Master Assignment BME

DETECTION OF DNA METHYLATION USING SURFACE-ENHANCED RAMAN SPECTROSCOPY

February 21, 2024

Master's assignment committee
Daily Supervisor: dr.ir. J.E. van Dongen
External member: dr. O.S. Ojambati
Chair: prof.dr.ir. L.I. Segerink

Jan Jacob de Waard
BIOS - University of Twente

Abstract

Cancer is one of the most impactful diseases known, with 18 million new cases in 2018 alone. This means that early detection is of great importance. DNA methylation is an epigenetic process, which can be used for the early detection of cancer. DNA methylation is responsible for the regulation of genes but is also associated with cancer. Surface-Enhanced Raman Spectroscopy (SERS) already shows promising results in the detection of methylation differences. SERS also has advantages over other methods, namely being more sensitive, quicker in analysis, and is a non-invasive method. This thesis aims to look at the application of detecting single base DNA methylation with SERS for the detection of cancer and whether SERS will be a feasible option to use for this application.

Firstly, regular Raman spectroscopy was used to differentiate between methylation levels. A DNA sequence was created, and three versions, each with different levels of cytosine methylation, were ordered. The Raman spectra of the three sequences were obtained, and differentiation was looked at with principal component analysis (PCA). The regular Raman showed very similar Raman spectra, with some subtle differences at around 1000 and 1460 cm^{-1} per sequence. PCA further showed the differences between methylation levels.

To translate the measurements to SERS, different substrates, and Raman setups were tested. The drain-to-deposit method in combination with an in-house build setup was chosen to conduct the SERS measurements. These SERS measurements were obtained by conducting an area scan of a gold nanoparticle (AuNP) aggregate on which DNA was desiccated. The SERS measurements consist of two samples, the DNA sequence which has no methylation, and the sequence which is fully methylated. Both sequences were hard to acquire, with a small amount of AuNP aggregates giving a spectrum. Of both area scans, 20 measurements were chosen randomly and PCA was conducted. The SERS measurements of the non- and fully-methylated DNA share some of the same peaks, but look different. Differences between measurements of the same sample would also occur due to fluctuations caused by, among others, picocavities. The 40 measurements were also analyzed with PCA, which was able to differentiate between the two samples.

Both regular Raman and SERS showed the ability to differentiate between methylation levels. Regular Raman was able to differentiate between single base differences in the DNA. For SERS however, only non-methylated and fully-methylated DNA were tested. Differentiation was possible between the two samples, even with the fluctuations between measurements. However, further research on substrate and setup should be conducted to make the SERS measurements more reliable and consistent. Further research should also look at lower DNA concentration levels, as this would be the case in a clinical application of the method. This could eventually lead to a microfluidic chip, on which DNA sample is captured at known hotspots, which can then be analyzed precisely.

Contents

Chapter 1

Introduction

Cancer is one of the most impactful diseases known to humankind. In 2018 alone, 18 million new cases were diagnosed [1]. The risk of someone on earth developing cancer during their lifetime is on average 20% [1]. The chance of curing increases in some types of cancer when caught early and treated correctly [2]. That is why it is important to catch the disease as early as possible.

Various biomarkers and techniques exist to find cancer in an early stage, which probe (epi)genetic changes that indicate cancer [3]. DNA methylation is one of the possible epigenetic changes researchers look at, and could be used to detect cancer [4]. DNA is methylated differently across different cell types to suppress certain genes. DNA methylation can also occur on genes that suppress cancer growth, which can lead to cancer.

Raman spectroscopy is already a proven method for detecting cytosine methylation and also has the advantage of being a non-invasive method [5, 6, 7, 8]. However, Raman scattering is very weak, being around 14 orders of magnitude smaller than fluorescent signal [9]. Surface-Enhanced Raman Spectroscopy (SERS) is a method to increase the intensity of the signal, with the ability to measure single molecules [10]. SERS uses rough metal surfaces such as gold or silver. When visible light excites the localized surface plasmon resonance of the metal, electromagnetic fields are generated [9]. This increases the dipole of the scatter, increasing the intensity of the signal. However, there are still challenges concerning SERS for label-free detection of DNA. These challenges range from similarity in spectra of DNA bases to the spacing of DNA-nanoparticles conjugates [11, 12].

This thesis aims to look at the application of detecting single base DNA methylation with SERS for the detection of cancer and whether SERS will be a feasible option to use for this application. Other research already shows promising results whilst also having some benefits compared to other methods [13, 14]. SERS is for example more sensitive, quicker in analysis, and is a non-invasive method. Research will be conducted on whether Surface-Enhanced Raman Spectroscopy could be used to determine the methylation level of DNA using gold nanoparticles.

Chapter 2

Theoretical background

2.1 DNA methylation

DNA is arguably the most critical biomolecule and is the basis of all organisms. DNA consists of nucleotides made of deoxyribose, a phosphate group, and one of four bases: adenine, guanine, thymine, or cytosine [15]. DNA contains all the genetic code for an organism, and the complete genome is present in every organism's cells. However, for different cells, different genes need to be expressed. One of the methods that DNA uses to regulate gene expression is DNA methylation [16].

DNA methylation is an epigenetic mechanism that adds methyl groups to the nucleobases of DNA. DNA methylation has been naturally observed in adenine and cytosine. However, most methylation in eukaryotes can be seen on the cytosine nucleotide with a guanine nucleotide following it, looking in the 5' $\rightarrow$ 3' direction [16]. These sites are called CpG sites and more than 98% of DNA methylation in somatic cells occurs here [17]. DNA methylation in cytosine is due to a covalent transfer of a methyl group on the C5 position with the use of DNA methyltransferases, which creates 5-methylcytosine [16, 17]. This change is shown in Figure 2.1.

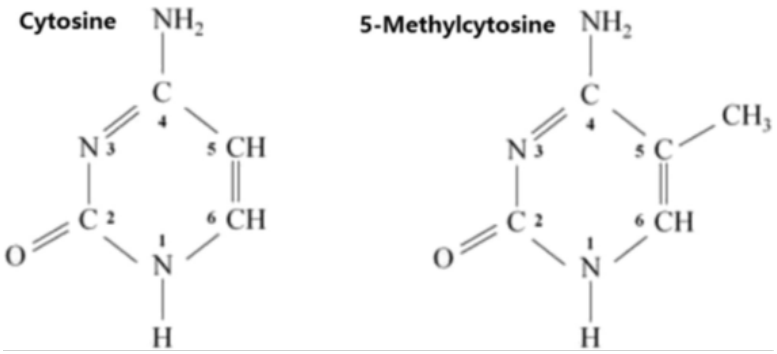


Figure 2.1: A non-methylated cytosine nucleotide (left) and a methylated cytosine nucleotide (right). The methyl group is attached to the carbon on the 5th position of the ring. [5]

The methylation can regulate gene expression by preventing the binding of transcription factors or by attracting certain proteins that are associated with gene repression [16]. When methylation occurs in a gene promoter, it can act as a suppressor for gene expression. If this occurs for a

tumor suppressor gene, this can lead to cancer, among other diseases [4, 17]. Over-methylation of DNA is called hypermethylation. However, less than average methylation levels in DNA, hypomethylation, has also been attributed to the development of cancer [18], albeit that it is often found in different parts of the DNA sequence than hypermethylation [19].

Because DNA methylation is a significant epigenetic factor, it can be used to identify whether the DNA is from healthy or cancerous cells [18]. The golden standard for the detection of DNA methylation currently is bisulfite treatment. This incorporates a chemical treatment step in the analysis [20]. The process turns cytosine into uracil but leaves 5-methylcytosine unaffected. PCR is used to amplify the DNA. It is expensive, takes a long time, and not a lot of DNA can be used for PCR templates due to degradation during the bisulfite conversion [21].

Another method is using methylation-sensitive restriction enzymes (MSRE) in combination with qPCR [22]. However, qPCR needs long amplicon lengths and has to include two MSRE sites, making it impossible to look at the methylation of single CpG couples.

A technique that shows promising results is Raman spectroscopy. This technique has the ability to distinguish between small changes in molecular structure and could be able to show DNA methylation in a label-free way [23].

2.2 Raman spectroscopy

Raman spectroscopy is a technique that uses inelastic scattering of photons, or Raman Scattering. When measured, this gives a Raman spectrum, which gives insight into the vibrational modes of molecules. It can also be used to identify a material [24], but also small differences between two molecules, such as methylation in DNA [23]. Measurements will give an intensity per wavenumber that is visualized in a spectrum that can be compared and analyzed.

Rayleigh scattering and the two types of Raman scattering, stokes, and anti-stokes begin when the absorption of a photon excites a molecule [9, 25]. This excitement will bring the energy to a virtual electronic energy level linked with the laser's photon energy. However, this energy state is short-lived and a photon is released, which causes the energy level to drop again. Rayleigh scattering, which is a type of elastic scattering, occurs when there is no energy exchange, the incident and scattered photons have the same energy. Raman scattering, which is inelastic scattering, occurs when there is a change in energy. Stokes Raman scattering is when the molecule absorbs energy and the scattered photon is lower in energy. Anti-Stokes Raman scattering occurs when the molecule loses energy and the scattered photon has more energy. The different types of scattering are shown in Figure 2.2.

Raman spectroscopy uses monochromatic light, often a laser. The laser photons interact with the vibrations of the molecules, causing Raman scattering. This can be used to give information about the vibrational modes. However, Raman gives a very weak signal, as it is much more likely for the photons to Rayleigh scatter. Raman scattering occurs for only around 1 in 10^8 of incident particles [25].

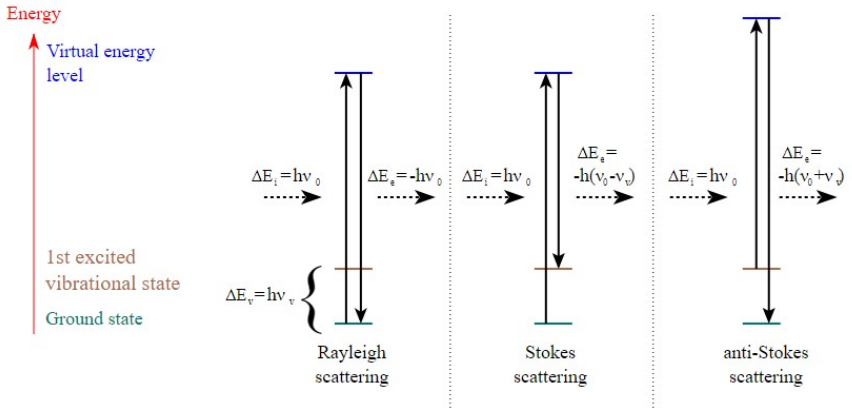


Figure 2.2: Rayleigh, Stokes, and anti-Stokes scattering [26]

When looking at the ideal laser wavelength to use for Raman spectroscopy, multiple aspects need to be considered. Using a different wavelength has no effect on the location of the peaks found in the Raman spectrum, but it does change the height of the peaks and the Raman scattering cross section [27]. Using different wavelengths can also cause other peaks to show up, while previously seen peaks could disappear. Also, the effect of wavelength on signal intensity plays a role. The Raman intensity, I, is proportional to the frequency of the incident light, v_0, to the power of 4 [28]. This means that the intensity is proportional to the inverse of the wavelength of the incident light to the power of 4, which can be seen in Equation 2.1. For an increasing wavelength of the laser, the intensity of the Raman scattering is lowered.

$$I = v_0^4 = 1/\lambda^4 \tag{2.1}$$

2.2.1 Raman spectrum

During measurements, a Raman spectrum will be recorded. This spectrum is a visualization of the Intensity, or number of photons measured, as a function of the wavenumber. Figure 2.3 shows a Raman spectrum. It can be seen that there are peaks, which are called the Raman bands. The fingerprint region in the spectrum, which can be used to identify the substance, can roughly be found between 400 and 1800 $[cm^{-1}]$ [29].

The Raman bands will say something about the vibrational modes of molecules. There will be specific peaks for the different interactions of individual atoms and bonds with each other. These interactions are the rocking, stretching, and bending of these parts of the molecule [30, 31]. All the bands can be explained by one, or multiple, interactions of the different atoms. This also makes Raman a technique that can be used to identify which substance is measured, through the process of fingerprinting.

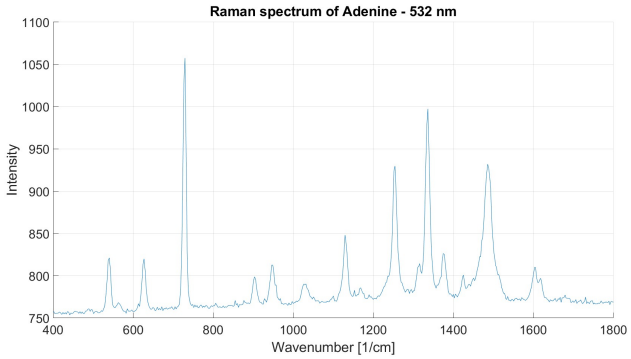


Figure 2.3: Raman spectrum of Adenine

2.2.2 Surface-Enhanced Raman Spectroscopy

Surface-enhanced Raman Spectroscopy (SERS) is a technique that can be used to enhance Raman scattering. Fleischmann et al. observed this effect in 1974 while studying molecules that were adsorbed on a rough silver surface, and found an intensity increase of 10^6 times [32]. Nowadays it is shown that under ideal conditions the SERS enhancement can be up to of up to 10^{12} times [33]. SERS is credited to its increase due to the combination of two effects: chemical impact and the electromagnetic effect [9]. The chemical impact says that this enhancement is achieved due to resonance Raman spectroscopy. The resonance of the incident photon and the electron transition are very similar, which boosts the Raman scattering intensity. The electromagnetic effect depends on the size, shape, and material of the used nanoparticle, which all have an effect on the resonance frequency of the metal. Localized surface plasmons are excited when light hits the metal, causing the resonant electromagnetic radiation to scatter, and the generation of strong electromagnetic fields at the rough surface of the metal, seen in figure 2.4 [9]. Raman-active molecules need to be present within the electromagnetic fields for the enhancement to take place [9].

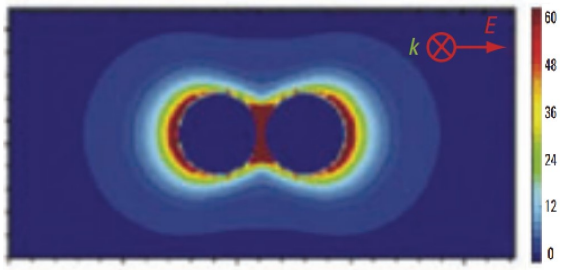


Figure 2.4: Electromagnetic field changes caused by localized surface plasmon resonance of two spherical nanoparticles [9]

SERS can be used in DNA detection while being a method with advantages such as being label-free, having a high sensitivity, and giving a detailed fingerprint [13]. These advantages are promising for DNA methylation detection and structural characteristic analysis [20, 34]. However, it remains hard to use SERS for label-free detection of DNA. This is due to some issues, which will be discussed further in chapter 2.3.

Gold Nano Particles

To enable SERS, a rough metal needs to be used, of which there are multiple choices. Candidates for SERS include but are not limited to, gold [20, 35], silver [7], or a gold-silver combination [8]. The close proximity of the AuNPs enables SERS, which causes a red shift of the signal when two AuNPs are in close proximity. The distance between nanoparticles for which SERS occurs depends on the size of the particles. In research done by Kumari et al. [36], it is shown that for silver nanoparticles (AgNPs) the distance for SERS to occur is 1 nm for AgNPs of 20 nm in diameter up to a distance of 5 nm for AgNPs of 90 nm in diameter. These results are shown in figure 2.5. Theoretical modeling shows that this is also the case for AuNPs [36]. Although giving more signal and can be used further apart, the AuNPs of $>$ 50nm are less stable in solution and tend to aggregate [36]. Polyethylene glycol (PEG) can be used to stabilize the gold nanoparticles, allowing for larger AuNPs [37]. The use of PEG might be a solution when larger molecules, such as DNA, are conjugated to the AuNPs for SERS measurements.

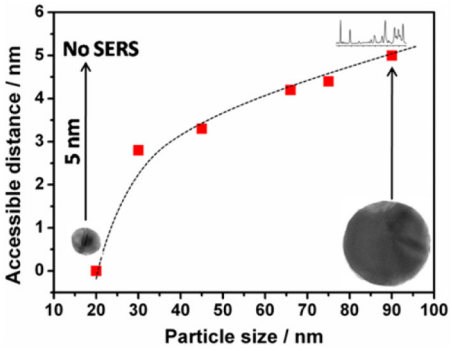


Figure 2.5: Graph showing the relation between particle size and particle distance for SERS to occur [36]

2.3 State of the art

The current state of the art in the detection of DNA methylation using Raman spectroscopy is focused on using SERS [38]. The main improvements in the field are in techniques to improve sample placement in hotspots. There are many possible techniques that can be used for this purpose. For example, the DNA can be captured by a complementary sequence already attached in an SERS active region [39]. A sandwich assay can be used, on which a piece of sample, needs to bind to a nanoparticle. This conjugate will be captured in a SERS active region, causing a sandwich of the sample [40]. Also, optical entrapment methods can be used, such as optical tweezers [41]. These methods ensure a more optimal signal that is obtained. This contributes to the ability to differentiate between a single nucleotide polymorphism, or single base difference, which has been achieved [23].

There are however some limitations of SERS. Firstly, distinguishing between methylated DNA and unmethylated DNA is hard to do. DNA consists of four bases giving similar Raman shift spectra for both unmethylated and methylated DNA [11]. Another problem is that the signals from single and double stranded DNA sequences are similar [38]. SERS is also notorious for having a random distribution of 'hotspots', causing poor reproducibility of the spectrum [42]. Lastly, SERS is very sensitive to spacing between the nanoparticles [12]. The spacing is on a

nm level, and with longer DNA strands, this could be an issue. These are problems that can be improved upon. This thesis will look at the ability of SERS to differ between methylated and unmethylated double stranded DNA by comparing multiple DNA sequences with different methylation levels. To have more reliable and reproducible SERS measurements, research will be done into a substrate on which the hotspot placement is more controlled. Multiple substrates will be manufactured and tested.

Chapter 3

Materials & Methods

The experiments conducted in this thesis can be classified into three parts. Part I is about measuring nucleobases and DNA with regular Raman spectroscopy. Part II is about testing SERS measurements and designing, making, and characterizing AuNPs. Part III will be about SERS measurements of DNA bound to the most ideal AuNP variation found.

3.1 Part I: Raman spectroscopy of nucleobases and DNA

This section will go into detail about the experiments that were conducted using regular Raman Spectroscopy on nucleobases and DNA.

3.1.1 Protocol for Raman spectroscopy

The following protocol was followed to measure the Raman spectra of samples such as nucleobases, DNA, and AuNPs. Two Raman spectrometers have been used for measurements. The commercial Raman spectrometer (WITec-alpha300 S, Oxford Instruments, Abingdon, United Kingdom) was used in combination with the program Control FIVE (WITec GmbH, Ulm, Germany) for the regular Raman measurements. For SERS measurements, both the commercial setup, as well as an in-house built setup by the Nanobiophysics group (NBP) of the University of Twente, were used.

To be able to measure the Raman spectrum, the sample needs to desiccate if dissolved, which will leave a crystalized substance. To do this, a sample of 2 to 10 μL was added to the desired substrate. The sample was allowed to desiccate until no moisture was left.

After the drying, the sample was placed under the Raman spectrometer. First, using the bright field camera, the location of the sample was located and the sample was brought into focus. Figure 3.1 shows a typical sample location, in this case, crystallized cytosine. Crystallized DNA will look similar. The crystal is where the sample will be, and thus it is the location that gives the Raman signal. Next, using the Raman modus, the Raman spectrum is measured. The measurement conducted is an integration of 10 measurements all with an integration time of 10 s. This is done for five locations on the sample. The final average Raman spectrum will be the average of the spectra at the five locations.

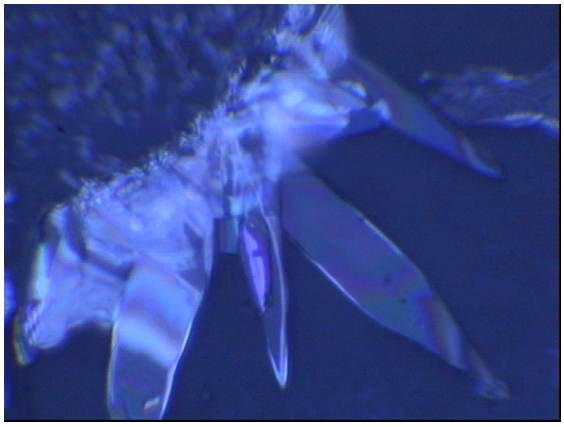


Figure 3.1: Crystallized cytosine seen with a brightfield microscope

Analysis

For analysis of the datasets gathered from the experiment, MATLAB (MathWorks, Natick, Massachusetts, USA) was used. A script was written that would load the Raman spectra measured. Firstly, a baseline correction was conducted, using the Polyfit method by Lieber & Mahadevan-Jansen [43]. When comparing measurements, normalization through the Standard Normal Variate (SVN) method was used afterward. The average intensity, $I_{average}$, was subtracted from the intensity, I. That value was divided by the standard deviation, SD_I, which gives the SNV:

$$SNV = \frac{I - I_{average}}{SD_I}$$

After normalization, a principal component analysis (PCA) was conducted using SPSS statistics 27 (IBM, Armonk, New York, USA).

Variations on the protocol have occurred and will be mentioned when applicable.

Raman spectrometer setups

As has been mentioned, there are two setups that were used for the experiments. The commercial WITec setup and the in-house built NBP setup. A figure of the WITec setup can be found in figure 3.2. The WITec has the choice of three laser wavelengths: 532 nm, 633 nm, and 785 nm, of which the latter has not been used. The program to control the WITec is control FIVE. The program also shows a real-time signal when in the Raman-mode, and is used to start measurements. The measurements can be saved manually into a .txt file or .jpg image. To setup the WITec correctly, it needs to be calibrated before each measurement session. This calibration is for maximization of Raman signal captured by the setup. However, because calibration is done by hand, there can be differences between measurement sessions. The WITec focuses on the sample manually using the Control FIVE program. A sample location is chosen by hand as well, using the translation knobs of a moving stage. The setup has brightfield microscopy available but lacks darkfield capabilities.

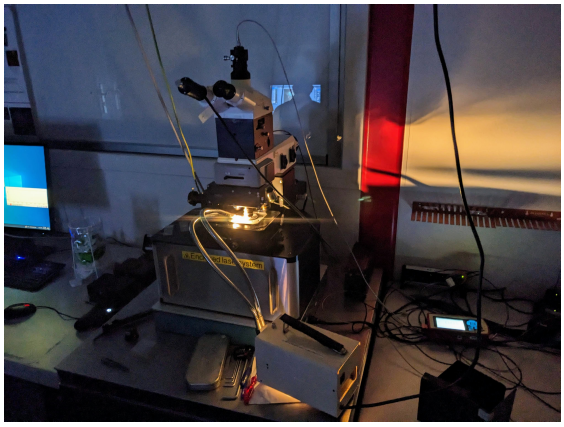


Figure 3.2: WITec setup with halogen lamp to allow for dark field microscopy

The in-house built setup of NBP can be seen in figure 3.3. The NBP setup uses a 633 nm laser for the conducted SERS measurements, but other laser wavelengths are available. Software was developed specifically for the setup using MATLAB. This code can control the spectroscopy setup, as well as show the data gathered, which is saved as an .h5 file. The NBP setup requires no calibration before each use of the system. Focusing can be done both by hand, or automatically using the software. The location on the sample can be chosen by hand on the setup, by hand on the program or by using a point-and-click feature. The setup has brightfield and darkfield capabilities.

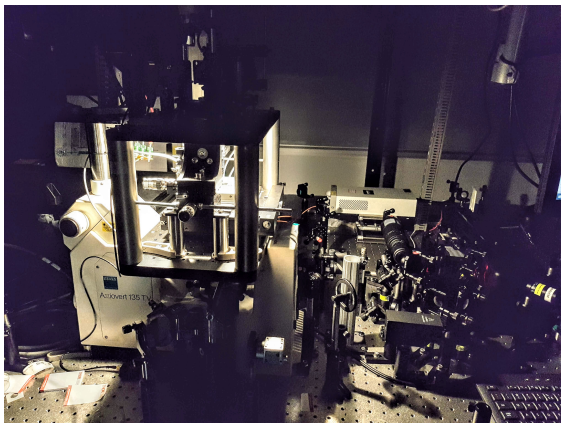


Figure 3.3: In-house built Raman spectrometer by NBP

3.1.2 Nucleobases

For the measurement of the individual nucleobases of DNA, the exact same protocol was used as described above, using the 532 nm laser, with a power used of 12,9 mW. The only difference is that the DNA sample is changed with Adenine (Thermo Fisher Scientific, Waltham, Massachusetts, USA), Guanine (Thermo Fisher Scientific, Waltham, Massachusetts, USA), Cytosine (Thermo Fisher Scientific, Waltham, Massachusetts, USA), methylated Cytosine (Thermo Fisher

Scientific, Waltham, Massachusetts, USA), or Guanine (Thermo Fisher Scientific, Waltham, Massachusetts, USA). Each sample was dissolved in nuclease-free water (New England Biolabs, Ipswich, Massachusetts, USA) to a desired concentration of 1 mM.

3.1.3 Simulated Raman spectrum

A MATLAB model for Raman spectra based on their nucleobase composition was created based on the results of the nucleobase spectra. This script uses average spectra of these nucleobases and requires user input telling the script out of which and how many nucleobases the sequences consists. The script can be found in Appendix I.

3.1.4 DNA

A DNA sequence (Eurofins Genomics, Ebersberg, Germany) has been designed and ordered:

FWD: 5' – TGC GGC GGA GAG GGG TAG AGC GAC GA – 3'
REV: 5' – TC GTC GCT CTA CCC CTC TCC GCC GCA – 3'

Three different versions of this sequence were ordered: the non-methylated sequence seen above, a sequence in which one CpG cytosine was methylated, and a sequence in which all CpG cytosines were methylated. The forward and reverse sequences were added together with nuclease-free water with a total concentration of 45 μM, and this mixture was put into a thermal cycler (Bio-Rad T100, Bio-Rad Laboratories, Hercules, California, USA). The length of a basepair of DNA is 0,34 nm [44]. That means that the length of this DNA sequence would be 8,84 nm.

The Raman spectrum was obtained for both the 532 nm and the 633 nm laser. The measurements used a laser power of 22,6 mW and 20,05 mW respectively. To compare data sets gathered on the DNA, PCA will be conducted. Prior to the PCA, a baseline correction and normalization will be done on the Raman spectra, for which the process is described in the chapter 3.1.1 **Analysis**.

3.2 Part II: SERS and AuNP testing

This section will talk about the experiments that were conducted to test SERS measurements as well as the creation of AuNP-DNA conjugates. Multiple different experiments have been conducted to achieve SERS.

3.2.1 Protocol for preparation of gold nanoparticles

This section describes the protocol for the preparation of AuNPs, which are coated with mPEG and with DNA. The following materials are needed: 20, 40, 60, or 80 nm AuNPs (BBI solutions, Newport, United Kingdom), 100 μM mPEG (Sigma Aldrich, Saint Louis, Missouri, USA), single stranded DNA with a thiol modification (Eurofins Genomics, Ebersberg, Germany), 1 wt% Tween (Sigma Aldrich, Saint Louis, Missouri, USA), 5M NaCl (Sigma Aldrich, Saint Louis, Missouri, USA), Phosphate-buffered saline Tween (PBST).

The protocol started with a 1000 μL of AuNPs, and 10 μL of 1wt% Tween was added. Then the desired concentration of mPEG was added and the substance was mixed, after which it stood for 20 minutes at room temperature. Next, the desired concentration of DNA was added. 400 μL

of 5M NaCl was added and mixed, after which the mixture was incubated for one hour. After the incubation, the washing of the AuNPs was done. The AuNPs are centrifuged and a pellet of AuNPs will form in the bottom of the Eppendorf. 800 μL of fluid is removed and 800 μL of PBST is added. This is repeated five times, after which the AuNPs are ready.

3.2.2 Creating SERS substrates

Different approaches were conducted to create a SERS substrate. The approaches are explained in this section.

Gold nanoparticles bound on glass

The first approach for creating a SERS substrate was by binding AuNPs to a glass surface. To be able to bind the AuNPs to the glass, the following materials are needed: a glass coverslip (Knittel Glass, Braunschweig, Germany), 60 nm AuNPs (BBI solutions, Newport, United Kingdom), ultrasonic cleaner (VWR International, Pennsylvania, USA), a Plasma System (Plasma Surface Treatment System CUTE, Femto Science, Hwaseong-si, South Korea). A schematic representation can be found in figure 3.4.

The container with the AuNPs was placed in the ultrasonic cleaner, to evenly distribute the AuNPs. The glass plates were put into the vacuum chamber of the plasma system to polarize. The system ran protocol #0 for five minutes. The glass plates were removed from the vacuum chamber and were now able to bind AuNPs. 50 μL of AuNPs were added to the middle of each glass plate, and distributed evenly over the surface. The plates with AuNPs were dried overnight, and a single-use glass plate with bound AuNPs was created.

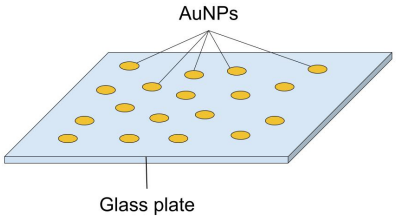


Figure 3.4: Schematic representation of the substrate with AuNPs bounded to glass.

Gold-sputtered plate

Another method of creating a SERS substrate was by creating a gold sputtered plate. This plate was made by a research engineer of the group. This plate consists of a silicon base on which a 4 nm layer of chrome is added because gold does not bind well to glass, but it does to chrome. On top of this chrome layer, a 40 nm thick gold layer is sputtered, trying not to make it as smooth as possible. A schematic representation can be found in figure 3.5.

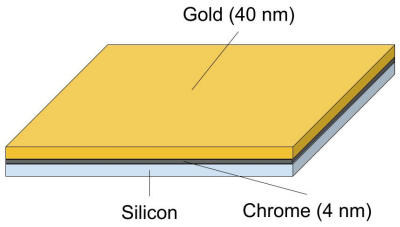


Figure 3.5: Schematic representation of the gold-sputtered plate substrate.

Drain-to-deposit method

The third way to create a SERS substrate was by using the drain-to-deposit method, created by Yang et al. [45, 46]. This method makes it possible to create AuNP monolayer films on a glass plate with a thin gold layer. The creation of the substrate was done with the help of NBP. The following materials are needed: gold thin layer substrate, AuNPs (80nm), hexane, ethanol, and butylamine. A schematic representation can be found in figure 3.6.

A gold thin layer substrate was cut out in a circle and put into a container. After this, a mixture of 0.5 mL ethanol and 0.5 mL butylamine, and a mixture of 1 mL AuNP and 1 mL ethanol was created. The AuNP-ethanol mixture was centrifuged, which ensured the AuNPs would sink to the bottom. After this, 1 mL of the transparent part of the mixture was removed and the remaining AuNP-ethanol part was mixed. 1 mL of this AuNP-ethanol mixture was added to the container with the gold thin layer substrate. After this, 1 mL of the ethanol-butylamine was added, and lastly, 1 mL of hexane was also added. Then the components were mixed with a pipet, limiting the mixing to two cycles.

The mixture was left undisturbed for two hours, whilst covered to eliminate contamination. In that time the AuNPs aggregate on the substrate. After this waiting period, a pump is put on to slowly remove the liquid from the container. After this, the substrate with AuNPs on it is placed inside a nitrogen box for drying, which takes 30 minutes. When the substrate is dried, an AuNP monolayer has successfully been created. Now the substrate was ready and DNA could be dropped on it. For this substrate, only DNA suffices, and no AuNP-DNA conjugate is necessary.

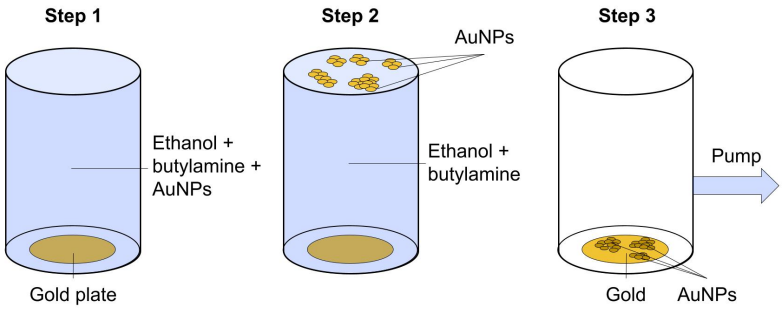


Figure 3.6: Schematic representation of the steps of the drain-to-deposit method substrate.

3.2.3 Measurement of AuNP-DNA conjugates on SERS substrates

Measurements have been conducted on each substrate to test the effectiveness of each method.

Measurements on gold nanoparticle bounded glass

AuNP-DNA conjugates were created following the protocol in 3.2.1. The AuNPs used were 40 nm in diameter. 4 μL of AuNP-DNA conjugate was pipetted on the AuNP bounded glass. For the rest of the experiment, the protocol in chapter 3.1.1 was used. The wavelength of the used laser was 532 nm, and the used laser power was 11,4 μW

Measurements on gold-sputtered plate

AuNP-DNA conjugates were also measured on the gold-sputtered plate. The protocol followed can be found in chapter 3.2.1, using an 80 nm diameter for the AuNPs. The measurement was done following the measurement protocol in chapter 3.1.1, with two additions. The sample volume used was 8 μL, and a halogen lamp was used under a bigger angle as a second light source, which is shown in figure 3.2. This lamp was used to create dark field capabilities on the WITec setup to be able to distinguish the AuNPs from the gold surface. A laser with a wavelength of 532 nm and power of 40,3 μW was used for this experiment.

3.3 Part III: Detection of DNA methylation using SERS

This section describes the experiments conducted with the final setup. These measurements are conducted by using the Drain-to-deposit method as well as the own NBP Raman setup.

The samples were prepared as described in chapter 3.2.2. A total of two samples were made, one for non-methylated DNA sequences and one for the fully methylated DNA, which is described in chapter 3.1.4. These samples were measured using the in-house build Raman spectrometer. The laser wavelength used was 633 nm, the laser power used was 100 μW, and the integration time of 10 seconds. No integration steps were used. With these settings, an area scan was conducted on an AuNP aggregate. This allowed for obtaining a big dataset, consisting of multiple points within the area in one session. Each point in the area gave a SERS spectrum.

Chapter 4

Results & Discussion

4.1 Part I: Raman spectroscopy of nucleobases and DNA

This chapter presents and discusses the measurements of the nucleobases and DNA sequences, as described in chapter 3.1. This part looks at regular Raman spectroscopy, which will be expanded on with SERS in part II, chapter 4.2. First, the Raman spectra of individual nucleobases will be looked at. A comparison will be done, specifically between cytosine and 5-methylcytosine to see whether there are differences in the Raman spectra. With the spectra of all nucleobases, a MATLAB script was written to try and predict the spectrum of a DNA sequence, based on which nucleobases are present. Lastly, Raman spectra of the DNA sequence with different methylation levels were obtained. These spectra were compared to look at whether differentiation between methylation levels is possible. The non-methylated DNA sequence was also compared to the MATLAB script to see whether the model could predict Raman spectra accurately.

4.1.1 Nucleobases

A Raman spectrum of each nucleobase was acquired through the protocol described in Chapter 3.1.1. These spectra show different Raman bands. These bands will be compared to the Raman bands found in the literature. Results of adenine, guanine, and thymine can be found in Appendix II. The tables with the Raman bands, assigned literature values, and experimentally found values can be found in Appendix III. The nucleobase of guanine has not been analyzed, due to the poor solubility in water of the substance.

Cytosine vs. 5-methylcytosine

The Raman spectrum of cytosine can be found in Figure 4.1, and for 5-methylcytosine in Figure 4.2. The numbers indicate the Raman bands, which assignment can be found in Appendix III. The figure shows the counts of the CCD, or Intensity, per wavenumber.

When comparing the cytosine with the 5-methylcytosine, it can be observed that the spectra are significantly different. Peak 2, in both cytosine spectra, shows the ring breathing. In the regular cytosine, the peak is shifted more towards 800 cm^{-1} compared with the 5-methylcytosine, this can also be seen in literature [47]. Peak 3 and 4 of both cytosines share a partly similar assignment. Peak 3 is located at similar wavenumbers in both graphs, this is not the case for peak 4. Peak 5 in regular cytosine has similar assignments to peak 5a and 5b in the methylated

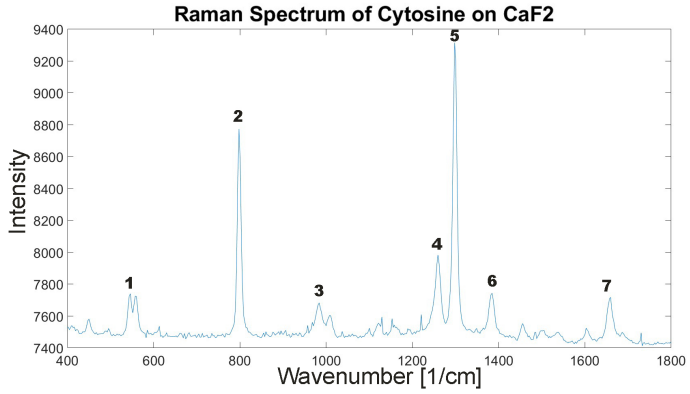


Figure 4.1: Average Raman spectrum of cytosine. The numbers correspond to the assignment found in Appendix 6

cytosine. Peak 5a is somewhat shifted to the left compared to peak 5 in regular cytosine, while peak 5b is at the same location.

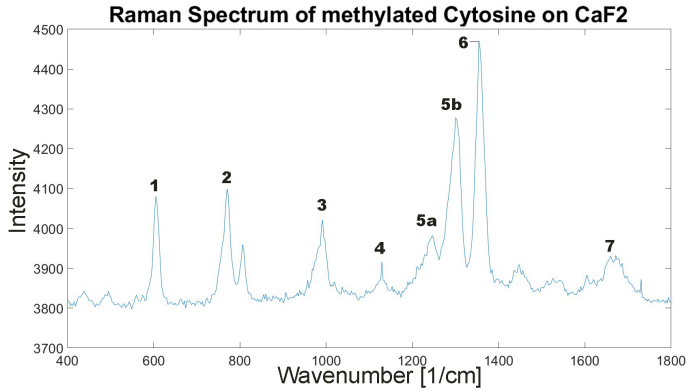


Figure 4.2: Average Raman spectrum of 5-methylcytosine. The numbers correspond to the assignment found in Appendix 6

The difference in intensity of the signal can be due to differences in concentration at the location on the sample. Although the concentration of the nucleobases is the same in solution, the crystallization of the sample could cause a non-homogeneous distribution of the nucleobase. However, peak location is more relevant when looking at differences between molecules. There are some peaks that are at roughly the same location, but have no similar assignment. This is the case for peak 1, 6, and 7 for both cytosines.

Spectra were obtained, which were in correspondence with cytosine spectra previously presented in literature [47, 48]. Peaks occur at roughly the same wavenumbers, and the overall shape of the peaks is also similar. The different peak locations and appearance between nucleobases show that Raman spectroscopy is able to differentiate between the two molecules. The next step is to show whether differences can be seen when looking at non-methylated DNA compared to methylated DNA.

4.1.2 Simulated Raman spectrum

The MATLAB script that was used has been written to get an indication of how the Raman spectrum of a DNA sequence will look, depending on the nucleobase composition. Figure 4.3 shows the spectrum created with the MATLAB script compared to the actually measured Raman spectrum of the non-methylated DNA sequence from chapter 3.1.4.

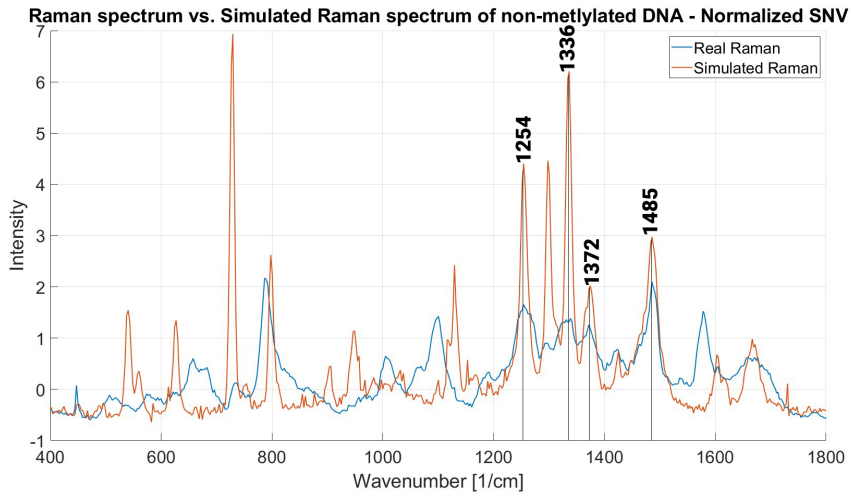


Figure 4.3: Actual Raman spectrum vs. Simulated Raman spectrum of non-metlylated DNA

The first half of the wavenumbers share not too many similarities. But in the second part of the plot it can be seen that, although the bands don't have the same intensity, the peaks are corresponding. Peaks that correspond can be found at 1254, 1336, 1372, and 1485 cm^{-1}. Although some peaks are corresponding, and thus similarities can be seen, it is not completely accurate. This could be due to multiple reasons. The spectrum of guanine was not acquired completely, due to poor solubility of the substance. This could also contribute to the spectrum not being completely accurate. Next to nucleobases, DNA also exists out of phosphate backbones, which can influence the Raman spectrum. However, only the nucleobases are used in the model, which could give a less accurate simulation. The simulated spectrum gives some useful insights into what the actual Raman signal could look but there are improvements that can be made.

4.1.3 Raman spectra of DNA

Figure 4.4 shows Raman spectra of the same DNA sequence with different degrees of methylation using a 532 nm laser. The Raman spectra acquired with the 633 nm laser can be found in Appendix IV.

When looking at figure 4.4, it can be observed that the three spectra look very similar. Significant peaks at 785, 1095, and 1484 cm^{-1} show these similarities clearly. This is contrary to the big differences that are observed comparing the cytosine with the 5-methylcytosine. Upon closer inspection of the spectra of the DNA sequences, it can be seen that there are subtle differences. The peak at 728 cm^{-1} appears to shrink and shift to the left when there are more methylated cytosines in the DNA. This change is named '**a**'. At 1005 cm^{-1} for the non-methylated DNA,

and shifted to 1015 cm^{-1} for the methylated DNA, change **b** can be found. Around 1460 cm^{-1} a side peak appears to emerge as methylation increases, named change '**c**'. A zoom-in of these changes can be seen in figure 4.5.

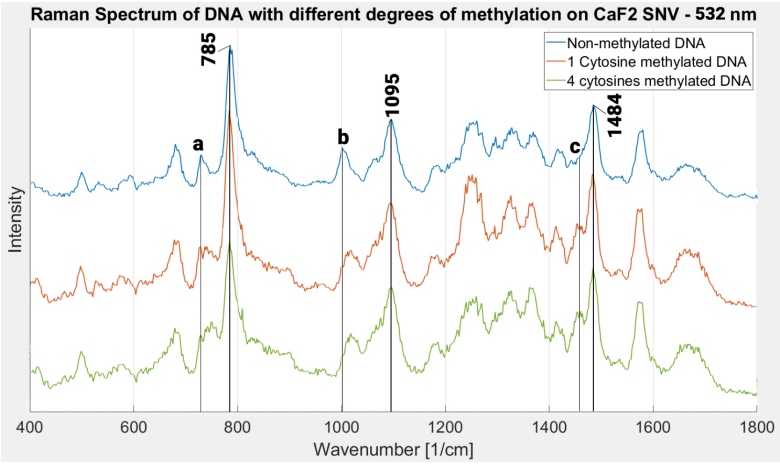


Figure 4.4: Raman spectra of the same DNA sequence, but with different methylation levels. Blue: non-methylated DNA, Orange: 1 cytosine methylated, Green: all cytosines methylated.

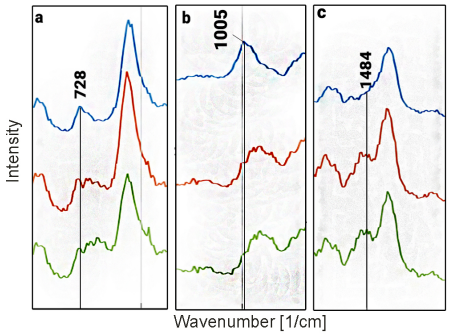


Figure 4.5: Zoom of changes **a**,**b**, and **c** from the plot in figure 4.4 Blue: non-methylated DNA, Orange: 1 cytosine methylated, Green: all cytosines methylated.

The peak at 1005 cm^{-1}, or change **b**, is interesting. This peak, seen at around 1000 cm^{-1} in other research [5, 49], corresponds to the methyl group for the cytosine. A bigger area under the curve means more methylated cytosines. It can be seen that the peak becomes broader and higher for higher methylation levels, which is what was expected. This change could be an interesting factor to look at when determining methylation levels without another reference. It should be checked whether this peak always follows the trend as methylation increases, as can be seen in figure 4.4. If this is the case, the peak around 1005 cm^{-1} is a good indicator for methylation in the DNA. A more profound peak at 1484 cm^{-1}, change **c** in figure 4.5, is also associated with methylation [5].

Principal Component Analysis

Because it can be difficult and subjective to compare Raman spectra with the naked eye, a principal component analysis (PCA) is conducted on the DNA samples with different methylation levels. PCA is chosen as a statistical analysis method because it reduces the dimensionality of large datasets, but preserves as much information as possible. PCA does this by principal components (PC), which are linear combinations of original features in the data. PC1 finds the maximum variance in the entire dataset, which gives the direction of the highest variability. This gives a line that is the closest to the data possible on which all variables are placed. For each consecutive PC, it will find the next maximum variance possible. These PCs can be plotted in a graph, that is able to identify trends and patterns which beforehand were hidden in the large dimensionality of the dataset. Two principal components are used because in each case this explains a significant part of the variance. In this case, a third PC would contribute very little to the variance explanation. But this would introduce a third dimension in which the data is plotted, making analysis harder without adding much relevance. So with two PCs, the results are plotted in two dimensions. Figure 4.6 shows PCA plots of Raman spectra achieved by the 532 nm laser respectively. The full elaboration of the steps taken for PCA can be found in Appendix V.

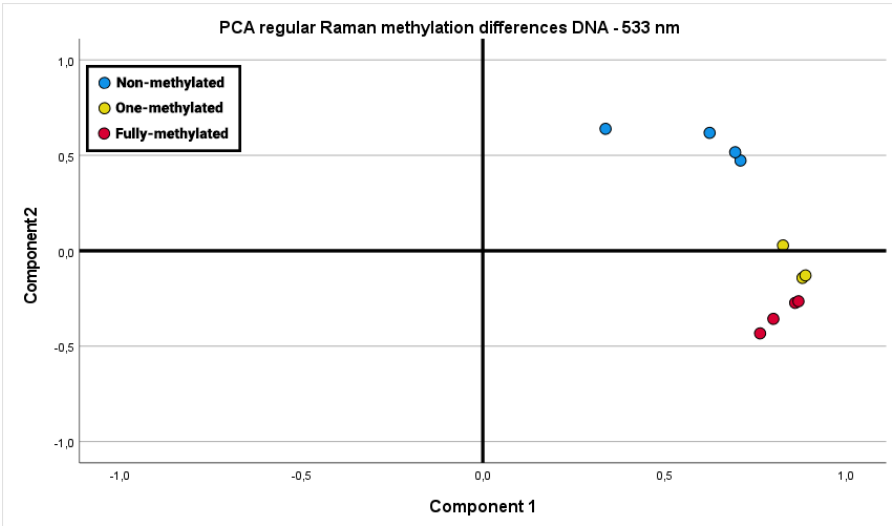


Figure 4.6: Principal component analysis of Raman spectra of DNA with different methylation levels. Non-methylated (blue), 1-methylated (yellow), fully-methylated (red). The spectrum has been normalized by the SNV method.

The PCA plot of the 532 laser clearly shows a difference between the methylated and unmethylated DNA sequences. This can be seen by the grouping of the points in the plot. The fully-methylated and single-methylated sequences seem to be close together. However, the measurements of the fully methylated sequence lie lower on the plot than the single-methylated sequence.

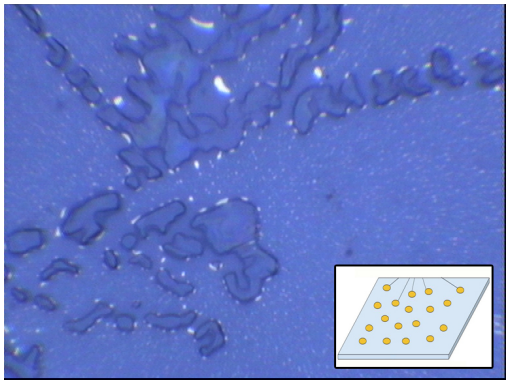
4.2 Part II: SERS and AuNP testing

The previous chapter already showed promising results in methylation differentiation in regular Raman. However, regular Raman uses relatively high laser power, and will only be useful for higher concentrations of DNA. SERS is a more sensitive technique and will be useful when scaling to lower concentrations. This section will present experiments conducted with SERS and look for the ideal substrate to use.

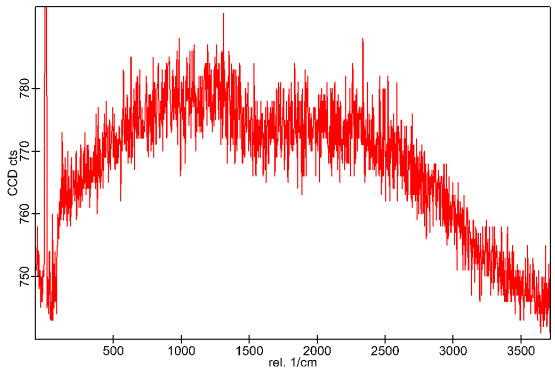
The results in this section were obtained by measuring AuNP-DNA conjugates as well as regular DNA on the selected substrates. Multiple substrates and production methods were tested to find the best substrate for the SERS experiment conducted in Part III, chapter 4.3. The first substrate is produced by binding AuNPs to glass. The second substrate is a gold sputtered plate. On these first two substrates, AuNP-DNA-PEG conjugates were used. The last substrate was created with the drain-to-deposit method. No AuNP-DNA-PEG conjugates are necessary with this method, and thus only DNA is used. At the end of this part, a conclusion will be drawn about which substrate will be the preferred one for the experiment in part III.

4.2.1 Gold nanoparticles bound on glass

A brightfield image of AuNP-DNA-PEG conjugates pipetted on glass with AuNPs bounded on it can be seen in figure 4.7a and the corresponding Raman spectrum can be seen in figure 4.7b. White dots can be observed in figure 4.7a. It is most likely that the aggregated white dots are the AuNP conjugates and the smaller, more evenly distributed dots are the bound AuNPs, but it is hard to be certain of this because no Raman signal was observed.



(a) Brightfield image of AuNP-DNA-PEG conjugates pipetted on glass with bounded AuNPs. Objective: 100x 0.9NA. A schematic of the substrate can be seen in the bottom right.



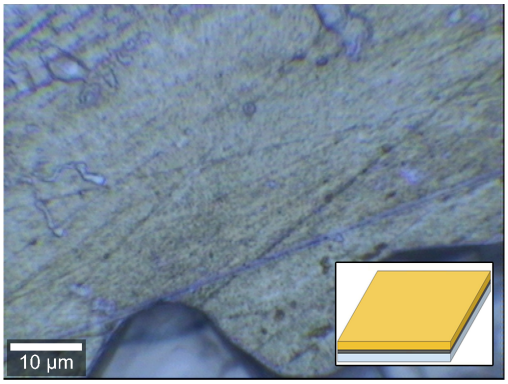
(b) Raman spectrum of AuNP-DNA-PEG conjugates pipetted on glass with bounded AuNPs.

Figure 4.7: AuNP-DNA-PEG conjugate on a AuNP-Glass substrate

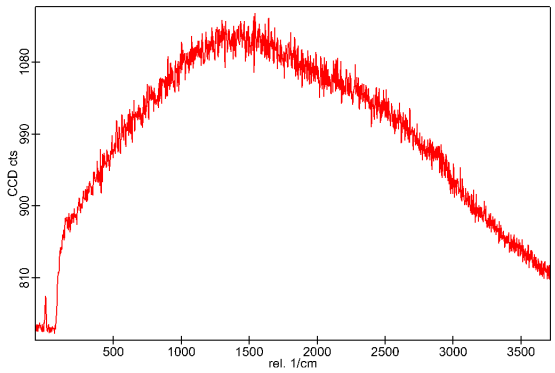
The signal is very noisy and no Raman bands can be seen in figure 4.7b. For all measurements using this method, with different laser power and integration time, similar Raman spectra were obtained. It is most likely that the density of AuNPs that are bound to glass is too low. That could explain the reason no SERS signal is observed because only at a distance of 1 to 5 nm between AuNPs SERS is able to occur.

4.2.2 Gold sputtered glass

Because the density of the AuNPs bound to glass was likely too low for SERS to occur, a gold sputtered plate was created, which makes sure the whole substrate area consists of gold. The AuNP-DNA-PEG conjugates were pipetted directly on the plate, of which a brightfield image can be seen in figure 4.8a. The corresponding Raman spectrum can be seen in figure 4.8b, which is obtained using a 532 nm laser.



(a) Brightfield image of AuNP-DNA-PEG conjugates on a gold sputtered plate. Objective: 100x 0.9NA. A schematic of the substrate can be seen in the bottom right.

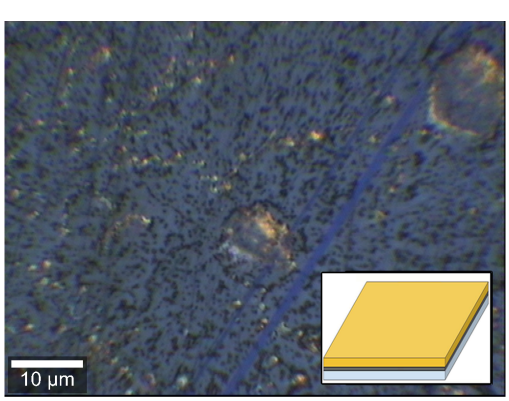


(b) Raman spectrum of AuNP-DNA-PEG conjugates on a gold sputtered plate.

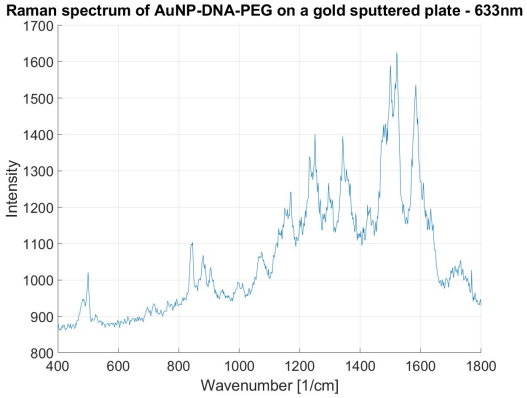
Figure 4.8: AuNP-DNA-PEG conjugate on a gold sputtered substrate

The Raman signal is noisy, and no Raman bands can be observed. It can be seen that the signal is bulging, which could be a sign of the laser power being too high. However, with a laser power of 40,3 μW, this seems unlikely. It could also be due to the high reflectivity of the gold sputtered plate. Because of this high reflectivity, it is also impossible to see the AuNP conjugates without a darkfield setup, which was not available at the time of this experiment. Another explanation for the lack of signal could be the size of the AuNP-DNA-PEG conjugates. In this experiment, 40 nm AuNPs were used as the basis of the conjugate. Bigger AuNPs have a higher chance of giving signal and can lay further apart for SERS to occur, as is shown in figure 2.5. Lastly, the laser wavelength could be of influence. A higher wavelength of the laser could improve SERS signal.

The experiment was redone with some optimization. Darkfield capabilities were added, AuNP-DNA-PEG conjugates were made using 80nm AuNPs, and a laser with 633 nm was used. The results can be seen in figure 4.9, where figure 4.9a shows an image obtained on the sample using both brightfield and darkfield simultaneous. Figure 4.9b shows the Raman spectrum for this sample location. The spectrum was hard to obtain and would fluctuate often. When SERS would occur, this was often located at the gold nanopatricles, which are clearly visible due to the darkfield microscopy. This was not always the case, and sometimes SERS signal would also appear when not on an AuNP. However, picking a precise spot where SERS would occur was hard, and tweaking of the sample position, which is done by hand, was required to get a clear signal. When a stable signal was obtained the measurement was started.



(a) Brightfield + Darkfield image of 80nm AuNP-DNA-PEG conjugates on a gold sputtered plate. Objective: 100x 0.9NA. A schematic of the substrate can be seen in the bottom right.



(b) Raman spectrum of 80nm AuNP-DNA-PEG conjugates on a gold sputtered plate.

Figure 4.9: AuNP-DNA-PEG conjugate on a gold sputtered substrate

The DNA which is used for the AuNP conjugates also was measured on CaF2, to be able to compare the SERS measurements. Figure 4.10 shows both the DNA on CaF2 and one of the few successful SERS measurements from the AuNP conjugates. It can be seen that there are differences between the two, peaks have shifted. However, taking the shift into account, similarities can be found. The peaks of the regular Raman signal around 1331, 1488, and on 1580 can be compared to the peaks of the SERS signal on 1342, 1500, and 1584 $[cm^{-1}]$ respectively. These peaks can be found in figure 4.10 as peaks **a**, **b**, and **c** respectively, with a subscript indicating whether the peak is for regular Raman or SERS. It can be seen that, although somewhat shifted, the SERS spectra show similar peaks with the regular Raman spectrum, giving this method the ability to measure SERS spectra, albeit with great difficulty.

4.2.3 Drain-to-deposit method

The drain to deposit method was tested to see whether the method would work for DNA. Due to the nature of the production of this method, only DNA was used, and not the AuNP-DNA-PEG conjugates because this was not necessary. Also important to note is that these measurements were conducted using the in-house built Raman setup of NBP. Only the non-methylated DNA sequence, described in chapter 3.1.4, was used for the test. The results of this measurement can be seen in figure 4.11. In this figure, the SERS signal is compared to the regular Raman signal obtained on the WITec setup of non-methylated DNA.

When comparing the SERS spectrum to the regular Raman spectrum of non-methylated DNA, at first there are some similarities that are observed. The SERS signal has less significantly large peaks. There are small peaks between 400 and 1200 $[cm^{-1}]$. These peaks do not correspond to the regular Raman signal. Two smaller SERS peaks around 1372 and 1411 $[cm^{-1}]$, peak **a** and **b** respectively, seem to be shifted to the right and correspond to the regular Raman peaks at 1327 and 1368 $[cm^{-1}]$ respectively, labeled **a** and **b** respectively again in figure 4.11. There is a big peak around 1610 $[cm^{-1}]$ for the SERS signal, this corresponds to **a** peak of the regular Raman signal at 1580 $[cm^{-1}]$, both labeled peak **c** with the respective subscript per technique.

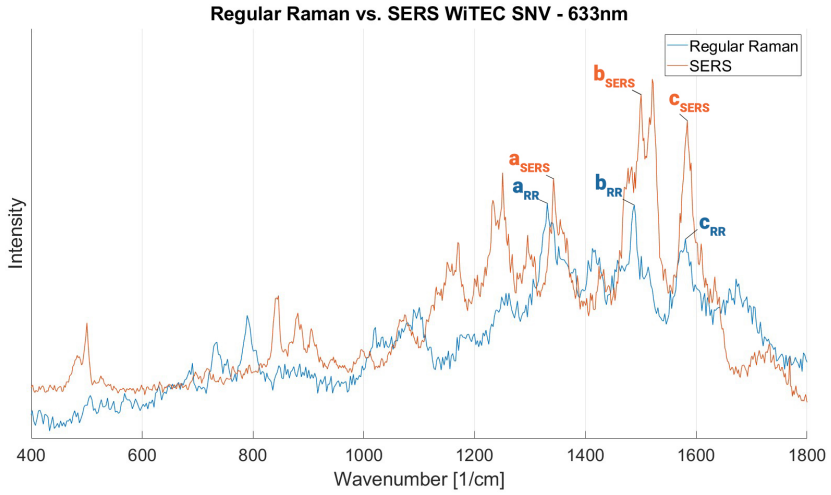


Figure 4.10: Regular Raman spectrum of DNA on a CaF_2 substrate vs. SERS spectrum of AuNP-DNA-PEG conjugates on a gold sputtered substrate. The same DNA sequence is used in both measurements. The spectra are normalized using the SNV method. Subscript: RR = regular Raman, SERS = Surface-Enhanced Raman spectroscopy

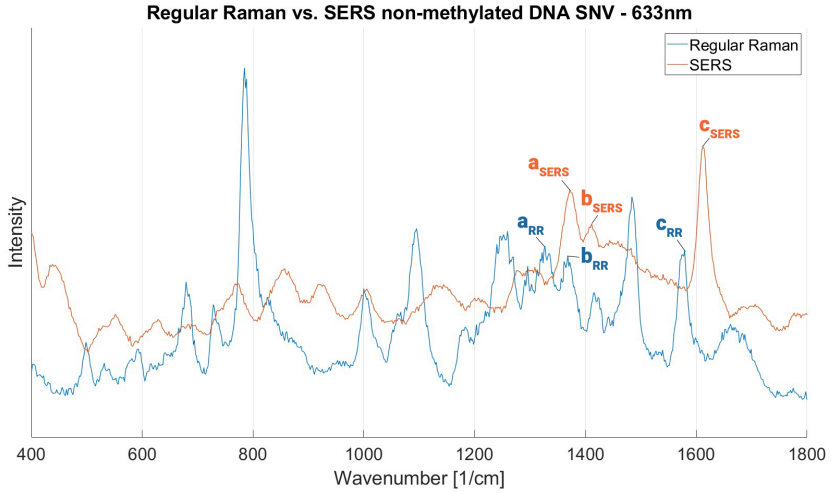


Figure 4.11: Regular Raman spectrum (blue) on WITec vs. SERS spectrum (orange) on NBP setup of non-methylated DNA. The spectra are normalized using the SNV method. Subscript: RR = regular Raman, SERS = Surface-Enhanced Raman spectroscopy

The drain-to-deposit method will be chosen to use in the SERS measurements of part III because it is the most controlled and reliable method. AuNP aggregations form in a relatively small but findable area. DNA in solution can be added to the whole surface and will dry. In that case, the chance of there being DNA on an aggregation is very likely. Also, the setup of NBP is much more preferable for SERS compared to the WITec. The WITec uses completely manual position controls. This is fine for regular Raman, but because the controls are imprecise, it is not great for SERS. Also, darkfield capability is not standard and needs to be improvised, which is also not ideal, making the in-house setup preferable.

4.3 Part III: Detection of DNA methylation using SERS

This part continues on the work of part II. The Drain-to-deposit method has been chosen as the method to us for the final measurements. This method has all the capabilities needed and is also created specifically for SERS. This part shows the results obtained by using the chosen method and compares this to the regular Raman measurements of non- and fully-methylated DNA respectively. This chapter also looks at the possibility of differentiating between non-methylated and fully-methylated SERS measurements of DNA using PCA.

4.3.1 Non-methylated DNA

During the setup of the SERS measurements with non-methylated DNA, an aggregate on which DNA was present, was found quickly. This was confirmed by some test measurements. When a big enough aggregate was found, a full area scan over 15 x 15 μm was conducted, and 5041 spectra were acquired. It must be noted that the SERS average consists of 5041 measurements, while the average regular Raman consists of 5 measurements, and is measured on the WITec setup.

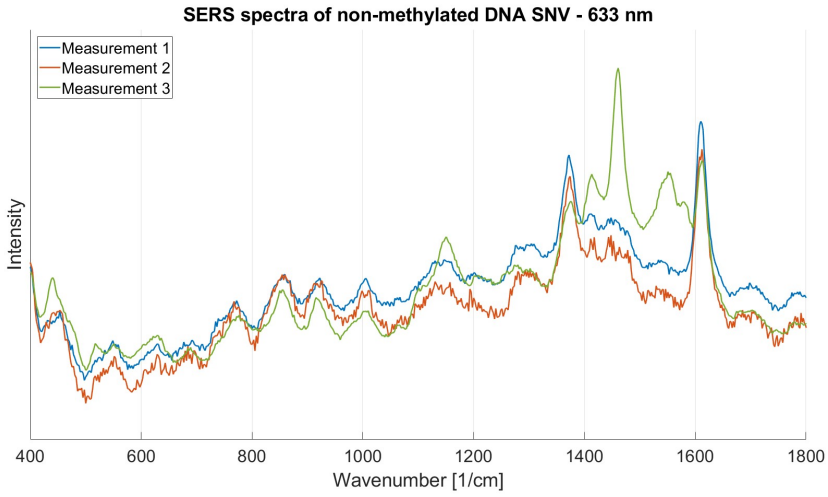


Figure 4.12: SERS spectra of three individual measurements of non-methylated DNA. The spectra are normalized using the SNV method.

When comparing individual measurements, differences in signal can occur. An example of this is shown in figure 4.12. All measurements have different intensities, with the highest difference being a factor of 4,5 between measurements 2 and 3. This difference is not shown in the figure due to the normalization, which allows for better direct comparison. Measurements 1 and 2 are fairly similar in shape. Measurement 3 is also very similar but differs vastly in the region between 1400 and 1600 cm^{-1}. In that same region for measurements 1 and 2, the signal barely shows peaks, but some resemblance with measurement 3 can be seen upon closer inspection. Temporal and spatial fluctuations can be a reason for this. These are fluctuations in the intensity of the Raman scattering generated on the SERS substrate [50]. The DNA could be located at different places in the electromagnetic field that is used to achieve the enhancement of SERS per measurement. Picocavities could also play a role in this. These are sub-nanometer cavities in the substrate surface in which light can be trapped [51]. The trapped light can cause a strong increase in the electromagnetic field due to more plasmonic resonance, creating hotspots. This impacts the uniformity of the signal. Zhu et al. also describes the random distribution and nonuniform enhancement of those hotspots as factors that cause poor spectral reproducability of SERS [42]. Because there can be differences between measurements, averaging the SERS measurements is important to get relevant results.

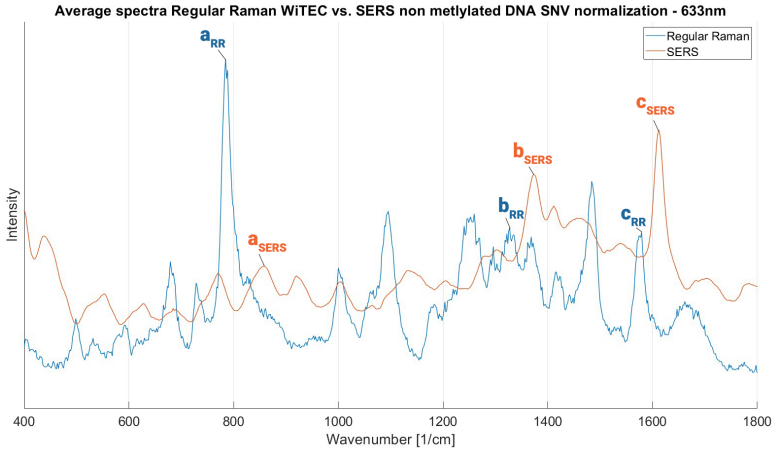


Figure 4.13: Plot of the average regular Raman (blue) vs. SERS (orange) spectra of non-methylated DNA. The spectra are normalized using the SNV method. Subscript: RR = regular Raman, SERS = Surface-Enhanced Raman spectroscopy

The average SERS measurement of the area compared to the average scan of the regular Raman measurement of non-methylated DNA on CaF2 can be seen in figure 4.13. It can be observed that the first part of the SERS spectrum, up to around 1300 cm^{-1}, has some differences compared to the regular Raman spectrum. There are some small peaks in the SERS spectrum that are corresponding, although shifted to the right. An example of this is peak **a**. This peak can be found at 857 cm^{-1} in the SERS spectrum and most likely corresponds with the peak at 785 cm^{-1} in the regular Raman spectrum. Due to the averaging, peaks that are larger in individual measurements can be suppressed somewhat overall. Another reason for a decrease in peaks could be due to plasmonic resonance caused by picocavities. These picocavities can cause fluctuations in the signal, as has been mentioned earlier. These fluctuations impact the signal seen between different measurements and could be a reason that the peaks are lower than expected. Another possible explanation is the unstable nature of picocavities at room temperature [52].

After 1300 cm^{-1}, similar peak locations occur for SERS, albeit somewhat shifted to the right. SERS peaks at 1372 and 1612 cm^{-1}, peaks **b** and **c** respectively, are examples of this. They seem to correspond with peaks at 1327 and 1579 cm^{-1} respectively in the regular Raman spectrum. The correspondence between the two spectra indicates that the SERS measurements of non-methylated DNA were successful.

4.3.2 Fully-methylated DNA

Acquiring the SERS spectrum of fully-methylated DNA was harder to achieve. This was due to the fact that on a very small number of gold aggregates, the signal could be measured. An explanation for this could be the lower concentration of the DNA compared to the concentrations normally used for this method, which is around 10 to 100 times larger. However, this does not explain why the non-methylated DNA spectra were obtained with much more ease. However, aggregates with DNA on them were found, and a signal was obtained at the end. For this sample two area scans were conducted. The first one was conducted over an area of 5 x 5 μm, and the second was conducted over an area of 2 x 2 μm, giving a total of 800 measurements. As with the SERS measurements of non-methylated DNA, individual measurements could differ significantly, so an average spectrum was plotted.

Figure 4.14 shows the average SERS measurement of the area compared to the average scan of the regular Raman measurement of non-methylated DNA on CaF2. The area, and thus the number of points, are lower than for the non-methylated DNA sample, because of the small number of aggregates that were giving signal. Also, the aggregates that gave signal were smaller, thus giving the smaller scan area, meaning less datasets.

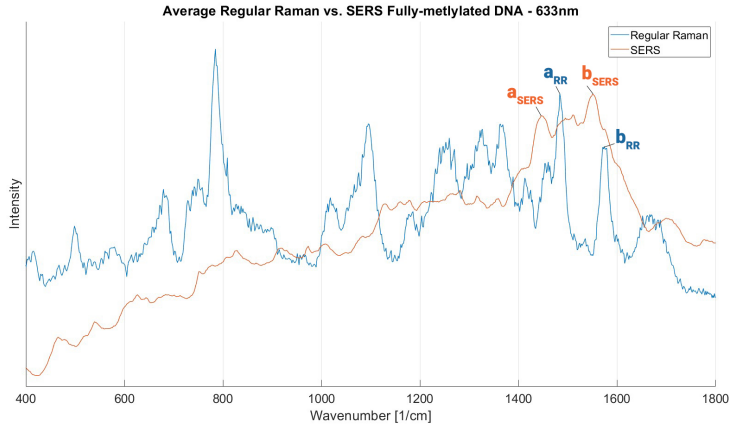


Figure 4.14: Plot of the average regular Raman (blue) vs. SERS (orange) spectra of fully-methylated DNA. Subscript: RR = regular Raman, SERS = Surface-Enhanced Raman spectroscopy

The SERS spectrum of fully-methylated DNA is vastly different when compared to the regular Raman spectrum of the same sample. The first part of the spectrum can be explained by the same reasoning as in chapter 4.3.1, namely due to temporal and spatial fluctuations. Two peaks, peak **a** and peak **b**, can be observed around 1448 and 1550 cm^{-1} respectively for the SERS

signal. It is most likely that these peaks represent the regular Raman peaks at 1485 cm^{-1} and 1575 cm^{-1}, but the SERS signal is shifted to the left. The SERS peaks are not very distinctive, but this is most likely due to the averaging of the signal. When looking at some of the individual measurements, large distinctive peaks are observed which correspond with the regular Raman signal. Appendix VI shows some of these measurements.

It is difficult to directly compare the results obtained with the literature. This is because a small custom-made piece of DNA is used, and no other research has used that exact sequence. For example, Moisoiu et al. uses DNA subtracted from an immortalized cancerous cell line and an immortalized non-cancerous cell line [47]. This means that the spectra will differ in appearance. What can be observed by looking at the results from Moisoiu et al. is that the spectra look different from one another, with some peaks corresponding, which is the same for the SERS spectra in figure 4.15.

4.3.3 DNA methylation comparison

Figure 4.15 shows an average of the SERS measurements of both non-methylated and fully-methylated DNA. From 400 to around 1000 cm^{-1}, both signals show peaks at similar wavenumbers. Peaks at wavenumbers 539, 917, and 1005 cm^{-1} are examples of this. Both spectra do have a different trend between 400 and 1000 cm^{-1}. This difference in trend could be due to the averaging of the signal. Between 1000 and 1600 cm^{-1} the signal is different. An increase in signal can be seen between 1400 and 1600 cm^{-1} for fully-methylated DNA. In this region, the signal for the non-methylated DNA drops somewhat. After 1600 cm^{-1} the signal of fully-methylated DNA drops again, whereas the signal of non-methylated DNA increases and forms a distinct peak at 1612 cm^{-1}. At around 1650 cm^{-1}, the last part of the two spectra seem to show similar peaks again, for example, the overlapping peaks at 1700 cm^{-1}.

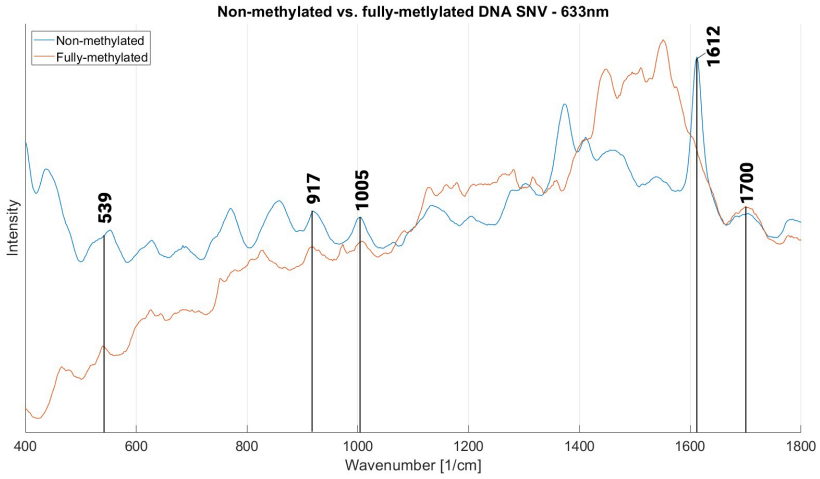


Figure 4.15: Average SERS spectra of non-methylated DNA (blue) and fully methylated DNA (orange).

Looking at the peak at 1005 cm^{-1}, it is expected that the area under the curve is larger for the higher methylation level [49]. This is not the case for the averages seen in figure 4.3.3. It could be due to the averaging of the signal that the peak is suppressed for the fully-methylated DNA.

When looking at the average signal of the fully-methylated DNA measurements, there were almost no distinct peaks. Upon closer inspection of individual measurements, it was observed that a lot of the measurements did not include relevant peaks but instead gave a noisy signal. For this reason, not all data was used. Instead, a selection of 50 measurements was used for the average signal. This gave an average signal where peaks could be distinguished from one another, which is seen in figure 4.15.

The shown plots in figure 4.15 are normalized to be able to conduct a direct comparison between the two. However, when looking at the raw data another observation can be made. This is the difference in intensity for the two measurements. The average intensity of the non-methylated signal is 5,5 times higher compared to the average intensity of the signal of fully-methylated DNA. The concentration of the DNA used is the same for both samples, however, local concentrations differ due to the crystallization of the sample. However, 5,5 times is a very steep decrease in signal intensity. The fully-methylated DNA sample was harder to measure, and a lot of aggregates did not show any signal at all. It is not clear why there is such a difference.

PCA analysis SERS

The two SERS spectra look vastly different, more so than when comparing the regular Raman spectra. This makes it easier to differentiate between the two methylation levels. PCA is still conducted to see whether these differences can be differentiated when reducing the dimensionality. Figure 4.16 shows PCA conducted on the SERS measurements. To keep the graph from cluttering, 20 successful measurements were chosen randomly for both non-methylated and fully-methylated DNA to be used for PCA.

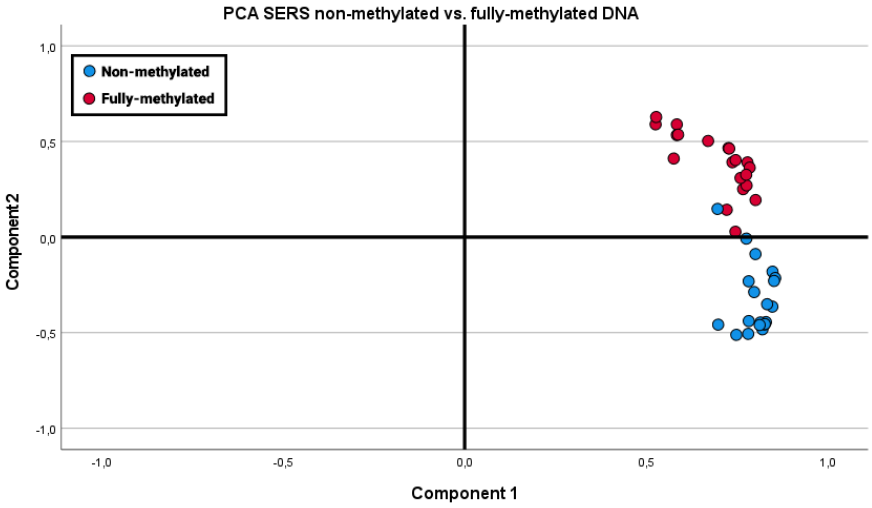


Figure 4.16: PCA conducted on SERS measurements of non-methylated (Blue) vs. fully-methylated (Red) DNA. The data has been normalized using the SNV method.

It can be seen that both methylation levels are mostly distributed in their own groups. There is one non-methylated measurement lying within the fully-methylated group, and one fully-methylated measurement is very close to the non-methylated group. Furthermore, the two groups are very distinct. It needs to be taken into account that even for the same sample, differences between measurements could be significant, which can be seen in figure 4.12. However, the measurements per sample were selected at random, meaning that the differences per measurement are taken into account. This indicates that it is possible to differentiate between non- and fully-methylated DNA using SERS.

4.3.4 Intensity of regular Raman vs. SERS

Looking at the average intensity, CCD counts, per sample for both regular Raman and SERS gives an insight into the efficiency of both methods. To compensate for power and integration time, the average counts per second per Watt will be calculated.

For regular Raman, the average intensity was 773 and 859 counts for non-methylated and fully-methylated DNA respectively. These measurements used a 633 nm laser with a power of 20,05 mW and an integration time of 10 s. This gives 3855 and 4284 $[counts/s/W]$ respectively.

For SERS this average intensity was 27948 and 5125 counts for non-methylated and fully-methylated DNA respectively. The SERS measurements used a 633 nm laser with a power of 80 μW and an integration time of 10 s. This gives 3,494 x 10^7 and 6,406 x 10^6 $[counts/s/W]$ respectively. Comparing that to the regular Raman intensities, there is an increase of counts per second per Watt of 9064 times for non-methylated DNA and 1495 times for fully-methylated DNA. This is already an enormous enhancement, however, increases of up to 10^{12} have been observed [33]. This shows that SERS is much more sensitive than regular Raman. For the large concentrations used in the experiments conducted in this research, that is not necessary, however, when going to much smaller concentrations of DNA, SERS will be preferable because of the enormous enhancement factor.

Chapter 5

Discussion & Conclusion

This chapter will consist of a discussion of the research conducted in this thesis. After that, a conclusion will be drawn based on the acquired results. An outlook will also be given for future research on this topic in the next chapter.

5.1 Discussion

Simulated Raman spectrum

The script used for the simulation of the expected Raman spectrum only used nucleobases. This already gave results that show similarities between the simulated spectrum and the real Raman spectrum. However, to get the signal to be more accurate, it could be beneficial to look into the use of nucleobases with the phosphate backbone. It is recommended to measure the nucleobase that is connected to a deoxiribose 5-phosphate molecule. These backbones could contribute to the signal and give a more correct representation of the real Raman spectrum. A more accurate simulation can be of importance so that it is known how the spectrum looks before doing any actual measurements. In addition to this, it is preferable to use the NBP setup for the measurements. Lastly, because there are differences between regular Raman and SERS spectra, such as shifting and peak proliferation, it is possible to measure the nucleobases with backbones for both techniques.

Substrate and setup

The drain-to-deposit method was chosen as the way to fabricate the substrate with the sample on it, in combination with the in-house built Raman setup of NBP. When looking at the results of this method, figure 4.11, it seems that the spectrum is less accurate compared to the regular Raman signal when looking at the AuNP-DNA-PEG on a gold sputtered plate result in combination with the WITec setup, seen in figure 4.10. Here many more similarities can be observed. However, the result seen for the AuNP-DNA-PEG conjugate was one of the few successful measurements conducted, as has been mentioned in the results. Some other measurements gave vastly different spectra, but most measurements consisted of noise, even when at the beginning of the measurement a spectrum could be observed. The vastly different spectra could be due to picocavities and the created resonance, or due to the random distribution of hotspots, which are known phenomenon [42, 51]. However, the inconsistency and the fact that very precise movement of the microscope had to be done by hand contributed to the fact that the drain-to-deposit

method on the NBP setup was chosen. Another factor contributing to this is the fact that the drain-to-deposit method is much more controlled than the gold sputtered plate method. This does leave the question if the AuNP-DNA-PEG on gold sputtered glass could be viable when using a more advanced setup, such as that of NBP, which is something further research can look at.

Next to see whether the AuNP-DNA-PEG on gold sputtered glass will work on a more advanced setup, further research can be done into other usable substrates for SERS. The research should look at SERS efficiency, reliability of measurements, ease of creating the substrate, and cost. Other materials, but also different substrate shapes can be used. Classic SERS materials such as silver [7] could be a cheaper option compared to gold. But there are also advances in combining the metal with other nanostructures to use for SERS. Gaidi et al. used silicon nanowires on silver nanoparticles to create the SERS substrate [53]. Other research used other methods to increase signal, such as zicronium ions which increase fingerprint signal in DNA [7], or graphene which increased SERS signal by positive interactions with the electromagnetic field [34]. When looking at the shape of a substrate, there are also other possibilities instead of a flat substrate. Luo et al. created a gold substrate using nanohole arrays. These are small, nanometer size gaps in the substrate, to act as hotspots [20]. These nanoholes can be seen as bigger controlled picocavities, that enhance SERS signal. When having a flat substrate, precisely placed nanoparticle arrays are also promising [54].

The drain-to-deposit method did have success after some tweaking. At times it was hard to find any signal, as was especially the case for the fully-methylated DNA sample. On many aggregates formed on the substrate, no signal, and thus most likely, no DNA was present. Even on aggregates where signal was present, this would often not be on the complete aggregate, causing not all data to be useful. A solution to this could be the immobilization of the DNA between AuNPs. This way, the DNA will be at a known location, because the AuNPs which are placed on the substrate can be seen using darkfield microscopy. This exact method has been shown recently by Li et al. with a limit of detection as low as 0.8 fM [55].

The regular Raman measurements on the DNA with different methylation levels were conducted using the WITec setup. The SERS measurements were done using the NBP setup. To counter variables, such as calibration differences, caused by the difference in setup, the regular Raman measurements were also conducted on the NBP setup. However, the measurements conducted on the NBP setup did not give usable results due to multiple reasons. The first reason is that the 633 nm laser was most likely not powerful enough for regular Raman. The laser has a maximum power of around 100 μW, which is around 200 times less power than the laser of the same wavelength used in the WITec. There was also a 640 nm laser that had a higher laser power, however, that laser was lacking a filter, which caused some irregularities in the signal. This would also mean using a slightly different laser wavelength, which also introduces new variables, even if small. Because regular Raman measurements could not be achieved on the NBP setup, it was decided that the comparison between regular Raman and SERS would be done over two setups. The signal acquired on the WITec was shifted somewhat due to the setup difference, but that is also something that can happen when switching from regular Raman to SERS. The shift would not decrease the capability to look for similarities and is thus accepted. The measurements could be repeated when these issues are fixed.

The NBP setup has multiple advantages over the WITec setup. The darkfield capabilities that the NBP setup has is one of the biggest reasons to use this setup. As has been shown in the SERS experiments conducted on the WITec setup, darkfield capabilities can be added by using

an external light source. However, this is not an optimal situation, because the light source needs to be setup for every individual measurement, and due to this, can vary. Another advantage is the ease the NBP setup can scan a whole area, this is more difficult with the WITec. When scanning individual points by hand, the WITec setup needs to be changed by hand, whilst the NBP setup can be positioned much more accurately using the software. All these reasons make the NBP setup much more suitable for SERS measurements. The WITec setup is a good commercial Raman spectrometer which gives decent results for regular Raman. However, the setup built in-house by NBP is more advanced and easier to use. There are some minor problems that need to be fixed, but when this is done it is preferable to use the NBP setup, or another setup available that has the advanced features of the NBP setup, for all measurements. This eliminates variables that occur when using two different setups.

Sample capturing

A problem that could occur during SERS measurements was that on some aggregates no signal was acquired. For fully-methylated DNA this was more significant than for non-methylated DNA. This problem could be tackled by looking at possibilities for the capture of DNA on the substrate. When DNA is captured and thus immobilized, the location of DNA is known, which makes measurement easier. Another advantage is that at the capturing location of the sample, a hotspot can be created, to increase the SERS efficiency. A possible solution to test this is the use of thiol groups, which can be attached to the DNA sequences with different methylation levels [56]. These groups make it so that DNA can bind to AuNPs. Because thiol groups are not attached to DNA naturally, these need to be made or ordered. When wanting to capture DNA out of, for example, a urine sample, then other techniques need to be searched for to immobilize the DNA. A possibility is to use a label-free hybridization sensor setup, where an immobilized single strand of DNA is linked to the substrate. A complementary target DNA strand can bind to this, which immobilizes the DNA [57]. This technique also has the capability to create a hotspot by binding an NP to the other side of the sequence. Figure 5.1 shows this technique, including the binding of another NP in the last step. The target sequence needs to be known, which excludes the possibility of capturing random cancerous DNA. Next to the capture of DNA, there are also other possibilities, such as micro RNA (miRNA). These miRNAs are also an indicator of cancer, and research has been conducted to capture these miRNAs using a sandwich method, after which SERS measurements are conducted [55]. Another option to capture DNA is to use physical immobilization techniques, such as optical tweezers [58]. When the sample is captured, this can ultimately be used to create a small chip on which the sample is placed, which then captures DNA. This chip can then be used for analyzing using SERS.

Methylation differentiation

For SERS, the non-methylated and fully-methylated DNA sequences were used to determine if those could be differentiated. The partly-methylated DNA was not used for these experiments, however, it could give more insight into the changes in the spectrum. For further research, the partly methylated DNA could be included and changes in spectra could be observed with the eye, or using methods such as PCA.

This research showed the possible differentiation between methylated and non-methylated DNA using SERS. However, it was known which sample was which. Further research could try and do the same differentiation, but without the knowledge of which sample is methylated and visa

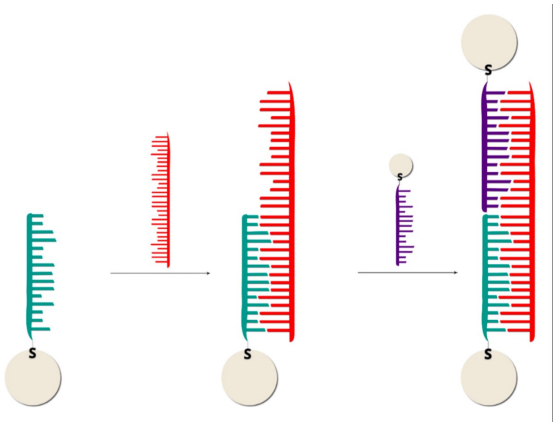


Figure 5.1: DNA capture method by using a complementary capture sequence (blue) which is immobilized on a substrate or particle. The target DNA (red), In the last step, another complementary sequence (purple) which also has an NP attached, binds to the target sequence, creating a potential hotspot. [59]

versa. The peak at 1006 cm^{-1} is one that corresponds with the methyl group added to the cytosine and becomes larger when more methylation is present in the DNA [49]. This is observed in the regular Raman measurements, but not in the SERS measurements. Further research into this needs to be conducted. It was still possible to see differences between the two different methylation levels, which indicates that other changes could also be important. It is important to look at whether the other observed changes seen in the used DNA sequence will also be observed in different DNA sequences, or that other parts of the spectrum will be changing. Due to the differences seen in cytosine and 5m-cytosine, it is hypothesized that it would be possible to observe the changes.

Concentration

In the results it can be seen that both regular Raman and SERS have the capability to differentiate between non-methylated and methylated DNA sequences. Why should a more expensive and difficult technique such as SERS be used when Raman spectroscopy is capable on its own? This can be explained by the much higher efficiency given by SERS, in combination with the much lower concentration of DNA available in a sample when translating the technique to a real-life case. SERS would give an enhancement of up to 10^{12} times compared to regular Raman [33]. When having a very small concentration of DNA that must be found in for example a urine sample, regular Raman would be much too weak for detection. The concentration of potential cell-free DNA in urine is very low [21]. Low concentrations are possible to measure using SERS, where even single-molecule sensitivity has been achieved, albeit in very controlled environments [60, 61]. When looking at concentration, this can also be lowered significantly. Ikegami et al. acquired SERS signal of Adenine, using a concentration of 10^{-11}M [62], and Li et al. going to a concentration of 10^{-12}M using miRNA [55].

All the measurements conducted in this thesis, both regular Raman and SERS, use a relatively high concentration of DNA. The research conducted was to see whether it was possible to see differences in methylation in DNA. As has been mentioned before, when is be translated into real-life detection of cancer, the concentration of DNA will be significantly smaller. Further re-

search needs to be conducted to find out what the lower limit of concentration is where a SERS spectrum still can be obtained. This can be done by lowering the concentration of DNA and increasing the laser power. Firstly, the highest possible laser power that can be used needs to be known. This can be done by trial and error on a sample. After this, the concentration can be lowered till there is barely any signal left to get the lower concentration threshold. When increasing the SERS efficiency with for example a substrate that captures the DNA at a hotspot, the concentration can be lowered again.

DNA

The length of the DNA used is 8,84 nm. Depending on the size of the AuNPs used the distance at which SERS occurs is 1 to 5 nm. This is a smaller distance than the entire DNA molecule takes up. Because DNA has a bend-persistence length of around 50 nm [63], it can be assumed that no additional part of the DNA appears in the same hotspot due to bending. This could influence the SERS signal of this DNA sequence negatively, especially when the DNA is deposited on the substrate. This could cause parts of the DNA to be outside of a hotspot. This is another reason, next to inconsistencies due to picocavities, why measuring an area and averaging the signal is important.

5.2 Conclusion

The goal of this thesis was to look at whether SERS would be a feasible option to detect methylation changes in DNA, and be an option to use for early detection of cancer. To do so, firstly the regular Raman spectra of the nucleobases were acquired. Specifically, the spectra of cytosine and methylated-cytosine were compared. These spectra were significantly different, showing that regular Raman spectroscopy is able to differentiate between the two nucleobases.

After this, DNA with different methylation levels were compared using regular Raman spectroscopy, to see whether differences would also show in the spectra. Comparison of the spectra with the naked eye and PCA showed that it is possible to differentiate between the different methylation levels. Especially when comparing the non-methylated and fully-methylated DNA. The single-methylated DNA sequence measurements would lay closer to the fully-methylated measurements when PCA was conducted and plotted, meaning that single base differences are harder to differentiate between, but still possible.

Multiple methods and setups were tested, and the drain-to-deposit method in combination with the NBP Raman setup was chosen to do the final SERS measurements. This was chosen because this method gave the best results in terms of repeatability and reliability. However, even with the best reliability overall, it could prove hard to find aggregates with DNA on it, making it hard to gather data. Improvements could be made, as has been described in the chapter 5.1.

Lastly, the step towards SERS was made. By combining the most ideal method and setup, the measurements of non-methylated and fully-methylated DNA were conducted. This means that single base differences in DNA have not been looked at. Instead, a difference of four bases was tested. It can be concluded that for the used substrate and setup, DNA methylation can be differentiated by looking at the SERS spectra. PCA also confirmed this, with clear differentiation between the two methylation levels. It must be said that differences between measurements for

the same sample can be vastly different, which can result in outliers or false interpretations of results. Averaging of the measurements is a possibility to counter the effects the vastly different measurements have. Because DNA methylation is a significant epigenetic factor that can be used for identifying healthy or cancerous DNA, and SERS can differentiate between this methylation, SERS is a feasible option to use for cancer detection when the measurements are more controlled and fluctuating effects are lessened. SERS will be especially useful when the concentration levels of the DNA will drop significantly, because of the much higher sensitivity of SERS compared to regular Raman spectroscopy. These levels will drop significantly when translating this to the clinic and for example, use it to study urine samples. Cancer needs to be detected and treated as early as possible. This means that, depending on the stage and type of cancer, concentrations could be as low as a few molecules per mL of urine. For the best results of any test to detect cancer as early as possible, it is preferable to be able to detect single DNA molecules. Single-molecule detection using SERS is already possible in controlled situations, which is hopeful for the viability of this technique translated into a clinical application.

Chapter 6

Outlook

This chapter gives an outlook on the further research that can be conducted based on this thesis. Recommendations will be given for further research using reflection based on this research. The outlook is split up into recommendations on where to continue this research, as well as a part that gives a longer-term future perspective.

Future perspective

Using SERS to differentiate between methylated and non-methylated DNA shows promising results. It is recommended to continue the research and work further to a practical application of this technique. This practical application could be a chip on which a sample will be placed, for example, urine. In the urine, a very low concentration of potentially cancerous DNA will be present. This can be only even a few molecules. This chip will then be able to catch the DNA molecules. The chip will be placed under a SERS setup which can systematically scan each possible DNA binding location, to look for cancerous DNA based on the methylation. Figure 6.1 shows a schematic of a possible microfluidic chip. The chip consists of two inputs, one for the sample containing the target DNA. The other input is for complementary DNA with an AuNP connected to it. An output is also fitted on the chip to dispose of excess fluid. The two inputs lead to a SERS substrate, in this case made of gold, on which another strand of complementary DNA is conjugated. The target DNA binds to the capture DNA connected to the substrate, and the AuNP-DNA conjugate of the second input binds to the remaining part of the target DNA, creating a SERS hotspot. The chip can now be placed in a SERS setup, where the area of the substrate will be scanned. This can be done very precisely, because of the known location of the hotspots.

The chip needs to be analyzed with a Raman setup. These can be very expensive to buy, costing a minimum of $12000 [64]. However, these prices can become much higher. This is a significant investment for a clinic and could be a reason not to implement this technique. A solution could be centralized laboratories to which all the sample chips will be sent. This will make it cheaper for the clinics but will increase test result time significantly. There is research done into the development of cheaper self-made Raman systems, but this does require significant knowledge of Raman [64]. The setup is also specifically built for regular Raman spectroscopy, and not for SERS. This can be seen in for example the laser power used in this research, which is 2 orders of magnitude too high for SERS application, with the laser being an expensive part of the setup. Commercial point-of-care devices using SERS are already in use, for example, to detect COVID-

19 [65]. Portable SERS devices are available commercially, however, detection limits are not low enough, with one study finding a lower detection limit of 10.87 mM [66].

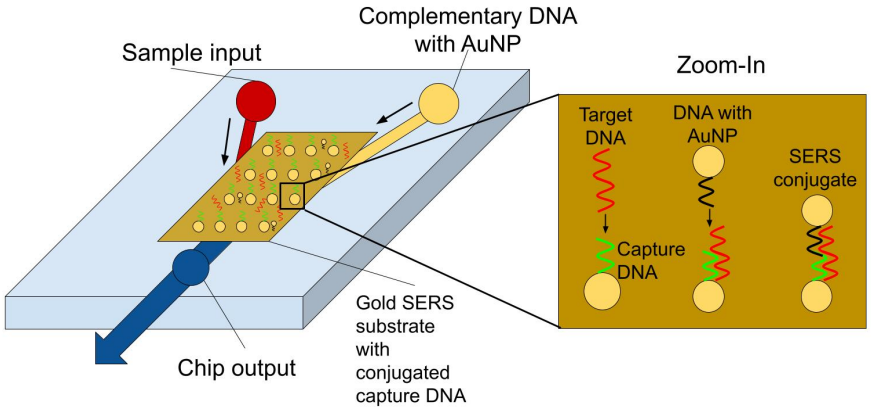


Figure 6.1: Schematics of a possible chip that captures target DNA (red) on a substrate with capture DNA (green) and adds an AuNP with complementary DNA (black), creating SERS conditions. The chip can be placed in a SERS setup and be measured.

In this research, the methylation level of the sample was known beforehand. It would ultimately be necessary to be able to know the level of methylation by spectra analysis itself. As has been shown in the results, changes can be seen in the spectra as methylation levels increase. It is important to look further into the seen changes and find which changes at which wavenumbers indicate more methylated cytosine. It might be necessary to develop pattern recognition using machine learning. PCA could still be an important tool for the reduction of dimensionality, as it shows the spectra as points in a 2 or 3 dimensional plot, which helps in determining the level of methylation. It will also be necessary to do more research into what level of methylation will be cause for concern. This could both be a level of methylation which is too low, or too high, given that hypo- and hypermethylation are both associated with cancer.

Research continuation

There are promising results in the SERS part of this research, for which three focus points are recommended for the continuation of this research. The intensities of the spectra are high, whilst the used laser power is low. Because DNA concentration will be significantly lower when translating the measurements to real-life samples, the first focus should be on testing lower samples and finding a lower limit. It can be observed that the intensities for SERS are significantly higher for lower laser powers. It can be assumed that a signal will still be obtained when decreasing the concentration. Some researchers have obtained SERS spectra using a concentration of Adenine of 10^{-11}M [62]. Translating the single molecule capabilities of SERS in other research to DNA will be an important step.

As has been mentioned in the discussion, other substrates are also a possibility. It is possible the drain-to-deposit method, which uses a gold plate with AuNPs on top, is not the best method in terms of efficiency and reliability. That is why it is recommended to look at different SERS substrates, both in material use and shape, as well as in fabrication method, and compare results to find an ideal substrate. This part can be combined with the last recommendation, the capturing

of the sample.

Another important step that needs further research is for a method to capture DNA, physically or chemically. This is practical for the envisioned chip and also helps with the efficient placement of sample DNA in the hotspots, which will also increase the efficiency of SERS, which will decrease the minimum concentration needed. It is recommended to look into different capture techniques viable, which will work with the chosen substrate.

Acknowledgements

First and foremost, I would like to thank dr. ir. Nienke van Dongen for her daily supervision and guidance. The many discussions we had and feedback or ideas you gave me were very valuable to me. I would also like to thank the rest of my master's assignment committee; Thank you prof. dr. ir. Loes Segerink for the help and the open door. Especially during the short time you took over the supervision from Nienke. The last committee member I would like to thank is dr. Femi Ojambati, who really helped me understand SERS, gave tips and ideas on the topic and made the setup of his group available.

I would like to thank the BIOS Lab-on-a-chip group for their help with the research. Thanks to Ing. Daniël Wijnperle who helped me with the sputtered gold substrate, Janike Bolter BSc who helped me bind AuNPs to glass, and everyone else who helped me with my questions.

Furthermore, I would like to thank the BioImaging Centrum of the University of Twente for making their Raman setup available. Special thanks to Tom Knops BSc for helping me understanding this setup and getting me up to speed with Raman spectroscopy measurements.

Lastly I would like to thank the Nanobiophysics group of the University of Twente for their help with the SERS measurements. The use of their setup helped this research further.

Bibliography

[1] Mattiuzzi C, Lippi G. Current Cancer Epidemiology. Journal of Epidemiology and Global Health. 2019 12;9:217. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7310786/.

[2] Al-Azri MH. Delay in Cancer Diagnosis: Causes and Possible Solutions. Oman Medical Journal. 2016 9;31:325. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4996960/.

[3] Calzone KA. Genetic Biomarkers of Cancer Risk. Seminars in Oncology Nursing. 2012 5;28:122-8.

[4] Qureshi SA, Bashir MU, Yaqinuddin A. Utility of DNA methylation markers for diagnosing cancer. International Journal of Surgery. 2010 1;8:194-8.

[5] Li L, Lim SF, Puretzky A, Riehn R, Hallen HD. DNA Methylation Detection Using Resonance and Nanobowtie-Antenna-Enhanced Raman Spectroscopy. Biophysical Journal. 2018 6;114:2498-506.

[6] Daum R, Brauchle EM, Berrio DAC, Jurkowski TP, Schenke-Layland K. Non-invasive detection of DNA methylation states in carcinoma and pluripotent stem cells using Raman microspectroscopy and imaging. Scientific Reports 2019 9:1. 2019 5;9:1-13. Available from: https://www.nature.com/articles/s41598-019-43520-z.

[7] Zhang Y, Zhan DS, Xu XY, Zhang Z, Hafez ME, He Y, et al. Label-free detection of DNA methylation by surface-enhanced Raman spectroscopy using zirconium-modified silver nanoparticles. Talanta. 2023 2;253:123941.

[8] Hasan SMA, He Y, Chang TW, Wang J, Gartia MR. Detecting DNA Methylation Using Surface-Enhanced Raman Spectroscopy. Journal of Physical Chemistry C. 2019 10;123:698-709. Available from: https://pubs.acs.org/doi/full/10.1021/acs.jpcc.8b10178.

[9] Haynes CL, Mcfarland AD, Duyne RPV. RAMAN SPECTROSCOPY. ANALYTICAL CHEMISTRY. 2005.

[10] Nie S, Emory SR. Probing single molecules and single nanoparticles by surface-enhanced Raman scattering. Science. 1997 2;275:1102-6. Available from: https://www.science.org/doi/10.1126/science.275.5303.1102.

[11] Barhoumi A, Halas NJ. Detecting Chemically Modified DNA Bases Using Surface Enhanced Raman Spectroscopy. The journal of physical chemistry letters. 2011 12;2:3118. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3888867/.

[12] Chen S, Zhang Y, Shih TM, Yang W, Hu S, Hu X, et al. Plasmon-Induced Magnetic Resonance Enhanced Raman Spectroscopy. 2018;18:40. Available from: https://www.mdpi.com/1424-8220/19/7/1712s.

[13] Chen C, Liu W, Tian S, Hong T. Novel Surface-Enhanced Raman Spectroscopy Techniques for DNA, Protein and Drug Detection. Sensors 2019, Vol 19, Page 1712. 2019 4;19:1712. Available from: https://www.mdpi.com/1424-8220/19/7/1712/htmhttps://www.mdpi.com/1424-8220/19/7/1712.

[14] Yu Z, Grasso MF, Cui X, Silva RN, Zhang P. Sensitive and Label-Free SERS Detection of Single-Stranded DNA Assisted by Silver Nanoparticles and Gold-Coated Magnetic Nanoparticles. ACS Applied Bio Materials. 2020 5;3:2626-32. Available from: https://pubs.acs.org/doi/full/10.1021/acsabm.9b01218.

[15] Neidle S. Nucleic acid structure and recognition. 2002:188.

[16] Moore LD, Le T, Fan G. DNA Methylation and Its Basic Function. Neuropsychopharmacology 2013 38:1. 2012 7;38:23-38. Available from: https://www.nature.com/articles/npp2012112.

[17] Jin B, Li Y, Robertson KD. DNA Methylation: Superior or Subordinate in the Epigenetic Hierarchy? Genes Cancer. 2011 6;2:607. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3174260/.

[18] Delpu Y, Cordelier P, Cho WC, Torrisani J. DNA methylation and cancer diagnosis; 2013.

[19] Ehrlich M. DNA methylation in cancer: too much, but also too little. Oncogene 2002 21:35. 2002 8;21:5400-13. Available from: https://www.nature.com/articles/1205651.

[20] Luo X, Xing Y, Galvan DD, Zheng E, Wu P, Cai C, et al. Plasmonic Gold Nanohole Array for Surface-Enhanced Raman Scattering Detection of DNA Methylation. 2019. Available from: https://pubmed.ncbi.nlm.nih.gov/31074265/.

[21] Dongen JEV, Dongen JEV. Novel Biosensors for Early Cancer Detection: Development of Gold Nanoparticle and CRISPR/Cas Assays for the Optical Detection of Cancer Biomarkers in Urine. 2022 11. Available from: https://research.utwente.nl/en/publications/novel-biosensors-for-early-cancer-detection-development-of-gold-n.

[22] Beikircher G, Pulverer W, Hofner M, Noehammer C, Weinhaeusel A. Multiplexed and sensitive DNA methylation testing using methylation-sensitive restriction enzymes "MSRE-qPCR". Methods in Molecular Biology. 2018;1708:407-24. Available from: https://link.springer.com/protocol/10.1007/978-1-4939-7481-8_21.

[23] Kim J, Park HJ, Kim JH, Chang B, Park HK. Label-free detection for a DNA methylation assay using Raman spectroscopy. Chinese Medical Journal. 2017 8;130:1961-7. Available from: www.cmj.org.

[24] Makanai H, Nishihara T, Tanabe K. Surface-Enhanced Raman Scattering Identification of Nucleic Acid Targets by Acetylene-Tagged Hoechst Molecule Binding with DNA-Tethered Gold Nanoparticles. ACS Applied Nano Materials. 2022 2;5:2935-42. Available from: https://pubs.acs.org/doi/abs/10.1021/acsanm.2c00213.

[25] Jones RR, Hooper DC, Zhang L, Wolverson D, Valev VK. Raman Techniques: Fundamentals and Frontiers. Nanoscale Research Letters. 2019 12;14. Available from: /pmc/articles/PMC6626094//pmc/articles/PMC6626094/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC6626094/.

[26] Resonance Raman spectroscopy - Wikipedia;. Available from: https://en.wikipedia.org/wiki/Resonance_Raman_spectroscopy.

[27] Gelder JD, Gussem KD, Vandenabeele P, Moens L. Reference database of Raman spectra of biological molecules. Journal of Raman Spectroscopy. 2007;38:1133-47.

[28] Albrecht AC. On the Theory of Raman Intensities. The Journal of Chemical Physics. 1961 5;34:1476-84. Available from: /aip/jcp/article/34/5/1476/79154/On-the-Theory-of-Raman-Intensities.

[29] Chau AH, Motz JT, Gardecki JA, Waxman S, Bouma BE, Tearney GJ. Fingerprint and high-wavenumber Raman spectroscopy in a human-swine coronary xenograft in vivo. Journal of biomedical optics. 2008;13:040501. Available from: /pmc/articles/PMC2715834//pmc/articles/PMC2715834/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC2715834/.

[30] Otto C, Tweel TJJVD, Mu1 FFMD, Greve J. Surf ace-Enhanced Raman Spectroscopy of DNA Bases. JOURNAL OF RAMAN SPECTROSCOPY. 1986;17:289-98.

[31] Nguyen DB, Nguyen TD, Kim S, Joo SW. Raman spectroscopy and quantum-mechanical analysis of tautomeric forms in cytosine and 5-methylcytosine on gold surfaces. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy. 2017 3;174:183-8. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1386142516306588.

[32] Fleischmann M, Hendra PJ, McQuillan AJ. Raman spectra of pyridine adsorbed at a silver electrode. Chemical Physics Letters. 1974 5;26:163-6.

[33] Pérez-Jiménez AI, Lyu D, Lu Z, Liu G, Ren B. Surface-enhanced Raman spectroscopy: benefits, trade-offs and future developments. Chemical Science. 2020 5;11:4563-77. Available from: https://pubs.rsc.org/en/content/articlehtml/2020/sc/d0sc00809ehttps://pubs.rsc.org/en/content/articlelanding/2020/sc/d0sc00809e.

[34] Ouyang L, Hu Y, Zhu L, Cheng GJ, Irudayaraj J. A reusable laser wrapped graphene-Ag array based SERS sensor for trace detection of genomic DNA methylation. Biosensors and Bioelectronics. 2017 6;92:755-62.

[35] van Dongen JE, Spoelstra LR, Berendsen JTW, Loessberg-Zahl JT, Eijkel JCT, Segerink LI. A Multiplexable Plasmonic Hairpin-DNA Sensor Based On Target-specific Tether Dynamics. ACS Sensors. 2021;6(12):4297-303. PMID: 34851614. Available from: https://doi.org/10.1021/acssensors.1c02097.

[36] Kumari G, Kandula J, Narayana C. How Far Can We Probe by SERS? Journal of Physical Chemistry C. 2015 8;119:20057-64. Available from: https://pubs.acs.org/doi/abs/10.1021/acs.jpcc.5b07556.

[37] Liu Y, Shipton MK, Ryan J, Kaufman ED, Franzen S, Feldheim DL. Synthesis, stability, and cellular internalization of gold nanoparticles containing mixed peptide-poly(ethylene glycol) monolayers. Analytical Chemistry. 2007 3;79:2221-9. Available from: https://pubs.acs.org/doi/full/10.1021/ac061578f.

[38] Adampourezare M, Hasanzadeh M, Seidi F. Optical bio-sensing of DNA methylation analysis: an overview of recent progress and future prospects. RSC Advances. 2022 9;12:25786-806. Available from: https://pubs.rsc.org/en/content/articlehtml/2022/ra/d2ra03630dhttps://pubs.rsc.org/en/content/articlelanding/2022/ra/d2ra03630d.

[39] Chiadò A, Novara C, Lamberti A, Geobaldo F, Giorgis F, Rivolo P. Immobilization of oligonucleotides on metal-dielectric nanostructures for miRNA detection. Analytical chemistry. 2016;88(19):9554-63.

[40] Guven B, Dudak FC, Boyaci IH, Tamer U, Ozsoz M. SERS-based direct and sandwich assay methods for mir-21 detection. Analyst. 2014 2;139:1141-7. Available from: https://pubs.rsc.org/en/content/articlehtml/2014/an/c3an01600ehttps://pubs.rsc.org/en/content/articlelanding/2014/an/c3an01600e.

[41] Pongor CI, Bianco P, Ferenczy GR, Kellermayer RR, Kellermayer MS. Optical Trapping Nanometry of Hypermethylated CPG-Island DNA. Biophysj. 2017;112:512-22. Available from: http://dx.doi.org/10.1016/j.bpj.2016.12.029.

[42] Zhu M, Li M, Su M, Liu J, Liu B, Ge Y, et al. Can "Hot Spots" Be Stable Enough for Surface-Enhanced Raman Scattering? Journal of Physical Chemistry C. 2021 6;125:13443-8. Available from: https://pubs.acs.org/doi/full/10.1021/acs.jpcc.1c03321.

[43] Lieber CA, Mahadevan-Jansen A. Automated method for subtraction of fluorescence from biological Raman spectra. Applied spectroscopy. 2003 11;57:1363-7. Available from: https://pubmed.ncbi.nlm.nih.gov/14658149/.

[44] Alberts B, Johnson A, Lewis J, Morgan D, Raff M, Roberts K, et al. Molecular Biology of the Cell. 500 Tips. Garland Science; 2014. Available from: https://books.google.nl/books?id=_NkpygAACAAJ.

[45] Yang G, Hallinan DT. Self-assembly of large-scale crack-free gold nanoparticle films using a 'drain-to-deposit' strategy. Nanotechnology. 2016 4;27:225604. Available from: https://iopscience.iop.org/article/10.1088/0957-4484/27/22/225604https://iopscience.iop.org/article/10.1088/0957-4484/27/22/225604/meta.

[46] Yang G, Hu L, Keiper TD, Xiong P, Hallinan DT. Gold Nanoparticle Monolayers with Tunable Optical and Electrical Properties. Langmuir. 2016 5;32:4022-33. Available from: https://pubs.acs.org/doi/full/10.1021/acs.langmuir.6b00347.

[47] Moisoiu V, Stefancu A, Iancu SD, Moisoiu T, Loga L, Dican L, et al. SERS assessment of the cancer-specific methylation pattern of genomic DNA: towards the detection of acute myeloid leukemia in patients undergoing hematopoietic stem cell transplantation. Analytical and Bioanalytical Chemistry. 2019 11;411:7907-13. Available from: https://link.springer.com/article/10.1007/s00216-019-02213-2.

[48] Florián J, Baumruk V, Leszczyński J. IR and Raman spectra, tautomeric stabilities, and scaled quantum mechanical force fields of protonated cytosine. Journal of Physical Chemistry. 1996 3;100:5578-89. Available from: https://pubs.acs.org/doi/full/10.1021/jp953284w.

[49] Lu D, Chen Y, Ke L, Wu W, Yuan L, Feng S, et al. Machine learning-assisted global DNA methylation fingerprint analysis for differentiating early-stage lung cancer from benign lung diseases. Biosensors and Bioelectronics. 2023 9;235:115235.

[50] Santos DPD, Temperini MLA, Brolo AG. Intensity Fluctuations in Single-Molecule Surface-Enhanced Raman Scattering. Accounts of Chemical Research. 2019 2;52:456-64. Available from: https://pubs.acs.org/doi/full/10.1021/acs.accounts.8b00563.

[51] Baumberg JJ. Picocavities: a Primer. Nano Letters. 2022 7;22:5859-65. Available from: https://pubs.acs.org/doi/full/10.1021/acs.nanolett.2c01695.

[52] Benz F, Schmidt MK, Dreismann A, Chikkaraddy R, Zhang Y, Demetriadou A, et al. Single-molecule optomechanics in "picocavities". Science (New York, NY). 2016 11;354:726-9. Available from: https://pubmed.ncbi.nlm.nih.gov/27846600/.

[53] Gaidi M, Daoudi K, Tlili A, Columbus S, Leblanc-Lavoie J, Ramachandran K, et al. Fast, highly sensitive and label free detection of small genetic sequence difference of DNA using novel Surface-Enhanced Raman Spectroscopy nanostructured sensor. Sensing and Bio-Sensing Research. 2021 6;32:100406. Available from: https://linkinghub.elsevier.com/retrieve/pii/S2214180421000118.

[54] Jonker D, Srivastava K, Lafuente M, Susarrey-Arce A, Stam WVD, Berg AVD, et al. Low-Variance Surface-Enhanced Raman Spectroscopy Using Confined Gold Nanoparticles over Silicon Nanocones. 2023. Available from: https://doi.org/10.1021/acsanm.3c01249.

[55] Li Y, Jiang L, Yu Z, Jiang C, Zhang F, Jin S. SPRi/SERS dual-mode biosensor based on ployA-DNA/ miRNA/AuNPs-enhanced probe sandwich structure for the detection of multiple miRNA biomarkers. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy. 2024 3;308:123664.

[56] Li Z, Jin R, Mirkin CA, Letsinger RL. Multiple thiol-anchor capped DNA–gold nanoparticle conjugates. Nucleic Acids Research. 2002 4;30:1558. Available from: /pmc/articles/PMC101851//pmc/articles/PMC101851/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC101851/.

[57] Kowalczyk A, Krajczewski J, Kowalik A, Weyher JL, Dziecielewski I, Chłopek M, et al. New strategy for the gene mutation identification using surface enhanced Raman spectroscopy (SERS). Biosensors and Bioelectronics. 2019 5;132:326-32.

[58] Dai X, Fu W, Chi H, Mesias VSD, Zhu H, Leung CW, et al. Optical tweezers-controlled hotspot for sensitive and reproducible surface-enhanced Raman spectroscopy characterization of native protein structures. Nature Communications 2021 12:1. 2021 2;12:1-9. Available from: https://www.nature.com/articles/s41467-021-21543-3.

[59] Pyrak E, Krajczewski J, Kowalik A, Kudelski A, Jaworska A. Surface Enhanced Raman Spectroscopy for DNA Biosensors—How Far Are We? Molecules 2019, Vol 24, Page 4423. 2019 12;24:4423. Available from: https://www.mdpi.com/1420-3049/24/24/4423/htmhttps://www.mdpi.com/1420-3049/24/24/4423.

[60] Chen C, Li Y, Kerman S, Neutens P, Willems K, Cornelissen S, et al. High spatial resolution nanoslit SERS for single-molecule nucleobase sensing. Nature Communications 2018 9:1. 2018 4;9:1-9. Available from: https://www.nature.com/articles/s41467-018-04118-7.

[61] Heck C, Michaeli Y, Bald I, Ebenstein Y. Analytical epigenetics: single-molecule optical detection of DNA and histone modifications. Current Opinion in Biotechnology. 2019 2;55:151-8.

[62] Ikegami K, Sugano K, Isono Y. Surface-enhanced Raman spectroscopy analysis of DNA bases using arrayed and single dimer of gold nanoparticle. Proceedings of the IEEE International Conference on Micro Electro Mechanical Systems (MEMS). 2017 2:408-11.

[63] Manning GS. The Persistence Length of DNA Is Reached from the Persistence Length of Its Null Isomer through an Internal Electrostatic Stretching Force. Biophysical Journal. 2006 11;91:3607. Available from: /pmc/articles/PMC1630458//pmc/articles/PMC1630458/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC1630458/.

[64] Emmanuel N, Nair RB, Abraham B, Yoosaf K. Fabricating a Low-Cost Raman Spectrometer to Introduce Students to Spectroscopy Basics and Applied Instrument Design. Journal of Chemical Education. 2021 6;98:2109-16. Available from: https://pubs.acs.org/doi/full/10.1021/acs.jchemed.0c01028.

[65] Leong SX, Leong YX, Tan EX, Sim HYF, Koh CSL, Lee YH, et al. Noninvasive and Point-of-Care Surface-Enhanced Raman Scattering (SERS)-Based Breathalyzer for Mass Screening of Coronavirus Disease 2019 (COVID-19) under 5 min. ACS Nano. 2022 2;16:2629-39. Available from: https://pubs.acs.org/doi/full/10.1021/acsnano.1c09371.

[66] Thayer E, Turner W, Blama S, Devadas MS, Hondrogiannis EM. Signal detection limit of a portable Raman spectrometer for the SERS detection of gunshot residue. MRS Communications. 2019 9;9:948-55. Available from: https://www.cambridge.org/core/journals/mrs-communications/article/signal-detection-limit-of-a-portable-raman-spectrometer-for-the-sers-detection-of-gunshot-re
BB444992838F623EA141BDBAF6BFF435.

[67] Kaiser HF. A second generation little jiffy. Psychometrika. 1970 12;35:401-15. Available from: https://link.springer.com/article/10.1007/BF02291817.

Appendix I: MATLAB script simulated Raman spectrum

The MATLAB script which is used to simulate the Raman spectrum can be found below:

```
%% Average Raman Spectrum: script that simulates a Raman spectra based on
% the number of nucleotides present
% Ian de Waard - 06-09-2023
clc, close all, clear all

% Open the text file of the Raman spectra of th average nucleotides
As = readtable("Path");
Cs = readtable("Path");
Gs = readtable("Path");
Ts = readtable("Path");
mCs = readtable("Path");

% Define the x-axis
X = table2array(As(:,1));

% Fill in the number of nucleotides as found in the DNA sequences
% The number of nucleotides are multipled with the spectra of each nucleotide
A = 8; As = table2array(As(:,2)).*A;
C = 18;  Cs = table2array(Cs(:,2)).*C;
G = 18;  Gs = table2array(Gs(:,2)).*G;
T = 8;  Ts = table2array(Ts(:,2)).*T;
mC = 0; mCs = table2array(mCs(:,2)).*mC;

% Total spectrum calculated by adding values and dividing it by the total nucleotides
SpecTot = (As + Cs + Gs + Ts + mCs)./(A + C + G + T + mC);

% Plot the results
figure(1)
plot(X,SpecTot);title('Simulated Spectrum non-Methylated DNA')
xlabel('Rel. 1/cm'); ylabel('CCD counts');
xlim([400,1800]);
```

Appendix II: Raman spectra of nucleobases

The tables with the Raman bands, literature value, and experimentally found values can be found in Appendix III.

Adenine

The Raman spectrum of adenine can be found in Figure 6.2.

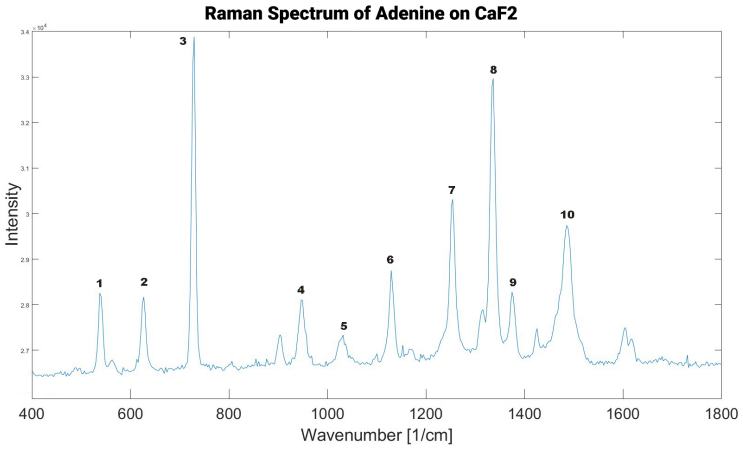


Figure 6.2: Average Raman spectrum of adenine. The numbers correspond to the assignment found in Appendix III

Guanine

The Raman spectrum of guanine can be found in Figure 6.3. As can be seen, there is a lot of noise in the signal. Only for the somewhat higher wavenumbers, peaks can be seen. No analysis was conducted on Guanine, because of the poor quality of the measurements.

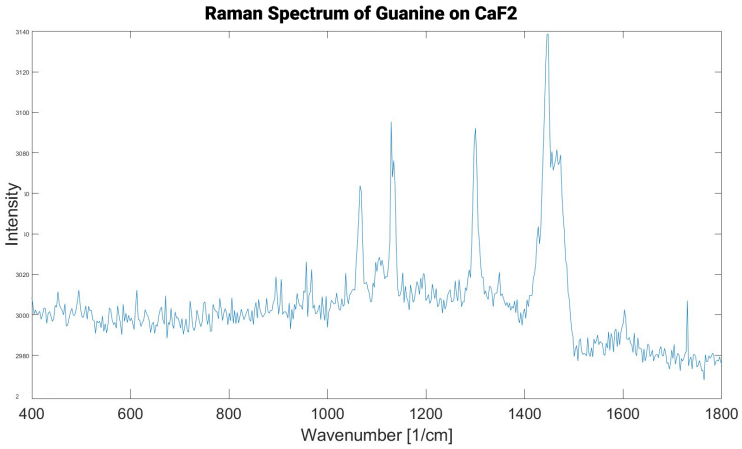


Figure 6.3: Average Raman spectrum of guanine. No analysis has been done, due to the poor results.

Thymine

The Raman spectrum of thymine can be found in Figure 6.4.

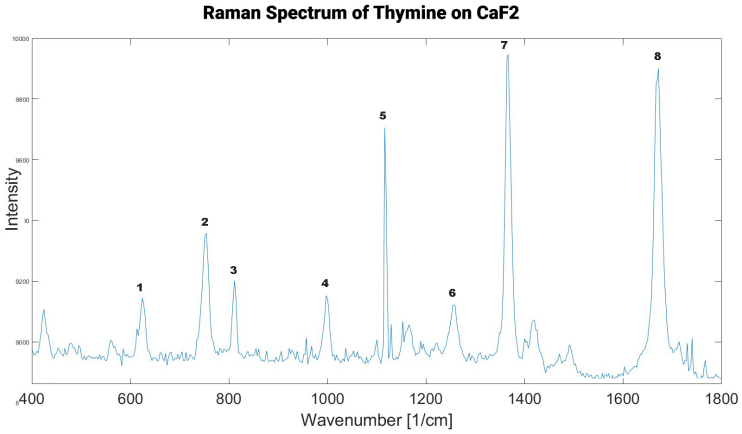


Figure 6.4: Average Raman spectrum of thymine. The numbers correspond to the assignment found in Appendix III

Appendix III: Tables for Raman bands of nucleobases

Adenine

Table 6.1: Raman bands for adenine own measurements compared to literature [30]. b: bending, s: stretching.

Number on graph	Assignment	Calculated wavenumber Literature $[cm^{-1}]$	Wavenumber Raman Spectrum $[cm^{-1}]$
1	$C_5C_4N^b - C_2N_1C_6^b$	540	537
2	$N_9R^b + C_6N_12^b - N_7C_5C_6^b$	623	627
3	Ring Stretching	718	729
4	$NH_2^r + N_1C_6^s$	966	946
5	$NH_2^r + N_9R^s$	1081	1031
6	-$N_3C_2^s + N_9R^s$	1119	1129
7	$N_1C_2^s + C_2H^b + N_9C_8^s + C_8N_7^s$	1259	1254
8	$-N_7C_5^s + C_8N_7^S$	1329	1336
9	$C_8N_9 + C_2N_3^s$	1354	1374
10	$C_2H^b - N_1C_2^s + N_3C_2^s$	1463	1485

Cytosine

Table 6.2: Raman bands for cytosine own measurements compared to literature [31]. b: bending, s: stretching, r: rocking.

Number on graph	Assignment	Calculated wavenumber Literature $[cm^{-1}]$	Wavenumber Raman Spectrum $[cm^{-1}]$
1	$C_4N_6C_1^b + C_3C_2N_7^b + N_6C_1N_7^b$	589	546
2	Ring breathing	778	797
3	$H_{10}N_8C_4^b + N_7C_2^sO_11C_1N_7^b$	979	984
4	$N_6C_4^s + H_9N_8C_4^b$	1219	1259
5	$N_7C_2^s + N_6C_1^s + N_7C_1^s$	1318	1298
6	$H_5C_2C_3^b + N_8C_4^s + H_{13}C_3C_2^b$	1390	1382
7	$C_3C_2^s + N_8C_4^s$	1696	1659

5methylated-Cytosine

Table 6.3: Raman bands for 5methylated-cytosine own measurements compared to literature [31]. b: bending, s: stretching.

Number on graph	Assignment	Calculated wavenumber Literature $[cm^{-1}]$	Wavenumber Raman Spectrum $[cm^{-1}]$
1	$C_4N_6C_1^b + C_11C_3C_4^b + C_3C_2N_7^b + N_6C_1^s$	598	605
2	Ring Breathing	760	771
3	$H_{10}N_8C_4^b + C_2N_7C_1^b + N_7C_1N_6^b + C_4N_6C_1^b$	958	992
4	$N_6C_4^s + H_16N_8C_4^b$	1170	1129
5a	$H_{16}N_7C_2^b + N_7C_1^s + N_7C_2^s$	1246	1244
5b	$N_6C_1^s + N_7C_1^s + H_9N_8C_4^b$	1318	1300
6	$C_3C - C_2^s + H_5C_2N_7^b$	1381	1354
7	$H_{10}N_6C_4^b + N_8C_4^s$	1633	1660

Thymine

Table 6.4: Raman bands for thymine own measurements compared to literature [30]. b: bending, s: stretching, r: rocking. *No calculated wavenumber, but wavenumber used from the researchers findings.

Number on graph	Assignment	Calculated wavenumber Literature $[cm^{-1}]$	Wavenumber Raman Spectrum $[cm^{-1}]$
1	$N_1C_2O^b + N_3C_4O^b$	614	624
2	Ring Breathing	795	754
3	$N_1C_2^s + N_1R^s + C_5C_4^s + N_1C_6^s + N_3C_4^s$	811*	811
4	$C_5 - Me^r$	986*	997
5	No assignment found	-	1116
6	$Ring^s + C_5 - Me^r$	1214*	1256
7	$N_3H^b - C_4 = O^s$	1375	1366
8	$C_2 = O^s$	1691*	1671

Appendix IV: DNA methylation results 633 nm laser

The Raman spectrum obtained with a 633 nm laser, seen in figure 6.5 looks very similar to the spectrum in figure 4.4, which uses the 532 nm laser. In both spectra the Raman bands are located at the similar wavenumbers. The spectrum obtained with the 633 nm laser has a more noisy signal than the spectrum of the 532 nm laser.

Where the 532 nm spectrum shows some subtle differences, it is even harder to see these differences for the 633 nm laser. Firstly, the peaks at 785, 1085 and 1494 cm^{-1} show up in the same location for each methylation level, which is also seen in the comparison of the spectra obtained with the 532 nm laser. The peak at around 728 cm^{-1}, change **a**, seems to become higher and shifts to the right for an increase in methylation. At 1000 cm^{-1}, change **b**, the peak shifts to the right for higher methylation levels. This peak is associated with the methyl group on the cytosine [49], and becomes larger when more cytosine is present. The area under the curve for is the smallest for the non-methylated DNA and largest for the fully-methylated DNA, which would be in-line with the literature. Around 1453 cm^{-1}, change **c**, a side peak appears to emerge, which is also seems to be the case for the 532 nm laser.

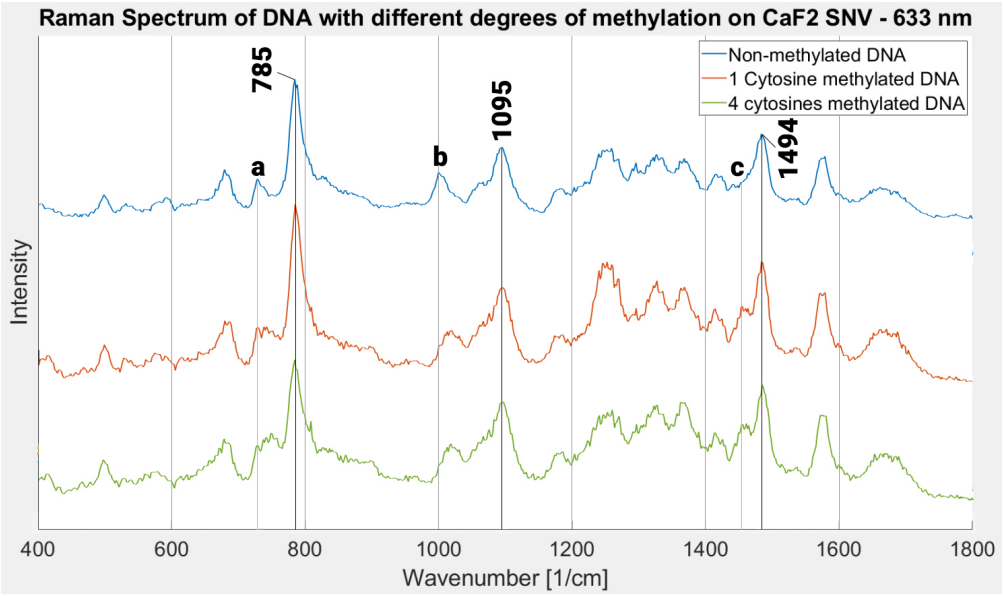


Figure 6.5: Raman spectra of the same DNA sequence, but with different methylation levels, acquired with a 633nm laser. Blue: non-methylated DNA, Orange: 1 cytosine methylated, Green: all cytosines methylated

Appendix V: Principal Component Analysis extended

This appendix goes more into depth what steps PCA uses, how choices about number of principal components (PC) were made, and whether the results are usable. It also show the steps for each PCA that is conducted.

Appendix V-A: Steps of PCA

After normalization, the data is loaded into SPSS. After this, the first step of the analysis is conducted and the data is put into a correlation matrix. The next step is to look at the sampling adequacy, using the Kaiser-Meyer-Olkin (KMO) test and the significance of the variables using Bartlett's test of sphericity. The KMO value shows the suitability of the data for factor analysis. A KMO value of less than 0.6 is considered not adequate, whilst a KMO value bigger than 0.9 is considered marvelous [67]. After this, significance is checked using the Bartlett's test. When this test has less than 5% significance, this indicates that there is a relation among variables. When both tests are passed, further analysis can be done. The next step is to determine the number of PC's that is going to be used in the analysis. Each PC will explain a percentage of the variance, with the first PC explaining the most, the second the most after PC1, etc. A way to determine the number of PCs is by using the scree plot of the eigenvalues in combination with the total percentage of variance explained. Using the scree plot, the value of PCs can be found around the elbow of the plot. This can be combined by looking at whether an extra PC will influence the percentage of explained variance significantly. This percentage can be calculated by dividing the Eigenvalue with the total number of components. It also is important to limit the PCs to a maximum of three if possible, because more gives you too many dimensions to visualize. After this, a number of PCs is chosen, and the last step of the PCA is conducted. This gives a location on a plot per sample measurement using the PCs as coordinates, which shows the final PCA plot.

Appendix V-B: Regular Raman 532 nm

KMO & Bartlett's test

The KMO and Bartlett's test results of the regular Raman measurements conducted with a 532 nm laser can be found in figure 6.6. It can be observed that the KMO value is 0,929, which is considered a marvelous sampling adequacy. The Bartlett's test of sphericity gives a significance of 0.000, which is smaller than 5%. This means that both tests are passed, and the PCA can be continued.

KMO and Bartlett's Test - Regular Raman 532 nm

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,929
Bartlett's Test of Sphericity	Approx. Chi-Square	14550,536
	df	55
	Sig.	,000

Figure 6.6: KMO and Bartlett's test results of PCA regular Raman 532 nm.

Number of Principal Components

The scree plot of the regular Raman measurements conducted with a 532 nm laser can be found in figure 6.7. The elbow is clearly formed by the time component number 3 is reached. This

means that component 1 and 2 are specifically significant, with a contribution of 58,7% and 16,2% respectively.

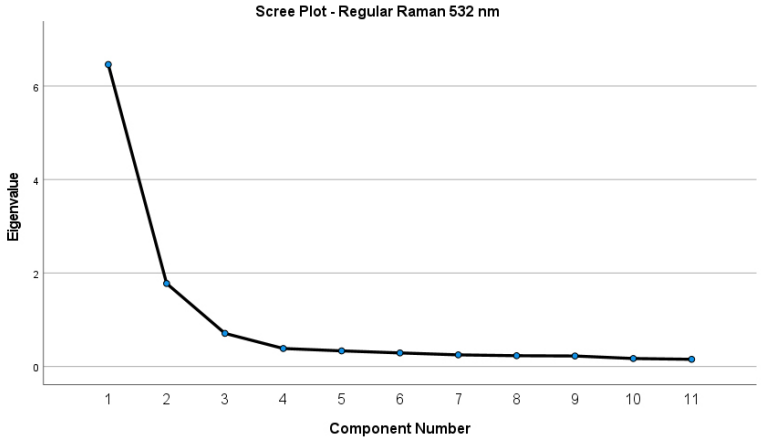


Figure 6.7: Scree plot of eigenvalues regular Raman 532 nm.

PCA plot

The PCA plot and analysis can be found in chapter 4.1.3 subsection **Principal Compenent Analysis**.

Appendix V-C: Regular Raman 633 nm

KMO & Bartlett's test

The KMO and Bartlett's test results of the regular Raman measurements conducted with a 633 nm laser can be found in figure 6.8. It can be observed that the KMO value is 0,968, which is considered a marvelous sampling adequacy. The Bartlett's test of sphericity gives a significance of 0,000, which is smaller than 5%. This means that both tests are passed, and the PCA can be continued.

KMO and Bartlett's Test - Regular Raman 633 nm

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,968
Bartlett's Test of Sphericity	Approx. Chi-Square	40409,901
	df	105
	Sig.	,000

Figure 6.8: KMO and Bartlett's test results of PCA regular Raman 633 nm.

Number of Principal Components

The scree plot of the regular Raman measurements conducted with a 633 nm laser can be found in figure 6.9. The elbow is clearly formed by the time component number 3 is reached. This means that component 1 and 2 are specifically significant, with a contribution of 80,3% and 7,4% respectively.

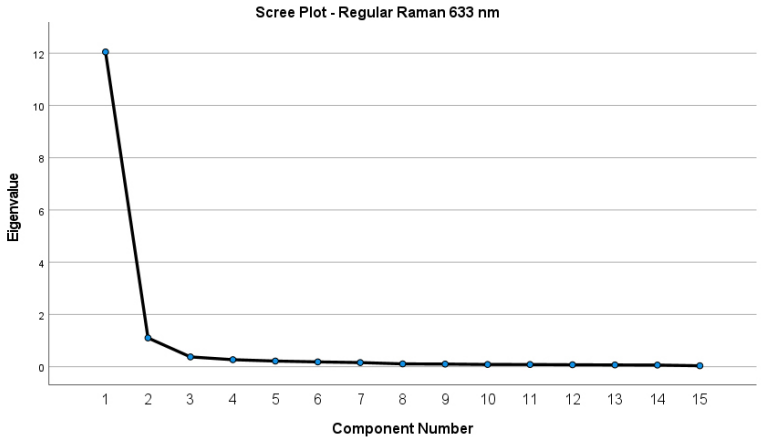


Figure 6.9: Scree plot of eigenvalues regular Raman 633 nm.

PCA plot

The PCA plot of the spectra using a 633 nm laser shows less difference, especially between the one- and fully-methylated DNA. The non-methylated DNA measurements are located above the other two clusters. Upon closer inspection, it can be seen the of the one- and fully-methylated sequence, one measurement is present in the non-methylated group, which seems to be outliers.

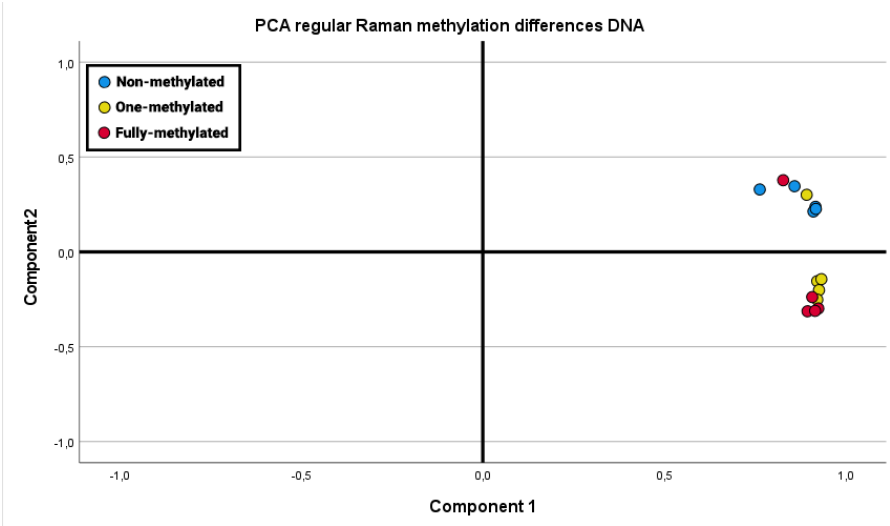


Figure 6.10: Principal component analysis of Raman spectra of DNA with different methylation levels. Non-methylated (blue), 1-methylated (yellow), fully-methylated (red). The spectrum has been normalized by the SNV method. $\lambda = 633$ nm.

The PCA plot of the spectra using a 532 nm laser can be found in figure 4.6.

Appendix V-D: SERS 633 nm

KMO & Bartlett's test

The KMO and Bartlett's test results of the SERS measurements conducted with a 633 nm laser can be found in figure 6.11. It can be observed that the KMO value is 0.980, which is considered a marvelous sampling adequacy. The Bartlett's test of sphericity gives a significance of 0.000, which is smaller than 5%. This means that both tests are passed, and the PCA can be continued.

KMO and Bartlett's Test - SERS 633 nm

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,980
Bartlett's Test of Sphericity	Approx. Chi-Square	82643,367
	df	780
	Sig.	,000

Figure 6.11: KMO and Bartlett's test results of PCA SERS 633 nm.

Number of Principal Components

The scree plot of the SERS measurements conducted with a 633 nm laser can be found in figure 6.12. The elbow is clearly formed by the time component number 3 is reached. This means that component 1 and 2 are specifically significant, with a contribution of 56,8% and 15,6% respectively.

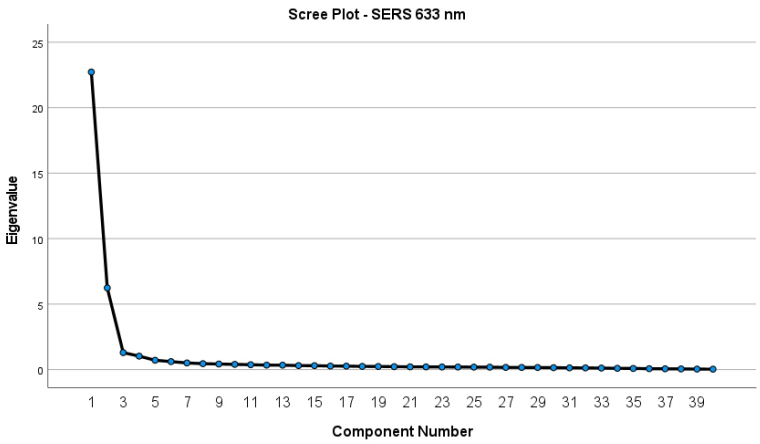


Figure 6.12: Scree plot of eigenvalues SERS 633 nm.

PCA plot

The PCA plot and analysis can be found in chapter 4.3.3 subsection **PCA analysis SERS**.

Appendix VI: Fully-methylated DNA measurements

The plot in figure 6.13 shows three randomly selected individual SERS measurements of the fully methylated DNA sequence. Measurement 3 shows the distinct peaks seen in the average signal the best. Especially between 1400 and 1600 $[cm^{-1}]$. As could be seen between the non-methylated spectra, each measurement differs quite from the other measurements.

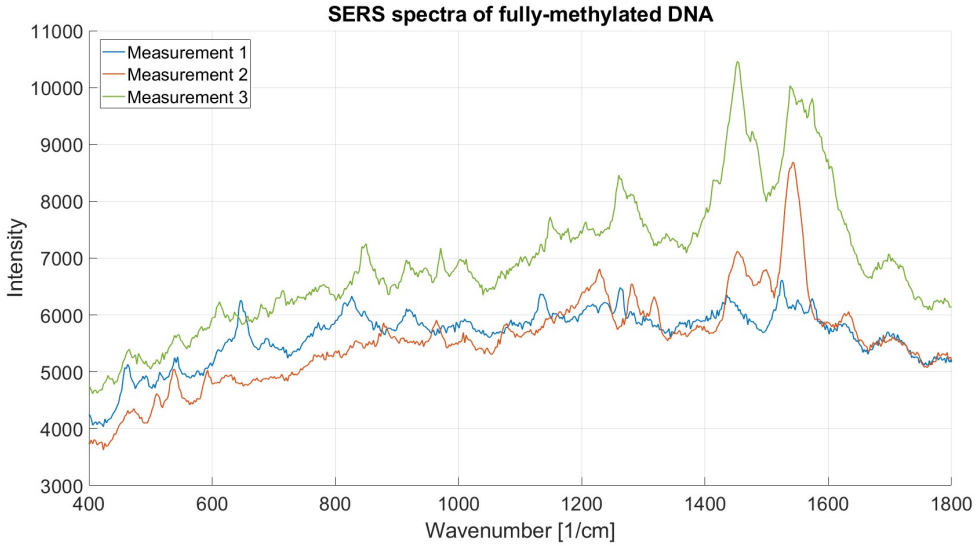


Figure 6.13: Three random SERS spectra from the fully-methylated DNA measurements, showing variation per measurement.