LABEL-FREE SERS DETECTION OF PROTEINS VIA CARBOHYDRATE

CONJUGATED NANOPARTICLES

by

SHARON JANILLE HURLEY MARTIN

(Under the Direction of Richard Dluhy)

ABSTRACT

Carbohydrates may be found inside or on the surface of cells. Understanding the structure and role carbohydrates play in regulating these processes is of great importance. This can be done by the development of carbohydrate detection methods. Surface enhanced Raman spectroscopy (SERS) can be used as a label free biosensing technique which yields a chemically sensitive signature of the analyte molecule. Carbohydrates were immobilized by copper (Cu) free click chemistry reactions to develop carbohydrate nanoparticles for detection and differentiation of proteins by SERS. Galectin proteins were detected with 95-100% sensitivity and specificity. Two galectin proteins were differentiated by SERS with 85-100% sensitivity and specificity. Then we demonstrate that sialylated carbohydrates can be immobilized to SERS substrates using Cu free click chemistry reactions. This created a sialylated carbohydrate sensing surface that was used to detect influenza hemagglutinin. Using SERS we were also able to detect H1N1 and H7N9 hemagglutinin binding to sialylated carbohydrates and differentiate H7N9 and H1N1 hemagglutinin. These findings demonstrate that SERS can be used as a label free method to detect biomolecular interactions.

INDEX WORDS: Carbohydrate arrays, surface-enhanced Raman spectroscopy
(SERS), click chemistry, strain promoted azide alkyne click
chemistry, chemometrics, galectin proteins, influenza
hemagglutinin

LABEL-FREE SERS DETECTION OF PROTEINS VIA CARBOHYDRATE

CONJUGATED NANOPARTICLES

by

SHARON JANILLE HURLEY MARTIN

B.S., Jackson State University, 2005

M.Ed., University of Arkansas, 2007

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial

Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2014

LABEL-FREE SERS DETECTION OF PROTEINS VIA CARBOHYDRATE

CONJUGATED NANOPARTICLES

by

SHARON JANILLE HURLEY MARTIN

Major Professor:	Richard Dluhy
Committee:	Geert-Jan Boons
	Shanta Dhar

Electronic Version Approved:

Julie Coffield
Interim Dean of the Graduate School
The University of Georgia
December 2014

DEDICATION

I give thanks to God Almighty for bringing me to this point and look forward to the next chapter in my life. Thank you my husband, Travis Martin, for your support through this process. Thanks to my parents, John and Emma Hurley, for implanting in me the importance of education and the tools to achieve it. This body of work is dedicated to the persons who were here when I started my journey and have since transitioned to the great beyond. My maternal grandmother, Margaret Jackson, my brother, Frederick Hurley, and my Father-in-law, Robert Martin, Sr.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

REVIEW OF THE LITERATURE

Carbohydrate Microarrays

In biology the major classes of macromolecules consist of DNA, proteins, lipids, and carbohydrates. These macromolecules govern many important biological processes and understanding the roles they play may lead to discoveries in cancer treatment, drug delivery systems, and disease screenings. DNA is important to living cells because it governs processes cells need to live. Proteins regulate many vital functions such as cell shape and participate in reactions with enzymes.[1] Carbohydrates participate in signaling and cell death. Carbohydrates are also known as glycans and can be further broken down to include monosaccharides, oligosaccharides, and polysaccharides. Monosaccharides are smaller carbohydrates and oligosaccharides and polysaccharides are polymeric saccharides which have been modified with proteins or lipids.

The study of carbohydrates in biology has been coined glycobiology and this field is dedicated to the study of glycoproteins, glycolipids, proteoglycans, and free oligosaccharides. In its early beginnings, carbohydrate research was limited because of structure complexity and difficulty in determining carbohydrate sequences. Some important discoveries in this field date back to the late 1800s to early 1900s.[2] This includes the discovery of glycosphingolipids, lectins and hemagglutinins, stereoisomeric structure of glucose and other monosaccharides, structure of ribose in RNA, sialic acids, and carbohydrate determinants of ABO blood group types.[2] Now, this is a rapidly

growing field which expands into the areas of biomedicine, biotechnology, chemistry of carbohydrates, glycan formation and degradation, and recognition of glycans by proteins.

Carbohydrates differ from other macromolecules in that they are branched molecules and are connected to each other by many different types of linkages.[3] This feature allows them to be attached to many different molecules such as proteins and lipids a process called glycosylation. Cells are surrounded by a layer of carbohydrates whose structures are determined by the cell type expression and sugar modifying enzymes.[4] Glycosylation is linked to protein folding, cell targeting, cell differentiation, and immune response. When attached to proteins or lipids, glycans are called glycoproteins or glycolipids. Carbohydrates attach proteins through nitrogen or oxygen atoms. When carbohydrates attach to proteins, they alter the structure and function of proteins.[5] On cell surfaces glycoproteins are presented as mixtures where the protein backbone are similar but the position of glycans are different.[6]

Research in this area has primarily focused on dietary sugars and their metabolic pathways however there is a desire to expand into other areas. Not much has been done or known about the functions of polysaccharides and oligosaccharides and their importance.[7] Glycosylation, which is an important form of post translation modification of proteins, has led to the characterization of an increasing amount of carbohydrate receptors such as effector proteins involved in the folding of nascent proteins, microbe host interactions which lead to infection, and mechanism of inflation and immunity.[8] In the past, there were difficulties with expanding carbohydrate research.

Some of these difficulties have been well documented. Characterization of carbohydrates ligand interactions is challenging because carbohydrates have diverse

structures and they are not readily synthesized. For instance, oligosaccharides have greater structural complexity than nucleic acids and proteins, due to branching and linkage variety. The complexity and diversity of the structure of oligosaccharides can be attributed to the glycosyltransferase genes which synthesize carbohydrates through nontemplated processes.[9] These processes are unique and not readily reproduced. Monitoring carbohydrate interactions require large amounts of sugars to be screened and reactions with carbohydrates are relatively weak in nature. Carbohydrate binding proteins generally have monomeric KD values in the micromolar to millimolar range.[10] Therefore, binding efficiency requires multivalent interactions, which can be hard to measure.[11]

Biophysical tools needed have not been fully developed to complement research in the field of glycomincs. Current techniques used to study carbohydrate interactions are labor intensive. Knock-out techniques which give insights to carbohydrate functions and interactions are time consuming and may interfere with the synthesis of some carbohydrates. There is a need for high throughput methods to detect and characterize these interactions. Purification of carbohydrates is challenging because many are closely related and possess similar physical properties. Advancements such as automated solid phase synthesizers and methods for enzymatic synthesis have led to pure chemically defied carbohydrates.[12,13] Hence, combining synthetic advances and new biochemical tools will increase the production of research in this area.

Glycan arrays can help overcome some of these challenges. Glycan arrays require small amount of product and many different compounds can be screened at on time. Already microarrays have been developed for DNA and proteins.[14,15] Work in DNA and protein arrays have been helped by a plethora of new laboratory techniques such as

automated synthesis of nucleic acids and peptides, rapid DNA and peptide sequencing, protein expression systems, and SiRNA gene silencing.[16] Glycan arrays combine carbohydrate libraries, attachment of saccharides to surface, high throughput expression of carbohydrate binding proteins, and analysis.[9]

Glycan arrays mimic naturally occurring events on cell surfaces so they are well suited for glycomic studies to examine binding profiles between carbohydrate and lectins. Arrays allow for the examination of multiple combinations of carbohydrates immobilized onto the surface. New technologies such as robotic arrays and high resolution imaging permit miniaturization of tens to thousands of carbohydrates which can be immobilized onto common glass microscope slides. Arrays reduce the amount of carbohydrates needed for binding experiments and this is important because often small amounts of carbohydrates are synthesized or purified. Carbohydrates can be isolated from cells, tissues, pathogens, or urine however, enzamayic or chemical synthesis methods of isolation are preferred because they allow for control of the structure being synthesized and produce homogeneous materials. Currently no amplification strategies such as polymerase chain reaction (PCR) or cloning exist to produce suitable quantities of oligosaccharides, so arrays are preferred because about 10^{-12} mol of compound can be used to perform several experiments.[17]

Large libraries of carbohydrates are needed for the fabrication of glycan arrays. Pure carbohydrates isolated from natural sources of chemical synthesis are difficult to obtain and they have to be derivatized with linkers to attach to array surfaces. Isolated or synthetic carbohydrates can be changed chemically or by enzymes. Functionalization and coupling reactions are important factors to designing carbohydrates. Functionalization of

glycans by chemical synthesis is generally easier than for natural occurring sugars. This involves reducing the terminal sugar end with the desirable linker for derivatization. Natural occurring sugars are present in smaller quantities and often require pre-existing groups such as amino groups, peptides, or lipids.[18] An alternative approach to the synthesis of glycans is enzymatic glycosylation using recombinant glycosyltransferases.[19] Glycosyltransferases are enzymes that synthesize glycosidic linkages. This process transfers a saccharide from a donor substrate to an acceptor substrate. Here, donors can be nucleotide sugars, sugar phosphates, sugar lipids, or saccharides. Acceptors tend to be other saccharides, proteins, lipids, or nucleic acids.[20] Realistically there are only a limited number of synthetic or natural glycans available. Libraries are generally created to focus on specific lectins so they tend to consist of a small number of carbohydrates.

Recognition between carbohydrates and lectins is dependent on how carbohydrates are presented on the surface. Because monovalent interactions are weak, multivalent interactions are desirable since there is an increase in binding sights. The orientation for optimal carbohydrate-lectin interactions is not easily established so several strategies have been employed to investigate glycan density. Strategies have included varying the spacing between glycan chains, printing multivariate glyconjugates onto surfaces, varying the context of the glycans, and spotting mixtures of glycans onto the surface.[21-23]

Spacing, orientation, linker length flexibility and ligand density are important for carbohydrate-lectin recognition to occur. Several factors affect ligand density. First, the maximum density is controlled by the capacity of the surface. Secondly saturation should not occur for materials that are printed onto the surface. Finally, some attachment

methods can cause carbohydrates to change new orientations once immobilized onto the surface creating densities that are not well defined. For carbohydrates that are printed on the surface, density can be controlled by varying the concentration of the carbohydrates in the print solution.

Since glycans must be present in large numbers to record a signal for analysis, immobilization methods are important factors to consider when developing microarrays. There a variety of immobilization chemistries used to fabricate arrays. Noncovalent site nonspecific methods assemble glycans to unmodified or modified surfaces. This method may lead to the washing away of glycans during purification steps. Another immobilization method is noncovalent site specific attachment. Examples of non-covalent attachment include immobilizing unmodified polysaccharides or lipid sugars to microtiter plates or binding of biotinylated sugars to streptavidin glass slides.[24,25] Other noncovalent immobilization methods include printing neoglycoplipids onto nitrocellulose of poly(vinylidenefluoride) (PVDF) sheets.[26] Large quantities of polysaccharides or lipids are needed for non-covalent immobilization strategies.

Covalent site nonspecific attachments of non-modified glycans involve a one-step procedure which can be time consuming and labor intensive. For covalent methods a glass slide is functionalized with a reactive group which is then coupled to a compatible functional group which contains the carbohydrate of interest. These reactions should be quick, specific, high yielding, and react with groups present on the carbohydrate.[4] Immobilization processes can be completed in several hours or days. Examples of this method include the recognition of 1,2 or 1,3 diols of sugars by boronic acid modified surfaces.[27] Covalent immobilization methods also involve amides and [3 + 2] Husigen

cycloadditon reactions.[28,29] Lectins gain access to immobilized carbohydrates because of the use of spacers which lie between the sugar and the array surface. The nature of the linker is important to prevent nonspecific binding on lectins. Oligo or poly(ethylene glycol) linkers are hydrophilic and help prevent nonspecific binding of lectins. In addition to the chemical makeup of the linker, the length is important as well, because it controls the access of glycans to its binding partner. [30]

Cellular assays, surface plasmon resonance (SPR), fluorescent imaging, and mass spectrometry (MS) are used as detection methods for glycan arrays. Fluorescent methods may exist in the form of dyes or labels, where the array is scanned for readout. Detection of fluorophores bound to glycans by a fluorescence scanner is the most common detection method for microarrays. Variations of this method include direct fluorophore labeling of the protein, labeling secondary reagents that bind to the protein, and labeling secondary reagents that bind to a tag such as biotin.[18] Common fluorescent dyes used in arrays are cyanine dyes because of their brightness and reduced complexity of labeling proteins. Fluorescence labels have high sensitivity and throughput and many different types of surfaces can be used such as glass, gold, and membranes. This method can be carried out with ELISA equipment making it inexpensive and readily available.

A typical fluorescence experiment begins before incubation with proteins. Fluorescent agents can be directly or indirectly bound to glycan binding proteins (GBP) with Fc fragments of human IgG, biotin, or anti-GBP antibodies. After incubation periods, probes are then reacted with binding partners such as streptavidin, or secondary antibodies. Washing of this complex removes any unbound protein complexes and the binding is detected with a fluorescence scanner. [31] Quantitative information can be

gathered because the fluorescence intensity is proportional to the amount of bound proteins. Advantages of using fluorescence include that many labels are commercially available, anti-glycan antibodies can be detected with labeled secondary antibodies, biotinylated proteins can be monitored by labeled streptavidin. Disadvantages of this method are insertion of labels can reduce activity or selective binding, labeled secondary regents are not always available, some regents can bind with carbohydrates, some captured GBPs can be removed during washing steps, and binding profiles may change due to binding of fluorescent labels .[32]

Because of these issues, there is a push in the research community to use label free methods of detection. Requirements for label free methods are techniques should be suitable for high throughput analysis, able to detect small molecules binding to immobilized targets, and detect interactions of carbohydrates with lectins at small concentrations. Label free techniques are acceptable alternatives to fluorescent based methods because they are not susceptible to interference from tags and the can provide real time kinetic information. SPR has been used as a label free method of detection for glycan arrays.

SPR occurs when an evanescent wave excites charge density oscillations along a metal dielectric interface. This leads to changes of the optical constants on the metal surface which influences the resonance condition.[33] SPR monitors optical parameters such as the angle at which photons resonantly couple with valence electrons on the metallic surface, wavelength at which resonance occurs, light phase, or modulation of light's polarization.[34] The change of refractive index is proportional to the mass and conformation of the analyte and immobilized constituent, number of analyte molecules

bound to the immobilized constituent, and distance of the bound analyte from the sensing surface. [34]

Flow systems combined with SPR allow for continuous monitoring of the formation of carbohydrate-lection complexes and the dissociation of analyte from the immobilized ligands.[35] There are two types of SPR sensors, angular and spectral. Angular SPR are based on angular interrogation and measure carbohydrate-lectin interactions by scanning angles of incidence at constant wavelength. Spectral SPR is based on wavelength interrogation and scans wavelengths at constant incidence angles.[33] SPR gives real time information of carbohydrates binding to lectins. SPR experiments proceeds first by immobilizing the analyte of interest onto a SPR chip. Next an interacting analyte is flowed over the chip and binds with the analyte on the surface causing a change in the refractive index of the chip. From this process association and dissociation constant can be calculated and there is no need for the use of fluorescent labels.[36] Sensitivity in SPR is dependent on interactions between the analytes being studied, surface chemistry of the array, density, activity of immobilized analyte, and availability of signal enhancers.[37] One challenge with SPR methods is that nonspecific binding can give false positive results. Nonspecific binding can occur on the sensor surface or to the analyte of interest and can be overcome by using gold surfaces and bioresistant surface chemistry.[34]

Signals of small peptides or proteins that are not detectable using traditional detection methods can be detected by MS and this technique can also provide chemical and structural information. MS can analyze a mixture of carbohydrate –lectin interactions provided that each interaction produces a unique mass to charge (m/z) peak. Analysis by

MS gives information on the species present in the reaction, the intended products, unintended components, and the substrate, thus creating a complete outlook of carbohydrate lectin activity. Surface enhanced laser desorption ionization time of flight mass spectrometry (SELDI-TOF-MS) has been used as a detection method for arrays. In this technique following the incubation of glycans with their binding partners, an energy absorb matrix is added to the substrate surface. SELDI-TOF-MS coupled with a UV nitrogen source laser causes the surface lectins to desorb and become ionized. The ions are moved by an electric field down a flight tube where they hit a detector and a signal is produced. The flight time is dependent on the m/z of the lectin, hence the mass of the lectin can be determined and used for identification.[38] Matrix assisted laser desorption ionization mass spectrometry is another MS (MALDI-TOF MS) technique that can be used as a detection method for arrays. MALDI can be combined with self-assembled monolayers to give specific interactions with carbohydrates that can be easily identified. MALDI can be used to monitor the immobilization of a peptide onto a monolayer, monitor enzymatic modification of an immobilized analyte, and provide kinetic data for carbohydrate-lectin interactions. Drawbacks of MS are the need for specialized equipment and sample preparation can be tedious and complex.

Other array detection methods not commonly used are quantum dots and atomic force microscopy (AFM). Quantum dots are nanocrystals made up of a semiconducting fluorescent core coated with a semiconducting shell which has a larger spectral band gap and light scattering capabilities. Quantum dots were shown to be 30 times more sensitive for the identification of cancer biomarkers.[39] They are advantageous over fluorescence labels because they have brighter fluorescence, stable against photo bleaching, and

tunable emission spectra.[40] They fall short in some areas because they have short lifetimes, are easily oxidized of photolyzed, and may be hazardous to the environment and humans.[41]

In AFM a nanometer sharp tip is attached to a cantilever that is sensitive to the force between the tip and sample surface. Disturbances to the cantilever are detected with a laser beam which is focused on the back of the cantilever, where the signal is then reflected onto a photodiode creating a topological image of the structure.[38] AFM gives information about surface features and it gives cantilever deflection with picometer resolution. It can detect carbohydrate-lectin interactions by measuring the height increase across an array surface. The technique will be limited in measuring biomolecules in aqueous solutions and artifacts are challenges as well.[42]

Applications of glycan arrays designed thus far have extended into many different areas. Their use is growing in the fields of biomedical research for diagnostic agents, high throughput analysis of glycan interactions with lectins, antibodies, and pathogens; and development of therapeutic agents. Ccarbohydrate-lectin binding is commonly detected and characterized using glycan arrays. Specific binding between mammalian, plant, viral, and, bacterial lectins has been analyzed by researchers at the Consotrium for Functional Glycomics. Amine derivatized glycans were coupled with a reactive surface containing N-hydroxysuccinimide using printing technology. Over 200 synthetic and natural glycans were placed onto the array. The glycans were then incubated with labeled proteins or sandwich procedures incorporating secondary antibodies. The lectin-carbohydrate interactions were analyzed by fluorescence scanning and imaging methods. This research demonstrated the advantages of using arrays. Only 0.1-2µg of protein was needed to get

an optimal fluorescence signal. The fluorescence intensity was greatly enhanced when the carbohydrates were presented as multivalent arrays.[43]

Pathogen detection is another application of glycan arrays. Pathogens are able to enter cells through proteins that bind to carbohydrates. These proteins regulate cell adhesion, binding, entry of toxins, and immune response. Understanding these interactions can lead to therapeutic treatments. Cancer associated carbohydrates are candidates for therapeutics because they are expressed on tumor cells in greater densities than many oncogenes and are located on proteins found in tumors.[44]

Glycan arras have been used to develop anticancer vaccines. It may be possible to target cancer cell glycans based on their altered glycosylation. Certain tumor glycans have embryonic origin, therefore, they are not perceived as foreign by the immune system so B cells expressing high affinity for these structures are not created during development. So anticancer vaccines can be designed to make the immune system respond to the tumor glycans as foreign invaders. Danishefsky, Livingston, and co-workers designed a glycan array by linking synthetic tumor glycans to the immunogenic carrier protein keyhole limpet haemocyanin (KLH). This created an immune response where peptide antigens required for T-cell help were generated. This method was employed into a vaccine form and when administered to mice and humans, antibody and cell mediated responses were generated against tumor glycans.[45]

Click Chemistry

As stated above there is a wide selection of synthetic transformations and immobilization techniques available to scientist to prepare highly complex molecular

structures. Most of these techniques have stringent reaction conditions that may require toxic solvents, stoichiometric reagents, extreme temperatures, and protocols for protective groups. Furthermore these reactions may not be selectively specific. These conditions make them unsuitable for reactions with biomolecules. Furthermore, modification and immobilization of biomolecules require for only a single functionality of interest to be selectively modified in the presence of functional groups; and this is hard to achieve. A solution is the use of bioorthogonal chemical species which form covalent bonds between two abiotic groups that have mutual reactivity. A group of reactions that fit this description are click chemistry reactions.[46]

The idea of click chemistry was first presented by Kolb, Finn, and Sharpless as a new strategy for organic chemistry. It is based on the idea of generating substances by joining small units together with heteroatom links. Several criteria must be met before a reaction may be labeled as click chemistry. Reactions must be modular, wide in scope, give high yields, create only one inoffensive byproduct which can be removed by nonchromatographic methods, and be stereospecific. Reaction conditions should have readily available starting materials, be simple, have very little use of solvent or none at all, and simple product recovery and isolation.[47]

Currently, five classes of reactions have been identified to meet these requirements. They are: Cycloaddition of unsaturated species: 1, 3-dipolar cycloaddition, cycloaddition of unsaturated species: [4+2]-cycloaddition (Diels-Alder), nucleophilic substitution/ring opening reactions, carbonyl reactions of the non-adol type, and addition to carbon-carbon multiple bonds.[48] Of these reactions, cycloaddition of unsaturated species has become the method of choice for click chemistry. These reactions fuse two

13

unsaturated reactants and provide access to a variety of five and six membered heterocycles.[49] The Cu^{I}-catalyzed 1, 3 dipolar cycloaddition of azides and alkynes (CuAAC) to yield 1, 2, 3 triazoles is considered to be the most useful cycloaddition reaction for click chemistry.

The usefulness of this reaction lies in the azide and alkynes moieties. Azides are an important moiety in click chemistry. They are abiotic in animals and absent from nearly all naturally occurring species. They do not react with water and are resistant to oxidation. The do not react with amines or other nucleophiles that are present in biological systems. Azides are 1,3-dipoles which can undergo reactions with dipolarophiles like activated alkynes.[50] The synthesis of alkyne-azide functionality is rather easy and the formation of a triazole between azides and alkynes is irreversible.[51]

Click chemistry has proven its worth by the immobilization of sugars, proteins, DNA, and cells in previous works. They are successful for bioconjugation specifically, attaching proteins and cell-adhesion ligands and attachment of cells.[52,53] Bioconjugation uses techniques which involve covalent attachment of synthetic labels to a biomolecular framework. Chaikof et al[54] demonstrated how click chemistry can be used to immobilize proteins and carbohydrates onto solid support substrates. This work showed that the attachment of such molecules to support substrates preserves their activity and this opens the door for the development of microarrays, microbeads, and biosensor chips.

There is one aspect of Cu^{I} catalyzed click chemistry this is unappealing. The Cu^{I} catalyst used for cycloaddition reactions is toxic to bacterial and mammalian cells.[55] The limitations of CuAAC reactions are highlighted when this method is applied to biological applications. Fokin, Sharpless, Finn, and co-workers noticed that cowpea mosaic virus

degraded or aggregated under CuAAC reaction conditions.[56] Cu causes oxidative stress in biomacromolecules which leads to the formation of reactive oxygen species (ROS).[57] This is problematic to the functionalization of proteins, nucleic acids, polysaccharides and lipids. In proteins, ROS cause cleavage of polypeptide chains and degradation of amino acids.[58]

Therefore, Cu^I-free cycloaddition reactions are more desirable. An alternative method for activating alkynes with azides for Cu^I-free cycloaddition involves ring strain. Wittig and Krebs showed that the reaction between cyclooctyne and azides can proceed like an explosion to give a triazole product.[59] Constraining an alkyne within an eight-membered ring creates 18 kcal mol^{-1} of ring strain which is released upon cycloaddition with an azide.[60] Attaching electron withdrawing groups to cyclooctynes will further increase the rate of this strain promoted reaction. Furthermore, attaching benzene rings to cyclooctyne rings impose additional ring strain hence, increasing the reactivity of the alkyne.

There are several types of strain promoted cycloaddition reactions (SPC). The SPC between alkenes and dipoles have been used for bioorthogonal conjugation techniques. Carrell et al, investigated SPC between norborenes and nitrile oxides. This reaction produces exoselective substituted 2-isoxazolines and it was used to label oligonucleotides in solution and on solid supports with high efficiency.[61] A bioorthogonal reaction based on strain promoted alkyne-nitrone cycloaddition (SPANC) was proposed by Boons et al. Benzofused cyclooctynes react rapidly with nitrones resulting in isoxazolines. Depending on the substituents, the reaction proceeded at rates 32 times

faster than those for SPAAC. This method has been used for site specific modification of peptides and proteins.[62]

Copper (I) free cycloaddition has been compared to Cu^I catalyzed azide alkyne cycloaddition in a study by Agard et. al. This study examined the effect of these types of reactions on live cell labeling. Significant cell death was observed when Cu^I catalyzed reactions were used. No cell death occurred when strain promoted cycloaddition was used. Because of the benefits of click chemistry, more specifically strain promoted click chemistry, this method will be used for the covalent attachment of carbohydrates to solid supports for the development of glycoarray technology.

Variations of SPC reactions have successfully immobilized carbohydrates onto microarray surfaces. In research by Wendel, Singh, Rinnen, and others β-$_D$-galactose was immobilized to an array surface using SPC. Oxime linked with cyclooctyne was printed onto an undecylazide modified glass surface using a microstructured elastomeric stamp. This produced an azide/oxime surface. AFM and X-ray photoelectron spectroscopy (XPS) were used to verify the immobilization of these moieties. After verification of the azide/oxime surface a β-$_D$-galactose norborene conjugate was synthesized and reacted with the modified surface in the presence of diacetoxy iodobenzene (DIB).[63] To this carbohydrate was bound TRITC-labeled peanut agglutinin. High surface coverage was verified with fluorescence microscopy. This paper demonstrated that SPC reactions can be carried out under mild conditions, with high yields, in an aqueous environment at room temperature.[64]

Surface Enhanced Raman Spectroscopy

The purpose of this dissertation research is to develop a glycan detection methods based on surface enhanced Raman spectroscopy (SERS). To that end, focus will shift to describing Raman spectroscopy and SERS. Raman spectroscopy is an optical technique based on the scattering of photons. Over the years this field of vibrational spectroscopy has seen a double digit growth in the number of articles published in scientific journals.[65] Raman spectroscopy has proven to be a powerful technique in the fields of pharmaceutics, forensics, bioscience, and biomedicine. Nontraditional uses of Raman spectroscopy has ventured into the field of art, archeology, pigments and paints. In this field, Raman spectroscopy has been used to identify the first occurrence of yellow pigment from Italian pottery from the 16th century.[66] Raman spectroscopy has also been used as a characterization technique of nanomaterials. The interest and growth of applications for Raman spectroscopy is expected to continually expand into more areas over the next half century.

Raman spectroscopy was discovered by Sir C.V. Raman in 1921.[67] Sir Raman observed that light from scattered molecules is accompanied by scattered light at a different wavelength from the incident light. Scattering is defined as the simultaneous absorption of an incident photon and emission of another photon. [68] The light observed by Sir Raman appeared very weakly and was difficult to see in his experiments. Since this discovery the scattering of light observed has been classified as either inelastic or elastic scattering. Rayleigh scattering, also known as elastic scattering; a molecule is struck by a photon with excitation energy $h\nu_0$. From this process, energy is emitted at the same $h\nu_0$ frequency. Inelastic scattering photons have energy at a different wavelength

from the incident photons. The photons of inelastic scattering may have a wavelength higher or lower than the incident photons. It is inelastic scattering that was observed by Sir C.V. Raman in 1921.

A diagram representing the quantum energy transitions for Rayleigh and Raman scattering is shown in Figure1.1.[69]
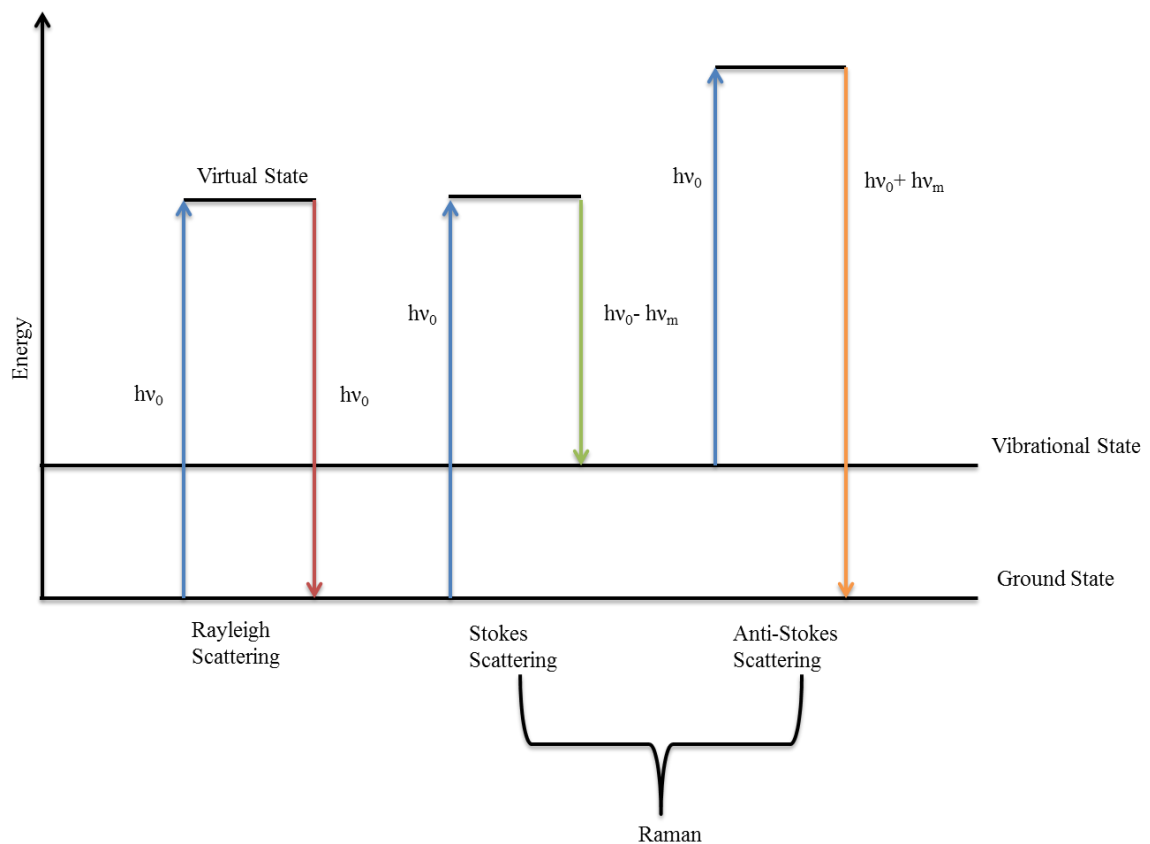
Figure 1.1 Diagram representing quantum energy transitions for Rayleigh and Raman Scattering.

Stokes scattering is used to describe scattered photons which have less energy than the incident photons. In this case a molecule is excited from the vibrational ground state to the first excited state. Therefore, absorbed energy at $h\nu_0$ frequency will be scattered a

frequency of $h\nu_0- h\nu_m$. Anti-Stokes scattering is used to describe scattered photons which have higher energy than the incident photons. This can be described as a vibrational excited electron relaxing into the vibrational ground state.[68] In this case the reemitted light is at a higher frequency $h\nu_{0+} h\nu_m$. Anti-Stokes scattering is generally much weaker than Stokes scattering and sometimes cannot be observed. This is because more molecules are in the ground state at room temperature.

Rayleigh and Raman scattering are weak processes. If total intensity of incident light is considered, only about one part in a thousand of the total intensity can be Rayleigh scattered. The value drops to one part in a million for Raman scattering. Stokes and anti-Stokes peaks are symmetrically positioned about the Rayleigh peak. Because Rayleigh scattering is so much stronger, Raman instruments must be designed to filter this intense light. Another competitive effect to Raman scattering is fluorescence. Some excitation sources used for Raman experiments can excite fluorescence of the analyte being measured. Even weak fluorescence can be stronger than Raman scattering thus overpowering the weak Raman signal.[70]

Raman instruments consist of an excitation source, sample illumination system, wavelength selector, and detector. Raman spectrometers are commonly combined with microscopes. This setup is allows for focus of excitation light into a spot several microns in diameter to capture a Raman spectrum. This allows for the Raman spectrum to be captured at high collection efficiency owing to the high numerical aperture of the microscope lens. Fluorescence can be limited using a confocal Raman setup which reduces the collection of fluorescence to photons emitted. Confocal describes a microscope where the sample is illuminated with a point source and the image of the

point is detected through a pinhole located in front of the detector.[71] Continuous wave

lasers (CW) are generally used as excitation sources for Raman instruments. These lasers

offer the advantage of narrow linewidth and are able to isolate a specific vibrational

transition. CW lasers can be tunable for Raman resonance experiments. Over 95% of the

commercially available lasers on Raman instruments are Argon Ion, frequency doubled

YAG, diode, and Helium-Neon CW lasers.[72]

Raman scattering intensity is proportional to $\tilde{\upsilon}^4$, where $\tilde{\upsilon}$ is the frequency of

radiation from the laser excitation. Because the Raman scattering is proportional to the

excitation power, the power of laser excitation should be selected so as to not thermally

decompose the sample. Resolution is given by:

$$\Delta x = 0.61 \text{ x } \lambda/NA \text{ } \textbf{Equation(1.1)}.$$

Δx = shortest distinguishable distance between two points

λ = wavelength

NA = numerical aperture of objective.

This equation indicates shorter excitation wavelengths yield higher lateral resolution.

Lateral resolution decreases at longer wavelengths and efficiency of Raman scattering

decreases. To optimize collection efficiency and spatial resolution, the highest numerical

aperture available should be used when collecting spectra.[71]

Gratings in Raman spectrometers disperse the collected signal to the charge

coupled device (CCD) detector by deflecting wavelengths at different angles. Dispersion

characteristics of gratings are determined by the number of grooves per millimeter,

hence, high number of grooves give high dispersion and resolution as well as distributing

the Raman signal over a large number of CCD pixels. CCD detectors are made of an

array of light sensitive silicon photodiodes, where each is connected to a capacitor. CCD detectors are graded by their quantum efficiency (QE), which is the percentage of detected photons to total incoming photons. QE determines the quality of the signal detected in Raman sampling and this can be affected by noise in the signal or noise in the detector. Common noise sources of noises are shot, thermal and readout noises. Thermal noise can be reduced by cooling the detector. Readout noise is dependent on the readout amplifier and readout speed. Ideal detectors have 100% QE, no thermal, readout, or any other sources of noise, or shot noise is limited.[71] Raman spectra generated from collected measurements are presented in units of wavenumber shift, which is the distance between the Raman scatter and the laser probe. Positive wavenumber shifts are Stokes shifts and negative wavenumbers are anti-Stokes shifts.

Carbohydrates and their interactions with various biomolecules have been analyzed with Raman spectroscopy. Raman was used to detect physiological concentrations of glucose in vitro from simulated aqueous human solutions.[73] However, these experiments required high laser powers and long acquisition times which are not desirable for point of care diagnosis similar to a commercially available glucose monitors. Another vibrational spectroscopy method that has been use to examine carbohydrate interactions is IR spectroscopy. This method was used to examine aromatic protein interactions with various sugars.[74] There are some advantages of using Raman over IR. Water can be used as a solvent for Raman analysis, which is helpful for analysis of biological and symmetric vibrations which are observable in Raman. However, due to the weak scattering cross section of Raman, there is a desire to increase this efficiency.

In 1977 Van Duyne and independently Albrecht observed very intense Raman signals measured from pyridine on a roughed silver electrode. They hypothesized that the intensely strong Raman signal was an enhancement of the Raman scattering efficiency and could not be explained by anything else, such as an increase of the adsorbed molecules. This effect observed by Van Duyne and Albrecht has become known as surface-enhanced Raman scattering (SERS). SERS has made advancements in systems for biomolecular detection, cancer cell detection and imaging, DNA and RNA detection, and food quality analysis.

A general description for the SERS effect begins with how an electromagnetic wave interacts with a roughened metal surface. When an electromagnetic wave interacts with roughened surfaces, plasmons on the surface are excited. Figure 1.2 diagrams the surface plasmon effect.
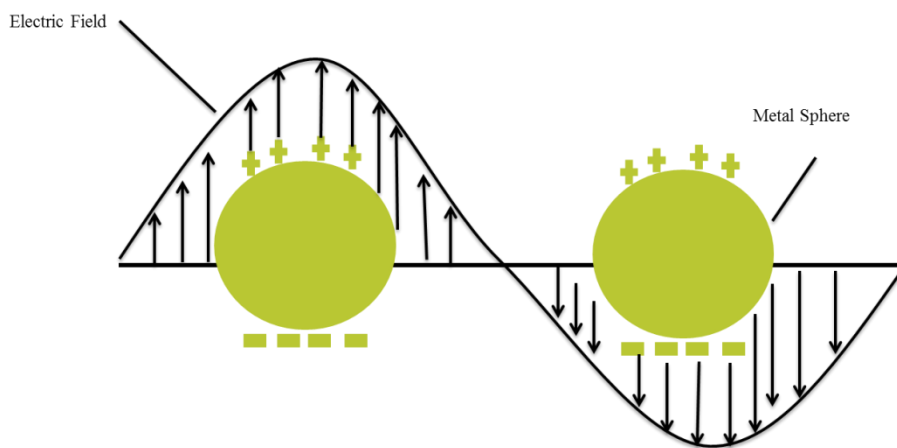
Figure 1.2. Diagram illustrating localized surface plasmon effect (a) surface propagating plasmon and (b) a localized surface plasmon.

Plasmons are excitations associated with high frequency collective oscillations of valence electrons.[75] Their excitation creates an oscillating dipole in the particles resulting in

22

enhancement of electromagnetic fields near the surface. When a laser excites electrons in these metals, they will oscillate collectively across the surface generating so-called surface plasmons. If the plasmons are excited to resonance, there is an increase in the local electromagnetic field which correlates to an enhanced Raman signal.[76] This can be seen by the dipole moment which is expressed as:

$$\mu(t) = P \cdot E_{inc}(t). \textbf{ Equation(1.2)}.$$

The incident electric field and molecular polarizability are E_{inc} and P respectively. Because the Raman intensity is proportional to the square of the dipole; an increase in polarizability or incident field enhances the Raman signal. Roughened surfaces used to achieve this effect have consisted of coinage metals such as platinum, ruthenium, iron, and nickel. However, very large enhancements have been reported from noble metals such as gold, silver, and copper.

In normal Raman, the Stokes Raman signal is proportional to the Raman cross section, the excitation laser intensity, and the number of molecules in the probed volume.[77] In SERS experiments where molecules are attached to metal colloids, the SERS Stokes Raman signal is proportional to the Raman cross section of the adsorbed molecule, the excitation laser, and the number of molecules which are involved in the SERS process.[77] There still exists great debate over what causes the SERS effect. The first experiments on SERS demonstrated how pyridine signals were enhanced when they were attached to roughened metallic surfaces. Methanol however shows no such enhancements, nor do molecules attached to flat metallic surfaces. This is an example of the electromagnetic enhancement effect. There is a gap of about two orders of magnitude between calculated electromagnetic SERS enhancements and experimentally observed

nonresonant SERS enhancement factors.[78] Findings from early experiments by Jean

Marie and VanDuyne point to electromagnetic enhancement as the mechanism for SERS

enhancement while Albrecht and Creighton proposed the chemical or charge transfer

enhancements for the reason behind the SERS effect. Therefore, the consensus is that

chemical and electromagnetic enhancements are the major contributors of the SERS

effect.

Chemical enhancements require direct contact between the molecule and the

metal and contribute an enhancement magnitude of $10\text{-}10^2$.[79] Chemical enhancement

provides information on chemisorption interactions between metal and adsorbate. Several

models exist to explain the chemical enhancement effect of SERS. One proposed model

is assigned to electronic coupling between the molecule and metal forming an adsorbate

complex which increases the Raman cross section of an adsorbed molecule.[78] This model

is called charge transfer and is linked with the overlapped electronic wavefunctions

between the metal and adsorbte leading to ground state and light induced charge transfer

processes.[80] In the charge transfer model, an electron of the metal which has been excited

by a laser, tunnels into a charge transfer excited state of the adsorbed molecule. This

induces a different equilibrium geometry for the adsorbate molecule electron. A

relaxation of the excited state returns the electron to the metal, which leads to a

vibrationally excited neutral molecule and emission of a Raman-shifted photon.[80]

The adatom model explains the chemical enhancement effect of molecules

adsorbed at different defect sites on the metal surface. This model concludes the charge

transfer enhancement mechanism receives contributions from adsorbates at special active

sites of atomic scale roughness.(ASR)[81] The adatom model was proposed by Otto et.al.

who theorized a strong Raman enhancement for an adsorbate on a silver surface was only possible when the adsorbate was bound to the ASR.[82] In this model, the electron laser interaction increases the Raman signal because there is an additional momentum available for scattering of electron hole pairs at ASR conditions. [82]

The electromagnetic enhancement effect is largely responsible for SERS. The electromagnetic enhancement effect is attributed to enhanced local optical fields between the molecule and metal surface due to excitation of electromagnetic resonances. Excitation of the surface plamson increases the local field experienced by the molecule adsorbed on the surface of the particle to roughly E^4. The resonances are due to excitation of surface plasmons. At the plasmon frequency, metal nanoparticles are polarized which results in a large local field on the surface and an increase in Raman enhancement.[83] The molecule adsorbed onto the surface experiences an effective magnetic field. However, the field decreases as it moves away from the surface a distance of $1/r^3$.[80]

The magnitude of SERS enhancement is dependent upon the design of the SERS substrate. There are no clear cut best practices to design them, but they should be designed to optimize chemical and electromagnetic enhancement effects. When designing SERS substrates; the composition of the metal, size, and shape of the nanomaterials is important. As stated earlier, silver and gold are the most commonly used materials for designing substrates. Furthermore, SERS signals from silver nanomaterials are 10-100 fold higher than for gold nanomaterials.[84] However, gold nanomaterials are generally used for biological samples because they have higher biocompatibility and better control over its shape and size.[85] The intensity of the SERS signal increases with size and nanomaterials size can range from 15-200nm. The optimum size was originally thought

to be between 20-100nm, however, it has been shown that SERS signals can be obtained from particles as large as 200nm in size.[86,87] Controlling the shape allows for tuning the optical and spectroscopic response of nanomaterials which is important for a wide range of applications.[88]

Many different platforms have been used as SERS substrates. They started out as roughened silver electrodes, moved to silver and gold colloids, metal films, and a combination of colloids and metal films.[89] Colloids are versatile, inexpensive, and cheap to fabricate. They are used for studies as a sol or dried onto a solid support which leads to the formation of a variety of shapes.[90] Colloids can be prepared by chemical reduction, laser ablation, or photoreduction. The reduction of gold or silver with sodium citrate or sodium borohydride is the most common chemical reduction method for the synthesis of nanoparticles. Particles made in this manner can be stable for months to a year.[91] Stability of the colloidal solutions is attributed to ions, introduced during the fabrication of colloids, absorbed to the metal surface.[90]

Self-assembly of nanoparticles into various patterns or arrays is another way to fabricate SERS substrates. Different assembly approaches include layer-by-layer (LBL), solvent induced evaporation, Langmuir-Blodgett (LB) assembly, external field-driven assembly, ion or small molecule or polymer induced assembly, and liquid-liquid (LL) interface assembly.[92] In our lab, we have used LBL, LB-assembly, and LL interface assembly to make SERS substrates. LBL assembly is a technique where multilayers are fabricated by alternating the deposition of positively and negatively charged colloid particles.[93] Substrates fabricated in this manner have reported SERS enhancement factors of $7x10^4$.[94] LB-assembly generate a large uniform metal nanoparticle film up to several

tens or hundreds of square centimeters in an area while controlling the optical and electrical properties of the nanoparticle film being fabricated.[95] A calculated enhancement factor for this technique was reported as 1.2×10^6. LL assembly allows for nanoparticles to assemble into three dimensional constructs while reserving their properties, and easy access to nanoparticles for post modification.[96] This technique can yield enhancement factors of 5.06×10^8.[97]

Immobilized nanorod assemblies are another kind of method that can be used to fabricate SERS substrates. These assemblies produce predictable and reproducible SERS enhancements, allow for surface plasmon tunability at any wavelength, and large area uniformity.[89] Our group has fabricated aligned silver nanorod arrays (AgNAs) by oblique angle deposition (OAD) as SERS substrates. OAD is a physical vapor deposition technique where a substrate is positioned close to the grazing angle resulting in the growth of nanorods in the direction of the deposition.[98] Characterization of these rods by scanning electron microscopy, transmission electron microscopy, and X-ray diffraction showed them to be cylindrical with some irregular shaped surface protrusions. This substrate can be patterned by a polymer molding technique to provide high throughput sensing and multiplexing.[99] Figure 1.3 show SEM images of tR array substrates. Enhancement factors of these substrates were reported to be around 5×10^8 and a uniform signal was achieved across the area of each substrate fabricated.[98] While there are many ways to fabricate SERS substrates several factors should be adhered to during the design process. They should be stable, reproducibly prepared, inexpensive, and easy to make for routine experiments.[100]
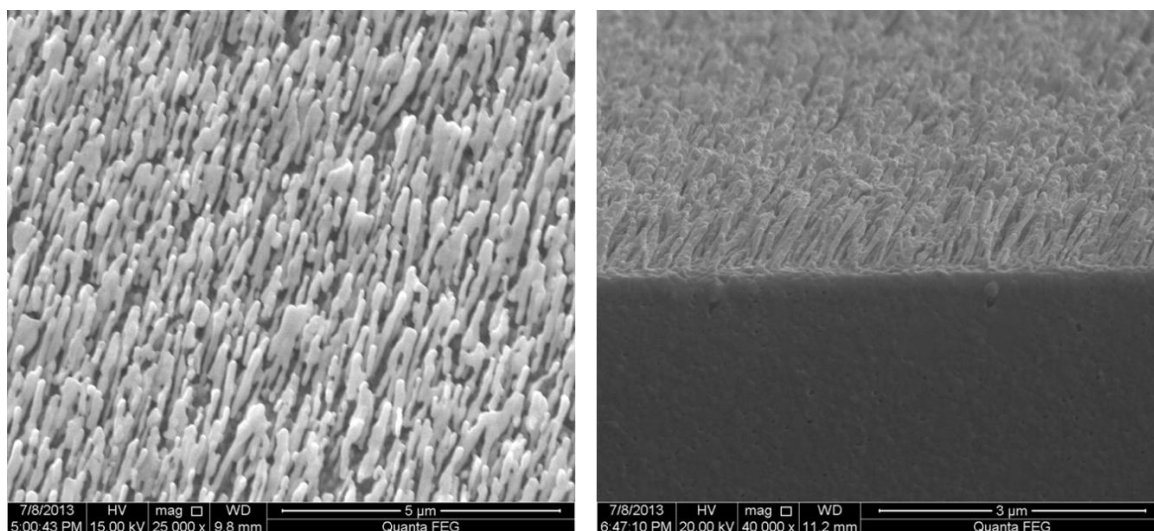
Figure 1.3 SEM images of Ag nanorod array substrate.

SERS substrates can be used as chemical sensors in a wide variety of applications. Various applied research areas include corrosion, catalysis, advanced materials, biology, and sensors. In environmental analysis SERS sensors have been used to identify herbicides, banned food dyes, and pesticides in drinking water. SERS sensors have also been used to detect pathogens such as adenovirus, bacterial meningitis, human immunodeficiency virus, and bacillus spores. Furthermore, there is a movement to implement SERS in clinical settings and research is being developed to probe in vivo analysis.[101] Since the interest of this dissertation is to develop microarray technology to detect carbohydrates and their interactions with various biomolecules a brief review of SERS related research is given.

Silver nanorod arrays used as SERS substrates have been developed by our group and applied to the detection of pathogens. We have fabricated a microwell arrayed SERS chip on standard glass microscope slides. Fabrication of this substrate included the use of the OAD technique to produce a SERS active surface, followed by well array pattering

with polydimethylsiloxane (PDMS). The use of such substrate has the advantage of saving on costs of deposition materials, requiring smaller amounts of samples, and the ability to perform multiple isolated analyses. One application of this array was to detect mycoplasma pneumonia in simulated and true clinical throat swab specimens. Three M. pneumonaie strains were reproducibly differentiated by this technique and were detected with a lower detection limit than standard PCR.[102] The array fabrication process shown here will serve as the basis for the development of carbohydrate microarrays based on SERS detection.

Another goal of this dissertation is to use SERS to detect carbohydrate interactions. Very little work has been done in this area, so it is still developing. However, recent work has demonstrated how SERS can be applied to this area.[103] A SERS aggregation system was designed to investigate glycan interactions with Concanavalin A (ConA), which has specificity towards α-$_D$-mannosyl and α-$_D$-glucosyl residues.[104] Glycans were immobilized onto nanoparticles and a SERS reporter compound, benzotriazole, dye was added to detect the signal. These glycannanoparticles were later incubated with ConA. Using information obtained from a calibration study which measured the SERS intensity obtained from the particles when mixed with ConA, a limit of detection was calculated from this method to be 40pM. Although an array was not created, this technique shows the potential to monitor carbohydrate interactions with various biomolecules.

Specific interactions we are interested in investigating are those between carbohydrates and influenza. More specifically we are interested in the interactions between hemagglutinin (HA) and carbohydrates. HA is a glycoprotein of influenza and it

binds with carbohydrates on the surfaces of cells. We have demonstrated the label free

detection of influenza proteins using the AgNR array. This study was carried out by

binding influenza protein to a polyvalent anti-influenza aptamer. Selective binding

between the aptamer-influenza protein complex could be identified based on collected

spectral signal.[105] Combining techniques and ideas from the applications presented

demonstrates the belief that SERS detection of carbohydrate interactions with various

biomolecules can be achieved.

Chemometrics

Once SERS spectra are collected, they can be analyzed by chemometrics.

Chemometrics uses statistical and mathematical methods to provide information about

SERS data. Mathematics and statistics are also used to design or select optimal

experimental procedures, analyze chemical data to give relevant information, and acquire

knowledge about chemical systems. It applies multivariate data analysis to chemical data

to determine important variables that are not readily measurable. Chemometrics has been

used to determine the concentration of a compound or compound mixture, predict the

property or activity of a chemical compound, and evaluate the process status in chemical

technology.[106]

The term chemometrics was first used by Svante Wold of the University of Umea

in1971. Early combinations of statistics and mathematics involved studying areas of

curve fitting and statistical control methods.[107] The field of chemometrics was developed

as advancements to the computer were made. In the late 1960s to early 1970s papers and

books were published on multivariate methods to interpret analytical chemistry

data.[108,109] Chemomertics really expanded as it was becoming increasingly applied to near infrared (NIR) techniques such as multivariate statistical methods principal component analysis (PCA) and partial least squares (PLS). The first paper with chemometrics in its title was submitted in 1975 by Kowalski in which he demonstrated the value of pattern recognition.

Kowalski viewed chemometrics as a tool to that can extract data from a wide variety of chemical fields. Kowalski's interest in chemometrics began when he became concerned that large amounts of data being collected from analytical measurements were not being used to their full potential. At this time, there were no computer programs being written for chemometrics so he created his own program for pattern recognition. Immediately he saw the benefits of using this technique. Using this program the combination of pattern recognition and display techniques gave solutions to problems readily and n-dimensional measurement analysis was an improvement over simple measurements by graphs. [110]Chemometrics met initial pushback from scientists who believed using this technique meant that experiments were being performed incorrectly. Furthermore, some persons suggested that the field would die out.[111,112] Now over a hundred journals publish papers with chemometric analysis.

Multivariate statistics is used because many analytical chemistry measurements require analysis of multiple species. This kind of analysis is divided into multivariate classification and multivariate regression techniques. Regression methods are used to analyze components of complex samples. A calibration model is built using data sets with known concentrations and samples with unknown concentrations can be predicted.[113]

Multivariate classification is a pattern recognition methods that can be further divided into unsupervised and supervised techniques.

Unsupervised methods require no a priori knowledge about the training samples, while a priori knowledge is required for supervised techniques. Samples which undergo analysis by unsupervised methods will be grouped based on their shared components and not qualifiers such as class membership. Therefore, this method is used for data exploration so trends can be observed. Supervised techniques draw boundaries for classification of samples into predefined boundaries.[113] Supervised pattern recognition methods can be summed up in the following manner: Training, calibration, and tests sets are created where class membership is known. Variables containing classification information are then selected. A mathematical model is derived using the training data that describes the classes based on the variables belonging to the sample. Finally, the model is validated to determine the reliability of the classification.[114] Common supervised and unsupervised statistical methods that will be discussed are PCA, hierarchial cluster analysis (HCA) and partial least squares discriminant analysis (PLS-DA) will be discussed as a supervised pattern recognition technique.

There are several advantages of using multivariate statistical analysis. Since multivariate analysis uses all measured data, there is a reduction in noise caused by the redundancy of measuring the same occurrence. Another advantage is that mathematics makes non-selective signals become selective. Information gained from a multivariate analysis can be used to validate, improve, and understand why generated models do not work. This can be gained from scores and loadings plots which provide information on samples that will give an understanding of the model. Also, the detection of data outliers

is increased. This is important because errors can cause misinterpretation of model information.[115]

Analysis of SERS spectra can be complicated. Fluorescence background noise and CCD background noise are just a few occurrences that may contribute to spurious SERS signals.[116] Because many biological samples contain similar components, band overlapping is also a problem in spectral analysis. Since SERS spectra can be complex, special preprocessing steps are applied to remove these affects. It is also done to improve exploratory analysis and classification models. Preprocessing steps also prevent researchers from making incorrect conclusions or generalizations about their data. Failure to preprocess data correctly can lead to the observance of noise or artefacts as real data. Generally preprocessing steps include normalization, spectral derivation, and mean centering. Preprocessing steps can be divided into sample base methods and variable based methods. Sample based methods acts on each sample one at a time to remove unwanted variance from individual samples. Normalization and derivitization are sample based methods. Variable based methods act on variables of the data. The order of preprocessing steps is important. Sample based preprocessing steps should occur prior to variable based methods.

Our group has developed an order of preprocessing step. The first step in preprocessing is baseline correction and it reduces uncontrolled background effects. Derivatives remove additive and multiplicative effects from spectra. The first derivative removes the baseline and the second derivative removes both the baseline and linear trend.[117] Derivative methods in spectroscopy are used for spectra discrimination, resolution enhancement, and quantitative analysis. Applied derivatives reveal small

spectral differences and overlapped bands between seemingly similar spectra in a visual way. It is recommended that derivative and smoothing take place in sync to provide useful information.[118]

One such method that both smooths and derivatizes spectra is Savitzky Golay (SG). This method was first popularized in 1964 by Abraham Savitzky and Marcel Golay. The algorithm for SG first smooths spectral data by fitting it to a spectral curve then derivatizes the fitted polynomial curve created.[119] The following smoothing equation is fit to spectral data:

$$Y_i = a_0 + a_1(v_j - v_i) + a_2(v_j - v_i)^2 + \ldots + a_p(v_j - v_i)^p \textbf{ Equation(1.3)}.$$

Here, $a_0, a_1, \ldots a_p$ fit the p-degree polynomial to the data within an interval of $2n + 1$ points centered on frequency v_i where $v_{i-n} \leq v_j \leq v_{i+n}$ are generated by a least squares criterion. Hence, the central data point of the smoothing interval is replaced and the data window is moved to the center of the next data point. Once spectral data has been smoothed, the derivatives of any order can be calculated.

After the derivative is calculated the next step is spectral normalization. Normalization methods identify a feature of spectral data that is consistent from sample to sample and scales all variables to this feature. This ensures that all a samples will have equal impact on the model constructed for analysis. Normalization methods are needed because the spectral intensity may vary from sample to sample. This process is carried out by dividing each data point $\textbf{X}_\lambda$ of a spectrum $\textbf{X}$ by a representative number $\textbf{A}$. Here $\textbf{A}$ can be the maximum value of the spectrum, the sum of all values in the spectra also called the unit area, or the sum of squares of all values which is the unit length.[120] The final preprocessing step that will be reviewed is mean centering. Mean centering

subtracts the mean spectrum of the data set from every spectrum in the data set. In doing so a sloping background is removed and variance within the spectral data set is not altered.[121]

After preprocessing spectra, data analysis can proceed by using exploratory methods such as HCA and PCA. They both allow for graphical visualization of data sets which may include high sample numbers and variables. Exploration of the presence or absence of class groupings is the type of information that can be gained from these methods.[122] HCA is the most widely used clustering method. HCA has two categories: agglomerative and partition methods. An agglomerative method starts each sample as its own cluster and then combines existing clusters into larger ones. Agglomerative methods include nearest neighbor, furthest neighbor, pair group average, centroid, median, and ward's method. In partition methods all samples are considered to be a single cluster and then they are divided into smaller clusters. Partitioning method most commonly used is K-means.[118,123]

When setting up an HCA experiment, the SERS spectral data is first preprocessed using the procedures above. Next the distance measure is specified. PLS toolbox uses PCA, the original X-variables, Euclidean distance, or Mahalanobis distance to make distance determinations. The number of PCs to use for distance measure is selected and using PCA will reduce noise and provides collinearity. Euclidean distance is the geometric distance in multidimensional space and it is the most common measure of distance. Mahalanobis allows for the dominate multivariate directions to be accounted for when performing clustering procedures. It accounts for the object points in the variable space and is independent from the scaling of variables.[106]

After distance measures are selected, the type of HCA method is selected. The partition method K-means selects K objects that are to be used as cluster targets. Here, K is determined a priori. As this method progresses objects are allocated to one of the clusters using the distance from each of the K targets. Next new cluster targets are calculated as the means of the objects in each cluster. This procedure is repeated until no objects are reassigned.[118]

Agglomerative method starts each sample as its own cluster and then combines existing clusters into larger ones. These methods differ on how the distance between the clusters is defined. For Nearest neighbor, the distance between clusters is the minimum of all possible pair wise distances of objects between the two clusters. Furthest neighbor distance is defined as the maximum of all pair wise distances of objects between the two clusters. Pair group averages distance between clusters as the average distance over all pair wise distances between objects in the clusters. Centroid is defined by the multivariate means of each cluster. Median distance is the difference in the weighted multivariate means of each cluster. Ward's method does not account for cluster centers. It joins existing clusters in a manner that minimizes within cluster variance.

HCA produces graphical information in the form of a dendrogram which is a two dimensional representation. Dendograms show how samples are grouped into successively larger and more inclusive categories and they show what groups are created when the clustering criteria changes. From the dendogram we can explore sample pattern analysis, grouping, decision making, data mining, and pattern classification.

PCA is a method for reorganizing data from a set of samples. This technique is a data reduction method from which information can be extracted from large data sets. The

individual samples are expressed as a linear summation of the original data multiplied by

a coefficient which describes the PC. The values generated from the linear summation are

known as scores. PCs explain a majority of the information contained within the data.

PCA dissolves data sets into orthogonal components where linear combinations

approximate the original data with a desired degree of accuracy. As PCs are calculated

each component describes the maximum possible amount of residual variance in the data

set.[124]

PCA identifies variables which describe major trends in data. Consider a data

matrix $\mathbf{X}$. Each sample is represented by rows and is designated by $\mathbf{m}$. The variables are

represented by columns and are designated by $\mathbf{n}$. PCA decomposes the matrix $\mathbf{X}$ as the

sum of r $\mathbf{t_i}$ and $\mathbf{p_i}$. The rank of the matrix $\mathbf{X}$ is denoted as r. The $\mathbf{t_i}$ and $\mathbf{p_i}$ pairs are ordered

by the amount of variance captured. Scores are $\mathbf{t_i}$ vectors and $\mathbf{p_i}$ vectors are loadings. The

score vector is a linear combination of the original $\mathbf{X}$ variables defined by $\mathbf{p_i}$. Loadings

contain information on how the variables correlate to each other. The $\mathbf{t_i}$, $\mathbf{p_i}$ pairs are

ordered in descending order according to the amount of variance described. So, the first

pair captures the largest amount of data for any pair. After so many components are

calculated the remaining small variance factors are put into a residual matrix $\mathbf{E}$.[125]

Scores can be visualized as line or scatter plots. Samples that have similar scores

will be grouped together, so classes formed from the PCA analysis can be readily

visualized. Scores plots are only specific for the fraction of variance they describe. For

instance, scores that describe five percent of the data do not have much meaning over the

entire data set. Loadings identify what a principal component represents therefore, what

the scores of the model represent can be described. Loading plots can also identify which

variables of the data cause groupings in the scores plots. The loadings and scores plots can be linked together. This allows for observed groupings to be explained.[126] Residuals provide information as to what spectral variation has not been explained.[126]

In addition to data reduction and classification, PCA can also be used to identify sample outliers. Outliers are describe as unusual samples and if they are not removed analysis of the PCA model can be incorrect. Identified outliers should be removed and replaced with data of the same type. The first step to identify outliers is to survey the raw data. Samples that contain cosmic rays or any other artefacts should not be included in the data set used to build a PCA model. Furthermore, cosmic rays can be removed with most software and after its removal spectra can be used. Hotelling's T^2 is an automated process of looking for outliers in scores plots.[126] The Q statistic measures the difference between the original data and the data reconstructed on the basis of the calibration model. They are used to examine how each sample conforms to the model. High Q values indicate samples have large out of model residuals. A plot of Q residuals and Hotelling's T^2 values are normally used to evaluate outliers.[127]

PCA has been used to evaluate information from carbohydrate arrays. PCA was applied to a ToF-SIMS study which investigated ways to immobilize saccharides onto metal surfaces with correct orientation. ToF-SIMS was used to monitor the immobilization of the saccharides in a stepwise manner. PCA applied to the data collected from ToF-SIMS allowed for the identification of key peaks that characterized the saccharide samples and contributed to differences between the surface chemistries. Scores and loadings plots were generated. The scores plot explained the relationship between the spectra in the collected data sets with each point on the plot. The loadings

plot identified the relationship between the PCs and the peaks in the ToF-SIMS data set. This information was used to identify which peaks were related to each sample types. Using PCA they were able to identify regions of spectral similarity and distinguish between similar chemical fragments. Combining PCA with ToF-SIMS researchers were able confirm immobilization of the saccharides as well as cell material interactions. This shows how beneficial it is to use carbohydrate microarrays combined with PCA to gather information about trends observed in data sets.[128]

Linear disciminant analysis (LDA) uses linear discriminant functions to maximize the ratio of between class variance and minimize the ratio of within class variance. It is a data reduction method because it identifies a smaller dimension hyperplane on which data points are projected from the higher dimension and this is done in a manner to achieve class separation. LDA is limited as a classification technique because the number of variables should not exceed the number of objects. This can be remedied by combining LDA with partial least squares (PLS). PLS models a relationship between dependent variables and independent variables; finding components in the data that describe as much as possible the variations within the data set. It does this by maximizing the correlations with values in the dependent variables and minimizing variations that are noisy or not important.[114]

PLS-DA is a classification method that combines partial least squares regression with discriminant analysis. It is based on PLS algorithm and identifies latent variables (LVs) that have a maximum covariance with the Y-variables.[129] LVs are linear combinations of original variables and they are used to identify data variability in SERS spectra. Similar to PCA, PLS-DA calculates scores and loadings, but they are done in a

different manner. Loadings are the linear combinations of variables which determine the LVs and they identify the influence of each variable on each LV. Scores show the projection of samples in the LV hyperspace.

Building a PLS-DA model requires several parts. A class vector is assembled containing memberships of samples the selected classes. Since PLS-DA is a supervised method, class membership is done before the analysis. For SERS, the model will also include the raw spectral data which has been input into an X-block and a Y-block which is a dummy matrix used to assign class membership. The Y-block assigns a 1 or 0 to each sample. One indicates the sample is a member of the class and 0 designates no class membership. The Y-block is a matrix that contains rows for every sample and columns for each class. The loadings of the X-block which is the preprocessed original data are calculated from the scores of the Y-block and the loadings of the Y-block are estimated from the scores of the X-block.

The model is calculated using a PLS2 regression which is calibrated on the Y-block. The model generates estimated values for each sample and class. Class assignments are made based on these estimated values. So samples whose values are close to one will be assigned class member and samples whose values are closer to 0 will not be given class membership. A threshold can also be use to assist with assessing class memberships. Thresholds are calculated using the Bayes theorem. The Bayes theorem attempts to minimize probability error that may occur when assigning samples to classes.[130] If the estimated value for sample is higher than the defined threshold for the class, then the sample will be assigned membership to the class. This also works in the reverse.[127]

The selection of LVs is based on the root mean square error cross validation (RMSECV) curves. RMSECV measures the model's ability to predict samples that were not used to build the model. CV methods divide samples into groups and each group is removed from the training set one at a time and the model is calibrated on the remaining training samples. Then class membership is predicted for the CV group. The model is then calculated based on the selected number of LVs. For PLS-DA the presence of outliers can be assessed by Q redisuls and Hotelling's T^2.

Scores and loadings plots are used to visually interpret the data. The scores plots graphs the calculated values as defined by the selected LVs. Classification of samples can be clearly observed using the PLS-DA scores plot. The loadings plot identifies which variables are important for classification. For SERS analysis, chemical information about the sample can be learned and this knowledge can be applied to make generalizations about similar systems. Models are evaluated based on their overall sensitivity and specificity. Specificity measures the portion of negative samples which are correctly identified as negative. Sensitivity measures the portion of positive samples that are correctly identified as positive.[127]

PLS-DA combined with SERS array technology was used to identify virulence determinants in influenza viruses. An oligonucleotide silver nanorod array was fabricated for rapid and sensitive detection of pathogenicity determinants isolated from pandemic influenza viruses. This method was used to directly identify RNA and genetic mutations in PB1-F2, a protein linked to influenza virulence, without amplification or labeling of the virus. PLS-DA was used to build classification models to discriminate high virulence RNA binding to the substrate. The Venetian Blinds cross validation method was used to

41

evaluate the model. Models were built to identify high and low virulence strains. The PLS-DA scores plots showed a clear separation between low and high virulence samples. The model reported 100% sensitivity and specificity. The current method used to assay for influenza is PCR. It has shown sensitivity around 75%. So in comparison SERS combined with PLS-DA shows much higher sensitivity and this demonstrates a powerful analysis method that can be readily applied to carbohydrate microarray analysis.[131]

The Objectives Of This Dissertation Were:

1. Determine if biotin can be immobilized on gold nanoparticles using strain promoted click chemistry. Then determine if SERS can be used to detect avidin binding to the modified surface.

2. Develop a carbohydrate array where carbohydrates are immobilized to gold nanoparticles using strain promoted click chemistry. Then detect protein binding and determine if SERS can differentiate between different proteins from the same family.

3. Immobilize 2,3 and 2,6 α sialylated carbohydrates to gold nanoparticles using strain promoted click chemistry. Then determine binding specificity of Influenza hemagglutinin to the α sialylated carbohydrates using SERS detection. Finally determine if SERS can differentiate hemagglutinin from different types of influenza viruses.

CHAPTER 2

SURFACE ENHANCED RAMAN SPECTROSCOPY DETECTION OF BIOTIN

AND AVIDIN INTERACTIONS

Introduction

Avidin is a tetrameric protein with a high affinity to biotin. The avidin-biotin molecular recognition complex has been studied extensively and applied to many applications. The concept behind its use is that avidin can recognize biotin in its natural or modified state.[132] Biotin is composed of an uriedo ring fused to a tetrahydrothiophene ring and it contains a valeric acid side chain which may be used for modification purposes. Avidin recognizes the bicyclic ring system of biotin and because avidin is tetrameric it can bind up to four biotin units at once. The reported dissociation constant for avidin biotin binding in a tetrameric assembly is $K_D=10^{-15}$M. [133] This is the highest known affinity in nature between a protein and a ligand.[134]

Because of the valeric acid side chain, biotin can be modified in a variety of ways. Through this modification, biotin can be immobilized to nanoparticles to create biosensors. Biotin can be modified by N-hydroxysuccinimide (NHS) esters to target amine groups. Variations of NHS esters that have been used for such linkage are sulfo-NHS and tetrafluorophenyl (TFP)-NHS. Biotin derivatives are also modified with amine groups directly. Biotin can also be modified to target sulfhydryl groups. This requires

biotin to be modified with malemide and pyridyl disulfide groups. An alternative method to immobilize biotin is to use strain promoted click chemistry reactions.

Therefore, to develop biotin nanoparticle sensors strain promoted azide alkyne click chemistry reactions were used to attach biotin derivatives to gold nanoparticles. Azide alkyne cycloaddition reactions are reversible, give high yields, stereospecific, produce no offensive by products and the incorporation of azide moieties into compounds is relatively simple. For this study azide linkers were synthesized containing an eleven carbon chain, PEG groups to reduce nonspecific protein binding, and disulfide moieties for attachment to gold nanoparticles. The hydrophobic alkane region encourages monolayer formation and the PEG moiety facilitates compatibility with an aqueous environment. After immobilization of the azide linker a biotinlyated DIBO was used to click to the azide groups on the nanoparticles. The ring strain generated using BDIBO is enough to catalyze the reaction between the azide and alkyne group without the need for a copper catalyst. After immobilization of biotin, the biosensing capability of the modified nanoparticles was tested to see if they could detect avidin.

The formation of the biotinlyated sensors and incubation with avidin protein was detected by UV-Vis analysis and surface enhanced Raman spectroscopy. SERS is as a powerful tool for chemical and biological sensing applications. This technique can be used to determine molecular structural information and provides ultrasensitive detection limits, which includes single molecule sensitivity.[135] Current applications of SERS relevant to this proposal include DNA and bacteria biosensing, detection of processes within living cells, and detection of different viral strains. SERS has also been used to monitor click chemistry reactions. In a study by Yoo et al., SERS was able to monitor in

real time the formation of a 1, 2, 3-triazole from the Cu[I] catalyzed azide alkyne cycloaddition.[136] Based on these bodies of work it is expected that SERS will be a suitable fit to detect biotin avidin binding interactions.

Raman and SERS studies have been used to investigate biotin avidin interactions. These studies relied on common protein bands that have high Raman cross sections to detect avidin. Tryptophan (Trp) residues of avidin are implicated in as being important for binding to biotin.[137] Studies have shown that Trp is protected by biotin from oxidation and removal of Trp renders the protein inactive. Raman spectra of the biotin-avidin complex yielded spectral bands of amino acids: Trp, phenylalanine (Phe), and tyrosine (Tyr), Lysine (Lys), threonine (Thr), and alanine (Ala). More protein bands observed in SERS spectra were Amide I and Amide III vibrations.[138,139] Therefore, using these studies as a guide, we were be able to see similar peaks using our method. To examine SERS's ability to extract structural information of avidin biotin binding hierarchical clustering analysis (HCA) and partial least squares analysis (PLS-DA) methods were used for chemometric data analysis.

Experimental Methods

Materials. Highest Grade V1 mica sheets and gold nanoparticles with a diameter of 60nm were purchased from Ted Pella. ACS grade ethanol, phosphate buffer saline (pH=7.4), and (11-Mercaptoundecyl)tetra(ethylene glycol) were purchased from Sigma Aldrich. Dimethylformamide was purchased from J.T. Baker and sterile water was purchased from Braun. Gold (99.999%) and Chromium (99.998%) evaporation pellets were

purchased from Kurt J. Lesker. Bovine Serum Albumin (BSA) was purchased from Thermo Scientific. Avidin-FITC was purchased from Invitrogen.

Preparation of Substrate. The substrate used for this study was a gold covered mica sheet. A Thermionics vacuum evaporator was used to prepare gold substrates. Highest Grade V1 mica sheets were cleaved using tweezers. The substrates were then placed into the Thermionics vacuum evaporator chamber. Ten nanometers of chromium was evaporated onto the freshly cleaved mica surface to provide an adhesion layer for gold. Then 300nm of gold was evaporated onto mica sheets at a rate of 1 Å/s. After thermal evaporation, substrates were flame annealed and cut into 1x1cm chips.

Preparation of Clicked Biotin Nanoparticles. Structures of analytes used in this study are shown in Figure 2.1.
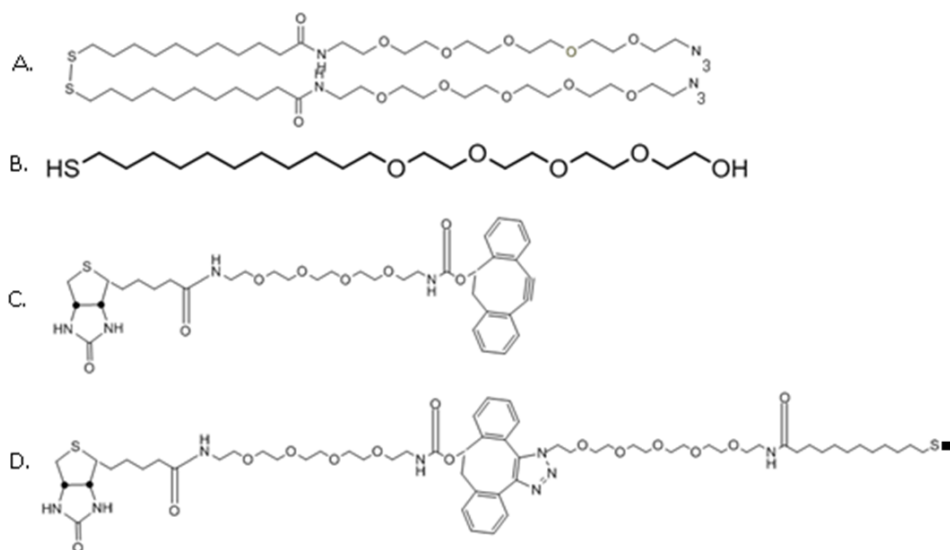
Figure 2.1. Analytes used in biotin –avidin studies. (A) N_3AT, (B) MEG, (C) BDIBO ,(D)Clicked Product.

Gold nanoparticles were functionalized with azides by placing them in a solution of the

azide (N_3AT) and spacer (MEG), which were dissolved in DMF, for 12h under stirring.

To ascertain the best binding condition MEG/N_3AT solutions were prepared in different

ratios. The ratios were 4:2, 6:2, and 8:2. After the binding time was completed, the gold

nanoparticle solution was purified by centrifugation to remove any excess MEG/N_3AT

that did not bind to the gold nanoparticle. Three rounds of centrifugation were done to

ensure all excess MEG/N_3AT had been washed away. Between each round, the

nanoparticle supernatant was removed and fresh nanopure water obtained from a

Barnstead Nanopure purification system was added. Centrifugation formed the

nanoparticles into a pellet so after the last cycle, the particles were sonicated to break up

the pellets. The MEG/N_3AT solution was divided into two halves. One half was kept for

SERS analysis and the second half was incubated with a 2mM solution of BDIBO. These

samples were reacted overnight under constant shaking. Following the overnight period,

this biotinylated gold nanoparticles were purified by centrifugation. The particles were

resuspended into nanopure water obtained from a Barnstead Nanopure water purification

system. SERS samples were prepared by removing 30µl of the biotinylated nanoparticles

from the solution. The particles were dropped onto a gold mica substrate, allowed to dry

overnight and analyzed by SERS.

Avidin-Biotin Binding Studies. All protein binding experiments were done using the

biotinylated nanoparticles that had been prepared with a MEG/N_3AT ratio of 8:2. Avidin-

FITC protein solution was prepared in a PBS buffer to a final stock concentration of

50µg/ml. For the timed binding experiments three centrifuge tubes were filled with 500µl

of the biotinylated nanoparticle solution. To the tubes were added 500µl of PBS buffer

and 250 µl of the avidin protein. The final concentration of the avidin protein in these solutions was 12.5µg/ml. The vials were wrapped into aluminum foil and placed into an incubator which had a temperature of 4ºC. Avidin was incubated with the biotinylated particles for 15mins. After this period, the tubes were purified by centrifugation. They were washed two times with buffer solution and a final wash with Braun water to remove salts from PBS from the solution. Samples were later sonicated to resuspend the particles in water. For SERS analysis 30µl of the avidin-biotin gold nanoparticle solution was removed from the centrifuge tubes and dropped onto the gold mica substrate and allowed to dry overnight.

In addition to avidin, neutravidin was also used in experiments. Neutravidin-FITC was purchased from Invitrogen. Neutravidin is a modified form avidin that does not contain carbohydrates. Therefore, it has lower nonspecific binding properties.[140] Binding experiments for neutravidin were prepared in a similar manner in comparison to the avidin samples. The same binding time of 15 mins was used and SERS, fluorescence, and UV-Vis response was measured.

UV-Vis Analysis. UV-visible absorption spectra were recorded on a Beckman Coulter DU800UV/Visible spectrophotometer using a 1cm path length quartz cuvette.

Fluorescence Measurements. Samples were spotted into a 96 well microplate from Greiner Bio-One. Fluorescence measurements were recorded on a POLARstat OPTIMA multidetection microplate reader. FITC has an excitation wavelength of 495 and an emission wavelength of 519.

SERS Measurements. SERS measurements were obtained using a Renishaw inVia confocal Raman microscope system equipped with a 785nm near-infrared diode laser as

the excitation source. The incident laser beam was delivered to the sample through a 20x

(NA=0.40) objective onto an automated sample stage. The laser spot is ~4.8 x 27.8µm.

The laser power was set to 1%, where the power at the sample surface was measured to

be ~3.5mW. Thirty microliters of each sample were spotted onto gold covered mica

substrates. Five-ten spectra were recorded from different locations on the substrate.

Acquisition settings were 5 accumulations for 10s each. The spectral range recorded was

between 2300 and 350 cm^{-1}.

Data Analysis. SERS spectra were averaged then baseline corrected using GRAMS32/A1

Version 6.0 spectral software package. Raw SERS spectra were pre-processed for

chemometric analysis in the following manner. Raw SERS spectra were derivatized using

the Savistky-Golay algorithm. The first derivative of each spectrum was taken using a

15point, 2nd order polynomial followed by normalization to unit vector length with

Matlab software. This allowed direct intensity comparison between samples and between

variables within a sample. HCA was carried out using the median agglomerative method

in PLS Toolbox 4.0 (Eigenvector, Wenatchee, WA). PLS-DA models were built to

classify samples as well. The robustness of the models created by PLS-DA was tested by

cross-validation (CV) with Venetian Blinds.

Results and Discussion

Determination of a Suitable Ratio of MEG and N$_3$AT. A suitable ratio of MEG and N$_3$AT

was determined to ensure clicking would take place. In the present scheme, MEG serves

as a spacer compound to dilute the number of azides on the surface. This is needed

because, if each of the self-assembling compounds contains an azide moiety, the

recognition and binding of BDIBO will less likely occur. Densely packed monolayer of all azides would block each other, preventing the binding of BDIBO.[141]

Molar ratios of MEG:N$_3$AT used were 8:2, 6:2, and 4:2. SERS spectra were taken of MEG followed by the mixed ratios of MEG:N$_3$AT. The spectra can be seen in the Figure 2.2.
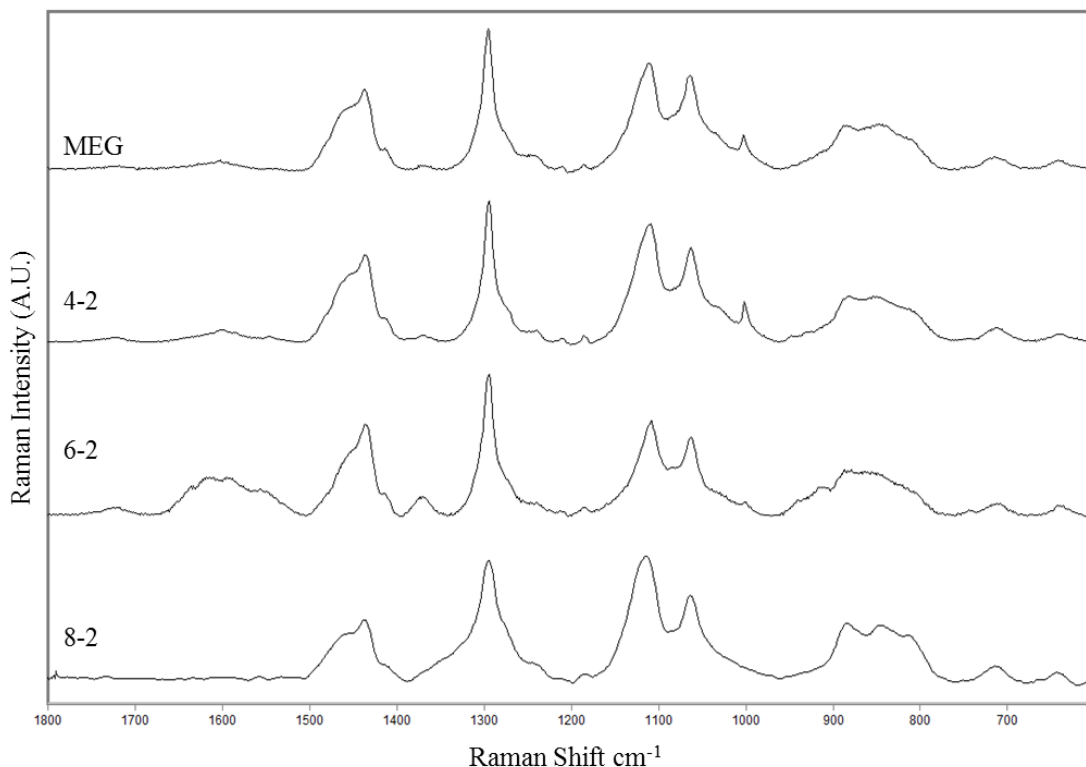
Figure 2.2 SERS spectra of MEG and different molar of MEG:N3AT.

MEG and N$_3$AT have similar structures. Therefore, it is not surprising that spectra MEG and MEG:N$_3$AT are similar. Although MEG and N$_3$AT have similar structures, because MEG is present in a higher ratio, its spectrum dominates. Assignments can be seen in Table 2.1. Azides generally have a band around 2200cm^{-1} however, it appears weak in Raman spectrum. Attachment to the nanoparticles is evidenced by the (C-S) bands at 638

and 712cm^{-1}. Furthermore, not shown in the spectra is the absence of the (-SH) band at 2500cm^{-1} and the (S-S) band ~500 indicate attachment of the linkers to the gold nanoparticle.

Table 2.1. Assignment of the SERS bands in the spectra of MEG and MEG:N3AT.

Wavenumber cm^{-1}	Assignment
638	v(C-S)G
712	v(C-S)T
803	CH$_2$ rocking
843	CH$_2$ rocking
885	CH$_2$ rocking
1000	v(C-C)
1062	v(C-O) stretching
1114	CH$_2$ twisting
1297	v(C-O) stretching
1438	v(C-H) stretching
1457	v(C-H) stretching

Biotin was clicked onto the functionalized surface using BDIBO. To understand what features can be attributed to the BDIBO moiety, a SERS spectrum was taken of this compound and is shown Figure 2.3. It should be noted, that the orientation of BDIBO as a free compound is different from how it would be once it is clicked onto the functionalized azide surface. Therefore, spectral features will appear differently once
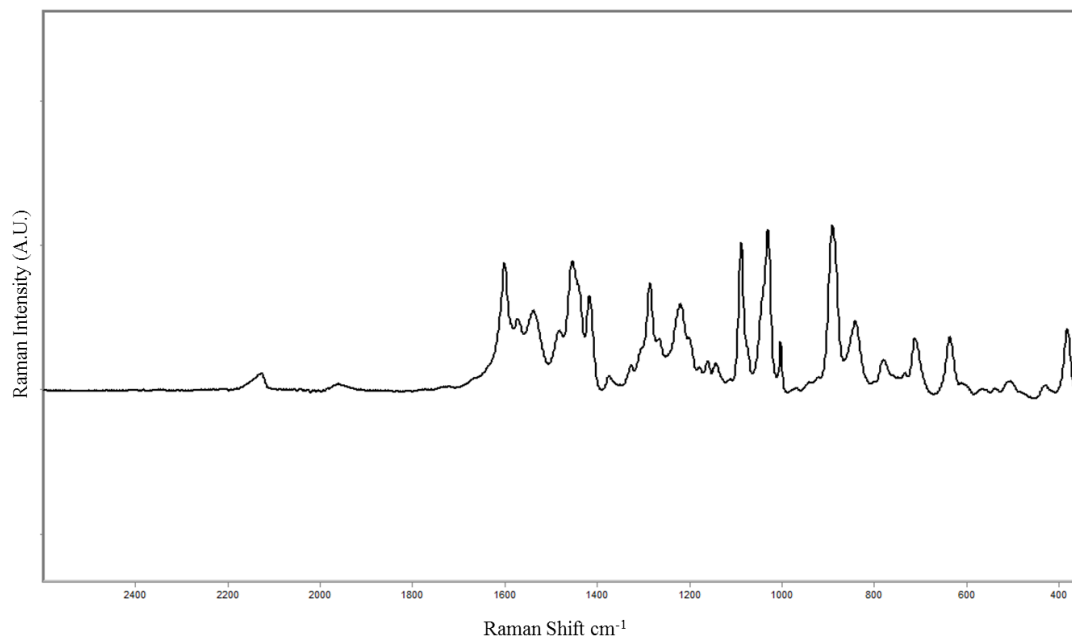
Figure 2.3. SERS spectrum of BDIBO.

BDIBO is clicked to the azide surface. However, there are some spectral characteristics that will be similar to both spectral orientations.

Characteristic spectral features that are of interest reside in the region between 1900-2200 wavenumber (cm^{-1}). These bands can be attributed to ($\equiv$C-C) vibrations. These peaks for ($\equiv$C-C) will disappear when BDIBO clicks to the azide functionalized surface.

The SERS spectra of the clicked biotin product are shown in Figure 2.4. Examination of the clicked biotin spectra reveals several things. First, in the spectra for

the ratios of 4:1 and 8:1, there is an appearance of bands around 679, 744, 951, 1021, 1563, and 1612 cm^{-1}. These bands can be assigned to the BDIBO compound. In these spectra, evidence that clicking occurred between the BDIBO and azides can be seen the disappearance of the bands for ($\equiv$C-C)vibrations and the appearance of bands for the formation of triazole around 1340 and 1530 cm^{-1}.
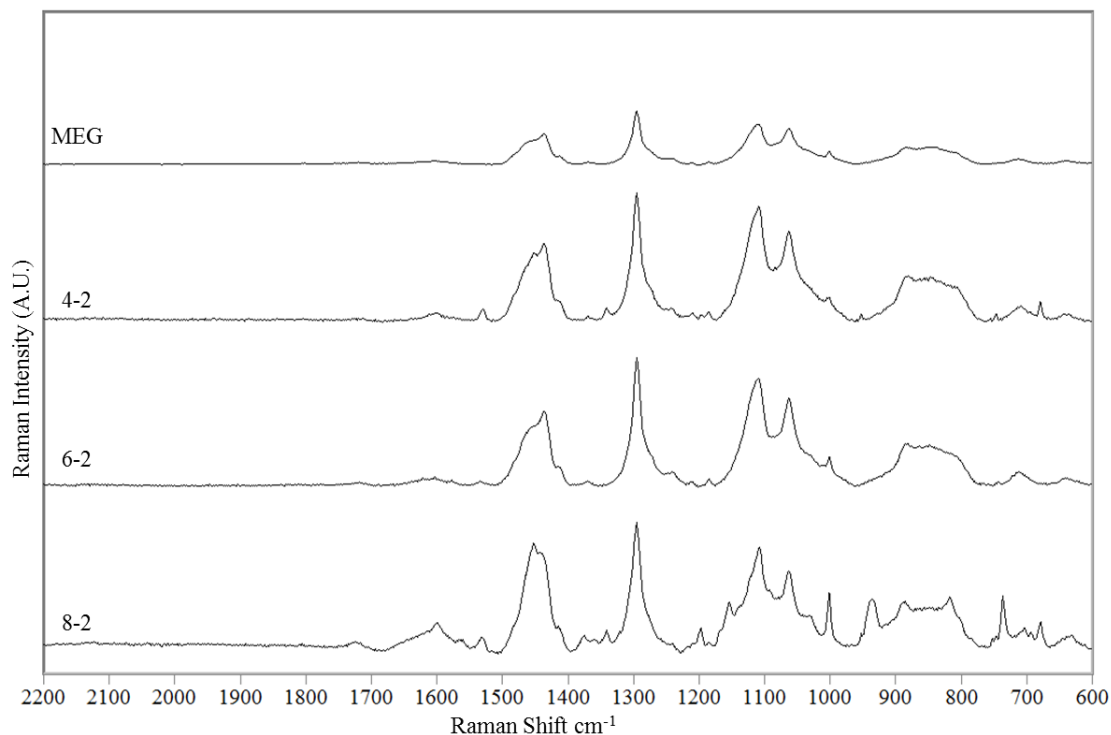
Figure 2.4. SERS spectra of the clicked biotinylated gold nanoparticles.

the disappearance of the bands for ($\equiv$C-C)vibrations and the appearance of bands for the formation of triazole around 1340 and 1530 cm-1. It should be noted that there is no evidence that clicking of BDIBO occurred in the ratio of 6:1.

The spectrum contains ($\equiv$C-C) vibrations and no bands for the BDIBO compound appear. Spectral assignment can be seen in Table 2.2.

To ensure that biotin was clicked onto the surface, our collaborators synthesized a model compound to simulate the clicked product between BDIBO and azide. This compound was prepared in an ethanolic solution and incubated with the gold nanoparticles. The spectrum for this compound is shown in Fig 2.5 and it is compared with the biotinylated spectra of 8:2, 6:2, and 4:2. Although there are dominant bands from the MEG group, comparison of the four spectra show similar bands. Furthermore,

Table 2.2. Assignment of the SERS bands in the spectra of the clicked biotin complex.

Wavenumber cm^{-1}	Assignment
641	MEG
679	BDIBO
709	MEG
744	BDIBO
850	MEG
887	MEG
950	BDIBO
1002	MEG
1021	BDIBO
1061	MEG
1103	MEG
1296	MEG
1335	Triazole
1436	MEG
1448	MEG
1529	Triazole
1563	BDIBO
1612	BDIBO

the bands denoted with asterisks identify bands for biotin. Based on these results the ratio of 8:2 was chosen for further experiments.

The biotinylation process for ratio of 8:2 samples were also analyzed by UV-Vis spectroscopy. The spectra can be seen in Figure 2.6. UV-Vis also allowed the derivation process to be followed step by step. Spectra were taken of the bare gold nanoparticles,
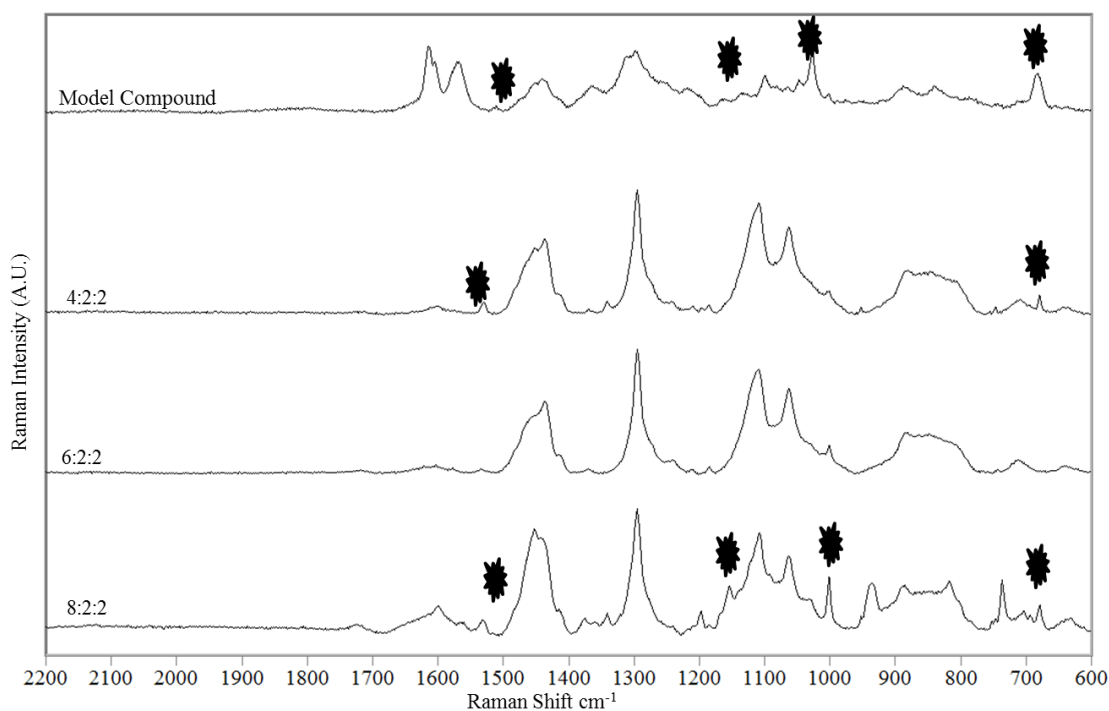
Figure 2.5. SERS spectra of the model compound and clicked products.

derivatization with MEG:N_3AT, addition of BDIBO. Upon addition of MEG:N_3AT to the gold nanoparticles, there is a slight red shift in the surface plasmon of the nanoparticle.

This shift indicates that the nanoparticle has been derivatized in some manner. These results go along with previous studies.[142] When biotynlation occurs, the peak is broadened and shortened.

Avidin Capture by Clicked Biotin Nanoparticles. Avidin-FITC was used to examine the sensor capability of the clicked biotin nanoparticles. Avidin[12.5μg/ml] was incubated with the nanoparticles for 15mins. The addition of avidin caused the nanoparticles to aggregate and in Figure 2.7 of the UV-Vis spectrum there is significant broadening of the
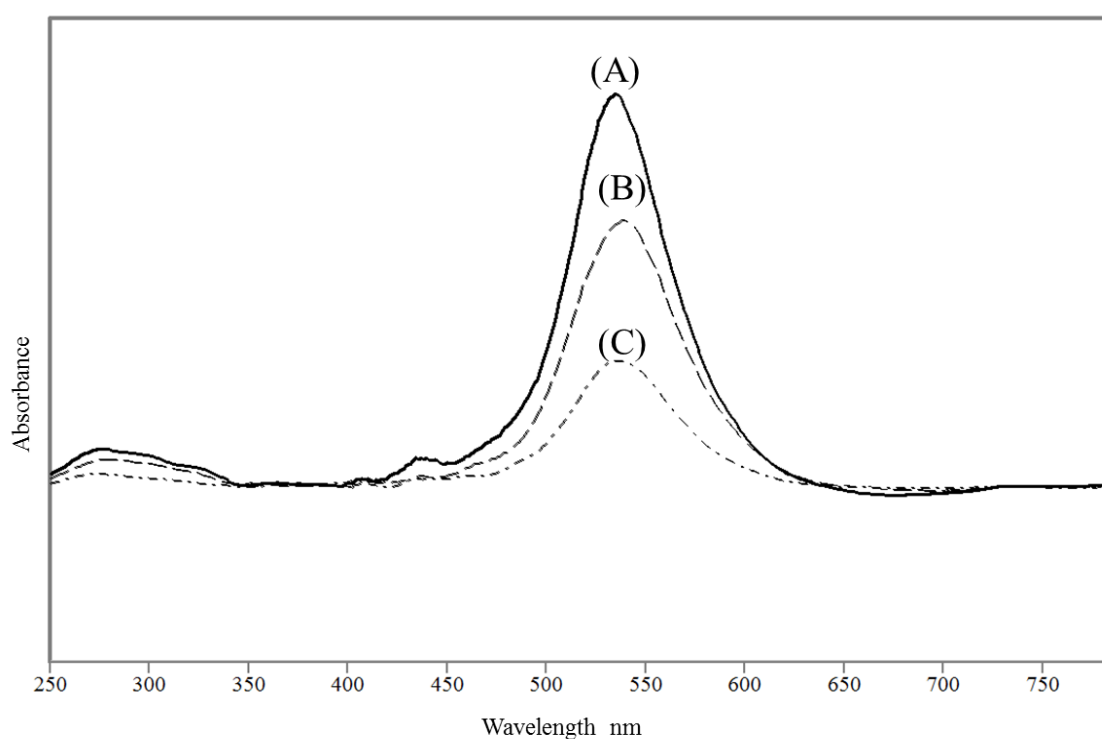
Figure 2.6. UV-Vis spectra Biotinylation process of gold nanoparticles. (A)Bare GNP (B) MEG/N$_3$AT, (C) biotinylated MEG/N$_3$AT, (D) Avidin binding to biotinylated nanoparticles.
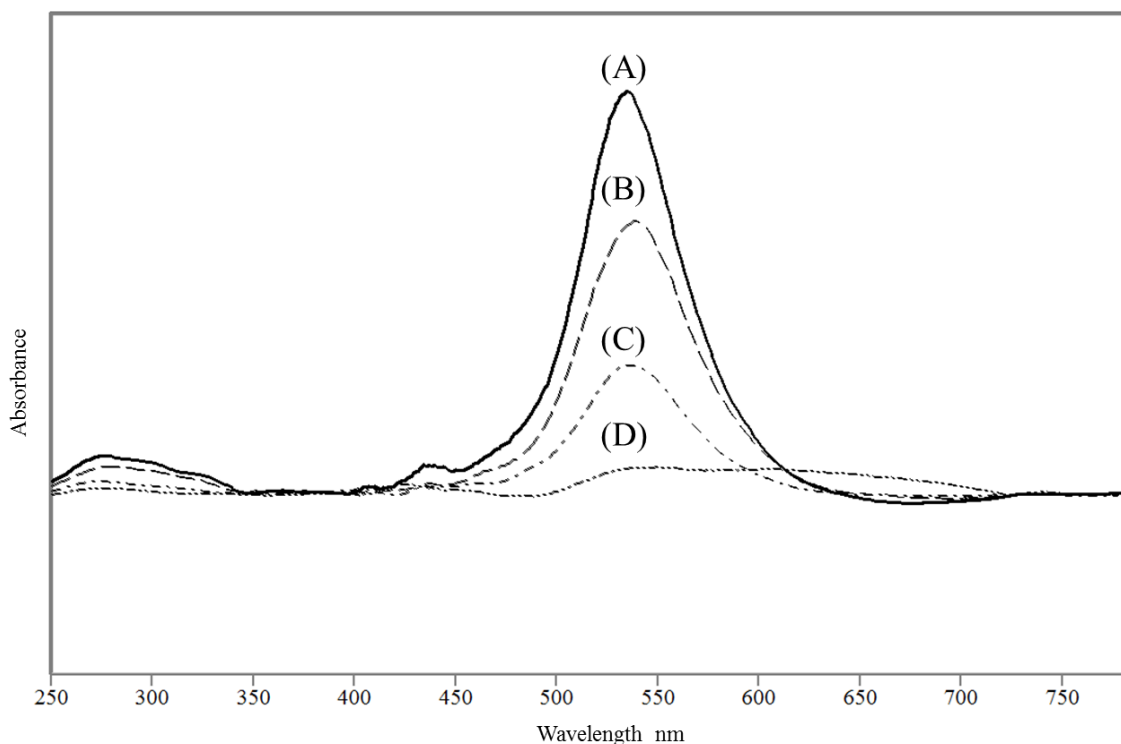
Figure 2.7. UV-Vis spectra of avidin timed Studies. (A)Bare GNP (B) MEG/N$_3$AT, (C) biotinylated MEG/N$_3$AT, (D) Avidin binding to biotinylated nanoparticles.

surface plasmon of the nanoparticle. This is expected because avidin has up to four

binding sights for biotin.Therefore, crosslinking between biotinylated nanoparticles can

occur as has been previously reported.[143] UV- VIS spectra were also taken for the

neutravidin samples and they are shown in Figure 2.8. A similar trend is observed when

neutravidin is added. The surface plasmon for the nanoparticles dampens and broadens.

This is indicative of the protein binding to the nanoparticles.

Since the avidin and neutravidin purchased contain the fluorescent label FITC at

a ratio of 1:1, the concentration of avidin captured by the biotinylated nanoparticles can

be estimated from a standard fluorescence curve. A standard fluorescence curve of

avidin-FITC was generated to estimate the amount of avidin and neutravidin captured
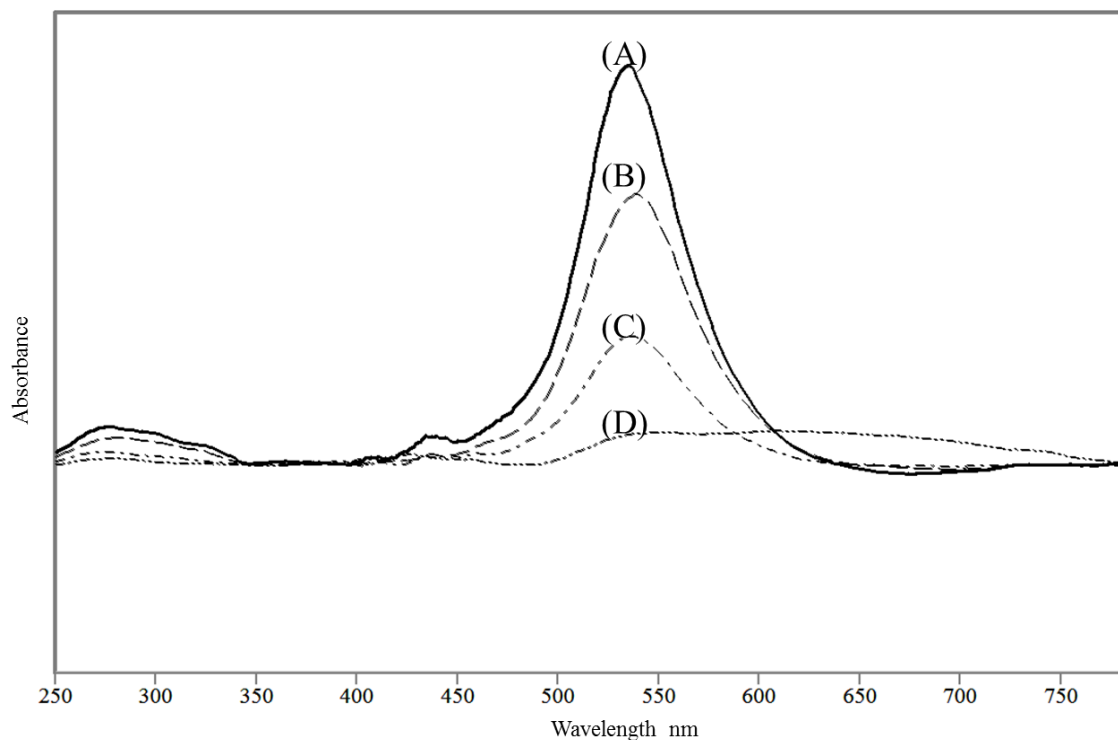
Figure 2.8. UV-Vis spectra of neutravidin timed studies. A)Bare GNP (B) MEG/N$_3$AT, (C) biotinylated MEG/N$_3$AT, (D) Neutravidin binding to biotinylated nanoparticles.

Table 2.3. Fluorescence intensity values for protein studies with estimated concentration.

Sample	Fluorescence Intensity	Estimated Protein Concentration (μM)
Biotin- Avidin	572	1.56E-2
MEG-Avidin	105	9.82E-3
Biotin-Neutravidin	316	1.39E-2
Biotin-BSA-Avidin	208	1.11E-2
MEG/N$_3$AT-BSA-Avidin	74	9.44E-3

onto the nanoparticles. A table detailing the fluorescence intensity values for the biotinylated nanoparticles and the standard curve is presented in Table 2.3. It is well known that a major problem with avidin is high nonspecific binding. This problem can be attributed to presence of sugars mannose and N-acetylglucosamine on the backbone structure of avidin and the high isoelectric point of avidin.[144] Therefore, nanoparticles containing only MEG were also reacted with avidin to examine nonspecific binding of avidin. The values of the fluorescence response and estimated concentration are also given in Table 2.3. The table shows that as expected, when biotin is present, the fluorescence response for avidin is enhanced. The fluorescence response recorded when avidin is present is over 5Xs greater when only MEG is present. Also, the fluorescence response for the MEG nanoparticles is over 5Xs higher than the control, PBS. This is due to the high nonspecific biding of avidin.

In addition to avidin, neutravidin was also used in similar binding experiments. Neutravidin is a modified form avidin that does not contain carbohydrates. The binding affinity of neutravidin to biotin is less than avidin, so a decrease in fluorescence intensity was expected.[140] The fluorescence responses and estimated concentration for neutravidin are shown in Table 2.3. It should be noted that the avidin and neutravidin concentrations and binding times are the same for both binding experiments. Comparison of the fluorescence response values for avidin and neutravidin indicate lower values for neutravidin.

One final fluorescence experiment included the use of BSA as an agent to block avidin from binding to the surface of the gold nanoparticles. BSA has a similar molecular

weight to avidin, no specific reaction with biotin, and is commonly used to block nonspecific interactions in ELISA procedures.[145] A solution of BSA was added to the nanoparticles after biotinylation for 15mins at 4°C. After purification by centrifugation, avidin was then added to the nanoparticles for 15mins. After this time the samples were washed using the same procedure in the methods section. The fluorescence response for the biotin-bsa-avidin and MEG:N_3AT-BSA-avidin nanoparticles are 208 and 74 respectively. These values are significantly lower than the biotin-avidin nanoparticles and MEG-BSA-avidin.

Further characterization of the derivatized nanoparticles was done by SERS. To begin, a SERS spectrum of avidin was taken as a reference for spectral assignments. This spectrum is seen in Figure 2.9.
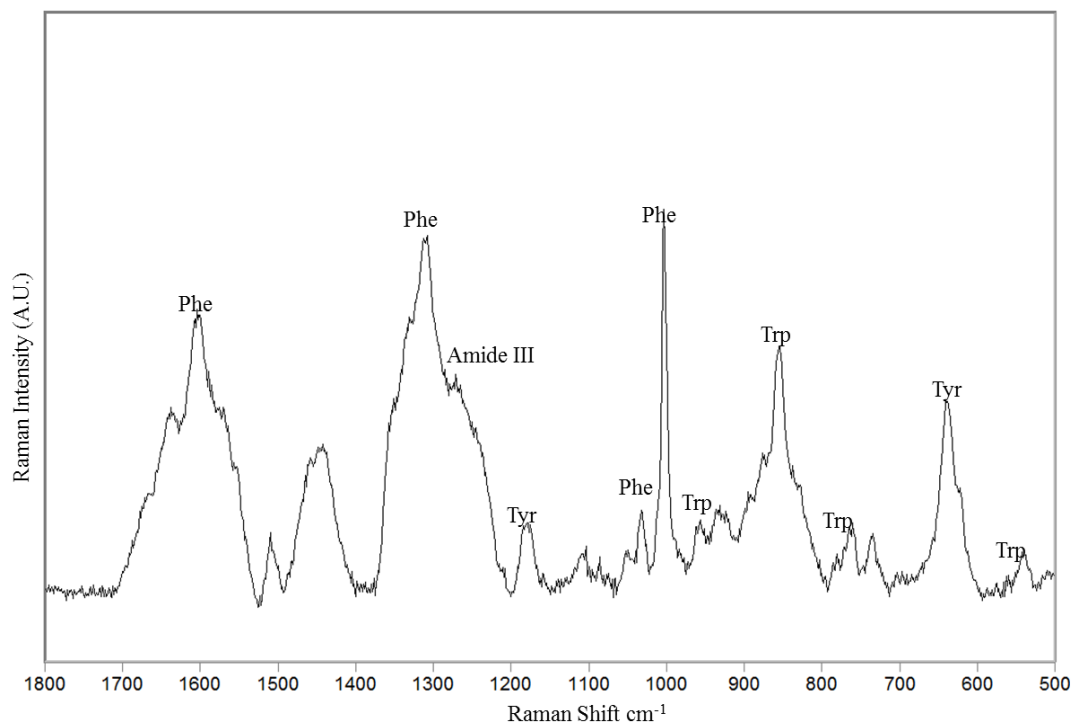
Figure 2.9. SERS spectrum of Avidin.

SERS spectra of biotinylated nanoparticles incubated with avidin, neutravidin, and MEG-avidin were taken as well. The averaged spectra are shown in the Figures 2.10 and 2.11. Several spectral comparisons were examined. The avidin bands which appear in the spectra are denoted by astericks.[146]

A number of observations can be made from the spectra presented. The SERS bands for avidin appear to be weak in the spectra. Previous studies have shown that the avidin signals can produce a very high background.[147] One reason for this observation is due to the small amount of avidin captured onto the biotinylated nanoparticles. The estimated concentration of the averaged biotin-avidin, biotin-neutravidin, and biotin-MEG is 15.6nM, 13.9nm, and 9.82nm respectively. Since the MEG and MEG:N_3AT concentrations are much higher, they mask some of the avidin signals.
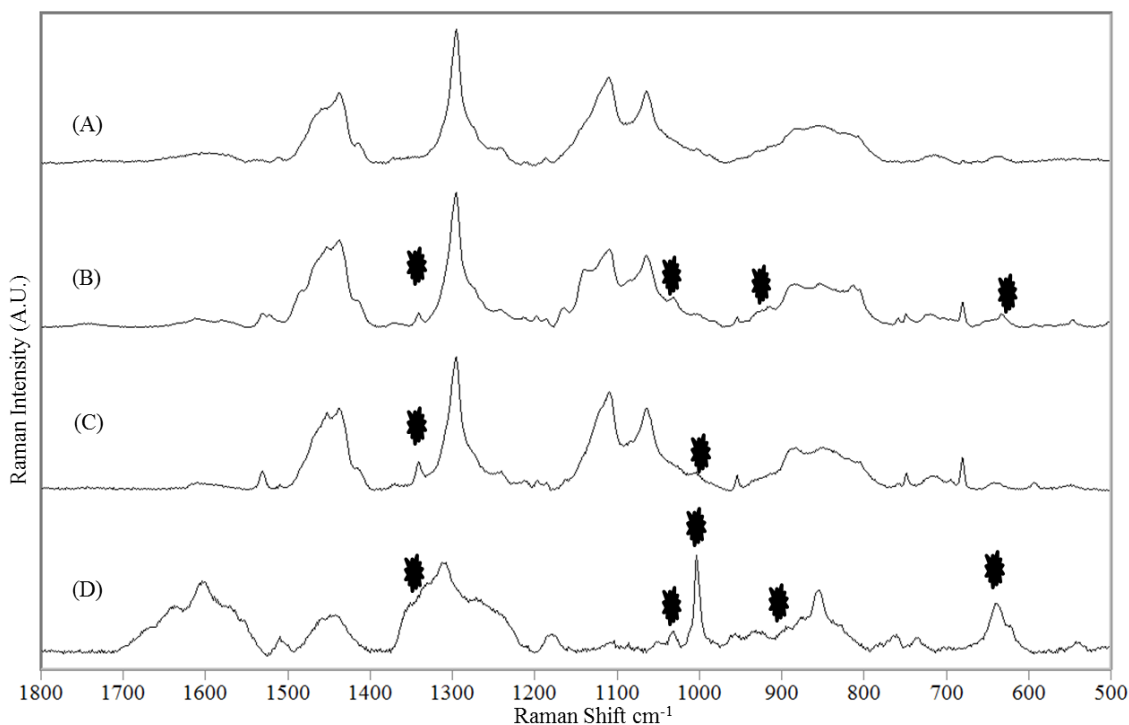
Figure 2.10. SERS spectrum of avidin and nentravidin binding. (A) Biotinylated nanoparticles, (B) Biotinylated nanoparticles + avidin , (C) Biotinylated nanoparticles + neutravidin, (D) avidin.
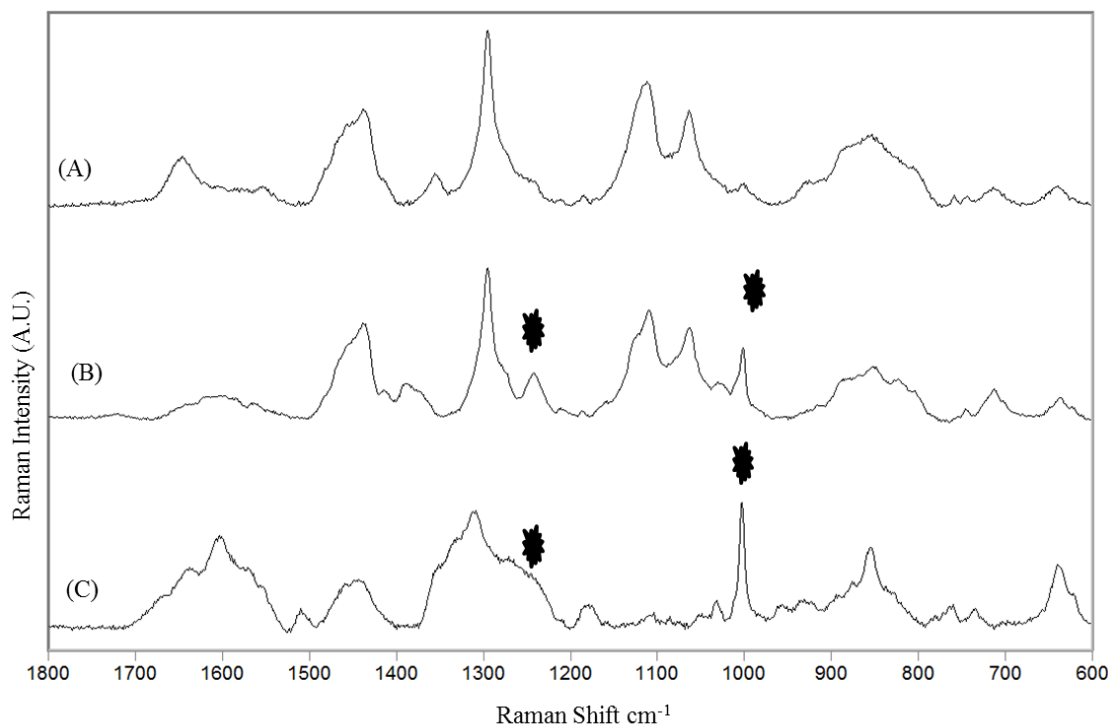
Figure 2.11. SERS spectrum of avidin and nonbiotinylated nanoparticles. (A) MEG nanoparticles, (B) MEG nanoparticles + avidin , (C) avidin.

Another explanation for the low SERS signal for Avidin may be explained by the SERS effect. Two mechanisms that are used to explain the SERS effect are charge transfer and electromagnetism. The electromagnetic mechanism occurs when the electromagnetic field is enhanced by a plasmon excitation through the incident laser on metal surfaces.[148] Charge transfer occurs through a charge transfer process between the adsorbed analyte and the metal surface.[79] These methods work simultaneously to contribute to the overall enhancement. A requirement for the charge transfer effect is that the molecules are in direct with the surface of the gold nanoparticles. Therefore, the first monolayer which consists of MEG or MEG:N_3AT will have higher enhancements than the clicked biotin and avidin, which are further from the surface.[149] It has been shown

that electromagnetic enhancements decrease exponentially with increasing chain length for groups chemically bound to roughened surfaces.[150] Hence, molecules closer to the surface will experience greater electromagnetic enhancements than those further away.

These observations can be seen in several ways be examining the SERS spectra. The avidin spectrum shows well defined spectral bands. This sample was prepared by adding avidin to bare gold nanoparticles. Hence, the avidin molecules are in direct contact with the gold nanoparticles, therefore avidin bands are greatly enhanced. If the spacer is added first and then avidin, there is a noticeable reduction in bands for avidin, which is due to its signal being masked by the MEG. In the spectra where avidin is added after biotynlation, there is a significant reduction in the signals for avidin. Avidin bands in the SERS spectra of the BSA coated nanoparticles appear weaker than the spectra where BSA is not used. This suggests that BSA was able block additional binding sights on the gold nanoparticles, therefore promoting further biding to biotin. The estimated concentration for MEG/N$_3$AT nanoparticles + BSA + avidin and biotinylated nanoparticles + BSA + avidin was 11.1nM and 9.44nM respectively. SERS spectra for these samples can be seen in Figure 2.12. Avidin bands in the spectra are indicated with astericks.
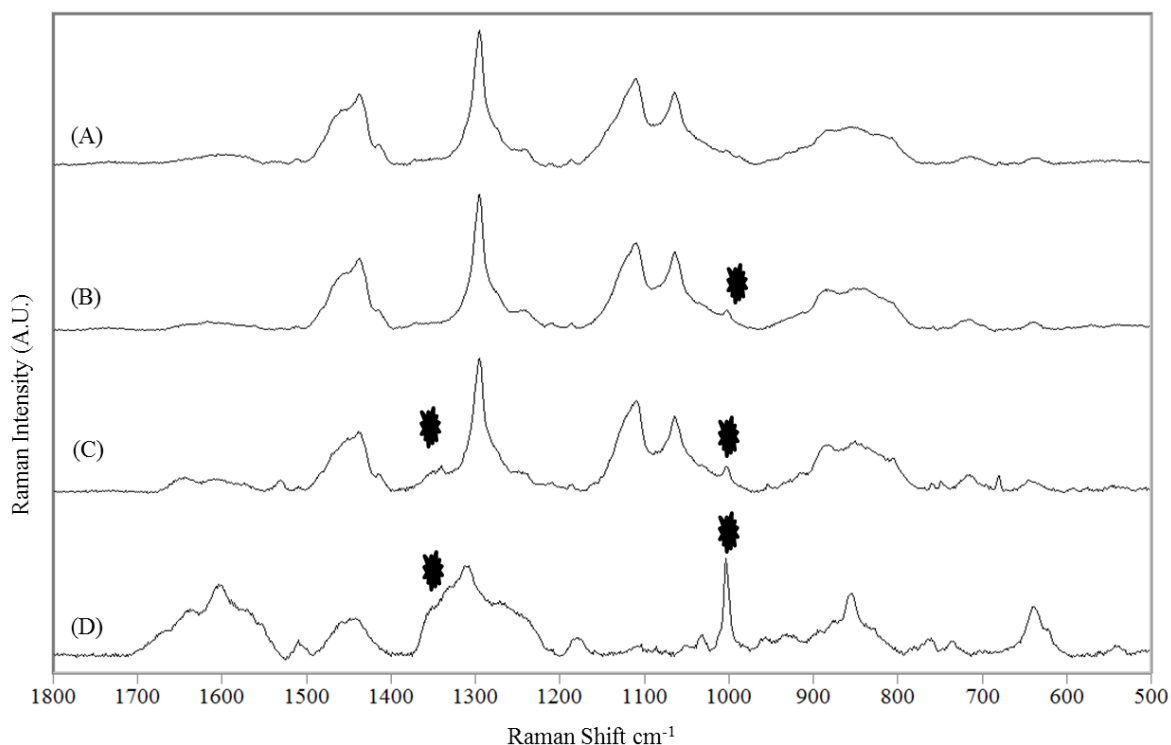
Figure 2.12. SERS spectrum of avidin and nentravidin binding. (A) Biotinylated nanoparticles, (B) MEG/N$_3$AT nanoparticles + BSA + avidin, (C) Biotinylated nanoparticles + BSA + avidin , (D) avidin.

Chemometric Analysis. Since the SERS signals of avidin appear very weakly in the SERS spectra, chemometric methods were used for classification purposes. Classification tools used were HCA and PLS-DA. HCA is an unsupervised statistical method for analysis which groups objects which have close distance to one another.[102] HCA is used to reduce dataset dimensionality, reduce noise, and cluster similar spectra into groups for classification. HCA was used to cluster data from the biotinylation of the gold nanoparticles followed by incubation with avidin. Because new chemical components are added to the nanoparticles at each step, the samples should cluster according to the new components added. The analysis contained SERS spectra of MEG:N$_3$AT,

MEG:N$_3$AT:BDIBO, and MEG:N$_3$AT:BDIBO-Avidin. The results of the HCA analysis are shown in Figure 2.13.
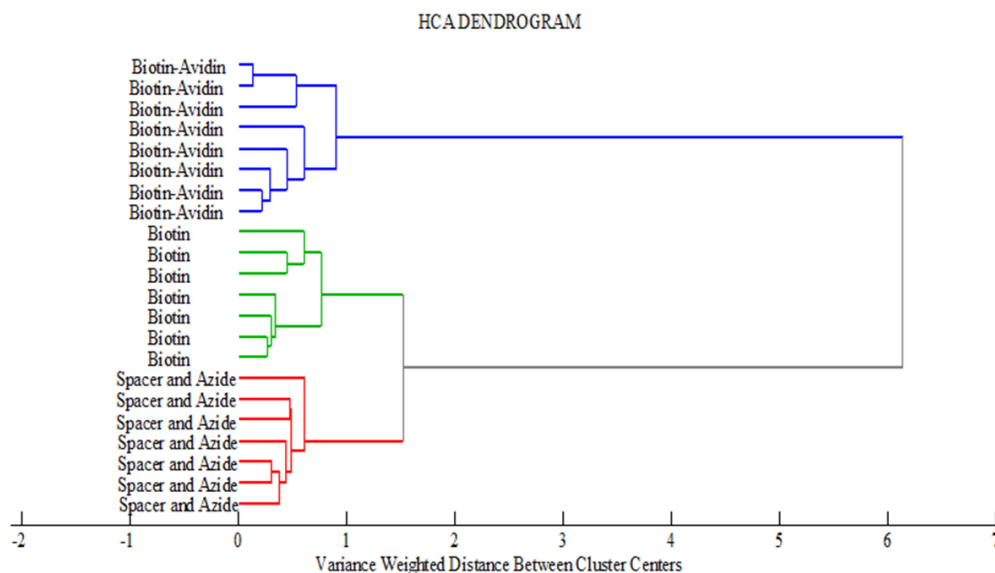
Figure 2.13 HCA dendrogram of the biotinylation process and avidin binding.

The dendogram shows clear separation between the classes. This indicates there is enough spectral differences for the samples to cluster together. Since this HCA is an unsupervised method of analysis class membership was not established beforehand.

Further analysis was done by PLS-DA, a supervised statistical method where prior knowledge of the classes is used for classification.[102] This method minimizes the contribution of spectral features which vary within a particular class and maximizes the contribution of spectral features which vary among the different classes. The first model was built to analyze the biotinylation and avidin incubation process. The model built

contained SERS spectra of MEG:N$_3$AT, MEG:N$_3$AT:BDIBO, and MEG:N$_3$AT:BDIBO-Avidin. The model built is shown in Figure 2.14.
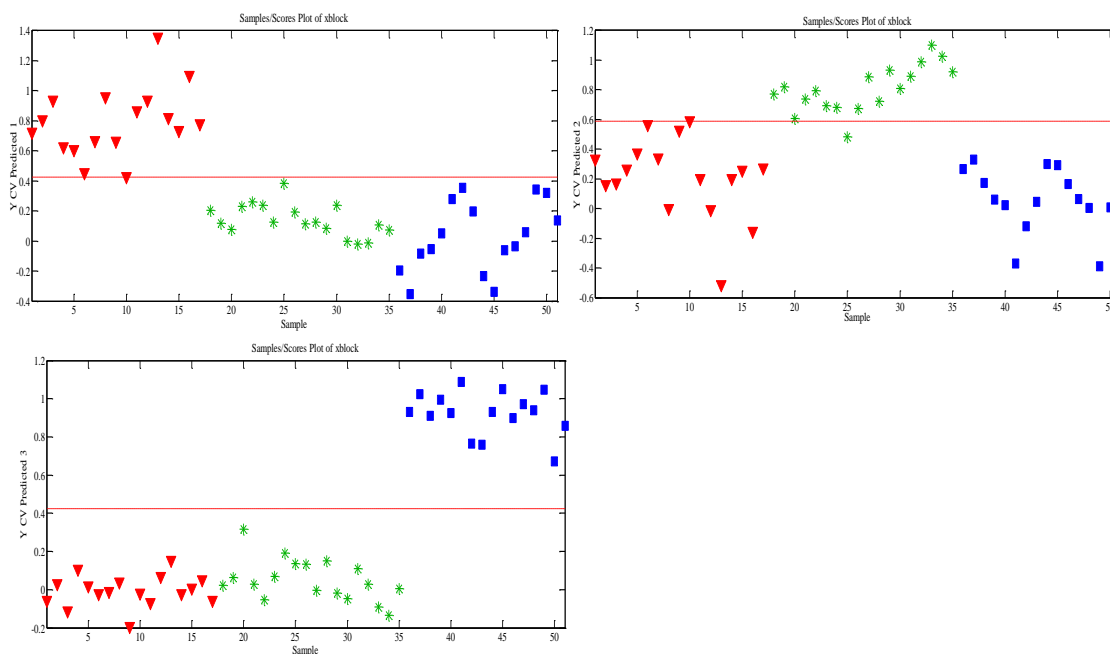
Figure 2.14 PLS-DA scores plot of the biotinylation process and avidin binding. Classes used to build the model were MEG:N$_3$AT ▼ MEG:N$_3$AT:BDIBO ✳ ,and MEG:N$_3$AT:BDIBO-Avidin ■ .

The model was able to correctly classify samples as belonging to each class. Using PLS-DA and HCA, we are able to identify samples that contain avidin. The next question to answer is using this method can SERS distinguish between samples containing avidin and nuetravidin. A PLS-DA model containing only samples from neutravidin and avidin was built for analysis. The PLS-DA scores plot for this model is shown in Figure 2.15.
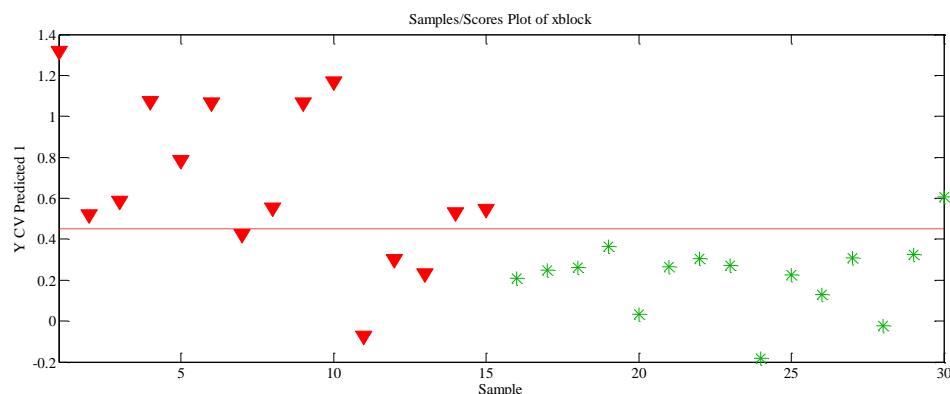
Figure 2.15 PLS-DA scores plot of the avidin and neutravidin. Classes used to build the model were MEG:N$_3$AT:BDIBO-Avidin ▼ and MEG:N$_3$AT:BDIBO-Neutravidi n ✳ .

As discussed earlier, neutravidin is an alternative biotin binding partner commonly used. Carbohydrates have been removed from the structure of neutravidin allowing it to be distinguished from avidin. PLS-DA was able to distinguish between neutravidin and avidin. This is important because many biomolecules have more than one binding partner. So the ability to distinguish binding partners will allow for an enhancement of information that can be gained.

In the figures the threshold is plotted as the dashed line. Cross validated predictions for each model tested are shown in Tables 2.4. Class designation is determined by the separation of Y prediction values which will appear either above or below the threshold. For example, in the Neutravidin vs. Avidin model, all the values that are above the threshold are classified as avidin and those below are classified as neutravidin. The PLS-DA model was evaluated for performance in terms of sensitivity and specificity. Sensitivity is defined as the number of samples assigned to the class divided by the actual number of samples belonging to the class. Specificity is defined as

Table 2.4. Cross Validated Prediction Scores for MEG:N$_3$AT, MEG:N$_3$AT:BDIBO, MEG:N$_3$AT:BDIBO-Avidin. Cross Validated Prediction Scores for MEG:N$_3$AT:BDIBO-Avidin, and MEG:N$_3$AT:BDIBO-Neutravidin.

Sample Name	Sensitivity	Specificity	Class Error CV	RMSECV
MEG:N$_3$AT	1.000	1.000	0	0.200
MEG:N$_3$AT:BDIBO	0.944	1.000	0.0278	0.207
MEG:N$_3$AT:BDIBO-Avidin	1.000	1.000	0	0.0848
MEG:N$_3$AT:BDIBO-Avidin	0.800	0.933	0.0667	0.133
MEG:N$_3$AT:BDIBO-Neutravidin	0.933	0.800	0.133	0.0667

the number of samples not assigned to the class defined by the number of samples not belonging to the class.[151] Sensitivity and specificity values for the model created to examine the biotinylation process were >94%. Sensitivity and specificity values for the model created to distinguish avidin from neutravidin were>80%.

Conclusions

A process for immobilizing biotin to nanoparticles was presented. This work demonstrated that copper free click chemistry can be used to immobilize biotin to nanoparticles. After biotin was immobilized, the biotinylated nanoparticles were incubated with avidin. The clicked particles were successful in capturing avidin. Due to the spectral uniqueness of the Cu^I-free click chemistry reaction, SERS was used to examine step-by step modification of the reaction. Using the nanoparticles as SERS

substrates, we were also able to characterize avidin binding by UV and fluorescence. Furthermore, we were able to quantify how much avidin was captured onto the particles. Nonspecific binding of avidin was shown to be minimal and further reduced using BSA as a blocking agent. Using chemometric analysis of SERS spectra classification of spectra according to group type was made. Using the copper free click chemistry platform, many different biomolecules and their binding partners can be analyzed. This work served as a proof of concept and results indicate that this method can be applied to carbohydrates, thus leading to the development of a carbohydrate array based on SERS detection.

CHAPTER 3

SURFACE ENHANCED RAMAN SPECTROSCOPY DETECTION OF

LACTOSE-GALECTIN PROTEIN INTERACTIONS

Introduction

Carbohydrates alter the functions of proteins and lipids to which they are attached. On cell surfaces they operate as ligands for proteins which govern cell trafficking, adhesion, and signaling.[152] Carbohydrate bimolecular interactions also play roles in tumor metastasis, viral and bacterial infection, and inflammation.[153] Hence, some biological processes cannot be fully understood without knowledge of the roles and functions carbohydrates play. Given this information, the study of carbohydrates may lead to the development of tools for clinical diagnostic technology as well as therapeutics.

Carbohydrate microarrays are used to identify and examine carbohydrate interactions with potential binding partners. Microarrays have the advantage of being high throughput tools which do not require large amounts of carbohydrate sample.[154] Carbohydrate microarrays have been used in immunology and clinical diagnosis, self-assembly, bacteria detection, and proteomics.[155] Common methods for the detection of carbohydrate microarrays include fluorescence, surface plasmon resonance (SPR), and mass spectrometry. Fluorescence based methods are more commonly used because they offer high sensitivity and throughput; however, this method is prone to denaturation of proteins and can cause interference with carbohydrate binding.[156] SPR, a label free

detection method, eliminates some of the problems caused by fluorescence methods namely the possibility of fluorescence labels binding to proteins. However, SPR is limited by its inability to simultaneously measure a large number of carbohydrate-lectin interactions.[157] Mass spectrometry has been used to detect carbohydrate modification and obtain structural information, but it is not suitable for detection of bound proteins on surfaces.[156]

Surface enhanced Raman spectroscopy (SERS) can be used as a label free method for carbohydrate microarray detection. SERS has already proven itself as a powerful tool for microarray detection. This technique can be used to determine molecular structural information and provides ultrasensitive detection limits, which includes single molecule sensitivity.[135] SERS has been used for detection in bacteria microarrays, protein microarrays, and DNA microarrays.[158-160] SERS has advantages over other label free detection schemes, namely SPR. SPR relies on a general response of a captured analyte, while SERS is molecule specific. There have been very few studies published to date which focus on carbohydrate and carbohydrate-lectin interactions detected by SERS.[103,161-163] Therefore, to add to the current literature on this subject we have developed a carbohydrate microarray based on SERS detection.

As a proof of concept interactions between lactose and galectin proteins were examined. Galectins are carbohydrate binding proteins that share a conserved sequence in the carbohydrate region domain (CRD).[164] Of important note is they all possess aromatic protein residues, which are strong Raman scatterers. Galectins can bind to the surface of viruses, bacteria, protista, and parasites.[165] We aimed to detect galectins binding to lactose using SERS. Lactose was immobilized to SERS substrates via strain promoted

azide alkyne click chemistry.[166] The lactose modified surface was then incubation with

galectins. Using SERS as a detection method allowed for monitoring carbohydrate

protein interactions to provide molecule specific analysis. Since carbohydrate protein

interactions are normally weak in nature, multivariate statistical analysis was employed to

identify spectral differences that were important indicators of protein binding and

provided a measure of sensitivity and specificity.[167]

Experimental Methods

Materials. Gold nanoparticles with a diameter of 60nm were purchased from Ted Pella.

Indium tin oxide (ITO) glass slides with square resistivity of 70-100 Ω/sq and (11-

Mercaptoundecyl)tetra(ethylene glycol) (95%) were purchased from Sigma Aldrich. Also

purchased from Sigma Aldrich were recombinant human galectin-1, recombinant human

galectin-3, streptavidin-cy3, and Albumin, biotin labeled bovine(BSA). Human galectin-3

alexa fluor 488 was purchased from R&D systems. Gold (99.999%) and Titanium

(99.999%) evaporation pellets were purchased from Kurt J. Lesker. Dimethylformamide

(DMF) was obtained from J.T. Baker. Sterile water purchased from Braun was used in all

aqueous solutions and rinsing procedures.

SERS Substrate Fabrication. ITO glass slides were cleaned for twenty minutes using

heated piranha solution containing a mixture of 4:1 H_2SO_4:H_2O_2, followed by rinsing

with nanopure water, and drying with a gentle nitrogen N_2 gas stream. Titanium (10nm)

and gold (300nm) were deposited at a rate of 0.5-1 Å/s onto the ITO glass slides using the

PVD75 electron beam deposition system where background pressure was maintained at

less than 5.5 x 10^{-6} Torr. Following deposition the gold coated substrates were removed

from the deposition chamber and stored inside a Petri dish sealed with parafilm to prevent

surface contamination. After fabrication of the gold film, a multi-well array was

assembled using a molding patterning process. In this process a stainless-steel mold was

pre-heated in an oven at 55ºC for 30 minutes. After pre-heating, a mixture of

polydimethylsiloxane (PDMS, Slygard 184 Elastomer Kit), curing agent, and accelerator

(10:1:0.5, w/w) were poured through the mold opening. This mixture cured at 70ºC for 45

minutes. After the curing process, the mold was removed from the oven and cooled down

to room temperature. The SERS substrate was detached from the stainless steel mold and

excess PDMS was removed with a razor blade. This process generated a uniform 4 x 10

multi-well array substrate which can be seen in Figure3.1.
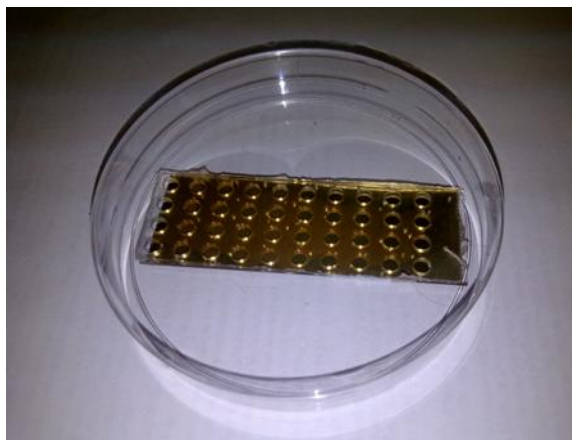
Figure 3.1 Picture of gold substrate for
SERS analysis.

Immobilization of Lactose to Gold Nanoparticles. Compounds used for experiments are

shown in Figure 3.2. Carbohydrates were attached to gold nanoparticles using strain

promoted azide-alkyne click chemistry. Azides used for experiments had a disulfide

moiety for attachment to gold nanoparticles, a hydrophobic alkane region for monolayer

formation, and a polar PEG moiety to facilitate compatibility with an aqueous

environment. A compound containing similar features was also added to the

nanoparticles to ensure proper spacing is achieved for carbohydrate attachment. Lactose

was synthesized to contain a dibenzocyclooctyne (DIBO) moiety for clicking properties.
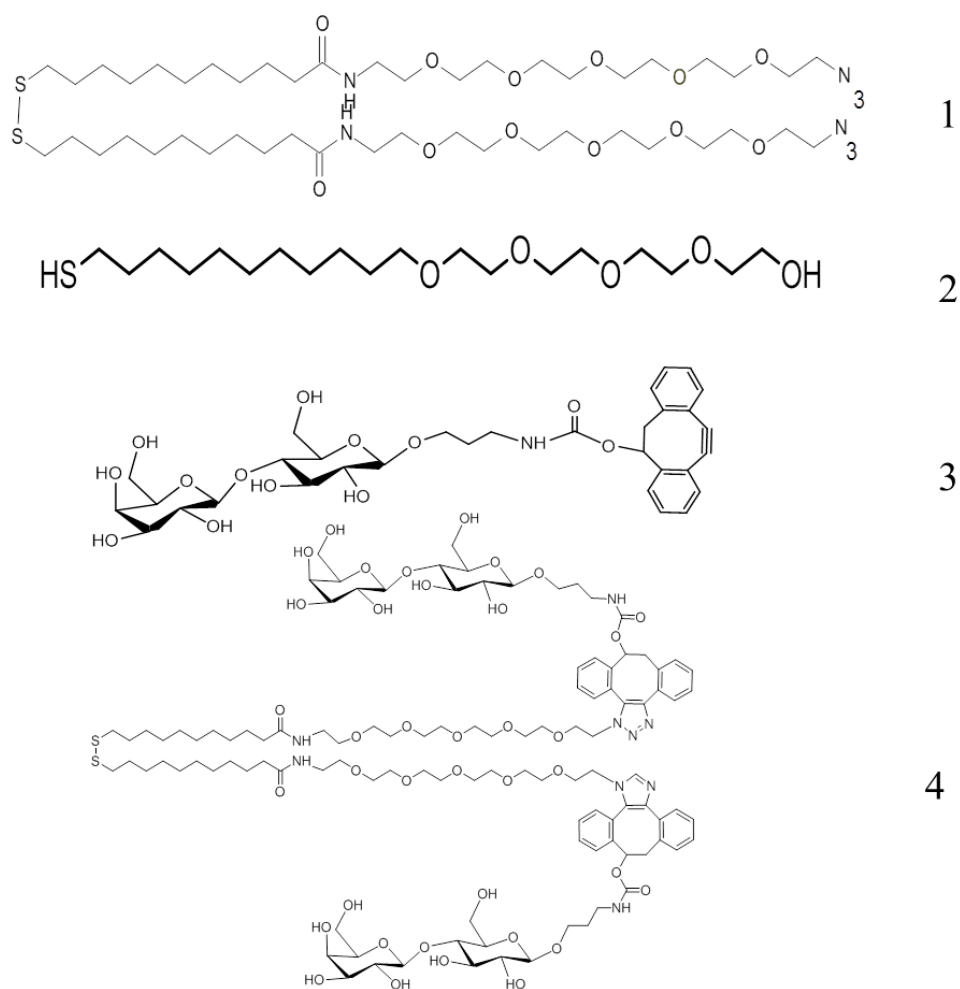
Figure 3.2. Compounds for SPAAC: (1) azide, (2) spacer (3) DIBO-Lactose, (4) model compound.

Derivatization of the gold nanoparticles occurred in a series of steps. First, solutions of the spacer and azide linkers were prepared in DMF. The spacer and azide total molar concentration was made to be 10mM. A total of 400µl of the spacer:azide mixture was added to 4ml of 60nm gold nanoparticles in a centrifuge tube. To reduce the exposure of light to the nanoparticle mixture, the centrifuge tube was wrapped in aluminum foil throughout all phases of the experiment. Then the mixture was placed onto a shaker for 8-12hrs. After this period of time, the solution was removed from the shaker and purified by centrifugation 3 times at 1050rpm for 10 minutes. Between centrifugation cycles, the supernatant was removed and the nanoparticle solution was rinsed with Braun water. After centrifugation, the nanoparticles were resuspended into water and sonicated for 20min to ensure particles were free from aggregation. Following sonication, 200µl of solution was removed for SERS analysis. To click with the azide modified nanoaprticles, lactose was modified with a dibenzocyclooctyne group.

DIBO-lactose (DIBO-Lac) was dissolved into water, 200µl was added to the azide gold nanoparticles, and this sample was placed onto the shaker for 8-12hrs. The ratio of DIBO-Lac added to the nanoparticle mixture was double the portion of azide. This was done to ensure immobilization of lactose throughout the substrate. After the click reaction, the azide-lactose nanoparticle mixture was purified using the same centrifugation and sonication procedure mentioned above. Following sonication, 200µl of the nanoparticle solution was removed for SERS analysis. A model compound was synthesized to simulate the clicked product. The model compound was dissolved in DMF and attached to gold nanoparticles via the gold-sulfur bond. After all samples were

prepared for SERS analysis 10μl of each sample were dropped into the multi-wells of a patterned gold film substrate and allowed to dry overnight. The addition of the modified nanoparticles to the gold film surface created a carbohydrate SERS sensing surface which was then used to probe for galectin binding.

Galectin-1 and Galectin-3 Binding Study. Galectin proteins chosen for SERS experiments were galectin-1 and galectin-3. Galectin-1 is a homodimer which has a carbohydrate binding region composed of 130 amino acids. Galectin-3 is not a homodimer and has a carbohydrate binding region consisting of 100 amino acids. It also has an N-terminal domain which consists of an additional 30 amino acids.[168] The differences in the structure and amino acid sequence of each protein allows for detection and classification of each subtype. Human galectin-1 and galectin-3 were purchased from Sigma Aldrich. The lyophilised galectin powders were dissolved into a PBS buffer and prepared in the following concentrations: 100μg/ml, 50μg/ml, 25μg/ml, 5μg/ml, and 1μg/ml. For protein binding experiments ten centrifuge tubes were filled with 25μl of the lactose modified nanoparticle solution. To the tube was added human galectin-1 or galectin-3 at a certain concentration. The vials were wrapped in aluminum foil and placed into an incubator for 2hrs at 37ºC. After this period, the tubes were purified by centrifugation. They were washed two times with buffer solution and a final wash with Braun water to removed salts from PBS from the solution. Samples were later sonicated to resuspend the particles into water. For SERS analysis 10μl of the galectin-lactose gold nanoparticle solution was placed onto the gold film multiwall substrate and allowed to dry overnight.

Nonspecific Protein Control Experiments. Control experiments were done to examine the specificity of the lactose-galectin interactions. Bovine serum albumim(BSA) has a low affinity for lactose, therefore, minimal binding is expected to occur. However, the binding of BSA with lactose is expected to be much less than lactose galectin binding. BSA was prepared at concentrations 100μg/ml, 50μg/ml, 25μg/ml, 5μg/ml, and 1μg/ml to be in line with concentrations used in the galectin experiments. Various concentrations of BSA were added to five centrifuge tubes containing 25μl of lactose modified gold nanoparticles. The vials were wrapped in aluminum foil and placed into an incubator for 2hrs at 37°C. After this period, the tubes were purified by centrifugation. They were washed two times with buffer solution and a final wash with Braun water to removed salts from PBS from the solution. Samples were then sonicated to resuspend the particles in water. For SERS analysis 10 μl of the BSA-lactose gold nanoparticle solution was placed onto the gold film multiwell substrate and allowed to dry overnight.

Raman Spectroscopy. All samples were prepared in duplicates. SERS spectra were measured using a confocal Raman Microscope (InVia, Renishaw, Inc. Gloucestershire, United Kingdom). Laser excitation was provided by a 785nm near-IR diode laser. Illumination of the sample was through a 20x (N.A. = 0.40) objective. This results in a spot size of approximately 4.8 x 27.8 μm. The laser power measured at the sample was ~3.62mW. SERS spectra were acquired from five different spots within a given micro-well using a 10s acquisition time with three accumulations. The spectral range recorded was between 2300 and 350 cm^{-1}.

Fluorescence Protein Binding Study. In addition to SERS analysis, fluorescence experiments were done to serve as a measure of comparison to SERS and to quantify how

much protein was captured onto the SERS substrate. Fluorescence measurements were recorded on a POLARstar OPTIMA multidetection microplate reader. Fluorescence experiments were designed to study the interactions of galectin-3 and BSA with lactose. BSA modified with biotin was purchased from Sigma Aldrich. Galectin-3 and BSA-biotin were prepared in concentration 100µg/ml. The freshly prepared lactose nanoparticle solution was separated into 6 centrifuge tubes. Either galectin-3 or BSA-biotin was added to the tubes and incubated with the lactose nanoparticles for 2hrs at 37ºC. Following this incubation period, the protein solutions were purified by the centrifugation process described above. Then 75µl of each solution was removed for SERS analysis.

96-well plates from Greiner Bio-One were used in fluorescence experiments. Cy3 streptavidin was purchased from Sigma Aldrich. Cy3 has an excitation and emission wavelengths at ~550nm and 570nm respectively. 50µl of 100µg/ml of Cy3 streptavidin was incubated with the lactose BSA-biotin nanoparticle solution for 1hr at 37ºC. Also, 50µl of 100µg/ml of Cy3 streptavidin was incubated with non-modified gold nanoparticles under the same experimental procedures. After the incubation period, the samples were purified using the centrifuge process described earlier. A standard curve of Cy3 streptavidin was prepared and used to estimate the concentration of BSA-biotin on the surface of the nanoparticles. In duplications, 80µl of each standard was placed into the microwells. Also added to the well plate were 80µl of Cy3 streptavidin-lactose BSA-biotin nanoparticle solution and 80µl of Cy3 streptavidin non-modified nanoparticle solution.

Human galectin-3 alexa fluor 488 affinity purified PAb was purchased from R&D Systems. Alexa fluor 488 has an excitation and emission wavelengths at ~493nm and 519nm respectively. Human galectin-3 alexa fluor 488 was prepared at a concentration of 100μg/ml. 100μl of this solution was added to the carbohydrate-galectin-3 nanoparticle solution for 1hr at 37ºC. After this exposure time, the nanoparticles were purified by the centrifugation process and analyzed by fluorescence. The alexa fluor 488 solution was also incubated with non-modified gold nanoparticles under the same experimental procedures. After the incubation period, the samples were purified using the centrifuge process described earlier. A fluorescence standard curve was prepared for the alexa fluor 488 solution and samples were prepared in duplicate.

Spectral Preprocessing. Ten SERS spectra were collected for each sample. To view spectral quality and increase signal-to-noise SERS spectra were averaged using GRAMS 32/AI spectral software package (Galactic industries, Nashua, NH). After averaging the spectra were baseline corrected using multipoint leveling in the GRAMS 32/AI software. In this step, the baseline is leveled at a value which is the average of the baseline points and offset so the new baseline is set to zero.

Preprocessing for multivariate statistical analysis was performed in PLS Toolbox 4.0 (eigenvector, Wenatchee, WA) operating in Matlab software (version R2013b). Preprocessing for multivariate analysis involved loading raw spectral data into an Excel workbook. The workbook contained several sheets which are used to characterize the SERS spectra: X-block, wavenumber, labels, classes, and Y-block. The X-block contains all the spectral intensities recorded for each sample across the data range. The wavenumber sheet contains data points of the spectral range that was recorded. The labels

and classes sheets report the name of the samples and class membership identified with a numerical value. The Y-block was created for use with partial least squares discriminant analysis(PLS-DA) and will be explained in a later section. The Excel workbook was then imported into the Matlab software, formed into a dataset array, and analyzed using PLS Toolbox 4.0. Using the Toolbox, preprocessing of the X-block data involved taking the first derivative of each spectrum using a 15 point, 2nd order polynomial Savitsky-Golay algorithm.[169] This was followed by normalization to unit vector length and mean centering. Preprocessing of the Y-block only involved mean centering.

Hierarchical Cluster Analysis (HCA). Hierarchical clustering is an unsupervised method which groups objects which have close distance to one another. HCA was done using the agglomerative method K-Nearest Neighbor (KNN). In agglomerative methods, each spectrum serves as its own cluster and similar spectra are combined into larger ones. HCA was performed to assess sample clustering.

Partial Least Squares Discriminant Analysis. PLS-DA is a supervised statistical classification method, where prior knowledge of the classes is used for classification.[102] This method minimizes the contribution of spectral features which vary within a particular class and maximizes the contribution of spectral features among the different classes.[151] PLS-DA incorporates both the X-block and Y-block. They Y-block contains variables of 0 or 1 which serve as a reference for each sample. This number indicates whether a sample belongs to a particular class. The models generated by PLS-DA are tested by cross-validation using the Venetian Blinds method. PLSDA was used to differentiate galectin proteins and to access nonspecific protein binding.

Results and Discussion

SERS spectra of the clicked lactose complex. SERS spectra of the clicked lactose

complex are shown in Figure 3.3. Lactose was immobilized onto gold nanoparticles using

a mixed monolayer approach. The total molar concentration of the components added to

the nanoparticles was 10mM. The absence of bands for thiols and disulfides in the SERS

spectra confirmed the attachment of azide and spacer components to the nanoparticles.

Gauche and trans bands for (C-S) vibrations at 646 and 713cm^{-1} are in agreement with

those reported in literature.[170] After azide immobilization lactose derivatized with DIBO
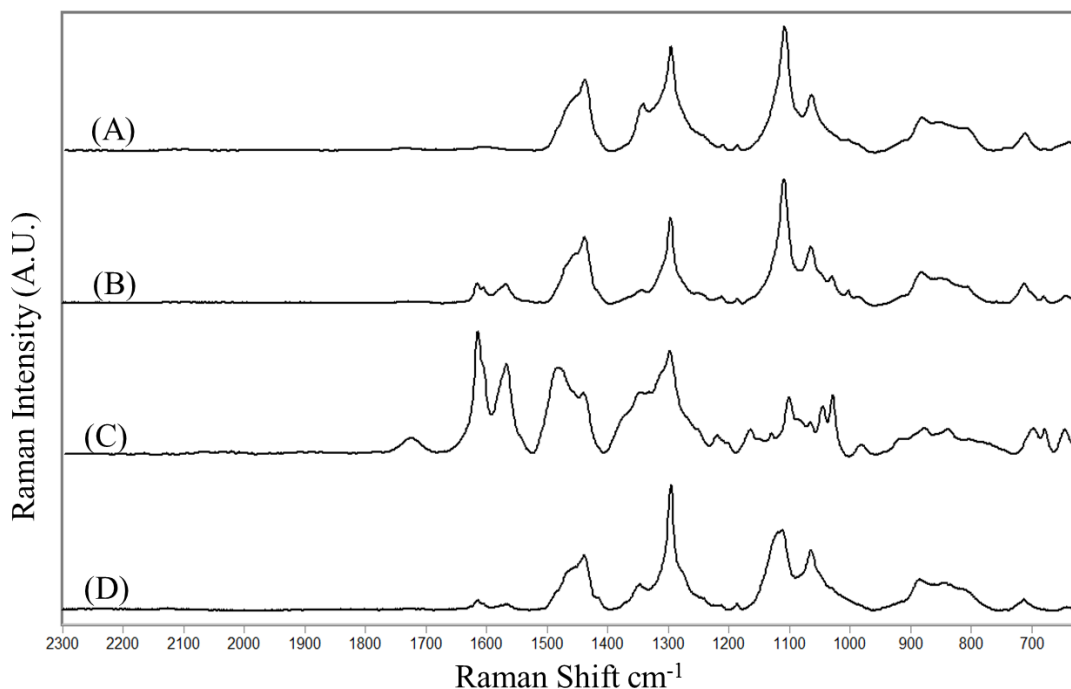
Figure 3.3. SERS spectra of: (A) spacer-azide, (B) spacer-azide-lactose, (C) model
compound, (D) spacer-model Compound. Each spectrum shown is an average of
10 individual spectra for each sample.

was added to the nanoparticles to produce the clicked lactose complex. The SERS spectra

show noticeable changes in the spectra after the addition of lactose.

Differences in the spectra are seen at spectral bands: 679, 1004, 1029, 1567, and 1615cm^1. These bands are consistent with those reported for benzene ring stretching and carbohydrates.[171,172] It was also noted that bands for the ($\equiv$C-C) group are absent in the spectra. This band would normally appear between 1900-2100 cm^{-1}.[136] Therefore, its absence in the spectra is further indication that clicking has taken place between the azide and DIBO components. We next evaluated SERS spectra of our in situ fabricated clicked nanoparticles with a model compound that was synthesized to simulate the clicked product. We examined the model compound as a mixed monolayer (spacer-model compound) and as a standalone monolayer. Examination of the spectra reveals clear similarities to the clicked nanoparticle product. Bands for ring stretches, C-S bonds, and carbohydrates can be seen in all the spectra. This gives further indication that we have successfully clicked lactose onto the gold nanoparticles.

Galectin Proteins Binding to the Clicked Lactose Complex. Galectin binding with lactose leads to the addition of protein on the surface of the modified nanoparticles which can be sensed by SERS. Only a small amount of protein is expected to bind to the lactose nanoparticles. Since bands from the spacer component dominate the spectra identifying protein bands can be somewhat complex. Therefore, SERS spectra of the lactose complex with and without protein were Savitzky-Golay derivatized and normalized using the procedure outlined in the methods section. Figure 3.4a shows the normalized derivatized spectrum for the clicked lactose complex before the addition of protein. Figure 3.4b shows the normalized derivatized spectrum for the lactose nanoparticles incubated with galectin-1 for 2 hours at 37ºC with 100μg/ml galectin-1 content. Figure 3.4c shows the normalized derivatized spectrum for the lactose nanoparticles incubated with lactose

galectin-3 for 2 hours at 37°C with 100µg/ml galectin-3 content. The derivatized spectra

reveal overlapped bands and changes in the spectral pattern are seen after the addition of

the galectin proteins. These changes can be used to make protein assignments. Aromatic

and amide vibrations are used in SERS to identify protein bands. Protein bands are

indicated in the figure by the box outlines and assignments are given in Table 3.1. The

normalized Savitzky Golay spectrum for galectin-1 and galectin-3 show a high degree of

specificity for binding only to the lactose nanoparticles.
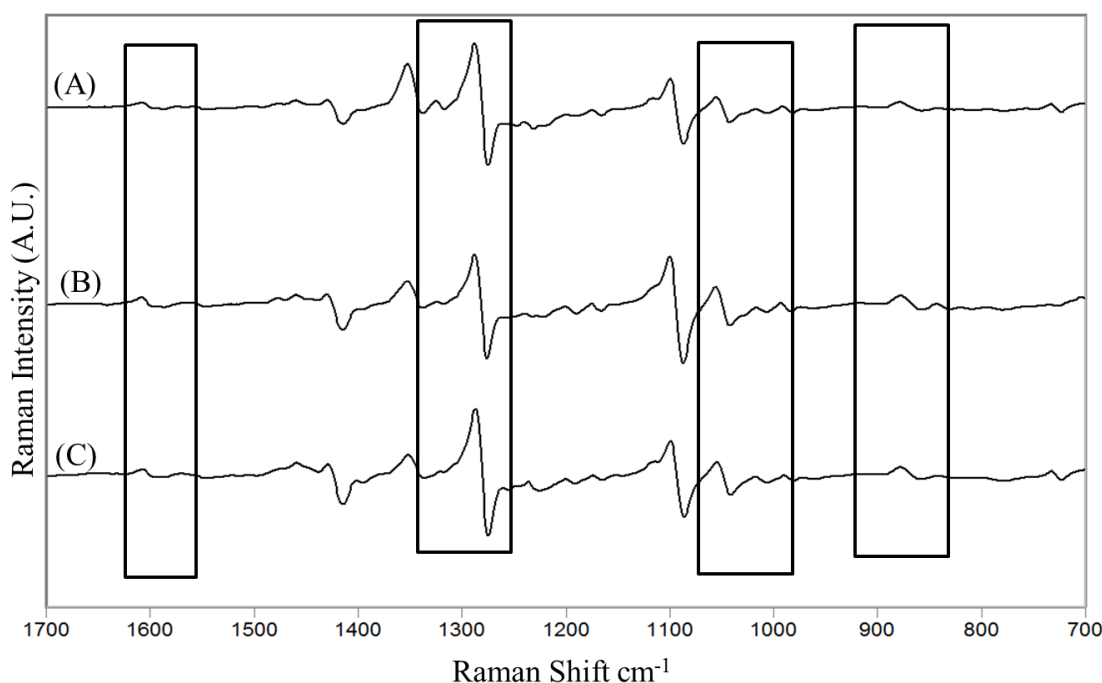
Figure 3.4. Normalized Savitzky-Golay SERS spectra of: (A) spacer-azide-dibo-lactose, (B) spacer-azide-dibo-lactose + galectin-1, (C) spacer-azide-dibo-lactose + galectin-3. Each spectrum shown is an average of 10 individual spectra for each sample.

Negative Control with Protein. We incorporated BSA as a negative control protein in

experiments to assess the specificity binding of galectin proteins to the lactose

nanoparticles. BSA is a protein derived from cows and often used as a protein

concentration standard in lab experiments. BSA prepared at various concentrations was

added to 25µl of lactose gold nanoparticles using identical binding conditions to those

Table 3.1. Assignments for galectin -1 galectin-3 on nanoparticles.

Assignment	Galectin-1 Assignment	Galectin-3 Assignment
Indole ring Trp		1558
Trp	1459	1449
Tyr	1352	1352
Trp, C-H		1325
Amide III	1237	1236
C-N str.		1128
Tyr	1086	1087
Phe	1036	1035
Phe	1003	
Phe	990	991
Trp	877	883

used for the incubation of galectin proteins. Figure 3.5b shows the normalized Savitzky

Golay spectrum of BSA-lactose nanoparticles.

This spectrum was compared with 3.5a which shows the normalized Savitzky Golay

spectrum of lactose nanoparticles. Minor spectral intensity changes are observed in the

figure. This observation was expected due to previous studies. The changes in the BSA

spectrum when compared to the spectral changes observed for galectin-1 and galectin-3

are minimal.
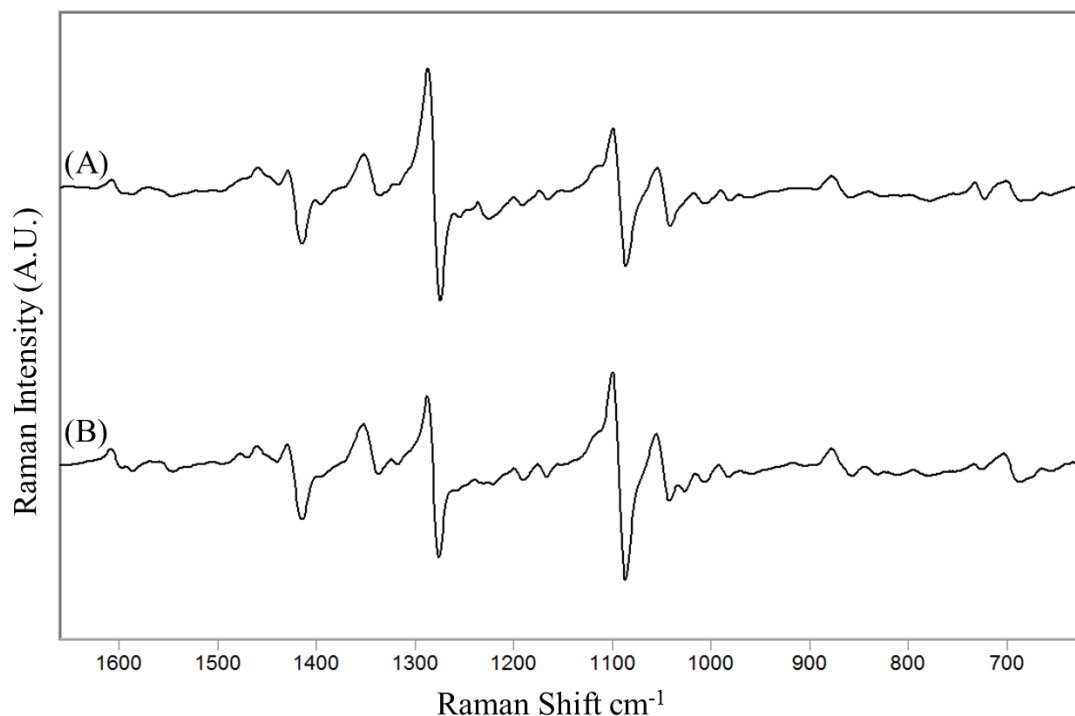
Figure 3.5. Normalized Savitzky-Golay SERS spectra of: (A) spacer-azide-dibo-lactose, (B) spacer-azide-dibo-lactose BSA, Each spectrum shown is an average of 10 individual spectra for each sample.

Fluorescence Analysis of Lactose Galectin Binding. Fluorescence experiments were done

to estimate how much protein was captured onto the DIBO-Lac nanoparticles.

Fluorescence samples included: DIBO-Lac + Galectin-3 Alexa fluor 488, and DIBO-Lac

+BSA-Cy-3. Nonspecific protein binding was assessed for samples containing the GNP

solution + fluorescence label and BSA samples. Results from the fluorescence studies are

shown in Figure 3.6.
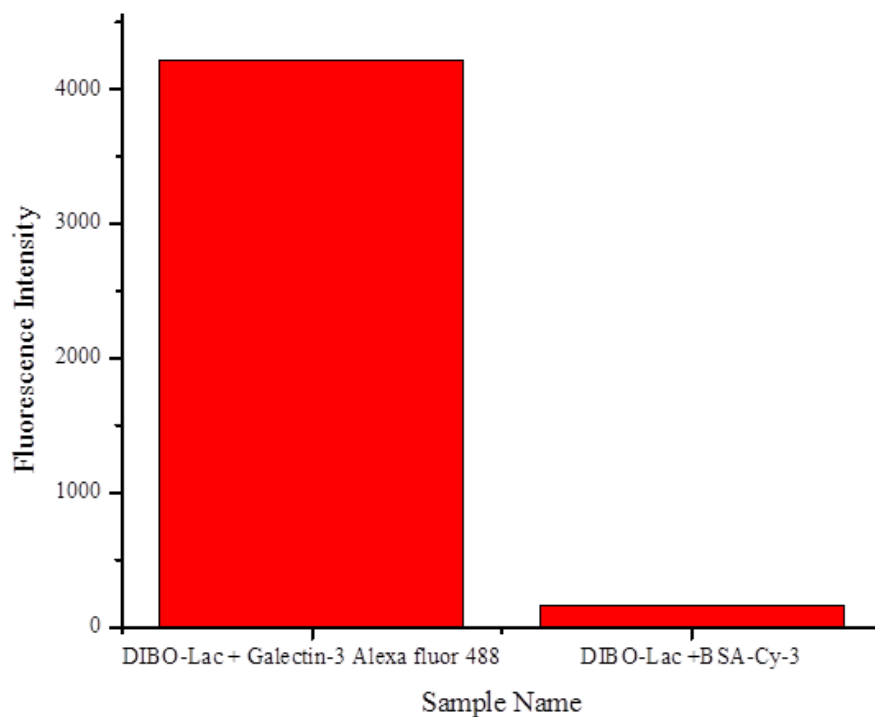
Figure 3.6. Fluorescence intensity values for galectin -3protein studies.

When BSA was added to the DIBO-Lac nanoparticles, the recorded fluorescence

intensity was 166. The signal is much weaker than the DIBO-Lac + Galectin-3 Alexa

fluor 488 sample. This suggests that lactose has stronger interactions with galectin

proteins and very minimal binding with BSA. Furthermore, when the fluorescence labels

were added to the bare gold nanoparticles, the signals is much less than when the GNPS

are modified.

Chemometric Aalysis of Lactose Galectin Spectra. Since the SERS spectra are dominated

by bands of the spacer component, multivariate statistical analysis was used to identify

the extent of galectin proteins binding to lactose nanoparticles. Statistically significant

differences between the SERS spectra are expected to be identified by the multivariate

analysis. Figure 3.7 shows dendograms produced by HCA based on the SERS spectra of

the lactose, lactose-galectin-1 binding, and lactose-galectin-3.
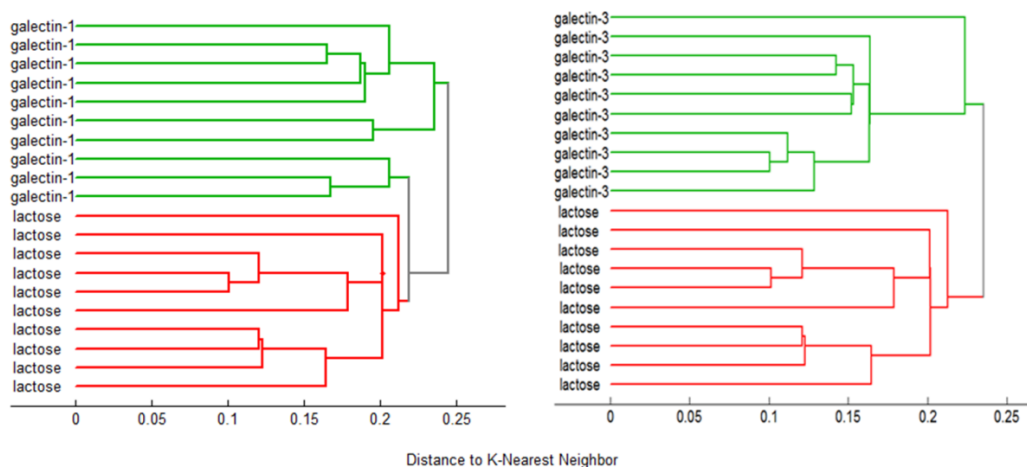
Figure 3.7. Dendogram produced by hierarchical cluster analysis (HCA) of SERS
spectra using K-Nearest Neighbor. A total of 20 spectra were used to create each
dendogram, corresponding to 10 spectra in each of the sample categories. (A) HCA
Dendogram for DIBO-Lac and DIBO-LAC + galecin-1. (B) HCA Dendogram for
DIBO-Lac and DIBO-LAC + galecin-3.

The concentration of galectin protein in the samples was 100µg/ml. HCA is an

unsupervised statistical data analysis method and separates data based on total variance.

The HCA dendogram was calculated based on distance to K-Nearest Neighbor. Twenty

spectra are shown, correlating to ten spectra from the lactose samples and 10 spectra from

either lactose-galectin-1 or lactose-galectin-3 samples. HCA grouped 18 out of 20

samples to their correct classes.

PLS-DA was used to further investigate binding of galectin proteins to lactose

nanoparticles. PLS-DA is a supervised statistical analysis method where prior knowledge

of the classes is given to return a robust discrimination, reducing variation within classes

while emphasizing latent variables between or among classes. This method identified

structural differences which can discriminate between the sample groups.[173] The model

was built with SERS spectra of the lactose nanoparticles, lactose + galectin1 or, lactose +

galectin3; where galectin was present in concentrations of 1μg/ml, 5μg/ml, and 100μg/ml.

A total of 70 spectra were included in the model, with 10 spectra corresponding to each

sample. The robustness of the generated model was tested by cross validation with

Venetian Blinds with 8 splits. During calibration of the model, 10% of the data is

withheld and then applied to "unknown" samples to assess its predictive power. The

model was generated with 6 latent variables as optimized by Matlab accounting for 88%

of total variance. Scores plots for the PLS-DA model generated are shown in Figure 3.6.

The data was classified with 83-100% sensitivity, and 85-98% sensitivity. Sensitivity is

the probability of classifying a SERS spectrum belonging to the correct class and

specificity is the probability of classifying a SERS spectrum not belonging to the correct

class.[164]

Lower sensitivity and specificity values can be explained by several reasons. First,

overall sensitivity and specificity affinity for galectin-1 are lower than galectin-3, so it

may not bind as tightly and can be washed away easier. Another factor that may explain

the lower sensitivity and specificity results is that this model was built with seven classes.

This may make classification difficult. So, binary models were built to contain samples

for lactose and lactose + galectin-1 or galectin-3 at a certain concentration. The cross

validated scores for those models generated sensitivity values of 100% and specificity

values ranged from 95-100% (Table 3.2).

Protein Binding Classification

	Lactose	Galectin-1	Galectin-3
Sensitivity	1.000	0.833	0.867
Specificity	0.983	0.850	0.900
Class Error	0.00833	0.142	0.117
RSMECV	0.176	0.330	0.314
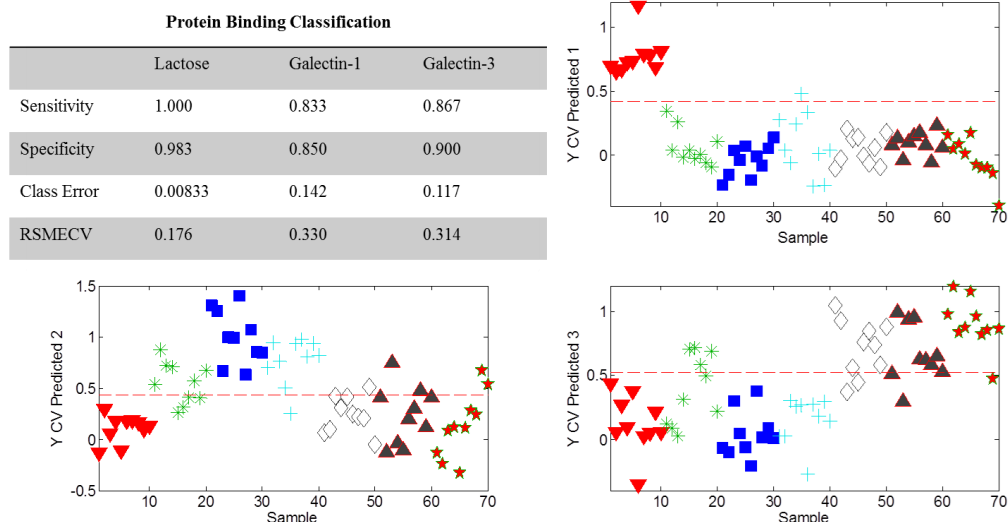
Figure 3.8. PLS-DA Y predicted plots. Classes used to build the model were lactose, lactose + galectin-1, and lactose + galectin-3. Galectin protein concentration used were 1μg/ml, 5μg/ml, and 100μg/ml. Each plot predicts a sample as belonging to or not belonging to the specified classes. Lactose ▼, 1μg/ml Gal-1 ✳,5μg/ml Gal-1 ■, 100μg/ml Gal-1 ◇ ,1μg/ml Gal-3 ✛,5μg/ml Gal-3 ▲ , 100μg/ml Gal-3 ★.

One benefit of PLS-DA is that it identifies variables that statistically explain the differences between the groups used to build the model. Therefore, a binary model was built to examine non protein and protein samples. The non protein samples contained 10 SERS spectra for lactose nanoparticles and the protein samples contained 20 spectra for lactose + galectin nanoparticles. The galectin sample contained protein concentrations of 100μg/ml for galectin-1 and galectin-3. Variables important in the projection (VIPs) as identified by the PLS-DA model were used to identify bands in the SERS spectra that were important for differentiation between non protein and protein classes. As a rule, it is generally accepted that VIP values greater than one are used for variable selection. The

Table 3.2. PLS-DA cross validation results of samples from binary models. Each model included Lactose and lactose + a specific concentration of galectin-1 and galectin-3.

89

Protein Binding Classification

	Lactose	1μg/ml Galectin-1	5μg/ml Galectin-1	100μg/ml Galectin-1	1μg/ml Galectin-3	5μg/ml Galectin-3	100μg/ml Galectin-3
Sensitivity	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Specificity	1.000	0.950	1.000	1.000	1.000	1.000	1.000
Class Error	0	0.025	0	0	0	0	0
RSMECV	0.0121	0.205	0.125	0.131	0.178	0.158	0.00951

variables correspond to bands in the SERS spectra and are listed in Table 3.3 with spectral assignments. It has been reported that amino acid residues of tryptophan, tyrosine, and phenylalanine participate in binding of galectin proteins to lactose.[174] These amino acids are aromatic and give very discernible bands in SERS spectra. The bands identified as VIPs correlate to vibrations for these amino acids.

Differentiation of Galectin Proteins. To further our investigation of SERS as a detection method for microarrays, we examined its ability to distinguish between proteins of the same family. Although galectin-1 and galectin-3 are from the same family of proteins, they have different structures. Galectin-1 is a homodimer composed of 130 amino acids which are considered to be the domain for carbohydrate binding.[168] Galectin-3 consists of the same 130 amino acids and an unrelated non-carbohydrate binding N-terminal domain of about 120 amino acids.[175] SERS spectra of Galectin-1 and galectin-3 prepared at concentrations of 100μg/ml, 5μg/ml, or 1μg/ml were used for this analysis. The averaged normalized Savitzky Golay spectrum (n=10) collected from the 100μg/ml samples are shown in Figure 3.9 as an example of the spectra collected.

Table 3.3. VIP scores generated as being important for differentiation of non protein and protein. Spectral bands that recorded a value greater than one are listed along with their spectral assignments.

Wavenumber cm^{-1}	VIP Value	Assignment
1498	1.08	CH$_2$ sciss.
1424	2.87	Trp
1413	16.67	Phe, Tyr, Trp
1352	34.90	Tyr
1235	9.07	Trp
1100	7.72	C-N str.
1090	15.11	Tyr
1038	3.75	Phe
1023	1.61	Phe
994	2.19	Phe
721	6.78	Phe, Trp, Tyr

The spectra show subtle distinctive spectral pattern differences between galectin-1 and galectin-3. Variations in band shapes and peak intensities are observed in the spectra. An example would be the band at 1056 cm^{-1}. This band appears to be broader in the spectra of galectin-1 than in the spectra of galectin-3. An intensity increase of the band around 1100 cm^{-1} in the spectrum of galectin-3 also shows variations in the SERS spectra. Furthermore, the band at 737 cm^{-1} does not appear in the SERS spectrum for galectin-3 but it does in the spectra of galectin-1. These pattern differences in the SERS spectra allow for multivariate methods to be used to discriminate proteins from the same family.
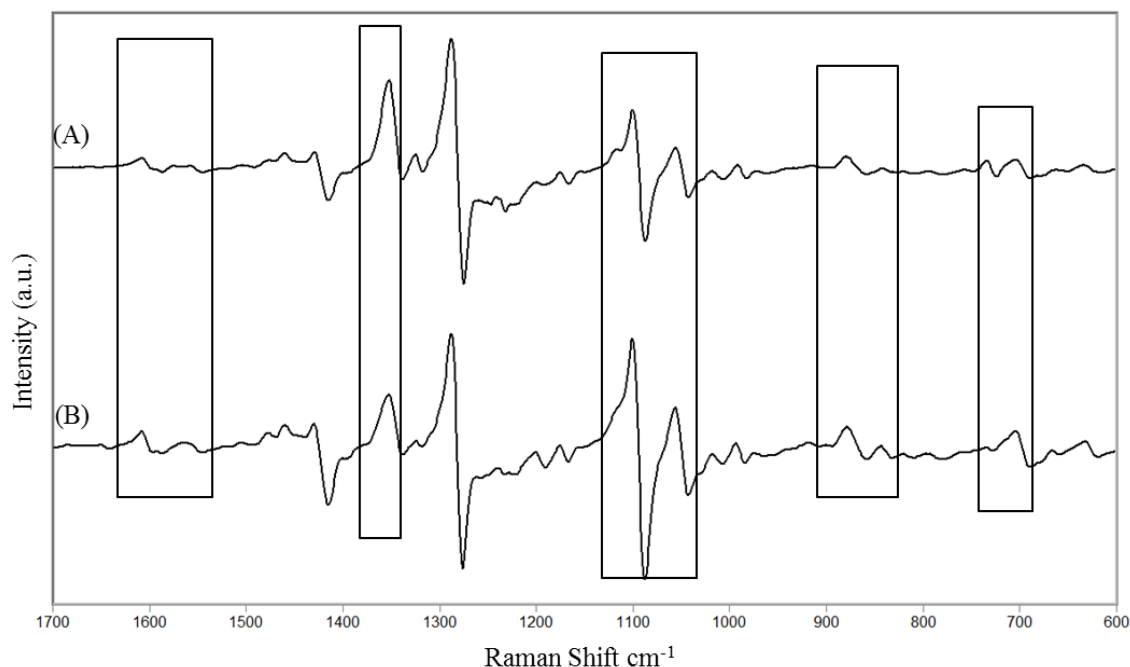
Figure 3.9. Normalized first derivative spectra of (A) Lactose + 100µg/ml galectin-1 and (B) Lactose + 100µg/ml galectin-3 showing similarities but also clear differences among the galectins. Differences are indicated by the boxes.

PLS-DA was used to determine if the differences observed in the SERS spectra are statistically significant. The galectin samples were grouped by concentration and family association. PLS-DA models were generated for each galectin protein concentration used, 100µg/ml, 5µg/ml, or 1µg/ml. Models contained 20 spectra, which included 10 spectra for each galectin type and the model was optimized using Venetian Blinds cross validation.

The models were generated with 4-5 LVs as optimized by the Matlab software, which accounted for 83-91% of the total variances. PLS-DA scores plots from the models generated for each concentration are shown in Figure 3.10. PLS-DA was able to distinguish between galectin-1 and galectin-3 at all concentrations. Cross validated sensitivity and specificity values are reported in Table 3.4. The cross validated sensitivity

and specificity ranged from 77-100%. An explanation for the drop in cross validated

sensitivity and specificity was expected for the galectin samples of 1µg/ml. Fluorescence

data indicated that the amount of protein captured is a very small amount which is not

readily distinguished between the background. However, multivariate statistical analysis

is able to identify enough spectral differences to distinguish between galectin-1 and

galectin-3 at this concentration.

Chemometric Analysis of Negative Control BSA. PLDSA models were also created to

investigate the negative control SERS spectra. The model was built with SERS spectra of

the lactose nanoparticles, lactose + galectin-1, lactose + galectin-3, and lactose + bsa.

Protein was present in concentration a of 100µg/ml. A total of 40 spectra were included

in the model, with 10 spectra corresponding to each sample. The model was created to

identify which samples contained protein, therefore two classes were created. They were

nonprotein, which included SERS spectra of the lactose nanoparticles and protein, which

included SERS spectra of the galectin samples.

Class values were not assigned the SERS spectra containing BSA. A class value

of 0 was used for these values. The designation of "0" effectively treats these samples as
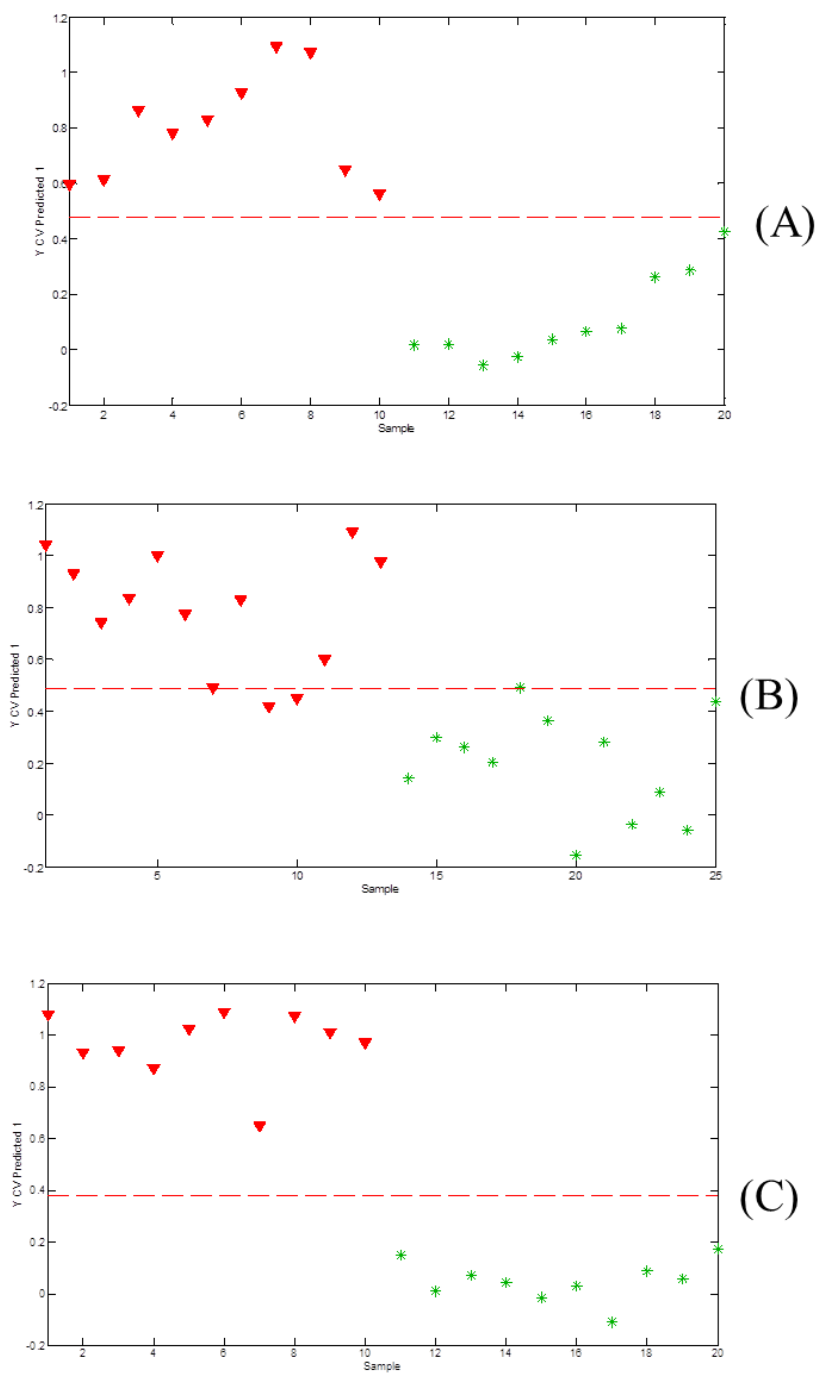
Figure 3.10. Cross-validated PLS-DA Y-prediction plots for galectin-1 and galectin-3 samples based on SERS spectra for each spectral class.The red triangles represent sample spectra of galectin-1 at various
concentrations. The green asterisks represent sample spectra of galectin-3 at various concentrations. (A) 100µg/ml, (B) 1µg/ml, (C) 5µg/ml.

Table 3.4. Summary of the PLS-DA cross validated results for classification of galectin samples at various concentrations.

Protein Classification

	100µg/ml Gal-1	100µg/ml Gal-3	5µg/ml Gal-1	5µg/ml Gal-3	1µg/ml Gal-1	1µg/ml Gal-3
Sensitivity	1.000	1.000	1.000	1.000	0.769	1.000
Specificity	0.900	0.900	1.000	1.000	1.000	0.769
Class Error CV	0	0	0	0	0.0758	0.0758
RMSECV	0.233	0.233	0.150	0.150	0.293	0.293

unknowns and they are not used when calibrating the model. Their class assignment will be predicted based on the calibrated model. The generated model was tested by cross validation with Venetian Blinds with 6 splits. The model was generated with 3 latent variables as optimized by Matlab accounting for 91% of total variance. Scores plots for the PLS-DA model generated are shown in Figure 3.11. Since, the BSA samples were entered as unknowns, they are represented by the black circles. Based on the results, the samples containing BSA classify with the lactose nanoparticle samples. This is expected because binding to lactose by BSA should be very minimal if at all. So, during the washing steps, BSA should be washed away leaving behind lactose on the nanoparticles. The model classified samples containing protein with 100% sensitivity and specificity and nonprotein samples with 100% sensitivity and specificity.
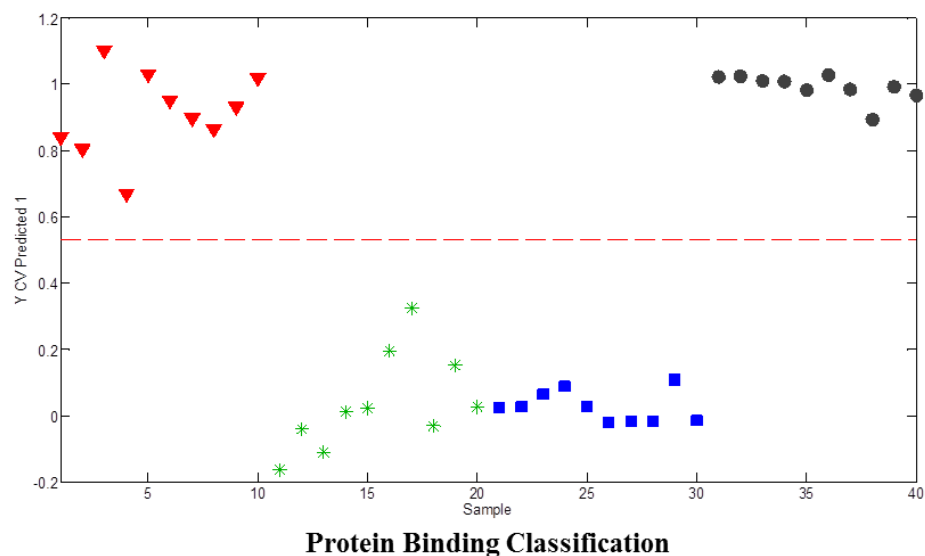
	No Protein	Protein
Sensitivity	1.000	1.000
Specificity	1.000	1.000
Class Error CV	0	0
RMSECV	0.109	0.109

Figure 3.11. PLS-DA Y-CV predicted scores plots. Classes used to build the model were lactose, lactose + galectin-1, lactose + galectin-3, and lactose + BSA . Galectin protein concentration used were 100µg/ml. Each plot predicts a sample as belonging to or not belonging to the specifie classes. The defined classes were no protein and protein. The no protein classes included lactose and BSA samples. The protein class included samples from galectin 1 and 3. Lactose ▼ 100µg/ml Gal-1 ✳ 100µg/ml Gal-3 ■ , 100µg/ml BSA ● .

Conclusions

We demonstrated that SERS is a rapid and direct method for detection of

carbohydrate-protein interactions. In this paper we reported the fabrication of lactose

nanoparticles using strain promoted alkyne azide click chemistry. Two galectin proteins

were used in this study to examine the interactions of lactose with galectin proteins. Because of the dominance of the spacer compound, multivariate statistical analysis was carried out to differentiate the extent of lactose-galectin binding. HCA showed a clear separation between the lactose nanoparticles and the lactose + galectin samples. Furthermore, we were able to identify bands in the SERS spectra which were important for distinguishing SRS spectra of lactose from lactose + galectin and correlate those peaks to proteins bands from the sequence of the galectin proteins. PLSD-DA results show high classification accuracy by having 95-100% sensitivity and specificity. We also demonstrated that SERS can be used to distinguish between two proteins from the same family. PLSD-DA classification results for distinguishing between galectin-1 and galectin-3 was 77-100% sensitivity and specificity. Finally we employed BSA as a negative control to examine the specificity of lactose binding. PLS-DA results showed that the BSA spectra were classified with the lactose nanoparticle SERS spectra. This indicated that BSA is washed off leaving behind the lactose nanoparticles. This work demonstrated the feasibility of SERS as a label free detection method for carbohydrate protein interactions. Future work will expand to study other carbohydrate lectin systems.

CHAPTER 4

SURFACE ENHANCED RAMAN SPECTROSCOPY DETECTION OF SIALYLATED CARBOHYDRATE-INFLUENZA HEMAGGLUTININ INTERACTIONS

Introduction

The influenza virus affects humans, pigs, and birds in large numbers. It manifests itself in humans as fever, head and body aches, and can cause other respiratory complications such as pneumonia and bronchitis.[176]   Yearly, an estimated 30 million people visit doctors' offices to seek relief from the flu virus. From this number, an estimated 300,000 persons are admitted for hospitalization and in the United States, nearly 30,000 deaths occur.[177] Flu outbreaks that turn into pandemics not only cause sickness, but also place strain on health and medical facilities, disrupts traveling and public gatherings, and significantly increases the number of deaths to over a million worldwide.[178]

Influenza is an RNA virus from the family *Orthomyxoviridae*. It has three subtypes, influenza A, B, and C.[179] Influenza A infects the widest range of host species and causes most of the worldwide pandemics. Subtypes of Influenza A are determined by the surface antigens hemagglutinin(HA) and neuraminidase(NA). HA plays a role in how the virus attaches to cells and NA determine how virus penetrates into cells.[180] HA binds to sialylated glycans on host cells. Sialic acids(SA) are nine carbon sugars and SA

molecules are characterized by glycosidic linkages between terminal SA and penultimate galactose residues. Common SA linkages that are important for interspecies transmission are α2,3 and α2,6 linkages.[181] It is generally believed that α2,6 SA linkages are predominately recognized by human and swine influenza viruses while avian and equine viruses bind to α2,3 SA linkages. However, recent data has shown that some avian viruses can bind to α2,6 linkages. Of note are avian H5,H7, and H9 which have caused high fatality in humans.[182] Therefore methods of detecting and understanding of HA receptor binding are needed.

HA receptor binding can be analyzed using glycan microarrays. Glycan microarrays have been used to probe the binding patterns of different influenza viruses' receptors and investigate the binding specificity of influenza HA. Glycan arrays are made by immobilizing carbohydrates to glass slides where their binding preferences are analyzed by different detection methods. They offer high throughput analysis because small quantities of analyte is used and many carbohydrates can be immobilized at one time. Therefore a large number of carbohydrate-receptor interactions can be screened.

Surface enhanced Raman spectroscopy (SERS) can be used as a detection method for glycan arrays. Our group has already demonstrated how SERS is a rapid, sensitive and specific detection method for glycan arrays. SERS was used to probe interactions of lactose and galectin proteins. From this study, we were able to detect unique molecular signatures for lactose and galectin proteins, examine the specificity or lactose-galectin binding, and differentiation between different galectin proteins. Furthermore, using SERS, we have been able to differentiate between different influenza viral strains, detect

genetic markers related to high pathogenicity in influenza, and detect viral nucleoprotein binding to anti influenza apatamers.[105,183,184]

This work aims to develop a glycan microarray to investigate sialylated-glycan binding to influenza HA. We demonstrate SERS active substrate manufactured from CTAB coated gold nanoparticle (Au) films can be used for rapid and sensitive detection of glycan-HA interactions without the use of labeling. These findings will add to the current literature of SERS detection of glycan microarrays.

Experimental Methods

<u>Materials.</u> Hexadecyltrimethylammonium bromide (CTAB) and (1-mercaptoundec-11-yl)tetra(ethyleneglycol) were purchased from Sigma Aldrich. Gold nanoparticles with a diameter of 60nm were purchased from Ted Pella. Gold (99.999%) and Titanium (99.999%) evaporation pellets were purchased from Kurt J. Lesker. Dimethylformamide(DMF) was obtained from J.T. Baker. Sterile water which was used to make HA binding buffers was purchased from Braun.

<u>Fabrication of SERS Substrate.</u> Glass microscope slides were cut into $1 \times 1 cm^2$ pieces and cleaned with heated Piranha solution (4:1 conc. H_2SO_4 : 30% H_2O_2) followed by rinsing with MilliQ 18.2 megaohm water and dried by a gentle N_2 stream. A custom designed thermal evaporator was used to produce layers of 20nm Cr and 200nm Au prior to the fabrication of the AuNP monolayer. The thickness of the film was determined by a quartz crystal microbalance (QCM) which was placed inside the thermal evaporator chamber. After fabrication of the gold film, 20µl of a disulfide amine linker was dropped onto each substrate for 3 hours to produce a modified surface that would allow for capturing of

AuNPs. The substrates were placed into a petri dish. Inside the petri dish humidity levels were maintained by placing water droplets around substrates, covering the petri dish with a Parafilm® sheet, and sealing the petri dish lid.

AuNPs monolayers were made using a water/hexane interface based on an ionic surfactant mediated Langmuir-Blodgett method. AuNPs were centrifuged for 10 minutes at 8000 rpm to produce a concentration of $1x10^{11}$ particles/ml. After centrifugation 7ml of the colloidal solution was placed into a 5cm petri dish. To this solution was added 350µl of CTAB, 3.5ml of hexane. Then 3.5ml of ethanol was added dropwise at a rate of 350µl/min. After the Au film was made modified gold substrates were dipped into the solution and then pulled out to transfer the Au monolayer film onto the modified Au coated glass slides. Substrates were then allowed to dry at room temperature.

Immobilization of Sialylated Carbohydrates. Alpha 2,3 and α2,6 sialylated LacNAc were immobilized to the fabricated SERS substrate using strained promoted click chemistry (SPAAC). In this method, dibenzocyclooctyne is reacted with azides to conjugate compounds in a selective and simplistic manner. A disulfide DIBO linker was dissolved in DMF to obtain a final concentration of 4mM. The sialylated carbohydrates, which were functionalized with an azide linker, were dissolved in 18.2 megaohm MilliQ to obtain a desired concentration of 2mM. The DIBO linker and sialylated carbohydrates were mixed in a ratio of 2:1 in a centrifuge tube and allowed to react for at least 8hrs under gentle stirring. To minimize nonspecific binding and to ensure adequate spacing (1-mercaptoundec-11-yl)tetra(ethyleneglycol) was used as a spacer linker. This compound was dissolved into DMF to produce a final concentration of 0.04mM. Then 10µl of the spacer and clicked sialylated carbohydrate mixture was then added to the

fabricated SERS substrate. The substrates were placed into a petri dish and allowed to

react for at least 8hrs in a humidity maintained environment. After the 8 hrs, the substrate

was washed with 50 µl of water 3xs, dried with a gentle N_2 stream, and analyze by SERS.

Binding of Influenza Hemagglutinin. Influenza HA was purchased from Abcam®. Avian

influenza A HA (Avian HA) full length protein from H7N9 (Shanghai/017/2013) and

Human influenza A (Human HA) full length protein H1N1 (New Caledonia/20/1999)

were chosen for experiments. Avian HA was lyophilized and dissolved into a sterile

buffer consisting of 0.6%Tris, 0.7% glycine, and pH of 7.5 to reach a stock concentration

of 100µg/ml. Human HA was received in the liquid form and mixed with a PBS buffer to

reach a stock concentration of 100µg/ml.

Concentrations of 50µg/ml, 25µg/ml, 10µg/ml, 5µg/ml, 1µg/ml, and 0.1µg/ml

were prepared for both types of HA. Binding of the sialylated carbohydrate SERS sensing

surface to HA was accomplished by adding 10µl of the desired Avian HA concentration

to α2,3 sialylated carbohydrates or Human HA to α2,6 sialylated carbohydrates. SERS

substrates were incubated for at least 8hrs at 37ºC inside of a petri dish humidity

chamber. After incubation, HA solutions were removed by washing twice with the

respective binding buffers and a final wash with 18.2 megaohm MilliQ water. The

substrates were then dried by a gentle N_2 stream and analyzed by SERS.

Negative Control Carbohydrate Experiments. Since, influenza HA bind to cells via

sialylated carbohydrates, non sialylated lactose(Lac) was used as a carbohydrate control

for HA binding experiments. Lac was dissolved in MilliQ water to a final concentration

of 2mM and added to a 4mM DIBO linker solution for at least 8hrs in a centrifuge tube

under gentle shaking. After 8 hrs, a 0.04mM solution of the spacer linker was added to

the Lac mixture and 10µl of this mixture was placed on the fabricated SERS substrate. After 8 hrs, the substrate was rinsed three times with MilliQ water and dried using a gentle N_2 stream. Avian or Human HA of concentrations 10µg/ml, 1µg/ml, and 0.1µg/ml were incubated with the lactose substrate using the same binding protocol described earlier.

Negative Control HA Experiments. As a negative control, HA from human H1N1 was incubated with α2,3 sialylated carbohydrates. Because human H1N1 HA is specific for α2,6 sialylated carbohydrates it is not expected that binding will occur with α2,3 sialylated carbohydrates. Avian HA from H7N9 was not chosen for control experiments because literature has shown that it can bind with α2,6 sialylated carbohydrates. α 2,3 sialylated LacNAc (SA- LacNAc) was immobilized using the strain promoted click chemistry protocol described above. HA concentrations of H1N1used for these experiments were 10µg/ml, 1µg/ml, and 0.1µg/ml. Incubation of H1N1 HA was done according to the binding conditions mentioned above.

Raman Spectroscopy. An In Via Raman Microscope (Renishaw, Inc. Gloucestershire, UK) was used to collect SERS Spectra. A 785nm near-infrared (NIR) diode laser was used for the laser excitation. Data was collected using a 20x objective (NA=0.40) which produced a laser illumination spot with dimensions of 4.8 x 27.8µm. Spectra was collected from 10 different spots using 5% power, 15s acquisition time, one accumulation and a spectral range of 2000-600 cm^{-1}.

Data Analysis. Spectral averaging and baseline correction of SERS spectra was performed using GRAMS 32/AI spectral software package (Galactic industries, Nashua, NH). Before data analysis, raw SERs spectra were preprocessed using the first derivative

of each spectrum and a 15 point, 2nd order polynomial Savitsky-Golay algorithm. This was followed by normalization to unit vector length and mean centering. Preprocessing steps eliminate non-analyte spectral variances such as baseline and substrate variations. Data analysis was done in PLS toolbox 4.0 (Eigenvector, Wenatchee, WA) operating in Matlab. Multivariate statistical analysis was carried out using Principal Component Analysis (PCA) and Partial Least Squares Discriminant Analysis (PLS-DA). Feature selection was done by Analysis of Variance (ANOVA) along with post-hoc Tukey Honestly Significant Difference (HSD) test. A test for spectra variability was based on spectral amplitude (D) and the Wilcoxon rank sum test was used to determine the source of variability.

Results and Discussion

Reproducibility Study on CTAB Coated AuNP Substrate. For fabrication of AuNP monolayer, concentration of AuNP, concentration of CTAB, and the amount of ethanol, were considered. The manipulation of these parameters determined the quality of the AuNP film. Previous work in our lab has shown that AuNPs with a size of 60nm and a concentration of $1x10^{11}$ particles/ml were sufficient to produce metal nanoparticle films. After concentration and size of AuNPs were selected the next step was to study the effects of CTAB on the particles. CTAB capping of AuNPs is one of the driving forces to make the nanoparticle film. As a surfactant, CTAB was functionalized onto AuNPs by forming self-assembled bilayers. This reduces the negative surface charge of the nanoparticles and leads to a net positive charge preventing disordered aggregation by having a net repulsive interaction between nanoparticles. In the water/hexane interface

CTAB molecules move their polar heads to the aqueous phase and non-polar tail in the organic phase. Furthermore, CTAB is easily replaced by a target molecule as it binds to the surface of AuNP by Coulombic forces.

Studies were done to determine which concentration of CTAB should be used because values above the critical micellar concentration (cmc) of CTAB will induce AuNP aggregation. The cmc of CTAB is 1mM. For these studies, the concentration of AuNps in all experiments was kept constant at 1×10^{11} particles/ml. The extent of aggregation was determined analyzing the AuNP plasmon by UV-Vis spectroscopy. For a 60nm sized particle, the plasmon band is normally at 520nm. Aggregation was indicated by elongated or shifted bands and concentrations showing these features were rejected for further analysis.

Another driving force to make the nanoparticle film is ethanol, which was used as a low dielectric solvent. Ethanol changes the polarity of the aqueous phase and traps AuNPs at water/hexane interface.[95] When ethanol is added to the aqueous phase, it decreases the nanoparticle surface charge density causing their movement to the water/hexane interface.[185] Therefore, the volume of ethanol and the rate at which ethanol was added were studied to optimize the nanoparticle film.

To access the reproducibility of the SERS substrate fabrication method used in this paper, SERS spectra collected from substrates fabricated from same batch and different batches were compared. SERS collection parameters used for measurements were 785nm laser excitation source, 0.1% laser power, 10 second acquisition time, 1 accumulation, and a spectral range of 1800-600cm[-1] was collected. Ten microliters of a 1 mM 4-aminothiphenol (4-ATP) solution in ethanol was dropped onto the AuNP substrate

and allowed to evaporate. Three substrates were made for each batch and SERS spectra

were collected from eight different spots on each substrate. Ten microliters of the 4-ATP

solution were also placed onto a bare silicon wafer that had been functionalized with (3-

mercaptopropyl)-trimethoxy-silane (MPTMS) without AuNP for control experiments.

SERS spectra collected are shown in Figure 4.1. The spectrum of 4-ATP taken on the

CTAB AuNp substrate agrees well with what has been reported in literature. The

spectrum shows great enhancements over the SERS spectrum of 4-ATP taken on

substrates that did not contain AuNPs. The SERS spectra were averaged and the relative

standard deviation (RSD) was calculated and used to characterize the substrates

reproducibility. The SERS band at $1084cm^{-1}$ represents the (C-C) bond in 4-ATP.
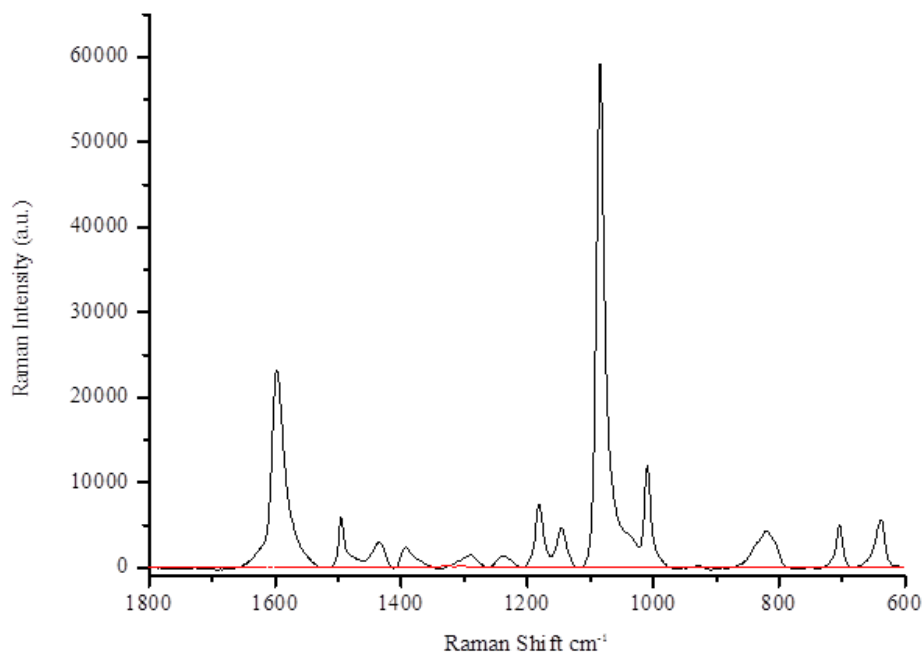
Figure 4.1 SERS spectra of 4-aminothiphenol on AuNP. The
black spectrum was taken of 4-ATP on the fabricated CTAB AuNP
substrate. The red spectrum was taken of 4-ATP on a substrate that
contained no nanoparticles.

This band is the most intense band in all the SERS spectra. Therefore, the intensity for

the band at 1084cm^{-1} was used as a measurement of the reproducibility of the SERS

substrates. The RSD across a signal substrate, from substrates fabricated in the same

batch, and substrates prepared in different batches are 9%, 16%, and 16% respectively.

Further characterization of the substrates was done by Scanning electron microscope

(SEM) imaging. Images of the AuNP substrate were taken from each substrate across the

four different batches. The images are shown in Figure 4.2.
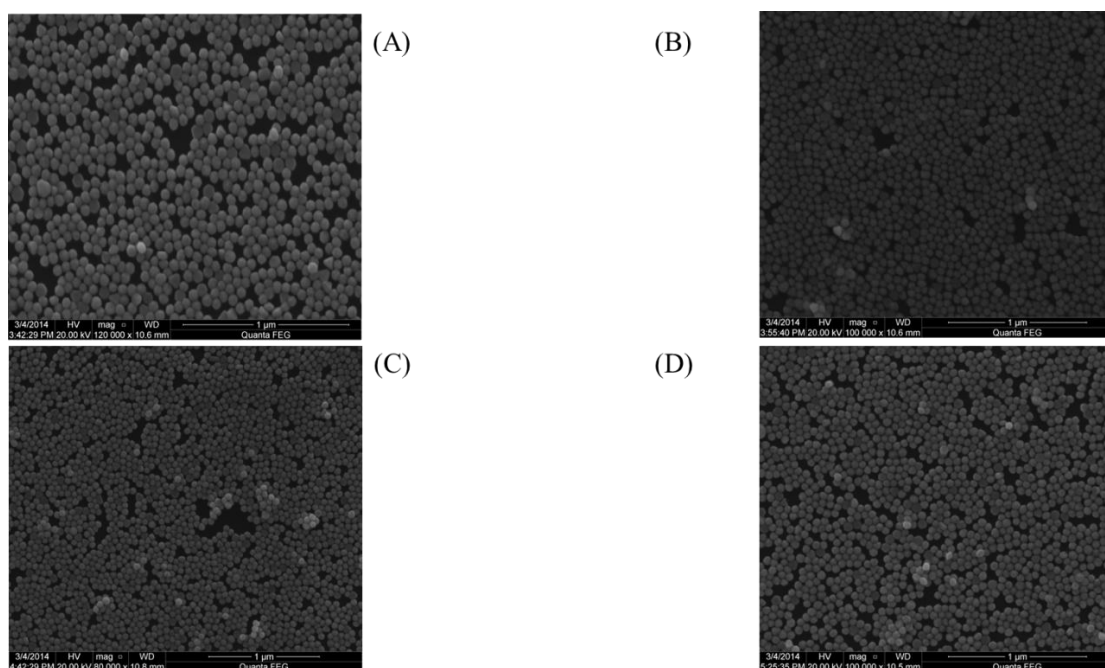
Figure 4.2 SEM Images of AuNP Batch to Batch comparison. (A) Batch 1, (B) Batch
2, (C) Batch 3, (D) Batch 4.

SEM images show the small presence of aggregates and gaps within the substrate as

expected. The aggregates can be removed with additional washing steps.

Immobilization of Sialylated Carbohydrates. Strain promoted azide alkyne click

chemistry was used to immobilize SA-LacNAc compounds to SERS substrates. Figure

4.3 shows the components used to make the influenza HA SERS sensing surface.

Figure 4.3. Compounds used in experiments: (A) DIBO, (B) spacer (C) 2,3 sialylated
LacNAc (D) 2,6 sialylated C, (E) Lac.

The azide group of SA-LacNAc reacts with the cyclooctyne. This effectively joins

carbohydrates to the DIBO linker and through the disulfide moiety of DIBO, SA-LacNAc

compounds are immobilized. The spacer molecule was immobilized with SA-LacNAc to fill in any empty spaces on the substrate. This was done to reduce nonspecific binding and ensure proper spacing between the SA-LacNAc on the surface. Samples for this study included: DIBO+Spacer, DIBO+Spacer+ α2,3SA- LacNAc, and DIBO-Spacer+ α2,6SA+ LacNAc. Ten SERS spectra were collected from different spots across each substrate. The average SERS spectra for each sample are shown in Figure 4.4.
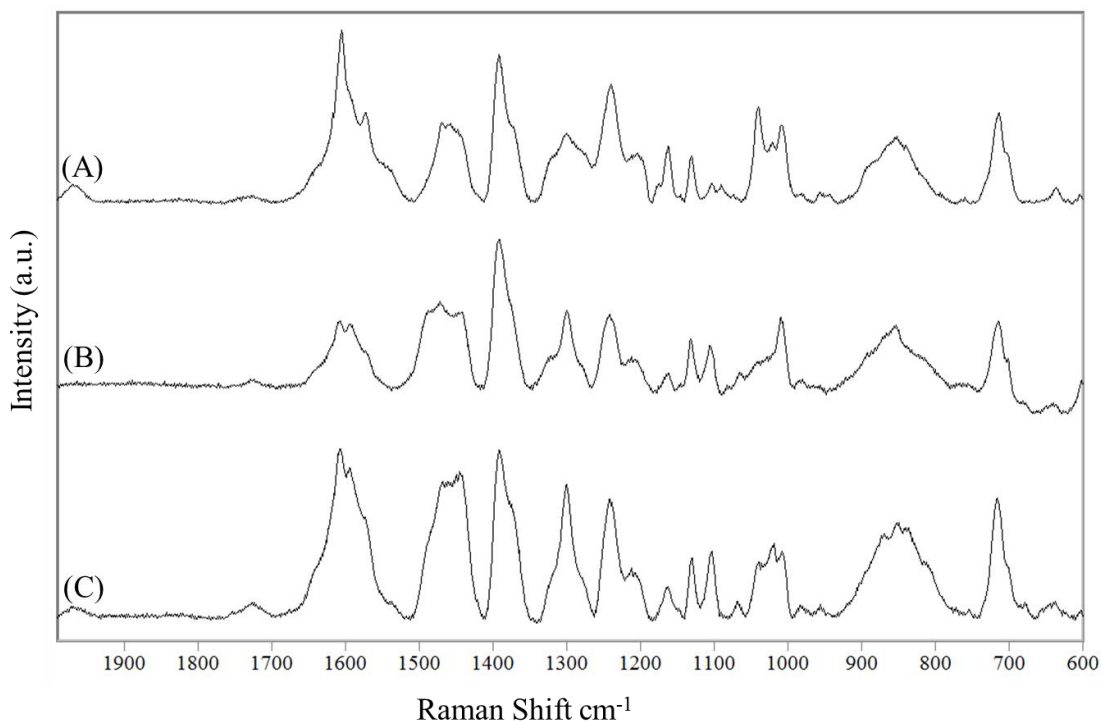
Figure 4.4. SERS spectra of: (A) DIBO+spacer, (B) DIBO+spacer+2,3sialylated LacNAc(C) DIBO+spacer+2,6sialylatedLacNAc. Each spectrum shown is an average of 10 individual spectra for each sample.

Carbohydrates consist mainly of carbon, hydrogen, and oxygen groups. These groups are not high Raman scatterers. Furthermore, precursors such as DIBO and the spacer group contain carbon, hydrogen, and oxygen as well. So, there will be some

overlap in the spectra. Therefore, to observe clicking between DIBO and LacNac, we monitored the alkyne band in DIBO. In the SERS spectrum for DIBO+spacer, the band around 1969cm^{-1} is assigned to the alkyne band. This band appears in part of the spectrum that is not crowded or overlapped with many other peaks. The reaction between azide and cyclooctyne causes the disappearance or reduction of the alkyne group. In Figure 4.4b, there is a noticeable reduction in the alkyne band after clicking the SA-LacNAc with DIBO. Because a higher concentration of DIBO was used in comparison to the concentration of SA-LacNAc, we do not expect the peak to completely disappear, however, a reduction is expected.

SERS Detection of Influenza Hemagglutinin. The SA-LacNAc nanoparticles were incubated with influenza HA. α2,3SA- LacNAc was incubated with Avian HA and α2,6SA- LacNAc was incubated with Human HA. Ten SERS spectra were collected for each sample. The concentration of HA ranged from 50μg/ml, 25μg/ml, 10μg/ml, 5μg/ml, 1μg/ml, and 0.1μg/ml for both types of HA. SERS spectra of the samples can be seen in Figure 4.5. In the figure, differences are identified in the spectra by the boxes. The differences created in the SERS spectra are caused by binding of the SA-carbohydrates. PLS-DA was used to analyze the SERS data. Models were built for each concentration. The models included samples for the DIBO-Spacer, DIBO-SA-LacNAc, and DIBO-SA-LacNAc-HA. Table 4.1 and 4.2 report the sensitivity and specificity scores from the PLS-DA models. The sensitivity scores for the Avian HA samples ranged from 80-100% and the specificity scores ranged from 89-100%. The sensitivity scores for the Human HA ranged from 97-100% and the specificity scores ranged from 89-100%. Therefore using the SERS method we can detect SA-LacNAc-HA binding.
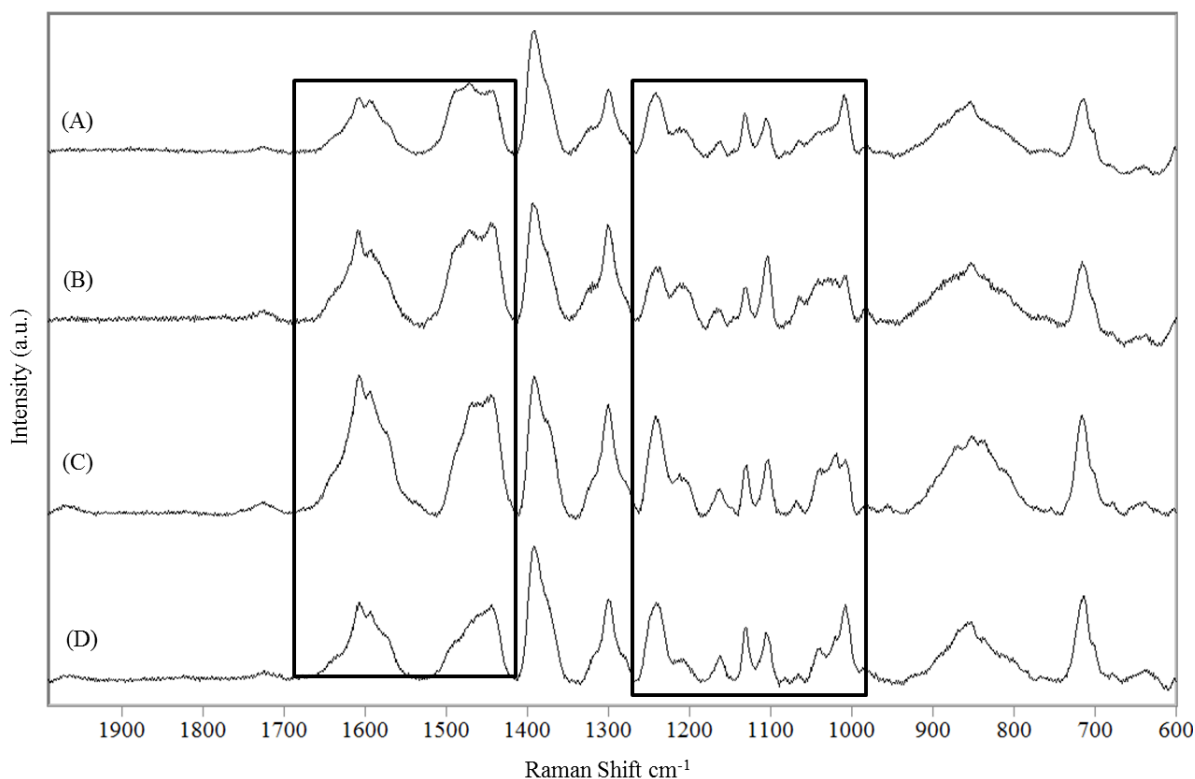
Figure 4.5. SERS spectra of: (A) DIBO+spacer+α 2,3sialylated-LacNAc, (B) DIBO +spacer+α 2,3sialylated-LacNAc-50μg/mlAvianHA (C)DIBO+spacer+α 2,6sialylated -LacNAc, (D) DIBO+spacer+α 2,6sialylated-LacNAc-50μg/mlHumanHA. Each spectrum shown is an average of 10 individual spectra for each sample.

Next, we identified which peaks from HA were important in discriminating SERS spectra. For this study, HA samples at 50μg/ml were used. Models built included the HA samples at 50μg/ml and DIBO-SA-LacNAc. A feature selection method based on the one-way ANOVA test was used. A feature is an individual measureable property of the SERS spectra that is observed. In the case of SERS spectra, the intensity values at each wavenumber are considered to be features. The spectral range of 2000-600 cm^{-1} collected 2902 data points for each spectrum. So feature selection was used identify which variables are important for predicting samples as influenza HA.[186]

Table 4.1. Summary of the PLS-DA cross validated results for classification of Avian HA with DIBO+Spacer, DIBO+ α2,3SA- LacNAc , and DIBO+Spacer+α2,3SA- LacNAc+ Avian HA.

Modeled Class	Sensitivity(CV)	Specificity(CV)
Dibo-spacer	1.000	1.000
α2,3 LacNAc(avg)	0.967	0.942
50µg/ml H7HA	1.000	1.000
25µg/ml H7HA	1.000	1.000
10µg/ml H7HA	0.800	0.850
1µg/ml H7HA	1.000	0.947
0.5µg/ml H7HA	1.000	0.950
0.1µg/ml H7HA	1.000	0.950

The one-way ANOVA test examines the statistical difference between the samples means. ANOVA tests the rejection of null hypothesis $H_o = \mu_1 = \mu_2 = \cdots = \mu_n$, where μ_n represents the mean spectra of nth class. A confidence level of 99% was used for the analysis. For data points that rejected the null hypothesis the post-hoc Tukey HSD test was performed. Tukey HSD test conducts a statistical pairwise comparison across

Table 4.2. Summary of the PLS-DA cross validated results for classification of Human HA with DIBO+Spacer, DIBO+α2,6SA- LacNAc , and DIBO+Spacer+α2,6SA-LacNAc+Human HA.

Modeled Class	Sensitivity(CV)	Specificity(CV)
Dibo-Spacer	1.000	1.000
Dibo-α2,6 LN(avg)	0.967	0.921
50µg/ml H1HA	1.000	0.950
25µg/ml H1HA	1.000	0.900
10µg/ml H1HA	1.000	0.895
1µg/ml H1HA	1.000	0.950
0.5µg/ml H1HA	1.000	1.000
0.1µg/ml H1HA	1.000	1.000

each wavenumber to identify the wavenumber locations in which all classes differ most from each other. Combining these two methods only wavenumbers that there were statistically different were retained and it was possible to identify wavenumbers that had the most significant difference between DIBO-SA-LacNAc and DIBO-SA-LacNAc-HA.[187]

For the model built using DIBO- α2,3-SA-LacNAc and DIBO-α2,3-SA-LacNAc-Avian HA, 181 wavenumbers were selected. The model built using DIBO- α2,6-SA-LacNAc and DIBO-α2,6-SA-LacNAc-Huma HA, returned 303 selected. Figures 4.6 and 4.7 show the spectrum created by the selected wavenumbers and peak assignments for each model. Peak assignments were made using wavenumbers common of amino acids of

(A)

(B)

Wavenumber cm^{-1}	Peak Assignment
639	Tyr
866	Tyr
1293	CH$_2$ wag
1454	CH
1614	Tyr, Trp, Phe

Intensity (a.u.)

1900 1800 1700 1600 1500 1400 1300 1200 1100 1000 900 800 700

Raman Shift cm^{-1}
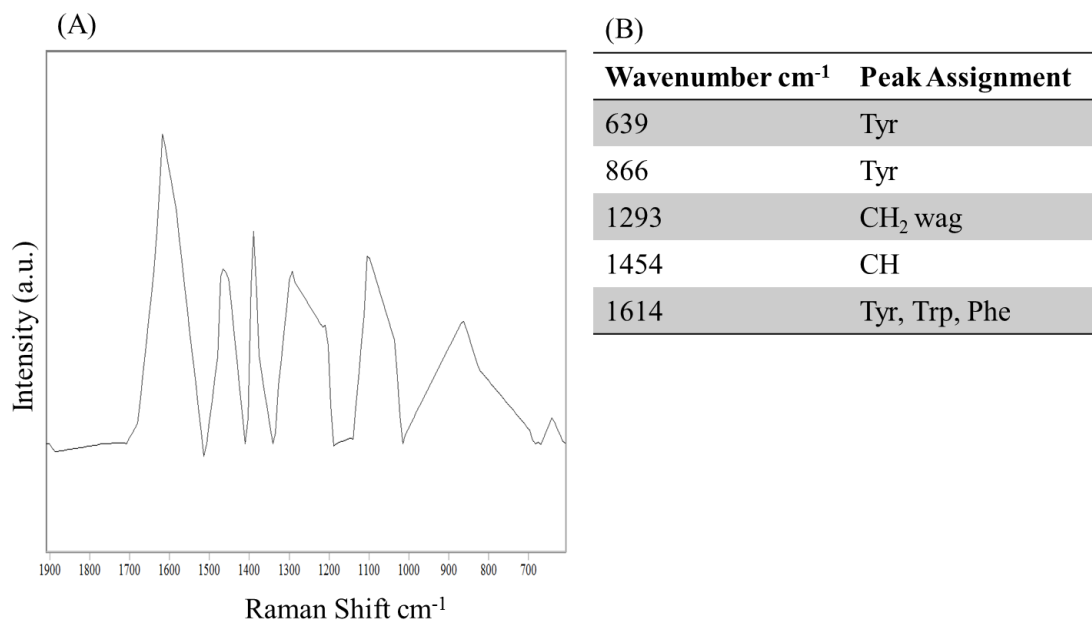
Figure 4.6. Summary Feature Selection for Avian HA with DIBO+α2,3SA- LacNAc. (A) Spectrum of the selected wavenumbers. (B) Table of peak assignments.

proteins. The peaks identified belonged to phenylalanine, tyrosine, tryptophan, CH$_2$ wagging and CH.

It was important to determine if SERS could distinguish between the two HAs at lower concentrations, therefore, influenza HA concentrations at 10µg/ml, 1µg/ml, and 0.1µg/ml were selected. The ability of SERS to differentiate between Avian and Human HA was analyzed. Two models were built for each concentration and samples included

Wavenumber cm^{-1}	Peak Assignment
639	Tyr
866	Tyr
1293	CH$_2$ wag
1454	CH
1614	Tyr, Trp, Phe
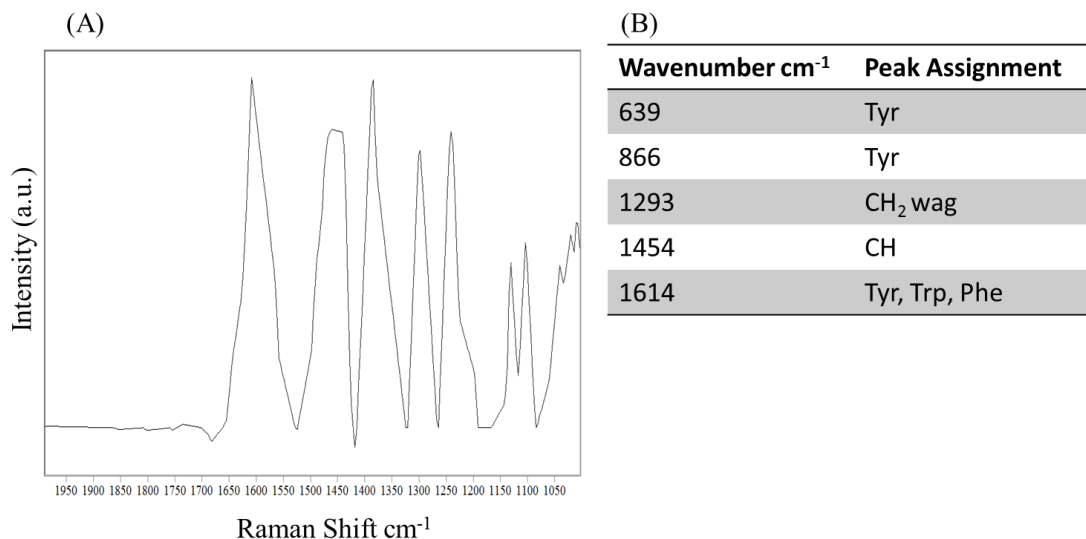
Figure 4.7. Summary Feature Selection for Human HA with DIBO+α2,6SA- LacNAc. (A) Spectrum of the selected wavenumbers. (B) Table of peak assignments.

DIBO-α2,3-SA-LacNAc-Avian HA and DIBO-α2,6-SA-LacNAc-Human HA. The unsupervised method PCA was used to view the data to determine if there was any clustering within the samples. The PCA scores plots for each model are shown in Figure 4.8. In the figure, the green squares represent the Human HA samples and the red diamonds represent the Avian HA samples. The PCA scores plots are plotted along PC1 and PC2 and represent 77-85% of the spectral variance. The plots show Human and Avian HA samples separate along PC1. This indicates that there is enough spectral difference between Avian and Human HA to cause this separation. From this data, we can conclude that SERS can be used to distinguish Avian and Human HA.

SERS Analysis of Nonspecific Binding. Two experiments were done to examine nonspecific binding. Non sialylated Lac was incubated with Avian or Human HA influenza samples. Non-SA-Lac was clicked on the DIBO nanoparticles in a similar
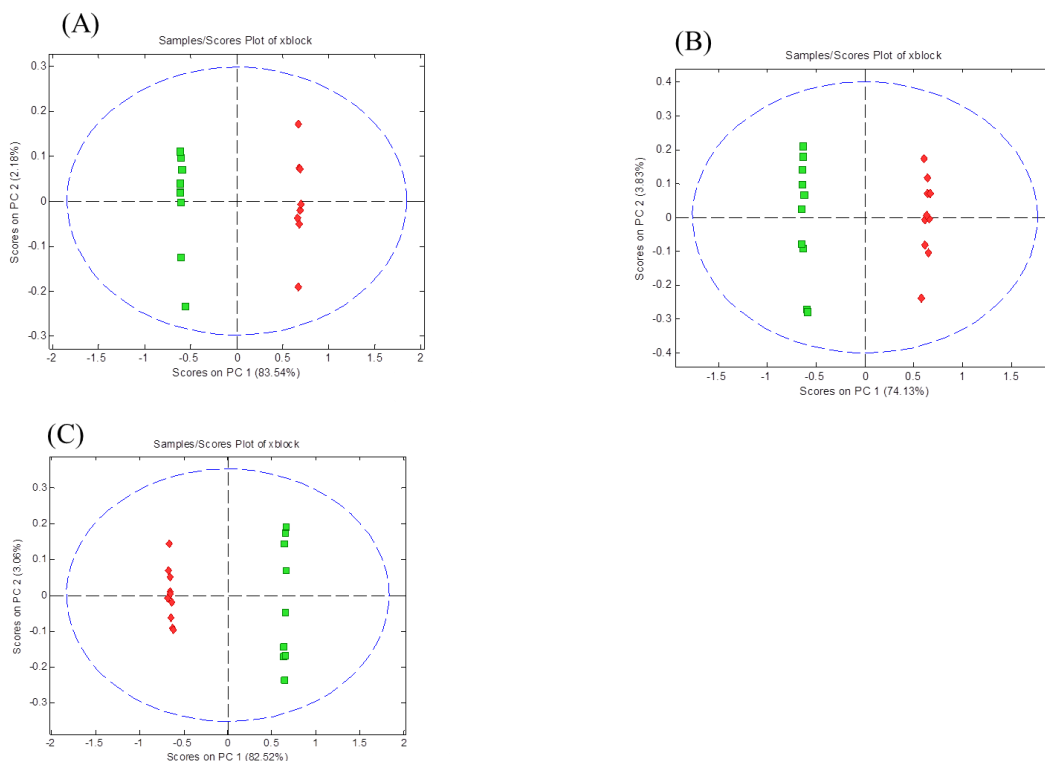
Figure 4.8. PCA scores plots for Avian HA vs Human HA. The ◆ represent Avian HA samples and ■ represents Human HA. (A) 10µg/ml, (B) 1µg/ml, and (C) 0.1µg/ml.

manner to the SA carbohydrates in previous experiments. Figure 4.9 shows the SERS spectra of DIBO-spacer and DIBO-spacer-non-SA-Lac. As indicated before, the alkyne peak disappeared when non-SA-Lac was clicked onto the nanoparticles. These nanoparticles were incubated with Avian or Human HA influenza samples at concentrations of 10µg/ml, 1µg/ml, and 0.1µg/ml. In the Figure 4.9C and 4.9D are the spectra that contain the influenza HA. The spectra are very similar to the SERS spectrum for DIBO-spacer-non-SA-Lac shown in 4.9B. Minor differences in the spectrum are indicated around 1000-1100cm^{-1}.
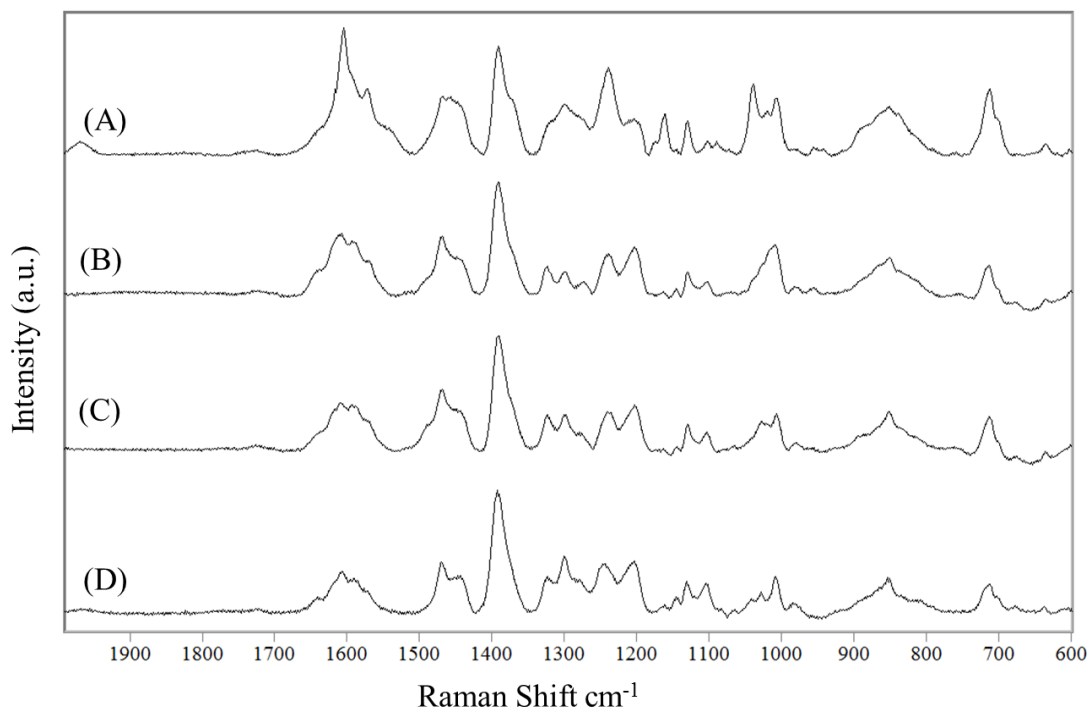
Figure 4.9. SERS spectra of: (A) DIBO+spacer and (B) DIBO+spacer+Lac,
(C) DIBO+spacer+Lac+AvianHA. (D)DIBO+spacer+Lac+HumanHA.
Each sample contained a HA concentration of 10µg/ml and the spectrum shown is an
average of 10 individual spectra for each sample.

PCA was used to examine clustering of the SERS spectrum for classification

purposes. The avian model was built using Avian HA, Human HA, DIBO-spacer-non-

SA-Lac, and DIBO-spacer-non-SA-Lac+Avian HA. PCA scores plots for these samples

are shown in Figure 4.10. The models captured 68-80% of the spectral variance. There

are a number of observations from the figure as it relates to the DIBO-spacer-non-SA-

Lac+Avian HA samples which are represented by the black circles in the figure. These

samples do not cluster with the avian or human HA samples as expected. The DIBO-

spacer-non-SA-Lac+Avian HA samples cluster with the DIBO-spacer-non-SA-Lac
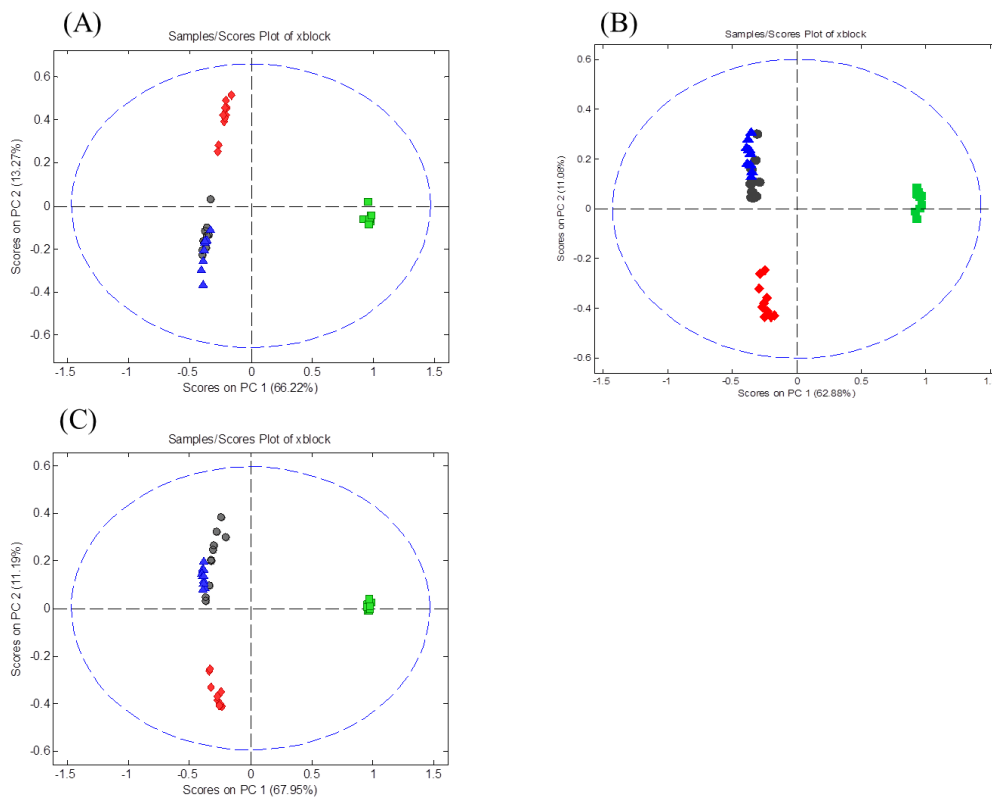
Figure 4.10. PCA scores plots for Avian HA vs Lac. The ☐ represents Human HA samples and ◆ represents Avian HA. The ▲ represents DIBO+spacer+ Lac, and ⬤ represents DIBO+spacer+ Lac +Avian. (A) 10µg/ml, (B) 1µg/ml, and (C) 0.1µg/ml.

samples. This indicates that the PCA does not identify Avian HA in the samples and the similarities between DIBO-spacer-non-SA-Lac+Avian HA and DIBO-spacer-non-SA-Lac are greater than the similarities between DIBO-spacer-non-SA-Lac+Avian HA and the Avian HA samples. As the concentration of the Avian HA to 0.1µg/ml there is some mixing of the samples with the Avian HA samples. This could indicate that the protein presence is very minimal causing them to cluster together. However, the samples that contain the Avian HA still do not cluster with the Human HA. So, as observed before, Avian and Human HA samples can still be separated.

A similar analysis was carried out for Non-SA-LacNAc with Human HA. The PCA scores plots are shown in Figure 4.11. The models captured 74-79% of the spectral variance. Samples included in these models included Avian HA, Human, HA, DIBO-spacer-non-SA-Lac, and DIBO-spacer-non-SA-Lac+Human HA. The DIBO-spacer-non-SA-Lac+Human HA samples in the plots do not cluster with Avian or Human HA. Furthermore, for all concentrations the DIBO-spacer-non-SA-Lac+Human HA samples clustered with the DIBO-spacer-non-SA-Lacfor every concentration. This may indicate that preference of H1N1 HA (New Caledonia/20/1999) to bind to sialylated carbohydrates is greater than H7N9 HA (Shanghai/017/2013).
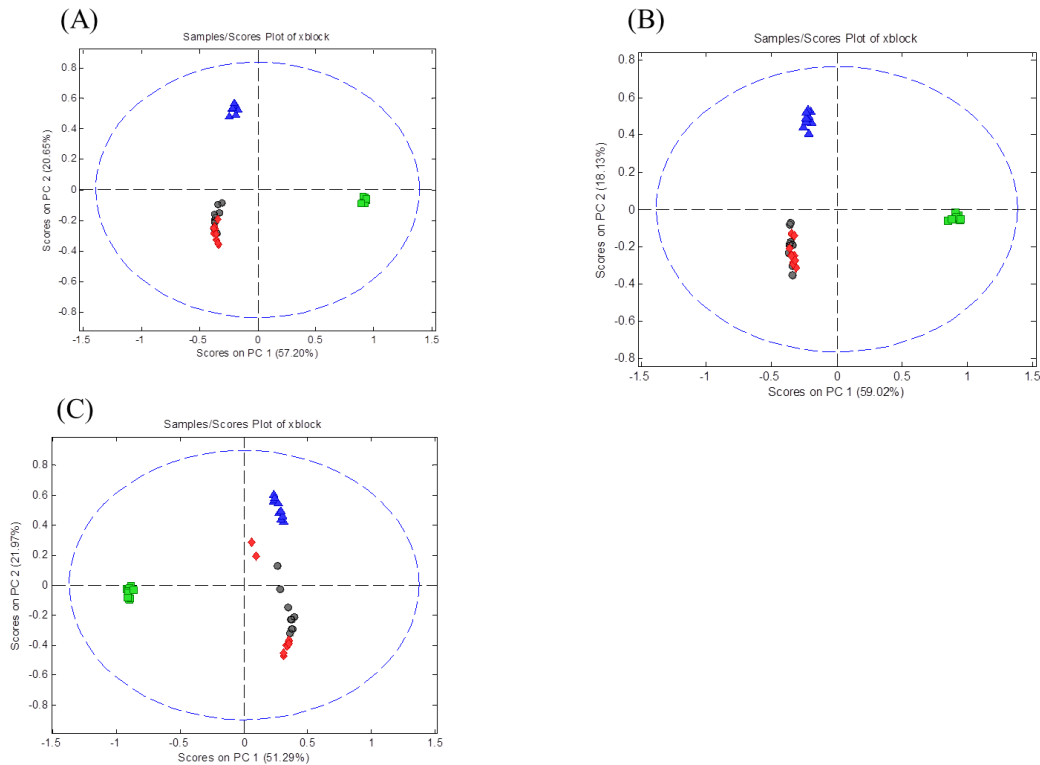
Figure 4.11. PCA scores plots for Human HA vs Lac. The ■ represent Avian HA samples and ◆ represents Human HA. The ▲ represents DIBO+spacer+Lac, and ● represents DIBO+spacer+Lac+Human HA. (A) 10μg/ml, (B) 1μg/ml, and (C) 0.1μg/ml.

Since, literature have proven that H7N9 HA (Shanghai/017/2013) can bind to both α2,3-SA-LacNAc and α2,6-SA-LacNAc the specificity of H1N1 HA (New Caledonia/20/1999) was examined.

Nanoparticles containing α2,3-SA-LacNAc was incubated with H1N1 HA using the same procedures outlined earlier. It is expected H1N1 HA will not bind with α2,3-SA-LacNAc nanoparticles. The PCA model was built using Avian HA, Human, HA, DIBO+spacer+ α2,3-SA-LacNAc, and DIBO+spacer+ α2,3-SA-LacNAc+Human HA. The models captured 72-77% of the spectral variance. In all the PCA models, the DIBO+spacer+ α2,3-SA-LacNAc+Human HA samples are indicated by the black circles and they do not cluster with Avian or Human HA samples. The samples cluster with the DIBO+spacer+ α2,3-SA-LacNAc samples. This indicates that H1N1 HA (New Caledonia/20/1999) does not bind with α2,3-SA-LacNAc.

Analysis of Spectral Variability. Up to this point we have used SERS spectra to detect influenza HA on sialylated nanoparticles as well as differentiate between Human and Avian HA. Models built using PCA and PLSDA analyze spectral variance within the spectra. It is important to ensure spectral variance is not caused by the substrate but by the SERS signals from the analyte. Therefore, a study of spectral variability was done to examine within sample variability and among sample variability. The spectral variability was measured according to the D metric developed by Price et al.[188] The D metric was as follows:

$$D = \left[\frac{1}{\lambda_b - \lambda_a} \int_{\lambda_a}^{\lambda_b} [S_1(\lambda) - S_2(\lambda)]^2 d\lambda\right]^{1/2} \text{ Equation(4.1).}$$

D is the root mean square difference between a pair of spectra (S_1 and S_2) averaged over the spectral range (λ_a toλ_b).[187] The D metric scores were calculated for the following

models: α2,3SA- LacNAc +Avian HA, α2,6SA- LacNAc+Human HA, and Human HA +Avian HA. These models underwent the same preprocessing steps that were outlined for the PCA and PLSDA studies.

D metric values were calculated for within samples and among samples for all pairwise combinations. Pairwise combinations were calculated as n[(n-1)/2] where n is the total number of samples. A total 90 within samples and 100 among samples pairwise combinations were obtained. After calculation of the D metric, Wilcoxon rank sum test (p value <0.05) was performed to analyze whether within samples and among samples were statistically significant. If the values are statistically significant, this indicated that spectra variability calculated in the PCA and PLSDA studies were from the analyte and not the SERS substrate. Tables 4.3-4.5 provide a summary of the D metric values calculated as well values from the Wilcoxon rank sum test. The p values for the within and among samples are <0.05 indicating that the values are statistically significant. Therefore, it was concluded that the spectral variability observed could not be attributed to the substrate but the analyte on the substrate.

Table 4.3. Summary D metric values for Avian HA.

Avian Influenza Spectral Variability Values

Category		Average D Values	Standard Deviation D Values	p-value
50ug/ml	within classes	1.068E-02	4.843E-03	2.04E-06
	among classes	1.169E-02	4.350E-03	
25ug/ml	within classes	7.611E-03	8.265E-04	1.53E-25
	among classes	9.060E-03	4.949E-04	
10ug/ml	within classes	6.632E-03	1.527E-03	5.50E-05
	among classes	7.815E-04	7.310E-04	
1ug/ml	within classes	5.995E-04	7.071E-04	1.30E-06
	among classes	9.397E-04	8.068E-04	
0.5ug/ml	within classes	8.350E-03	1.190E-03	1.31E-18
	among classes	9.998E-03	1.001E-03	
0.1ug/ml	within classes	8.221E-03	1.873E-03	8.74E-12
	among classes	1.022E-02	1.022E-02	

Table 4.4. Summary D metric values for Human HA.

Human Influenza Spectral Variability Values

Category		Average D Values	Standard Deviation D Values	p-value
50ug/ml	within classes	9.023E-03	1.569E-03	3.41E-10
	among classes	1.031E-02	1.226E-03	
25ug/ml	within classes	8.293E-03	1.714E-03	4.75E-04
	among classes	9.427E-03	2.192E-03	
10ug/ml	within classes	6.620E-03	1.500E-03	0.0280
	among classes	7.128E-03	1.412E-03	
1ug/ml	within classes	6.937E-03	1.964E-03	9.53E-04
	among classes	7.941E-03	2.148E-03	
0.5ug/ml	within classes	5.703E-03	1.899E-03	6.28E-06
	among classes	6.734E-03	1.892E-03	
0.1ug/ml	within classes	5.300E-03	1.800E-03	3.12E-04
	among classes	6.167E-03	1.836E-03	

Table 4.5. Summary D metric values for Avian HA and Human HA.

	Avian vs Human Influenza Spectral Variability Values			
Category		Average D Values	Standard Deviation D Values	p-value
10ug/ml	within classes	8.41E-03	2.312E-03	1.36E-32
	among classes	2.50E-02	2.350E-04	
1ug/ml	within classes	9.98E-03	2.429E-03	1.36E-32
	among classes	2.52E-02	2.938E-04	
0.1ug/ml	within classes	8.10E-03	2.000E-03	1.36E-32
	among classes	2.55E-02	3.153E-04	

Conclusions

We demonstrated the use of CTAB AuNP substrate for SER detection of SA-Lac-NAc and influenza HA. We tuned parameters of AuNP concentration, CTAB concentration, and volume of ethanol to produce uniform SERS substrates. Reproducibility of the fabrication method was evaluated by comparing SERS signals across a single substrate from substrates fabricated in the same batch, and substrates prepared in different batches. The RSD across a signal substrate, from substrates fabricated in the same batch, and substrates prepared in different batches were 9%, 16%, and 16% respectively. Furthermore, we established a method to evaluate spectral variability within the substrate. D metric values calculated indicated that the SERS substrates used in these experiments did not contribute to spectral variability examined in PCA and PLSDA.

A method based on SPAAC immobilization of SA-LacNAc onto SERS substrates was established to identify influenza HA binding and differentiation between Avian and

Human HA. Avian HA H7N9 (Shanghai/017/2013) and Human HA H1N1 (New Caledonia/20/1999) were used in this study. PCA and PLSDA were used to assess binding between influenza HA and SA-Lac-NAc. PLSDA returned sensitivity and specificity scores of 80-100% and 89-100% for Avian HA samples. Sensitivity and specificity score of 97-100% and the specificity scores ranged from 89-100% for Human HA samples. Furthermore, PCA was able to differentiate between H7N9 HA and H1N1 HA. Also, feature selection was used to identify which regions of the spectral range contributed to the most difference between the different samples. From this, we were able to identify that those differences were attributed to amino acids in the influenza HA. This study demonstrated the feasibility of using SERS as a detection method for carbohydrates and their interactions with various biomolecules. Future work will examine SERS analysis of longer chained SA-Lac-NAc compounds which are supposed to have higher affinities to influenza HA.

CHAPTER 5

CONCLUSIONS

The principal objective for this project was to develop a method for SERS detection of carbohydrate and their interactions with proteins. To develop this method, we examined the immobilization of carbohydrates by strain promoted click chemistry (SPCC). The practical use of this study was investigated by immobilization of biotin derivatives to gold nanoparticles. The azide and cyclooctyne constituents of the SPCC reaction provided a way to monitor the reaction via SERS. Furthermore, unique bands from biotin were also observed in the SERS spectra. The presence of avidin was characterized by SERS, UV-Vis analysis, and fluorescence. This project served as a proof of concept. From it, we developed appropriate mixtures of spacer and azide derivatives that were needed to click biotin to the substrate and used them in other projects. Using gold nanoparticles as SERS substrates also allowed for versatility in characterization methods. The surface plasmon of nanoparticles in UV-Vis can be used to monitor nanoparticle derivatization. Also, fluorescence is the most common method of detection for carbohydrate arrays. So combining these three techniques a greater combination of information was able to be obtained.

After we determined parameters for SPCC reactions, we next examined the immobilization of carbohydrates to SERS substrates and studied their interactions with various lectins. It was important to immobilize carbohydrates to the nanoparticles but detection of lectins using the carbohydrate nanoparticles as a sensor was important.

Lactose was immobilized to nanoparticles using SPCC. SERS monitoring of lactose immobilization was based on observing the alkyne peak in the SERS spectra. Galectin-1 and galectin-3 were later incubated with the surface. Using SERS we were able to detect the galectin proteins as well as differentiate between the two. Furthermore, we were able to quantify how much protein was captured onto the substrate. SERS and fluorescence measurements were able to determine that proteins not specific for lactose bound to the nanoparticles with less affinity. This study also demonstrated how an array can be fabricated. To develop an array, a gold slide was patterned to create a high throughput multiwell substrate. We were able to analyze all three proteins used in this study at one time thus conducting time efficient measurements.

After the previous study, we sought to develop a SERS substrate where the nanoparticles were already immobilized onto a solid surface and later functionalized with the SPCC constituents. This was done to increase substrate fabrication reproducibility. Therefore, a method was developed to create gold nanoparticle monolayers which were later transferred to a substrate for SERS. We were able to make small areas of monolayers that were sufficient for SERS detection. The reproducibility of these substrates was tested from spot to spot across a single substrate, from substrate to substrate, and from batch to batch. The standard deviation ranged 9-16% which is much lower in comparison to the silver nanorods fabricated in our lab for SERS substrates. The standard deviation values for those substrates ranged from 24-40%.[189]

The final application demonstrated by this body of work immobilized sialylated carbohydrates to a SERS substrate to create sensors for influenza hemagglutinin. Influenza hemagglutinin binds to cells via sialylated carbohydrates. Specificity of the

type of influenza hemagglutinin (Avian or Human) is determined by the sialic acid linkage. To investigate this system α2,3SA- LacNAc and α2,6SA- LacNAc were immobilized to nanoparticles using SPCC. Then they were incubated with influenza HA depending on their specificity. This study showed that α2,3SA- LacNAc which is specific for Avian HA also bound to Human HA. These results were not shown but were verified by previous literature. In our study, Human HA only bound to α2,6SA- LacNAc. When non-SA Lac was incubated with Avian or Human HA, the binding was very minimal if at all.

This project began a body of work to develop SERS as a detection method for carbohydrate-lectin interactions. The study of carbohydrates by SERS is not widely done. So this work can serve as a starting point into the field. There are some improvements and future studies that can be done to increase this body of knowledge. There is a need to increase the sampling size. While we are able to carry out most experiments in duplicates, a larger sampling class is still needed to be able to make generalizations about the data analysis. The synthesis of carbohydrates is a lengthy process and often times very little sample is available to carry out extensive SERS experiments. Additionally, the average experiment including sample prep time, SERS collection, and data analysis took as much as one week to carry out. Therefore improvements can be made to determine the minimal sample requirement and sample prep time needed to carry out the experiments.

Clicking of the SPCC constituents was done in solution and not on the SERS substrate. The SA-carbohydrate experiments were done on the CTAB substrate, however, we were not able to click to the DIBO linker on the substrate. In the biotin and lactose studies, clicking was carried out on nanoparticles that were in solution. There was no

evidence of problems with the reaction between the alkyne and azide when click in solution. The surface roughness of nanoparticles immobilized onto a solid support may be problematic for clicking to occur. This is evident by earlier attempts to click DIBO-biotin onto silver nanorod arrays. Like the CTAB substrate, the silver nanorods are fixed onto a solid support. In these experiments, clicking between azide and DIBO was difficult to obtain, hence the move to gold colloids in solution. To reach the full potential of this method, it is important to establish that this reaction can occur onto nanoparticles that have been immobilized onto solid surfaces. So, this phenomenon should be examined further.

The CTAB AuNP SERS substrate developed for the final project has not been developed into an array format. Since, little carbohydrate samples were available arrays are needed. They have the advantage of being able to carry out multiple experiments at once so more data will be collected and minimal sample is required. Currently, our group is able to fabricate patterned arrays on 1 x 3 in. glass slides. To reach a comparable size, the CTAB AuNP monolayer would require scaling up fabrication materials. This could include greater volume of AuNPs, ethanol, hexane, and CTAB. Furthermore the size of the crystalizing dish used for fabrication would have to be increased as well to allow the monolayer to grow to a bigger size. After the substrate is prepared, then establishing conditions for creating the PDMS multiwell substrate would have to be examined. Alternatively, patterning molds could be designed to fit substrates that are smaller than 1 x 3 iniches.

The influenza study can be expanded to include longer chained carbohydrates which have higher specificity for influenza HA due to their multivalent nature.

Furthermore, examination of other subtypes of Human and Avian influenza HA are needed as well to determine if the trend observed in this study can be generalized across different influenza subtypes. A lit search demonstrated that H7N9 (Avian influenza strain) used was able to bind to both α2,6SA- LacNAc and α2,3SA- LacNAc. It would be interesting to determine if other such trends are observable for other subtypes. Avian influenza viruses H5,H7, and H9 can be transmitted between mammals. How this occurs is currently not well understood but receptor binding specificity plays a role.[190]

Increasing the length of the carbohydrates can be problematic for SERS analysis. SERS enhancement decreases as the nanoparticle distance from the surface is increased. Liver at al. reported that molecules placed between two spherical particles experience greatly enhanced fields.[191] Ketating et al. demonstrated the dependence of the SERS signal increasing distance from the substrate can be overcome by placing the analyte between two nanoparticles.[192] In the Keating study, the protein cytochrome C was conjugated to gold colloids and later adsorbed onto aggregated silver colloids. This trapped cytochrome C between silver and gold nanoparticles. Several features of this method are of importance. The two colloidal nanoparticles could be placed a known distance apart and bimetallic nanoparticles allow for electromagnetic coupling to be identified through wavelength-dependent measurements. SERS spectra of cytochrome C conjugated to silver nanoparticles were compared to SERS spectra of cytochrome C sandwiched between the silver and gold nanoparticles. They observed that SERS signals from the sandwiched complex were weaker however, they were still observable. This method can be tried an improved upon to gain further insights into analyses where the target analyte is approaching the optimum distance for the SERS effect to be observed.

This method has practical applications to study the binding of longer chained SA-carbohydrates. After incubation of the SA-carbohydrates with influenza HA, SERS spectra should be taken as a standard protocol. After SERS spectra are taken, particles can be placed on top of the influenza HA surface and taken again to see if there are any enhancements of the HA signal. Because in this orientation, the HA signals are closer to the top layer of nanoparticles, their signals should be increased. An alternative to this method could be conjugating HA to the second layer of nanoparticles. This may involve some protein derivitation methods. Protein conjugation methods are similar to carbohydrate immobilization schemes. After protein modification the HA-nanoparticles would be incubated with the SA-nanoparticles and SERS spectra can be recorded.

REFERENCES

1   Malacinski, G. M. *Essentials of Molecular Biology*. 3rd ed. edn, (Boston, Mass., 1998).

2   Varki, A. *Essentials of Glycobiology*. 2nd ed. edn, (Cold Spring Harbor, N.Y., 2009).

3   Dwek, R. A. Glycobiology: Toward Understanding the Function of Sugars. *Chem Rev* **96**, 683-720 (1996).

4   Horlacher, T. & Seeberger, P. H. Carbohydrate Arrays as Tools for Research and Diagnostics. *Chemical Society Reviews* **37**, 1414-1422 (2008).

5   Struwe, W., Cosgrave, E. J., Byrne, J., Saldova, R. & Rudd, P. *Glycoproteomics in Health and Disease* 1-38 (Springer Netherlands, 2011).

6   Haase, C. & Seitz, O. *Chemical Synthesis of Glycopeptides*. Vol. 267 43 (Springer Berlin Heidelberg, 2007).

7   Roseman, S. Reflections on Glycobiology. *Journal of Biological Chemistry* **276**, 41527-41542 (2001).

8   Feizi, T., Fazio, F., Chai, W. C. & Wong, C. H. Carbohydrate Microarrays - a New Set of Technologies at the Frontiers of Glycomics. *Current Opinion in Structural Biology* **13**, 637-645 (2003).

9   Liang, P.-H., Wu, C.-Y., Greenberg, W. A. & Wong, C.-H. Glycan Arrays: Biological and Medical Applications. *Current Opinion in Chemical Biology* **12**, 86-92 (2008).

10   Weis, W. I. & Drickamer, K. Structural Basis of Lectin-Carbohydrate Recognition. *Annual Review of Biochemistry* **65**, 441-473 (1996).

11   Kiessling, L. L., Gestwicki, J. E. & Strong, L. E. Synthetic Multivalent Ligands as Probes of Signal Transduction. *Angewandte Chemie-International Edition* **45**, 2348-2368 (2006).

12   Plante, O. J., Palmacci, E. R. & Seeberger, P. H. Automated Solid-Phase Synthesis of Oligosaccharides. *Science* **291**, 1523-1527 (2001).

13   Chen, X. *et al.* Reassembled Biosynthetic Pathway for Large-Scale Carbohydrate Synthesis: A-Gal Epitope Producing "Superbug". *Chembiochem* **3**, 47-53 (2002).

14   Brown, P. O. & Botstein, D. Exploring the New World of the Genome with DNA Microarrays. *Nature Genetics* **21**, 33-37 (1999).

15   Zhu, H. & Snyder, M. Protein Chip Technology. *Current Opinion in Chemical Biology* **7**, 55-63 (2003).

16   Ratner, D. M., Adams, E. W., Disney, M. D. & Seeberger, P. H. Tools for Glycomics: Mapping Interactions of Carbohydrates in Biological Systems. *Chembiochem* **5**, 1375-1383 (2004).

17   Werz, D. B. & Seeberger, P. H. Carbohydrates as the Next Frontier in Pharmaceutical Research. *Chemistry – A European Journal* **11**, 3194-3206 (2005).

18   Park, S., Gildersleeve, J. C., Blixt, O. & Shin, I. Carbohydrate Microarrays. *Chemical Society Reviews* **42**, 4310-4326 (2013).

19   Palcic, M. M. Glycosyltransferases as Biocatalysts. *Current Opinion in Chemical Biology* **15**, 226-233 (2011).

20   Lairson, L. L., Henrissat, B., Davies, G. J. & Withers, S. G. Glycosyltransferases: Structures, Functions, and Mechanisms. *Annual Review of Biochemistry* **77**, 521-555 (2008).

21   Rinaldi, S. *et al.* Analysis of Lectin Binding to Glycolipid Complexes Using Combinatorial Glycoarrays. *Glycobiology* **19**, 789-796 (2009).

22   Houseman, B. T. & Mrksich, M. Carbohydrate Arrays for the Evaluation of Protein Binding and Enzymatic Modification. *Chemistry & Biology* **9**, 443-454 (2002).

23   Kilcoyne, M. *et al.* Construction of a Natural Mucin Microarray and Interrogation for Biologically Relevant Glyco-Epitopes. *Analytical Chemistry* **84**, 3330-3338 (2012).

24   Alvarez, R. A. & Blixt, O. *Identification of Ligand Specificities for Glycan‑Binding Proteins Using Glycan Arrays*. Vol. Volume 415 (Academic Press, 2006).

25   Wong, C.-H. & Bryan, M. C. *Sugar Arrays in Microtiter Plates* Vol. Volume 362 218-225 (Academic Press, 2003).

26   Fukui, S., Feizi, T., Galustian, C., Lawson, A. M. & Chai, W. Oligosaccharide Microarrays for High-Throughput Detection and Specificity Assignments of Carbohydrate-Protein Interactions. *Nat Biotech* **20**, 1011-1017 (2002).

27   Hsiao, H.-Y. *et al.* Fabrication of Carbohydrate Microarrays through Boronate Formation. *Chemical Communications* **47**, 1187-1189 (2011).

28   Nilsson, U., Striker, R. T., Hultgren, S. J. & Magnusson, G. Papg Adhesin from E. Coli J96 Recognizes the Same Saccharide Epitope When Present on Whole

Bacteria and as Isolated Protein. *Bioorganic & Medicinal Chemistry* **4**, 1809-1817 (1996).

29   Feizi, T., Fazio, F., Chai, W. & Wong, C.-H. Carbohydrate Microarrays — a New Set of Technologies at the Frontiers of Glycomics. *Current Opinion in Structural Biology* **13**, 637-645 (2003).

30   Park, S., Lee, M.-r., Pyo, S.-J. & Shin, I. Carbohydrate Chips for Studying High-Throughput Carbohydrate−Protein Interactions. *Journal of the American Chemical Society* **126**, 4812-4819 (2004).

31   Gruber, K. *et al.* Cantilever Array Sensors Detect Specific Carbohydrate−Protein Interactions with Picomolar Sensitivity. *ACS Nano* **5**, 3670-3678 (2011).

32   Fei, Y. *et al.* Fluorescent Labeling Agents Change Binding Profiles of Glycan-Binding Proteins. *Molecular BioSystems* **7**, 3343-3352 (2011).

33   Yuk, J. S. *et al.* Sensitivity of Ex Situ and in Situ Spectral Surface Plasmon Resonance Sensors in the Analysis of Protein Arrays. *Biosensors and Bioelectronics* **20**, 2189-2196 (2005).

34   Ramachandran, N., Larson, D. N., Stark, P. R. H., Hainsworth, E. & LaBaer, J. Emerging Tools for Real-Time Label-Free Detection of Interactions on Functional Protein Microarrays. *FEBS Journal* **272**, 5412-5425 (2005).

35   Karlsson, R. Spr for Molecular Interaction Analysis: A Review of Emerging Application Areas. *Journal of Molecular Recognition* **17**, 151-161 (2004).

36   Houseman, B. T., Gawalt, E. S. & Mrksich, M. Maleimide-Functionalized Self-Assembled Monolayers for the Preparation of Peptide and Carbohydrate Biochips†. *Langmuir* **19**, 1522-1531 (2002).

37   Kodoyianni, V. Label-Free Analysis of Biomolecular Interactions Using Spr Imaging. *Biotechniques* **50**, 32-40 (2011).

38   Yu, X., Xu, D. & Cheng, Q. Label-Free Detection Methods for Protein Microarrays. *PROTEOMICS* **6**, 5493-5503 (2006).

39   Ghazani, A. A. *et al.* High Throughput Quantification of Protein Expression of Cancer Antigens in Tissue Microarray Using Quantum Dot Nanocrystals. *Nano Letters* **6**, 2881-2886 (2006).

40   Sun, Y.-P. *et al.* Quantum-Sized Carbon Dots for Bright and Colorful Photoluminescence. *Journal of the American Chemical Society* **128**, 7756-7757 (2006).

41   Gonzalez-Gonzalez, M. *et al.* Nanotechniques in Proteomics: Protein Microarrays and Novel Detection Platforms. *European Journal of Pharmaceutical Sciences* **45**, 499-506 (2012).

42   Ray, S., Mehta, G. & Srivastava, S. Label-Free Detection Techniques for Protein Microarrays: Prospects, Merits and Challenges. *PROTEOMICS* **10**, 731-748 (2010).

43   Blixt, O. *et al.* Printed Covalent Glycan Array for Ligand Profiling of Diverse Glycan Binding Proteins. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 17033-17038 (2004).

44   Cunto-Amesty, G., Monzavi-Karbassi, B., Luo, P., Jousheghany, F. & Kieber-Emmons, T. Strategies in Cancer Vaccines Development. *International Journal for Parasitology* **33**, 597-613 (2003).

45   Dube, D. H. & Bertozzi, C. R. Glycans in Cancer and Inflammation [Mdash] Potential for Therapeutics and Diagnostics. *Nat Rev Drug Discov* **4**, 477-488 (2005).

46   Debets, M. F. *et al.* Bioconjugation with Strained Alkenes and Alkynes. *Accounts of Chemical Research* **44**, 805-815 (2011).

47   Kolb, H. C., Finn, M. G. & Sharpless, K. B. Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angewandte Chemie-International Edition* **40**, 2004-+ (2001).

48   Nandivada, H., Jiang, X. W. & Lahann, J. Click Chemistry: Versatility and Control in the Hands of Materials Scientists. *Advanced Materials* **19**, 2197-2208 (2007).

49   Banert, K. Reactions of Unsaturated Azides .6. Synthesis of 1,2,3-Triazoles from Propargyl Azides by Rearrangement of the Azido Group - Indication of Short-Lived Allenyl Azides and Triazafulvenes. *Chemische Berichte* **122**, 911-918 (1989).

50   Prescher, J. A. & Bertozzi, C. R. Chemistry in Living Systems. *Nature Chemical Biology* **1**, 13-21 (2005).

51   Moses, J. E. & Moorhouse, A. D. The Growing Applications of Click Chemistry. *Chemical Society Reviews* **36**, 1249-1262 (2007).

52   Yousaf, M. N. & Mrksich, M. Diels-Alder Reaction for the Selective Immobilization of Protein to Electroactive Self-Assembled Monolayers. *Journal of the American Chemical Society* **121**, 4286-4287 (1999).

53   Yousaf, M. N., Houseman, B. T. & Mrksich, M. Turning on Cell Migration with Electroactive Substrates. *Angewandte Chemie-International Edition* **40**, 1093-+ (2001).

54   Sun, X. L., Stabler, C. L., Cazalis, C. S. & Chaikof, E. L. Carbohydrate and Protein Immobilization onto Solid Surfaces by Sequential Diels-Alder and Azide-Alkyne Cycloadditions. *Bioconjugate Chemistry* **17**, 52-57 (2006).

55   Link, A. J. & Tirrell, D. A. Cell Surface Labeling of Escherichia Coli Via Copper(I)-Catalyzed 3+2 Cycloaddition. *Journal of the American Chemical Society* **125**, 11164-11165 (2003).

56   Wang, Q. *et al.* Bioconjugation by Copper(I)-Catalyzed Azide-Alkyne [3 + 2] Cycloaddition. *Journal of the American Chemical Society* **125**, 3192-3193 (2003).

57   Brewer, G. J. Risks of Copper and Iron Toxicity During Aging in Humans. *Chemical Research in Toxicology* **23**, 319-326 (2009).

58   Lallana, E., Riguera, R. & Fernandez-Megia, E. Reliable and Efficient Procedures for the Conjugation of Biomolecules through Huisgen Azide–Alkyne Cycloadditions. *Angewandte Chemie International Edition* **50**, 8794-8804 (2011).

59   Wittig, G. & Krebs, A. Zur Existenz Niedergliedriger Cycloalkine .1. *Chemische Berichte-Recueil* **94**, 3260-3275 (1961).

60   Turner, R. B., Goebel, P., Mallon, B. J. & Jarrett, A. D. Heats O Hydrogenation .9. Cyclic Acetylenes and Some Miscellaneous Olefins. *Journal of the American Chemical Society* **95**, 790-792 (1973).

61   Gutsmiedl, K., Wirges, C. T., Ehmke, V. & Carell, T. Copper-Free "Click" Modification of DNA Via Nitrile Oxide−Norbornene 1,3-Dipolar Cycloaddition. *Organic Letters* **11**, 2405-2408 (2009).

62   Ning, X. *et al.* Protein Modification by Strain-Promoted Alkyne–Nitrone Cycloaddition. *Angewandte Chemie International Edition* **49**, 3065-3068 (2010).

63   Sanders, B. C. *et al.* Metal-Free Sequential [3 + 2]-Dipolar Cycloadditions Using Cyclooctynes and 1,3-Dipoles of Different Reactivity. *Journal of the American Chemical Society* **133**, 949-957 (2010).

64   Wendeln, C. *et al.* Orthogonal, Metal-Free Surface Modification by Strain-Promoted Azide-Alkyne and Nitrile Oxide-Alkene/Alkyne Cycloadditions. *Chemical Science* **3**, 2479-2484 (2012).

65   Nafie, L. A. Recent Advances in Linear and Nonlinear Raman Spectroscopy. Part Iv. *Journal of Raman Spectroscopy* **41**, 1566-1586 (2010).

66   Sandalinas, C., Ruiz-Moreno, S., Lopez-Gil, A. & Miralles, J. Experimental Confirmation by Raman Spectroscopy of a Pb-Sn-Sb Triple Oxide Yellow Pigment in Sixteenth-Century Italian Pottery. *Journal of Raman Spectroscopy* **37**, 1146-1153 (2006).

67   Raman, C. V. & Krishnan, K. S. A New Type of Secondary Radiation (Reprinted from Nature, Vol 121, Pg 501-502, 1928). *Current Science* **74**, 381-381 (1998).

68   Le Ru, E. C. *Principles of Surface-Enhanced Raman Spectroscopy and Related Plasmonic Effects*. 1st ed. edn, (Amsterdam ;, 2009).

69   Inc., B. *Raman Knowledge*, <http://bwtek.com/raman-theory-of-raman-scattering/> (2014).

70   McCreery, R. L. *Raman Spectroscopy for Chemical Analysis*. (New York, 2000).

71   *Confocal Raman Microscopy*. (Berlin ;, 2011).

72   Kilmer, J., Iadevaia, A. & Yin, Y.   803914-803914-803911.

73   McNichols, R. J. & Cote´, G. L. Optical Glucose Sensing in Biological Fluids: An Overview. *Journal of Biomedical Optics* **5**, 5-16 (2000).

74   Screen, J. *et al.* Ir-Spectral Signatures of Aromatic-Sugar Complexes: Probing Carbohydrate-Protein Interactions. *Angewandte Chemie-International Edition* **46**, 3644-3648 (2007).

75   LeRu, E. C. & Etchegoin, P. G. *Principles of Surface-Enhanced Raman Spectroscopy: And Related Plasmonic Effects*. (Elsevier Science Bv, 2009).

76   Nima, Z. A. *et al.* Applications of Surface-Enhanced Raman Scattering in Advanced Bio-Medical Technologies and Diagnostics. *Drug Metabolism Reviews* **46**, 155-175 (2014).

77   Kneipp, K., Kneipp, H., Itzkan, I., Dasari, R. R. & Feld, M. S. Ultrasensitive Chemical Analysis by Raman Spectroscopy. *Chemical Reviews* **99**, 2957-+ (1999).

78   Kneipp, K., Kneipp, H., Itzkan, I., Dasari, R. R. & Feld, M. S. Surface-Enhanced Raman Scattering and Biophysics. *Journal of Physics-Condensed Matter* **14**, R597-R624 (2002).

79   Campion, A. & Kambhampati, P. Surface-Enhanced Raman Scattering. *Chemical Society Reviews* **27**, 241-250 (1998).

80   Vo-Dinh, T. Surface-Enhanced Raman Spectroscopy Using Metallic Nanostructures. *Trac-Trends in Analytical Chemistry* **17**, 557-582 (1998).

81   Pettinger, B. & Krischer, K. Comparison of Cross-Sections for Absorption and Surface-Enhanced Resonance Raman-Scattering for Rhodamine 6g at Coagulated Silver Colloids. *Journal of Electron Spectroscopy and Related Phenomena* **45**, 133-142 (1987).

82   Otto, A., Timper, J., Billmann, J., Kovacs, G. & Pockrand, I. Surface-Roughness Induced Electronic Raman-Scattering. *Surface Science* **92**, L55-L57 (1980).

83   Ritchie, R. H. Plasma Losses by Fast Electrons in Thin Films. *Physical Review* **106**, 874-881 (1957).

84   de Abajo, F. J. G. Colloquium: Light Scattering by Particle and Hole Arrays. *Reviews of Modern Physics* **79**, 1267-1290 (2007).

85   Murphy, C. J. *et al.* Gold Nanoparticles in Biology: Beyond Toxicity to Cellular Imaging. *Accounts of Chemical Research* **41**, 1721-1730 (2008).

86   Moskovits, M. Surface-Enhanced Raman Spectroscopy: A Brief Retrospective. *Journal of Raman Spectroscopy* **36**, 485-496 (2005).

87   Rodriguez-Fernandez, J. *et al.* The Effect of Surface Roughness on the Plasmonic Response of Individual Sub-Micron Gold Spheres. *Physical Chemistry Chemical Physics* **11**, 5909-5914 (2009).

88   Abalde-Cela, S. *et al.* Surface-Enhanced Raman Scattering Biomedical Applications of Plasmonic Colloidal Particles. *Journal of the Royal Society Interface* **7**, S435-S450 (2010).

89   Sharma, B. *et al.* High-Performance Sers Substrates: Advances and Challenges. *Mrs Bulletin* **38**, 615-624 (2013).

90   Aroca, R. F., Alvarez-Puebla, R. A., Pieczonka, N., Sanchez-Cortez, S. & Garcia-Ramos, J. V. Surface-Enhanced Raman Scattering on Colloidal Nanostructures. *Advances in Colloid and Interface Science* **116**, 45-61 (2005).

91   Kimling, J. *et al.* Turkevich Method for Gold Nanoparticle Synthesis Revisited. *The Journal of Physical Chemistry B* **110**, 15700-15707 (2006).

92   Cao, Y. Q., Li, D., Jiang, F., Yang, Y. & Huang, Z. R. Engineering Metal Nanostructure for Sers Application. *Journal of Nanomaterials* (2013).

93   Zhang, X., Chen, H. & Zhang, H. Layer-by-Layer Assembly: From Conventional to Unconventional Methods. *Chemical Communications*, 1395-1405 (2007).

94   Mulvaney, S. P., He, L., Natan, M. J. & Keating, C. D. Three-Layer Substrates for Surface-Enhanced Raman Scattering: Preparation and Preliminary Evaluation. *Journal of Raman Spectroscopy* **34**, 163-171 (2003).

95   Pienpinijtham, P., Han, X. X., Ekgasit, S. & Ozaki, Y. An Ionic Surfactant-Mediated Langmuir-Blodgett Method to Construct Gold Nanoparticle Films for Surface-Enhanced Raman Scattering. *Physical Chemistry Chemical Physics* **14**, 10132-10139 (2012).

96   Lin, Y., Skaff, H., Emrick, T., Dinsmore, A. D. & Russell, T. P. Nanoparticle Assembly and Transport at Liquid-Liquid Interfaces. *Science* **299**, 226-229 (2003).

97   Guo, S. J., Dong, S. J. & Wang, E. K. Rectangular Silver Nanorods: Controlled Preparation, Liquid-Liquid Interface Assembly, and Application in Surface-Enhanced Raman Scattering. *Crystal Growth & Design* **9**, 372-377 (2009).

98   Driskell, J. D. *et al.* The Use of Aligned Silver Nanorod Arrays Prepared by Oblique Angle Deposition as Surface Enhanced Raman Scattering Substrates. *Journal of Physical Chemistry C* **112**, 895-901 (2008).

99   Abell, J. L., Driskell, J. D., Dluhy, R. A., Tripp, R. A. & Zhao, Y. P. Fabrication and Characterization of a Multiwell Array Sers Chip with Biological Applications. *Biosensors and Bioelectronics* **24**, 3663-3670 (2009).

100   Jung, D. S. *et al.* Facile Fabrication of Large Area Nanostructures for Efficient Surface-Enhanced Raman Scattering. *Journal of Materials Chemistry* **16**, 3145-3149 (2006).

101   Harper, M. M., McKeating, K. S. & Faulds, K. Recent Developments and Future Directions in Sers for Bioanalysis. *Physical Chemistry Chemical Physics* **15**, 5312-5328 (2013).

102   Hennigan, S. L. *et al.* Detection of Mycoplasma Pneumoniae in Simulated and True Clinical Throat Swab Specimens by Nanorod Array-Surface-Enhanced Raman Spectroscopy. *Plos One* **5** (2010).

103   Craig, D., Simpson, J., Faulds, K. & Graham, D. Formation of Sers Active Nanoparticle Assemblies Via Specific Carbohydrate-Protein Interactions. *Chemical Communications* **49**, 30-32 (2013).

104   Wang, X., Ramstrom, O. & Yan, M. D. Quantitative Analysis of Multivalent Ligand Presentation on Gold Glyconanoparticles and the Impact on Lectin Binding. *Analytical Chemistry* **82**, 9082-9089 (2010).

105   Negri, P. *et al.* Direct Optical Detection of Viral Nucleoprotein Binding to an Anti-Influenza Aptamer. *Analytical Chemistry* **84**, 5501-5508 (2012).

106   Varmuza, K. & Filzmoser, P. *Introduction to Multivariate Statistical Analysis in Chemometrics / Kurt Varmuza, Peter Filzmoser*. (Boca Raton : CRC Press, c2009., 2009).

107   Hopke, P. K. The Evolution of Chemometrics. *Analytica Chimica Acta* **500**, 365-377 (2003).

108   Jurs, P. C., Kowalski, B. R. & Isenhour, T. L. Computerized Learning Machines Applied to Chemical Problems. Molecular Formula Determination from Low Resolution Mass Spectrometry. *Analytical Chemistry* **41**, 21-27 (1969).

109   Weiner, P. H. & Malinowski, E. R. Investigation of the Van Der Waals Effect in Nuclear Magnetic Resonance Spectroscopy by Factor Analysis and the Prediction of Diamagnetic and Paramagnetic Susceptibilities. *The Journal of Physical Chemistry* **75**, 3160-3163 (1971).

110   Kowalski, B. R. Chemometrics: Views and Propositions. *Journal of Chemical Information and Computer Sciences* **15**, 201-203 (1975).

111   Geladi, P. & Hopke, P. K. Is There a Future for Chemometrics? Are We Still Needed? *Chemometrics and Intelligent Laboratory Systems* **91**, 99-100 (2008).

112   Geladi, P. & Hopke, P. K. Editorial: Is There a Future for Chemometrics? Are We Still Needed? *Journal of Chemometrics* **22**, 289-290 (2008).

113   Wang, L. Q. & Mizaikoff, B. Application of Multivariate Data-Analysis Techniques to Biomedical Diagnostics Based on Mid-Infrared Spectroscopy. *Analytical and Bioanalytical Chemistry* **391**, 1641-1654 (2008).

114 Berrueta, L. A., Alonso-Salces, R. M. & Héberger, K. Supervised Pattern Recognition in Food Analysis. *Journal of Chromatography A* **1158**, 196-214 (2007).

115 Bro, R. Multivariate Calibration: What Is in Chemometrics for the Analytical Chemist? *Analytica Chimica Acta* **500**, 185-194 (2003).

116 Bocklitz, T., Walter, A., Hartmann, K., Rösch, P. & Popp, J. How to Pre-Process Raman Spectra for Reliable and Stable Models? *Analytica Chimica Acta* **704**, 47-56 (2011).

117 Rinnan, Å., Berg, F. v. d. & Engelsen, S. B. Review of the Most Common Pre-Processing Techniques for near-Infrared Spectra. *TrAC Trends in Analytical Chemistry* **28**, 1201-1222 (2009).

118 Wise, B. G. N. B., R. Shaver, J.; Windig W.; Koch, R. *Chemometrics Tutorial for Pls_Toolbox and Solo*. (Eigenvector Research Inc.).

119 DeNoyer, L. K. & Dodd, J. G. *Smoothing and Derivatives in Spectroscopy*. (John Wiley & Sons, Ltd, 2006).

120 Gendrin, C., Roggo, Y. & Collet, C. Pharmaceutical Applications of Vibrational Chemical Imaging and Chemometrics: A Review. *Journal of Pharmaceutical and Biomedical Analysis* **48**, 533-553 (2008).

121 Heise, H. M. Donald A. Burns, Emil W. Ciurczak (Eds.): Handbook of near-Infrared Analysis, 3rd Ed. *Analytical & Bioanalytical Chemistry* **393**, 1387-1389 (2009).

122   Correia, P. R. M. & Ferreira, M. M. C. Reconhecimento De Padrões Por Métodos

Não Supervisionados: Explorando Procedimentos Quimiométricos Para

Tratamento De Dados Analíticos. *Química Nova* **30**, 481-487 (2007).

123   (Wenatchee, WA, 2011).

124   Ghosh, S. Online Measurement of Polyvinyl-Alcohol Size on Warp Yarns Using

a near-Ir Diffuse Reflectance Spectroscopy Method. *Journal of the Textile

Institute* **84**, 85-98 (1993).

125   Wise, B. G. N. B., R. Shaver, J.; Windig W.; Koch, R. *Chemometrics Tutorial for

Pls_Tookbos and Solo*. (Eigenvector Research Incorporated 2006).

126   Bro, R. & Smilde, A. K. Principal Component Analysis. *Analytical Methods* **6**,

2812-2831 (2014).

127   Ballabio, D. & Consonni, V. Classification Tools in Chemistry. Part 1: Linear

Models. Pls-Da. *Analytical Methods* **5**, 3790-3798 (2013).

128   Leone, G., Consumi, M., Lamponi, S. & Magnani, A. Combination of Static Time

of Flight Secondary Ion Mass Spectrometry and Infrared Reflection–Adsorption

Spectroscopy for the Characterisation of a Four Steps Built-up Carbohydrate

Array. *Applied Surface Science* **258**, 6302-6315 (2012).

129   Höskuldsson, A. Pls Regression Methods. *Journal of Chemometrics* **2**, 211-228

(1988).

130   Pérez, N. F., Ferré, J. & Boqué, R. Calculation of the Reliability of Classification

in Discriminant Partial Least-Squares Binary Classification. *Chemometrics and

Intelligent Laboratory Systems* **95**, 122-128 (2009).

131   Negri, P. *et al.* Identification of Virulence Determinants in Influenza Viruses. *Analytical Chemistry* **86**, 6911-6917 (2014).

132   Wilchek, M. & Bayer, E. A. The Avidin-Biotin Complex in Bioanalytical Applications. *Analytical Biochemistry* **171**, 1-32 (1988).

133   Wilchek, M., Bayer, E. A. & Livnah, O. Essentials of Biorecognition: The (Strept)Avidin–Biotin System as a Model for Protein–Protein and Protein–Ligand Interaction. *Immunology Letters* **103**, 27-32 (2006).

134   Green, N. M. *Avidin*. Vol. Volume 29 (Academic Press, 1975).

135   Tripp, R. A., Dluhy, R. A. & Zhao, Y. P. Novel Nanostructures for Sers Biosensing. *Nano Today* **3**, 31-37 (2008).

136   Yoo, B. K. & Joo, S. W. In Situ Raman Monitoring Triazole Formation from Self-Assembled Monolayers of 1,4-Diethynylbenzene on Ag and Au Surfaces Via "Click" Cyclization. *Journal of Colloid and Interface Science* **311**, 491-496 (2007).

137   Torreggiani, A. & Fini, G. Raman Spectroscopic Studies of Ligand–Protein Interactions: The Binding of Biotin Analogues by Avidin. *Journal of Raman Spectroscopy* **29**, 229-236 (1998).

138   Galarreta, B. C., Norton, P. R. & Lagugné-Labarthet, F. o. Sers Detection of Streptavidin/Biotin Monolayer Assemblies†. *Langmuir* **27**, 1494-1498 (2011).

139   Fagnano, C., Fini, G. & Torreggiani, A. Raman-Spectroscopic Study of the Avidin-Biotin Complex. *Journal of Raman Spectroscopy* **26**, 991-995 (1995).

140   Vermette, P. *et al.* Immobilization and Surface Characterization of Neutravidin Biotin-Binding Protein on Different Hydrogel Interlayers. *Journal of Colloid and Interface Science* **259**, 13-26 (2003).

141   Knoll, W. *et al.* Streptavidin Arrays as Supramolecular Architectures in Surface-Plasmon Optical Sensor Formats. *Colloids and Surfaces a-Physicochemical and Engineering Aspects* **161**, 115-137 (2000).

142   Mayya, K. S., Patil, V. & Sastry, M. On the Stability of Carboxylic Acid Derivatized Gold Colloidal Particles: The Role of Colloidal Solution Ph Studied by Optical Absorption Spectroscopy. *Langmuir* **13**, 3944-3947 (1997).

143   Sastry, M., Lala, N., Patil, V., Chavan, S. P. & Chittiboyina, A. G. Optical Absorption Study of the Biotin-Avidin Interaction on Colloidal Silver and Gold Particles. *Langmuir* **14**, 4138-4142 (1998).

144   Delange, R. J. & Huang, T. S. Egg White Avidin .3. Sequence of 78-Residue Middle Cyanogen Bromide Peptide - Complete Amino Acid Sequence of Protein Subunit. *Journal of Biological Chemistry* **246**, 698-& (1971).

145   Yam, C. M., Pradier, C. M., Salmain, M., Fischer-Durand, N. & Jaouen, G. Molecular Recognition of Avidin on Biotin-Functionalized Gold Surfaces Detected by Ft-Irras and Use of Metal Carbonyl Probes. *Journal of Colloid and Interface Science* **245**, 204-207 (2002).

146   Han, X. X., Huang, G. G., Zhao, B. & Ozaki, Y. Label-Free Highly Sensitive Detection of Proteins in Aqueous Solutions Using Surface-Enhanced Raman Scattering. *Analytical Chemistry* **81**, 3329-3333 (2009).

147   Ahern, A. M. & Garrell, R. L. Protein Metal Interactions in Protein-Colloid Conjugates Probed by Surface-Enhanced Raman-Spectroscopy. *Langmuir* **7**, 254-261 (1991).

148   Hering, K. *et al.* Sers: A Versatile Tool in Chemical and Biochemical Diagnostics. *Analytical and Bioanalytical Chemistry* **390**, 113-124 (2008).

149   Sarkar, U. K., Pal, A. J., Chakrabarti, S. & Misra, T. N. Classical and Chemical Effects of Sers from 2,2'-5,2'' Terthiophene Adsorbed on Ag-Sols. *Chemical Physics Letters* **190**, 59-63 (1992).

150   Kennedy, B. J., Spaeth, S., Dickey, M. & Carron, K. T. Determination of the Distance Dependence and Experimental Effects for Modified Sers Substrates Based on Self-Assembled Monolayers Formed Using Alkanethiols. *Journal of Physical Chemistry B* **103**, 3640-3646 (1999).

151   Driskell, J. D. *et al.* Rapid and Sensitive Detection of Rotavirus Molecular Signatures Using Surface Enhanced Raman Spectroscopy. *Plos One* **5** (2010).

152   Varki, A. Glycan-Based Interactions Involving Vertebrate Sialic-Acid-Recognizing Proteins. *Nature* **446**, 1023-1029 (2007).

153   Park, S., Lee, M. R. & Shin, I. Chemical Tools for Functional Studies of Glycans. *Chemical Society Reviews* **37**, 1579-1591 (2008).

154   Oyelaran, O. & Gildersleeve, J. C. Glycan Arrays: Recent Advances and Future Challenges. *Current Opinion in Chemical Biology* **13**, 406-413 (2009).

155   Wan, A. J., Wang, K., Zhang, H. C., Li, H. L. & Wang, D. N. Modern Carbohydrate Microarray Biochip Technologies. *Chinese Journal of Analytical Chemistry* **40**, 1780-U1746 (2012).

156   Park, S., Lee, M. R. & Shin, I. Carbohydrate Microarrays as Powerful Tools in
        Studies of Carbohydrate-Mediated Biological Processes. *Chemical*
        Communications, 4389-4399 (2008).

157   Unfricht, D. W., Colpitts, S. L., Fernandez, S. M. & Lynes, M. A. Grating-
        Coupled Surface Plasmon Resonance: A Cell and Protein Microarray Platform.
        Proteomics **5**, 4432-4442 (2005).

158   Knauer, M., Ivleva, N. P., Liu, X. J., Niessner, R. & Haisch, C. Surface-Enhanced
        Raman Scattering-Based Label-Free Microarray Readout for the Detection of
        Microorganisms. *Analytical Chemistry* **82**, 2766-2772 (2010).

159   Yan, J. *et al.* Nano Rolling-Circle Amplification for Enhanced Sers Hot Spots in
        Protein Microarray Analysis. *Analytical Chemistry* **84**, 9139-9145 (2012).

160   Zhang, H., Harpster, M. H., Wilson, W. C. & Johnson, P. A. Surface-Enhanced
        Raman Scattering Detection of Dnas Derived from Virus Genomes Using Au-
        Coated Paramagnetic Nanoparticles. *Langmuir* **28**, 4030-4037 (2012).

161   Vangala, K. *et al.* Sensitive Carbohydrate Detection Using Surface Enhanced
        Raman Tagging. *Analytical Chemistry* **82**, 10164-10171 (2010).

162   Mrozek, M. F. & Weaver, M. J. Detection and Identification of Aqueous
        Saccharides by Using Surface-Enhanced Raman Spectroscopy. *Analytical*
        Chemistry **74**, 4069-4075 (2002).

163   Mrozek, M. F., Zhang, D. & Ben-Amotz, D. Oligosaccharide Identification and
        Mixture Quantification Using Raman Spectroscopy and Chemometric Analysis.
        Carbohydrate Research **339**, 141-+ (2004).

164   Meynier, C., Guerlesquin, F. & Roche, P. Computational Studies of Human Galectin-1: Role of Conserved Tryptophan Residue in Stacking Interaction with Carbohydrate Ligands. *Journal of Biomolecular Structure & Dynamics* **27**, 49-57 (2009).

165   Vasta, G. R. Roles of Galectins in Infection. *Nature Reviews Microbiology* **7**, 424-438 (2009).

166   Ning, X. H., Guo, J., Wolfert, M. A. & Boons, G. J. Visualizing Metabolically Labeled Glycoconjugates of Living Cells by Copper-Free and Fast Huisgen Cycloadditions. *Angewandte Chemie-International Edition* **47**, 2253-2255 (2008).

167   Saraboji, K. *et al.* The Carbohydrate-Binding Site in Galectin-3 Is Preorganized to Recognize a Sugarlike Framework of Oxygens: Ultra-High-Resolution Structures and Water Dynamics. *Biochemistry* **51**, 296-306 (2012).

168   Barondes, S. H., Cooper, D. N. W., Gitt, M. A. & Leffler, H. Galectins - Structure and Function of a Large Family of Animal Lectins. *Journal of Biological Chemistry* **269**, 20807-20810 (1994).

169   Savitzky, A. & Golay, M. J. E. Smoothing + Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry* **36**, 1627-& (1964).

170   Kudelski, A. Characterization of Thiolate-Based Mono- and Bilayers by Vibrational Spectroscopy: A Review. *Vibrational Spectroscopy* **39**, 200-213 (2005).

171   Rubim, J. C., Gutz, I. G. R., Sala, O. & Orvillethomas, W. J. Surface Enhanced Raman-Spectra of Benzotriazole Adsorbed on a Copper Electrode. *Journal of Molecular Structure* **100**, 571-583 (1983).

172   De Gelder, J., De Gussem, K., Vandenabeele, P. & Moens, L. Reference Database of Raman Spectra of Biological Molecules. *Journal of Raman Spectroscopy* **38**, 1133-1147 (2007).

173   Ledin, P. A., Kolishetti, N. & Boons, G. J. Multifunctionalization of Polymers by Strain-Promoted Cycloadditions. *Macromolecules* **46**, 7759-7768 (2013).

174   Stephan, M. T., Moon, J. J., Um, S. H., Bershteyn, A. & Irvine, D. J. Therapeutic Cell Engineering with Surface-Conjugated Synthetic Nanoparticles. *Nature Medicine* **16**, 1035-U1135 (2010).

175   J. Seetharaman, A. K., Rita Slaaby, Hakon Leffler, Samuel H. Barondes, & Rini, a. J. M. X-Ray Crystal Structure of the Human Galectin-3 Carbohydrate

Recognition Domain at 2.1-Å Resolution. *The Journal of Biological Chemistry* **273**, 13047–13052 (1998).

176   Rothberg, M. B. & Haessler, S. D. Complications of Seasonal and Pandemic Influenza. *Critical Care Medicine* **38**, E91-E97 (2010).

177   Nafziger, A. N. & Pratt, D. S. Seasonal Influenza Vaccination and Technologies. *Journal of Clinical Pharmacology* **54**, 719-731 (2014).

178   Santibañez, S., Fiore, A. E., Merlin, T. L. & Redd, S. A Primer on Strategies for Prevention and Control of Seasonal and Pandemic Influenza. *American Journal of Public Health* **99**, S216-S224 (2009).

179   Clancy, S. Genetics of the Influenza Virus. *Nature Education* **1** (2008).

180   eds., A. W. W. C. H. J. in *Influenza*   151–172 (Public Health Foundation, CDC, 2012).

181   Ge, S. Q. & Wang, Z. L. An Overview of Influenza a Virus Receptors. *Critical Reviews in Microbiology* **37**, 157-165 (2011).

182   Lu, X. S. *et al.* Structure and Receptor Binding Specificity of Hemagglutinin H13 from Avian Influenza a Virus H13n6. *Journal of Virology* **87**, 9077-9085 (2013).

183   Negri, P. & Dluhy, R. A. Detection of Genetic Markers Related to High Pathogenicity in Influenza by Sers. *Analyst* **138**, 4877-4884 (2013).

184   Negri, P., Kage, A., Nitsche, A., Naumann, D. & Dluhy, R. A. Detection of Viral Nucleoprotein Binding to Anti-Influenza Aptamers Via Sers. *Chemical Communications* **47**, 8635-8637 (2011).

185   Park, Y.-K., Yoo, S.-H. & Park, S. Assembly of Highly Ordered Nanoparticle Monolayers at a Water/Hexane Interface. *Langmuir* **23**, 10505-10510 (2007).

186   Chandrashekar, G. & Sahin, F. A Survey on Feature Selection Methods. *Computers & Electrical Engineering* **40**, 16-28 (2014).

187   Ferreira, M. P., Grondona, A. E. B., Rolim, S. B. A. & Shimabukuro, Y. E. Analyzing the Spectral Variability of Tropical Tree Species Using Hyperspectral Feature Selection and Leaf Optical Modeling. *Journal of Applied Remote Sensing* **7**, 13 (2013).

188   Price, J. C. How Unique Are Spectral Signatures. *Remote Sensing of Environment* **49**, 181-186 (1994).

189   Abell, J. L., Driskell, J. D., Dluhy, R. A., Tripp, R. A. & Zhao, Y. P. Fabrication and Characterization of a Multiwell Array Sers Chip with Biological Applications. *Biosensors & Bioelectronics* **24**, 3663-3670 (2009).

190   de Graaf, M. & Fouchier, R. A. M. Role of Receptor Binding Specificity in Influenza a Virus Transmission and Pathogenesis. *Embo Journal* **33**, 823-841 (2014).

191   Liver, N., Nitzan, A. & Gersten, J. I. Local-Fields in Cavity Sites of Rough Dielectric Surfaces. *Chemical Physics Letters* **111**, 449-454 (1984).

192   Keating, C. D., Kovaleski, K. K. & Natan, M. J. Heightened Electromagnetic Fields between Metal Nanoparticles: Surface Enhanced Raman Scattering from Metal-Cytochrome C-Metal Sandwiches. *Journal of Physical Chemistry B* **102**, 9414-9425 (1998).