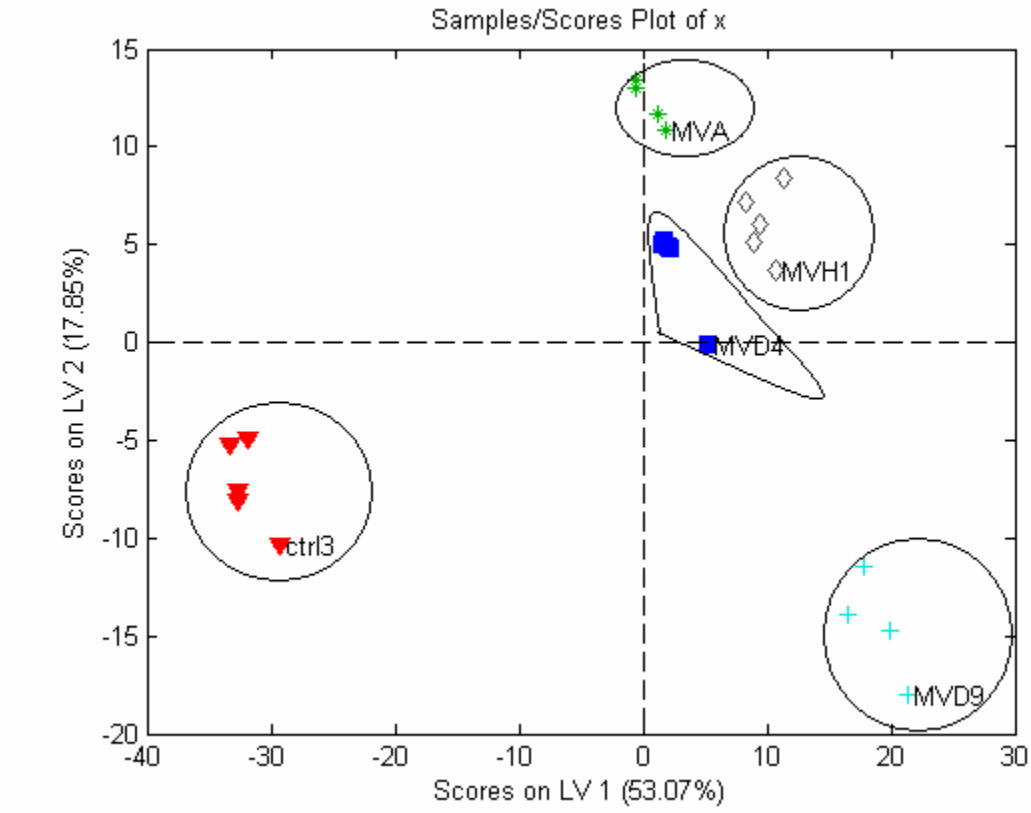

Samples/Scores Plot of x
Scores on LV 2 (17.85%)
Scores on LV 1 (53.07%)
MVA
MVH1
MVD4
ctrl3
MVD9
15
10
5
0
-5
-10
-15
-20
-40
-30
-20
-10
0
10
20
30

Figure 3.7 Y predicted plot for the ctrl 3 (SLAM) class prediction of IR data. The red dash line represented the threshold for the prediction model. The threshold value defined the cut-off point for the class assignment. Samples that belonged to the ctrl 3 (SLAM) class had a calculated y-predicted value that falls between the threshold (0.3801) and a value that is appromixately close to 1. Samples with y-predicted values below the threshold were identified as not belonging to the ctrl 3 (SLAM) class. According to the y predicted plot, all the samples from the ctrl 3 (SLAM) class were correctly identified by PLS-DA.

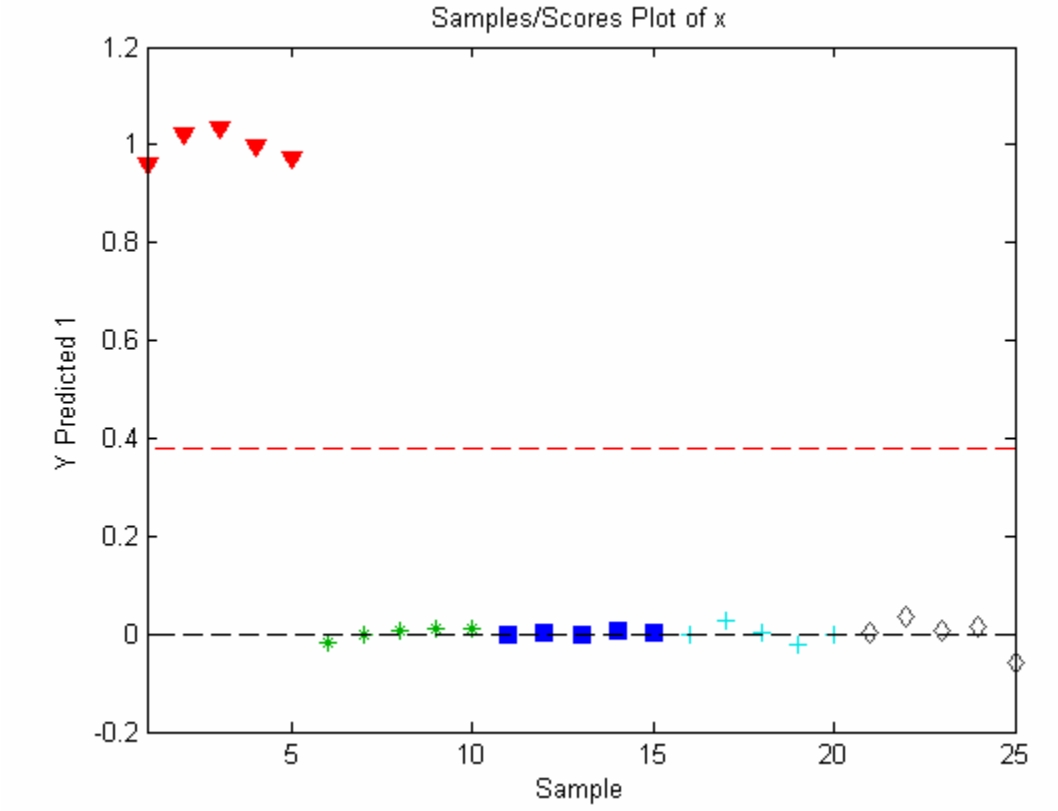

Samples/Scores Plot of x
Y Predicted 1
Sample
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
5
10
15
20
25

Figure 3.8 Y predicted plot for the A class prediction of IR data. Samples with calculated y-predicted values above the 0.7042 threshold were assigned to the A class. In the model, all the A samples were correctly assigned to the A class.

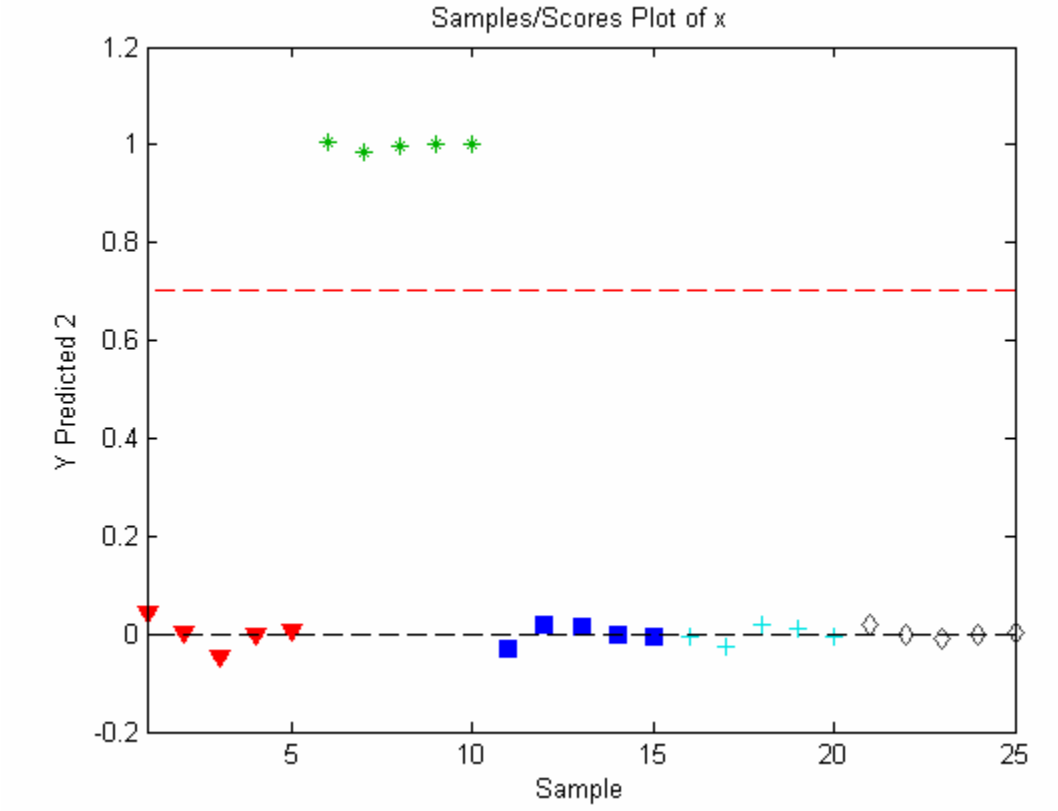

Samples/Scores Plot of x
Y Predicted 2
Sample
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
5
10
15
20
25

Figure 3.9 Y predicted plot for the D4 class prediction of IR data. Samples with calculated y-predicted values above the 0.5212 threshold were assigned to the D4 class. In the model, all the D4 samples were correctly assigned to the D4 class.

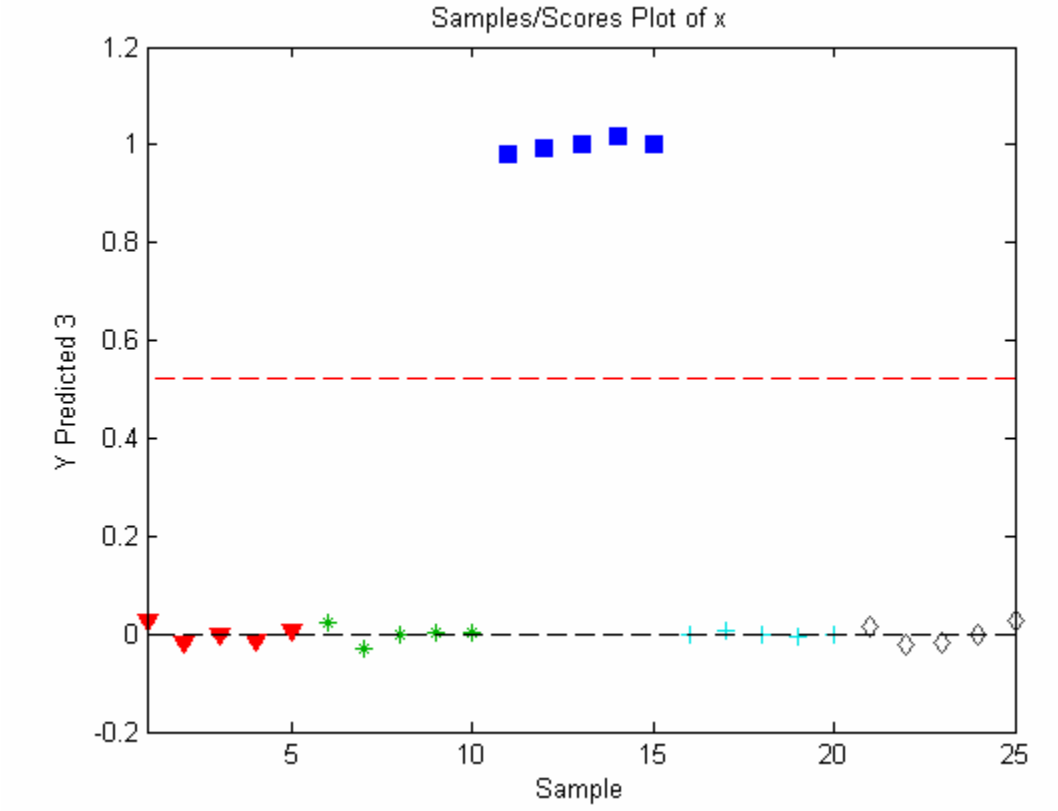

Samples/Scores Plot of x
Y Predicted 3
Sample
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
5
10
15
20
25

Figure 3.10 Y predicted plot for the D9 class prediction of IR data. Samples with calculated y-predicted values above the 0.5923 threshold were assigned to the D9 class. In the model, all the D9 samples were correctly assigned to the D9 class.

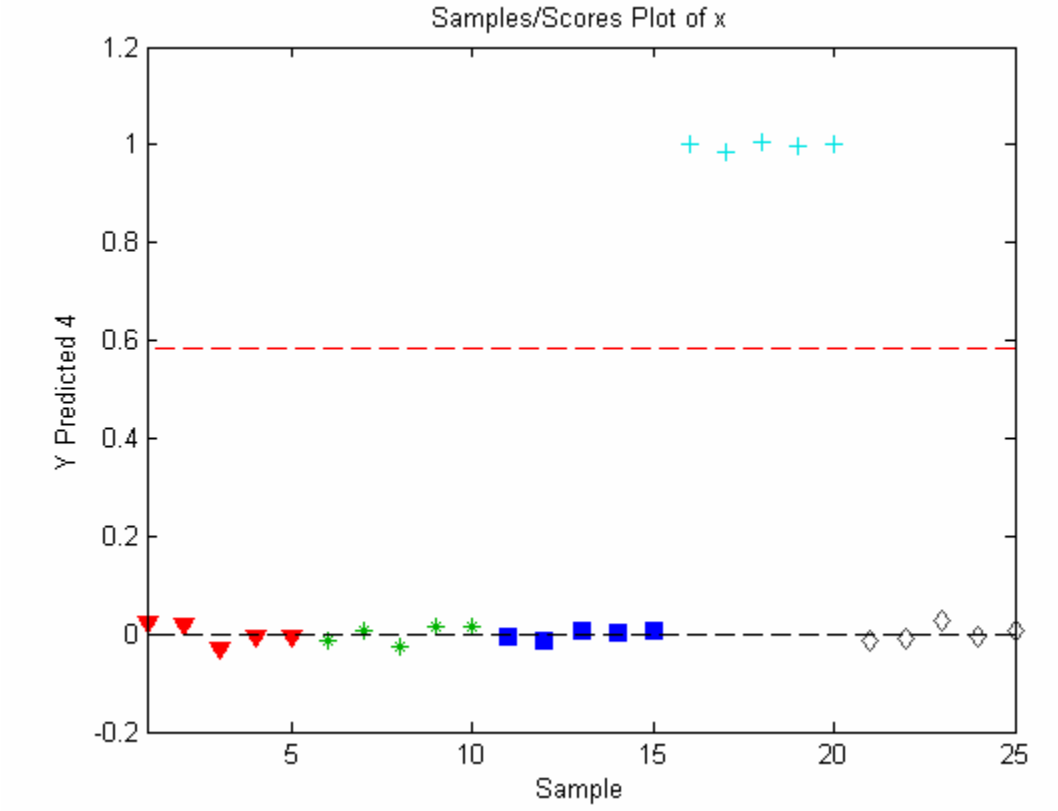

Samples/Scores Plot of x
Y Predicted 4
Sample
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
5
10
15
20
25

Figure 3.11 Y predicted plot for the H1 class prediction of IR data. Samples with calculated y-predicted values above the 0.6500 threshold were assigned to the H1 class. In the model, all the H1 samples were correctly assigned to the H1 class.

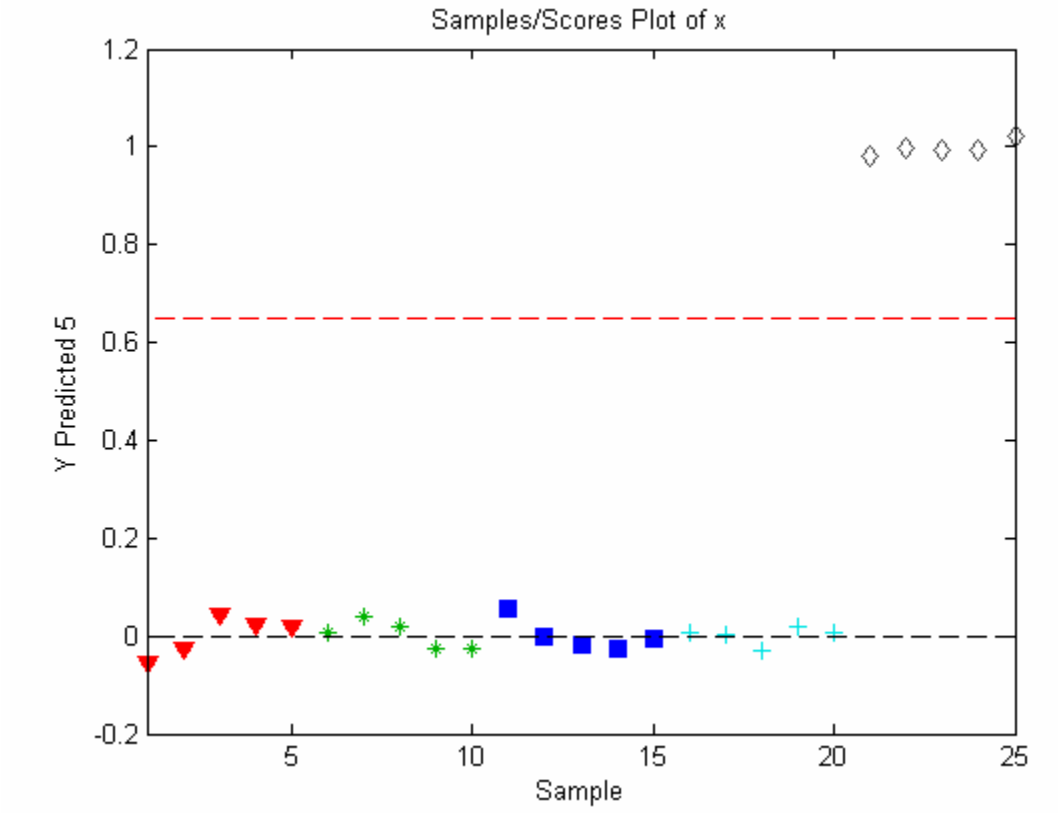

Samples/Scores Plot of x
Y Predicted 5
Sample
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
5
10
15
20
25

Figure 3.12 Y CV predicted plot for the ctrl 3 (SLAM) class prediction of IR data after cross-validation (CV). After cross-validation with the Venetian blinds method, the PLS-DA model correctly predicted the samples from the ctrl 3 (SLAM) class.

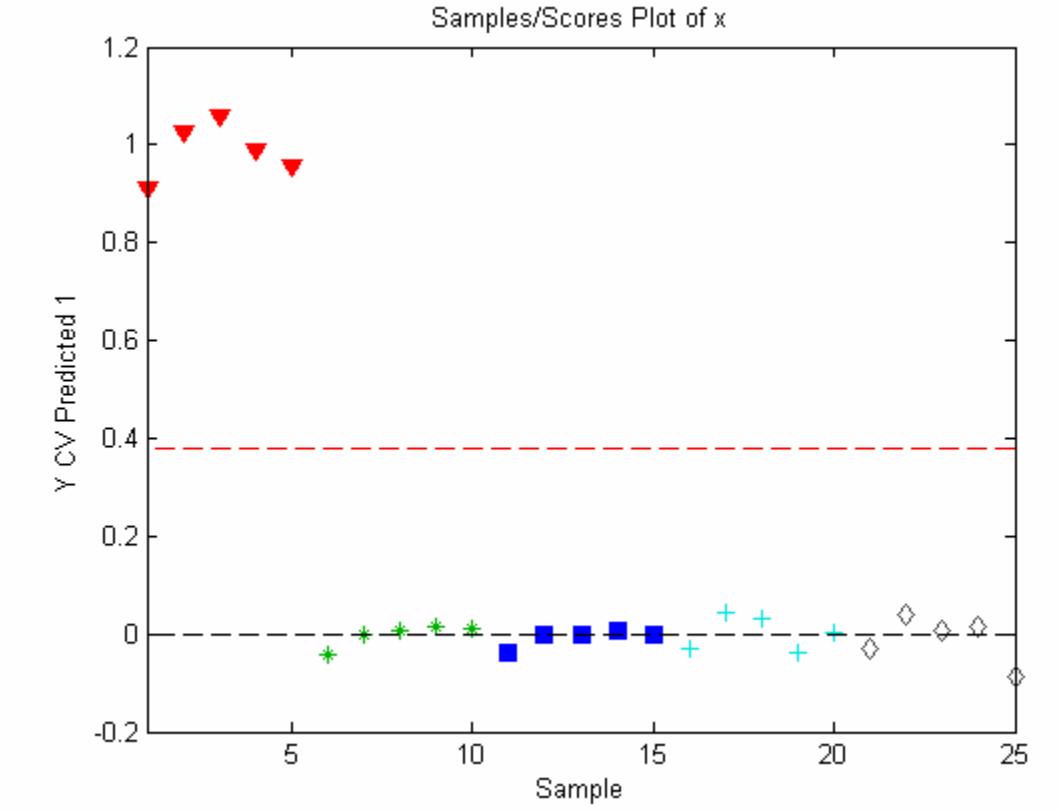

Samples/Scores Plot of x
Y CV Predicted 1
Sample
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
5
10
15
20
25

Figure 3.13 Y CV predicted plot for the A class prediction of IR data after cross-validation (CV). The plot indicated that the A samples remained in the correct class after cross-validation.

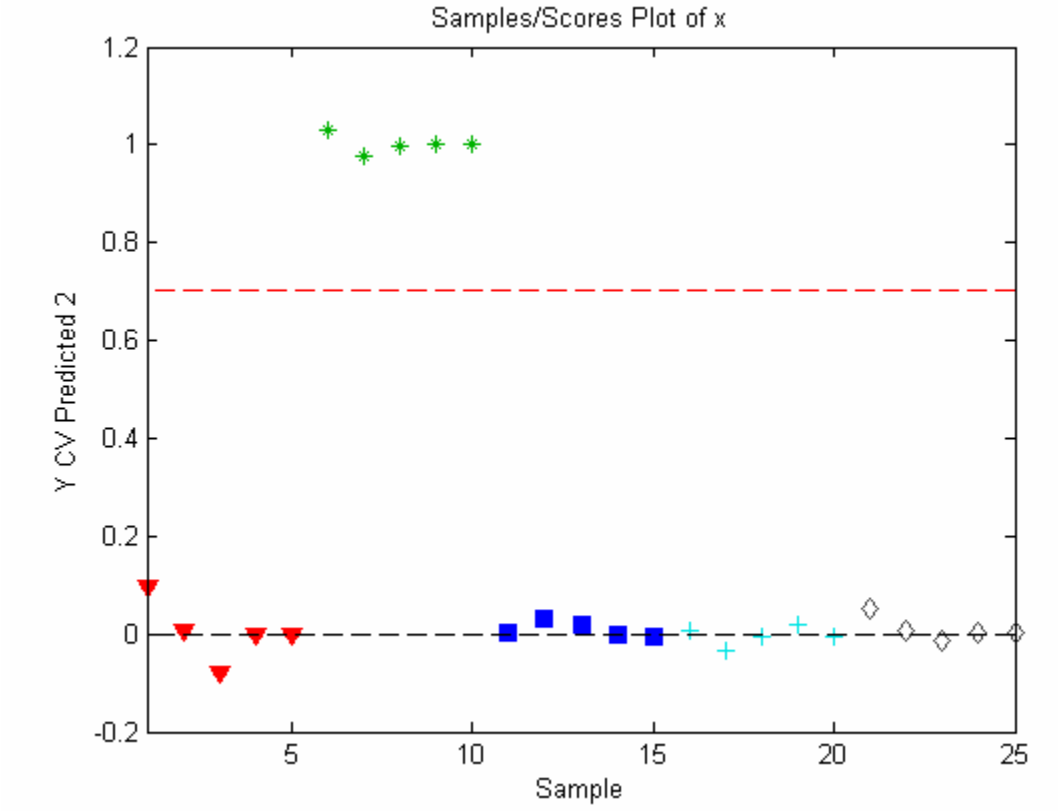

Samples/Scores Plot of x
Y CV Predicted 2
Sample
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
5
10
15
20
25

Figure 3.14 Y CV predicted plot for the D4 class prediction of IR data after cross-validation (CV). The plot indicated that the D4 samples remained in the correct class after cross-validation.

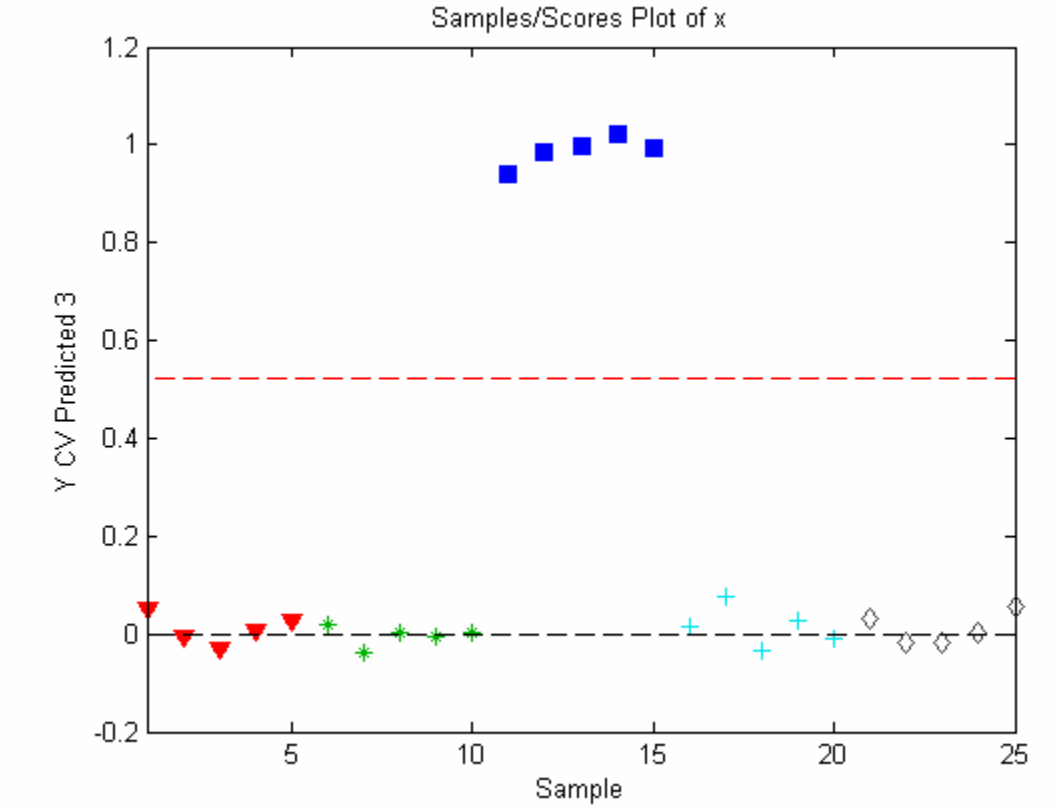

Samples/Scores Plot of x
Y CV Predicted 3
Sample
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
5
10
15
20
25

Figure 3.15 Y CV predicted plot for the D9 class prediction of IR data after cross-validation (CV). The plot indicated that the D9 samples remained in the correct class after cross-validation.

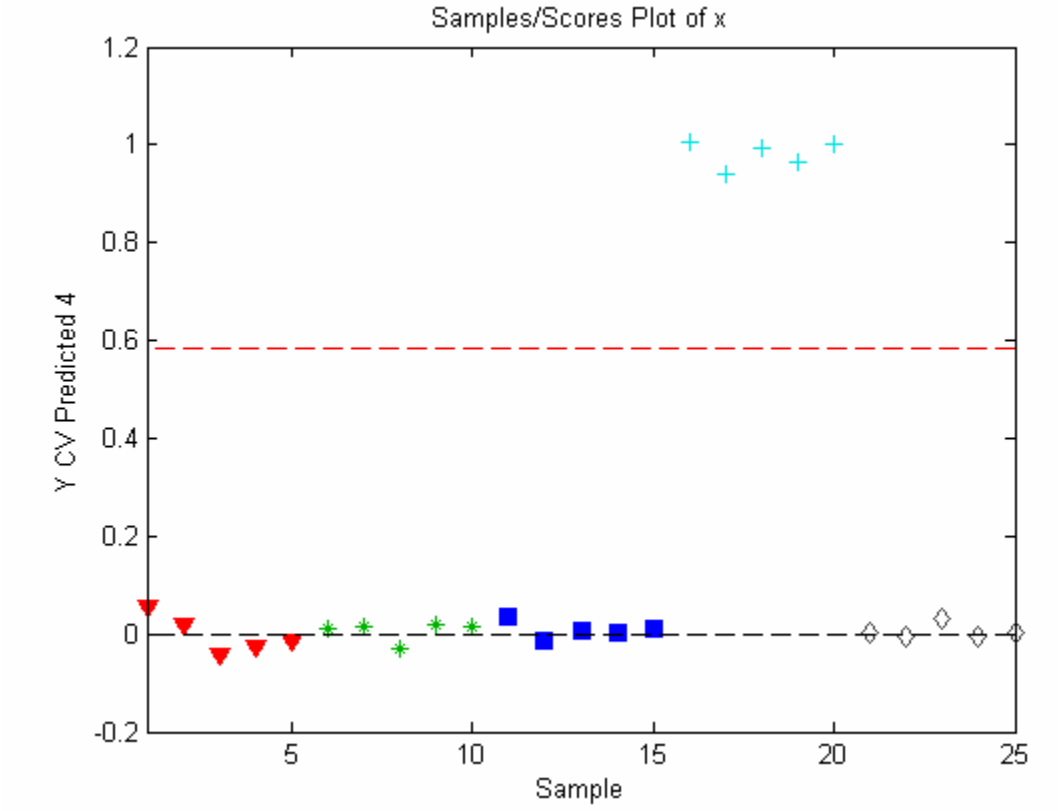
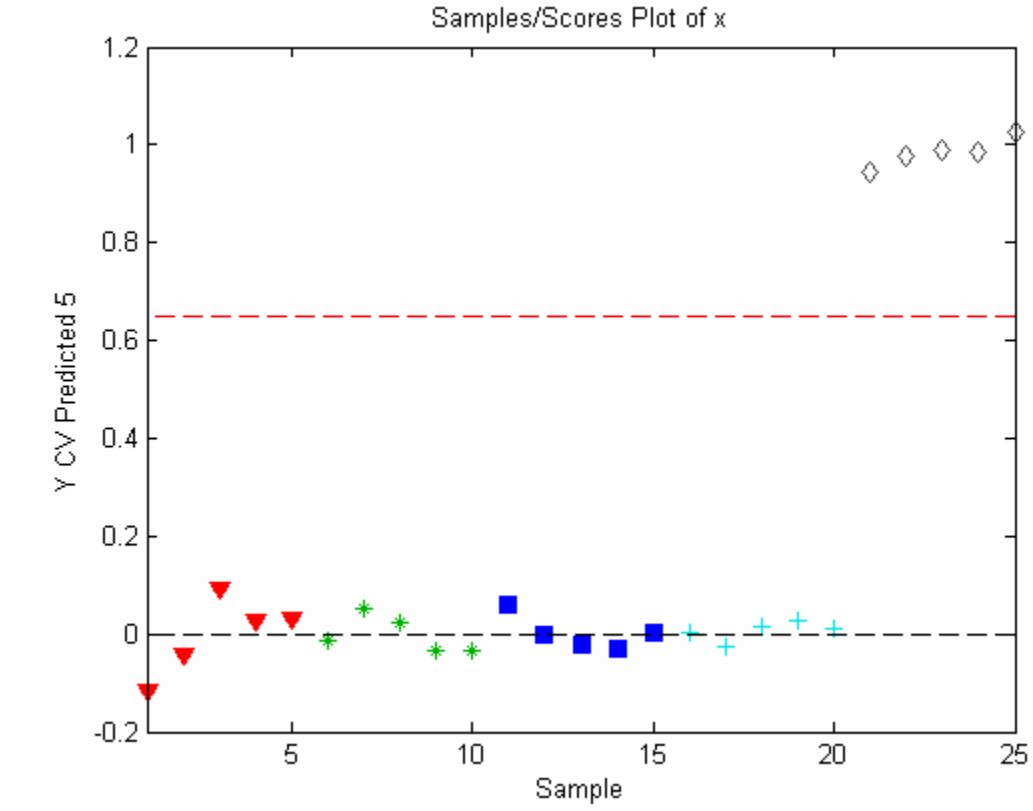

Samples/Scores Plot of x
Y CV Predicted 4
Sample
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
5
10
15
20
25

Figure 3.16 Y CV predicted plot for the H1 class prediction of IR data after cross-validation (CV). The plot indicated that the H1 samples remained in the correct class after cross-validation.


Samples/Scores Plot of x
Y CV Predicted 5
Sample
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
5
10
15
20
25

Chapter IV

Conclusions and Future Work:

Measles is known as one of the leading causes of childhood deaths [2]. According to the CDC, in 2006, measles caused approximately 242,000 deaths worldwide [1]. In countries where measles is no longer endemic, new outbreaks can occur from virus importation and reduced vaccination coverage [4]. Thus, there is still a need for rapid detection of measles to facilitate diagnosis of suspected cases and to improve the global measles surveillance.

Currently, the common diagnostic methods for measles detection are IgM and IgG enzyme-linked immunoassay (ELISA) tests [2]. However, IgM tests are not effective for early phase diagnosis because of the low sensitivity [19]. IgM tests that are conducted within 72 hours of the rash onset can produce up to 20 % false negatives. Alternatively, IgG tests are also used to diagnose measles. However, IgG assay time is dependent on the availability of the specimens. Thus, IgG test results cannot be obtained until the second specimen is collected 10 – 30 days after exposure [1]. In addition to ELISA tests, immunofluorescent antibody assays are also common measles detection techniques. However, because of the high detection limit, these assays require reverse transcriptase-polymerase chain reaction (RT-PCR) amplification and other necessary sample preparation [20]. In general, the current ELISA tests and immunofluorescent antibody assays available for measles detection are time-consuming and have low sensitivity.

Unlike the previously described immunoassays, the great signal enhancement in surface-enhanced Raman scattering (SERS) spectroscopy significantly improves the detection limit of the technique and favors its application in low level viral detection [33, 40]. Because the measured Raman shifts were characteristic of the chemical composition of the samples, SERS provided an effective method for vibrational analysis of the virus samples. In this study, the

spectral results proved that SERS can detect different components of the measles virus and the cell culture media. For example, characteristic Raman shifts were assigned to proteins, amino acids, and RNA in the samples.

On the SERS spectra, reproducible spectral variations were observed for the virus samples and the media control. These variations showed that the sensitivity provided by the AgNR substrates enabled SERS detection of virus signal in a complex sample mixture. Furthermore, even though the percentage of sequence diversity was only 4.2 – 12.3 % for A, D4, D9, and H1 genotypes, the high sensitivity of the substrates allowed for detection of measles from these different genotypes [6, 16, 17, 18]. Because of the signal enhancement, SERS measles detection was achieved at a low sample concentration.

The reproducible variations in the spectral data facilitated sample discrimination by chemometric analysis. The results from the chemotric analysis indicated that PCA and HCA can differentiate measles samples from the media controls and also distinguish A and H1 from D4 and D9 samples. However, PLS-DA can discriminate samples of all 4 genotypes. Even though the intergroup variability was low for D4 and D9 classes, PLS-DA correctly classified the samples from these 2 classes. The accurate classification by PCA, HCA, and PLS-DA showed that SERS spectroscopy can be combined with chemometric analysis to identify different genotypes of measles in cell culture media.

The second method for measles detection was FTIR spectroscopy. In the past, FTIR has been the common spectroscopic method for detection of pathological agents [52, 53, 54, 55]. Thus, FTIR was applied in this study to prove that measles detection can be achieved with FTIR spectroscopy. Similar to SERS, FTIR also provided rapid viral detection without cumbersome sample preparation. However, because of the high detection limit, FTIR measurements were

performed on concentrated viral samples. The spectral results showed that FTIR can detect protein, amino acids, viral RNA, and membrane components in the samples. However, the broad and overlapping bands made it difficult to visually separate the samples for different classes. Nonetheless, even though the variations between the virus and the control spectra were not very observable, PCA, HCA, and PLS-DA were able to differentiate all the samples accurately. Because of the data reproducibility, HCA and PLS-DA classification of the IR data were performed with 100 % accuracy.

In conclusion, the results from the SERS and FTIR studies proved that SERS and FTIR can be potential spectroscopic techniques for measles detection. In particular, because of its high sensitivity, SERS is the more favorable method for measles detection at low viral concentration. In the SERS study, the virus signal was detected in a complex and diluted sample mixture. The detection of measles virus at those conditions demonstrated the ability of SERS AgNR substrates to provide the sensitivity essential for low level viral detection. The successful differentiation of the measles samples in both SERS and FTIR studies proved the combination of these spectroscopic techniques with chemometric analysis provides a powerful tool for measles identification. For future works, SERS will be applied for detection of measles samples from all of the current genotypes. The SERS spectra will be used to create a spectral database for different measles genotypes. The spectra from the database will serve as the training set for PLS-DA classification of measles samples.

References for **Conclusions and Future Work** were provided on pages 66 – 73.